IFIC
INSTITUT DE FÍSICA CORPUSCULAR

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

VNIVERSITAT DE VALÈNCIA

EXCELENCIA SEVERO OCHOA

GENERALITAT VALENCIANA

European Commission
Horizon 2020
European Union funding
for Research & Innovation

Stefano Gariazzo

IFIC, Valencia (ES)
CSIC – Universitat de Valencia

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

Bayesian statistics in neutrino cosmology: towards model-independent constraints

On prior effects, marginalization, model selection, quantifying tensions, . . .

Oscar Klein Center, Stockholm (SE), 19/06/2019

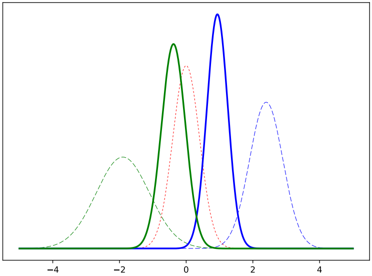
1 ***Basics of Bayesian probability***
- Parameter inference
- Bayesian model comparison
- Information gain, model dimensionality and quantifying tensions

2 ***Cosmological tensions***
- Local Universe versus CMB
- Quantifying tensions in Bayesian statistics

3 ***Neutrino mass ordering***
- How to constrain the mass ordering
- Subtleties in the Bayesian analysis
- Constraints on the mass ordering

4 ***Neutrino masses from cosmology***
- The current status
- One step forward
- Non-probabilistic limits

5 ***Conclusions***



Probability

[Trotta, arxiv:0803.4089]

What is probability?

a frequency

"the number of times
the event occurs over
the total number of trials, in
the limit of an infinite series
of equiprobable repetitions"

another subtle point:
"randomness" of the trial series

what is really "random"?

do we properly know the initial
state (and do not cheat)?

[Trotta, arxiv:0803.4089]

What is probability?

a frequency

"the number of times
the event occurs over
the total number of trials, in
the limit of an infinite series
of equiprobable repetitions"

another subtle point:
"randomness" of the trial series

what is really "random"?

do we properly know the initial
state (and do not cheat)?

Probability

[Trotta, arxiv:0803.4089]

What is probability?

a frequency

"the number of times
the event occurs over
the total number of trials, in
the limit of an infinite series
of equiprobable repetitions"

another subtle point:
"randomness" of the trial series

what is really "random"?

do we properly know the initial
state (and do not cheat)?

Probability

[Trotta, arxiv:0803.4089]

What is probability?

a frequency

"the number of times
the event occurs over
the total number of trials, in
the limit of an infinite series
of equiprobable repetitions"

another subtle point:
"randomness" of the trial series

what is really "random"?

do we properly know the initial
state (and do not cheat)?

a degree of belief

"probability is
a measure of the degree
of belief about a preposition"

Probability

[Trotta, arxiv:0803.4089]

What is probability?

a frequency	a degree of belief
"the number of times the event occurs over the total number of trials, in the limit of an infinite series of equiprobable repetitions"	"probability is a measure of the degree of belief about a preposition"

another subtle point:
"randomness" of the trial series

what is really "random"?

do we properly know the initial state (and do not cheat)?

Advantages:

- recovers frequentist on the long run;
- can be applied when frequentist cannot;
- no need to assume a distribution of possible data;
- deals effortlessly with nuisance parameters (*marginalization*);
- relies on *prior information*.

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

parameter inference

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

parameter inference

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:

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

model comparison

(Bayesian) Parameter inference

Parameter inference = what we learn on the parameters, given:

data d

model $\mathcal{M}_0$

build $\mathcal{L}(\theta)$ from d & $\mathcal{M}_0$

parameters θ (& priors)

(physical ϕ + nuisance ψ)

Full posterior:

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

(Bayesian) Parameter inference

Parameter inference = what we learn on the parameters, given:

data d → build $\mathcal{L}(\theta)$ from d & $\mathcal{M}_0$

model $\mathcal{M}_0$ → parameters θ (& priors)

(physical ϕ + nuisance ψ)

Full posterior:

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

Marginalize over nuisance to obtain posterior for physical:

$$p(\phi|d,\mathcal{M}_0) \propto \int_{\Omega_\psi} \mathcal{L}(\phi,\psi)\pi(\phi,\psi|\mathcal{M}_0)d\psi$$

marginalize over all the parameters except one (two)

→ 1D (2D) posterior

Credible intervals from the posterior

Credible interval α?

range of values within which an unobserved parameter value falls with a particular subjective probability α

Credible intervals from the posterior

Credible interval α?

range of values within which an unobserved parameter value falls
with a particular subjective probability α

Analogous to frequentist confidence intervals α

Bayesian credible interval:

- bounds as fixed;
- estimated parameter as a random variable.

Frequentist confidence interval:

- bounds as random variables;
- estimated parameter as fixed value.

Credible intervals from the posterior

Credible interval α?

range of values within which an unobserved parameter value falls with a particular subjective probability α

Analogous to frequentist confidence intervals α

Bayesian credible interval:

- bounds as fixed;
- estimated parameter as a random variable.

Frequentist confidence interval:

- bounds as random variables;
- estimated parameter as fixed value.

Credible intervals are not uniquely defined!

highest posterior density interval: narrowest interval, includes values of highest probability density

equal-tailed interval: same probability of being below or above the interval

interval for which the mean is the central point

Credible intervals from the posterior

Credible interval α?

range of values within which an unobserved parameter value falls with a particular subjective probability α

Analogous to frequentist confidence intervals α

Bayesian credible interval:

- bounds as fixed;
- estimated parameter as a random variable.

Frequentist confidence interval:

- bounds as random variables;
- estimated parameter as fixed value.

Credible intervals are not uniquely defined!

highest posterior density interval: narrowest interval, includes values of highest probability density

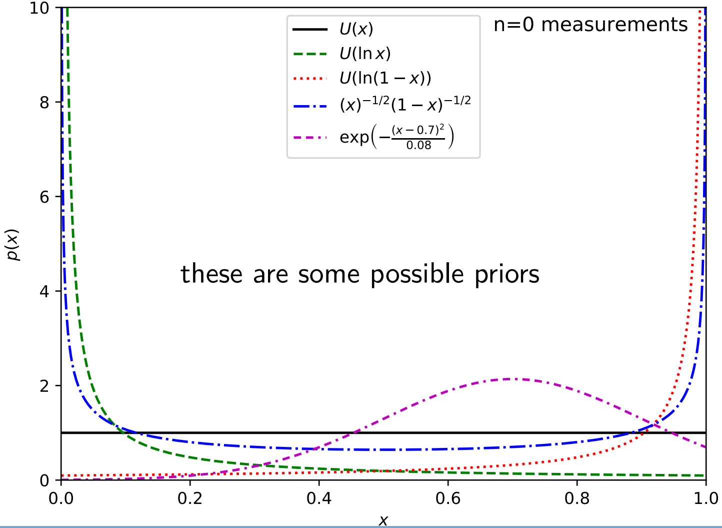
equal-tailed interval: same probability of being below or above the interval

interval for which the mean is the central point

Prior dependence in parameter estimation - I

example: need to measure $0 < x < 1$

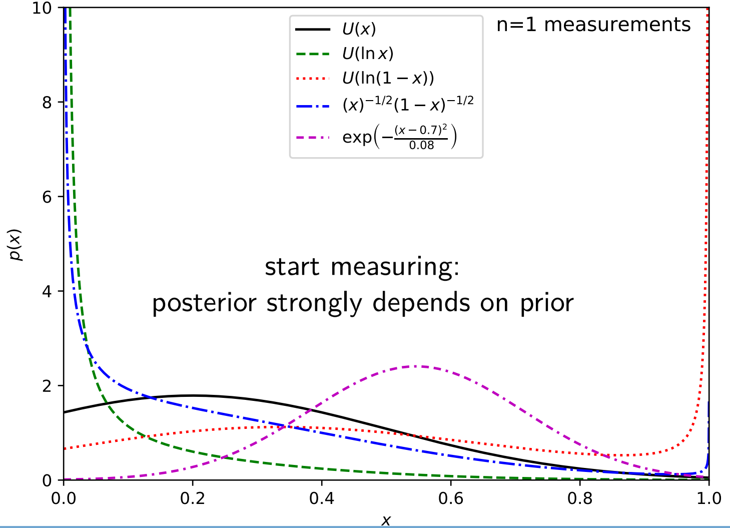
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2 \cdot 0.3^2)]$



Prior dependence in parameter estimation - I

example: need to measure $0 < x < 1$

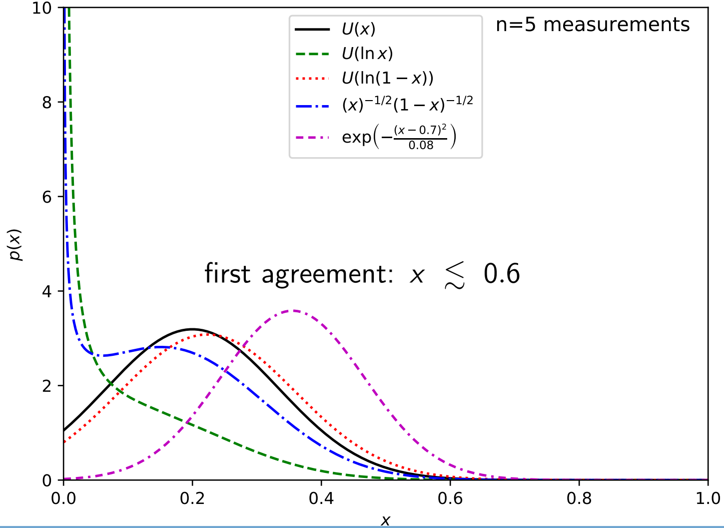
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



Prior dependence in parameter estimation - I

example: need to measure $0 < x < 1$

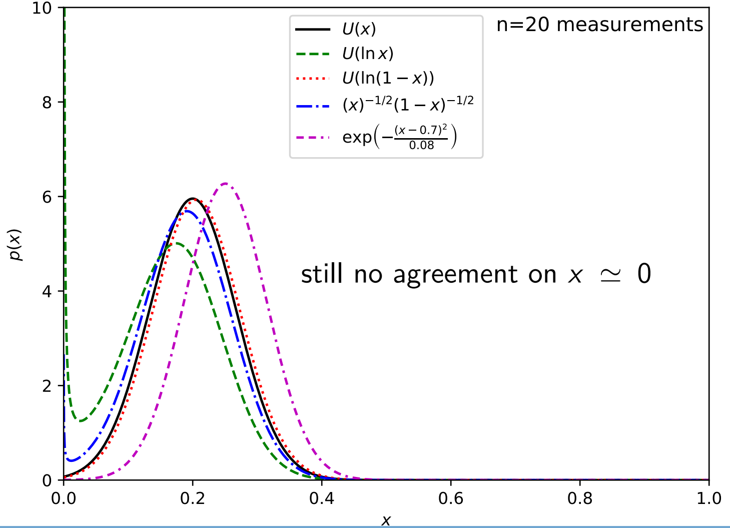
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



Prior dependence in parameter estimation - I

example: need to measure $0 < x < 1$

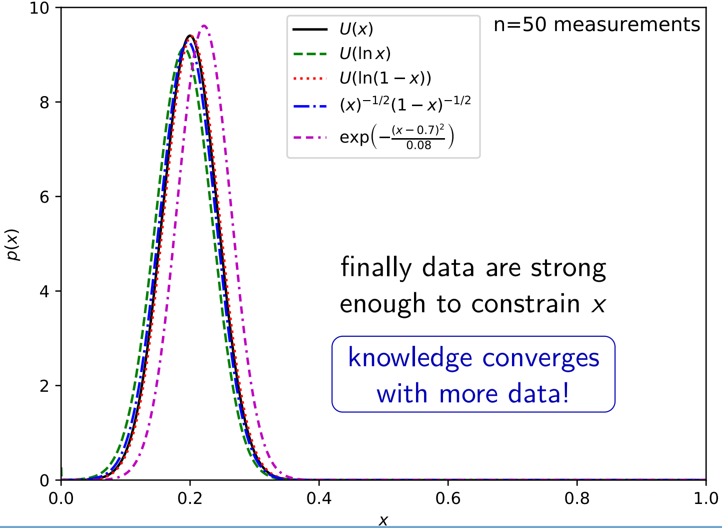
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



Prior dependence in parameter estimation - I

example: need to measure $0 < x < 1$

likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



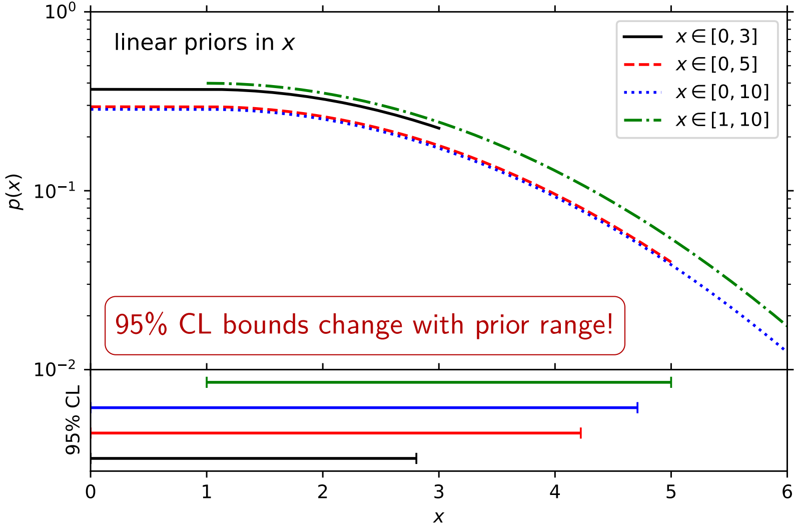
finally data are strong enough to constrain x

knowledge converges with more data!

Prior dependence in parameter estimation - II

other example: need to measure $x > 0$ (Σm_ν?)

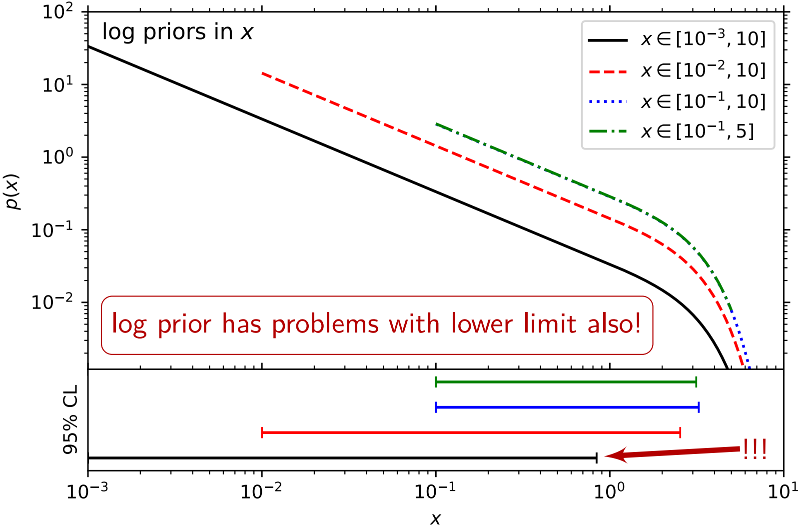
likelihood $\mathcal{L}(x) \propto \exp[-(x-1)^2/(2\cdot 1^2)]$ for $x > 1$, constant otherwise



Prior dependence in parameter estimation - II

other example: need to measure $x > 0$ (Σm_ν?)

likelihood $\mathcal{L}(x) \propto \exp[-(x-1)^2/(2\cdot 1^2)]$ for $x > 1$, constant otherwise



Bayesian evidence

"Bayesian evidence" or "Marginal likelihood"

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

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:

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

← proportional to constant that depends only on data

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

Bayes factor

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

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

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 $[\pi(\mathcal{M}_1) = \pi(\mathcal{M}_2)]$,
$B_{1,2}$ 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 → Occam's razor ← Predictivity

the simplest theory that fits data is preferred

model with more parameters → better fit (usually)

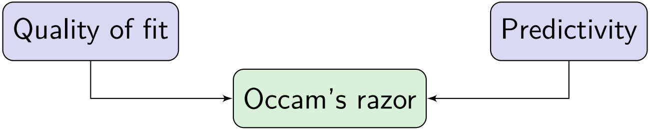
→ are all the parameters needed?

Bayes factor penalizes unnecessarily complex models!

Occam's razor

what the Bayesian model comparison tells us?

Best model strikes optimum balance between



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

Prior dependence in the Bayesian evidence

Bayesian evidences depend on priors!

likelihood: $\mathcal{L}(x) \propto \begin{cases} 1 & \text{for } x \leq 1 \\ \exp[-(x-1)^2/(2\cdot 1^2)] & \text{for } x > 1 \end{cases}$

linear prior		log prior	
range	Z	range	Z
$0 \leq x \leq 3$	0.180	$10^{-3} \leq x \leq 10$	0.192
$0 \leq x \leq 5$	0.135	$10^{-2} \leq x \leq 10$	0.172
$0 \leq x \leq 10$	0.070	$10^{-1} \leq x \leq 10$	0.151
$1 \leq x \leq 10$	0.056	$10^{-1} \leq x \leq 5$	0.177

linear prior $x \in [a, b]$ is $\propto 1/(b-a)$

irrelevant for Bayes factor
if the compared models
have the parameter x in common

Prior dependence in the Bayesian evidence

Bayesian evidences depend on priors!

likelihood: $\mathcal{L}(x) \propto \begin{cases} 1 & \text{for } x \leq 1 \\ \exp[-(x-1)^2/(2\cdot 1^2)] & \text{for } x > 1 \end{cases}$

linear prior		log prior	
range	Z	range	Z
$0 \leq x \leq 3$	0.180	$10^{-3} \leq x \leq 10$	0.192
$0 \leq x \leq 5$	0.135	$10^{-2} \leq x \leq 10$	0.172
$0 \leq x \leq 10$	0.070	$10^{-1} \leq x \leq 10$	0.151
$1 \leq x \leq 10$	0.056	$10^{-1} \leq x \leq 5$	0.177

linear prior $x \in [a, b]$ is $\propto 1/(b-a)$

irrelevant for Bayes factor
if the compared models
have the parameter x in common

towards Lindley's paradox:

use $\pi(x) \propto \exp[-x^2/(2\Sigma^2)]$,
$\mathcal{L}(x) \propto \exp[-(x - N\sigma_t)^2/(2\sigma^2)]$,
with $\sigma_t = \sqrt{\sigma^2 + \Sigma^2}$

$$Z = \exp(-N^2/2) \Big/ \left(\sqrt{2\pi}\,\sigma_t\right)$$

Prior dependence in the Bayesian evidence

Bayesian evidences depend on priors!

likelihood: $\mathcal{L}(x) \propto \begin{cases} 1 & \text{for } x \leq 1 \\ \exp[-(x-1)^2/(2\cdot 1^2)] & \text{for } x > 1 \end{cases}$

max evidence for a given likelihood $\mathcal{L}(x)$?

Select a Dirac delta centered on the $\hat{x}$
that gives the maximum of the likelihood

useful estimate of the **max Bayes factor**, in particular for nested models

$\mathcal{M}_1$: free x
$\mathcal{M}_0$: $\mathcal{M}_1 |\ x = x_0$

$$B_{01} = \frac{\mathcal{L}(x_0)}{\int \mathrm{d}x\, \mathcal{L}(x)\, \pi(x)} \geq \frac{\mathcal{L}(x_0)}{\mathcal{L}(\hat{x})} = \frac{\mathcal{L}(x_0)}{\int \mathrm{d}x\, \mathcal{L}(x)\, \delta(x-\hat{x})}$$

maximum likelihood ratio

you will never find a prior that gives a better B_{01} than this!

useful for prior-independent estimates of B_{01}

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
$\in [5.0, 10]$	$(150 - 2.2 \times 10^4):1$	$2.7 - 4.1$	strong
$\in [10, 15]$	$(2.2 \times 10^4 - 3.3 \times 10^6):1$	$4.1 - 5.1$	very strong
> 15	$> 3.3 \times 10^6 : 1$	> 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+, PRD 99 (2019) 021301]

Assume N models, equal model prior probabilities:

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

Compute model posterior probabilities:

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

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

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

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

the sum includes $B_{00} = 1$

How to compute the model posterior

[SG+, PRD 99 (2019) 021301]

Assume N models, equal model prior probabilities:

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

Compute model posterior probabilities:

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

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

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

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

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+, PRD 99 (2019) 021301]

Assume N models, equal model prior probabilities:

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

Compute model posterior probabilities:

$p_i \equiv p(\mathcal{M}_i|d)$ $\qquad p_i = A\pi_i Z_i$ with A 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: $N = 2$

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

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

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!

Model posterior with many models

$$p_i = Z_i \Big/ \sum_j^N Z_j = B_{i0} \Big/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

Model posterior with many models

$$p_i = Z_i \left/ \sum_j^N Z_j = B_{i0} \right/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

Model posterior with many models

$$p_i = Z_i \Bigg/ \sum_j^N Z_j = B_{i0} \Bigg/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

ΛCDM ⟵ this will probably be the favorite one

Model posterior with many models

$$p_i = Z_i \Bigg/ \sum_j^N Z_j = B_{i0} \Bigg/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

ΛCDM

+1 parameter

$+r$ $\quad +\Sigma m_\nu$ $\quad +N_{\rm eff}$ $\quad +w$ $\quad +\Omega_k$ $\quad +Y_p$ $\quad +A_{\rm lens}$ $\quad +\ldots$

Model posterior with many models

$$p_i = Z_i \Big/ \sum_j^N Z_j = B_{i0} \Big/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

ΛCDM

+1 parameter

$+r$ $\quad +\Sigma m_\nu$ $\quad +N_{\rm eff}$ $\quad +w$ $\quad +\Omega_k$ $\quad +Y_p$ $\quad +A_{\rm lens}$ $\quad +\ldots$

+2 parameters

$+\Sigma m_\nu + N_{\rm eff}$ $\quad +N_{\rm eff} + m_s^{\rm eff}$ $\quad +w_0 + w_a$ $\quad +\alpha_s + \beta_s$ $\quad +Y_p + N_{\rm eff}$

$+r + \alpha_s$ $\quad +A_{\rm lens} + \Sigma m_\nu$ $\quad +\alpha_s + N_{\rm eff}$ $\quad +\ldots$

Model posterior with many models

$$p_i = Z_i \Big/ \sum_j^N Z_j = B_{i0} \Big/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

ΛCDM

+1 parameter

$+r$ $+\Sigma m_\nu$ $+N_{\rm eff}$ $+w$ $+\Omega_k$ $+Y_p$ $+A_{\rm lens}$ $+\ldots$

+2 parameters

$+\Sigma m_\nu + N_{\rm eff}$ $+N_{\rm eff} + m_s^{\rm eff}$ $+w_0 + w_a$ $+\alpha_s + \beta_s$ $+Y_p + N_{\rm eff}$

$+r + \alpha_s$ $+A_{\rm lens} + \Sigma m_\nu$ $+\alpha_s + N_{\rm eff}$ $+\ldots$

+3 parameters (and so on...)

Model posterior with many models

$$p_i = Z_i \Big/ \sum_j^N Z_j = B_{i0} \Big/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models, starting from the simplest one under consideration (e.g. ΛCDM in cosmology) and then extending it

ΛCDM

+1 parameter

$+r$ $+\Sigma m_\nu$ $+N_{\rm eff}$ $+w$ ... $+\ldots$

+2 parameters

$+\Sigma m_\nu + N_{\rm eff}$ $+N_{\rm eff} + m_s^{\rm eff}$ $+w_0 + w_a$ ... $N_{\rm eff}$

$+r + \alpha_s$ $+A_{\rm lens} + \Sigma m_\nu$ $+\alpha_s + N_{\rm eff}$ $+\ldots$

+3 parameters (and so on...)

Complexity increases: more and more penalized by Occam's razor

Model posterior with many models

$$p_i = Z_i \Bigg/ \sum_j^N Z_j = B_{i0} \Bigg/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models,
starting from the simplest one under consideration
(e.g. ΛCDM in cosmology) and then extending it

ΛCDM

$+1$ parameter

the number of relevant models is not infinite!

$+\Sigma m_\nu + N_{\rm eff}$ $+N_{\rm eff} + m_s$ $+w_0 + w_a$ $+\alpha_s + \beta_s$ $+r_p + N_{\rm eff}$

$+r + \alpha_s$ $+A_{\rm lens} + \Sigma m_\nu$ $+\alpha_s + N_{\rm eff}$ $+\ldots$

$+3$ parameters (and so on...)

Model posterior with many models

$$p_i = Z_i \Big/ \sum_j^N Z_j = B_{i0} \Big/ \sum_j^N B_{j0}$$

Do the result depend on N?

Does $p_0 \to 0$ when $N \to \infty$?

in principle one should consider all the possible models, starting from the simplest one under consideration (e.g. ΛCDM in cosmology) and then extending it

ΛCDM

$+1$ parameter

the number of relevant models is not infinite!

$+\sum m_\nu + N_{\rm eff}$ $+N_{\rm eff} + m_s$ $+w_0 + w_a$ $+\alpha_s + \beta_s$ $+r_p + N_{\rm eff}$

but beware: unconstrained parameters...

$+3$ parameters (and so on...)

Kullback-Leibler divergence

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

Shannon information: $\mathcal{I}(\theta) = \log \frac{p(\theta)}{\pi(\theta)}$

parameters θ
prior $\pi(\theta)$
posterior $p(\theta)$

- Encodes information gain provided by data for a given θ
- Additive for independent parameters:
  $\mathcal{I}(\theta_1, \theta_2) = \log \frac{p(\theta_1)p(\theta_2)}{\pi(\theta_1)\pi(\theta_2)} = \mathcal{I}(\theta_1) + \mathcal{I}(\theta_2)$

Kullback-Leibler divergence

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

Shannon information: $\mathcal{I}(\theta) = \log \frac{p(\theta)}{\pi(\theta)}$

parameters θ
prior $\pi(\theta)$
posterior $p(\theta)$

- Encodes information gain provided by data for a given θ
- Additive for independent parameters:
  $\mathcal{I}(\theta_1, \theta_2) = \log \frac{p(\theta_1)p(\theta_2)}{\pi(\theta_1)\pi(\theta_2)} = \mathcal{I}(\theta_1) + \mathcal{I}(\theta_2)$

Kullback-Leibler divergence: $\mathcal{D}(\mathcal{M}) = \int_{\Omega_\mathcal{M}} p(\theta) \log \frac{p(\theta)}{\pi(\theta)} \mathrm{d}\theta$

- Average Shannon information, weighted over the posterior
- Total information provided by data, independent of parameter values
- Also additive for independent parameters

Kullback-Leibler divergence

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

Shannon information: $\mathcal{I}(\theta) = \log \frac{p(\theta)}{\pi(\theta)}$

parameters θ
prior $\pi(\theta)$
posterior $p(\theta)$

- Encodes information gain provided by data for a given θ
- Additive for independent parameters:
  $\mathcal{I}(\theta_1, \theta_2) = \log \frac{p(\theta_1)p(\theta_2)}{\pi(\theta_1)\pi(\theta_2)} = \mathcal{I}(\theta_1) + \mathcal{I}(\theta_2)$

Kullback-Leibler divergence: $\mathcal{D}(\mathcal{M}) = \int_{\Omega_{\mathcal{M}}} p(\theta) \log \frac{p(\theta)}{\pi(\theta)} d\theta$

- Average Shannon information, weighted over the posterior
- Total information provided by data, independent of parameter values
- Also additive for independent parameters

prior dependent

depend both on prior shape *and range*. . .

Model dimensionality

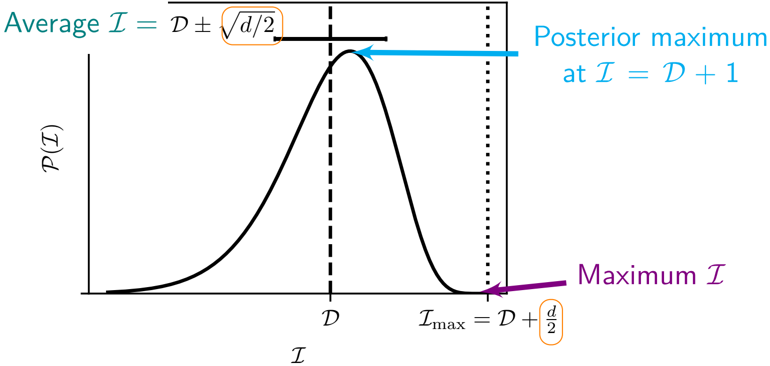
[Handley+, arxiv:1903.06682]

How many parameters are constrained by data?

$$\frac{d}{2} = \int p(\theta) \left(\log \frac{p(\theta)}{\pi(\theta)} - \mathcal{D} \right)^2 \mathrm{d}\theta = \langle \mathcal{I}^2 \rangle_p - \langle \mathcal{I} \rangle_p^2 = \langle \log(\mathcal{L})^2 \rangle_p - \langle \log \mathcal{L} \rangle_p^2$$

Variance of the Shannon information

Consider a Gaussian:



($d = 16$)

Model dimensionality

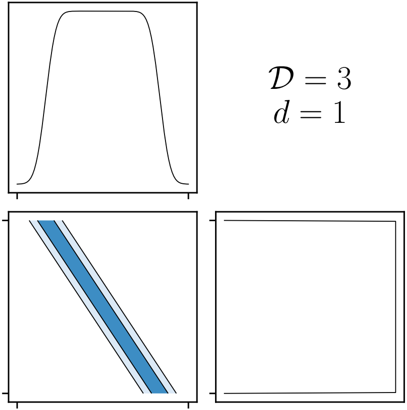
[Handley+, arxiv:1903.06682]

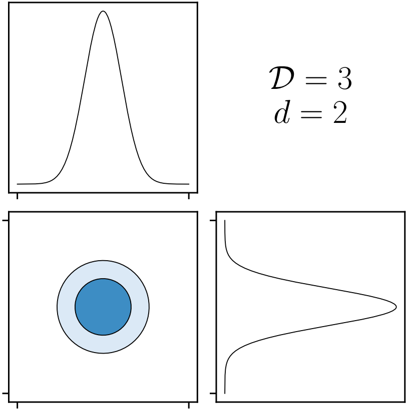
How many parameters are constrained by data?

$$\frac{d}{2} = \int p(\theta)\left(\log\frac{p(\theta)}{\pi(\theta)} - \mathcal{D}\right)^2 \mathrm{d}\theta = \langle \mathcal{I}^2\rangle_p - \langle \mathcal{I}\rangle_p^2 = \langle \log(\mathcal{L})^2\rangle_p - \langle \log\mathcal{L}\rangle_p^2$$

Variance of the Shannon information

Adds information over the KL divergence (mean of Shannon information)





Model dimensionality

[Handley+, arxiv:1903.06682]

How many parameters are constrained by data?

$$\frac{d}{2} = \int p(\theta) \left(\log \frac{p(\theta)}{\pi(\theta)} - \mathcal{D}\right)^2 \mathrm{d}\theta = \langle \mathcal{I}^2 \rangle_p - \langle \mathcal{I} \rangle_p^2 = \langle \log(\mathcal{L})^2 \rangle_p - \langle \log \mathcal{L} \rangle_p^2$$

Variance of the Shannon information

Adds information over the KL divergence (mean of Shannon information)

Applications:

Common parameters:

$$d_{A \cap B} = d_A + d_B - d_{AB}$$

(number of parameters constrained by both experiments A and B)

Model priors with penalisation based on d

$$\pi_i(\lambda) = \lambda e^{-\lambda d_i}$$

and

$$\log B_{ij} = (\log Z_i - \lambda d_i) - (\log Z_j - \lambda d_j)$$

e.g. $\lambda = 1$

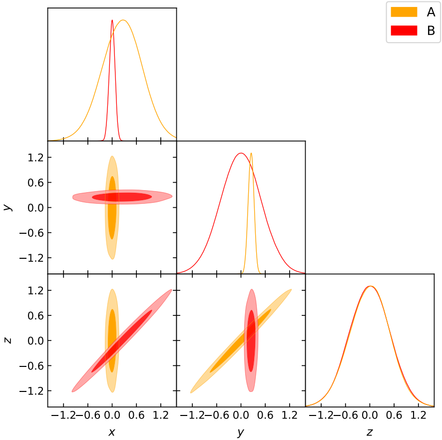
(favor models with fewer parameters)

Tensions, R test and suspiciousness

[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

How to determine if they are in agreement?



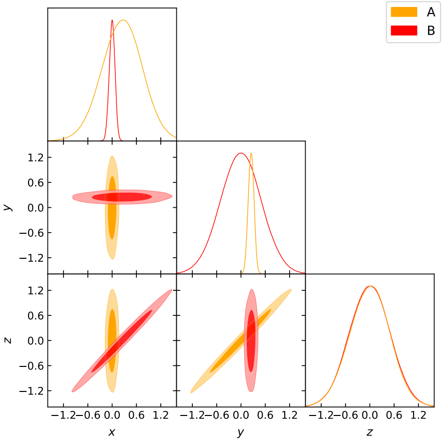
they seem in agreement...

Tensions, R test and suspiciousness

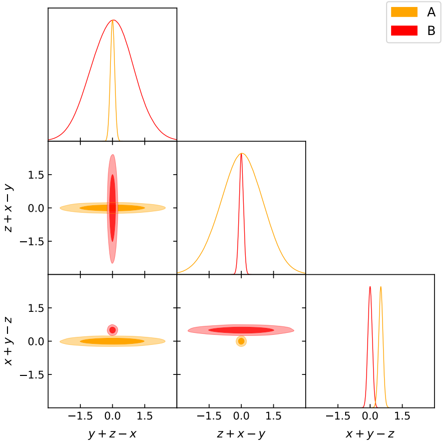
[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

How to determine if they are in agreement?



they seem in agreement...



Not anymore after a linear transformation!

Tensions, R test and suspiciousness

[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

$$R = \frac{Z_{AB}}{Z_A Z_B} = \frac{P(A,B)}{P(A)P(B)} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

Tensions, R test and suspiciousness

[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

$$R = \frac{Z_{AB}}{Z_A Z_B} = \frac{P(A,B)}{P(A)P(B)} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

B (A) has strengthened/weakened our confidence in A (B) by a factor R

$R > 1$: consistency — $R < 1$: inconsistency

Tensions, R test and suspiciousness

[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

$$R = \frac{Z_{AB}}{Z_A Z_B} = \frac{P(A,B)}{P(A)P(B)} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

B (A) has strengthened/weakened our confidence in A (B) by a factor R

$R > 1$: consistency — $R < 1$: inconsistency

Rewrite: $$R = \int \frac{\mathcal{L}_A \mathcal{L}_B \pi}{Z_A Z_B} \mathrm{d}\theta = \int \frac{p_A p_B}{\pi} \mathrm{d}\theta \qquad (p_i = \mathcal{L}_i \pi / Z_i)$$

Depends on the prior of *shared* parameters! (not of nuisance)

Tensions, R test and suspiciousness

[Handley+, arxiv:1902.04029]

Consider independent datasets A, B

$$R = \frac{Z_{AB}}{Z_A Z_B} = \frac{P(A,B)}{P(A)P(B)} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

B (A) has strengthened/weakened our confidence in A (B) by a factor R

$R > 1$: consistency — $R < 1$: inconsistency

Rewrite: $R = \int \frac{\mathcal{L}_A \mathcal{L}_B \pi}{Z_A Z_B} \mathrm{d}\theta = \int \frac{p_A p_B}{\pi} \mathrm{d}\theta$ $\quad$ ($p_i = \mathcal{L}_i \pi / Z_i$)

Depends on the prior of *shared* parameters! (not of nuisance)

Divide R in two parts: information and suspiciousness

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB}$$

prior dependent!

$S = R/I$ or $\log S = \log R - \log I$

Interpreting the suspiciousness

[Handley+, arxiv:1902.04029]

$$S = R/I \text{ or } \log S = \log R - \log I$$

$$\log S = \log Z_{AB} + \mathcal{D}_{AB} - (\log Z_A + \mathcal{D}_A) - (\log Z_B + \mathcal{D}_B)$$

prior independent! (opposite prior dependence for $\log Z$ and $\mathcal{D}$)

Interpreting the suspiciousness

[Handley+, arxiv:1902.04029]

$$S = R/I \text{ or } \log S = \log R - \log I$$

$$\log S = \log Z_{AB} + \mathcal{D}_{AB} - (\log Z_A + \mathcal{D}_A) - (\log Z_B + \mathcal{D}_B)$$

prior independent! (opposite prior dependence for $\log Z$ and $\mathcal{D}$)

Consider a Gaussian: d dimensions, μ central value, Σ covariance, V_π prior volume

- $\log R = -\frac{1}{2}(\mu_A - \mu_B)(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B) - \frac{1}{2}\log|2\pi(\Sigma_A + \Sigma_B)| + \log V_\pi$
- $\log I = -\frac{d}{2} - \frac{1}{2}\log|2\pi(\Sigma_A + \Sigma_B)| + \log V_\pi$
- $\log S = \frac{d}{2} - \frac{1}{2}(\mu_A - \mu_B)(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B)$

one can demonstrate that $d - 2\log S$ has a χ^2_d distribution!

expected value: $\log S = 0 \pm \sqrt{d/2}$

$\log S \ll 0$ suspicious discordance

$\log S \gg 0$ suspicious concordance

Interpreting the suspiciousness

[Handley+, arxiv:1902.04029]

$$S = R/I \text{ or } \log S = \log R - \log I$$

$$\log S = \log Z_{AB} + \mathcal{D}_{AB} - (\log Z_A + \mathcal{D}_A) - (\log Z_B + \mathcal{D}_B)$$

prior independent! (opposite prior dependence for $\log Z$ and $\mathcal{D}$)

Consider a Gaussian:

one can demonstrate that $d - 2\log S$ has a χ^2_d distribution!

Tension probability

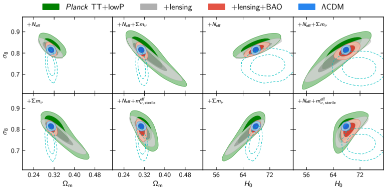
use inverse cumulative χ^2_d distribution
to determine if two datasets are discordant by chance:

$$p = \int_{d-2\log S}^{\infty} \chi^2_d(x)\mathrm{d}x = \int_{d-2\log S}^{\infty} \frac{x^{d/2-1} e^{-x/2}}{2^{d/2}\Gamma(d/2)}\mathrm{d}x$$

if $p \lesssim 5\%$, datasets are in tension

1 *Basics of Bayesian probability*
- Parameter inference
- Bayesian model comparison
- Information gain, model dimensionality and quantifying tensions

2 ***Cosmological tensions***
- Local Universe versus CMB
- Quantifying tensions in Bayesian statistics



3 *Neutrino mass ordering*
- How to constrain the mass ordering
- Subtleties in the Bayesian analysis
- Constraints on the mass ordering

4 *Neutrino masses from cosmology*
- The current status
- One step forward
- Non-probabilistic limits

5 *Conclusions*

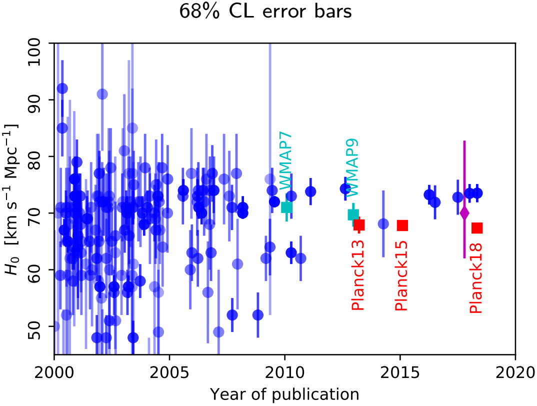
Tension I: the Hubble parameter H_0

[Planck Collaboration, 2018]

$$v = H_0 d,$$
with $H_0 = H(z=0)$

Local measurements:
$H(z=0)$,
local and independent on evolution (model independent, but **systematics?**)

CMB measurements
(probe $z \simeq 1100$):
H_0 from the cosmological evolution (model dependent, well controlled systematics)



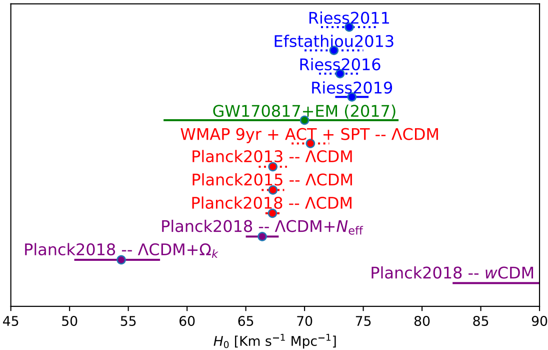
Tension I: the Hubble parameter H_0

$$v = H_0 d, \quad \text{with } H_0 = H(z=0)$$

Local measurements:
$H(z=0)$,
local and independent on evolution (model independent, but **systematics?**)

CMB measurements
(probe $z \simeq 1100$):
H_0 from the cosmological evolution (model dependent, well controlled systematics)

68% CL error bars



Using HST Cepheids:
[Efstathiou 2013] $H_0 = 72.5 \pm 2.5$ Km s^{-1} Mpc^{-1}
[Riess+, 2019] $H_0 = 74.03 \pm 1.42$ Km s^{-1} Mpc^{-1}

GW: [Abbott et al., 2017] $H_0 = 70^{+12}_{-8}$ Km s^{-1} Mpc^{-1}

(ΛCDM model - CMB data only)
[Planck 2013]: $H_0 = 67.3 \pm 1.2$ Km s^{-1} Mpc^{-1}
[Planck 2018]: $H_0 = 67.27 \pm 0.60$ Km s^{-1} Mpc^{-1}

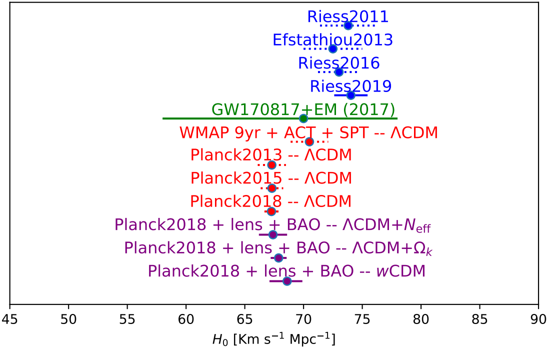
Tension I: the Hubble parameter H_0

$$v = H_0 d,$$
with $H_0 = H(z=0)$

Local measurements:
$H(z=0)$,
local and independent on evolution (model independent, but **systematics?**)

CMB measurements
(probe $z \simeq 1100$):
H_0 from the cosmological evolution (model dependent, well controlled systematics)

68% CL error bars



Using HST Cepheids:
[Efstathiou 2013] $H_0 = 72.5 \pm 2.5$ Km s^{-1} Mpc^{-1}
[Riess+, 2019] $H_0 = 74.03 \pm 1.42$ Km s^{-1} Mpc^{-1}

GW: [Abbott et al., 2017] $H_0 = 70^{+12}_{-8}$ Km s^{-1} Mpc^{-1}

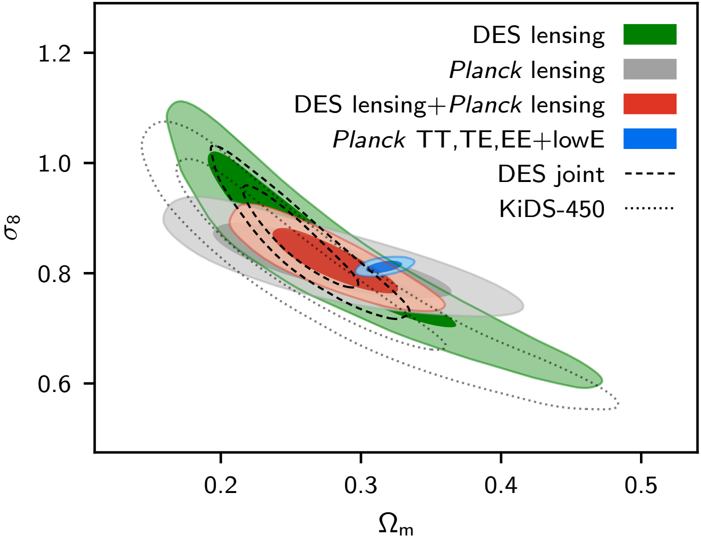
(ΛCDM model - CMB data only)
[Planck 2013]: $H_0 = 67.3 \pm 1.2$ Km s^{-1} Mpc^{-1}
[Planck 2018]: $H_0 = 67.27 \pm 0.60$ Km s^{-1} Mpc^{-1}

Tension II (?): the matter distribution at small scales

Assuming ΛCDM model:

σ_8: rms fluctuation in total matter (baryons + CDM + neutrinos) in $8h^{-1}$ Mpc spheres, today;

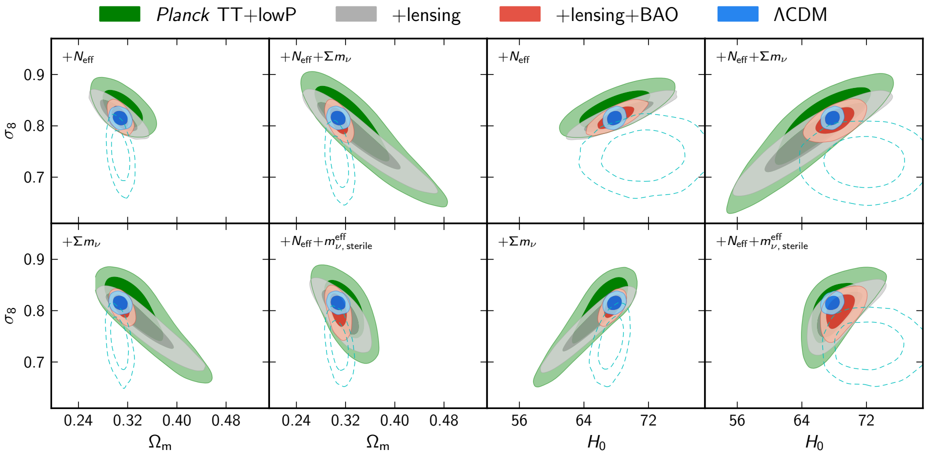
Ω_m: total matter density today divided by the critical density



[Planck collaboration, 2018]

Do neutrinos help?

[Planck Collaboration, 2015]



dashed: local measurements – ΛCDM model, ΛCDM + $\nu_{a,s}$ models: full cosmological dataset

H_0 increases ⇒ σ_8 increases (and viceversa)!

The correlations do not help.

Quantifying tensions

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

$$R = \frac{Z_{AB}}{Z_A Z_B} \text{ or } \log R = \log Z_{AB} - \log Z_A - \log Z_B \longleftarrow \text{prior!}$$

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \longleftarrow \text{prior!}$$

$$S = R/I \text{ or } \log S = \log R - \log I \longleftarrow \text{no prior!}$$

$$d = 2\int p(\theta)\left(\log\frac{p(\theta)}{\pi(\theta)} - \mathcal{D}\right)^2 \mathrm{d}\theta \qquad p = \int_{d-2\log S}^{\infty} \frac{x^{d/2-1}e^{-x/2}}{2^{d/2}\Gamma(d/2)}\mathrm{d}x$$

Quantifying tensions

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

$$R = \frac{Z_{AB}}{Z_A Z_B} \text{ or } \log R = \log Z_{AB} - \log Z_A - \log Z_B \longleftarrow \text{prior!}$$

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \longleftarrow \text{prior!}$$

$$S = R/I \text{ or } \log S = \log R - \log I \longleftarrow \text{no prior!}$$

$$d = 2\int p(\theta)\left(\log\frac{p(\theta)}{\pi(\theta)} - \mathcal{D}\right)^2 \mathrm{d}\theta \qquad p = \int_{d-2\log S}^{\infty} \frac{x^{d/2-1}e^{-x/2}}{2^{d/2}\Gamma(d/2)}\mathrm{d}x$$

Dataset	Prior	$\log R$	$\log I$	$\log S$	d	$p(\%)$
BOSS-*Planck*	default	6.24 ± 0.30	6.15 ± 0.29	0.09 ± 0.29	2.69 ± 0.23	41.58 ± 4.38
	medium	4.49 ± 0.30	4.03 ± 0.29	0.46 ± 0.29	3.48 ± 0.24	54.80 ± 4.16
	narrow	1.30 ± 0.23	0.69 ± 0.23	0.61 ± 0.23	2.11 ± 0.23	66.31 ± 5.19
DES-*Planck*	default	2.91 ± 0.35	6.18 ± 0.35	-3.27 ± 0.35	2.50 ± 0.32	1.91 ± 0.58
	medium	0.51 ± 0.36	3.98 ± 0.36	-3.47 ± 0.36	2.03 ± 0.33	1.22 ± 0.43
	narrow	-1.88 ± 0.31	0.92 ± 0.30	-2.80 ± 0.30	1.18 ± 0.31	1.31 ± 0.60
SH_0ES-*Planck*	default	-2.00 ± 0.31	1.98 ± 0.31	-3.98 ± 0.31	1.26 ± 0.23	0.39 ± 0.14
	medium	-2.50 ± 0.29	1.55 ± 0.28	-4.05 ± 0.28	1.12 ± 0.23	0.31 ± 0.12
	narrow	-2.01 ± 0.25	1.43 ± 0.23	-3.44 ± 0.23	2.35 ± 0.23	1.48 ± 0.35

Quantifying tensions

[Handley+, arxiv:1902.04029]
[Handley+, arxiv:1903.06682]

$$R = \frac{Z_{AB}}{Z_A Z_B} \text{ or } \log R = \log Z_{AB} - \log Z_A - \log Z_B \longleftarrow \text{prior!}$$

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \longleftarrow \text{prior!}$$

$$S = R/I \text{ or } \log S = \log R - \log I \longleftarrow \text{no prior!}$$

$$d = 2 \int p(\theta) \left(\log \frac{p(\theta)}{\pi(\theta)} - \mathcal{D} \right)^2 \mathrm{d}\theta \qquad p = \int_{d-2\log S}^{\infty} \frac{x^{d/2-1} e^{-x/2}}{2^{d/2}\Gamma(d/2)} \mathrm{d}x$$

Dataset	Prior	$\log R$	$\log I$	$\log S$	d	$p(\%)$
BOSS-*Planck*	default	6.24 ± 0.30	6.15 ± 0.29	0.09 ± 0.29	2.69 ± 0.23	41.58 ± 4.38
	medium	4.49 ± 0.30	4.03 ± 0.29	0.46 ± 0.29	3.48 ± 0.24	54.80 ± 4.16
	narrow	1.30 ± 0.23	0.69 ± 0.23	0.61 ± 0.23	2.11 ± 0.23	66.31 ± 5.19
DES-*Planck*	default	2.91 ± 0.35	6.18 ± 0.35	-3.27 ± 0.35	2.50 ± 0.32	1.91 ± 0.58
	medium	0.51 ± 0.36	3.98 ± 0.36	-3.47 ± 0.36	2.03 ± 0.33	1.22 ± 0.43
	narrow	-1.88 ± 0.31	0.92 ± 0.30	-2.80 ± 0.30	1.18 ± 0.31	1.31 ± 0.60
SH_0ES-*Planck*	default	-2.00 ± 0.31	1.98 ± 0.31	-3.98 ± 0.31	1.26 ± 0.23	0.39 ± 0.14
	medium	-2.50 ± 0.29	1.55 ± 0.28	-4.05 ± 0.28	1.12 ± 0.23	0.31 ± 0.12
	narrow	-2.01 ± 0.25	1.43 ± 0.23	-3.44 ± 0.23	2.35 ± 0.23	1.48 ± 0.35

BOSS-Planck: agreement

DES-Planck: moderate tension

SH_0ES-Planck: strong tension

1 *Basics of Bayesian probability*

- Parameter inference
- Bayesian model comparison
- Information gain, model dimensionality and

2 *Cosmological tensions*

- Local Universe versus CMB
- Quantifying tensions in Bayesian statistics

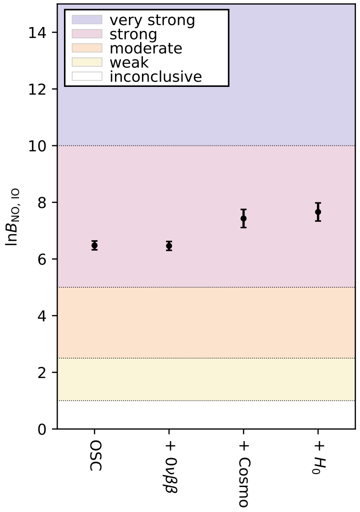
3 ***Neutrino mass ordering***

- How to constrain the mass ordering
- Subtleties in the Bayesian analysis
- Constraints on the mass ordering

4 *Neutrino masses from cosmology*

- The current status
- One step forward
- Non-probabilistic limits

5 *Conclusions*



Neutrinos and their masses

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

Normal ordering (NO)

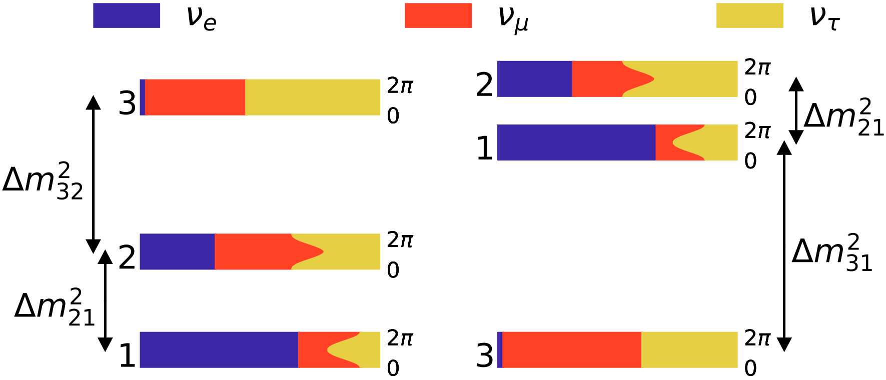
$m_1 < m_2 < m_3$

$\sum m_k \gtrsim 0.06$ 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 $\sum m_\nu$?

Neutrino masses from β decay

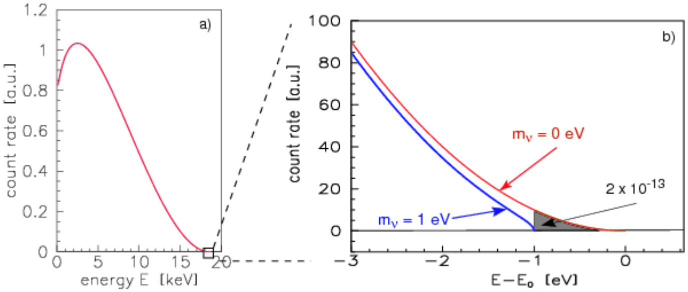
Must measure β decay endpoint → $m^2_{\nu_e} = \sum_k |U_{ek}|^2\, m_k^2$

U_{ek} mixing matrix

Mainz/Troitsk limits, $m_{\nu_e} \lesssim 2$ eV

Katrin, (expected) $m_{\nu_e} \lesssim 0.2$ eV

[Katrin L.o.I., 2001]



Neutrino masses from β decay

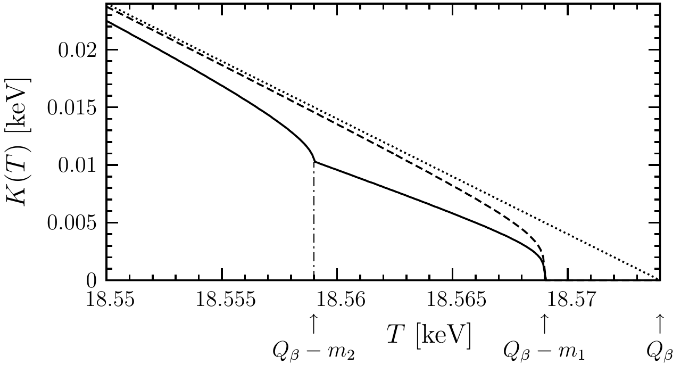
Must measure β decay endpoint → $m^2_{\nu_e} = \sum_k |U_{ek}|^2 m_k^2$

Mainz/Troitsk limits, $m_{\nu_e} \lesssim 2$ eV

U_{ek} mixing matrix

Katrin, (expected) $m_{\nu_e} \lesssim 0.2$ eV

[Giunti&Kim, 2007]

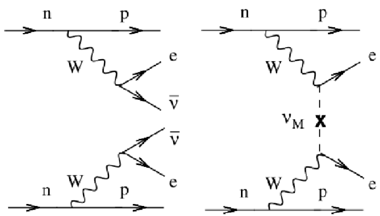


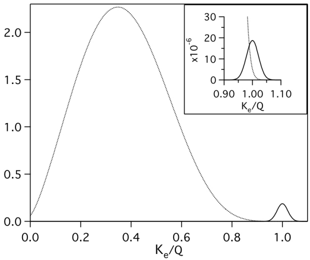
Neutrino masses from neutrinoless double β decay

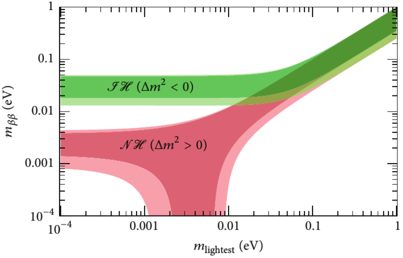
(if neutrino is Majorana)

[Schechter&Valle, 1982]

figure from [NEXT] webpage







[Dell'Oro et al., 2016]

Measure $T^{0\nu}_{1/2}$

m_e electron mass,
$G_{0\nu}$ phase space,
$\mathcal{M}'^{\nu}$ matrix element

convert into $m_{\beta\beta} = \dfrac{m_e}{\mathcal{M}'^{\nu}\sqrt{G_{0\nu}\,T^{0\nu}_{1/2}}}$

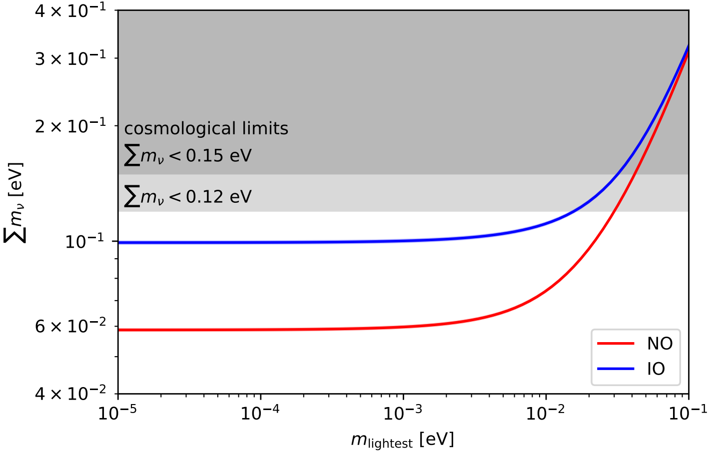
α_k Majorana phases

and then use $m_{\beta\beta} = \left|\sum_k e^{i\alpha_k} U^2_{ek}\, m_k\right|$

From cosmology...

Warning: model dependent content!

How the limit change when considering extensions of the ΛCDM model?



Warning: $\sum m_\nu \lesssim 0.1$ eV at 95% CL
does not mean IO disfavored at 95% CL!

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

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

Should we use linear or logarithmic priors on m_k (m_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 $m_{\rm lightest}$, Δm^2_{21}, $|\Delta m^2_{31}|$ (B)

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

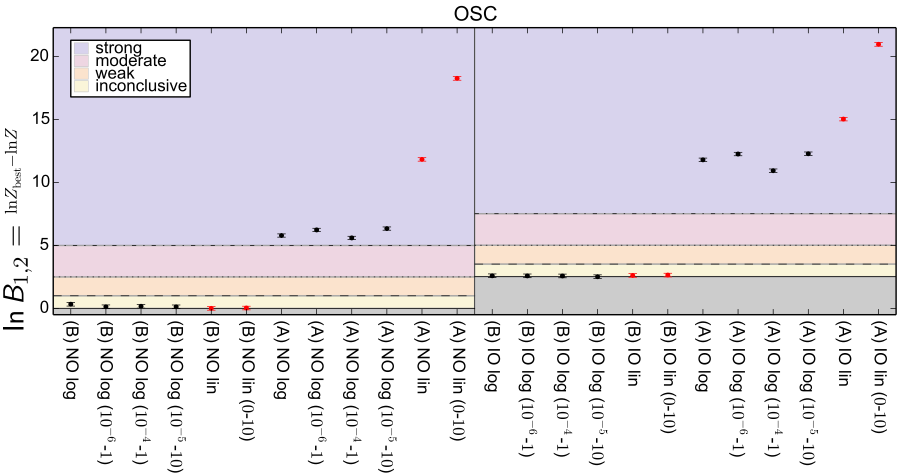
Should we use linear or logarithmic priors on m_k ($m_{\rm lightest}$)?

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

Case A			Case B		
Parameter	Prior	Range	Parameter	Prior	Range
m_1/eV	linear log	$0-1$ $10^{-5}-1$	$m_{\rm lightest}/\mathrm{eV}$	linear log	$0-1$ $10^{-5}-1$
m_2/eV	linear log	$0-1$ $10^{-5}-1$	$\Delta m^2_{21}/\mathrm{eV}^2$	linear	$5\times10^{-5}-10^{-4}$
m_3/eV	linear log	$0-1$ $10^{-5}-1$	$\lvert\Delta m^2_{31}\rvert/\mathrm{eV}^2$	linear	$1.5\times10^{-3}-3.5\times10^{-3}$

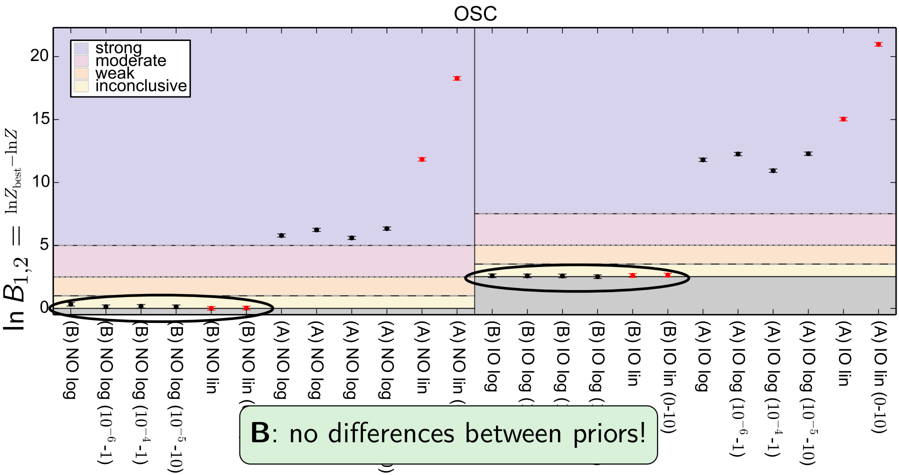
Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

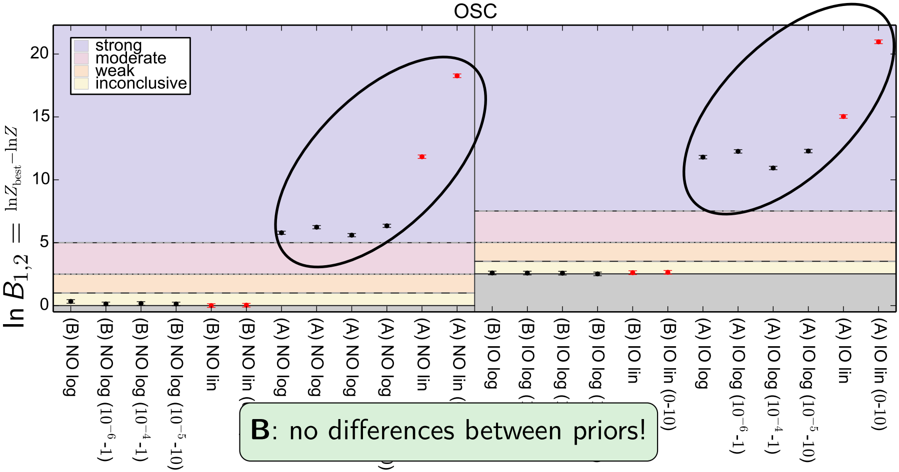


B: no differences between priors!

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

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



B: no differences between priors!

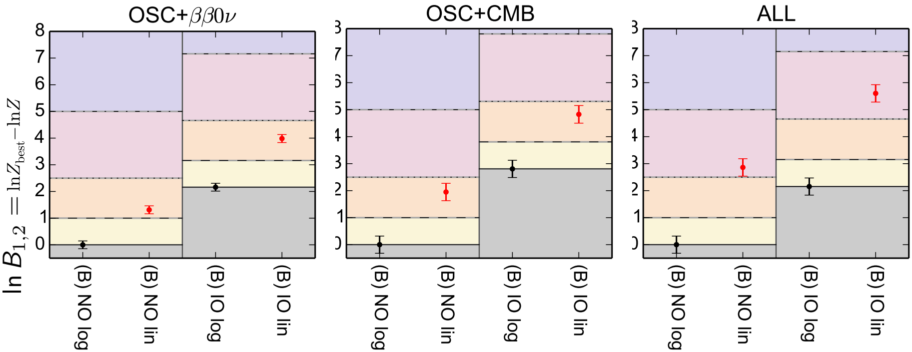
(no penalty for unconstrained parameter $m_{\rm lightest}$!)

A: always strongly disfavored!

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

Comparing parameterizations/priors

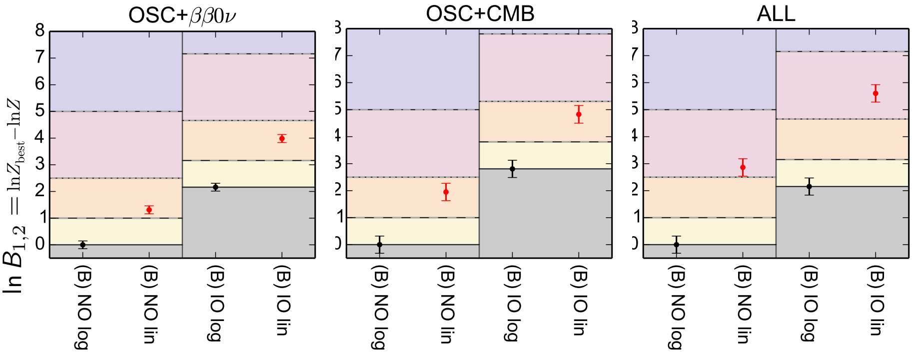
[SG+, JCAP 03 (2018) 11]



compare linear versus **logarithmic**

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

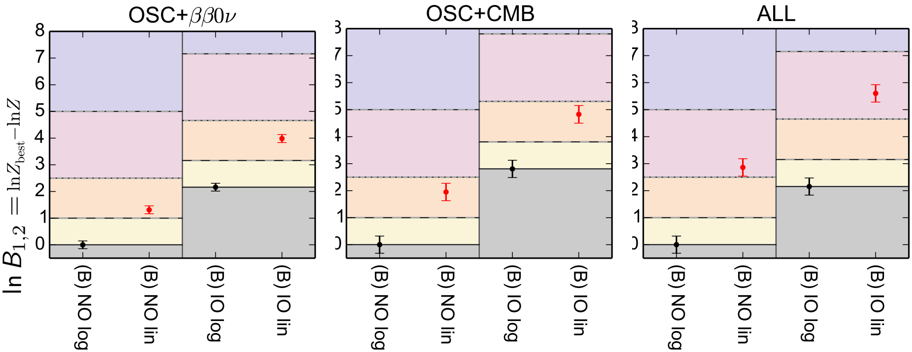


compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



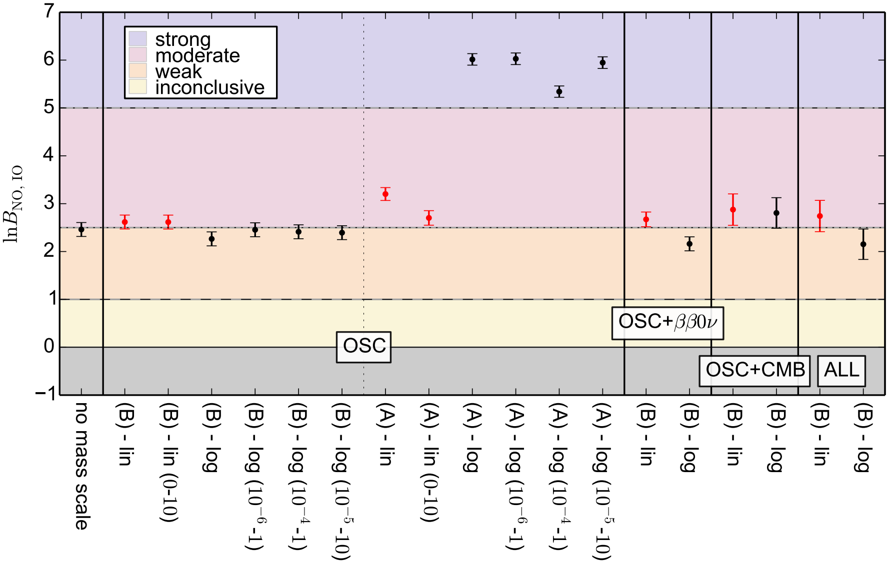
compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

summary: case B, log prior is better!

Comparing the mass orderings

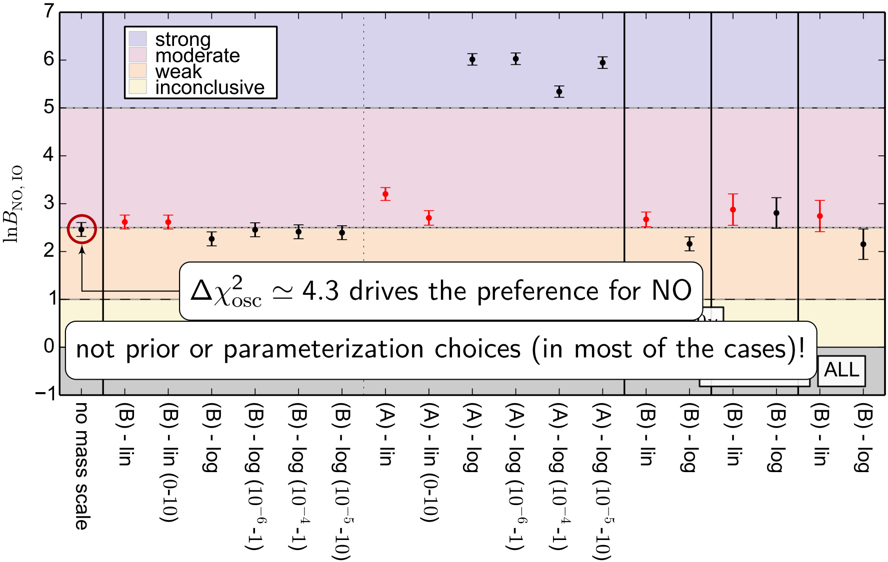
[SG+, JCAP 03 (2018) 11]



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

Comparing the mass orderings

[SG+, JCAP 03 (2018) 11]



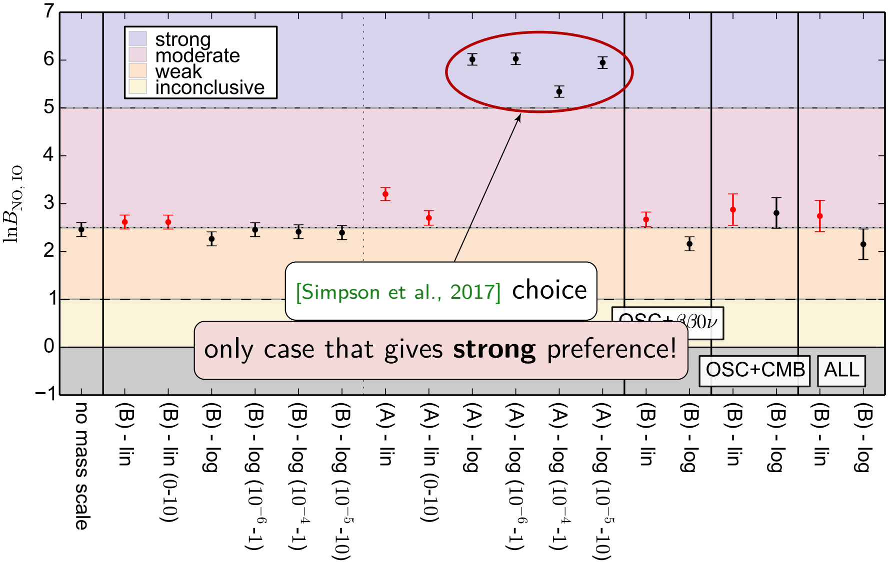
$\Delta\chi^2_{\rm osc} \simeq 4.3$ drives the preference for NO

not prior or parameterization choices (in most of the cases)!

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

Comparing the mass orderings

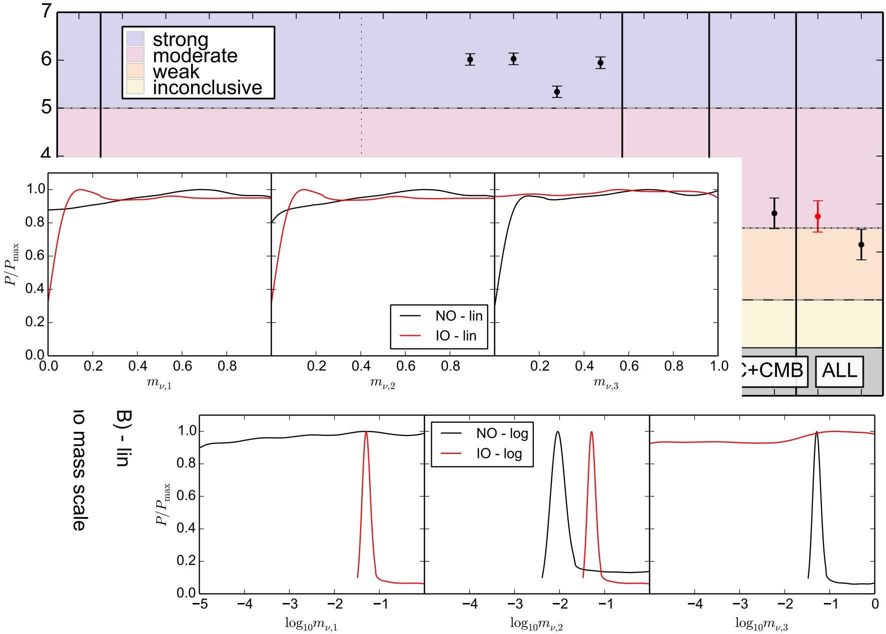
[SG+, JCAP 03 (2018) 11]



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

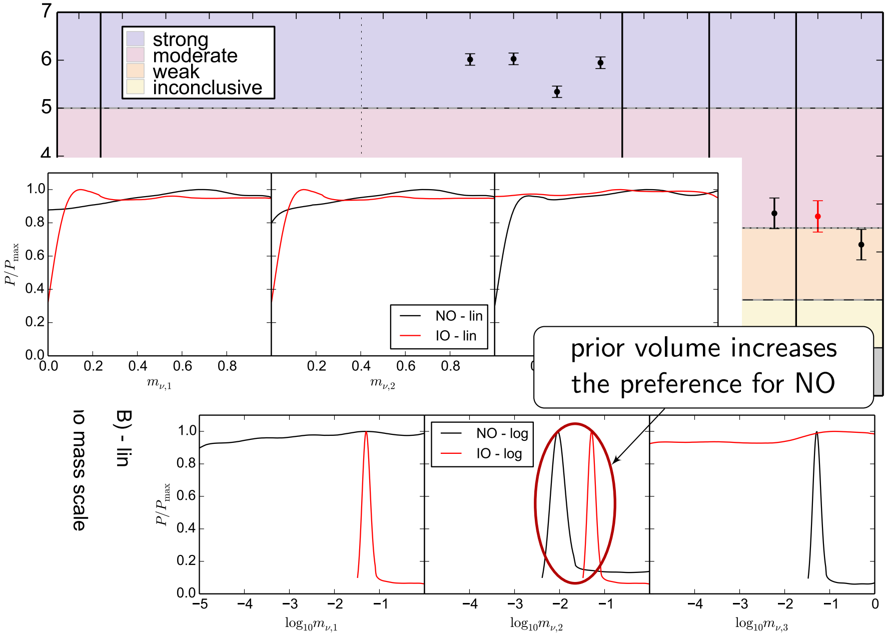
Comparing the mass orderings

[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:

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

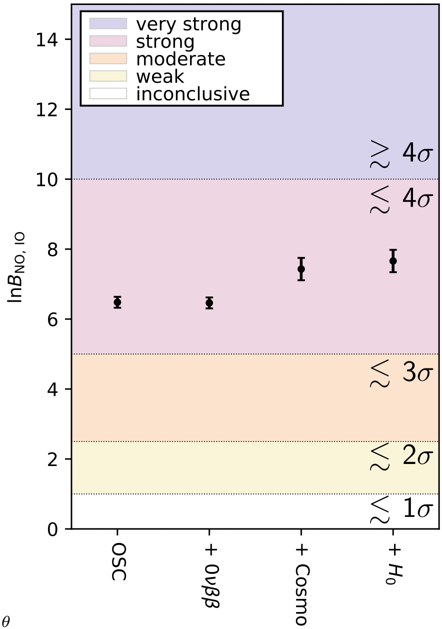
Bayes factor NO vs IO:

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

Posterior probability:

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

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



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

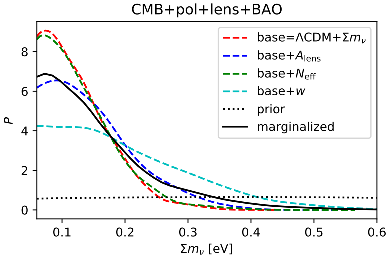
1 *Basics of Bayesian probability*
- Parameter inference
- Bayesian model comparison
- Information gain, model dimensionality and quantifying tensions

2 *Cosmological tensions*
- Local Universe versus CMB
- Quantifying tensions in Bayesian statistics

3 *Neutrino mass ordering*
- How to constrain the mass ordering
- Subtleties in the Bayesian analysis
- Constraints on the mass ordering

4 ***Neutrino masses from cosmology***
- The current status
- One step forward
- Non-probabilistic limits

5 *Conclusions*



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

Playing with priors

Bayes theorem:

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

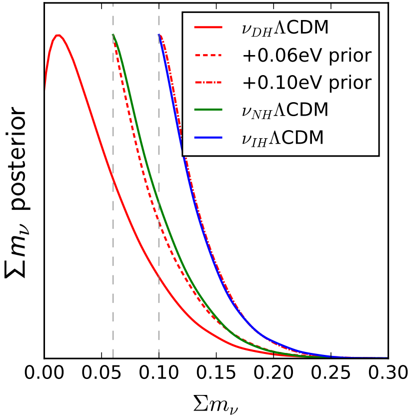
posterior depends on prior!

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

[Wang+, 2017]

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

(i.e. change the prior lower bound)



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:

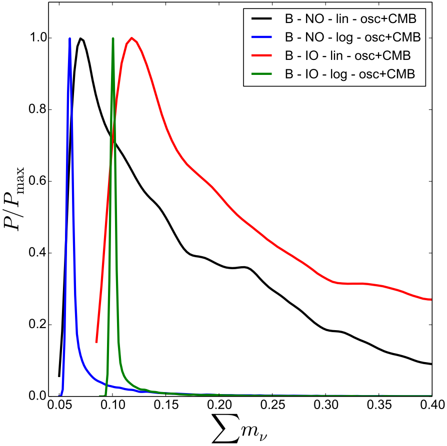
$$p(\theta|d,\mathcal{M}) = \mathcal{L}(\theta)\frac{\pi(\theta|\mathcal{M})}{Z_{\mathcal{M}}}$$

posterior depends on prior!

You can artificially tighten the bounds on Σm_ν with different priors...

[SG+, 2018]

logarithmic vs **lin**ear prior on $m_{\rm lightest}$



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**
ΛCDM [Planck 2015]	$\Sigma m_\nu < 0.68$ eV $\Sigma m_\nu < 0.49$ eV		$\Sigma m_\nu < 0.25$ 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**	Model	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**
ΛCDM [Planck 2015]	$\Sigma m_\nu < 0.68$ eV $\Sigma m_\nu < 0.49$ eV		$\Sigma m_\nu < 0.25$ 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]
ΛCDM+$A_{\rm lens}$ [Planck 2015] (free phenomenological lensing amplitude $A_{\rm lens} \neq -1$)	-		$\Sigma m_\nu < 0.41$ eV
eCDM [Di Valentino+, 2015] (12-parameters cosmological model, ΛCDM based)	$\Sigma m_\nu < 0.96$ eV		$\Sigma m_\nu < 0.53$ eV

Marginalize over models?

We usually marginalize over parameters:

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

Can we marginalize over models?

Marginalize over models?

[SG+, PRD 99 (2019) 021301]

We usually marginalize over parameters:

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

Can we marginalize over models?

Yes, if we know the model posteriors:

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

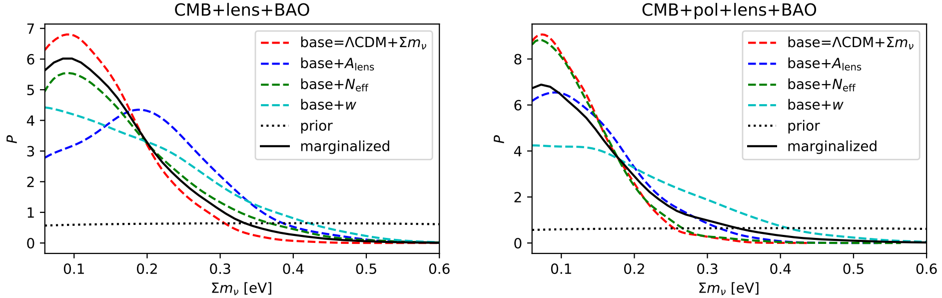
Select a model $\mathcal{M}_0$ and use $p_i = Z_i/(\sum Z_j) = B_{i0}/(\sum B_{j0})$:

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

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

Model-marginalization applied to Σm_ν

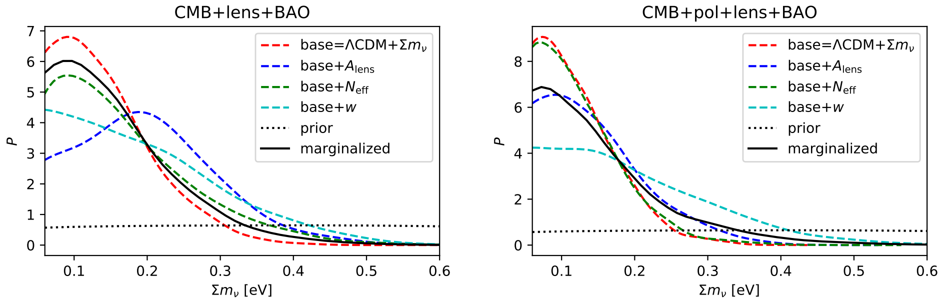
[SG+, PRD 99 (2019) 021301]



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	

Model-marginalization applied to Σm_ν

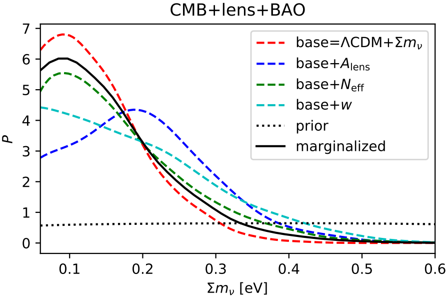
[SG+, PRD 99 (2019) 021301]

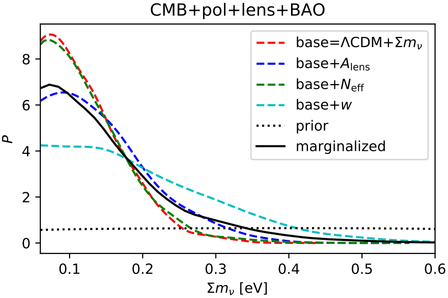


model	CMB+lens+BAO $\ln B_{i0}$	CMB+lens+BAO Σm_ν [eV]	CMB+pol+lens+BAO $\ln B_{i0}$	CMB+pol+lens+BAO Σm_ν [eV]
base=ΛCDM+Σm_ν	0.0	< 0.28	0.0	< 0.23
base+$A_{\rm lens}$	-2.6	< 0.38	-2.4	< 0.29
base+$N_{\rm eff}$	-1.5	< 0.37	-2.3	< 0.25
base+w	-1.4	< 0.42	-0.1	< 0.42
marginalized	–	< 0.33	–	< 0.35
p_0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

Model-marginalization applied to Σm_ν

[SG+, PRD 99 (2019) 021301]





	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	

Prior-independent Bayesian parameter constraints

relative belief updating ratio
[Astone, 1999]
[D'Agostini, 2000]

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

independent of $\pi(x)$!

Prior-independent Bayesian parameter constraints

relative belief updating ratio
[Astone, 1999]
[D'Agostini, 2000]

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

independent of $\pi(x)$!

see $\frac{p(\mathcal{M}_i|d)}{p(\mathcal{M}_j|d)} = B_{ij}\frac{\pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j)}$

Rewrite in a more familiar form:

$$\frac{p(x|d)}{p(x_0|d)} = \mathcal{R}(x, x_0|d)\frac{\pi(x)}{\pi(x_0)}$$

Prior-independent Bayesian parameter constraints

relative belief updating ratio
[Astone, 1999]
[D'Agostini, 2000]

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

independent of $\pi(x)$!

Rewrite in a more familiar form:

see $\frac{p(\mathcal{M}_i|d)}{p(\mathcal{M}_j|d)} = B_{ij}\frac{\pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j)}$

$$\frac{p(x|d)}{p(x_0|d)} = \mathcal{R}(x, x_0|d)\frac{\pi(x)}{\pi(x_0)}$$

→ it's the same as a Bayes factor!

not a probability distribution!!

DON'T USE FOR PROBABILISTIC LIMITS

Prior-independent Bayesian parameter constraints

relative belief updating ratio
[Astone, 1999]
[D'Agostini, 2000]

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

independent of $\pi(x)$!

Rewrite in a more familiar form:

see $\frac{p(\mathcal{M}_i|d)}{p(\mathcal{M}_j|d)} = B_{ij}\frac{\pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j)}$

$$\frac{p(x|d)}{p(x_0|d)} = \mathcal{R}(x, x_0|d)\frac{\pi(x)}{\pi(x_0)}$$

→ it's the same as a Bayes factor!
not a probability distribution!!

DON'T USE FOR PROBABILISTIC LIMITS

x_0 is limit to which data are insensitive to x, e.g. $x_0 = 0$ (if x is Σm_ν)

$\mathcal{R}(x, x_0|d)$ describes how **data** update our initial beliefs on x

→ $\mathcal{R} \simeq 1$ ($x \to x_0$): data are insensitive to x
→ $\mathcal{R} \to 0$ ($x \gg x_0$): data disfavor x, regardless of prior

Prior-independent Bayesian parameter constraints

relative belief updating ratio
[Astone, 1999]
[D'Agostini, 2000]

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

independent of $\pi(x)$!

Rewrite in a more familiar form:

see $\frac{p(\mathcal{M}_i|d)}{p(\mathcal{M}_j|d)} = B_{ij}\frac{\pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j)}$

$$\frac{p(x|d)}{p(x_0|d)} = \mathcal{R}(x, x_0|d)\frac{\pi(x)}{\pi(x_0)}$$

→ it's the same as a Bayes factor!
not a probability distribution!!

DON'T USE FOR PROBABILISTIC LIMITS

x_0 is limit to which data are insensitive to x, e.g. $x_0 = 0$ (if x is Σm_ν)

$\mathcal{R}(x, x_0|d)$ describes how **data** update our initial beliefs on x

- → $\mathcal{R} \simeq 1$ ($x \to x_0$): data are insensitive to x
- → $\mathcal{R} \to 0$ ($x \gg x_0$): data disfavor x, regardless of prior

we can use $\mathcal{R}$ to derive a (non-probabilistic) "sensitivity bound x_s"

$x > x_s$ disfavored because $\mathcal{R}(x, x_0|d) < s$, with $s = 5\%$ or 1%

x_s is a hedge "which separates the region in which we are, and where we see nothing, from the the region we cannot see" [D'Agostini, 2000]

An example with Planck 2018

relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

An example with Planck 2018

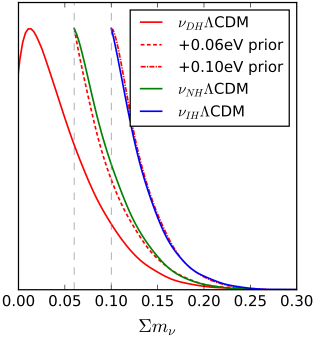
relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

Note: 1D plots in cosmology are already close to show $\mathcal{R}$
as for linear priors, the shape of $\mathcal{R}(x, x_0|d)$ is equal to the one of $p(x|d)$!

e.g. [Wang+, 2017]



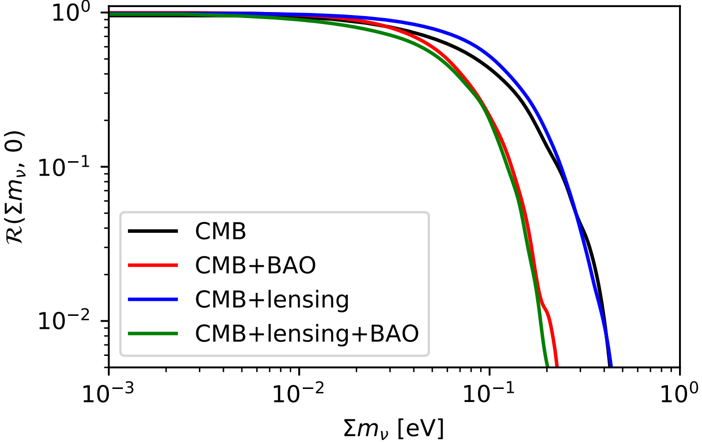
An example with Planck 2018

relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

Example with [Planck 2018] chains from PLA, ΛCDM+Σm_ν



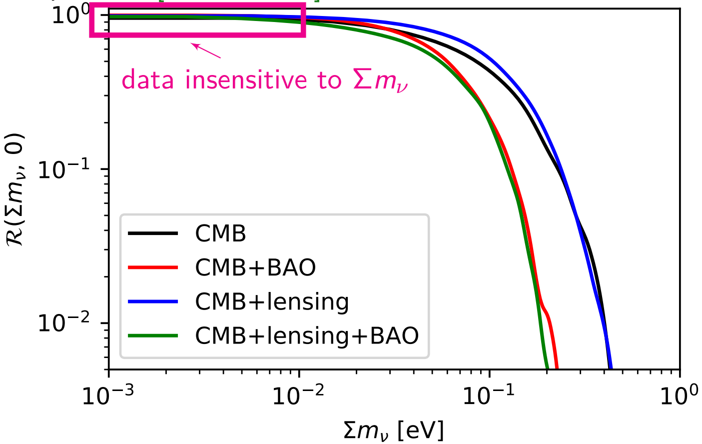
An example with Planck 2018

relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

Example with [Planck 2018] chains from PLA, ΛCDM+Σm_ν



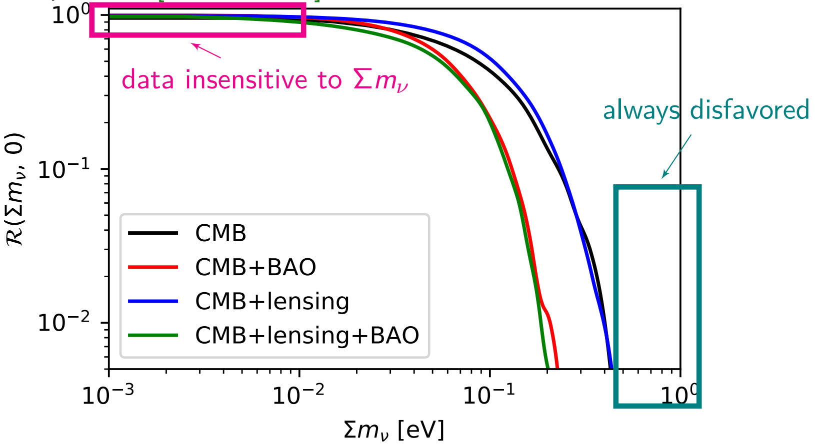
An example with Planck 2018

relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

Example with [Planck 2018] chains from PLA, ΛCDM+Σm_ν



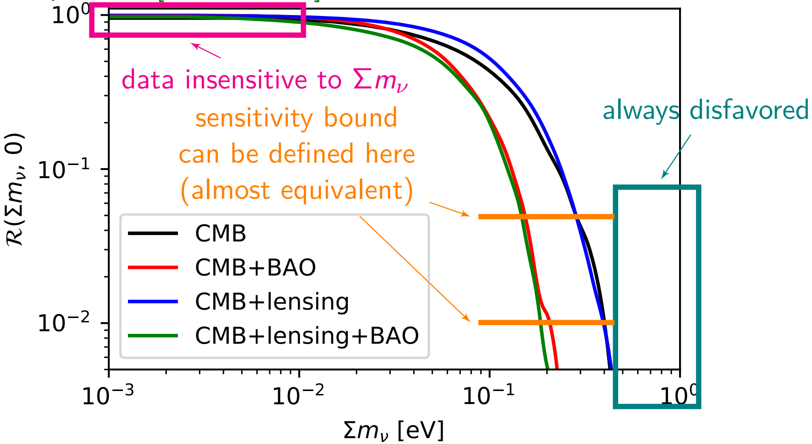
An example with Planck 2018

relative belief updating ratio

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)}$$

Numerically easy to compute: fix $\pi(x)$, get $p(x|d)$ normally and divide

Example with [Planck 2018] chains from PLA, ΛCDM+Σm_ν



The two ways of ~~the Force~~ Bayesianism

prior dependence is intrinsic of Bayesian statistics

two ways to deal with this

The two ways of ~~the Force~~ Bayesianism

prior dependence is intrinsic of Bayesian statistics

two ways to deal with this

Subjective "dark side"?

- priors depend on the researcher
- state your assumptions and present your results
- results *may* be different
- they will converge with more data

The two ways of ~~the Force~~ Bayesianism

prior dependence is intrinsic of Bayesian statistics

two ways to deal with this

Subjective "dark side"?

- priors depend on the researcher
- state your assumptions and present your results
- results *may* be different
- they will converge with more data

Objective "light side"?

- mathematics can help to minimize subjectivity
- priors from objective criteria (e.g. maximize information gain)
- *still, dependence on prior ranges may remain* (see later)

The two ways of ~~the Force~~ Bayesianism

prior dependence is intrinsic of Bayesian statistics

two ways to deal with this

Subjective "dark side"?

- priors depend on the researcher
- state your assumptions and present your results
- results *may* be different
- they will converge with more data

Objective "light side"?

- mathematics can help to minimize subjectivity
- priors from objective criteria (e.g. maximize information gain)
- *still, dependence on prior ranges may remain* (see later)

Balance is the way

sensitivity analysis: try different priors+ranges, see if results are stable

Conclusions

1. Be **careful** when you play
with **priors in Bayesian analysis!**
(and always declare your model completely)

2. Bayesian techniques allow
to **marginalize over different models/priors**
and to present
(nearly) model- and prior-independent results!

3. Bayesian techniques allow to quantify
number of constrained parameters and
amount of tensions between datasets

Conclusions

1

Be **careful** when you play
with **priors in Bayesian analysis!**
(and always declare your model completely)

2

Bayesian techniques allow
to **marginalize over different models/priors**
and to present
(nearly) model- and prior-independent results!

3

Bayesian techniques allow to quantify
number of constrained parameters and
amount of tensions between datasets

Thank you for the attention!