On the Bayesian approach to neutrino mass ordering

Mattias Blennow

Department of Theoretical Physics, School of Engineering Sciences,
KTH Royal Institute of Technology,
AlbaNova University Center, 106 91 Stockholm, Sweden

E-mail: emb@kth.se

ABSTRACT: We study the framework of Bayesian statistics for analyzing the capabilities and results of future experiments looking to solve the issue of the neutrino mass ordering. Starting from the general scenario, we then give examples of the procedure for experiments with Gaussian and non-Gaussian distributions for the indicator. We describe in detail what can and cannot be said about the neutrino mass ordering and a future experiment's capabilities to determine it. Finally, we briefly comment on the application to other binary measurements, such as the determination of the octant of θ_{23}.

Contents

1 Introduction

With the lepton mixing angle θ_{13} being discovered to be large [1–4], there has been a surge of interest in the possibility of determining whether the neutrino mass ordering is normal (NO) or inverted (IO) in the next or next-to-next generation of neutrino experiments. Many studies of the capabilities for doing this in different types of experiments have been performed, including atmospheric [5–28], reactor [27, 29–46], and long baseline [14, 19, 32, 47–66] neutrino experiments. Most of these studies take an approach where the neutrino mass ordering is determined in a frequentist manner with the typical square root of the test statistic used as a measure of the sensitivity as if the distribution of the test statistic was a χ^2 distribution with one degree of freedom. The issue of how to deal with the fact that Wilks' theorem [67] does not apply to a binary measurement has also been dealt with in different ways in several studies [40, 45, 68, 69]. The statistical analysis of the ordering measurement can be performed in either a frequentist or a Bayesian statistics setting. Although the Bayesian method of model selection is ideally suited for the task, the neutrino community is traditionally frequentist and more accustomed to interpreting frequentist results. Thus, in choosing which approach to take, these two facts have to be weighted against each other. It should be mentioned that the frequentist analysis can lead to some results that may be considered unappealing, such as the rejection of both hierarchies at high confidence, there is nothing wrong in performing it as long as proper care is taken in interpreting the results. In this text we will concentrate on how to correctly perform the Bayesian analysis, both as a method of interpreting actual results as well as for judging the capabilities of future experiments. The frequentist approach will be discussed elsewhere [70].

The remainder of this paper has been organized as follows: we start by quickly reviewing the Bayesian approach to model selection in section 2. In section 3 we then specialize to the situation where we have two models that we wish to compare, including definitions that we will use later as well as an analytical treatment of the case where a Gaussian approximation is valid. We continue by giving an example of a non-Gaussian situation in section 4, where we also briefly discuss a semi-analytic approach for the case when the distributions are essentially sums of different Gaussian distributions. In this section we also discuss the prior dependence of Bayesian analyses and its implications for decision making in future neutrino oscillation experiments. Finally, in section 5, we discuss and summarize our results.

2 Bayesian model selection

Unlike frequentist statistics, which are only concerned with how probable outcomes were given a hypothesis, Bayesian statistics deal with our degree of belief in a hypothesis. This has both advantages and disadvantages, the main disadvantage being the introduction of a final result which may depend on the prior knowledge that is inserted into the analysis. However, in some cases it may be of interest to actually consider the prior and changes to it as valuable tools at the design level of an experiment (see, e.g., ref. [71]). Bayesian model selection is performed as follows: consider a situation where we have to select among several different hypotheses H_i, all of which are mutually exclusive. Before any experiment is performed, we assign a *prior* $\pi_i = P(H_i)$ to each hypothesis. These priors are quantifications of our degree of belief in the different hypotheses before having any experimental input and thus are typically chosen such that $\pi_i = \pi_j$ unless some arguments can be applied for having greater belief in one hypothesis than in the others (or if some other experiment has already provided an indication that this is the case). In general

$$\sum_i \pi_i \leq 1, \tag{2.1}$$

where the equality only holds if we are absolutely convinced that one of the hypotheses is true. Once we have performed an experiment, we wish to update our degree of belief in the different hypotheses by computing how likely each hypothesis is, this quantity is the *posterior probability* $P(H_i; t)$, i.e., the probability that H_i is true given an observation t. What we are really interested in are the *posterior odds*

$$\mathrm{Odds}(H_i, H_j; t) = \frac{P(H_i; t)}{P(H_j; t)} \equiv \frac{p_{H_i}}{p_{H_j}}. \tag{2.2}$$

Central to Bayesian statistics, we now make use of Bayes' theorem

$$P(A, B) = P(A; B)P(B) = P(B; A)P(A) \quad \Rightarrow \quad p_{H_i} \propto \mathcal{L}_{H_i}(t)\pi_i \tag{2.3}$$

to rewrite the odds as

$$\mathrm{Odds}(H_i, H_j; t) = \frac{\mathcal{L}_{H_i}(t)}{\mathcal{L}_{H_j}(t)} \frac{\pi_i}{\pi_j}. \tag{2.4}$$

Here, $\mathcal{L}_{H_i}(t)$ is the likelihood of producing the observation t if the hypothesis H_i is true. Thus, the posterior odds are simply updates of the prior odds such that each hypothesis is weighted by how likely the outcome was given the hypothesis.

In many situations, the hypotheses that are being treated are not simple, but rather composed of a general model with one or several model parameters θ. In these cases, we also need to assign a parameter prior $\pi_{H_i}(\theta_i)$ quantifying how likely each realization of parameters are. The likelihood of the data is then in general dependent on the parameters θ_i. However, the model likelihood is easily computable by weighting the likelihoods with the parameter prior

$$\mathcal{L}_{H_i}(t) = \int \pi_{H_i}(\theta_i)\mathcal{L}_{H_i}(t;\theta_i)\,d\theta_i. \tag{2.5}$$

In general, the Bayesian approach would consider the data as is without referencing a test statistic. However, by only considering a single observable T, we are going to simplify the analysis significantly and this is just what we will do below. Since we are not dealing with the frequentist notion of hypothesis testing, we will refrain from calling this T a test statistic and instead refer to it as an *indicator*. This indicator will be well suited for telling two hypotheses, H_i and H_j, apart if its distribution under the different hypotheses differ significantly. Keeping this in mind, we now concentrate on the case at hand, where we want to distinguish two given hypotheses (normal and inverted neutrino mass ordering). When discussing the neutrino mass ordering at future neutrino oscillation experiments below, we will assume that the indicator used is the quantity typically referred to as the $\Delta\chi^2$:

$$T = \Delta\chi^2 \equiv \min_{\text{IO}} \chi^2 - \min_{\text{NO}} \chi^2, \tag{2.6}$$

where χ^2 for given parameters is related to the likelihood $\mathcal{L}$ of the data according to $\chi^2 = -2\log(\mathcal{L})$, thus making T equivalent to the likelihood ratio of the best fit points in NO and IO. We will continue calling this quantity T, since it does typically not follow a χ^2 distribution.

3 Selecting between two hypotheses

In particular, Bayesian model selection is very well suited for selecting between two mutually exclusive hypotheses where we are essentially certain that one of them has to be true. Such is the case of the neutrino mass ordering as was discussed by Qian et al. [68]. In this section, we will largely follow their approach to analyze the Gaussian situation analytically in detail and also discuss the modifications introduced in a non-Gaussian scenario. We will later comment on the different possibilities of evaluating the merits future oscillation experiments based on Bayesian measures of sensitivity.

Let us first consider the general case, where the parameter space may be extended and the distribution of the indicator typically displays non-gaussianities regardless of the gaussianity of the underlying data. Once an experiment has been performed and resulted in a result $T = t$, we need to compute the posterior odds of the hypotheses H and $\bar{H}$. By Bayes' theorem, this will be given by

$$\frac{P(H;t)}{P(\bar{H};t)} = \frac{\mathcal{L}_H(t)}{\mathcal{L}_{\bar{H}}(t)}\frac{\pi_H}{\pi_{\bar{H}}} = \frac{p}{1-p}, \tag{3.1}$$

Strength of evidence for H	κ	Posterior odds	Degree of belief
Barely worth mentioning	0 to 2	ca 1 to 3	$< 73.11\%$
Positive	2 to 6	ca 3 to 20	$> 73.11\%$
Strong	6 to 10	ca 20 to 150	$> 95.26\%$
Very strong	> 10	$\gtrsim 150$	$> 99.33\%$

Table 1. The Kass-Raftery scale [72] for wording the strength of evidence for a hypothesis H.

where $p = p_H$ is the degree of belief in NO after making the observation t and we have assumed $p_H + p_{\bar{H}} = 1$. Thus, in order to compute the posterior odds, compute the general integral in eq. (2.5). In general, this may be a non-trivial task to perform analytically. However, it can be done numerically in a very straightforward fashion through simple Monte Carlo simulation as follows:

1. If H is a composite hypothesis, sample parameter values θ from the parameter prior $\pi_H(\theta)$.

2. Generate a set of data under the assumption that H is true with parameters θ.

3. Compute and store the value of the indicator T given the data generated in 2.

4. Repeat 1 to 3 until a sufficient number of samples of T have been generated.

This procedure will result in a sample of the distribution of T under the assumption that H is true. In some ways, this procedure is very similar to computing the distribution of a test statistic in a frequentist analysis. The big difference is that we are allowed to compute the distribution for H as a whole rather than being restricted to doing it for a particular selection of parameters θ. The reason we can do this is that the prior $\pi_H(\theta)$ has introduced a valid way of properly weighting the contributions from different parameter values.

Once the likelihoods are known as a function of t and the posterior odds computed, we can phrase our degree of belief in H by using either the Kass-Raftery [72] or Jeffrey [73] scales. For comparison, it is useful to define the quantity

$$\kappa(t) = 2\log\left(\frac{P(H;t)}{P(\bar{H};t)}\right) \tag{3.2}$$

For the remainder of this text, we will use the Kass-Raftery scale, see table 1, but also give the degree of belief for reference.

3.1 Evaluating the performance of a future experiment

The Bayesian framework gives us the freedom of computing a series of interesting probabilities already before an experiment has been performed. If we assume that H and $\bar{H}$ are the only two possible hypotheses which are equally probable before performing the experiment, then $P(H) = P(\bar{H}) = 0.5$. As $\kappa(T)$ can be seen as the distribution of posterior odds

once the experiment is performed. We can, for example, compute the probability of the experiment giving at least strong evidence ($> 95.3\%$) for *either* ordering

$$P(|\kappa| > 6) = P(|\kappa| > 6; H)P(H) + P(|\kappa| > 6; \bar{H})P(\bar{H}). \qquad (3.3)$$

This of course also includes the possibility of obtaining strong evidence for the wrong ordering (corresponding to a degree of belief $< 4.7\%$ for the correct ordering), so the more interesting quantity is

$$P(\text{Strong correct}) = P(\kappa > 6; H)P(H) + P(\kappa < -6; \bar{H})P(\bar{H}), \qquad (3.4)$$

where we have exchanged the conditioned probabilities for obtaining strong evidence for the probability of obtaining strong evidence *for the correct hypothesis*. Obviously $P(\text{Strong correct}) \leq P(|\kappa| > 6)$ as it should. Just as the sensitivity can be used to evaluate the performance of future experiments in the frequentist approach, probabilities such as this may be used to judge the merits of future experiments within a Bayesian approach. For example, we may judge experiments on the basis of their probability to give a certain level of evidence or on the evidence that the median experiment will provide. Below we will give examples of such performance indicators.

Note that eq. (3.4) does not assume H or $\bar{H}$ to be true (as was done in ref. [68]), but is rather a weighted average of the results from both hypotheses with the weights given by the prior degree of belief in H and $\bar{H}$, respectively. In the remainder of this paper, we will consider the situation where an equal prior probability of 0.5 has been assigned to both hypotheses. Furthermore, it should be pointed out that it is important to refrain from interpreting κ or posterior degree of belief p in terms of a number of σ, which is an inherently frequentist concept not present in a Bayesian analysis. The number of σ at which a frequentist analysis would reject a hypothesis is dependent on the p-value, the probability of obtaining a result more extreme than the observed one if the hypothesis is true. In contrast, the p computed in this Bayesian framework is the degree of belief that the hypothesis is true after an experiment is performed. These two concepts are fundamentally different and should not be confused.

3.2 Normal distributed indicators

In some limiting cases, the indicator T will follow a normal distribution for both H and $\bar{H}$ with the same standard deviation. This limit is well fulfilled, e.g., for future reactor neutrino experiments sensitive to the neutrino mass ordering as discussed in refs. [68, 69]. For easy applicability to this scenario, we will assume that $T = N(\pm T_0, 2\sqrt{T_0})$, where $N(\mu, \sigma)$ is a normal distribution with mean μ and standard deviation σ and the $+$ $(-)$ is for NO (IO). Clearly, any indicator which supplies normal distributions for T can be brought to this form by translations and rescalings as long as they have the same standard deviation. For the particular case of the neutrino mass ordering at reactors, we also have $T_0 = \overline{\Delta\chi^2}$, i.e., the $\Delta\chi^2 = \min_{\bar{H}} \chi^2 - \min_H \chi^2$ for the Asimov data set for H [74], i.e., the data set where all observables are given by their expectation values.

Since the cumulative distribution and probability density functions of the normal distribution are known, we can obtain simple analytic expressions for several of the quantities and distributions mentioned above, in particular

$$\kappa(t) = t, \tag{3.5}$$

$$P(\kappa > \kappa_0; H) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\kappa_0 - T_0}{2\sqrt{2T_0}}\right)\right], \tag{3.6}$$

$$F(p) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{2\log\frac{p}{1-p} - T_0}{2\sqrt{2T_0}}\right)\right], \tag{3.7}$$

$$f(p) = \frac{1}{\sqrt{2\pi T_0}}\frac{1}{p(1-p)}\exp\left[-\frac{\left(2\log\frac{p}{1-p} - T_0\right)^2}{8T_0}\right]. \tag{3.8}$$

Here, $F(p)$ and $f(p)$ are the cumulative distribution and probability density functions for p (before the experiment is performed, once an observation has been made p is fixed) assuming that H is true, respectively. Thus, by definition

$$P(\kappa > \kappa_0; H) = 1 - F\left(\frac{1}{1 + e^{-\kappa_0/2}}\right) = P(\kappa < -\kappa_0; \bar{H}), \tag{3.9}$$

where the last equality follows from symmetry, which agrees with eq. (3.6). The probability density function $f(p)$ is the analytic form of what is shown in the left panel of figure 3 in ref. [68]. Note that since the probability of obtaining evidence at least at level κ_0 for the true hypothesis in this case is independent of the hypothesis (this is true for all cases where the indicator is anti-symmetric under the exchange of H and $\bar{H}$) and thus

$$P(\text{evidence at least } \kappa_0 \text{ for true hypothesis}) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\kappa_0 - T_0}{2\sqrt{2T_0}}\right)\right]. \tag{3.10}$$

Note that this quantity is the probability of obtaining evidence at least at level κ_0 for the true hypothesis *before* the experiment has been performed. Of course, once the experiment has been performed, we either have found the evidence or not. In particular, the quantity

$$P(\kappa > 0; H) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{\frac{T_0}{8}}\right)\right] \tag{3.11}$$

is the probability that the experimental outcome will favor the correct ordering. A priori, this quantity has little to do with the actual posterior odds assigned to H and $\bar{H}$ as claimed in ref. [69]. If the experiment then measures $t = 0 + \varepsilon$, the posterior odds will be close to one (equal degree of belief in H and $\bar{H}$) regardless of T_0. In other terms, the experiment did not provide us with significant discrimination between H and $\bar{H}$ although we might or might not have expected it to do so. This is neither strange nor unwanted. In fact, this is how we want the posterior odds to behave. If the results are essentially as likely in H as in $\bar{H}$, then a small shift in the experimental outcome should not significantly change our interpretation of the experiment. Of more interest is the probability of actually obtaining strong (95.3%) or very strong (99.3%) evidence for the correct hypothesis, we show these probabilities as a function of T_0 in figure 1.
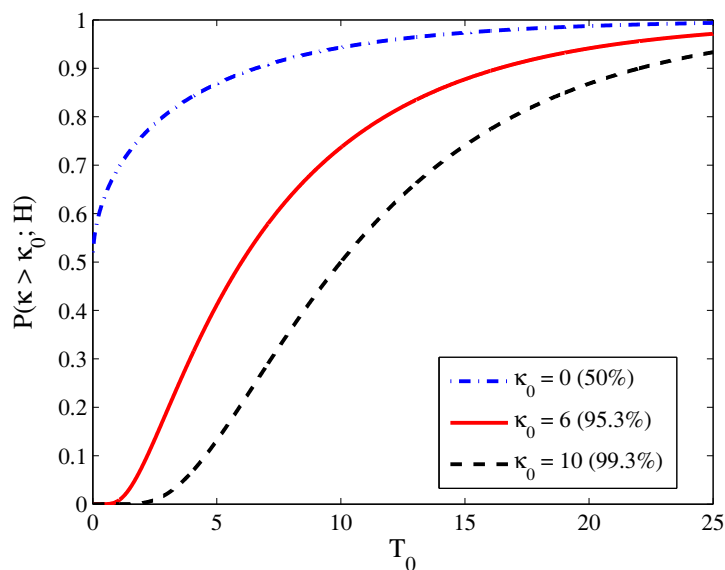
Figure 1. The probability of obtaining strong or very strong evidence (corresponding to a posterior degree of belief of 95.3% and 99.3%, respectively, according to the Kass-Raftery scale) for the correct ordering in an experiment with Gaussian distributions for the indicator T as a function of T_0. For reference, the dash-dotted blue curve shows the probability of actually favoring the correct ordering.

To conclude this section, let us finally note that, since $\mathrm{erf}(0) = 0$, the median experiment will give evidence of strength $\kappa = T_0$. This can be understood from the facts that $\kappa(t) = t$ and T is symmetric around $\pm T_0$.

4 Example of a non-Gaussian situation

As mentioned earlier, the general case is typically not analytically solvable but instead we must apply numerical simulations in a fashion similar to that found in frequentist statistics, where one in general must simulate the distribution of the test statistic. However, contrary to the frequentist framework, the appearance of a probability measure on the parameters means that we will only need to perform one simulation for each hypothesis, rather than once for every parameter set. This detail in itself offers a significant reduction in the computing power needed to perform a test. In particular, the actual testing once the experiment has been performed turns into a simple matter of averaging the likelihood of the acquired data over each hypothesis to find the hypothesis likelihood and, through it, the posterior odds.

In this section, we will give an example of a situation where non-gaussianities play a role in the form of a simulation of the NOνA experiment [75]. In order to perform this analysis we will utilize the GLoBES software [76, 77] along with selected parts of the MonteCUBES plug in [78]. Since we are mainly concerned with the statistical challenges, we simply use the predefined NOνA glb files from the GLoBES homepage [53, 79] to describe

Parameter	Central value	Prior
θ_{23}	$45°$	Constant
θ_{12}	$33°$	Constant
$\sin^2(2\theta_{13})$	0.1	Constant
Δm_{21}^2	$7.9 \cdot 10^{-5}\,\mathrm{eV}^2$	Constant
Δm_{31}^2	$\pm 2.4 \cdot 10^{-3}\,\mathrm{eV}^2$	Gaussian $\pm 10\%$
δ	$-$	Flat, cyclic

Table 2. Priors on the oscillation parameters used in our simulated non-Gaussian scenario studying the NOνA experiment. Note that, although the oscillation probabilities depend on $\cos\delta$ and $\sin\delta$, the prior should rather be chosen according to theoretical prejudice based on an underlying model. We have chosen a flat prior in δ as this corresponds to an invariant measure on the space of possible mixing matrices.

the experiment. For the neutrino oscillation parameters and their priors we impose the conditions displayed in table 2. We only allow Δm_{31}^2 and δ to vary and fix the rest of the parameters. Again, this is done mainly for illustration purposes and in fact the deviation of θ_{23} from its maximal value is an additional challenge for future oscillation experiments, which we will comment on more later. In addition, note that we use these priors only for sampling the parameters in order to find the indicator distributions. We then follow the steps described in the previous section in order to compute these distributions.

In figure 2, we show the cumulative distribution and probability density functions of the indicator $T = \min_{\mathrm{IO}} \chi^2 - \min_{\mathrm{NO}} \chi^2$ (we here use the built-in GLoBES χ^2 function, which is $-2\log\mathcal{L}$ for a given parameter set). As can be seen from the figure, the main difference from a Gaussian distribution is the appearance of large tail distributions in the direction of the simulated ordering. These tails arise mainly from regions of parameter space where the hierarchies are well separated. In this case, values of $\delta = \pm\pi/2$ are either close to the other ordering or the ones furthest removed from it. If the value of δ is such that it will be easy to separate the hierarchies, then we have a significant chance of obtaining evidence for the true ordering. However, this is where the prior on the parameter space of the models come into play and properly weights the probabilities of being in different regions of parameter space. This is in some sense similar to the gain fraction argument for CP-violation presented in ref. [71], with the difference that we are now performing a full Bayesian analysis. From the probability density functions, it is now easy to construct the value of $\kappa(t)$ and we show this quantity in figure 3. As can be seen in this figure, the posterior odds for a given value of the indicator T will typically be larger than the Gaussian t. This effect is due to the tails of the probability density functions. Since κ is related to the ratio of the probability density functions and these fall off much slower than in the Gaussian case on the side of the true ordering, it follows that this will generally be the case. If instead we had a situation where the tails were on the other side, then we would typically obtain weaker posterior odds for the same t.

It should be stressed that the results shown in figure 3 is the only important piece of information once the experiment has been performed as relating the measured value of
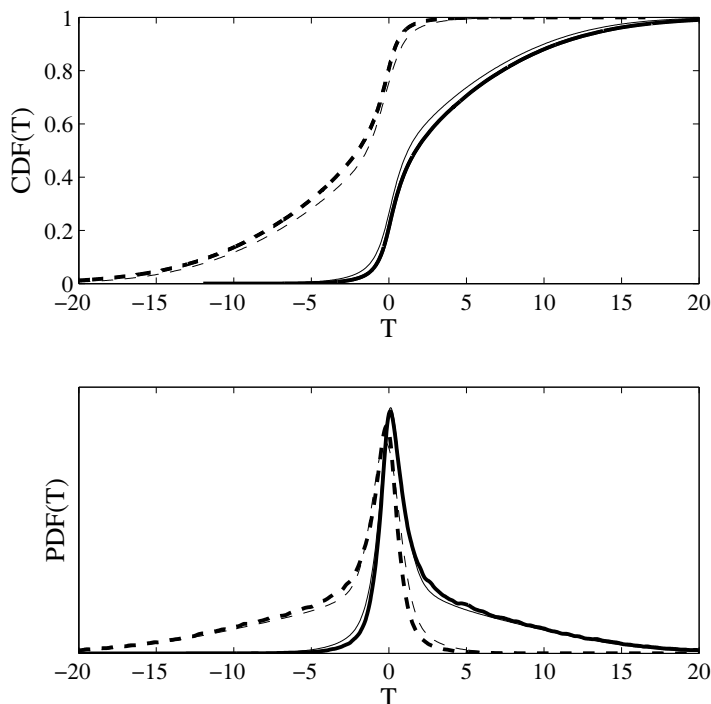
Figure 2. Cumulative distribution (top panel) and probability density (bottom panel) functions for T_{NO} (solid curves) and T_{IO} (dashed curves), respectively, in our simulation of the NOνA experiment as described in the text. The non-gaussianities are clearly visible in asymmetry in the top panel and the large tails of the bottom panel. The thick curves correspond to the result of full simulation while the thin curves correspond to the distributions obtained in the semi-analytic Gaussian approximation (see section 4.1).

the indicator with the posterior odds is straightforward using this figure. As expected, a measured value around $t = 0$ adds very little information on whether NO or IO is the true one.

Finally, we will now discuss how to judge the capabilities of a future experiment within the Bayesian framework and argue that the full distribution of the expected evidence should be considered rather than reducing it to a single number. This will roughly correspond to the typical sensitivity analysis that is usually performed in the frequentist setting. In order to do this, we compute the distributions of κ under the assumptions of NO and IO. This is done by simply inserting the indicator distributions for NO and IO, respectively, and thus obtaining the corresponding distributions of κ. These distributions, whose cumulative distribution functions are shown in figure 4, give us an estimate of what we can expect for the outcome of the experiment. For example, we see that $P(\kappa > 10; \mathrm{NO}) = 1 - \mathrm{CDF}_{\kappa,\mathrm{NO}}(10) \simeq 0.2$. Thus, the probability of NOνA presenting very strong evidence ($> 99.3\%$) for NO if NO is true is around 20%. Since the situation is relatively symmetric, the same is true for IO and thus the probability of NOνA presenting very strong evidence
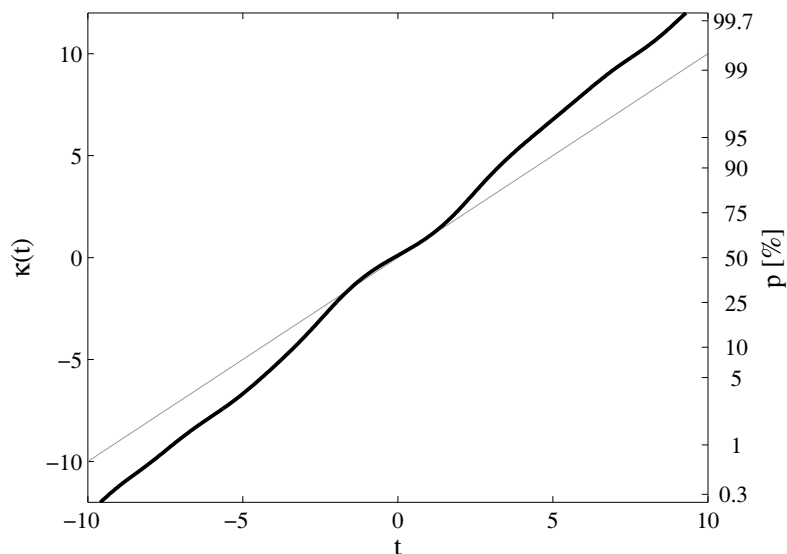
Figure 3. The solid curve shows the value of κ as a function of the measured value of the indicator $T = \min_{\mathrm{IO}} \chi^2 - \min_{\mathrm{NO}} \chi^2$ for the simulated NOνA experiment. The thin gray curve shows the corresponding result for a Gaussian distribution. The scale on the right shows the corresponding posterior degree of belief p in normal neutrino mass ordering.

for the true ordering is also around 20%. Also note that the probability of presenting very strong evidence for the false ordering ($< 0.7\%$ degree of belief in the correct ordering) is miniscule. In order to compare different possibilities for future experiments, one can either construct the equivalent of figure 4 for all experiments and use them to judge the pros and cons of the different experiments. It may be possible that experiment A provides a greater chance of obtaining at least strong evidence, while experiment B provides a greater chance of obtaining at least very strong evidence. The alternative is to pick a desired level of evidence and just compare the experiments' probabilities of reaching at least that level. While the disadvantage of this procedure is a certain loss of information, the advantage is that this single number can be plotted against, e.g., experimental parameters such as the neutrino energy, baseline, or running time. We wish to stress the fact that, since the shape of the distribution of κ in general will vary depending on the experiment, a single number will typically be insufficient to describe the full situation. For example, a Gaussian experiment with $T_0 = 4$ would have a larger median evidence for the true ordering than our simulated NOνA experiment. However, the chances of getting lucky and obtain very strong evidence in such an experiment would be miniscule. Thus, rather than tabulating several predefined numbers describing the distributions, we would advocate simply showing the distribution and basing any comparison on how the distributions compare.

4.1 Semi-analytic approximation

It was observed in ref. [70], that for fixed values of the oscillation parameters (in particular δ), the distribution of T would still take on a Gaussian form with mean and standard
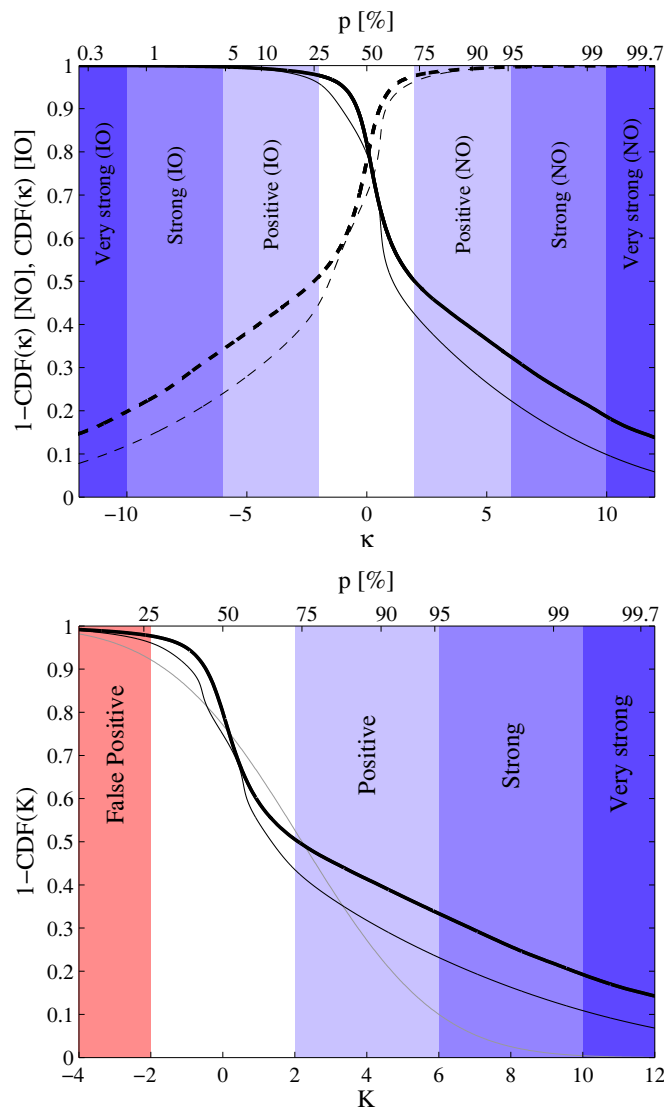
Figure 4. Top panel: probability of obtaining posterior odds corresponding to κ *or higher* in favor of the correct ordering. The solid curve assumes NO, while the dashed curve assumes IO. Note that the curves are tilted in opposite directions as IO becomes more likely as the value of κ decreases. Bottom panel: the probability of obtaining posterior odds corresponding to K *or higher* in favor of the correct ordering. We use $K = \kappa$ for true NO and $K = -\kappa$ for true IO in order to collect the probability for obtaining evidence for the true ordering. As reference, the thin gray curve shows the corresponding quantity for a Gaussian experiment with $T_0 = 2.2$, which gives the same median posterior odds. The scale on the top of the upper (lower) panel show the corresponding posterior degree of belief in the normal (correct) neutrino mass ordering. The thin black curves show the corresponding results of the semi-analytic approximation presented in section 4.1.

deviation now given by the $T_0(\theta)$ produced by that particular parameter set θ. If this holds and $T_0(\theta)$ is known, then the probability density function of T will be given by

$$\mathcal{L}_{\text{NO}}(T) = \int \pi_{\text{NO}}(\theta)\mathcal{L}_{T_0(\theta)}(T)d\theta, \tag{4.1}$$

where

$$\mathcal{L}_{T_0}(t) = \frac{1}{\sqrt{8\pi T_0}} \exp\left(-\frac{(t-T_0)^2}{8T_0}\right) \tag{4.2}$$

is the probability density function of $N(T_0, 2\sqrt{T_0})$ and the corresponding expressions hold for inverted ordering. If $T_0(\theta)$ has the same distribution in both hierarchies, i.e.,

$$\pi_{\mathrm{NO}}(\theta)\frac{d\theta}{dT_0}dT_0 = \pi_{\mathrm{IO}}(\bar{\theta})\frac{d\bar{\theta}}{dT_0}dT_0 \equiv \pi(T_0)dT_0 \tag{4.3}$$

for all T_0, then

$$\mathcal{L}_{\mathrm{NO}}(t) = \exp\left(\frac{t}{4}\right) \int \pi(T_0)\frac{1}{\sqrt{8\pi T_0}} \exp\left(-\frac{t^2 + T_0^2}{8T_0^2}\right) dT_0, \tag{4.4}$$

$$\mathcal{L}_{\mathrm{IO}}(t) = \exp\left(-\frac{t}{4}\right) \int \pi(T_0)\frac{1}{\sqrt{8\pi T_0}} \exp\left(-\frac{t^2 + T_0^2}{8T_0^2}\right) dT_0. \tag{4.5}$$

It follows directly that also in this case,

$$\kappa(t) = 2\log\left(\frac{\mathcal{L}_{\mathrm{NO}}(t)}{\mathcal{L}_{\mathrm{IO}}(t)}\right) = t. \tag{4.6}$$

This is a generalization of the result found in eq. (3.5) and any deviations from this result must be caused by non-Gaussianities or different distributions of $T_0(\theta)$ in the different orderings.

The largest deviation from this appeared for the case of the NOνA experiment, which is the example we have been using so far in this work. For comparison, we show the expectation of this semi-analytic approximation with thin curves in the lower panel of figure 2. As can be seen, the approximation overestimates the value of the probability density function where the ordering is not the assumed one, while it underestimates it in the other tail. For producing this result, we have computed T_0 as a function of δ for both hierarchies independently and the results do not provide the same T_0 distributions although it is close enough for $\kappa(t)$ not to show any significant deviations from t for $|t| < 20$. Thus, the deviations from this rule seen in figure 3 originate mainly from the non-Gaussianity of the NOνA distributions. Thus, using this semi-analytic approach for NOνA will therefore lead to conservative estimates of the NOνA capabilities. We also show the result of the approximation as thin black curves in figure 4. This figure confirms our expectation that the approximation will underestimate the capability of NOνA, although the main features such as the larger tail probability are still present, the probability of reaching a good degree of belief may be underestimated by as much as 10%.

4.2 Prior dependence

Just as with any Bayesian statement, the distribution of the posterior odds prior to an experiment is performed will typically be dependent on the parameter prior within the models. In the situation described above, the main prior impact arises from the flat prior on δ, while the actual prior chosen for Δm_{31}^2 does not change the prediction significantly. This mainly occurs since the probability of generating data that will separate the hierarchies
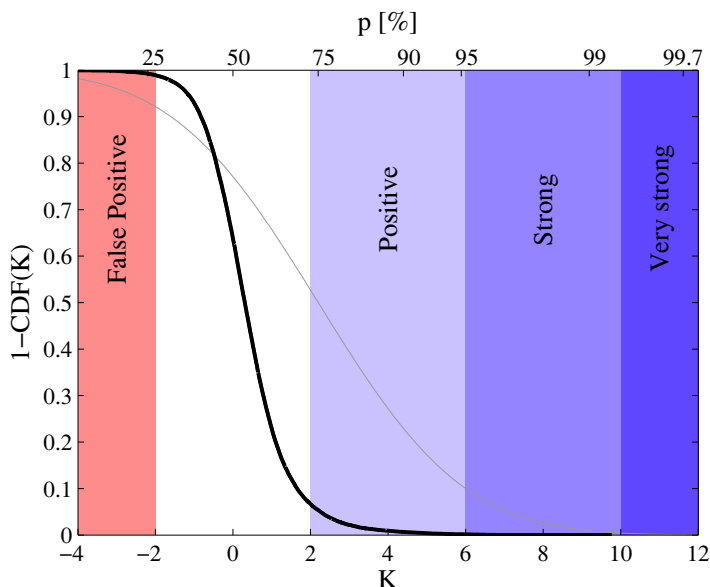
Figure 5. The same as the lower panel of figure 4, but with a prior that fixes δ to the most pessimistic values for the purposes of mass ordering determination.

is highly δ dependent. This only reflects the fact that when considering what future experiment to built, the current knowledge of the model parameters should also be taken into account (see also ref. [71]). For the CP-violating phase, most people would agree that a flat prior is a reasonable starting point as long as no data is available to favor one region of δ or the other. However, once there is a hint for a value of δ, it is only natural that the perceived probability of making a given measurement should change accordingly.

The prior dependence would play an even bigger role in a situation where there is no lower bound on the effect that is being searched for. In the case of the mass ordering, $\Delta m_{31}^2 = 0$ has long since been ruled out. However, for the determination of the octant of θ_{23}, the data of global fits are still very compatible with maximal mixing. Of course, the possibility to discriminate between the octants is crucially dependent on how far from maximal θ_{23} is allowed to be and a prior which allows larger deviations will typically give a more optimistic prediction for the capabilities of an experiment to discover the octant. Let us also note that if the parameter knowledge is not very precise, leading to good perceived chances of a discovery, and a discovery is later not made once the experiment is performed, the experiment should still typically not be considered a failure as it then will tend to disfavor those parts of parameter space where a discovery would be easy, thus providing a better prior as input for the next experiment.

As an example of the dependence on the prior, we show the case where δ is fixed to $90°$ in NO and $-90°$ in IO in figure 5. Due to these points being almost degenerate, essentially no chance remains for obtaining any level of reliable evidence for which ordering is the true one through the NOνA experiment.

5 Summary and discussion

We have discussed and given concrete examples of how to apply a Bayesian analysis to the problem of neutrino mass ordering determination. In doing so, we have described the correct procedures to obtain the Bayesian posterior odds for any experiment and discussed how Bayesian methods may be used to judge the capabilities of future neutrino oscillation experiments by extending the approach and argumentation presented by Qian et al. [68]. In short, before an experiment is performed, one may speak only about the expected distribution of the posterior odds of the two different orderings. Here, the actual value of the indicator, typically taken to be $\Delta\chi^2 = \min_{IO}\chi^2 - \min_{NO}\chi^2$, and not only its sign is of importance. We want to point out that this is precisely what we would expect in a Bayesian setting, as changing a measurement by a small amount from $\Delta\chi^2 = \varepsilon$ to $-\varepsilon$ should not change the results significantly. We contrast this to the situation presented by Ciuffoli et al. ref. [69], where the ordering is identified as normal if $\Delta\chi^2 > 0$ and the probability of obtaining $\Delta\chi^2 > 0$ is taken as the confidence level. This corresponds to attempting to determine the confidence level in a frequentist fashion, i.e., asking the question of how large the ratio of infinitely repeated experiments would arrive at the correct ordering. In the frequentist nomenclature, this is simply a computation of the confidence level provided by a test which takes zero as the critical value. As we have shown, in the symmetric Gaussian approximation, this probability only corresponds to the chance of the correct ordering being favored after the experiment is performed and says nothing about how favored it has to be (just that it has to be above 50%). A more interesting question than this would be to ask for the probability that an experiment will give at least strong or very strong evidence for the correct ordering, corresponding to a posterior degree of belief of at least 95.3% and 99.3%, respectively, which we have also computed for the case of a Gaussian distribution of the indicator.

The second question of ref. [69] regards the confidence level obtained by a typical experiment. This is somewhat misguided, since confidence is an inherent frequentist concept and what is interesting in the Bayesian framework is simply the posterior odds. In particular, what is interpreted in ref. [69] as the probability of success is actually the posterior probability of the correct ordering, which the authors then go on to interpret as a frequentist p-value. While the p-value in a frequentist setting is the probability of obtaining a more extreme result, the posterior probability evaluated for the median experiment is the degree of belief in the correct ordering if the data would turn out to be the median expected data. Here we have also argued against the use of a number of σ for describing the strength of the evidence, since in the frequentist nomenclature the neutrino community is familiar with, this represents a measure of how much the data deviates from the expectation and not the degree of belief in a model. The third question raised in ref. [69] is asking what the probability of achieving a certain level of evidence (although it is referred to as confidence), we have discussed this extensively throughout this text and the main results include eq. (3.10) for the Gaussian approximation and figure 4 for the simulation of NOνA.

In addition to the above, we have given examples of how to perform the full Bayesian analysis both analytically for the case where a Gaussian distribution can be assumed as well as numerically in the case of non-Gaussian distributions. Non-Gaussian distributions will

typically result from appearance experiments where the value of the CP-violating phase δ is of importance. We have illustrated this by performing a simplified analysis of what may be expected from the NOνA experiment and seen that the appearance of large tails in the distribution functions imply that values of $\Delta\chi^2$ typically give stronger evidence than for the Gaussian case. For the case of composite hypotheses composed of simple hypotheses with Gaussian distributions of the indicator, we have also provided a semi-analytic approach to slightly simplify the computation of the indicator distributions for the full hypothesis. In the very end, we discussed the impact of priors in the Bayesian analysis and argued that the prior dependence is not necessarily a bad thing when it comes to decision making for future experiments, since their capabilities should be based on our current knowledge and beliefs about the true values of the underlying physics.

Let us finally remark that this analysis by no means is specific for the neutrino mass ordering. It is equally appropriate also for other binary measurements such as which octant θ_{23} belongs to, or any other binary (or even larger degeneracies) measurement in physics. In particular, it should be expected that the octant degeneracy of θ_{23} displays even more non-Gaussian behavior since, unlike for the neutrino mass ordering, the degenerate solutions are very weakly separated.

Acknowledgments

References

[1] DAYA-BAY collaboration, F. An et al., *Observation of electron-antineutrino disappearance at Daya Bay*, *Phys. Rev. Lett.* **108** (2012) 171803 [arXiv:1203.1669] [INSPIRE].

[2] DAYA BAY collaboration, F. An et al., *Improved Measurement of Electron Antineutrino Disappearance at Daya Bay*, *Chin. Phys.* **C 37** (2013) 011001 [arXiv:1210.6327] [INSPIRE].

[3] RENO collaboration, J. Ahn et al., *Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment*, *Phys. Rev. Lett.* **108** (2012) 191802 [arXiv:1204.0626] [INSPIRE].

[4] DOUBLE CHOOZ collaboration, Y. Abe et al., *Reactor electron antineutrino disappearance in the Double CHOOZ experiment*, *Phys. Rev.* **D 86** (2012) 052008 [arXiv:1207.6632] [INSPIRE].

[5] M. Banuls, G. Barenboim and J. Bernabeu, *Medium effects for terrestrial and atmospheric neutrino oscillations*, *Phys. Lett.* **B 513** (2001) 391 [hep-ph/0102184] [INSPIRE].

[6] T. Tabarelli de Fatis, *Prospects of measuring $sin^2 2\Theta_{13}$ and the sign of Δm^2 with a massive magnetized detector for atmospheric neutrinos*, *Eur. Phys. J.* **C 24** (2002) 43 [hep-ph/0202232] [INSPIRE].

[7] J. Bernabeu, S. Palomares Ruiz and S. Petcov, *Atmospheric neutrino oscillations, theta(13) and neutrino mass hierarchy*, *Nucl. Phys.* **B 669** (2003) 255 [hep-ph/0305152] [INSPIRE].

[8] S. Palomares-Ruiz and S. Petcov, *Three-neutrino oscillations of atmospheric neutrinos, θ_{13}, neutrino mass hierarchy and iron magnetized detectors*, *Nucl. Phys.* **B 712** (2005) 392 [hep-ph/0406096] [INSPIRE].

[9] D. Indumathi and M. Murthy, *A question of hierarchy: Matter effects with atmospheric neutrinos and anti-neutrinos*, *Phys. Rev.* **D 71** (2005) 013001 [hep-ph/0407336] [INSPIRE].

[10] S. Petcov and T. Schwetz, *Determining the neutrino mass hierarchy with atmospheric neutrinos*, *Nucl. Phys.* **B 740** (2006) 1 [hep-ph/0511277] [INSPIRE].

[11] A. Samanta, *The mass hierarchy with atmospheric neutrinos at INO*, *Phys. Lett.* **B 673** (2009) 37 [hep-ph/0610196] [INSPIRE].

[12] J. Kopp and M. Lindner, *Detecting atmospheric neutrino oscillations in the ATLAS detector at CERN*, *Phys. Rev.* **D 76** (2007) 093003 [arXiv:0705.2595] [INSPIRE].

[13] R. Gandhi et al., *Mass Hierarchy Determination via future Atmospheric Neutrino Detectors*, *Phys. Rev.* **D 76** (2007) 073012 [arXiv:0707.1723] [INSPIRE].

[14] A. Donini et al., *Neutrino hierarchy from CP-blind observables with high density magnetized detectors*, *Eur. Phys. J.* **C 53** (2008) 599 [hep-ph/0703209] [INSPIRE].

[15] O. Mena, I. Mocioiu and S. Razzaque, *Neutrino mass hierarchy extraction using atmospheric neutrinos in ice*, *Phys. Rev.* **D 78** (2008) 093003 [arXiv:0803.3044] [INSPIRE].

[16] R. Gandhi, P. Ghoshal, S. Goswami and S.U. Sankar, *Resolving the Mass Hierarchy with Atmospheric Neutrinos using a Liquid Argon Detector*, *Phys. Rev.* **D 78** (2008) 073001 [arXiv:0807.2759] [INSPIRE].

[17] A. Samanta, *Discrimination of mass hierarchy with atmospheric neutrinos at a magnetized muon detector*, *Phys. Rev.* **D 81** (2010) 037302 [arXiv:0907.3540] [INSPIRE].

[18] E. Fernandez-Martinez, G. Giordano, O. Mena and I. Mocioiu, *Atmospheric neutrinos in ice and measurement of neutrino oscillation parameters*, *Phys. Rev.* **D 82** (2010) 093011 [arXiv:1008.4783] [INSPIRE].

[19] M. Blennow and T. Schwetz, *Identifying the Neutrino mass Ordering with INO and NOvA*, *JHEP* **08** (2012) 058 [*Erratum ibid.* **1211** (2012) 098] [arXiv:1203.3388] [INSPIRE].

[20] E.K. Akhmedov, S. Razzaque and A.Y. Smirnov, *Mass hierarchy, 2–3 mixing and CP-phase with Huge Atmospheric Neutrino Detectors*, *JHEP* **02** (2013) 082 [*Erratum ibid.* **1307** (2013) 026] [arXiv:1205.7071] [INSPIRE].

[21] A. Ghosh, T. Thakore and S. Choubey, *Determining the neutrino mass hierarchy with INO, T2K, NOvA and reactor experiments*, *JHEP* **04** (2013) 009 [arXiv:1212.1305] [INSPIRE].

[22] S.K. Agarwalla, T. Li, O. Mena and S. Palomares-Ruiz, *Exploring the Earth matter effect with atmospheric neutrinos in ice*, arXiv:1212.2238 [INSPIRE].

[23] D. Franco et al., *Mass hierarchy discrimination with atmospheric neutrinos in large volume ice/water Cherenkov detectors*, *JHEP* **04** (2013) 008 [arXiv:1301.4332] [INSPIRE].

[24] M. Ribordy and A.Y. Smirnov, *Improving the neutrino mass hierarchy identification with inelasticity measurement in PINGU and ORCA*, *Phys. Rev.* **D 87** (2013) 113007 [arXiv:1303.0758] [INSPIRE].

[25] W. Winter, *Neutrino mass hierarchy determination with IceCube-PINGU*, *Phys. Rev.* **D 88** (2013) 013013 [arXiv:1305.5539] [INSPIRE].

[26] A. Ghosh and S. Choubey, *Measuring the Mass Hierarchy with Muon and Hadron Events in Atmospheric Neutrino Experiments*, *JHEP* **10** (2013) 174 [arXiv:1306.1423] [INSPIRE].

[27] M. Blennow and T. Schwetz, *Determination of the neutrino mass ordering by combining PINGU and Daya Bay II*, *JHEP* **09** (2013) 089 [arXiv:1306.3988] [INSPIRE].

[28] S.-F. Ge, K. Hagiwara and C. Rott, *A Novel Approach to Study Atmospheric Neutrino Oscillation*, arXiv:1309.3176 [INSPIRE].

[29] S. Petcov and M. Piai, *The LMA MSW solution of the solar neutrino problem, inverted neutrino mass hierarchy and reactor neutrino experiments*, *Phys. Lett.* **B 533** (2002) 94 [hep-ph/0112074] [INSPIRE].

[30] S. Schonert, T. Lasserre and L. Oberauer, *The HLMA project: Determination of high Δm^2 LMA mixing parameters and constraint on $|U_{e3}|$ with a new reactor neutrino experiment*, *Astropart. Phys.* **18** (2003) 565 [hep-ex/0203013] [INSPIRE].

[31] S. Choubey, S. Petcov and M. Piai, *Precision neutrino oscillation physics with an intermediate baseline reactor neutrino experiment*, *Phys. Rev.* **D 68** (2003) 113006 [hep-ph/0306017] [INSPIRE].

[32] H. Nunokawa, S.J. Parke and R. Zukanovich Funchal, *Another possible way to determine the neutrino mass hierarchy*, *Phys. Rev.* **D 72** (2005) 013009 [hep-ph/0503283] [INSPIRE].

[33] J. Learned, S.T. Dye, S. Pakvasa and R.C. Svoboda, *Determination of neutrino mass hierarchy and θ_{13} with a remote detector of reactor antineutrinos*, *Phys. Rev.* **D 78** (2008) 071302 [hep-ex/0612022] [INSPIRE].

[34] L. Zhan, Y. Wang, J. Cao and L. Wen, *Determination of the Neutrino Mass Hierarchy at an Intermediate Baseline*, *Phys. Rev.* **D 78** (2008) 111103 [arXiv:0807.3203] [INSPIRE].

[35] L. Zhan, Y. Wang, J. Cao and L. Wen, *Experimental Requirements to Determine the Neutrino Mass Hierarchy Using Reactor Neutrinos*, *Phys. Rev.* **D 79** (2009) 073007 [arXiv:0901.2976] [INSPIRE].

[36] P. Ghoshal and S.T. Petcov, *Neutrino mass hierarchy determination using reactor antineutrinos*, *JHEP* **03** (2011) 058 [arXiv:1011.1646].

[37] X. Qian et al., *Mass hierarchy resolution in reactor anti-neutrino experiments: Parameter degeneracies and detector energy response*, *Phys. Rev.* **D 87** (2013) 033005 [arXiv:1208.1551].

[38] E. Ciuffoli, J. Evslin and X. Zhang, *The neutrino mass hierarchy at reactor experiments now that θ_{13} is large*, *JHEP* **03** (2013) 016 [arXiv:1208.1991].

[39] P. Ghoshal and S.T. Petcov, *Addendum: Neutrino mass hierarchy determination using reactor antineutrinos*, *JHEP* **09** (2012) 115 [arXiv:1208.6473].

[40] S.-F. Ge, K. Hagiwara, N. Okamura and Y. Takaesu, *Determination of mass hierarchy with medium baseline reactor neutrino experiments*, *JHEP* **05** (2013) 131 [arXiv:1210.8141].

[41] E. Ciuffoli, J. Evslin and X. Zhang, *Mass hierarchy determination using neutrinos from multiple reactors*, *JHEP* **12** (2012) 04 [arXiv:1209.2227].

[42] E. Ciuffoli, J. Evslin and X. Zhang, *Neutrino mass hierarchy from nuclear reactor experiments*, *Phys. Rev.* **D 88** (2013) 033017 [arXiv:1302.0624].

[43] Y.-F. Li, J. Cao, Y. Wang and L. Zhan, *Unambiguous determination of the neutrino mass hierarchy using reactor neutrinos*, *Phys. Rev.* **D 88** (2013) 013008 [arXiv:1303.6733].

[44] A.B. Balantekin et al., *Neutrino mass hierarchy determination and other physics potential of medium-baseline reactor neutrino oscillation experiments*, arXiv:1307.7419.

[45] F. Capozzi, E. Lisi and A. Marrone, *Neutrino mass hierarchy and electron neutrino oscillation parameters with one hundred thousand reactor events*, arXiv:1309.1638.

[46] E. Ciuffoli et al., *Advantages of Multiple Detectors for the Neutrino Mass Hierarchy Determination at Reactor Experiments*, arXiv:1308.0591.

[47] A. Cervera et al., *Golden measurements at a neutrino factory*, *Nucl. Phys.* **B 579** (2000) 17 [hep-ph/0002108].

[48] V. Barger, S. Geer, R. Raja and K. Whisnant, *Determination of the pattern of neutrino masses at a neutrino factory*, *Phys. Lett.* **B 485** (2000) 379 [hep-ph/0004208].

[49] J. Burguet-Castell, M.B. Gavela, J.J. Gómez-Cadenas, P. Hernández and O. Mena, *Superbeams plus neutrino factory: the golden path to leptonic CP-violation*, *Nucl. Phys.* **B 646** (2002) 301 [hep-ph/0207080].

[50] P. Huber, M. Lindner and W. Winter, *Synergies between the first-generation JHF-SK and NuMI superbeam experiments*, *Nucl. Phys.* **B 654** (2003) 3 [hep-ph/0211300].

[51] H. Minakata, H. Nunokawa and S. Parke, *Complementarity of eastern and western hemisphere long-baseline neutrino oscillation experiments*, *Phys. Rev.* **D 68** (2003) 013010 [hep-ph/0301210].

[52] P. Coloma et al., θ_{13}, δ *and the neutrino mass hierarchy at a* $\gamma = 350$ *double baseline Li/B* β-*beam*, *JHEP* **05** (2008) 050 [arXiv:0712.0796].

[53] P. Huber, M. Lindner, T. Schwetz and W. Winter, *First hint for CP-violation in neutrino oscillations from upcoming superbeam and reactor experiments*, *JHEP* **11** (2009) 044 [arXiv:0907.1896].

[54] S.K. Agarwalla, T. Li and A. Rubbia, *An incremental approach to unravel the neutrino mass hierarchy and CP-violation with a long-baseline superbeam for large* θ_{13}, *JHEP* **05** (2012) 154 [arXiv:1109.6526].

[55] P. Coloma and E. Fernandez-Martinez, *Optimization of neutrino oscillation facilities for large* θ_{13}, *JHEP* **04** (2012) 089 [arXiv:1110.4583].

[56] S. Prakash, S.K. Raut and S.U. Sankar, *Getting the best out of T2K and NOνA*, *Phys. Rev.* **D 86** (2012) 033012 [arXiv:1201.6485].

[57] P. Coloma, E. Fernández-Martínez and L. Labarga, *Physics reach of CERN-based SuperBeam neutrino oscillation experiments*, *JHEP* **11** (2012) 069 [arXiv:1206.0475].

[58] P. Coloma, T. Li and S. Pascoli, *A comparative study of long-baseline superbeams within LAGUNA for large* θ_{13}, arXiv:1206.4038.

[59] S. Dusini et al., *CP violation and mass hierarchy at medium baselines in the large* θ_{13} *era*, *Eur. Phys. J.* **C 73** (2013) 2392 [arXiv:1209.5010].

[60] S.K. Agarwalla, S. Prakash, S.K. Raut and S.U. Sankar, *Potential of optimized NOνA for large* θ_{13} & *combined performance with a LArTPC* & *T2K*, *JHEP* **12** (2012) 075 [arXiv:1208.3644].

[61] S.K. Agarwalla, S. Prakash and S.U. Sankar, *Exploring the three flavor effects with future superbeams using liquid argon detectors*, arXiv:1304.3251.

[62] V. Barger, A. Bhattacharya, A. Chatterjee, R. Gandhi, D. Marfatia and M. Masud, *Configuring the Long-Baseline Neutrino Experiment*, arXiv:1307.2519.

[63] X. Qian, J.J. Ling, R.D. McKeown, W. Wang, E. Worcester and C. Zhang, *A Second Detector Focusing on the Second Oscillation Maximum at an Off-axis Location to Enhance the Mass Hierarchy Discovery Potential in LBNE10*, arXiv:1307.7406.

[64] M.D. Messier and for the NOvA Collaboration, *Extending the NOvA Physics Program*, arXiv:1308.0106.

[65] M. Ghosh, P. Ghoshal, S. Goswami and S.K. Raut, *Synergies between neutrino oscillation experiments: An 'adequate' configuration for LBNO*, arXiv:1308.5979.

[66] M. Bass et al., *Baseline optimization for the measurement of CP-violation and mass hierarchy in a long-baseline neutrino oscillation experiment*, arXiv:1311.0212.

[67] S.S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Ann. Math. Statist.* **9** (1938) 60.

[68] X. Qian et al., *Statistical evaluation of experimental determinations of neutrino mass hierarchy*, *Phys. Rev.* **D 86** (2012) 113011 [arXiv:1210.3651].

[69] E. Ciuffoli, J. Evslin and X. Zhang, *Sensitivity to the Neutrino Mass Hierarchy*, arXiv:1305.5150.

[70] M. Blennow, P. Coloma, P. Huber and T. Schwetz, *Quantifying the sensitivity of oscillation experiments to the neutrino mass ordering*, arXiv:1311.1822.

[71] M. Blennow, P. Coloma, A. Donini and E. Fernández-Martínez, *Gain fractions of future neutrino oscillation facilities over T2K and NOvA*, *JHEP* **07** (2013) 159 [arXiv:1303.0003].

[72] R.E. Kass and A.E. Raftery, *Bayes factors*, *J. Am. Statist. Assos.* **90** (1995) 773.

[73] H. Jeffreys, *The Theory of Probability*, 3rd edition, Oxford University Press, Oxford U.K. (1961).

[74] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C 71** (2011) 1554 [arXiv:1007.1727].

[75] NOvA collaboration, D. Ayres et al., *NOvA Proposal to Build a 30 Kiloton Off-Axis Detector to Study Neutrino Oscillations in the Fermilab NuMI Beamline*, hep-ex/0503053.

[76] P. Huber, M. Lindner and W. Winter, *Simulation of long-baseline neutrino oscillation experiments with GLoBES. (General Long Baseline Experiment Simulator)*, *Comput. Phys. Commun.* **167** (2005) 195 [hep-ph/0407333].

[77] P. Huber, J. Kopp, M. Lindner, M. Rolinec and W. Winter, *New features in the simulation of neutrino oscillation experiments with GLoBES 3.0. (General Long Baseline Experiment Simulator)*, *Comput. Phys. Commun.* **177** (2007) 432 [hep-ph/0701187].

[78] M. Blennow and E. Fernandez-Martinez, *Neutrino oscillation parameter sampling with MonteCUBES*, *Comput. Phys. Commun.* **181** (2010) 227 [arXiv:0903.3985].

[79] P. Huber, J. Kopp, M. Lindner, M. Rolinec and W. Winter, *GLoBES homepage*, http://www.mpi-hd.mpg.de/personalhomes/globes/.