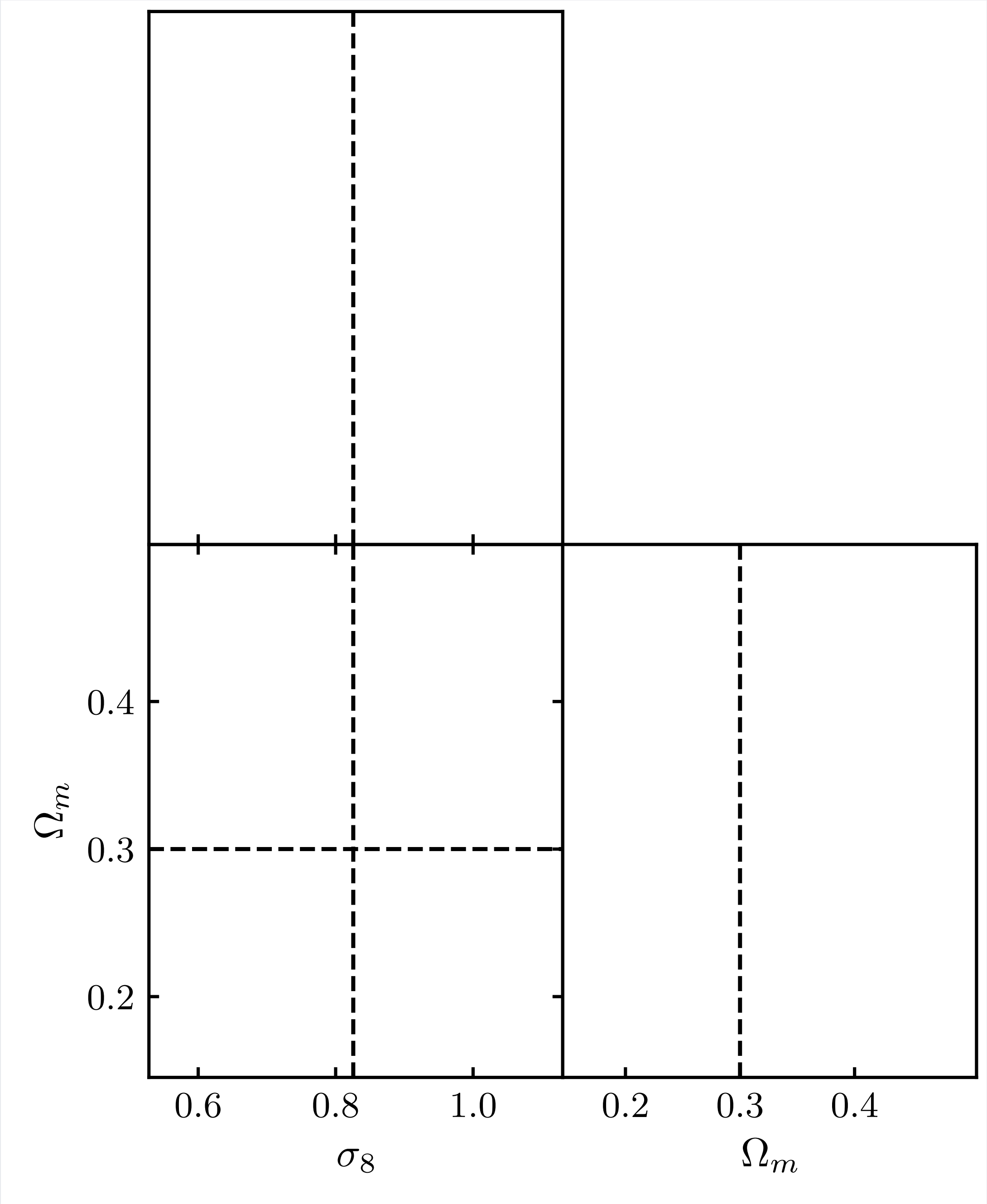
Understanding Posterior Projection Effects With Normalizing Flows

Marco Raveri

University of Genova

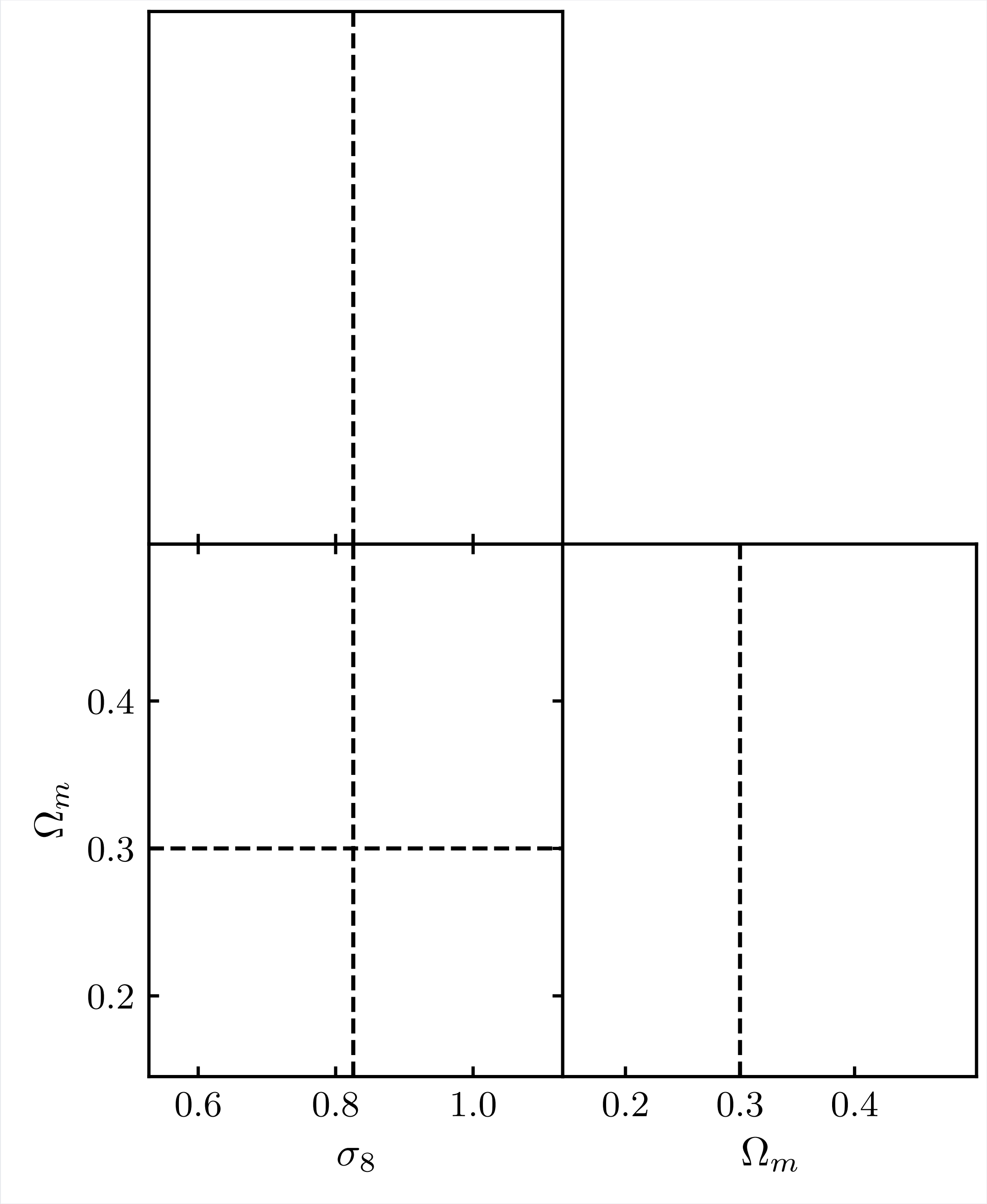
What's the problem?

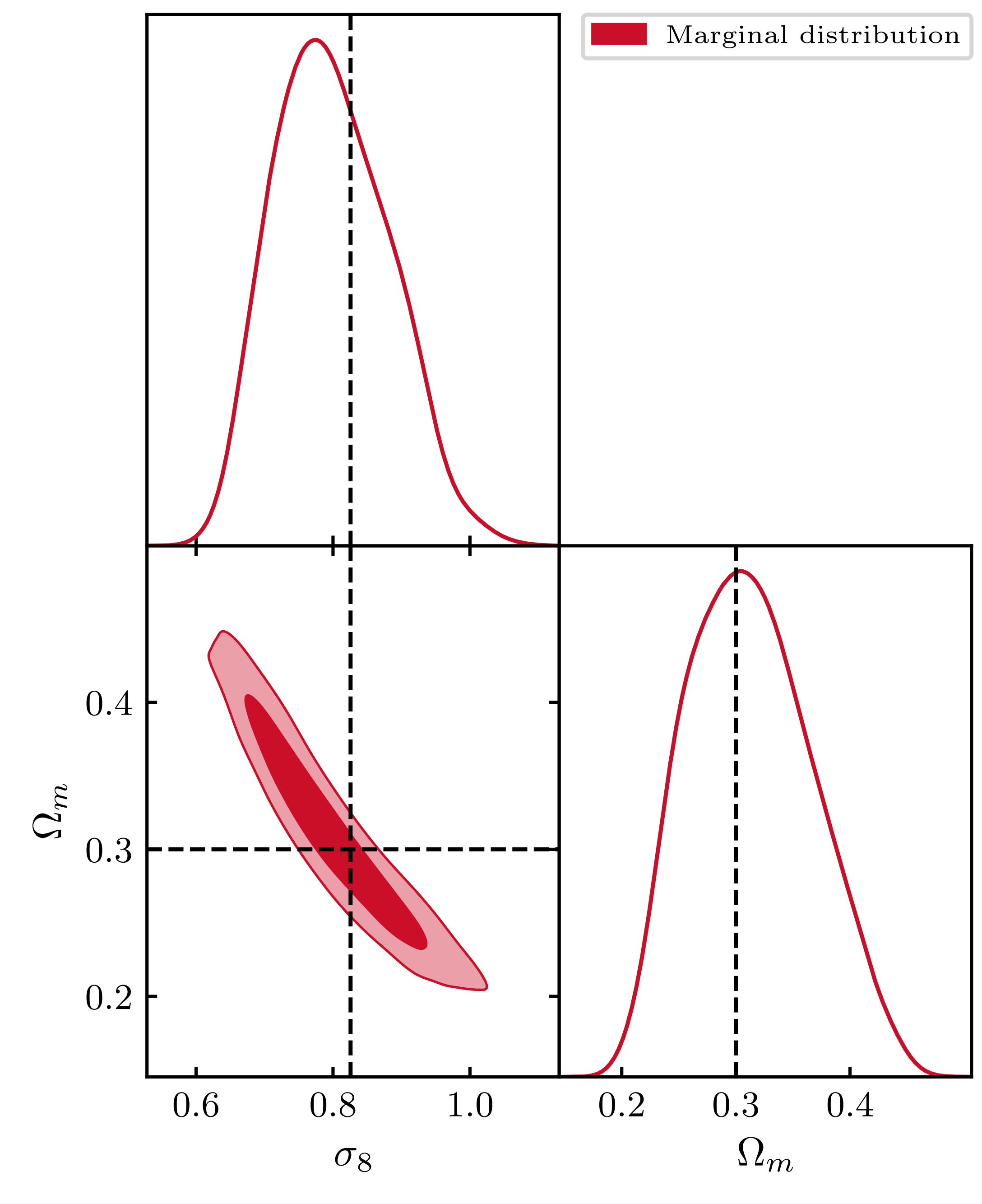
Take your favorite experiment - DES lensing 19D



What's the problem?

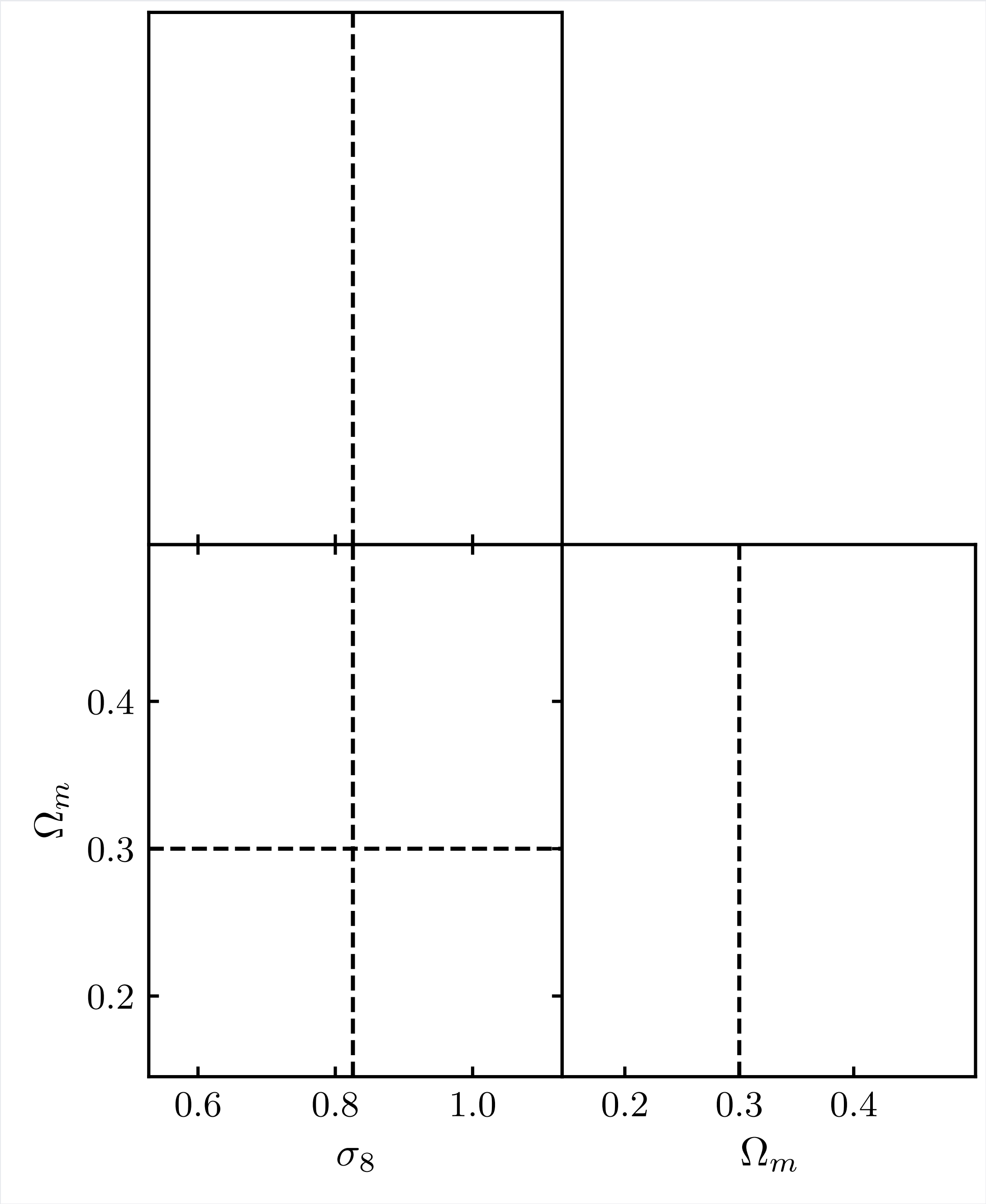
Take your favorite experiment - DES lensing 19D

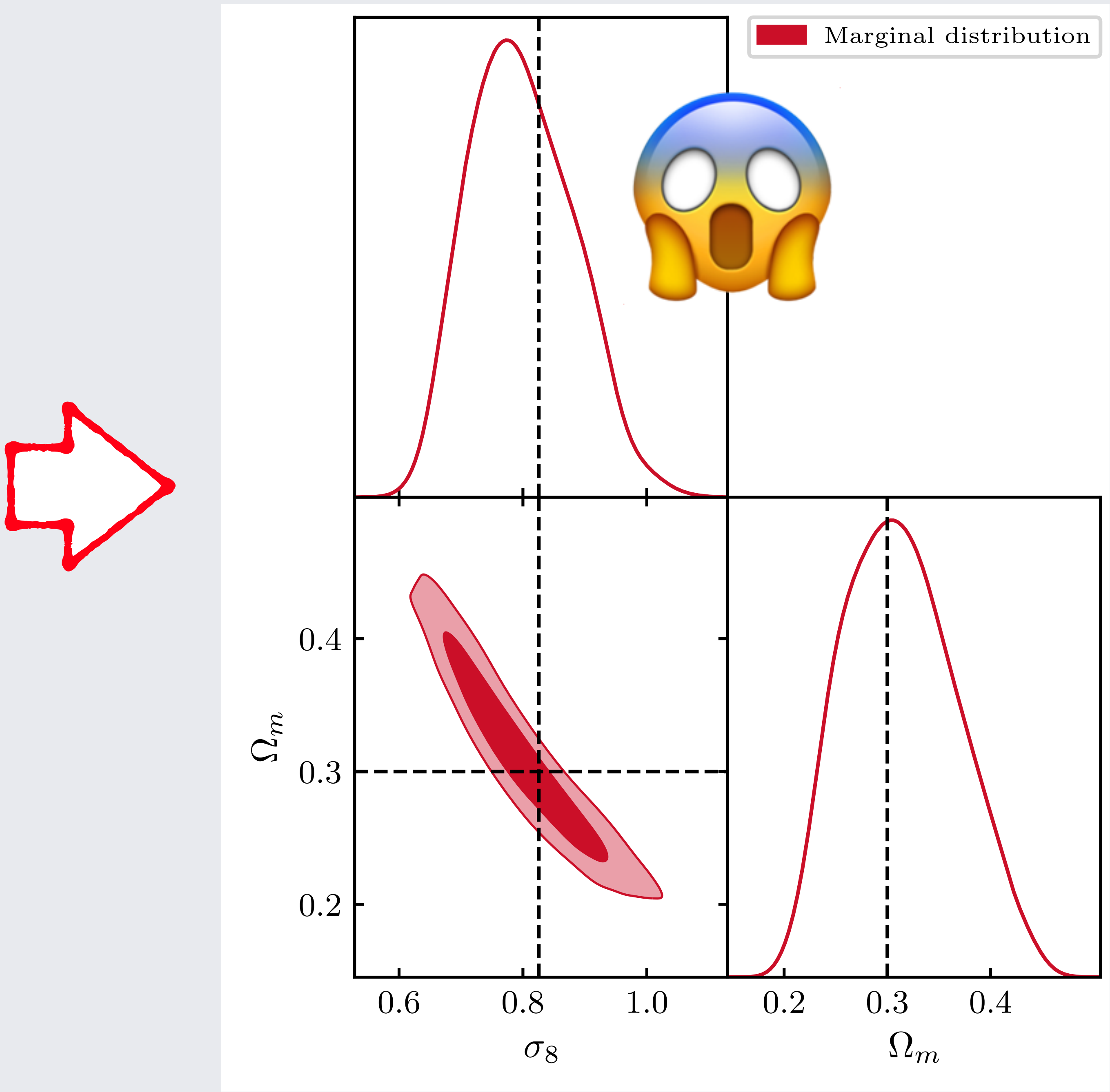




What's the problem?

Take your favorite experiment - DES lensing 19D





Marginalization over parameters

Marginalization

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

Integrate out parameters that we are not looking at

This usually gives a puzzling picture of the distribution

Profiling over parameters

Profiling

$$\hat{\mathcal{P}}(\theta_1|x) \equiv \max_{\theta_2} \mathcal{P}(\theta_1, \theta_2|x) \, .$$

Maximize over parameters that we are not looking at

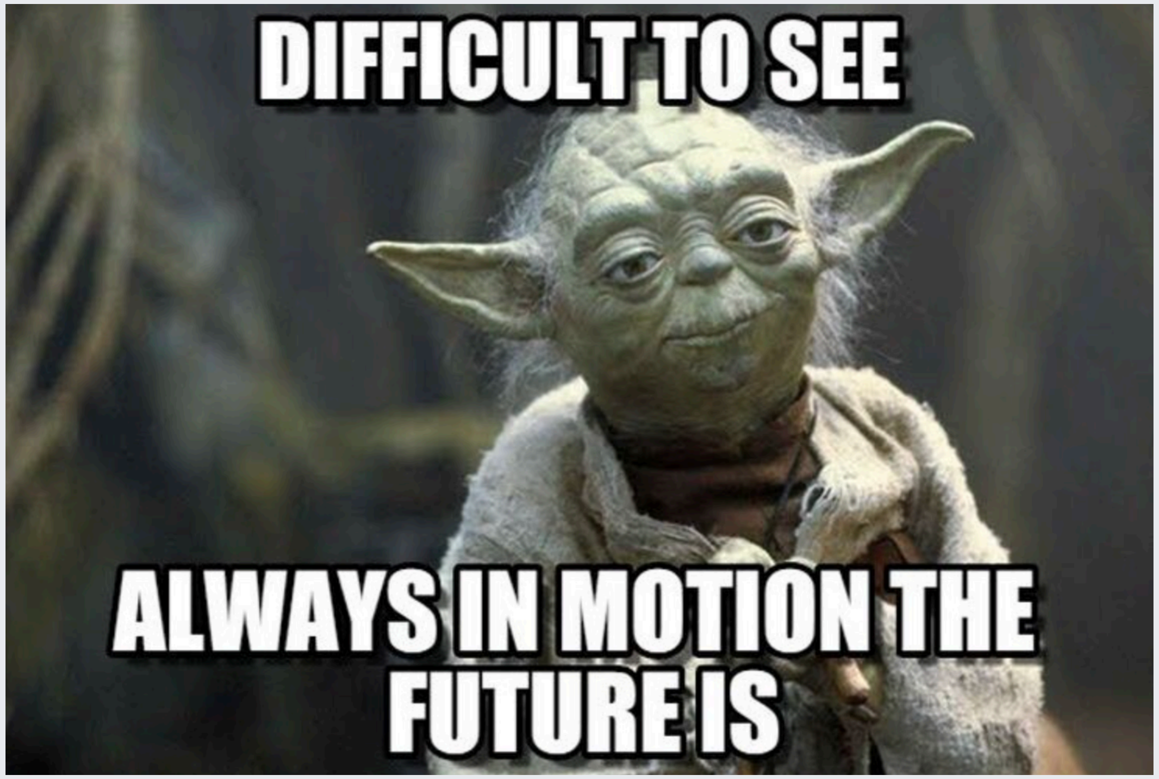
Less puzzling but statistical interpretation harder

Bayesian/Frequentist disclaimer

Approaches diverge when interpreting what guarantees these distributions give about the **future**

Bayesian/Frequentist disclaimer

Approaches diverge when interpreting what guarantees these distributions give about the **future**



We fail to understand the structure of the distribution of the data we have…

Two pictures of the same distribution

Marginal

↓

Mass along the line of sight

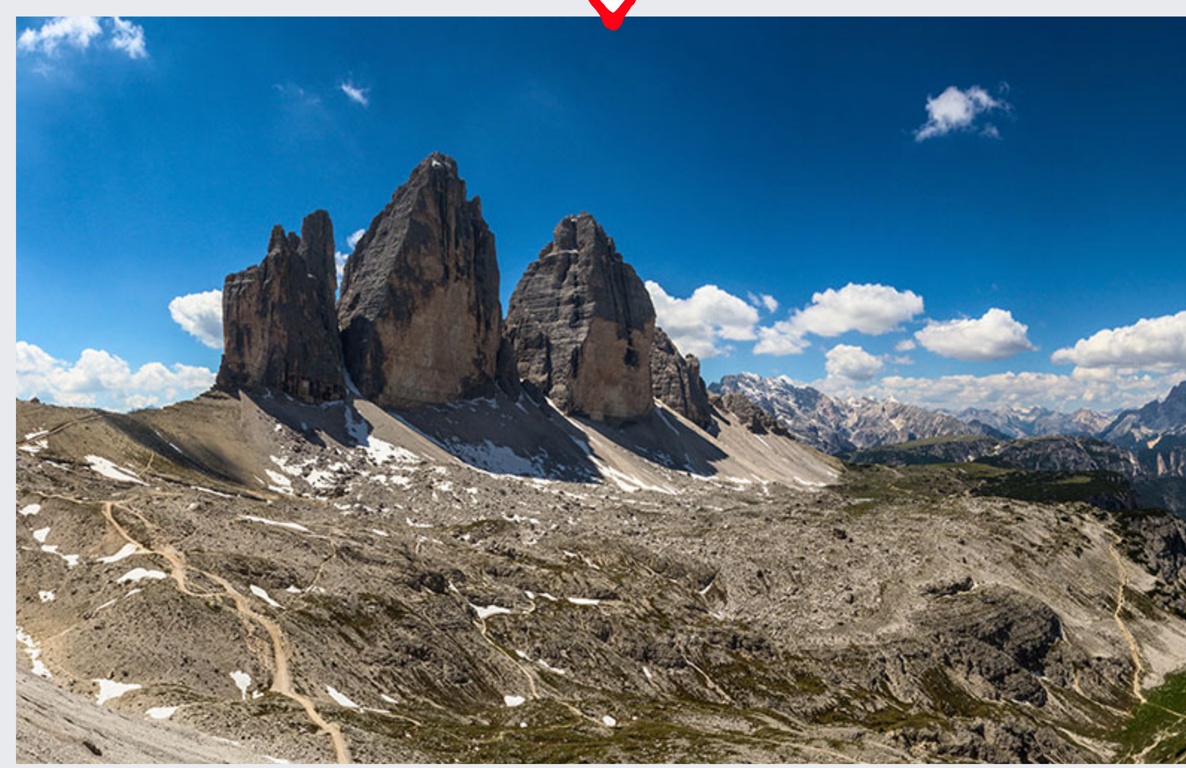
↓

Profile

↓

Outline of the distribution

↓



Relationship between the two

Take the difference between the two

$$\log P(\theta_1) = \log \hat{P}(\theta_1) + \log \int \frac{P(\theta_1, \theta_2)}{\hat{P}(\theta_1)} \, d\theta_2$$

Marginal Profile

Assume Gaussianity in the marginalized direction

$$\log P(\theta_1) - \log \hat{P}(\theta_1) \leq -\frac{1}{2} \log \det F^{(2)}_{\rm data}(\theta_1) + \log V_2(\theta_1)$$

Relationship between the two

Take the difference between the two

$$\log P(\theta_1) = \log \hat{P}(\theta_1) + \log \int \frac{P(\theta_1, \theta_2)}{\hat{P}(\theta_1)} d\theta_2$$

Marginal　　Profile

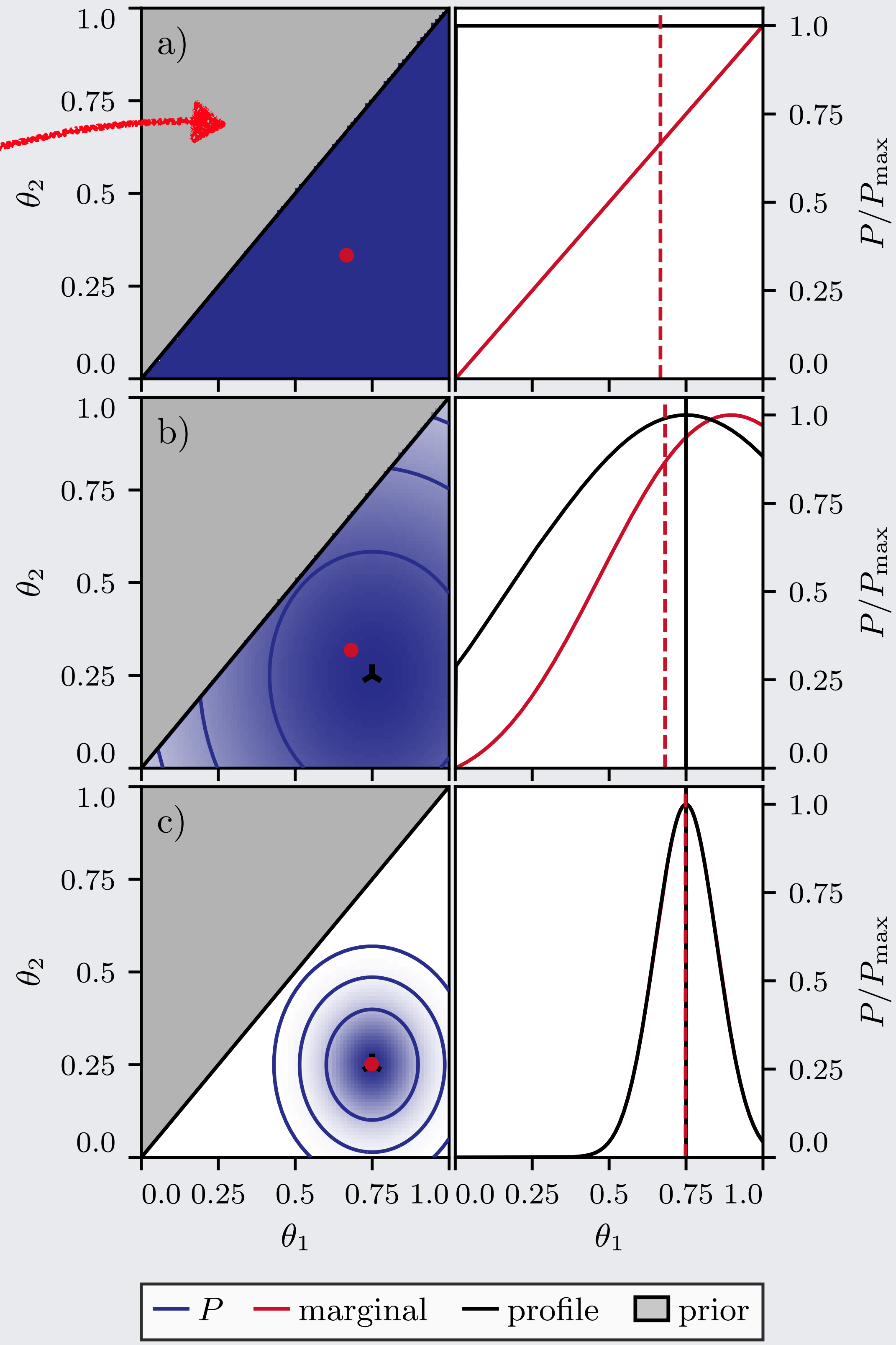
Assume Gaussianity in the marginalized direction

$$\log P(\theta_1) - \log \hat{P}(\theta_1) \leq -\frac{1}{2} \log \det F^{(2)}_{\rm data}(\theta_1) + \log V_2(\theta_1)$$

Projection effects　　**Volume effects**

Example of a volume effect

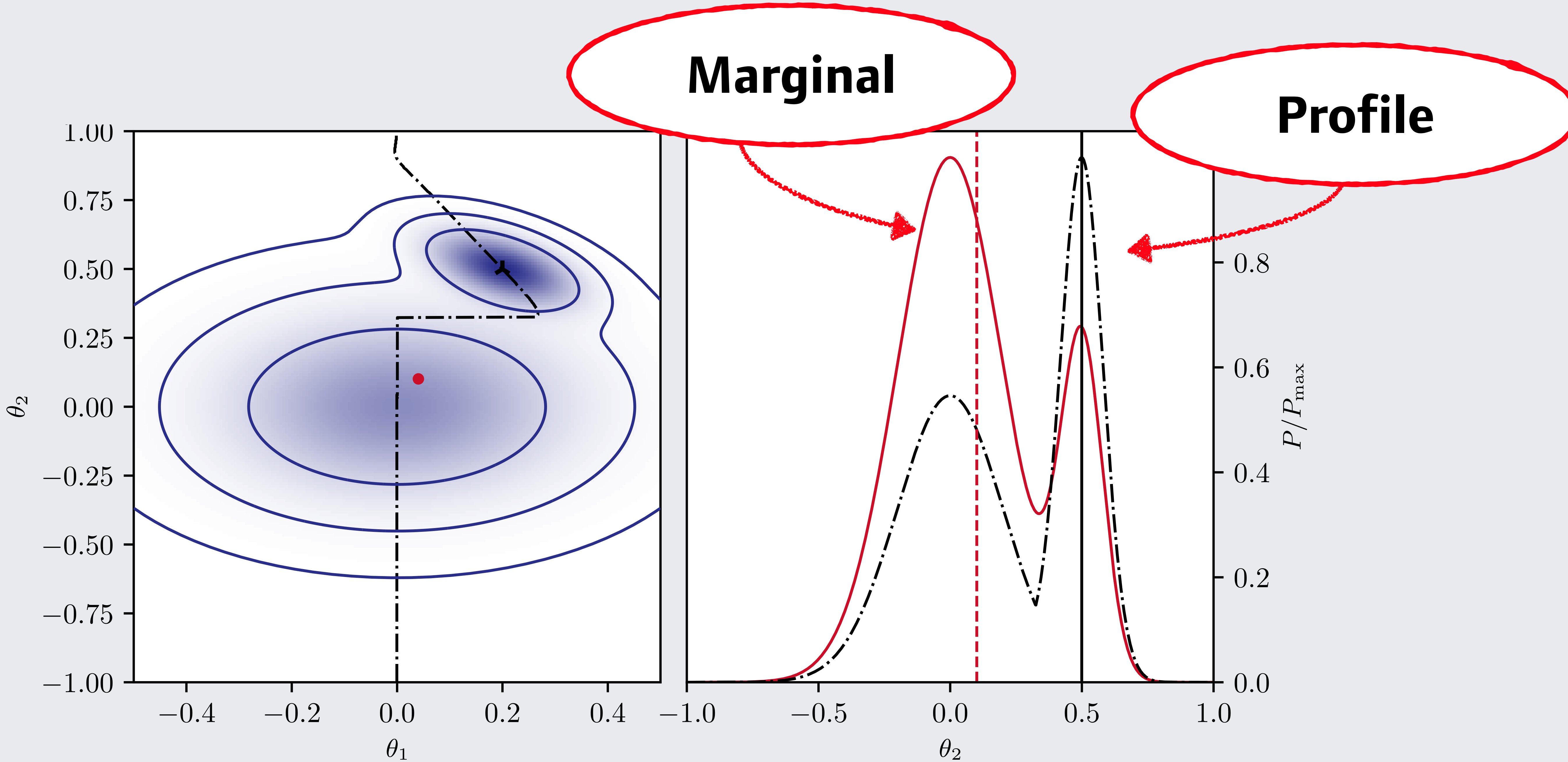
Prior



Increasing data power

Better localized

Example of a projection effect



Local Fisher matrix changes abruptly

Few lessons

Profiles are usually more conservative when we have flat portions of parameter space

Discontinuities are related to multimodality

Profiles are good as they preserve “height”
(i.e. the top of a marginal is the top of the full distribution)

Marginals are good as they preserve probabilities
(i.e. a marginal distribution is a distribution)

When the two differ: don’t trust what you see!

Practical Problems and Solution

Problem:

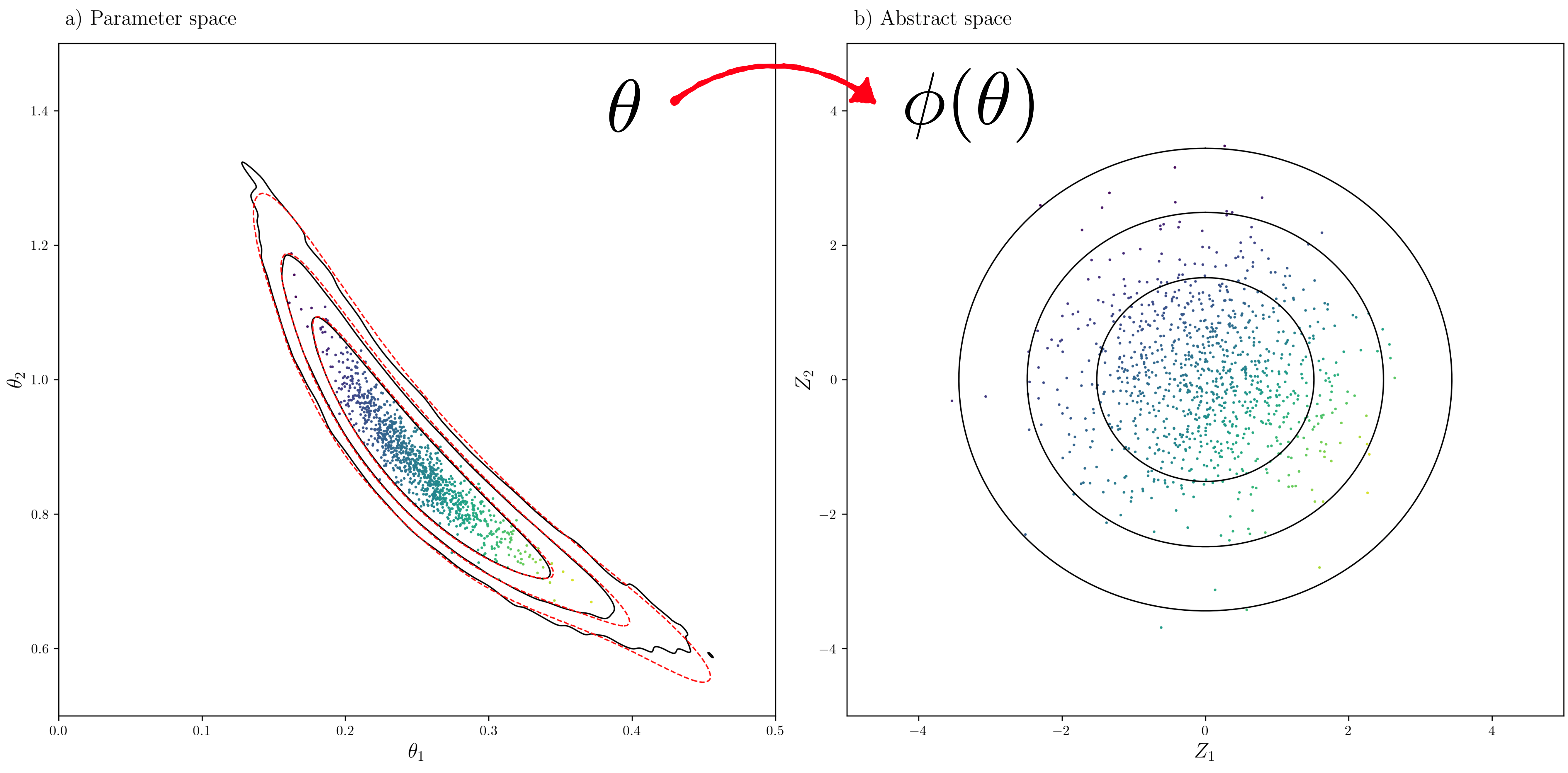
Marginals are easy to get, **profiles require (lots of) high dimensional maximizations**

High dimensional -> need Jacobian
Lots of -> need fast evaluation of posterior

Solution:

Normalizing flow models of posterior distributions

Normalizing Flow Models



Learn the distribution as a mapping to a Gaussian

Lots of engineering…

Code implementation available

```
~ pip install tensiometer
```

New version out today!

Cyrille Doux
(CNRS)

Shivam Pandey
(Columbia U)

MR, Cyrille Doux and Shivam Pandey “Understanding posterior projection effects with normalizing flows” to appear on the arxiv soon

Tensiometer

Industry-standard for tension calculations

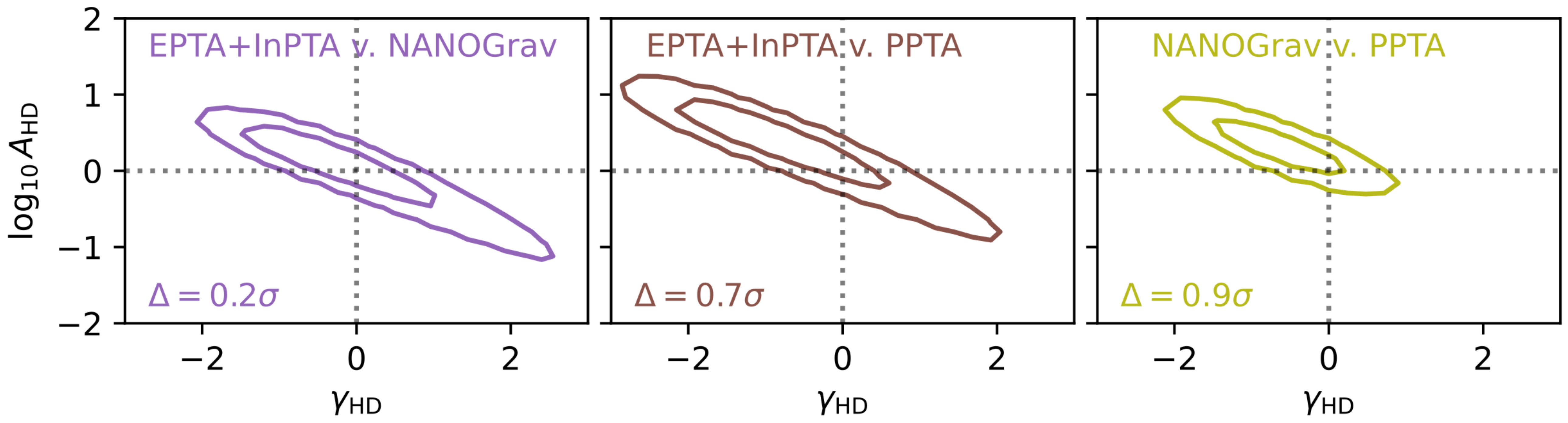
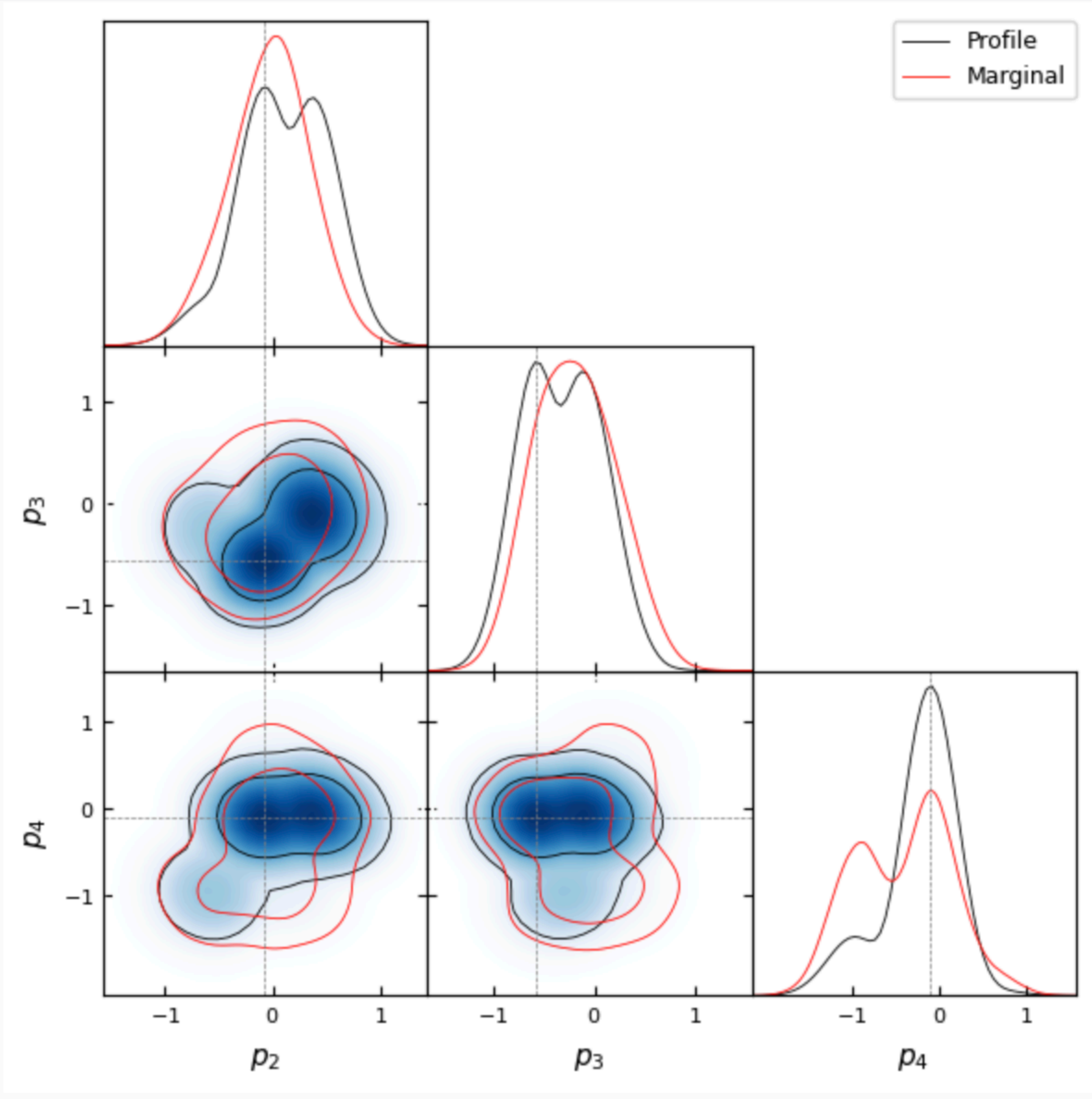


Figure 2. Difference distributions for GWB parameters between pairs of PTAs as computed by `tensiometer`. The contours show 68 and 95% of the distribution mass.

Adopted by PTA collaborations - NOT including MR…

From arXiv:2309.00693 "Comparing recent PTA results on the nanohertz stochastic gravitational wave background"

Toy example

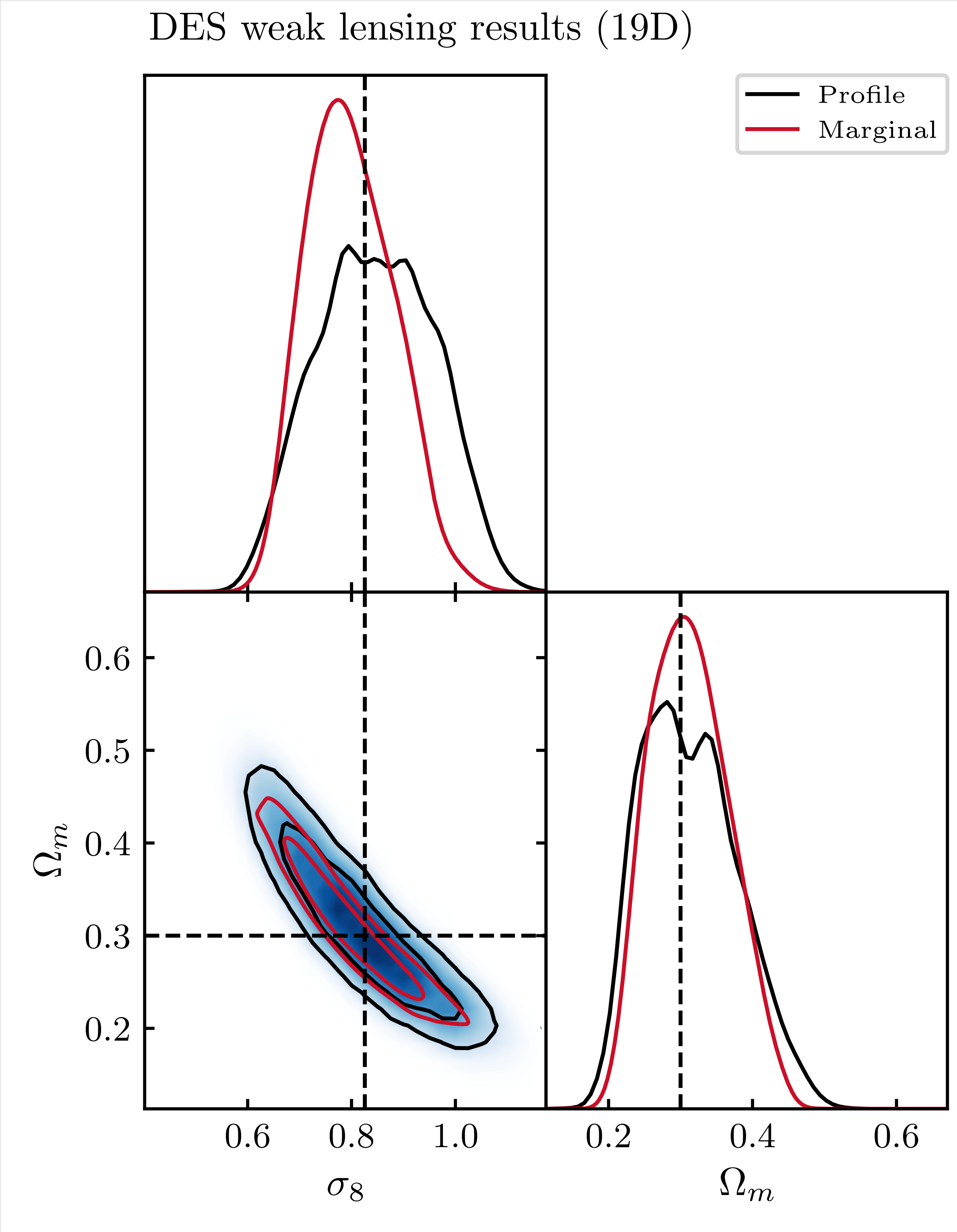


With an accurate flow model we can afford lots of maximizations

High (32D) dimensional Gaussian mixture

1D and 2D profiles in minutes

Back to the DES example



The sigma8/omegam degeneracy is flatter than expected

Noise may move posteriors by more than what the marginal implies

Partially known - especially in extended models

What can we do about it?

Projection effects may arise because of two effects:
1- non-Gaussian likelihood
2- informative shaped (along the los) prior

N. 2 can be minimized by looking at best constrained parameters

Best constrained parameters maximize the difference between prior and posterior
See Dacunha + (2112.05737)

What can we do about it?

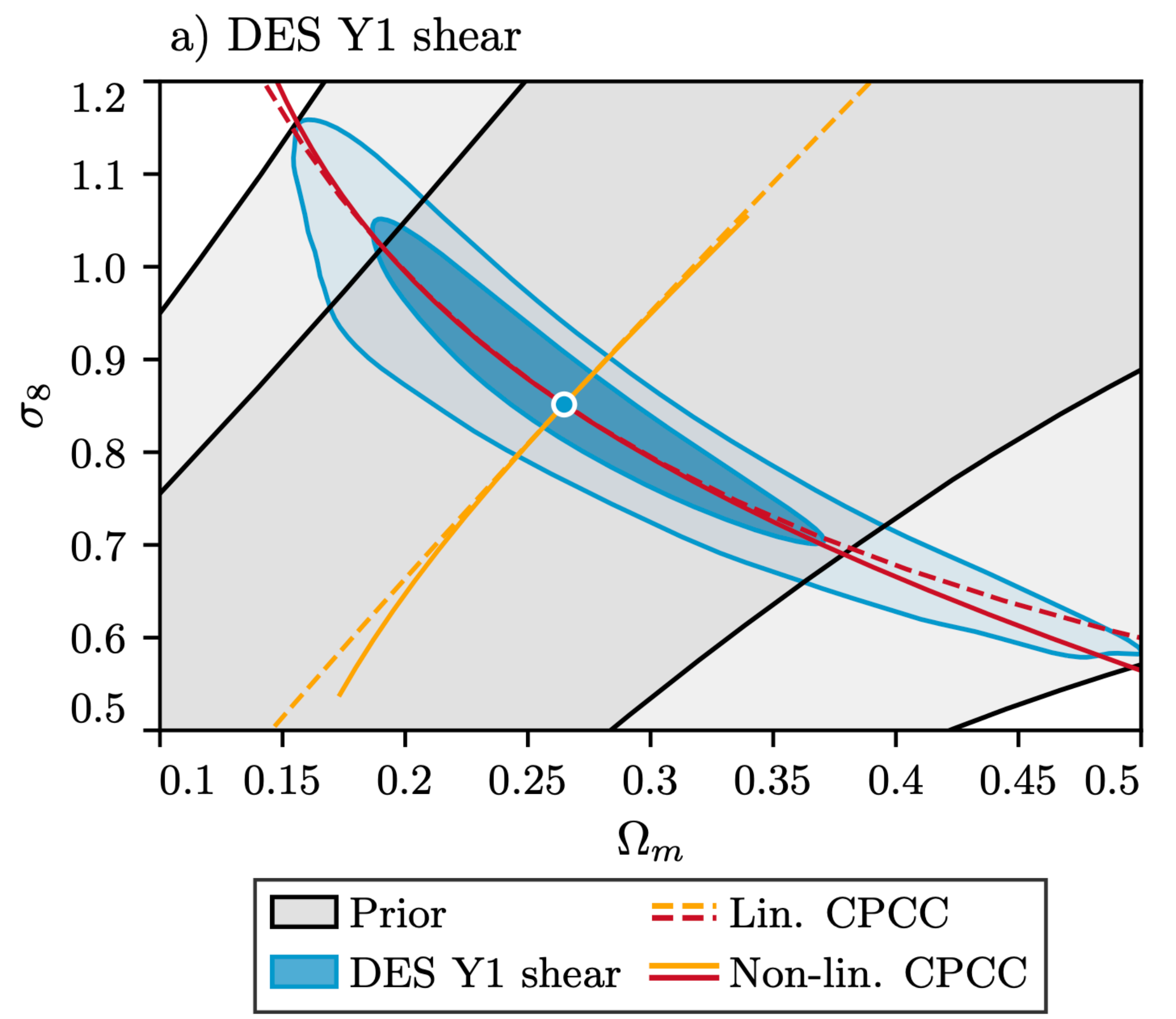
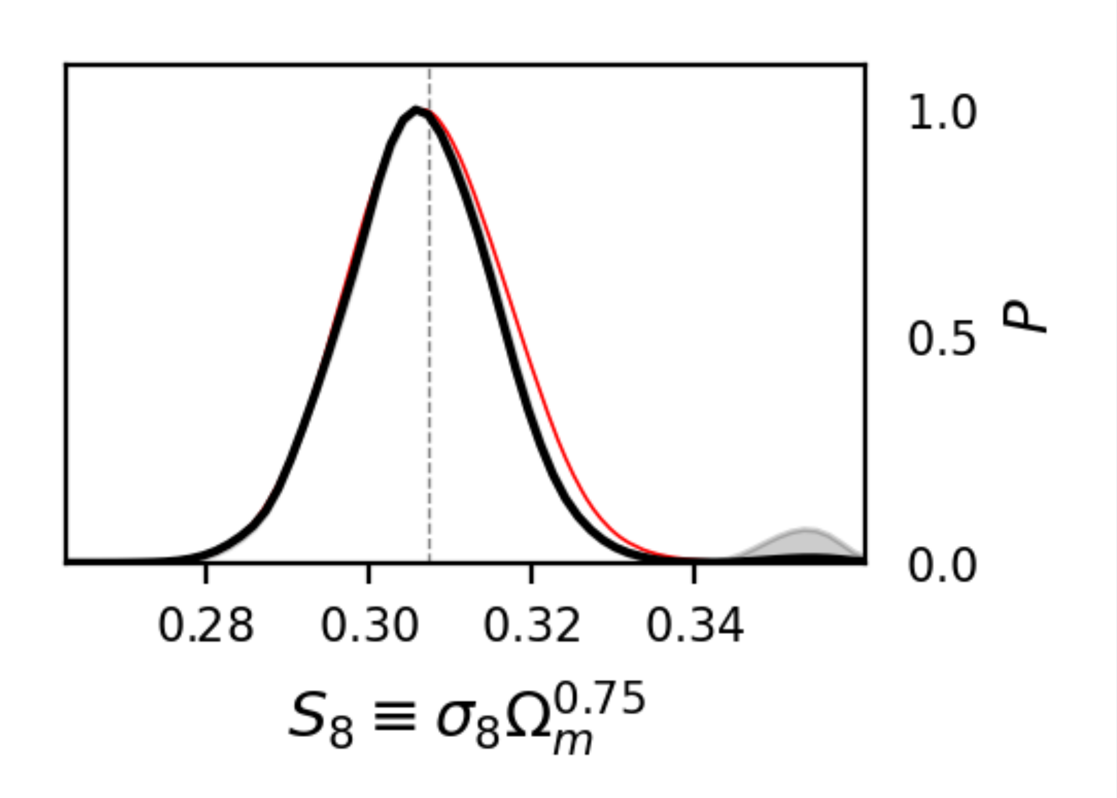


Figure out best constrained parameters

Tensiometer has you covered (See tutorial)

If marginal/profile agree believe those…



Conclusions

* In many cosmology examples, we see projection effects that complicate the interpretation of the posterior we have

* These effects arise because of either weak data constraints or genuinely non-Gaussian likelihoods

* A difference between marginal and profiled distributions is a warning sign -> if found understand what the data is measuring before looking at the parameters

Conclusions

* Systematic profiling was unfeasible/extremely computationally intensive until today

* Tensiometer has tools to produce profiled triangles in minutes. And many tutorials to show you how to do it!

* Have fun!