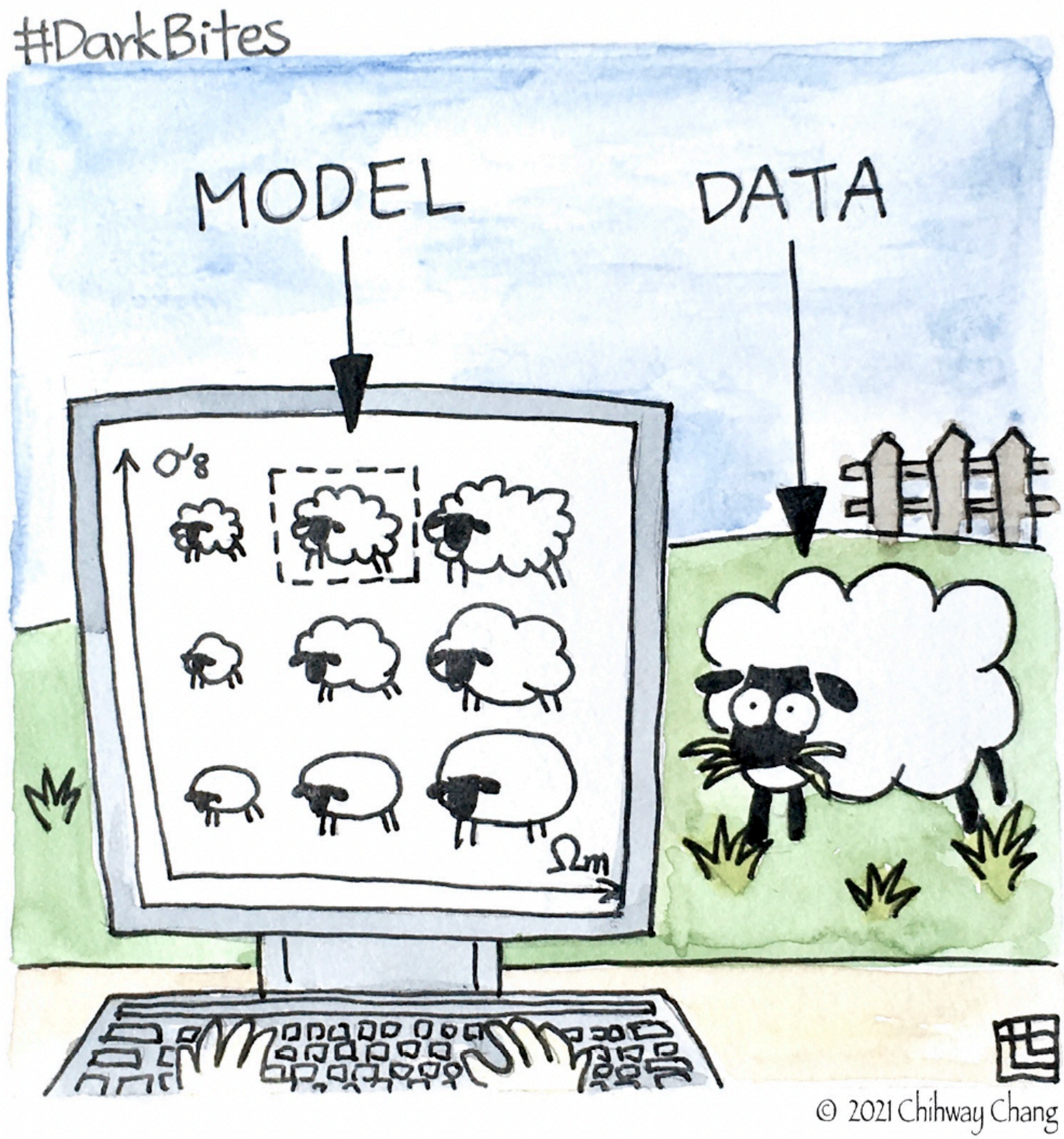
From Summary Statistics to Parameters



What is Probability?

Classical: Probability as frequency.

Probability of an event := *the number of times the event occurs over the total number of trials, in the limit of an infinite series of equiprobable repetitions.*

model is fixed, data are repeatable

Bayesian: Probability as degree of belief.

Probability is a measure of the degree of belief about a proposition.

data are fixed, model is repeatable

Trotta: Bayes in the sky, 0803.4089

Bayesian and Frequentist statistics

Frequentist: model is fixed, data are repeatable
Bayesian: data are fixed, model is repeatable

Say H0 = (72 ± 2) km/s/Mpc. Then:

Frequentist: Performing the same procedure with independent data will cover the real value of H0 within the limits 68% of the time.
(Limited practicability in cosmology…)

Bayesian: the posterior distribution for H0 has 68% if its integral between 70 and 74 km/s/Mpc. The posterior can be used as a prior for future analyses of independent data.

Bayesian Parameter Inference

Bayes Law:

Posterior

$$P(p|dM) = \frac{P(d|pM)P(p|M)}{P(d|M)}$$

Likelihood

Prior

$$\propto P(d|pM)P(p|M)$$

Model

Observed data

Parameters

What you know after the experiment (posterior)
= what you knew before (prior)+ what you learn (likelihood)

Priors

- Priors quantify what you knew about the parameters before the experiment

    Theoretical limits, preferences, things that must be true (e.g., from previous experiments)

- In regions where the likelihood is zero your prior doesn't matter for parameter estimation, *but can for more advanced model selection*

- It is common practice in cosmology to use uniform priors for most parameters

    easy to write down, hard to justify

→**Sensitivity analysis**: change priors, check how your conclusions change!

Transformed Priors

- In particular note that a uniform prior in one parametrisation may not be uniform in another - probability mass is conserved, not density:

$$\int P(x)\mathrm{d}x = \int P(u)\mathrm{d}u$$

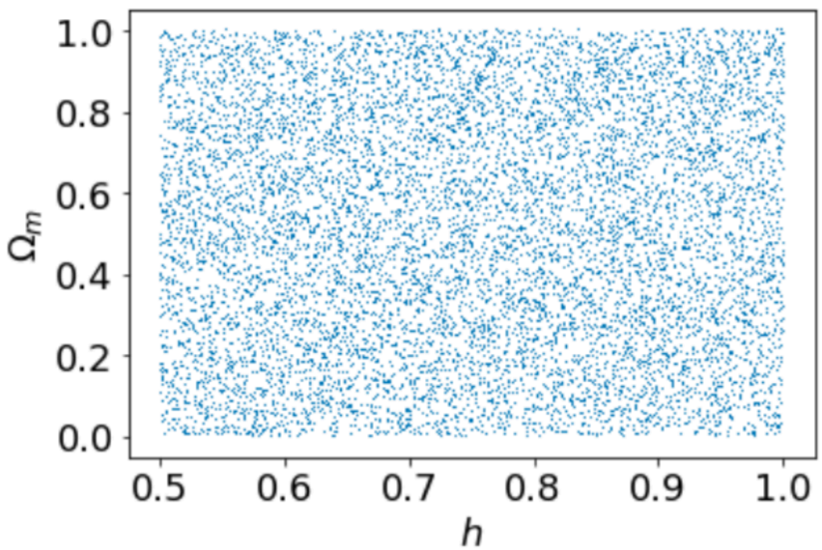
- So when we transform to new variables:

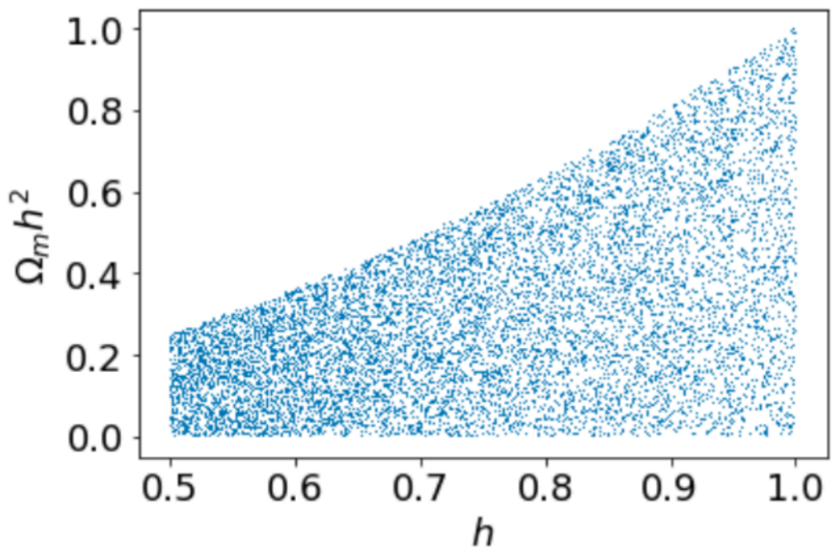
$$P(u) = P(x) \;/\; \left|\frac{\mathrm{d}u}{\mathrm{d}x}\right|$$

- or in higher dimensions:

$$P(\mathbf{u}) = P(\mathbf{x})/|J| \qquad\qquad J_{ij} = \frac{\partial u_i}{\partial x_j}$$

Transformed Priors





Jointly uniform priors on Ω_m - h

Implied priors on $\Omega_m h^2$ - h

Likelihoods

Most existing cosmological analyses assume Gaussian likelihood

$$\ln \mathcal{L}(\mathbf{D}|\mathbf{p}) \propto -\frac{1}{2}\left[(\mathbf{D}-\mathbf{M}(\mathbf{p}))^{\tau}\,\mathbf{C}^{-1}\,(\mathbf{D}-\mathbf{M}(\mathbf{p}))\right]$$

Assumes data points are Gaussian-distributed around the truth – reasonableness depends on type of measurement and sources of noise.

Alternatives:

- non-Gaussian likelihood (explored in e.g. Lin et al. 2019, Hall & Taylor 2022) – low on the priority list *for 2pt statistics.*
- Likelihood-free Inference (LFI), Simulation-base Inference (SBI)

Sampling the Likelihood

For most data sets, likelihoods cannot be written in a simple closed form equation.

We cannot just evaluate/plot posteriors directly, but instead musts indirect methods.

Most obvious solution is to evaluate at every point in the space, on a grid. Impossible for high-dimensional parameter spaces!

$\rightarrow$ sampling methods like Monte-Carlo Markov Chains.

each element of Markov Chain depends only on the previous one

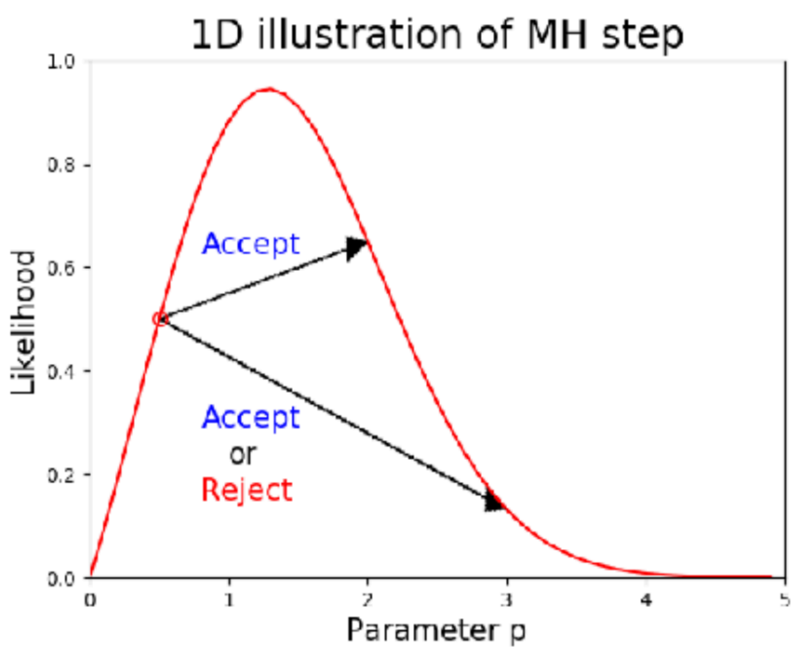
basic algorithm: Metropolis–Hastings

improved in widely used packages `Emcee, Zeus`

limitations: lack of definitiveness that the chain has converged

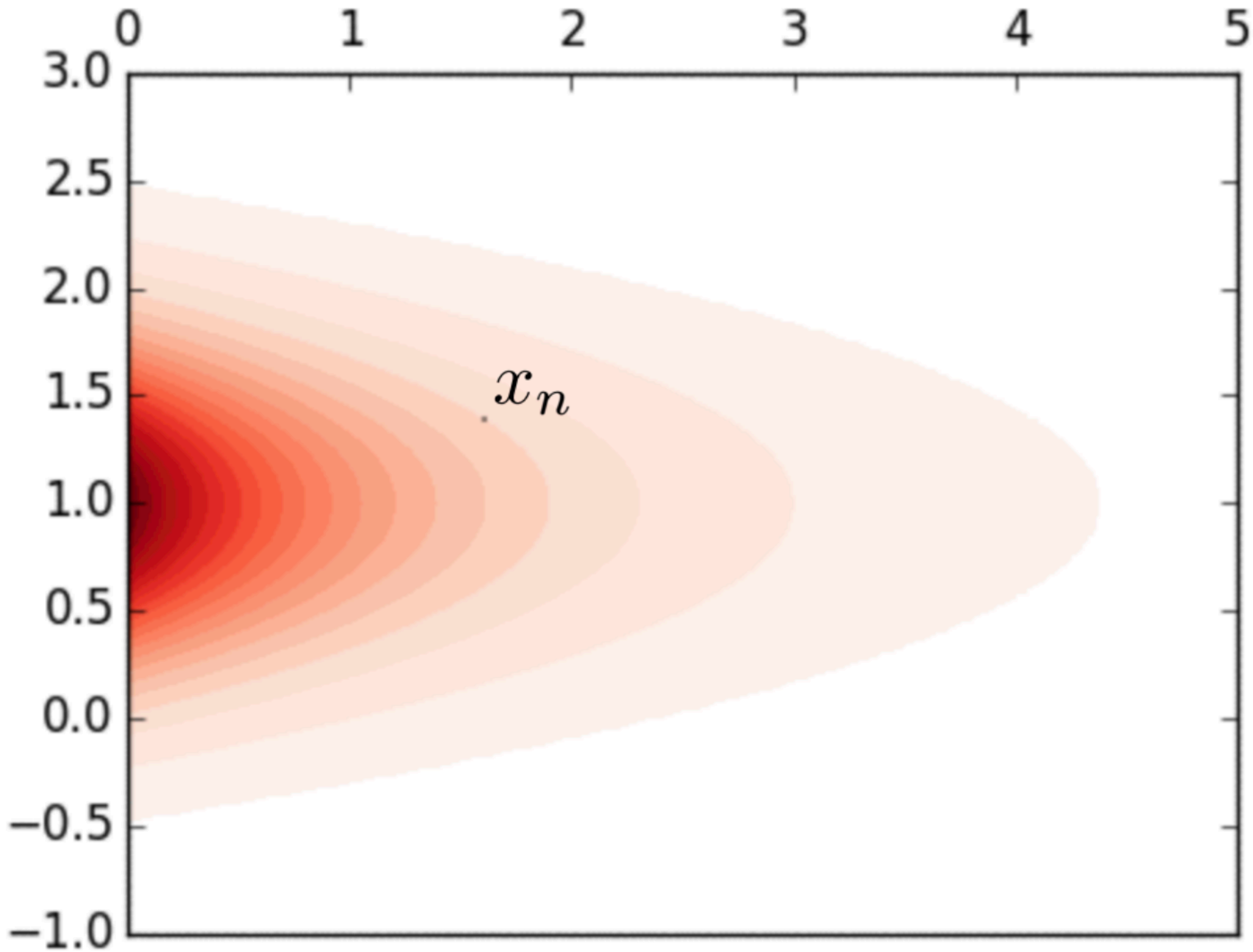
The Metropolis-Hastings algorithm

- at step t, at some parameters p_t
- propose move to $p_t' = p_t + \Delta p_t$ (randomly draw Δp_t)
- evaluate $r = L(p_t')/L(p_t)$
- MH step:
  - if $r > 1$ accept move
  - if $r < 1$ generate a random number $\alpha \in [0, 1]$
    - if $\alpha < r$, accept move
    - if $\alpha > r$, reject move
- t=t+1

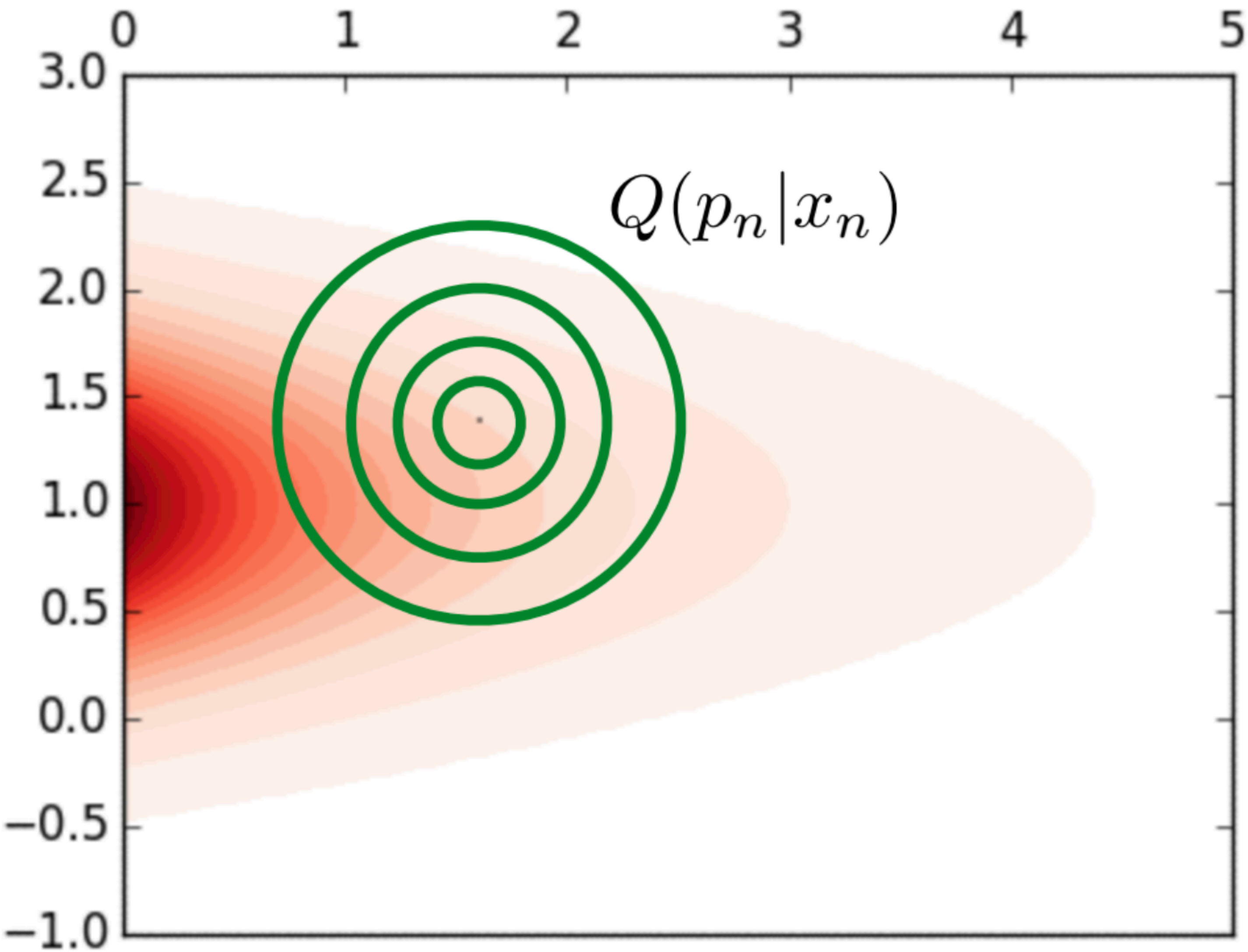


One can prove that this algorithm asymptotically recovers the true posterior

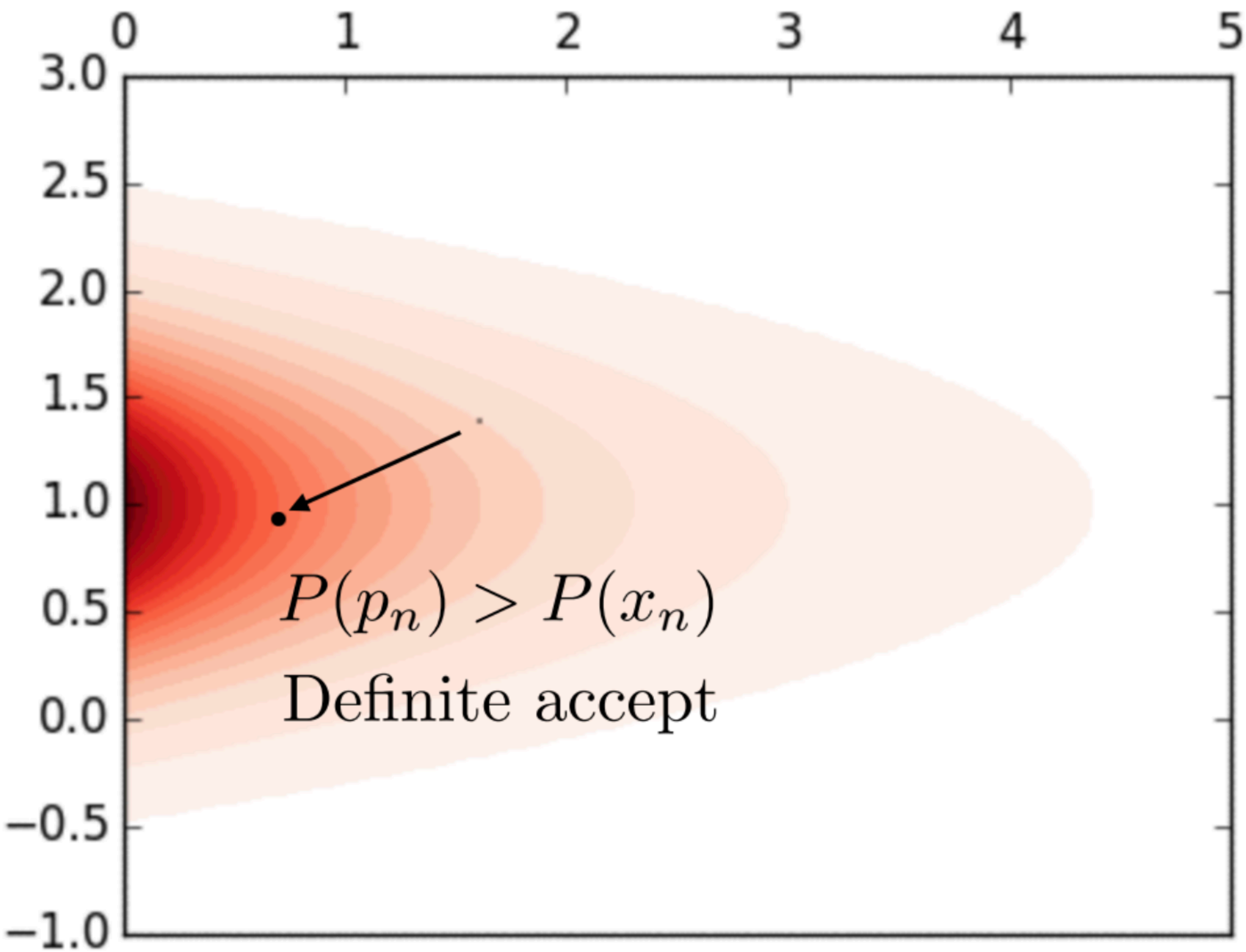
The Metropolis-Hastings algorithm



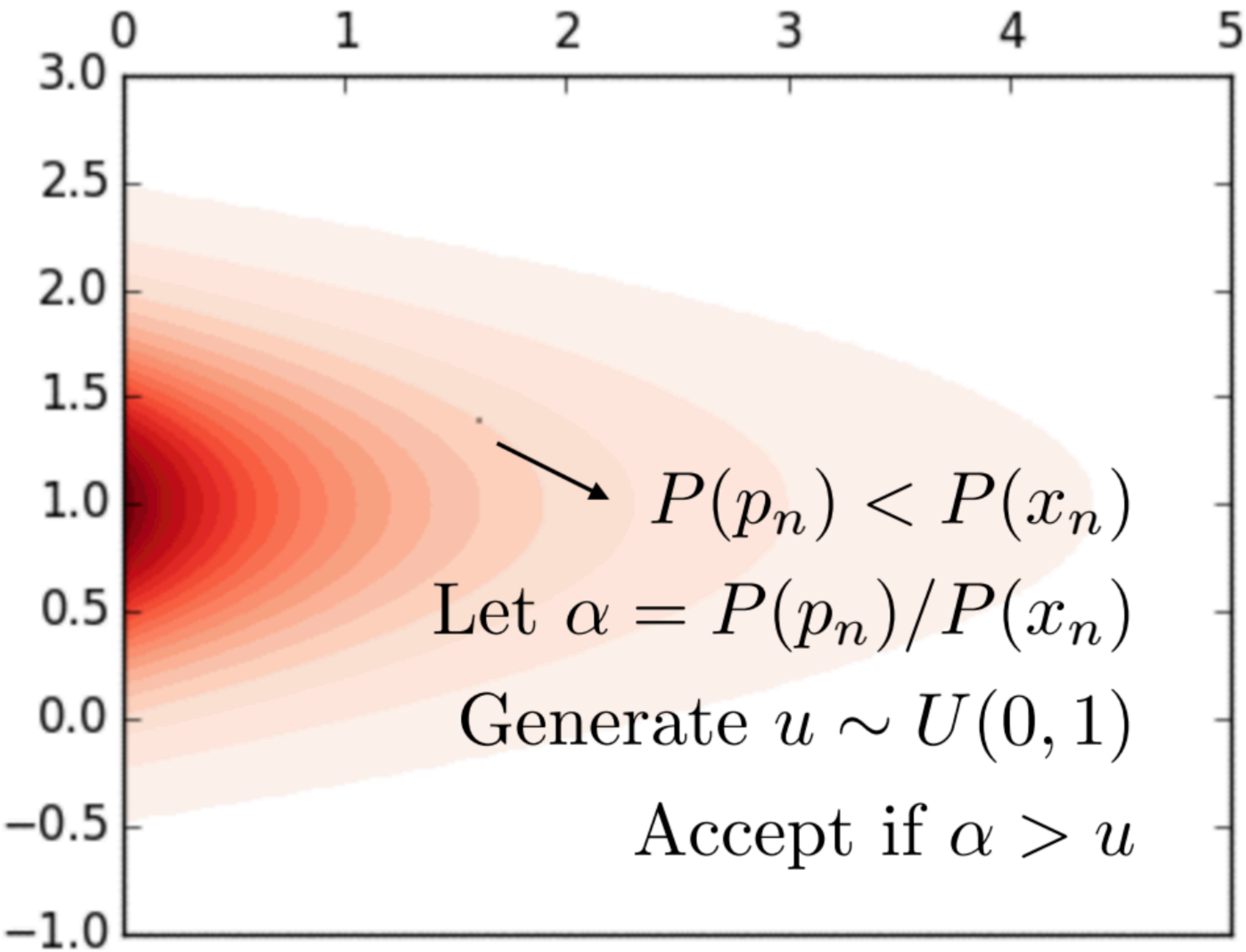
The Metropolis-Hastings algorithm



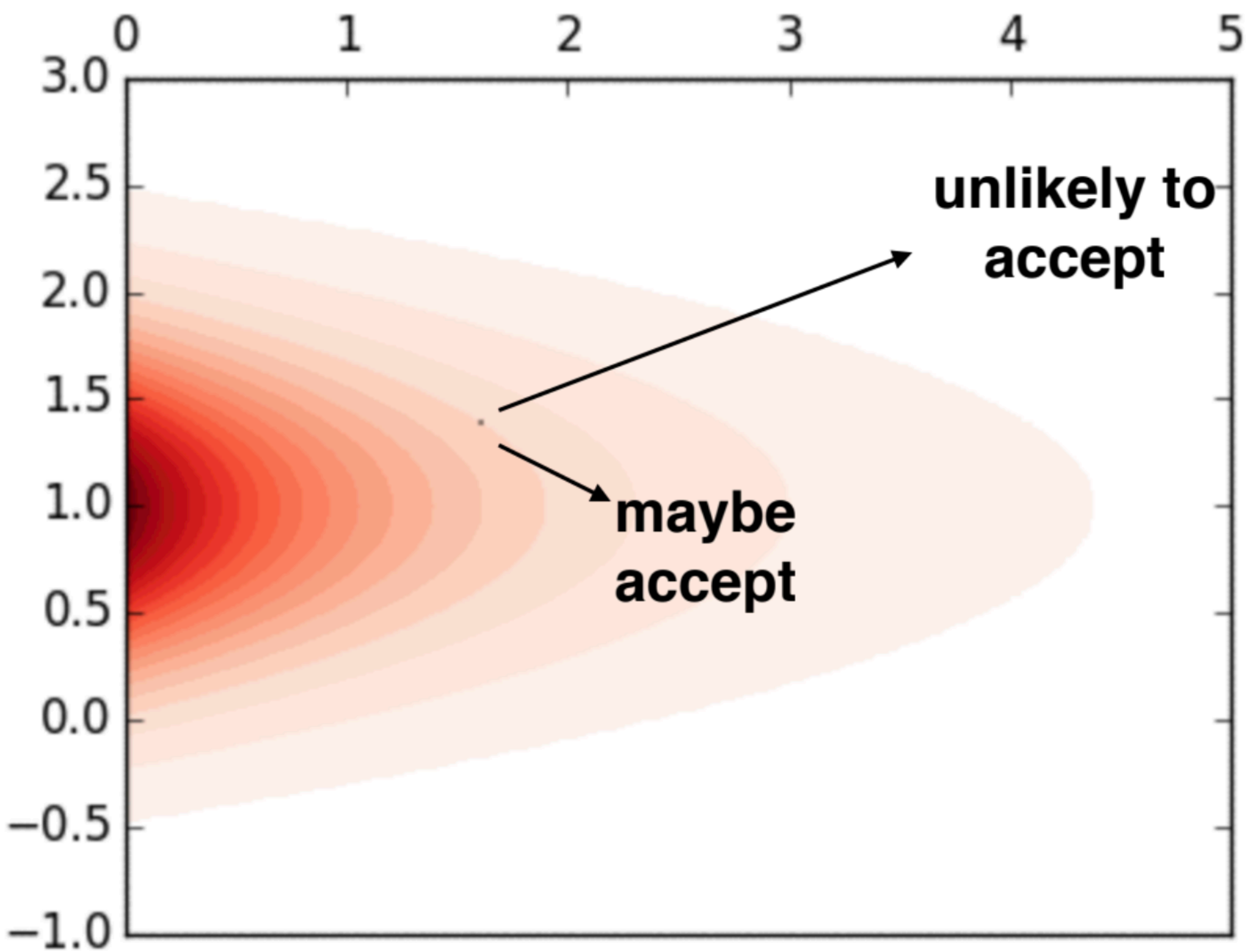
The Metropolis-Hastings algorithm



The Metropolis-Hastings algorithm

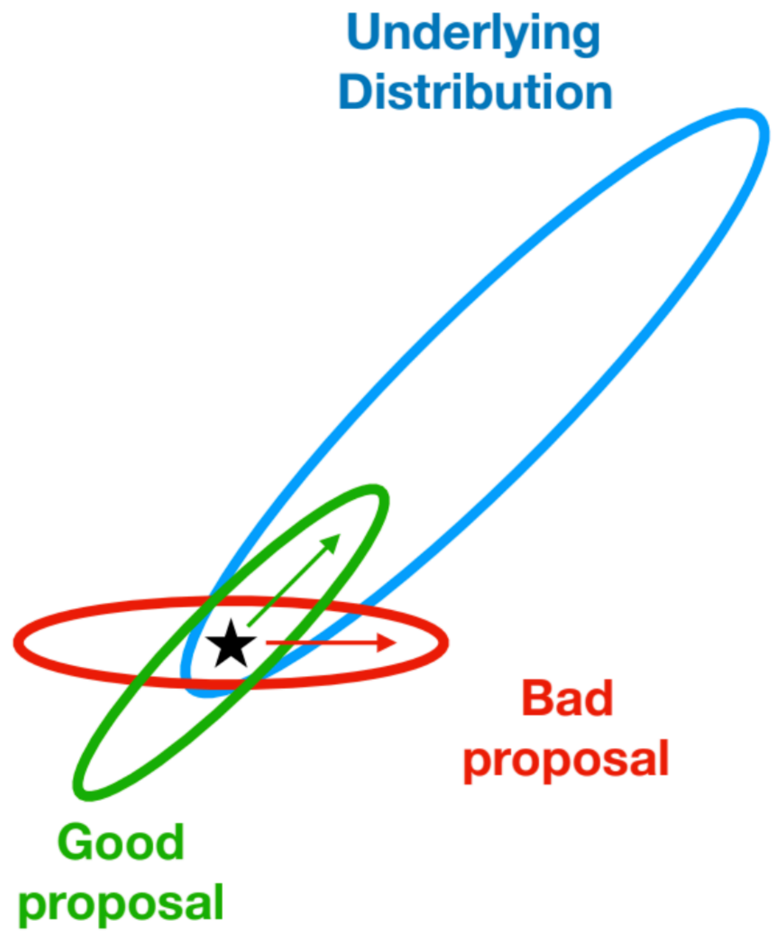


The Metropolis-Hastings algorithm



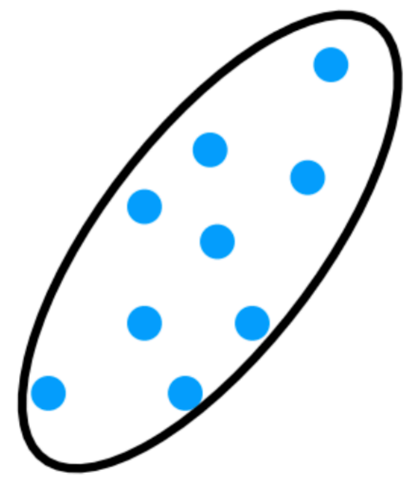
MCMC: Proposal Distribution

- Efficiency of MH depends dramatically on how good the proposal is
- A bad proposal will not converge in any practical length of time
- The ideal proposal matches the shape of the underlying distribution
    - We don't know this, but can look for best approximation



MCMC: Proposal Distribution

- One way to get a good proposal is by tuning
  - Run a short initial chain to estimate covariance
  - Use this covariance to initialise the next iteration
- You have to throw away the first chain, and only use samples from when your tuning was finished
  - Detailed balance broken
  - There are specific algorithms that do let you do a variant of this, but not standard MH



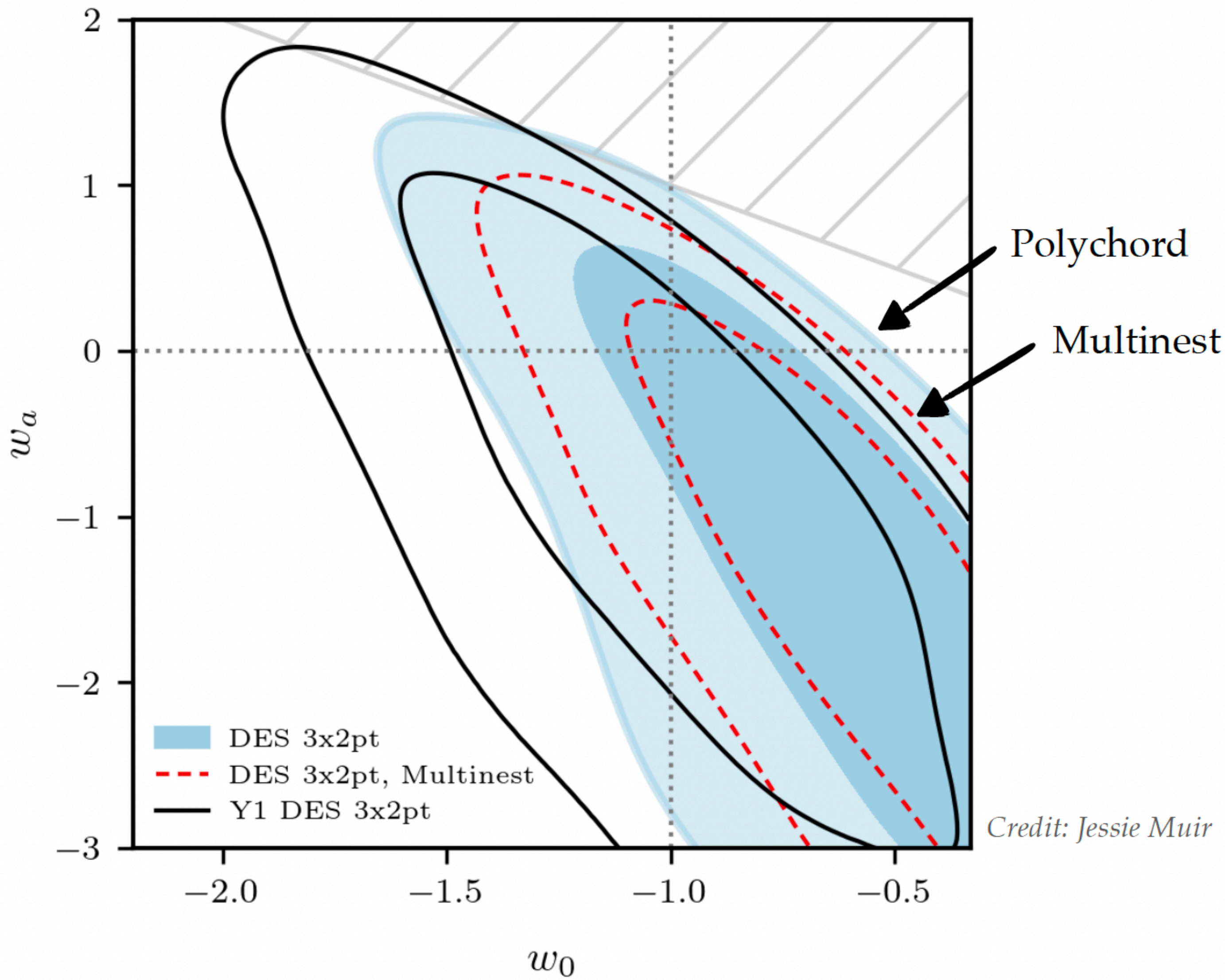
Intricacies of High-Dimensional Sampling

Nested sampling: starts with a large number of points, and repeatedly eliminates and find new replacement points

- e.g. Multinest, PolyChord
- calculate Bayesian evidence simultaneously

Choosing the right sampler to accurate sample your parameter space is an art - *and hard validation work.*

Intricacies of High-Dimensional Sampling



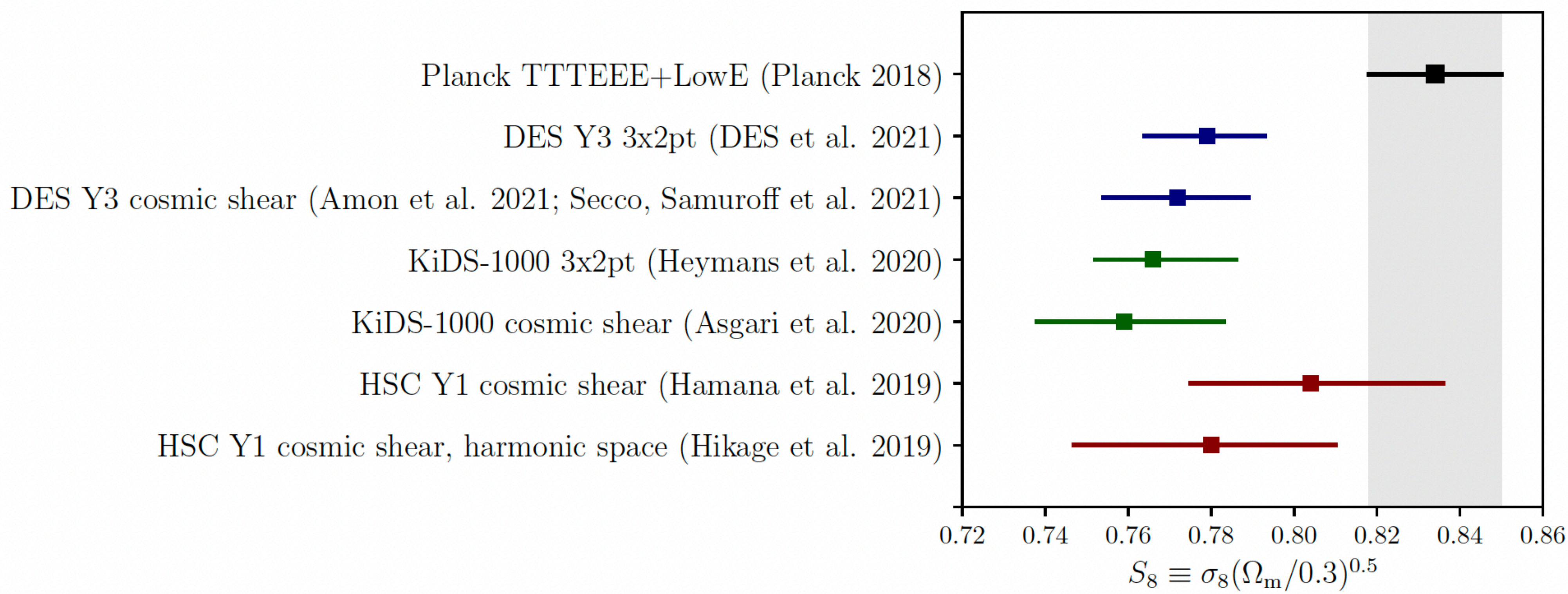
Interpreting chains

- Check to see if we actually found a good fit
- Quote the cosmological constraints, check to see if we've broken ΛCDM yet
- Compare with other similar measurements
- Compare with other independent measurements

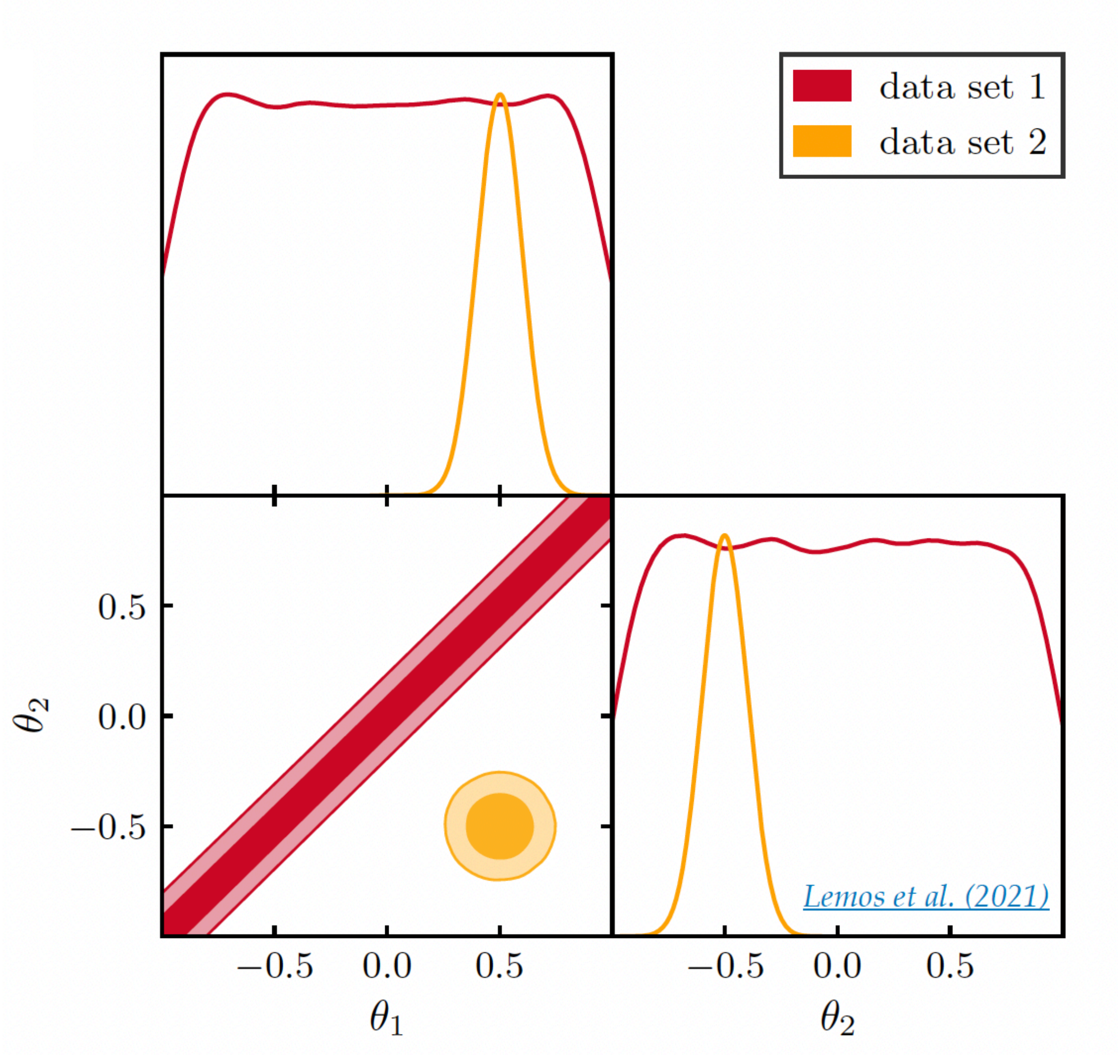
Interpreting chains

Can only plot 1D/2D results - report marginalized constraints.

$$P(\theta_1|d) = \int d^{n-1}\theta_{2..n} P(\boldsymbol{\theta}|d)$$



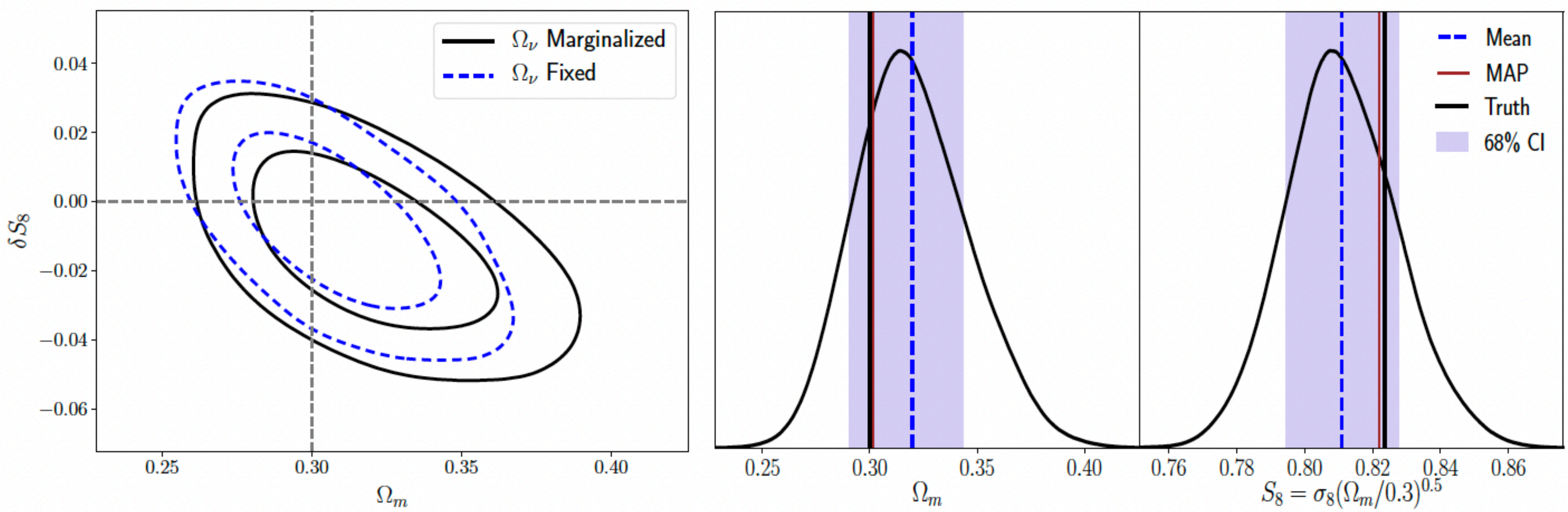
Marginalized Parameters



Beware of Projection/Prior Volume Effects!

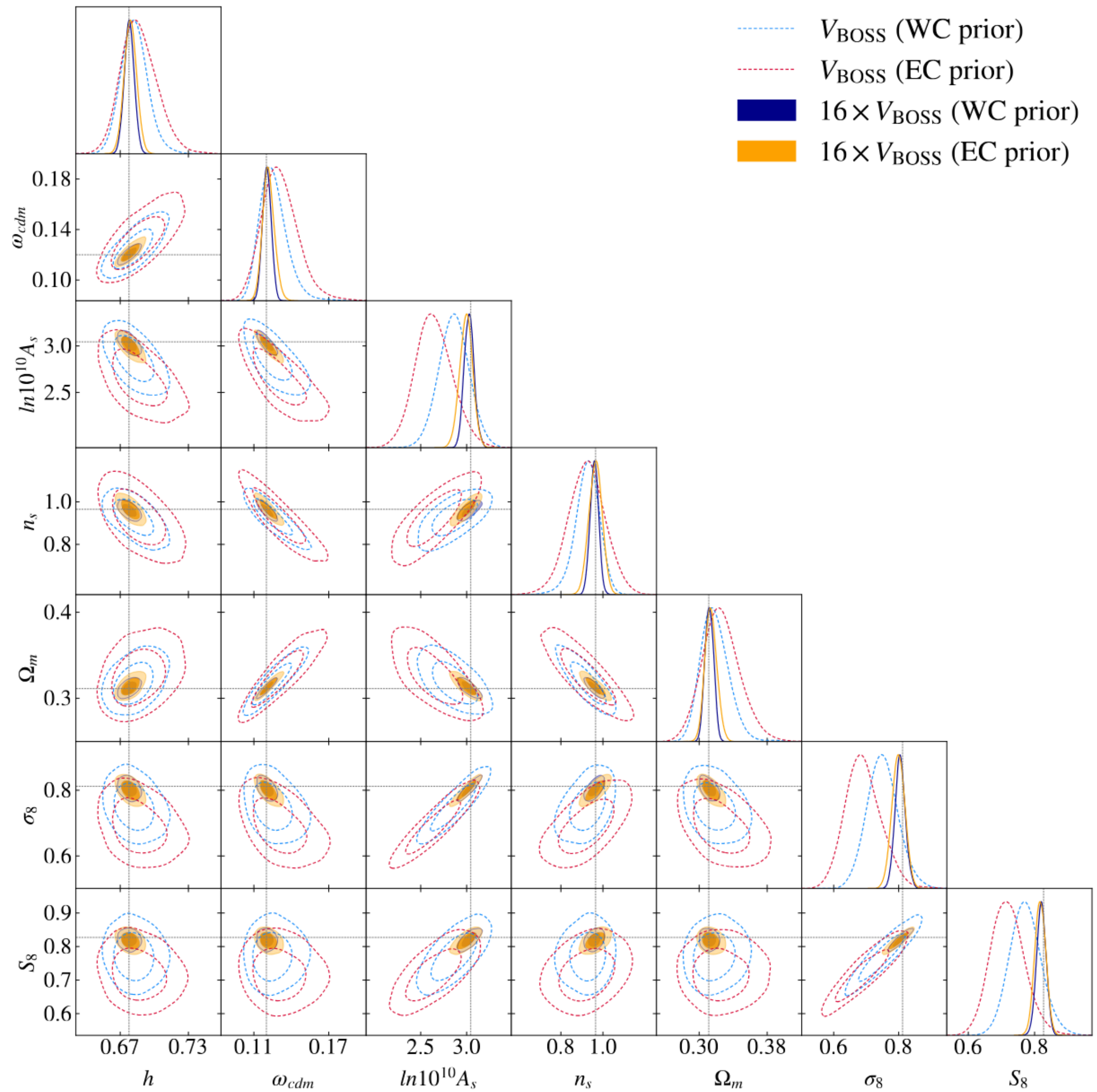
Parameters of interest may be correlated with poorly constrained "nuisance parameters".

Marginalization may introduce projection effects, skew marginalized posteriors away from best fit.



This effect can be characterized on synthetic data!

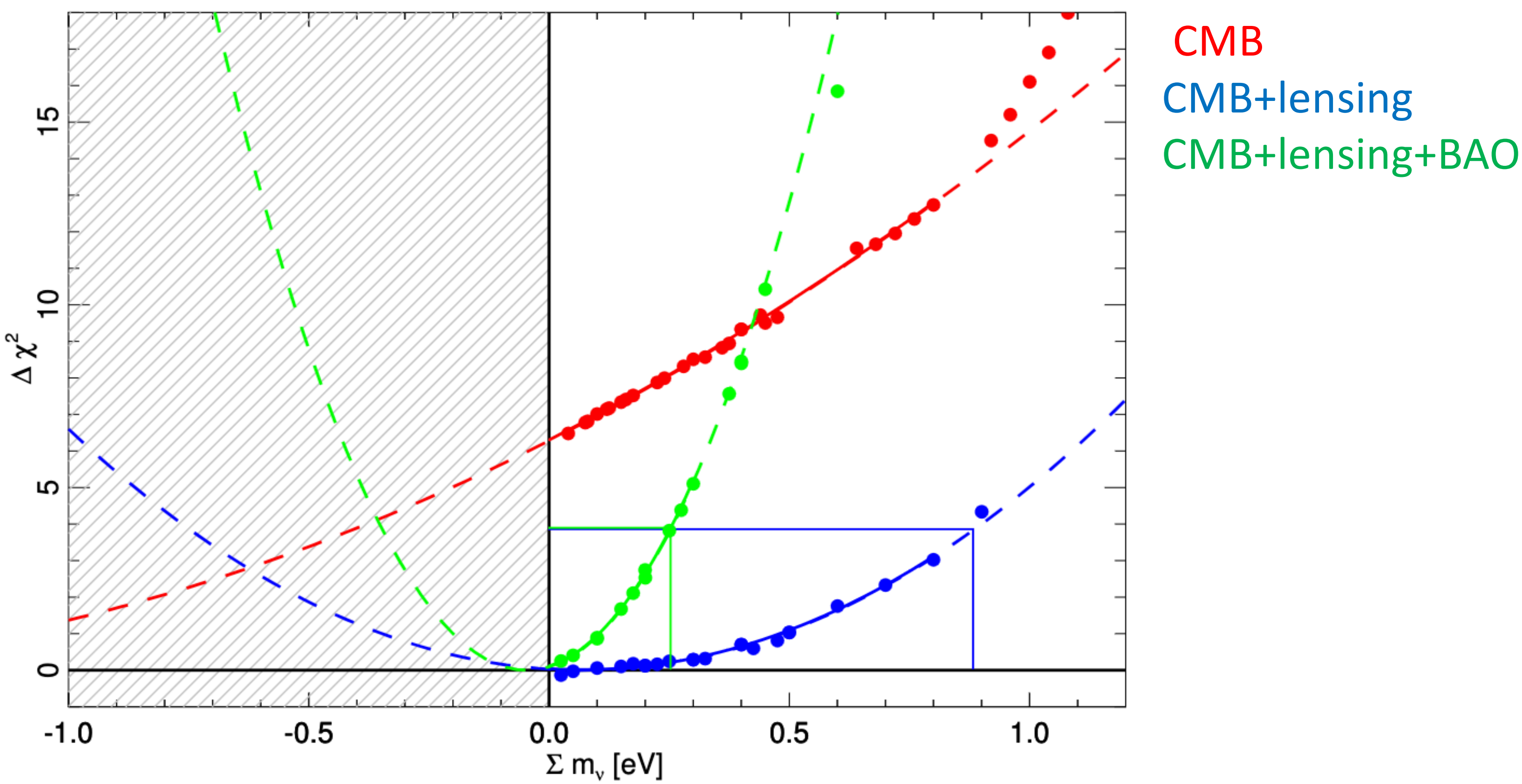
Beware of Projection/Prior Volume Effects!



Simon et al. 2023: EFTofLSS analyses of BOSS data with different nuisance parameter priors.

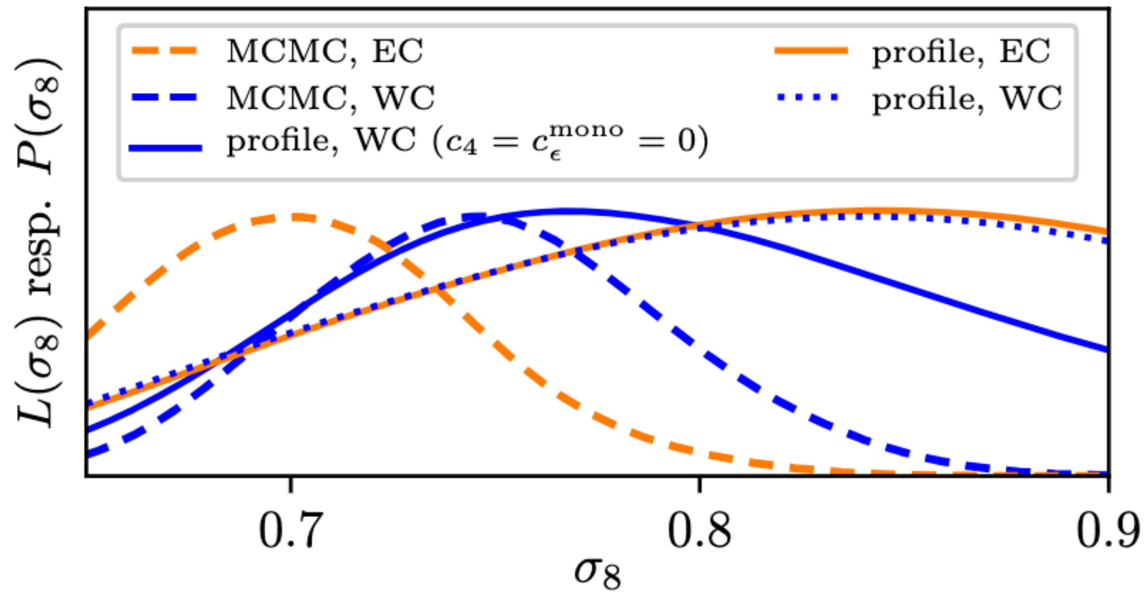
Profile Likelihoods

Frequentists' way to treat nuisance parameters ν $L(\boldsymbol{\theta}) = \max_{\boldsymbol{\nu}} L(\boldsymbol{\theta}, \boldsymbol{\nu})$

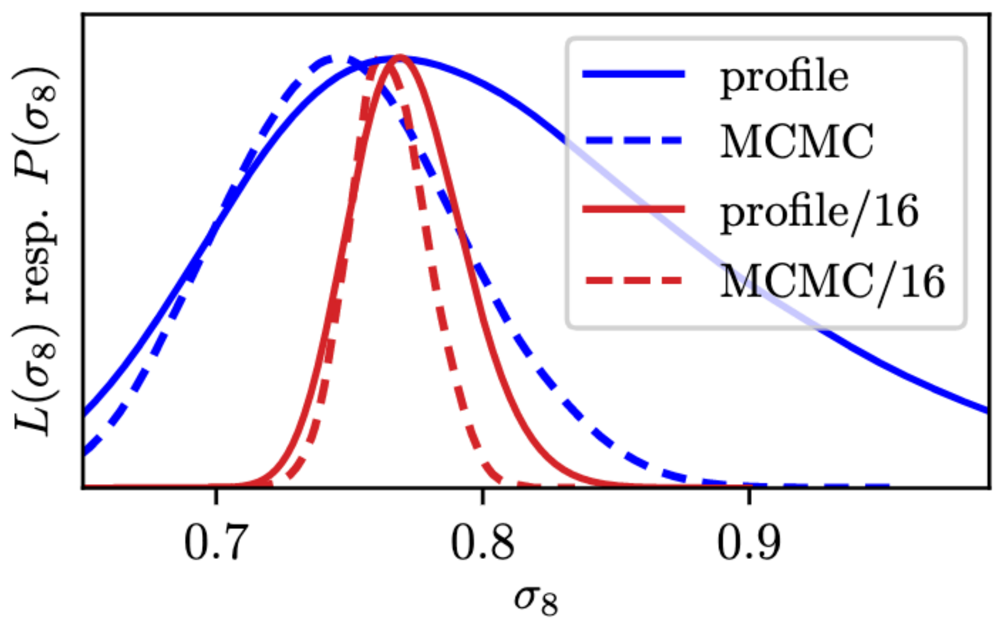


Planck Profile Likelihood (37 parameters!) 1311.1657

Profile Likelihoods



Holm+ 2023
Comparison of MCMC, profile likelihoods for EFTofLSS BOSS analyses



Improved constraining power will reduce difference between frequentist and Bayesian statistics.

Model Comparison/Selection

- Given two models, how can we decide which fits the data better, overall?
- Simplest approach: compare best fit points
    - Does not include uncertainty or Occam's Razor
- Recall that all our probabilities have been conditional on the model, as in Bayes:

$$P(p|M) = \frac{P(d|pM)P(p|M)}{P(d|M)}$$

Model Comparison/Selection

- Evidence is the bit we ignored before when doing parameter estimation

- Given by an integral over prior space

$$P(d|M) = \int P(d|pM)P(p|M)\mathrm{d}p$$

- Hard to evaluate - posterior usually small compared to prior

Model Comparison/Selection

- Can use Bayes Theorem again, on model level:

$$P(M|d) = \frac{P(d|M)P(M)}{P(d)}$$

- Only really meaningful when comparing models:

Bayesian Evidence Ratio

$$R = \frac{P(M_1|d)}{P(M_2|d)} = \frac{P(d|M_1)}{P(d|M_2)} \frac{P(M_1)}{P(M_2)}$$

Model Priors

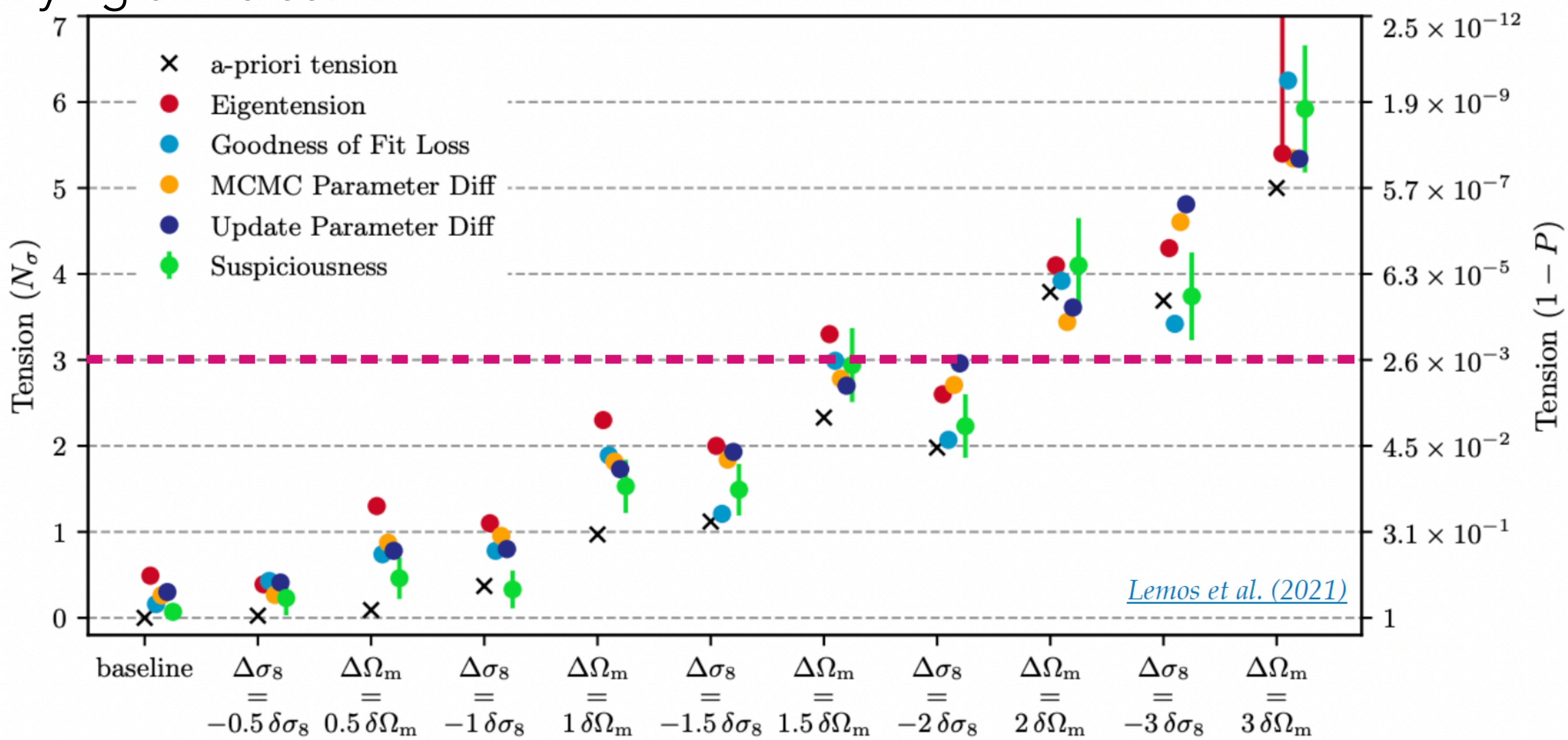
Bayesian Evidence Values

Comparing Experiments

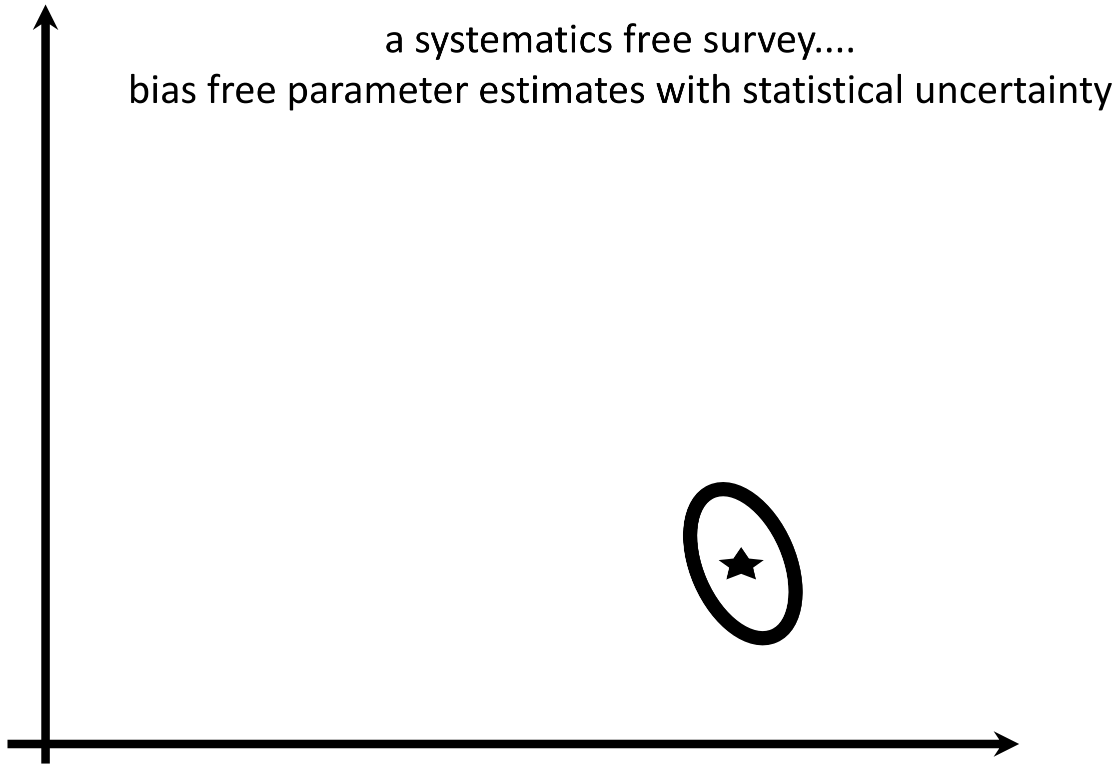
We like to quantify to what extend our results are consistent with other experiments.

Complicated since we are comparing two chains in very high dimension, and the effect of priors are non-trivial.

In the past few years, many have devised certain statistics ("tension metrics") to quantify how likely the two experiments are realizations drawn from the same underlying universe.



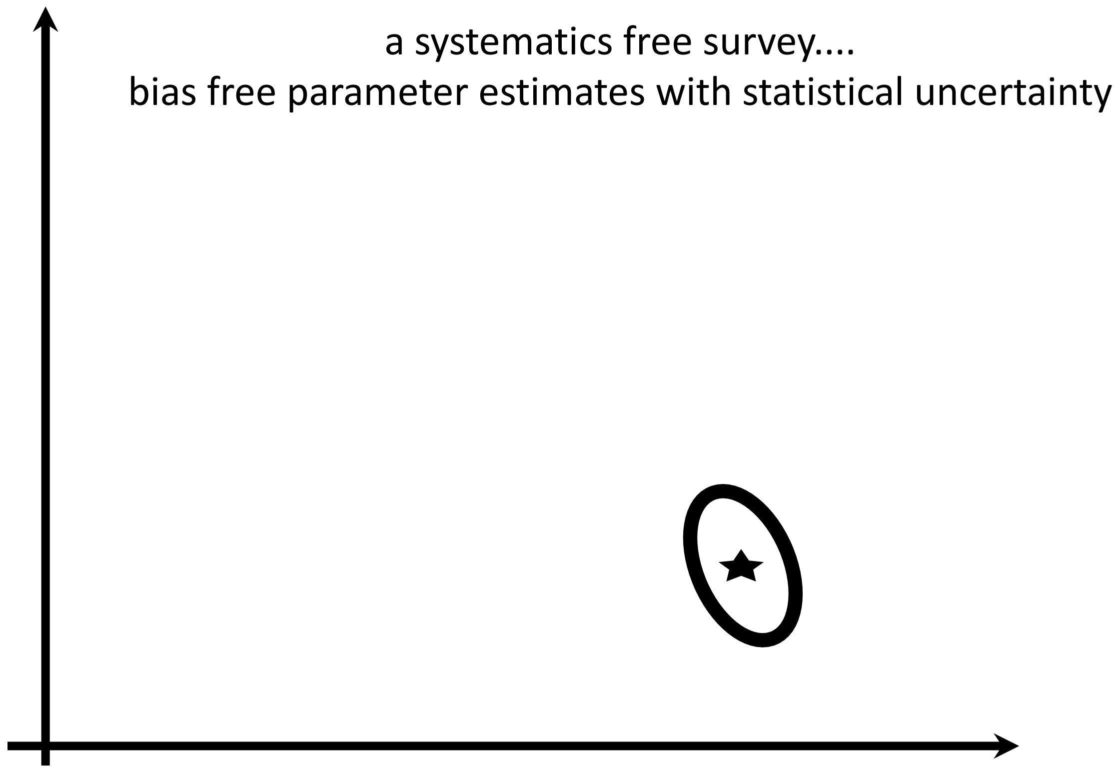
Model Incompleteness



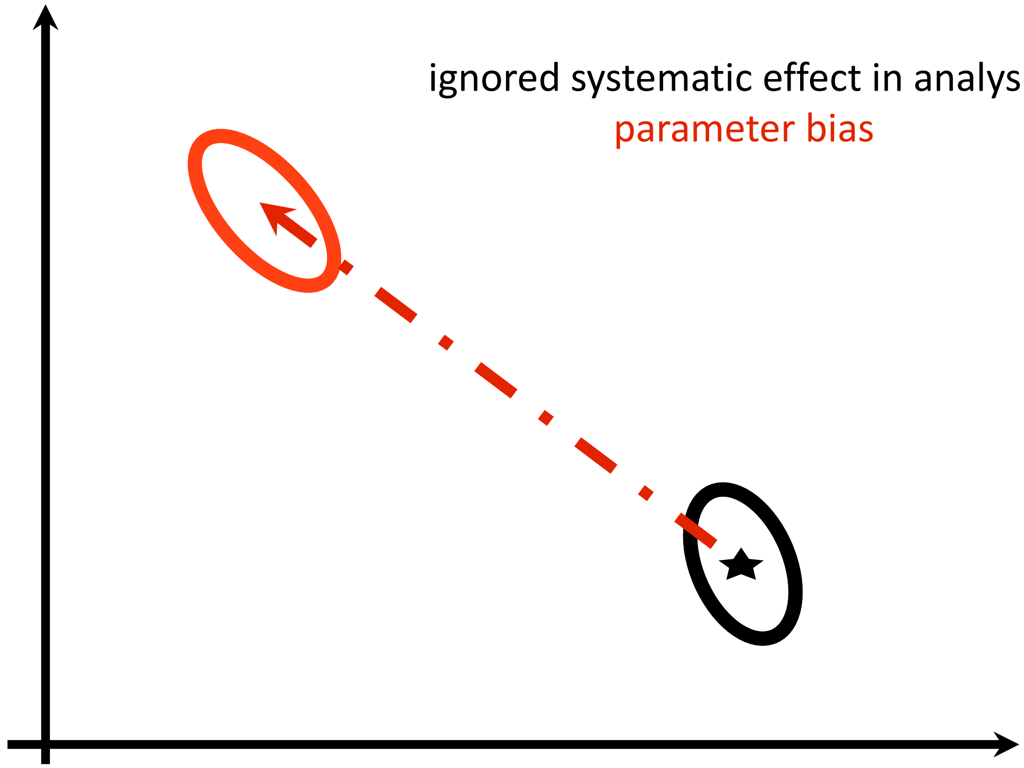
So far, we haven’t discussed the model itself yet…

While details depend on the cosmological probe and survey, we can say that no model will perfect.

Model Incompleteness

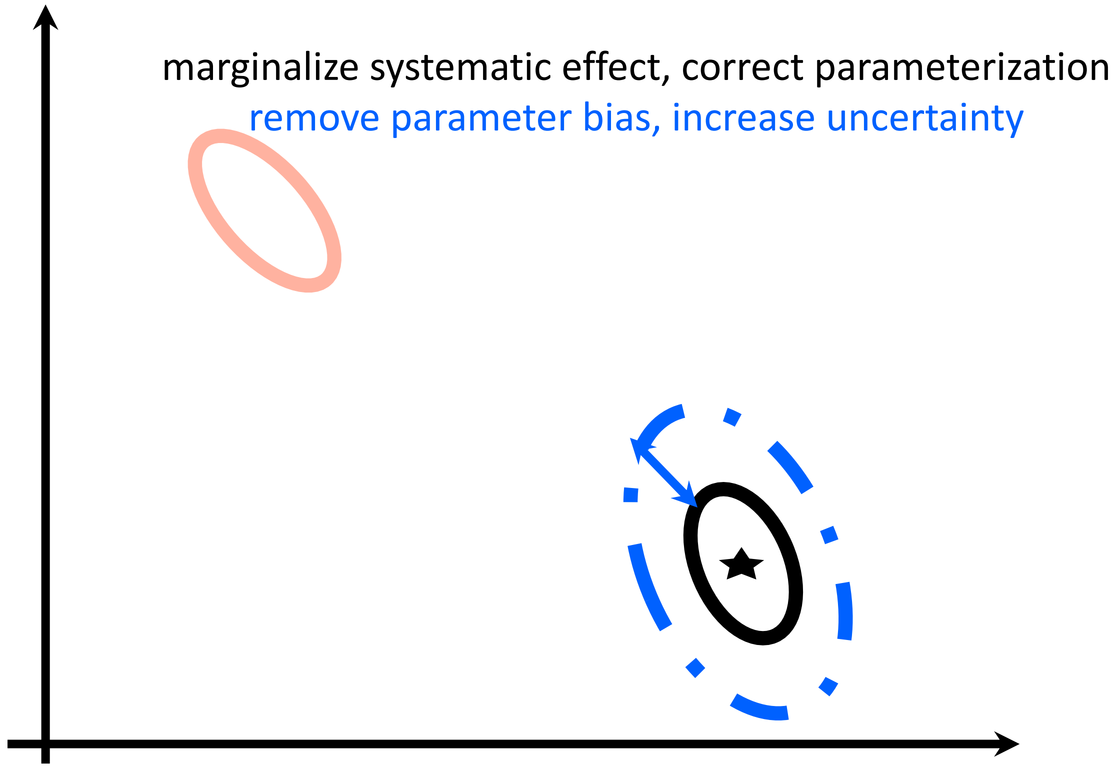


Model Incompleteness



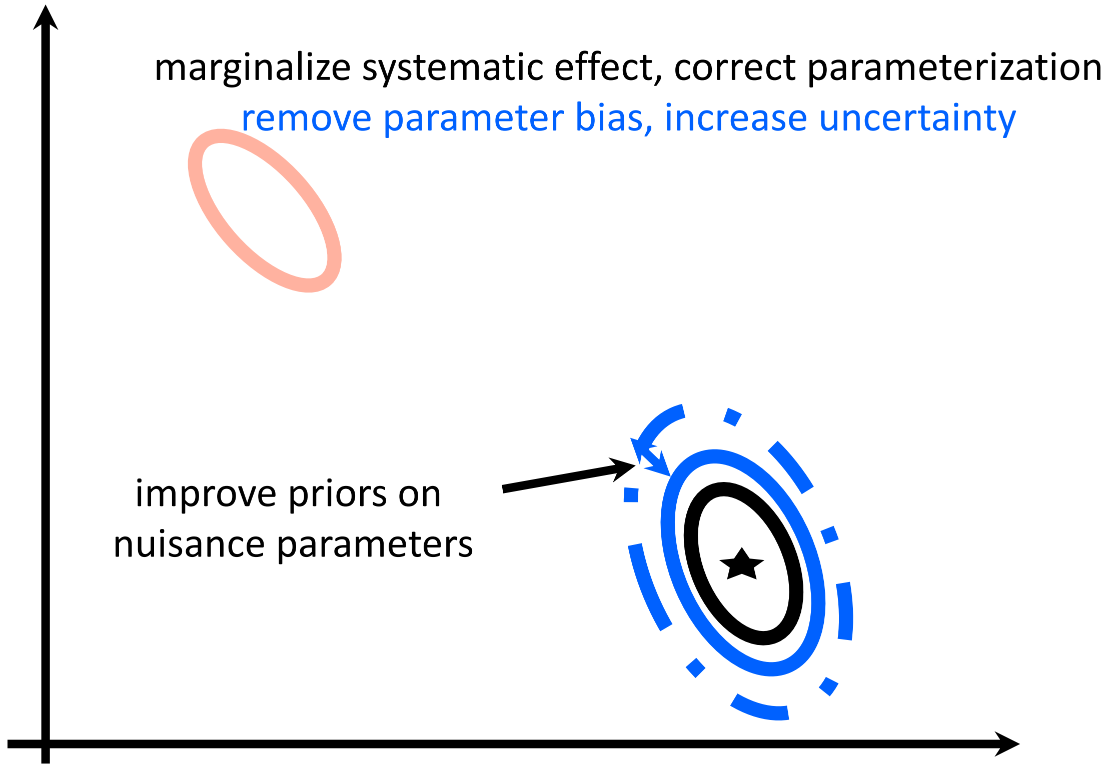
It’s research time!

Model Incompleteness

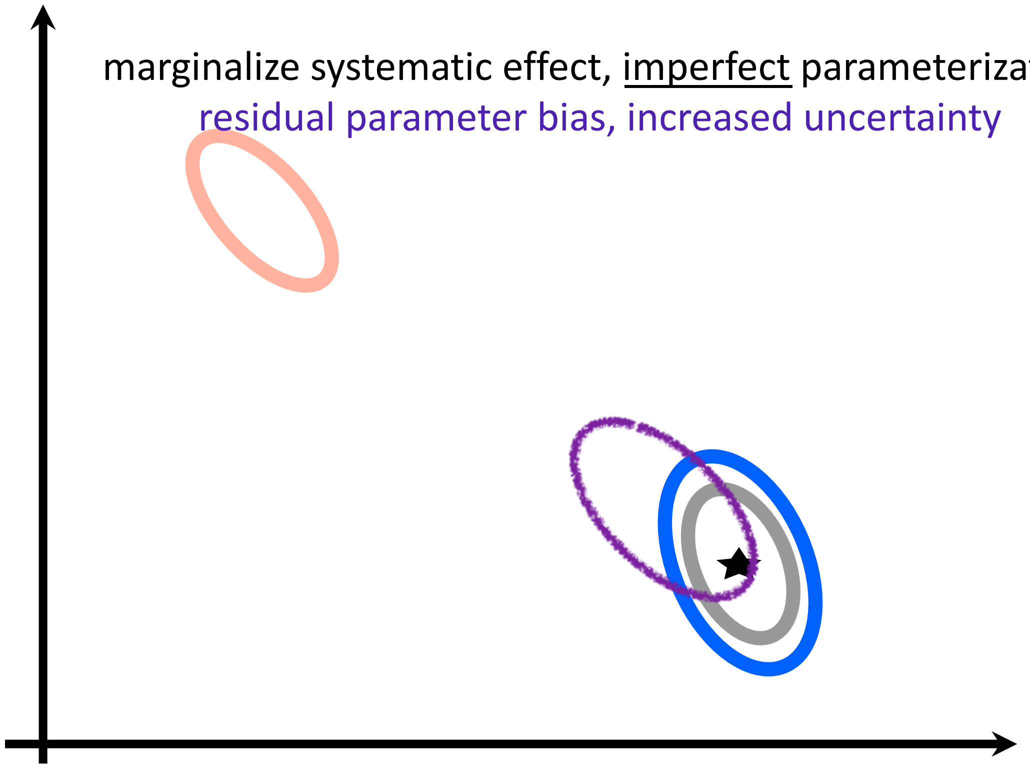


It's research time!

Model Incompleteness



Model Incompleteness + Misspecification



It's research time!

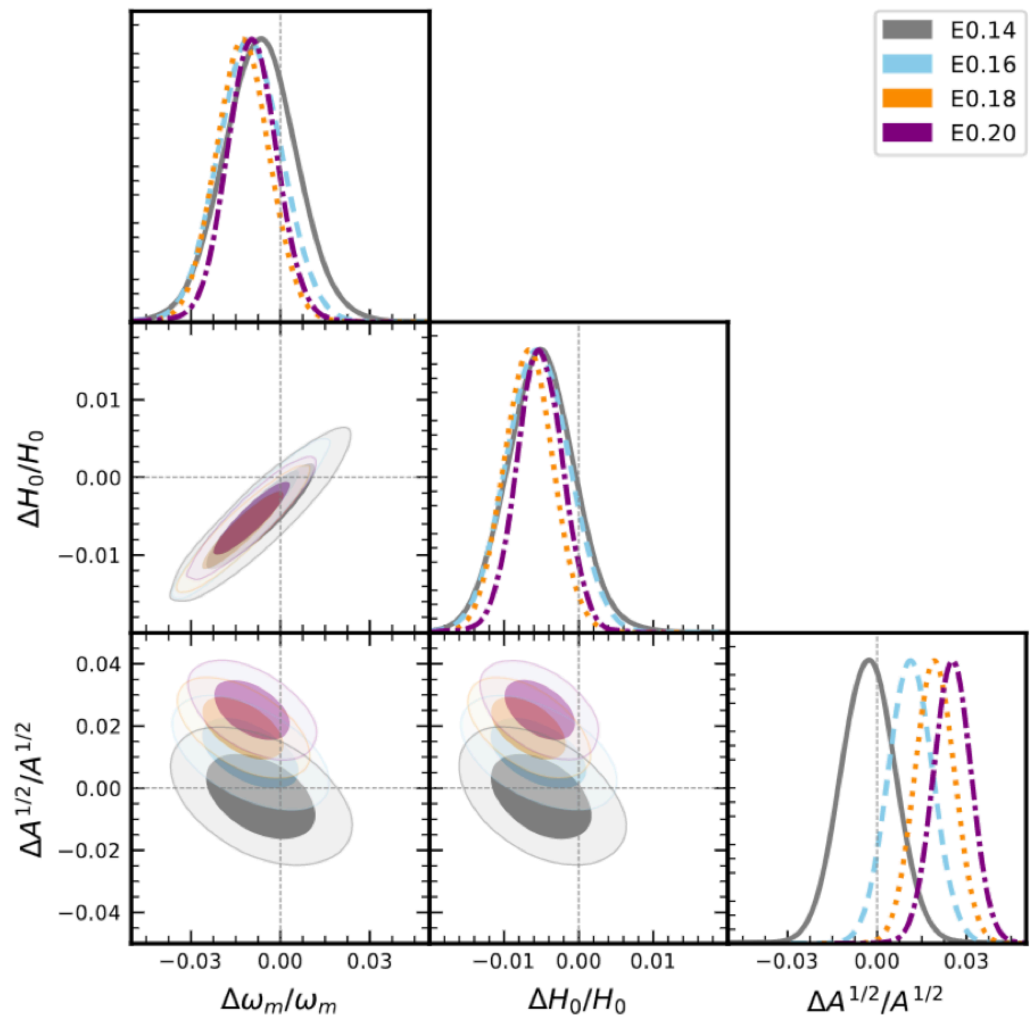
We need more research…

And in the end, we may throw out some data points

Scale Cuts

k_{max} largely determined by non-linear (bias, RSD) modeling accuracy

often determined by parameter drift tests (on mocks, if sufficiently accurate; on data, if allowed by blinding)



Nishimishi+21 (PT challenge)
Parameter constraints (East Coast team)
for different k_{max} (in *h*/Mpc)

On Systematics-limited Constraints

- Statistics-limited: parameter constraint $\sigma(\theta|\hat{\mathbf{d}}) \propto 1/\sqrt{V_s}$

  for fixed galaxy sample, model, summary statistic (+Gaussian likelihood?)

- Systematics may slow improvement in constraints:

$$p(\theta|\hat{\mathbf{d}}, k_{\min}, k_{\max}) \propto \int d\boldsymbol{\theta}_{\rm sys} p(\hat{\mathbf{d}}|\theta, \boldsymbol{\theta}_{\rm sys}, k_{\min}, k_{\max}) p(\boldsymbol{\theta}_{\rm sys})$$

scale cuts

often determined to limit parameter bias Δθ due to systematics/model misspecification

at fixed tolerance on Δθ/σ(θ), improved statistics partially compensated by more restrictive scale cuts

-> improve model (with as few additional parameters as possible)

prior-dominated systematics

-> improve calibration with external data/sims

good news: statistical power isn't everything - some reduction in volume/number density (e.g., easier-to-model galaxy sample, more regular mask) may be worth it!