PHYSICAL REVIEW C **85**, 068501 (2012)

Calculating error bars for neutrino mixing parameters

H. R. Burroughs,[1] B. K. Cogswell,[1] J. Escamilla-Roa,[1] D. C. Latimer,[2] and D. J. Ernst[1]
[1]*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235*
[2]*Department of Physics, Reed College, Portland, Oregon 97202*
(Received 5 April 2012; published 19 June 2012)

One goal of contemporary particle physics is to determine the mixing angles and mass-squared differences that constitute the phenomenological constants that describe neutrino oscillations. Of great interest are not only the best-fit values of these constants but also their errors. Some of the neutrino oscillation data is statistically poor and cannot be treated by normal (Gaussian) statistics. To extract confidence intervals when the statistics are not normal, one should not utilize the value for $\Delta\chi^2$ versus confidence level taken from normal statistics. Instead, we propose that one should use the normalized likelihood function as a probability distribution; the relationship between the correct $\Delta\chi^2$ and a given confidence level can be computed by integrating over the likelihood function. This allows for a definition of confidence level independent of the functional form of the χ^2 function; it is particularly useful for cases in which the minimum of the χ^2 function is near a boundary. We point out that the question of what is the probability that a parameter is not zero is more precisely worded as what is the maximum confidence level at which the value of zero is not included. We present two pedagogic examples and find that the proposed method yields confidence intervals that can differ significantly from those obtained by using the value of $\Delta\chi^2$ from normal statistics. For example, we find that for the T2K experiment the value of $\Delta\chi^2$ corresponding to a confidence level of 90% is 3.57 rather than the normal statistics value of 2.71.

DOI: 10.1103/PhysRevC.85.068501 PACS number(s): 14.60.Pq, 06.20.Dk

Neutrino oscillation is a unique, experimentally observed phenomenon that goes beyond the standard model of the electroweak interaction. Assuming the observations can be understood within the context of three neutrino flavors, a coherent picture of the global data is sought in terms of two mass-squared differences, three mixing angles, and one CP phase. For an individual experiment, the experimentalist assesses the values of measured parameters (and their associated errors) through a detailed Monte Carlo simulation of the experiment. For one who wishes to assess the oscillation parameters from the global neutrino oscillation data set, this procedure is not feasible. Instead, one develops a model of each experiment and compares the model results to the data through a choice of a particular statistic, often expressed as a χ^2 function. For a sufficiently large data set, normal (Gaussian) statistics can be assumed, and the χ^2 function is defined as

$$\chi^2(\{a_j\}) =: \sum_i \frac{\left[n_i^{\rm th}(\{a_j\},\{c_k\}) - n_i^{\rm exp}\right]^2}{\sigma_i^2} + \sum_k \frac{\left(c_k - c_k^{\rm th}\right)^2}{\sigma_k^2}, \tag{1}$$

where $\{a_j\}$ is a set of parameters, the mixing angles and mass-squared differences, to be determined; $\{c_k\}$ is a set of systematic errors; $n_i^{\rm exp}$ are the experimental data points; $n_i^{\rm th}(\{a_j\},\{c_k\})$ are the theoretical predictions of the data; σ_i are the statistical errors for the data points; $c_k^{\rm th}$ are the best estimates of the systematic errors; and σ_k the errors for the systematics. The systematic error parameters are usually treated as nuisance parameters and the $\chi^2(\{a_j\})$ is minimized with respect to these parameters, often using the pull method [1], for each set of the parameters $\{a_j\}$. The best-fit parameters are then the values of the a_j which minimize $\chi^2(\{a_j\})$.

Neutrino oscillations require that we must also deal with small statistical samples. In particular, the recent T2K results [2] report a total of six observed neutrino events, binned by energy into sets containing zero, one, or two counts each. Despite this paucity, the data is a significant indicator that θ_{13} is nonzero. The Super-Kamiokande (Super-K) atmospheric data afford another example. Though they provide relatively stringent bounds upon the mixing angle θ_{23} and the "atmospheric" mass-squared difference, the data also impact the determination of θ_{13}. The Super-K experiment provides an upper bound for the angle and shows a slight preference for negative values of θ_{13} [3,4]. The sensitivity of the data to θ_{13} can be traced to sub-GeV neutrinos with very long baselines [5] and the MSW resonances that occur for the normal hierarchy in the 3 to 7 GeV range [4]. The statistical significance of the data in these two regions is low, and the resulting χ^2 is not well represented by a quadratic, so that the assumption of Gaussian statistics is tenuous.

For small sample sizes, it is standard usage to employ a χ^2 function defined in terms of Poisson statistics,

$$\begin{aligned}\chi^2(\{a_j\}) =: \sum_i 2\left[n_i^{\rm th}(\{a_j\},\{c_k\}) + b_i - n_i^{\rm exp}\right] \\ + n_i^{\rm exp}\ln\left(\frac{n_i^{\rm exp}}{n_i^{\rm th}(\{a_j\},\{c_k\}) + b_i}\right) \\ + \sum_k \frac{\left(c_k - c_k^{\rm th}\right)^2}{\sigma_k^2},\end{aligned} \tag{2}$$

where b_i is a theoretical estimate of background events. The best-fit parameters remain the values of the a_j at the minimum value of $\chi^2(\{a_j\})$. In addition to being valid for small sample sizes, this χ^2 allows for the treatment of the situation where it is not possible to cleanly separate the signal from the background. Background estimates are usually assessed through Monte Carlo simulations of the experimental detection and then inserted into Eq. (2). For large sample sizes, the Poisson χ^2 limits to the normal statistic χ^2, thus allowing its use for data where some bins have good statistics but some have poor statistics, as is the case for atmospheric data.

Herein, we address the question as to how one should extract the errors on these parameters at a given confidence level. A common practice is to use the value of $\Delta\chi^2 =: \chi^2 - \chi^2_{\rm min}$ that corresponds to the desired confidence level as found from normal statistics, and then define the allowed region for the parameter a as lying within the interval $[a_o - \delta_1, a_o + \delta_2]$ where $\chi^2(a_o \pm \delta_{1,2}) = \chi^2_{\rm min} + \Delta\chi^2$ with a_o corresponding to the best fit. For example, in a review on θ_{13} phenomenology [6], the authors quote the 90% CL for $\sin^2\theta_{13}$ computed by several groups. As this mixing angle is small and the parametrization of the mixing angle is strictly positive, it is near zero, the boundary of the parameter space. By observation, it is apparent that the χ^2 for this parameter is manifestly not a quadratic and thus does not correspond to normal statistics. The authors state that their quoted 90% confidence levels on $\sin^2\theta_{13}$ is found using the value $\Delta\chi^2 = 2.71$, but for the reasons cited above, caution must be employed in using this value. Indeed, the authors of Ref. [6] admonish us that "the results on θ_{13} ... should be taken with some grain of salt, and in particular the numbers given for various confidence levels ... have to be considered only as approximate, and should always be understood in terms of the $\Delta\chi^2$ value."

We propose a method for extracting allowed regions for a single parameter at a given confidence level that does not depend on the use of normal statistics. Instead, we take a Bayesian approach and interpret the normalized likelihood function with a flat prior probability distribution function (pdf) as a posterior pdf. The likelihood function $\mathcal{L}$ is defined in terms of the χ^2 function by

$$\chi^2(\{a_i\}) =: -2\,\ln\,\mathcal{L}(\{a_i\}). \tag{3}$$

For a single parameter a, normal statistics give $\chi^2 = (a - a_o)^2/\sigma^2$ and $\mathcal{L} = \exp[-(a-a_o)^2/2\sigma^2]$, where σ is the one-standard-deviation error for a. For a compact parameter space, as with the mixing angles, one can assuredly normalize $\mathcal{L}$; for the mass-squared differences, the likelihood function falls off rapidly enough so that normalization is possible for these parameters as well. We will hereafter work with a normalized likelihood function, $\overline{\mathcal{L}}$.

The concept of marginalization is intricately connected to the calculation of allowed regions and error bars. We begin with a quick review of marginalization. Generally, the χ^2 function and the maximum likelihood function are functions of n parameters, $\{a_i\}$. Here these are the two mass-squared differences and the three mixing angles. Suppose we wish to extract information about one particular parameter, say a_1, in light of the knowledge of the remaining $n-1$ parameters. Marginalization tells us how to do so:

$$\overline{\mathcal{L}}(a_1) = \int da_2\,da_3\cdots da_n\,\overline{\mathcal{L}}(\{a_i\}). \tag{4}$$

This follows simply because with the choice of a constant prior pdf the normalized likelihood function is a pdf; hence, $\overline{\mathcal{L}}(a_1)$ is also a pdf.

Dropping the subscript 1 for simplicity, we note that the probability $\mathcal{P}$ that the parameter a lies between $a_{\rm min}$ and $a_{\rm max}$ is

$$\mathcal{P}(a_{\rm min}, a_{\rm max}) = \int_{a_{\rm min}}^{a_{\rm max}} \overline{\mathcal{L}}(a)\,da. \tag{5}$$

There exists an ambiguity in what we propose. If we make a nonlinear change in variables, such as $\theta \to \sin^2 2\theta$, and use a constant prior pdf for both, the results obtained are not independent [8] of which variable we use. In the limit of large statistics, this dependence disappears as normal statistics then become valid. In many cases, the work done by the experimentalists will resolve this ambiguity. They will have used a more sophisticated methodology to extract allowed regions and error bars, often the Feldman-Cousins [9] approach. To do this they will have chosen a variable with which to work and will have provided results in terms of that variable. In constructing a model for the global analysis, we must closely reproduce the results of the experimentalists. To do this, we must choose the same variable as they have used and use a constant prior pdf for this variable so that we are making a meaningful comparison of our results to their results.

We choose two pedagogic examples to demonstrate this procedure. For Example 1, we consider the extraction of θ_{13} with $-\pi/2 \leqslant \theta_{13} \leqslant \pi/2$ from the global analysis in Ref. [3]. (Note that this analysis does not contain the recent data from Super-K III [10], T2K [2], MINOS neutrino disappearance [11], anti-neutrino dissappearance [12] or neutrino appearance [13], Double Chooz [14], or Daya Bay [15] experiments and is used here purely for illustrative purposes.) In Fig. 1, we plot $\Delta\chi^2$ versus θ_{13} and assume the prior is flat for θ_{13}; note that $\Delta\chi^2$ is clearly not a quadratic function. In Example 2, we consider the extraction $\sin^2 2\theta_{13}$ from an analysis [7] of the recent T2K data [2]; we assume the prior to be flat for $\sin^2 2\theta_{13}$. The T2K results are dependent on the hierarchy and the sign of θ_{13}; we show the results for normal hierarchy and positive θ_{13}. In Fig. 2, we show $\Delta\chi^2$ versus $\sin^2 2\theta_{13}$. Note

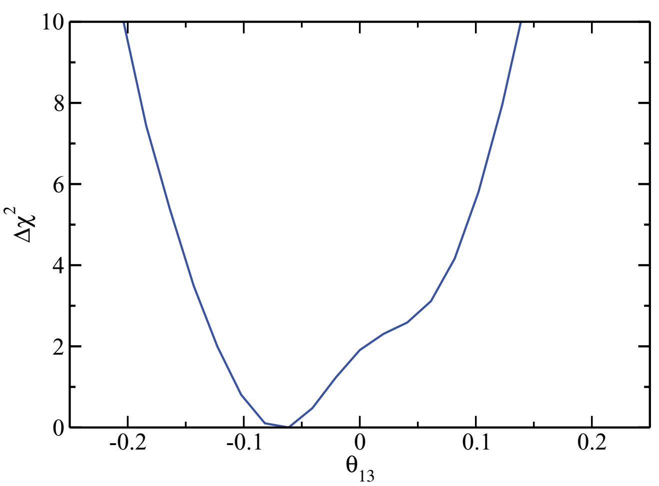

FIG. 1. (Color online) $\Delta\chi^2$ versus the mixing angle θ_{13} as taken from the global analysis given in Ref. [3].

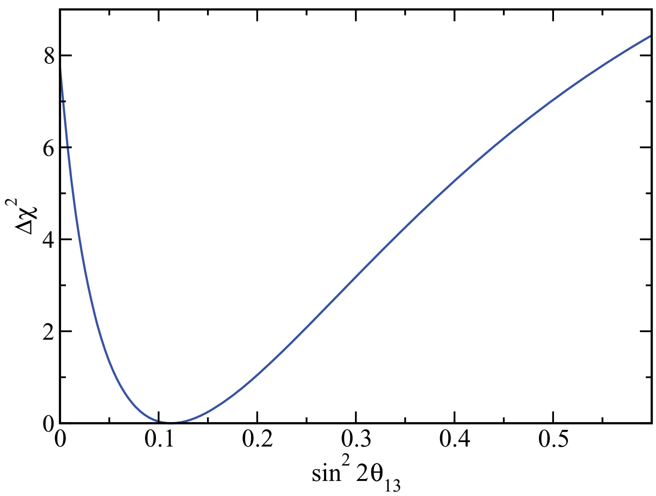

FIG. 2. (Color online) $\Delta\chi^2$ versus $\sin^2 2\theta_{13}$ for the T2K first data release [2] as taken from the analysis in Ref. [7]. The curve depicted is calculated for positive θ_{13} and normal hierarchy.

that not only is $\Delta\chi^2$ not quadratic, but the minimum is near the lower bound of zero for $\sin^2 2\theta_{13}$.

A simple application of Eq. (5) would be to ask what is the probability calculated from Fig. 1 that θ_{13} is less than zero. The result is 80%. Similarly for Fig. 2 we can find that there is a 90% probability that $\sin^2 2\theta_{13} \leqslant 0.31$.

To define a confidence level for the parameter a, we choose a value for $\Delta\chi^2$, find the two points $a_o \pm \delta_{1,2}$ that correspond to the chosen $\Delta\chi^2$, and integrate the likelihood function $\overline{\mathcal{L}}(a)$ from $a_o - \delta_1$ to $a_o + \delta_2$. The integral yields the confidence level associated with the particular value of $\Delta\chi^2$. If you desire a particular confidence level, pick an initial guess for $\Delta\chi^2$, such as the value from normal statistics, calculate the actual confidence level for this value and then repeat the process until you find the appropriate $\Delta\chi^2$ that produces the desired confidence level. The process is not computationally difficult

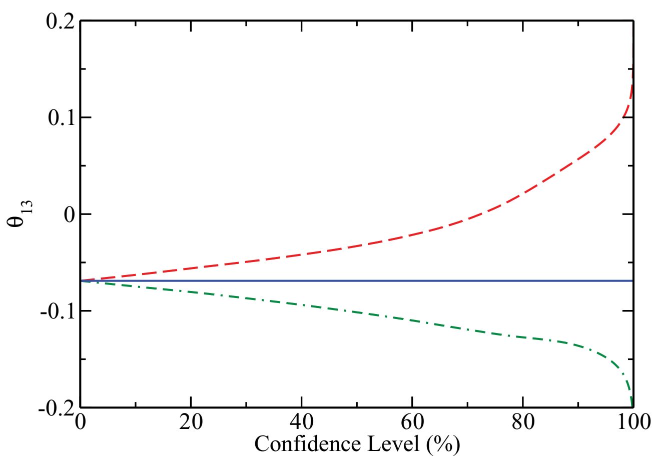

FIG. 3. (Color online) The error bars as a function of confidence level for the $\Delta\chi^2$ from Ref. [3] as depicted in Fig. 1. The solid straight (blue) horizontal line is the minimum value of θ_{13}, the dashed (red) line is the upper end of the upper error bar, while the dot-dash (green) curve is the lower end of the lower error bar.

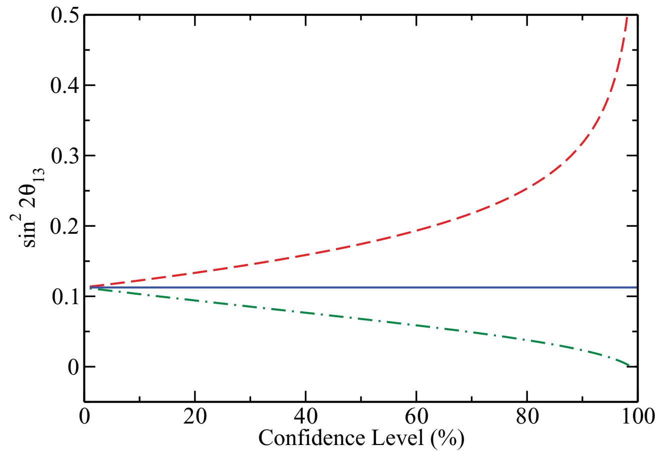

FIG. 4. (Color online) The error bars as a function of confidence level for the $\Delta\chi^2$ for T2K [2] as depicted in Fig. 2. The curves are the same as in Fig. 3.

nor computationally intensive. Note that the concept of a standard deviation applies only to normal statistics, while confidence level is universal.

For our two examples, we plot in Figs. 3 and 4 the error bars on θ_{13} and $\sin^2 2\theta_{13}$, respectively, as they vary with the confidence level. Notice the errors are asymmetric in both cases. In Fig. 4, we see that the lower error bar for $\sin^2 2\theta_{13}$ extends to zero near 98% and then remains there as the confidence level increases. This demonstrates the point that, if the best fit parameter is near a boundary of the parameter space, the confidence level will not be well approximated by the normal statistic, as $\Delta\chi^2$ is not quadratic.

In Figs. 5 and 6, we examine the relationship between $\Delta\chi^2$ and the confidence level for our two examples, comparing our results with those from normal statistics. In both figures, the (red) dashed curves utilize the normalized likelihood function, while the (blue) solid curves employ normal statistics. In Table I, we present the same information for some commonly

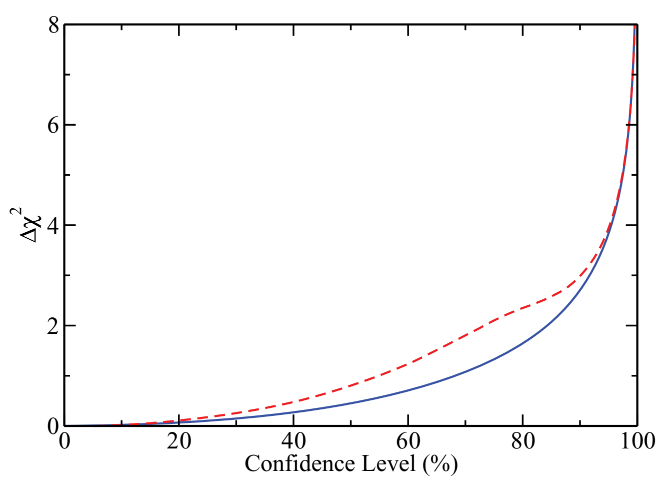

FIG. 5. (Color online) The relationship of $\Delta\chi^2$ to the confidence level. The solid (blue) curve is for normal statistics and the dashed (red) curve is calculated for the $\Delta\chi^2$ from the global analysis in Ref. [3] as depicted in Fig. 1.

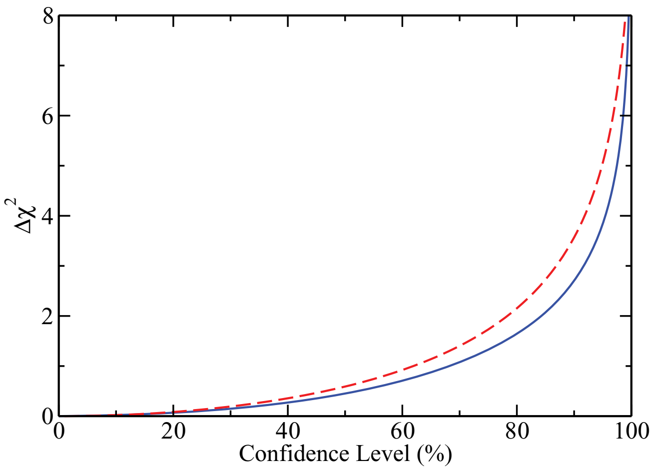

FIG. 6. (Color online) The relationship of $\Delta\chi^2$ to the confidence level. The solid (blue) curve is for normal statistics and the dashed (red) curve is calculated for the $\Delta\chi^2$ for the T2K experiment [2] in Ref. [7] as depicted in Fig. 2.

used confidence levels. We see that at low confidence levels there is a large difference between either example and the normal statistics result. For example, from Table I we see that for Example 1 the 68% confidence level corresponds to a $\Delta\chi^2$ that is a factor of 1.7 larger than the normal statistics value of 1.00, and for Example 2 the $\Delta\chi^2$ is a factor of 1.3 larger than the normal statistics. For Example 1, we can understand why the correct $\Delta\chi^2$ is larger than the normal statistics values up to the 99% confidence level. This is because the $\Delta\chi^2$ curve in Fig. 1 is more pointed than a quadratic, and it thus takes a higher value of $\Delta\chi^2$ to get a given percentage below that value. Also for Example 1, the correct and the normal statistics value are nearly equal at a confidence level of 99%, but this is accidental as the two confidence level curves intersect at a single point in this region.

The question that remains to be answered is "What is the probability that θ_{13} is or is not zero?" The correct answer to this question is that the probability that $\theta_{13} = 0$ is zero; the probability that it is not zero is one. Notice that θ_{13} can be taken to lie between $-\pi/2$ and $+\pi/2$, and zero is a single point out of the continuum. Thus the question is an ill-posed one. The more meaningful question is "What is the maximum confidence level at which zero is not an allowed value?" Consider Example 1: for normal statistics, we find $\Delta\chi^2(0) = 2.0$ at $\theta_{13} = 0$ so that we might claim that the mixing angle is nonzero at a confidence level of 84%. Using the likelihood function for Example 1, we find θ_{13} is nonzero at the 72% confidence level, knowing that we are here using the language somewhat loosely.

TABLE I. The relationship of confidence level to $\Delta\chi^2$ for some commonly used confidence levels. Three examples are given: (1) normal statistics, (2) the $\Delta\chi^2$ for θ_{13} taken from a global analysis [3] and shown in Fig. 1, and (3) the $\Delta\chi^2$ for $\sin^2 2\theta_{13}$ taken from Ref. [7] for the recent T2K data [2] and shown in Fig. 2.

Confidence level (%)	$\Delta\chi^2$		
	Normal statistics	Fig. 1 θ_{13}	Fig. 2 $\sin^2 2\theta_{13}$
68.27	1.00	1.70	1.30
90.00	2.71	3.00	3.57
95.00	3.84	3.95	5.06
95.42	4.00	4.09	5.24
99.00	6.63	6.65	8.05
99.73	9.00	8.90	9.42

For Example 2, the likelihood function excludes $\theta_{13} = 0$ as an allowed value at the 98.8% confidence level. The $\Delta\chi^2$ value at $\theta_{13} = 0$ of 7.97 would give, using normal statistics, 99.5%. Why do we find this overestimation? From Fig. 2 we see that below the minimum $\Delta\chi^2$ rises quite rapidly while above the minimum $\Delta\chi^2$ rises slowly. This combination will always yield an overestimation of the confidence level extracted from a single point on the lower, rapidly rising curve.

In summary, we propose that confidence level and error bars be calculated based on the understanding that the normalized likelihood function is a probability distribution function for whatever statistic is chosen to do the analysis. The confidence level is then given by an integral over the normalized likelihood function. We find that this alters the error bars we assign to parameters, and that in the case of the minimum being near to an end point of the independent variable, such as in the case of $\sin^2 2\theta_{13}$, the change that this procedure makes can be particularly significant. Further, we note that the question of what is the probability that θ_{13} is not zero is more carefully worded as "What is the maximum confidence level at which the allowed region for θ_{13} does not include the value zero?" Using this definition, some published confidence levels for nonzero θ_{13} based on the normal statistics relationship of $\Delta\chi^2$ to confidence level are found to be overestimations.

The authors are grateful for helpful suggestions from Johannes Bergstrom. The work of B.K.C. is supported, in part, by US Department of Education Grant No. P200A090275, the work of D.J.E. is supported, in part, by US Department of Energy Grant No. DE-FG02-96ER40975, and the work of J.E-R. is supported, in part, by CONACyT, Mexico.

[1] G. L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo, Phys. Rev. D **66**, 053010 (2002).
[2] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **85**, 031103(R) (2012).
[3] J. Escamilla-Roa, D. C. Latimer, and D. J. Ernst, Phys. Rev. C **81**, 015501 (2010).
[4] J. E. Roa, D. C. Latimer, and D. J. Ernst, Phys. Rev. Lett. **103**, 061804 (2009).
[5] D. C. Latimer and D. J. Ernst, Phys. Rev. C **71**, 062501 (2005).
[6] M. Mezzetto and T. Schwetz, J. Phys. G **37**, 103001 (2010).
[7] B. K. Cogswell, H. R. Burroughs, J. Escamilla-Roa, D. C. Latimer, and D. J. Ernst (unpublished).
[8] R. D. Cousins, Am. J. Phys. **63**, 398 (1995).
[9] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).

[10] R. Wendell *et al*. (Super-Kamiokande Collaboration), Phys. Rev. D **81**, 092004 (2010).
[11] P. Adamson *et al*. (MINOS Collaboration), Phys. Rev. Lett. **106**, 181801 (2011).
[12] P. Adamson *et al*. (MINOS Collaboration), Phys. Rev. D **84**, 071103 (2011).
[13] P. Adamson *et al*. (MINOS Collaboration), Phys. Rev. Lett. **107**, 181802 (2011).
[14] Y. Abe *et al*. (DOUBLE-CHOOZ Collaboration), Phys. Rev. Lett. **108**, 131801 (2012).
[15] F. An *et al*. (DAYA-BAY Collaboration), Phys. Rev. Lett. **108**, 171803 (2012).