Stefano Gariazzo

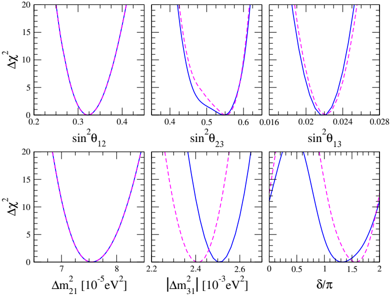
IFIC, Valencia (ES)
CSIC – Universitat de Valencia

gariazzo@ific.uv.es
http://ific.uv.es/~gariazzo/

Fits to large and combined data sets

Neutrino properties
from oscillations and cosmology

PHYSTAT-nu 2019, CERN, 23/01/2019



$\Delta\chi^2$
$\sin^2\theta_{12}$
$\sin^2\theta_{23}$
$\sin^2\theta_{13}$
Δm^2_{21} [10^{-5}eV^2]
$|\Delta m^2_{31}|$ [10^{-3}eV^2]
δ/π

Three Neutrino Oscillations

$$\nu_\alpha = \sum_{k=1}^{3} U_{\alpha k} \nu_k \quad (\alpha = e, \mu, \tau)$$

$U_{\alpha k}$ described by 3 mixing angles θ_{12}, θ_{13}, θ_{23} and one CP phase δ_{CP}

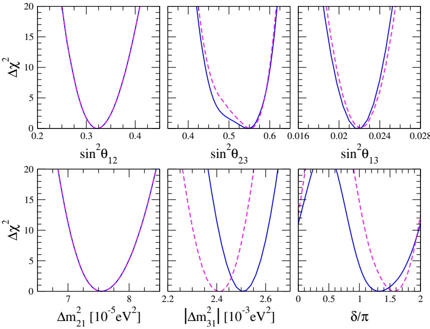
Current knowledge of the 3 active ν mixing: [de Salas et al. (2018)]

NO: Normal Ordering, $m_1 < m_2 < m_3$ IO: Inverted Ordering, $m_3 < m_1 < m_2$

$$\begin{aligned}
\Delta m^2_{21} &= (7.55^{+0.20}_{-0.16}) \cdot 10^{-5}\ \mathrm{eV}^2 \\
|\Delta m^2_{31}| &= (2.50 \pm 0.03) \cdot 10^{-3}\ \mathrm{eV}^2\ (\mathrm{NO}) \\
&= (2.42^{+0.03}_{-0.04}) \cdot 10^{-3}\ \mathrm{eV}^2\ (\mathrm{IO})
\end{aligned}$$

$$\begin{aligned}
\sin^2(\theta_{12}) &= 0.320^{+0.020}_{-0.016} \\
\sin^2(\theta_{13}) &= 0.0216^{+0.008}_{-0.007}\ (\mathrm{NO}) \\
&= 0.0222^{+0.007}_{-0.008}\ (\mathrm{IO}) \\
\sin^2(\theta_{23}) &= 0.547^{+0.020}_{-0.030}\ (\mathrm{NO}) \\
&= 0.551^{+0.018}_{-0.030}\ (\mathrm{IO})
\end{aligned}$$

First hints for $\delta_{\mathrm{CP}} \simeq 3/2\pi$



see also: `http://globalfit.astroparticles.es`

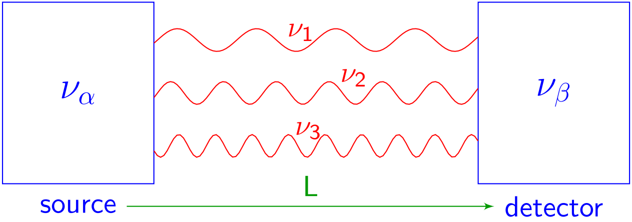
Two types of neutrinos

flavor neutrinos ν_α

$|\nu_\alpha\rangle = U_{\alpha k}|\nu_k\rangle$

massive neutrinos ν_k

$$|\nu(t=0)\rangle = |\nu_\alpha\rangle = U_{\alpha 1}|\nu_1\rangle + U_{\alpha 2}|\nu_2\rangle + U_{\alpha 3}|\nu_3\rangle$$



$$|\nu(t>0)\rangle = |\nu_\beta\rangle = U_{\alpha 1}\, e^{-iE_1 t}|\nu_1\rangle + U_{\alpha 2}\, e^{-iE_2 t}|\nu_2\rangle + U_{\alpha 3}\, e^{-iE_3 t}|\nu_3\rangle \neq |\nu_\alpha\rangle$$

$E_k^2 = p^2 + m_k^2$ ⟵ define ⟶ $t = L$

$$P_{\nu_\alpha \to \nu_\beta}(L) = |\langle \nu_\alpha|\nu(L)\rangle|^2 = \sum_{k,j} U_{\beta k} U^*_{\alpha k} U^*_{\beta j} U_{\alpha j} \exp\left(-i\frac{\Delta m^2_{kj} L}{2E}\right)$$

$$\Delta m^2_{ij} = m_i^2 - m_j^2$$

Three-neutrino oscillation data

Solar + LBL reactors

Experiments:	Parameters:
SuperK	θ_{12}
SNO	Δm^2_{21}
Borexino	(θ_{13})
KamLAND	
...	

baseline defined by $\Delta m^2_{kj} \cdot L/E$ LBL: long baseline ($E/L \gtrsim \Delta m^2_{31}$) SBL: short baseline ($E/L \sim \Delta m^2_{21}$)

Three-neutrino oscillation data

Solar + LBL reactors

Experiments:	Parameters:
SuperK	θ_{12}
SNO	Δm^2_{21}
Borexino	(θ_{13})
KamLAND	
...	

SBL reactors

Experiments:	Parameters:
DayaBay	θ_{13}
RENO	Δm^2_{31}
DoubleChooz	(θ_{12})
...	(Δm^2_{21})

baseline defined by $\Delta m^2_{kj} \cdot L/E$ LBL: long baseline ($E/L \gtrsim \Delta m^2_{31}$) SBL: short baseline ($E/L \sim \Delta m^2_{21}$)

Three-neutrino oscillation data

Solar + LBL reactors

Experiments:	Parameters:
SuperK	θ_{12}
SNO	Δm^2_{21}
Borexino	(θ_{13})
KamLAND	
...	

SBL reactors

Experiments:	Parameters:
DayaBay	θ_{13}
RENO	Δm^2_{31}
DoubleChooz	(θ_{12})
...	(Δm^2_{21})

Atmospheric

Experiments:	Parameters:
Antares	θ_{23}
IceCube	Δm^2_{31}
SuperK	(θ_{13})
...	$(\delta_{\rm CP})$

baseline defined by $\Delta m^2_{kj} \cdot L/E$ LBL: long baseline ($E/L \gtrsim \Delta m^2_{31}$) SBL: short baseline ($E/L \sim \Delta m^2_{21}$)

Three-neutrino oscillation data

Solar + LBL reactors

Experiments:	Parameters:
SuperK	θ_{12}
SNO	Δm^2_{21}
Borexino	(θ_{13})
KamLAND	
...	

SBL reactors

Experiments:	Parameters:
DayaBay	θ_{13}
RENO	Δm^2_{31}
DoubleChooz	(θ_{12})
...	(Δm^2_{21})

Atmospheric

Experiments:	Parameters:
Antares	θ_{23}
IceCube	Δm^2_{31}
SuperK	(θ_{13})
...	$(\delta_{\rm CP})$

LBL accelerators

Experiments:	Parameters:
NOνA	θ_{13}
T2K	θ_{23}
MINOS	Δm^2_{31}
...	$\delta_{\rm CP}$

baseline defined by $\Delta m^2_{kj} \cdot L/E$ LBL: long baseline ($E/L \gtrsim \Delta m^2_{31}$)   SBL: short baseline ($E/L \sim \Delta m^2_{21}$)

Studying the χ^2

We have to combine all the experiments to study the global picture

Use total $\chi^2 = \sum_i \chi_i^2$ information

Experiments as independent!

Find best-fit (χ^2 minimum)
in D-dimensional parameter space

↑

Minimization problem,
in principle not difficult

Studying the χ^2

We have to combine all the experiments to study the global picture

Use total $\chi^2 = \sum_i \chi^2_i$ information

Experiments as independent!

Find best-fit (χ^2 minimum) in *D*-dimensional parameter space

↑

Minimization problem, in principle not difficult

This is expensive!

Find bounds/regions defined by various $\Delta\chi^2$ values

e.g. bounds for 1 parameter at 1σ (68.3% CL): $\Delta\chi^2 = 1$

e.g. bounds for 2 parameters at 1σ (68.3% CL): $\Delta\chi^2 = 2.3$

Studying the χ^2

We have to combine all the experiments to study the global picture

Use total $\chi^2 = \sum_i \chi_i^2$ information

Experiments as independent!

Find best-fit (χ^2 minimum) in D-dimensional parameter space

↑

Minimization problem, in principle not difficult

This is expensive!

Find bounds/regions defined by various $\Delta\chi^2$ values

e.g. bounds for 1 parameter at 1σ (68.3% CL): $\Delta\chi^2 = 1$

e.g. bounds for 2 parameters at 1σ (68.3% CL): $\Delta\chi^2 = 2.3$

If D is small, you can create a grid of χ^2 points, and then analyse 1/2-dimensional sections of the grid

Given N points per dimension, the grid requires N^D χ^2 calculations...

This way will become unfeasible for large D!

What if the number of parameters increases?

χ^2 of ν oscillation experiments depends on 3/4 *physical* parameters

BUT

Nuisance parameters sometimes enter!

(flux models, propagation model, detector response, ...)

New physics?

(NSI, Lorentz violation, non-unitarity, sterile neutrino, ...)

Combined analyses?

(coherent scattering, cosmology, mass measurements, $0\nu\beta\beta$, (multimessenger) astrophysics, ...)

What if the number of parameters increases?

χ^2 of ν oscillation experiments depends on 3/4 *physical* parameters

BUT

Nuisance parameters sometimes enter!

(flux models, propagation model, detector response, ...)

New physics?

(NSI, Lorentz violation, non-unitarity, sterile neutrino, ...)

Combined analyses?

(coherent scattering, cosmology, mass measurements, $0\nu\beta\beta$, (multimessenger) astrophysics, ...)

scanning χ^2 in a grid not feasible, too many parameters!

also: single χ^2 computation may become expensive (e.g.: cosmology)

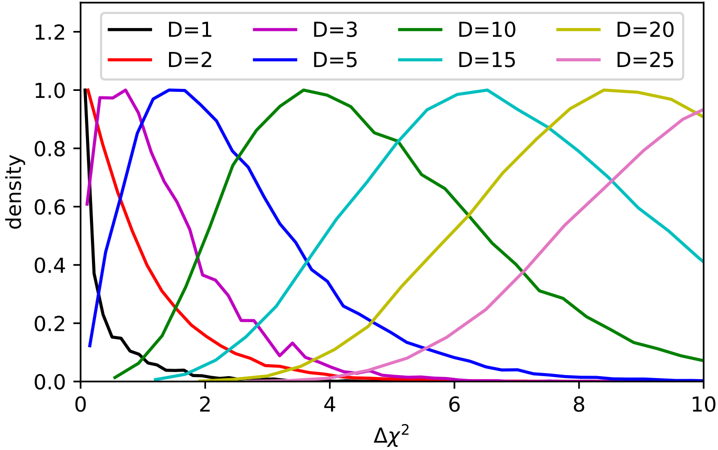
Possibility: use Monte Carlo scan, only study χ^2 contours → χ^2 profiling

(find best-fit and build contours using random points instead of regular spaced ones)

This is not the Bayesian way!

MCMC in a Bayesian context

Problem!



"frequentist" MCMC method not good for exploring around the best-fit!

point density near $\chi^2_{\rm min}$ may be too small, difficult to profile well the χ^2

MCMC in a Bayesian context

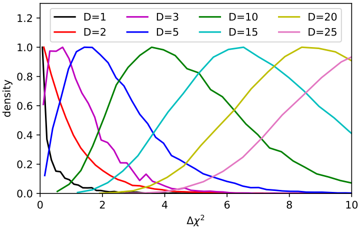
Problem!

"frequentist" MCMC method

not good

for exploring around the best-fit!

point density near $\chi^2_{\rm min}$ may be too small,
difficult to profile well the χ^2



$\chi^2 = -2\ln\mathcal{L}$ conversion

Solution?

use Bayesian methods to analyse MCMC output!

no need to find the real $\chi^2_{\rm min}$,
as what matters most is the parameter space *around* the best-fit

→ Marginalization, not profiling

MCMC in a Bayesian context

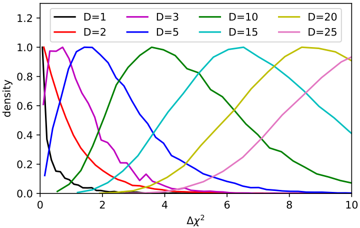
Problem!

"frequentist" MCMC method

not good

for exploring around the best-fit!

point density near $\chi^2_{\min}$ may be too small, difficult to profile well the χ^2



$\chi^2 = -2 \ln \mathcal{L}$ conversion

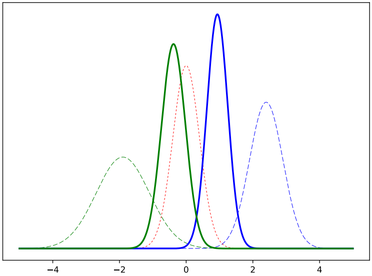
Solution?

use Bayesian methods to analyse MCMC output!

no need to find the real $\chi^2_{\min}$, as what matters most is the parameter space *around* the best-fit

→ Marginalization, not profiling

study the distribution of points in the parameter space, not single points


−4
−2
0
2
4

Bayes' theorem

how to deal with **Bayesian probability**?

given hypothesis H, data d, some information I (true):

$p(\theta)$

Posterior probability:

what we know after

Bayes theorem:

$$p(H|d,I) = \frac{p(d|H,I)\,p(H|I)}{p(d|I)}$$

$\pi(\theta)$

Prior probability:

what we knew before

Likelihood: $\mathcal{L}(\theta)$

sampling distribution of data, given that H is true

Marginal likelihood:

or "Bayesian evidence",

$$p(d|I) \equiv \sum_H p(d|H,I)\,p(H|I)$$

Bayes theorem:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Bayes' theorem

how to deal with **Bayesian probability**?

given hypothesis H, data d, some information I (true):

$p(\theta)$

Posterior probability:

what we know after

Bayes theorem:

$$p(H|d,I) = \frac{p(d|H,I)\,p(H|I)}{p(d|I)}$$

$\pi(\theta)$

Prior probability:

what we knew before

Likelihood: $\mathcal{L}(\theta)$

sampling distribution of data, given that H is true

Marginal likelihood:

or "Bayesian evidence",

$$p(d|I) \equiv \sum_H p(d|H,I)\,p(H|I)$$

parameter inference

Bayes theorem:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Bayes' theorem

how to deal with **Bayesian probability**?

given hypothesis H, data d, some information I (true):

$p(\theta)$

Posterior probability:

what we know after

$\pi(\theta)$

Prior probability:

what we knew before

Bayes theorem:

$$p(H|d,I) = \frac{p(d|H,I)\, p(H|I)}{p(d|I)}$$

Likelihood: $\mathcal{L}(\theta)$

sampling distribution of data, given that H is true

Marginal likelihood:

or "Bayesian evidence",

$$p(d|I) \equiv \sum_H p(d|H,I)\, p(H|I)$$

parameter inference

Bayes theorem:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

model comparison

MCMC with Metropolis-Hastings

[Metropolis et al., 1953]
[Hastings, 1970]

MCMC = build a series of points θ_i in the parameter space
(they should be independent, as much as possible)

The main point: how to go from θ_n to θ_{n+1}

(sampling the points with a density proportional to the posterior $p(\theta)$)

Key idea: Use a *proposal density distribution* $q(\theta_n, \theta_{n+1})$

Acceptance probability: $\alpha(\theta_n, \theta_{n+1}) = \min\left\{1, \frac{p(\theta_{n+1})\, q(\theta_{n+1}, \theta_n)}{p(\theta_n)\, q(\theta_n, \theta_{n+1})}\right\}$

↓

Transition probability: $T(\theta_n, \theta_{n+1}) = \alpha(\theta_n, \theta_{n+1})\, q(\theta_n, \theta_{n+1})$

↓

Detailed balance holds: $p(\theta_{n+1})\, T(\theta_{n+1}, \theta_n) = p(\theta_n)\, T(\theta_n, \theta_{n+1})$

→ $p(\theta)$ is the equilibrium distribution of the chain

Bayesian evidence

"Bayesian evidence" or "Marginal likelihood"

$$p(d|\mathcal{M}) = Z = \int_{\Omega_{\mathcal{M}}} p(d|\theta, \mathcal{M})\, p(\theta|\mathcal{M})\, d\theta$$

integrate over all possible (continuous) parameters of model $\mathcal{M}$
(given that $\mathcal{M}$ is true)

What if there are several possible models $\mathcal{M}_i$?

use Z_i to perform bayesian model comparison

Warning: compare models given the same data!

Model posterior:

$$p(\mathcal{M}_i|d) \propto p(\mathcal{M}_i)\, Z_i$$

← proportional to constant that depends only on data

given model prior $p(\mathcal{M}_i)$

Bayes factor

Posterior odds of $\mathcal{M}_1$ versus $\mathcal{M}_2$:

$$\frac{p(\mathcal{M}_1|d)}{p(\mathcal{M}_2|d)} = B_{1,2}\,\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}$$

Bayes factor:

$$B_{1,2} = \frac{Z_1}{Z_2} \quad \Rightarrow \quad \ln B_{1,2} = \ln Z_1 - \ln Z_2$$

if priors are the same $[p(\mathcal{M}_1) = p(\mathcal{M}_2)]$,
$B_{1,2}$ tells which model is preferred:

$B_{1,2} > 1$ ($\ln B_{1,2} > 0$)

$\mathcal{M}_1$ preferred

$B_{1,2} < 1$ ($\ln B_{1,2} < 0$)

$\mathcal{M}_2$ preferred

$\exp(|\ln B_{1,2}|)$ tells the odds in favor of preferred model

Occam's razor

what the Bayesian model comparison tells us?

Best model strikes optimum balance between

Quality of fit

Predictivity

Occam's razor

the simplest theory that fits data is preferred

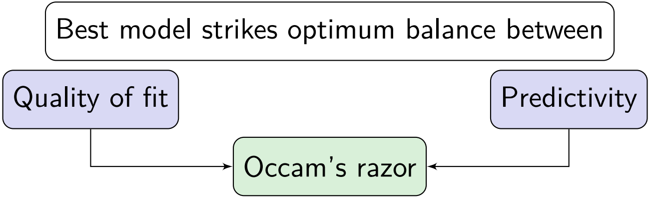
model with more parameters ⟶ better fit (usually)

⟶ are all the parameters needed?

Bayes factor penalizes unnecessarily complex models!

Occam's razor

what the Bayesian model comparison tells us?



the simplest theory that fits data is preferred

what if we compare same model and different priors?

Bayesian evidence depends on priors!

Bayes factor penalizes unnecessarily wide priors!

Occam's razor

what the Bayesian model comparison tells us?

Best model strikes optimum balance between

Quality of fit → Occam's razor ← Predictivity

the simplest theory that fits data is preferred

what if we compare same model and different priors?

Bayesian evidence depends on priors!

Bayes factor penalizes unnecessarily wide priors!

Bayes factor DOES NOT penalize models with parameters that are unconstrained by the data

Jeffreys' scale

[Trotta, arxiv:0803.4089]

odds in favor of the preferred model:

$$\exp(|\ln B_{1,2}|) : 1$$

strength of preference according to Jeffreys' scale:

$\lvert\ln B_{1,2}\rvert$	Odds	$N\sigma$	strength of evidence
< 1.0	$\lesssim 3:1$	< 1.1	inconclusive
$\in [1.0, 2.5]$	$(3-12):1$	$1.1-1.7$	weak
$\in [2.5, 5.0]$	$(12-150):1$	$1.7-2.7$	moderate
$\in [5.0, 10]$	$(150-2.2\times10^4):1$	$2.7-4.1$	strong
$\in [10, 15]$	$(2.2\times10^4-3.3\times10^6):1$	$4.1-5.1$	very strong
> 15	$> 3.3\times10^6:1$	> 5.1	decisive

odds & strength always valid

$N\sigma$ correspondence is valid only given equal model priors
and that only two models are possible
(see e.g. neutrino mass ordering: normal OR inverted)

Jeffreys' scale

[Trotta, arxiv:0803.4089]

odds in favor of the preferred model:

$$\exp(|\ln B_{1,2}|) : 1$$

strength of preference according to Jeffreys' scale:

$\|\ln B_{1,2}\|$	Odds	$N\sigma$	strength of evidence
< 1.0	$\lesssim 3:1$	< 1.1	inconclusive
$\in [1.0, 2.5]$	$(3-12):1$	$1.1-1.7$	weak
$\in [2.5, 5.0]$	$(12-150):1$	$1.7-2.7$	moderate
$\in [5.0, 10]$	$(150-2.2\times10^4):1$	$2.7-4.1$	strong
$\in [10, 15]$	$(2.2\times10^4-3.3\times10^6):1$	$4.1-5.1$	very strong
> 15	$> 3.3\times10^6:1$	> 5.1	decisive

odds & strength always valid

$N\sigma$ correspondence is valid only given equal model priors
and that only two models are possible
(see e.g. neutrino mass ordering: normal OR inverted)

Can we extend to more than two (mutually exclusive) models?

How to compute the model posterior

[SG+, arxiv:1812.05449]

Assume N models, equal model prior probabilities:

$$\pi_i \equiv p(\mathcal{M}_i) \qquad \pi_i = \pi_j \quad \forall i,j \qquad \sum_i \pi_i = 1 \rightarrow \pi_i = 1/N$$

Compute model posterior probabilities:

$$p_i \equiv p(\mathcal{M}_i|d) \qquad p_i = A\pi_i Z_i \quad \text{with } A \text{ constant} \qquad \sum_i p_i = 1$$

$$\sum_i^N p_i = A\sum_i^N \pi_i Z_i = 1 \quad \Longrightarrow \quad p_i = \pi_i Z_i \Big/ \sum_j^N \pi_j Z_j = \pi_i \Big/ \sum_j^N \pi_j B_{ji}$$

Selecting a generic $\mathcal{M}_0$ as a reference, we have:

$$p_0 = \left(\sum_i^N B_{i0}\right)^{-1}$$

the sum includes $B_{00} = 1$

How to compute the model posterior

[SG+, arxiv:1812.05449]

Assume N models, equal model prior probabilities:

$$\pi_i \equiv p(\mathcal{M}_i) \qquad \pi_i = \pi_j \quad \forall i, j \qquad \sum_i \pi_i = 1 \rightarrow \pi_i = 1/N$$

Compute model posterior probabilities:

$$p_i \equiv p(\mathcal{M}_i|d) \qquad p_i = A\pi_i Z_i \quad \text{with } A \text{ constant} \qquad \sum_i p_i = 1$$

$$\sum_i^N p_i = A \sum_i^N \pi_i Z_i = 1 \quad \Longrightarrow \quad p_i = \pi_i Z_i \Big/ \sum_j^N \pi_j Z_j = \pi_i \Big/ \sum_j^N \pi_j B_{ji}$$

Selecting a generic $\mathcal{M}_0$ as a reference, we have:

$$p_0 = \left(\sum_i^N B_{i0} \right)^{-1}$$

the sum includes $B_{00} = 1$

example 1: $N = 2$

$$p_0 = 1/(1 + B_{10})$$
$$p_1 = B_{10}/(1 + B_{10})$$

How to compute the model posterior

[SG+, arxiv:1812.05449]

Assume N models, equal model prior probabilities:

$\pi_i \equiv p(\mathcal{M}_i)$ $\quad$ $\pi_i = \pi_j \quad \forall i, j$ $\quad$ $\sum_i \pi_i = 1 \rightarrow \pi_i = 1/N$

Compute model posterior probabilities:

$p_i \equiv p(\mathcal{M}_i | d)$ $\quad$ $p_i = A \pi_i Z_i$ with A constant $\quad$ $\sum_i p_i = 1$

$$\sum_i^N p_i = A \sum_i^N \pi_i Z_i = 1 \quad \Longrightarrow \quad p_i = \pi_i Z_i \Bigg/ \sum_j^N \pi_j Z_j = \pi_i \Bigg/ \sum_j^N \pi_j B_{ji}$$

Selecting a generic $\mathcal{M}_0$ as a reference, we have:

$$p_0 = \left(\sum_i^N B_{i0} \right)^{-1}$$

the sum includes $B_{00} = 1$

example 1: $N = 2$

$p_0 = 1/(1 + B_{10})$

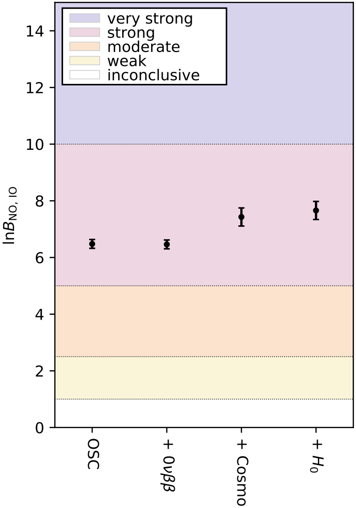
$p_1 = B_{10}/(1 + B_{10})$

example 2: $N = 8$

assume $B_{i0} \simeq e^{-5}$ $(i \neq 0)$ to get

$p_0 = 1/(1 + \sum_{i \neq 0} B_{i0}) \simeq 0.955$

strong? no, only 2σ!


very strong
strong
moderate
weak
inconclusive
$\ln B_{\mathrm{NO,\,IO}}$
0
2
4
6
8
10
12
14
OSC
+ $0\nu\beta\beta$
+ Cosmo
+ H_0

Neutrinos and their masses

[de Salas+, Frontiers 5 (2018) 36]

Normal ordering (NO)

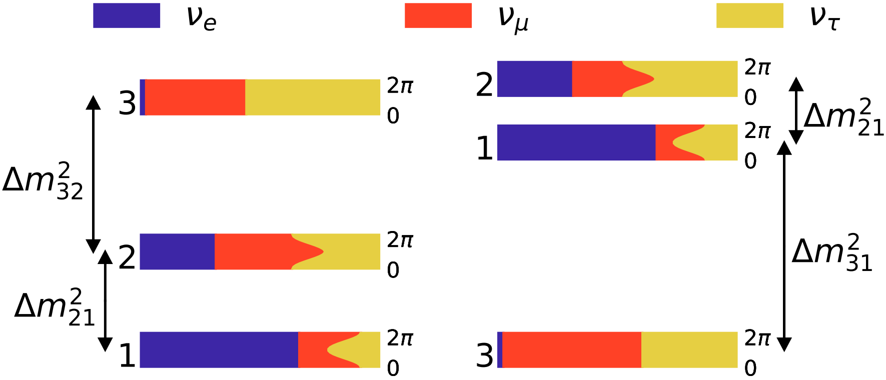
$m_1 < m_2 < m_3$

$\sum m_k \gtrsim 0.06$ eV

Inverted ordering (IO)

$m_3 < m_1 < m_2$

$\sum m_k \gtrsim 0.1$ eV



Absolute scale unknown!

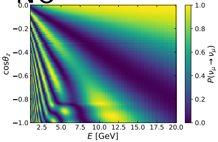
Can we constrain the mass ordering using bounds on $\sum m_\nu$?

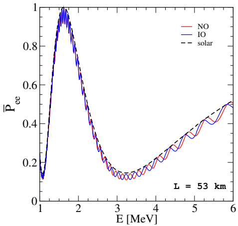
Constraining the mass ordering

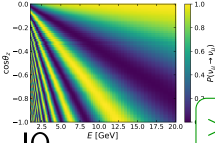
[de Salas+, Frontiers 5 (2018) 36]

NO

oscillations







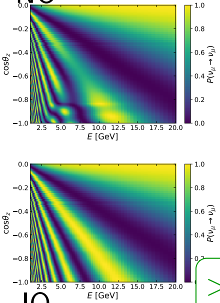
IO

$> 3\sigma$ preference for NO

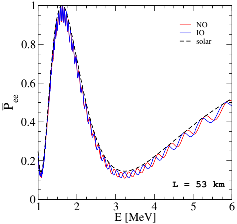
Constraining the mass ordering

[de Salas+, Frontiers 5 (2018) 36]

NO oscillations

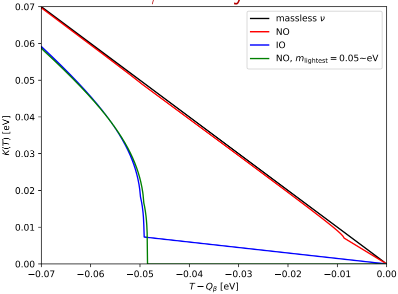


IO

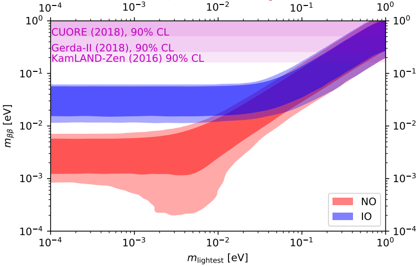


$> 3\sigma$ preference for NO

β decay

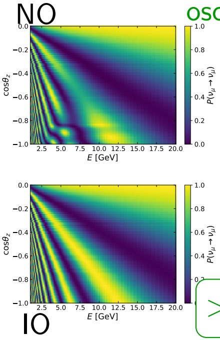


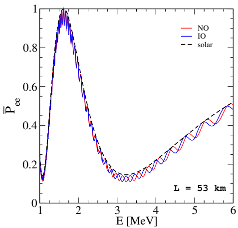
$0\nu\beta\beta$ decay



Constraining the mass ordering

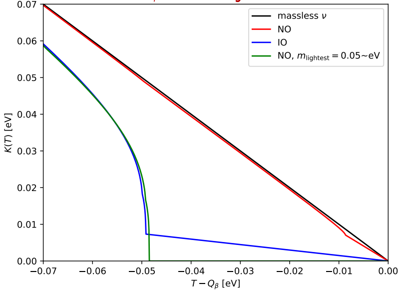
oscillations



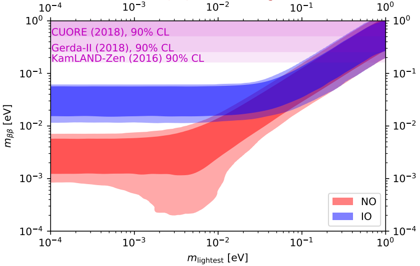


$> 3\sigma$ preference for NO

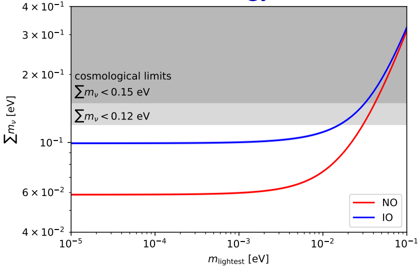
β decay



$0\nu\beta\beta$ decay



cosmology



Can current data tell us the neutrino mass ordering?

1. [Hannestad, Schwetz, 2016]: extremely weak (2:1, 3:2) preference for NO (cosmology + [Bergstrom et al., 2015] neutrino oscillation fit) Bayesian approach;
2. [Gerbino et al, 2016]: extremely weak (up to 3:2) preference for NO (cosmology only), Bayesian approach;
3. [Simpson et al., 2017]: strong preference for NO (cosmological limits on $\sum m_\nu$ + constraints on Δm^2_{21} and $|\Delta m^2_{31}|$) Bayesian approach;
4. [Schwetz et al., 2017], "Comment on ..."[Simpson et al., 2017]: effect of prior?
5. [Capozzi et al., 2017]: 2σ preference for NO (cosmology + [Capozzi et al., 2016, updated 2017] neutrino oscillation fit) frequentist approach;
6. [Caldwell et al., 2017] very mild indication for NO (cosmology + neutrinoless double-beta decay + [Esteban et al., 2016] readapted oscillation results) Bayesian approach;
7. [Wang, Xia, 2017]: Bayes factor NO vs IO is not informative (cosmology only).

Can current data tell us the neutrino mass ordering?

1. [Hannestad, Schwetz, 2016]: extremely weak (2:1, 3:2) preference for NO (cosmology + [Bergstrom et al., 2015] neutrino oscillation fit) Bayesian approach;
2. [Gerbino et al, 2016]: extremely weak (up to 3:2) preference for NO (cosmology only), Bayesian approach;
3. [Simpson et al., 2017]: strong preference for NO (cosmological limits on $\sum m_\nu$ + constraints on Δm^2_{21} and $|\Delta m^2_{31}|$) Bayesian approach;
4. [Schwetz et al., 2017], "Comment on ..."[Simpson et al., 2017]: effect of prior?
5. [Capozzi et al., 2017]: 2σ preference for NO (cosmology + [Capozzi et al., 2016, updated 2017] neutrino oscillation fit) frequentist approach;
6. [Caldwell et al., 2017] very mild indication for NO (cosmology + neutrinoless double-beta decay + [Esteban et al., 2016] readapted oscillation results) Bayesian approach;
7. [Wang, Xia, 2017]: Bayes factor NO vs IO is not informative (cosmology only).

Parameterizing neutrino masses

[SG+, JCAP 03 (2018) 11]

[Simpson et al, 2017]

using m_1, m_2, m_3 (A)

[Caldwell et al, 2017]

using $m_{\rm lightest}$, Δm^2_{21}, $|\Delta m^2_{31}|$ (B)

intuition says: (B) is closer to observable quantities! Better than (A)?

Should we use linear or logarithmic priors on m_k ($m_{\rm lightest}$)?

Can data help to select (A) or (B), linear or log?

Parameterizing neutrino masses

[SG+, JCAP 03 (2018) 11]

[Simpson et al, 2017]

using $m_1,\ m_2,\ m_3$ (A)

[Caldwell et al, 2017]

using $m_{\rm lightest},\ \Delta m^2_{21},\ |\Delta m^2_{31}|$ (B)

intuition says: (B) is closer to observable quantities! Better than (A)?

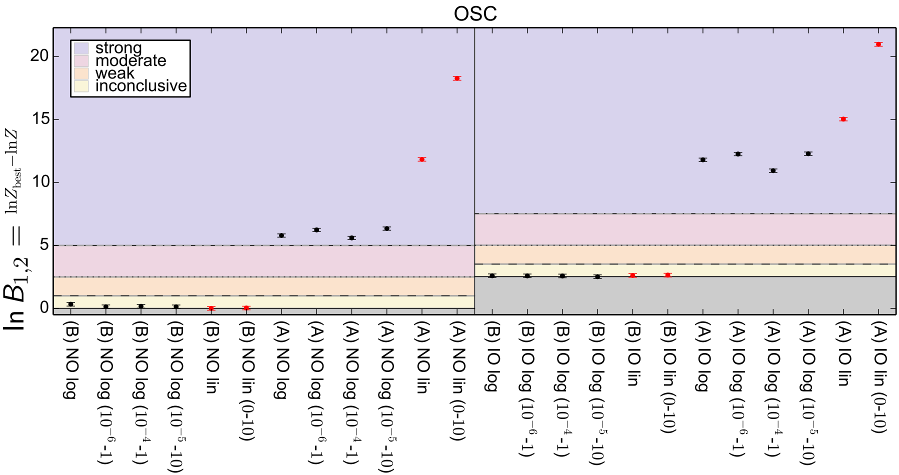
Should we use linear or logarithmic priors on m_k ($m_{\rm lightest}$)?

Can data help to select (A) or (B), linear or log?

Case A			Case B		
Parameter	Prior	Range	Parameter	Prior	Range
m_1/eV	linear log	$0 - 1$ $10^{-5} - 1$	$m_{\rm lightest}/\rm eV$	linear log	$0 - 1$ $10^{-5} - 1$
m_2/eV	linear log	$0 - 1$ $10^{-5} - 1$	$\Delta m^2_{21}/\rm eV^2$	linear	$5 \times 10^{-5} - 10^{-4}$
m_3/eV	linear log	$0 - 1$ $10^{-5} - 1$	$\|\Delta m^2_{31}\|/\rm eV^2$	linear	$1.5 \times 10^{-3} - 3.5 \times 10^{-3}$

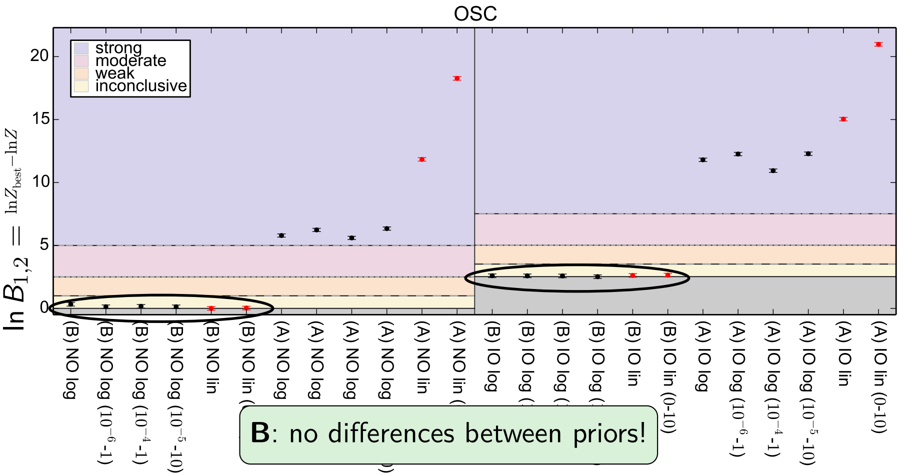
Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

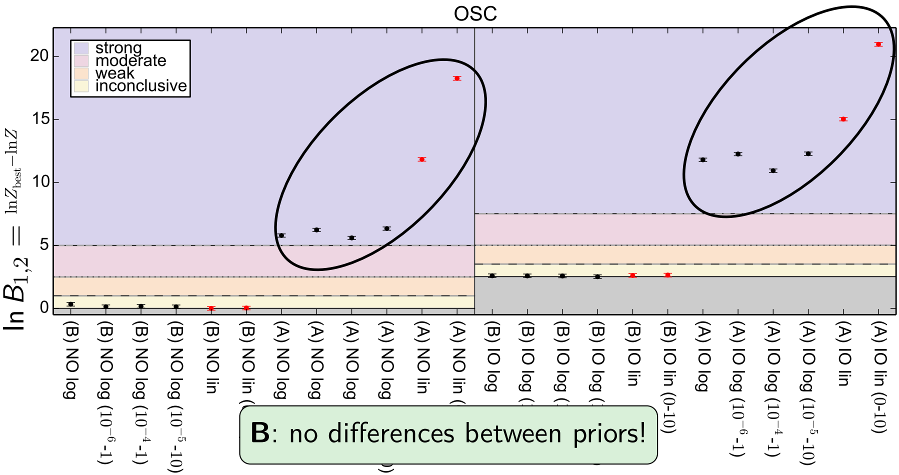


B: no differences between priors!

(no penalty for unconstrained parameter $m_{\rm lightest}$!)

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



B: no differences between priors!

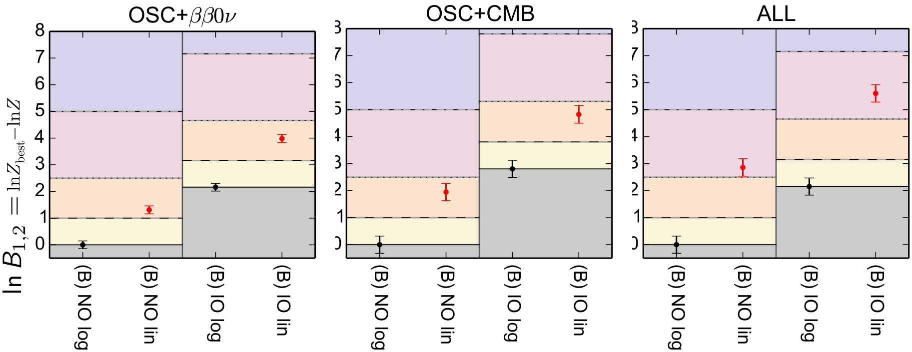
(no penalty for unconstrained parameter $m_{\rm lightest}$!)

A: always strongly disfavored!

(waste of parameter space, no unconstrained parameters due to Δm^2_{i1}!)

Comparing parameterizations/priors

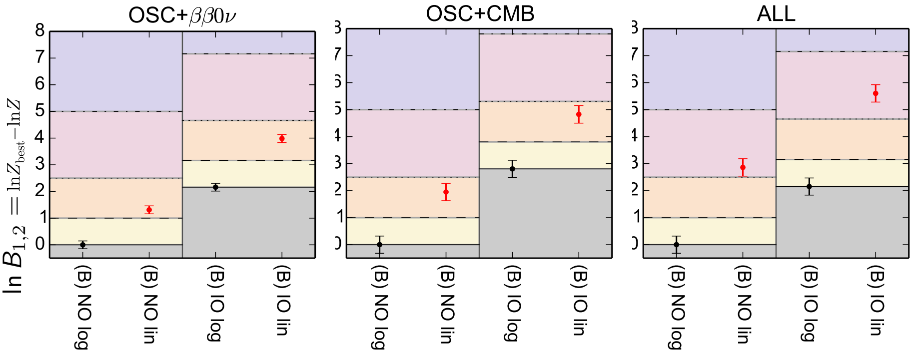
[SG+, JCAP 03 (2018) 11]



compare linear versus **logarithmic**

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

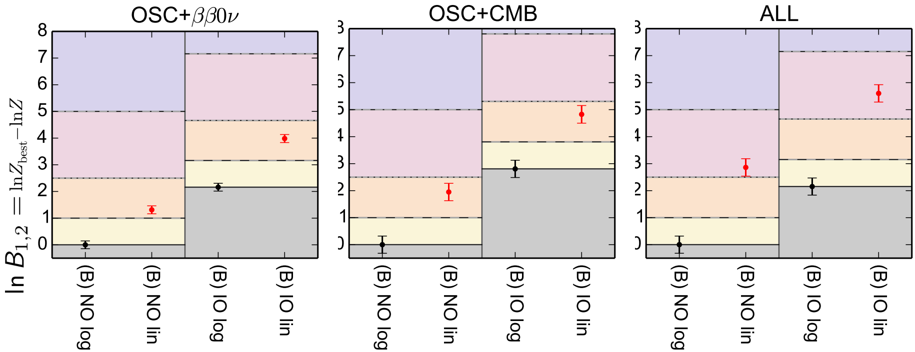


compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



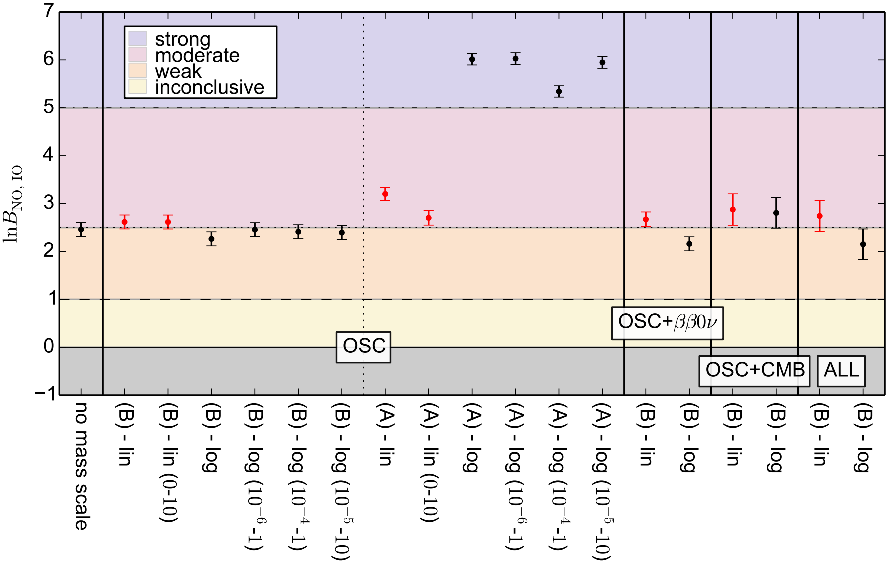
compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

summary: case B, log prior is better!

Comparing the mass orderings

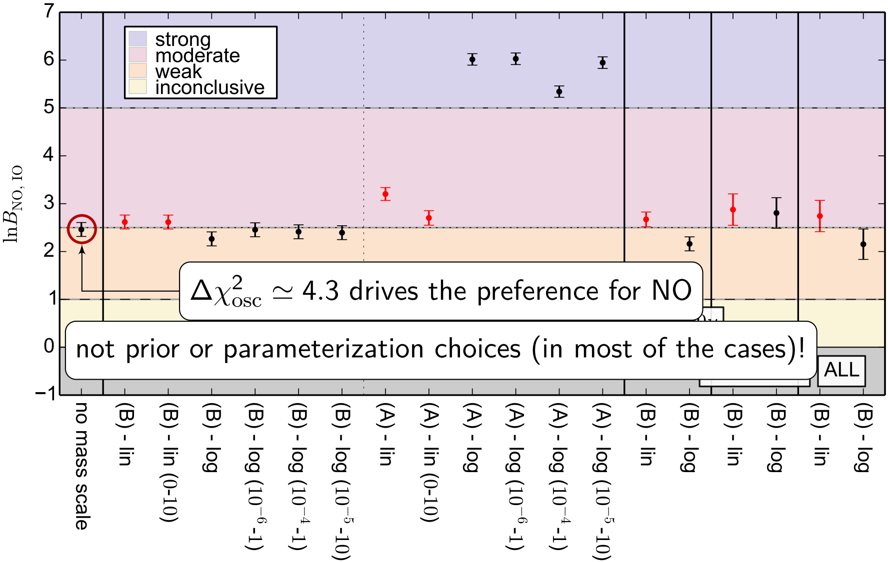
[SG+, JCAP 03 (2018) 11]



Note: only oscillation data until the end of 2017 are included!

Comparing the mass orderings

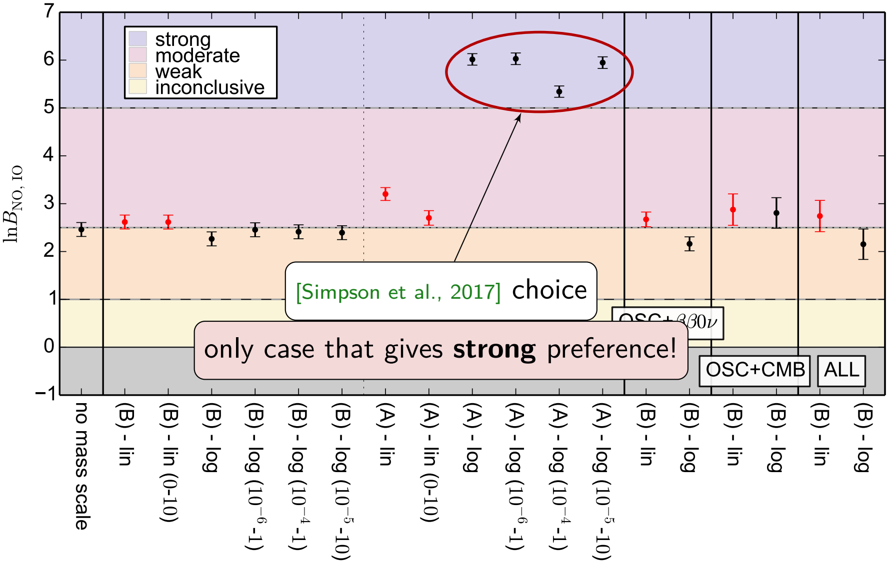
[SG+, JCAP 03 (2018) 11]



Note: only oscillation data until the end of 2017 are included!

Comparing the mass orderings

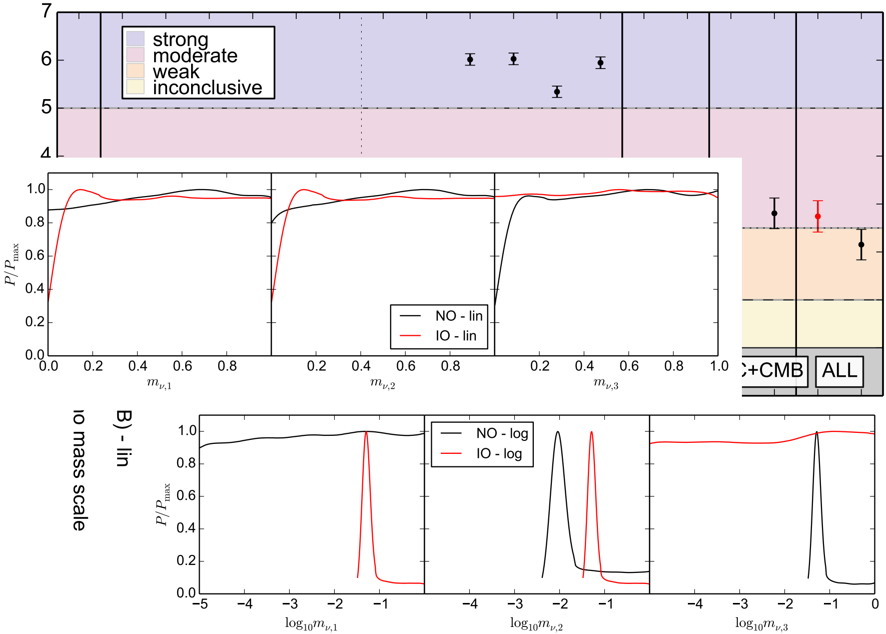
[SG+, JCAP 03 (2018) 11]



Note: only oscillation data until the end of 2017 are included!

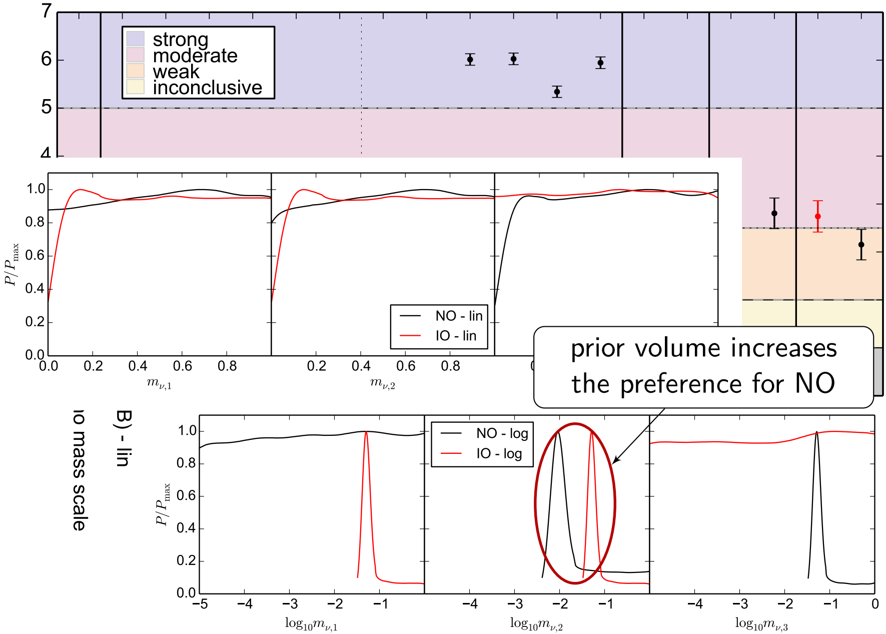
Comparing the mass orderings

[SG+, JCAP 03 (2018) 11]



Comparing the mass orderings

[SG+, JCAP 03 (2018) 11]



Results in 2018

[de Salas+, Frontiers 5 (2018) 36]
http://globalfit.astroparticles.es/

Bayes theorem for models:

$$p(\mathcal{M}|d) \propto Z_{\mathcal{M}} \pi(\mathcal{M})$$

Bayesian evidence:

$$Z_{\mathcal{M}} = \int_{\Omega_{\mathcal{M}}} \mathcal{L}(\theta)\, \pi(\theta)\, d\theta$$

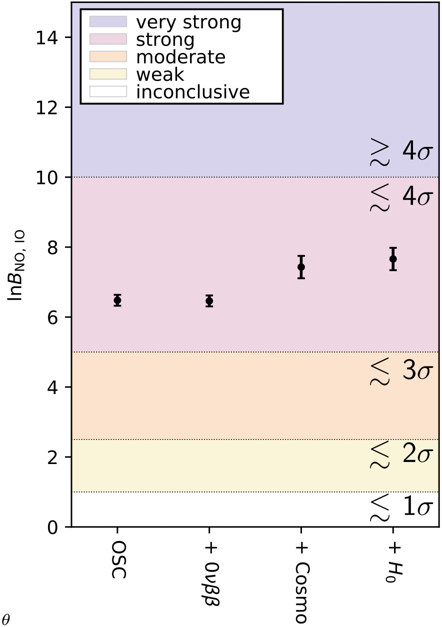
Bayes factor NO vs IO:

$$B_{\rm NO,IO} = Z_{\rm NO}/Z_{\rm IO}$$

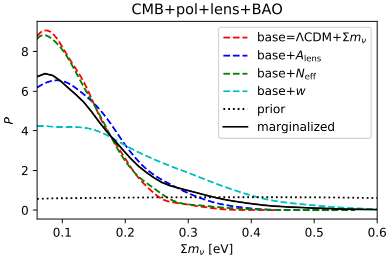
Posterior probability:

$$\begin{aligned} P_{\rm NO} &= B_{\rm NO,IO}/(B_{\rm NO,IO}+1) \\ P_{\rm IO} &= 1/(B_{\rm NO,IO}+1) \end{aligned}$$

$N\sigma$ from $P_{\rm NO} = {\rm erf}(N/\sqrt{2})$



$\pi(\mathcal{M})$ model prior
$p(\mathcal{M}|d)$ model posterior
$\mathcal{L}(\theta)$ likelihood
$\Omega_{\mathcal{M}}$ parameter space, for parameters θ


CMB+pol+lens+BAO
base=ΛCDM+Σm_ν
base+A_lens
base+N_eff
base+w
prior
marginalized
P
Σm_ν [eV]
0
2
4
6
8
0.1
0.2
0.3
0.4
0.5
0.6

Playing with priors

Bayes theorem:

$$p(\theta|d,\mathcal{M}) = \mathcal{L}(\theta)\frac{\pi(\theta|\mathcal{M})}{Z_{\mathcal{M}}}$$

posterior depends on prior!

parameters θ, model $\mathcal{M}$, data d $\pi(\theta|\mathcal{M})$ prior $p(\theta|d,\mathcal{M})$ posterior $\mathcal{L}(\theta)$ likelihood $Z_{\mathcal{M}}$ Bayesian evidence

Playing with priors

Bayes theorem:

$$p(\theta|d,\mathcal{M}) = \mathcal{L}(\theta)\frac{\pi(\theta|\mathcal{M})}{Z_\mathcal{M}}$$

posterior depends on prior!

[Planck 2018]: prior
$0 < \Sigma m_\nu < \mathcal{O}(1)$ eV

strongest upper limit (95%):
$\Sigma m_\nu < 113$ meV
(CMB+lens+BAO+SN)

corresponding to
$\Sigma m_\nu < 53.6$ meV (68%)

below minimum for NO!
does it make sense?

parameters θ, model $\mathcal{M}$, data d $\pi(\theta|\mathcal{M})$ prior $p(\theta|d,\mathcal{M})$ posterior $\mathcal{L}(\theta)$ likelihood $Z_\mathcal{M}$ Bayesian evidence

Playing with priors

Bayes theorem:

$$p(\theta|d,\mathcal{M}) = \mathcal{L}(\theta)\frac{\pi(\theta|\mathcal{M})}{Z_{\mathcal{M}}}$$

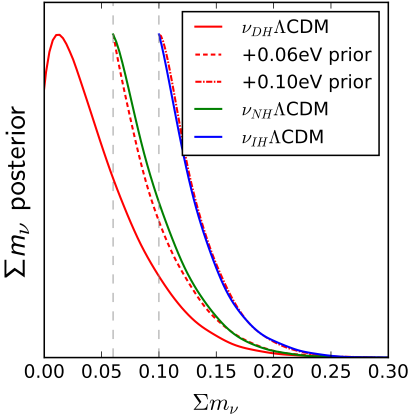
posterior depends on prior!

[Wang+, 2017]

degenerate (DH)
vs normal (NH)
vs inverted (IH) hierarchy

(i.e. change the prior lower bound)

Different limits if you consider simply $\Sigma m_\nu > 0$ or you take into account oscillation results...



parameters θ, model $\mathcal{M}$, data d $\pi(\theta|\mathcal{M})$ prior $p(\theta|d,\mathcal{M})$ posterior $\mathcal{L}(\theta)$ likelihood $Z_{\mathcal{M}}$ Bayesian evidence

Playing with the baseline model

what if we release the assumption of the ΛCDM model?

CMB TT + lens		CMB TT + lens + BAO
CMB TT,TE,EE		**CMB TT,TE,EE + BAO**
	[Planck 2015]	
$\Sigma m_\nu < 0.68$ eV	ΛCDM	$\Sigma m_\nu < 0.25$ eV
$\Sigma m_\nu < 0.49$ eV		$\Sigma m_\nu < 0.17$ eV

Playing with the baseline model

what if we release the assumption of the ΛCDM model?

CMB TT + lens / **CMB TT,TE,EE**		CMB TT + lens + BAO / **CMB TT,TE,EE + BAO**
$\Sigma m_\nu < 0.68$ eV $\Sigma m_\nu < 0.49$ eV	[Planck 2015] ΛCDM	$\Sigma m_\nu < 0.25$ eV $\Sigma m_\nu < 0.17$ eV
-	wCDM	$\Sigma m_\nu < 0.37$ eV [Planck 2015] $\Sigma m_\nu < 0.27$ eV [Wang+, 2016]

free dark energy equation of state $w \neq -1$

Playing with the baseline model

what if we release the assumption of the ΛCDM model?

		CMB TT + lens **CMB TT,TE,EE + BAO**
CMB TT + lens **CMB TT,TE,EE**		CMB TT + lens + BAO **CMB TT,TE,EE + BAO**
$\Sigma m_\nu < 0.68$ eV $\Sigma m_\nu < 0.49$ eV	[Planck 2015] ΛCDM	$\Sigma m_\nu < 0.25$ eV $\Sigma m_\nu < 0.17$ eV
-	wCDM free dark energy equation of state $w \neq -1$	$\Sigma m_\nu < 0.37$ eV [Planck 2015] $\Sigma m_\nu < 0.27$ eV [Wang+, 2016]
-	[Planck 2015] ΛCDM+$A_{\rm lens}$ free phenomenological lensing amplitude $A_{\rm lens} \neq -1$	$\Sigma m_\nu < 0.41$ eV

Playing with the baseline model

what if we release the assumption of the ΛCDM model?

CMB TT + lens **CMB TT,TE,EE**		CMB TT + lens + BAO **CMB TT,TE,EE + BAO**
$\Sigma m_\nu < 0.68$ eV $\Sigma m_\nu < 0.49$ eV	[Planck 2015] ΛCDM	$\Sigma m_\nu < 0.25$ eV $\Sigma m_\nu < 0.17$ eV
-	wCDM free dark energy equation of state $w \neq -1$	$\Sigma m_\nu < 0.37$ eV [Planck 2015] $\Sigma m_\nu < 0.27$ eV [Wang+, 2016]
-	[Planck 2015] ΛCDM+$A_{\rm lens}$ free phenomenological lensing amplitude $A_{\rm lens} \neq -1$	$\Sigma m_\nu < 0.41$ eV
$\Sigma m_\nu < 0.96$ eV	[Di Valentino+, 2015] eCDM 12-parameters cosmological model, ΛCDM based	$\Sigma m_\nu < 0.53$ eV

Marginalize over models?

[SG+, arxiv:1812.05449]

We usually marginalize over parameters:

$$p(\theta|d,\mathcal{M}_0) \propto \int \mathcal{L}(\theta,\psi)p(\theta,\psi|\mathcal{M}_0)d\psi$$

Can we marginalize over models?

Marginalize over models?

[SG+, arxiv:1812.05449]

We usually marginalize over parameters:

$$p(\theta|d,\mathcal{M}_0) \propto \int \mathcal{L}(\theta,\psi)p(\theta,\psi|\mathcal{M}_0)d\psi$$

Can we marginalize over models?

Yes, if we know the model posteriors:

$$p(\theta|d) = \sum_i^N p(\theta|d,\mathcal{M}_i)\, p_i$$

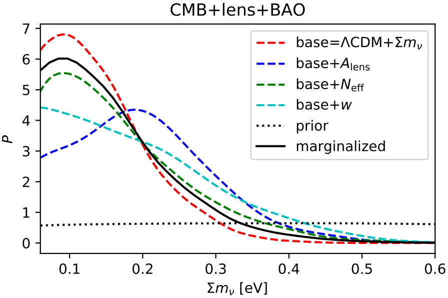
Select a model $\mathcal{M}_0$ and use $p_i = Z_i/(\sum Z_j) = B_{i0}/(\sum B_{j0})$:

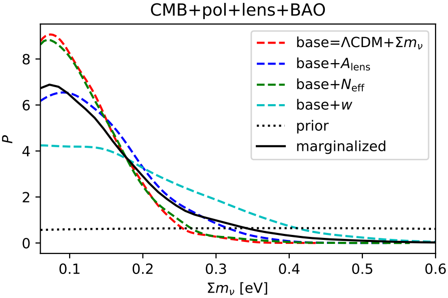
$$p(\theta|d) = \sum_i^N p(\theta|d,\mathcal{M}_i)Z_i \Bigg/ \sum_j^N Z_j$$

$p(\theta|d)$ is a model-marginalized posterior for θ, given the data d

Model-marginalization applied to Σm_ν

[SG+, arxiv:1812.05449]

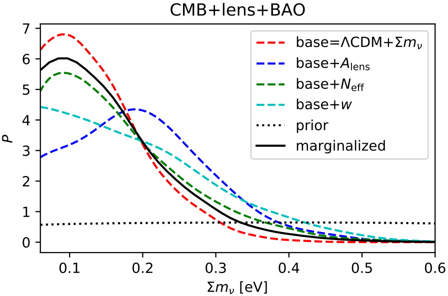


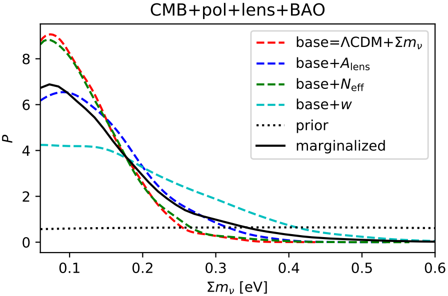


	CMB+lens+BAO		CMB+pol+lens+BAO	
model	$\ln B_{i0}$	Σm_ν [eV]	$\ln B_{i0}$	Σm_ν [eV]
base=ΛCDM+Σm_ν	0.0	< 0.28	0.0	< 0.23
base+$A_{\rm lens}$	-2.6	< 0.38	-2.4	< 0.29
base+$N_{\rm eff}$	-1.5	< 0.37	-2.3	< 0.25
base+w	-1.4	< 0.42	-0.1	< 0.42
marginalized	–	< 0.33	–	< 0.35
p_0	0.65		0.48	

Model-marginalization applied to Σm_ν

[SG+, arxiv:1812.05449]

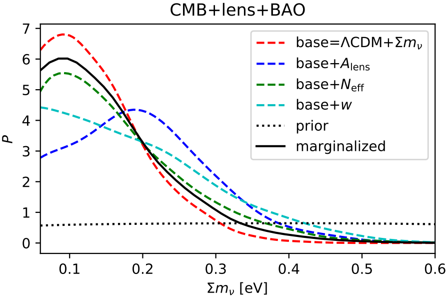




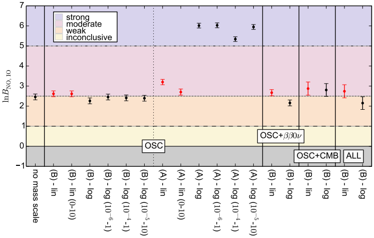
	CMB+lens+BAO		CMB+pol+lens+BAO	
model	$\ln B_{i0}$	Σm_ν [eV]	$\ln B_{i0}$	Σm_ν [eV]
base=ΛCDM+Σm_ν	0.0	< 0.28	0.0	< 0.23
base+$A_{\rm lens}$	-2.6	< 0.38	-2.4	< 0.29
base+$N_{\rm eff}$	-1.5	< 0.37	-2.3	< 0.25
base+w	-1.4	< 0.42	-0.1	< 0.42
marginalized	–	< 0.33	–	< 0.35
p_0	0.65		0.48	

Model-marginalization applied to Σm_ν

[SG+, arxiv:1812.05449]



model	CMB+lens+BAO		CMB+pol+lens+BAO	
	$\ln B_{i0}$	Σm_ν [eV]	$\ln B_{i0}$	Σm_ν [eV]
base=ΛCDM+Σm_ν	0.0	< 0.28	0.0	< 0.23
base+$A_{\rm lens}$	-2.6	< 0.38	-2.4	< 0.29
base+$N_{\rm eff}$	-1.5	< 0.37	-2.3	< 0.25
base+w	-1.4	< 0.42	-0.1	< 0.42
marginalized	–	< 0.33	–	< 0.35
p_0	0.65		0.48	


strong
moderate
weak
inconclusive
$\ln B_{\rm NO,IO}$
OSC
OSC+$\beta\beta0\nu$
OSC+CMB
ALL
no mass scale
(B) - lin
(B) - lin (0-10)
(B) - log
(B) - log (10^{-6}-1)
(B) - log (10^{-4}-1)
(B) - log (10^{-5}-10)
(A) - lin
(A) - lin (0-10)
(A) - log
(A) - log (10^{-6}-1)
(A) - log (10^{-4}-1)
(A) - log (10^{-5}-10)
(B) - lin
(B) - log
(B) - lin
(B) - log
(B) - lin
(B) - log

Conclusions

1. studying the χ^2 with regular grid only feasible for few parameters, Monte Carlo otherwise

2. Combined analyses will be more and more important in the future...number of parameters increase!

3. Prior dependence is intrinsic of Bayesian statistics! Careful when choosing the parameterizations/priors! *Do not influence the results with your choice...*

4. Constraints also depend on the model you define... **Marginalize over models** is possible!

Conclusions

1. studying the χ^2 with regular grid only feasible for few parameters, Monte Carlo otherwise

2. Combined analyses will be more and more important in the future...number of parameters increase!

3. Prior dependence is intrinsic of Bayesian statistics!
Careful when choosing the parameterizations/priors!
Do not influence the results with your choice...

4. Constraints also depend on the model you define...
Marginalize over models is possible!

Thank you for the attention!