Sensitivity to the Neutrino Mass Hierarchy

Emilio Ciuffoli, Jarah Evslin and Xinmin Zhang

Theoretical Physics Division, IHEP, CAS, YuQuanLu 19B, Beijing 100049, China

In the next decade, a number of experiments will attempt to determine the neutrino mass hierarchy. Feasibility studies for such experiments generally determine the statistic $\overline{\Delta\chi^2}$. As the hierarchy is a discrete choice, $\Delta\chi^2$ does not obey a one degree of freedom χ^2 distribution and so the number of σ's of confidence of the hierarchy determination is not the square root of $\overline{\Delta\chi^2}$. We present a simple Bayesian formula for the sensitivity to the hierarchy that can be expected from the median experiment as a function of $\overline{\Delta\chi^2}$.

In the next two decades a number of reactor, accelerator and atmospheric neutrino experiments will attempt to determine the neutrino mass hierarchy, which is the sign of the mass difference $\Delta M_{31}^2 = M_3^2 - M_1^2$ where M_i is the ith eigenvalue of the neutrino mass matrix. If the sign is positive (negative), one says that the hierarchy is normal (inverted). Most of these experiments are still in the planning stages, where the key role is played by studies of the sensitivity of a given design to the hierarchy.

Such studies determine, either analytically or via Monte Carlo simulations,

$$\Delta\chi^2 = \chi_I^2 - \chi_N^2 \tag{1}$$

where χ_N^2 (χ_I^2) is the χ^2 statistic equal to a weighted sum of the squares of the differences between the data and predictions given the normal (inverted) hierarchy, choosing all of the nuisance parameters so as to minimize χ_N^2 (χ_I^2). The goal of these experiments is not to determine whether each of the hierarchies is consistent with the data, as would be usual in a frequentist approach. Rather, as it is already well accepted that precisely one of the hierarchies is manifested in nature, the goal of these experiments is to determine *which* of the hierarchies provides the best fit to the data. In this paper we will use the test statistic $\Delta\chi^2$ to answer this question as follows. We will define the best fit hierarchy to be that which yields the lowest value of χ^2, and so the hierarchy determined by the experiment simply corresponds to the sign of $\Delta\chi^2$.

The critical question is then, given $\Delta\chi^2$, what is the sensitivity of a typical experiment to the hierarchy? In Ref. [1] the authors showed that the most naive answer, the p value that would be obtained if $\Delta\chi^2$ satisfied a one degree of freedom χ^2 distribution, gives the incorrect answer. Indeed $\Delta\chi^2$ is not necessarily positive and so such a prescription would not even always be defined. In this note we will provide an analytic answer (13) to this question and will compare our answer to the results of simulations of Daya Bay II and disappearance data at NOνA.

Nested hypotheses

To begin, we will describe just why the p value is not the answer to the question stated above. Consider N data points $\{y_i\}$ generated by an experiment trying to determine an unknown quantity x. We will use the approximation in which these data points y_i follow a Gaussian distribution peaked at $y_i^{(0)}(x)$ with variance $\sigma_i^2(x)$. Both $y_i^{(0)}(x)$ and $\sigma_i(x)$ are known functions of x. An experimenter is interested in two hypotheses. Hypothesis (A) states that x is a real number. Hypothesis (B) states that $x = x_0$, for a particular real number x_0. Clearly hypothesis (B) is a special case of hypothesis (A), so these hypotheses are said to be nested. In particular, (B) is obtained from (A) by fixing one, otherwise unconstrained, real number, the number x.

For any given value of x, the experimenter can define a statistic $\chi^2(x)$ by simulating the experiment with that value of x and calculating the weighted sum of the squares of differences between his measured and simulated results

$$\chi^2(x) = \sum_i \frac{(y_i - y_i^{(0)}(x))^2}{\sigma_i(x)^2}. \tag{2}$$

The experimenter then determines a best fit $\overline{x}$, for which $\chi^2(x)$ is minimized. He then asks how compatible his results are with the hypothesis (B). To determine this, he calculates

$$\delta\chi^2 = \chi^2(x_0) - \chi^2(\overline{x}). \tag{3}$$

Unlike $\Delta\chi^2$ defined in Eq. (1), $\delta\chi^2$ is manifestly non-negative, because $\overline{x}$ is defined so as to give the lowest value of χ^2.

Just what value of $\delta\chi^2$ should the experimenter expect? 75 years ago Wilks proved [2] that if hypothesis (B) is true then $\delta\chi^2$ will obey a χ^2 distribution with a single degree of freedom. The experimenter can then determine a conditional probability that given (B), the experiment would have gone as badly as it did

$$p_W(\delta\chi^2) = \frac{1}{2}\left(1 - \mathrm{erf}\left(\sqrt{\frac{\delta\chi^2}{2}}\right)\right). \qquad (4)$$

For example, if he found $\delta\chi^2 = 9$, then p_W would only be about 0.13%, and so a frequentist experimenter might conclude that he has ruled out (B) with 3σ of confidence.

Non-nested Hypotheses

As described in Ref. [1], the determination of the hierarchy is qualitatively different. The two hypotheses are the normal hierarchy (NH) and the inverted hierarchy (IH). These hypotheses are not nested, and they correspond to a discrete choice, not the fixing of a degree of freedom. So the conditions for Wilks' theorem are strongly violated. As was observed in general in Ref. [3] and in this context in Ref. [1], this means that the statistic $\Delta\chi^2$ defined in Eq. (1) does not follow a χ^2 distribution.

Just what distribution does $\Delta\chi^2$ follow? Let us begin with the simple case in which there are no nuisance parameters, which was applied to a toy model of the hierarchy determination in Ref. [1].

An experiment will produce a set of numbers $\{y_i\}$, which we assemble into a vector $\mathbf{y}$. The normal and inverted hierarchies yield two theoretical estimates of this vector which we will denote $\mathbf{y}^N$ and $\mathbf{y}^I$ respectively. Again let us assume that the measured numbers y_i are normally distributed about their mean with a variance σ_i^2, which for simplicity we take to be independent of the hierarchy. Without loss of generality, let us assume for the moment that the true hierarchy is normal. Then the measured numbers will be

$$y_i = y_i^N + \sigma_i g_i \qquad (5)$$

where g_i is a standard Gaussian random variable.

The statistic $\Delta\chi^2$ is then easily determined to be

$$\begin{aligned}
\Delta\chi^2 &= \chi_I^2 - \chi_N^2 \qquad (6)\\
&= \sum_i \frac{(y_i - y_i^I)^2}{\sigma_i^2} - \sum_i \frac{(y_i - y_i^N)^2}{\sigma_i^2}\\
&= \sum_i \frac{(y_i^N + \sigma_i g_i - y_i^I)^2 - (y_i^N + \sigma_i g_i - y_i^N)^2}{\sigma_i^2}\\
&= \sum_i \frac{(y_i^N - y_i^I)^2}{\sigma_i^2} + \sum_i \frac{2(y_i^N - y_i^I)}{\sigma_i} g_i.
\end{aligned}$$

This identifies $\Delta\chi^2$ as a Gaussian distributed random variable with mean given by the first term on the right hand side

$$\overline{\Delta\chi^2} = \sum_i \frac{(y_i^N - y_i^I)^2}{\sigma_i^2} \qquad (7)$$

and standard deviation given by the second term [1]

$$\sigma_{\Delta\chi^2} = \sqrt{\sum_i \frac{4(y_i^N - y_i^I)^2}{\sigma_i^2}} = 2\sqrt{\overline{\Delta\chi^2}}. \qquad (8)$$

Note that $\overline{\Delta\chi^2}$ is the $\Delta\chi^2$ statistic without statistical fluctuations, for example it may be given by the theoretical spectra of $\overline{\nu}_e$ observed at a reactor experiment, of ν_μ and $\overline{\nu}_\mu$ at an iron calorimeter atmospheric neutrino experiment, or of ν_e ($\overline{\nu}_e$) appearance at an accelerator experiment running in the neutrino (antineutrino) mode. In an atmospheric neutrino experiment one may use the spectra as a function of energy, zenith angle and even the inelasticity of the events [4]. $\overline{\Delta\chi^2}$ is the statistic most often reported in the literature. We will now use Eq. (8) to relate $\overline{\Delta\chi^2}$ to three quantities of interest.

What is the probability that the hierarchy which yields the lowest χ^2 is indeed the true hierarchy?

Let us first consider the case in which the normal hierarchy is manifested in nature. The correct hierarchy will be determined by the experiment if $\Delta\chi^2 > 0$. The statistic $\Delta\chi^2$ is centered on the positive value $\overline{\Delta\chi^2}$ with a standard deviation of $2\sqrt{\overline{\Delta\chi^2}}$ and so the closest negative value is $\sqrt{\overline{\Delta\chi^2}}/2$ σ's from the mean, on one side of the distribution. For example, if $\overline{\Delta\chi^2} = 9$ then a negative value will be excluded at 1.5σ's on one side, yielding a probability of successfully determining the hierarchy of 93.3%, considerably less than the 99.7% that one may naively suspect just by taking the square root of $\overline{\Delta\chi^2}$. More

generally, the probability of correctly determining the hierarchy is

$$p_c(\overline{\Delta\chi^2}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\sqrt{\frac{\overline{\Delta\chi^2}}{8}}\right)\right). \qquad (9)$$

In a more standard terminology, p_c is the sensitivity to the hierarchy of the binary classification test whose classification function is the sign of $\Delta\chi^2$. We will refer to it simply as the "probability of success" in what follows. In Ref. [5] the authors obtained a similar result which differs as a result of their formula (5.11) for the probability of success for a given $\Delta\chi^2$.

If instead the inverted hierarchy is correct, the calculation proceeds identically. As we have approximated σ_i to be hierarchy independent, the probability of success is identical for both hierarchies. This is the quantity quoted in a number of studies such as Refs. [6–8].

Second, what is the sensitivity of a typical experiment to the hierarchy?

A "typical experiment" is one in which $|\Delta\chi^2|$ obtains its average value $|\overline{\Delta\chi^2}|$. As the probability of successfully determining the hierarchy is a monotonic function of $\Delta\chi^2$, the average value of $\Delta\chi^2$ corresponds to the median value of the probability of success and so we will refer to such experiments as median experiments. The sensitivity of the $\mathrm{sign}(\Delta\chi^2)$ test to the hierarchy is the probability that a fit to the correct hierarchy yields a lower value of χ^2 than one to the wrong hierarchy. Since $|\Delta\chi^2|$ is fixed, this is simply the probability that $\Delta\chi^2$ has the correct sign.

Again the calculation will proceed identically for both hierarchies, so we may restrict our attention to the case in which the normal hierarchy is correct. Therefore the question is, given that $\overline{\Delta\chi^2}$ is positive and $\Delta\chi^2$ is equal to either $\overline{\Delta\chi^2}$ or $-\overline{\Delta\chi^2}$, what is the probability p_v that $\Delta\chi^2 = \overline{\Delta\chi^2}$.

Let $L_\pm$ be the likelihood, given the normal hierarchy, that $\Delta\chi^2 = \pm\overline{\Delta\chi^2}$, which is easily found using the fact that $\Delta\chi^2$ obeys a normal distribution centered at $\overline{\Delta\chi^2}$ with standard deviation $2\sqrt{\overline{\Delta\chi^2}}$. Using the fact that the distribution of $\Delta\chi^2$ is odd with respect to a change in the hierarchy, the Bayes

factor for the normal hierarchy is

$$\frac{L_+}{L_-} = e^{\overline{\Delta\chi^2}/2}. \qquad (10)$$

In particular, symmetric Bayesian priors assigning a 50% chance to each hierarchy yield a probability of success of

$$p_v = \frac{L_+}{L_+ + L_-} = \frac{1}{1 + e^{-\overline{\Delta\chi^2}/2}} \qquad (11)$$

for median experiments, those in which $|\Delta\chi^2| = |\overline{\Delta\chi^2}|$. For example, if $\overline{\Delta\chi^2} = 9$ then the probability that a median experiment correctly determines the hierarchy will be 98.9%. While this is better than the mean probability of success 93.3%, it still falls noticeably short of the 99.7% which one might expect from Wilks' theorem. In Ref. [9] it was noted that the sensitivity (11) is equal to the posterior probability of determining the correct hierarchy.

Given $\overline{\Delta\chi^2}$ determined either from Monte Carlo simulations or from Asimov data, one may express the sensitivity to the hierarchy expected at a median experiment in terms of a number s of standard deviations σ. We will convert probabilities into standard deviations using the one-sided Gaussian distribution

$$p_v(\overline{\Delta\chi^2}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{s}{\sqrt{2}}\right)\right). \qquad (12)$$

While the double-sided Gaussian is also often used in the literature, we have checked that this choice of convention has a small effect on our results.

Using Eq. (11) one now finds that the number of σ's of sensitivity is

$$s(\overline{\Delta\chi^2}) = \sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1 - e^{-\overline{\Delta\chi^2}/2}}{1 + e^{-\overline{\Delta\chi^2}/2}}\right). \qquad (13)$$

This function is plotted in Fig. 1. For example, if $\overline{\Delta\chi^2} = 9$ then a median experiment determines the hierarchy with a sensitivity of 2.3σ instead of the 3σ which might be expected. Had we insteaded opted for the double-sided Gaussian convention for s, we would have instead found 2.5σ.

A general Bayesian prior of b and $1 - b$ for the normal and inverted hierarchies leads to a sensitivity

$$s(\overline{\Delta\chi^2}) = \sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1 + \left(1 - \frac{1}{b}\right)e^{-\overline{\Delta\chi^2}/2}}{1 + \left(\frac{1}{b} - 1\right)e^{-\overline{\Delta\chi^2}/2}}\right). \qquad (14)$$
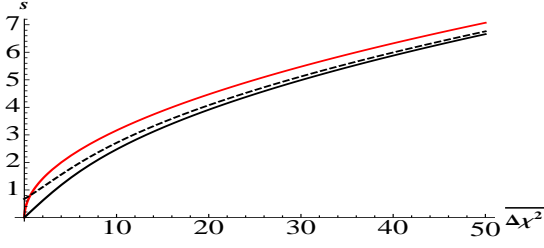
FIG. 1: For a given $\overline{\Delta\chi^2}$ statistic determined from theoretical spectra, the black curve is the number s of σ's of sensitivity of the determination of the mass hierarchy by a median experiment. A median experiment is one in which $|\Delta\chi^2|$ obtains its median value. For comparison, the dashed curve uses the two-sided definition of s and the red curve is the square root of $\overline{\Delta\chi^2}$.

FIG. 2: The black, red, blue, purple and green curves are the probability of a hierarchy determination with 1σ, 2σ, 3σ, 4σ and 5σ of sensitivity as a function of $\overline{\Delta\chi^2}$. The dashed line represents a median experiment, and its intersections with the curves yield the same information as Fig. 1.

Third, what is the probability $p(s)$ that the hierarchy will be determined with a sensitivity of at least $s\sigma$?

Note first that for a general experimental outcome $\Delta\chi^2$, the probability of success

$$
\begin{aligned}
p_v &= \frac{L_+}{L_+ + L_-} \\
&= \frac{e^{-(\overline{\Delta\chi^2}-\Delta\chi^2)^2/8\overline{\Delta\chi^2}}}{e^{-(\overline{\Delta\chi^2}-\Delta\chi^2)^2/8\overline{\Delta\chi^2}} + e^{-(\overline{\Delta\chi^2}+\Delta\chi^2)^2/8\overline{\Delta\chi^2}}} \\
&= \frac{1}{1+e^{-\Delta\chi^2/2}}
\end{aligned}
\tag{15}
$$

is independent of $\overline{\Delta\chi^2}$. Using this fact, an argument similar to those above leads to

$$
p(s) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\overline{\Delta\chi^2} - 4\mathrm{arctanh}\left(\mathrm{erf}\left(\frac{s}{\sqrt{2}}\right)\right)}{\sqrt{8\overline{\Delta\chi^2}}}\right)\right).
\tag{16}
$$

This function is plotted in Fig. 2.

Parallel nuisance parameters

In reality there is no single experimental result $\mathbf{y}^{\mathrm{N}}$ or $\mathbf{y}^{\mathrm{I}}$ which is predicted by a given hierarchy. The results also depend on a number of nuisance parameters, such as the neutrino mass matrix parameters and the flux normalization of the source. We will assemble these nuisance parameters into a vector $\mathbf{x} = \{x_i\}$.

If the final data consists of N numbers, such as the number of events in N energy bins, and if there are K nuisance parameters, then each hierarchy cor-
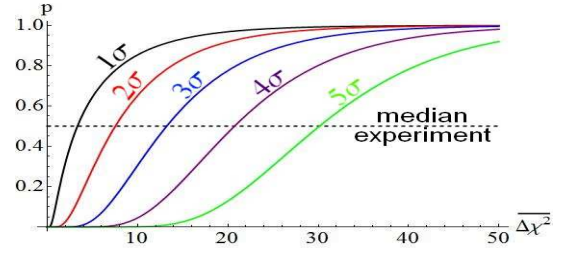
responds not to a point but to a K-dimensional subset of the N-dimensional vector space in which $\mathbf{y}$ is valued. The nuisance parameters x_i are coordinates on these subsets. If the standard deviations σ_i vary sufficiently slowly, then the inverse covariance matrix defines a metric on this space. Recall that, in the case of the normal (inverted) hierarchy, the nuisance parameters $\mathbf{x}$ are chosen to minimize χ^2_N (χ^2_I). Geometrically, this minimization corresponds to choosing the point in each subset which is closest to $\mathbf{y}$, the coordinates of the point are the nuisance parameters which minimize the corresponding χ^2 statistic.

In this framework, it is easy to combine data from multiple experiments. They can simply be added to $\mathbf{y}$ as new components. For example, one can combine a forecast spectrum of Daya Bay II with a value of the nuisance parameter θ_{13} determined at Daya Bay and RENO by letting the first $N - 2$ components of $\mathbf{y}$ correspond to the $\overline{\nu_e}$ spectrum at Daya Bay II and the next two to the relative survival probabilities observed at Daya Bay and RENO. The single nuisance parameter θ_{13} yields a curve in the N-dimensional space of observations for each hierarchy. The curve is parameterized by θ_{13}. The last two coordinates of this curve are simply the relative survival probabilities expected at Daya Bay and RENO as a function of the parameter θ_{13}. The χ^2 to be minimized is the distance in the full N dimensional space, so it automatically combines determinations of θ_{13} at RENO, Daya Bay and Daya Bay II without the need for any penalty terms.

Now let us make two approximations. First, we approximate $\mathbf{y}^{\mathrm{N}}$ and $\mathbf{y}^{\mathrm{I}}$ to be linear (or affine) functions of the nuisance parameters $\mathbf{x}$, so that the subspaces corresponding to theoretical predictions are hyperplanes. The resulting models are called linear regression models. Model selection in one dimensional non-nested linear regression models was first studied in Ref. [10]. Ref. [3] presented a statistic, generalizing $\Delta\chi^2$, which is Gaussian distributed and distinguishes the models. The properties of this statistic, in the case of linear regression models, were determined in Ref. [11].

One may object that the spectra are not indeed linear functions of the neutrino mass matrix. However the essential point is that they be approximately linear in a regime whose size is the precision to which an experiment can determine the nuisance parameters. This is a much easier criterion. Later we will compare our analytical results to simulations in which no such approximation is made, and we will see that the resulting error is small.

For now we will make the further approximation that one obtains the same value of $\Delta\chi^2$ for any value of the nuisance parameters chosen for the normal hierarchy if the nuisance parameters for the inverse hierarchy are chosen so as to minimize $\Delta\chi_I^2$. In other words, $\overline{\Delta\chi^2}$, is independent of the choice of the nuisance parameters so long as each χ^2 is properly minimized. Geometrically this means that the hyperplanes corresponding to the theoretical values $\mathbf{y}^{\mathrm{N}}$ and $\mathbf{y}^{\mathrm{I}}$ are parallel.

Again assume that the normal hierarchy is correct. If $\mathbf{x}^{\mathrm{T}}$ is the true value of the nuisance parameters, then the theoretical values of the observables $\mathbf{y}^{\mathrm{N}}$ will be linear functions y_i^{N} of $\mathbf{x}^{\mathrm{T}}$. χ_N^2 (χ_I^2) is just the minimum distance squared from the observations $y_i = y_i^{\mathrm{N}} + \sigma_i g_i$ to the hyperplane corresponding to the normal (inverted) hierarchy. The statistical fluctuation vector $\mathbf{g} = \sigma_i g_i$ can be decomposed into a two vectors, $\mathbf{g}^\perp$ and $\mathbf{g}^\parallel$ such that $\mathbf{g}^\perp$ is perpendicular to the hyperplanes and $\mathbf{g}^\parallel$ is parallel.

To determine χ_N^2 or χ_I^2, one must choose the nuisance parameters $\mathbf{x}$ at which it is minimized. χ^2 will be minimized for the choice of nuisance parameters $\mathbf{x}^{\mathrm{T}} + \mathbf{g}^\parallel$. In other words, the parallel part of $\mathbf{g}$ yields the statistical error in the determination of the nuisance parameters. We have assumed that this error is the same for both hierarchies. For this choice of
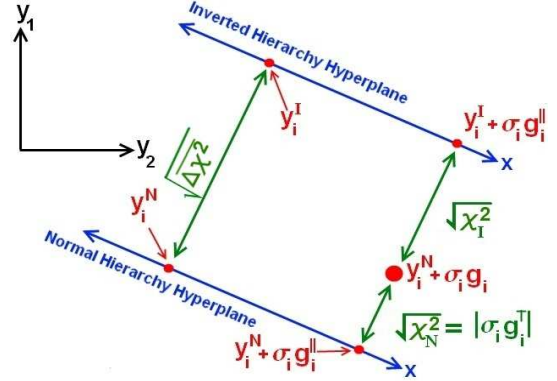

FIG. 3: In this figure the hierarchy is normal and $\overline{\Delta\chi^2}$ is independent of the nuisance parameters. The two parallel lines are the expected measurements corresponding to various values of the nuisance parameters for the two hierarchies. As a result of statistical fluctuations $y_i^{\mathrm{N}} + \sigma_i g_i$ is measured instead of the theoretical value y_i^{N}. The parallel part of $\mathbf{g}$ determines the effect of this fluctuation on the best fit nuisance parameters and the perpendicular part its effect on $\Delta\chi^2$.

nuisance parameters, the theoretical predictions for y_i are $y_i^{\mathrm{N}} + \sigma_i g_i^\parallel$ and $y_i^{\mathrm{I}} + \sigma_i g_i^\parallel$ in the cases of the two hierarchies.

Now we are ready to calculate

$$
\begin{aligned}
\Delta\chi^2 &= \chi_I^2 - \chi_N^2 \qquad\qquad (17)\\
&= \sum_i \frac{(y_i - y_i^{\mathrm{I}} - \sigma_i g_i^\parallel)^2}{\sigma_i^2} - \sum_i \frac{(y_i - y_i^{\mathrm{N}} - \sigma_i g_i^\parallel)^2}{\sigma_i^2}\\
&= \sum_i \frac{(y_i^{\mathrm{N}} + \sigma_i g_i^\perp - y_i^{\mathrm{I}})^2 - (y_i^{\mathrm{N}} + \sigma_i g_i^\perp - y_i^{\mathrm{N}})^2}{\sigma_i^2}\\
&= \sum_i \frac{(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})^2}{\sigma_i^2} + \sum_i \frac{2(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})}{\sigma_i} g_i^\perp.\\
&= \sum_i \frac{(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})^2}{\sigma_i^2} + \sum_i \frac{2(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})}{\sigma_i} g_i.
\end{aligned}
$$

In the last step we used the identity

$$
\sum_i \frac{2(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})}{\sigma_i} g_i^\parallel = 0 \qquad (18)
$$

which follows from the fact that, using the metric $1/\sigma_i^2$, the vector $(y_i^{\mathrm{N}} - y_i^{\mathrm{I}})$ is perpendicular to the hyperplanes and so to $\sigma_i g_i^\parallel$.

Just as in Eq. (6), Eq. (17) describes a normal distribution centered at $\overline{\Delta\chi^2}$ and with standard deviation $2\sqrt{\overline{\Delta\chi^2}}$. As a result, Eqs. (9) and (13) for

the probability of success and number of σ's of sensitivity in the median experiment remain correct.

General nuisance parameters

Of course, $\overline{\Delta\chi^2}$ does depend on the nuisance parameters, and so the hyperplanes corresponding to the theoretical data are not parallel and the above results are only approximate. This fact was first noted in Ref. [3], where it was concluded that as a result $\Delta\chi^2$ is not normally distributed. Its distribution leptokurtic.

This observation can be intuitively understood as follows. Imagine that $\overline{\Delta\chi^2}$ depends so strongly upon the nuisance parameters that a 1σ change in the nuisance parameters can reduce the sensitivity to the hierarchy by several σ's. As a result, most of the experiments in which the hierarchy determination is incorrect will be those in which the nuisance parameter is such that $\overline{\Delta\chi^2}$ is much smaller. Thus the tails of the distribution of $\Delta\chi^2$ will grow as a result of those simulations in which the nuisance parameters take a nonstandard value. Clearly, this effect is only present in simulations in which the nuisance parameters are allowed to vary, and so simulations that fix the nuisance parameters will yield values of $\Delta\chi^2$ which, upon using Eq. (13), overestimate the sensitivity to the hierarchy.

In Ref. [3] the author proposed a new statistic which does follow a Gaussian distribution even in this more general setting. However, in the case of the hierarchy determinations planned in the near future, the angle between the hypersurfaces is actually quite small. This is reflected in the observation [13] that even a 1σ variation in θ_{13} only leads to about a one third of a σ variation in the confidence. Therefore the approximate treatment of the $\Delta\chi^2$ statistic above is quite precise.

To illustrate this point, in Fig. 4 we present the distribution of the $\Delta\chi^2$ statistic in simulations which combine the $\overline{\nu}_e$ spectrum measured at Daya Bay II with MINOS' 4% determination of the atmospheric mass difference [12] and also with an optimistic 1% forecast determination at an upgraded NOνA . All of the nuisance parameters are fixed except for $|\Delta M_{32}^2|$, which is chosen to minimize χ_I^2 and χ_N^2 as described above. Following [14] we have considered 6 years of exposure at a 20 kton detector for Daya Bay II which detects $\overline{\nu}_e$ via inverse β decay on the 10% of
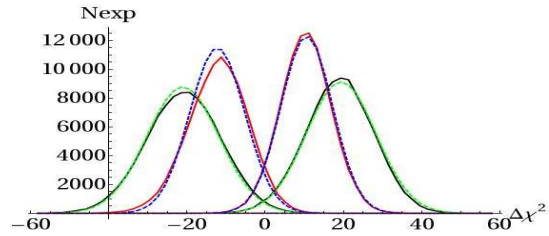

FIG. 4: The distribution of $\Delta\chi^2$ in 50,000 experiments with each hierarchy is shown, combining the data with MINOS' 4% determination of the atmospheric mass splitting (red curve) and with an optimistic 1% determination at NOνA (black curve). The dashed curves are the corresponding Gaussian distributions centered at $\overline{\Delta\chi^2}$ with width $2\sqrt{\overline{\Delta\chi^2}}$.

its mass consisting of free protons. The baselines and reactor fluxes are identical to Ref. [14]. The leptonic CP-violating angle δ is set to $\pi/2$.

We find that the distribution of $\Delta\chi^2$ is indeed well approximated by a Gaussian distribution centered at $\overline{\Delta\chi^2}$ with standard deviation $2\sqrt{\overline{\Delta\chi^2}}$. $\overline{\Delta\chi^2} \sim$ 11 (20) for Daya Bay II with MINOS (NOνA) yielding 2.6σ (3.9σ) of sensitivity at the median experiment, with a rate of successfully determining the hierarchy of 94.6% (98.5%) in good agreement with Eq. (9).

In Fig. 5 we present the distribution of $\Delta\chi^2$ in simulations in which $\delta = 0$ and π, although we always fit to a $\delta = \pi/2$ theoretical mode as the appearance mode at T2K and NOνA cannot distinguish 0 and π [13, 15]. At $\delta = 0$ (π) we find $\overline{\Delta\chi^2} = 17$ (22) yielding 3.5σ (4.2σ) of sensitivity, confirming the expectations of Ref. [16]. Despite the fact that the model used for fitting differs from that used to generate the data, the distribution of $\Delta\chi^2$ described in this paper approximates the simulated data well.

Frequentist confidence

A frequentist notion of confidence can be made well defined even in this context [17, 18]. Imagine that an experiment measures $\Delta\chi^2$. This differs from the expected $\overline{\Delta\chi^2}$ for the normal (inverted) hierarchy by $|\overline{\Delta\chi^2} \mp \Delta\chi^2|$ which corresponds to a frequentist incompatibility of

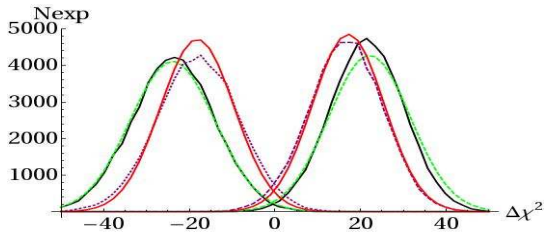$$\frac{|\overline{\Delta\chi^2} \mp \Delta\chi^2|}{2\sqrt{\overline{\Delta\chi^2}}} \tag{19}$$

FIG. 5: As in Fig. 4, but using only a 1% determination of the atmospheric mass splitting. The simulations reported in the red and black curves use $\delta = 0$ and π respectively, although the fitting is always performed assuming $\delta = \pi/2$. As can be seen, if $\delta = \pi$, the hierarchy determination will be more reliable [13, 14].

σ's.

In particular, in the case of the median experiment with the true hierarchy, $\Delta\chi^2 = \overline{\Delta\chi^2}$. Therefore the inverted hierarchy is excluded at a confidence of $\sqrt{\Delta\chi^2}$ σ's. In this sense it might be tempting to ignore the results of this paper and to identify the frequentist incompatibility $\sqrt{\Delta\chi^2}$ with the confidence in the hierarchy determination expected in a median experiment.

While such a definition of confidence is well-defined, it has a very unattractive feature. Consider an experiment with an expected $\overline{\Delta\chi^2} = 16$. The general arguments in this note imply that if the hierarchy is normal (inverted) then $\Delta\chi^2$ will follow a Gaussian distribution centered on 16 (-16) with a width of $\sigma = 8$. In the frequentist sense, the median experiment will yield $|\Delta\chi^2| = 16$ and so is incompatible with the false hierarchy with 4σ of confidence while the 98th percentile experiment will yield $\Delta\chi^2 = 0$ and so is incompatible with the false hierarchy with 2σ of confidence. An identification of the sensitivity to the hierarchy with the frequentist incompatibility would therefore imply that even the 98th percentile of experimental outcomes will yield a 2σ sensitivity to the hierarchy.

Now consider the somewhat unlikely case in which due to statistical fluctuations, the results of this experiment are indeed in the 98th percentile, so that $\Delta\chi^2 = 0$. Now the experimentalist will be asked to provide the hierarchy with 2σ of confidence. Of course he cannot, the experiment has not yielded any preference for either hierarchy, even at the 2σ level that was promised for a 98th percentile experi-

ment when the funding was requested. In this sense, the identification of the frequentist incompatibility with the confidence in the hierarchy determination is misleading: the confidence can be nonzero even when no information is obtained.

The basic problem with the application of the frequentist notion of confidence in this example is that both hierarchies have been ruled out with equal confidence. Ruling out both hierarchies can be useful when searching for new physics, testing assumptions regarding backgrounds, etc. Although in that case one would use a χ^2 test and not a $\Delta\chi^2$ test, as the latter is insensitive to effects that affect both hierarchies similarly. However, for the purpose of determining which hierarchy is manifested in nature it is reasonable to assume that one of the hierarchies is indeed correct. In this case one is led to the Bayesian constructions described in this note.

Acknowledgement

[1] X. Qian, A. Tan, W. Wang, J. J. Ling, R. D. McKeown and C. Zhang, "Statistical Evaluation of Experimental Determinations of Neutrino Mass Hierarchy," Phys. Rev. D **86** (2012) 113011 [arXiv:1210.3651 [hep-ph]].

[2] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses" Ann. Math. Stat. **9**, No. 1 (1938), 60-62.

[3] D. R. Cox, "Tests of Separate Families of Hypotheses," Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. **1** (1961), 105-123.

[4] M. Ribordy and A. Y. Smirnov, "Improving the neutrino mass hierarchy identification with inelasticity measurement in PINGU and ORCA," arXiv:1303.0758 [hep-ph].

[5] S. -F. Ge, K. Hagiwara, N. Okamura and Y. Takaesu, "Determination of mass hierarchy with

medium baseline reactor neutrino experiments," arXiv:1210.8141 [hep-ph].

[6] 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 [hep-ex]].

[7] X. Qian, D. A. Dwyer, R. D. McKeown, P. Vogel, W. Wang and C. Zhang, "Mass Hierarchy Resolution in Reactor Anti-neutrino Experiments: Parameter Degeneracies and Detector Energy Response," PRD, 87, **033005** (2013) [arXiv:1208.1551 [physics.ins-det]].

[8] E. Ciuffoli, J. Evslin and X. Zhang, "Mass Hierarchy Determination Using Neutrinos from Multiple Reactors," JHEP **1212** (2012) 004 [arXiv:1209.2227 [hep-ph]].

[9] M. Blennow, "On the Bayesian approach to neutrino mass ordering," arXiv:1311.3183 [hep-ph].

[10] H. Hotelling, "The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters," Ann. Math. Statist. Volume 11, **3** (1940) 271-283

[11] M. H. Pesaran, "On the General Problem of Model Selection," Rev. of Econ. Stud. **41** (1974) 153-171.

[12] 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].

[13] E. Ciuffoli, J. Evslin and X. Zhang, "Optimizing Medium Baseline Reactor Neutrino Experiments," arXiv:1302.0624 [hep-ph].

[14] Y. -F. Li, J. Cao, Y. Wang and L. Zhan, "Unambiguous Determination of the Neutrino Mass Hierarchy Using Reactor Neutrinos," arXiv:1303.6733 [hep-ex].

[15] H. Minakata and S. J. Parke, "Correlated, Precision Measurements of θ_{23} and Δ using only the Electron Neutrino Appearance Experiments," arXiv:1303.6178 [hep-ph].

[16] H. Minakata, H. Nunokawa, S. J. Parke and R. Zukanovich Funchal, "Determining neutrino mass hierarchy by precision measurements in electron and muon neutrino disappearance experiments," Phys. Rev. D **74** (2006) 053008 [hep-ph/0607284].

[17] Y. Takaesu, "Mass Hierarchy Using Medium Baseline Detector," presented at the International Workshop on RENO-50.

[18] A. U. Abeysekara, R. Alfaro, C. Alvarez, J. D. lvarez, R. Arceo, J. C. Arteaga-Velzquez, H. A. A. Solares and A. S. Barber *et al.*, "Sensitivity of the High Altitude Water Cherenkov Detector to Sources of Multi-TeV Gamma Rays," arXiv:1306.5800 [astro-ph.HE].