Fast likelihood-free inference in the LSS Stage-IV era

Guillermo Franco Abellán

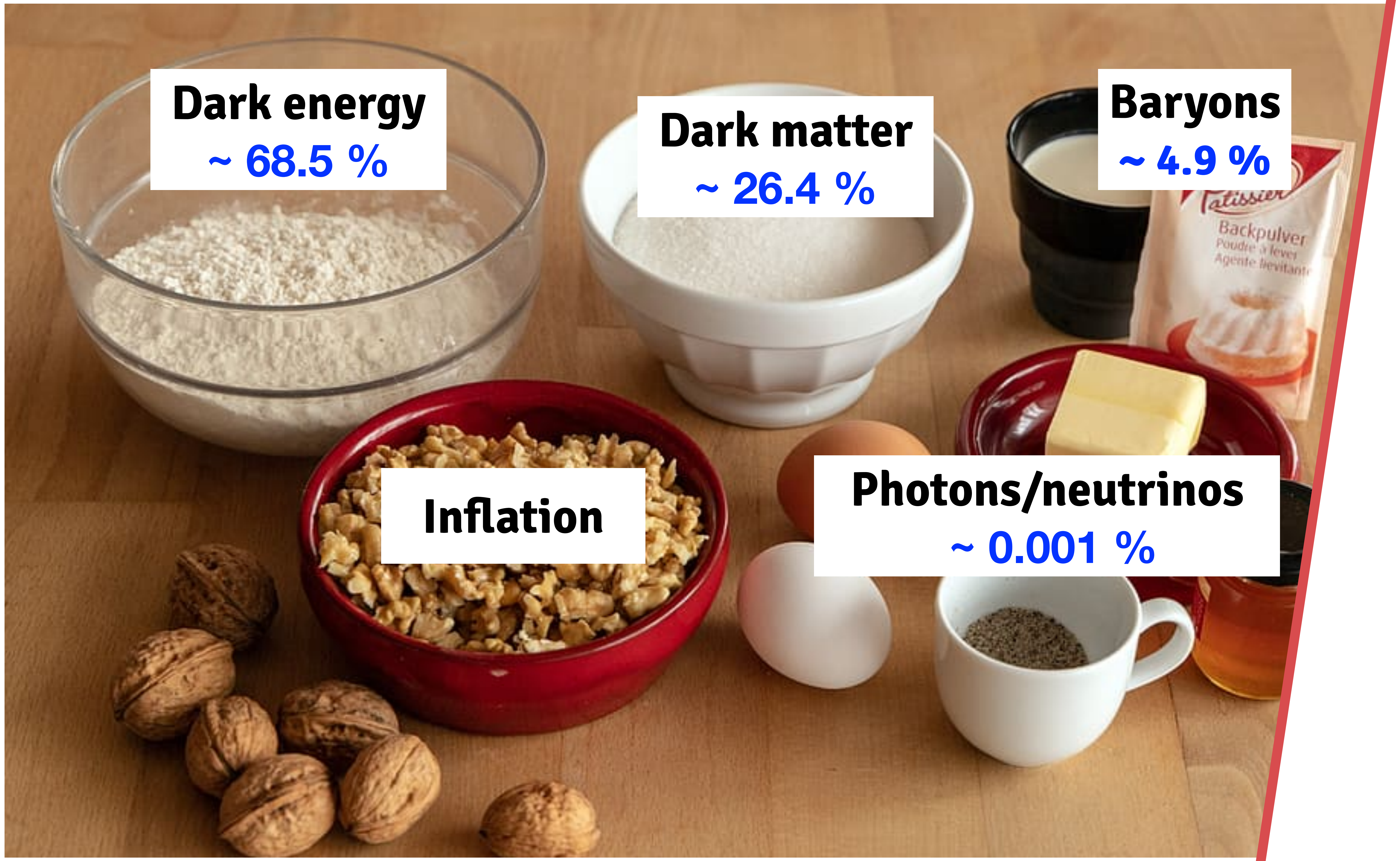
GRAPPA
GRavitation AstroParticle Physics Amsterdam

Based on arXiv:2403.14750

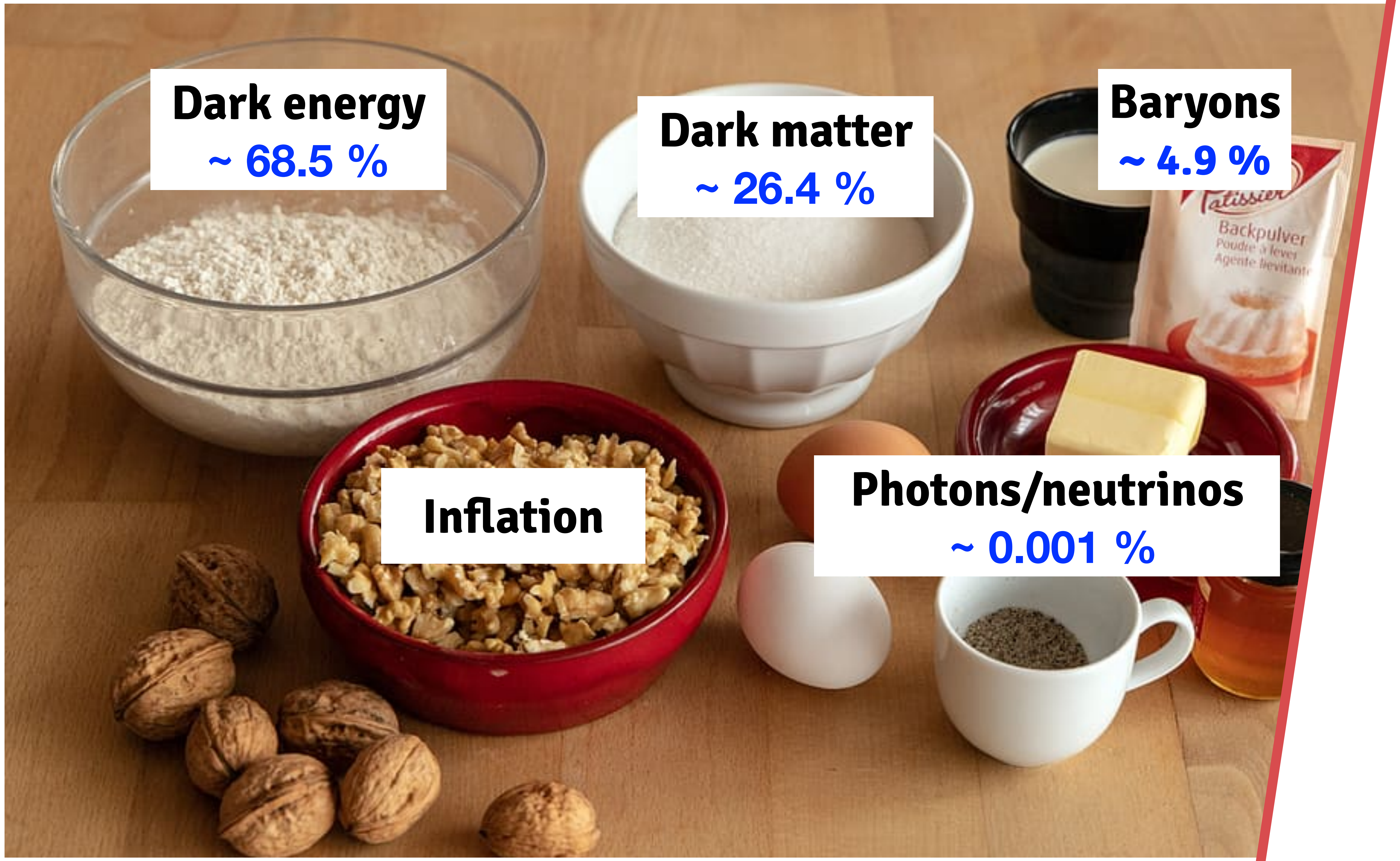
with Guadalupe Cañas-Herrera,
Matteo Martinelli,
Oleg Savchenko,
Davide Sciotti,
& Christoph Weniger

Euclid TWG/WP17 - 26th April 2024

Concordance ΛCDM model of cosmology:

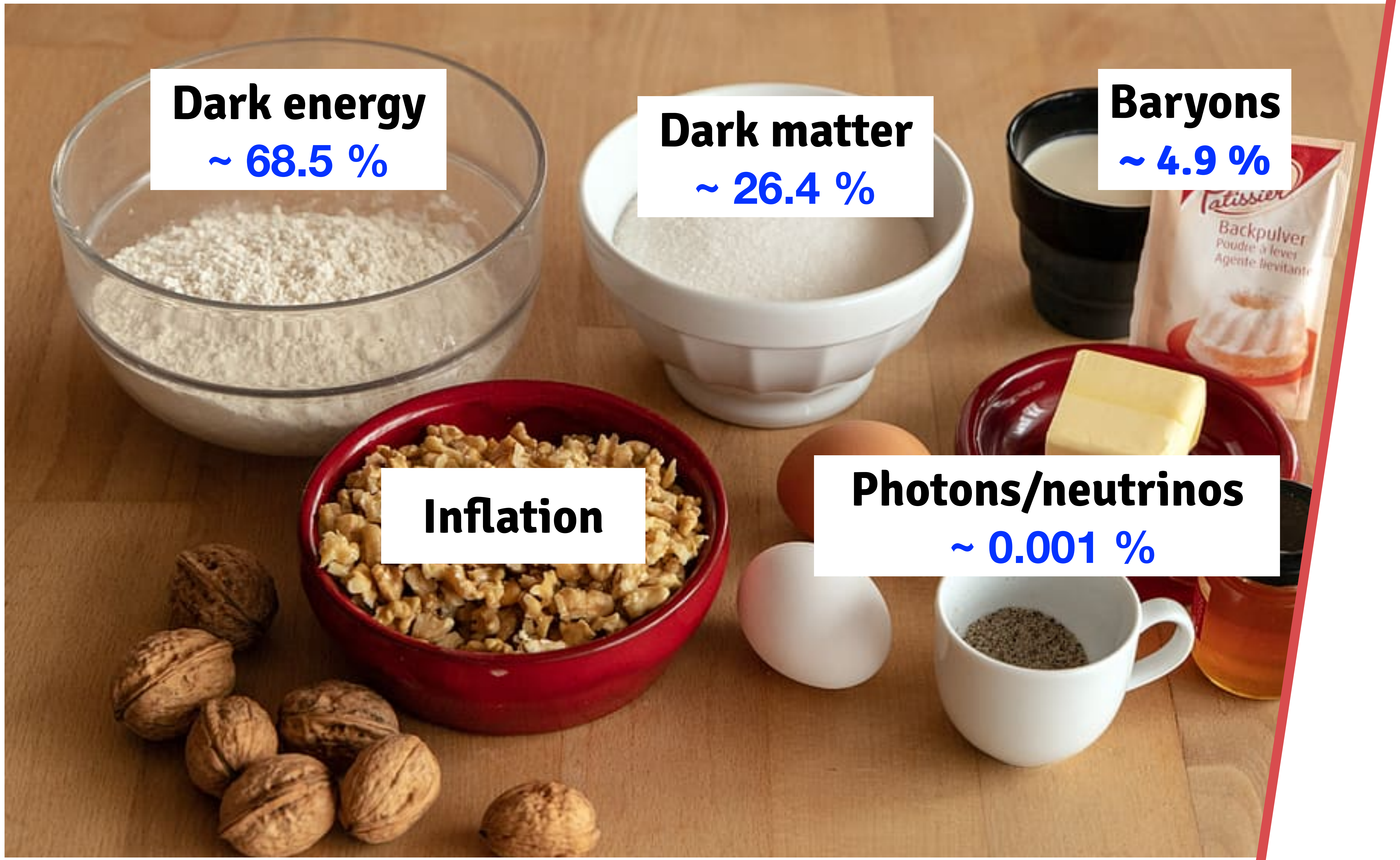


Concordance ΛCDM model of cosmology:



Some open problems:

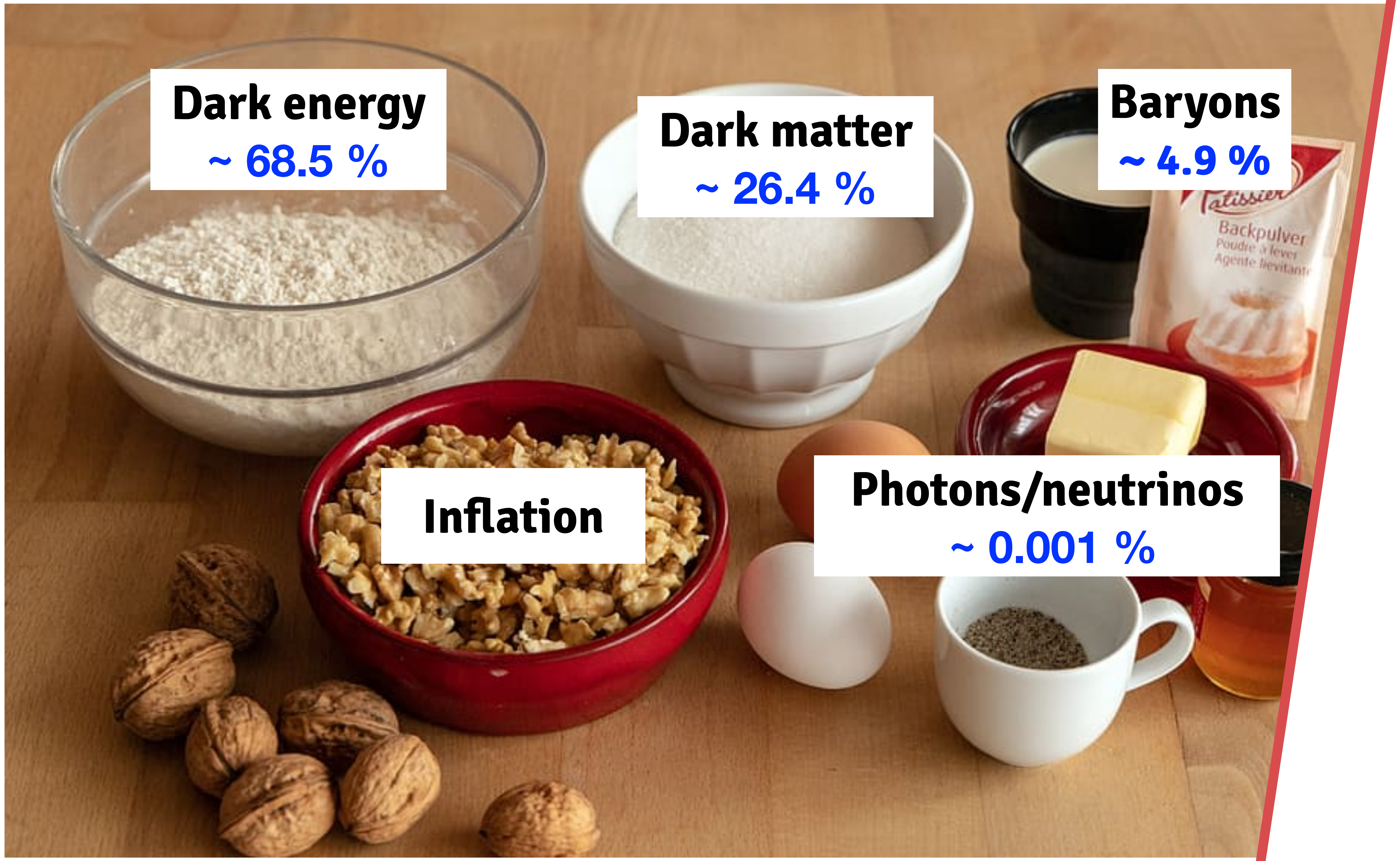
Concordance ΛCDM model of cosmology:



Some open problems:

- **Nature of dark sector?**

Concordance ΛCDM model of cosmology:



Some open problems:

- **Nature of dark sector?**
- **H_0/S_8 tensions?**

Growing interest in testing ΛCDM extensions...

Growing interest in testing ΛCDM extensions…

…but still no smoking-gun signature of new physics

Next-generation cosmological data is becoming available

Next-generation cosmological data is becoming available

Analysing these high-quality data will be extremely challenging with traditional methods

Outline

I. Why we need to go **beyond MCMC**

II. Our new approach: **Marginal Neural Ratio Estimation**

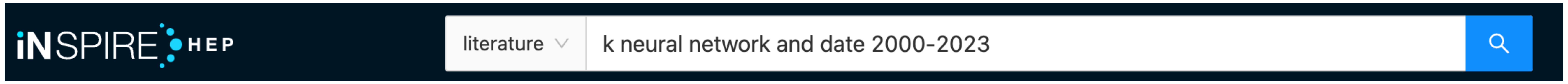
III. Applying MNRE to **Stage IV** photometric observables

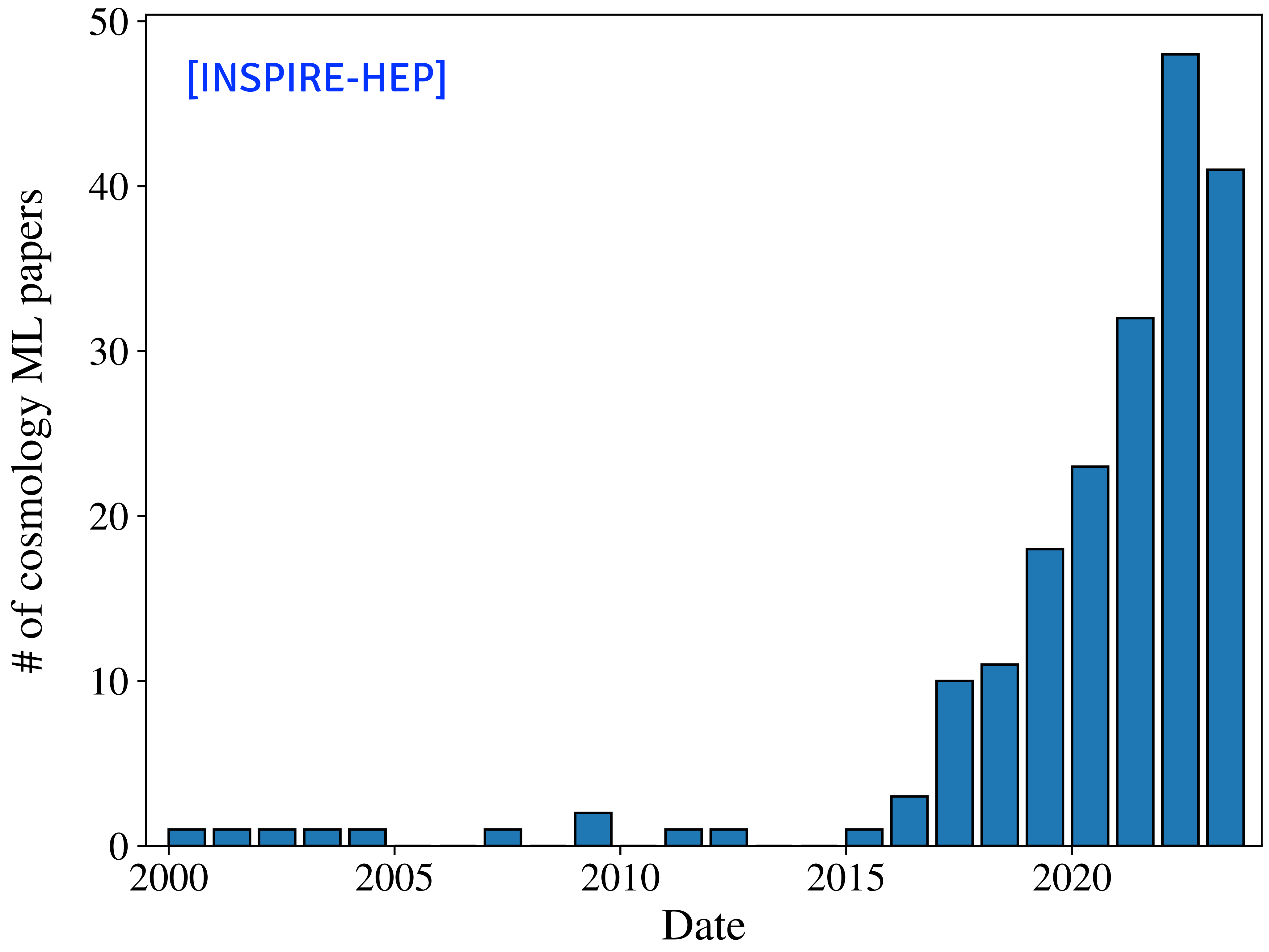
Outline

I. Why we need to go **beyond MCMC** ✓

II. Our new approach: **Marginal Neural Ratio Estimation**

III. Applying MNRE to **Stage IV** photometric observables





Machine learning is having a **strong impact** in cosmology

Two main approaches in ML

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

New statistical methods

to improve the sampling in **high-dimensional** parameter spaces

Two main approaches in ML

Emulators

to achieve **ultra-fast** evaluations of cosmological observables

New statistical methods

to improve the sampling in **high-dimensional** parameter spaces

This talk

Bayesian inference

Likelihood

Posterior $\quad$ Prior

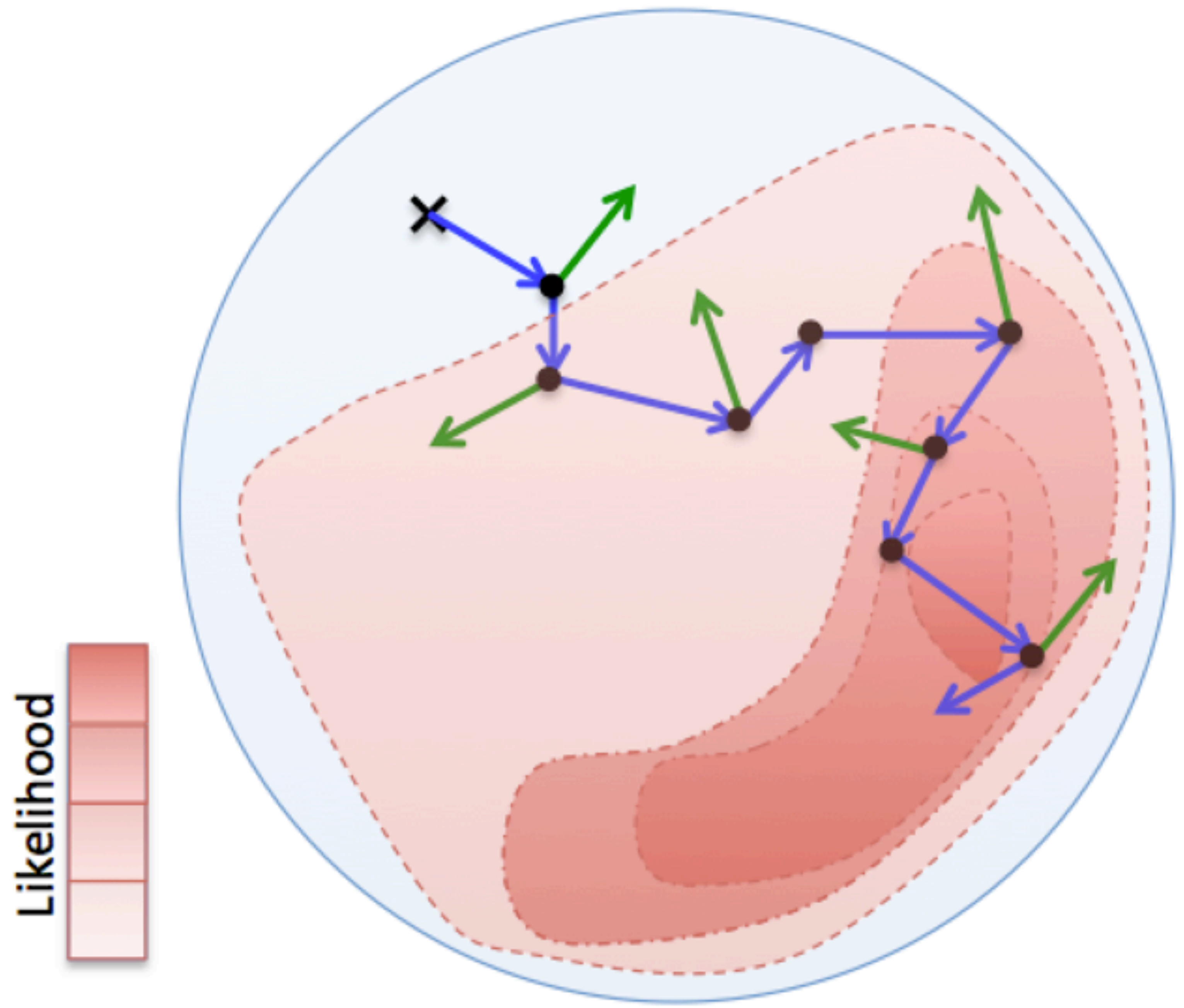
$$p(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \boldsymbol{\theta})}{p(\mathbf{x})} p(\boldsymbol{\theta})$$

Evidence

$\mathbf{x}$: Data

$\boldsymbol{\theta}$: Parameters

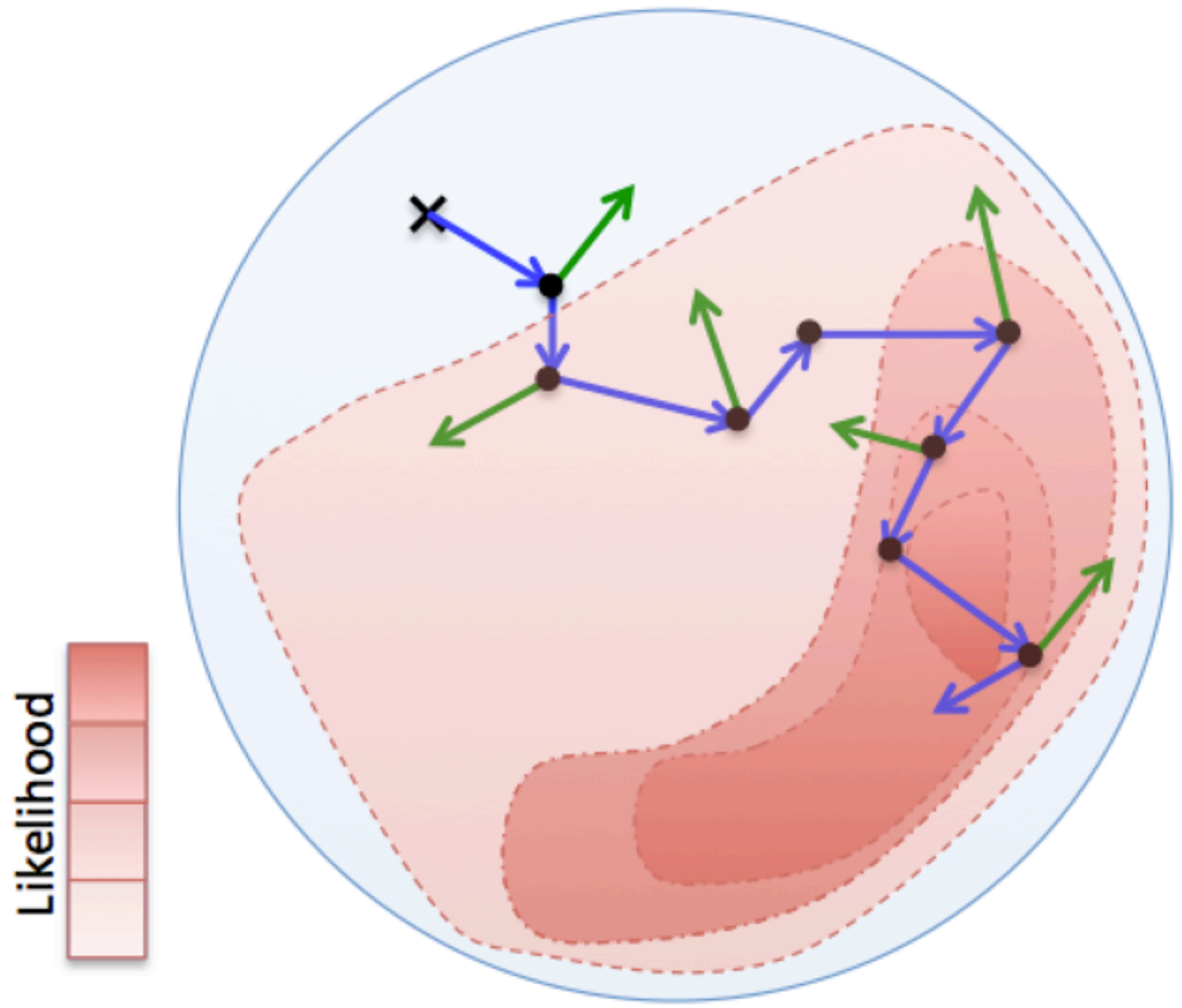
Metropolis-Hastings algorithm



Traditional likelihood-based methods (MCMC, Nested Sampling,...) allow to get samples from the full joint posterior

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \,|\, \mathbf{x}), \qquad \boldsymbol{\theta} \in \mathbb{R}^D$$

Metropolis-Hastings algorithm



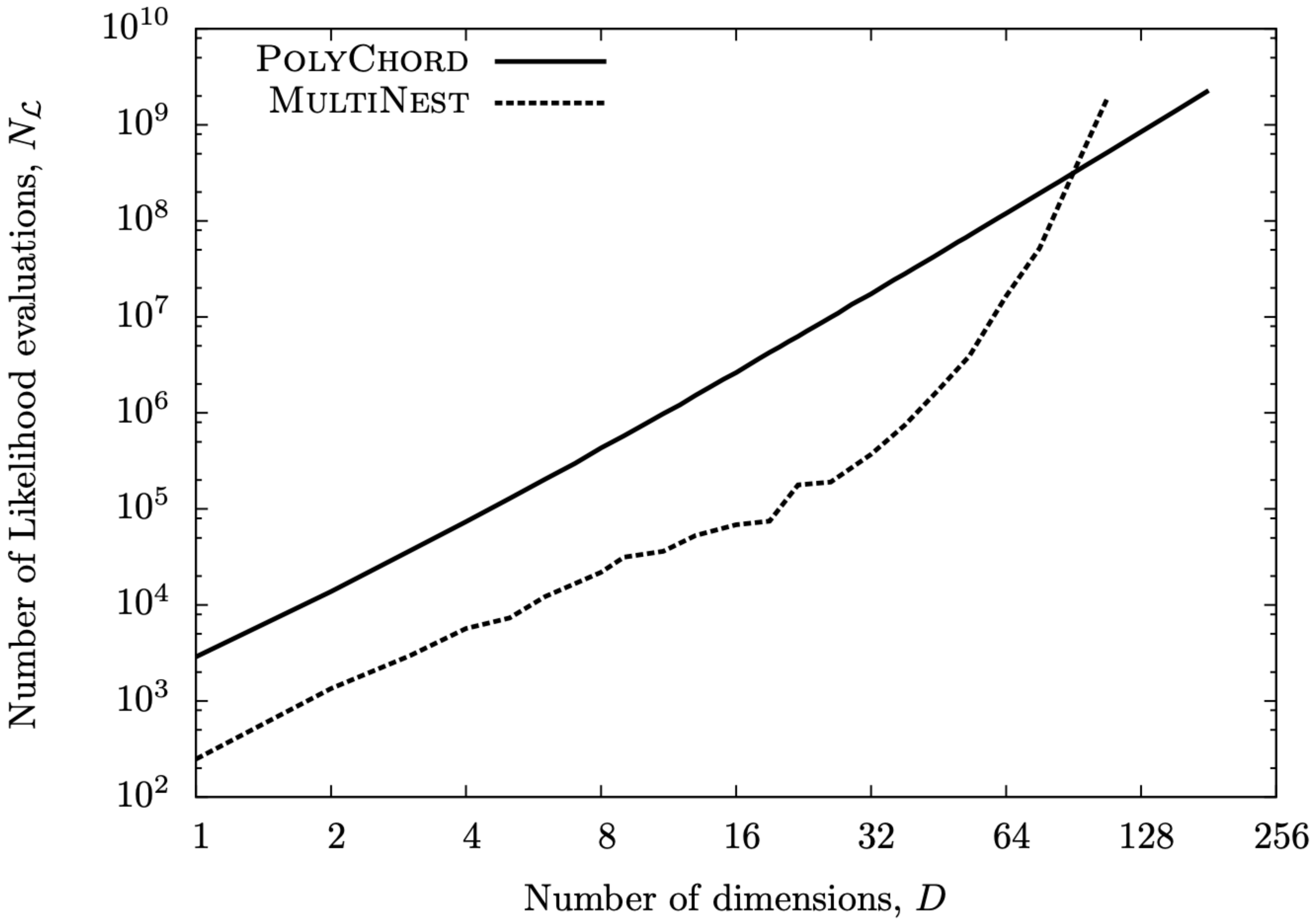
- Traditional likelihood-based methods (MCMC, Nested Sampling,...) allow to get samples from the full joint posterior

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \,|\, \mathbf{x}), \qquad \boldsymbol{\theta} \in \mathbb{R}^D$$

- Then we marginalise to get posteriors of interest

The curse of dimensionality

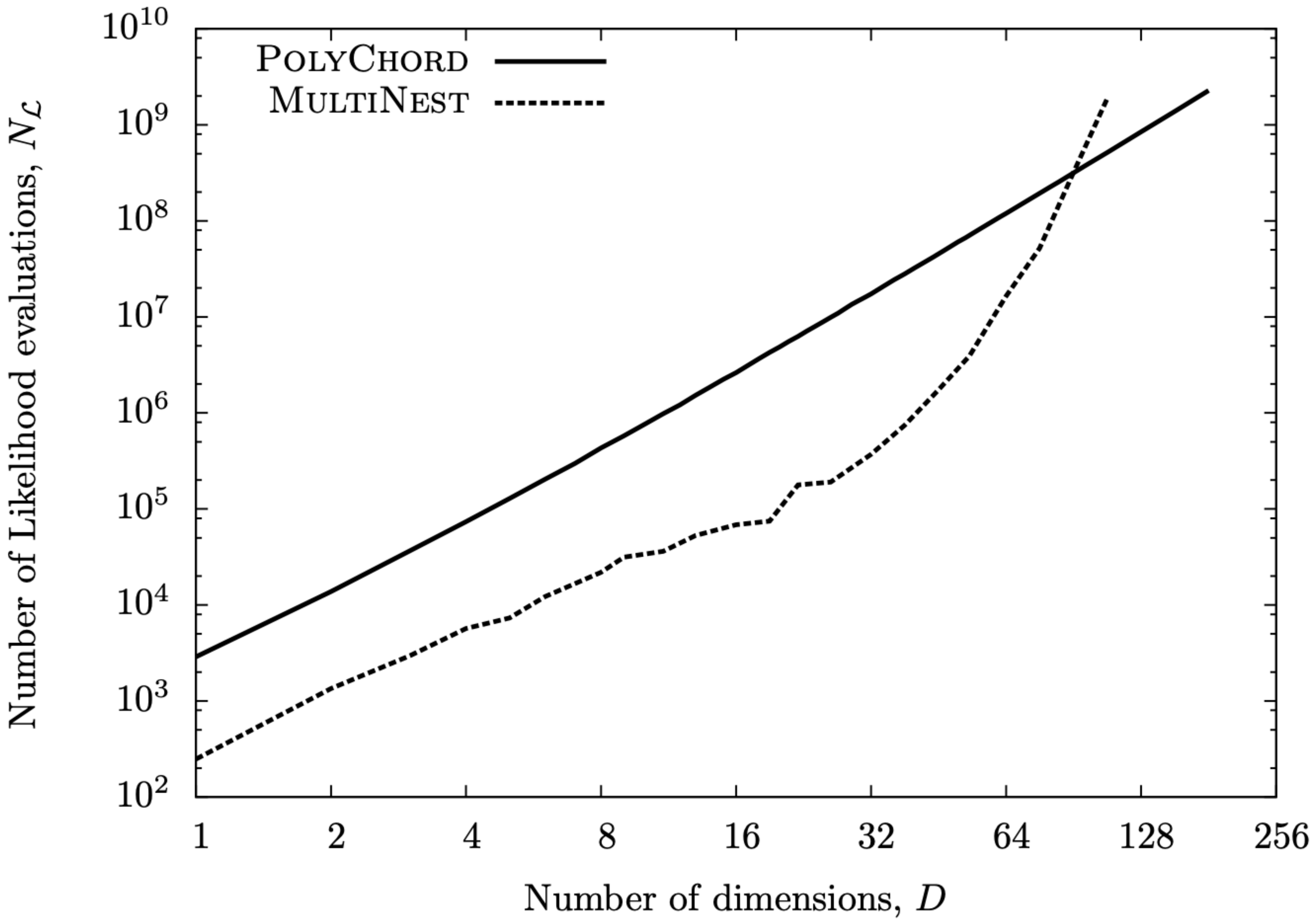
These methods scale poorly with the dimensionality of the parameter space



Handley+ 15

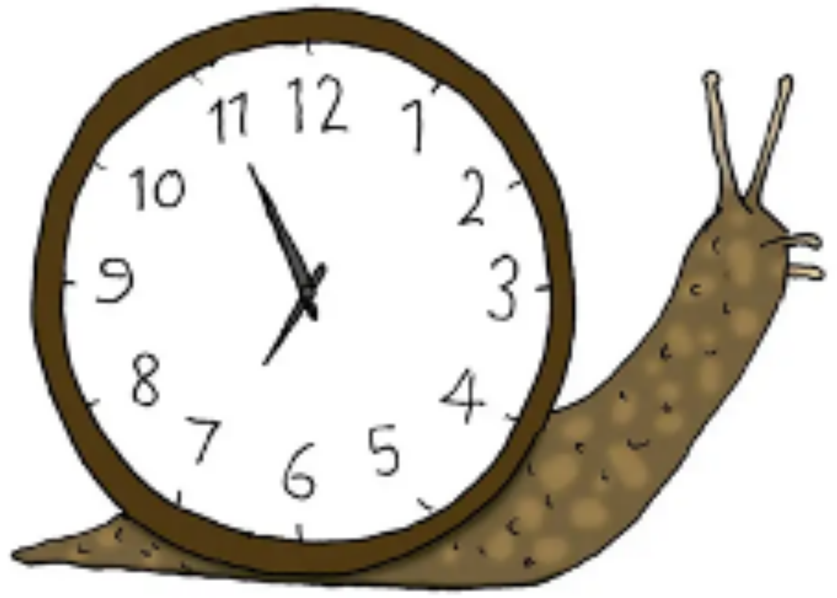
The curse of dimensionality

These methods scale poorly with the dimensionality of the parameter space



Handley+ 15

Ex: For Stage IV surveys, we expect **~100 nuisance parameters**



Weeks...

Are there methods to overcome this problem?

Are there methods to overcome this problem?

Can machine learning be helpful?

A NEW HOPE

MNRE = **M**arginal **N**eural **R**atio **E**stimation

Implemented in Swyft* [Miller+ 20]

* Stop Wasting Your Precious Time

I. Why we need to go **beyond MCMC**

II. Our new approach: **Marginal Neural Ratio Estimation** ✓

III. Applying MNRE to **Stage IV** photometric observables

1. Simulation

(generate training data)

2. Inference

(train networks to get posteriors)

1. Simulation

Simulation-based inference
(or likelihood-free inference)

1. Simulation

Simulation-based inference
(or likelihood-free inference)

↓

Stochastic simulator that maps from model parameters **θ** to data **x**

$$\mathbf{x} \sim p(\mathbf{x} \,|\, \boldsymbol{\theta})$$

(implicit likelihood)

We can simulate N samples that can be used as training data for a neural network

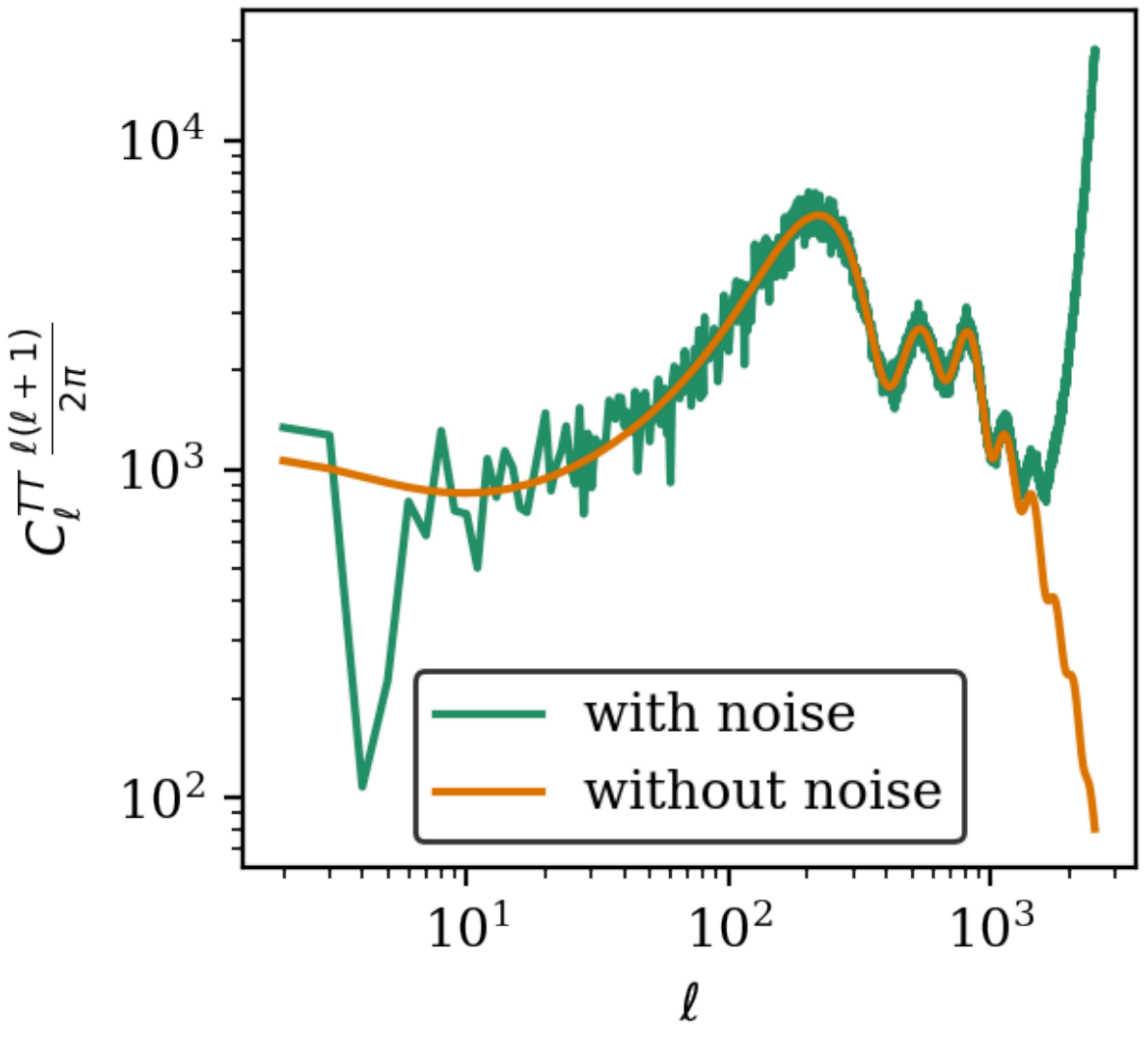
$$\{(\mathbf{x}^{(1)}, \boldsymbol{\theta}^{(1)}), (\mathbf{x}^{(2)}, \boldsymbol{\theta}^{(2)}), \ldots, (\mathbf{x}^{(N)}, \boldsymbol{\theta}^{(N)})\}$$

We can simulate N samples that can be used as training data for a neural network

$$\{(\mathbf{x}^{(1)}, \boldsymbol{\theta}^{(1)}), (\mathbf{x}^{(2)}, \boldsymbol{\theta}^{(2)}), \ldots, (\mathbf{x}^{(N)}, \boldsymbol{\theta}^{(N)})\}$$

Ex: CMB simulator

$$\boldsymbol{\theta} \;\rightarrow\; C_\ell(\boldsymbol{\theta}) \;\rightarrow\; C_\ell(\boldsymbol{\theta}) + N_\ell$$



Cole+ 22

Neural Ratio Estimation

Instead of directly estimating the posterior, estimate:

$$r(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})}$$

2. Inference

Neural Ratio Estimation

Instead of directly estimating the posterior, estimate:

$$r(\mathbf{x};\boldsymbol{\theta}) = \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})}$$

It's easy to show that:

$$\frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta}\,|\,\mathbf{x})}{p(\boldsymbol{\theta})}$$ (posterior-to-prior ratio)

2. Inference

$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ **(ex:** Ω_b and Ω_c **)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ **(ex:** CMB spectra**)**

$$\frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta} \mid \mathbf{x})}{p(\boldsymbol{\theta})}$$

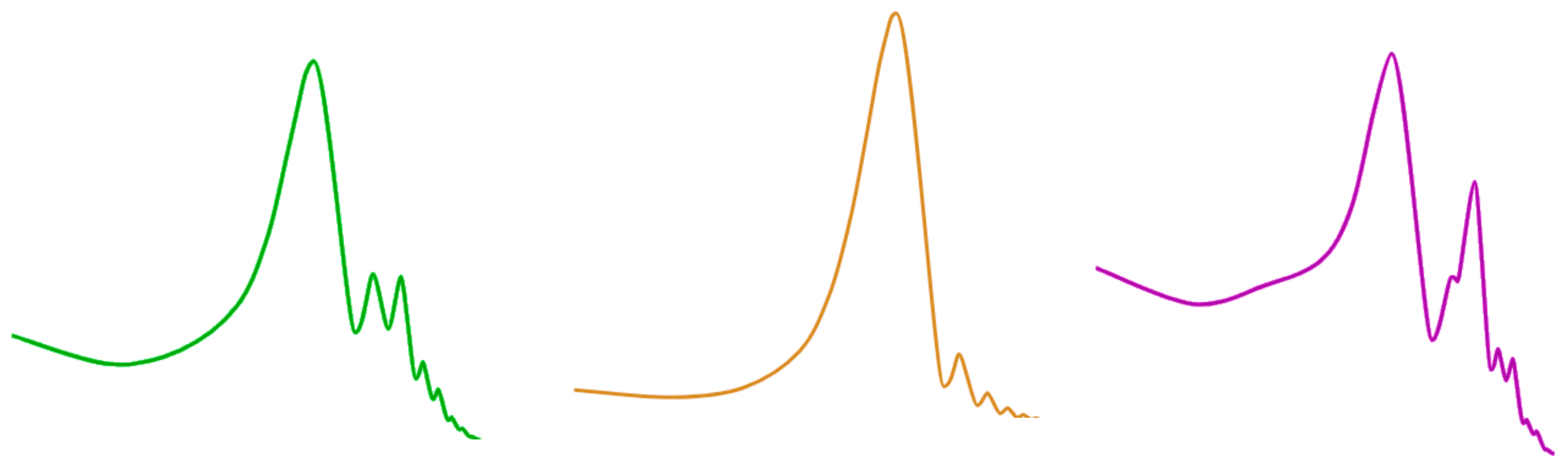
$$\frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta}\,|\,\mathbf{x})}{p(\boldsymbol{\theta})}$$

$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ **(ex:** Ω_b and Ω_c **)**

$\boldsymbol{x} \sim p(\boldsymbol{x})$ **(ex:** CMB spectra**)**

$(\mathbf{x},\boldsymbol{\theta}) \sim p(\mathbf{x},\boldsymbol{\theta})$ (jointly drawn)

Ω_b =	5%	30%	5%
Ω_c =	25%	0%	95%

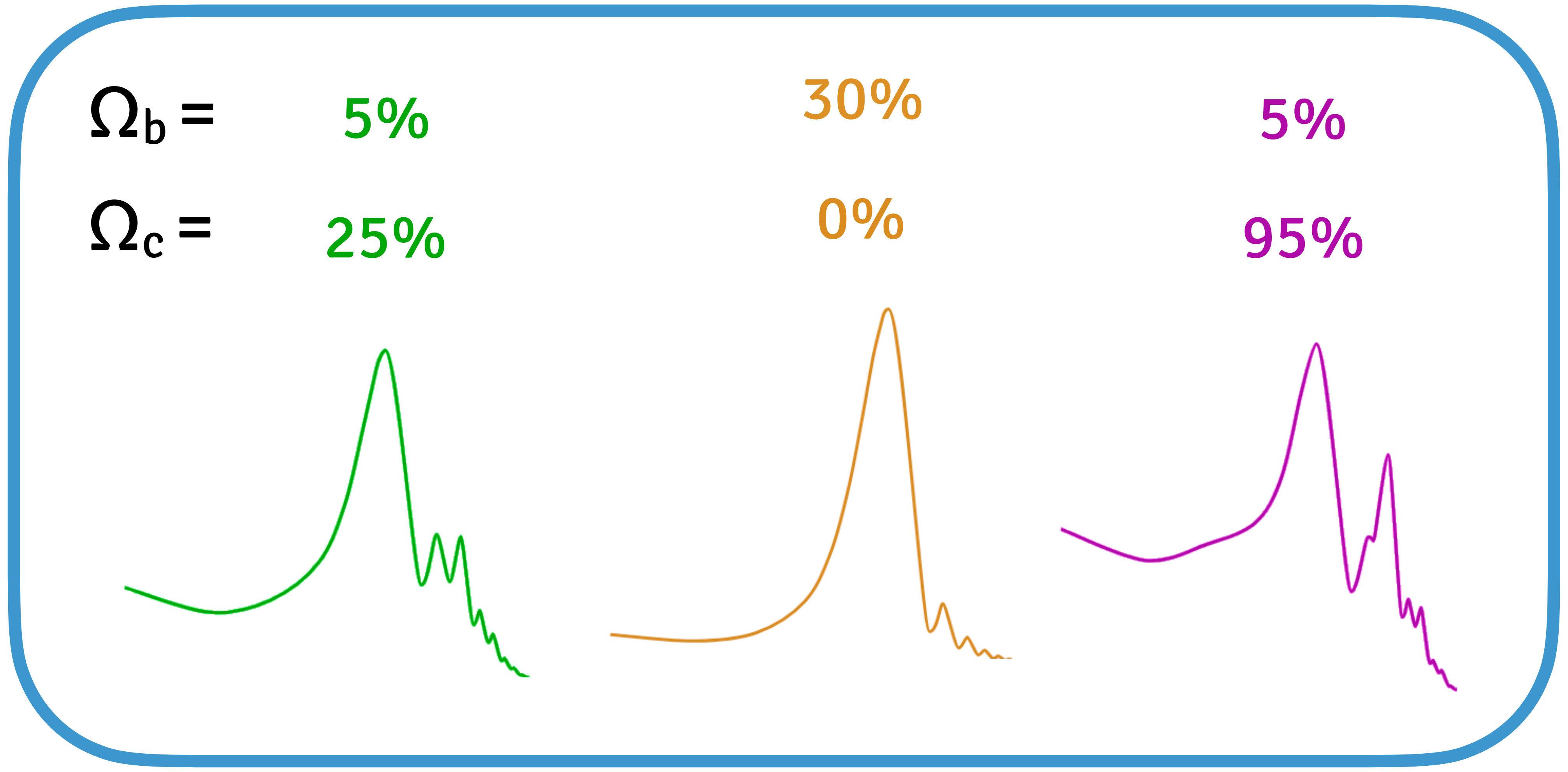


$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ **(ex:** Ω_b and Ω_c)

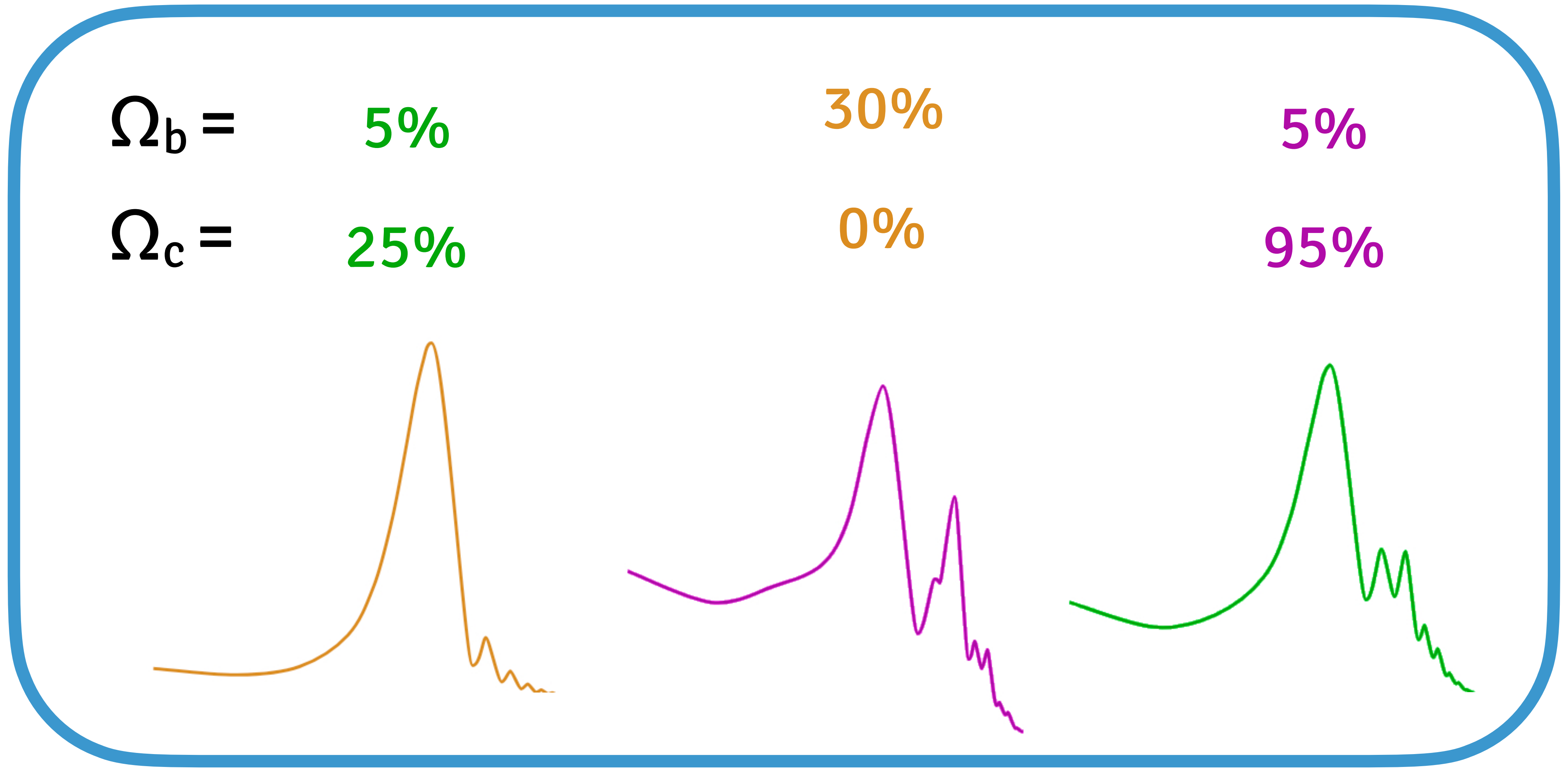
$\boldsymbol{x} \sim p(\boldsymbol{x})$ **(ex:** CMB spectra)

$$\frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta} \mid \mathbf{x})}{p(\boldsymbol{\theta})}$$

$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)

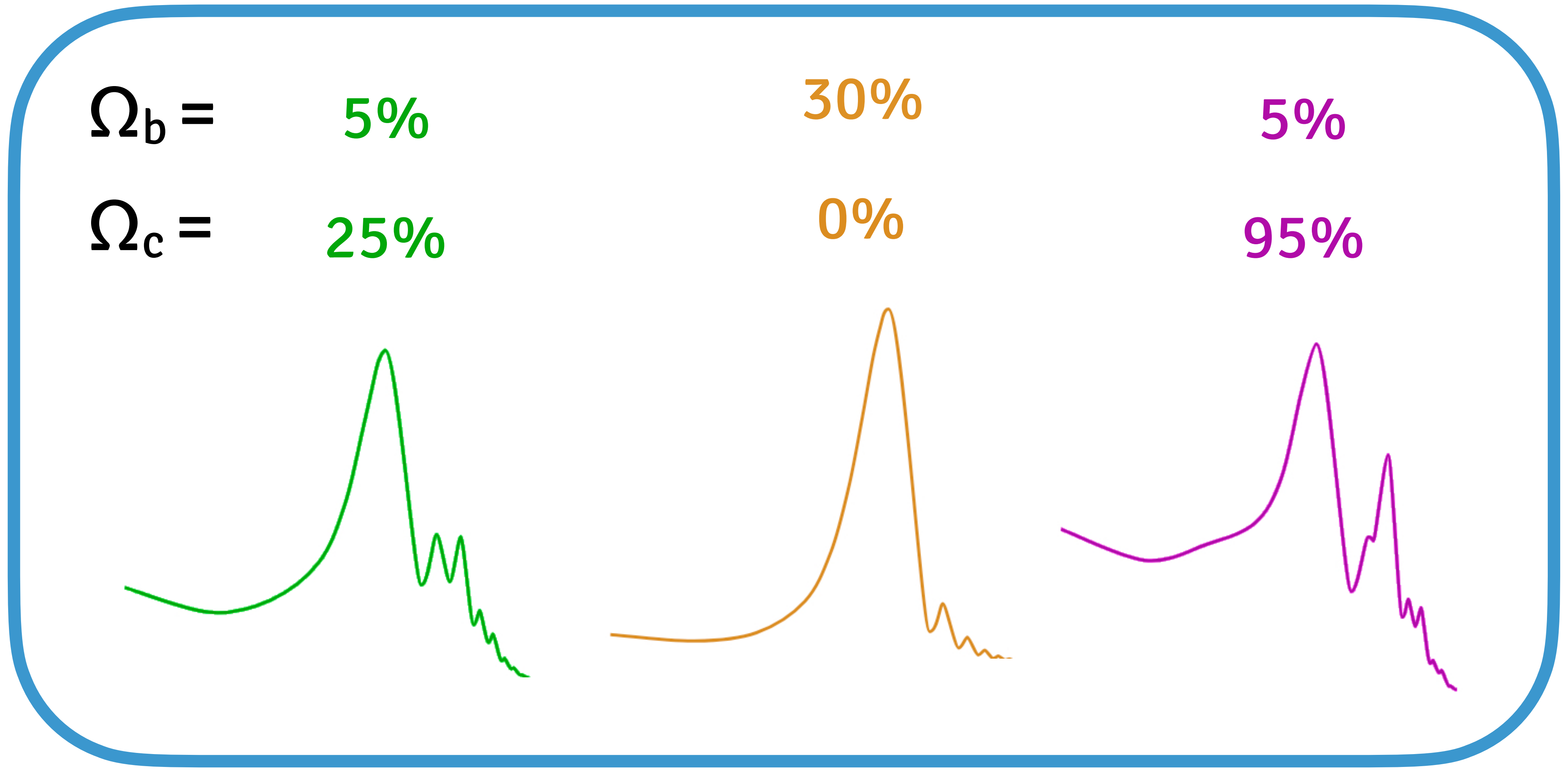


$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$$ **(ex:** Ω_b and Ω_c **)**

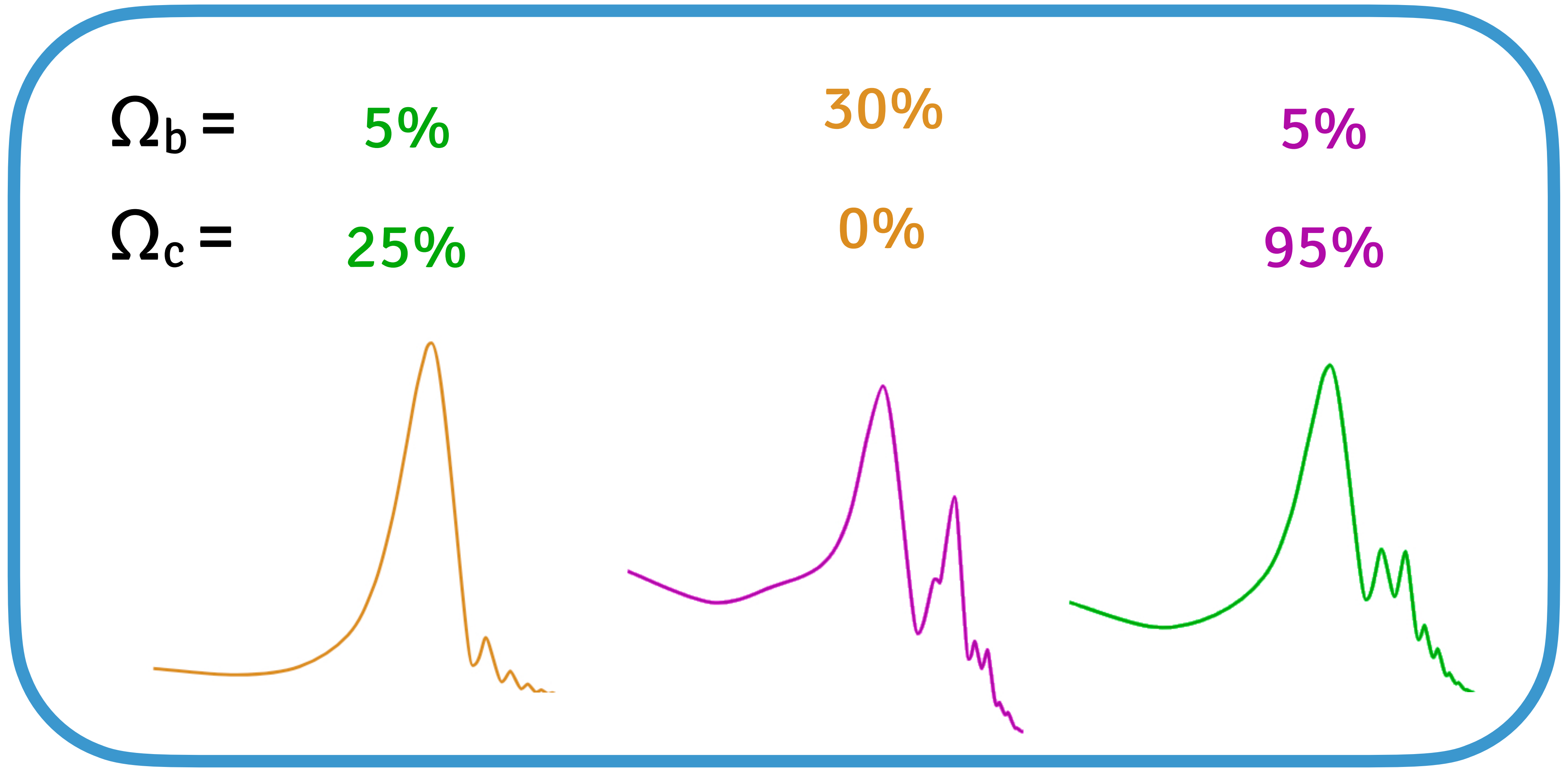
$$\boldsymbol{x} \sim p(\boldsymbol{x})$$ **(ex:** CMB spectra**)**

$$\frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta} \mid \mathbf{x})}{p(\boldsymbol{\theta})}$$

$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)



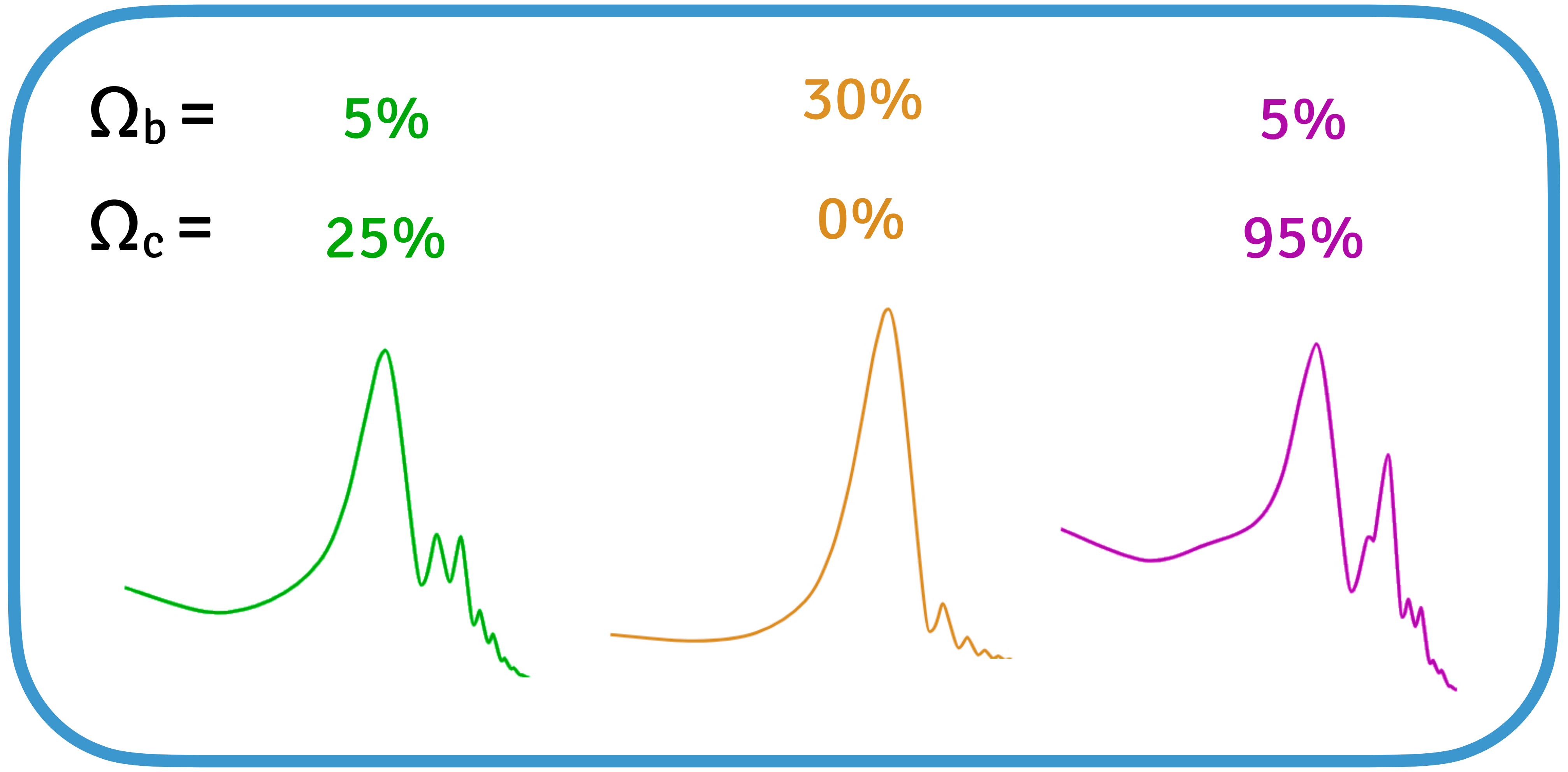
Given some (x, θ) pair, are they drawn jointly or marginally?

$\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ **(ex:** Ω_b and Ω_c **)**

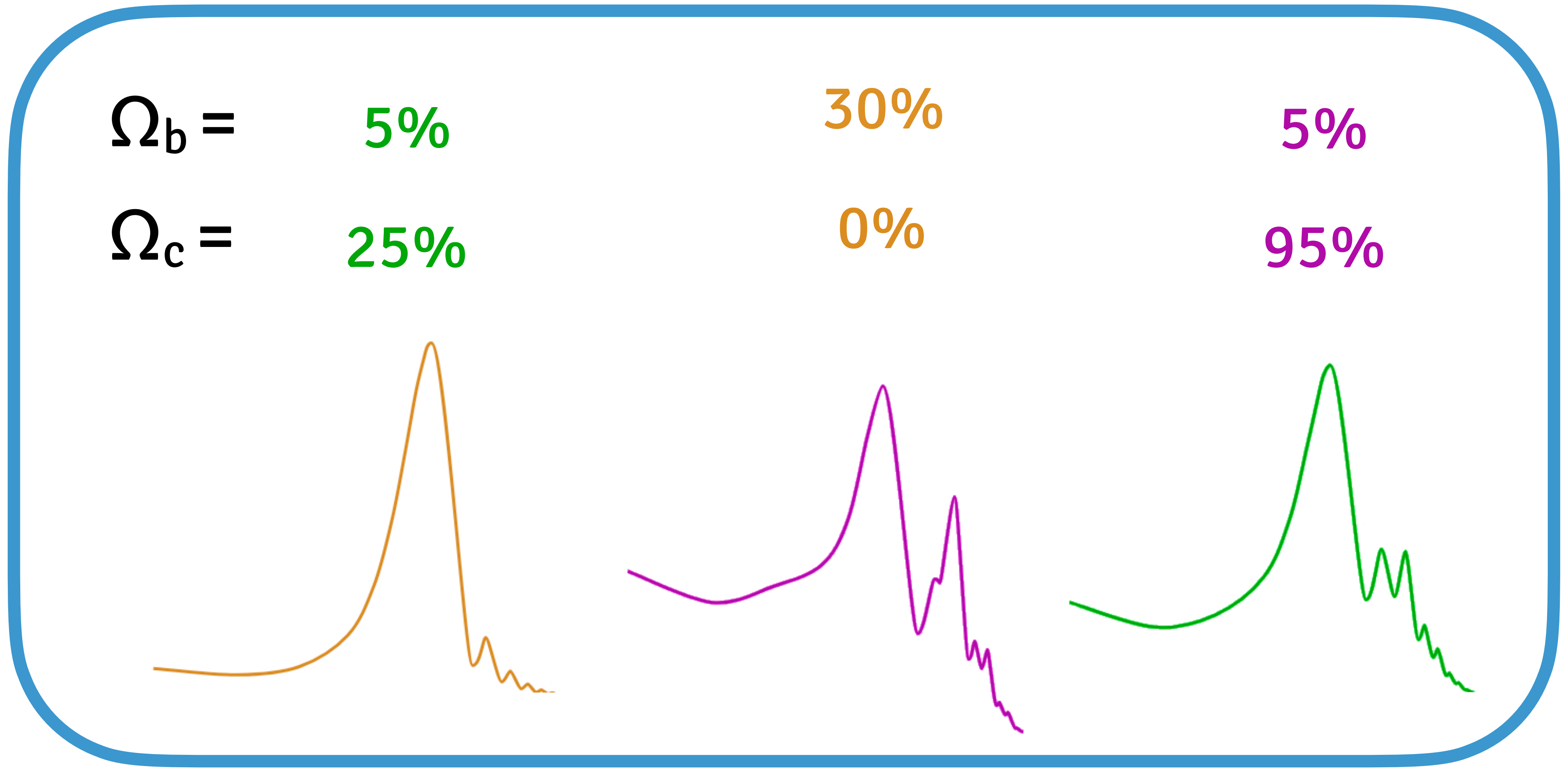
$\boldsymbol{x} \sim p(\boldsymbol{x})$ **(ex:** CMB spectra**)**

$$\frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\boldsymbol{\theta} \mid \mathbf{x})}{p(\boldsymbol{\theta})}$$

$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta})$ (jointly drawn)



$(\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x})p(\boldsymbol{\theta})$ (marginally drawn)

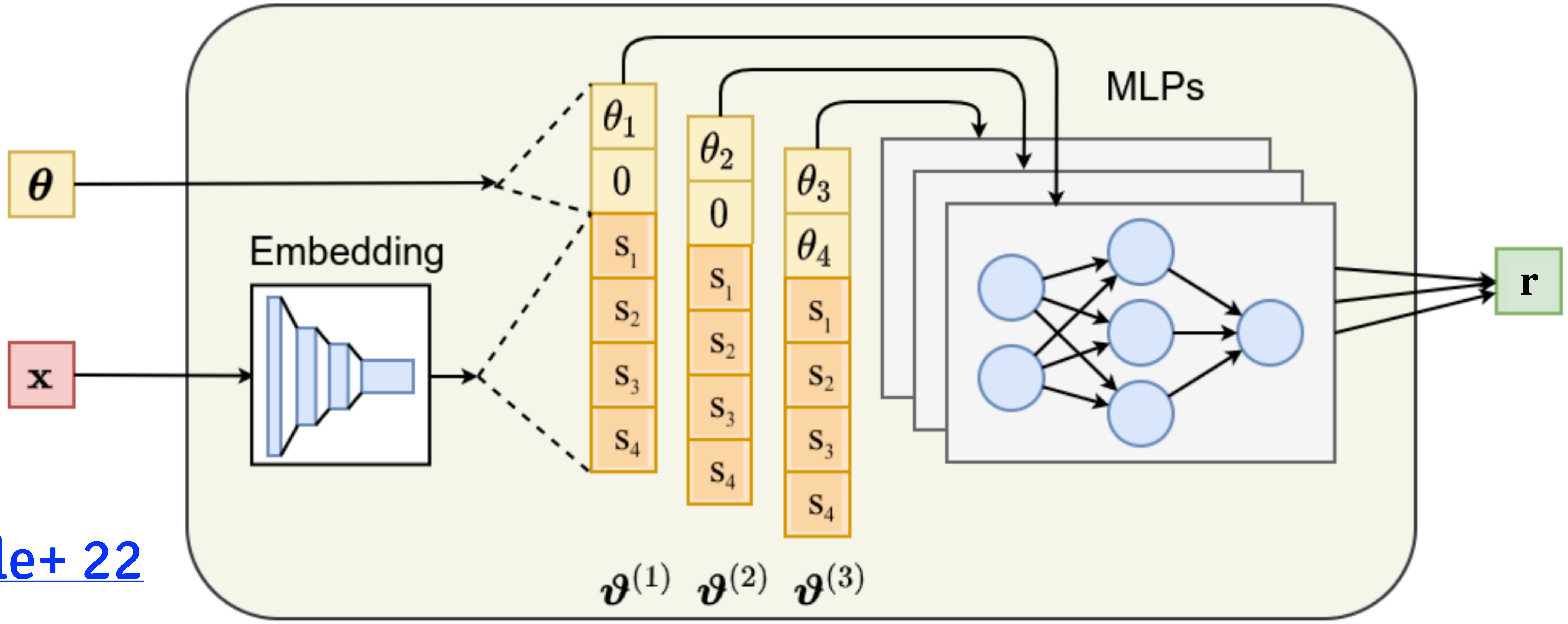


Given some (x, θ) pair, are they drawn jointly or marginally?

→ Rephrase inference as a **binary classification problem**

We can directly target marginal posteriors of interest,
and forget about the rest

We can directly target marginal posteriors of interest, and forget about the rest



Cole+ 22

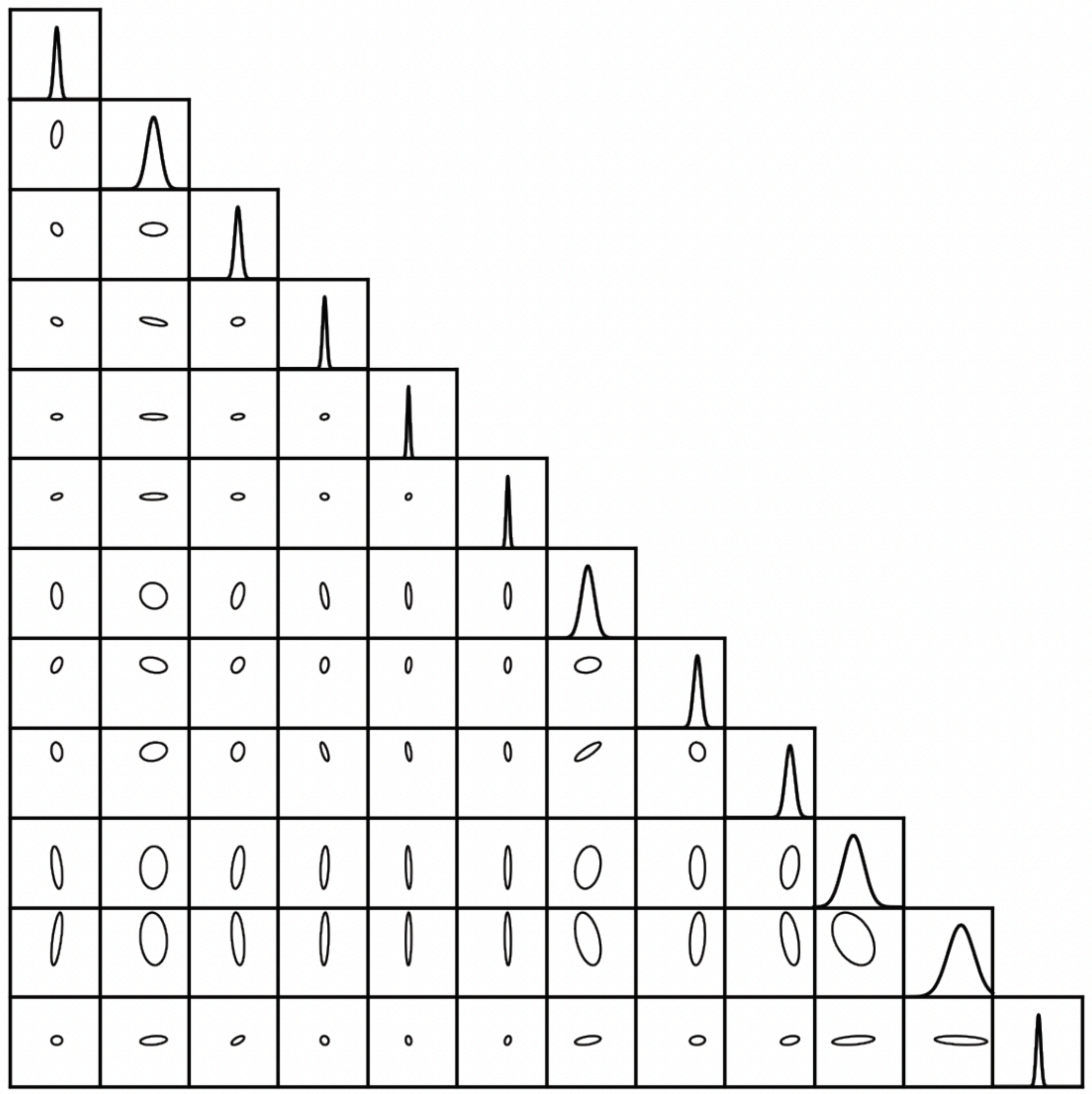
Estimates

$p(\theta_1 \,|\, \mathbf{x}), \quad p(\theta_2 \,|\, \mathbf{x}), \quad p(\theta_3, \theta_4 \,|\, \mathbf{x})$

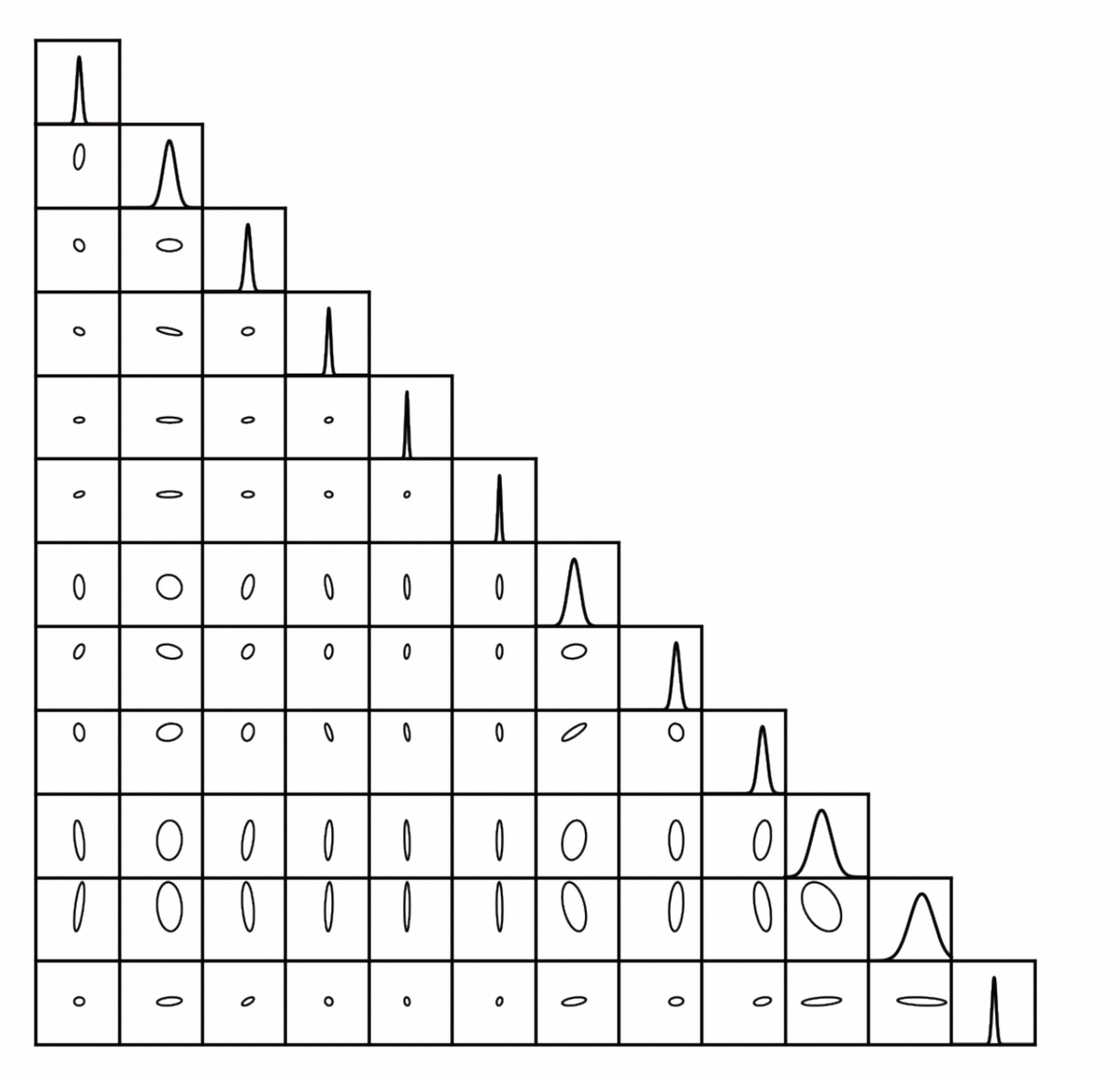
Does not estimate

$p(\theta_1, \theta_2 \,|\, \mathbf{x}), \quad p(\theta_1, \theta_2, \theta_3 \,|\, \mathbf{x}), \ldots$

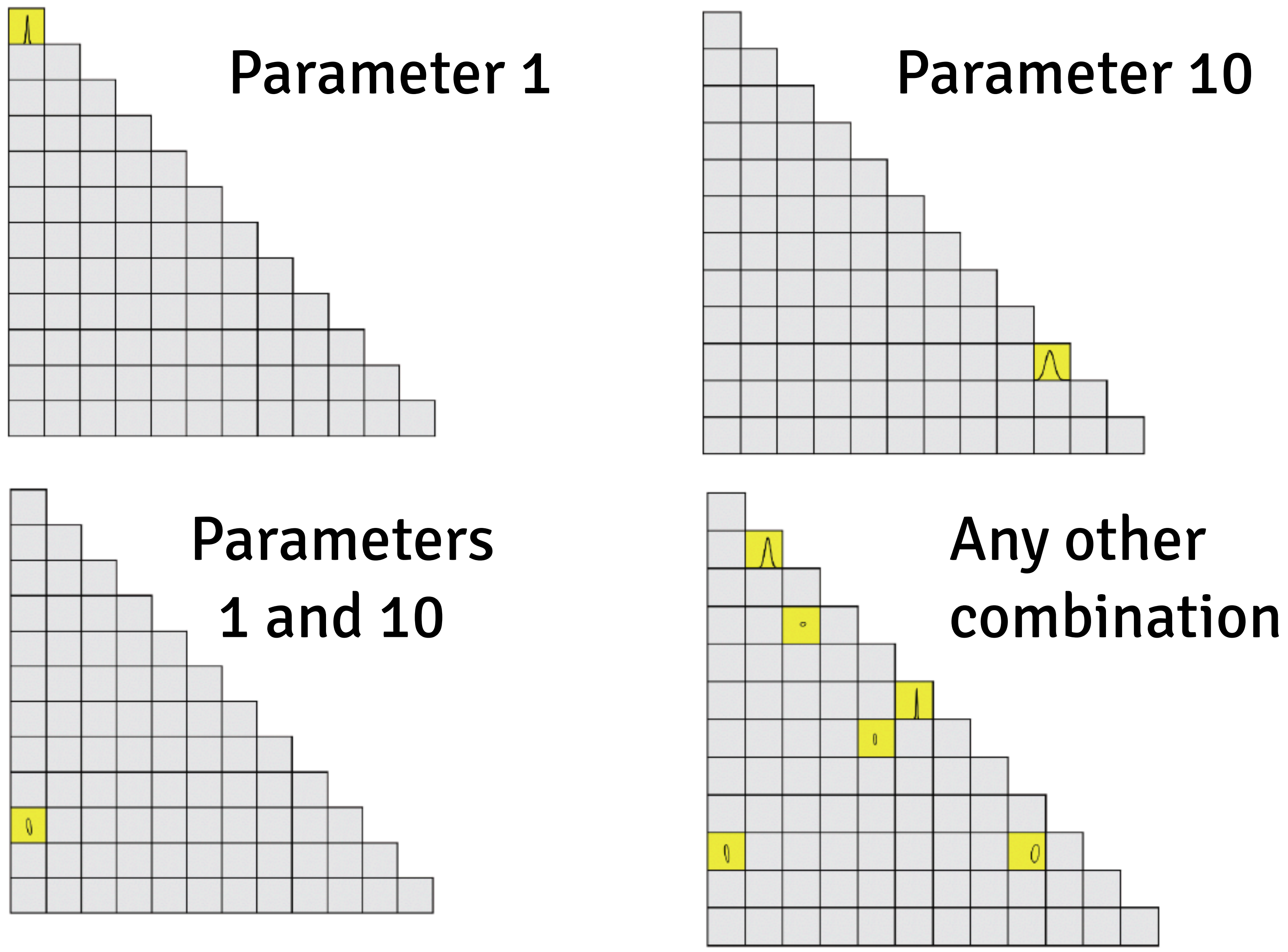
Instead of estimating all parameters...



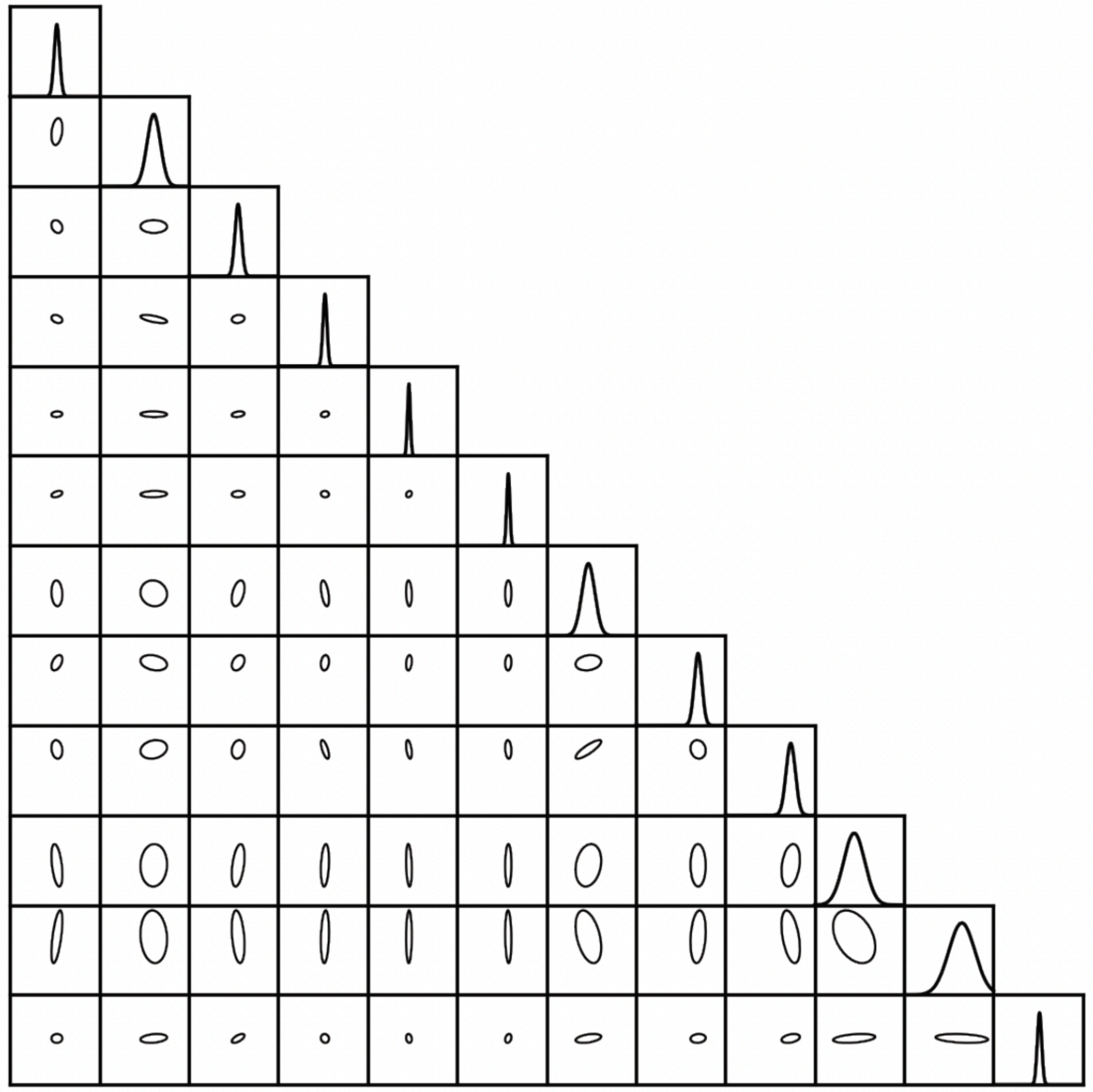
Instead of estimating all parameters...



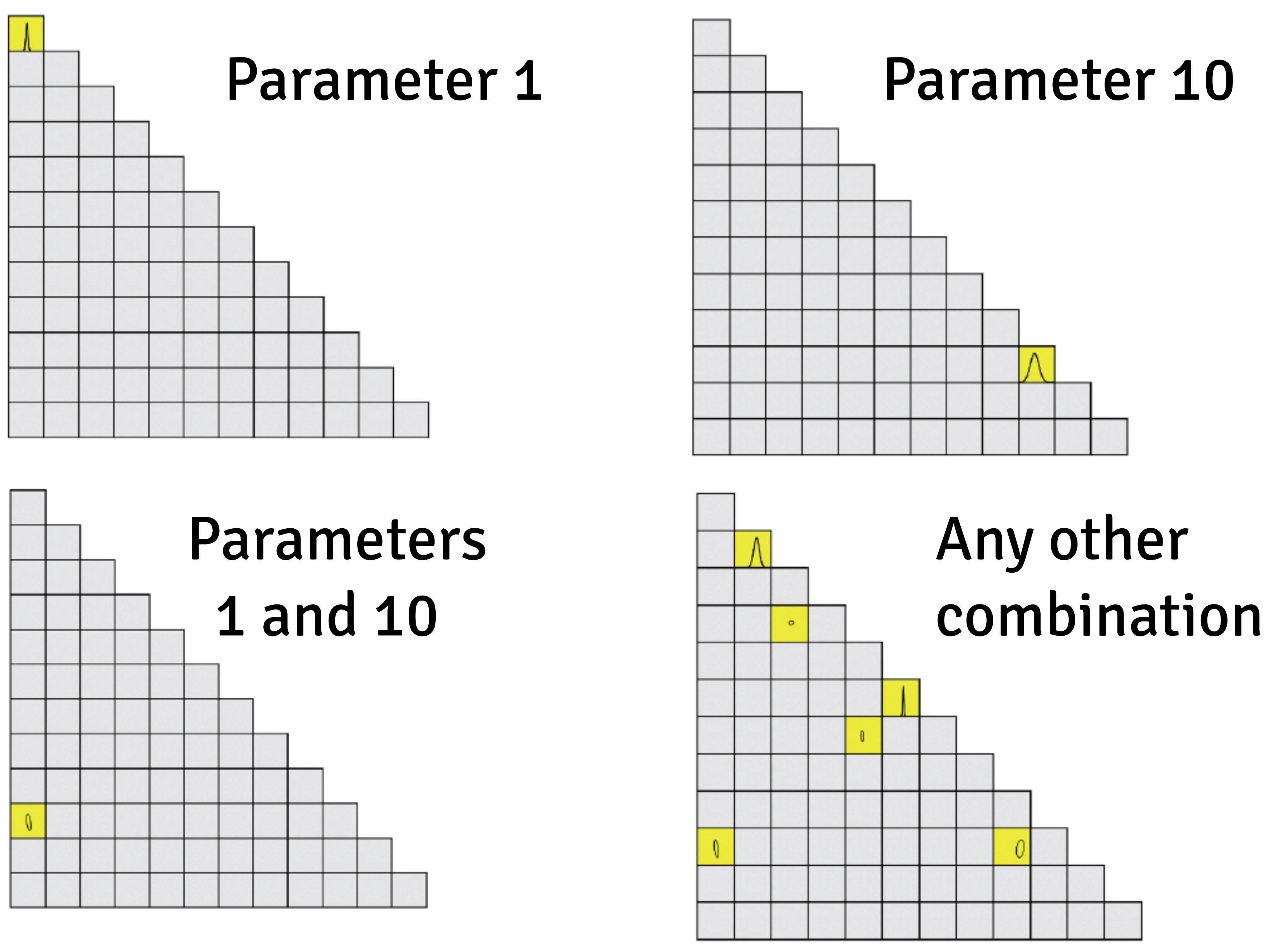
... we can cherry-pick what we care about



Instead of estimating all parameters...



... we can cherry-pick what we care about



Much more flexible much more efficient!

MNRE has been successfully applied in many contexts:

- **Strong lensing** [Montel+ 22]
- **Stellar Streams** [Alvey+ 23]
- **Gravitational Waves** [Bhardwaj+ 23] [Alvey+ 23]
- **CMB** [Cole+ 22]
- **21-cm** [Saxena+ 23]

MNRE has been successfully applied in many contexts:

- **Strong lensing** [Montel+ 22]
- **Stellar Streams** [Alvey+ 23]
- **Gravitational Waves** [Bhardwaj+ 23] [Alvey+ 23]
- **CMB** [Cole+ 22]
- **21-cm** [Saxena+ 23]

Our goal: apply MNRE to **Stage IV** photometric observables

I. Why we need to go **beyond MCMC**

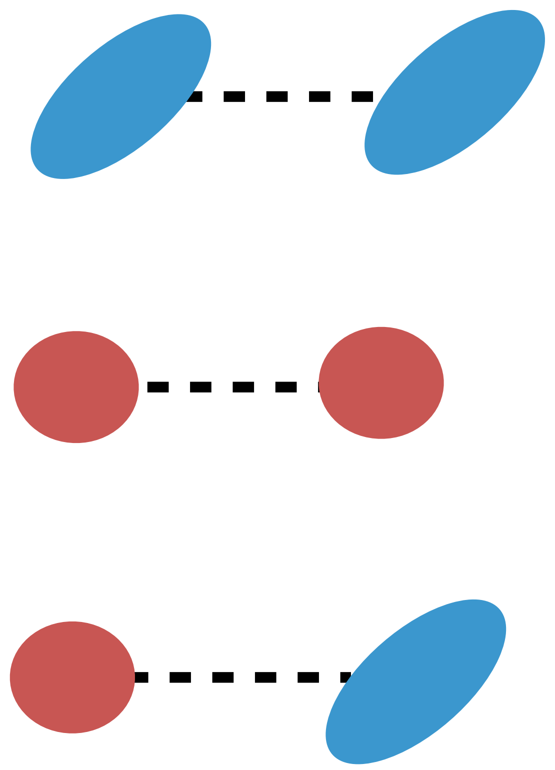
II. Our new approach: **Marginal Neural Ratio Estimation**

III. Applying MNRE to **Stage IV** photometric observables ✓

3x2pt photometric probes

Summarise maps of galaxy positions/shapes using three 2-point statistics **(3x2pt)** measured at 10 tomographic redshift bins

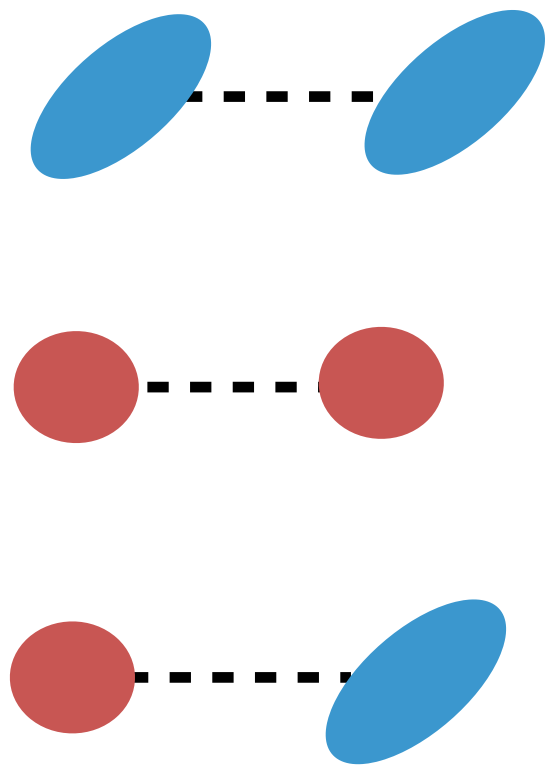
- **Cosmic Shear**
- **Galaxy clustering**
- **Galaxy-Galaxy lensing**



3x2pt photometric probes

Summarise maps of galaxy positions/shapes using three 2-point statistics **(3x2pt)** measured at 10 tomographic redshift bins

- **Cosmic Shear**
- **Galaxy clustering**
- **Galaxy-Galaxy lensing**



...described by angular power spectra $$C_{ij}^{XY}(\ell) = \int dz\ W_i^X(z) W_j^Y(z)\ P_m(k_\ell, z)$$

Swyft 3x2pt analysis

1. Simulator

Swyft 3x2pt analysis

1. Simulator

We generate 50k realisations of 3x2pt spectra with gaussian noise

$$\hat{C}_{ij}^{AB}(\ell) = C_{ij}^{AB}(\ell) + n_{ij}^{AB}(\ell)$$

$n_{ij}^{AB}(\ell) \leftarrow \mathcal{N}(0,\mathbf{C})$

Swyft 3x2pt analysis

1. Simulator

We generate 50k realisations of 3x2pt spectra with gaussian noise

$$\hat{C}_{ij}^{AB}(\ell) = C_{ij}^{AB}(\ell) + n_{ij}^{AB}(\ell)$$

$n_{ij}^{AB}(\ell) \leftarrow \mathcal{N}(0,\mathbf{C})$

12 nuisance params

$$\{A_{\mathrm{IA}}, \eta_{\mathrm{IA}}, b_1, \ldots, b_{10}\}$$

Swyft 3x2pt analysis

1. Simulator

We generate 50k realisations of 3x2pt spectra with gaussian noise

$$\hat{C}_{ij}^{AB}(\ell) = C_{ij}^{AB}(\ell) + n_{ij}^{AB}(\ell)$$

$n_{ij}^{AB}(\ell) \leftarrow \mathcal{N}(0,\mathbf{C})$

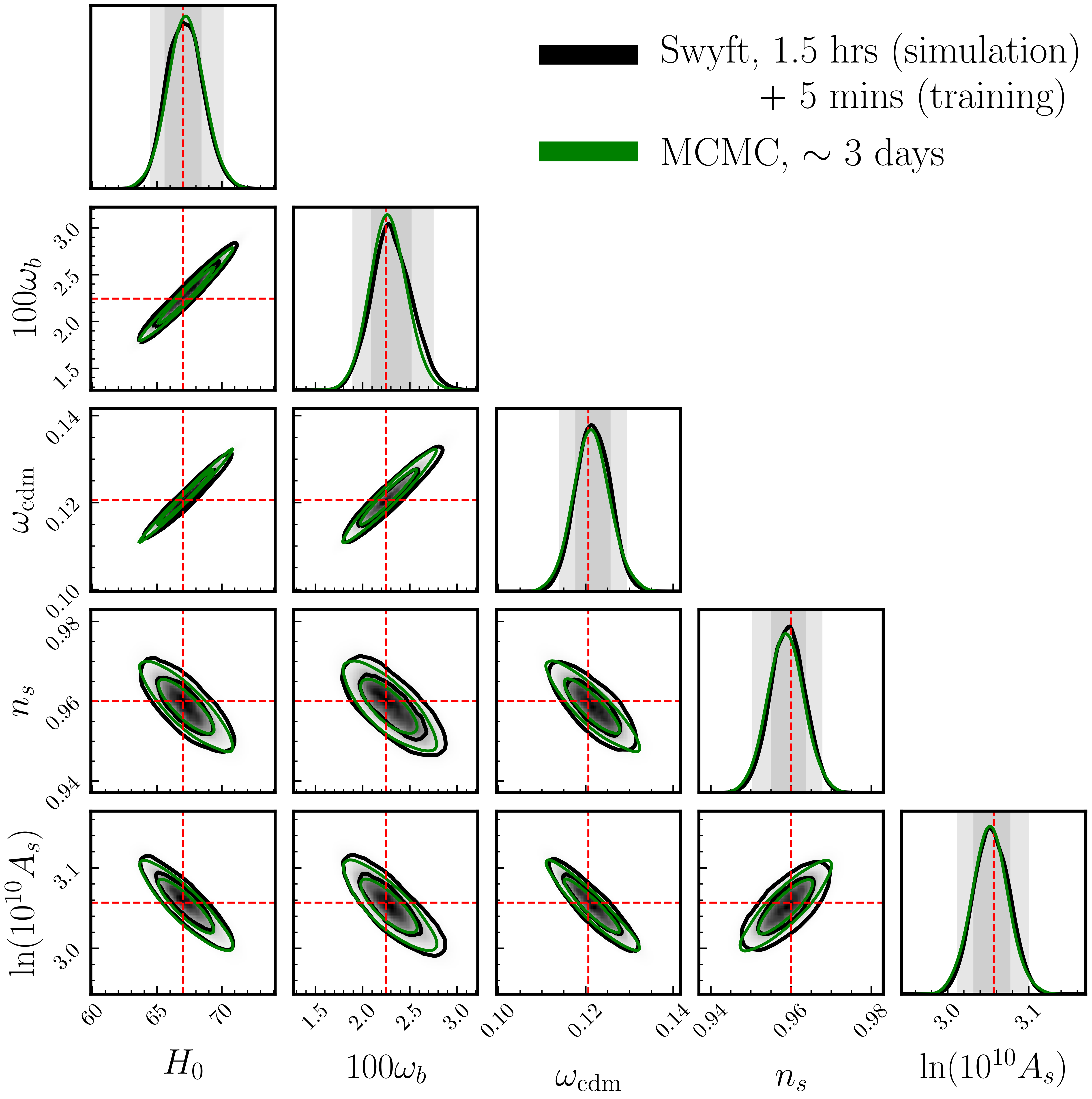
12 nuisance params

$$\{A_{\rm IA}, \eta_{\rm IA}, b_1, \ldots, b_{10}\}$$

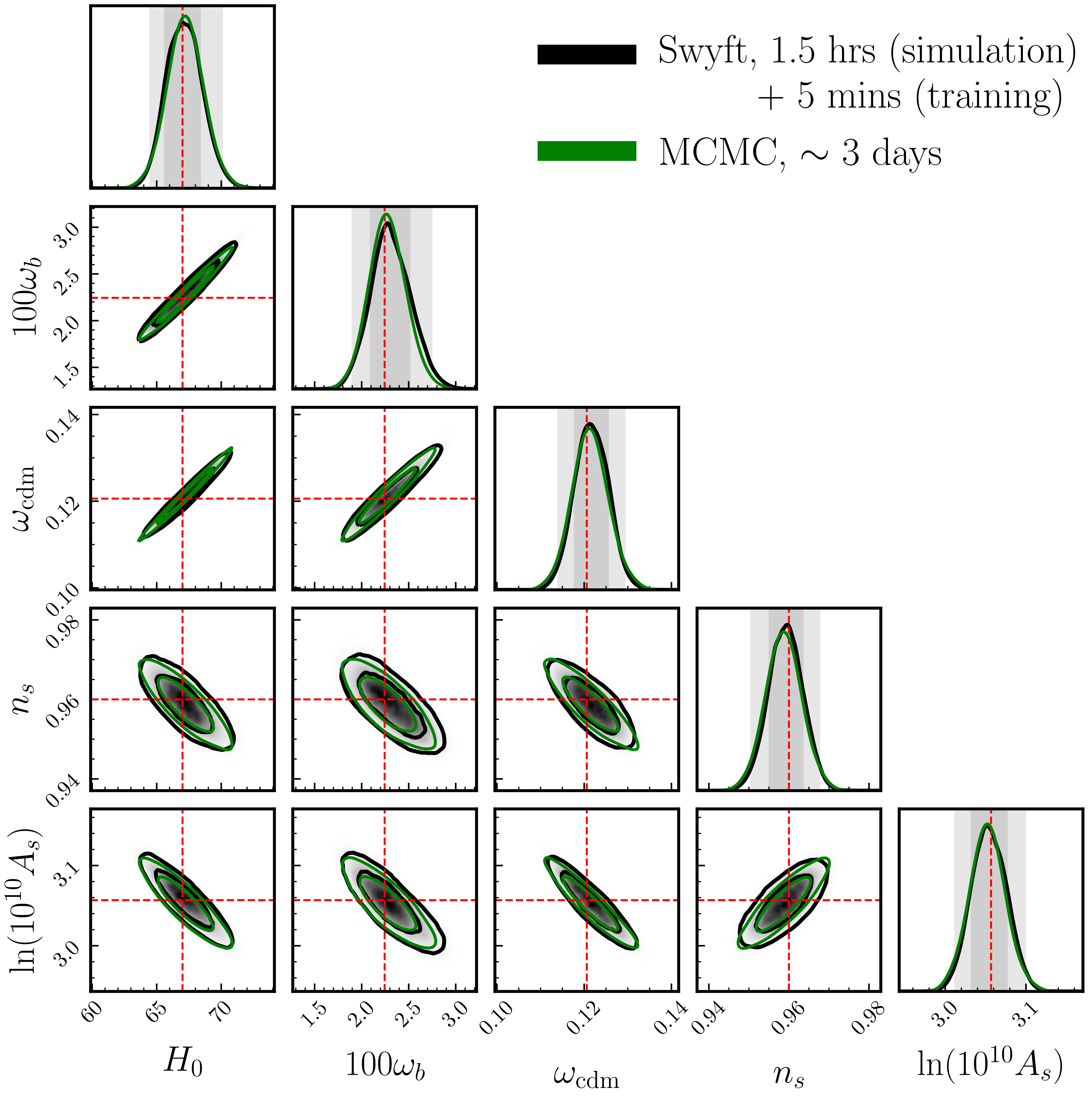
2. Network

We do a **pre-compression** of data using **PCA** and parameter-specific data summaries

Forecast ΛCDM posteriors

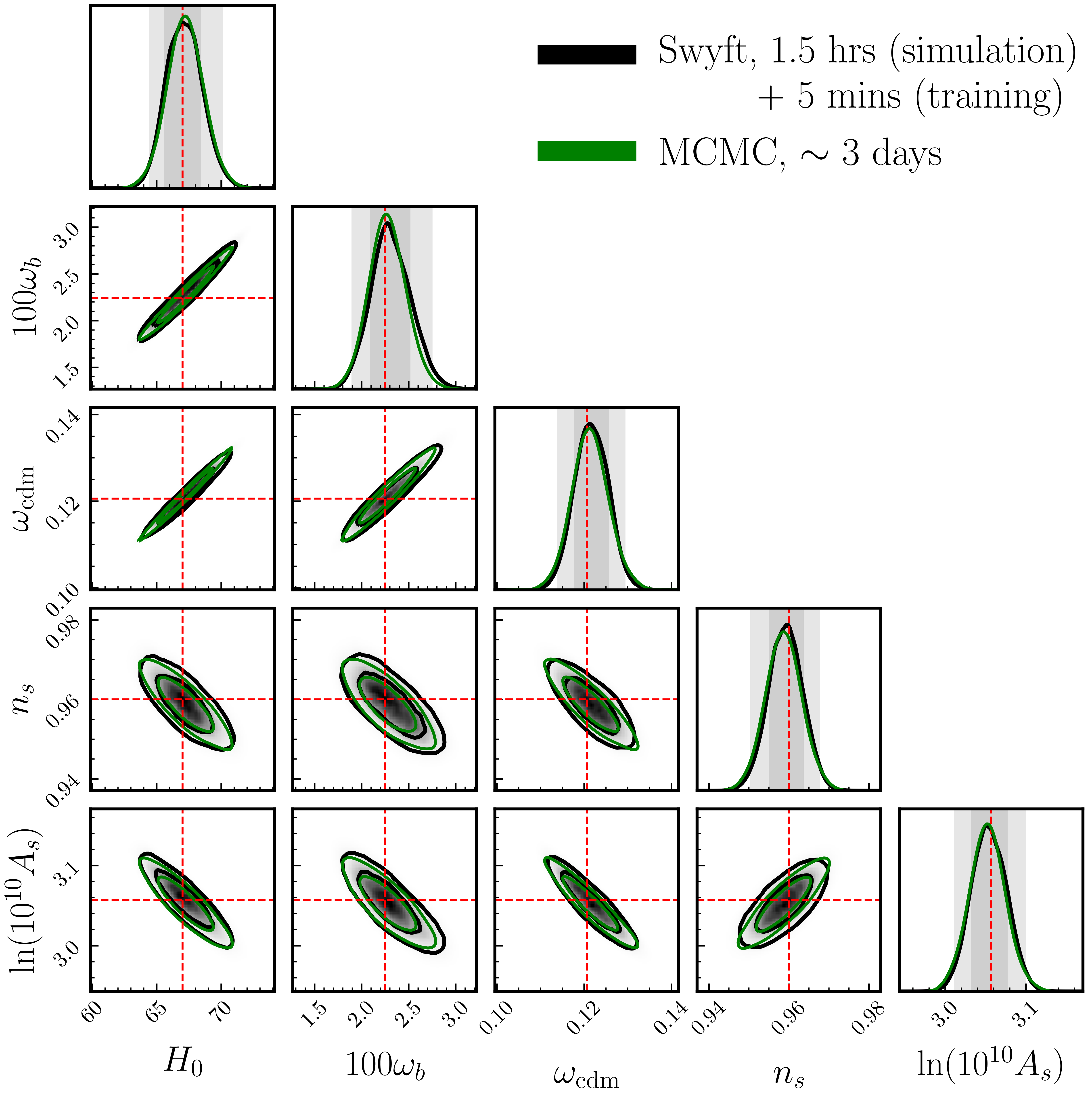


Forecast ΛCDM posteriors



MNRE & MCMC are in excellent agreement!

Forecast ΛCDM posteriors



Dramatic reduction in CPU time!

MNRE & MCMC are in excellent agreement!

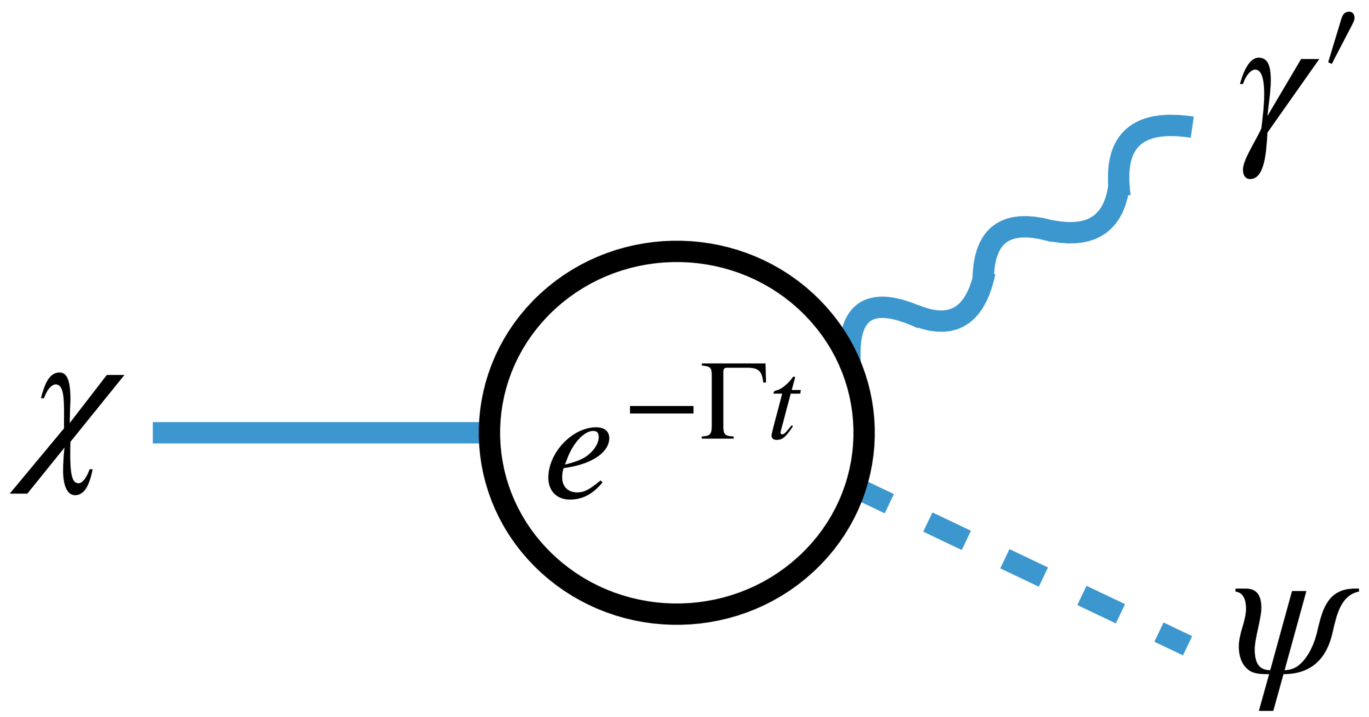
With MNRE, we can also test models with highly non-Gaussian posteriors

- With MNRE, we can also test models with highly non-Gaussian posteriors

- As an example, we consider a model of **CDM decaying to DR + WDM** (proposed to explain the S_8 tension)

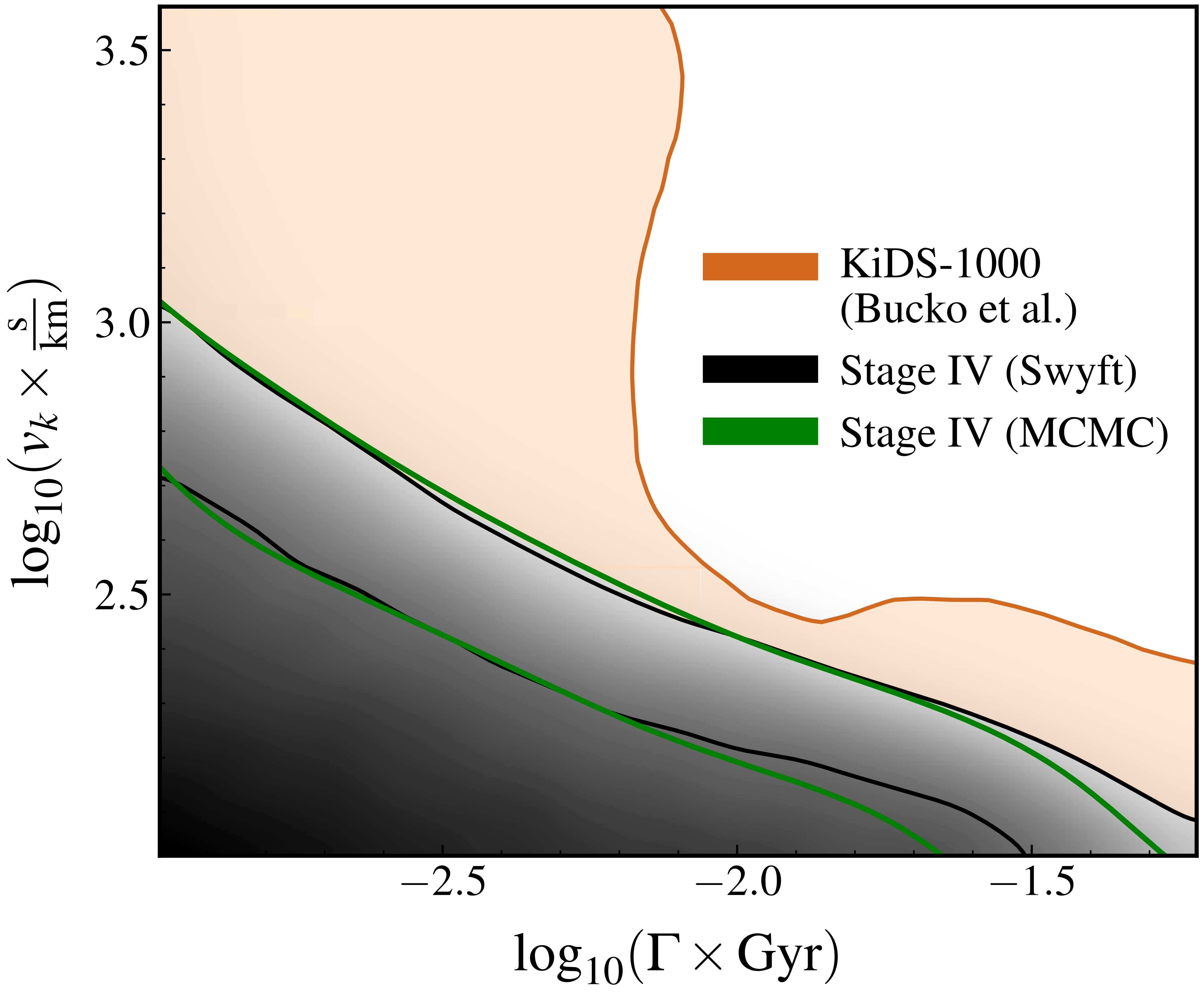
[Abellan+ 21]

[Bucko+ 23]

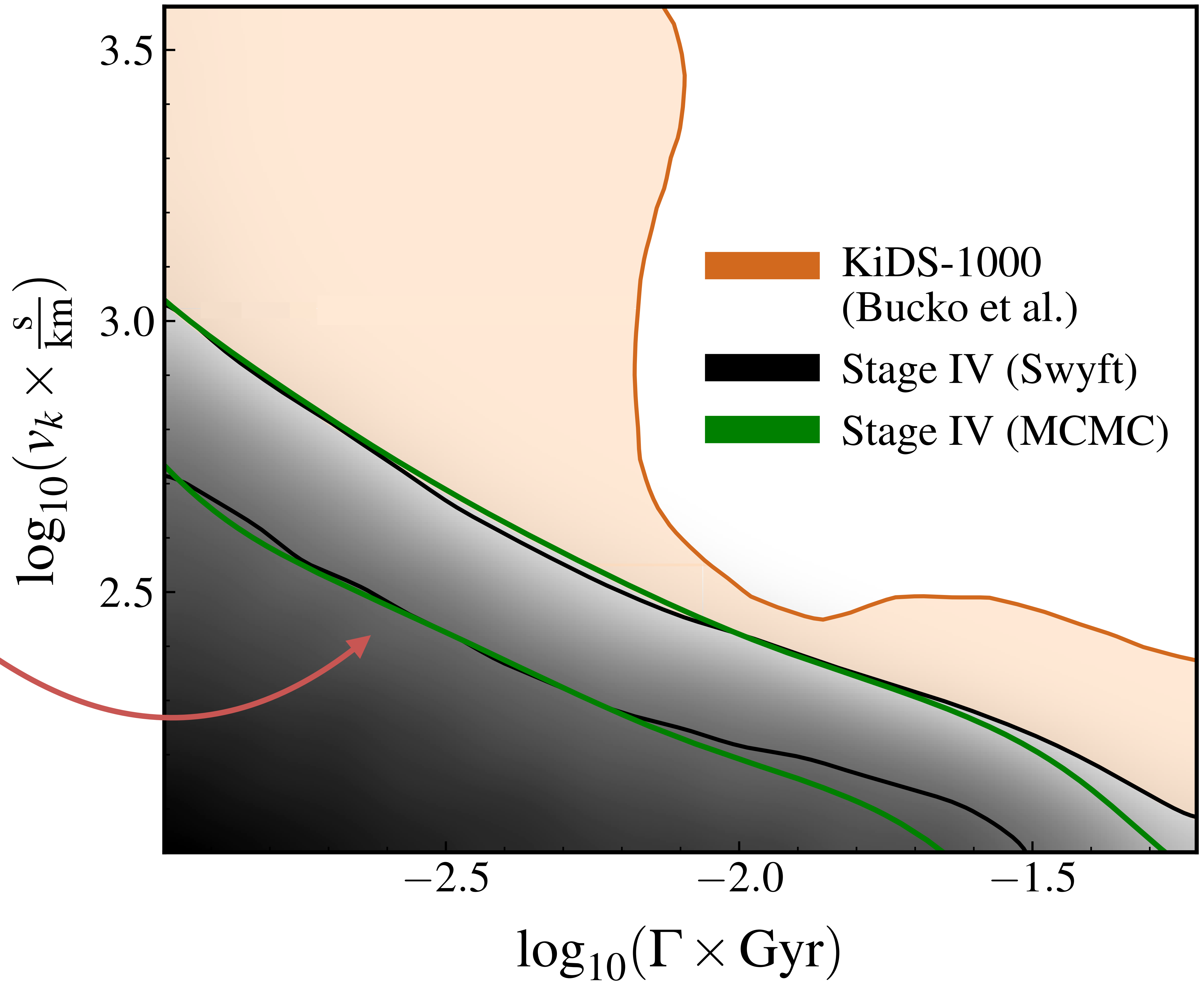


Decay rate Γ
WDM velocity kick v_k

Forecast constraints on decaying DM

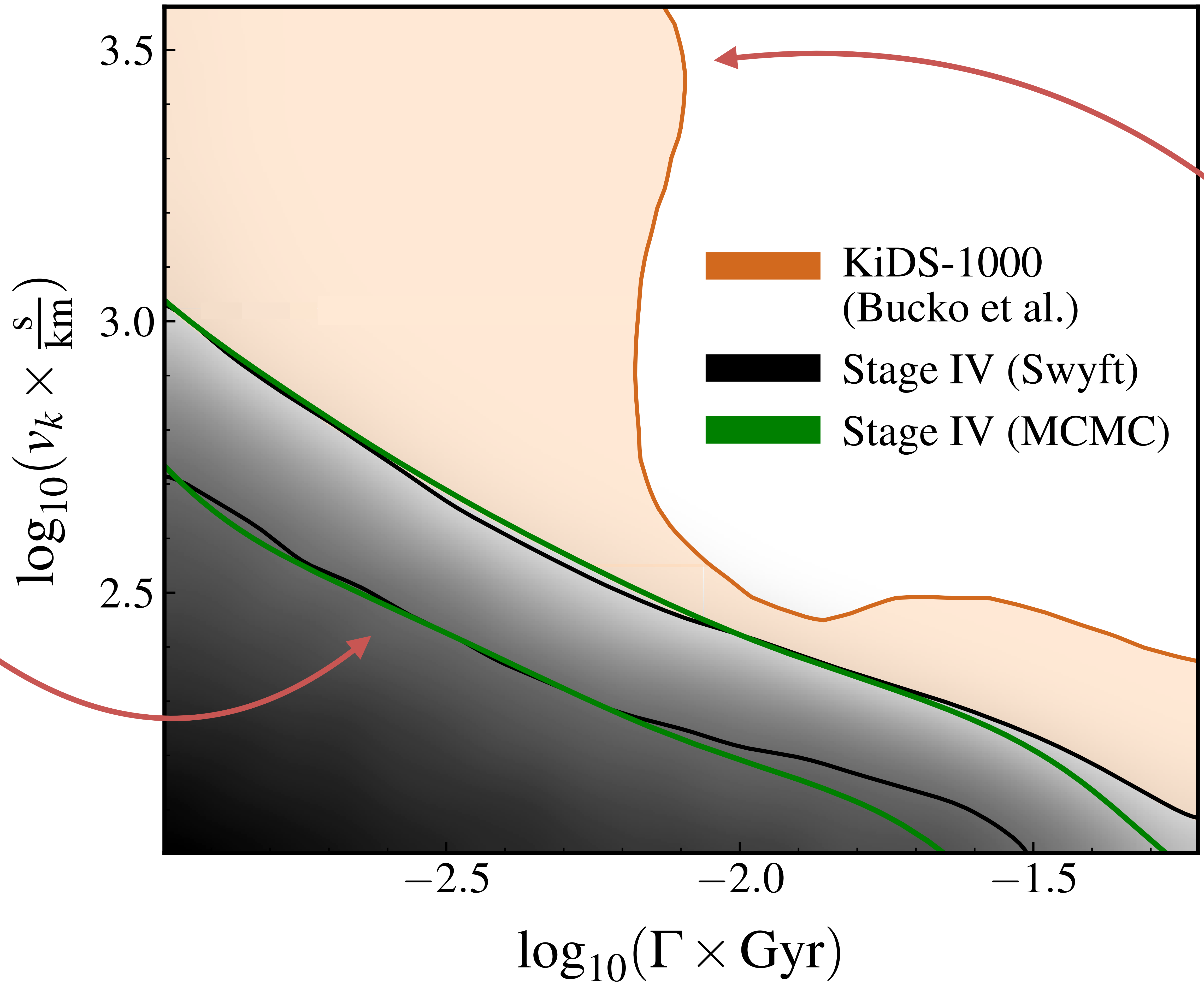


Forecast constraints on decaying DM



MNRE & MCMC in good agreement

Forecast constraints on decaying DM



MNRE & MCMC in good agreement

Improve current limits by ~1 order of magnitude!

Next steps

- Use simulator based on CLOE (many more nuisance params.)

Next steps

- Use simulator based on CLOE (many more nuisance params.)
- Consider other observables, like spectroscopic galaxy clustering

Next steps

- Use simulator based on CLOE (many more nuisance params.)
- Consider other observables, like spectroscopic galaxy clustering
- Perform field-level inference to extract all possible information

[Lemos+ 23]

[Jeffrey+ 24]

Conclusions

- To learn as much as we can about the dark sector from future data, we need to go beyond traditional methods

Conclusions

- To learn as much as we can about the dark sector from future data, we need to go beyond traditional methods

- MNRE provides a powerful framework to constrain ΛCDM and its extensions with next-generation LSS surveys (like Euclid)

Conclusions

- To learn as much as we can about the dark sector from future data, we need to go beyond traditional methods

- MNRE provides a powerful framework to constrain ΛCDM and its extensions with next-generation LSS surveys (like Euclid)

THANKS FOR YOUR ATTENTION

g.francoabellan@uva.nl

BACK-UP

Strategy: train a neural network $d_\phi(\mathbf{x}, \boldsymbol{\theta}) \in [0,1]$ as a binary classifier, so that

$$d_\phi(\mathbf{x}, \boldsymbol{\theta}) \simeq 1 \quad \text{if} \quad (\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x} \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$d_\phi(\mathbf{x}, \boldsymbol{\theta}) \simeq 0 \quad \text{if} \quad (\mathbf{x}, \boldsymbol{\theta}) \sim p(\mathbf{x}) p(\boldsymbol{\theta})$$

Note: Φ denotes all the network parameters

We have to minimise a **loss function** w.r.t. the network params. Φ

$$L[d_\phi(\mathbf{x},\boldsymbol{\theta})] = -\int d\mathbf{x}d\boldsymbol{\theta}\left[p(\mathbf{x},\boldsymbol{\theta})\ln(d_\phi(\mathbf{x},\boldsymbol{\theta})) + p(\mathbf{x})p(\boldsymbol{\theta})\ln(1 - d_\phi(\mathbf{x},\boldsymbol{\theta}))\right]$$

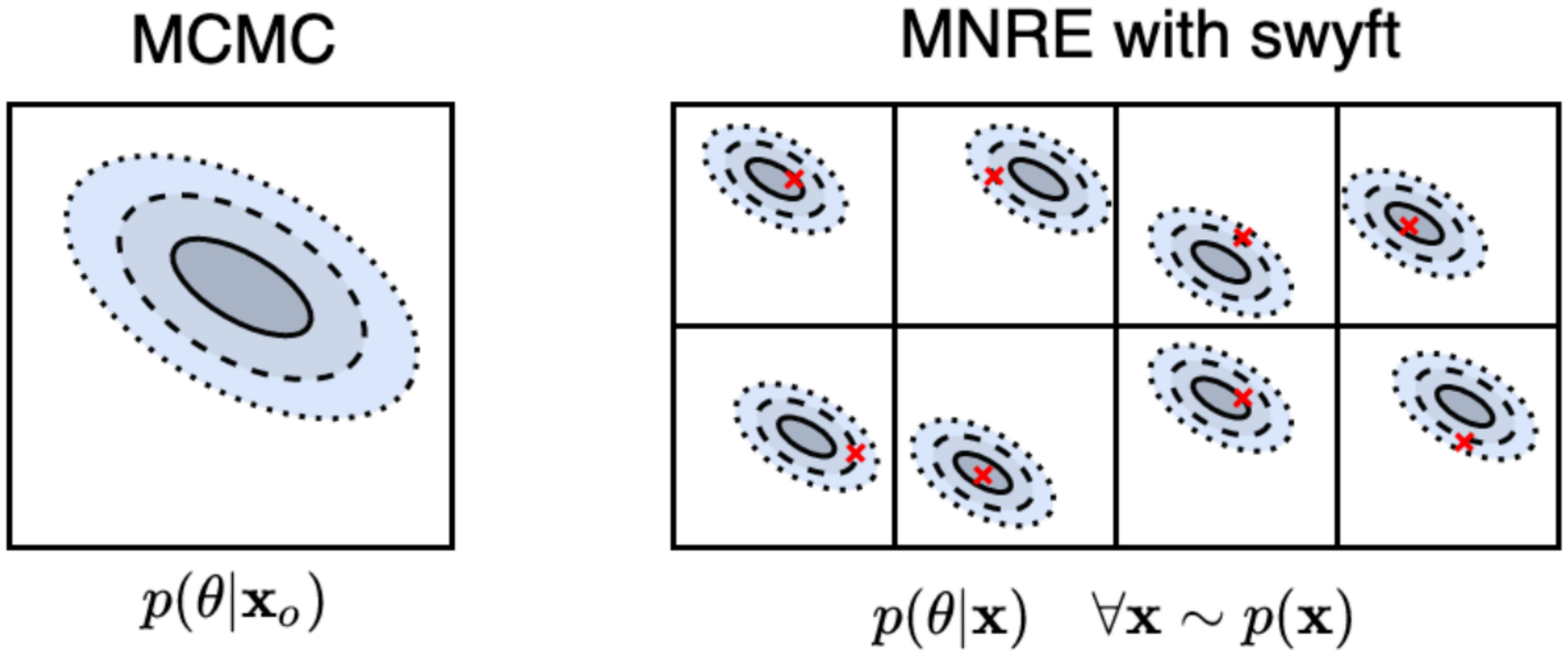
which yields

$$d_\phi(\mathbf{x},\boldsymbol{\theta}) \simeq \frac{p(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x},\boldsymbol{\theta}) + p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{r(\mathbf{x};\boldsymbol{\theta})}{r(\mathbf{x};\boldsymbol{\theta}) + 1}$$

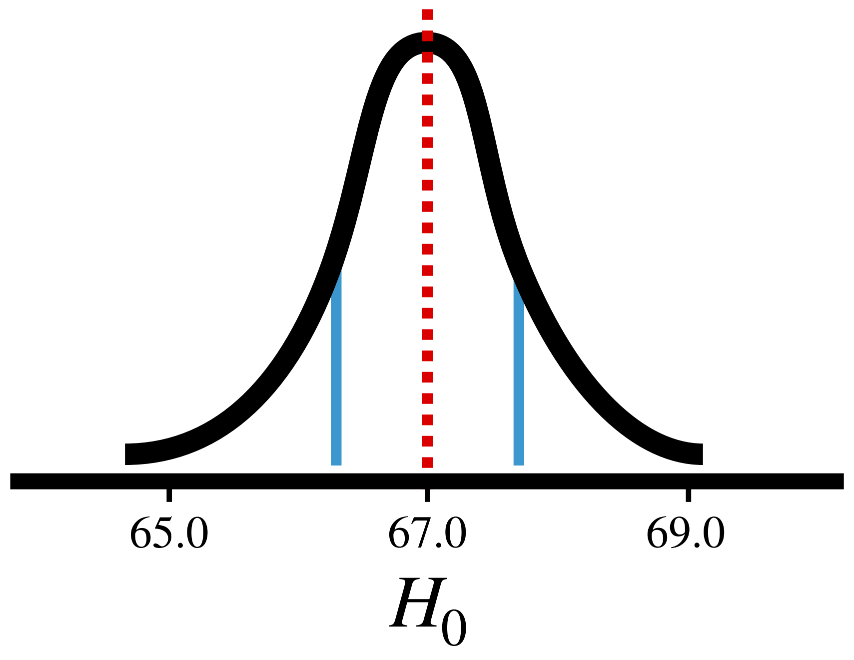
But can we trust our results?

...even if NNs are often seen as "black boxes”, it is possible to perform statistical consistency tests which are impossible with MCMC

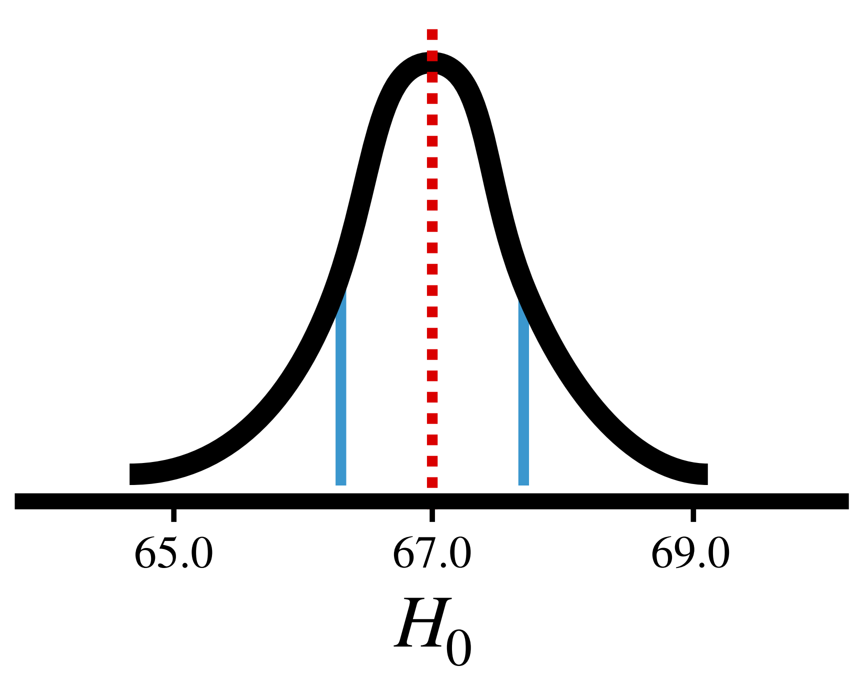
Trained networks can estimate effortlessly
the posteriors for all simulated observations

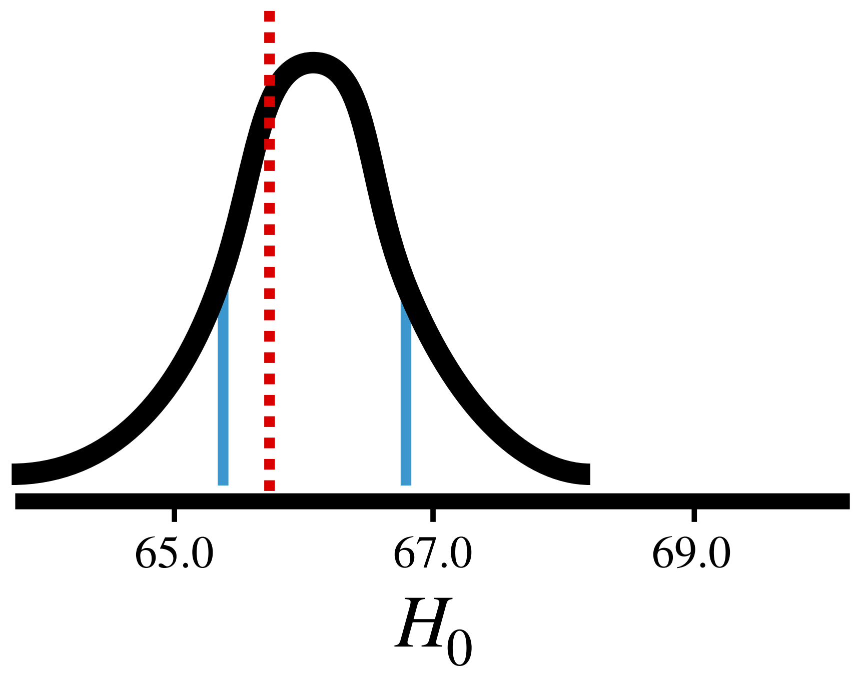


Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

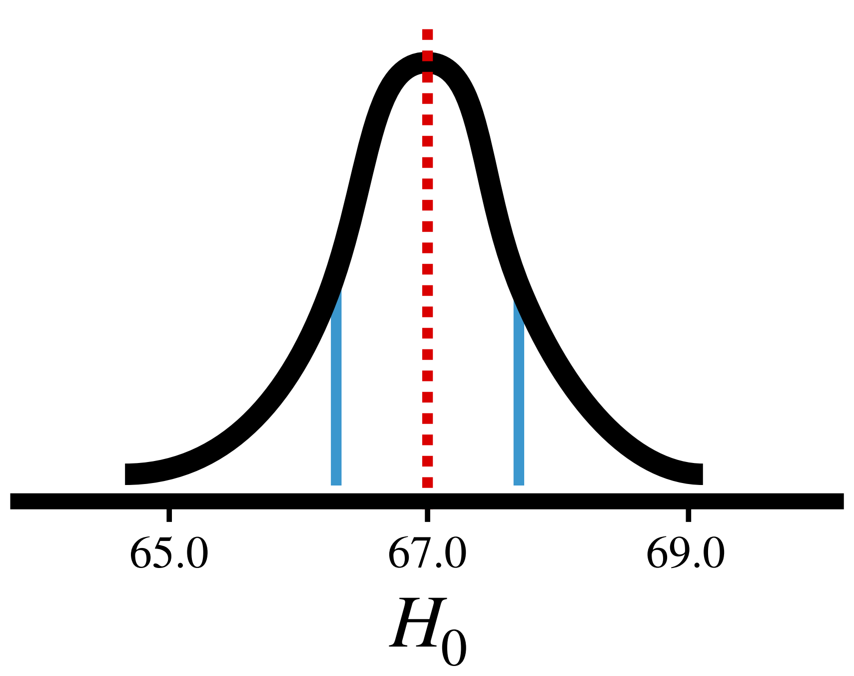


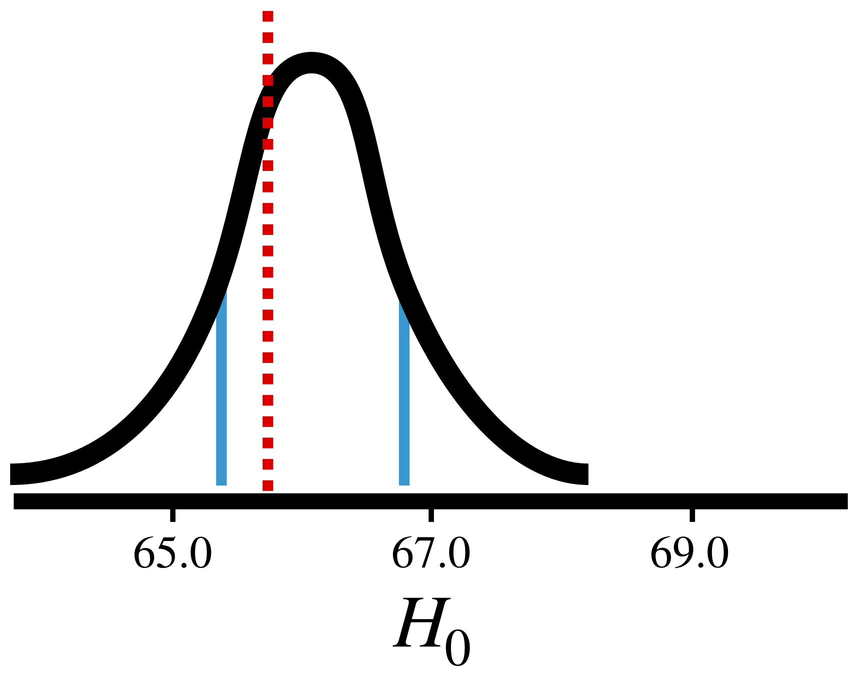
Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

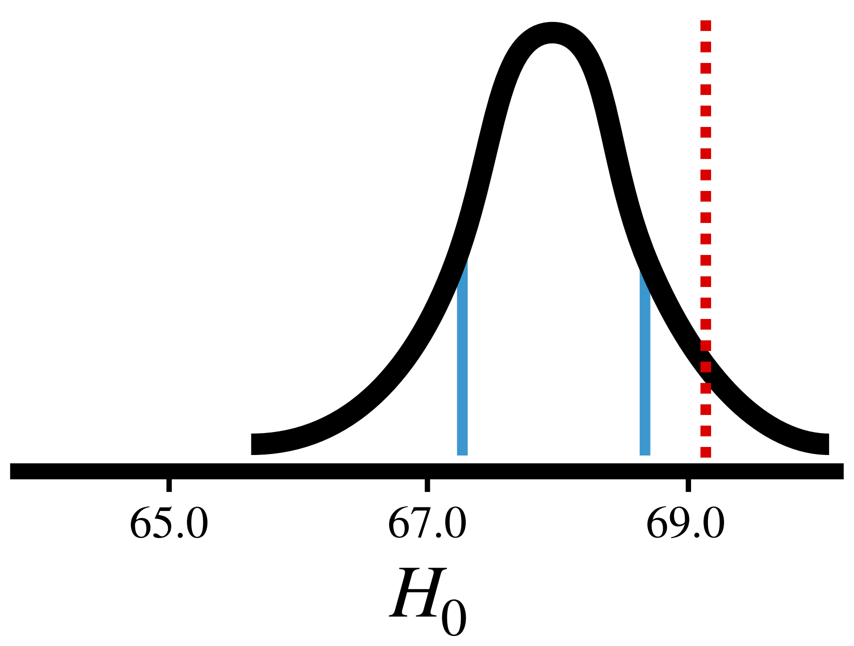




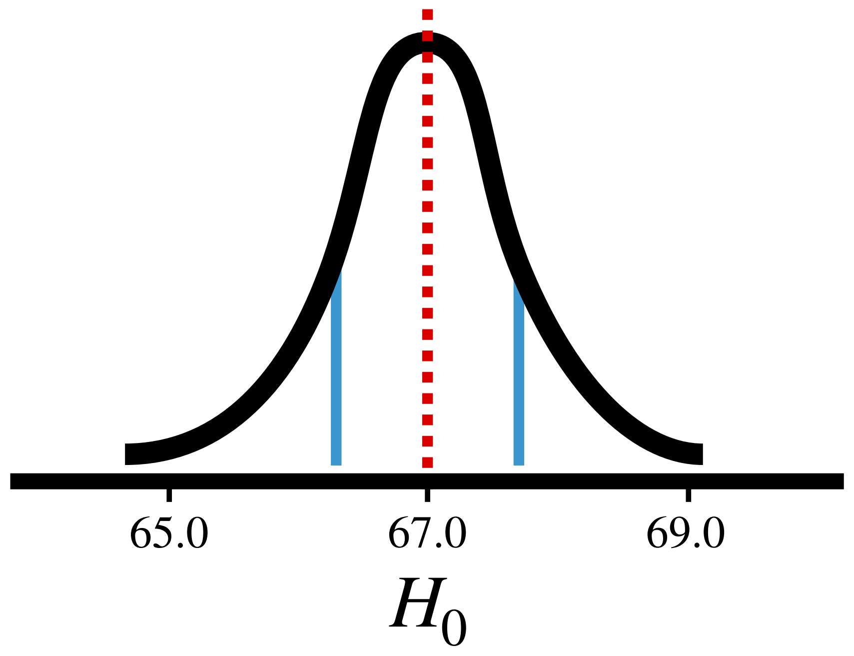
Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

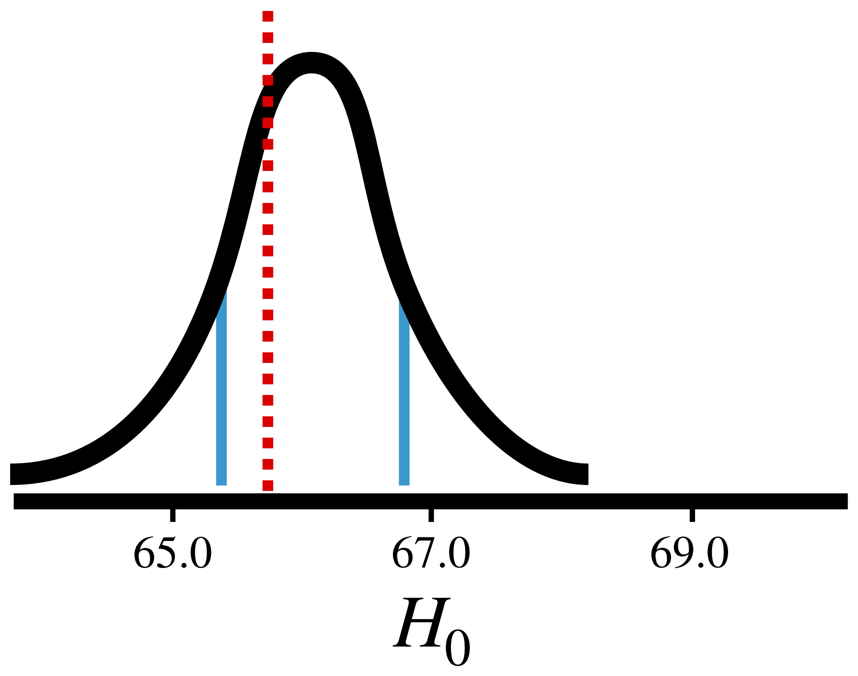


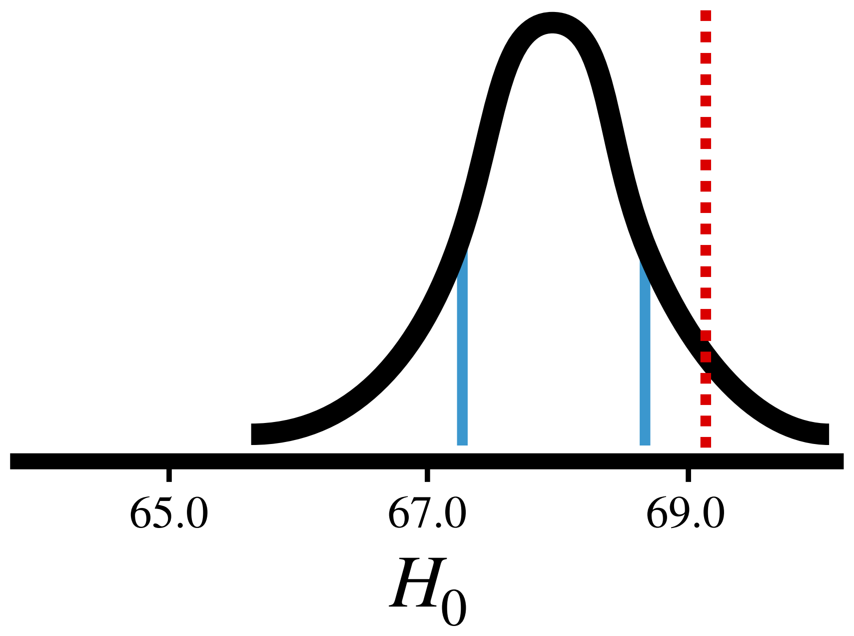


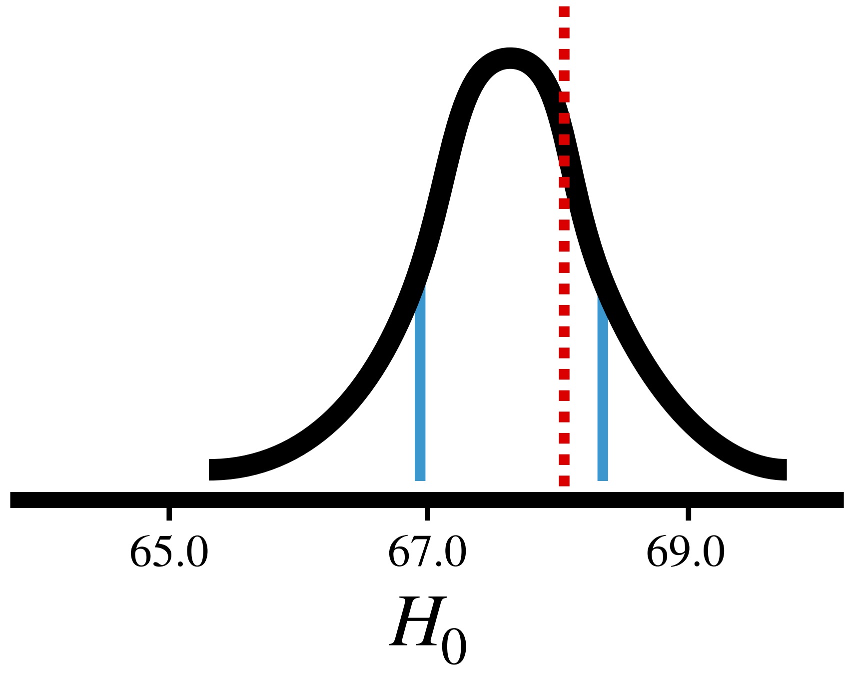


Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

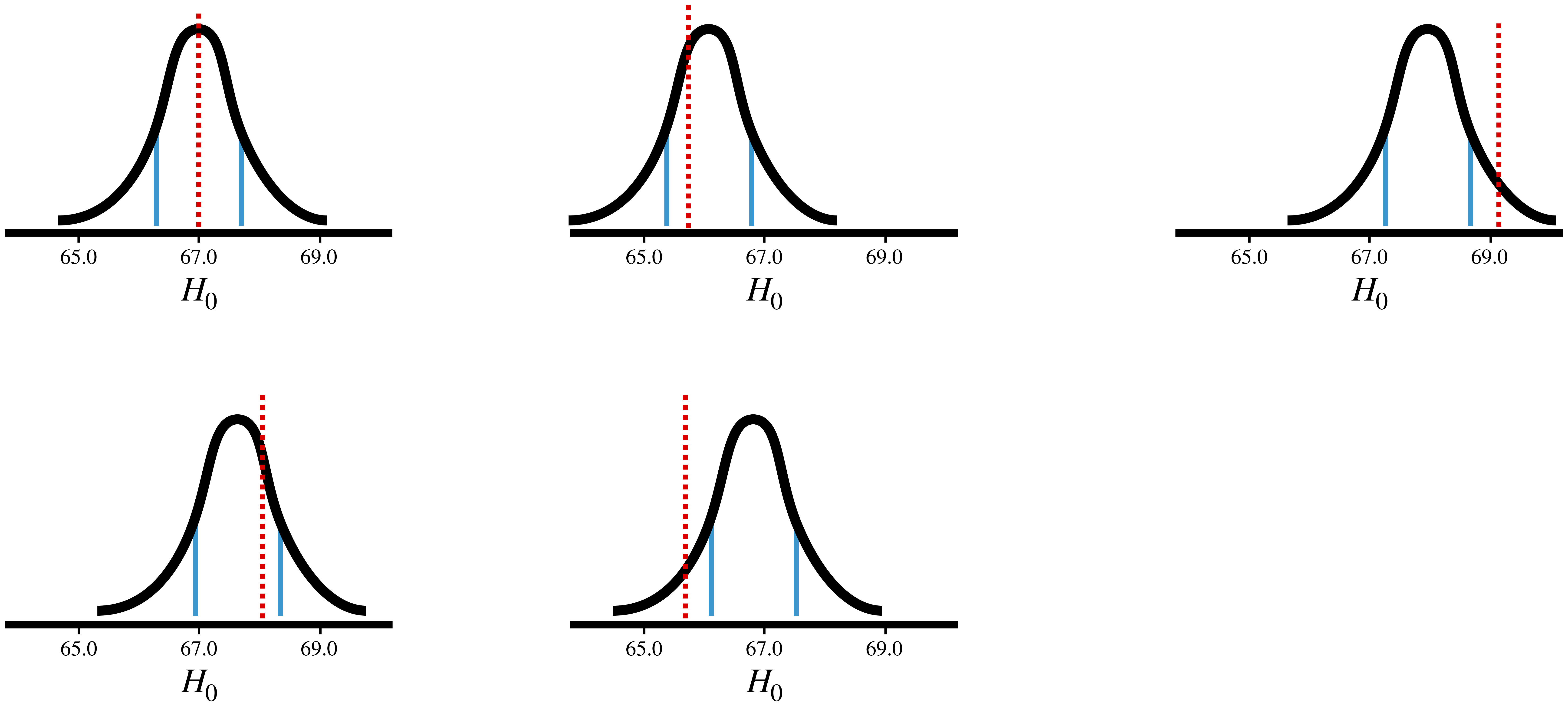




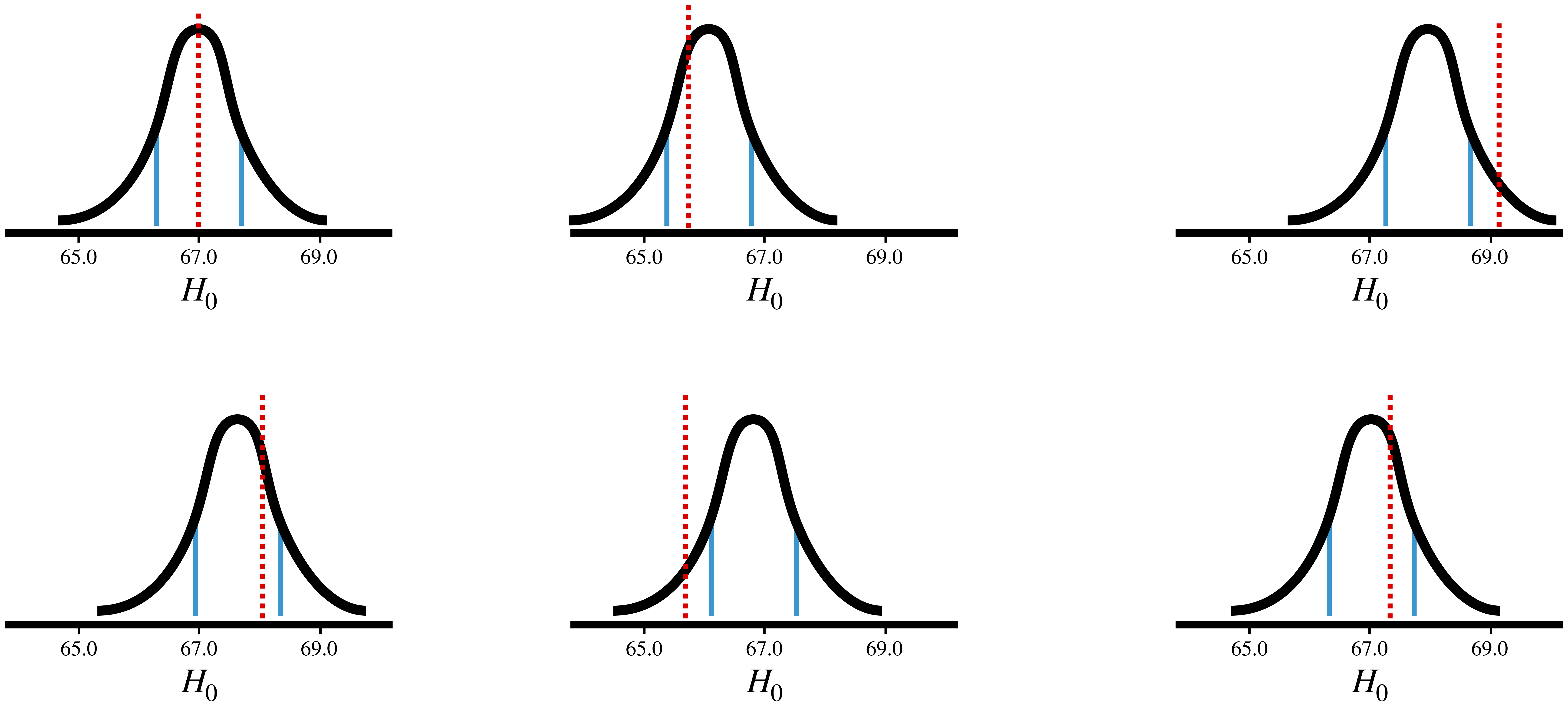




Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

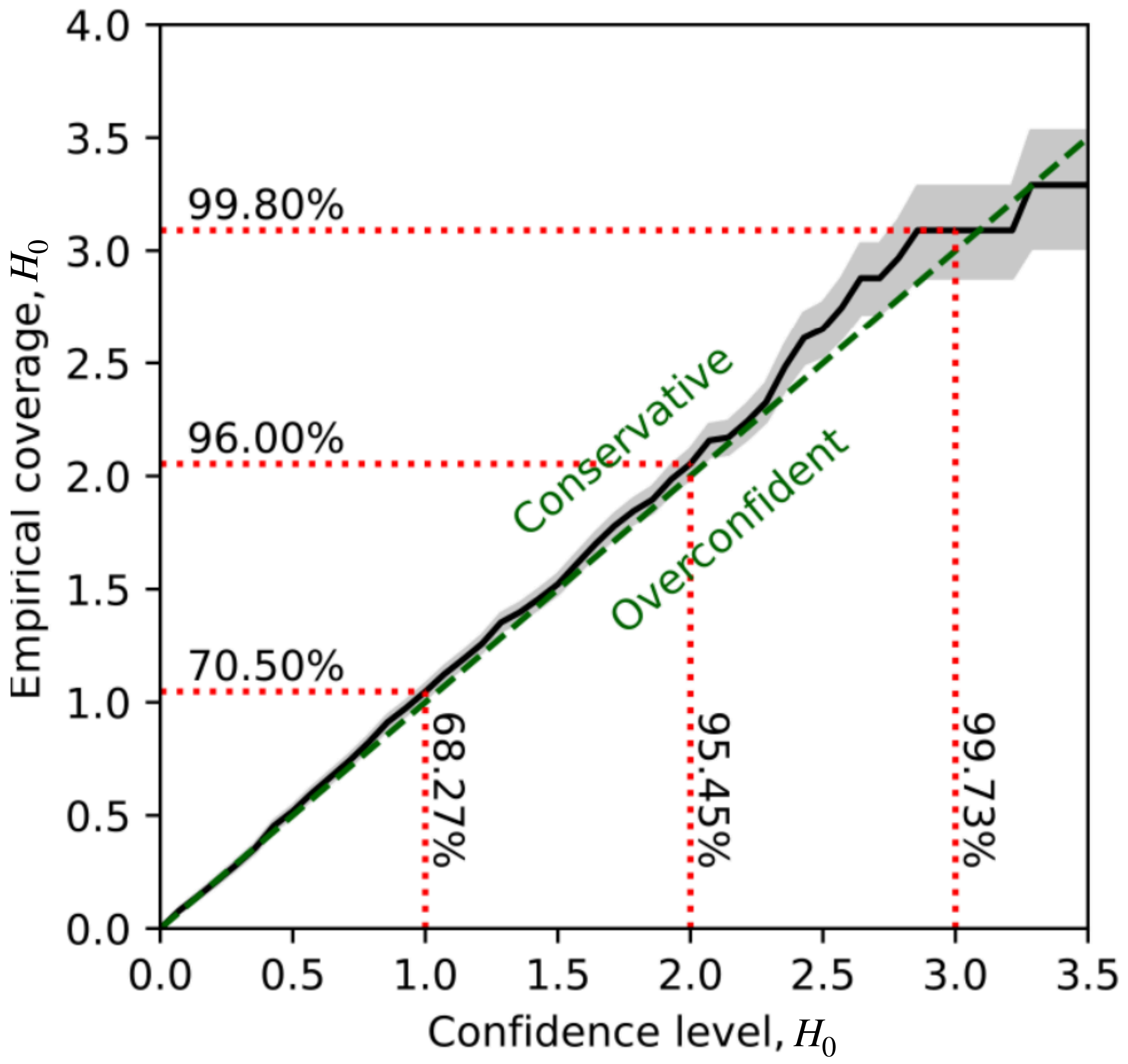


Ex: Is the estimated 68.27% interval covering the ground truth in ~68% of the cases?

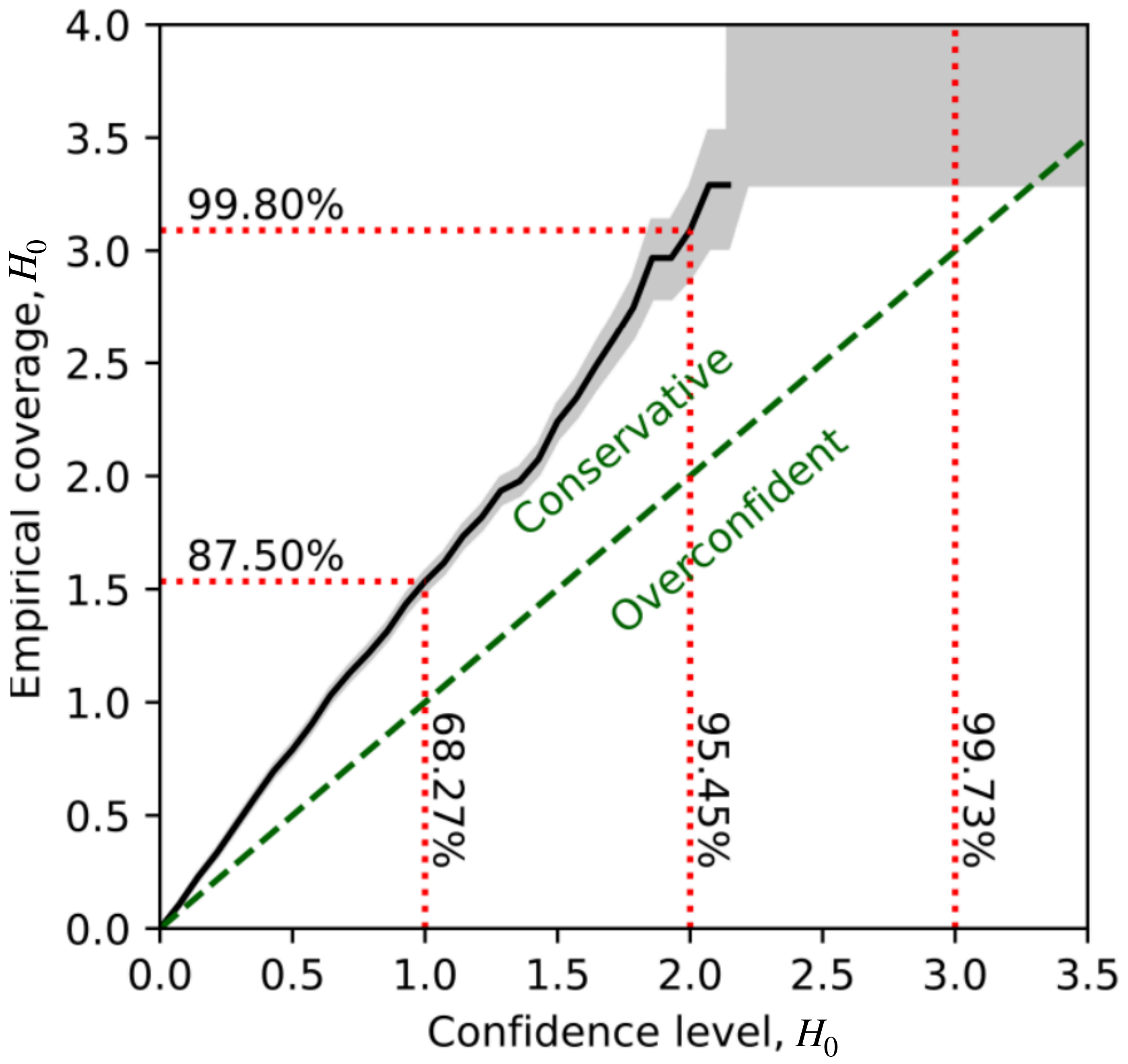


We can empirically estimate the Bayesian coverage

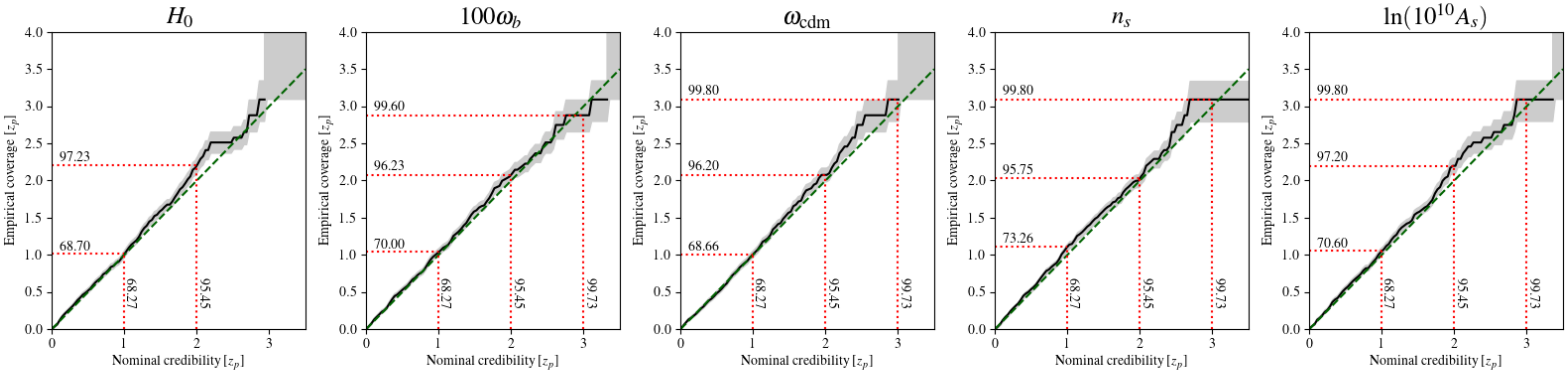
Converged network



Non-converged network

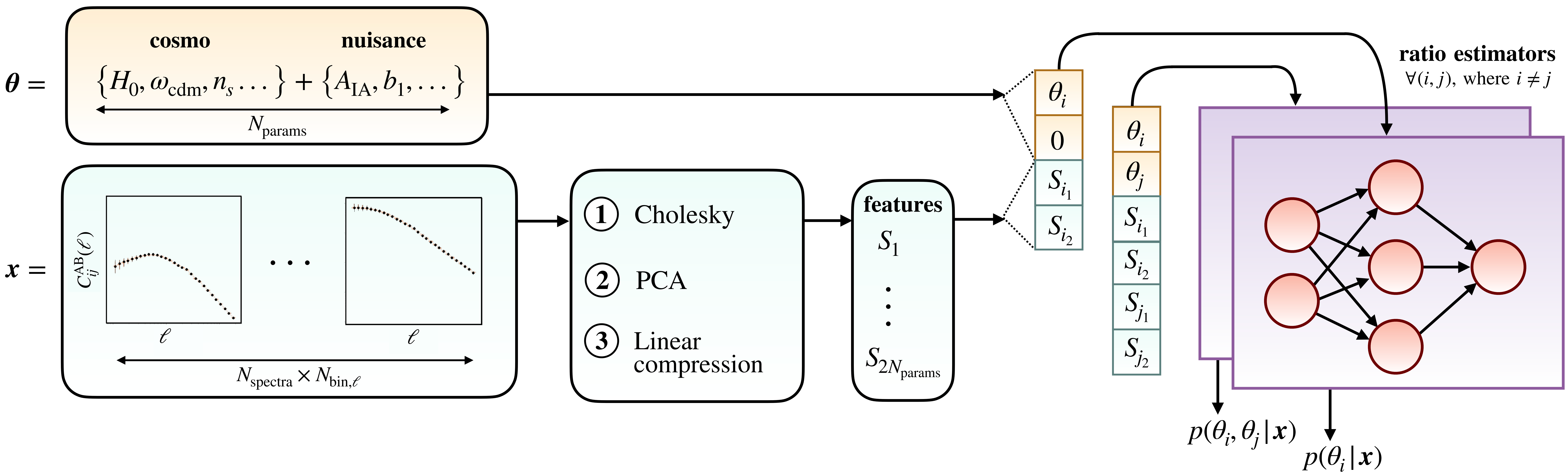


Coverage test for Euclid 3x2pt

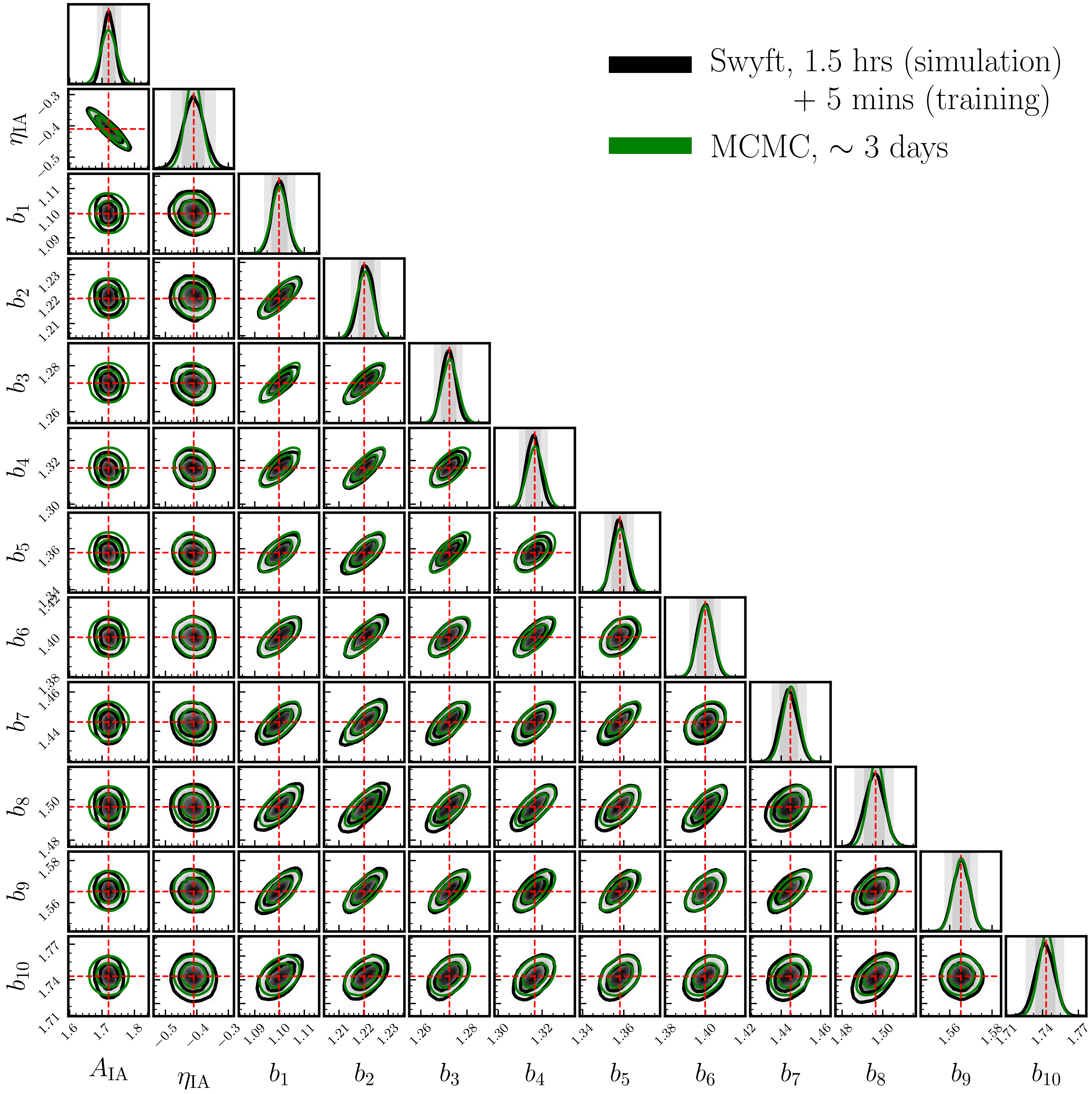


Empirical coverage and confidence level match to excellent precision!

Network architecture



Posteriors for nuisance parameters



MNRE & MCMC are again in good agreement

Another big advantage of MNRE: simulation re-use

It is interesting to see how constraints change with different data combinations

It is interesting to see how constraints change with different data combinations

BUT

with MCMC, one has to **restart chains** for each experimental configuration

It is interesting to see how constraints change with different data combinations

BUT

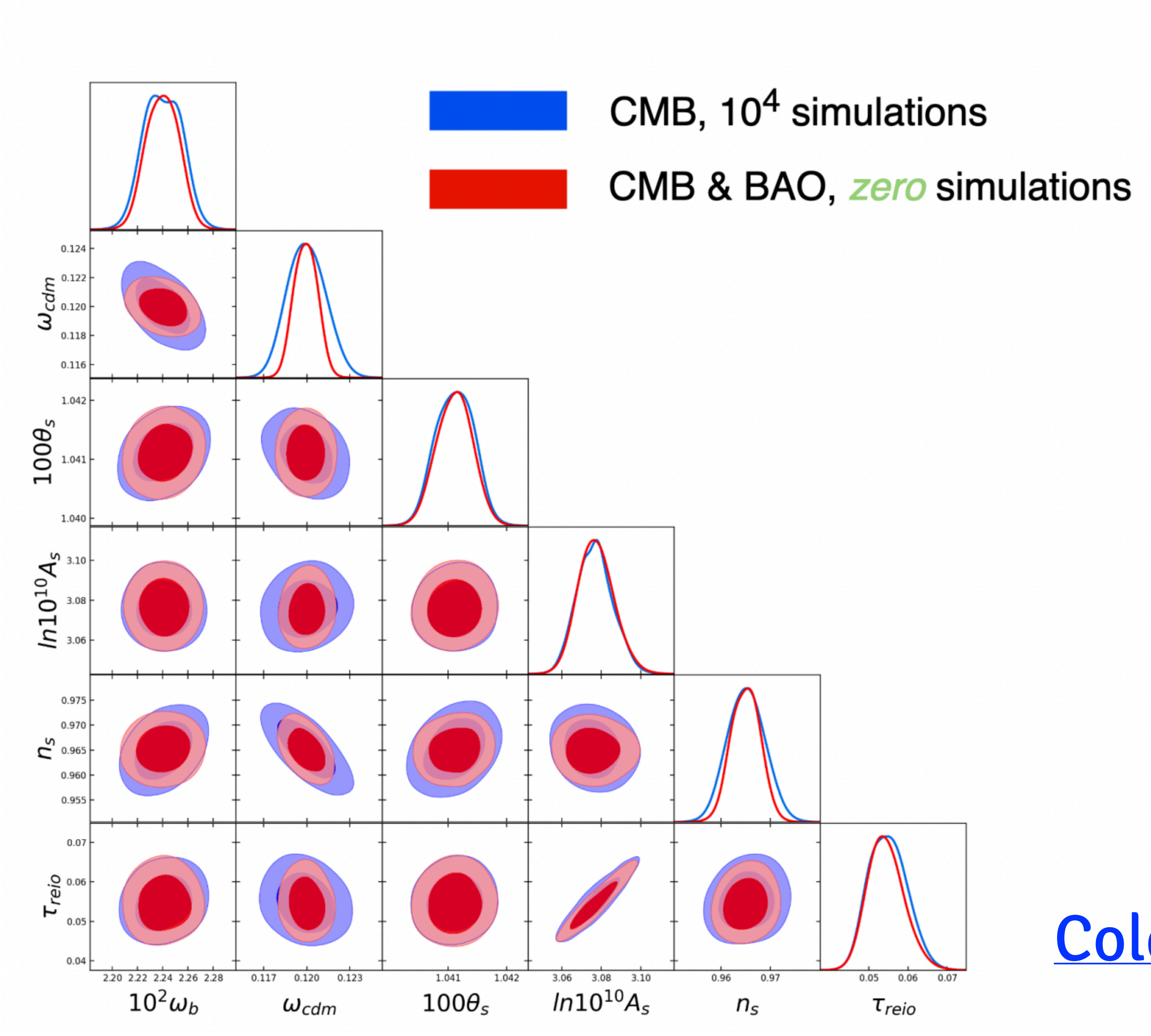
with MCMC, one has to **restart chains** for each experimental configuration

In MNRE it is possible to re-use simulations for different data combinations

The idea is to simulate all the data at once, and then train different inference networks for different data combinations

The idea is to simulate all the data at once, and then train different inference networks for different data combinations

Ex: Planck+BAO



Cole+ 22