# How to assess MH statistical significance in future projects

Mattias Blennow
emb@kth.se

KTH Theoretical Physics

January 8–10, 2014, ICFA Neutrino Panel, Paris

1 The usual deal

2 Frequentist methods

# Parameter estimation sensitivity

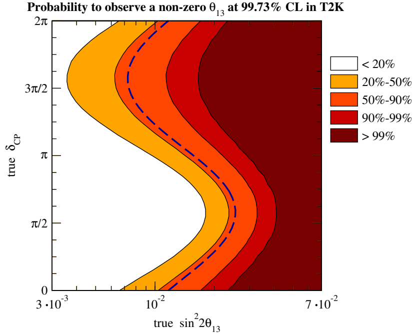
- Define the test statistic "$\Delta\chi^2$"

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

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

# The interpretation

- $\Delta\chi^2$ *is* asymptotically $\chi^2$ (Wilk's theorem)
- The Asimov data is representative
- Works for *nested hypotheses*
- Some additional requirements




Schwetz, Phys.Lett. B648 (2007) 54

# Not a nested hypothesis

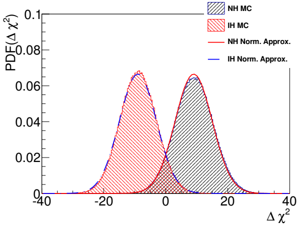
- Mass ordering is not nested
- Wilk's theorem not applicable
- Some different choices of test statistic

$$\begin{aligned} \Delta\chi^2 &= \chi^2_{\rm NO} - \min\chi^2 \\ T &= \chi^2_{\rm IO} - \chi^2_{\rm NO} \\ T' &= \chi^2(\theta) - \min\chi^2 \end{aligned}$$

- All have different distributions
- We concentrate on $T$

$$T \simeq \mathcal{N}(T_0, 2\sqrt{T_0})$$

$T_0$ = value for Asimov data



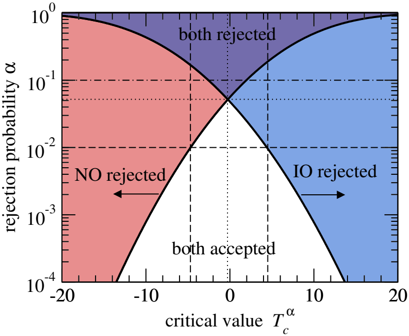

Qian, *et al.*, Phys.Rev. D86 (2012) 113011

# Back to basics

- Hypothesis testing
- Test hypothesis $H_0$, alternative hypothesis $H_1$
- Critical region defined by $T_c$
- Reject $H_0$ if $T < T_c$
- Test significance: $1 - \alpha = 1 - P(T < T_c | H_0)$
- Test power: $1 - \beta = P(T < T_c | H_1)$
- *Both* $\alpha$ and $\beta$ are relevant

# Testing both orderings

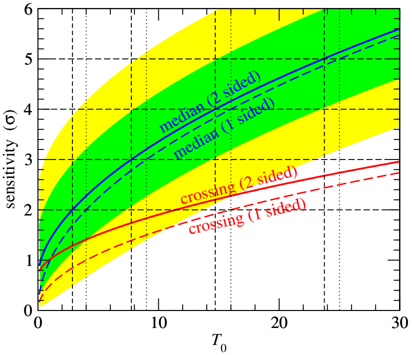
- Depending on the significance it may be possible to:
  - Reject exactly one hypothesis
  - Reject both hypotheses
  - Not reject any hypothesis
- It is natural, the true hypotheses *should* be rejected with probability $\alpha$




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

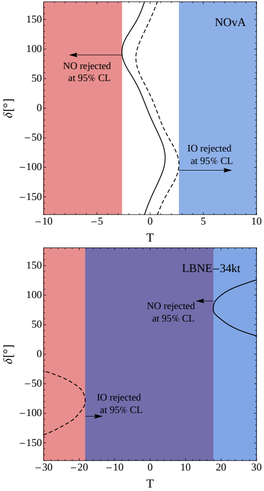
# Simple hypotheses

- Hypotheses are not parameter dependent
- Nyman-Pearson lemma applies
- Straight forward
- Applicable to reactor experiments




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

# Composite hypotheses




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

- Hypotheses that have parameter dependence
- To reject = Rejecting all parameter sets
- Distribution of $T$ is parameter dependent

$$T_c = \min_\theta T_c(\theta)$$

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

$$T(\theta) = \mathcal{N}(T_0(\theta), 2\sqrt{T_0(\theta)})$$

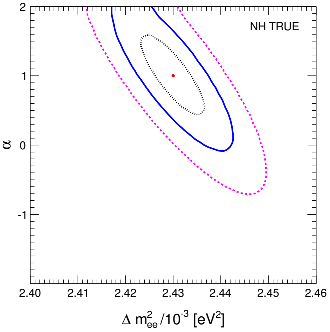
# Interpolating parameters

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

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

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

- Wilk's theorem now applies
- Put bounds on $\alpha$
- If $\alpha = 1$ $(-1)$ is allowed, NO (IO) is allowed



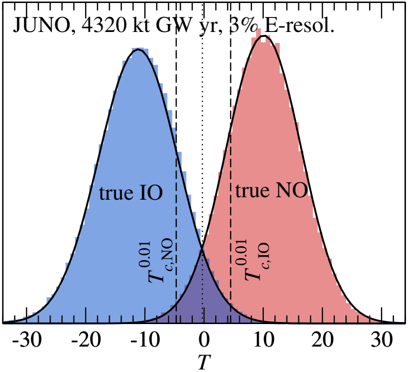

Capozzi, Lisi, Marrone, arXiv:1309.1638

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

# Reactor experiments

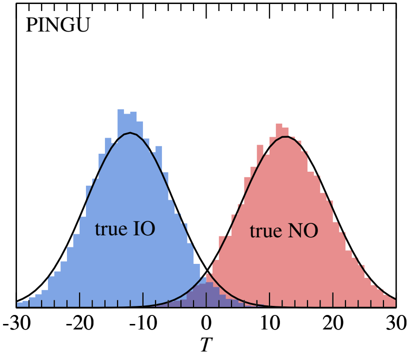
- Distribution of $T$ essentially parameter independent
- Simple hypotheses
- Gaussian limit well satisfied
- Median sensitivity $n\sigma$:

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




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

# Atmospheric experiments



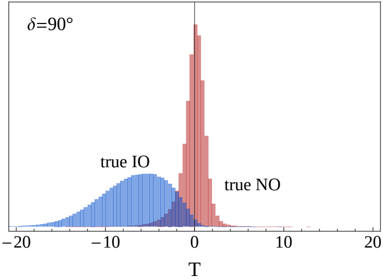

MB, Coloma, Huber, Schwetz, arXiv:1311.1822

- 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: $T_0$

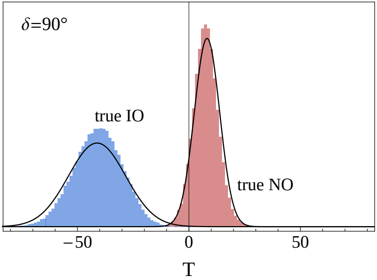
# Accelerator experiments - distributions

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

NO$\nu$A

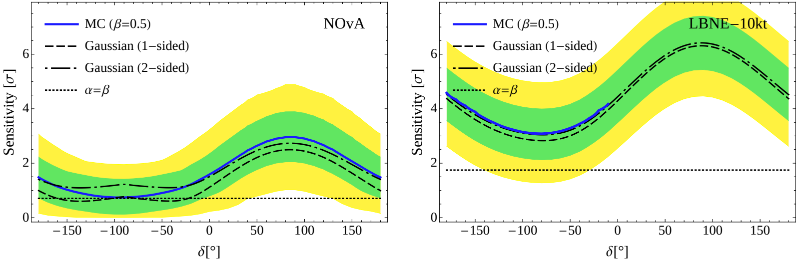



LBNE



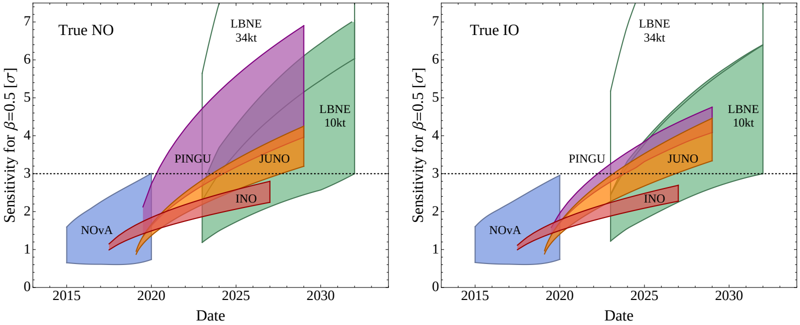

MB, Coloma, Huber, Schwetz, arXiv:1311.1822

# Accelerator experiments - results




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

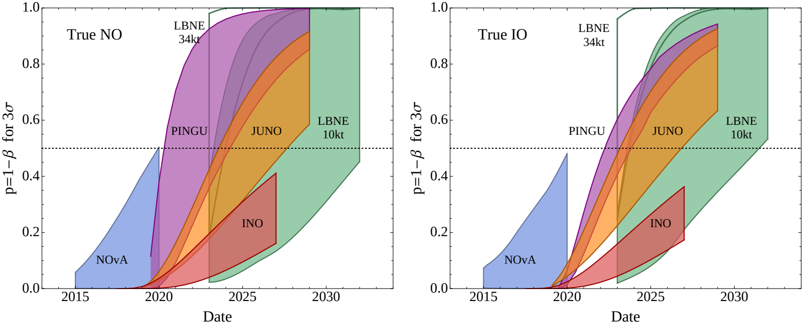
# Comparison of experiments - specific $\beta$




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

*Note:* Bands have different meanings!

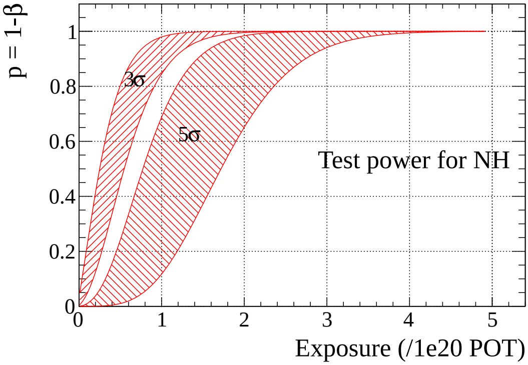
# Comparison of experiments - specific $\alpha$




MB, Coloma, Huber, Schwetz, arXiv:1311.1822

*Note:* Bands have different meanings!

# LBNO predictions



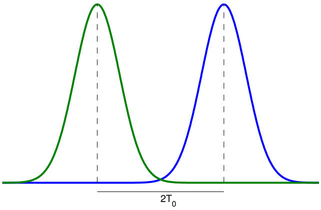

LBNO collaboration, arXiv:1312.6520

# 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



# 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

# 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 $d$

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

# 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:

| Strength of evidence for H | $\kappa$ | Posterior odds | Degree of belief |
|---|---|---|---|
| Barely worth mentioning | 0 to 2 | ca 1 to 3 | $< 73.11\%$ |
| Positive | 2 to 6 | ca 3 to 20 | $> 73.11\%$ |
| Strong | 6 to 10 | ca 20 to 150 | $> 95.26\%$ |
| Very strong | $> 10$ | $\gtrsim 150$ | $> 99.33\%$ |

# What can be said about the future?

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

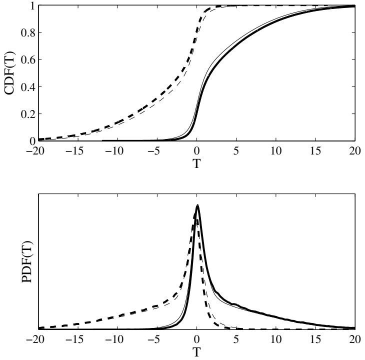
$$P(\kappa_0) = P(\kappa > \kappa_0; H_1)P(H_1) + P(\kappa < -\kappa_0; H_2)P(H_2)$$

- 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

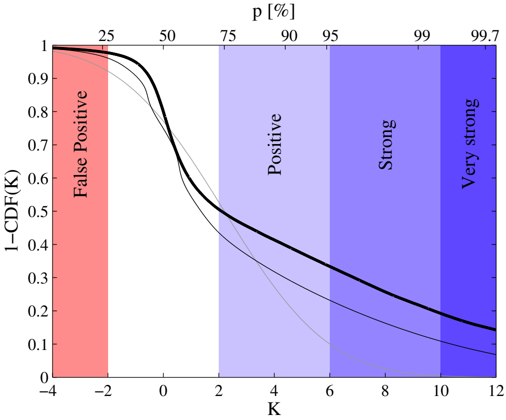
# Example: NO$\nu$A




MB, arXiv:1311.3183

- NO$\nu$A experiment
- GLoBES implementation
- Only $\delta$ and $\Delta m^2_{31}$ varying
- Flat and 10 % Gaussian priors, respectively

# Example: NO$\nu$A, results




MB, arXiv:1311.3183

# Summary and conclusions

- Wilk's theorem is not applicable, the test statistic is not $\chi^2$ distributed
- Regardless, $\sqrt{T_0}$ is still a good approximation of the (median) sensitivity
- Frequentist methods are perfectly applicable
- Bayesian methods equally applicable, matter of preference
- Important not to mix the concepts and be aware of the proper interpretation