Statistical issues in long baseline neutrino physics

Alessandra Tonazzo* for the LBNO Collaboration

*APC, Univ Paris Diderot, CNRS/IN2P3, CEA/Irfu, Obs de Paris, Sorbonne Paris Cité, France

E-mail: tonazzo@in2p3.fr

Abstract. An animated debate has been ongoing in the neutrino community on how to estimate and quote the expected sensitivity of future long-baseline neutrino experiments to key parameters such as Mass Hierarchy or CP violation. We will present an overview of some items covered by recent papers and will detail the approach chosen by the LBNO Collaboration to present its results.

1. Introduction

Future neutrino experiments using long-baseline beams and very large detectors 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. In particular, the LBNO Collaboration [1] has proposed the installation of a liquid argon (LAr) TPC [2] in an extension of the Pyhäsalmi mine in Finland, which will be able to cover important issues in particle physics (proton decay) and astrophysics (neutrinos from supernovae and from the sun). A neutrino beam from CERN, located at a distance of 2300 km, would allow assessment of the MH and exploration of CPV. The experiment envisages a staged approach with two phases: a first phase with 20 kt (fiducial volume) of LAr and a 750 kW beam based on the existing SPS accelerator complex, providing about $1.2 \times 10^{20}$ p.o.t. per year, and a second phase with a total LAr fiducial mass of 70 kt and a 2 MW beam from a newly-built high-power accelerator (HPPS), providing an intensity 20 times higher.

Since the assessment of CPV depends on the knowledge of the MH, the LBNO strategy is to ensure that MH is known “with high confidence level” in the first few years of running. How to assess this confidence level has recently been at the heart of animated debates.

Papers published in the past couple of years have raised questions about the assumptions on the analytical form of the distribution functions to be used for the computation of the confidence level [3, 4], on how to present the information on sensitivities in a complete and unambiguous way [5] and on the difference between a Bayesian and a frequentist approach [6]. After a short review of some of these topics, we will present the approach chosen by the LBNO collaboration to present its results [7] following the guidelines of [5].

2. Hypothesis test: some basics

The aim of a test of hypothesis is to test a “null hypothesis” $H_0$ against an “alternative hypothesis” $H_1$, using a given set of data. The hypotheses are “simple” if they do not contain unspecified parameters, otherwise they are “composite”.

For simple hypotheses, the frequentist method 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 each hypothesis;
3. define a “critical region” \( \Omega_C \) such that values of \( T \) in \( \Omega_C \) suggest \( H_0 \) to be true: \( H_0 \) will be accepted if \( T \) falls in \( \Omega_C \), otherwise \( H_1 \) will be accepted;
4. evaluate the probability to give the wrong answer, which you can do in two ways:
   \begin{itemize}
   \item reject \( H_0 \) when it is true: this is the error “of type I” or “loss”, usually denoted as \( \alpha \);
   \item accept \( H_0 \) when \( H_1 \) is true: this is the error “of type II” or “contamination”, \( \beta \).
\end{itemize}
This is illustrated, for a simple example, in Figure 1.

\[ PFD(T|H_0) \]

\[ PFD(T|H_1) \]

\[ T_C \]

\[ T^\sigma \]

\[ -30, -20, -10, 0, 10, 20, 30 \]

\[ 0, 0.05, 0.1, 0.07 \]

\( \sigma \)

\( CL \)

Figure 1. Illustration of the basic ingredients of a hypothesis test in a simple case: PDF of the test statistic \( T \) under the two hypotheses, definition of the critical value \( T_C \) (the critical region is \( T < T_C \)), \( \alpha \) and \( \beta \) errors.

The “Confidence Level” (CL) for rejecting the null hypothesis is defined as \( CL = 1 - \alpha \) and the “Power” to test \( H_1 \) against \( H_0 \) is \( p = 1 - \beta \). For simple 1-dimensional cases such as the one depicted in Figure 1, the critical region is defined by a threshold \( T_C \) on the value of \( T \) such that \( H_0 \) is accepted when \( T < T_C \) and rejected when \( T > T_C \); the CL and the power for rejection \( H_0 \) are given by

\[
CL = 1 - \alpha = \int_{-\infty}^{T_C} PDF(T|H_0) \, dT \\
p = 1 - \beta = \int_{T_C}^{+\infty} PDF(T|H_1) \, dT
\]

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 “3\( \sigma \) CL” should be interpreted as \( \int_{-\infty}^{T_C} PDF(T|H_0) \, dT = 99.73\% \).

Only if the two values of \( \alpha \) and \( \beta \), or equivalently \( CL \) and Power, are quoted is the hypothesis test fully quantified.

Two approaches are frequently used:

(i) set \( T_C \) such that \( p = 0.5 \). The corresponding \( CL \) is often referred to as “median sensitivity”.
(ii) set \( T_C \) such that \( \alpha = \beta \), or \( CL = p \). This is called “crossing sensitivity” in [5].

They will be discussed below for the case of the MH.

When the hypotheses are composite, the \( CL \) and the power depend on some parameters, whose values are in general unknown.
3. Mass Hierarchy as a hypothesis test

The case of mass hierarchy determination can be treated as a test of hypotheses. The two hypotheses are Normal Hierarchy (NH, $\Delta m_{32}^2 > 0$) and Inverted Hierarchy (IH, $\Delta m_{32}^2 < 0$). Both of them are composite, as in general the distributions of test statistic depend on other oscillation parameters (mixing angles, CP violating phase, etc.) or on detector details. This will be taken into account in the following discussion.

The test statistic $T$ can be constructed thanks to the Neyman-Pearson lemma, which implies that, for simple hypotheses, the most powerful test for any given $CL$ is the likelihood ratio. This can be extended to the case of composite hypotheses. For the MH problem, the likelihood ratio is easily proven to be equivalent to the difference of the $\chi^2$'s between data and expectations, where the $\chi^2$ is computed on a single- or multi-dimensional distribution of kinematic variables of the data (such as energy, transverse momentum, etc.), possibly including systematic effects as nuisance parameters. The test statistic $T$ is thus defined as

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

The exact definition of the $\chi^2$ used for the LBNO experiment is provided in [7].

3.1. Test statistic distributions and confidence levels

If $H_0$ and $H_1$ are nested hypotheses, i.e., if $H_1$ depends on $n$ parameters and $H_0$ can be obtained from $H_1$ by fixing $m$ of these parameters, Wilks’ theorem [8] states that the PDF of $T$ when $H_0$ is true follows a $\chi^2$ distribution with $n-m$ degrees of freedom. In the past, the $CL$ for the MH problem was often computed assuming that the PDF for a given hierarchy was distributed as a $\chi^2$ (1 d.o.f.).

However, as pointed out in [3, 4], the MH problem is clearly NOT a case of nested hypotheses! The correct distributions for the PDF of the test statistic $T$ under the two hypotheses must be evaluated. It can be demonstrated from simple arguments [3, 4] that the two distributions are well approximated by Gaussians:

$$PDF(T|NH) = N(|T_{0}^{NH}|, 2\sqrt{|T_{0}^{NH}|}) ; \quad PDF(T|IH) = N(-|T_{0}^{IH}|, 2\sqrt{|T_{0}^{IH}|}).$$

This approximation is verified at the percent level by generating the distribution with a toy MonteCarlo simulation, as shown for the LBNO experiment in Figure 2.

![Figure 2. Distribution of the test statistic for the MH test in the LBNO experiment (left) and ratio between width and mean of the distributions as a function of exposure (right).](image-url)
In the simple approximation $|T_0^{NH}| = |T_0^{IH}|$, which holds for many experiments, it is easy to compute the significance of the test in terms of $T_0$ with the two prescriptions for $CL$ and power quoted in the previous section. Let us consider, to fix ideas, $H_0=IH$. Then,

- to have $p = 0.5$, one should set $T_C = T_0^{NH}$ and one gets $CL = N\sigma$ with $N = \sqrt{|T_0^{IH}|}$;
- to have $p = CL$, one should set $T_C = 0$ and one gets $CL = N\sigma$ with $N = \sqrt{|T_0^{IH}|}/2$.

The same conclusion holds with $H_0=NH$.

What we wish to stress here is the factor 2 that appears between the two cases. This has led to a lot of misunderstanding, which is however fully resolved if the power of the test is quoted along with the $CL$. We note in passing that the factor 2 would also appear if the “crossing sensitivity” was computed with a $PDF$ having $\sigma = \sqrt{T_0}$, which is however completely wrong for the MH case!

3.2. Construction of a hypothesis test for MH

In order to construct properly a hypothesis test for MH, the LBNO Collaboration has chosen the approach described in [5]. To fix ideas, suppose that the IH is taken as $H_0$. Define the desired $CL$ for its exclusion and set the critical value $T_C^{IH}$ such that

$$\int_{-\infty}^{T_C^{IH}} PFD(T|IH) \ dt = CL,$$

then compute the corresponding power as

$$p = \int_{T_C^{IH}}^{+\infty} PFD(T|NH) \ dt.$$

This can be done for a given value of $T_0^{IH}, T_0^{NH}$ or for all possible values of $T_0^{IH}, T_0^{NH}$: typically $|T_0|$ increases with exposure of the experiment, so the power can also be provided as a function of exposure or time. The procedure should then be repeated with NH as $H_0$, which in general will yield a different critical value $T_C^{NH}$.

Up to now we have neglected the fact that the hypotheses are composite. The distributions of $T$ depend on additional parameters, related to neutrino mixing or to detector performance. In particular, for the LBNO experiment, they depend on the value of the $CP$ violating phase $\delta_{CP}$, as shown in figure 3. This parameter will not be known at the time of the experiment.

As pointed out in [5], the desired $CL$ must be ensured for all values of the (unknown) parameters, and in general the power will depend on the parameters. For the MH case, one should thus set $T_C$ in the most pessimistic case, i.e. for the smallest absolute value of $T_0$ in either case. We thus compute

$$T_C^{IH} \text{ for } T_0^{IH}(\delta_{CP} = 3\pi/2)$$
$$T_C^{NH} \text{ for } T_0^{NH}(\delta_{CP} = \pi/2).$$

The power can then be evaluated as a function of exposure for all the possible values of $\delta_{CP}$, as in the plots shown on Figure 4.

From these plots one can easily read the exposure needed to achieve the given $CL$ whatever value of $\delta_{CP}$ (lower end of the bands) and with a given power (for example, 0.5 or 1). It is common practice to quote the exposure that provides a 50% chance to achieve the desired $CL$, however this may be risky when implying the decision to invest time and money in a future experiment. LBNO has chosen to quote its sensitivity with the exposure ensuring a probability close to 100% to achieve the desired $CL$, i.e. for power close to 1.

Either choice is correct, as long as the power is quoted along with the $CL$. LBNO has $\sim 100\%$ probability (power $\sim 1$) to achieve MH discrimination at $5\sigma$ $CL$, for either ordering, after 4-5 years of its first phase.
Figure 3. Test statistic values for MH as a function of $\delta_{CP}$, under the hypothesis of IH (left) and of NH (right). The exposure corresponds to about 2 years of running of LBNO 1st phase.

Figure 4. Statistical power as a function of exposure for the test of NH (left) and IH (right) for $3\sigma$ and $5\sigma$ CL. The shaded bands correspond to the variation of $\delta_{CP}$.

4. 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 [6], 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 normalisation 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. $P(NH) = P(IH) = 0.5$, and we will use this in the following.

One then sets a threshold $t$ and computes two critical values

\[ T_{NH}^C \quad \text{such that} \quad \frac{PDF(T|NH)}{PDF(T|NH) + PDF(T|IH)} > t \quad \text{for} \quad T > T_{NH}^C \]

\[ T_{IH}^C \quad \text{such that} \quad \frac{PDF(T|IH)}{PDF(T|NH) + PDF(T|IH)} > t \quad \text{for} \quad T < T_{IH}^C \]

Essentially, $t$ represents the “odds” or ratio of posterior probabilities: if an experiment gives a results $T$, the probability that nature is actually NH (or IH) is larger than $t$ in the two critical regions. 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 as
\[
p = \int_{-\infty}^{T_C} PFD(T|IH) \, dT + \int_{T_C}^{+\infty} PFD(T|NH) \, dT
\]
representing the overall probability to obtain a value of \( T \) such that one of the hypotheses has a posterior probability of at least \( t \).

This approach has been taken, for example, in [9] for the study of future large underwater or under-ice detectors.

5. CPV as a test of hypothesis

For the case of CP violation in the leptonic sector, the null hypothesis \( H_0 \) is the absence of CPV (\( \delta_{CP} = 0 \) or \( \pi \)) 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 [7] for the exact definition of the \( \chi^2 \) in the case of the LBNO experiment. The distribution of the test statistic is strongly dependent on the assumptions on oscillation parameters and experimental errors, however the main dependence is by far the one on the value of \( \delta_{CP} \) itself.

This test of CPV is thus a case of nested hypotheses with one free parameter, \( \delta_{CP} \). According to Wilks’ theorem [8], the test statistic follows a \( \chi^2 \) distribution with 1 degree of freedom, and this is confirmed by toy Monte Carlo simulation in the case of the LBNO experiment. An example is shown in Figure 5 (left): this distribution is independent of the exposure, unlike for the MH case.

The PDF of the test statistic under the alternative hypothesis is not easy to predict, it can be estimated with toy Monte Carlos for each value of \( \delta_{CP} \) and of the exposure. An example is shown in Figure 5 (right). A simple parametrisation with a skewed Gaussian, depending on the exposure, is found to be a good approximation for this distribution. The power of the test can be computed, for each value of \( \delta_{CP} \), by integrating numerically or analytically the distribution. The LBNO collaboration has chosen, for the first time, to present the potential for CPV discovery in a similar way to the MH case, i.e. quoting the power of the test as a function of exposure.

![Figure 5](image_url)

*Figure 5.* Distribution of the test statistic for the CPV test in the LBNO experiment, under the hypothesis of no CPV (left) and with \( \delta_{CP} = \pi/2 \) (right), with an exposure of \( 15 \times 10^{30} \) p.o.t.

Since the value of the test statistic is strongly dependent on \( \delta_{CP} \), as shown in Figure 6 (left) for the LBNO experiment, the power of the test will depend strongly on this parameter. The
power of the LBNO experiment to exclude null CPV at two different CL is shown in Figure 6 (right) as a function of the exposure, for the two extreme cases $\delta_{CP} = \pi/2$ and $3\pi/2$, and it can be as low as zero for other values of the phase.

Figure 6. LBNO experiment. Left: average value of the test statistic for CPV test as a function of the $\delta_{CP}$ phase. Right: Power for exclusion of the non-existence of CPV as a function of exposure, for two CLs and for the two extreme values of $\delta_{CP}$.

6. Conclusions
A review of some recent debate on how to evaluate and quote the sensitivity of future long-baseline neutrino experiment has been provided. The key conclusion is that, when dealing with problems of hypothesis testing, both CL and power of the test should be explicitly quoted.

The LBNO Collaboration has chosen to quote its sensitivity to Mass Hierarchy with a statistical power close to 1. A $5\sigma$ CL can be achieved by the experiment in at most 4-5 years of running, for either ordering.

A proposal to quote sensitivity to CP violation in a similar way has been shown.

A Bayesian approach to the Mass Hierarchy test has also been presented.

7. Acknowledgments
We are grateful to the European Commission for the financial support of the project through the FP7 Design Study LAGUNA-LBNO (Project Number 284518). We acknowledge the financial support of the UnivEarthS LabEx programme at Sorbonne Paris Cité (ANR-10-LABX-0023 and ANR-11-IDEX-0005-02).

References
[1] A. Stahl et al., Expression of Interest for a "Very long baseline neutrino oscillation experiment (LBNO)."; submitted to CERN SPSC, June, 2012. (CERN-SPSC-2012-021 (SPSC-EOI-007)).
[2] A. Rubbia, “Experiments for CP violation: A Giant liquid argon scintillation, Cerenkov and charge imaging experiment?”, hep-ph/0402110; A. Rubbia et al., J. Phys. Conf. Ser. 171 (2009) 012020.
[3] 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]].
[4] E. Ciuffoli, J. Evslin and X. Zhang, JHEP 1401 (2014) 095 [arXiv:1305.5150 [hep-ph]].
[5] M. Blennow, P. Coloma, P. Huber and T. Schwetz, JHEP 1403 (2014) 028 [arXiv:1311.1822 [hep-ph]].
[6] M. Blennow, JHEP 1401 (2014) 139 [arXiv:1311.3183 [hep-ph]].
[7] S.K. Agarwalla et al. [LAGUNA-LBNO Collaboration], “The mass-hierarchy and CP-violation discovery reach of the LBNO long-baseline neutrino experiment,” arXiv:1312.6520 [hep-ph].
[8] S. S. Wilks, The Annals of Mathematical Statistics, 9, 1938.
[9] 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]].