New Statistical Analyses for the Neutrino Mass Hierarchy Determination: Quantify the Sensitivity to the Neutrino Mass Hierarchy

Emilio Ciuffoli - Institute of Modern Physics, CAS

8 September 2017

Astro@Stat2017

Outline

- Neutrino Oscillations
- Statistical Issue
- Frequentist Approach
- Bayesian Approach
- Summary

3 Flavor Oscillations

The mass matrix is not diagonal in the interaction eigenbase $\Rightarrow$ The neutrinos are created in a superpositions of three different energy eigenstate.

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \end{pmatrix} = U \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \end{pmatrix} \quad \Rightarrow \quad |\nu_a\rangle = \sum_\alpha U_{a\alpha}|\nu_\alpha\rangle$$

$$a = e, \mu, \tau, \; \alpha = 1, 2, 3$$

$$|\nu_a(t)\rangle = \sum_\alpha U_{a\alpha}|\nu_\alpha(t)\rangle = \sum_\alpha U_{a\alpha}e^{-iE_\alpha t}|\nu_\alpha\rangle = \sum_\alpha U_{a\alpha}e^{-iE_\alpha t}U^\dagger_{\alpha a'}|\nu_{a'}\rangle$$

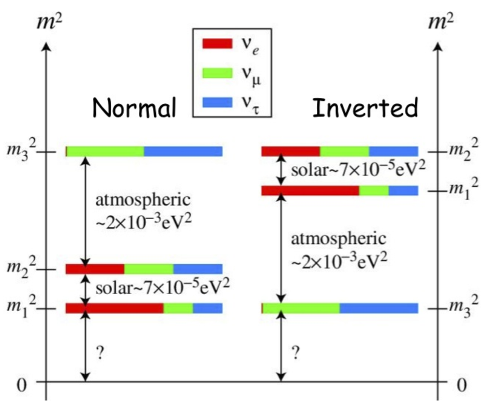
$$P_{ee} = |\langle\nu_e(0)||\nu_e(t)\rangle|^2 = |c_1 e^{-iE_1 t} + c_2 e^{-iE_2 t} + c_3 e^{-iE_3 t}|^2$$

The neutrino mass hierarchy

There are three light, mostly-active neutrino mass eigenstates called ν_1, ν_2 and ν_3 with masses are m_1, m_2 and m_3. Define

$$\Delta m^2_{ij} = m^2_i - m^2_j$$

- Vacuum oscillations of ultrarelativistic $\nu \Rightarrow |\Delta m^2_{ij}|$
  $|\Delta m^2_{ij}|$ now measured with good precision
- Matter effect are sensitive to the sign of Δm^2.
  From solar neutrinos $\Rightarrow \Delta m^2_{21} > 0$



The neutrino mass hierarchy (MH) is Sign(Δm^2_{31}).

The hierarchies are two disjoint hypotheses

One hierarchy is true (realized in nature), the other is false

MH from Reactor Neutrinos

Since in vacuum oscillations we can observe only the absolute values of Δm^2's, we have

$$|\Delta m^2_{31}| = |\Delta m^2_{32}| \pm |\Delta m^2_{21}| = |\Delta m^2_{32}|(1 \pm 0.03)$$

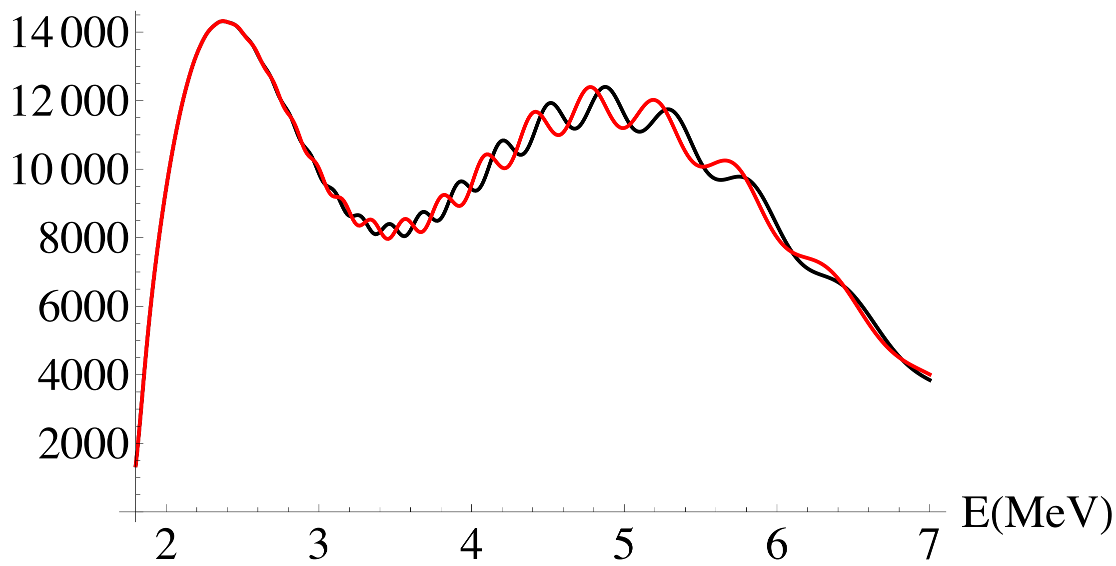
Reactor neutrino experiments (like JUNO, RENO 50) will measure the MH by studying the survival probability $P(\bar{\nu}_e \to \bar{\nu}_e)$.
(S. T. Petcov and M. Piai, Phys. Lett. B 2002)

$$P_{ee} = 1 - 0.81 \sin^2\left(\frac{1.27\Delta m^2_{21} L}{E}\right) - 0.06 \sin^2\left(\frac{1.27\Delta m^2_{13} L}{E}\right) - 0.03 \sin^2\left(\frac{1.27\Delta m^2_{23} L}{E}\right)$$

The beating between the 1-3 and 2-3 oscillations determines $\text{sign}(|\Delta m^2_{32}| - |\Delta m^2_{31}|)$ and so the hierarchy (positive $\leftrightarrow$ IH)

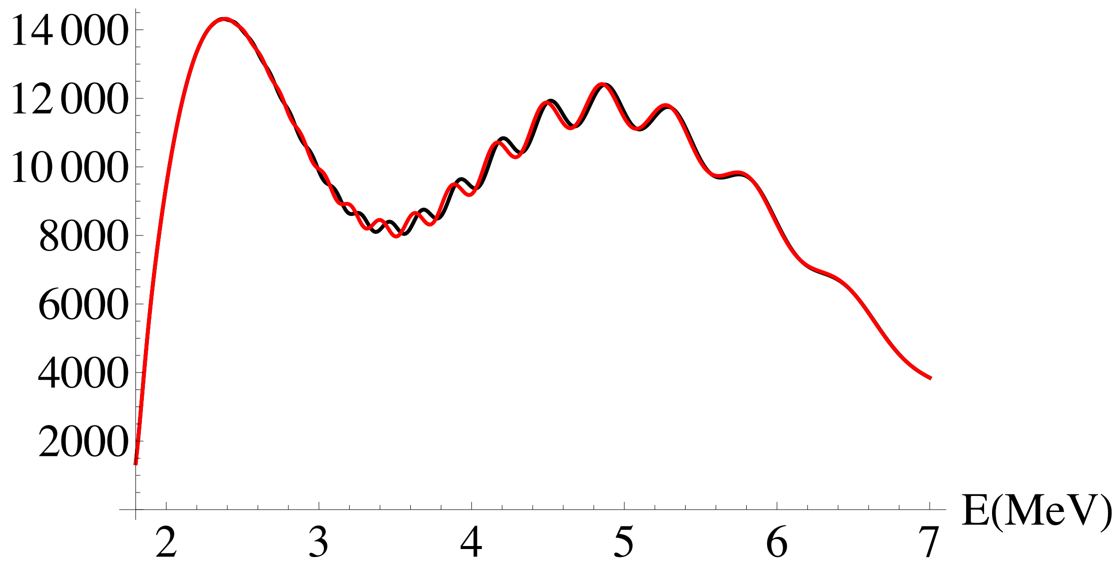
However, degeneracy between a change of hierarchy and a shift of Δm^2_{32} $\Rightarrow$ NOT simple vs. simple case

MH Determination in Reactor Neutrino Experiments



Expected spectra for normal and inverted hierarchy at 58km, using the best fit values of Δm^2_{32} for NH and IH (from PDG)

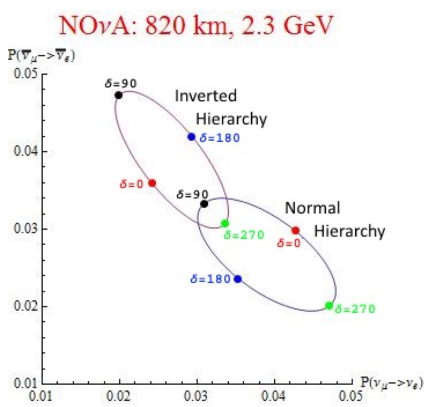
MH Determination in Reactor Neutrino Experiments



Expected spectra for normal and inverted hierarchy at 58km. Inverted hierarchy: Δm^2_{23} shifted (by $\simeq 0.7\sigma$'s).

MH from Accelerator Neutrinos

Accelerator neutrino experiments (like NOνA, T2K, etc...) can measure the MH by comparing the oscillation probability in the neutrino and antineutrino modes



- Vacuum oscillations: total degeneracy between a change of hierarchy and $\delta_{CP} \to \pi - \delta_{CP}$
- Degeneracy (partially) broken by the matter effect, that depends on $\mathrm{Sign}(\Delta m^2_{32})$. However still a residual degeneracy between (NH,$\delta_{CP} \simeq 90^\circ$) and (IH,$\delta_{CP} \simeq 270^\circ$)

Test Statistic

For the mass hierarchy, we define the test statistic

$$\Delta\chi^2 = \chi^2_{\mathrm{IH}} - \chi^2_{\mathrm{NH}}$$

Where $\chi^2_{\mathrm{NH/IH}}$ are the χ^2 values of the data to NH/IH:

1) Pull parameters minimized for *each* hierarchy separately
2) A penalty term in χ^2 for each pull parameter is considered

NOT the only possible choice!!
Alternative test statistics available: see, for example, Luca Stanco's talk

IMPORTANT

Note that $\Delta\chi^2$ is *not* the quantity in Wilks' theorem, because the last term is not necessarily the best fit:

It is the difference between two *disjoint* hypotheses, not two *nested* hypotheses

Additional Parameter

A possible way to avoid this problem consists in introducing an additional pull parameter, without any physical meaning that will reduce the problem to parameter fitting.

- For the MH, this was first suggested for reactor neutrino experiments **(Capozzi, Lisi and Marrone PRD 2014)**, writing

$$|\Delta m^2_{31}| = |\Delta m^2_{32}| + (2\eta - 1)|\Delta m^2_{21}|$$

$\eta = 1 \rightarrow$ Normal Hierarchy; $\eta = 0 \rightarrow$ Inverted Hierarchy

- A more general approach, that can be applied also to accelerator neutrinos is described in **S. Algeri, J. Conrad and D.A. van Dyk; MNRAS: Letters, 2016**: Let $f(E,\theta)$ and $g(E,\theta)$ be the expected spectra for the normal and inverted hierarchy, then one can consider

$$\eta f(E,\theta) + (1-\eta) g(E,\theta)$$

Additional Parameter

It is possible now to define **two** $\Delta\chi^2$'s, one for hierarchy. Calling $\hat{\eta}$ the best-fit value for η, we have

$$\Delta\chi^2_{NH} = \chi^2(1) - \chi^2(\hat{\eta}) \qquad \Delta\chi^2_{IH} = \chi^2(0) - \chi^2(\hat{\eta})$$

Both follow a one-degree-of-freedom chi-square distribution, but $\Delta\chi^2$ defined before is the **difference** between these two quantities

$$\Delta\chi^2 = \Delta\chi^2_{IH} - \Delta\chi^2_{NH} = \chi^2(0) - \chi^2(1) = \chi^2_{IH} - \chi^2_{NH}$$

We know the distribution for $\Delta\chi^2_{NH}/\Delta\chi^2_{IH}$m but not for $\Delta\chi^2$

- This method gives us a very compact way to express the compatibility of each hierarchy with the data (e.g. $\eta = 0.8 \pm 0.1$)
- On the other side, no physical meaning for $\eta \neq 0, 1$

Distribution of $\Delta\chi^2$

Since χ^2_{NH} is not always the best fit, our test statistic does not follow a one-degree-of-freedom χ^2 distribution
(for example: it is not always > 0)

Under certain assumptions, to a good approximation it follows a Gaussian distribution, with

$$\mu = \pm\overline{\Delta\chi^2} \qquad \sigma = 2\sqrt{\overline{\Delta\chi^2}}$$

Qian et al. PRD 2012; EC, Evslin and Zhang JHEP 2014; Blennow et al. JHEP 2014

Conditions for Gaussianity

n_i expected number of events in each bin for a certain experiment (N=number of bins). In general, they will be function of a certain number P of pull parameters θ_j: $n_i = n_i(\theta_j)$.

Conditions for Gaussianity:

1. $n_i(\theta_j)$ can be approximated as a linear function of θ_j in **the region of interest**. This define a P-dimensional hyperplane in the N-dimenstional space.
   $\Rightarrow \chi^2_{NH/IH}$ is described by a (hyper)-parabola.
2. The hyperplanes for the normal and the inverted hierarchies are parallel around the minima

I will discuss more in detail the statistical distribution of $\Delta\chi^2$ using, as examples, two toy models inspired by reactor and accelerator neutrino experiments

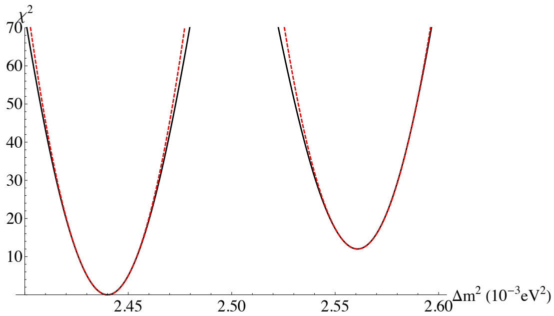
Two Examples

Two examples: MH from reactors and accelerator neutrinos

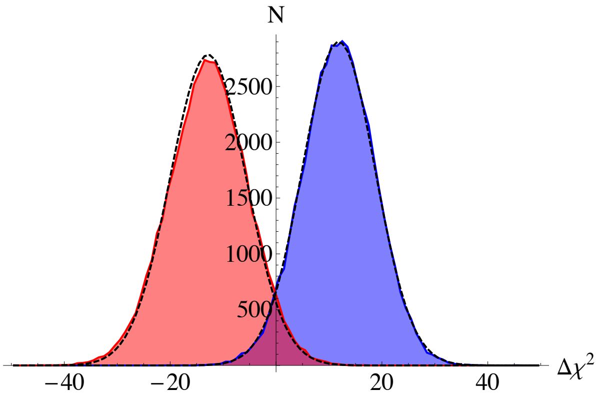
- Very simplified models considered
- Only one pull parameter: Δm^2_{32} and δ_{CP}, respectively
- In case of accelerator neutrinos, no spectral information
- No background or possible systematic errors considered

$\Delta\chi^2$ distribution (Reactor Neutrinos)

In the case of the reactor neutrino experiments, the statistical distribution of $\Delta\chi^2$ can be approximated with excellent precision with a Gaussian distribution

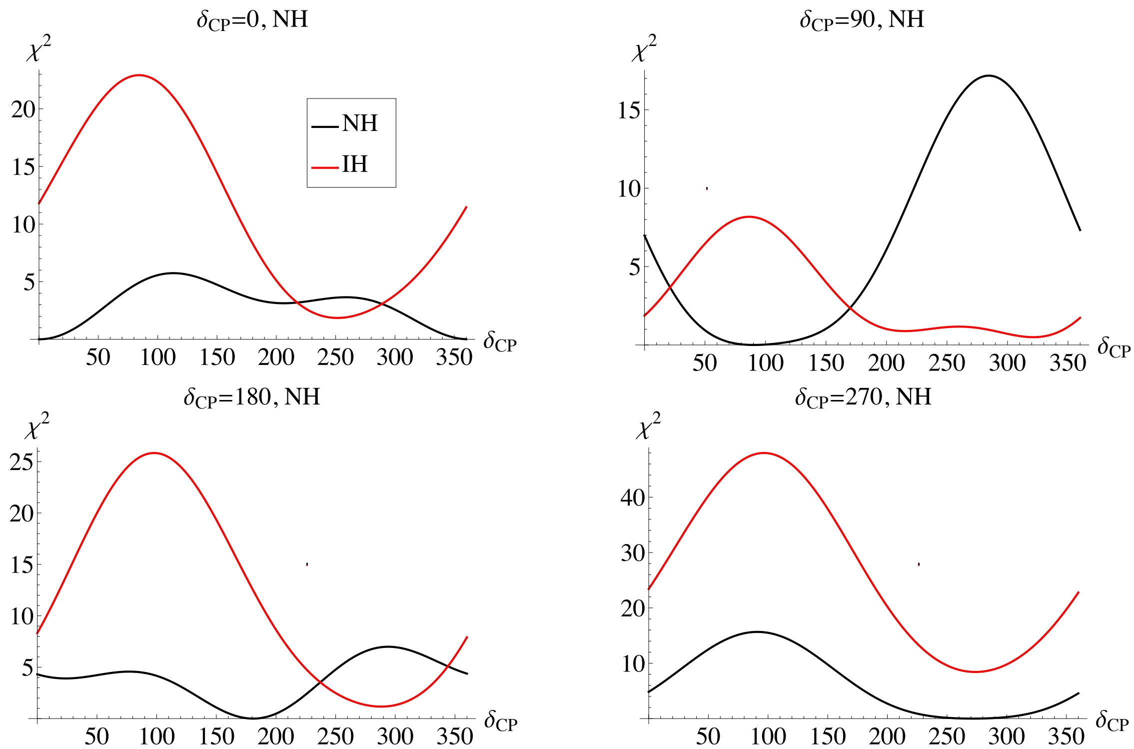


Solid black: Asimov χ^2; dashed red: parabolic fit

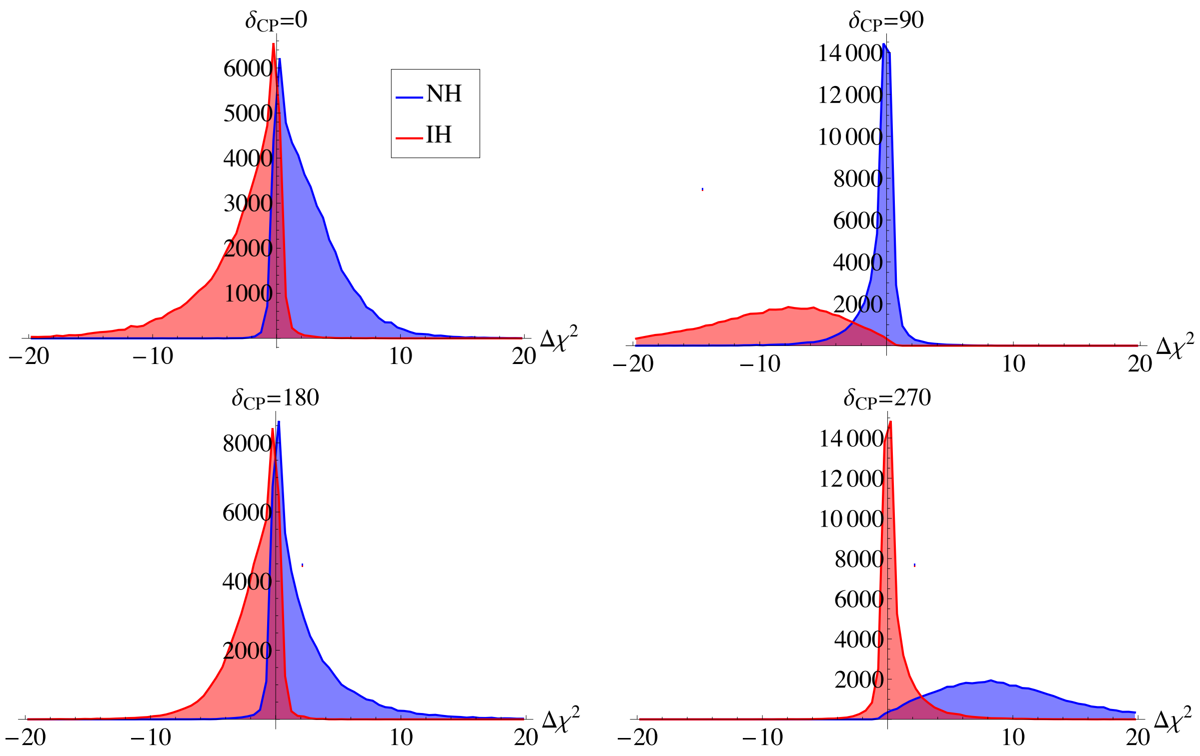


P.d.f. for $\Delta\chi^2$: solid curves: MC results, dashed curves: Gaussian fit

Asimov χ^2 (Accelerator Neutrinos)



$\Delta\chi^2$ distribution (Accelerator Neutrinos)



How to Quantify Sensitivity

Significance Level and Power

Hypothesis test (frequentist test)

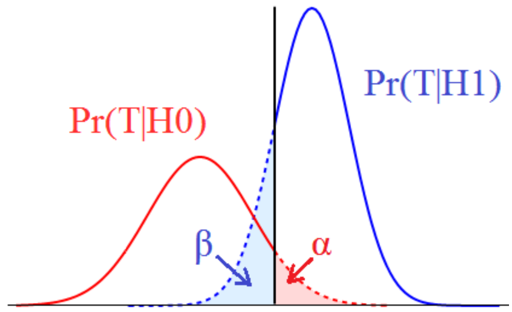
We want to test an hypothesis H_0, with respect to an alternative hypothesis H_1. We define a test statistic T: our experiment will give, as result, T_{obs}. If $T_{obs} \in w$ (ex: $T_{obs} > T_C$), H_0 is rejected, otherwise it's accepted.

- $1-\alpha$="Confidence Level". α=probability of rejecting H_0 if it's true (*type-I error*)

$$\alpha = \Pr(T > T_C | H_0)$$

- $1-\beta$= "Power". β=probability of not rejecting H_0 if H_1 is true (*type-II error*)

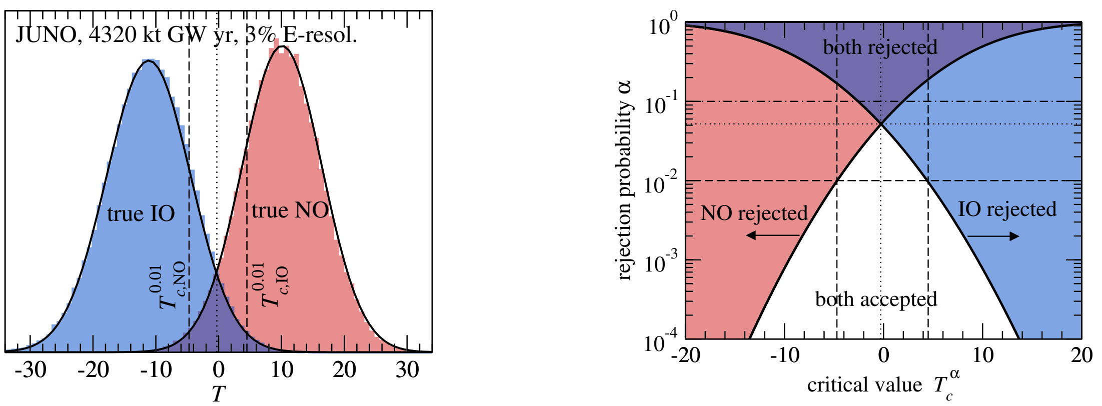
$$\beta = \Pr(T < T_C | H_1)$$



Hypothesis Test for the MH

Frequentist approach to the MH determination

- We test BOTH hierarchies SEPARATELY
- We define two thresholds, $T_{C,NH}$ and $T_{C,IH}$
- If $T_{obs} < T_{C,NH}$ the normal hierarchy is rejected, if $T_{obs} > T_{C,IH}$, the inverted hierarchy is rejected
- It is possible to accept or reject both hierarchies



(Plots from **Blennow et al. JHEP 2014**, see also **Qian et al. PRD 2012**)

Hypothesis Test for the MH

We can express the CL as the number s of σ's (Gaussian standard deviations) using the relation (one-sided CL)

$$\alpha = \frac{1}{2}\mathrm{Erfc}\left(\frac{s}{\sqrt{2}}\right)$$

Hypothesis Test: CL is defined **before** the experiment (the result tells only if it's achieved or not): convenient to define the sensitivity of future experiments. For the Gaussian, symmetric case:

- **Median Sensitivity**: $T_{C,NH(IH)} = -(+)\overline{\Delta\chi^2}$; $s = \sqrt{\overline{\Delta\chi^2}}$; $\beta = 0.5$
- **Crossing Sensitivity**: $T_{C,NH(IH)} = 0$; $s = \sqrt{\overline{\Delta\chi^2}/2}$; $\beta = \alpha$

If the pdf of T depends strongly on pull parameters, problems for defining the CL. Possible solution: a given CL is achieved only if it's achieved for all the values of the pull parameters

Blennow et al. JHEP 2014

P-value

P-value

Probability of finding a "more extreme" value of T than T_{obs}

$$p(\theta) = \Pr(T > T_{obs}|H_0, \theta)$$

- Depends on the "true" value of the pull parameters!
  Possible solutions: $\max_\theta p(\theta)$, $p(\hat{\theta})$ ($\hat{\theta}$ = best-fit value), etc...
- All the methods indicated so far rely on the knowledge of the distribution of T. When it's not known?
  MC simulations are a possible solution (but precise at 5σ's?)
- The frequentist approach cannot give us the probability that the hierarchy is normal or inverted, only the compatibility of EACH hierarchy with the data

$$\text{Ex. 1: NH excluded at } 5\sigma \begin{cases} \text{IH excluded at } 1\sigma \\ \text{IH excluded at } 5\sigma \end{cases}$$

$$\text{Ex. 2:} \quad \text{NH excluded at } 5\sigma \text{ and IH at } 3\sigma \quad \text{vs} \quad \text{NH excluded at } 4\sigma \text{ and IH at } 1\sigma$$

Bayesian Approach

The frequentist approach allows us to determine only $\Pr(\mathbf{D}|MH)$. However in order to "determine the mass hierarchy" we are want to know $\Pr(MH|\mathbf{D})$.

Bayes Theorem

$$\Pr(NH|\mathbf{D}) = \frac{\Pr(\mathbf{D}|NH)\pi(NH)}{\Pr(\mathbf{D}|NH)\pi(NH) + \Pr(\mathbf{D}|IH)\pi(IH)}$$

Bayesian Method (**D. van Dyk, Neutrino 2016**)

- Provide a single quantity $\rightarrow$ choose between H_0 and H_1
- BUT the result depends on priors: $\pi(MH)$!
  (however, natural choice in the case of the hierarchy: $\pi(NH) = \pi(IH) = 0.5$).
- Bayesian and frequenstist approach answer different questions, but they are not exclusive: why not use both?
  Ex: $(NH \rightarrow 5\sigma, IH \rightarrow 3\sigma)$ and $(NH \rightarrow 4\sigma, IH \rightarrow 1\sigma)$ give roughly the same $\Pr(MH|\mathbf{D})$

Bayes Factor

$$\text{Bayes Factor:} \quad K = \frac{\Pr(\mathbf{D}|NH)}{\Pr(\mathbf{D}|IH)} = \frac{\int \Pr(\mathbf{D}|NH,\theta)\pi(\theta)\mathrm{d}\theta}{\int \Pr(\mathbf{D}|IH,\theta)\pi'(\theta)\mathrm{d}\theta} = e^{\Delta\chi^2/2}$$

Marginalization, not Minimization!

$$\Delta\chi^2 = -2\ln(\Pr(\mathbf{D}|IH)/\Pr(\mathbf{D}|NH))$$

$\Pr(\mathbf{D}|MH) \neq \min_\theta \Pr(\mathbf{D}|MH,\theta)$: **marginalization**, not **minimization**

The Bayes factor can be used to determine the posterior probability (**Qian et al. PRD 2012; EC, Evslin and Zhang JHEP 2014; Blennow JHEP 2014)**

$$\mathrm{P}(NH|\mathbf{D}) = \frac{\mathrm{P}_{NH}(\mathbf{D})\pi(NH)}{\mathrm{P}_{NH}(\mathbf{D})\pi(NH) + \mathrm{P}_{IH}(\mathbf{D})\pi(IH)} = \frac{\pi(NH)}{\pi(NH) + \pi(IH)K^{-1}}$$

Does not depend on the statistical distribution of $\Delta\chi^2$

Laplace Method

In the Bayesian approach, the eventual pull parameter must be **marginalized** (*i.e.* integrated over), not **minimized**

$$\Delta\chi^2_B = -2\ln\frac{\Pr(\mathbf{D}|IH)}{\Pr(\mathbf{D}|NH)} \qquad \Pr(\mathbf{D}|MH) = \int \Pr(\mathbf{D}|\theta, MH)\pi(\theta)\mathrm{d}\theta$$

$$\Delta\chi^2_F = -2\ln\frac{\min_\theta \Pr(\mathbf{D}|\theta, IH)}{\min_\theta \Pr(\mathbf{D}|\theta, NH)}$$

If many pull parameters are present, the computation of the multidimensional integrals involved in the marginalization may be very difficult.

Laplace Method (Kass and Raftery, 1995) $\Rightarrow$ If

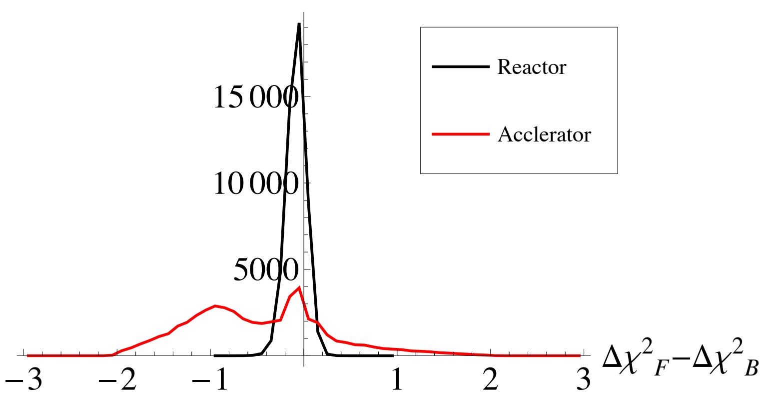
- $I(\theta, MH) = \Pr(\mathbf{D}|\theta, MH)\pi(\theta)$ is highly peaked around its maximum
- the determinants of the Hessian matrices for $I(\theta, NH)$ and $I(\theta, IH)$ calculated in the minima are the same

then $\Delta\chi^2_B = \Delta\chi^2_F$

Laplace Method

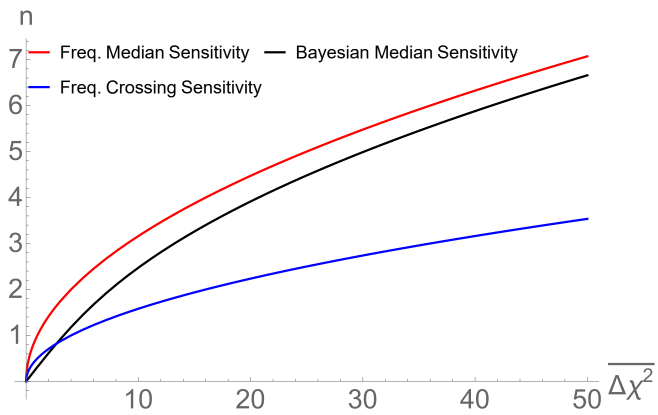
This approximation works with very good precision for reactor neutrinos, but not for accelerator neutrinos.

- Reactors: $\overline{\Delta\chi}^2 \simeq 12$; Accelerators $\overline{\Delta\chi}^2 \simeq 4$
- Valid also with additional pull parameters?
- Many other methods available, as Markov chain Monte Carlo (MCMC), nested sampling algorithms, etc...



In the symmetric, Gaussian case we can define the "median experiment" as the experiment where we have $\Delta\chi^2 = \overline{\Delta\chi^2}$. Using symmetric priors, we can define the "median bayeisian sensitivity" as

$$\Pr(MH|\mathbf{D}) = \frac{1}{1 + e^{-\overline{\Delta\chi^2}}}$$



Red: Median Frequentist Sensitivity, Black: Median Bayesian Sensitivity, Blue: Crossing Sensitivity

(Plot from **EC, Evslin and Zhang JHEP 2014**)

Summary

- Some results on the statistical distribution of $\Delta\chi^2$, however not always valid
- Bayesian approach gives only one quantity: more suitable to choose between H_0 and H_1
- Frequentist approach gives two quantities: both must be reported! Different and complementary information
- Why not use both?
- Different approaches available: there is no "right" or "wrong" choice, but it is important to specify the convention used