IFIC INSTITUT DE FÍSICA CORPUSCULAR
CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
VNIVERSITAT DE VALÈNCIA
EXCELENCIA SEVERO OCHOA
GENERALITAT VALENCIANA
European Commission

Stefano Gariazzo

IFIC, Valencia (ES)
CSIC – Universitat de Valencia

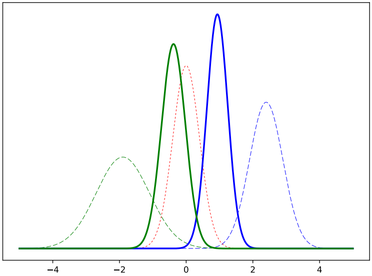
gariazzo@ific.uv.es
http://ific.uv.es/~gariazzo/

Horizon 2020
European Union funding
for Research & Innovation

Towards model-independent constraints on neutrino properties from cosmology

Including several concepts of Bayesian statistics

RWTH Aachen, Informal Cosmology Seminar, 11/02/2019



−4
−2
0
2
4

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)?

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

"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"

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

$\downarrow$

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

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:

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

model comparison

(Bayesian) Parameter inference

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

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

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

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

(physical ϕ + nuisance ψ)

Full posterior:

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

(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

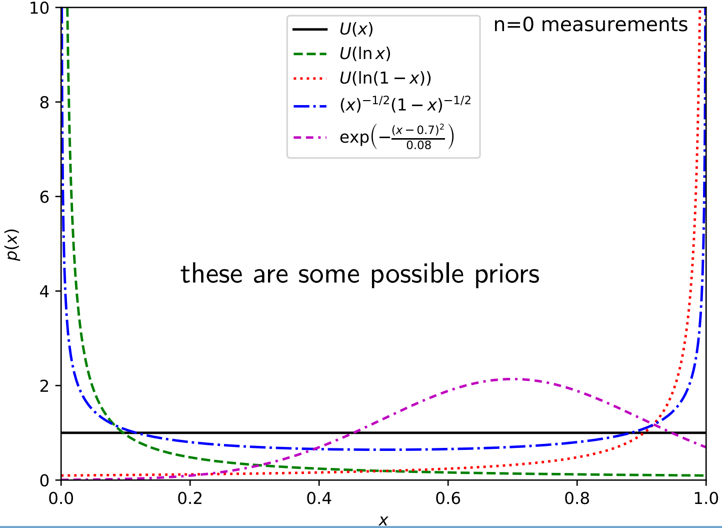
equal-tailed interval: same probability of being below or above the interval

interval for which the mean is the central point

Prior dependence in parameter estimation - I

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

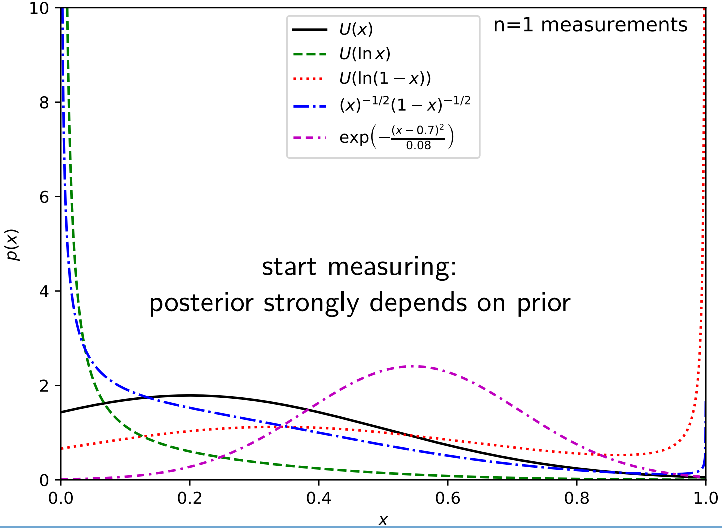
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2 \cdot 0.3^2)]$



Prior dependence in parameter estimation - I

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

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

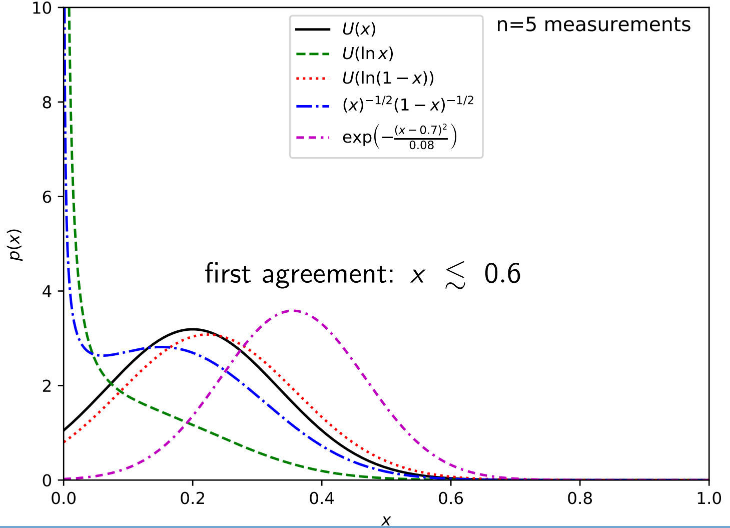


start measuring:
posterior strongly depends on prior

Prior dependence in parameter estimation - I

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

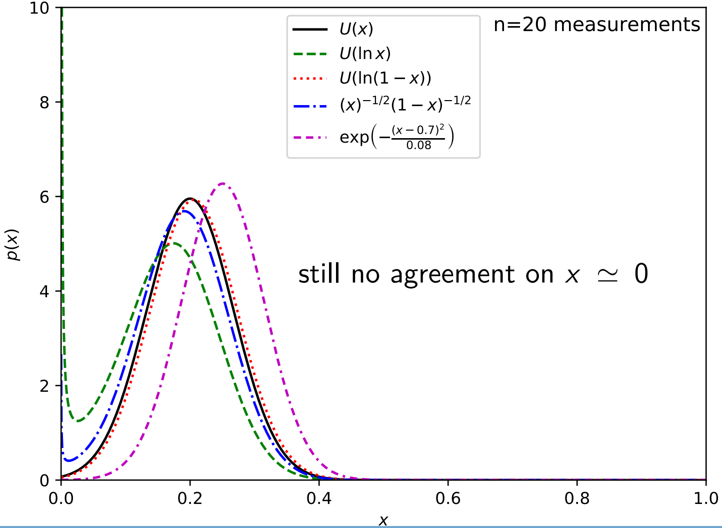
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



Prior dependence in parameter estimation - I

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

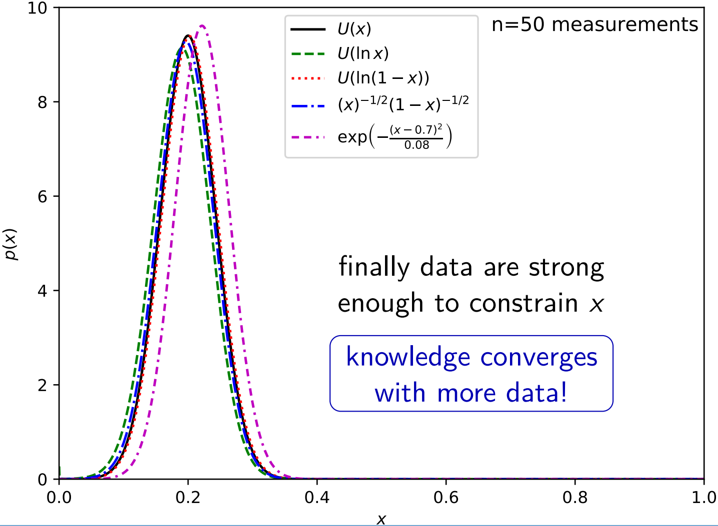
likelihood $\mathcal{L}(x) \propto \exp[-(x-0.2)^2/(2\cdot 0.3^2)]$



Prior dependence in parameter estimation - I

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

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



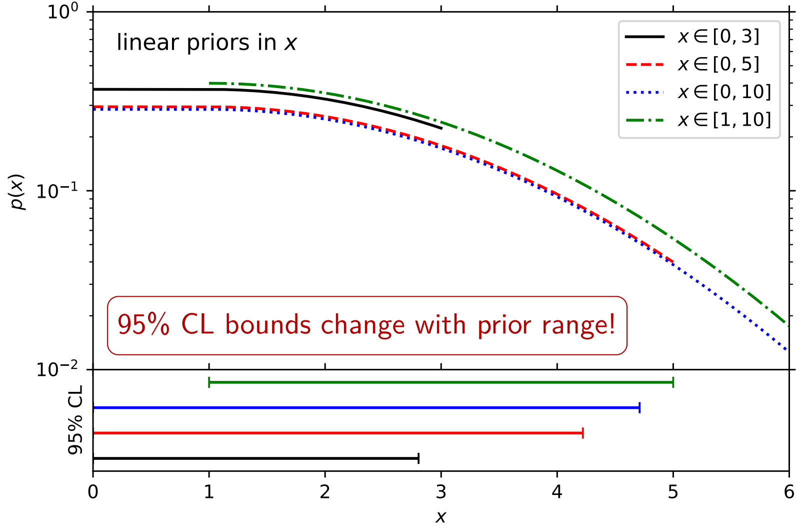
finally data are strong enough to constrain x

knowledge converges with more data!

Prior dependence in parameter estimation - II

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

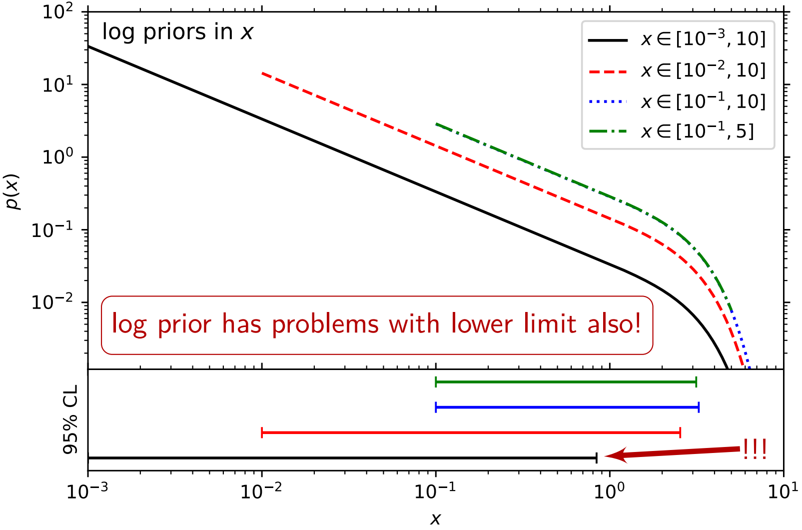
likelihood $\mathcal{L}(x) \propto \exp[-(x-1)^2/(2\cdot 1^2)]$ for $x > 1$, constant otherwise



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)

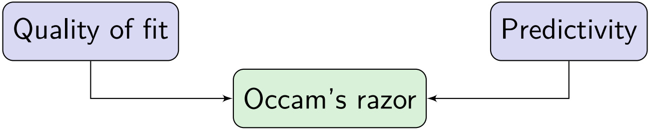
→ are all the parameters needed?

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

Bayes factors depend on priors!

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

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

Bayes factors depend on priors!

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

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

Bayes factors depend on priors!

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

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
$\in [5.0, 10]$	$(150-2.2\times10^4):1$	$2.7-4.1$	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:

$\lvert\ln B_{1,2}\rvert$	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)

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:

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

Compute model posterior probabilities:

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

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

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+, PRD 99 (2019) 021301]

Assume N models, equal model prior probabilities:

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

Compute model posterior probabilities:

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

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

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+, 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:

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

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

Do the result depend on N?

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

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

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 ⟵ this will probably be the favorite one

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

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

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

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

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

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

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

+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

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

$+2$ parameters

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

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

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

$+3$ 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!

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

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

$+3$ 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!

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

but for a complete analysis one should consider
also modified gravity models and so on...

$+3$ pa

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"?

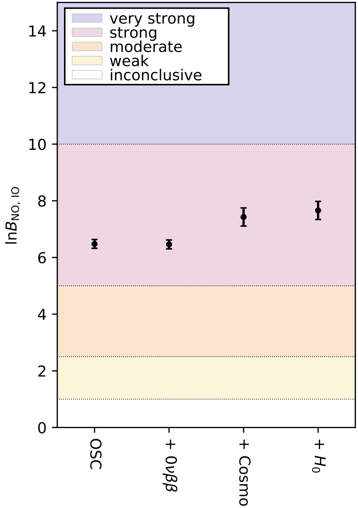
- priors depend on the researcher
- state your assumptions and present your results
- results *may* be different
- they will converge with more data

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


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

Normal ordering (NO)

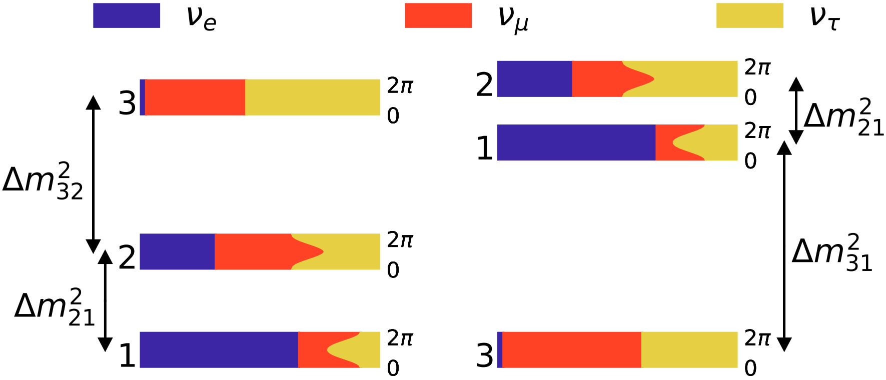
$m_1 < m_2 < m_3$

$\sum m_k \gtrsim 0.06$ 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

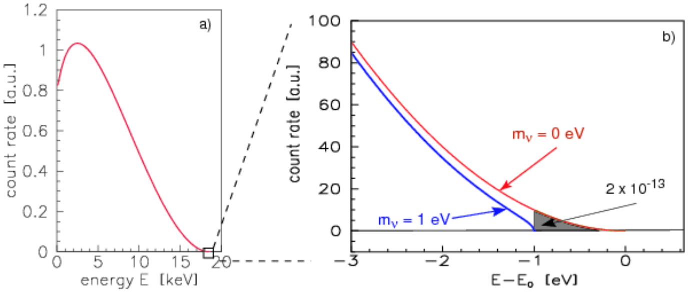
Must measure β decay endpoint → $m^2_{\nu_e} = \sum_k |U_{ek}|^2 m_k^2$

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

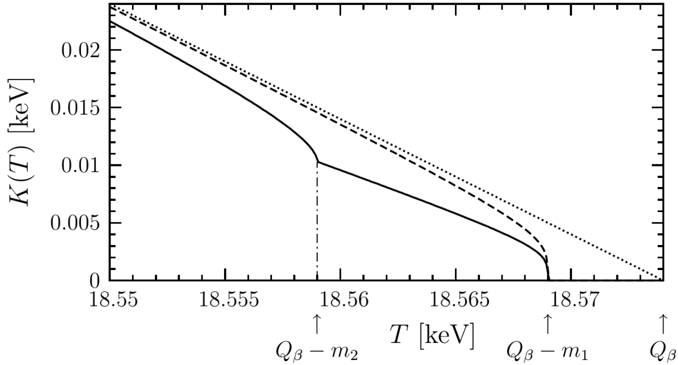
Must measure β decay endpoint → $m^2_{\nu_e} = \sum_k |U_{ek}|^2 m^2_k$

U_{ek} mixing matrix

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

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]

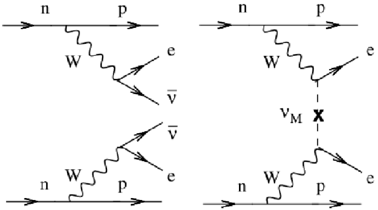
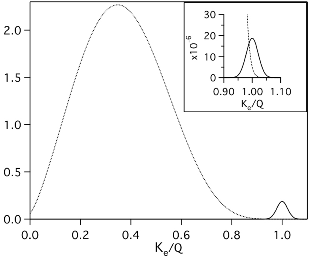
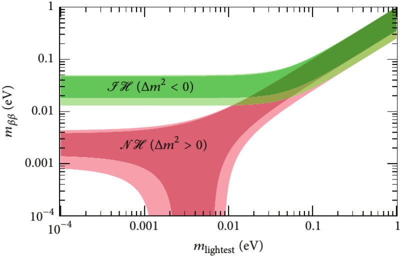


figure from [NEXT] webpage





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

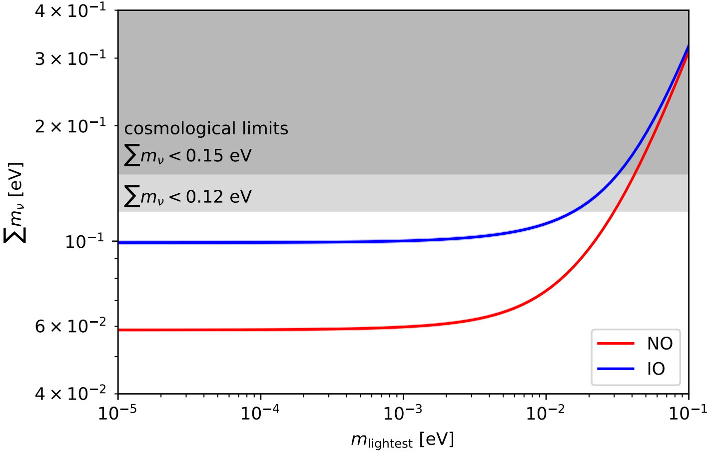
α_k Majorana phases

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

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)?

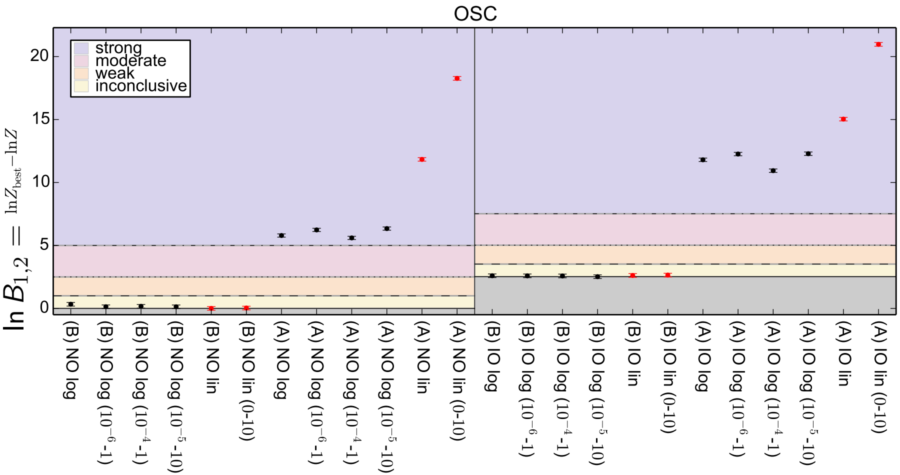
Should we use linear or logarithmic priors on m_k ($m_{\rm lightest}$)?

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	$0 - 1$ $10^{-5} - 1$	$\lvert\Delta m^2_{31}\rvert/\rm eV^2$	linear	$1.5 \times 10^{-3} - 3.5 \times 10^{-3}$

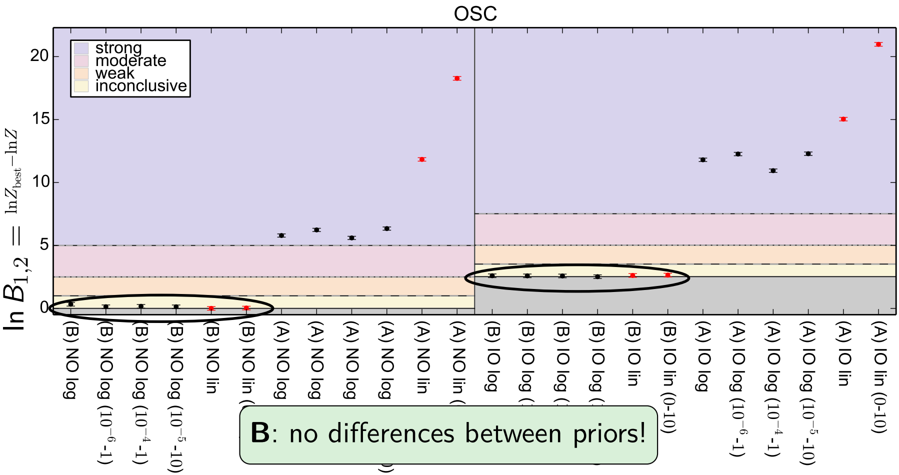
Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

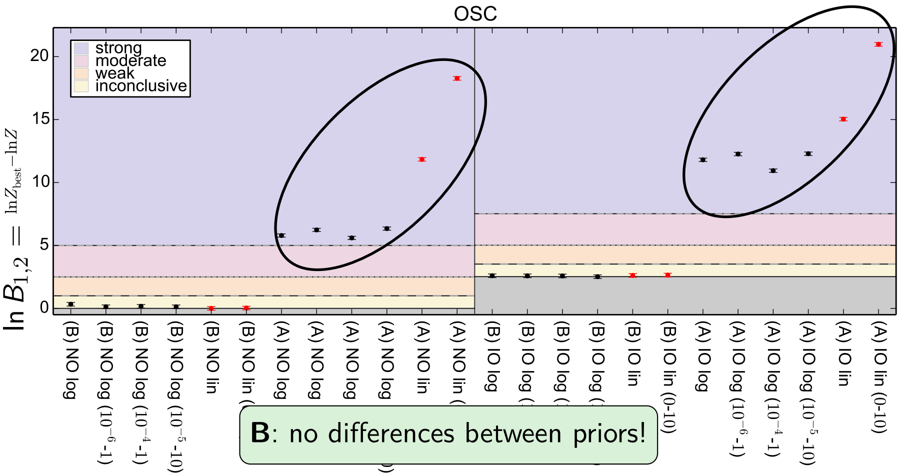


B: no differences between priors!

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

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]



B: no differences between priors!

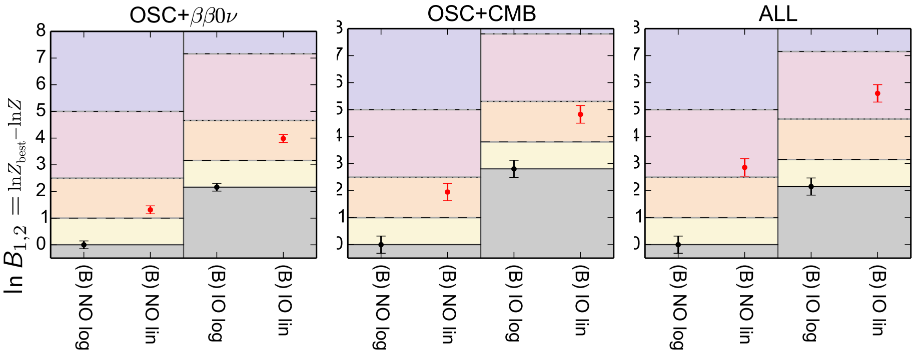
(no penalty for unconstrained parameter $m_{\rm lightest}$!)

A: always strongly disfavored!

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

Comparing parameterizations/priors

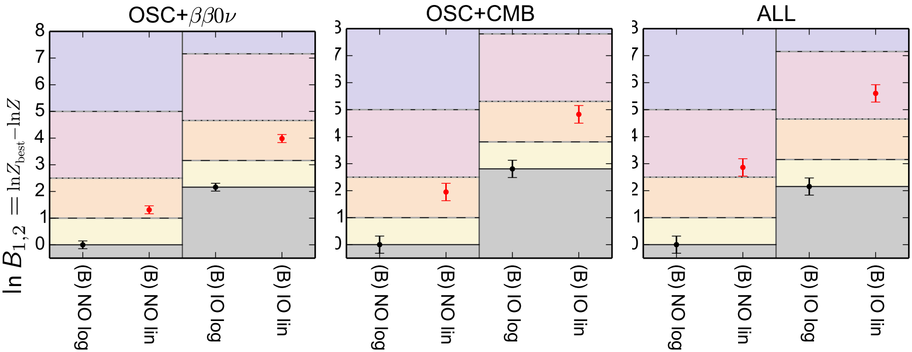
[SG+, JCAP 03 (2018) 11]



compare linear versus **logarithmic**

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

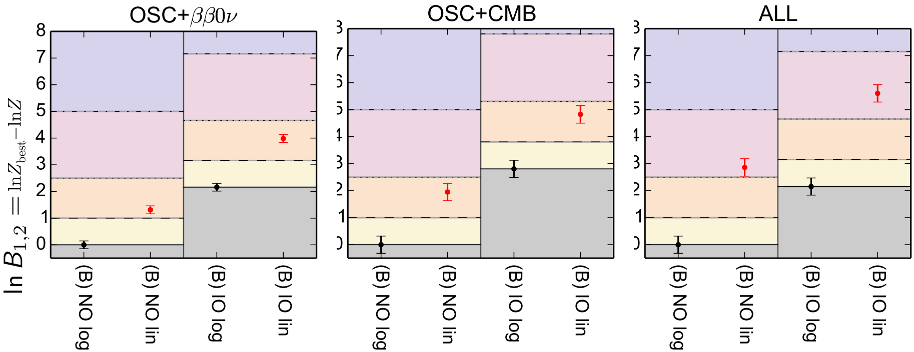


compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

Comparing parameterizations/priors

[SG+, JCAP 03 (2018) 11]

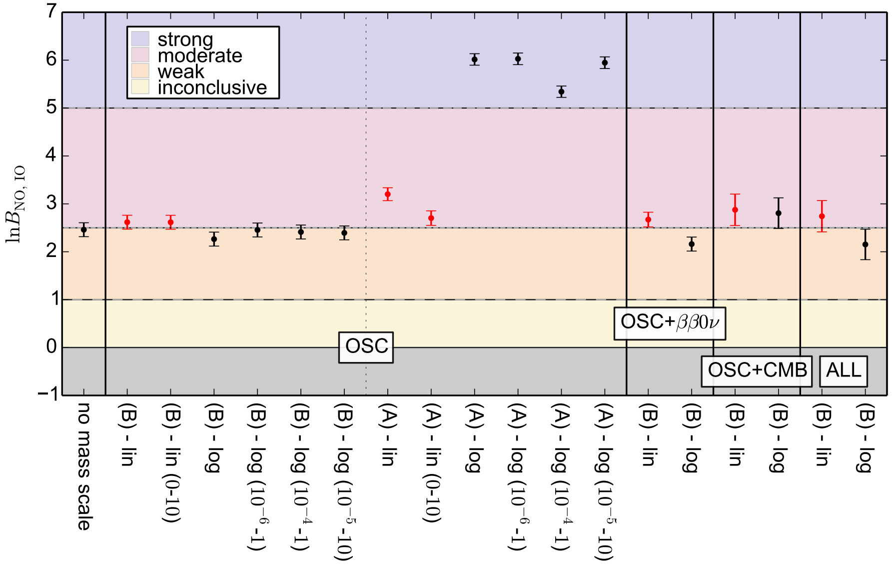


compare linear versus **logarithmic**

log priors are
weakly-to-moderately more efficient

summary: case B, log prior is better!

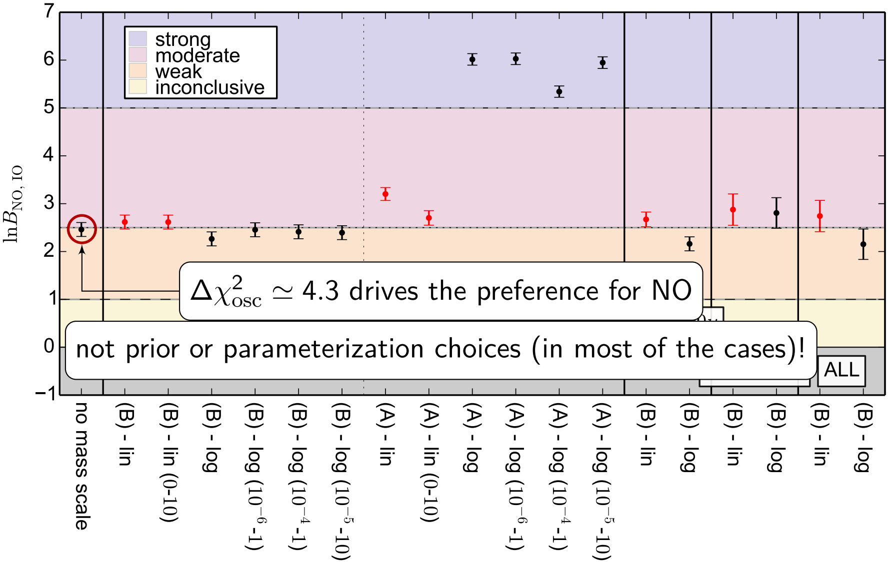
Comparing the mass orderings



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

Comparing the mass orderings

[SG+, JCAP 03 (2018) 11]



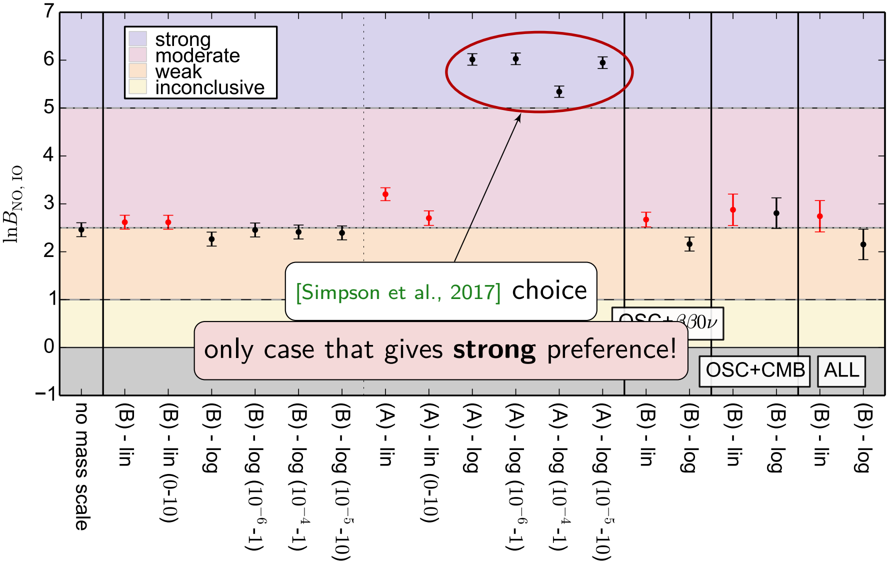
$\Delta\chi^2_{\rm osc} \simeq 4.3$ drives the preference for NO

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

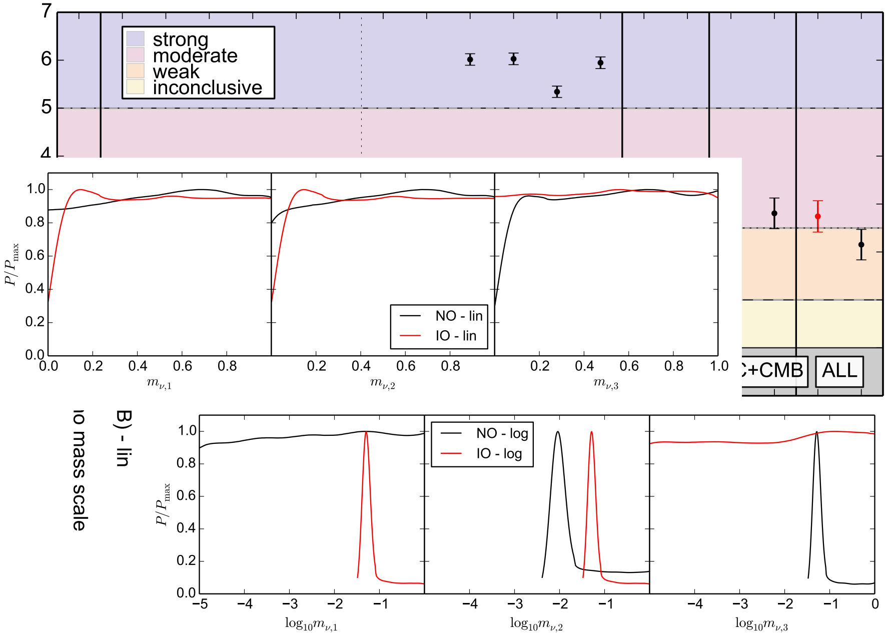
[SG+, JCAP 03 (2018) 11]



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

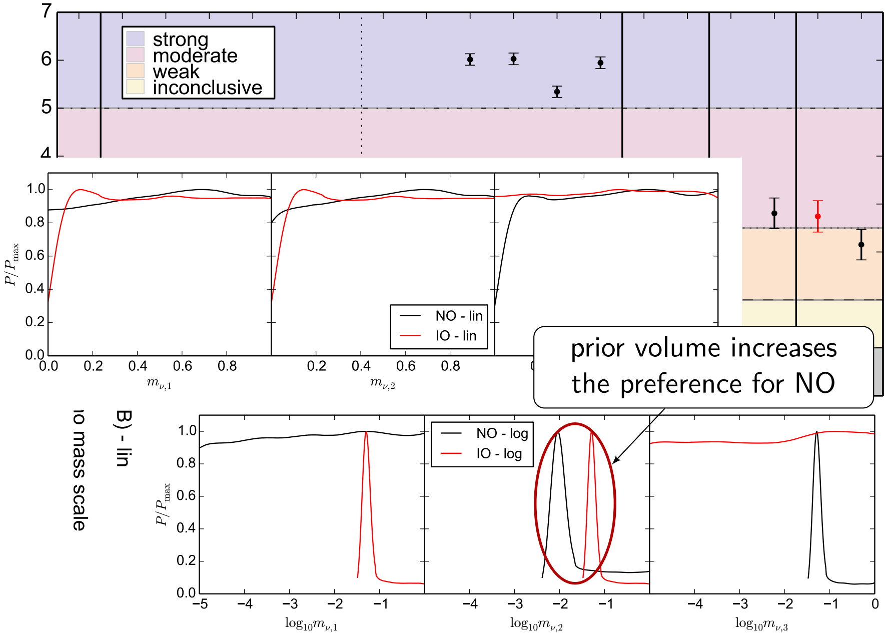
Comparing the mass orderings

[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:

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

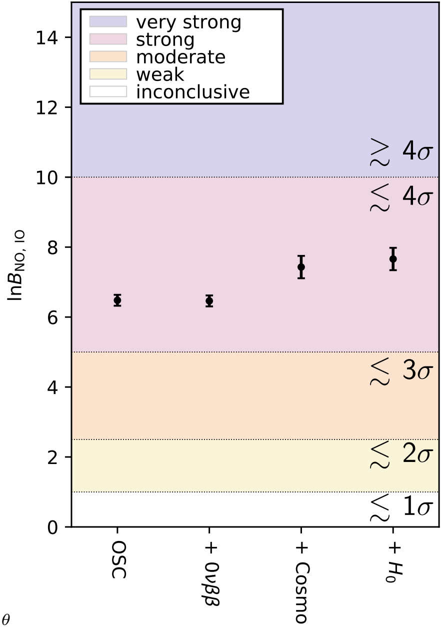
Bayes factor NO vs IO:

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

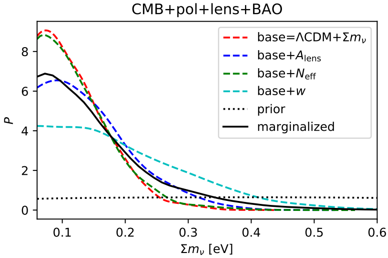
Posterior probability:

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

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



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


CMB+pol+lens+BAO
base=ΛCDM+Σm_ν
base+A_{lens}
base+N_{eff}
base+w
prior
marginalized
P
0
2
4
6
8
0.1
0.2
0.3
0.4
0.5
0.6
Σm_ν [eV]

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:

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

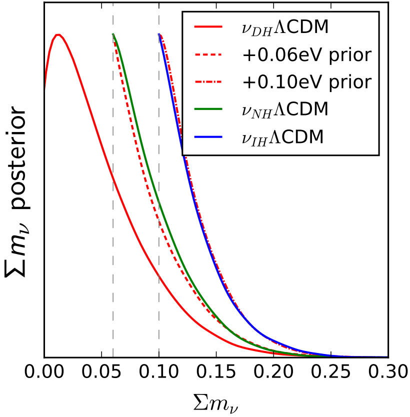
posterior depends on prior!

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:

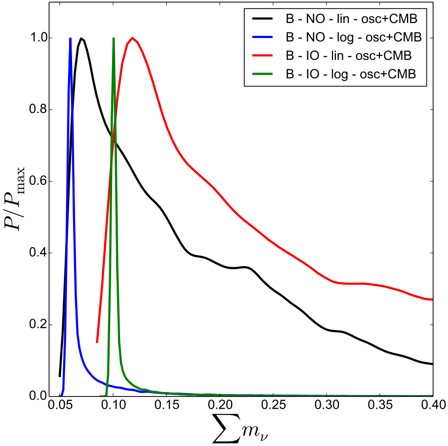
$$p(\theta|d,\mathcal{M}) = \mathcal{L}(\theta)\frac{\pi(\theta|\mathcal{M})}{Z_\mathcal{M}}$$

posterior depends on prior!

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

[SG+, 2018]

logarithmic vs linear prior on $m_{\rm lightest}$



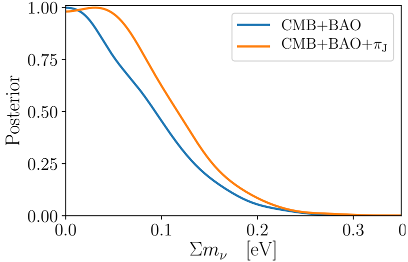
parameters θ, model $\mathcal{M}$, data d $\pi(\theta|\mathcal{M})$ prior $p(\theta|d,\mathcal{M})$ posterior $\mathcal{L}(\theta)$ likelihood $Z_\mathcal{M}$ Bayesian evidence

Playing with priors

Bayes theorem:

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

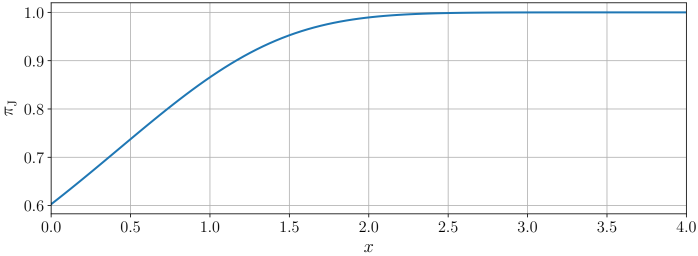
posterior depends on prior!



[Hannestad+, 2017]

Jeffreys prior (π_J) for Σm_ν

π_J makes the posterior maximally sensitive to data
for constrained parameter, compensate border effect



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**		CMB TT + lens + BAO **CMB TT,TE,EE + BAO**
[Planck 2015] ΛCDM	$\Sigma m_\nu < 0.68$ 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]
[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**	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
$\Sigma m_\nu < 0.96$ eV	[Di Valentino+, 2015] eCDM 12-parameters cosmological model, ΛCDM based	$\Sigma m_\nu < 0.53$ eV

Marginalize over models?

[SG+, 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?

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:

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

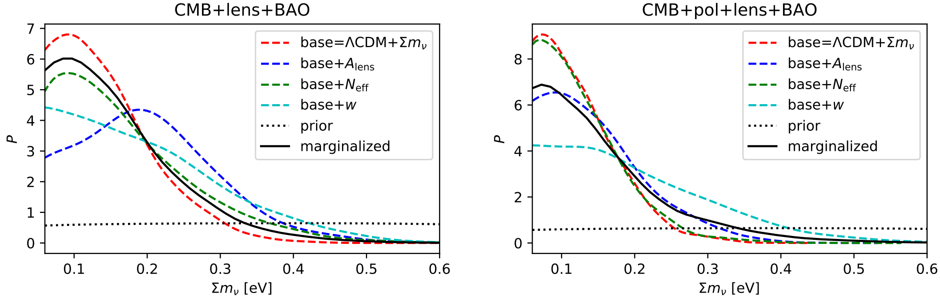
Select a model $\mathcal{M}_0$ and use $p_i = Z_i/(\sum Z_j) = B_{i0}/(\sum B_{j0})$:

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

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

Model-marginalization applied to Σm_ν

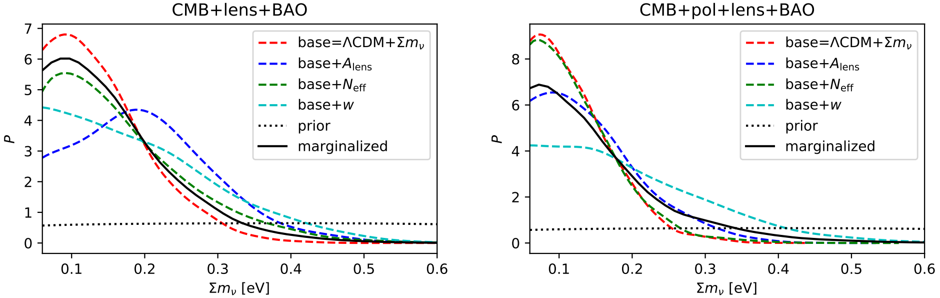
[SG+, PRD 99 (2019) 021301]



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

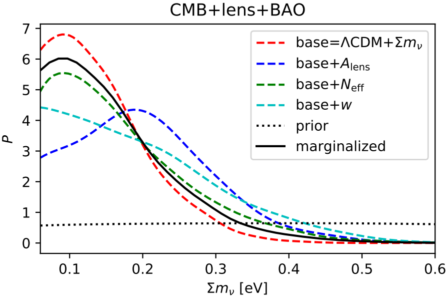
[SG+, PRD 99 (2019) 021301]

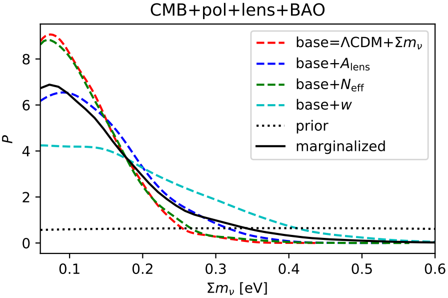


model	CMB+lens+BAO $\ln B_{i0}$	CMB+lens+BAO Σm_ν [eV]	CMB+pol+lens+BAO $\ln B_{i0}$	CMB+pol+lens+BAO Σm_ν [eV]
base=ΛCDM+Σm_ν	0.0	< 0.28	0.0	< 0.23
base+$A_{\rm lens}$	-2.6	< 0.38	-2.4	< 0.29
base+$N_{\rm eff}$	-1.5	< 0.37	-2.3	< 0.25
base+w	-1.4	< 0.42	-0.1	< 0.42
marginalized	–	< 0.33	–	< 0.35
p_0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

Bayes theorem (again!): $p(\theta|d,\mathcal{M}_i) = \pi(\theta|\mathcal{M}_i)\,\mathcal{L}_{\mathcal{M}_i}(\theta)/\,Z_i$

⟶ function of x

We usually present 1-dim marginalized posterior distributions:

⟶ over params ψ

$$p(x|d,\mathcal{M}_i) = \int_{\Omega_\psi} d\psi\, p(x,\psi|\mathcal{M}_i,d)$$

Prior-independent Bayesian parameter constraints - I

Bayes theorem (again!): $p(\theta|d,\mathcal{M}_i) = \pi(\theta|\mathcal{M}_i)\,\mathcal{L}_{\mathcal{M}_i}(\theta)/\,Z_i$

We usually present 1-dim marginalized posterior distributions:

$$p(x|d,\mathcal{M}_i) = \int_{\Omega_\psi} d\psi\, p(x,\psi|\mathcal{M}_i,d)$$

Assume that prior is separable: $\pi(\theta|\mathcal{M}_i) = \pi(x|\mathcal{M}_i)\cdot\pi(\psi|\mathcal{M}_i)$
and that $\pi(x) \equiv \pi(x|\mathcal{M}_i)$ does not depend on $\mathcal{M}_i$

$$p(x|d,\mathcal{M}_i) = \frac{\pi(x)}{Z_i}\int_{\Omega_\psi} d\psi\,\pi(\psi|\mathcal{M}_i)\mathcal{L}_{\mathcal{M}_i}(x,\psi)$$

$\equiv Z_i^x$ Bayesian evidence of model $\mathcal{M}_i|_{\text{fixed}\,x}$
independent of $\pi(x)$ but not of x

Prior-independent Bayesian parameter constraints - I

Bayes theorem (again!): $p(\theta|d,\mathcal{M}_i) = \pi(\theta|\mathcal{M}_i)\,\mathcal{L}_{\mathcal{M}_i}(\theta)/\,Z_i$

We usually present 1-dim marginalized posterior distributions:

$$p(x|d,\mathcal{M}_i) = \int_{\Omega_\psi} d\psi\, p(x,\psi|\mathcal{M}_i,d)$$

Assume that prior is separable: $\pi(\theta|\mathcal{M}_i) = \pi(x|\mathcal{M}_i)\cdot\pi(\psi|\mathcal{M}_i)$
and that $\pi(x) \equiv \pi(x|\mathcal{M}_i)$ does not depend on $\mathcal{M}_i$

$$p(x|d,\mathcal{M}_i) = \frac{\pi(x)}{Z_i}\underbrace{\int_{\Omega_\psi} d\psi\,\pi(\psi|\mathcal{M}_i)\mathcal{L}_{\mathcal{M}_i}(x,\psi)}_{\equiv Z_i^x}$$

[SG+, PRD 99 (2019) 021301]

Model marginalization: $p(x|d) = \sum_i p(x|\mathcal{M}_i,d)\,Z_i \Big/ \sum_j Z_j$

Replace $p(x|\mathcal{M}_i,d)$: $p(x|d) = \pi(x)\underbrace{\sum_i Z_i^x \Big/ \sum_j Z_j}_{\text{independent of } \pi(x)?}$

Prior-independent Bayesian parameter constraints - I

Bayes theorem (again!): $p(\theta|d,\mathcal{M}_i) = \pi(\theta|\mathcal{M}_i)\,\mathcal{L}_{\mathcal{M}_i}(\theta)/\,Z_i$

We usually present 1-dim marginalized posterior distributions:

$$p(x|d,\mathcal{M}_i) = \int_{\Omega_\psi} d\psi\, p(x,\psi|\mathcal{M}_i,d)$$

Assume that prior is separable: $\pi(\theta|\mathcal{M}_i) = \pi(x|\mathcal{M}_i)\cdot\pi(\psi|\mathcal{M}_i)$
and that $\pi(x) \;\equiv\; \pi(x|\mathcal{M}_i)$ does not depend on $\mathcal{M}_i$

$$p(x|d,\mathcal{M}_i) = \frac{\pi(x)}{Z_i}\underbrace{\int_{\Omega_\psi} d\psi\,\pi(\psi|\mathcal{M}_i)\mathcal{L}_{\mathcal{M}_i}(x,\psi)}_{\equiv Z_i^x}$$

[SG+, PRD 99 (2019) 021301]

Model marginalization: $p(x|d) = \sum_i p(x|\mathcal{M}_i,d)\,Z_i \Big/ \sum_j Z_j$

Replace $p(x|\mathcal{M}_i,d)$: $p(x|d) = \pi(x)\sum_i Z_i^x \Big/ \sum_j Z_j$

independent of $\pi(x)$!

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

$$\mathcal{R}(x,x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)} = \frac{\sum_i Z_i^x}{\sum_j Z_j^{x_0}}$$

Prior-independent Bayesian parameter constraints - II

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

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)} = \frac{\sum_i Z_i^x}{\sum_j Z_j^{x_0}}$$

independent of $\pi(x)$!

Rewrite in a more familiar form:

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

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

Prior-independent Bayesian parameter constraints - II

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

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)} = \frac{\sum_i Z_i^x}{\sum_j Z_j^{x_0}}$$

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 - II

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

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)} = \frac{\sum_i Z_i^x}{\sum_j Z_j^{x_0}}$$

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 - II

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

$$\mathcal{R}(x, x_0|d) \equiv \frac{p(x|d)/\pi(x)}{p(x_0|d)/\pi(x_0)} = \frac{\sum_i Z_i^x}{\sum_j Z_j^{x_0}}$$

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

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

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

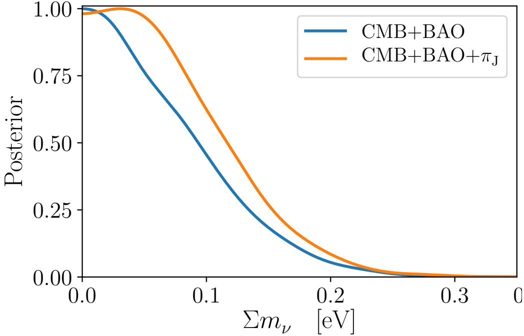
An example with Planck 2018

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

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. [Hannestad+, 2017]

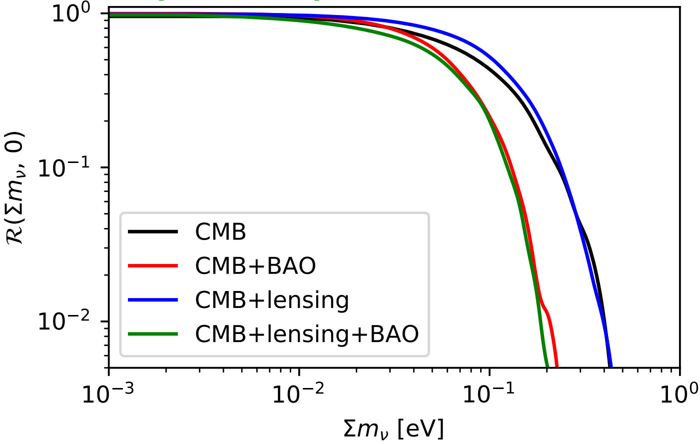


An example with Planck 2018

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

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

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

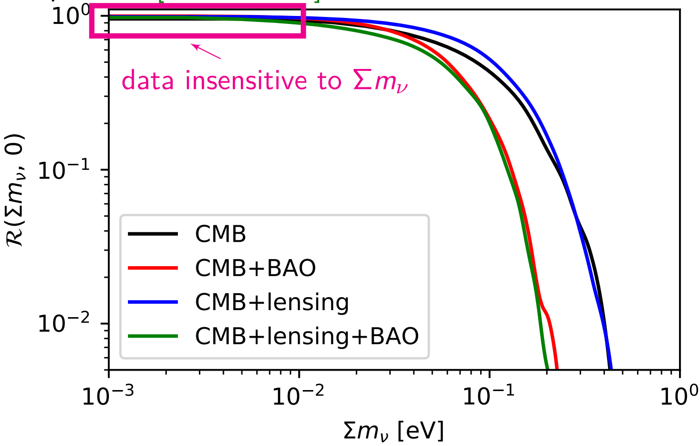


An example with Planck 2018

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

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

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

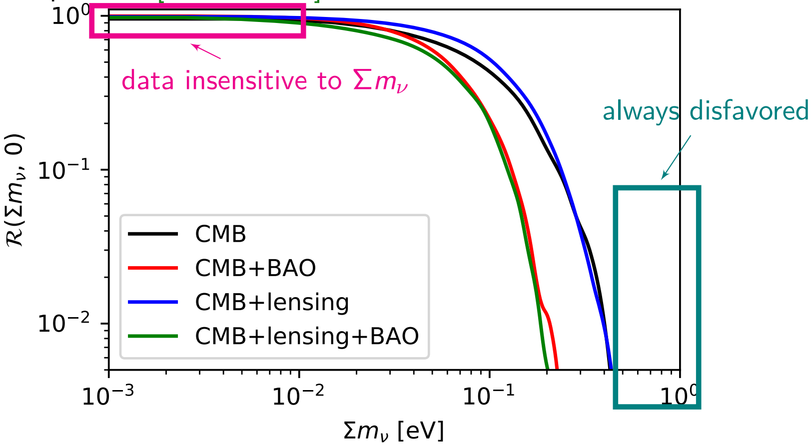


An example with Planck 2018

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

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

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

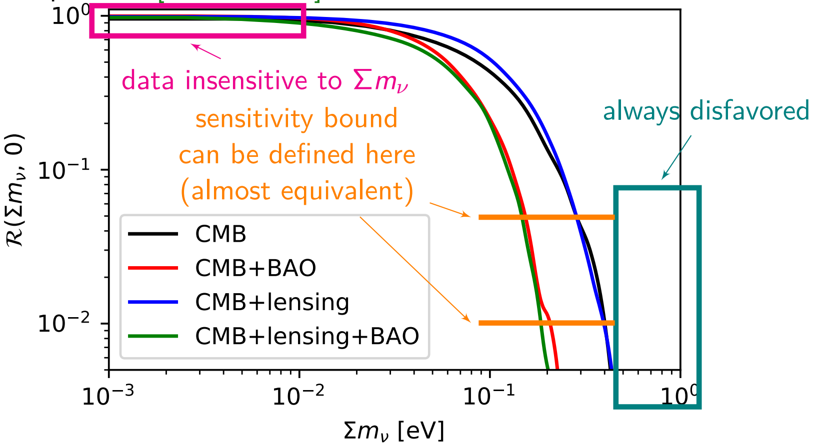


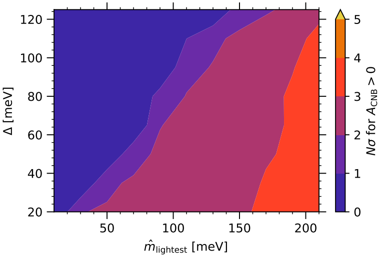
An example with Planck 2018

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

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

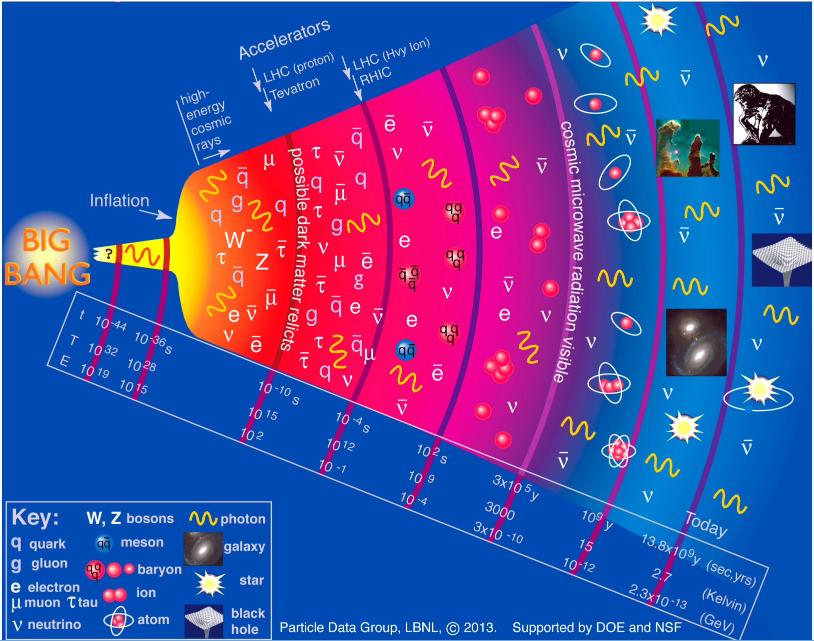
Example with [Planck 2018] chains from PLA, ΛCDM+Σm_ν



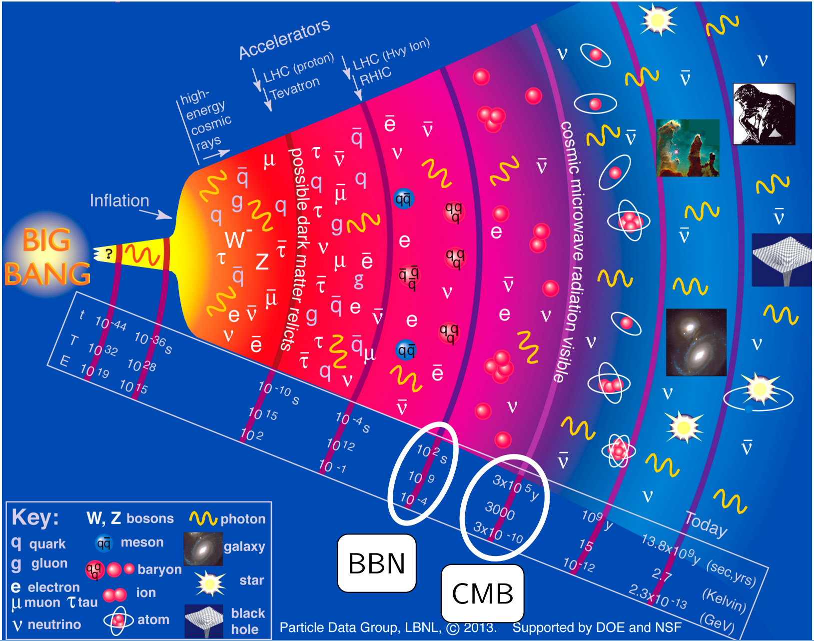


120
100
80
60
40
20
Δ [meV]
50
100
150
200
$\hat{m}_{\rm lightest}$ [meV]
0
1
2
3
4
5
$N\sigma$ for $A_{\rm CNB} > 0$

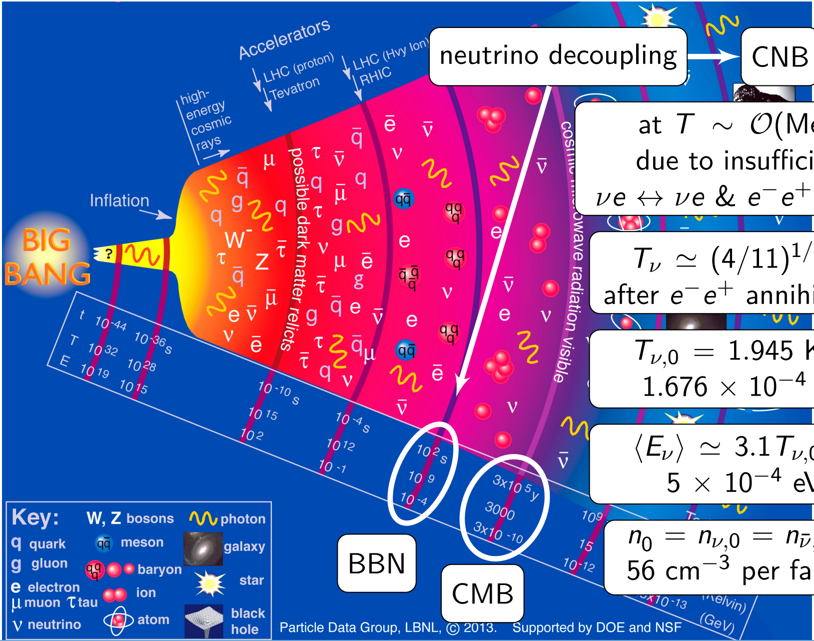
History of the universe



History of the universe



History of the universe



neutrino decoupling → CNB

at $T \sim \mathcal{O}(\text{MeV})$ due to insufficient $\nu e \leftrightarrow \nu e$ & $e^- e^+ \leftrightarrow \nu\bar\nu$

$T_\nu \simeq (4/11)^{1/3} T_\gamma$ after $e^- e^+$ annihilation

$T_{\nu,0} = 1.945$ K $\simeq 1.676 \times 10^{-4}$ eV

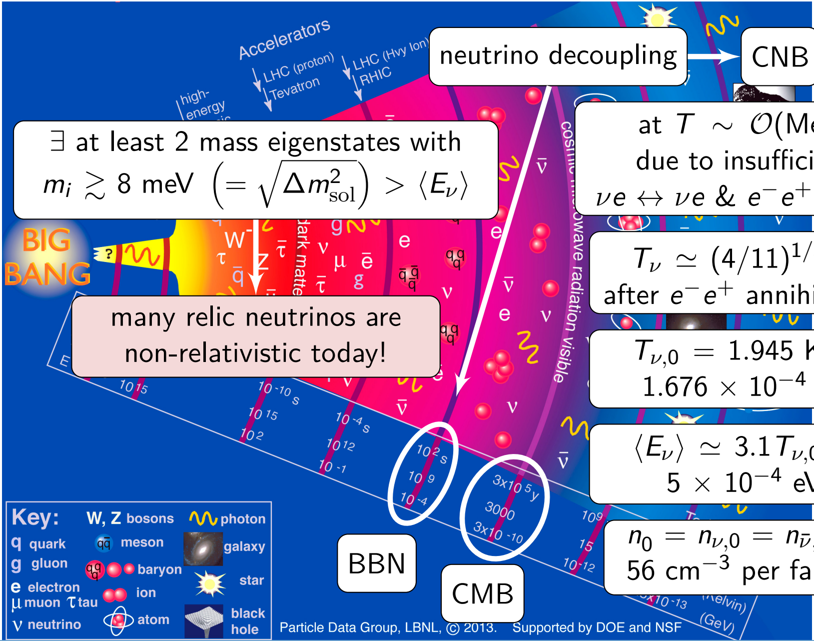
$\langle E_\nu \rangle \simeq 3.1\, T_{\nu,0} \simeq 5 \times 10^{-4}$ eV

$n_0 = n_{\nu,0} = n_{\bar\nu,0} \simeq 56$ cm^{-3} per family

BBN

CMB

History of the universe



neutrino decoupling → CNB

at $T \sim \mathcal{O}(\text{MeV})$
due to insufficient
$\nu e \leftrightarrow \nu e$ & $e^- e^+ \leftrightarrow \nu\bar{\nu}$

$T_\nu \simeq (4/11)^{1/3} T_\gamma$
after $e^- e^+$ annihilation

$T_{\nu,0} = 1.945$ K $\simeq$
1.676×10^{-4} eV

$\langle E_\nu \rangle \simeq 3.1\, T_{\nu,0} \simeq$
5×10^{-4} eV

$n_0 = n_{\nu,0} = n_{\bar{\nu},0} \simeq$
56 cm^{-3} per family

$\exists$ at least 2 mass eigenstates with
$m_i \gtrsim 8$ meV $\left(= \sqrt{\Delta m^2_{\rm sol}}\right) > \langle E_\nu \rangle$

many relic neutrinos are
non-relativistic today!

BBN

CMB

A viable detection method

[Long et al., JCAP 08 (2014) 038]

How to directly detect non-relativistic neutrinos?

Remember that $\langle E_\nu \rangle \simeq \mathcal{O}(10^{-4})$ eV today $\longrightarrow$ a process without energy threshold is necessary

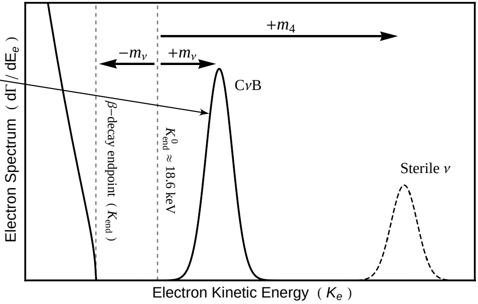
[Weinberg, 1962]: neutrino capture in β–decaying nuclei $\nu + n \to p + e^-$

Main background: β decay $n \to p + e^- + \bar{\nu}$!

signal is a peak at $2m_\nu$ above β–decay endpoint

only with a lot of material

need a very good energy resolution



PTOLEMY

[Long et al., JCAP 08 (2014) 038]
[Betts et al., arxiv:1307.4738]
[PTOLEMY LoI, arxiv:1808.01892]

PonTecorvo Observatory for Light, Early-universe, Massive-neutrino Yield (PTOLEMY)

expected resolution $\Delta \simeq 0.1$ eV? 0.05 eV?

built mainly for CNB

can probe $m_\nu \simeq 1.4\Delta \simeq 0.1$ eV

$M_T = 100$ g of atomic ^{3}H

$$\Gamma_{\rm CNB} = \sum_{i=1}^{3} |U_{ei}|^2 [n_i(\nu_{h_R}) + n_i(\nu_{h_L})]\, N_T\, \bar\sigma \quad \sim \mathcal{O}(10)\ \text{yr}^{-1}$$

N_T number of ^{3}H nuclei in a sample of mass M_T $\quad \bar\sigma \simeq 3.834 \times 10^{-45}$ cm^2 $\quad n_i$ number density of neutrino i

(without clustering)

PTOLEMY

[Long et al., JCAP 08 (2014) 038]
[Betts et al., arxiv:1307.4738]
[PTOLEMY LoI, arxiv:1808.01892]

PonTecorvo Observatory for Light, Early-universe, Massive-neutrino Yield (PTOLEMY)

expected resolution $\Delta \simeq 0.1$ eV? 0.05 eV? ↔ built mainly for CNB

can probe $m_\nu \simeq 1.4\Delta \simeq 0.1$ eV

$M_T = 100$ g of atomic ^{3}H

ehnancement from ν clustering in the galaxy?

$$\Gamma_{\text{CNB}} = \sum_{i=1}^{3} |U_{ei}|^2 [n_i(\nu_{h_R}) + n_i(\nu_{h_L})] \, N_T \, \bar{\sigma} \quad \sim \mathcal{O}(10) \text{ yr}^{-1}$$

N_T number of ^{3}H nuclei in a sample of mass M_T $\bar{\sigma} \simeq 3.834 \times 10^{-45}$ cm^2 n_i number density of neutrino i

(without clustering)

ν clustering with N-one-body simulations

[JCAP 09 (2017) 034]

Milky Way (MW) matter attracts neutrinos!

clustering →

$$\Gamma_{\rm CNB} = \sum_{i=1}^{3} |U_{ei}|^2 f_c(m_i) \left[n_{i,0}(\nu_{h_R}) + n_{i,0}(\nu_{h_L})\right] N_T \bar{\sigma}$$

$f_c(m_i) = n_i/n_{i,0}$ clustering factor ⟶ How to compute it?

Idea from [Ringwald & Wong, 2004] ⟶ N-one-body= N × single ν simulations

→ each ν evolved from initial conditions at $z = 3$

→ spherical symmetry, coordinates $(r,\ \theta,\ p_r,\ l)$

→ need $\rho_{\rm matter}(z) = \rho_{\rm DM}(z) + \rho_{\rm baryon}(z)$

Assumptions:

- νs are independent
- only gravitational interactions
- νs do not influence matter evolution ($\rho_\nu \ll \rho_{\rm DM}$)

how many νs is "N"?

→ must sample all possible r, p_r, l

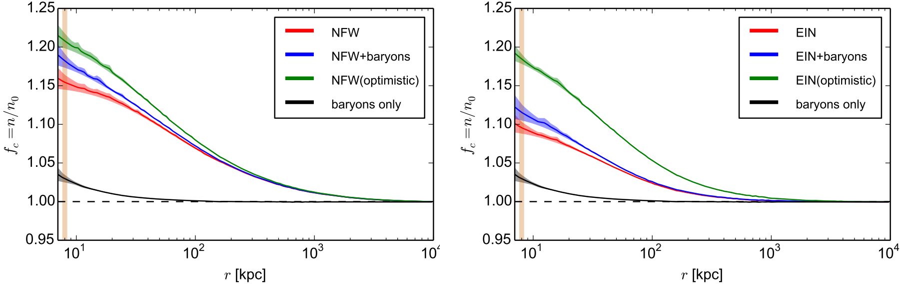
→ must include all possible νs that reach the MW (fastest ones may come from several (up to $\mathcal{O}(100)$) Mpc!)

given N ν:

→ weigh each neutrinos

→ reconstruct final density profile with kernel method from [Merritt&Tremblay, 1994]

Overdensity when $m_{\text{heaviest}} \simeq 60$ meV



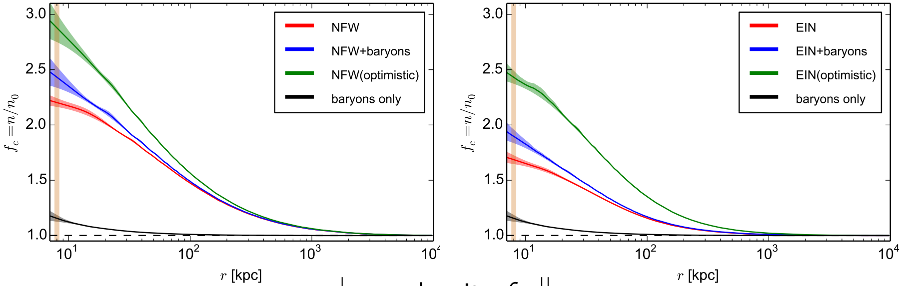
masses	ordering	matter halo	overdensity f_c: $f_1 \simeq f_2$	f_3	Γ_{tot} (yr^{-1})
any	any	any	no clustering		4.06
$m_3 = 60$ meV	NO	NFW(+bar)	$\sim$1	1.15 (1.18)	4.07 (4.08)
		NFW optimistic		1.21	4.08
		EIN(+bar)		1.09 (1.12)	4.07 (4.07)
		EIN optimistic		1.18	4.08
$m_1 \simeq m_2 = 60$ meV	IO	NFW(+bar)	1.15 (1.18)	$\sim$1	4.66 (4.78)
		NFW optimistic	1.21		4.89
		EIN(+bar)	1.09 (1.12)		4.42 (4.54)
		EIN optimistic	1.18		4.78

ordering dependence from $\Gamma_{\text{CNB}} = \sum_{i=1}^{3} |U_{ei}|^2 f_i \left[n_i(\nu_{h_R}) + n_i(\nu_{h_L})\right] N_T \bar{\sigma}$

Overdensity when $m_\nu \simeq 150$ meV

[JCAP 09 (2017) 034]

$\Longrightarrow$ minimal mass detectable by PTOLEMY if $\Delta \simeq 100$–150 meV



matter halo	overdensity f_c $f_1 \simeq f_2 \simeq f_3$	$\Gamma_{\rm tot}$ (yr^{-1})
any	no clustering	4.06
NFW(+bar)	2.18 (2.44)	8.8 (9.9)
NFW optimistic	2.88	11.7
EIN(+bar)	1.68 (1.87)	6.8 (7.6)
EIN optimistic	2.43	9.9

no ordering dependence: $m_1 \simeq m_2 \simeq m_3 \Longrightarrow f_1 \simeq f_2 \simeq f_3$

Additional clustering due to other galaxies

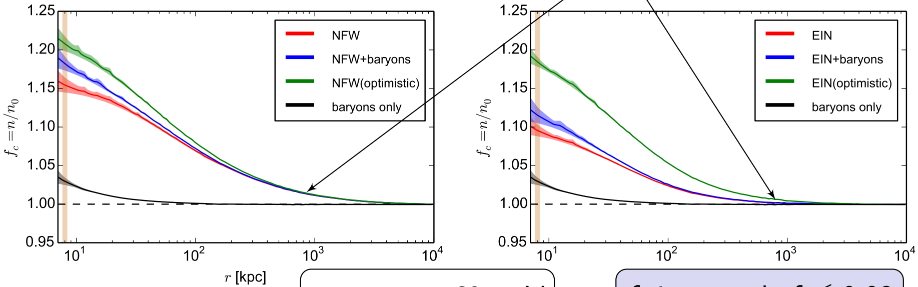
nearest galaxies: various MW satellites

with $M_{\rm sat} \ll M_{\rm MW}$ → negligibly small ν halo

nearest big galaxy:

Andromeda

$$M_{\rm Andromeda} = M_{\rm MW} \times \mathcal{O}(1) \;\text{—}\; d_{\rm Andromeda} \simeq 800 \text{ kpc}$$



$m_{\rm heaviest} \simeq 60$ meV → f_c increased of $\lesssim 0.03$

Additional clustering due to other galaxies

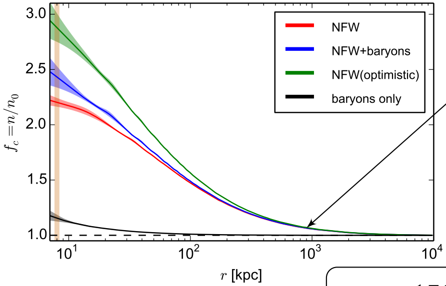
nearest galaxies: various MW satellites

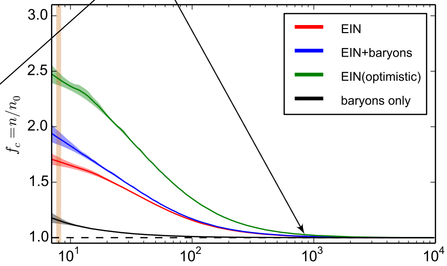
with $M_{\rm sat} \ll M_{\rm MW}$ ⟶ negligibly small ν halo

nearest big galaxy:

Andromeda

$$M_{\rm Andromeda} = M_{\rm MW} \times \mathcal{O}(1) \text{ — } d_{\rm Andromeda} \simeq 800 \text{ kpc}$$





$m_\nu \simeq 150$ meV ⟶ f_c increased of $\lesssim 0.1$

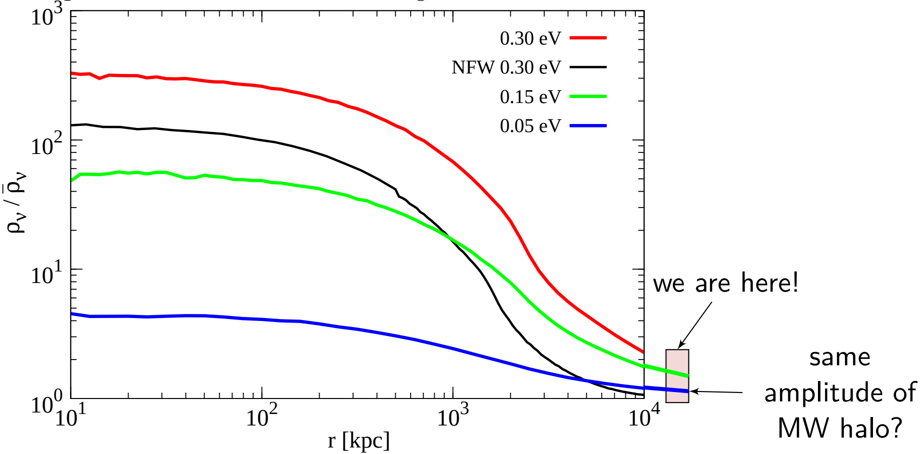
(halo is less diffuse for higher ν masses)

Additional clustering due to Virgo cluster

nearest galaxy cluster:

Virgo cluster → very wide ν halo, may reach Earth

$M_{\rm Virgo} = M_{\rm MW} \times \mathcal{O}(10^3)$ — $d_{\rm Virgo} \simeq 16$ Mpc



[Villaescusa-Navarro et al., JCAP 1106 (2011) 027]

Simulations - I

[PTOLEMY, in preparation]

Events in **bin** i, centered at E_i:

$$N^i_\beta = T\int_{E_i-\Delta/2}^{E_i+\Delta/2}\frac{d\widetilde{\Gamma}_\beta}{dE_e}dE_e \qquad N^i_{\rm CNB} = T\int_{E_i-\Delta/2}^{E_i+\Delta/2}\frac{d\widetilde{\Gamma}_{\rm CNB}}{dE_e}dE_e$$

fiducial number of events: $\hat{N}^i = N^i_\beta(\hat{E}_{\rm end},\hat{m}_i,\hat{U}) + N^i_{\rm CNB}(\hat{E}_{\rm end},\hat{m}_i,\hat{U})$

add **background** $\hat{N}_b = \hat{\Gamma}_b T$ with $\hat{\Gamma}_b \simeq 10^{-5}$ Hz $\longrightarrow$ $N^i_t = \hat{N}^i + \hat{N}_b$

simulated experimental spectrum: $N^i_{\rm exp}(\hat{E}_{\rm end},\hat{m}_i,\hat{U}) = N^i_t \pm \sqrt{N^i_t}$

repeat for theory spectrum, free **amplitudes** and **endpoint position**:

$$N^i_{\rm th}(\boldsymbol{\theta}) = \boldsymbol{A_\beta} N^i_\beta(\hat{E}_{end} + \boldsymbol{\Delta E_{end}}, m_i, U) + \boldsymbol{A_{\rm CNB}} N^i_{\rm CNB}(\hat{E}_{end} + \boldsymbol{\Delta E_{end}}, m_i, U) + N_b$$

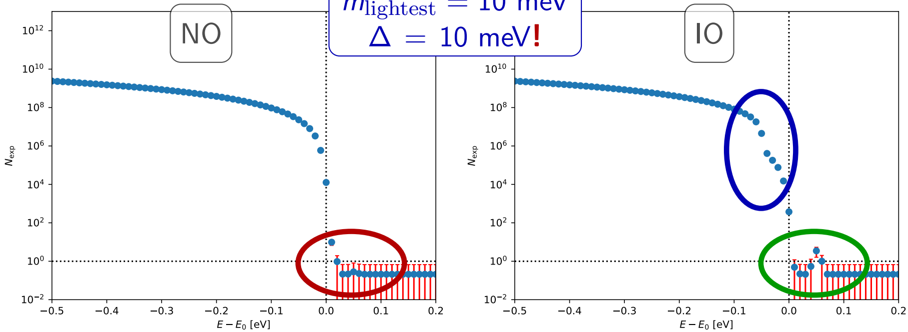
fit $\longrightarrow$ $\chi^2(\boldsymbol{\theta}) = \sum_i \left(\frac{N^i_{\rm exp}(\hat{E}_{\rm end},\hat{m}_i,\hat{U}) - N^i_{\rm th}(\boldsymbol{\theta})}{\sqrt{N^i_t}}\right)^2$ or $\log\mathcal{L} = -\frac{\chi^2}{2}$

T exposure time – $(\hat{E}_{\rm end}, \hat{m}_i, \hat{U})$ fiducial endpoint energy, masses, mixing matrix – $\boldsymbol{\theta} = (A_\beta, N_b, \Delta E_{end}, A_{\rm CNB}, m_i, U)$

Simulations - II

no random noise?

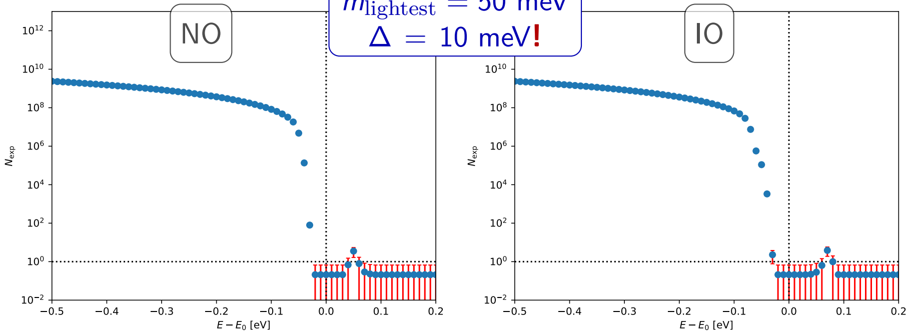
$m_{\text{lightest}} = 10$ meV
$\Delta = 10$ meV!



Simulations - II

no random noise?

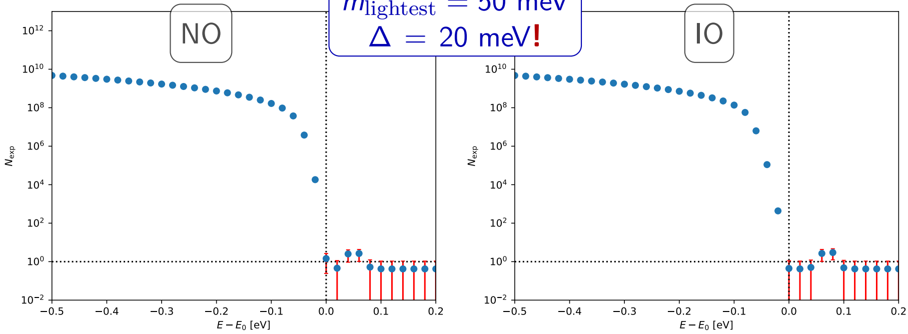
$m_{\text{lightest}} = 50$ meV
$\Delta = 10$ meV!



Simulations - II

no random noise?

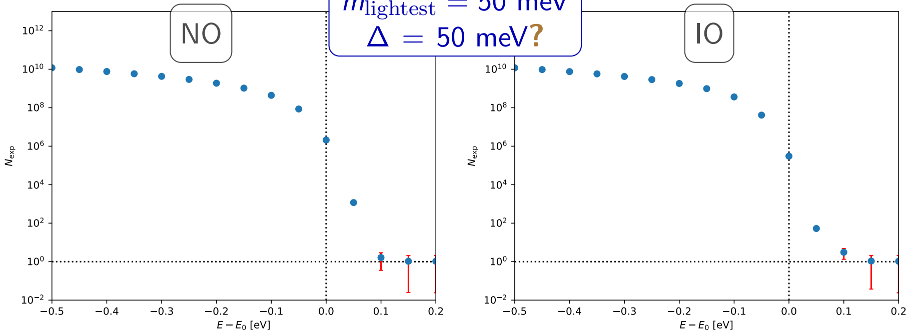
$m_{\rm lightest} = 50$ meV
$\Delta = 20$ meV!



Simulations - II

no random noise?

$m_{\text{lightest}} = 50$ meV
$\Delta = 50$ meV?

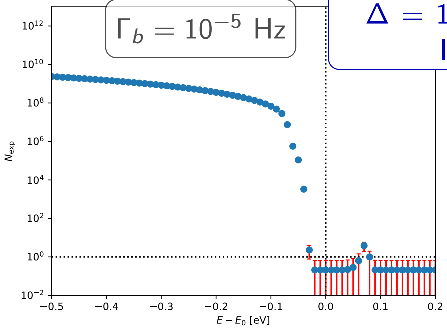


Simulations - II

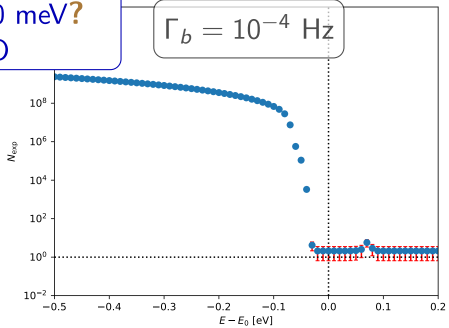
no random noise?

$m_{\text{lightest}} = 50$ meV
$\Delta = 10$ meV?
IO

$\Gamma_b = 10^{-5}$ Hz

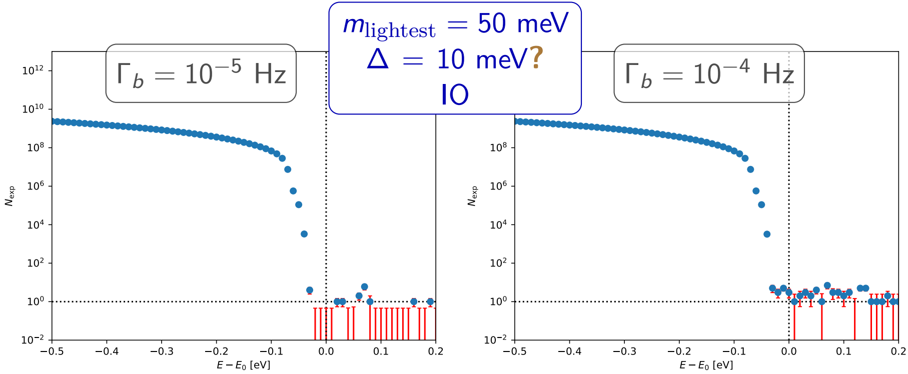


$\Gamma_b = 10^{-4}$ Hz



Simulations - II

with random noise!



things are more complicated in this way...low background needed!

Perspectives for the mass determination

[PTOLEMY, in preparation]

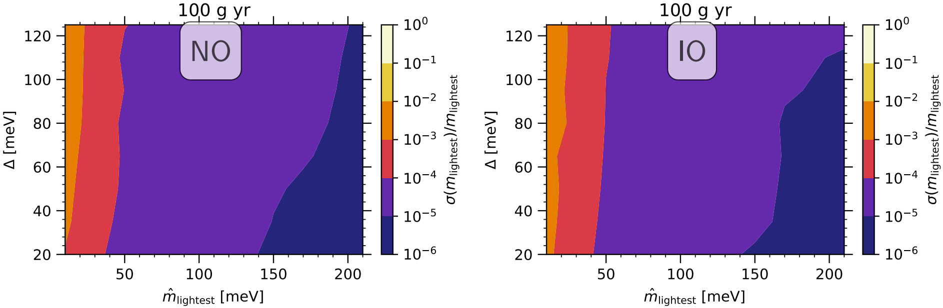
statistical only!

relative error on $m_{\rm lightest}$
as a function of $\hat{m}_{\rm lightest}$, Δ

Perspectives for the mass determination

statistical only!

relative error on $m_{\rm lightest}$
as a function of $\hat{m}_{\rm lightest}$, Δ



wonderful precision in determining the neutrino mass

(well, yes, with 100 g of tritium...)

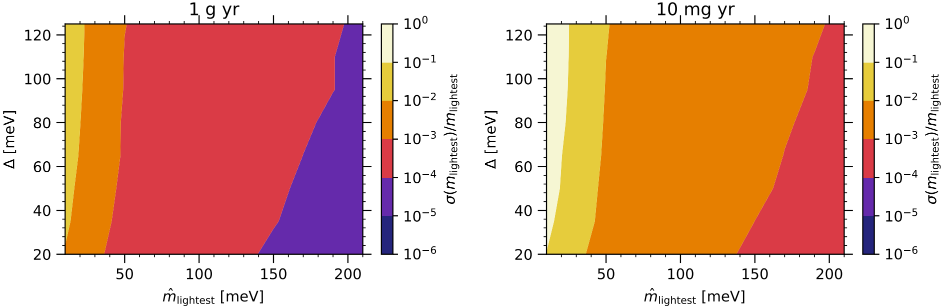
Perspectives for the mass determination

[PTOLEMY, in preparation]

statistical only!

relative error on $m_{\rm lightest}$

as a function of $\hat{m}_{\rm lightest}$, Δ



wonderful precision in determining the neutrino mass

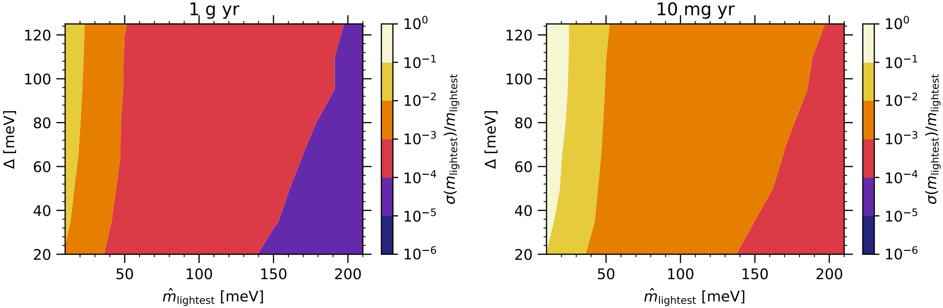
(mass detection already with 10 mg of tritium!)

Perspectives for the mass determination

[PTOLEMY, in preparation]

statistical only!

relative error on $m_{\rm lightest}$
as a function of $\hat{m}_{\rm lightest}$, Δ



wonderful precision in determining the neutrino mass

(mass detection already with 10 mg of tritium!)

Δ has almost no impact

Detection of the relic neutrinos

[PTOLEMY, in preparation]

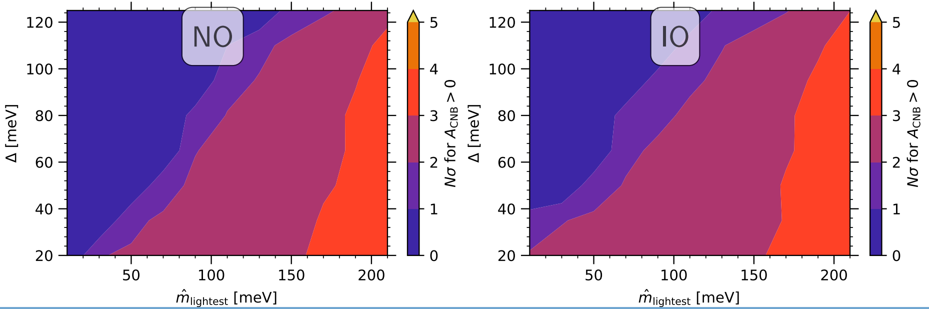
using the definition:

$$N^i_{\rm th}(\boldsymbol{\theta}) = A_\beta N^i_\beta(\hat{E}_{end} + \Delta E_{end}, m_i, U) + \boldsymbol{A}_{\rm CNB} N^i_{\rm CNB}(\hat{E}_{end} + \Delta E_{end}, m_i, U) + N_b$$

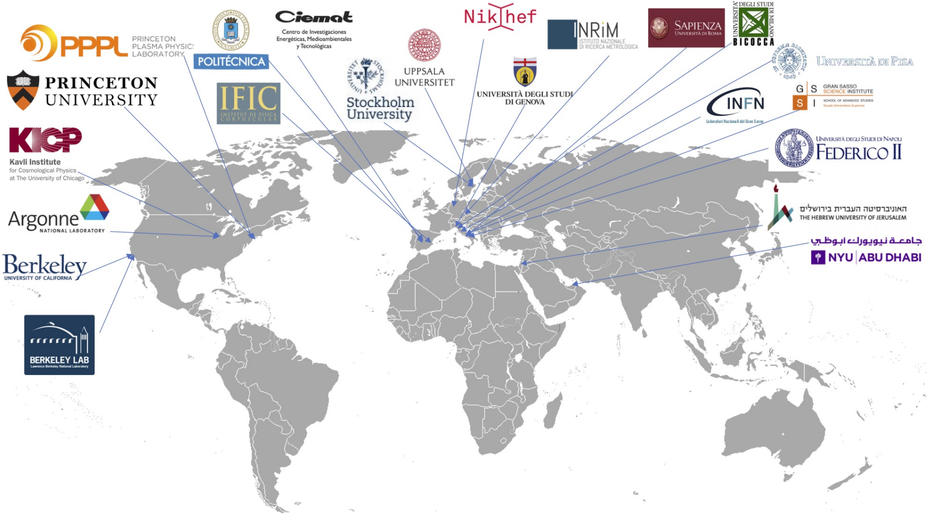
if $\boldsymbol{A}_{\rm CNB} > 0$ at $N\sigma$, direct detection of CNB accomplished at $N\sigma$

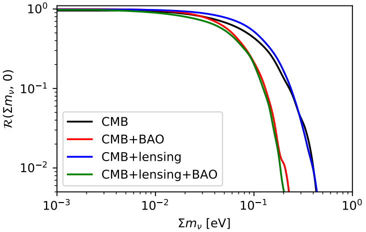
statistical only!

significance on $A_{\rm CNB} > 0$
as a function of $\hat{m}_{\rm lightest}$, Δ



PTOLEMY collaboration




$\mathcal{R}(\Sigma m_\nu, 0)$
Σm_ν [eV]
CMB
CMB+BAO
CMB+lensing
CMB+lensing+BAO

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

For the (far) future:
model independent neutrino properties (and more!)
from **direct detection of relic neutrinos**

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. For the (far) future:
model independent neutrino properties (and more!)
from **direct detection of relic neutrinos**

Thank you for the attention!