Three-Neutrino Mixing Fits

Johannes Bergström

May 31, 2016

(Based on arXiv:1507.04366, with M.C. Gonzalez-Garcia, M. Maltoni, T. Schwetz)

PhyStat-ν Workshop, Kavli IPMU

Universitat de Barcelona

Outline

Outline

Global fits

- Oscillation wavelengths $4\pi E/\Delta m_{ij}^2$
- Different experiments sensitive to different sets of parameters $\Rightarrow$ Global fits

s_{12}^2	s_{23}^2	s_{13}^2	$\Delta m_{21}^2/10^{-5}\mathrm{eV}^2$	$\lvert m_{31(32)}^2\rvert/10^{-3}\mathrm{eV}^2$
$0.27-0.34$	$0.38-0.64$	$0.019-0.025$	$7.0-8.1$	$2.3-2.6$

Gonzalez-Garcia, et al., arXiv:1409.5439, nu-fit.org

Fogli et al., arXiv:1312.2878

Forero et al., arXiv:1405.7540

Global fits – statistical method?

Standard likelihood/χ^2

- Quite simplified
- Easy, reasonably well understood
- Aims for frequentist confidence intervals and hypothesis tests
- Assuming asymptotic (χ^2) distributions
  - Assumption typically not tested – in important cases known to be inaccurate
  - Evaluate through simulation – difficult
  - Or just eliminate need for them

Bayesian model comparison

Model comparison

- Update model probabilities using data
- Or consider the odds

$$\begin{aligned}
\frac{\Pr(A|\mathbf{D})}{\Pr(B|\mathbf{D})} &= \frac{\Pr(\mathbf{D}|A)}{\Pr(\mathbf{D}|B)} \frac{\Pr(A)}{\Pr(B)} \\
\text{Posterior odds} &= \text{Bayes factor} \cdot \text{Prior odds}
\end{aligned}$$

- Usually prior odds $= 1$

Bayesian model comparison

Model comparison

- Update model probabilities using data
- Or consider the odds

$$
\begin{aligned}
\frac{\Pr(A|\mathbf{D})}{\Pr(B|\mathbf{D})} &= \frac{\Pr(\mathbf{D}|A)}{\Pr(\mathbf{D}|B)} \frac{\Pr(A)}{\Pr(B)} \\
\text{Posterior odds} &= \text{Bayes factor} \cdot \text{Prior odds}
\end{aligned}
$$

- Usually prior odds $= 1$

Evidence

- Marginal (model) likelihood

$$
\begin{aligned}
\mathcal{Z} &= \Pr(\mathbf{D}|H) = \int \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})\mathrm{d}^N\mathbf{\Theta} \\
\text{Evidence} &= \text{Average likelihood of model parameters}
\end{aligned}
$$

- Balances quality of fit and model complexity – can favour simpler model

Bayesian model comparison

Model comparison

- Update model probabilities using data
- Or consider the odds

$$\begin{aligned} \frac{\Pr(A|\mathbf{D})}{\Pr(B|\mathbf{D})} &= \frac{\Pr(\mathbf{D}|A)}{\Pr(\mathbf{D}|B)} \frac{\Pr(A)}{\Pr(B)} \\ \text{Posterior odds} &= \text{Bayes factor} \cdot \text{Prior odds} \end{aligned}$$

- Usually prior odds $= 1$

Interpretation?? – Jeffreys scale

odds	Interpretation?
~ 5	Weak evidence
~ 20	Moderate evidence
~ 100	Strong evidence

Lindley's paradox

- Gaussian data with mean μ
- Testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$
- Bayes factor

$$\mathcal{B} \sim \frac{\sigma_\pi}{\sigma_\mathcal{L}} e^{-\lambda^2/2}$$

  $\lambda =$ number of st. dev. from H_0

- Classical test can strongly reject H_0 (large λ) while supported in Bayesian analysis
- "Paradox" in large-sample limit ($\sigma_\mathcal{L} \to 0$), where Bayes and frequentist usually agree
- BUT – Bayes will select right model in this limit!

Problem in practice

- Not so much $n \to \infty$ limit
- Problem if no good value of σ_π
  - not for neutrino parameters
- No difference if H_0 replaced by diffuse hypothesis if width small ($\ll \sigma_\mathcal{L}$)

Data model

$$\Pr(\mathbf{D}|\boldsymbol{\Theta}) = \prod \text{Loads of Poisson distributions}$$

Neutrino parameter space

Neutrino parameter space (Majorana)

- Neutrino mass matrix

$$M_\nu \in \mathbb{C}^{3\times 3}, \quad M_\nu = M_\nu^T$$

- Change variables to $U \in$U(3), $M_\mathrm{d} = \mathrm{diag}(m_1, m_2, m_3)$

$$M_\nu = U M_\mathrm{d} U^T$$

- A priori invariance under basis rotations $M_\nu \to V M_\nu V^T$, $U \to VU$
- Unique uniform prior on U, invariant under basis rotations – Haar measure
  Kass, Wasserman J. Am. Stat. Ass. 91, 1343 (1996)

  Neutrinos: Haba, Murayama hep-ph/0009174

In terms of mixing parameters

- $\pi(s_{12}^2, c_{13}^4, s_{23}^2, \delta) = 1/2\pi$
- Shape and scale fixed – c.f. Lindley's paradox
- Only possible prior without a priori basis preference
  – Why not $\sin\delta$? $\cos\delta$?
- Prior on m_i undetermined – fixed (Gaussian) imposing independence of M_ν elements

Outline

Mass ordering – model comparison

Testing mass ordering

- Bayesian testing of mass ordering should work really well

$$\mathcal{B} = \frac{\int \mathcal{L}_{\mathrm{NO}}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}}{\int \mathcal{L}_{\mathrm{IO}}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}}$$

  - Plausible assumption on their own
  - Actually just interval hypotheses on Δm^2_{31}
  - All parameters are common and retain their meaning
    - No prior sensitivity
    - More evidence for less $\Delta\chi^2$ – but also info in $\mathcal{L}$ shape!
  - Global data: no preference ($\mathcal{B} \sim 1$)
- All other inference can be averaged over mass ordering, e.g. posterior

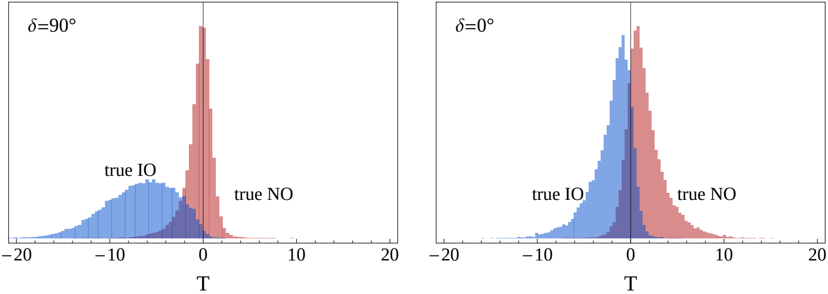
$$\Pr(\boldsymbol{\Theta}|\mathbf{D}) = \sum_{O=\mathrm{NO},\mathrm{IO}} \Pr(\boldsymbol{\Theta}|\mathbf{D}, O)\Pr(O|\mathbf{D})$$

Frequentist test

- Frequentist tests, using e.g.

$$T = \chi^2_{\text{min,IO}} - \chi^2_{\text{min,NO}}$$

- Need to make symmetric tests with NO $\leftrightarrow$ IO
- Distr. of T asymptotic Gaussian – but depends on values of mixing params!



Blennow, 1311.1822 – simulation of NOνA experiment

Mass ordering extension?

- Can smth be learned by nesting mass orderings in extended parameter space
  NO$\leftrightarrow \lambda = 0$, IO$\leftrightarrow \lambda = 1$
- Extend by predicting Poisson means – Trivial?

$$n = \lambda n_{\mathrm{IO}}(\mathbf{\Theta}) + (1-\lambda) n_{\mathrm{NO}}(\mathbf{\Theta})$$

- Comprehensive model

$$\begin{aligned}
\Pr(\mathbf{D}|\mathbf{\Theta}_{\mathrm{NO}}, \mathbf{\Theta}_{\mathrm{IO}}, \lambda) &= \Pr(\mathbf{D}|\mathbf{\Theta}_{\mathrm{IO}})^{\lambda} \Pr(\mathbf{D}|\mathbf{\Theta}_{\mathrm{NO}})^{1-\lambda} / Z(\mathbf{\Theta}_{\mathrm{NO}}, \mathbf{\Theta}_{\mathrm{IO}}, \lambda) \\
&\text{OR} \\
&= \Pr(\mathbf{D}|\mathbf{\Theta}_{\mathrm{IO}})\lambda + \Pr(\mathbf{D}|\mathbf{\Theta}_{\mathrm{NO}})(1-\lambda)
\end{aligned}$$

  - Bayesian fit of λ?
  - Put $\mathbf{\Theta}_{\mathrm{IO}} = \mathbf{\Theta}_{\mathrm{NO}}$?

  See Algeri, Conrad, van Dyk: 1509.01010 and refs
- Free to make up any "embbedding" one wants – physics motivated
  Capozzi, 1309.1638: non-L/E contribution to L/E oscillation phase

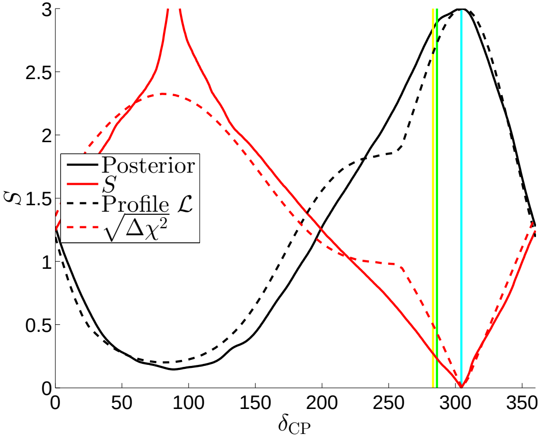
Outline

Likelihoods

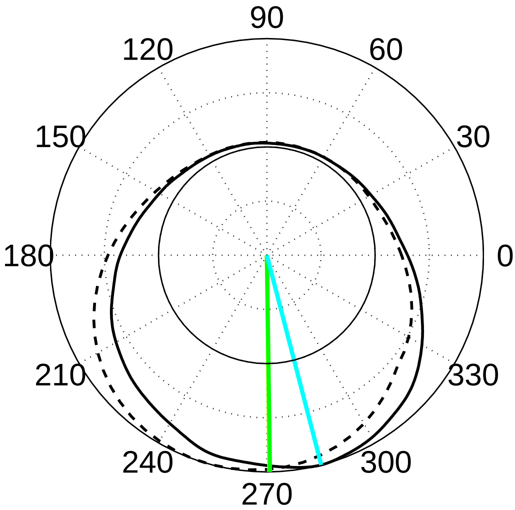
- Posterior/marginal $\mathcal{L}$ vs. profile $\mathcal{L}$

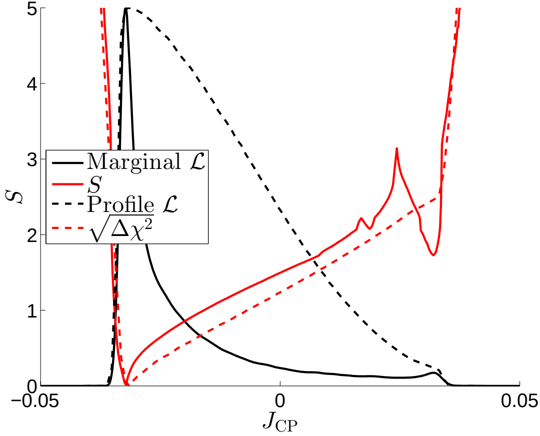
"Significance"

- S – Bayesian cred. lev.
- $\sqrt{\Delta\chi^2}$
- Don't know dist. of test statistic



δ – Posterior and profile $\mathcal{L}$





Jarlskog invariant $J_{\rm CP} = c_{12}s_{12}c_{23}s_{23}c_{13}^2 s_{13} \sin\delta$ – strength of CP-violation

δ – Bayesian summaries

Point estimates

- Circular nature of δ does not affect the posterior distributions or its interpretation
- But summaries – mean, median, st. dev., correlation coefficients – depend on arbitrary choice of origin

  Ex: mean(10°, 350°) = 180°

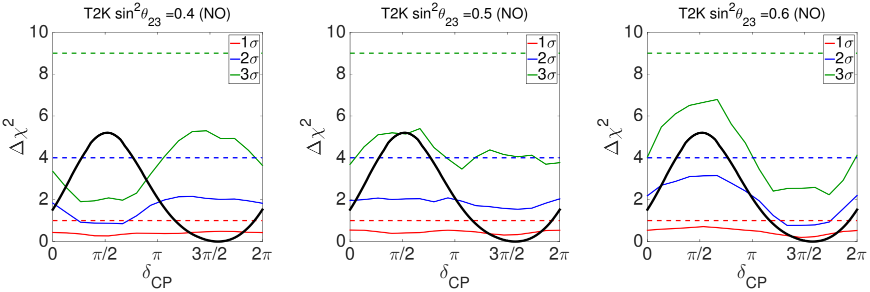
  Should be 0°
- Always need invariant measures
- Circular mean

$$\overline{\delta} = \arg\langle e^{i\delta}\rangle$$

- Circular median: endpoint closer to mean of diameter that splits prob 50/50

δ - Frequentist analyses

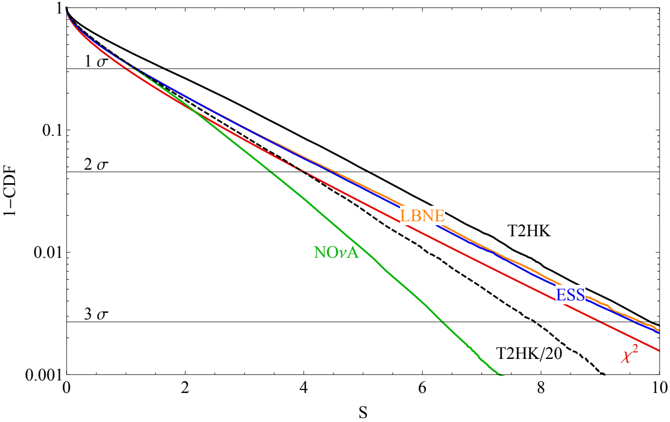
- Asymptotic distributions do not hold (small sample size, bounded parameter space)
- Confidence levels, p-values etc. can depend a lot on assumed s_{23}^2
- Smaller p-value than using asymp. dist.



Elevant, Schwetz, 1506.07685: Value of test statistic corresponding to $n\sigma$. Black = observed

χ^2

- Global data: $\Delta\chi^2 \simeq 1.5 - 3.5$ – impossible to simulate
- In general not quite asymptotic distributions



Blennow, 1407.3274: Distribution of $S = \min_{\delta=0,\pi}\chi^2 - \min_{\text{global}}\chi^2$

Testing for CPV – Bayes

Possbile assumptions

- $\delta = 0$
- $\delta = \pi$
- CPC: $\delta = 0$ or $\delta = \pi$
- CPV: δ free

Question: Reasonable treat $\delta = 0, \pi$ as plausible hypotheses – and put finite prior?

- No? Those are just any parameter values - not "special"
- But in many frequentist analyses treated as very special indeed
- Models of neutrino flavor predicting CP conservation
  - – Approx. symmetry is OK too!
- And IF so – model comparison follows

Testing for CPV – Bayes

Model comparison

- For example, Bayes factor

$$\frac{\Pr(\mathbf{D}|\delta=0)}{\Pr(\mathbf{D}|\mathrm{CPV})} = \frac{\mathcal{L}_\mathrm{m}(\delta=0)}{\int \mathcal{L}_\mathrm{m}(\delta)\pi(\delta)\mathrm{d}\delta} = \left.\frac{P(\delta)}{\pi(\delta)}\right|_{\delta=0}$$

  – Calculable from samples from full posterior

- Global data: No evidence for or against CPV, all $\mathcal{B} < 2.5$
- Prior fixed ⇒ Small and unique penalty for additional parameter
- Bayesian analysis stronger preference for CVP than "typically"
  – More evidence for less $\Delta\chi^2$

Outline

θ_{23}

θ_{23} – octant comparison

- Is θ_{23} larger or smaller than $\pi/4$?
  - – Interval H : model comparison straightforward
- Octants not nested, degeneracies – no χ^2-distribution for frequentist test

θ_{23} – maximal mixing

- Also consider $\theta_{23} = \pi/4$ as a valid assumption (exact or approximate)?
- Similar arguments as for CP-conservation
  - – Test independently? Also prediction of some models

Global data

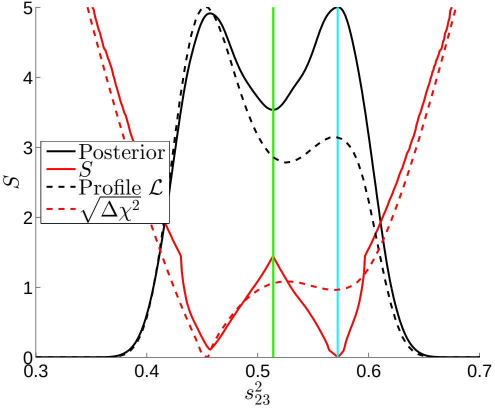
- 2nd octant vs. 1st octant:

  $\mathcal{B} \sim 1.5/3.5$ (NO/IO)

- Maximal mixing vs. non-maximal:

  $\mathcal{B} \sim 4.5/3.5$ (NO/IO)

- Non-maximal punished for additional complexity – but unique and small



Outline

Bayesian experimental design

- Uncertainties predicting data and capabilities of future experiment
  - Which model?
  - Which parameter values?
  - Which data realization?
- *Expected utility* of experiment e

$$\mathrm{EU}(e) = \int \Pr(D_f | D_p, e) U(D_f, e) \mathrm{d}D_f$$

- Utility $U =$ usefulness of observed data D_f in experiment e
- Future data D_f predicted using present data D_p, e.g.,

$$\Pr(D_f | D_p, e) = \int \Pr(D_f | \mathbf{\Theta}, e) \Pr(\mathbf{\Theta} | D_p) \mathrm{d}\mathbf{\Theta}$$

  – Minimizing prior sensitivity
- Frequentist requires performing tons of analyses for fixed parameter values – no way to combine them

Bayesian experimental design

Choice of utility

- Estimation
  - Size of error/Fisher matrix
  - Information gain on parameters
- Testing/signal detection
  - $\Delta\chi^2$
  - Prob observing $\Delta\chi^2 > C$
  - $\log B$
  - Prob observing $\log B > C$
  - Information gain on model space (MO!)

$$\sum_{O=\mathrm{NO,IO}} \Pr(O|\mathbf{D}) \log \frac{\Pr(O|\mathbf{D})}{\Pr(O)}$$

Trotta, Kunz, Liddle: arXiv:1012.3195

Loredo: astro-ph/0409386

von Toussaint, Rev. Mod. Phys. 83 (2011)

Conclusions

- In Bayesian analysis no need for distribution of test statistic
- Unique prior on mixing matrix
- Parameter estimation: no big difference between Bayes and χ^2 – as usual
- Testing: potentially differing conclusions – but not enough data
  - Mass orderings: have equal complexity – model comparison straightforward
  - CPV: weak punishment for extra parameter
- Look at Bayesian experimental design?
- Hopefully you will soon have more data to learn from

Thanks for listening!

EXTRA SLIDES

δ – dispersion

Standard deviation

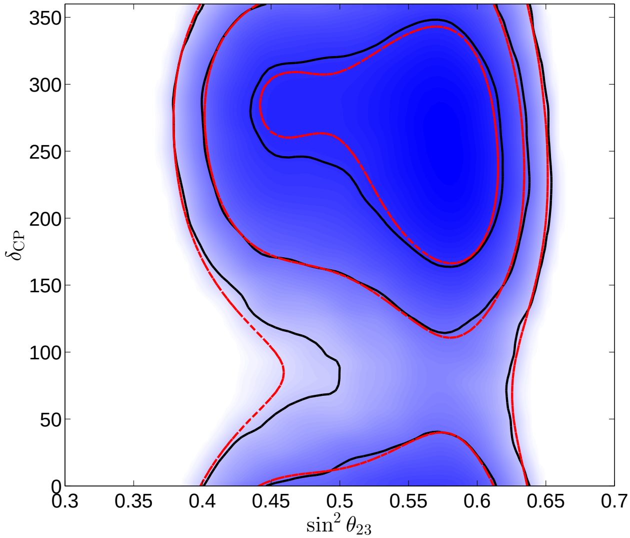
- Standard deviation also not invariant under choice of origin
- Make invariant by using $V = \langle d^2(\delta, \bar{\delta}) \rangle$
  - – σ as an angle
  - – Reproduce linear value as width $\to 0$
- Invariant metric on circle: $d(\alpha, \beta) =$ minimum arc length
- Also Euclidean embedding

$$\begin{aligned} d'(\alpha, \beta)^2 &= |e^{\mathrm{i}\alpha} - e^{\mathrm{i}\beta}|^2 = 2(1 - \cos(\alpha - \beta)) \\ \Rightarrow \sigma' &= \arccos |\langle e^{i\delta} \rangle| \end{aligned}$$

- Global data:

$$\begin{aligned} \sigma/\sigma' &= 65^\circ/58^\circ \text{ (NO)} \\ &= 56^\circ/51^\circ \text{ (IO)} \end{aligned}$$

$s_{23}^2 - \delta$ posterior, IO



Black = Bayes, red = χ^2

s_{23}^2–δ correlation

Linear correlation

- Linear correlation coefficient (of posterior)

$$\rho = \frac{\langle (x - \overline{x})(y - \overline{y}) \rangle}{\sigma_x \sigma_y}$$

- Measures linear association
- Not invariant under parameter transformations
- Not circular-invariant

Correlation with circular variables

- Various coefficients with circular variables, e.g.

$$\rho_{\mathrm{cc}} = \frac{\langle \sin(\Theta - \overline{\Theta}) \sin(\Phi - \overline{\Phi}) \rangle}{\sqrt{\langle \sin^2(\Theta - \overline{\Theta}) \rangle \langle \sin^2(\Phi - \overline{\Phi}) \rangle}}$$

  - In $[-1, 1]$
  - 0 if variables independent
  - Agrees with linear version for concentrated distributions
- Still only sensitive to specific kind of correlation/dependence

Mutual information

- **How much more info in $P(x,y)$ than $P(x)P(y)$?**

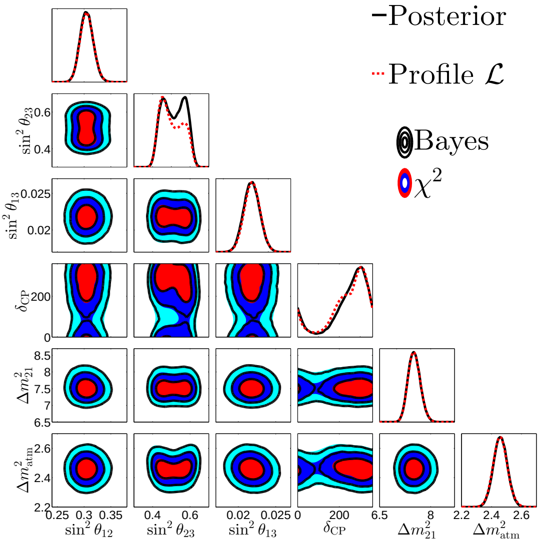
$$I(X,Y) = \int P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\mathrm{d}x\mathrm{d}y$$

- Equals 0 if and only if x and y independent
- Invariant under transformations, boundary conditions
- For Gaussian $I = \log(1/\sqrt{1-\rho^2}) \Rightarrow$

$$|\rho_I| \equiv \sqrt{1-e^{-2I}}$$

- Global data:

	NO	IO
$\rho_{\rm cc}$	-0.20	-0.15
$\lvert\rho_I\rvert$	0.30	0.18



Small difference between chi2 and Bayes