Frequentist statistics & profile likelihoods

Tanvi Karwal

THE UNIVERSITY OF CHICAGO KICP

GGI workshop: Exploring New Frontiers in Cosmology

Parameter estimation in cosmology II

9th July 2026

Open notebook 2 now — we build on your sampler

1. Scan the QR → opens the repo
2. Open `ggi_write_profile_likelihood.ipynb` in Google Colab
3. It re-provides your notebook-1 functions, so it stands alone
4. Run the first cell to check your kernel

Frequentist method

When and how to use profiles

Build the profiler

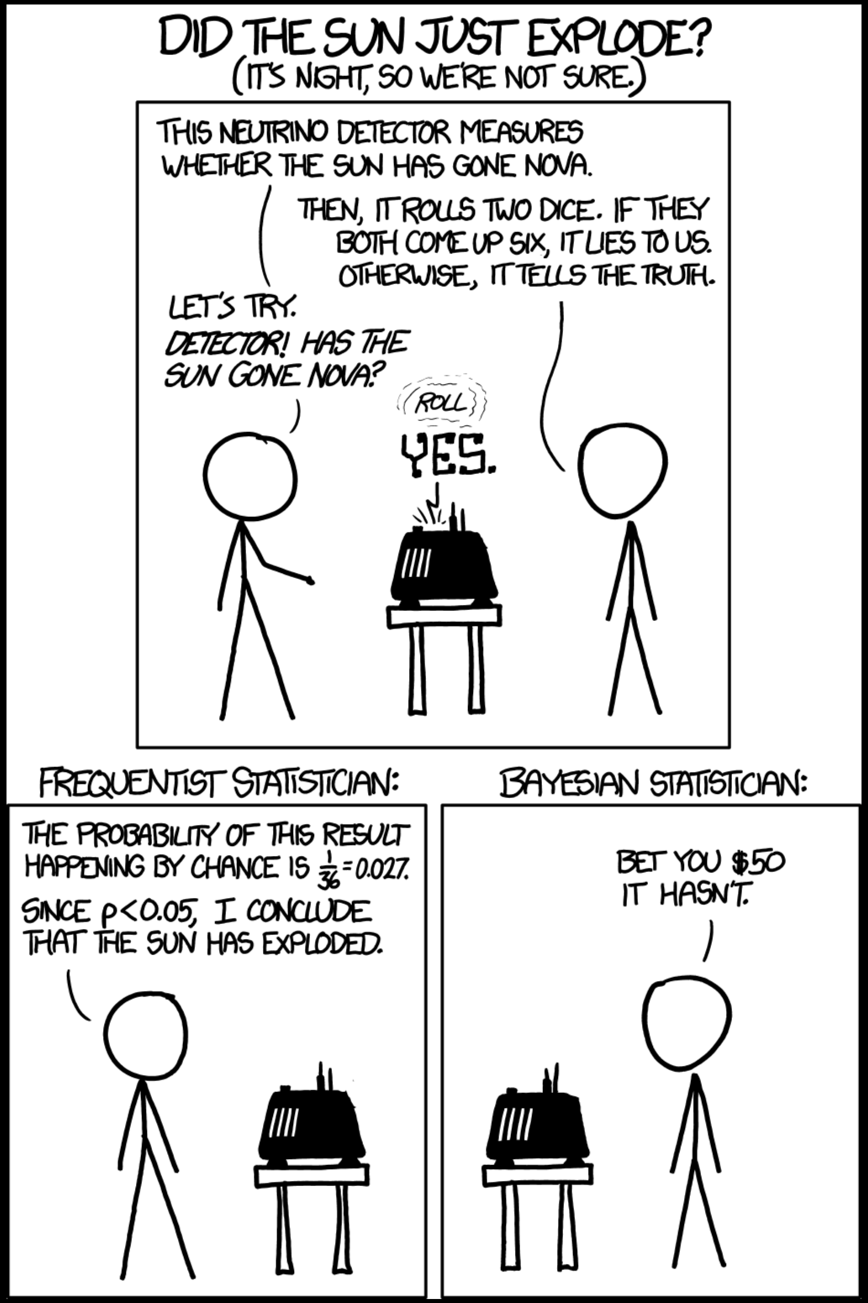
Profile gains

Frequentist method

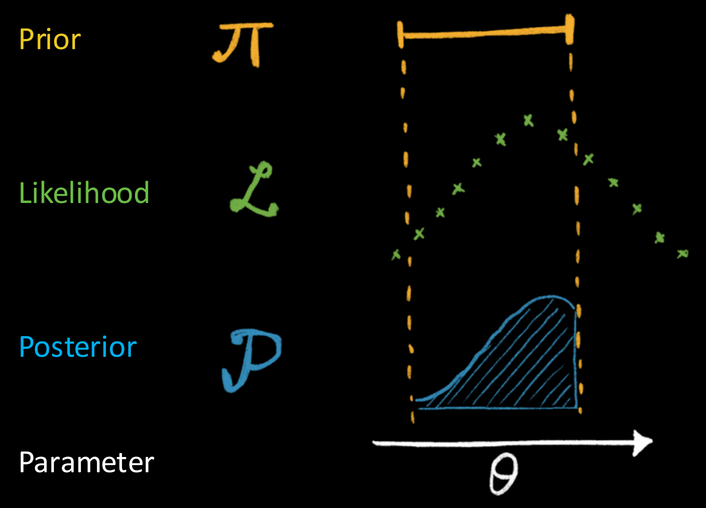
When and how to use profiles

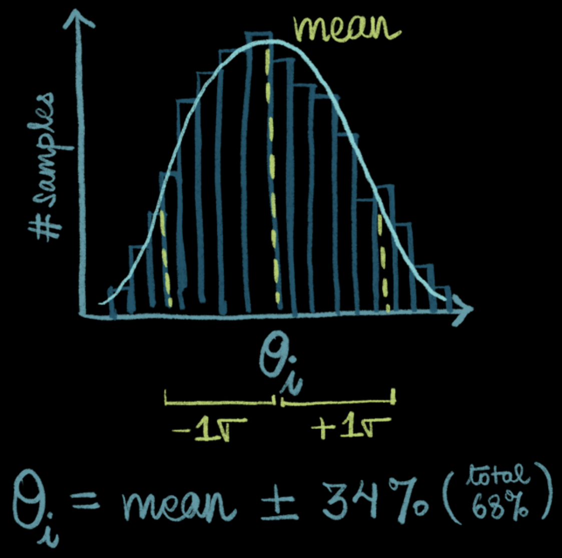
Build the profiler

Profile gains


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS $\frac{1}{36}$ = 0.027. SINCE $p < 0.05$, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

xkcd [1132]





Bayes theorem → **Sampling algorithm** → **Diagnostics** → **Constraints**

$$P(\theta|d) = \mathcal{L}(d|\theta)\Pi(\theta)$$

Prior dependence

Metropolis-Hastings MCMC

Existing infrastructure

Credible intervals from density of samples

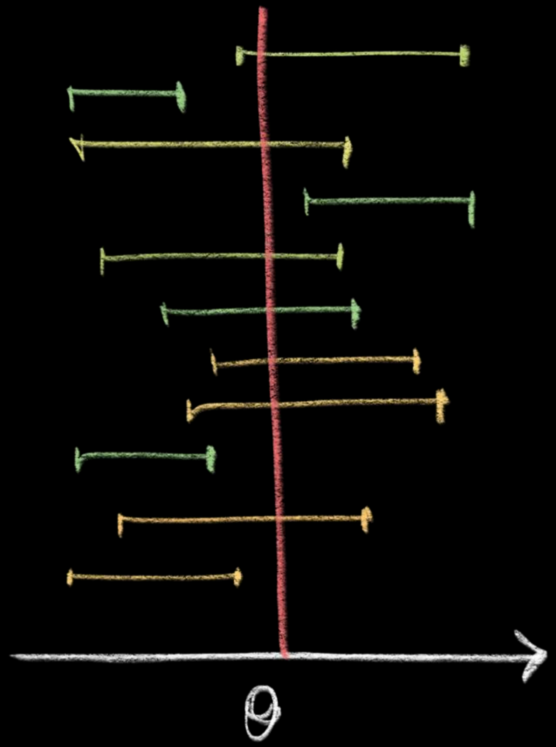
Statement about belief

Even with flat priors, marginalising can shift a constraint — prior-volume effects

Today: a tool with no prior at all. →

The frequentist philosophy
Probability = long-run frequency

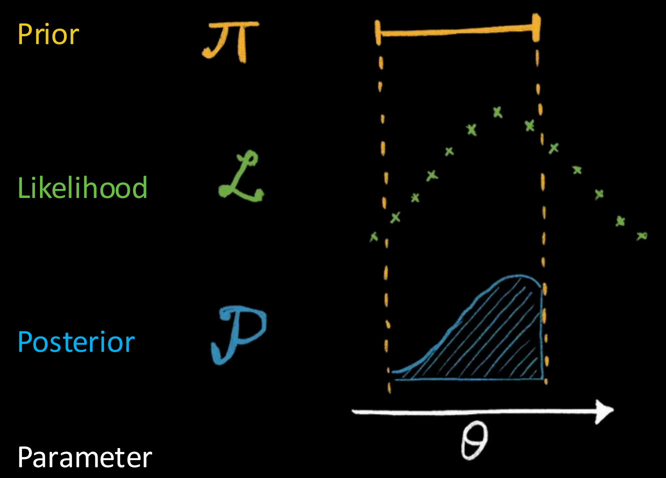
- The parameter is fixed but unknown. The data are random.
- We ask how estimates would behave over many repetitions of the experiment — not what we believe.

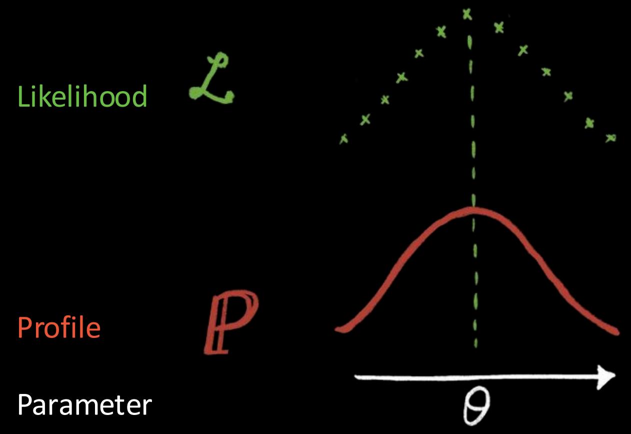


Bayesian: we quantify belief about the parameter.

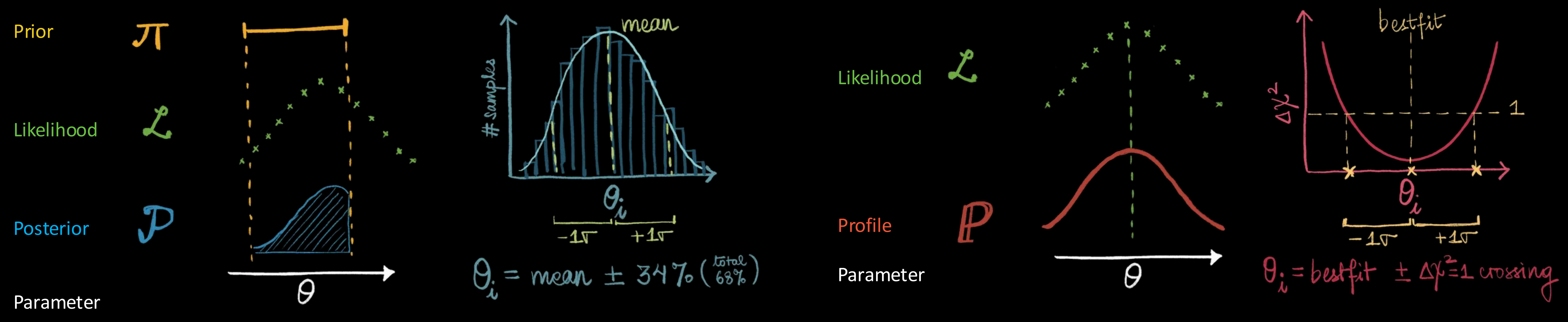
Frequentist: the parameter is fixed; the data are the random thing.

	Bayesian	Frequentist
Main object	$\mathcal{P}(\theta\|d) = \dfrac{\Pi(\theta)\,\mathcal{L}(d\|\theta)}{\mathcal{E}(d)}$	$\mathbb{P}(\theta_i) = \max \mathcal{L}(\boldsymbol{d}\|\boldsymbol{\theta})\|(\theta_i = \theta_i') \;\forall\; \theta_i{}'$
Role of prior	Central — must choose one	None required
Nuisance params	Marginalise (integrate)	Profile (maximise)
Interval meaning	Credible — degree of belief	Confidence — coverage of true parameter
Constraints from	Density of samples	Shape near the maximum





	Bayesian	Frequentist
Main object	$\mathcal{P}(\theta\|d) = \dfrac{\Pi(\theta)\,\mathcal{L}(d\|\theta)}{\mathcal{E}(d)}$	$\mathbb{P}(\theta_i) = \max \mathcal{L}(\boldsymbol{d}\|\boldsymbol{\theta})\|(\theta_i = \theta_i')\ \forall\ \theta_i{}'$
Role of prior	Central — must choose one	None required
Nuisance params	Marginalise (integrate)	Profile (maximise)
Interval meaning	Credible — degree of belief	Confidence — coverage of true parameter
Constraints from	Density of samples	Shape near the maximum



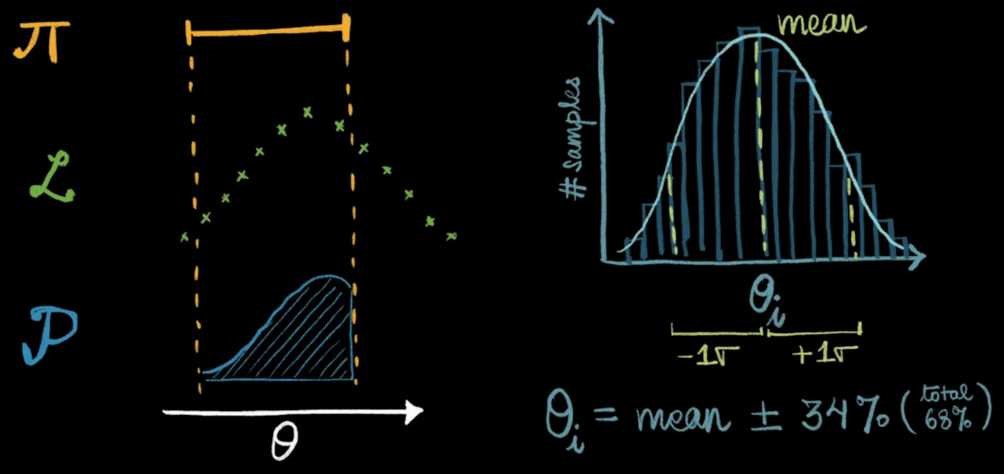
	Bayesian	Frequentist
Main object	$\mathcal{P}(\theta\|d) = \dfrac{\Pi(\theta)\ \mathcal{L}(d\|\theta)}{\mathcal{E}(d)}$	$\mathbb{P}(\theta_i) = \max \mathcal{L}(\boldsymbol{d}\|\boldsymbol{\theta})\|(\theta_i = \theta_i') \ \forall\ \theta_i{}'$
Role of prior	Central — must choose one	None required
Nuisance params	Marginalise (integrate)	Profile (maximise)
Interval meaning	Credible — degree of belief	Confidence — coverage of true parameter
Constraints from	Density of samples	Shape near the maximum

Prior

Likelihood

Posterior

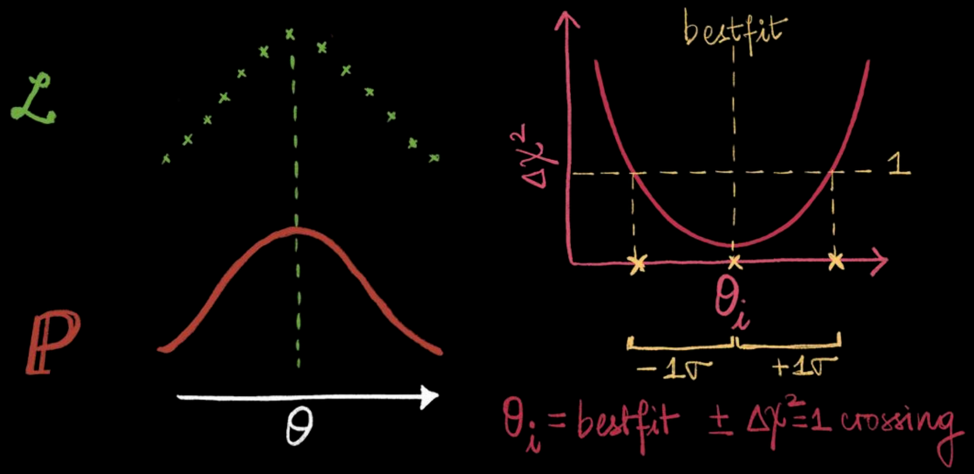
Parameter



Likelihood

Profile

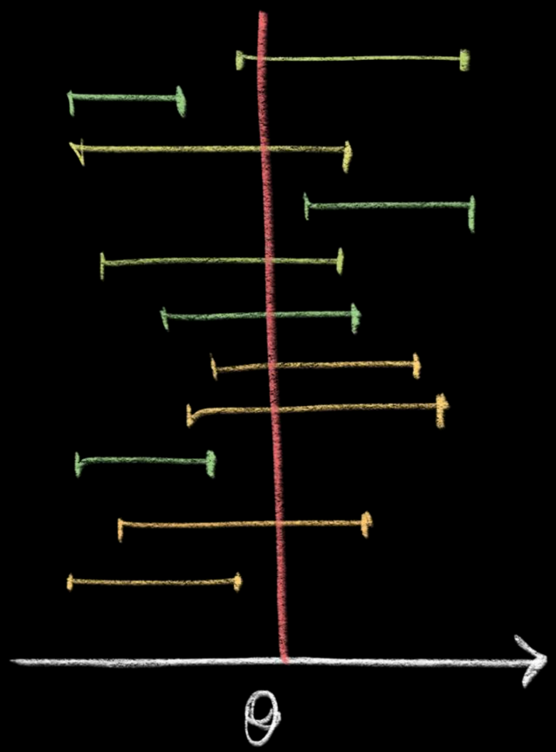
Parameter



What “coverage” means

A 68% confidence interval is a procedure that, repeated over many datasets, brackets the true value 68% of the time.

It is a property of the method, not of this one interval.



Frequentist method

When and how to use profiles

Build the profiler

Profile gains

The likelihood is prior-independent

- Frequentist inference uses only $\mathcal{L}$ — no prior to choose
- It is reparametrisation-invariant: A_s vs $\ln(10^{10}A_s)$, θ_* vs H_0 — same answer

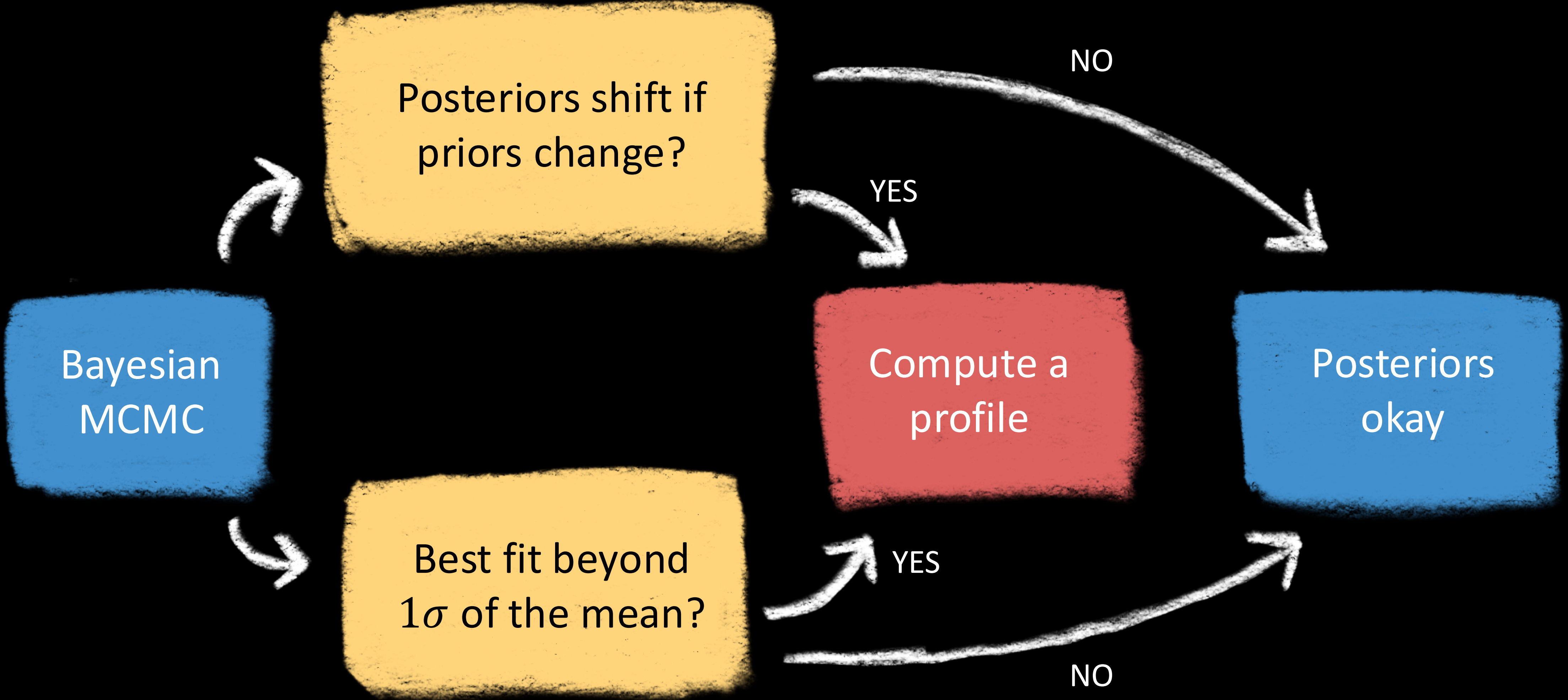
So it is insensitive to the things that can wobble a posterior

Profile Likelihoods in Cosmology:
When, Why and How illustrated with ΛCDM, Massive Neutrinos and Dark Energy

Laura Herold,[1,*] Elisa G. M. Ferreira,[2] and Lukas Heinrich[3]

[2408.07700]

When to compute a profile



Prior-volume effects

- Marginalising integrates over nuisance volume
- Values that open up more good-fit room collect more posterior mass

This behaves like an "effective prior" that we never wrote down

Maximising (profiling) only sees the best fit — regardless of the width of the good-fit region

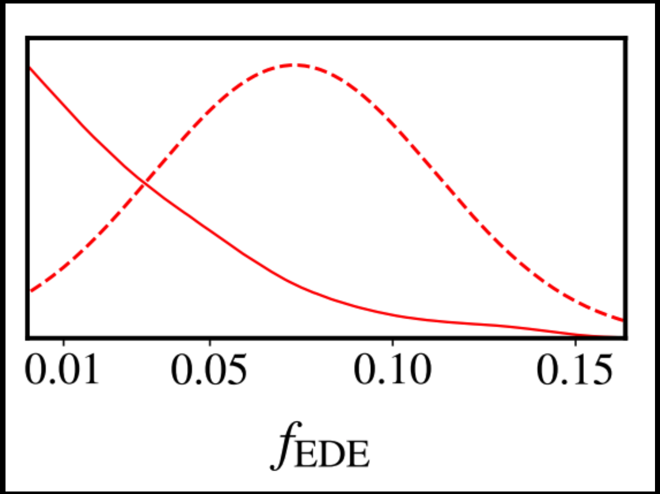
Early dark energy

f_{ede} - how much EDE
z_c - when EDE appears
w_f - how fast is disappears

—— Bayesian posteriors
- - - - Likelihood profile

Prior volume increases as $f_{ede} \rightarrow 0$

z_c and w_f become unconstrained



EDE review
Poulin, Smith and TK [2302.09032]

Prior-volume effects

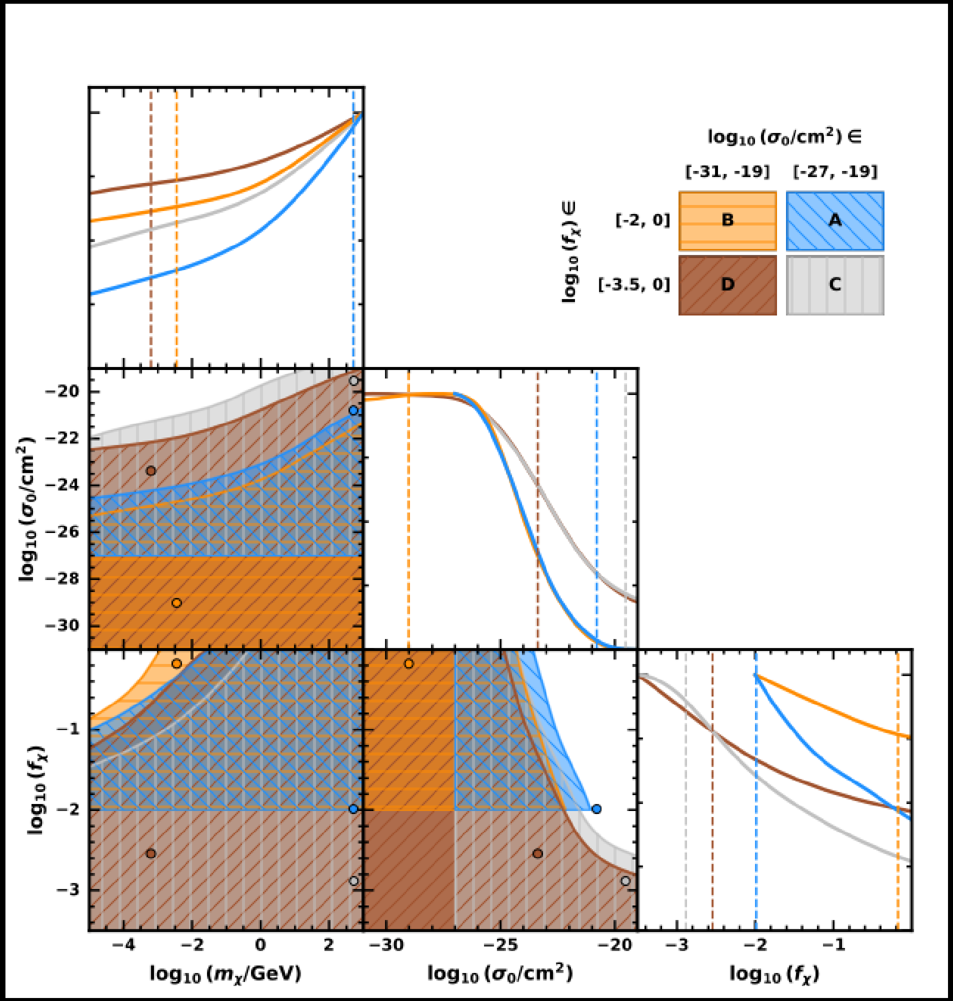
- Marginalising integrates over nuisance volume
- Values that open up more good-fit room collect more posterior mass

This behaves like an "effective prior" that we never wrote down

Maximising (profiling) only sees the best fit — regardless of the width of the good-fit region

Dark matter scattering with protons

Straight, TK et al [2603.25731]



- Dark matter mass m_χ
- Interaction strength σ_0
- Interacting fraction f_χ

2D prior-volume effects as $\sigma_0 \to 0$ or $f_\chi \to 0$

Prior-volume effects

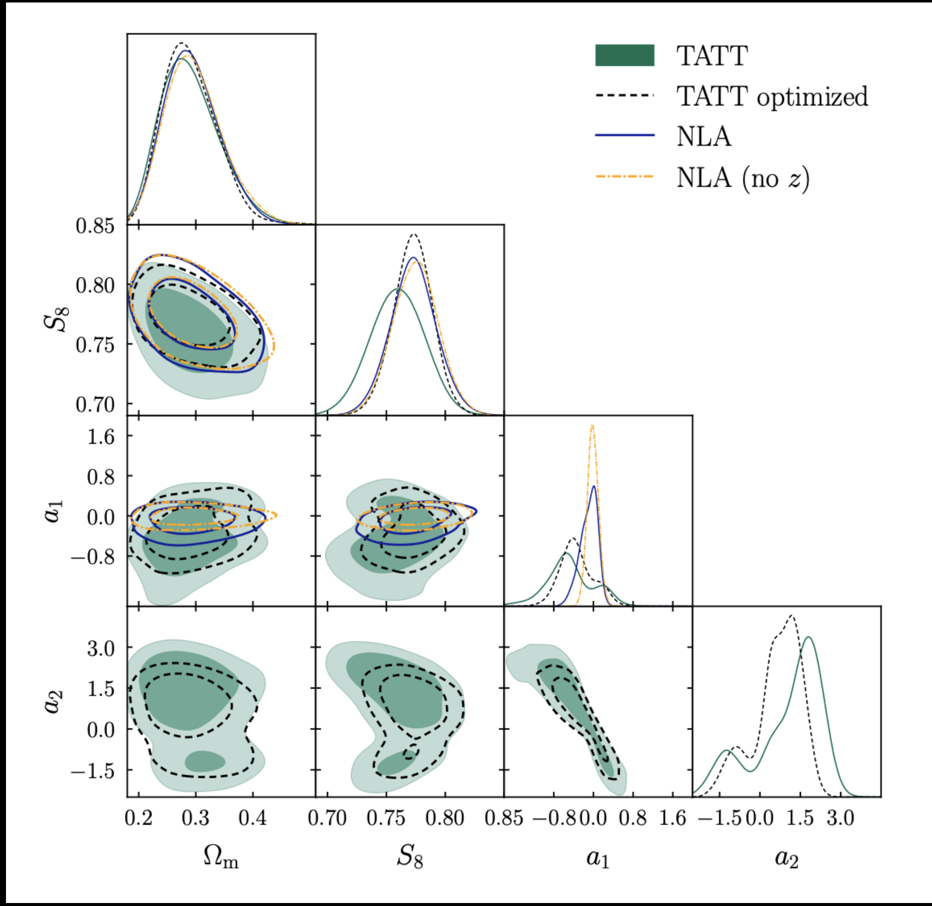
- Marginalising integrates over nuisance volume
- Values that open up more good-fit room collect more posterior mass

This behaves like an "effective prior" that we never wrote down

Maximising (profiling) only sees the best fit — regardless of the width of the good-fit region

Galaxy weak lensing nuisance parameters

DES Y3 Secco et al [2105.13544]



TATT is nested within the NLA model for intrinsic alignment. The likelihood peaks in the NLA limit but TATT parameters add prior volume and shift posteriors

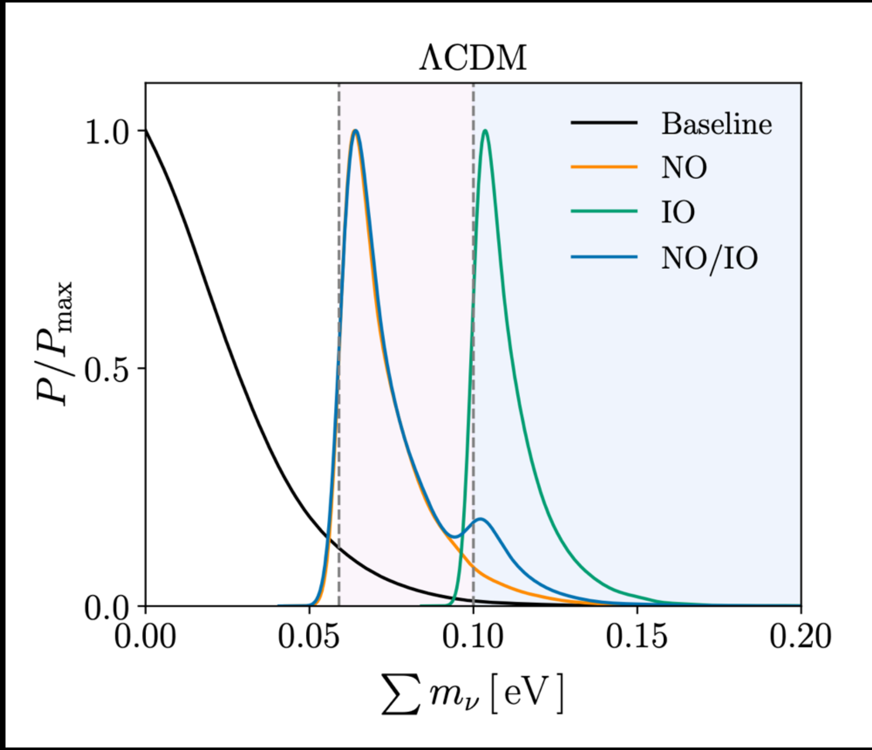
Boundaries & unconstrained direction

Parameters near a physical boundary eg.

$$\Sigma m_\nu \geq 0, \quad r \geq 0, \quad f_{ede} \to 0, \quad \sigma_0 \to 0$$

or flat likelihood directions where data don't constrain

These are where posteriors get most prior-sensitive — and where profiles help most.



DESI DR2 neutrinos

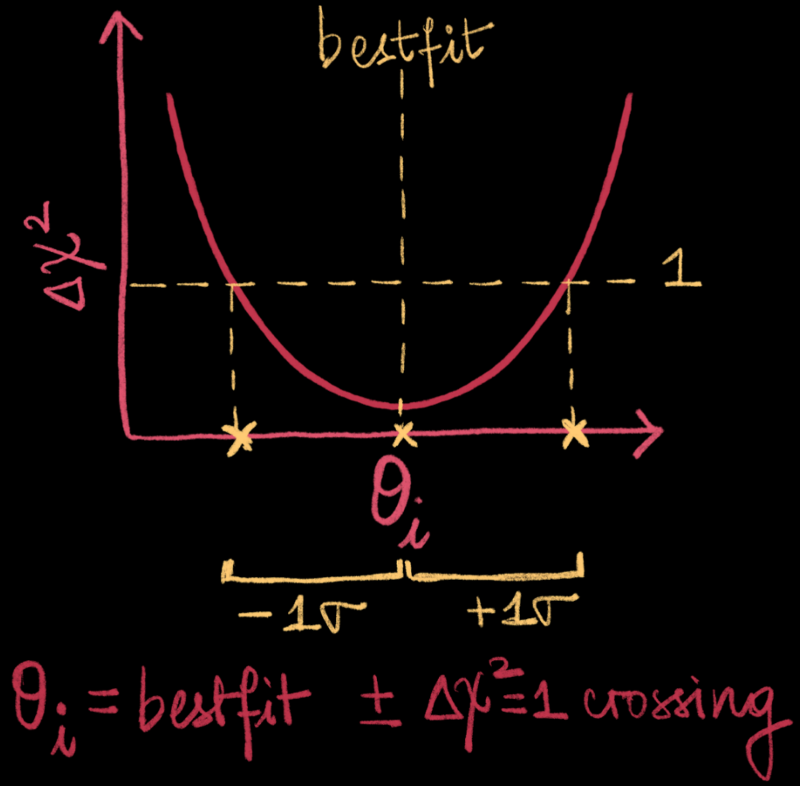
Profile likelihood caveats I
Wilks' theorem – graphical construction

$$\chi^2 \equiv -2\ln\mathcal{L}$$

Lower χ^2 = better fit and only differences matter $\Delta\chi^2 = -2\,\Delta\ln\mathcal{L}$

Graphical $\Delta\chi^2$ construction assumes Wilks / a Gaussian limit

- In the asymptotic (large-data) limit, $-2\Delta\ln\mathcal{L}$ follows a χ^2 distribution
- Confidence from $\Delta\chi^2$ thresholds — the “graphical” method
  $\Delta\chi^2 = \;1\;(1\sigma, 1D)\;,\quad 4\;(2\sigma, 1D)\;,\quad 6.18\;(2\sigma, 2D)\;\ldots$
- Needs a Gaussian likelihood for all parameter values & hypothetical data — not just the observed data



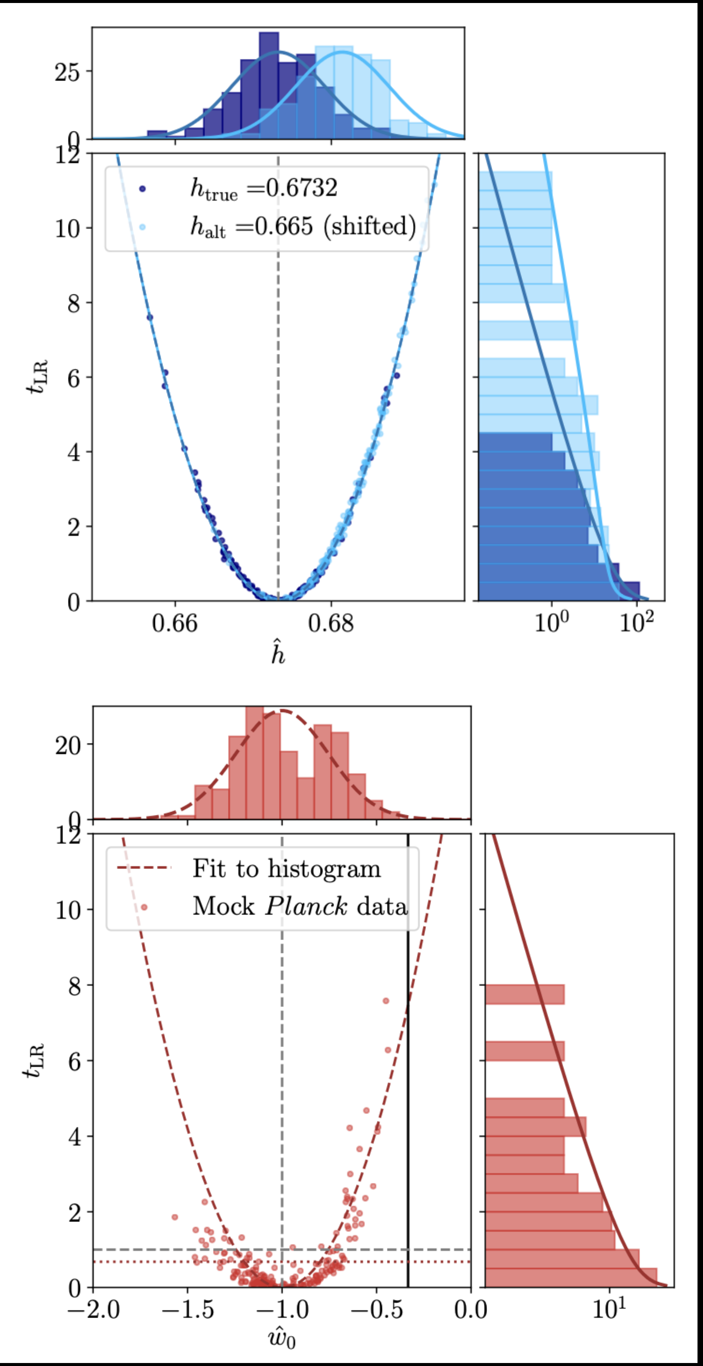
Profile likelihood caveats I
Wilks' theorem – graphical construction

$$\chi^2 \equiv -2 \ln \mathcal{L}$$

Lower χ^2 = better fit and only differences matter $\Delta\chi^2 = -2\,\Delta\ln\mathcal{L}$

Graphical $\Delta\chi^2$ construction assumes Wilks / a Gaussian limit

- In the asymptotic (large-data) limit, $-2\Delta\ln\mathcal{L}$ follows a χ^2 distribution
- Confidence from $\Delta\chi^2$ thresholds — the "graphical" method
  $\Delta\chi^2 = \; 1 \;(1\sigma, 1D)\;, \quad 4\;(2\sigma, 1D)\;, \quad 6.18\;(2\sigma, 2D)\;\;\ldots$
- Needs a Gaussian likelihood for all parameter values & hypothetical data — not just the observed data



Herold et al [2408.07700]

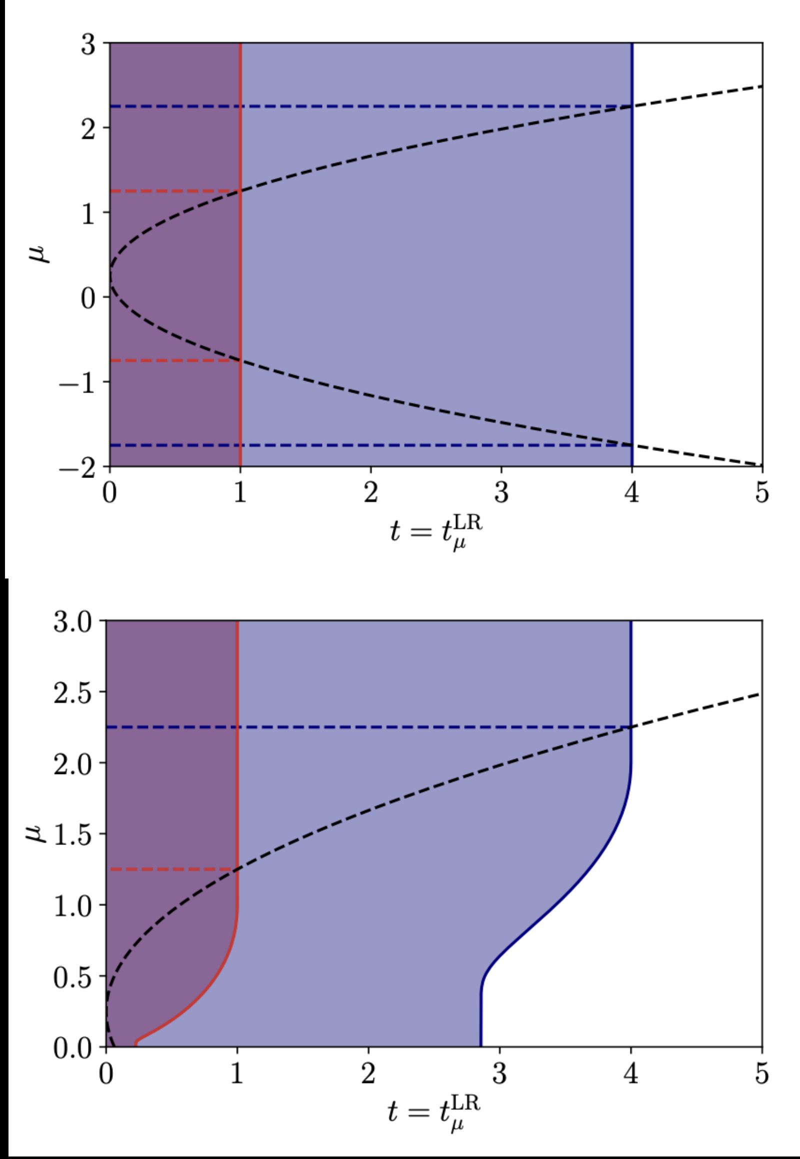
Profile likelihood caveats II
Feldman–Cousins correction

Near boundaries you need corrections to the graphical construction

$$\Sigma m_\nu \geq 0, \qquad r \geq 0, \qquad f_{ede} \to 0, \qquad \sigma_0 \to 0$$

- A naïve $\Delta\chi^2 = 1, 4, \ldots$ interval can under/over-cover or go unphysical
- Feldman–Cousins [9711021]: an ordering rule for building the interval

Guarantees physical, correctly-covering intervals and a smooth transition between upper limits and two-sided intervals.



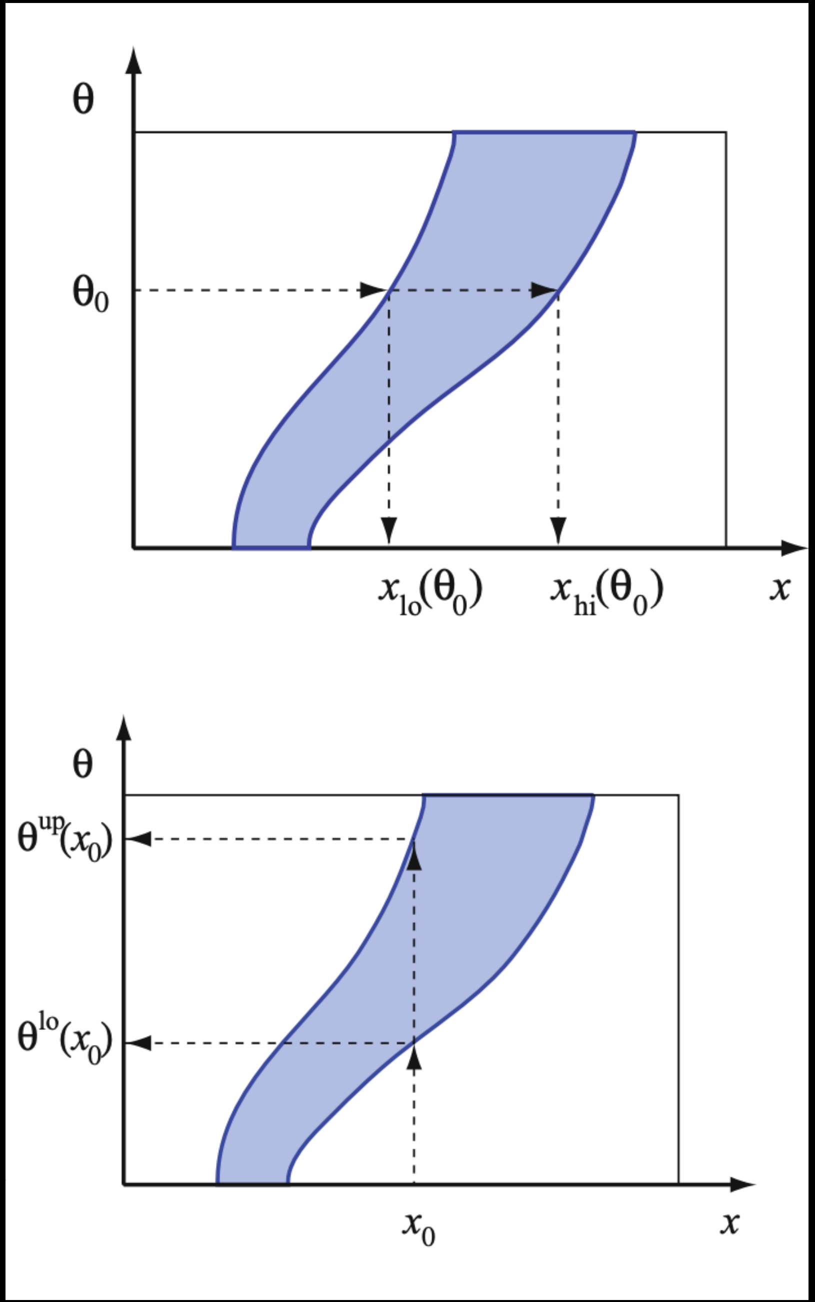
Herold et al [2408.07700]

Profile likelihood caveats III
Full Neyman construction

The gold standard: guarantees correct coverage under all conditions but very expensive

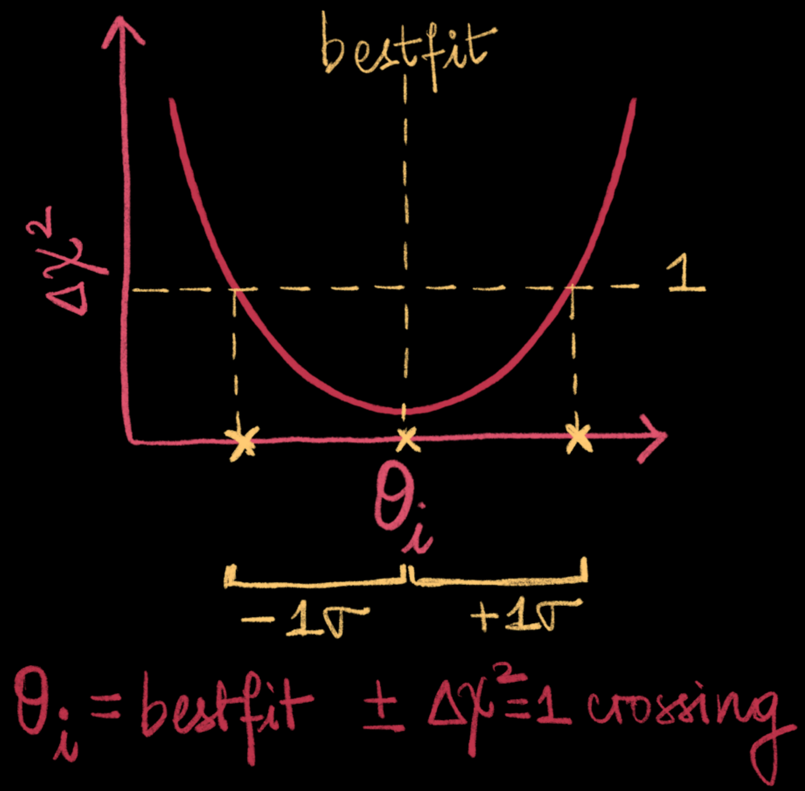
- For each candidate θ, simulate many mock datasets x
- Find the acceptance region per θ
- Then invert

The graphical profile method is the cheap approximation to this in the Gaussian limit



Profile likelihood caveats

- Graphical $\Delta\chi^2$ construction assumes Wilks / a Gaussian limit
- Near boundaries you need Feldman–Cousins corrections
- Full Neyman construction is correct but very expensive
- Profiling can "fine-tune" into tiny-volume solutions (it ignores volume), making it important to complement with Bayesian constraints



Frequentist method

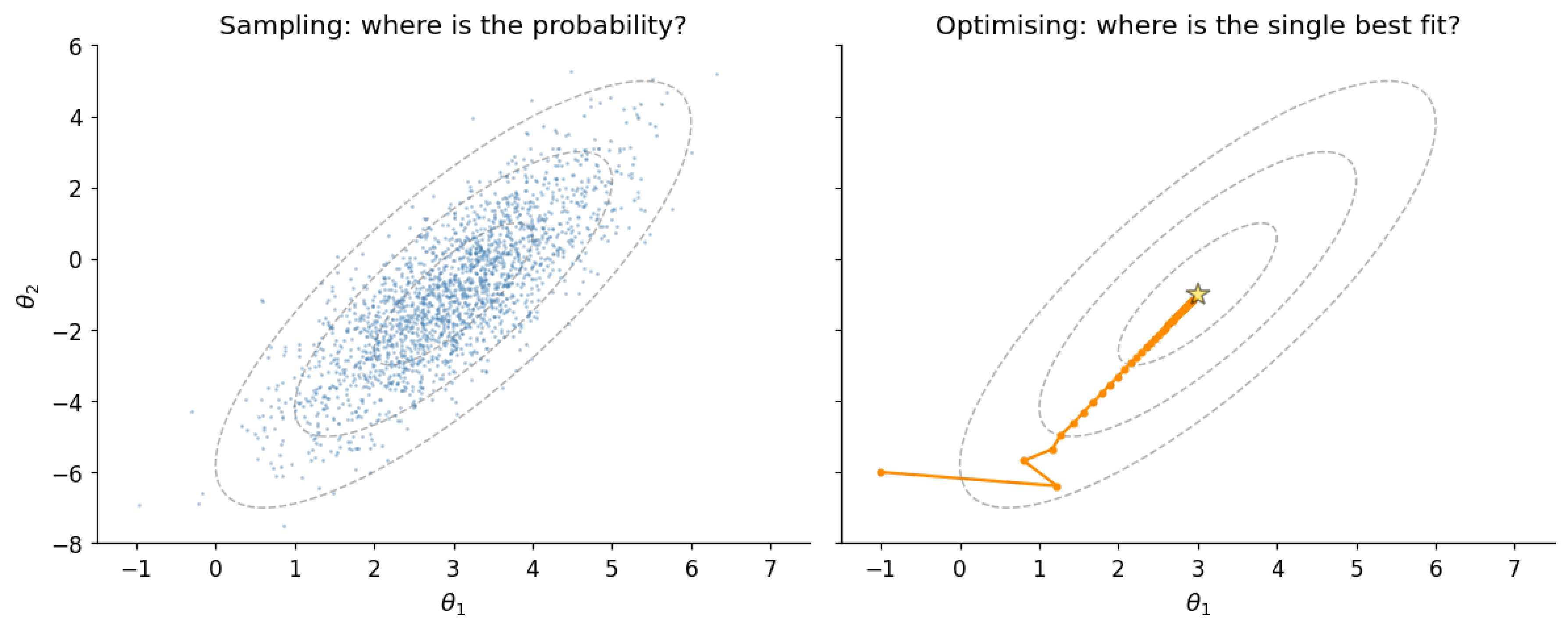
When and how to use profiles

Build the profiler

Profile gains

Optimisation is not sampling

Run Section 1



Sampling maps where the probability is

Optimisation finds the single best point and the shape of $\mathcal{L}$ around it

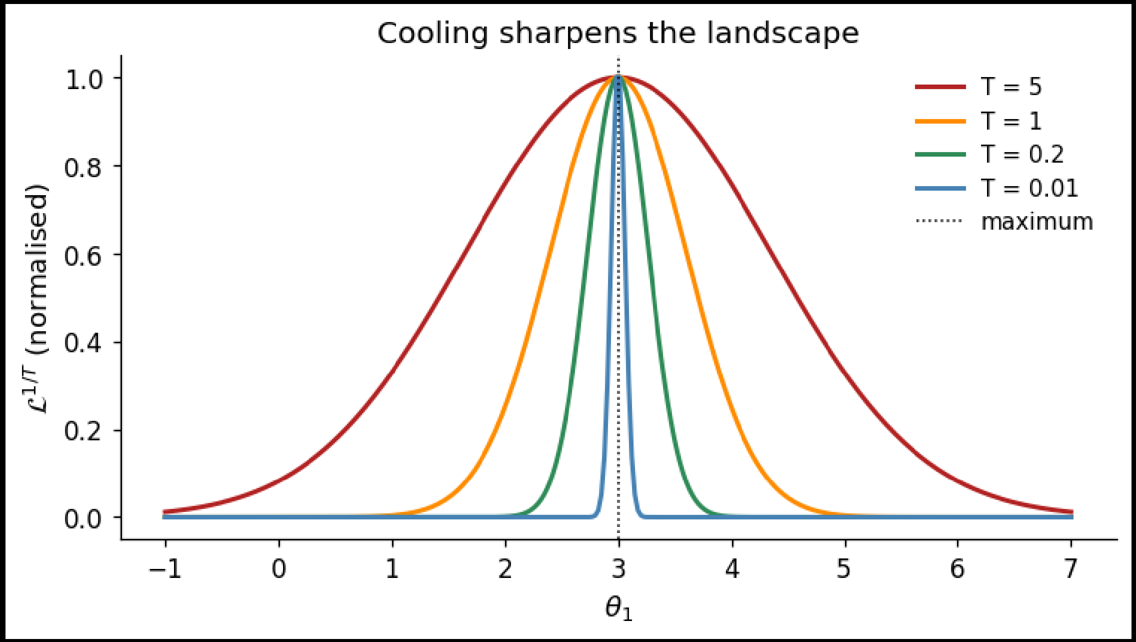
Frequentist constraints come from that shape, not from mass

The temperature knob

Run Section 2

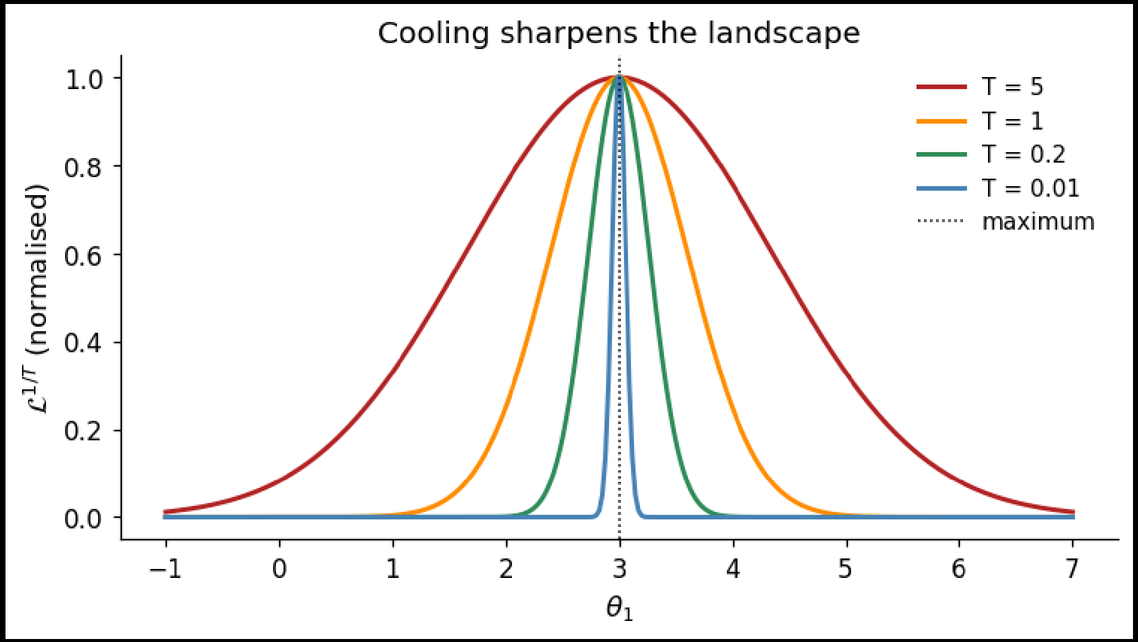
- $T > 1$ flattens the landscape → wide exploration
- $T = 1$ is ordinary Metropolis–Hastings (notebook 1)
- $T < 1$ sharpens → only good moves survive → hone in
- $T \to 0$ accepts only improvements → greedy hill-climber

This is the only code change from Lecture 1's sampler



The temperature knob turns the sampler into an optimiser

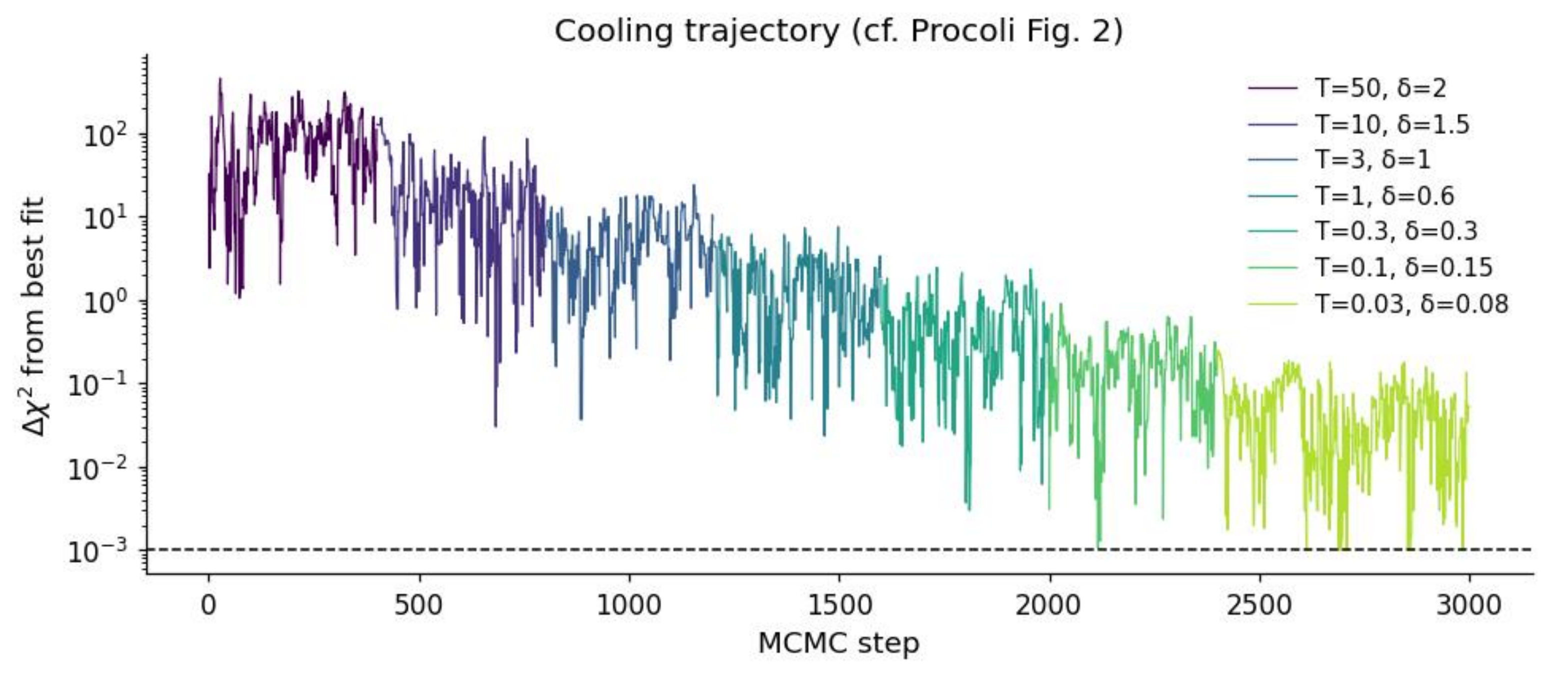
```
1.  start at  θ_current
2.  propose   θ*  ~  Normal(θ_current, step)
3.  log_α = 1/T*( log_post(θ*) − log_post(θ_current) )
4.  draw  u ~ Uniform(0,1)
5.  if  log(u) < log_α :
        ACCEPT  → θ_current = θ*
6. record just the best fit θ and χ²
```

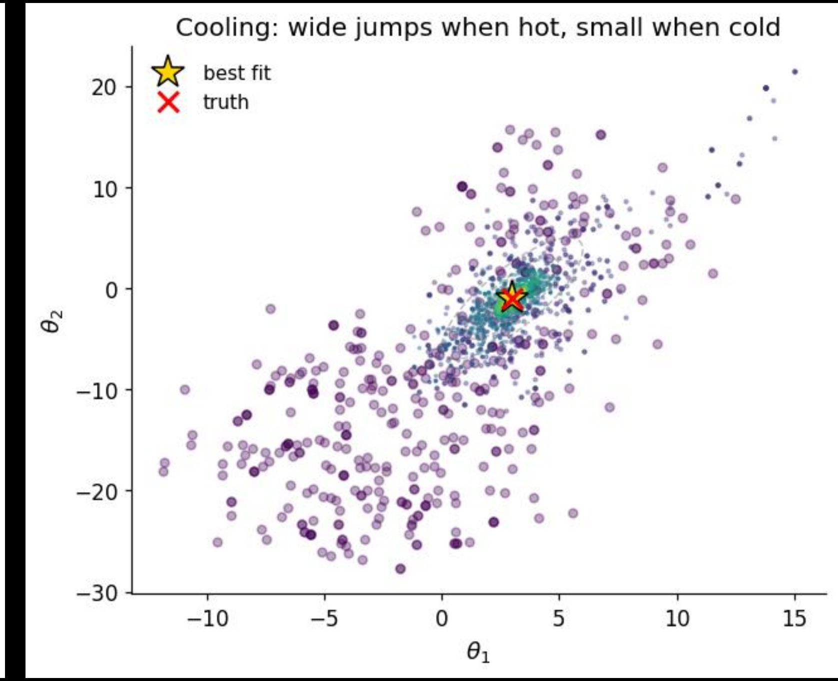


Build the optimiser

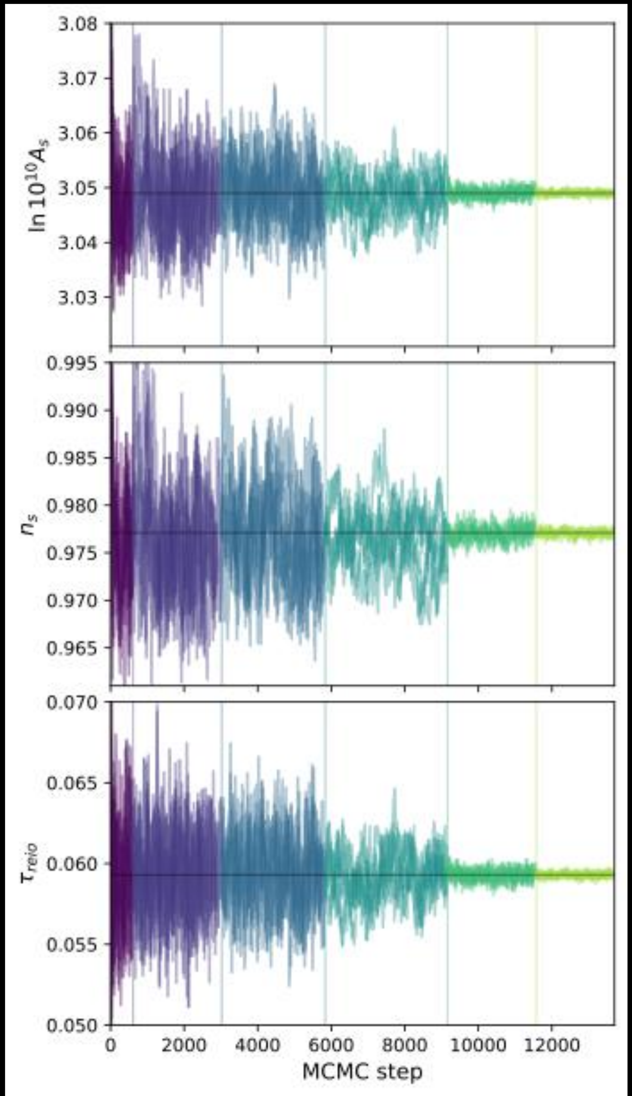
Do Sections 3–5

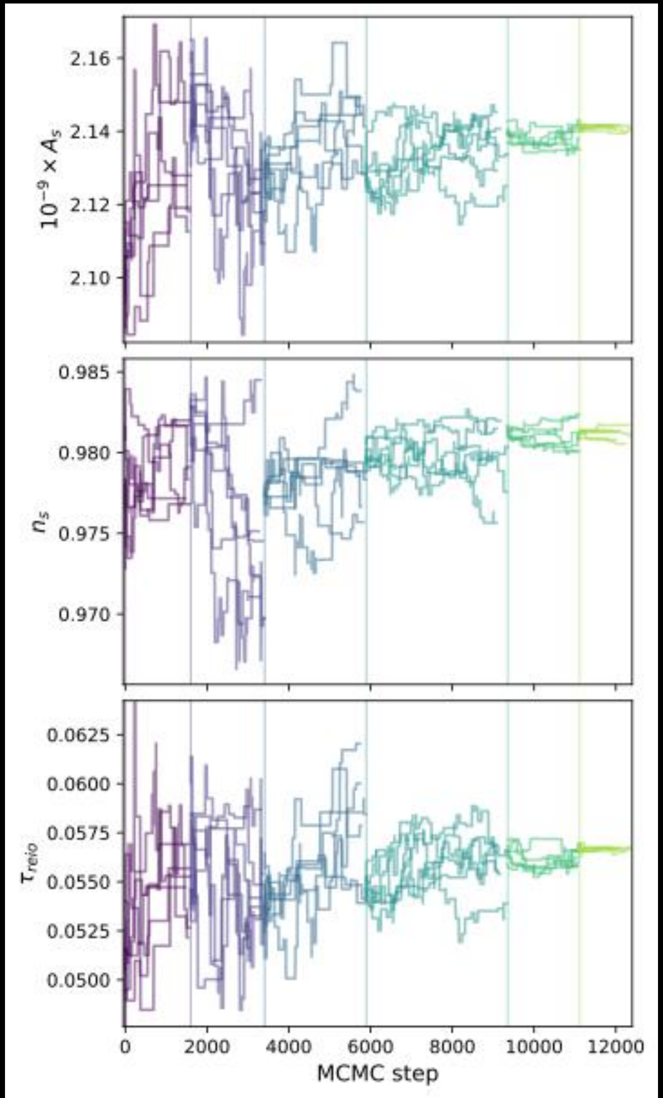
Produce the cooling trajectory in 1 and 2D





Similar optimisation with mock and real data

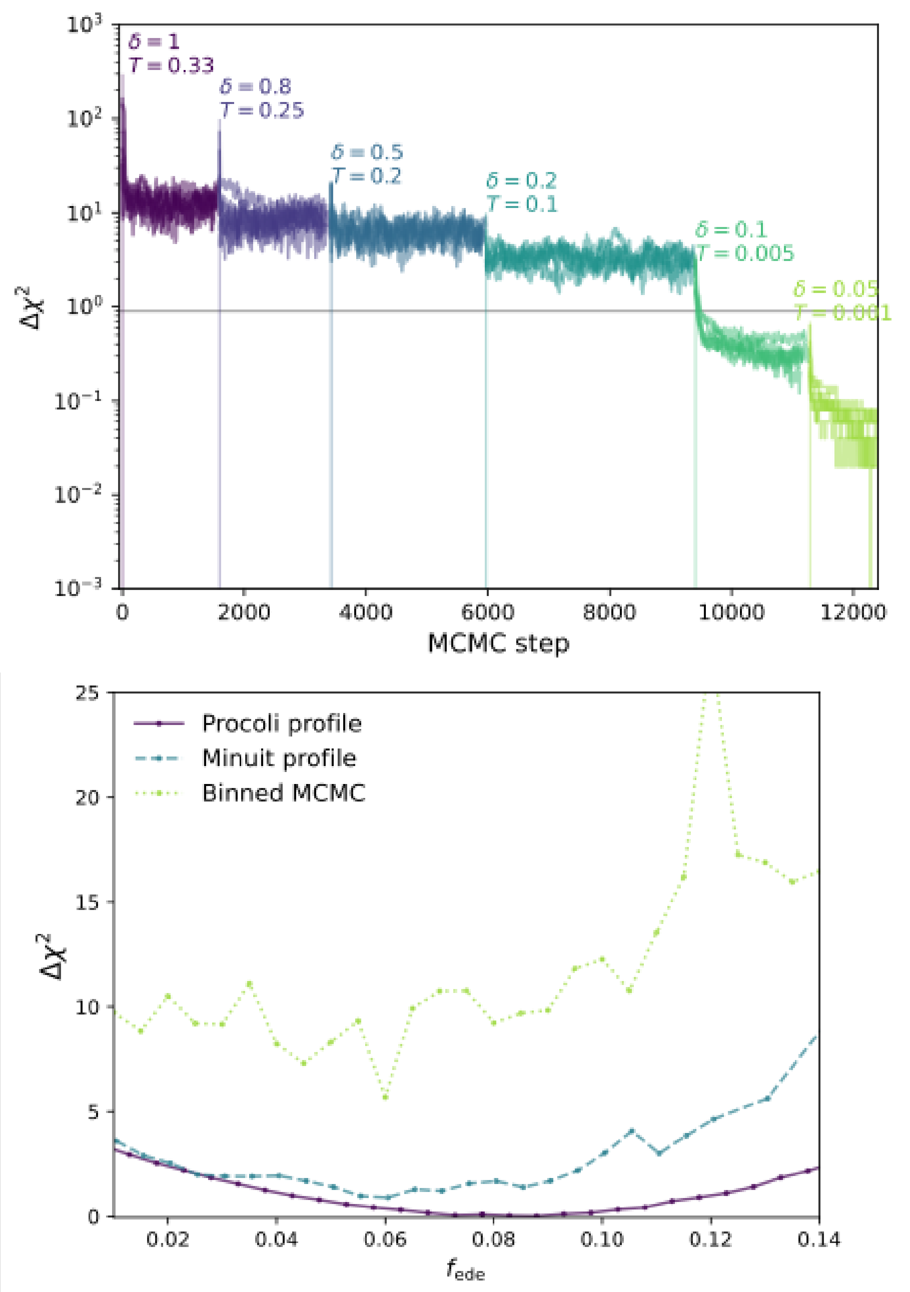




Procoli: TK et al [2401.14225]

Stochastic vs deterministic optimisers

- Cosmological likelihoods are high-dimensional, noisy, and can be multimodal
- Gradient / quasi-Newton methods stall on noise and fall into the nearest local minimum
- A hot stochastic process crosses barriers to the true, global optimum



Procoli: TK et al [2401.14225]

Cooling too fast → stuck in a local minimum

Real likelihoods can have several minima (e.g. EDE z_c bimodality)

- Cool too quickly from a far start → freeze in whichever basin you began in
- Spend time hot first → cross the barrier to the true minimum

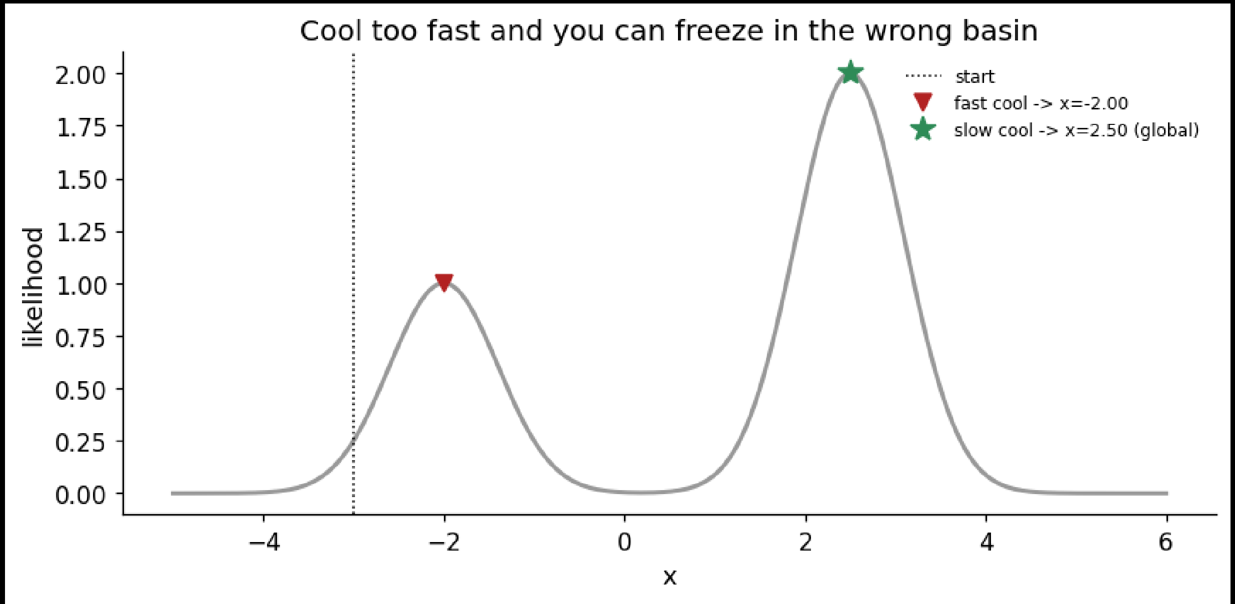
Do Section 6

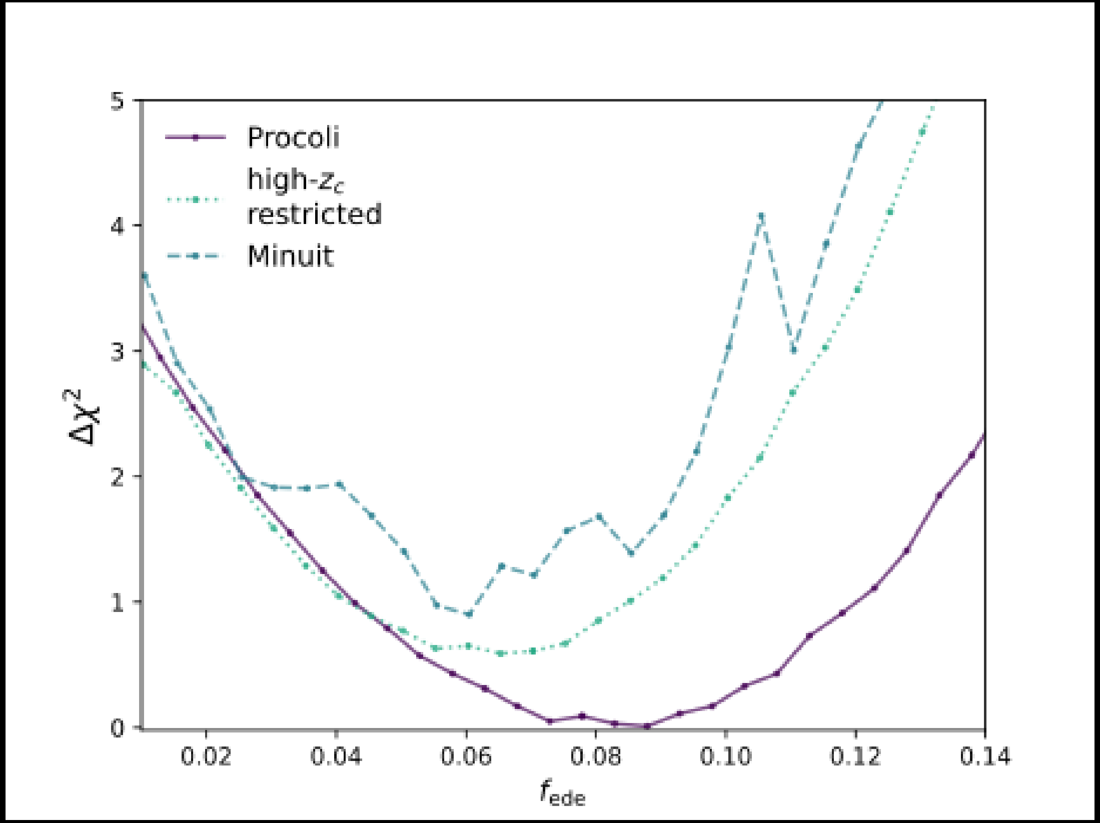
Cooling too fast → stuck in a local minimum

Real likelihoods can have several minima (e.g. EDE z_c bimodality)

- Cool too quickly from a far start → freeze in whichever basin you began in
- Spend time hot first → cross the barrier to the true minimum

Do Section 6





TK et al [2401.14225]

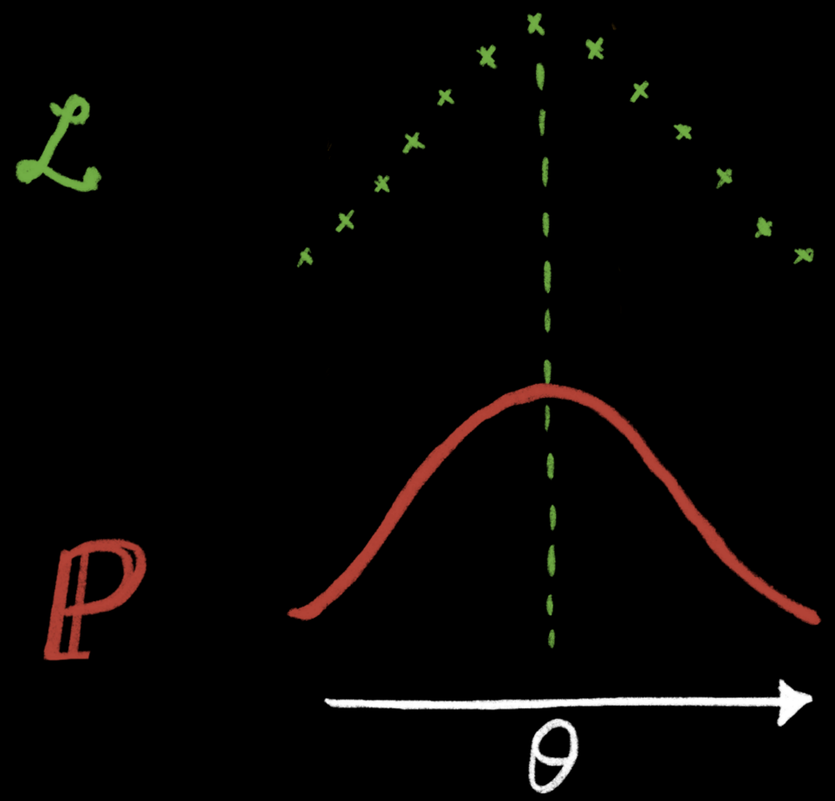
A simple profiling recipe

For the parameter of interest θ_i

1. Fix θ_i at a grid value
2. Optimise all other parameters at that fixed θ_i
3. Record the best-fit χ^2
4. Step to the next grid value, repeat

The curve ${\chi_{\min}}^2(\theta_i)$ is the profile. Read intervals at $\Delta\chi^2 = 1, 4, \ldots$

Do Section 7

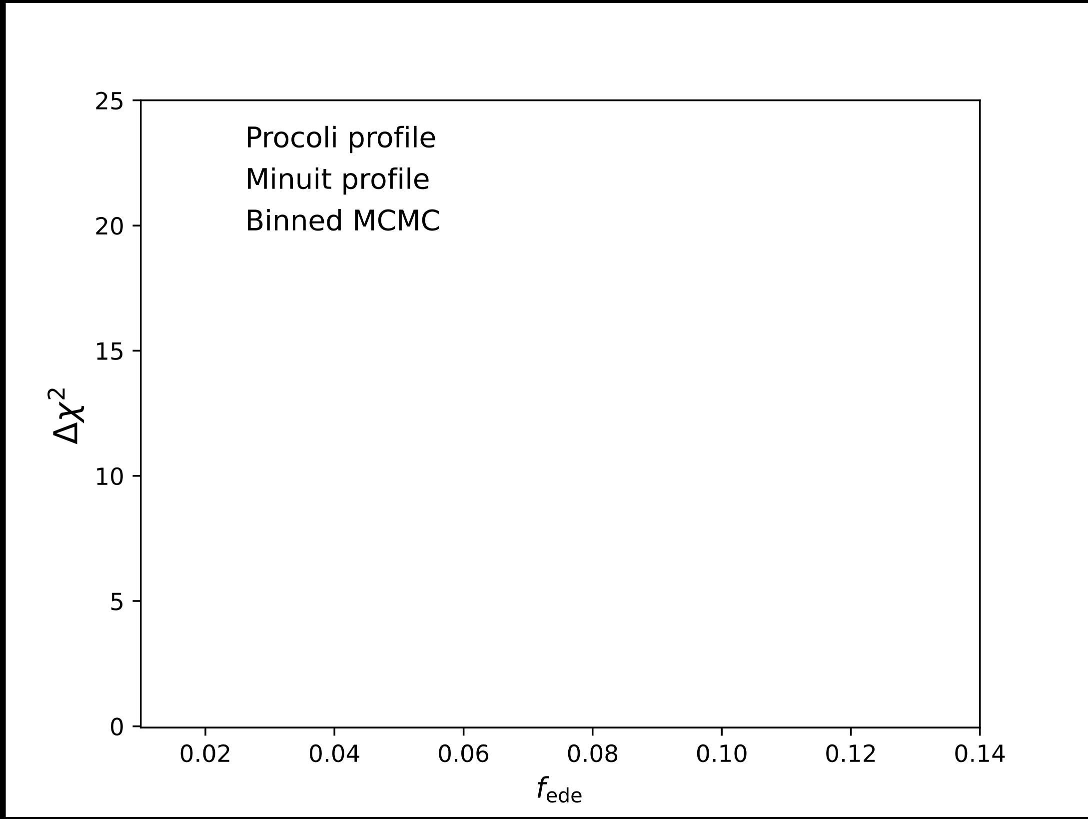


Sequential profiling — why it's efficient

- Neighbouring grid points share almost the same best fit
- Start each optimisation from the previous point's answer
- Cheap after the first rigorous (global) minimisation

This is Procoli's core efficiency trick, it builds a sequential profile

- Start with a robust global minimisation
- Follow with faster, more efficient profile



TK et al [2401.14225]

Frequentist method

When and how to use profiles

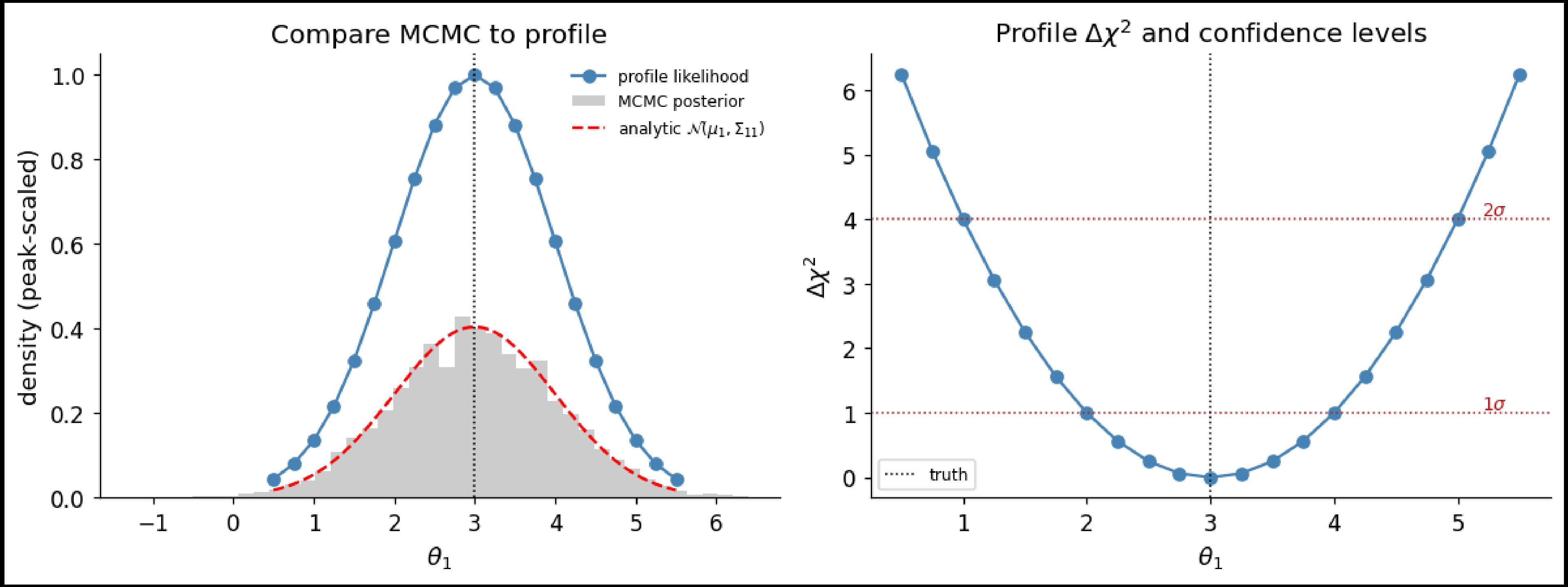
Build the profiler

Profile gains

Example 1 — flat priors, profile = posterior

Section 8 – The sanity check

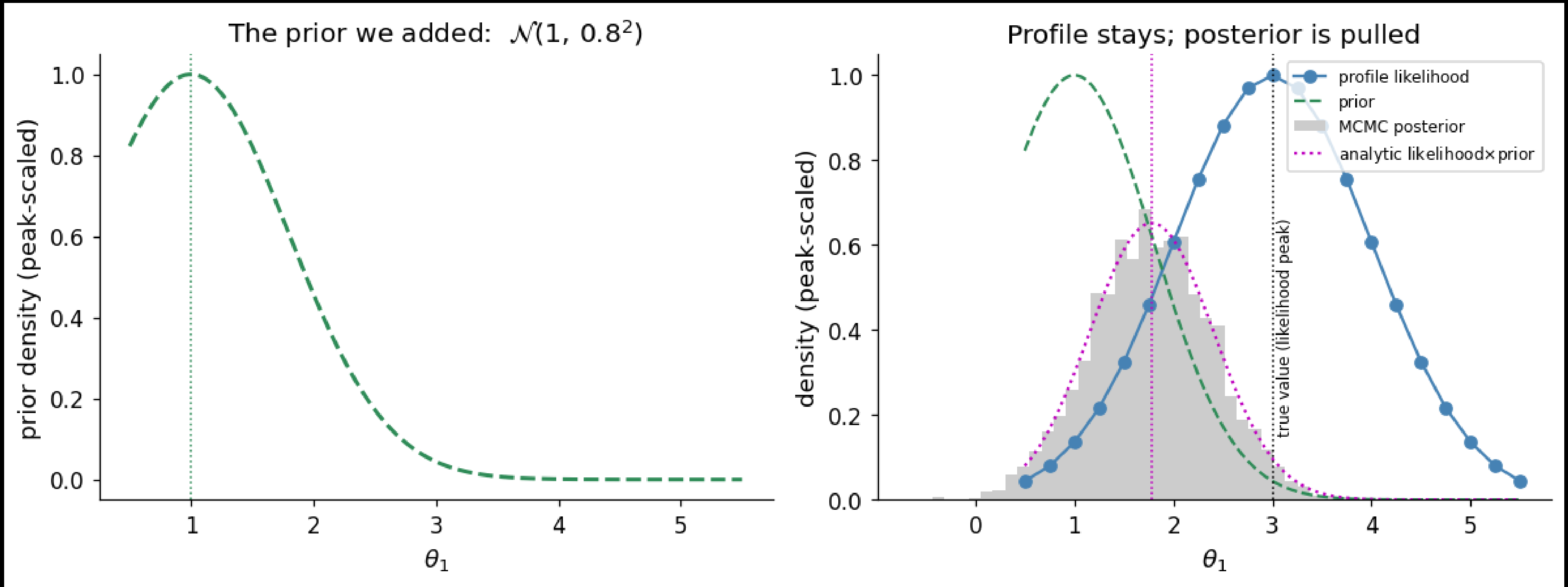
With flat priors: profile likelihood, MCMC posterior, and the analytic Gaussian should all coincide



Example 2 — an informative prior pulls the posterior

Same likelihood but we add a Gaussian prior on θ_1

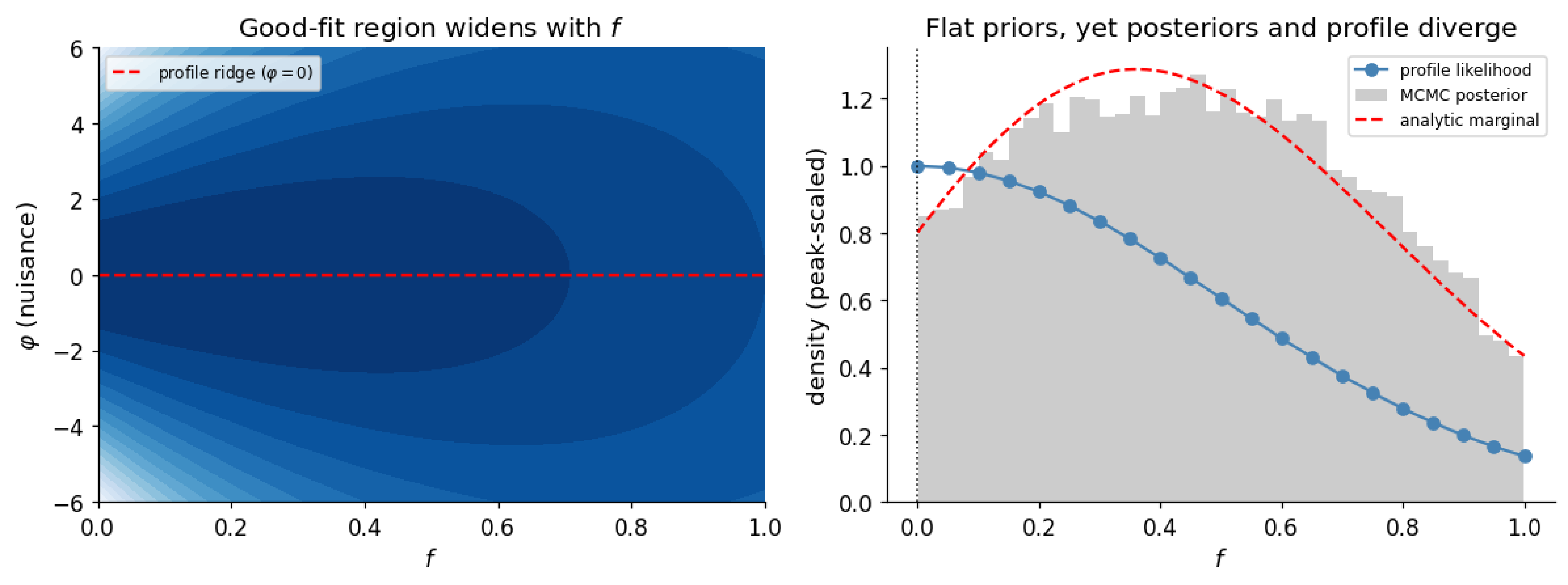
- The posterior peak shifts toward the prior
- The profile does not move — it follows the likelihood and ignores the prior



Example 3

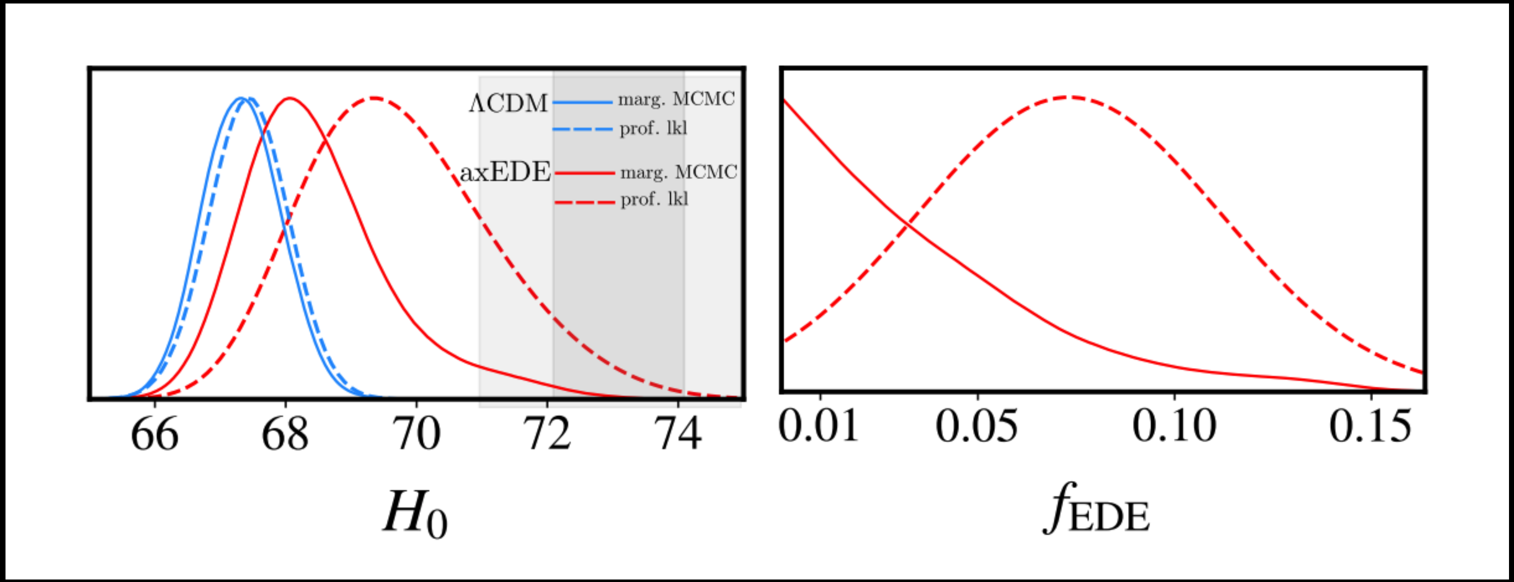
Example 3 — flat priors, hidden prior-volume effect

- A nuisance direction whose good-fit width grows with θ_i
- Maximising ignores the width but integrating collects the extra volume
- Posterior disagrees with the profile — even with flat priors

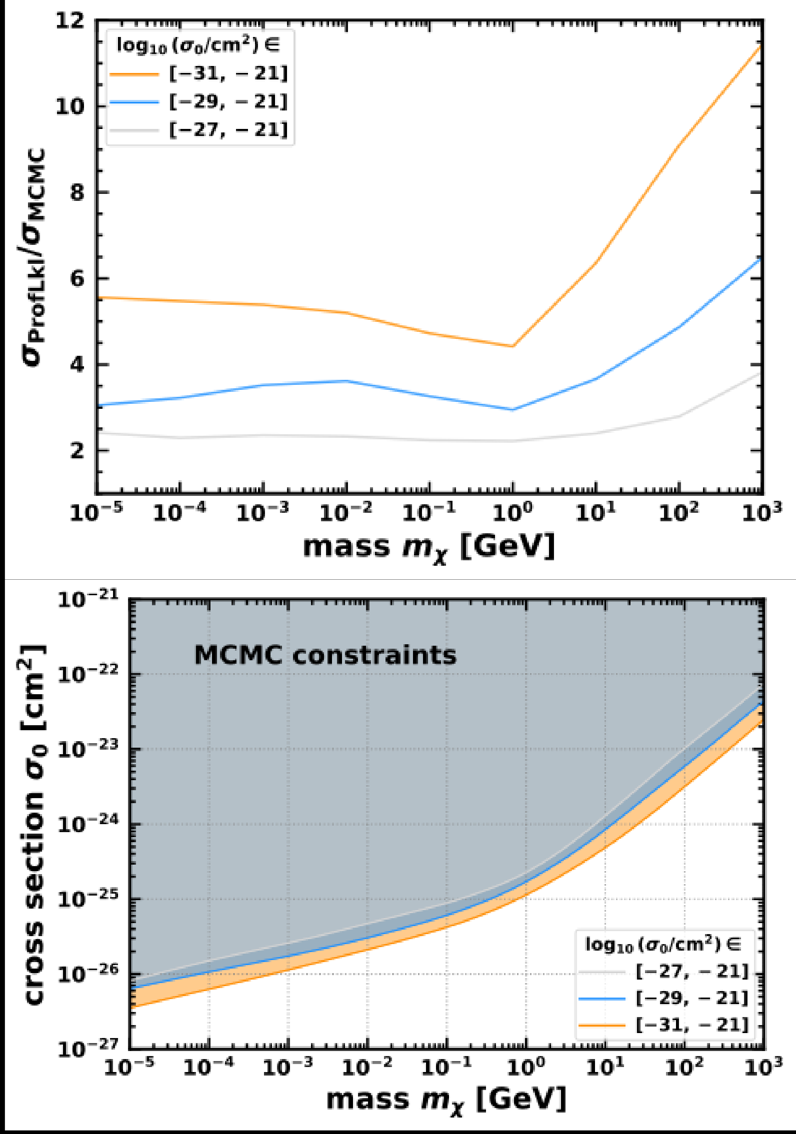


Example 3 — flat priors, hidden prior-volume effect

- A nuisance direction whose good-fit width grows with θ_i
- Maximising ignores the width but integrating collects the extra volume
- Posterior disagrees with the profile — even with flat priors
- Real cosmology: EDE ($f_{ede} \to 0$) and DM–proton ($\sigma_0 \to 0, f_\chi \to 0$).

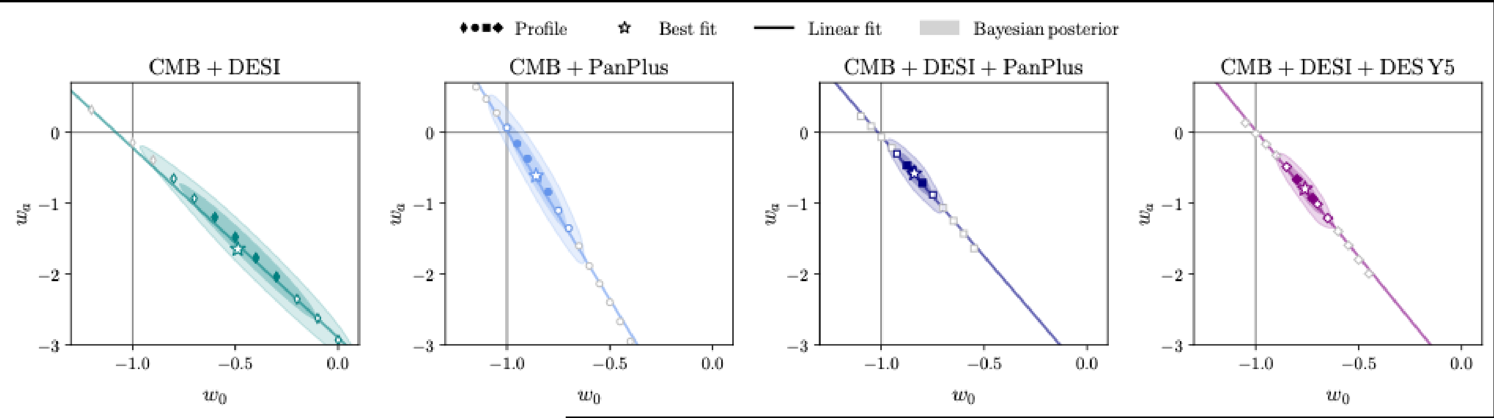


Poulin, Smith and TK [2302.09032]

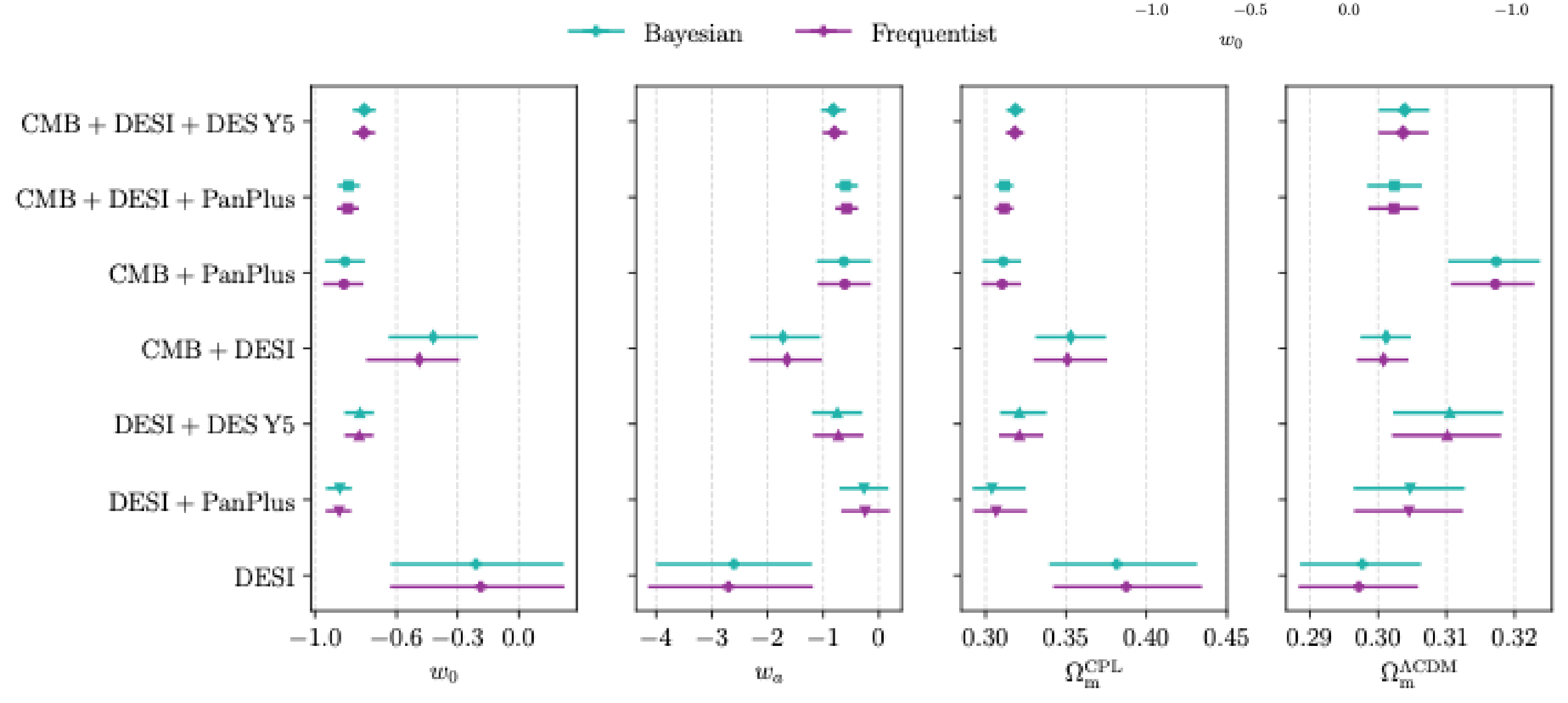


Straight, TK et al [2603.25731]

Counter-example —they agree on $w_0 w_a$ dark energy



Herold and TK [2506.12004]



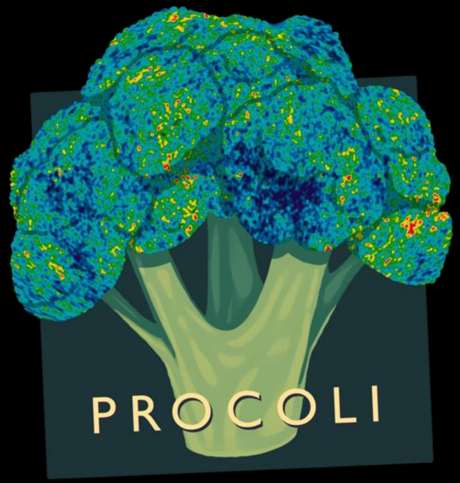
- DESI hints at dynamical dark energy (w_0, w_a) within a Bayesian analysis
- Profile likelihoods agree with the posteriors for both means and error bars
- Opening the CPL parameter space genuinely improves the fit $\Delta\chi^2 \approx -10$ to -20

Profile a cosmological parameter

Section 12 — profile a ΛCDM CMB parameter with the covariance-aware optimiser

This is most of Procoli's engine:

- Robust global optimisation via simulated annealing
- Read previous point as guess for next point
- Faster optimisation for subsequent points
- Sequential profile
- Record χ^2 for each individual experiment



Takeaways

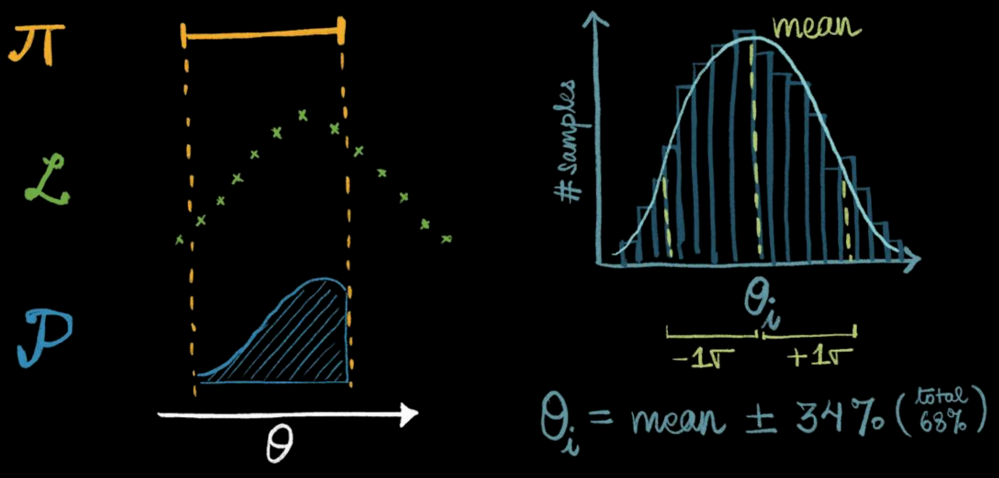
Bayesian posteriors

Prior

Likelihood

Posterior

Parameter

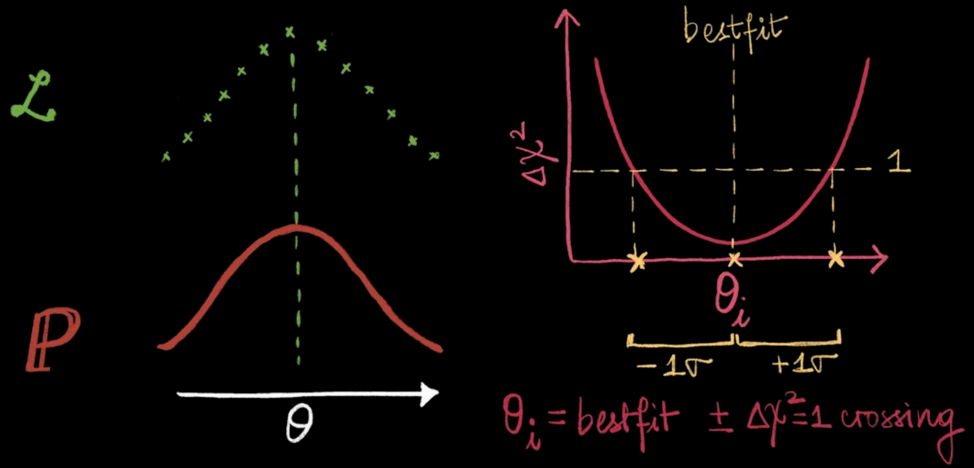


Frequentist profiles

Likelihood

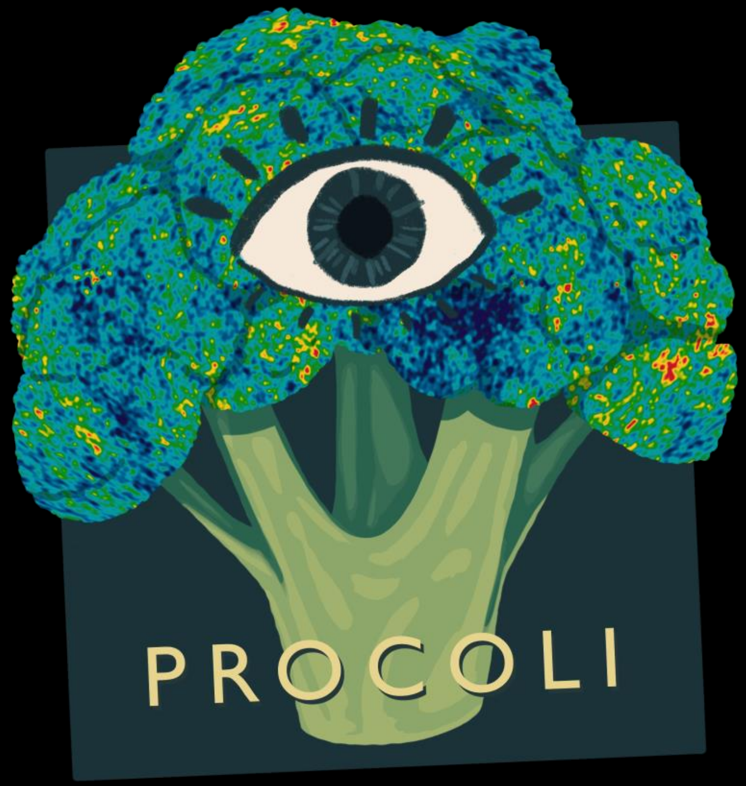
Profile

Parameter



- Profiles are prior-independent by design and provide a check on prior-volume effects
- Agreement (eg. $w_0 w_a$) → trust the Bayesian result. Disagreement (EDE, DM–proton) → prior-volume impacts
- Use both frameworks together — not Bayes vs freq, but Bayes and freq.

Open your eyes to the impact of priors!



Frequentist–Bayesian hybrids

- Common in LSS / EFTofLSS: profile the nuisance parameters, stay Bayesian on cosmology
- Or add prior information into a frequentist fit via a joint likelihood of current + past experiments
- Keeps prior-volume out of the parameters you care about, without discarding real prior knowledge

Coverage tests

- To check that intervals provide correct coverage, inject a known truth into mock data
- Build the interval many times over independent mocks; count how often it brackets the truth
- Should equal the expected confidence level at $\Delta\chi^2 = 1\ (68\%)\ ,\ \ 4(95\%)\ ,\ \ldots$