Cosmological inference tools

Marginal statistics and fully Bayesian forecasts

Will Handley
<wh260@cam.ac.uk>

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

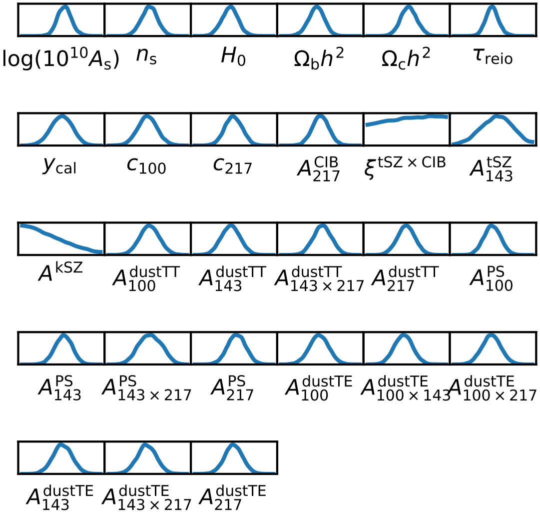
12th October 2023

Marginal inference

- Many cosmological likelihoods come with nuisance parameters that have limited relevance for onward inference.
- Notation: $\mathcal{L} = P(D|\theta, \alpha, M)$

$\mathcal{L}$	Likelihood	(e.g. `plik`),
D	Data	(e.g. CMB),
θ	Cosmological parameters	(e.g. Ω_m, $H_0 \ldots$),
α	Nuisance parameters	(e.g. $A_{\text{planck}} \ldots$),
M	Model	(e.g. ΛCDM).

- Some marginal statistics (e.g. marginal means, posteriors. . .) are easy to compute.
- More machinery is needed for e.g. nuisance marginalised likelihoods and marginal KL divergences $\mathcal{D}_{\text{KL}}$.



Nuisance marginalised likelihoods: Theory [2207.11457]

- Bayes theorem

$$\mathcal{L}(\theta,\alpha) \times \pi(\theta,\alpha) = \mathcal{P}(\theta,\alpha) \times \mathcal{Z} \tag{1}$$

$$\text{Likelihood} \times \text{Prior} = \text{Posterior} \times \text{Evidence}$$

α: nuisance parameters, θ: cosmo parameters.

- Marginal Bayes theorem

$$\mathcal{L}(\theta) \times \pi(\theta) = \mathcal{P}(\theta) \times \mathcal{Z} \tag{2}$$

- Non-trivially gives **nuisance-free likelihood**

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}} = \frac{\int \mathcal{L}(\theta,\alpha)\pi(\theta,\alpha)d\alpha}{\int \pi(\theta,\alpha)d\alpha} \tag{3}$$

Key properties

- Given datasets A and B, each with own nuisance parameters α_A and α_B:
- If you use $\mathcal{L}_A(\theta)$, you get the same (marginal) posterior and evidence if you had run with nuisance parameters α_A (ditto B).
- If you run inference on $\mathcal{L}_A(\theta) \times \mathcal{L}_B(\theta)$, you get the same (marginal) posterior and evidence if you had run with all nuisance parameters α_A, α_B on.

  (weak marginal consistency requirements on joint $\pi(\theta,\alpha_A,\alpha_B)$ and marginal priors)

Nuisance marginalised likelihoods: Practice

Harry Bevins

PhD→JRF

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

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

$\boxed{\mathcal{L}(\theta,\alpha)}$ $\boxed{\pi(\theta,\alpha)}$

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

$\boxed{\mathcal{L}(\theta,\alpha)}$ $\boxed{\pi(\theta,\alpha)}$ $\downarrow$ $\boxed{\text{Nested Sampling}}$

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

Harry Bevins

PhD→JRF

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

$$\begin{tikzcd}
\boxed{\mathcal{L}(\theta,\alpha)} \arrow[dr] & & \boxed{\pi(\theta,\alpha)} \arrow[dl] \\
& \boxed{\text{Nested Sampling}} \arrow[d] & \\
& \boxed{\{\theta,\alpha\}_\mathcal{P}} &
\end{tikzcd}$$

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

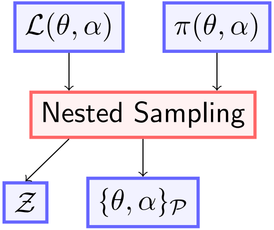
Nuisance marginalised likelihoods: Practice

Harry Bevins

PhD→JRF

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

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities



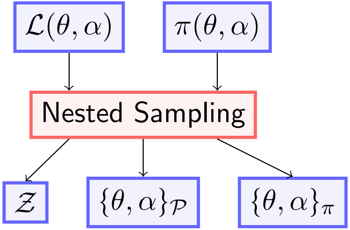
1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_{\mathcal{P}}$ and $\{\theta, \alpha\}_{\pi}$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities



1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

Harry Bevins

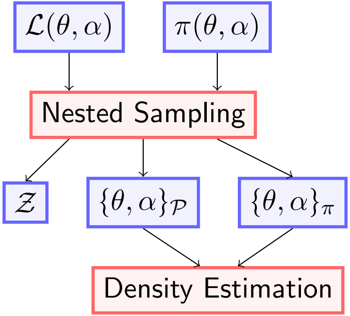
PhD→JRF

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

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

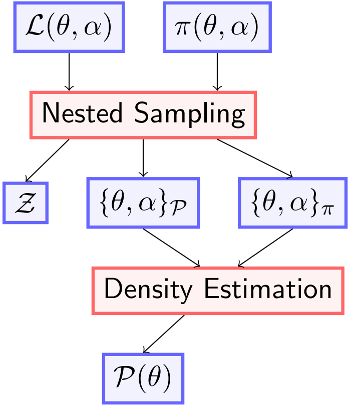
Harry Bevins
PhD→JRF

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

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

Harry Bevins

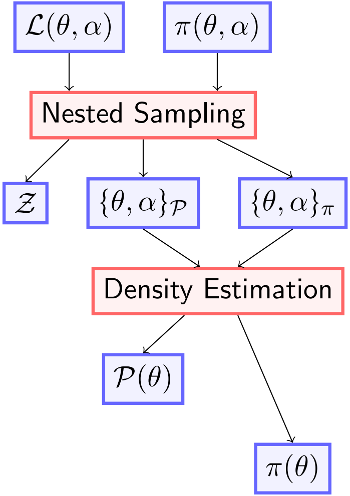
PhD→JRF

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

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_{\mathcal{P}}$ and $\{\theta, \alpha\}_{\pi}$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

Harry Bevins

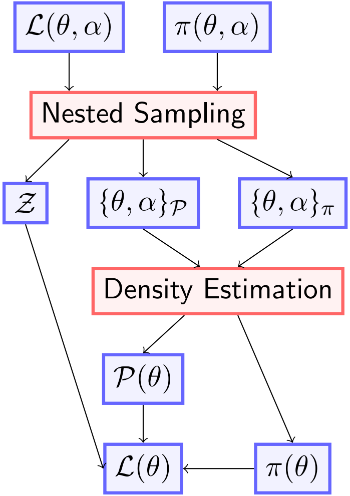
PhD→JRF

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_{\mathcal{P}}$ and $\{\theta, \alpha\}_{\pi}$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



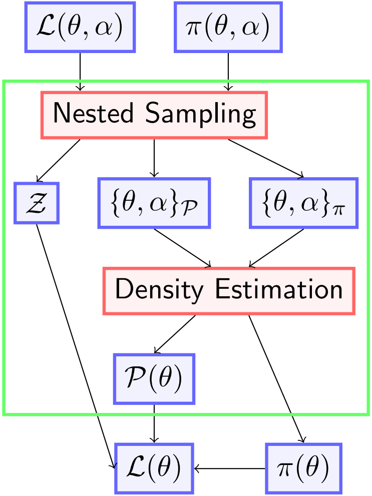
Nuisance marginalised likelihoods: Practice

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

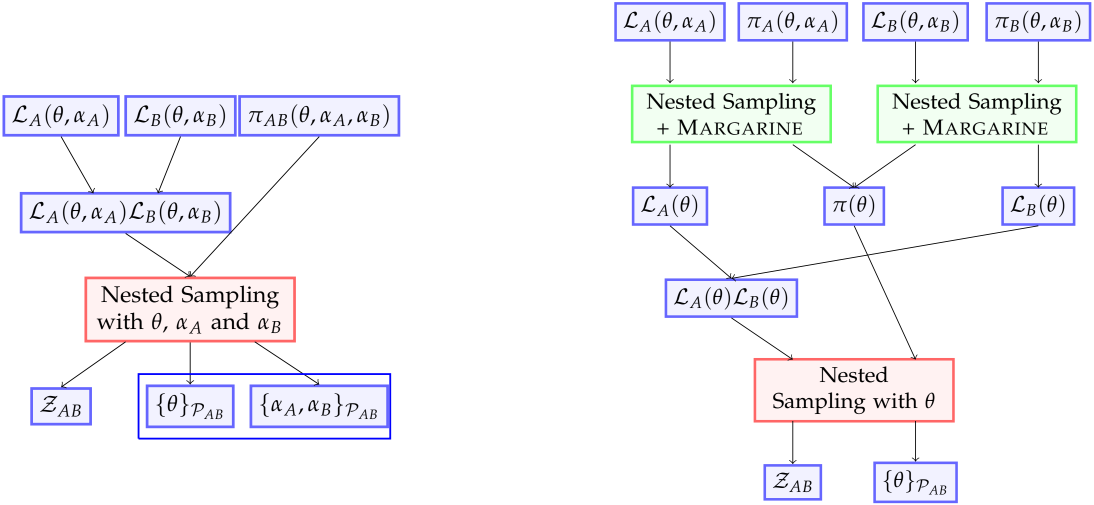
- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Combination



History of margarine

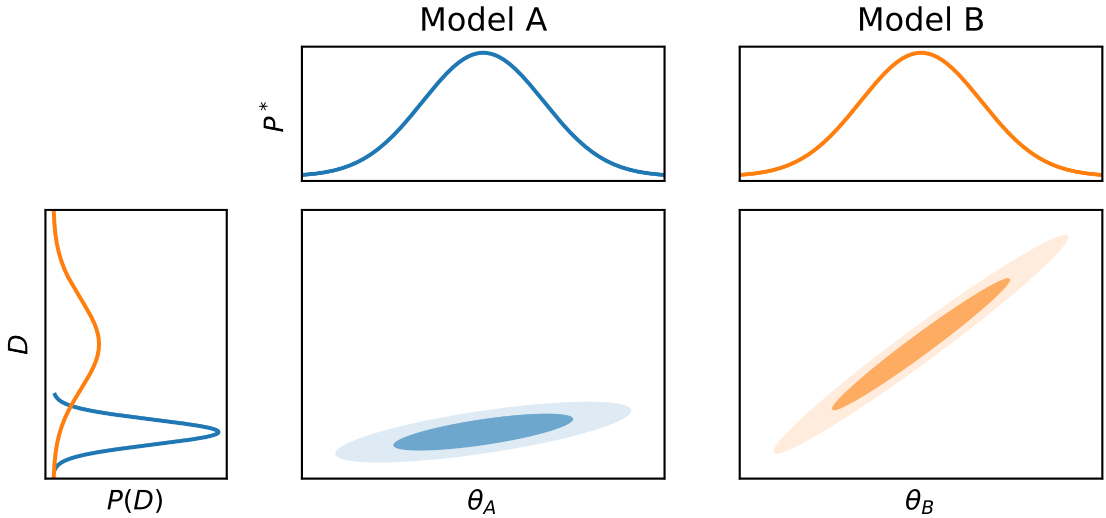
Dily Ong
PhD

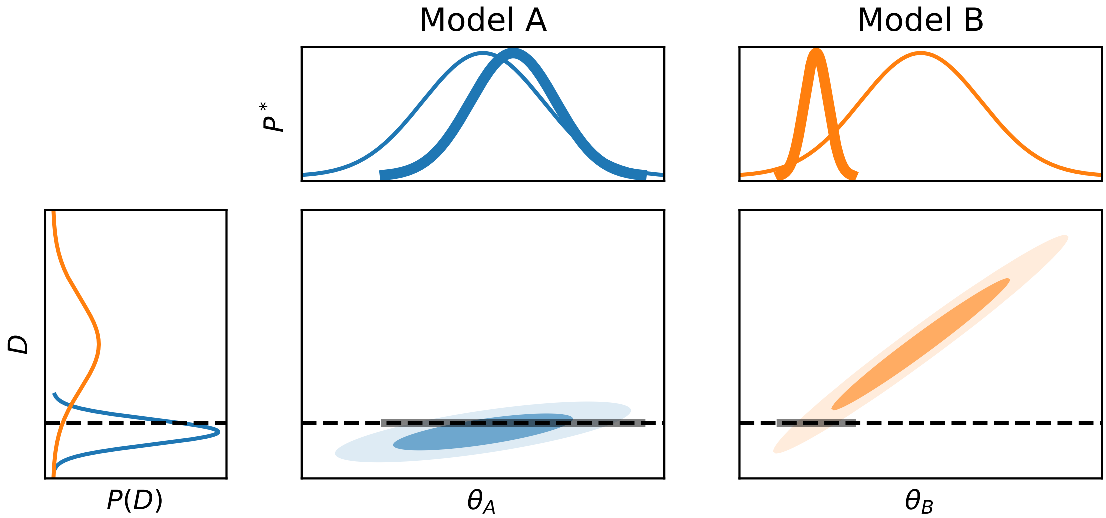
- Papamakarios et al [1912.02762] (normalising flows)
- Alsing et al [1903.00007] (Delfi)
- Nested sampling with any prior you like (Alsing & Handley) [2102.12478]
- `margarine` (theory) Bevins et al [2207.11457]
- `margarine` (practice) Bevins et al [2205.12841]
- Next step: `unimpeded`
    - pip-installable download system for DiRAC chains

Cosmological forecasting

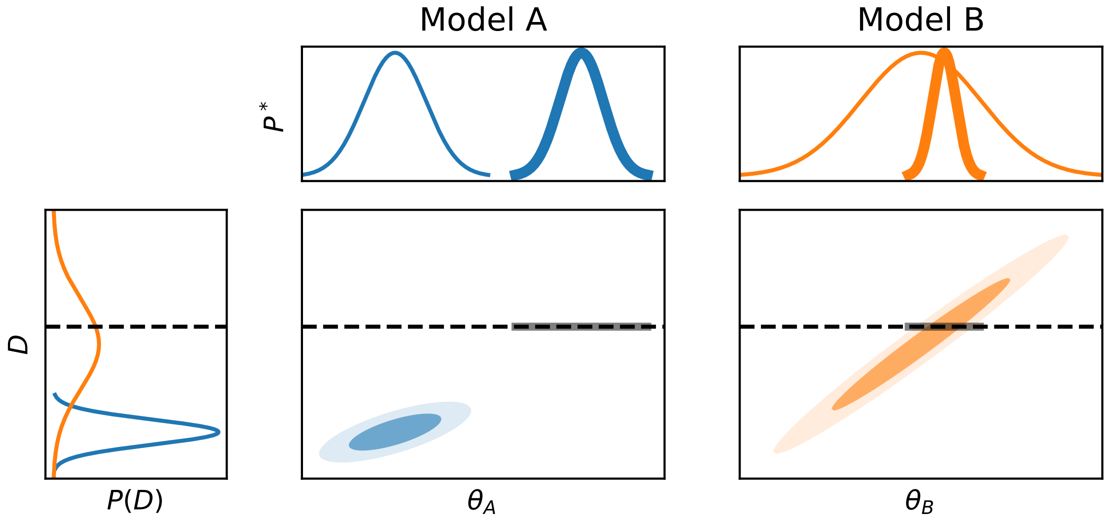
Have you ever done a Fisher forecast, and then felt Bayesian guilt?

- Cosmologists are interested in forecasting what a Bayesian analysis of future data might produce.
- Useful for:
  - white papers/grants,
  - optimising existing instruments/strategies,
  - picking theory/observation to explore next.
- To do this properly:
  1. start from current knowledge $\pi(\theta)$, derived from current data
  2. Pick potential dataset D that might be collected from $P(D)$ $(= \mathcal{Z})$
  3. Derive posterior $P(\theta|D)$
  4. Summarise science (e.g. constraint on θ, ability to perform model comparison)
- This procedure should be marginalised over:
  1. All possible parameters θ (consistent with prior knowledge)
  2. All possible data D
- i.e. marginalised over the joint $P(\theta, D) = P(D|\theta)P(\theta)$.
- Historically this has proven very challenging.
- Most analyses assume a fiducial cosmology θ_*, and/or a Gaussian likelihood/posterior (c.f. Fisher forecasting).
- This runs the risk of biasing forecasts by baking in a given theory/data realisation.


Model A
Model B
P^*
D
$P(D)$
θ_A
θ_B

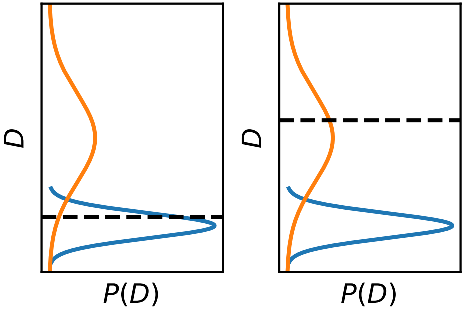

Model A
Model B
P^*
D
$P(D)$
θ_A
θ_B

Simulation-based inference & model comparison



Evidence networks [2305.11241]

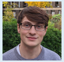
- Procedure proposed by Jeffreys & Wandelt:
  1. Generate labelled data from model A and model B.
  2. Train a probabilistic classifier to distinguish between the two.
  3. Use neural ratio trick to extract Bayes Factor $B = P(D|A)/P(D|B)$.
- Fully marginalises out parameters
- Only works in the data space
- Model comparison without nested sampling!



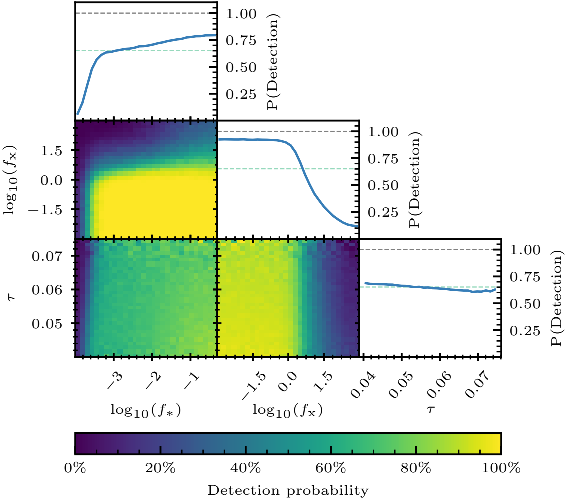
Fully Bayesian Forecasting [2309.06942]

Thomas Gessey-Jones

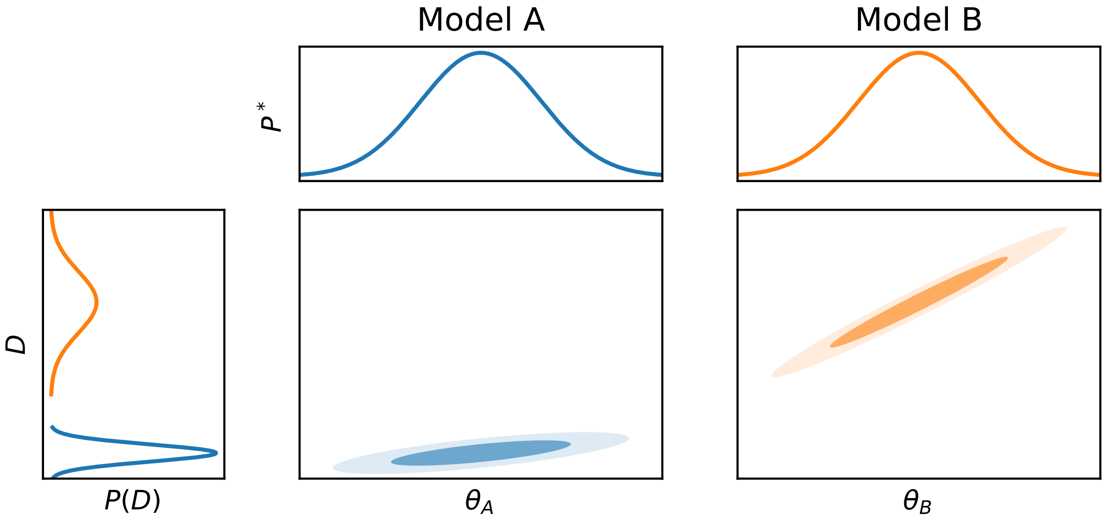
PhD



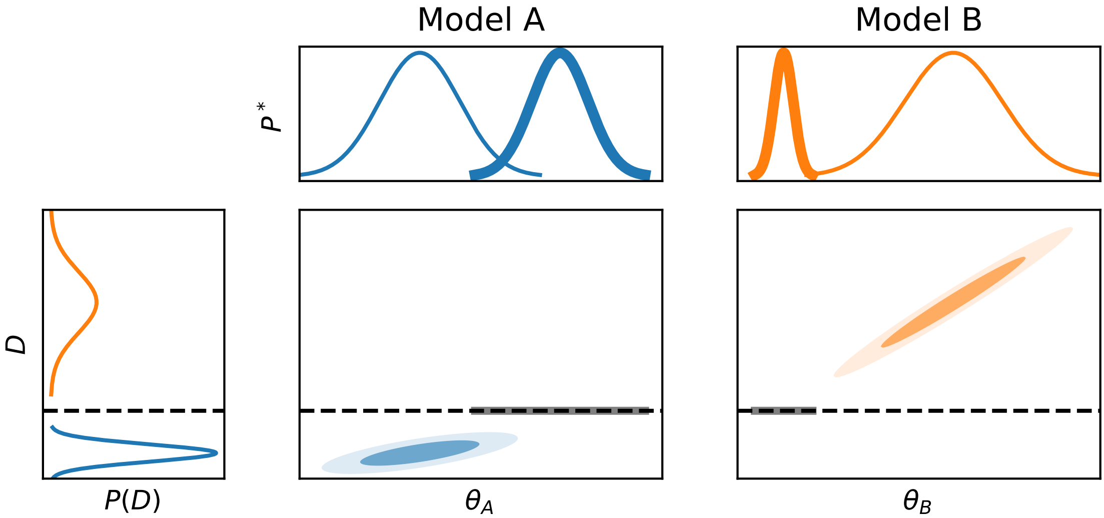
- Simulation based inference gives us the language to marginalise over parameters θ and possible future data D.
- Evidence networks give us the ability to do this at scale for forecasting.
- Demonstrated in 21cm global experiments, marginalising over:
  - theoretical uncertainty
  - foreground uncertainty
  - systematic uncertainty
- Able to say "at 67mK radiometer noise", have a 50% chance of 5σ Bayes factor detection.
- Can use to optimise instrument design
- Re-usable package: `prescience`



A word of caution on evidence networks



A word of caution on evidence networks



A word of caution on evidence networks

- Does not give evidence/partition function (which can be useful), only Bayes factor.
- Only valid if the true data lies within domain of extrapolation of neural network
  - True for forecasting.
  - False for real data.
- By throwing away parameter θ fitting, model cannot respond to mis-specified data.
- This criticism applies to any method claiming "amortization"
  - Amortized methods claim to train a system which works for all possible observed D.
  - As Bayesians we should be suspicious, since the only truth we know is the observed data.
- Traditional SBI interchange:

  audience What if your simulator is missing (X,Y,Z,...)?

  speaker The exact same thing affects likelihood-based analysis
- Here, the audience is implicitly making a query about the danger of working in data space, whilst the speaker's comment only applies to parameter space θ.
- We should therefore focus on SBI approaches which have tunable parameter spaces (i.e. interpretable posteriors).

Conclusions

github.com/handley-lab

- `margarine` as a tool for marginal cosmological inference
- `unimpeded` as a tool for distributing fast reusable inference products
- `prescience` as a tool for fully Bayesian forecasting