Accelerated nested sampling with β-flows

Metha Prathaban
myp23@cam.ac.uk

May 13, 2025

About Me

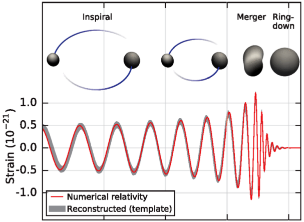
- 3rd year PhD student in Will Handley's group
- Work on gravitational waves and Bayesian numerical method development
- Office in K34

The Handley group!

\+ Part III students + others!

Current work is in collaboration with Will Handley and Harry Bevins.

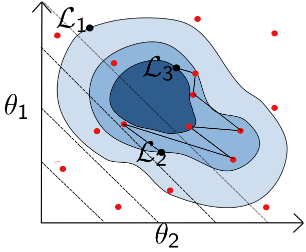

Given some model $\mathcal{M}$ and observed signal $\mathcal{D}$, Bayes' theorem enables us to relate the posterior probability of the set of parameters θ which generated the signal to the likelihood of the $\mathcal{D}$ given θ and the prior probability of θ given $\mathcal{M}$:



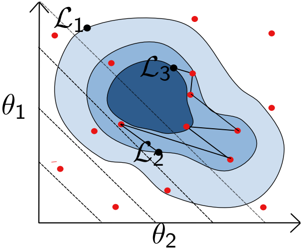
$$P(\theta|D,\mathcal{M}) = \frac{P(D|\theta,\mathcal{M})P(\theta|\mathcal{M})}{P(D|\mathcal{M})} = \frac{\mathcal{L}(D|\theta)\pi(\theta)}{\mathcal{Z}}$$

Abbott et al.
arXiv:1602.03837

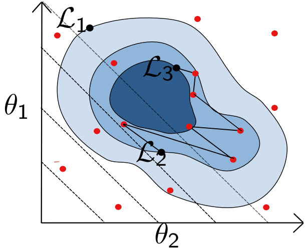
The evidence, $\mathcal{Z}$, plays a key role in model comparison.



- Nested sampling first and foremost calculates the evidence (MultiNest 0809.3437, dynesty 1904.02180, nessai 2102.11056 and PolyChord 1506.00171)

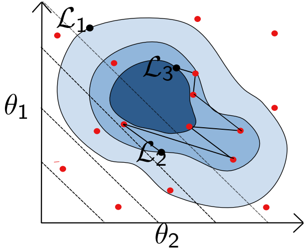


- Nested sampling first and foremost calculates the evidence (MultiNest 0809.3437, dynesty 1904.02180, nessai 2102.11056 and PolyChord 1506.00171)
- Prior is populated with set of 'live points'.



- Nested sampling first and foremost calculates the evidence (MultiNest 0809.3437, dynesty 1904.02180, nessai 2102.11056 and PolyChord 1506.00171)
- Prior is populated with set of 'live points'.
- At each iteration i, point with lowest likelihood is deleted and new live point is drawn, which must have a **likelihood higher than that of the deleted point.**

Nested sampling (NS)



- Nested sampling first and foremost calculates the evidence (MultiNest 0809.3437, dynesty 1904.02180, nessai 2102.11056 and PolyChord 1506.00171)
- Prior is populated with set of 'live points'.
- At each iteration i, point with lowest likelihood is deleted and new live point is drawn, which must have a **likelihood higher than that of the deleted point.**
- Evidence is calculated as weighted sum over deleted ('dead') points.

2404.16428

Time of convergence of NS

$$T \propto n_{\mathrm{live}} \times T_{\mathcal{L}} \times f_{\mathrm{sampler}} \times D_{\mathrm{KL}} \tag{1}$$

Uncertainty in $\log \mathcal{Z}$

$$\sigma \propto \sqrt{D_{\mathrm{KL}}/n_{\mathrm{live}}} \tag{2}$$

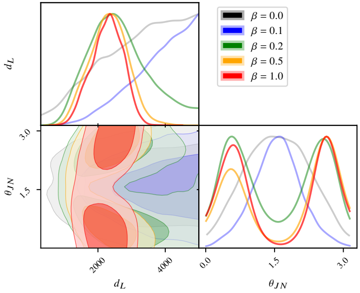
Precision-normalized runtime has quadratic dependence on KL divergence – want to reduce amount of compression from prior to posterior. Petrosyan & Handley 212.01760

- REACH collaboration do this already (see e.g. Anstey et al. 2010.09644):
  - Perform a low-res pass of NS
  - Identify rough region of posterior and set box prior around this
  - Full high-res run with narrower prior
  - Need to correct the evidence as the prior has changed.

- REACH collaboration do this already (see e.g. Anstey et al. 2010.09644):
  - Perform a low-res pass of NS
  - Identify rough region of posterior and set box prior around this
  - Full high-res run with narrower prior
  - Need to correct the evidence as the prior has changed.
- Can use **machine learning** to learn the rough posterior distribution from pass 1, and use as proposal for pass 2.

- REACH collaboration do this already (see e.g. Anstey et al. 2010.09644):
  - Perform a low-res pass of NS
  - Identify rough region of posterior and set box prior around this
  - Full high-res run with narrower prior
  - Need to correct the evidence as the prior has changed.
- Can use **machine learning** to learn the rough posterior distribution from pass 1, and use as proposal for pass 2.
- Posterior repartitioning (PR) (Chen et al. 1803.06387 & 1908.04655) allows one to do this **without needing to correct evidence** (see e.g. Petrosyan & Handley 2212.0176).

β-flows vs typical NFs



- NFs can perform density estimation of posteriors from NS/other sampling method.
- Nested sampling gives you a lot more information than just the posterior samples - we have access to samples at every "temperature".
- In statistical mechanics, this temperature is used in context of partition function.
- Idea of β-flows is to explore this arm of NS to learn posterior better.

Thank you for listening!

Unlike other sampling algorithms, such as Metropolis-Hastings or Hamiltonian Monte Carlo, NS distinguishes between $\mathcal{L}$ and π by 'sampling from the prior, subject to the hard likelihood constraint, $\mathcal{L}$'.

But evidence and posteriors only depend on product of $\mathcal{L}$ and π:

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta \tag{3}$$

$$\mathcal{P}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\mathcal{Z}} \tag{4}$$

Therefore, we are free to redefine the likelihood and prior however we like - as long as the product is the same! arXiv:1908.04655

$$\tilde{\mathcal{Z}} = \int \tilde{\mathcal{L}}(\theta)\tilde{\pi}(\theta)d\theta = \int \mathcal{L}(\theta)\pi(\theta)d\theta = \mathcal{Z} \tag{5}$$