PoS(EPS-HEP2015)085

Statistical issues in future neutrino oscillation experiments

Alessandra Tonazzo[*]

APC, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu, Observatoire de Paris, Sorbonne Paris Cité, Paris 75205, France

E-mail: tonazzo@in2p3.fr

The neutrino community has been debating on how to assess the potential of future oscillation experiments for determining the neutrino Mass Hierarchy and for establishing CP violation in the leptonic sector. A review of some basic concepts and of the approach chosen by different projects to present their results will be shown. The key issues relevant for the future will be discussed.

---

[*]Speaker.

1. Introduction

Present and future neutrino oscillation experiments will address key questions that are still open, such as the ordering of the three known neutrino mass eigenstates (Mass Hierarchy, MH) and the existence of CP violation (CPV) in the leptonic sector. Assessing a "discovery" on these items requires a proper treatment of the frequentist test of hypothesis, which will be presented in section 2. The case of discovery of one given MH will be discussed in section 3, while the application to CPV will be the subject of section 4. A Bayesian approach to the MH problem will be presented in 3.1.

This review focuses on the statistical issues relevant for MH and CPV discovery, not on the physics potential of the experiments. The physics results shown below or in the relevant references might thus not be the most recent ones. The reader is invited to refer to other talks presented at this conference or to more recent papers for updates on the physics reach of the experiments mentioned below.

2. Assessing a discovery: hypothesis testing

A "discovery" can be assessed when some "null hypothesis" H_0 can be excluded at a given Confidence Level (*CL*). Examples of null hypotheses are statements such as "MH is inverted" or "there is no CPV".

The frequentist method for simple hypotheses proceeds as follows:

1. define a "test statistic" T, which is a function of your data;

2. construct the probability density function (*PDF*) of T under the null hypothesis H_0, $PDF(T|H_0)$;

3. define a "critical region" Ω_C such that values of T in Ω_C suggest H_0 to be true and compute the probability for T to fall in that region if H_0 is true:

$$CL = \int_{-\infty}^{T_C} PDF(T|H_0) \, dT$$

4. exclude H_0 at the given *CL* if your experiment has provided $T > T_C$ [1].

It is common practice to use the terminology proper to two-sided Gaussian integrals even if one-sided integrals are actually involved, thus for instance "3σ CL" should be interpreted as $\int_{-\infty}^{T_C} PFD(T|H_0) \, dT = 99.73\%$.

In order to assess the *CL* correctly, it is obviously crucial to know $PDF(T|H_0)$, either on sound theoretical basis on from MonteCarlo simulations.

2.1 Hypothesis testing in future experiments

When dealing with a future experiment, the value of T that will result from the data is still unknown and its possible statistical fluctuations must be considered.

---

[1]The complementary integral and the region $T < T_C$ may obviously be considered, if appropriate.

The possible experimental outcomes if the "alternative hypothesis" H_1 is true will be distributed according to $PDF(T|H_1)$. The probability content of this distribution should be taken into account, or alternatively some assumption should be made on what the actual result of the experiment will be.

Two approaches are usual:

1) for a given *CL*, consider the possible fluctuations of the experimental outcome and quote the probability to do "as least as good", i.e. the probability that $T > T_C$. This is the "power" p of the hypothesis test:

$$p = \int_{T_C}^{+\infty} PDF(T|H_1) \, dT$$

2) quote the *CL* with a given, fixed outcome of the experiment. Common choices are

   2.a) the "typical" or "Asimov" experiment, i.e. the one maximizing $PDF(T|H_1)$;

   2.b) the "median" experiment, corresponding to $p = 0.5$. It coincides with the "typical" experiment if $PDF(T|H_1)$ is a symmetric function around this value.

With the second approach, only if the assumption on the outcome of a future experiment is specified can its discovery potential be quoted meaningfully. In either case, $PDF(T|H_1)$ or at least some of its properties must be known.

3. Assessing discovery of a Mass Hierarchy

For assessing the MH, a χ^2 function is built using some experimental observables (energy, missing transverse momentum, etc.) and the test statistic is built from the difference between the values obtained when each of the two hierarchies is assumed:

$$T = \Delta\chi^2 = \chi^2_{\text{IH}} - \chi^2_{\text{NH}}.$$

The *PDF*s of T with both NH and IH are shown [1, 2] to be, in most cases, well approximated by Gaussian distributions with a standard deviation equal to twice the absolute value of the mean (T_0):

$$PDF(T|\text{NH}) = N(|T_0^{\text{NH}}|, 2\sqrt{|T_0^{\text{NH}}|}) \quad ; \quad PDF(T|\text{IH}) = N(-|T_0^{\text{IH}}|, 2\sqrt{|T_0^{\text{IH}}|}).$$

Thus, the "median" or "typical" sensitivity , in terms of "number of sigmas", is given by that same number times $\sqrt{T_0}$ and the power p for a given confidence level can be computed with simple analytical formulae [3].

The test of mass hierarchy is in general not a case of "simple hypotheses", in that the values of T depends on parameters [3] such as the mixing angles, the value of the CP violating phase δ_{CP}, etc. For future long-baseline neutrino beam experiments, such as DUNE or LBNO, the most important dependence is on δ_{CP}, while for reactor experiments such as JUNO it is the θ_{23} mixing angle. For such so-called "composite hypotheses", one has to marginalize over the parameters to compute the critical value T_C to assess the *CL*. The possible range of (median) sensitivity or of power for the different values of the parameters must then be shown. Examples are presented in Figure 1.
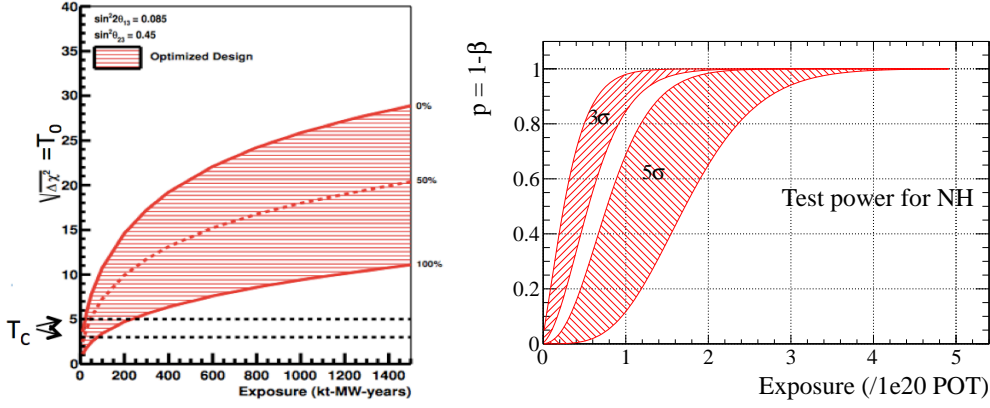
Figure 1: Left (from [4]): Median sensitivity to NH as a function of exposure in the DUNE experiment. The band covers the variation of the δ_{CP} phase. Right (from [5]): Power for the discovery of NH, at two *CL*s, as a function of exposure in the LBNO experiment. The shaded bands correspond to the variation of δ_{CP}.

An additional complication arises from the fact that, in general, $|T_0^{\text{NH}}| \neq |T_0^{\text{IH}}|$ and two separate sets of information must be provided, one for NH discovery (i.e., IH exclusion) and one for IH discovery (NH exclusion). All these issues were beautifully clarified in [3].

Finally, systematic uncertainties play an important role in determining the sensitivity to MH, as shown for instance in [6]. They must be correctly accounted for in the *PDF*s.

3.1 The Bayesian approach to MH determination

In the frequentist approach to the MH test described so far, two sets of information must be provided: *CL* and power for NH and for IH. With a Bayesian approach [7], a single set of values can contain all the information on the test.

In the Bayesian approach, one needs to define a "prior" *P* on each hypothesis, which effectively provides a relative normalization of the two *PDF*s. It can be prove that the most conservative assumption is to consider the two hypotheses to be equally likely, i.e. $P(\text{NH}) = P(\text{IH}) = 0.5$.

One then computes the "odds" or ratio of posterior probabilities: the probability that nature is actually NH (or IH), for a given experimental outcome. This is a question that it makes no sense to ask in the frequentist approach.

Finally, one can define a "*p*-value" for the test, representing the overall probability to obtain a value of *T* such that one of the hypotheses has a posterior probability larger than a given value.

This approach has been chosen, for example, in [8] for the study of future large underwater or under-ice detectors.

4. Assessing discovery of CP violation

For the case of CP violation in the leptonic sector, the null hypothesis H_0 is the absence of CPV ($\delta_{CP} = 0$ or π) and the alternative hypothesis is any other value of the CPV phase. The test statistic is defined as

$$\Delta\chi^2 = min(\Delta\chi^2(\delta_{CP} = 0), \Delta\chi^2(\delta_{CP} = \pi))$$

One can refer again to [5] or [4] for the exact definition of the χ^2 in the LBNO or the DUNE experiments respectively. This test of CPV is a case of nested hypotheses with one free parameter, δ_{CP}. According to Wilks' theorem [9], the test statistic follows a χ^2 distribution with 1 degree of freedom, and this can be confirmed by toy MonteCarlo simulations, as shown in Figure 2 (left) for the LBNO experiment. The distribution of $\Delta\chi^2$ in the absence of CPV is in general independent of the exposure, unlike for the MH case. The critical values to assess CPV discovery with a given *CL* can thus be computed easily as $n\sigma = \sqrt{\Delta\chi^2}$, as shown in Figure 2 (right).
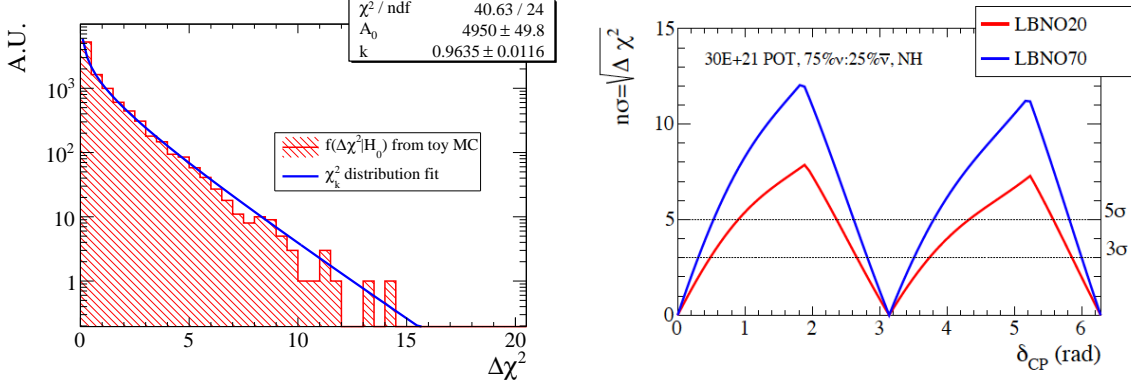

Figure 2: Left (from [5]): Distribution of the test statistic for CPV in the LBNO experiment, under the hypothesis of no CPV. Right (from [10]): Median sensitivity to CPV of the LBNO experiment, for two different exposures, compared with the critical values to assess 3σ and 5σ discovery.

However, it has been shown in [11] that Wilks' theorem does not always hold when the experiment has low sensitivity to δ_{CP} or low statistics. $PDF(\Delta\chi^2|H_0)$ is not distributed according to a know analytical function and (toy-)MonteCarlo simulations are needed to obtain its correct distributions and thus compute the *CL* correctly.

The deviations of $PDF(\Delta\chi^2|H_0)$ from $\chi^2(1)$ also affect existing results, as shown in [12]. In particular, more stringent limits on δ_{CP} are obtained from T2K and MINOS data.

5. Conclusions

There is some degree of arbitrariness in how to quote the sensitivity of neutrino oscillation experiments to key parameters, such as Mass Hierarchy and CP Violation. It is crucial to use the correct probability distributions of the test statistic and to specify the assumptions on the outcome of future experiments. If a Bayesian approach is chosen, it should be explicitly stated.

6. Acknowledgments

References

[1] X. Qian, A. Tan, W. Wang, J. J. Ling, R. D. McKeown and C. Zhang, Phys. Rev. D **86** (2012) 113011 [arXiv:1210.3651 [hep-ph]].

[2] E. Ciuffoli, J. Evslin and X. Zhang, JHEP **1401** (2014) 095 [arXiv:1305.5150 [hep-ph]].

[3] M. Blennow, P. Coloma, P. Huber and T. Schwetz, JHEP **1403** (2014) 028 [arXiv:1311.1822 [hep-ph]].

[4] DUNE Conceptual Design Report - Refresh 1, http://www.dunescience.org

[5] S. K. Agarwalla *et al.* [LAGUNA-LBNO Collaboration], JHEP **1405** (2014) 094 [arXiv:1312.6520 [hep-ph]].

[6] F. Capozzi, E. Lisi and A. Marrone, Phys. Rev. D **91** (2015) 073011 [arXiv:1503.01999 [hep-ph]].

[7] M. Blennow, JHEP **1401** (2014) 139 [arXiv:1311.3183 [hep-ph]].

[8] D. Franco, C. Jollet, A. Kouchner, V. Kulikovskiy, A. Meregaglia, S. Perasso, T. Pradier and A. Tonazzo *et al.*, JHEP **1304** (2013) 008 [arXiv:1301.4332 [hep-ex]].

[9] S. S. Wilks, The Annals of Mathematical Statistics, **9**, 1938.

[10] S. K. Agarwalla *et al.* [LAGUNA-LBNO Collaboration], arXiv:1412.0593 [hep-ph].

[11] M. Blennow, P. Coloma and E. Fernandez-Martinez, JHEP **1503** (2015) 005 [arXiv:1407.3274 [hep-ph]].

[12] J. Elevant and T. Schwetz, JHEP **1509** (2015) 016 [arXiv:1506.07685 [hep-ph]].