Statistical issues for future neutrino experiments

Mattias Blennow
emb@kth.se
KTH Theoretical Physics

July 4, 2014, ICHEP, Valencia, Spain
1. The usual deal
1 The usual deal

2 The alternative
1 The usual deal

2 The alternative

3 Summary and conclusions
1 The usual deal

2 The alternative

3 Summary and conclusions
Parameter estimation sensitivity

- Define the test statistic \( \Delta \chi^2 \)

\[
\Delta \chi^2(\theta) = -2 \log \left[ \frac{\mathcal{L}(\theta|d)}{\sup_{\theta'} \mathcal{L}(\theta'|d)} \right]
\]

- Assume it is \( \chi^2 \) distributed with \( n \) degrees of freedom
- Use the data set without statistical fluctuations (Asimov data)
- Quote result
The usual deal

The alternative

Summary and conclusions

The interpretation

- $\Delta \chi^2$ is asymptotically $\chi^2$ (Wilks’ theorem)
- The Asimov data is representative
- Several requirements, not always fulfilled

Not a nested hypothesis

- Mass ordering is not nested
- Wilks’ theorem not applicable
- Some different choices of test statistic

\[ \Delta \chi^2 = \chi^2_{\text{NO}} - \min \chi^2 \]
\[ T = \chi^2_{\text{IO}} - \chi^2_{\text{NO}} \]
\[ T' = \chi^2(\theta) - \min \chi^2 \]

- All have different distributions
- We concentrate on \( T \)

\[ T \sim \mathcal{N}(T_0, 2\sqrt{T_0}) \]

\( T_0 = \) value for Asimov data

<table>
<thead>
<tr>
<th></th>
<th>The usual deal</th>
<th>The alternative</th>
<th>Summary and conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The usual deal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The alternative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Summary and conclusions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Back to basics

- Test hypothesis NO, alternative hypothesis IO \( \rightarrow \) “Can we reject NO if IO is true?”
- Critical region defined by \( T_c \)
  
  Reject NO if \( T < T_c \)
- Test significance (CL): \( 1 - \alpha = 1 - P(T < T_c | NO) \)
  
  \( \alpha \): Probability to reject NO if NO is true
- Test power: \( 1 - \beta = P(T < T_c | IO) \)
  
  \( \beta \): Probability of failing to reject false NO if IO is true
- Both \( \alpha \) and \( \beta \) are relevant
What is sensitivity?

Sensitivity (median)
What is the expected rejection of a false ordering? (Given a parameter set)

Other
- What is the probability of ruling out a false hypothesis at $x\sigma$?
- At what CL is exactly one ordering ruled out?
Simple hypotheses

- Distribution of test statistic independent of parameter values
- Straight forward
- Applicable to reactor experiments
Composite hypotheses

- Distributions have parameter dependence
- To reject = Rejecting all parameter sets ($\theta_{13}, \Delta m_{31}^2, \delta, \text{etc}$)
- Distribution of $T$ is parameter dependent

\[ T_c = \min_{\theta} T_c(\theta) \]

- Extensive Monte Carlo simulations
- Approximation (must check validity)

\[ T(\theta) = \mathcal{N}(T_0(\theta), 2\sqrt{T_0(\theta)}) \]
Accelerator experiments - results

**NOvA**

Green: 68 % of outcomes
Yellow: 95 % of outcomes

**LBNE – 10kt**

**MB, Coloma, Huber, Schwetz, JHEP 03(2014)028**
Comparison of experiments - specific $\beta$

MB, Coloma, Huber, Schwetz, JHEP 03(2014)028

*Note:* Bands have different meanings!
Comparison of experiments - specific $\alpha$

**Note:** Bands have different meanings!
How to interpret the median sensitivity

- It is *representative* for how well the experiment will do
- 50 % probability of not reaching it
- 50 % probability of *doing better*
- *Not* 50 % probability of “being wrong”
- Not the only relevant quantity, distribution matters (do Brazilian bands!)
- Personal preference: Quote the power $1 - \beta$ for a target sensitivity
Other measurements

**CP violation**
- Nested hypothesis
- Does not mean that Wilk’s theorem holds
- Work in progress

**θ_{23} octant**
- Degeneracies closer
- Wilk’s theorem still violated
- A priori, a dedicated study is needed
1. The usual deal

2. The alternative

3. Summary and conclusions
Summary and conclusions

- Wilks’ theorem is not applicable to the neutrino mass ordering, the test statistic is not $\chi^2$ distributed.
- Regardless, $\sqrt{T_0}$ is still a good approximation of the (median) sensitivity.
- Frequentist (as well as Bayesian – not discussed) methods are perfectly applicable.
- Critical values will depend on the experiment and underlying assumptions.
- More results on other measurements soon.
4 Backup frequentist

5 Backup distributions and results

6 Backup Bayesian basics

7 Bayesian methods
Testing both orderings

- Depending on the significance it may be possible to:
  - Reject exactly one hypothesis
  - Reject both hypotheses
  - Not reject any hypothesis
- It is natural, the true hypotheses should be rejected with probability $\alpha$
At the end of the day

For the simple hypotheses:

- Two Gaussians,
  \[ H_{\pm} : \mathcal{N}(\pm T_0, 2\sqrt{T_0}) \]
- For \( H_+ \), typical (median) result is \(+ T_0\)
- \(+ T_0\) is \( T_0 - (-T_0) = 2T_0 \) away from the expected \( H_- \) result
- \( 2T_0/(2\sqrt{T_0}) = \sqrt{T_0} \)

See also: Vitelis, Read, 1311.4076
Interpolating parameters

- Introduce an interpolating continuous parameter $\alpha$ such that

$$P(\alpha = \pm 1, \theta) = P(\theta, \text{NO/IO})$$

see, e.g., Capozzi, Lisi, Marrone, arXiv:1309.1638

- Wilks’ theorem now applies

- Put bounds on $\alpha$

- If $\alpha = 1 (-1)$ is allowed, NO (IO) is allowed

- Personal comment: $\alpha = 0$ is not special, it being allowed a priori does not affect the rejection of NO/IO
LBNO predictions

Test power for NH

LBNO collaboration, arXiv:1312.6520
4 Backup frequentist

5 Backup distributions and results

6 Backup Bayesian basics

7 Bayesian methods
Reactor experiments

- Distribution of $T$ essentially parameter independent
- Simple hypotheses
- Gaussian limit well satisfied
- Median sensitivity $n\sigma$:

$$n \simeq \sqrt{2} \text{erfc}^{-1} \left[ \frac{1}{2} \text{erfc} \left( \sqrt{\frac{T_0}{2}} \right) \right]$$

JUNO, 4320 kt GW yr, 3% E-resol.

MB, Coloma, Huber, Schwetz, JHEP 03(2014)028
Atmospheric experiments

- Distribution of $T$ mainly depends on $\theta_{23}$
- Gaussianity is still well satisfied for each $\theta_{23}$
- Analytic as function of $T_0(\theta_{23})$
- Boils down to evaluating the Asimov data: $T_0$

MB, Coloma, Huber, Schwetz, JHEP 03(2014)028
Accelerator experiments - distributions

- Not always Gaussian!
- Typical for low-sensitivity experiments
- Need to perform Monte Carlo studies for accuracy
- Rejection power depends on the true parameters

**NOνA**

\[ \delta = 90^\circ \]

true IO

true NO

\( T \)  

\(-20\) \(-10\) \( 0 \) \( 10 \) \( 20 \)

**LBNE**

\[ \delta = 90^\circ \]

true IO

true NO

\( T \)  

\(-50\) \( 0 \) \( 50 \)
4 Backup frequentist

5 Backup distributions and results

6 Backup Bayesian basics

7 Bayesian methods
Bayesian basics

- Assign a degree of belief in each hypothesis $P(H)$
- Update the degree of belief depending on observations
- Bayes’ theorem

$$P(A, B) = P(A; B)P(B) = P(B; A)P(A)$$

$$P(A; B) = \frac{P(B; A)P(A)}{P(B)}$$

- Take $A =$ hypothesis $H$, $B =$ data $d$

$$P(H; d) = \frac{\mathcal{L}_H(d)P(H)}{P(d)}$$
4 Backup frequentist

5 Backup distributions and results

6 Backup Bayesian basics

7 Bayesian methods
Bayesian hypothesis testing

- Study the relative degrees of belief in two hypotheses

\[
\frac{P(H_1; d)}{P(H_2; d)} = \frac{\mathcal{L}_{H_1}(d)}{\mathcal{L}_{H_2}(d)} \frac{P(H_1)}{P(H_2)}
\]

- Strength of the evidence for \( H_1 \):

\[
\kappa = 2 \log \left[ \frac{P(H_1; d)}{P(H_2; d)} \right]
\]

- Kass-Raftery scale:

<table>
<thead>
<tr>
<th>Strength of evidence for ( H )</th>
<th>( \kappa )</th>
<th>Posterior odds</th>
<th>Degree of belief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barely worth mentioning</td>
<td>0 to 2</td>
<td>ca 1 to 3</td>
<td>&lt; 73.11%</td>
</tr>
<tr>
<td>Positive</td>
<td>2 to 6</td>
<td>ca 3 to 20</td>
<td>&gt; 73.11%</td>
</tr>
<tr>
<td>Strong</td>
<td>6 to 10</td>
<td>ca 20 to 150</td>
<td>&gt; 95.26%</td>
</tr>
<tr>
<td>Very strong</td>
<td>&gt; 10</td>
<td>( \geq 150 )</td>
<td>&gt; 99.33%</td>
</tr>
</tbody>
</table>
What can be said about the future?

- Can compute probability of obtaining evidence at least strength $\kappa_0$ for the true ordering

$$P(\kappa_0) = P(\kappa > \kappa_0; H_1)P(H_1) + P(\kappa < -\kappa_0; H_2)P(H_2)$$

- Typical choice $P(H_1) = P(H_2) = 0.5$

- Takes into account information on oscillation parameters

$$\mathcal{L}_H(d) = \int \mathcal{L}_H(\theta)(d)\pi(\theta)d\theta$$

- Compactifies all of the available information to one number

- Easy to simulate through Monte Carlo methods

- Prior dependent
Example: NOνA

- NOνA experiment
- GLoBES implementation
- Only $\delta$ and $\Delta m^2_{31}$ varying
- Flat and 10% Gaussian priors, respectively

MB, JHEP 01(2014)139
Example: NOνA, results

- Probability of obtaining evidence with at least strength $K$ for the correct ordering

MB, JHEP 01(2014)139