Statistical methods in cosmology

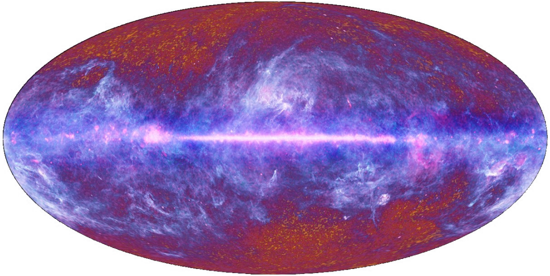
Will Handley
<wh260@cam.ac.uk>

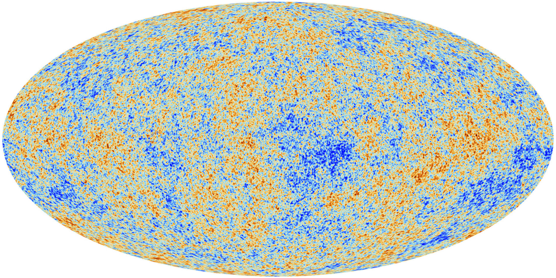
Royal Society University Research Fellow & Turing Fellow
Astrophysics Group, Cavendish Laboratory, University of Cambridge
Kavli Institute for Cosmology, Cambridge
Gonville & Caius College
willhandley.co.uk

12^{th} April 2022

Lessons from the CMB

- 21cm global signal detection is superficially similar to detecting the primordial CMB.
- Both are attempted to detect tiny $\sim 10^{-5}$ cosmological signals hidden beneath a large "uninteresting" foreground.
- Both measurements are frustrated by complicated contaminants.
- Both are amenable to a Bayesian analysis.





The three pillars of Bayesian inference

- Parameter estimation: "What do the data tell me about my model?":

$$P(\theta|D,M) = \frac{P(D|\theta,M)P(\theta|M)}{P(D|M)}, \qquad \mathcal{P} = \frac{\mathcal{L}\times\pi}{\mathcal{Z}}, \qquad \text{Posterior} = \frac{\text{Likelihood}\times\text{Prior}}{\text{Evidence}}.$$

- Model comparison: "Which model best fits the data?":

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}, \qquad \frac{\mathcal{Z}_\mathcal{M}\Pi_\mathcal{M}}{\sum_m Z_m \Pi_m}, \qquad \text{Model Posterior} = \frac{\text{Evidence}\times\text{Model Prior}}{\text{Normalisation}}.$$

- Tension quantification: "Are datasets consistent within a given model?" [1902.04029]

$$\mathcal{R} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A\mathcal{Z}_B}, \qquad \log\mathcal{S} = \langle\log\mathcal{L}_{AB}\rangle_{\mathcal{P}_{AB}} - \langle\log\mathcal{L}_B\rangle_{\mathcal{P}_A} - \langle\log\mathcal{L}_B\rangle_{\mathcal{P}_B}$$

What is a model?

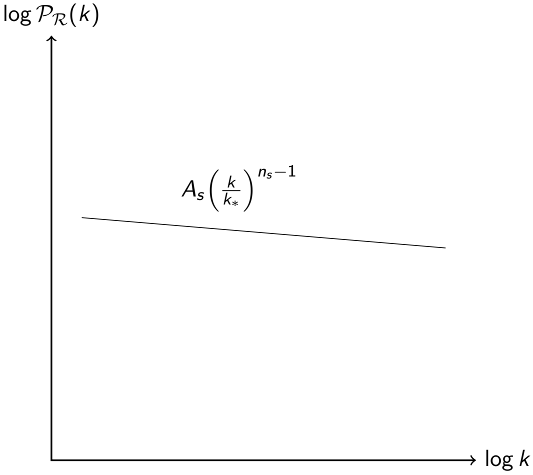
- Model comparison in its purest form answers question such as:
  - "Is the universe ΛCDM?"
  - "Are neutrinos in a normal or inverted hierarchy?"
  - "Is there a detectable global signal in this data?"
- However model $\mathcal{M}$ is likelihood $\mathcal{L} = P(D|\theta, M)$ and priors $\pi = P(\theta|M)$, $\Pi = P(M)$
- Can use the evidence $\mathcal{Z}$ to decide on which out of a set of likelihoods best describe data (e.g. Gaussian, Cauchy, Poisson, radiometric).
- Can also use it for antenna selection [2106.10193] [2109.10098].
- In principle can use it to decide between theoretically motivated priors (care needed)
- It can also be used for non-parametric reconstruction:
  - "How many polynomial terms best describe the data?"
  - "How complicated a sky model do I need?"
  - "Which is the best sky model?"

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction [1908.00906]

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s\left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

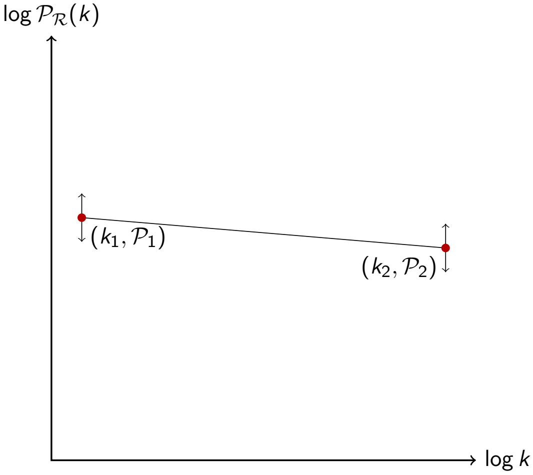


Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction [1908.00906]

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

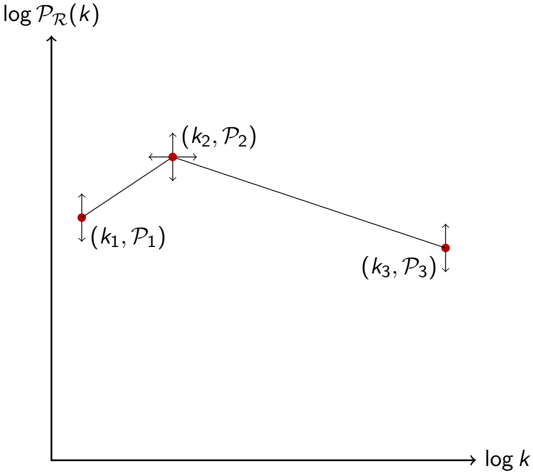


Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction [1908.00906]

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

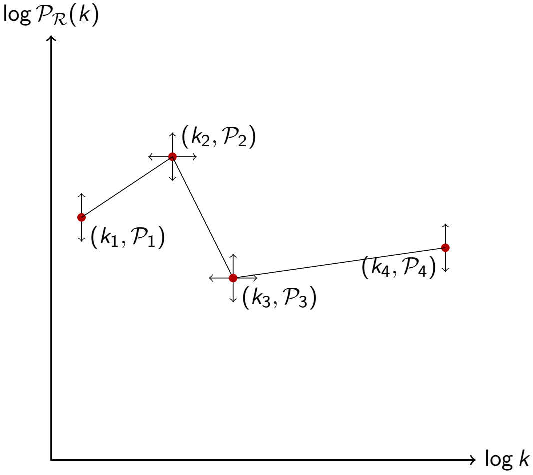


Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction [1908.00906]

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

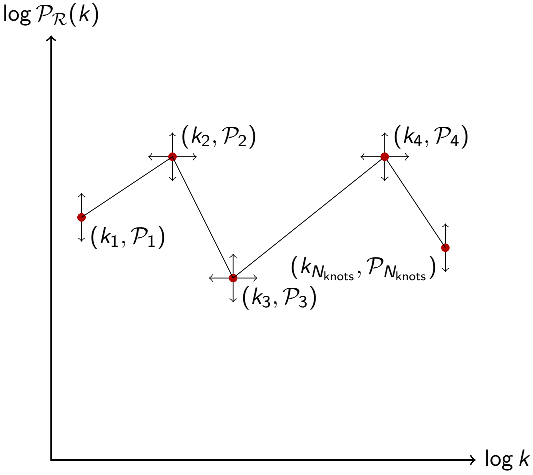


Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction [1908.00906]

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

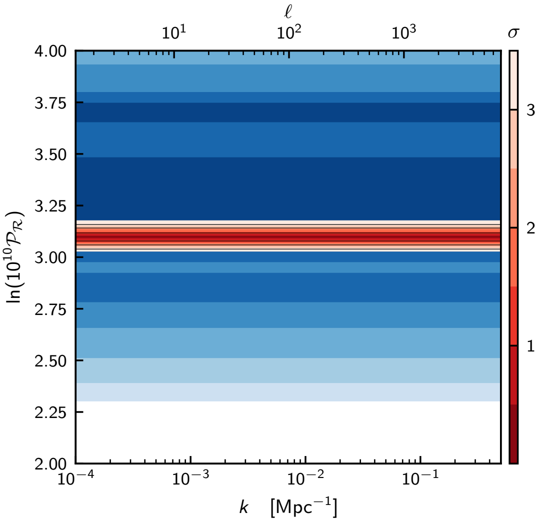


0 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

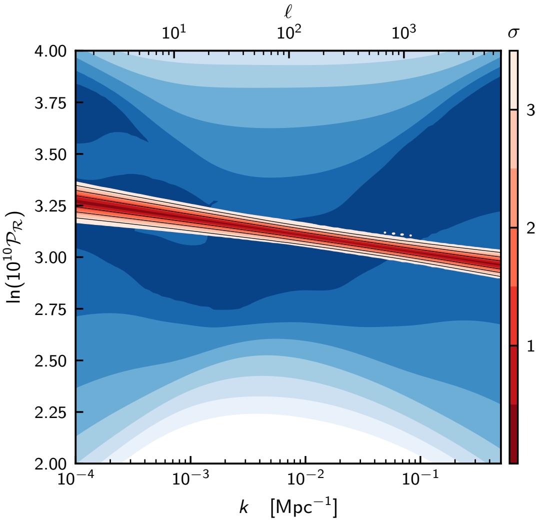


1 internal knot

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

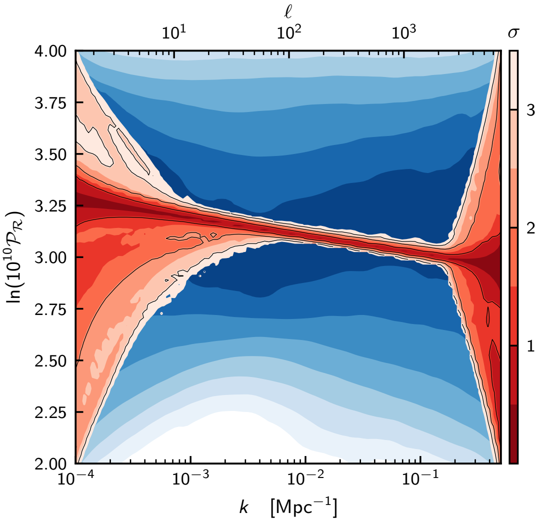


2 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

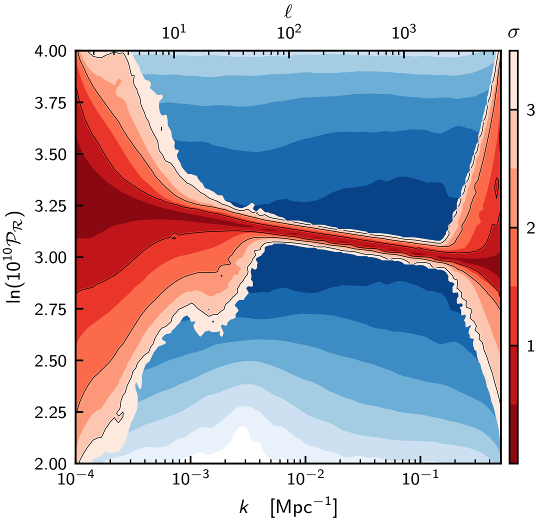


3 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

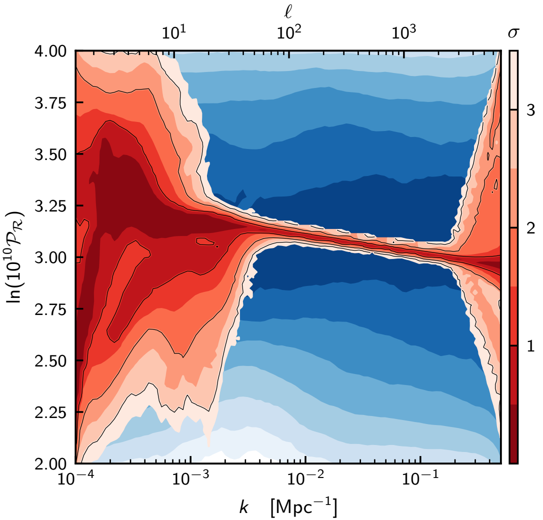


4 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

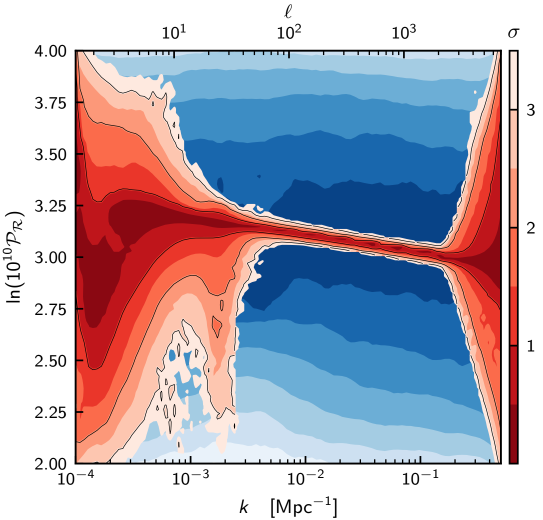


5 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

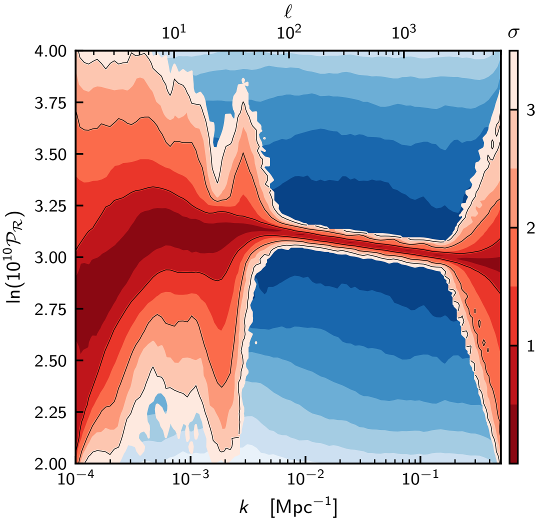


6 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

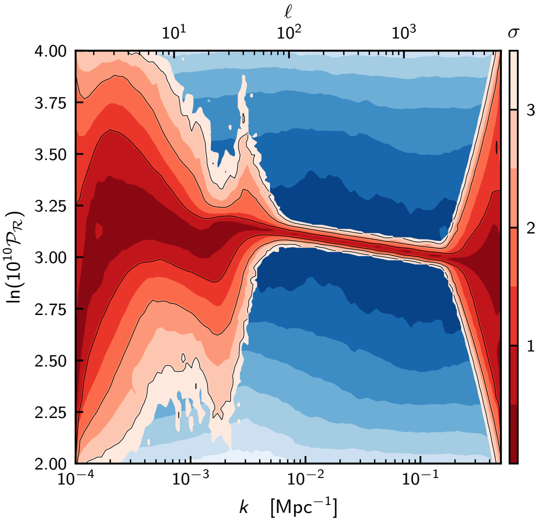


7 internal knots

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

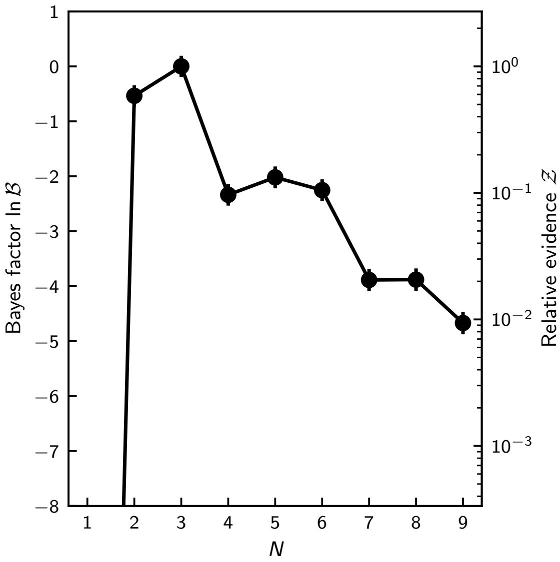


Bayes Factors

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

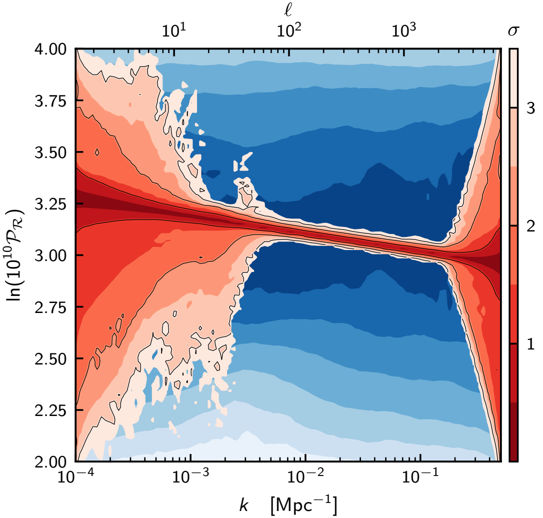


Marginalised plot

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_run (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters

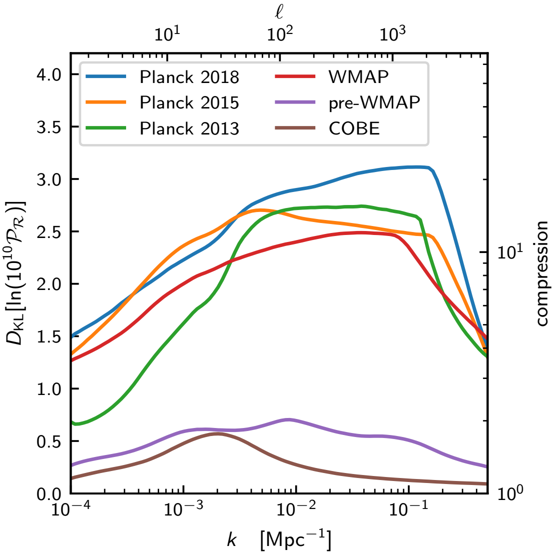


Kullback-Liebler divergences

- Traditionally parameterise the primordial power spectrum with (A_s, n_s)

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s-1}$$

- To add more degrees of freedom, can add "running" parameters n_{run} (higher order polynomial in index)
- Alternative non-parametric technique introduces a more flexible phenomenological parameterisation: "FlexKnots"
- Let the Bayesian evidence decide when you've introduced too many parameters



Occam's Razor [2102.11511]

- Bayesian inference quantifies Occam's Razor:
  - *"Entities are not to be multiplied without necessity"* — William of Occam
  - *"Everything should be kept as simple as possible, but not simpler"* — "Albert Einstein"
- Properties of the evidence: rearrange Bayes' theorem for parameter estimation

$$\mathcal{P}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\mathcal{Z}} \qquad \Rightarrow \qquad \log\mathcal{Z} = \log\mathcal{L}(\theta) - \log\frac{\mathcal{P}(\theta)}{\pi(\theta)}$$

- Evidence is composed of a "goodness of fit" term and "Occam Penalty"

- RHS true for all θ. Take max likelihood value θ_*:

$$\log\mathcal{Z} = -\chi^2_{\mathrm{min}} - \text{Mackay penalty}$$

- Be more Bayesian and take posterior average to get the "Occam's razor equation"

$$\boxed{\log\mathcal{Z} = \langle\log\mathcal{L}\rangle_{\mathcal{P}} - \mathcal{D}_{\mathrm{KL}}}$$

- Natural regularisation which penalises models with too many parameters.

Kullback Liebler divergence

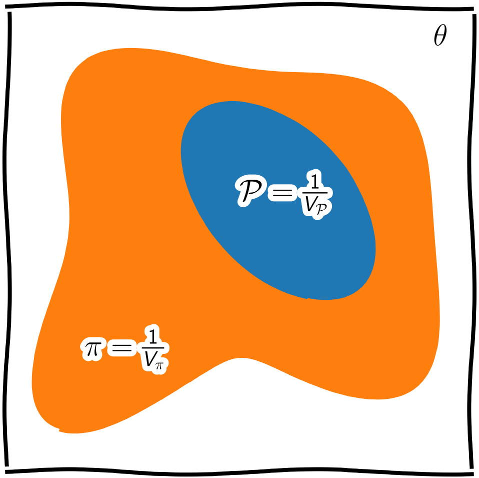
- The KL divergence between prior π and posterior $\mathcal{P}$ is is defined as:

$$\mathcal{D}_{\mathrm{KL}} = \left\langle \log \frac{\mathcal{P}}{\pi} \right\rangle_{\mathcal{P}} = \int \mathcal{P}(\theta) \log \frac{\mathcal{P}(\theta)}{\pi(\theta)} d\theta.$$

- Whilst not a distance, $\mathcal{D} = 0$ when $\mathcal{P} = \pi$.
- Occurs in the context of machine learning as an objective function for training functions.
- In Bayesian inference it can be understood as a log-ratio of "volumes":

$$\mathcal{D}_{\mathrm{KL}} \approx \log \frac{V_\pi}{V_{\mathrm{P}}}.$$

(this is exact for top-hat distributions).

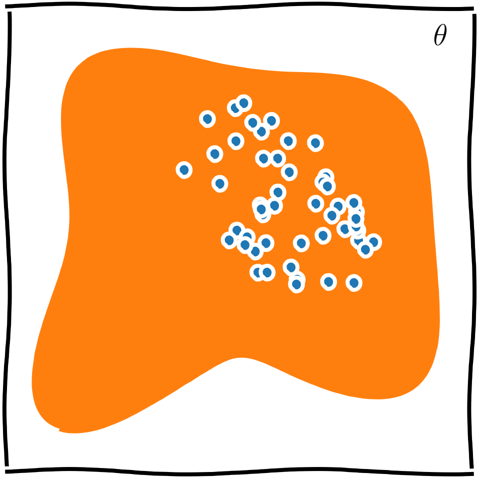


Why do sampling?

- The cornerstone of numerical Bayesian inference is working with samples.
- Generate a set of representative parameters drawn in proportion to the posterior $\theta \sim \mathcal{P}$.
- The magic of marginalisation $\Rightarrow$ perform usual analysis on each sample in turn.
- The golden rule is stay in samples until the last moment before computing summary statistics/triangle plots because

$$\boxed{f(\,\langle X\rangle\,) \ne \langle\, f(X)\,\rangle}$$

- Generally need $\sim \mathcal{O}(12)$ independent samples to compute a value and error bar.



How to generate samples

- MCMC!
- chi-feng.github.io/mcmc-demo/

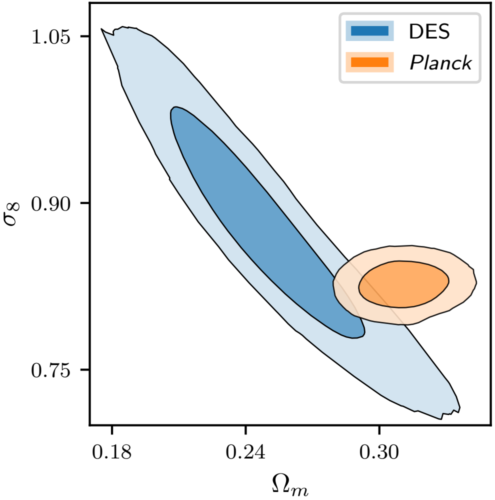
Nested Sampling: Benefits and drawbacks

Relative to traditional numerical posterior samples (Metropolis Hastings, HMC, emcee), nested sampling:

+ Can calculate evidence (and therefore perform model comparison).
+ Can calculate KL divergence.
+ Can handle multi-modal distributions.
+ Requires little tuning for an a-priori unseen problem.
+ Highly parallelisable ($n_\mathrm{cores} \sim n_\mathrm{live} \gg 4$).
+ Does not require gradients
− Slower than a well-tuned posterior sampler.
− Run time is dependent on prior choice, and priors must be proper (some people view this as a feature rather than a bug).

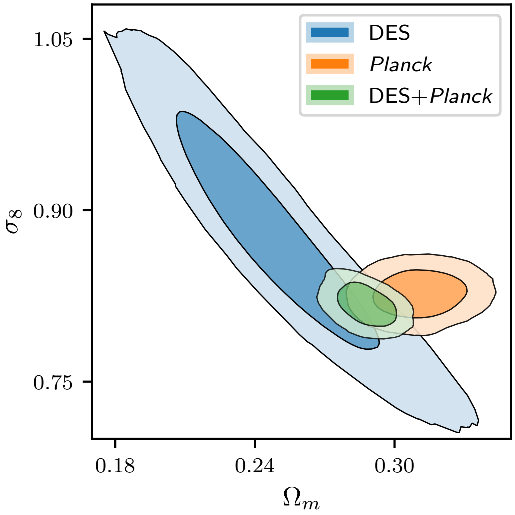
The importance of global measures of tension

- Hubble tension [1907.10625]
  - *Planck*: $H_0 = 67.4 \pm 0.5$
  - SH_0ES: $H_0 = 74.0 \pm 1.4$
- In other situations the discrepancy doesn't exist in a single interpretable parameter
- For example: DES+*Planck* [1902.04029]
- Are these two datasets in tension?
- There are a lot more parameters – are we sure that tensions aren't hiding? Are we sure we've chosen the best ones to reveal the tension?
- Should use "Suspiciousness" statistic $\mathcal{S}$, or Bayes ratio $\mathcal{R}$ to determine global tension.



The importance of global measures of tension

- Hubble tension [1907.10625]
  - *Planck*: $H_0 = 67.4 \pm 0.5$
  - SH_0ES: $H_0 = 74.0 \pm 1.4$
- In other situations the discrepancy doesn't exist in a single interpretable parameter
- For example: DES+*Planck* [1902.04029]
- Are these two datasets in tension?
- There are a lot more parameters – are we sure that tensions aren't hiding? Are we sure we've chosen the best ones to reveal the tension?
- Should use "Suspiciousness" statistic $\mathcal{S}$, or Bayes ratio $\mathcal{R}$ to determine global tension.



Future extensions for REACH

- Tension quantification for cross validation
  - Between experiments
  - Between REACH antennae
  - Between different subsets of the REACH timestream
- Model marginalisation rather than comparison
- FlexKnot reconstructions
- Likelihood selection
- Occam factors on evidence plots.
- Integration of calibration and cosmology pipelines

FAQs

- What was that awesome website?

  Full credit to Chi-feng for this incredible online demonstration tool

  chi-feng.github.io/mcmc-demo/

- How do you make your plots look hand-drawn?

```
import matplotlib.pyplot as plt
plt.xkcd()
```