Next generation inference tools for cosmology and beyond

Will Handley
<wh260@cam.ac.uk>

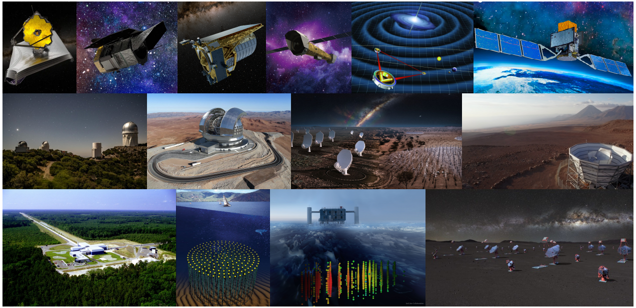
Royal Society University Research Fellow
Astrophysics Group, Cavendish Laboratory, University of Cambridge
Kavli Institute for Cosmology, Cambridge
Gonville & Caius College
willhandley.co.uk/talks

23rd Jan 2024

The golden age of cosmology data

- Over our research lifetimes we will see next-generation data rates across the electromagnetic spectrum & beyond:

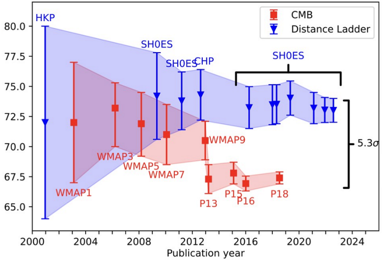
Radio	SKA *et al*
Micro	SO/CMB-S4
IR	JWST, Roman (WFIRST)
Optical	Euclid, DESI, Rubin (LSST), EELT
X-ray	Athena
Gamma-ray	e-ASTROGAM
Gravitational	LIGO/Virgo/Kagra + LISA
Particle	CTA, IceCube, KM3NeT

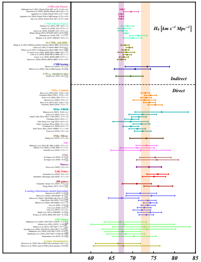


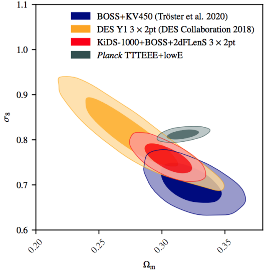
- We are moving from an age of **precision** cosmology to **accurate** cosmology.
- **Systematics** $\gtrsim$ **statistics**.
- Tools risk lagging behind hardware

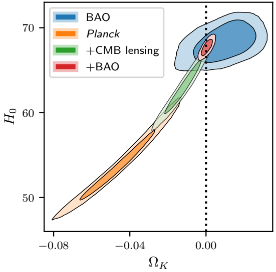
Tensions in cosmology

- **Data**: H_0, S_8, A_L/Ω_K, Li
- **Theory**: Initial conditions, Entropy, Dark energy, dark matter, quantum gravity
- **Analysis**:
- Disentangling systematics from new physics is challenging!
- Almost all cosmological analyses pragmatically assume a fiducial flat ΛCDM assumption during their analyses.
- Unless this is resolved, we risk confirmation bias in the analysis of next-generation data.









Structure of talk

1. Likelihood-based inference
2. Nested sampling
3. Marginal hierarchical inference
4. Simulation-based inference
5. Fully Bayesian Forecasts

Bayesian notation

- A "generative" model M, with tunable parameters θ, describing (compressed) data D.
  - e.g. $M = \Lambda\mathrm{CDM}$, $\theta = \{\Omega_b, \Omega_c, \tau, H_0, A_s, n_s\}$, $D = \{C_\ell\}$.
- Described by simulation process $\theta \to D$, or likelihood $P(D|\theta, M)$.
- Frequentists & Bayesians agree on the likelihood.
- Bayesians treat parameter space θ the same as data space D.
- Quantifying uncertainty with probability using Bayes theorem:

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

- Follows from the oft-forgotten Joint (the probability of everything):

$$\mathcal{P} \times \mathcal{Z} = \mathcal{J} = \mathcal{L} \times \pi, \qquad P(D, \theta|M) = \text{Joint} = \mathcal{J}$$

- Also relevant (in many overlapping contexts) is the dimensionless ratio

$$r = \frac{\mathcal{J}}{\pi \times \mathcal{Z}} = \frac{\mathcal{L}}{\mathcal{Z}} = \frac{\mathcal{P}}{\pi}$$

The three pillars of Bayesian inference

Parameter estimation

What do the data tell us about the parameters of a model? *e.g. the size or age of a ΛCDM universe*

$$P(\theta|D,M) = \frac{P(D|\theta,M)P(\theta|M)}{P(D|M)}$$

$$\mathcal{P} = \frac{\mathcal{L}\times\pi}{\mathcal{Z}}$$

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

Model comparison

How much does the data support a particular model? *e.g. ΛCDM vs a dynamic dark energy cosmology*

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

$$\frac{\mathcal{Z}_\mathcal{M}\Pi_\mathcal{M}}{\sum_m Z_m\Pi_m}$$

$$\text{Posterior} = \frac{\text{Evidence}\times\text{Prior}}{\text{Normalisation}}$$

Tension quantification

Do different datasets make consistent predictions from the same model? *e.g. CMB vs Type IA supernovae data*

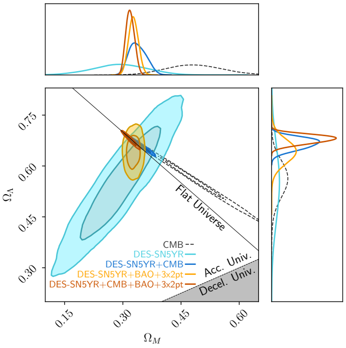
$$\mathcal{R} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A\mathcal{Z}_\mathcal{B}}$$

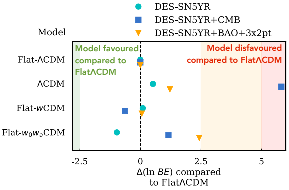
$$\log\mathcal{S} = \langle\log\mathcal{L}_{AB}\rangle_{\mathcal{P}_{AB}} - \langle\log\mathcal{L}_A\rangle_{\mathcal{P}_A} - \langle\log\mathcal{L}_B\rangle_{\mathcal{P}_B}$$

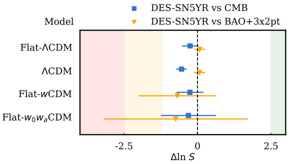
LBI: Likelihood-based inference

The standard approach if you are fortunate enough to have a likelihood function $\mathcal{L}(D|\theta)$: e.g recent DES analysis [2401.02929]

1. Define prior $\pi(\theta)$
   - spend some time being philosophical
2. Sample posterior $\mathcal{P}(\theta|D)$
   - use out-of-the-box MCMC tools such as `emcee` or `MultiNest`
   - make some triangle plots
3. Optionally compute evidence $\mathcal{Z}(D)$
   - e.g. nested sampling or parallel tempering
   - do some model comparison (i.e. science)
4. Optionally talk about tensions
   - Bayes ratio (Nested Sampling)
   - Suspiciousness (MCMC) [2007.08496]





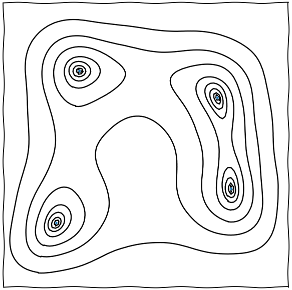


What is Nested Sampling?

- Nested sampling is a radical, multi-purpose numerical tool.
- Given a (scalar) function f with a vector of parameters θ, it can be used for:

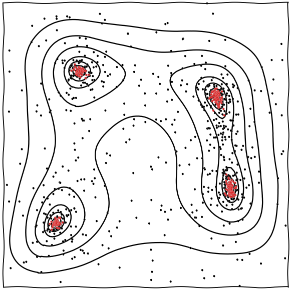
Optimisation

$$\theta_{\max} = \max_{\theta} f(\theta)$$



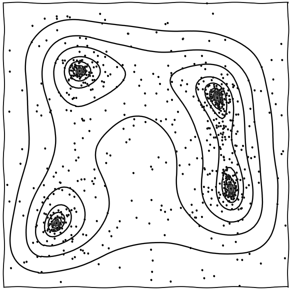
Exploration

$$\text{draw/sample} \quad \theta \sim f$$

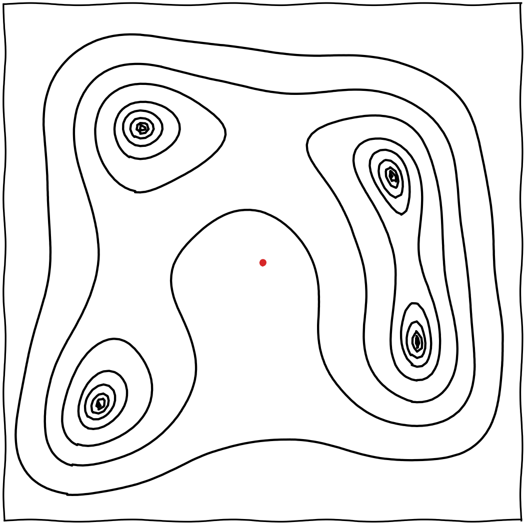


Integration

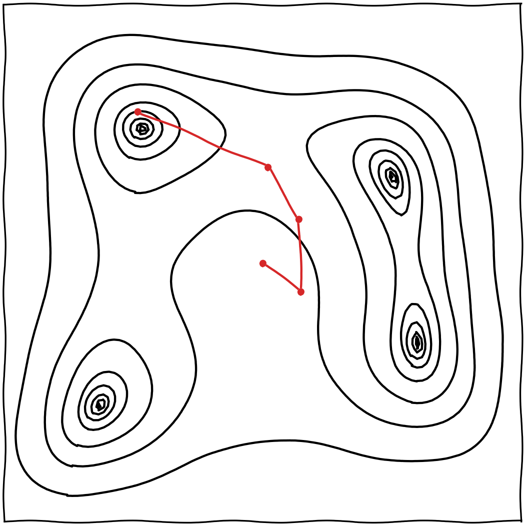
$$\int f(\theta) dV$$



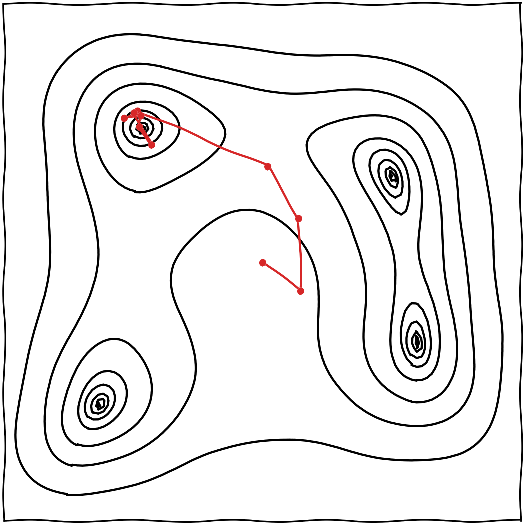
MCMC



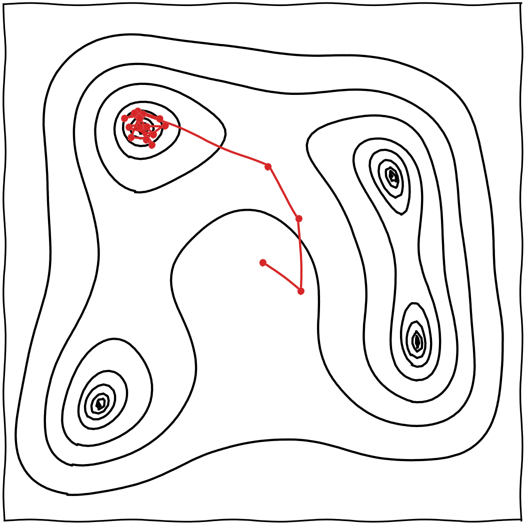
MCMC



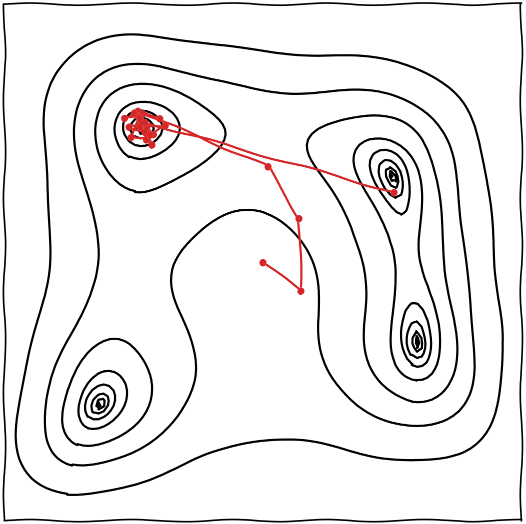
MCMC



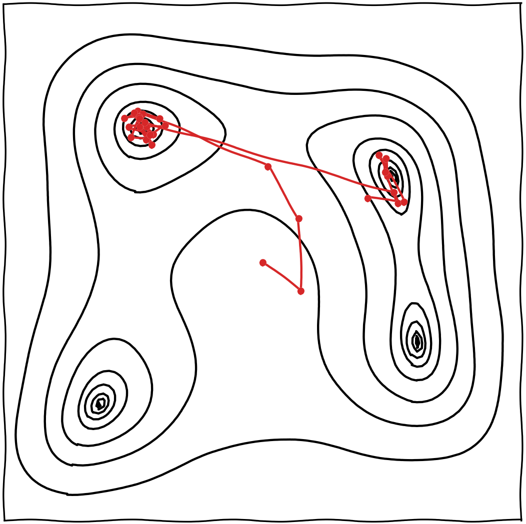
MCMC



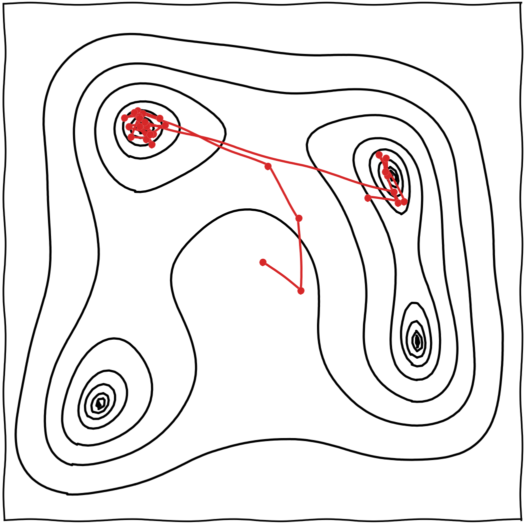
MCMC



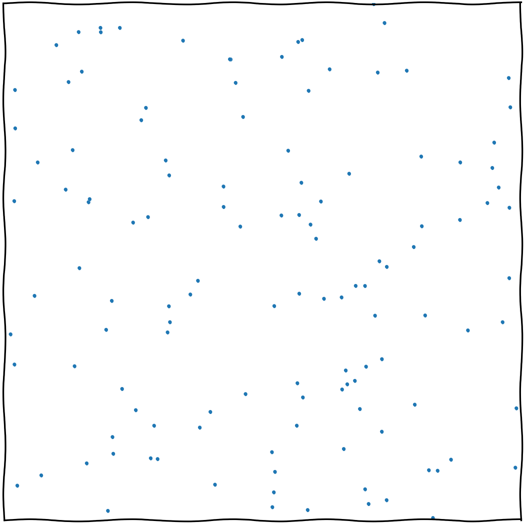
MCMC



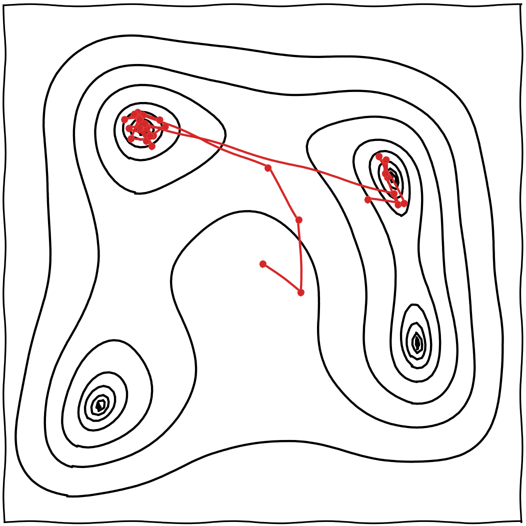
MCMC



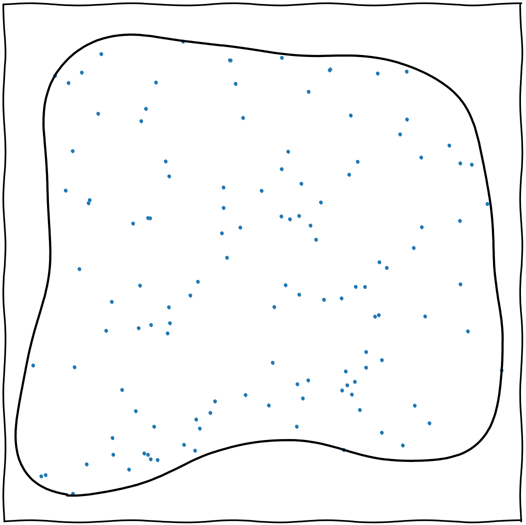
Nested sampling



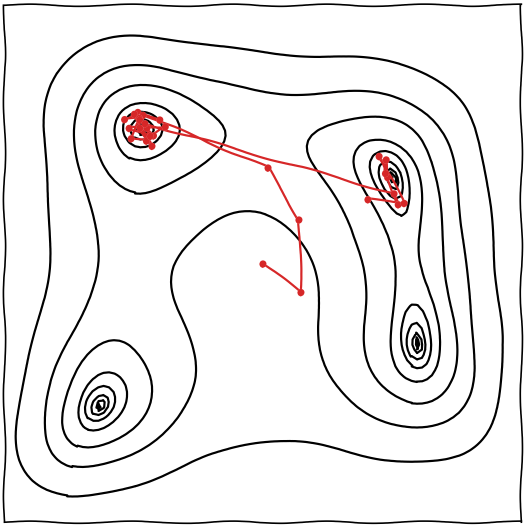
MCMC



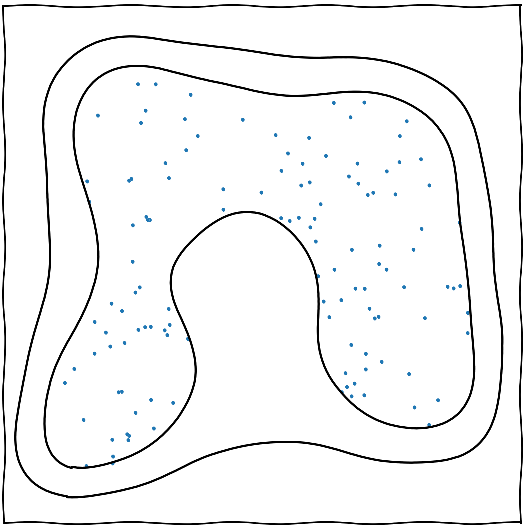
Nested sampling



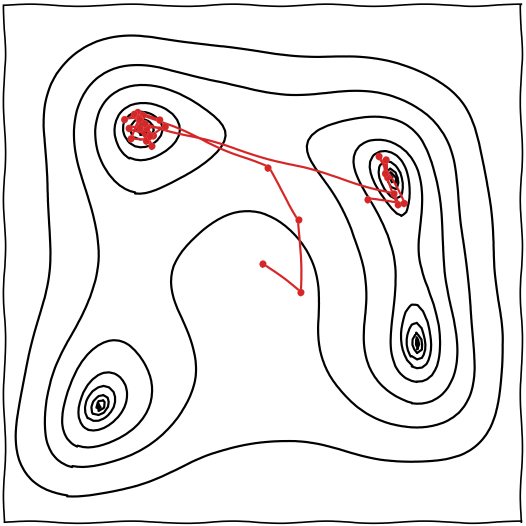
MCMC



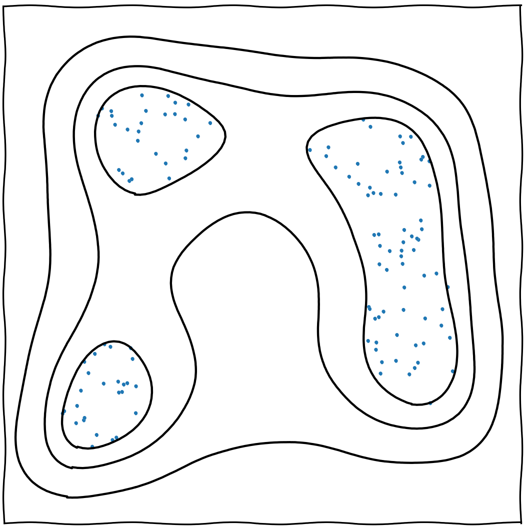
Nested sampling



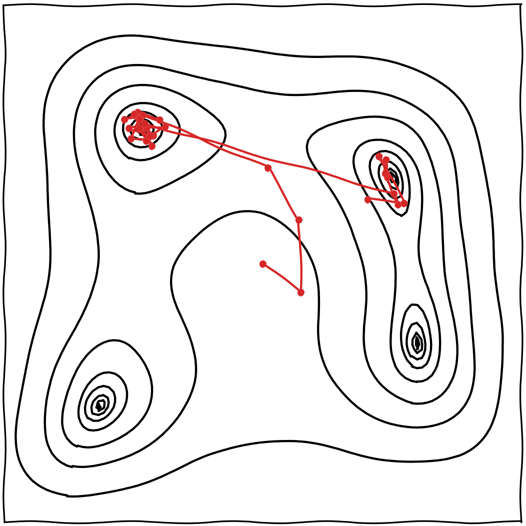
MCMC



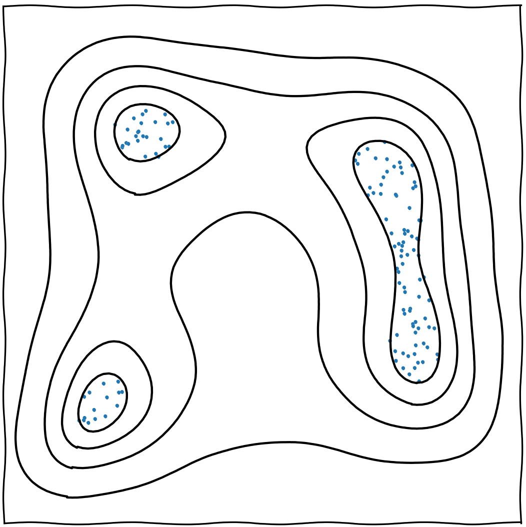
Nested sampling



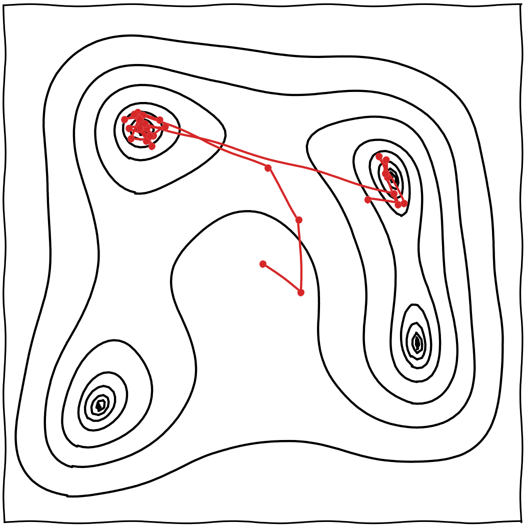
MCMC



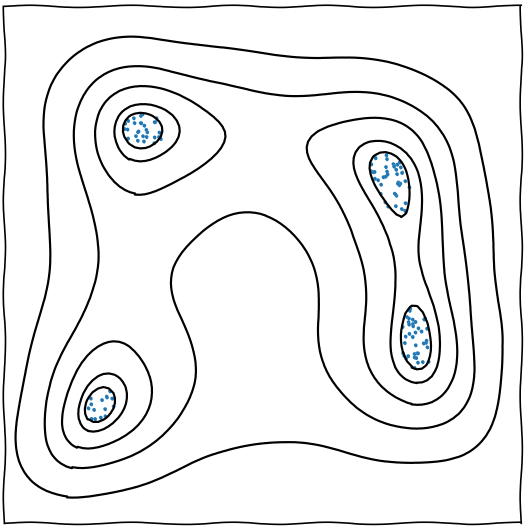
Nested sampling



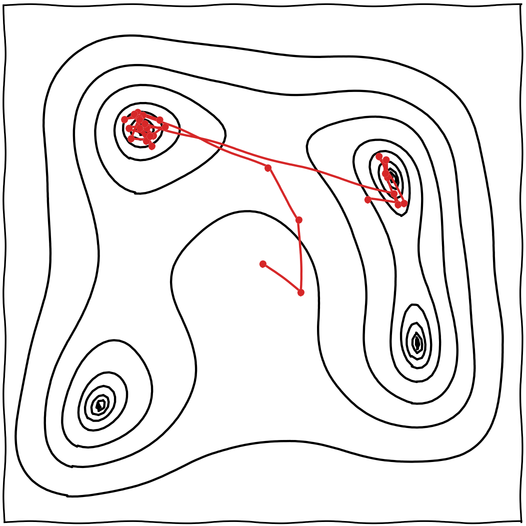
MCMC



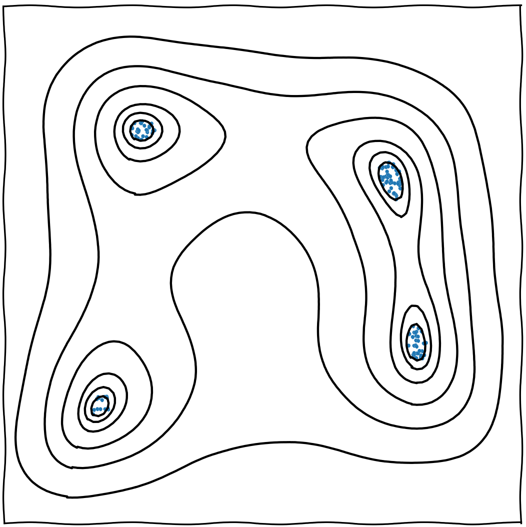
Nested sampling



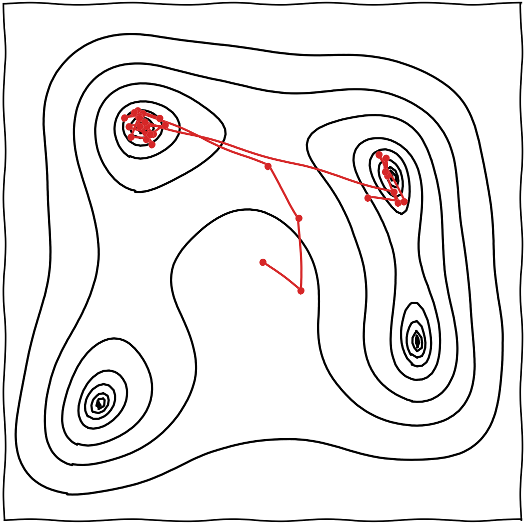
MCMC



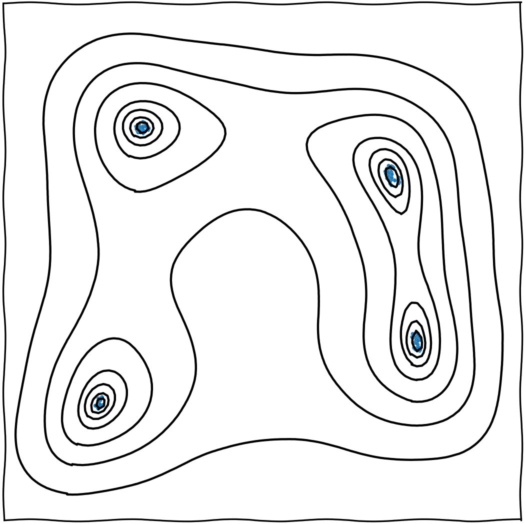
Nested sampling



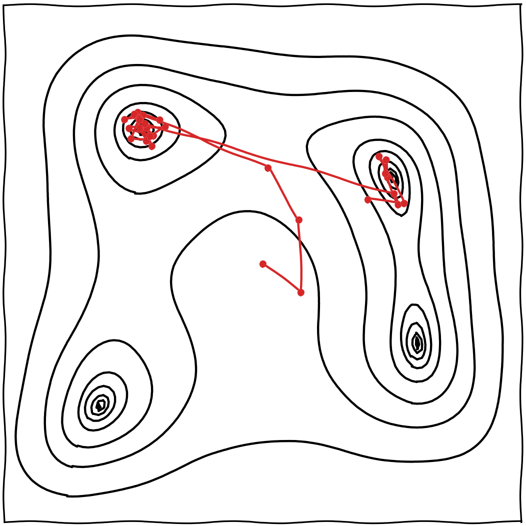
MCMC



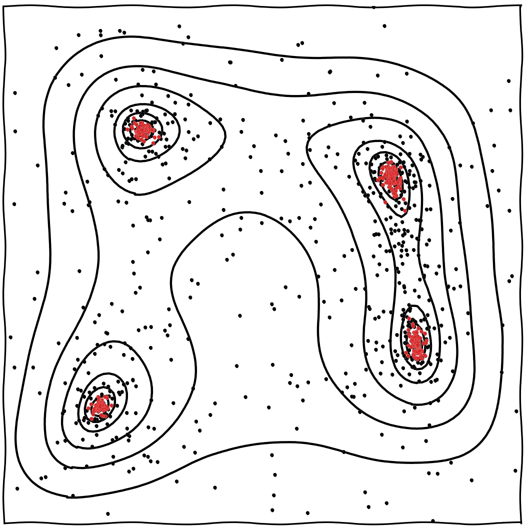
Nested sampling



MCMC



Nested sampling

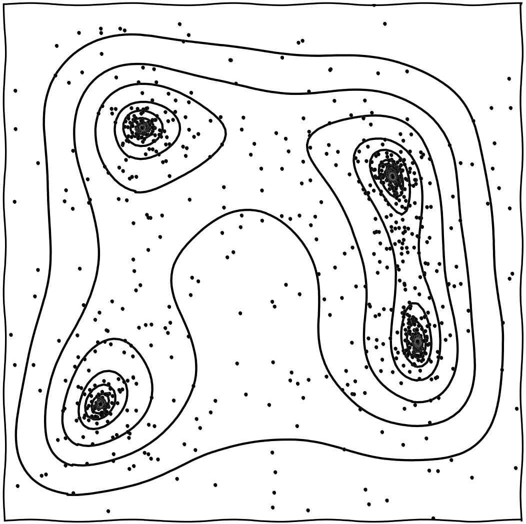


The nested sampling meta-algorithm: Lebesgue integration

- Full dead-point coverage of tails enables integration.
- Can be weighted to form posterior samples, prior samples, or anything in between.
- Nested sampling estimates the **density of states** and calculates partition functions

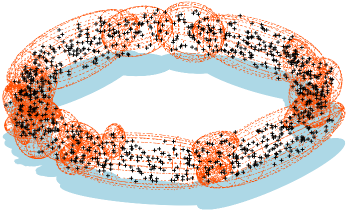
$$Z(\beta) = \sum_i f(x_i)^\beta \Delta V_i.$$

- The evolving ensemble of live points allows:
  - implementations to self-tune
  - exploration of multimodal functions
  - global and local optimisation

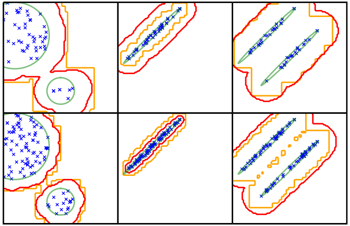


Implementations of Nested Sampling [2205.15570](NatReview)

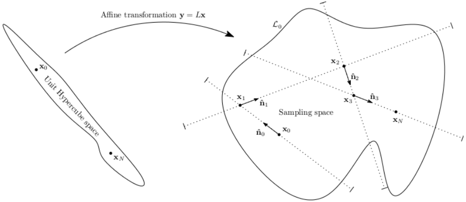
MultiNest [0809.3437]



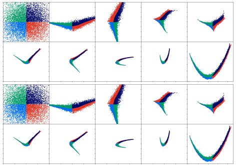
UltraNest [2101.09604]

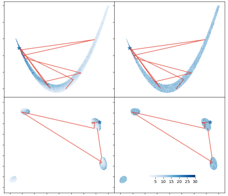


PolyChord [1506.00171]



NeuralNest [1903.10860]

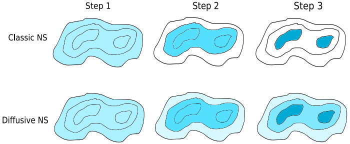




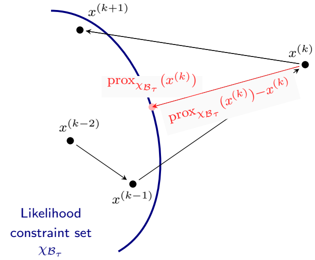
nessai [2102.11056]

nora [2305.19267]

DNest [1606.03757]



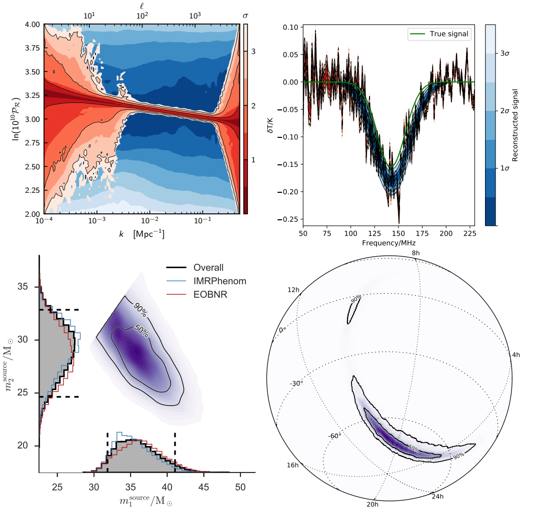
ProxNest [2106.03646]



dynesty [1904.02180]

Standard uses of nested sampling

- Battle-tested in Bayesian cosmology on
  - Parameter estimation: multimodal alternative to MCMC samplers.
  - Model comparison: using integration to compute the Bayesian evidence
  - Tension quantification: using deep tail sampling and suspiciousness computations.
- Plays a critical role in major cosmology pipelines: Planck, DES, KiDS, BAO, SNe.
- The default ΛCDM cosmology is well-tuned to have Gaussian-like posteriors for CMB data.
- Less true for alternative cosmologies/models and orthogonal datasets, so nested sampling crucial.
- Also used in Gravitational Waves & Exoplanets
- Often taken as “ground truth to beat”.



Marginal inference

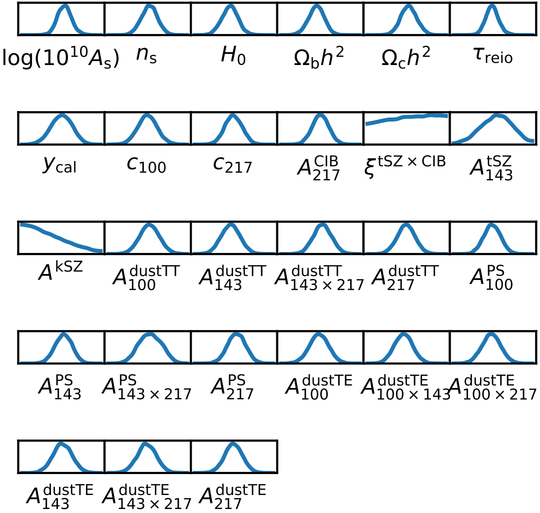
Harry Bevins

PhD→JRF

- Many cosmological likelihoods come with nuisance parameters that have limited relevance for onward inference.
- Notation: CMB cosmology

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

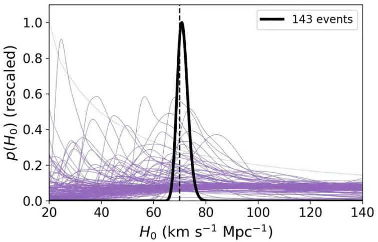


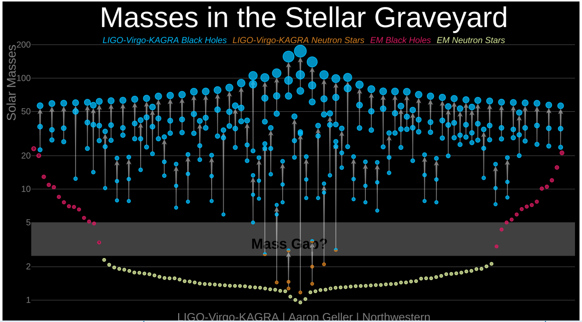
Marginal inference

- Many cosmological likelihoods come with nuisance parameters that have limited relevance for onward inference.
- Notation: GW cosmology

$\mathcal{L}$	Likelihood	(e.g. LAL),
D	Data	(e.g. GW170817),
θ	Cosmological parameters	(e.g. , $H_0\ldots$),
α	Nuisance parameters	(e.g. m_1, $m_2\ldots$),
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 $\mathcal{D}_{\mathrm{KL}}$.





Nuisance marginalised likelihoods: Theory [2207.11457]

- Bayes theorem

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

Likelihood × Prior = Posterior × Evidence

α: nuisance parameters, θ: cosmo parameters.

- Marginal Bayes theorem

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

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

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

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 [2205.12841]

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. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Practice [2205.12841]

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**

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

↓ ↓

Nested Sampling

1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

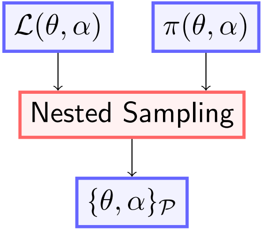
- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

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. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

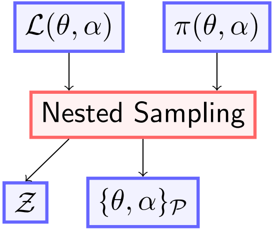
- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Practice [2205.12841]

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**



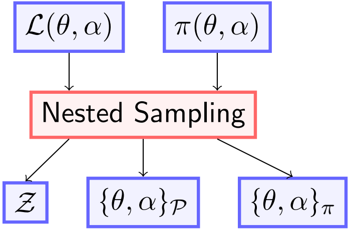
1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Practice [2205.12841]

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

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



1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

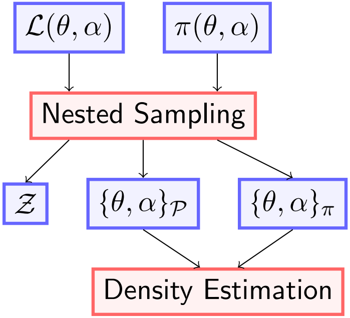
- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Practice [2205.12841]

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**



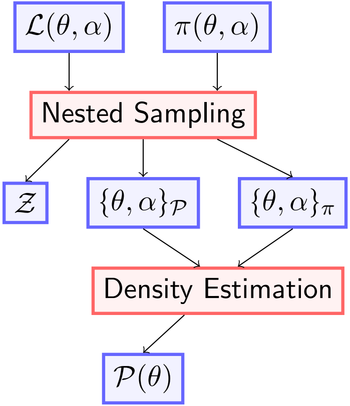
1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

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

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

1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...



- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

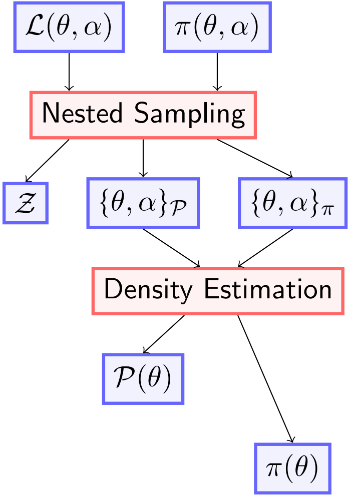
Nuisance marginalised likelihoods: Practice [2205.12841]

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. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...



- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Practice [2205.12841]

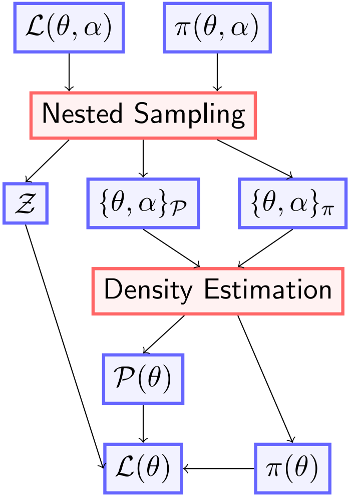
Harry Bevins
PhD→JRF

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

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

1. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...

- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine



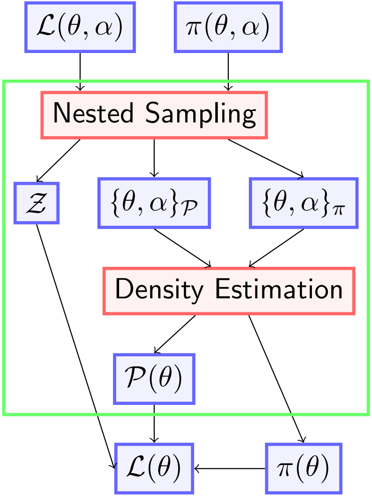
Harry Bevins

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. Bayesian evidence $\mathcal{Z}$: g
   - Nested sampling
   - Parallel tempering (pocomc, ptmcmc)
   - Sequential Monte Carlo (SMC)
   - MCEvidence
2. Marginal prior $\pi(\theta)$ and posterior $\mathcal{P}(\theta)$ densities:
   - Histograms of samples
   - Kernel density estimation
   - Normalising flows / Diffusion models
   - ...



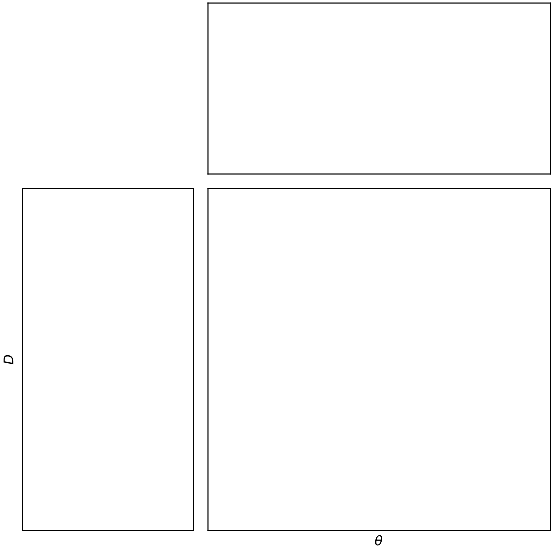
- Emulators usually much faster than original likelihoods
- `margarine`: PyPI, github.com/htjb/margarine

Nuisance marginalised likelihoods: Example uses

- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods (DiRAC allocation `unimpeded`)
- 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.
- Hierarchical modelling:
  - Usually, have N objects, each with nuisance parameters α_i, and shared parameters of interest θ.
  - Likelihood $\mathcal{L}(\{D_i\}|\theta,\{\alpha_i\}) = \prod_i^N \mathcal{L}_i(D_i|\theta,\alpha_i)$ has $N\times \texttt{len}(\alpha_i) + \texttt{len}(\theta)$ parameters
  - Instead, break problem down into N runs on $\texttt{len}(\theta) + \texttt{len}(\alpha_i)$ parameters, and one final one on $\texttt{len}(\theta)$ parameters, using nuisance marginal likelihoods $\mathcal{L}_i(\theta)$.
  - In addition to computational tractability, also can perform model comparison with nuisance marginalised likelihoods.

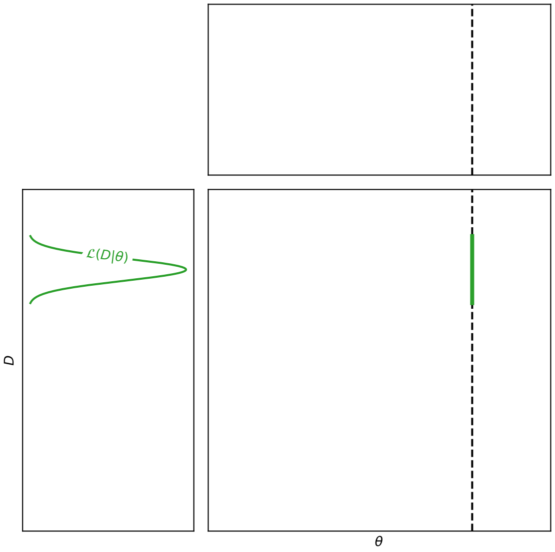
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



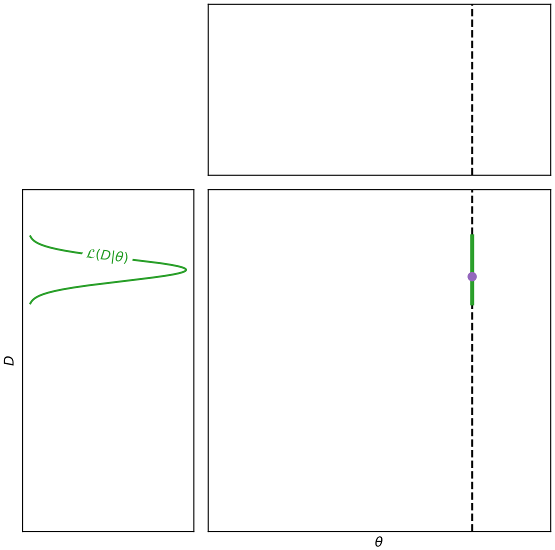
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



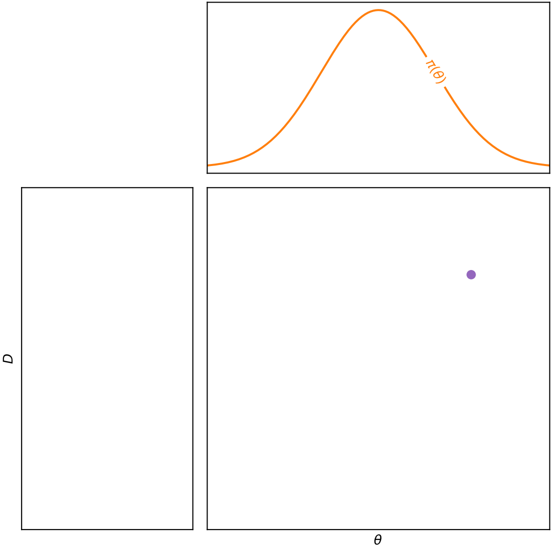
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



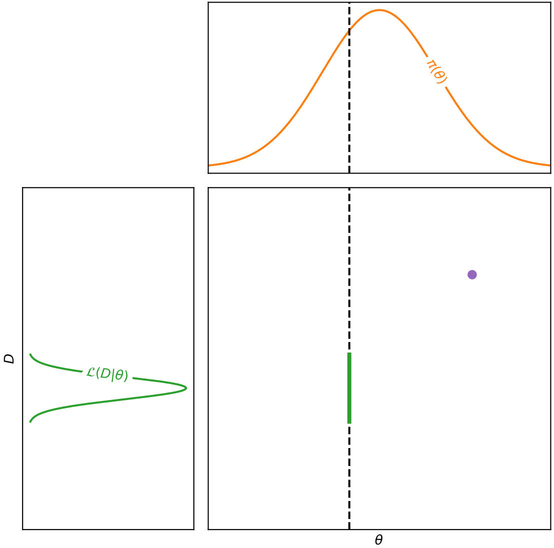
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



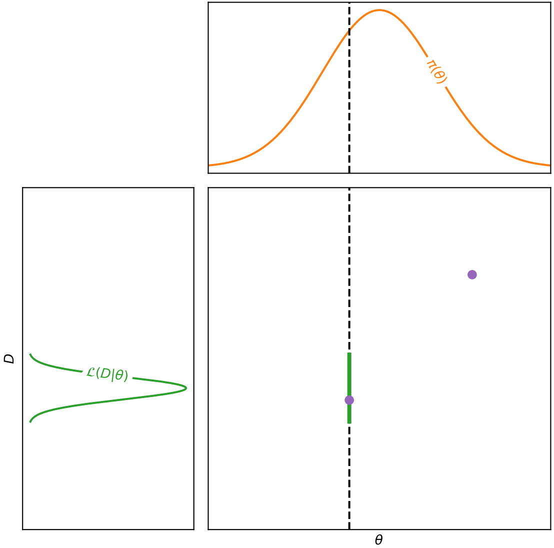
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



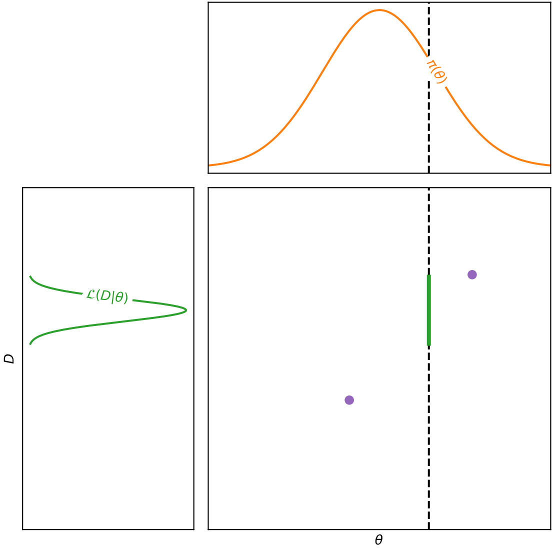
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



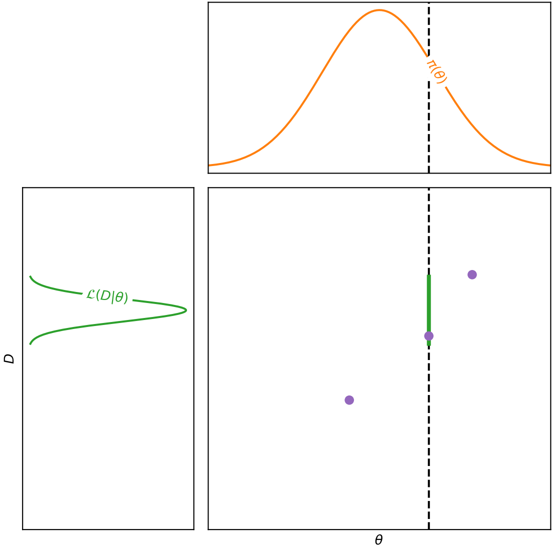
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



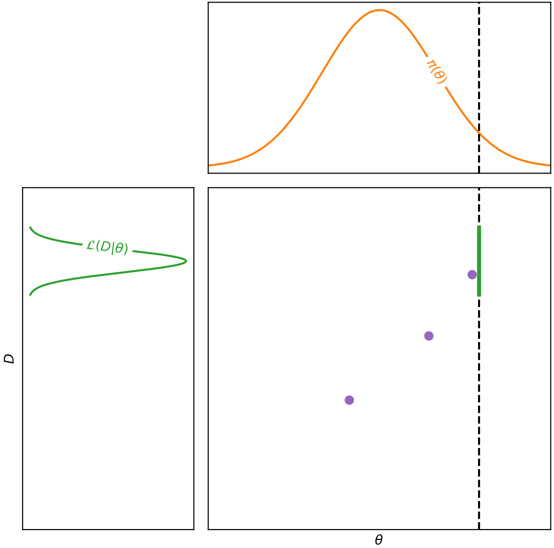
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



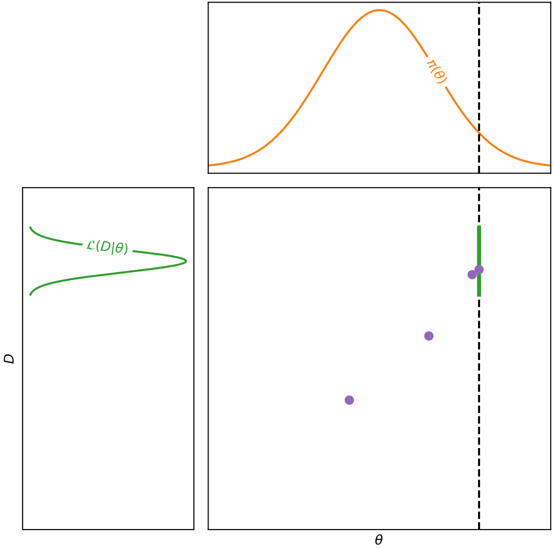
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



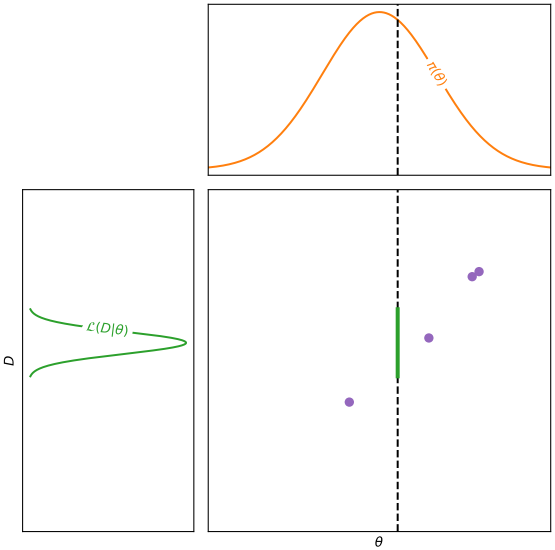
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



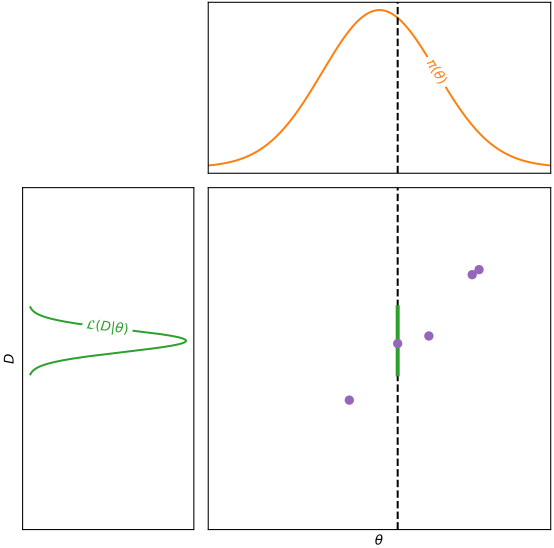
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



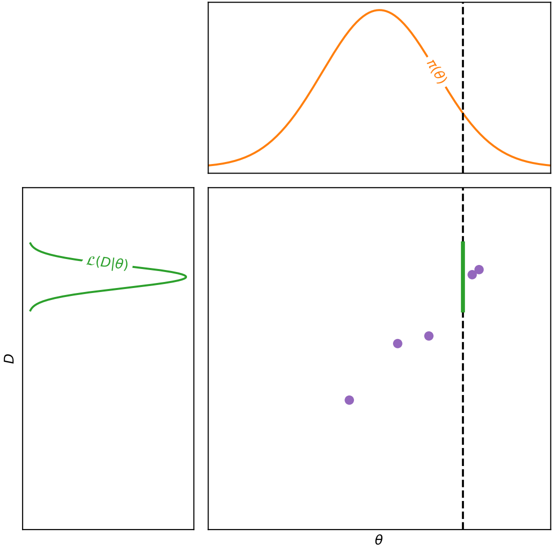
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



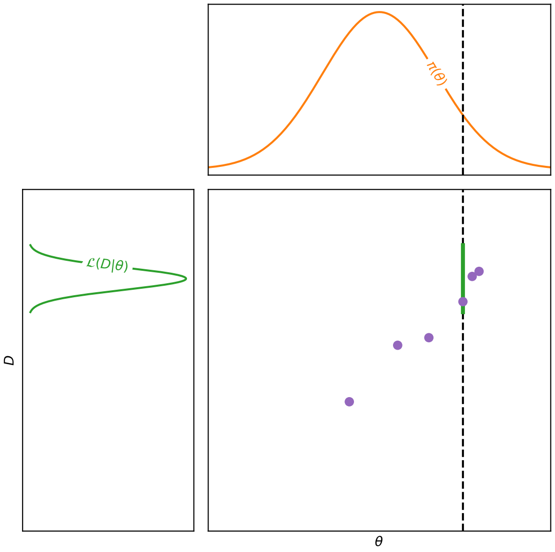
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



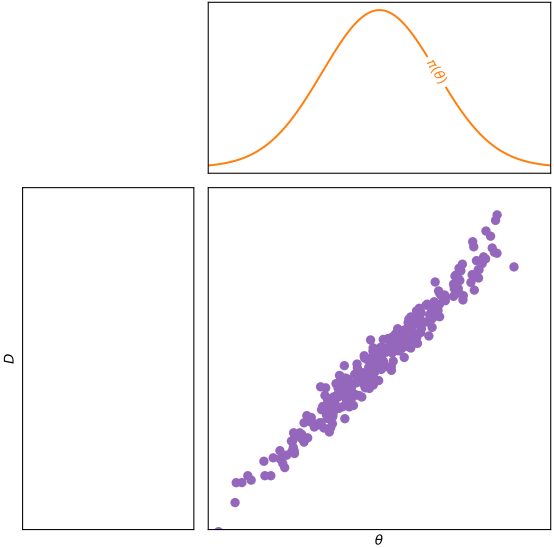
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$
  the "probability of everything".
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



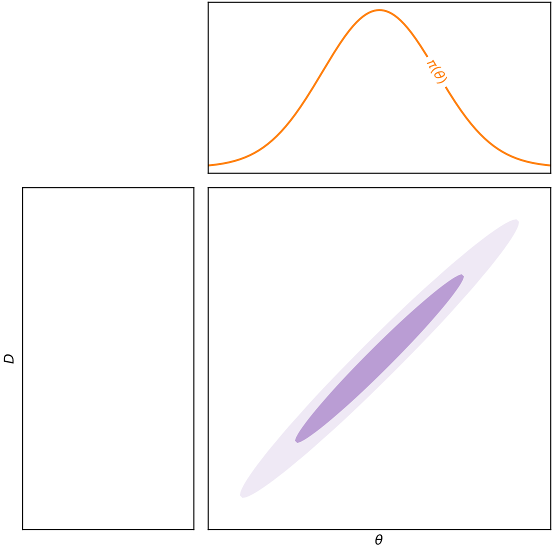
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



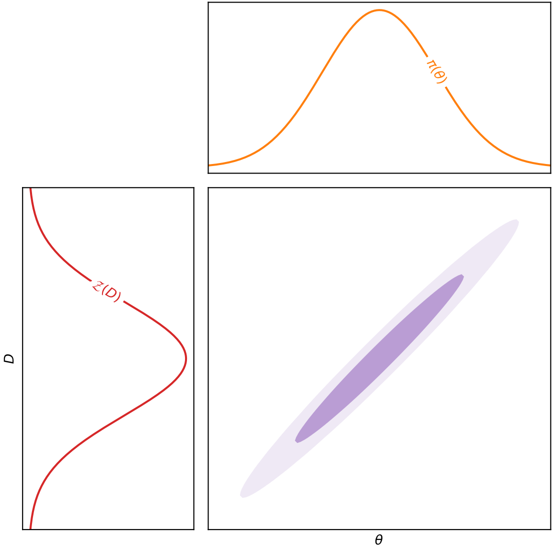
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



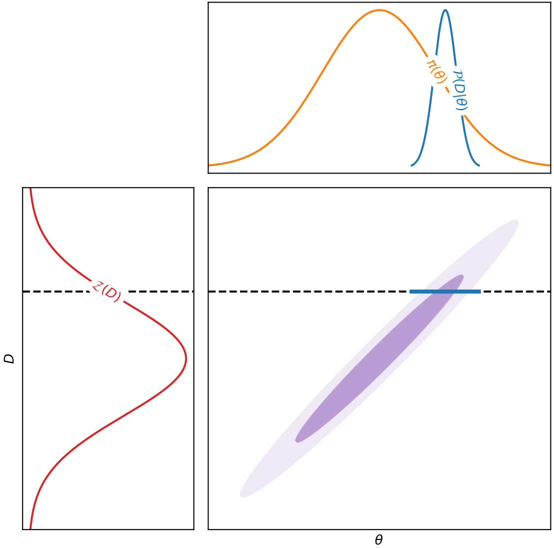
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



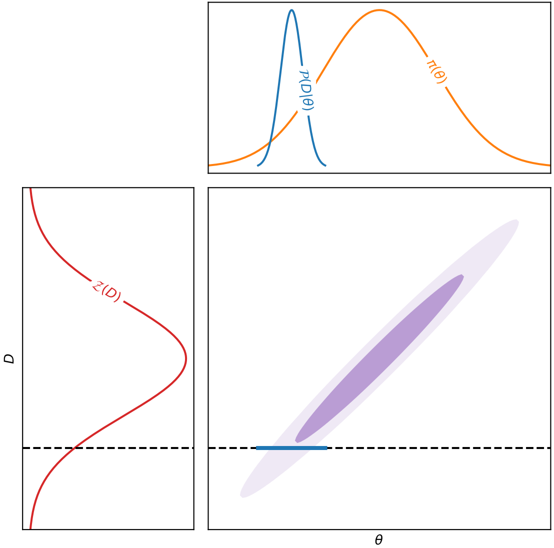
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



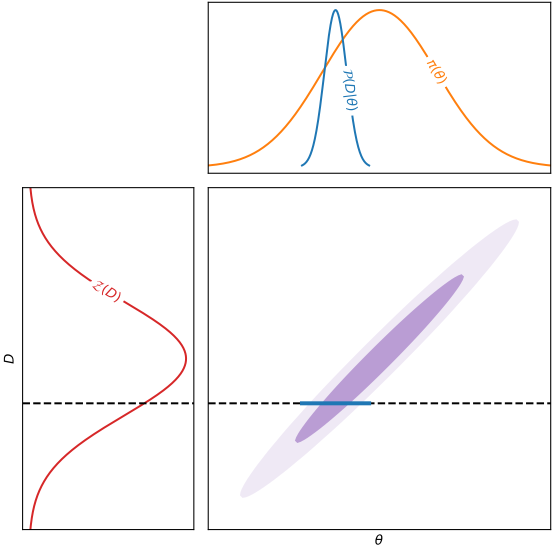
SBI: Simulation-based inference

- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



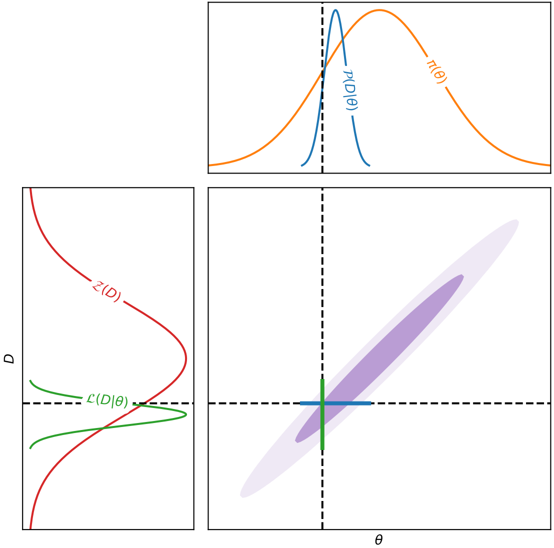
SBI: Simulation-based inference

- Only have access to a forward model $\theta \rightarrow D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



SBI: Simulation-based inference

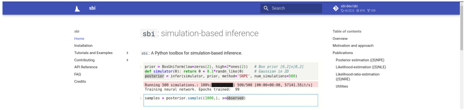
- Only have access to a forward model $\theta \to D$.
- (θ, D) plane gives a more expansive theoretical view of inference.
- Forward model defines *implicit* likelihood $\mathcal{L}$:
- Simulator generates samples from $\mathcal{L}(D|\theta)$.
- With a prior $\pi(\theta)$ can generate samples from joint distribution $\mathcal{J}(\theta, D) = \mathcal{L}(D|\theta)\pi(\theta)$ *the "probability of everything"*.
- Task of SBI is then to go from joint $\mathcal{J}$ samples to posterior $\mathcal{P}(\theta|D)$ and evidence $\mathcal{Z}(D)$ – and possibly likelihood $\mathcal{L}(D|\theta)$.
- Present SotA: NPE, NLE, NJE, NRE
- SBI & forward modelling force us to think about data space D & parameter space θ.



Why SBI?

SBI is useful because:

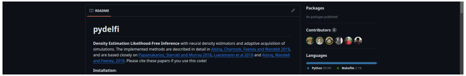
1. If you don't have a likelihood, you can still do inference
    - The usual case beyond CMB cosmology
2. Faster than LBI
    - emulation – also applies to LBI in principle
3. No need to pragmatically encode fiducial cosmologies
    - Covariance computation implicitly encoded in simulations
4. Equips AI/ML with Bayesian interpretability
5. Lower barrier to entry than LBI
    - Much easier to forward model a systematic
    - Emerging set of plug-and-play packages
    - For this reason alone, it will come to dominate scientific inference



github.com/sbi-dev

github.com/undark-lab/swyft

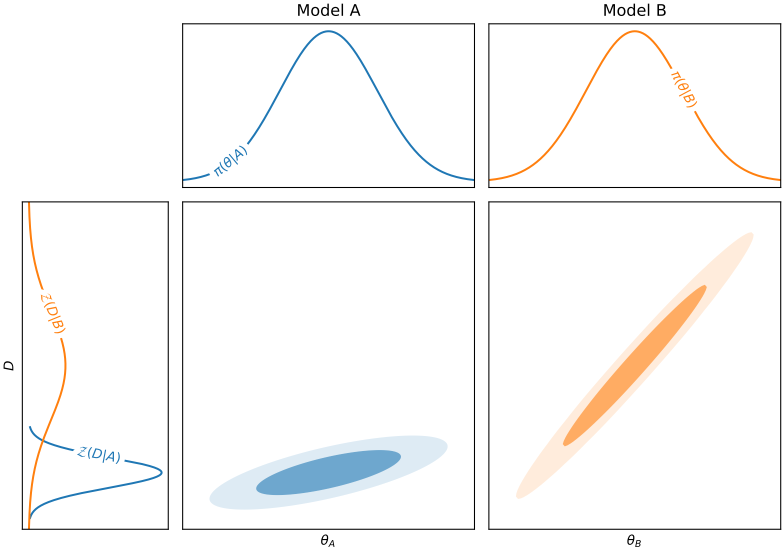
github.com/florent-leclercq/pyselfi



github.com/justinalsing/pydelfi

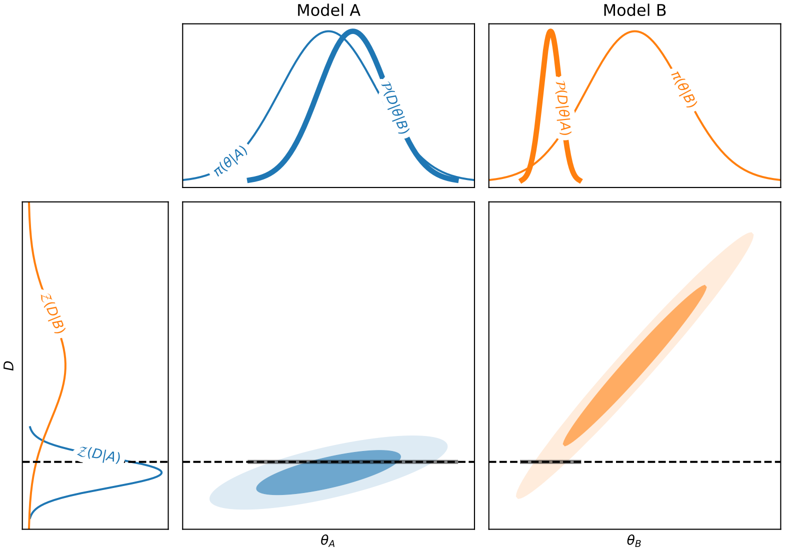
SBI & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor: more predictive $\equiv$ more probable (due to normalisation).



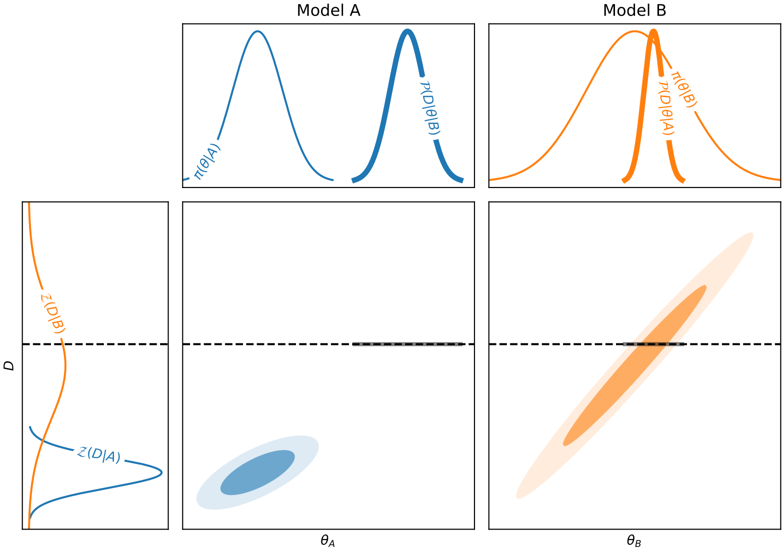
SBI & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor:
  more predictive
  $\equiv$ more probable
  (due to normalisation).



SBI & model comparison

- Extend: models A and B.
- Each with own separate parameters θ_A and θ_B (can be same).
- The evidence $\mathcal{Z}(D|M)$ compares models
- Occams razor: more predictive $\equiv$ more probable (due to normalisation).



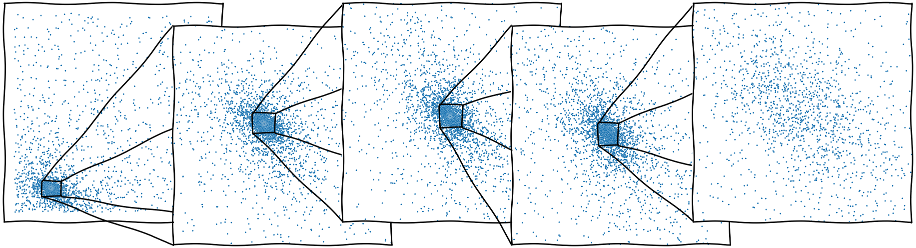
Kilian Scheutwinkel
PhD

Weak SBI

- Use model comparison to choose between likelihoods
- Use flexible likelihood (e.g. unknown noise scale σ, non-gaussian shape, mixture components [1809.04598])
- 21cm [2204.04491], SNe [2312.02075]

Strong SBI

- Develop "likelihood-free nested sampling"
- Use dead points to train NRE
- Replaces/enhances current SotA of truncation techniques



Cosmological forecasting

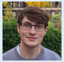
Have you ever done a Fisher forecast, and then felt Bayesian guilt?

- 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.

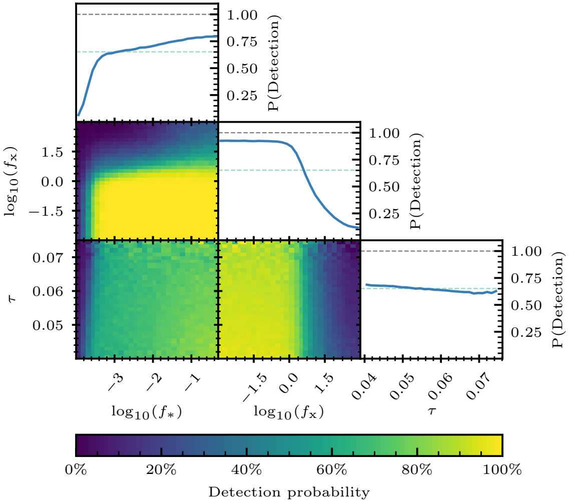
Fully Bayesian Forecasting [2309.06942]

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`



Conclusions

github.com/handley-lab

Covered a suite of tools for next-generation "generative cosmology"

- Nested sampling for cosmological...
  - model comparison
  - parameter estimation
  - tension quantification
- Nuisance-marginalised cosmology $\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$
- Simulation-based inference
- Fully Bayesian forecasting