Bayesian analysis of neutrino oscillation data

Johannes Bergström

June 22, 2015

(Based on work with M.C. Gonzalez-Garcia, M. Maltoni, T. Schwetz)

Invisibles15 Worskhop, Madrid

Outline

Outline

Neutrino oscillations

Neutrino oscillations

- Appearance/disappearance of neutrinos observed: solar, reactor, accelerator, atmospheric
- Neutrino oscillations $\Rightarrow$ neutrinos massive and flavours mixed
- No color nor electromagnetic charge $\Rightarrow$ neutrinos Majorana or Dirac particles

3 neutrinos – mixing described by unitary matrix

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \overbrace{\mathrm{diag}\left(e^{i\rho}, e^{i\sigma}, 1\right)}^{\text{only if Majorana}}$$

$s_{ij} = \sin\theta_{ij}, c_{ij} = \cos\theta_{ij}$

Global fits

Global fits

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

s^2_{12}	s^2_{23}	s^2_{13}	$\Delta m^2_{21}/10^{-5}\text{eV}^2$	$\lvert m^2_{31(32)}\rvert/10^{-3}\text{eV}^2$
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 $\Rightarrow$ data used here
Fogli et al., arXiv:1312.2878
Forero et al., arXiv:1405.7540

Global fits

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

Gonzalez-Garcia, et al., arXiv:1409.5439, nu-fit.org $\Rightarrow$ data used here
Fogli et al., arXiv:1312.2878
Forero et al., arXiv:1405.7540

Global fits

- Large mixing, different from quarks
- Some info on δ
- Ordering of masses unknown:
  - Normal (NO): $m_3 > m_1, m_2$
  - Inverted (IO): $m_3 < m_1, m_2$

Global fits – statistical method?

Standard likelihood/χ^2/frequentist fit

- Easy, commonly used, reasonably well understood

Global fits – statistical method?

Standard likelihood/χ^2/frequentist fit

- Easy, commonly used, reasonably well understood
- Does not obey rules of consistent inference
- Depends on data that was never observed ("significance")
- Distributions of test statistics not always known
    - Find out through simulations, but limited computing resources

Global fits – statistical method?

Standard likelihood/χ^2/frequentist fit

- Easy, commonly used, reasonably well understood
- Does not obey rules of consistent inference
- Depends on data that was never observed ("significance")
- Distributions of test statistics not always known
  - Find out through simulations, but limited computing resources

Let's do a Bayesian one! :)

Outline

1. Introduction: oscillations, global fits
2. Bayesian inference
3. Results
   - Posterior distributions
   - Mass ordering
   - s^2_{23}
   - CP-violation
4. Conclusions

Bayesian inference

Bayesian inference

- Proposition A associated with probability (plausibility) $\Pr(A)$
- Related by laws of probability theory
- Update odds using data

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

- Usually prior odds $= 1$

Bayesian inference

Bayesian inference

- Proposition A associated with probability (plausibility) $\Pr(A)$
- Related by laws of probability theory
- Update odds using data

$$\frac{\Pr(A|\mathbf{D})}{\Pr(B|\mathbf{D})} = \frac{\Pr(\mathbf{D}|A)}{\Pr(\mathbf{D}|B)} \frac{\Pr(A)}{\Pr(B)}$$

$$\text{Posterior odds} = \text{Likelihood ratio (Bayes factor)} \cdot \text{Prior odds}$$

- Usually prior odds $= 1$

Evidence

- Model likelihood – evidence

$$\mathcal{Z} = \int \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})\mathrm{d}^N\mathbf{\Theta}$$

$$\text{Model likelihood} = \text{Average likelihood of model parameters}$$

- Evidence balances quality of fit and model complexity – can favour simpler model

Bayesian inference

Bayesian inference

- Proposition A associated with probability (plausibility) $\Pr(A)$
- Related by laws of probability theory
- Update odds using data

$$\frac{\Pr(A|\mathbf{D})}{\Pr(B|\mathbf{D})} = \frac{\Pr(\mathbf{D}|A)}{\Pr(\mathbf{D}|B)} \frac{\Pr(A)}{\Pr(B)}$$

$$\text{Posterior odds} = \text{Likelihood ratio (Bayes factor)} \cdot \text{Prior odds}$$

- Usually prior odds $= 1$

Jeffreys scale: translation into English

$\lvert\log(\text{odds})\rvert$	Interpretation
< 1.0	Inconclusive
1.0	Weak evidence
2.5	Moderate evidence
5.0	Strong evidence

Oscillation parameters and priors

Infer parameters a fixed model

Posterior distribution

$$\Pr(\mathbf{\Theta}|\mathbf{D}) \propto \Pr(\mathbf{D}|\mathbf{\Theta})\Pr(\mathbf{\Theta}) = \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})$$

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Oscillation parameters and priors

Infer parameters a fixed model

Posterior distribution

$$\Pr(\mathbf{\Theta}|\mathbf{D}) \propto \Pr(\mathbf{D}|\mathbf{\Theta})\Pr(\mathbf{\Theta}) = \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})$$

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Oscillation parameters and priors

- A priori invariance under flavor transformations $\Rightarrow$

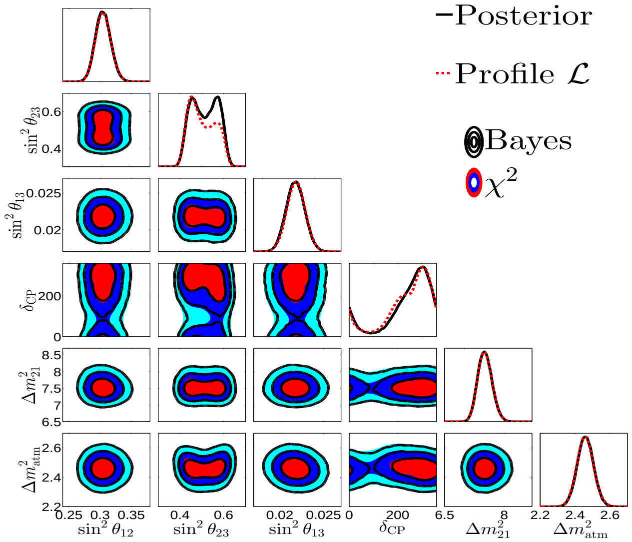
$$\pi(s_{12}^2, c_{13}^4, s_{23}^2, \delta) = \frac{1}{2\pi}$$

- Haar measure, Majorana and unphysical phases marginalized (Haba, Murayama, hep-ph/0009174)
- $\Delta m_{21}^2, \Delta m_{31}^2$, experimental nuisance params,. . .
- Most interesting:
  - s_{23}^2
  - δ
  - mass ordering

Outline

1. Introduction: oscillations, global fits
2. Bayesian inference
3. Results
    - Posterior distributions
    - Mass ordering
    - s^2_{23}
    - CP-violation
4. Conclusions

Posterior distributions: NO



Mass ordering

Mass ordering

- Don't know the ordering ⇒ include its uncertainty
- MO – Mixed ordering: Either NO or IO with equal priors
- Posterior distributions in MO = weighted average of NO and IO posteriors

Mass ordering

Mass ordering

- Don't know the ordering ⇒ include its uncertainty
- MO – Mixed ordering: Either NO or IO with equal priors
- Posterior distributions in MO = weighted average of NO and IO posteriors

But data says very little

- But data says very little:

$$\text{Posterior of IO} \simeq 0.55, \qquad \log \text{odds} \simeq 0.2$$

- Neither ordering preferred
- Compare with $\Delta\chi^2 = 1$
  - but no χ^2-distribution

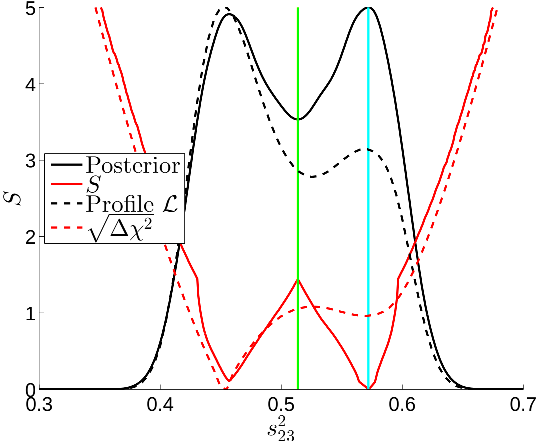
s_{23}^2 – estimation (NO)

Likelihoods: Bayesian vs. maximized

- Marginalization over δ increases probability of second octant

"Significance"

- S – Bayesian
- $\sqrt{\Delta\chi^2}$
- Frequentist significance depends on assumed value of δ



s_{23}^2 – model comparison

s_{23}^2 – octant comparison

- Octants not nested – no χ^2-distribution for frequentist test
- Bayesian analysis straightforward – just do the integration
- Can also consider maximal mixing $s_{23}^2 = 0.5$ as a valid assumption (exact or approximate)

s^2_{23} – model comparison

s^2_{23} – octant comparison

- Octants not nested – no χ^2-distribution for frequentist test
- Bayesian analysis straightforward – just do the integration
- Can also consider maximal mixing $s^2_{23} = 0.5$ as a valid assumption (exact or approximate)

		NO	IO
2nd octant vs. 1st	$\log \mathcal{B}$	0.3	1.2
($>$ 0 prefers 2nd oct)	$\Delta\chi^2$	-0.9	2.0

Conclusions

- Second octant weakly preferred over the first for IO

s_{23}^2 – model comparison

s_{23}^2 – octant comparison

- Octants not nested – no χ^2-distribution for frequentist test
- Bayesian analysis straightforward – just do the integration
- Can also consider maximal mixing $s_{23}^2 = 0.5$ as a valid assumption (exact or approximate)

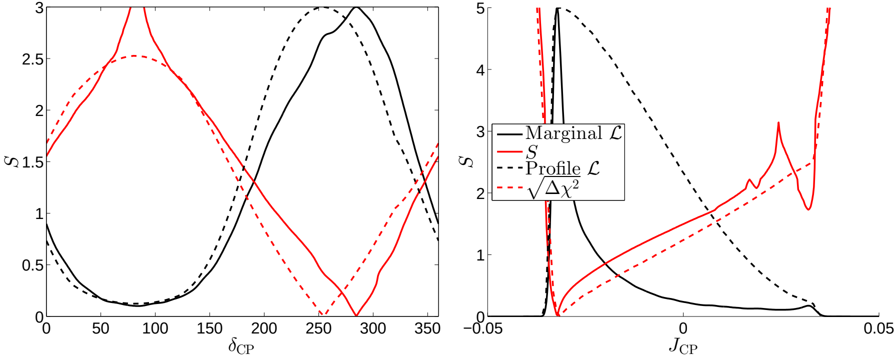
		NO	IO
2nd octant vs. 1st	$\log \mathcal{B}$	0.3	1.2
(> 0 prefers 2nd oct)	$\Delta\chi^2$	-0.9	2.0
Non-maximal vs. Maximal	$\log \mathcal{B}$	-1.4	-1.2
(> 0 prefers non-maximal)	$\Delta\chi^2$	0.9	2.0

Conclusions

- Second octant weakly preferred over the first for IO
- No evidence for non-maximal mixing – maximal weakly preferred
- Non-maximal punished for additional complexity – but unique and small

CP-violation – estimation

MO:

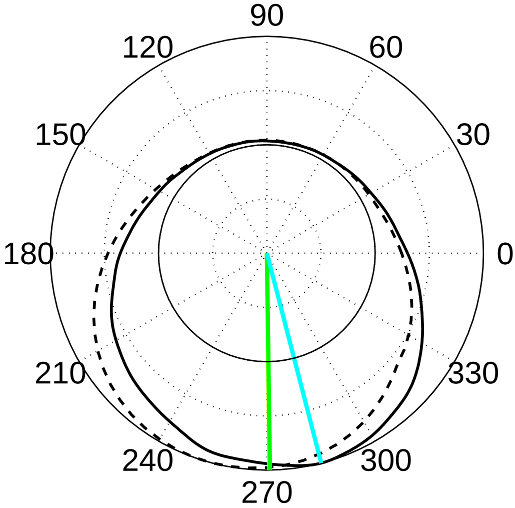


Jarlskog invariant $J_{\rm CP} = c_{12}s_{12}c_{23}s_{23}c_{13}^2 s_{13} \sin\delta$

Frequentist analysis

- Asymptotic distributions do not hold
- Statements regarding δ depends on assumed s_{23}^2

δ – circularity



CP-violation - model comparison

Possbile assumptions

- $\delta = 0^\circ$
- $\delta = 180^\circ$
- CPV: δ free

CP-violation - model comparison

Possbile assumptions

- $\delta = 0°$
- $\delta = 180°$
- CPV: δ free

Results

- Weak penalty for additional parameter
- Bayesian analysis more powerful than normally, and than χ^2
- Compared to CPV:

	NO	IO
$\delta = 0°$	-0.1	-0.8
$\delta = 180°$	-0.4	-0.1

- No evidence for or against CPV
- $\Delta\chi^2 \simeq 1.5 - 3.5$

Conclusions

Conclusions

- Consistent Bayesian analysis – no need for distribution of test statistic etc.
- Neither ordering preferred
- s_{23}^2 – difference compared to χ^2, but no evidence for non-maximal mixing, or any octant
- δ – difference compared to χ^2, no evidence for CP-violation
- Hopefully we will soon have better data to learn more

Thank you!

Thank you!

EXTRA SLIDES

δ – circularity

Point estimates

- Mean, median of δ not well defined – depend on arbitrary choice of origin (Ex: mean of 10° and 350° is 180°. Should be 0°)
- Always need invariant measures
- Circular mean

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

- Circular median : endpoint closer to mean of the diameter that splits the probability equally
- Also applies to standard deviation, correaltions, ...

δ – dispersion

Standard deviation

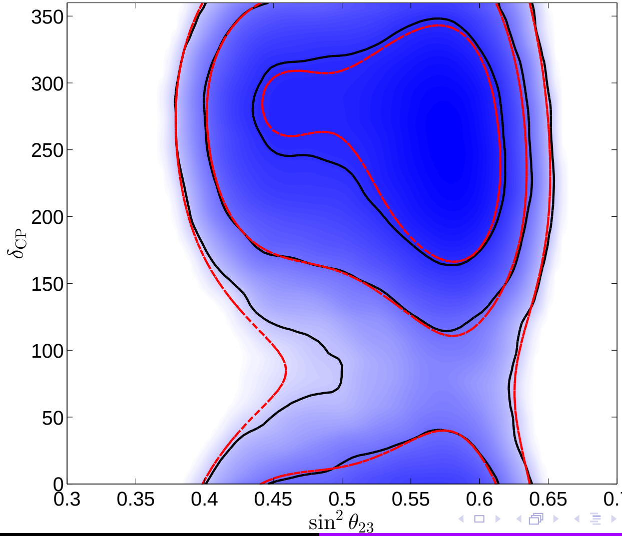
- Standard deviation also not invariant under choice of origin
- Make invariant by using $V = \langle d^2(\delta, \overline{\delta})\rangle$
- Invariant metric on circle: $d(\alpha, \beta) =$ minimum arc length, or
- Or from Euclidean embedding

$$d'(\alpha, \beta)^2 = |e^{i\alpha} - e^{i\beta}|^2 = 2(1 - \cos(\alpha - \beta))$$

- Results:

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

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



s_{23}^2–δ correlation

Linear correlation

- χ^2 only gives "local" correlation at best-fit
- Bayes gives global, but

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

Not circular-invariant

s_{23}^2–δ correlation

Linear correlation

- χ^2 only gives "local" correlation at best-fit
- Bayes gives global, but

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

Not circular-invariant

Correlation with circular variables

- Between two circular variables

$$r_{\mathrm{cc}} = \frac{\langle \sin(x-\overline{x})\sin(y-\overline{y})\rangle}{\sqrt{\langle \sin^2(x-\overline{x})\rangle\langle \sin^2(y-\overline{y})\rangle}}$$

- Circular-linear

$$r_{\mathrm{cl}}^2 = \frac{r_{xc}^2 + r_{xs}^2 - 2 r_{xs} r_{xc} r_{cs}}{1 - r_{cs}^2}$$

$r_{xc} = r(x, \cos y)$, $r_{xs} = r(x, \sin y)$, $r_{cs} = r(\cos y, \sin y)$.

- Still only sensitive to specific kind of correlation/dependence

s_{23}^2–δ correlation

Mutual information

- **How much is learned about x by knowing 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 redefinitions, boundary conditions
- For Gaussian $I = \log(1/\sqrt{1-r^2})$, define

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

	NO	IO	MO
$r_{\rm cc}$	-0.20	-0.15	-0.21
$r_{\rm cl}$	0.27	0.16	0.23
r_I	0.30	0.18	0.26