Fast test to assess the impact of marginalization in Monte Carlo analyses and its application to Cosmology

based on Phys. Rev. D 106, 063506 (2022) [arXiv:2203.16285]

Adrià Gómez-Valent

Lisbon, 1 June 2023

INFN
Istituto Nazionale di Fisica Nucleare

TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

Outline

I
- Impact of volume effects: marginalized posterior vs profile distribution
- A simple 2D example to illustrate the problem

II
- Profile distributions from the Monte Carlo Markov chains

III
- Application to cosmology:
  - ΛCDM
  - Early Dark Energy
  - Coupled Quintessence
  - Brans-Dicke with Λ

IV
- Conclusions

Why do we need Monte Carlo analyses?

Bayes' theorem

$$P(\theta|D) = \frac{\mathcal{L}(D|\theta)\pi(\theta)}{P(D)}$$

Likelihood

Prior

Posterior

Evidence

Constraints are extracted from the posterior.

Grid to evaluate its shape

$$\#\,\text{evaluations} = \prod_{i=1}^{\dim(\theta)} n_i \sim n^{\dim(\theta)}$$

e.g. CMB analysis with ΛCDM: 10^{20}-10^{30} evaluations → UNFEASIBLE even in relatively small parameter spaces

Monte Carlo sampling

Preliminary output

MC Markov chains

Marginalization method to obtain the posterior in parameter subspaces of the theory/model

$\theta=\{\theta_1,\theta_2\}$

$$\mathcal{P}(\theta_1) = \int \mathcal{P}(\theta_1, \theta_2) d\theta_2$$

Final output: visualization

1D posteriors

2D contour plots

The computation of marginalized posteriors from Markov chains is trivial

Just two steps:

1) Make a grid for the parameter(s) of interest.

2) Count the number of points in the Markov chain that fall inside each bin of the grid and use this information to make a histogram.

For sufficiently long chains, this histogram will be a good approximation of the exact posterior.

But marginalization does not conserve in general the statistical information content

$$\mathcal{P}(\theta_1) = \int \mathcal{P}(\theta_1, \theta_2) d\theta_2$$

e.g. If $P(\theta_1)$ is large at a particular location θ_1^* it can be due to:

→ $P(\theta_1^*, \theta_2)$ is small, but non-null in a big volume of θ_2
→ $P(\theta_1^*, \theta_2)$ is big, but confined in a much smaller volume of θ_2

Take-home message

> There can be points in parameter space that are able to fit very well the data, but have a very small marginalized posterior due to volume effects

Profile Distribution (PD)

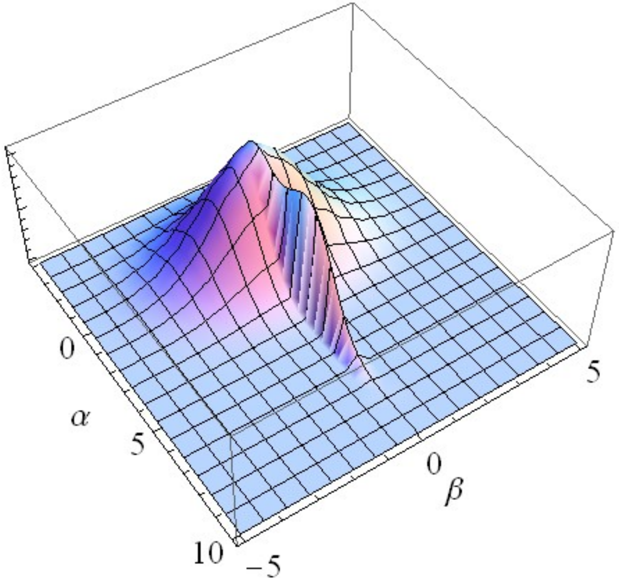
$$\tilde{\mathcal{P}}(\theta_1) = \max_{\theta_2} \mathcal{P}(\theta_1, \theta_2)$$

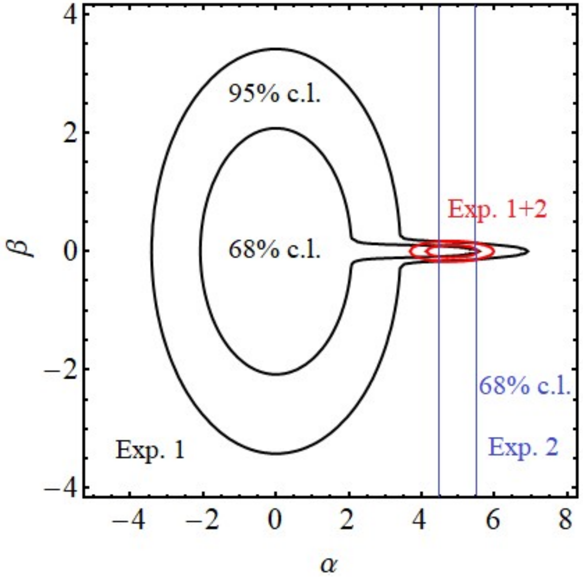
- The maximum of the one-dimensional PDs is located at the value of the parameter that also maximizes the original posterior.
- PDs are not subject to volume effects.
- PDs lead to more robust constraints, especially when the original posterior has important non-Gaussian features.

An example in 2D

Experiment 1 $$\mathcal{P}(\alpha,\beta) = \frac{5}{21\pi}\left[e^{-\frac{1}{4}(\alpha^2+\beta^2)} + e^{-\frac{1}{4}(\alpha-3.5)^2-100\beta^2}\right]$$

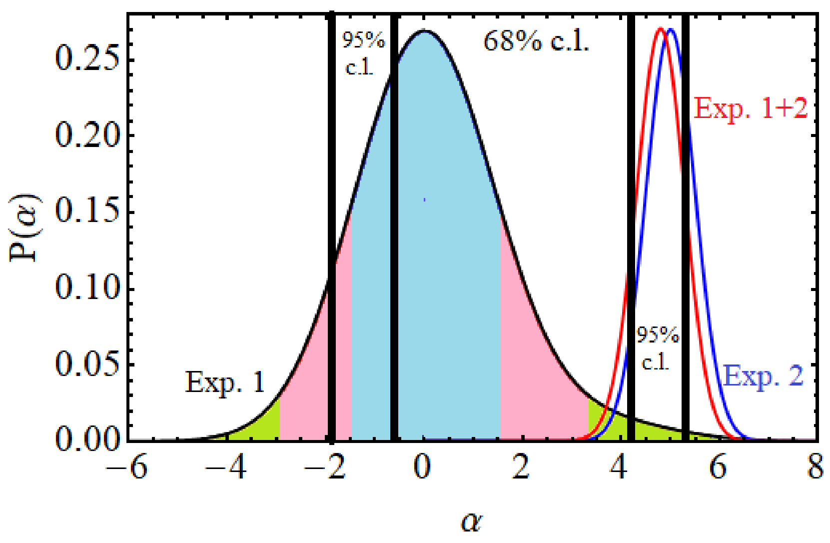
Experiment 2 $\alpha = 5.0 \pm 0.5$

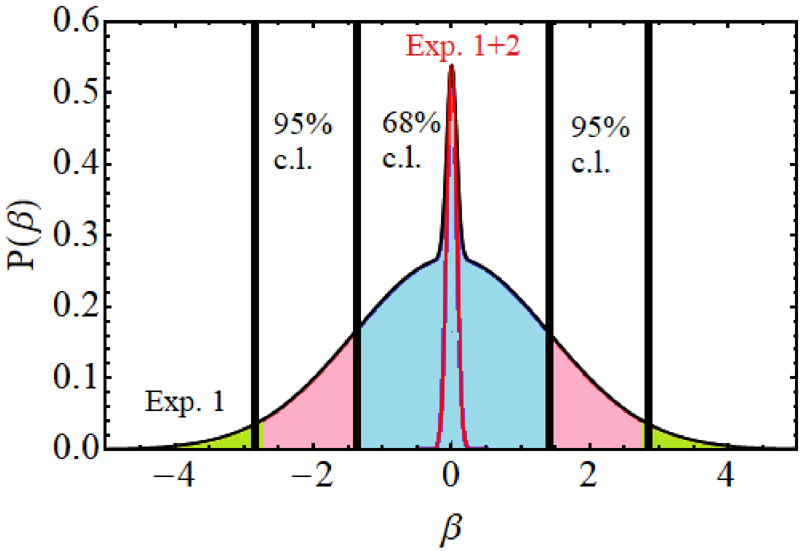




What happens if we study the compatibility of the two experiments using the 1D marginalized posteriors?

$$\mathcal{P}(\alpha) = \frac{5}{21\sqrt{\pi}}\left[2e^{-\frac{\alpha^2}{4}} + 0.1e^{-\frac{1}{4}(\alpha-3.5)^2}\right] \qquad \mathcal{P}(\beta) = \frac{10}{21\sqrt{\pi}}\left[e^{-\frac{\beta^2}{4}} + e^{-100\beta^2}\right]$$





Some remarks:

- In a 2D parameter space it is very easy to study the existence of tensions, since the original posteriors from different experiments can be directly plotted.
- The problem becomes more involved in higher-dimensional parameter spaces.
- PDs are more reliable than the marginalized posteriors to quantify tensions, due to the removal of volume effects
- But the computation of the PDs in high-dimensional spaces is very demanding

To obtain all the one-dimensional PDs of our model we have (in principle) to build a grid and perform a minimization in a parameter space of dimension dim(θ)-1 at each of its knots.

To decide the ranges of the grid for the PD method it is still useful to perform a preliminary MCMC analysis

Can we use the resulting Markov chain to estimate somehow the PDs?

This would allow us to estimate the impact of volume effects and assess the need of a more accurate PD analysis

Estimation of 1D PDs from the Markov chains

1) Build a grid for the parameter of interest.
2) At each knot, search for the maximum value of the posterior and save it (no need of removing the points of the burn-in phase).

$$\tilde{\mathcal{P}}(\theta_1) = \max_{\theta_2} \mathcal{P}(\theta_1, \theta_2)$$

3) Plot the result and compute confidence intervals.

The estimation of the PDs with this method does not require additional computational time

Application to cosmology

Models:

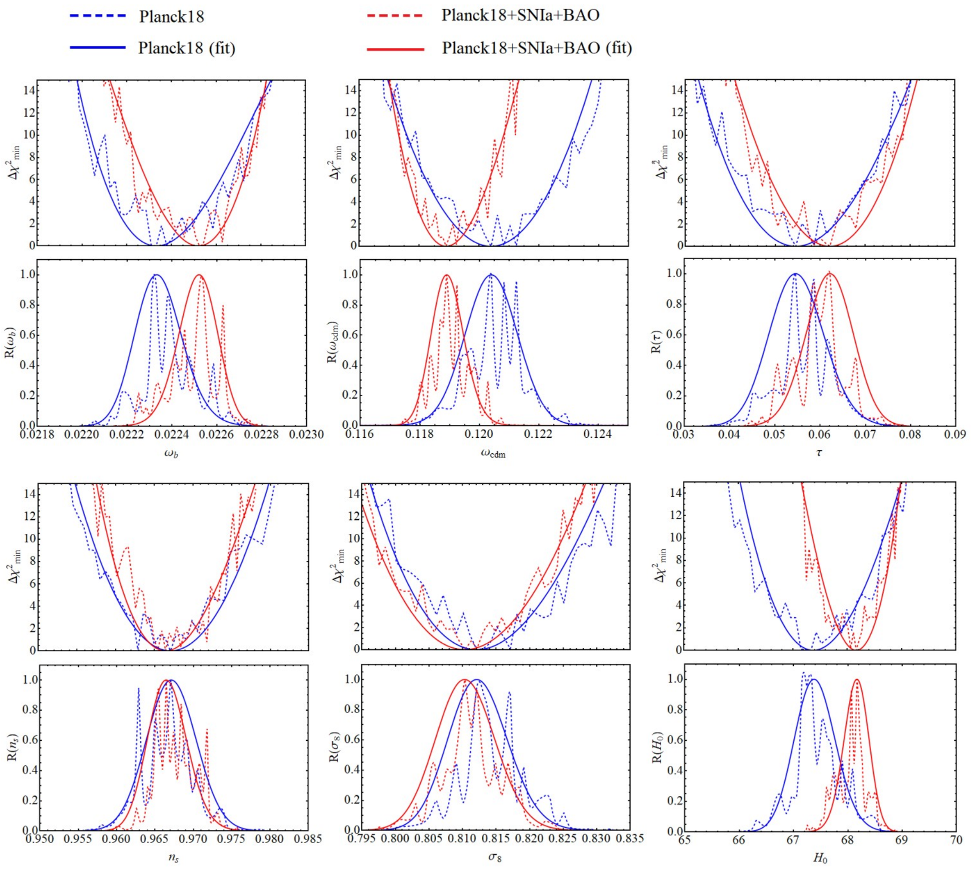
ΛCDM
Ultralight Axionlike (ULA) Early dark energy
Coupled quintessence (with dilatonic conformal coupling)
Brans-Dicke with a cosmological constant

Data:

Planck 2018 TT,TE,EE+lowE+lensing
Pantheon compilation of supernovae of Type Ia
BAO (6dFGS, SDSS MGS, WiggleZ, DES, BOSS, eBOSS)

ΛCDM

Benchmark test: the Planck Collaboration studied the impact of volume effects in this model with the first data release P. A. R. Ade et al., Astron. Astrophys. 566, A54 (2014)

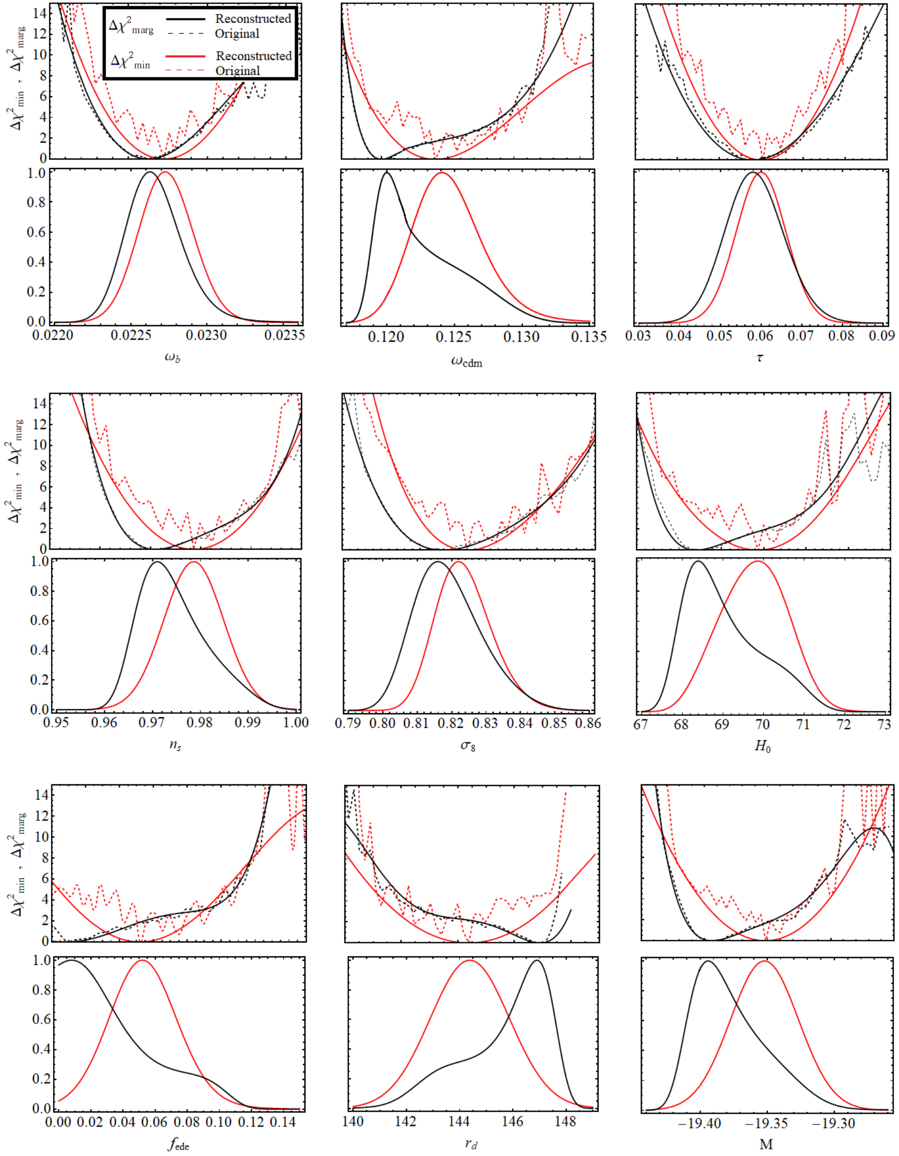


ΛCDM	Planck 2018	
Parameter	Marginalization	PD
ω_b	0.02238 ± 0.00015	$0.02233^{+0.00014}_{-0.00013}$
ω_{cdm}	$0.1204^{+0.0012}_{-0.0013}$	0.1204 ± 0.012
n_s	0.966 ± 0.004	0.967 ± 0.004
τ	$0.055^{+0.007}_{-0.008}$	0.055 ± 0.008
σ_8	0.814 ± 0.006	$0.812^{+0.006}_{-0.007}$
H_0 (km/s/Mpc)	67.4 ± 0.6	67.4 ± 0.6
r_d (Mpc)	146.92 ± 0.27	147.01 ± 0.26
M	...	...
S_8	$0.832^{+0.013}_{-0.014}$	$0.833^{+0.013}_{-0.012}$
σ_{12}	$0.807^{+0.008}_{-0.009}$	$0.808^{+0.008}_{-0.009}$
S_{12}	$0.814^{+0.010}_{-0.011}$	$0.813^{+0.010}_{-0.009}$

ULA

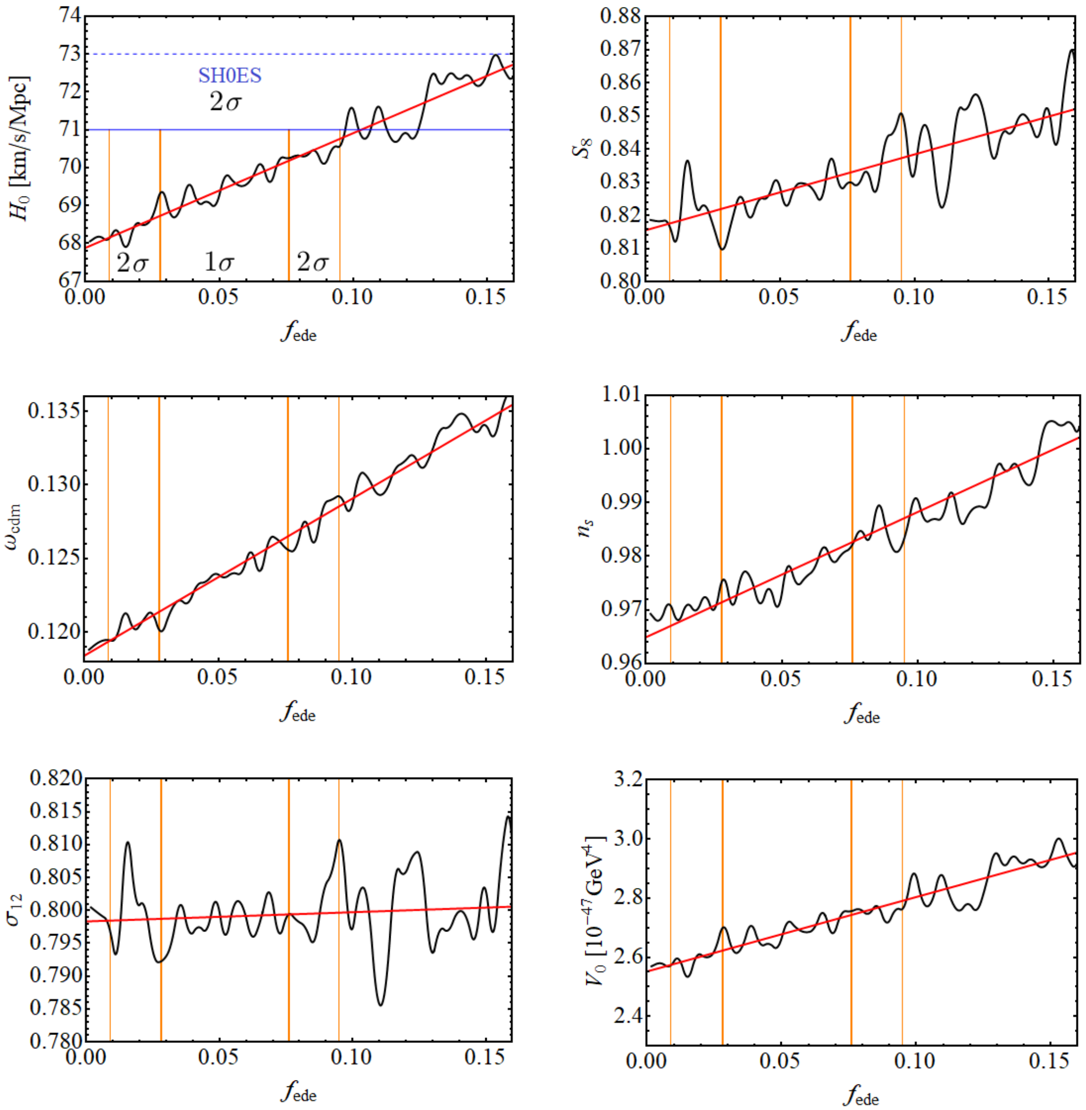
■ Marg.
■ PD

EDE	Planck18 + SNIa + BAO	
Parameter	Marginalization	PD
ω_b	$0.02265^{+0.00019}_{-0.00017}$	$0.02273^{+0.00021}_{-0.00020}$
$\omega_{\rm cdm}$	$0.1203^{+0.0039}_{-0.0016}$	$0.1241^{+0.0027}_{-0.0025}$
n_s	$0.971^{+0.009}_{-0.005}$	0.979 ± 0.008
τ	0.058 ± 0.007	0.060 ± 0.007
σ_8	$0.816^{+0.011}_{-0.009}$	$0.822^{+0.011}_{-0.010}$
H_0 (km/s/Mpc)	$68.4^{+1.3}_{-0.5}$	$69.9^{+0.9}_{-1.0}$
$f_{\rm ede}$	< 0.048	$0.052^{+0.022}_{-0.021}$
$\theta_{\rm ini}$	> 2.2	> 2.3
$z_{\rm max}$	$(3.9^{+4.2}_{-1.0}) \times 10^3$	$(4.2^{+3.0}_{-0.7}) \times 10^3$
$\log_{10}(m/{\rm eV})$	$-26.66^{+0.66}_{-0.54}$	$-27.24^{+0.38}_{-0.49}$
$\log_{10}(f/{\rm eV})$	$26.56^{+0.44}_{-0.33}$	$26.56^{+0.44}_{-0.40}$
r_d (Mpc)	$146.9^{+0.5}_{-2.4}$	144.4 ± 1.5
M	$-19.394^{+0.033}_{-0.018}$	-19.352 ± 0.025
S_8	$0.825^{+0.012}_{-0.011}$	$0.823^{+0.012}_{-0.009}$
σ_{12}	$0.798^{+0.006}_{-0.007}$	$0.797^{+0.006}_{-0.005}$
S_{12}	$0.806^{+0.013}_{-0.009}$	$0.812^{+0.012}_{-0.009}$



See e.g. Poulin et al., Phys. Rev. D 98, 083525 (2018) [arXiv:1806.10608]
Exact PD results of $f_{\rm ede}$ firstly reported by
Herold et al. ApJL 929, L16 (2022), [arXiv:2112.12140]

Correlations with PDs in ULA

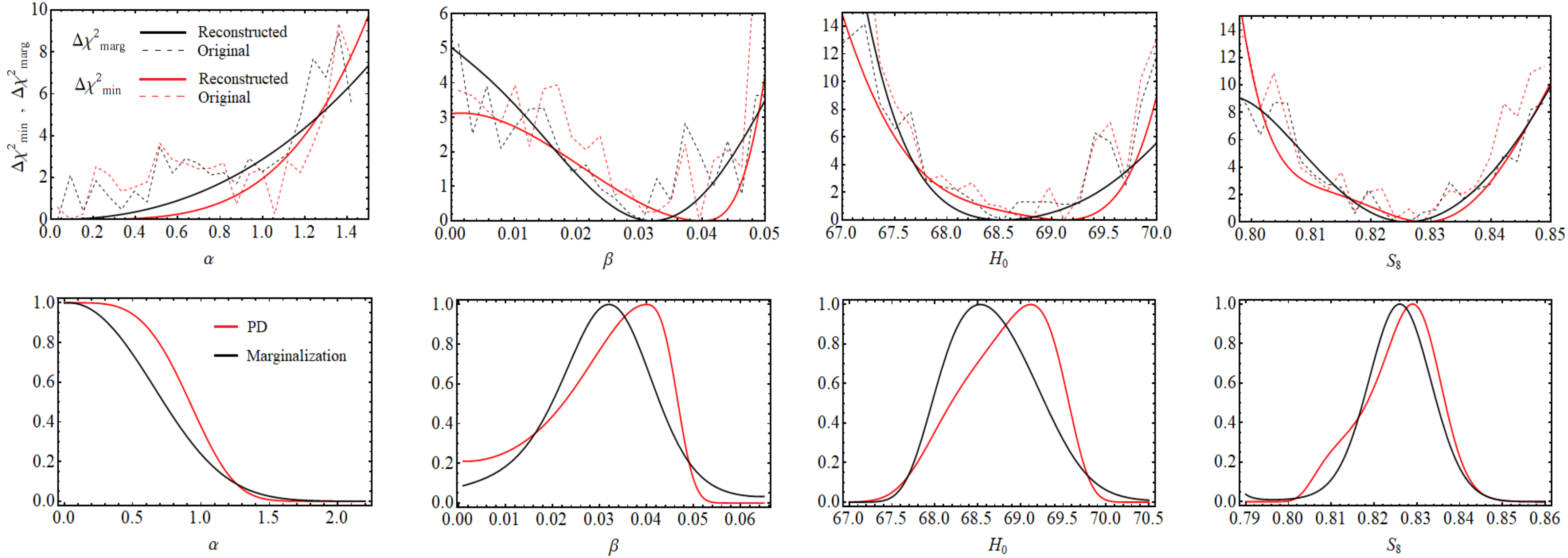


$\sigma_{12} = \sigma(R=12$ Mpc)

A. Sanchez, Phys. Rev. D 102, 123511 (2020) [arXiv:2002.07829]

Coupled quintessence

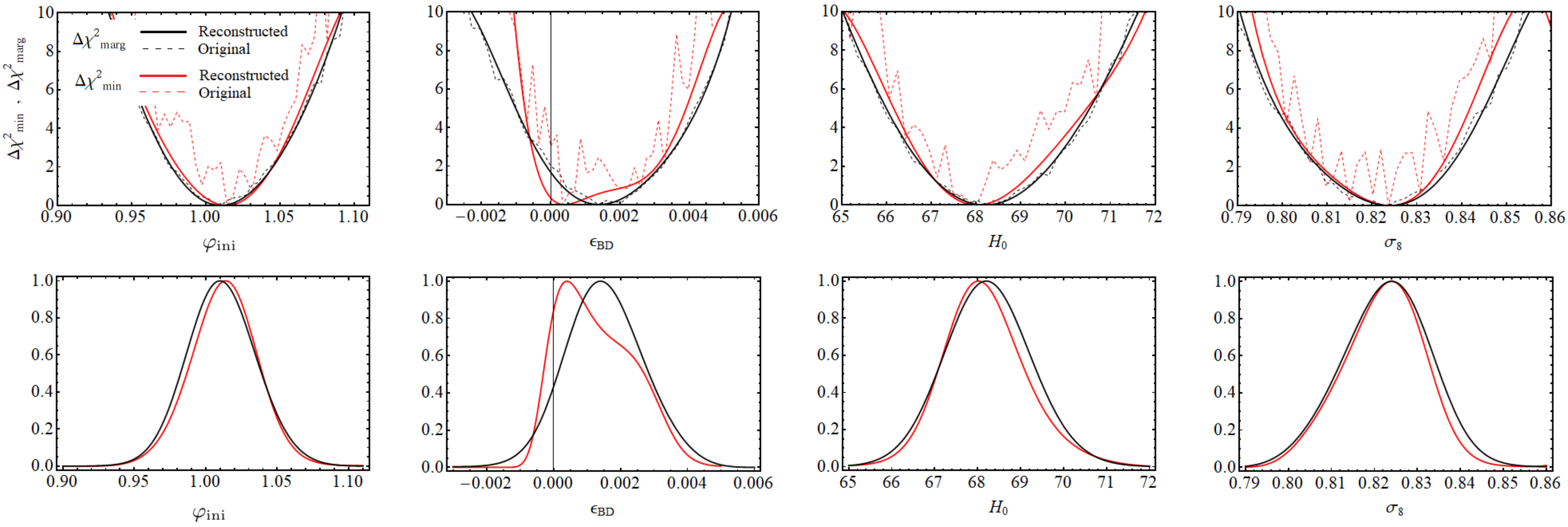
■ Marg.
■ PD



C. Wetterich, Astron. Astrophys. 301, 321 (1995); L. Amendola, Phys. Rev. D 62, 043511 (2000) [arXiv:astro-ph/9908023]

For recent pheno. Analyses: C. van der Bruck et al. Phys. Rev. D 95, 043513 (2017) [arXiv:1609.09855]
Gómez-Valent et al. Phys Rev. 101, 123513(2020) [arXiv:2004.00610]; 106, 103522 (2022) [arXiv:2207.14487]
Goh et al. Phys. Rev. D 107, 083503 (2023) [arXiv:2211.13588]

Brans-Dicke with Λ



See e.g. A. Avilez and C. Skordis, Phys. Rev. Lett. 113, 011101 (2014) [arXiv:1303.4330]
J. Solà Peracaula et al., ApJL 886, L6 (2019) [arXiv:1909.02554]; CQG 37, 245003 (2020) [arXiv:2006.04273]

Conclusions

- The marginalization of posterior distributions with important non-Gaussian features can introduce biases in our constraints.
- These effects cannot be “visualized” in parameter spaces with dimension larger than 2.
- The profile distribution allows us to obtain more robust constraints, which do not suffer from marginalization issues, but their computation is very time-consuming.
- I have proposed a fast method to estimate the PDs directly from the Markov chains, with no additional computational cost.
- This efficient method allows us to obtain not only one PD, but the whole set of PDs for all the main cosmological, nuisance and derived parameters.
- We have applied it to four interesting cosmological models. We have been able to: (1) recover the expected results for the ΛCDM; (2) find the important volume effects in ULA for f_{ede}; (3) obtain the complete set of PD constraints for ULA, CQ and BD-ΛCDM for the first time in the literature.

Other recent works about PDs:

- A. Reeves et al., "Restoring cosmological concordance with early dark energy and massive neutrinos?", MNRAS 520, 3688 (2023) [arXiv:2207.01501]
- Holm et al., "Decaying dark matter with profile likelihoods", Phys. Rev. D 107, L021303 (2023) [arXiv:2211.01935]
- J.S. Cruz et al., "Profiling Cold New Early Dark Energy", [arXiv:2302.07934]

Using the PDs from MCMC:

- G. Galloni et al. "Updated constraints on amplitude and tilt of the tensor primordial spectrum", JCAP 04, 062 (2023) [arXiv:2208.00188]