BAYESIAN PRESENTATION OF NEUTRINO OSCILLATION RESULTS

M. Doucet
CERN, Geneva, Switzerland

Abstract

We try a Bayesian approach to present neutrino oscillation results. To make this presentation exercise, we are using data published earlier this year. Two data samples are treated in a Bayesian approach based on the ratio of likelihoods. The combination of these two samples is also considered. To be able to appreciate the case where a signal is observed, we also apply the technique to a modified sample with added observed events. Bayesian credible intervals are obtained.

1. INTRODUCTION

It has been pointed out on several occasions that a Bayesian approach would provide a correct and consistent way to report results of searches, when the experiments are at the limit of their sensitivity [1, 2]. In the field of neutrino oscillation physics, where some experiments are excluding oscillations while others are claiming to see oscillation signals, a reliable technique to compare and interpret the results of various experiments is mandatory. In this paper, we use the Bayesian approach advocated in Reference [1] to interpret neutrino oscillation results and to combine them. For this purpose, we use the results presented during this past year [3] by the CHORUS [4] experiment at CERN. This experiment is searching for $\nu_\mu \to \nu_\tau$ oscillations in a ν_μ beam, by looking for tau decays in an emulsion target. We use this experiment as an example because it has two separate data samples that we can combine, corresponding to two channels of the tau decay: the muon channel $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ (which we denote $\tau \to \mu$) and the single charged hadron channel $\tau^- \to h^- (nh^\circ)\nu_\tau$ (which we denote $\tau \to h$). The details of the analyses of these samples are described in Reference [4]. For the present exercise, it suffices to recall that due to a higher efficiency of the tau detection, the $\tau \to \mu$ sample is more sensitive to oscillations than the $\tau \to h$ sample in spite of the fact that the $\tau \to h$ branching ratio is larger than the $\tau \to \mu$ branching ratio. The $\tau \to \mu$ sample also has less expected background than the $\tau \to h$ sample, although in both cases the expected number of background events is below unity. CHORUS has reported no candidate so far. The subject of this paper is restricted to the presentation of neutrino oscillation results, and not to the results themselves. After recalling a few Bayesian notions that we have used, we will first present each sample separately, and we will afterwards combine them. We will close the discussion by considering the Bayesian credible intervals.

2. BAYESIAN PRESENTATION OF RESULTS

Given a process having an unknown rate of occurrence, Bayes's theorem states that the probability that this rate has a given value r is related to the observed rate n by the following relation:

$$f(r|n) = \frac{f(n|r)f_\circ(r)}{\int f(n|r)f_\circ(r)\mathrm{d}r}, \tag{1}$$

where $f_\circ(r)$ is the *prior*; the probability attributed to r before the actual measurement. For a Poisson process in the presence of background, we have:

$$f(r|n) \propto \frac{e^{-(r+r_b)\mathcal{L}}\left((r+r_b)\mathcal{L}\right)^n}{n!}f_\circ(r), \tag{2}$$

where r_b is the background rate. We write the equation in terms of a *luminosity* factor $\mathcal{L}$, which relates the total number of events expected by the experiment to the rate of events: $n_{\text{expected}} = (r + r_b)\mathcal{L}$. According to Bayes's theorem, we cannot infer any probability about r from the observation n without taking into account the prior knowledge $f_\circ(r)$ we have about r. A convenient way of presenting the experimental results without having to infer such probabilities is to present the ratio of likelihoods

$$\mathcal{R}(r; n, r_b) = \frac{f(n|r, r_b)}{f(n|r = 0, r_b)}. \tag{3}$$

This is the ratio of the probability to observe n given the background r_b and a hypothetical signal r, to the probability to observe n given the background r_b alone. This ratio tends to unity in the region where the experiment has no sensitivity (where the signal r would be too weak) and zero in the region where the signal is excluded (where the expected signal would be too large to be compatible with the observations). For a Poisson process with background, we have

$$\mathcal{R}(r; n, r_b) = e^{-r\mathcal{L}} \left(1 + \frac{r}{r_b}\right)^n. \tag{4}$$

By introducing a *prior* $f_\circ(r)$, this ratio can be related to a probability about r. In particular, for the case of a constant *prior* $f_\circ(r) = constant$ with a null observation $n = 0$, the credible interval for a 90% confidence level limit can be retrieved by putting $\mathcal{R} = 0.1$.

To take into account the systematic errors on the number of events expected, the likelihoods $f(n|r, r_b)$ can be convoluted with the probability distribution of the number of events expected given the oscillation parameters and the systematic error. In the present case, we assumed the 17% systematic error presented by the CHORUS Collaboration.

3. PRESENTATION OF INDIVIDUAL SAMPLES

Neutrino oscillations are described by two parameters: a mixing angle θ and the squared mass difference Δm^2 between the neutrino mass states. The oscillation probability is given by

$$P_{\nu_\mu \to \nu_\tau} \simeq \sin^2 2\theta \sin^2(1.27\Delta m^2 L/E_\nu), \tag{5}$$

where L is the flight length of the neutrinos and E_ν their energy. Therefore, the expected rate of events will depend on these two variables ($r = r[\sin^2 2\theta, \Delta m^2]$), and so will the ratio $\mathcal{R}(r[\sin^2 2\theta, \Delta m^2]; n, r_b)$.

According to recent data, the CHORUS experiment would expect to observe a maximum of $N_{\max}^{\tau \to \mu} = 4003$ events in its $\tau \to \mu$ sample assuming complete conversion of the ν_μ neutrinos from the beam into ν_τ neutrinos ($P_{\nu_\mu \to \nu_\tau} = 1$). This expected number of events will vary with the oscillation parameters according to equation 5. It will also be further modified by the change in detection efficiency as a function of the energy, so as a function of Δm^2 which modifies the energy distribution of the oscillated neutrinos. In the present exercise, we assumed a constant detection efficiency as a function of energy [1]. For the $\tau \to \mu$ sample, CHORUS expects an average background of $n_b^{\tau \to \mu} = 0.1$ event. Taking into account the dependence of the expected number of events on the oscillation parameters, the energy spectrum of the neutrinos and the flight length of the neutrinos, one obtains values of $\mathcal{R}(r[\sin^2 2\theta, \Delta m^2]; n, r_b)$ for the full oscillation parameter space. Figure 1 shows a 3-D representation of $\mathcal{R}$ as a function of $\sin^2 2\theta$ and Δm^2. The value of $\mathcal{R}$ varies from unity in the region where the CHORUS experiment is insensitive, to zero in the region it excludes. The gradient of colour indicates the change from the excluded region to the region of insensitivity. The region in-between is the one for which CHORUS has difficulties concluding about the existence of neutrino oscillations.

---

[1] The information about the variation of the detection efficiency in CHORUS as a function of the energy is not available publicly at present.
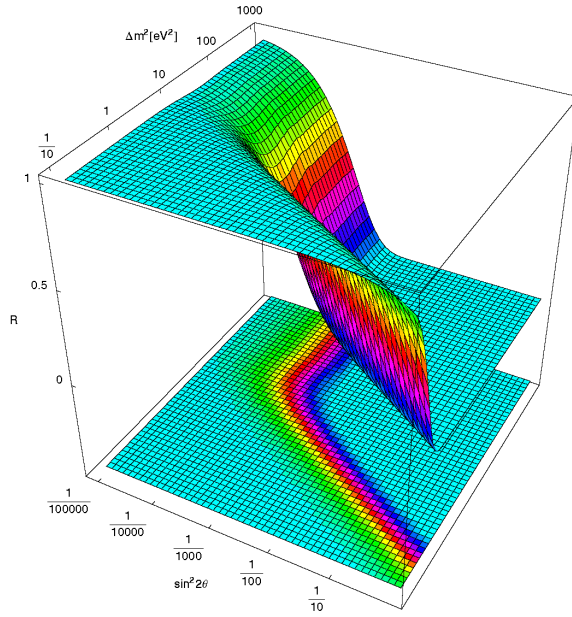
Fig. 1: Tri-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the CHORUS $\tau \to \mu$ sample.

Figure 2 presents the same kind of information for the $\tau \to h$ sample. The region where CHORUS excludes neutrino oscillations is in this case smaller than with the $\tau \to \mu$ sample, which reflects the fact that the $\tau \to h$ sample is smaller and has more expected background. CHORUS expects to observe a maximum of $N_{\max}^{\tau \to h} = 1149$ events in this sample for $P_{\nu_\mu \to \nu_\tau} = 1$, with an average expected background of $n_b^{\tau \to h} = 0.5$ event.
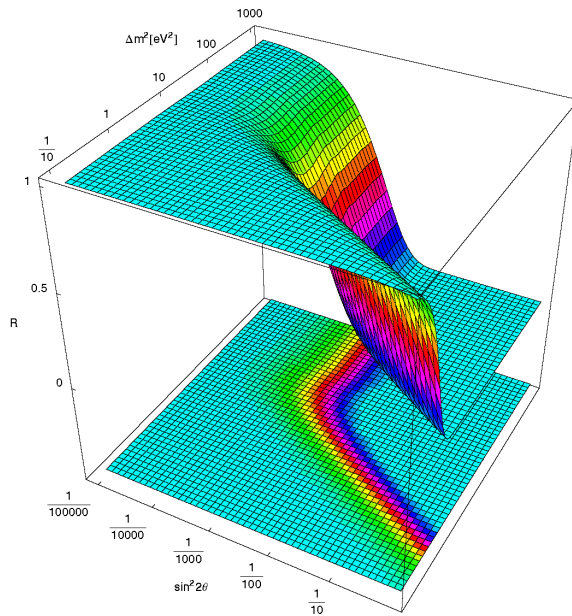

Fig. 2: Tri-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the CHORUS $\tau \to h$ sample.

4. PRESENTATION OF COMBINED SAMPLES

The $\mathcal{R}$ function for the combination of several samples is given by the multiplication of the $\mathcal{R}$ functions of the different samples:

$$\mathcal{R}(r; \text{N samples}) = \Pi_i^N \mathcal{R}(r; \text{sample } i). \qquad (6)$$

Figures 3 and 4 show the kind of plots presented in the preceding section, for the combination of the $\tau \to \mu$ and $\tau \to h$ samples of CHORUS. Since the $\tau \to h$ sample is statistically less significant, the overall result is very similar to the one of the $\tau \to \mu$ sample.

Having presented the recent CHORUS results in a Bayesian way, we can turn to the question of what would happen in the case of an observation different from zero for one of the samples. For instance, let us consider the hypothetical case of an observed number of events in the $\tau \to h$ sample of $n_{\tau \to h} = 3$. In this particular case, without taking into account the energy of the tau candidates, Fig. 2 would look like Fig. 5. We now see a rise above unity of $\mathcal{R}$ for certain values of the oscillation parameters, corresponding to the region where the observation of $n_{\tau \to h} = 3$ is more probable in the case of neutrino oscillations than in the case of the absence of neutrino oscillations. The actual interpretation of this rise of $\mathcal{R}$ in terms of neutrino oscillations will depend on our knowledge of the problem, so on the *prior*. In this particular example, further information can be obtained by combining the $\tau \to h$ sample with the $\tau \to \mu$ sample. For the case where a $\tau \to \mu$ sample with $n_{\tau \to \mu} = 0$ and a $\tau \to h$ sample with $n_{\tau \to h} = 3$ would be combined, Fig. 6 would be obtained. We clearly see that the observed rise in the $\tau \to h$ sample is attenuated by the null result of the $\tau \to \mu$ sample, which is more sensitive to oscillations.

To better appreciate the effect of combining two samples, Figs. 7, 8 and 9 show the value of $\mathcal{R}$ as a function of $\sin^2 2\theta$ for a given value of Δm^2. We arbitrarily chose $\Delta m^2 = 3.6$ eV^2. The transition between exclusion and insensitivity for the $\tau \to \mu$ sample with $n_{\tau \to \mu} = 0$ is clearly seen in Fig. 7, whereas the indication of signal in the $\tau \to h$ sample with $n_{\tau \to h} = 3$ is seen in Fig. 8. Figure 9 shows the attenuation of the evidence obtained as we combine the $\tau \to h$ sample with the $\tau \to \mu$ sample.
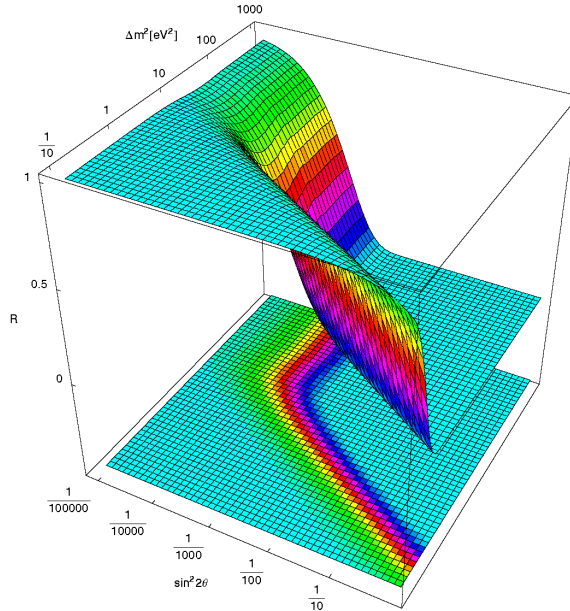

Fig. 3: Tri-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the combined CHORUS sample.
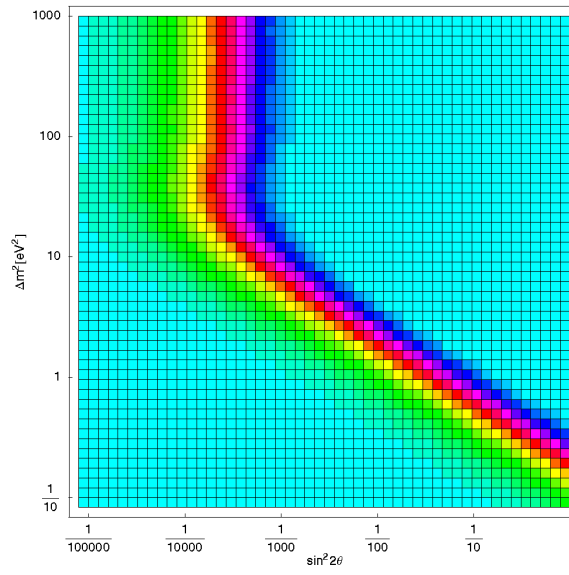
Fig. 4: Bi-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the combined CHORUS sample.
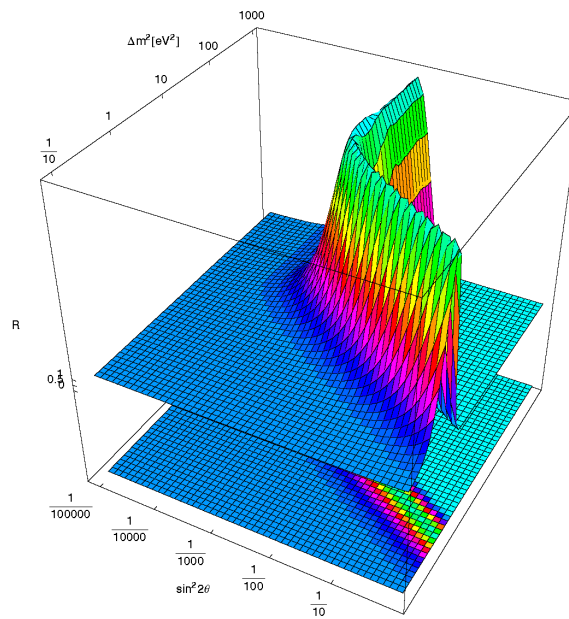

Fig. 5: Tri-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the modified CHORUS $\tau \rightarrow h$ sample with added observed events.
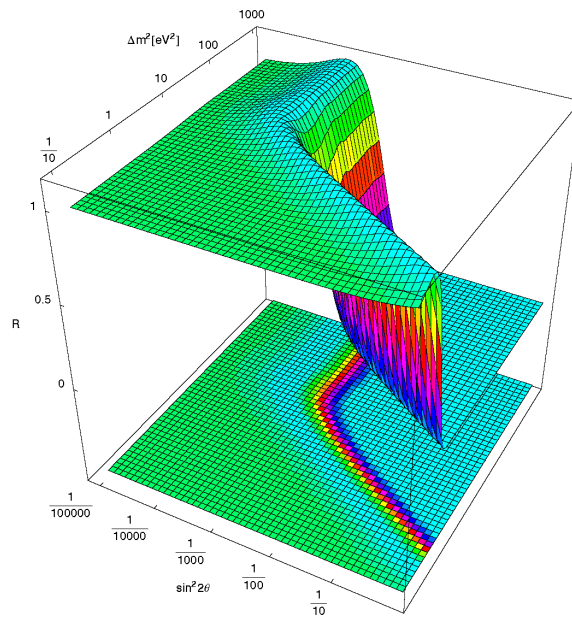
Fig. 6: Tri-dimensional representation of $\mathcal{R}$ as a function of the oscillation parameters for the combined modified CHORUS sample.
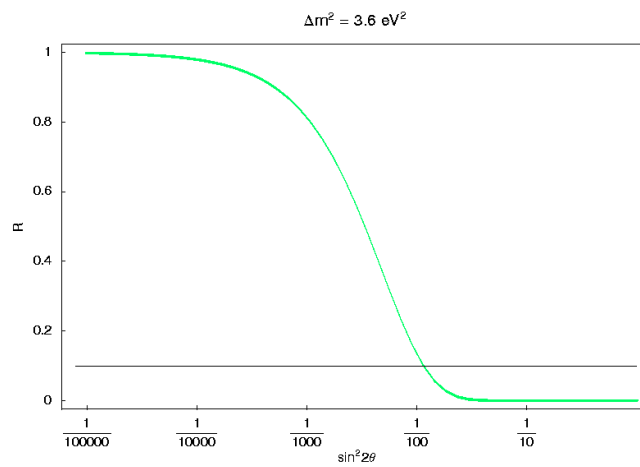

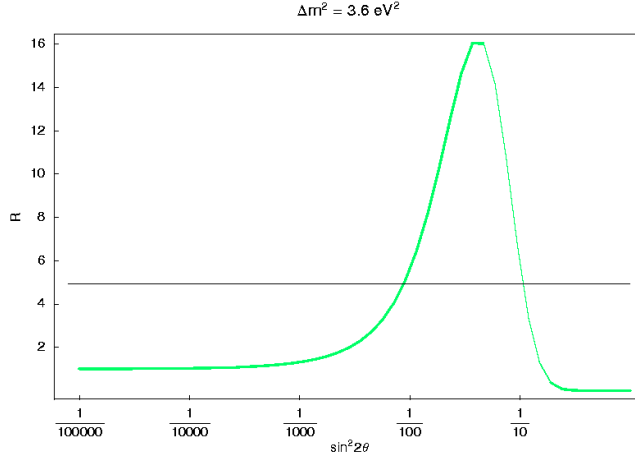Fig. 7: $\mathcal{R}$ as a function of $\sin^2\theta$ for $\Delta m^2 = 3.6\,\mathrm{eV}^2$ for the CHORUS $\tau \to \mu$ sample.

Fig. 8: $\mathcal{R}$ as a function of $\sin^2\theta$ for $\Delta m^2 = 3.6$ eV^2 for the modified CHORUS $\tau \to h$ sample.
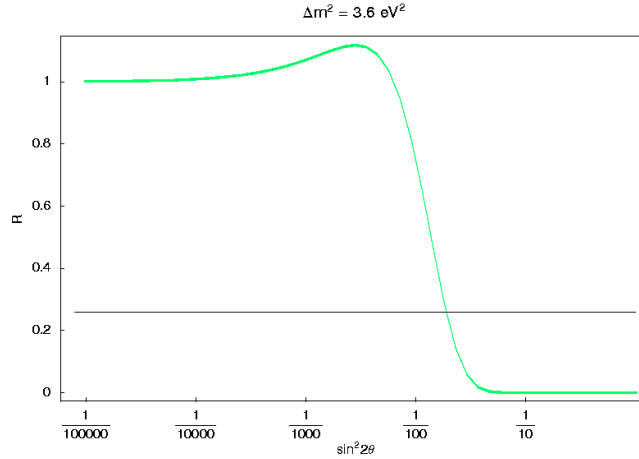

Fig. 9: $\mathcal{R}$ as a function of $\sin^2\theta$ for $\Delta m^2 = 3.6$ eV^2 for the combined modified CHORUS sample.

5. BAYESIAN CREDIBLE INTERVALS

Given the observations of a single experiment, the probability distribution of the true rate of events r of the process we are searching for is given by equation 2. For a uniform *prior*, this equation becomes

$$f(r|n, r_b, f_\circ = constant) = \frac{e^{-r\mathcal{L}}\left((r + r_b)\mathcal{L}\right)^n}{\int_0^\infty e^{-r\mathcal{L}}\left((r + r_b)\mathcal{L}\right)^n \, dr}. \tag{7}$$

In the case of $n = 0$, this equation is simplified to

$$f(r|n = 0, r_b, f_\circ = constant) = \mathcal{L}e^{-r\mathcal{L}} = \mathcal{L}\mathcal{R}, \tag{8}$$

so that credible intervals are easily recovered in terms of $\mathcal{R}$. For instance, in Fig. 7, the values of $\sin^2 2\theta$ are excluded at 90% confidence level between unity and the value crossing the horizontal line at $\mathcal{R} = 0.1$.

In general, one must compute the value of r_{CL} for which the integral of $f(r|n, r_b, f_\circ)$ between zero and r_{CL} gives the desired confidence level. The corresponding value of $\mathcal{R}$ at $r = r_{\mathrm{CL}}$ can then be

193

obtained from equation 4. As an example, the 90% confidence level exclusion in the case of Fig. 8 is the region between unity and the point on the right where $\mathcal{R}$ crosses the horizontal line at $\mathcal{R} = 4.9$.

The relation between the $\mathcal{R}$ distribution and the credible intervals is slightly more involved when several experiments are combined. For two experiments, equation 7 now becomes

$$f(r|n_1, n_2, r_{b1}, b_{b2}, f_\circ = constant) = \frac{e^{-r\mathcal{L}_1} \left((r + r_{b1})\mathcal{L}_1\right)^{n_1} e^{-r\mathcal{L}_2} \left((r + r_{b2})\mathcal{L}_2\right)^{n_2}}{\int_0^\infty e^{-r\mathcal{L}_1} \left((r + r_{b1})\mathcal{L}_1\right)^{n_1} e^{-r\mathcal{L}_2} \left((r + r_{b2})\mathcal{L}_2\right)^{n_2} \, \mathrm{d}r}, \quad (9)$$

where the indices correspond to the two experiments. In this case, we define the probability f of the true rate r of the process, which is common to both experiments. Each experiment nonetheless expects a different number of events for a particular value of r, given by $r\mathcal{L}_i$. The factorization of the number of expected events into a rate and a luminosity is arbitrary up to a constant factor. It the present case, we can for example choose to define the luminosity relative to the number of events in the $\tau \to \mu$ channel, so that $\mathcal{L}_{\tau \to \mu} = 1$ and $\mathcal{L}_{\tau \to h} = N_{\max}^{\tau \to h}/N_{\max}^{\tau \to \mu} = 1149/4003$. The background rates should then be scaled accordingly: $r_b^{\tau \to \mu} = n_b^{\tau \to \mu} = 0.1$ and $r_b^{\tau \to \mu} = n_b^{\tau \to h} N_{\max}^{\tau \to h}/N_{\max}^{\tau \to \mu} = 1.7$. The rate r is then defined as $r = N_{\max}^{\tau \to \mu} P_{\nu_\mu \to \nu_\tau}$.

Integrating equation 9 on r, one can calculate credible intervals and in turn the corresponding limit values of $\mathcal{R}$. Figure 10 shows the resulting 90% confidence level exclusion contour for the case of CHORUS. The value of $\mathcal{R}(r_{\mathrm{CL}})$ in this case is 0.10. The exclusion contour of Fig. 10 is comparable to the combined exclusion contour shown by the CHORUS Collaboration [4].
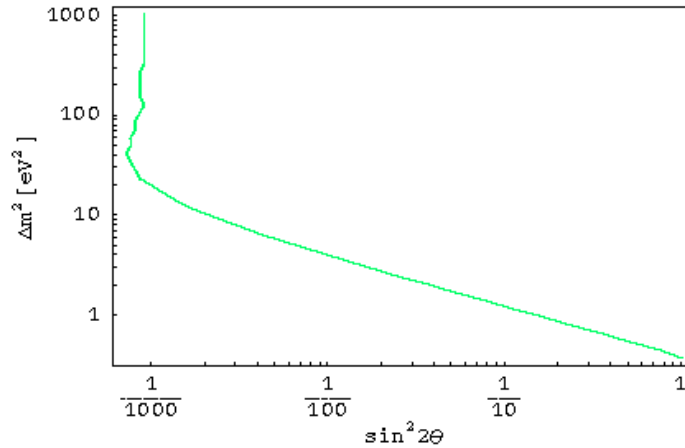

Fig. 10: 90% confidence level exclusion contour for the CHORUS data.

6. CONCLUSIONS

We have tried the Bayesian approach advocated in Reference [1] to present the neutrino oscillation results of CHORUS. The results of two different samples from the experiment were presented. These two samples were combined, both for the actual CHORUS results and for modified results having observed events. The relation between the presentation of the ratio of likelihoods ($\mathcal{R}$) and the credible intervals was discussed for a uniform *prior*. Combining different samples, at least in the present case, is an easy task. For a *prior*-less presentation of the results, the ratio of likelihoods need only to be multiplied. No additional information is required.

References

[1] P. Astone and G. D'Agostini, CERN-EP/99-126, submitted to *Annals of Physics*.

[2] G. D'Agostini, Yellow Report CERN 99-03, and references therein.

[3] P. Righini, "Recent results from the CHORUS experiment", Lake Louise Winter Institute 1999.

[4] E. Eskut *et al.*, *Nucl. Inst. and Meth.* **A401**, 7 (1997);
E. Eskut *et al.*, *Phys. Lett.* **B424**, 202 (1998);
E. Eskut *et al.*, *Phys. Lett.* **B434**, 205 (1998).

Discussion after talk of Mathieu Doucet. Chairman: Peter Igo-Kemenes.

H. Prosper

Just a comment. I very much liked these arguments which are very intuitive, but I should note that it's only priorless if in fact you know everything exactly, you know the background and so on, but if you've other parameters that are not known very well then you have to basically integrate over those unknown parameters and then the ratio becomes dependent on the prior.

M. Doucet

That is the case if you have systematic errors for example. This is what I've actually done.

H. Prosper

...but still I think the presentation of $\mathcal{R}$ is rather useful.

J. Linnemann

Seeing the two talks together I'm still a little bit puzzled. Do the two prescriptions really suggest a different normalization for the likelihood function? That might be a problem if we want to publish likelihood functions. The previous speaker talked about the difficulty in combining the two experiments.

M. Doucet

The normalization is a little different from what was done by Eitel. Here, for each sample, I normalized to the probability of seeing what you see assuming no oscillation signal. I don't see any problem in using different normalisations.

G. D'Agostini

As you said, the overall normalization is not relevant. It's only if you rescale to 1 that you get this function which has intuitive interpetation which we explained in our paper, we even give it a name, now I don't go into detail. What is important is the use we make of the function. Mathieu has used it to evaluate some confidence regions - I would rather call them credibility probability intervals - assuming some priors. What I now prefer, for example, it is not to give these probability intervals anymore, just sensitivity bound, and from this plot you see what is the sensitivity bound; you have a wall. You say: There I don't know, here I am and I've seen nothing, and here is the wall, so we just need to report the position of the wall. There is no problem of prior dependence or of interpretation.

W. Murray

Just continuing that discussion, I think there is a problem when the wall has some thickness. When in your plot the wall is rather narrow, it doesn't really matter what you do, you only get a band that is rather the same. When we went through this problem in the Higgs working group we did not use the Bayesian integral, but rather the classical confidence level construction, because it moves the wall slightly further down and left, and excludes a larger part of the area than in the frequentist definition. It's a small effect but why be conservative?

M. Doucet

Actually, in the present case, the wall is not that narrow. It has a width of about an order of magnitude. This figure has a logarithmic scale.

G. D'Agostini

You have an infinite order of magnitude to your left, so it's very narrow. [Laughter] Anyhow, there is no problem to have this function somewhere in a web page, parametrized with wavelets, as you like. Then just for the purpose of saying to your friends what roughly has been seen, we can report the result with a single number. But the complete result is that.