Statistical Estimation for Physics

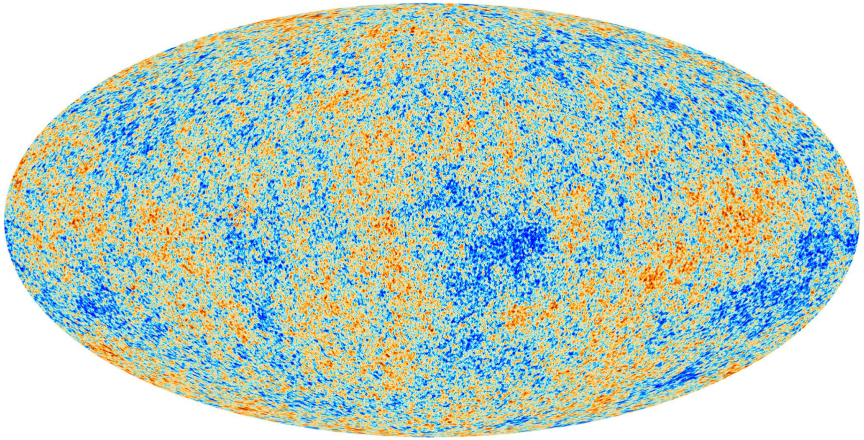
Bayesian Statistical Inference*

Boris Bolliet
Senior Research Associate
(*DAMTP*)

*[Link]() to GitHub repository

The Cosmic Microwave Background (CMB) radiation

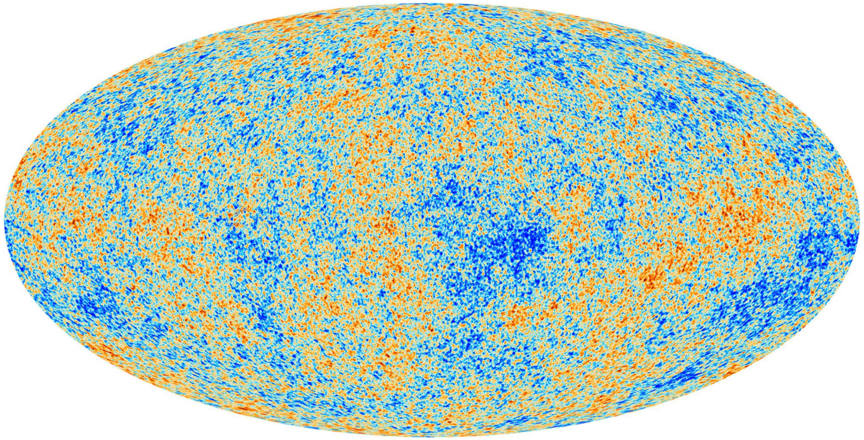
Full-sky map of the CMB temperature fluctuations



- Relic radiation from the Big Bang
- Average temperature $\bar{T} = 2.7\mathrm{K}$
- Fluctuations $\dfrac{\Delta T}{\bar{T}} \approx 10^{-5}$

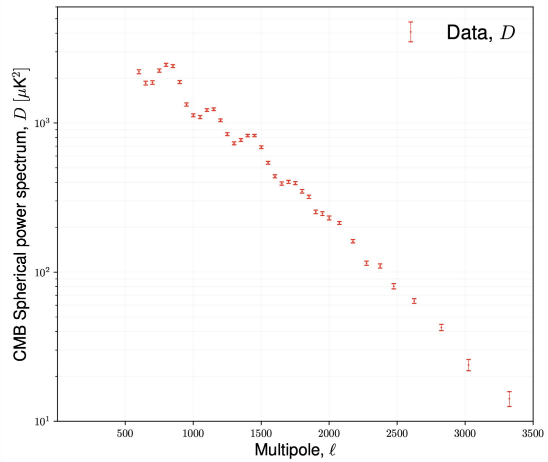
The Cosmic Microwave Background (CMB) radiation

Full-sky map of the CMB temperature fluctuations



- Relic radiation from the Big Bang
- Average temperature $\bar{T} = 2.7\mathrm{K}$
- Fluctuations $\dfrac{\Delta T}{\bar{T}} \approx 10^{-5}$

"Data compression"



large scales

small scales

A two-parameter model $D^{\mathrm{th}} = f(\mathrm{age}, M_\nu)$

- **Age of the universe** $\boxed{\mathrm{age}}$
    - Parameter of interest

A two-parameter model $D^{\mathrm{th}} = f(\mathrm{age}, M_\nu)$

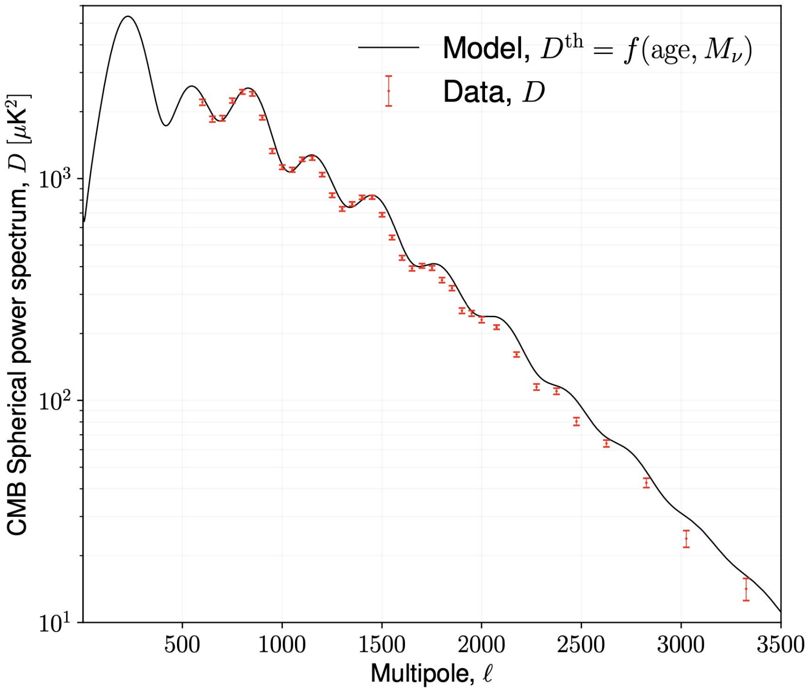
- **Age of the universe** $\boxed{\mathrm{age}}$
    - Parameter of interest
- **Total neutrino mass** $\boxed{M_\nu}$
    - *"Nuisance"* parameter

→ $\theta = \{\mathrm{age}, M_\nu\}$

A two-parameter model $D^{\mathrm{th}} = f(\mathrm{age}, M_\nu)$

- **Age of the universe** $\boxed{\mathrm{age}}$
  - Parameter of interest
- **Total neutrino mass** $\boxed{M_\nu}$
  - *"Nuisance"* parameter

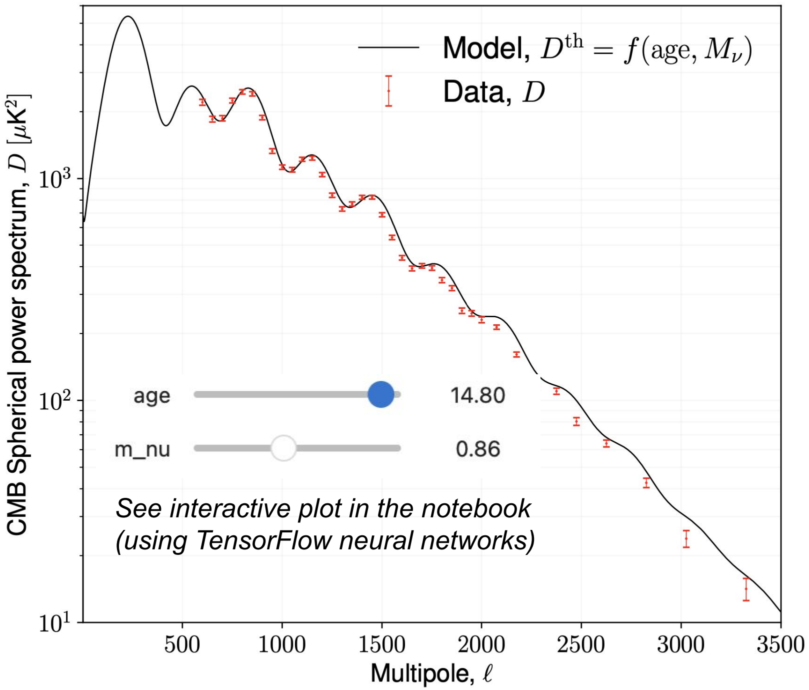
$$\rightarrow \quad \theta = \{\mathrm{age}, M_\nu\}$$



A two-parameter model $D^{\text{th}} = f(\text{age}, M_\nu)$

- **Age of the universe** $\boxed{\text{age}}$
    - Parameter of interest
- **Total neutrino mass** $\boxed{M_\nu}$
    - *"Nuisance"* parameter

$$\rightarrow \quad \theta = \{\text{age}, M_\nu\}$$

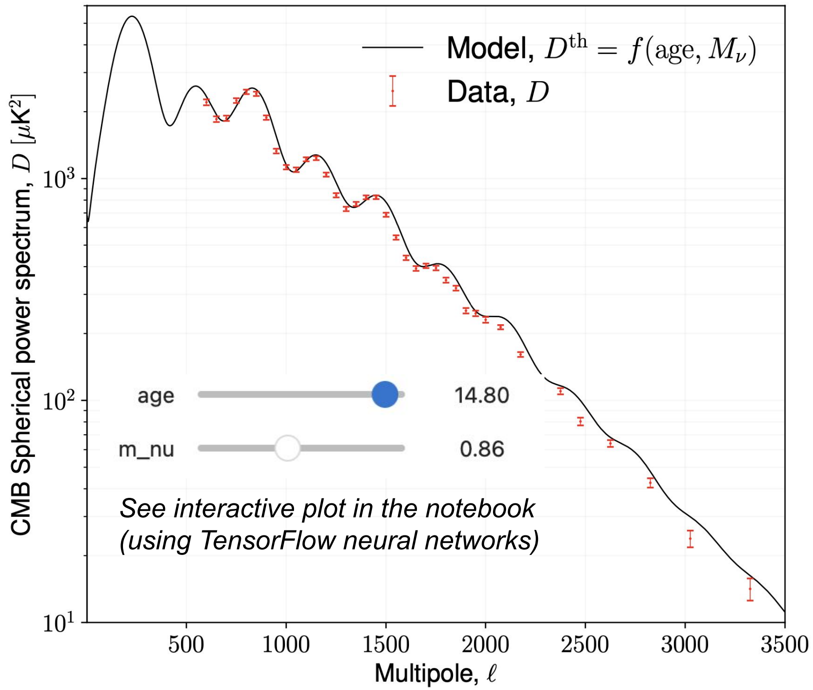


A two-parameter model $D^{\mathrm{th}} = f(\mathrm{age}, M_\nu)$

- **Age of the universe** $\boxed{\mathrm{age}}$
  - Parameter of interest
- **Total neutrino mass** $\boxed{M_\nu}$
  - *"Nuisance"* parameter

$$\rightarrow \quad \theta = \{\mathrm{age}, M_\nu\}$$

How old is the universe?



Bayes' theorem for Bayesian inference

- Bayes' theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

- Probability P: "*degree of belief*"

Bayes' theorem for Bayesian inference

- Bayes' theorem: $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Probability P: "*degree of belief*"

- For inference: $$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

D : data

$D^{\mathrm{th}} = f(\theta)$: model

$\boldsymbol{\theta}$: parameters

Bayes' theorem for Bayesian inference

- Bayes' theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

- Probability P: "*degree of belief*"

- For inference: $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$

- Evidence: normalizing constant

D : data

$D^{\mathrm{th}} = f(\theta)$: model

θ : parameters

Bayes' theorem for Bayesian inference

- Bayes' theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

- Probability P: "*degree of belief*"

- For inference: $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$

- Evidence: normalizing constant $\longrightarrow$ $\boxed{P(\theta|D) \propto P(D|\theta)P(\theta)}$

D : data

$D^{\mathrm{th}} = f(\theta)$: model

θ : parameters

The likelihood function

$$P(\theta|D) \propto P(D|\theta) P(\theta)$$

- *Definition:* probability of data for a given θ, a function of θ, $\mathcal{L}(\theta) = P(D|\theta)$
- Summarises experimental data
- Main focus of classical/frequentist analysis

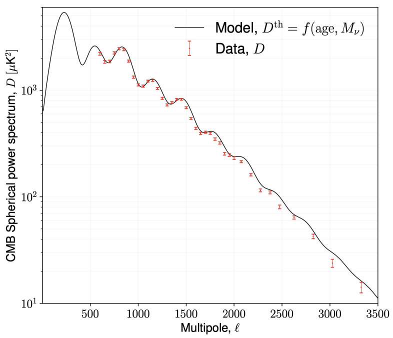
The CMB likelihood

- Gaussian approximation $P(D|\theta) = \frac{1}{\sqrt{2\pi|C|}} \exp\{-\frac{1}{2}(D^{\mathrm{th}}(\theta) - D)^T C^{-1} (D^{\mathrm{th}}(\theta) - D)\}$

The CMB likelihood

- Gaussian approximation $P(D|\theta) = \frac{1}{\sqrt{2\pi|C|}}\exp\{-\frac{1}{2}(D^{\rm th}(\theta) - D)^T C^{-1}(D^{\rm th}(\theta) - D)\}$

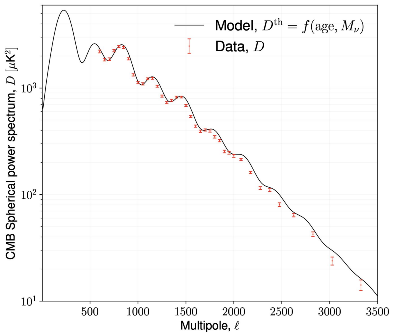
Model and data



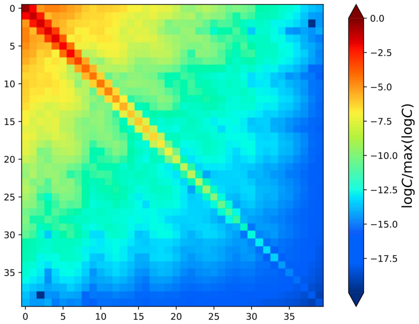
The CMB likelihood

- Gaussian approximation $$P(D|\theta) = \frac{1}{\sqrt{2\pi|C|}} \exp\{-\frac{1}{2}(D^{\rm th}(\theta) - D)^T C^{-1}(D^{\rm th}(\theta) - D)\}$$

Model and data



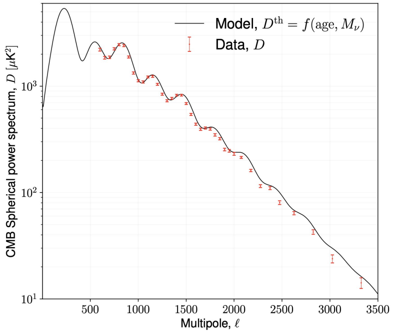
Covariance matrix (part of experimental data)



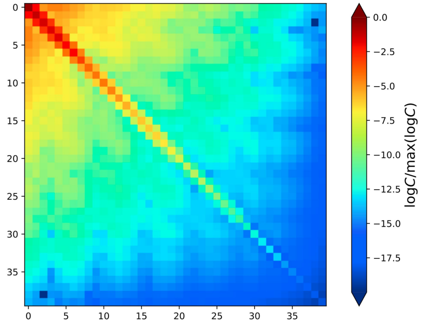
The CMB likelihood

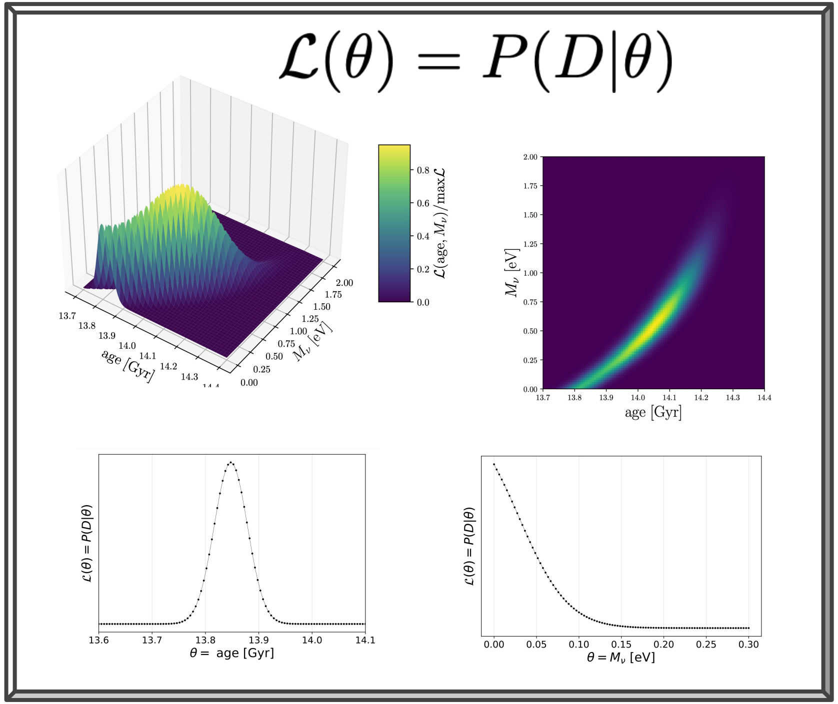
- Gaussian approximation $$P(D|\theta) = \frac{1}{\sqrt{2\pi|C|}}\exp\{-\frac{1}{2}(D^{\rm th}(\theta) - D)^T C^{-1}(D^{\rm th}(\theta) - D)\}$$

Model and data



Covariance matrix
(part of experimental data)





The prior

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

- *Definition:* probability distribution for the range of possible parameter values
- Key advantage of Bayesian approach
- **Summarises prior knowledge** on the model

The prior

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

- *Definition:* probability distribution for the range of possible parameter values
- Key advantage of Bayesian approach
- **Summarises prior knowledge** on the model
- Prior choice
  - Guiding principles
  - Educated intelligence

The posterior

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

- *Definition:* multivariate/joint probability distribution for the model parameters (conditioned on D and model, given prior)

- Goal of Bayesian inference: solution to the problem

- ***Degree of belief***

The posterior

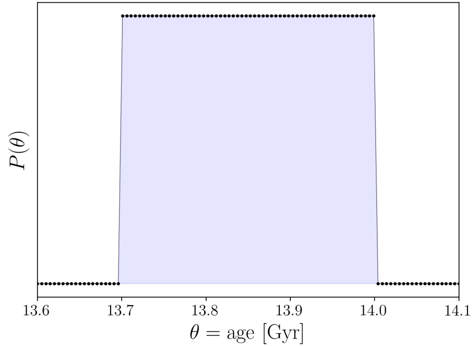
$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

- *Definition:* multivariate/joint probability distribution for the model parameters (conditioned on D and model, given prior)
- Goal of Bayesian inference: solution to the problem
- ***Degree of belief***
- Does not exist in the classical/frequentist approach

Evaluating the posterior, a 1-parameter example

- Pick a **uniform prior**
- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

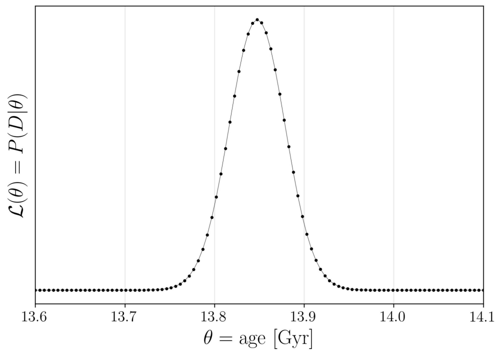
$$P(\mathrm{age}|D) \propto P(D|\mathrm{age})P(\mathrm{age})$$



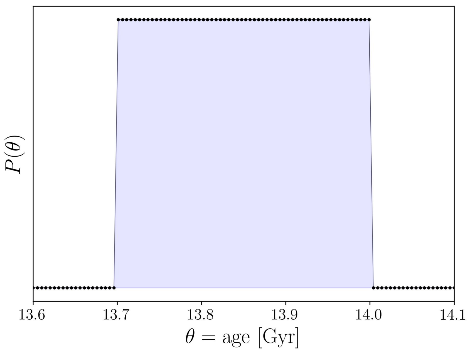
Evaluating the posterior, a 1-parameter example

- Pick a **uniform prior**
- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

$$P(\text{age}|D) \propto P(D|\text{age})P(\text{age})$$



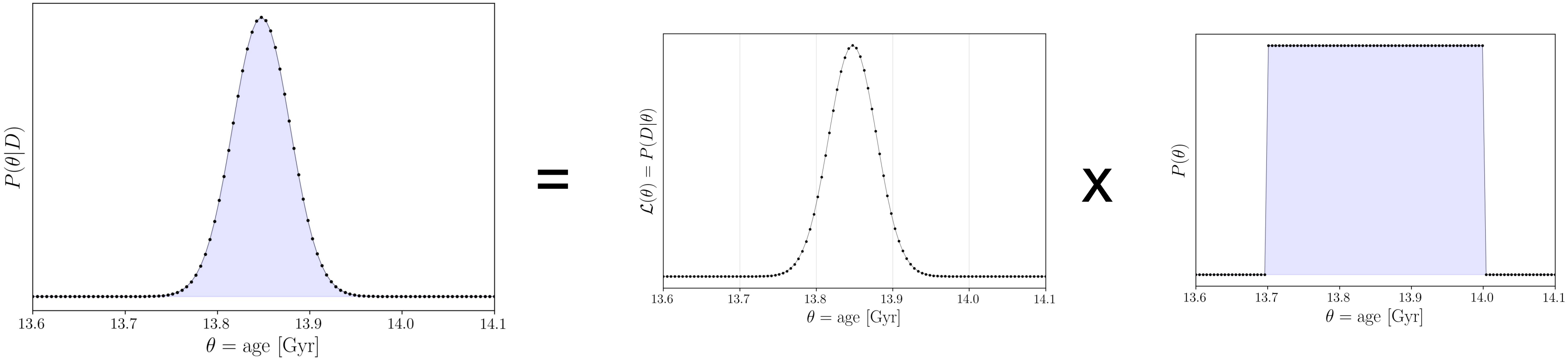
X



Evaluating the posterior, a 1-parameter example

- Pick a **uniform prior**
- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

$$P(\text{age}|D) \propto P(D|\text{age})P(\text{age})$$

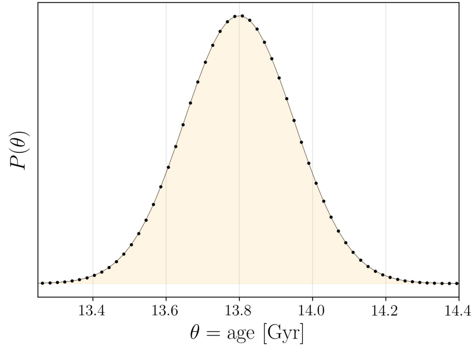


Evaluating the posterior, a 1-parameter example

- Pick a **wide Gaussian prior**

- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

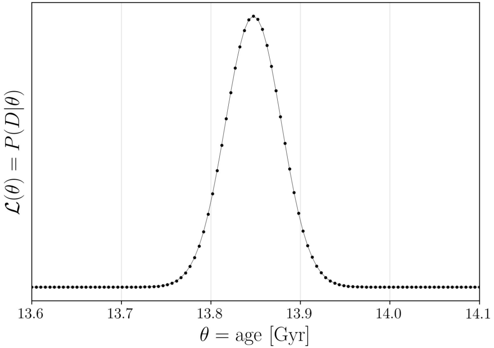
$$P(\text{age}|D) \propto P(D|\text{age})P(\text{age})$$



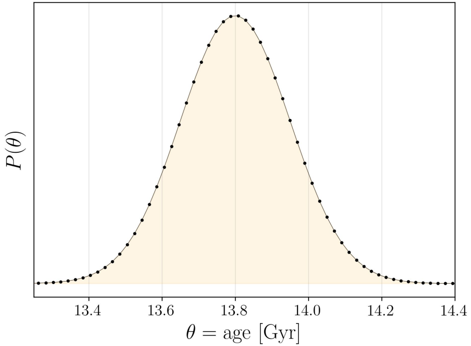
Evaluating the posterior, a 1-parameter example

- Pick a **wide Gaussian prior**
- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

$$P(\mathrm{age}|D) \propto P(D|\mathrm{age})P(\mathrm{age})$$



X

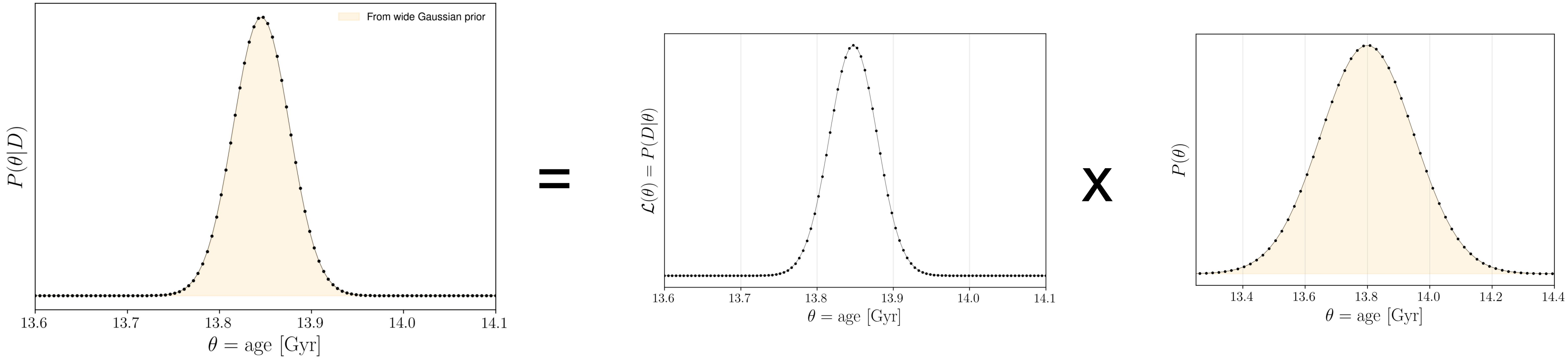


Evaluating the posterior, a 1-parameter example

- Pick a **wide Gaussian prior**

- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

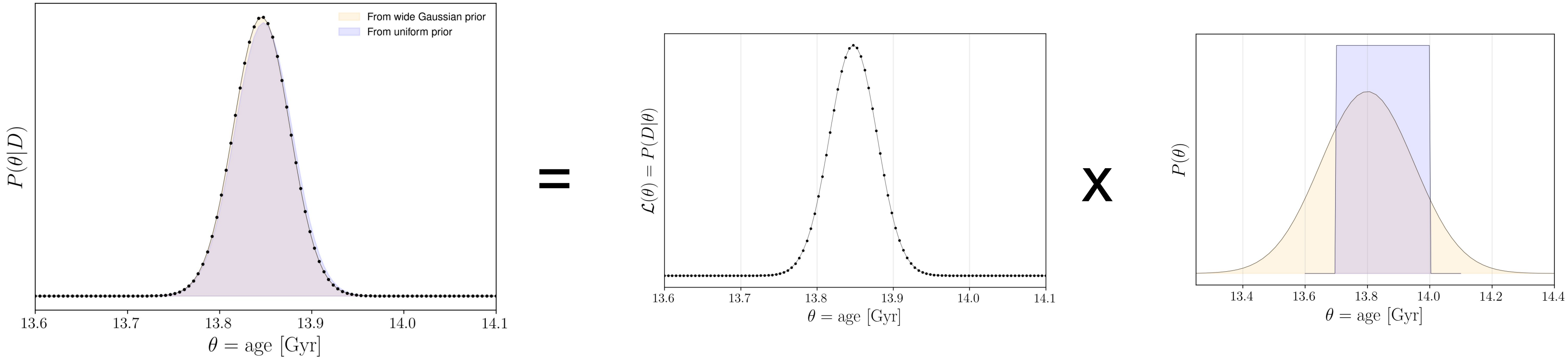
$$P(\text{age}|D) \propto P(D|\text{age})P(\text{age})$$



Evaluating the posterior, a 1-parameter example

- Pick a **wide Gaussian prior**
- Idea: explore parameter space (sampling), compute prior and multiply by likelihood

$$P(\text{age}|D) \propto P(D|\text{age})P(\text{age})$$



Two very different priors lead to the same posterior!

Bayesian credible intervals

- Set a X% Confidence Level (CL), e.g., X=68 or 95
- Find narrowest interval $t_1 < \theta < t_2$ such that:

$$X = 100 \int_{t_1}^{t_2} P(\theta|D) d\theta$$

Bayesian credible intervals

- Set a X% Confidence Level (CL), e.g., X=68 or 95

- Find narrowest interval $t_1 < \theta < t_2$ such that:

$$X = 100 \int_{t_1}^{t_2} P(\theta|D)d\theta$$

- Bayesian interpretation: **degree of belief** that the true value of theta lies within this interval (given D and model)

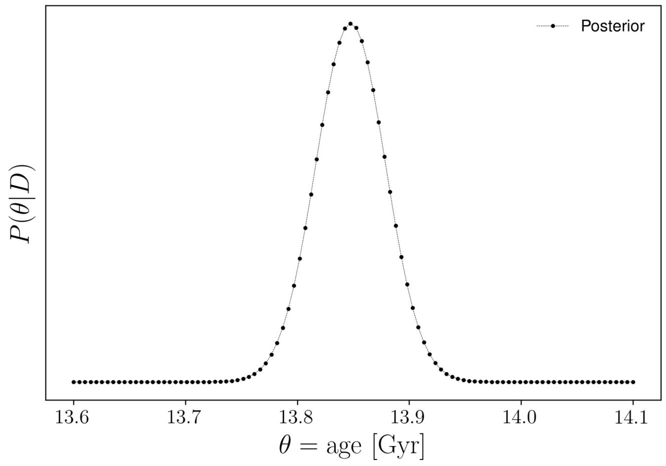
- Different from frequentist view!

Obtaining the 68% and 95%CL credible intervals

- General method: Kernel Density Estimation (KDE)

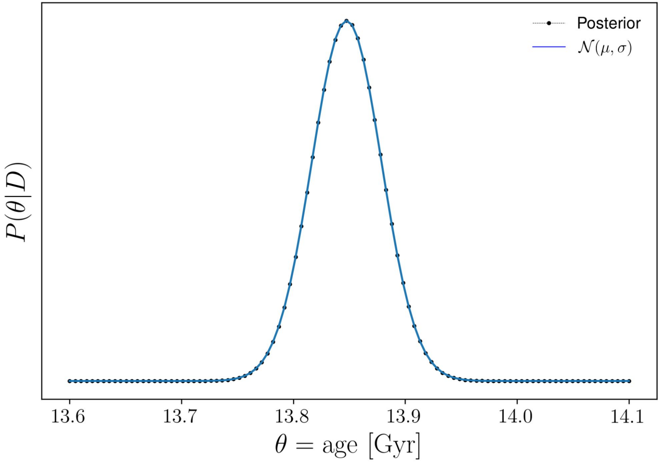
Obtaining the 68% and 95%CL credible intervals

- General method: Kernel Density Estimation (KDE)
- Straightforward when Posterior ~ Normal distribution



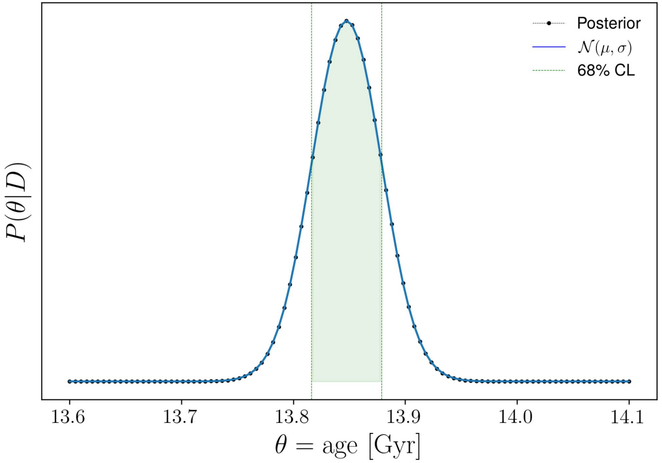
Obtaining the 68% and 95%CL credible intervals

- General method: Kernel Density Estimation (KDE)
- Straightforward when Posterior ~ Normal distribution
- Fit posterior with a Gaussian and quote the 1 and 2-σ intervals



Obtaining the 68% and 95%CL credible intervals

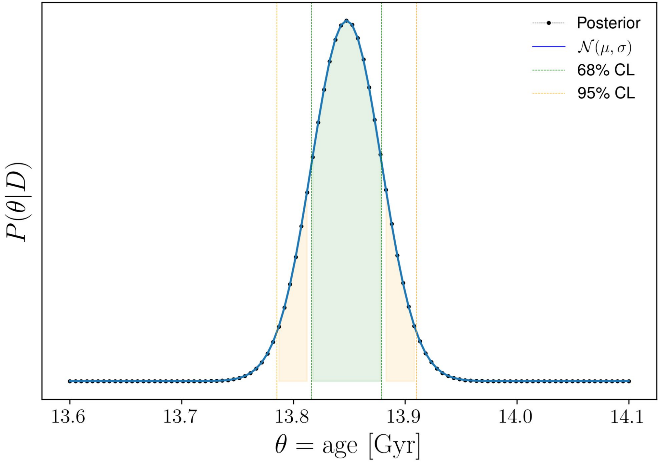
- General method: Kernel Density Estimation (KDE)
- Straightforward when Posterior ~ Normal distribution
- Fit posterior with a Gaussian and quote the 1 and 2-σ intervals



$$\mathrm{age} = 13.847 \pm 0.031\,\mathrm{Gyr}\ \ (68\%\mathrm{CL})$$

Obtaining the 68% and 95%CL credible intervals

- General method: Kernel Density Estimation (KDE)
- Straightforward when Posterior ~ Normal distribution
- Fit posterior with a Gaussian and quote the 1 and 2-σ intervals



$$\text{age} = 13.847 \pm 0.031\,\text{Gyr} \;\; (68\%\text{CL})$$

$$\text{age} = 13.847 \pm 0.062\,\text{Gyr} \;\; (95\%\text{CL})$$

General case

$$P(\mathrm{age}, M_\nu|D) \propto P(D|\mathrm{age}, M_\nu)P(\mathrm{age}, M_\nu)$$

General case

$$P(\text{age}, M_\nu | D) \propto P(D | \text{age}, M_\nu) P(\text{age}, M_\nu)$$

- Uniform priors
  - $0.06 \text{ eV} < M_\nu < \text{few eV}$
  - $t_{\text{sun}} < \text{age} < \text{few}\, 10\text{Gyr}$

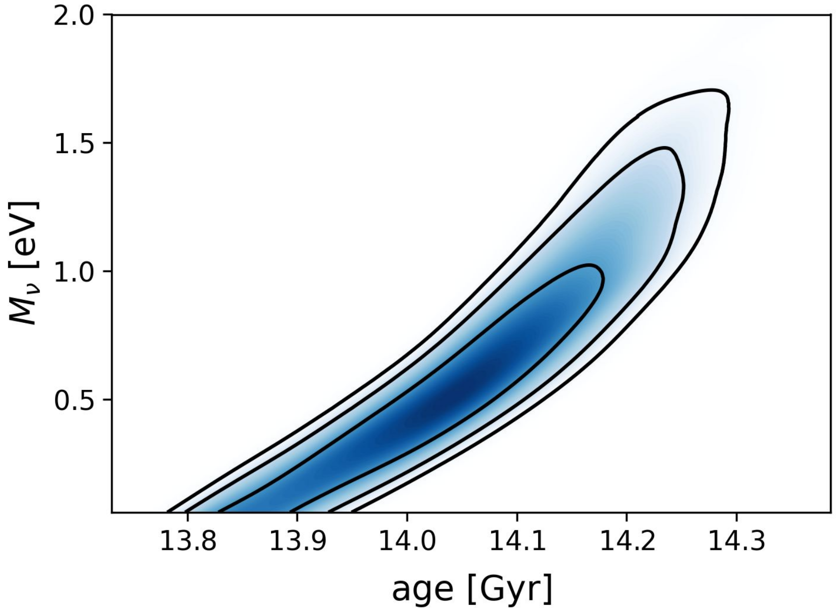
General case

$$P(\mathrm{age}, M_\nu | D) \propto P(D | \mathrm{age}, M_\nu) P(\mathrm{age}, M_\nu)$$

- Uniform priors
  - $0.06 \text{ eV} < M_\nu < \text{few eV}$
  - $t_{\text{sun}} < \text{age} < \text{few } 10\text{Gyr}$

➔ 2D **posterior probability distribution**

➔ Obtained with MCMC* and KDE*
(see notebook)

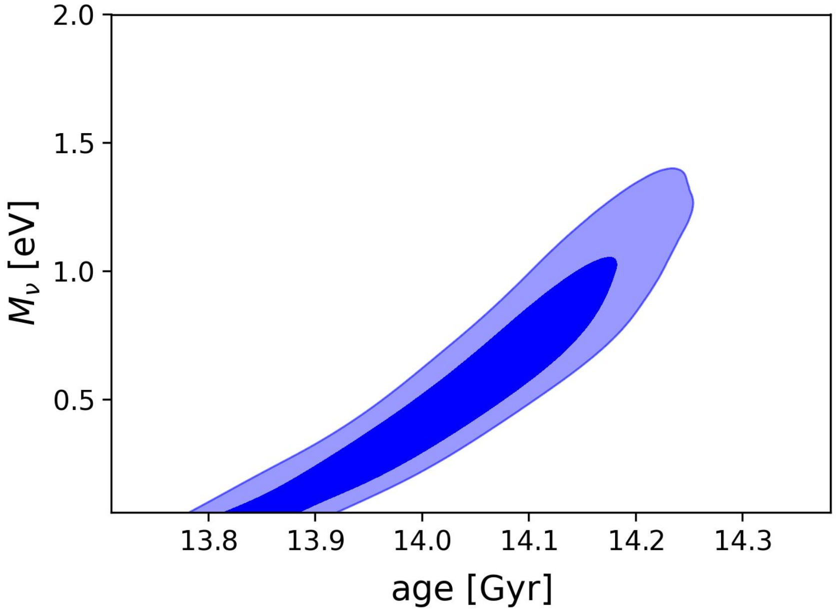


*MCMC: Monte Carlo Markov Chains, KDE: Kernel Density Estimation

General case

$$P(\text{age}, M_\nu | D) \propto P(D|\text{age}, M_\nu) P(\text{age}, M_\nu)$$

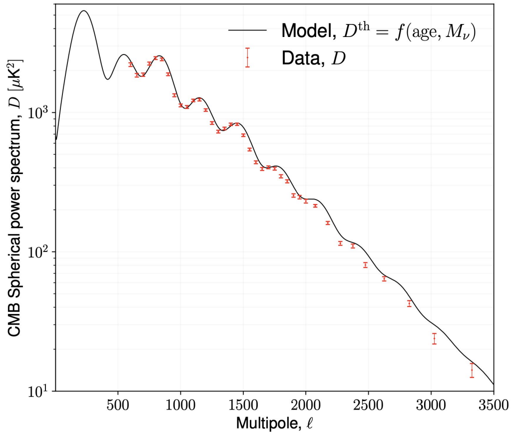
- ***Bayesian credible regions*** (68% and 95%CL)

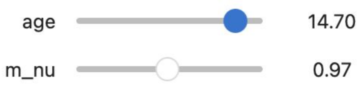


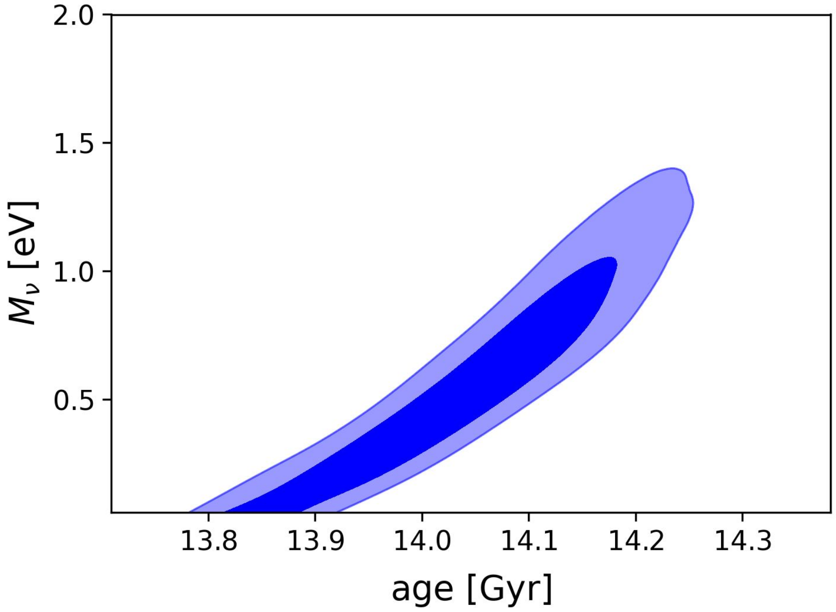
General case

$$P(\text{age}, M_\nu | D) \propto P(D|\text{age}, M_\nu) P(\text{age}, M_\nu)$$

- ***Bayesian credible regions*** (68% and 95%CL)
- Parameter degeneracies



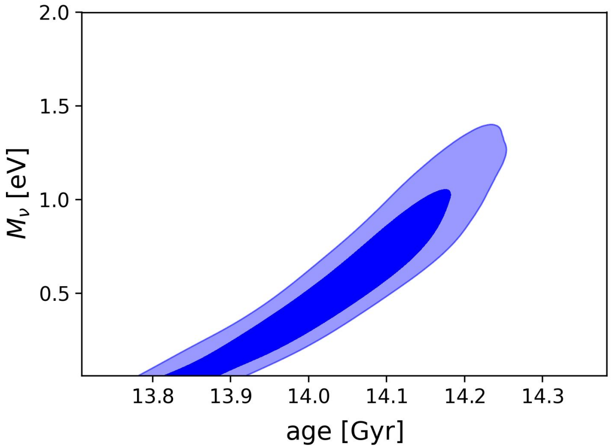




Eliminating a “*nuisance*” parameter

- Give-up hope of knowing neutrino mass precisely

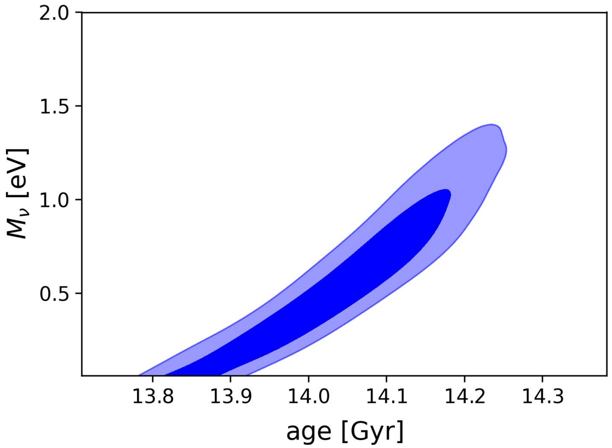
- How old is the universe?



Eliminating a “*nuisance*” parameter

- Give-up hope of knowing neutrino mass precisely

- How old is the universe?

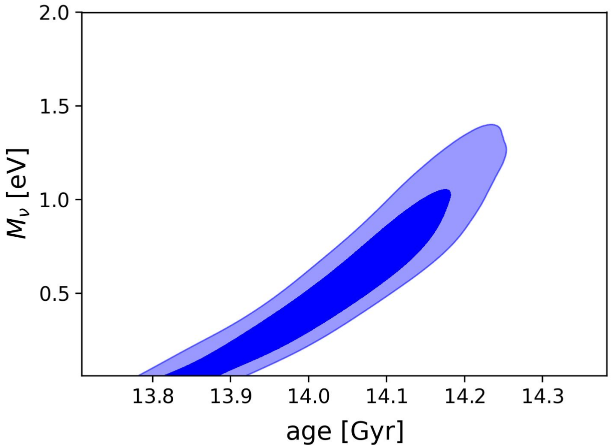


Marginalisation

$$P(\text{age}|D) = \int P(\text{age}, M_\nu|D)\,dM_\nu$$

Eliminating a "*nuisance*" parameter

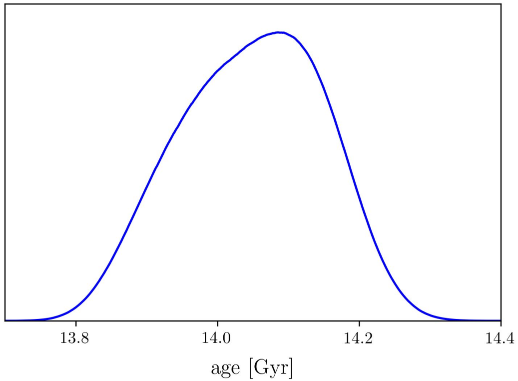
- Give-up hope of knowing neutrino mass precisely
- How old is the universe?



Marginalisation

$$P(\mathrm{age}|D) = \int P(\mathrm{age}, M_\nu|D)dM_\nu$$

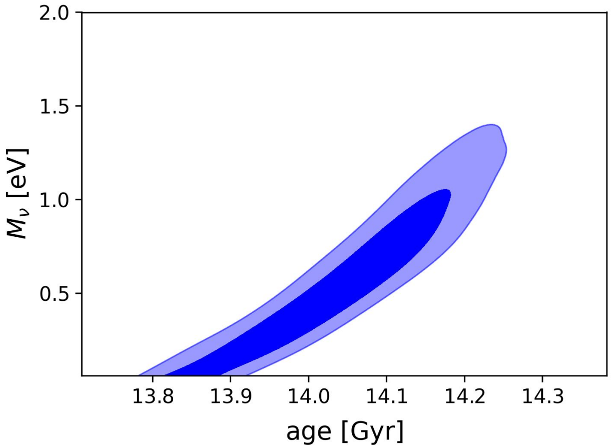
1D marginalised posterior PDF



Eliminating a “*nuisance*” parameter

- Give-up hope of knowing neutrino mass precisely

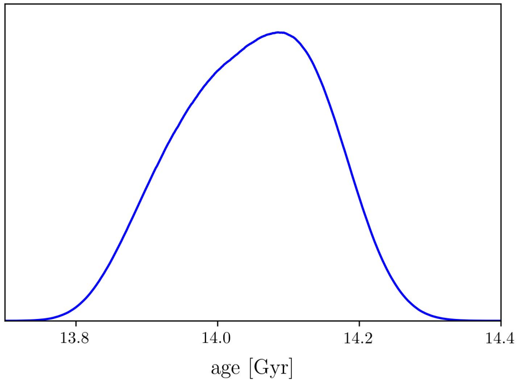
- How old is the universe?



Marginalisation

$$P(\text{age}|D) = \int P(\text{age}, M_\nu|D)dM_\nu$$

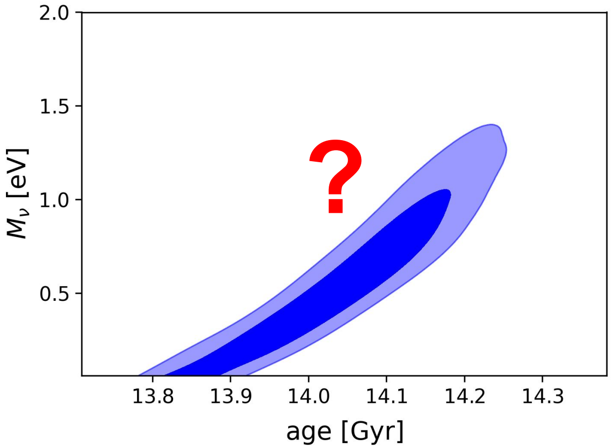
1D marginalised posterior PDF



$$\boxed{\text{age} = 14.05 \pm 0.1\ \text{Gyr}(68\%\text{CL})}$$

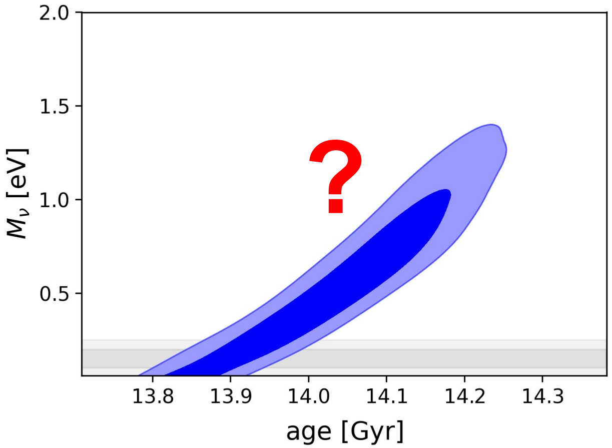
Combining experiments

- A new neutrino mass measurements is available $P(M_\nu) \sim \mathcal{N}(\mu = 0.15, \sigma = 0.05)$



Combining experiments

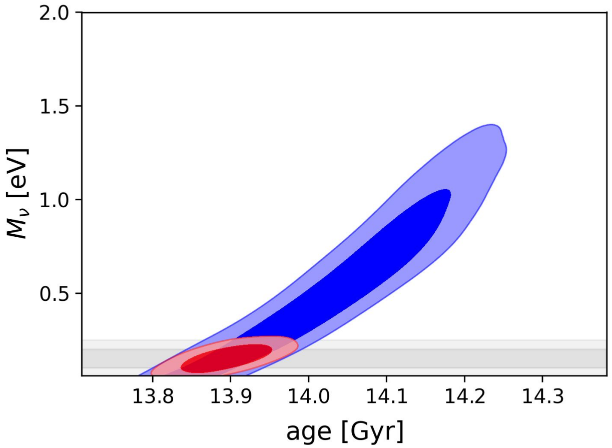
- A new neutrino mass measurements is available $P(M_\nu) \sim \mathcal{N}(\mu = 0.15, \sigma = 0.05)$



Combining experiments

- A new neutrino mass measurements is available $P(M_\nu) \sim \mathcal{N}(\mu = 0.15, \sigma = 0.05)$

- $P(\mathrm{age}, M_\nu | D) \propto P(D | \mathrm{age}, M_\nu) P(\mathrm{age}, M_\nu)$

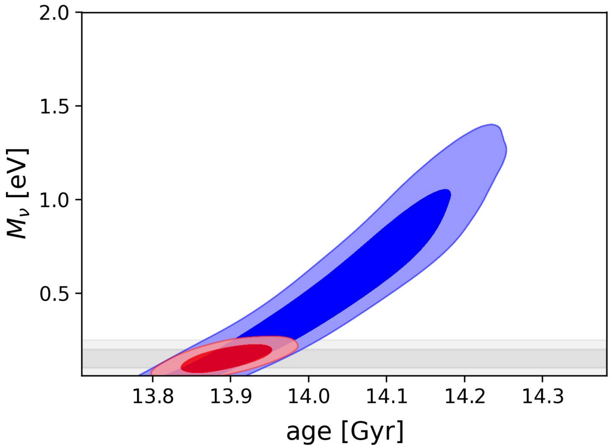


Combining experiments

- A new neutrino mass measurements is available $P(M_\nu) \sim \mathcal{N}(\mu = 0.15, \sigma = 0.05)$

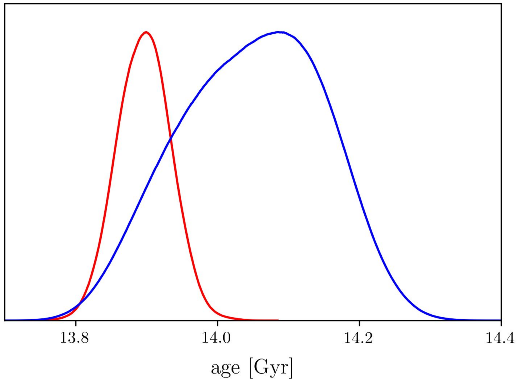
- $P(\text{age}, M_\nu | D) \propto P(D | \text{age}, M_\nu) P(\text{age}, M_\nu)$

1D marginalised posterior PDF



Marginalisation

$$P(\text{age}|D) = \int P(\text{age}, M_\nu | D) dM_\nu$$

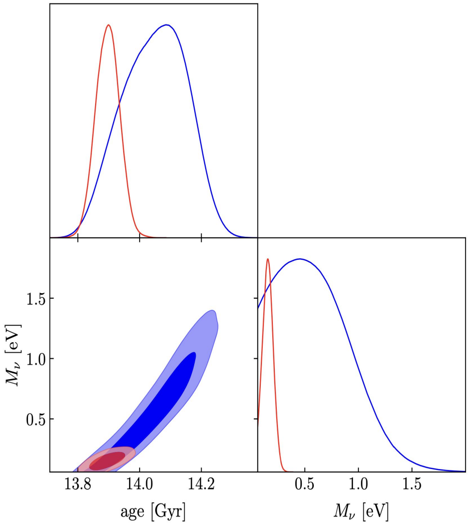


$\text{age} = 14.90 \pm 0.04\ \text{Gyr}(68\%\text{CL})$

Bayesian inference summary and key points

"What you know about θ after the data arrive is what you knew before [P(θ)], and what the data told you [P(D|θ)]." (MacKay)

- Interpretation of Bayes theorem $P(\theta|D) = \dfrac{P(D|\theta)P(\theta)}{P(D)}$
- Effects of prior choices
- Calculating the posterior probability distribution
- Bayesian credible intervals
- Marginalisation over nuisance parameters
- Adding new information/Combining experiments
- Approach directly applicable to large number of free parameters and priors



References

- CMB Data/likelihood: Aiola et al 2020 (eprint:2007.07288)
- CMB Model: Bolliet et al 2023 (eprint:2303.01591)

- F. James, Statistical Methods in Experimental Physics, Second Edition, WS, 2006
- A. Heavens, ICIC Data Analysis Workshop (private communication)
- R. Trotta, Bayesian Methods in Cosmology (eprint:1701.01467)
- S. Dodelson and F. Schmidt, Modern Cosmology, Second Edition, AP, 2020
- G. Cowan, Statistical Data Analysis, Oxford, 1998
- D. Hogg, J. Bovy & D. Lang, Data analysis recipes: Fitting a model to data (eprint:1008.4686)
- E. T. Jaynes, Probability Theory, The Logic of Science, CUP, 2003
- Hobson et al, Bayesian Methods in Cosmology, CUP, 2009
- T. Loredo, Bayesian Inference for the Physical Sciences (online)
- R. Cousins, Why isn’t every physicist a Bayesian?, 1994
- D. MacKay, Information Theory, Inference and Learning Algorithms, CUP, 2003