Efficient Cosmological Model Selection using Bayesian Optimisation

Nathan Cohen, Jan Hamann and **Ameek Malhotra**

Physics Department, Swansea University, SA2 8PP, United Kingdom

email: `ameek.malhotra@swansea.ac.uk`

Bayesian model comparison

Model M_0 vs M_1 given data D, the Bayesian way:

$$\frac{P(M_0|D)}{P(M_1|D)} = B_{01}\frac{P(M_0)}{P(M_1)}, \quad B_{01} = \frac{P(D|M_0)}{P(D|M_1)}$$

$$\underbrace{\mathcal{Z} = P(D|M)}_{\text{Bayesian evidence}} = \int \underbrace{P(D|M,\theta)P(\theta|M)}_{\text{Posterior = Likelihood} \times \text{Prior}} \, d\theta$$

$\mathcal{Z}$ can be computed from MCMC chains (Savage–Dickey density ratio) or using nested sampling.

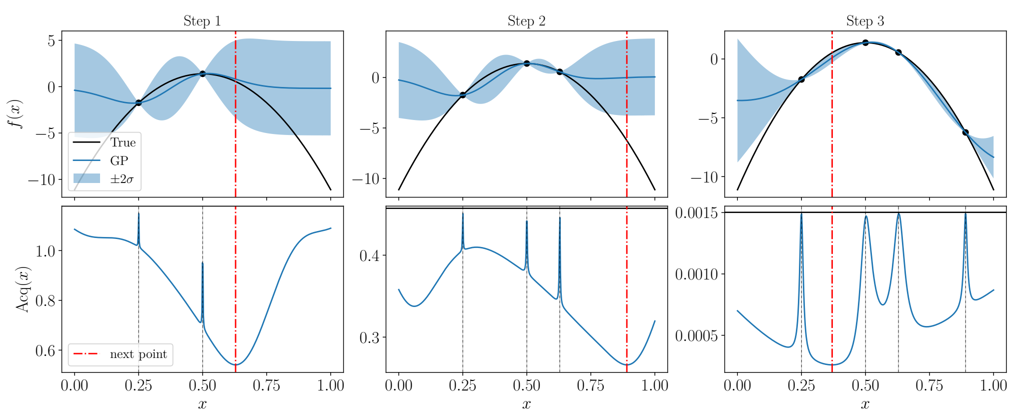
Bayesian Optimisation (BO)

When to use it?

MCMC/nested sampling not feasible for *slow to evaluate* likelihoods – instead *train a fast surrogate of the likelihood using BO*, then run MCMC/nested sampling on surrogate.

How does it work?

Uses *mean of Gaussian process* (GP) [1] as surrogate for the (log)-likelihood + *active acquisition strategy* to train the GP. Uncertainty of GP → $\Delta \log \mathcal{Z}$ for *convergence control*.



Bayesian Optimisation in action.

Algorithm

Input: GP μ, acquisition α, $\Delta \log \mathcal{Z}$ threshold ϵ, max steps N
Initialise training set $\mathcal{D}$ with random Sobol points
while $i < N$ and $\neg$converged **do**
  $x_i \leftarrow \arg\min_x \alpha(x \mid \mu, \mathcal{D})$, $y_i \leftarrow f(x_i)$
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x_i, y_i)\}$
  Update μ with $\mathcal{D}$
  converged=True if $\Delta \log Z \leq \epsilon$ else False
  $i \leftarrow i + 1$
end while
return samples from *surrogate* posterior μ, final $\log \mathcal{Z}$, $\Delta \log \mathcal{Z}$

Pros

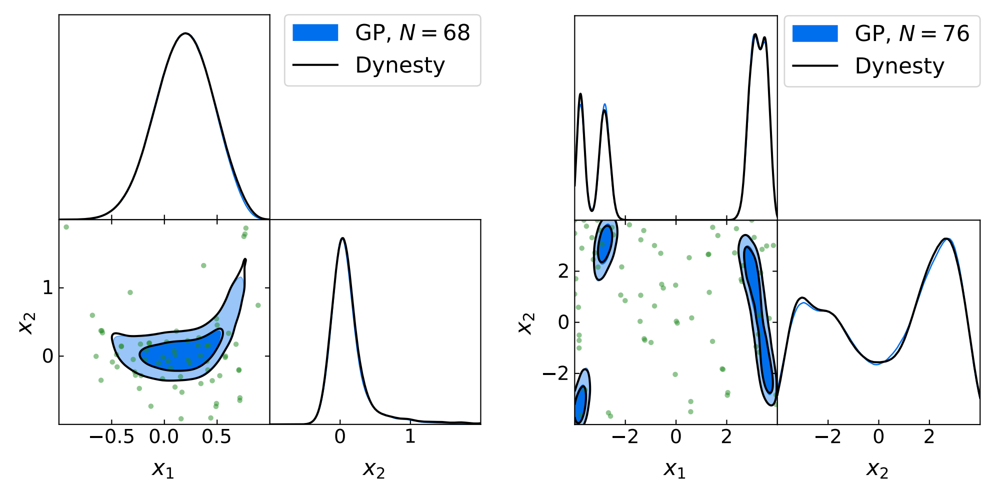
Fast GP surrogate ($\gtrsim$ 1000 eval/sec). 10–100 times *fewer likelihood evals* required compared to traditional methods, *overall faster for slow likelihoods* ($t \gtrsim$ 1s).

Cons

Greater computational overhead of intermediate steps.
Limited to $N_{\text{dim}} \lesssim 20$.

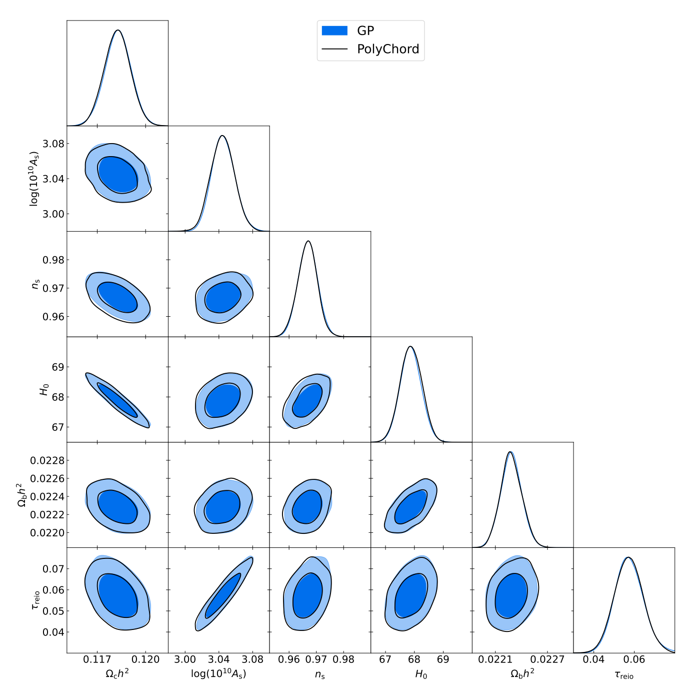
Results

Test functions



BO vs nested sampler dynesty [2]. Green dots = GP training points (N) from BO.

Cosmology



LCDM from DESI+Planck CamSpec (9 nuisance params). BO with $N \sim 900$ likelihood evals and runtime ~ 10 hours, $\log \mathcal{Z} = -5529.48^{+2.5}_{-0.6}$, excellent agreement with PolyChord [3].

Prev. applications in cosmology – for global optimisation [4], and parameter posteriors [5]. *We build on these but differ in GP, acquisition and convergence choices.*

Paper (more examples) + code (JAX based) out soon. If interested in applications, please get in touch!

References

[1] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT Press Cambridge, MA, 2006.

[2] J. S. Speagle, *dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences*, *Monthly Notices of the Royal Astronomical Society* **493** (Feb., 2020) 3132–3158.

[3] W. J. Handley, M. P. Hobson and A. N. Lasenby, *PolyChord: nested sampling for cosmology*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) L61–L65, [1502.01856].

[4] J. Hamann and J. Wons, *Optimising inflationary features the Bayesian way*, *JCAP* **03** (2022) 036, [2112.08571].

[5] J. E. Gammal, N. Schöneberg, J. Torrado and C. Fidler, *Fast and robust Bayesian inference using Gaussian processes with GPry*, *JCAP* **10** (2023) 021, [2211.02045].

LaTeX TikZposter