

# Nested sampling: powering next-generation inference and machine learning tools for astrophysics, cosmology, particle physics and beyond

Will Handley

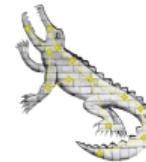
[wh260@cam.ac.uk](mailto:wh260@cam.ac.uk)

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

18<sup>th</sup> July 2024



UNIVERSITY OF  
CAMBRIDGE

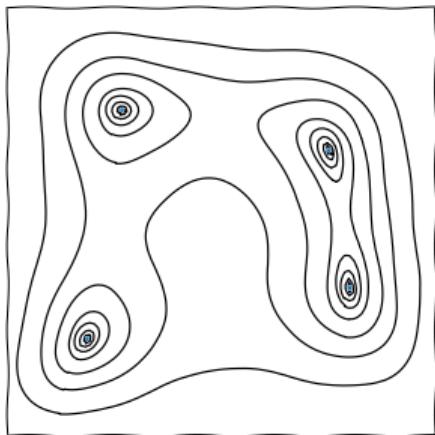


# What is Nested Sampling?

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

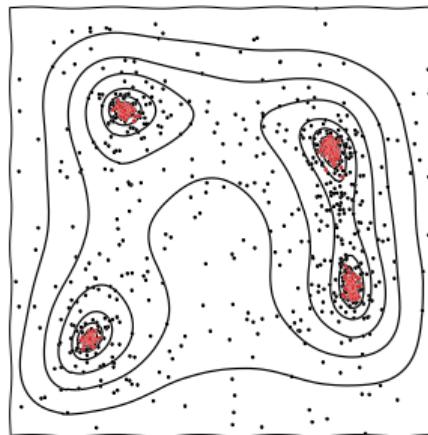
## Optimisation

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



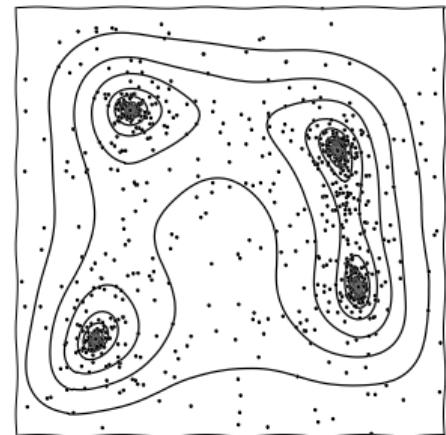
## Exploration

draw/sample  $\theta \sim f$



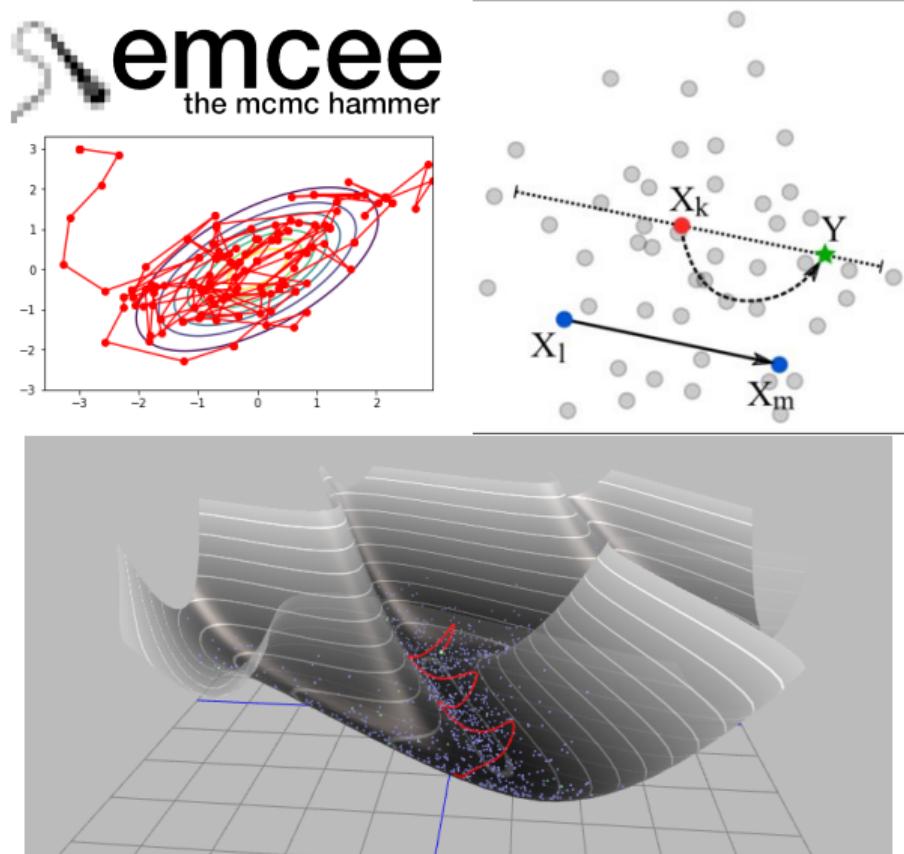
## Integration

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



# Where is Nested Sampling?

- ▶ For many purposes, in your Neural Net you should group Nested Sampling with (MCMC) techniques such as:
  - ▶ Metropolis-Hastings (PyMC, MontePython)
  - ▶ Hamiltonian Monte Carlo (Stan, blackjax)
  - ▶ Ensemble sampling (emcee, zeus).
  - ▶ Variational Inference (Pyro)
  - ▶ Sequential Monte Carlo
  - ▶ Thermodynamic integration
  - ▶ Genetic algorithms
- ▶ You may have heard of it branded form:
  - ▶ MultiNest
  - ▶ PolyChord
  - ▶ dynesty
  - ▶ ultranest



# Integration in Physics

- ▶ Integration is a fundamental concept in physics, statistics and data science:

## Partition functions

$$Z(\beta) = \int e^{-\beta H(q,p)} dq dp$$

## Path integrals

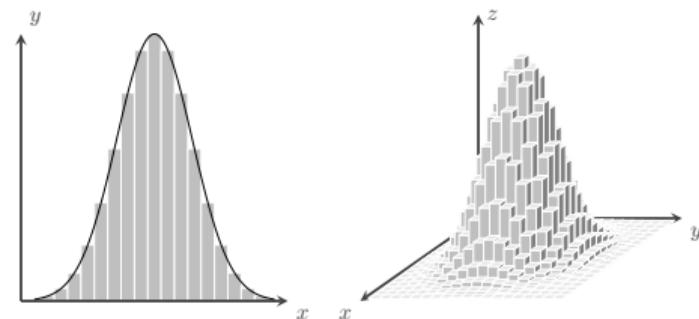
$$\Psi = \int e^{iS} \mathcal{D}x$$

## Bayesian marginals

$$\mathcal{Z}(D) = \int \mathcal{L}(D|\theta) \pi(\theta) d\theta$$

- ▶ Need numerical tools if analytic solution unavailable.
- ▶ High-dimensional numerical integration is hard.
- ▶ Riemannian strategy estimates volumes geometrically:

$$\int f(x) d^n x \approx \sum_i f(x_i) \Delta V_i \sim \mathcal{O}(e^n)$$



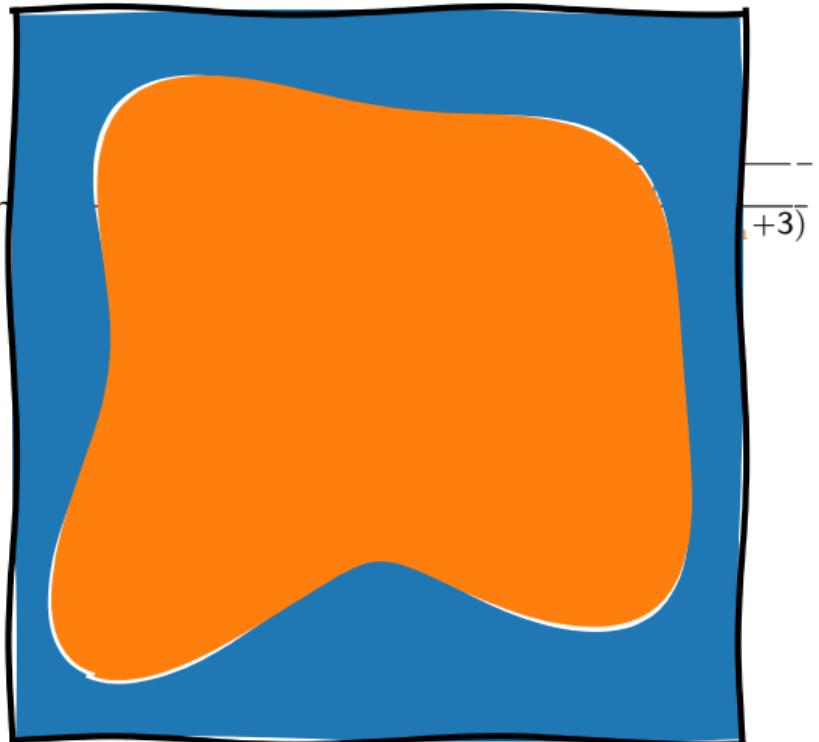
- ▶ Curse of dimensionality  $\Rightarrow$  exponential scaling.

# Probabalistic volume estimation

- ▶ Key idea in NS: estimating volumes probabilistically

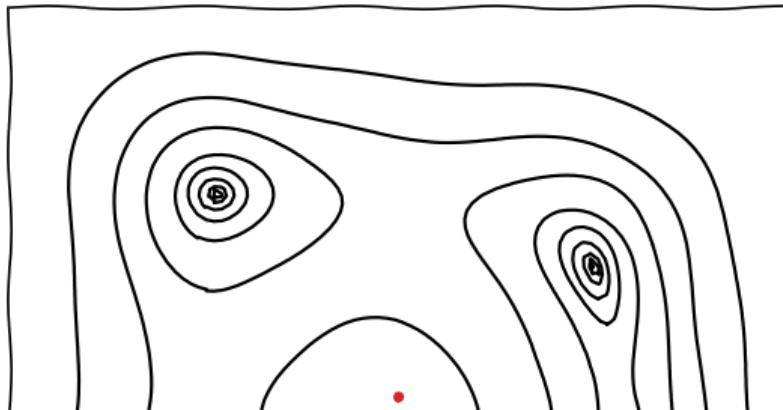
$$\frac{V_{\text{after}}}{V_{\text{before}}} \approx \frac{n_{\text{in}}}{n_{\text{out}} + n_{\text{in}}} \quad \frac{V_{\text{after}}}{V_{\text{before}}} \approx \frac{n_{\text{in}} + 1}{n_{\text{out}} + n_{\text{in}} + \sqrt{2}}$$

- ▶ This is the **only** way to calculate volume in high dimensions  $d > 3$ .
  - ▶ Geometry is exponentially inefficient.
- ▶ This estimation process does not depend on geometry, topology or dimensionality
- ▶ Basis of all Monte-Carlo integration
- ▶ Nested Sampling uniquely uses a nested framework to couple together MC integrals in a robust, scalable manner.



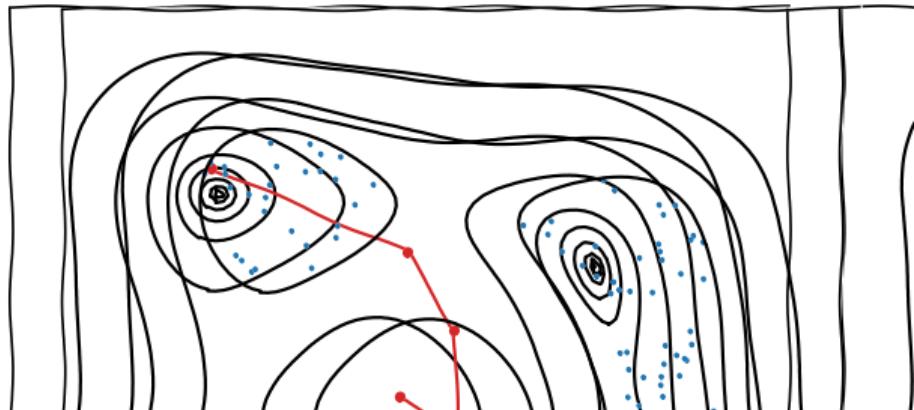
## MCMC

- ▶ Single “walker”
- ▶ Explores posterior
- ▶ Fast, if proposal matrix is tuned
- ▶ Parameter estimation, suspiciousness calculation
- ▶ Channel capacity optimised for generating posterior samples



## Nested sampling

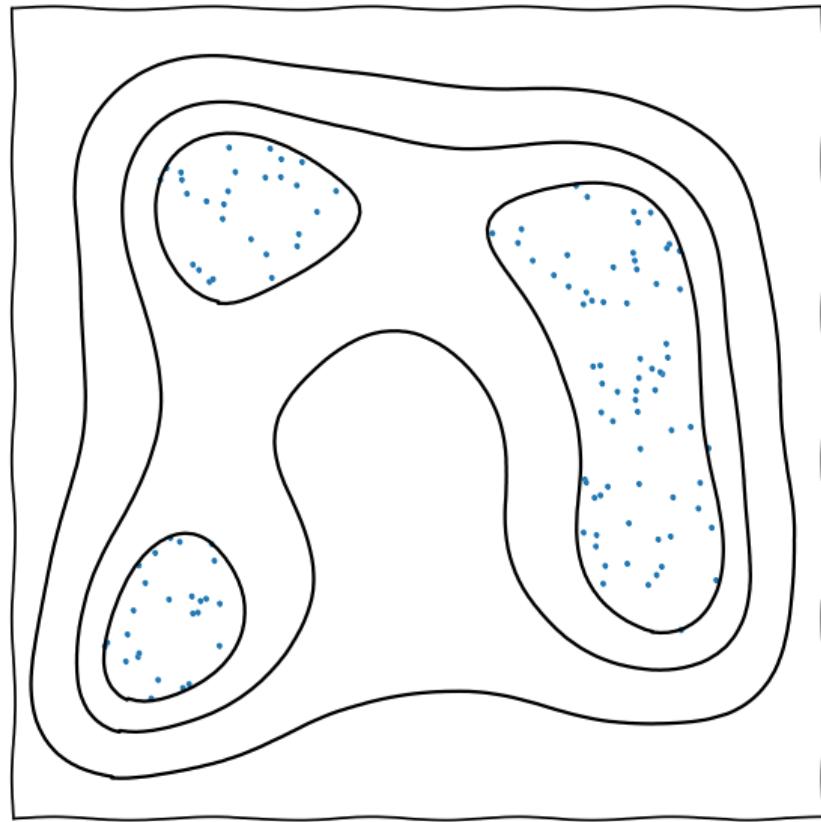
- ▶ Ensemble of “live points”
- ▶ Scans from prior to peak of likelihood
- ▶ Slower, no tuning required
- ▶ Parameter estimation, model comparison, tension quantification
- ▶ Channel capacity optimised for computing partition function



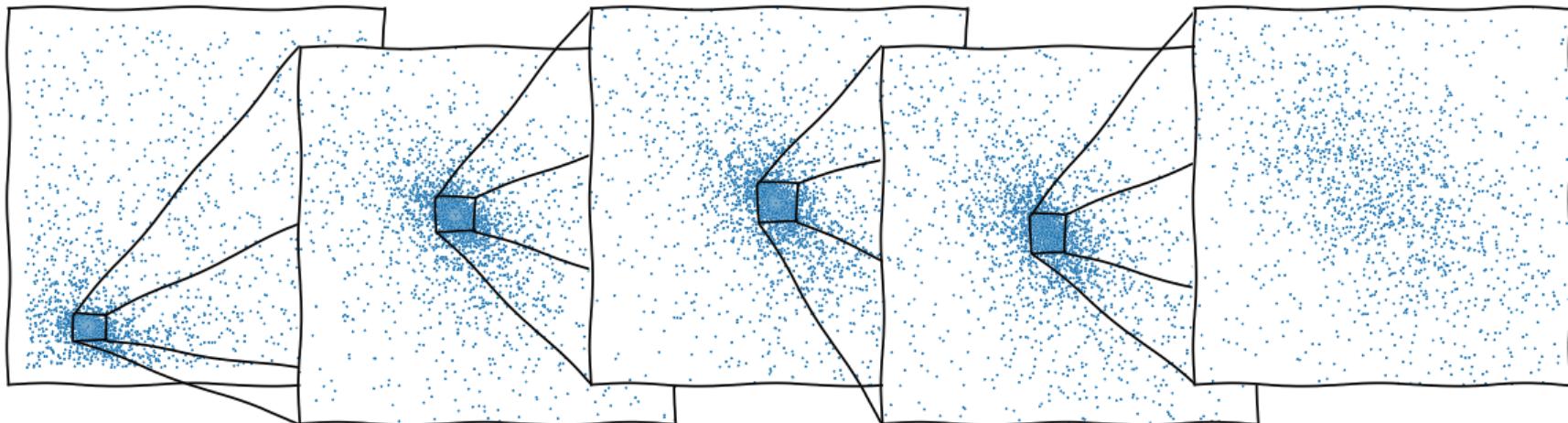
## The nested sampling meta-algorithm: live points

- ▶ Start with  $n$  random samples over the space.
- ▶ Delete outermost sample, and replace with a new random one at higher integrand value.
- ▶ The “live points” steadily contract around the peak(s) of the function.
- ▶ We can use this evolution to estimate volume *probabilistically*.
- ▶ At each iteration, the contours contract by  $\sim \frac{1}{n} \pm \frac{1}{n}$  of their volume.
- ▶ This is an exponential contraction, so

$$\int f(x)dV \approx \sum_i f(x_i)\Delta V_i, \quad V_i = V_0 e^{-(i \pm \sqrt{i})/n}$$



# The nested sampling meta-algorithm: dead points



- ▶ At the end, one is left with a set of discarded “dead” points.
- ▶ Dead points have a unique scale-invariant distribution  $\propto \frac{dV}{V}$ .
- ▶ Uniform over original region, exponentially concentrating on region of interest (until termination volume).
- ▶ Good for training emulators (HERA [[2108.07282](#)]).

## Applications

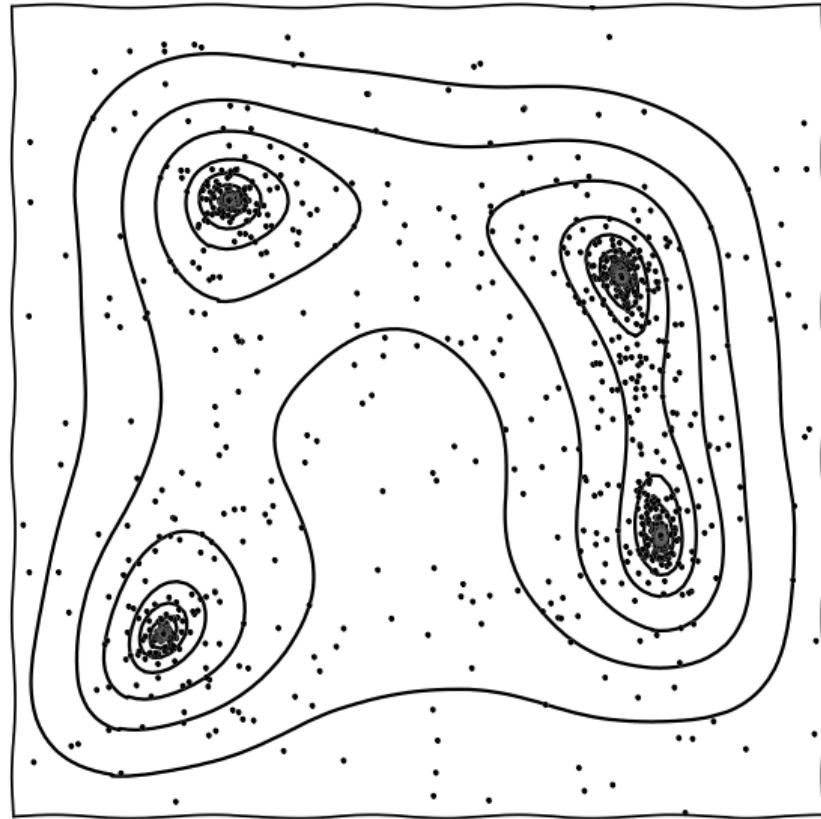
- ▶ training emulators.
- ▶ gridding simulations
- ▶ beta flows
- ▶ “dead measure”

# 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



# Sampling from a hard likelihood constraint

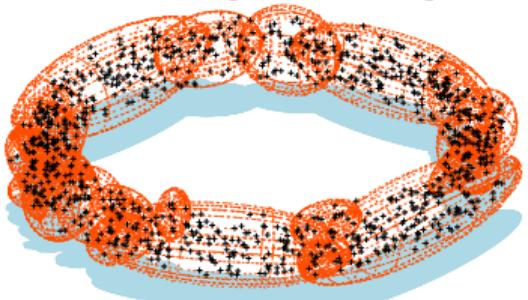
*"It is not the purpose of this introductory paper to develop the technology of navigation within such a volume. We merely note that exploring a hard-edged likelihood-constrained domain should prove to be neither more nor less demanding than exploring a likelihood-weighted space."*

— John Skilling

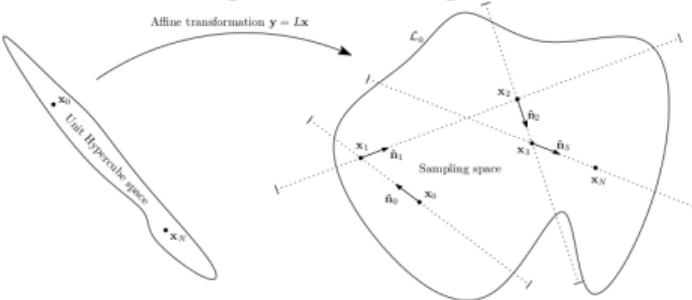
- ▶ A large fraction of the work in NS to date has been in attempting to implement a hard-edged sampler in the NS meta-algorithm  $\{\theta \sim \pi : \mathcal{L}(\theta) > \mathcal{L}_*\}$ .
- ▶ <https://projecteuclid.org/euclid.ba/1340370944>.
- ▶ There has also been much work beyond this (see 'Frontiers of nested sampling' talk)
  - ▶ [willhandley.co.uk/talks](http://willhandley.co.uk/talks)

# Implementations of Nested Sampling [2205.15570](NatReview)

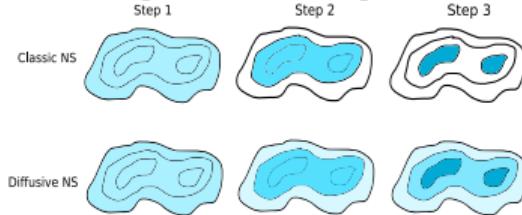
MultiNest [0809.3437]



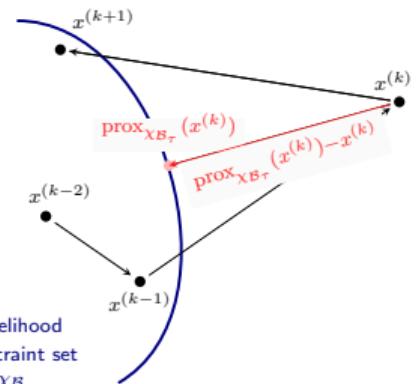
PolyChord [1506.00171]



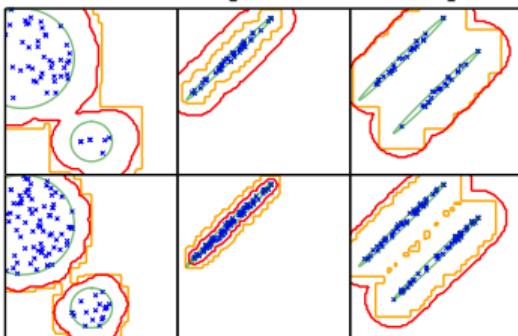
DNest [1606.03757]



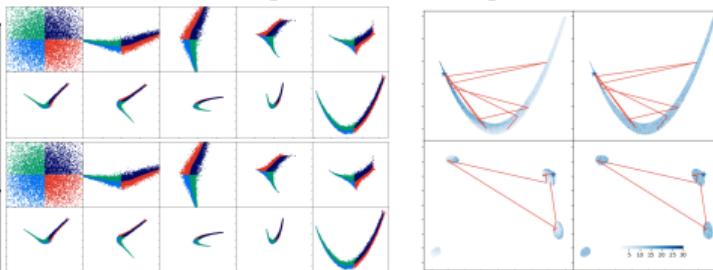
ProxNest [2106.03646]



UltraNest [2101.09604]



NeuralNest [1903.10860]



nessai [2102.11056]

nora [2305.19267]

jaxnest [2012.15286]

nautilus [2306.16923]

<wh260@cam.ac.uk>

willhandley.co.uk/talks

dynesty [1904.02180]

# Types of nested sampler

- ▶ Broadly, most nested samplers can be split into how they create new live points.
- ▶ i.e. how they sample from the hard likelihood constraint  $\{\theta \sim \pi : \mathcal{L}(\theta) > \mathcal{L}_*\}$ .

## Rejection samplers

- ▶ e.g. MultiNest, UltraNest.
- ▶ Constructs bounding region and draws many invalid points until  $\mathcal{L}(\theta) > \mathcal{L}_*$ .
- ▶ Efficient in low dimensions, exponentially inefficient  $\sim \mathcal{O}(e^{d/d_0})$  in high  $d > d_0 \sim 10$ .

- ▶ Nested samplers usually come with:

- ▶ *resolution* parameter  $n_{\text{live}}$  (which improve results as  $\sim \mathcal{O}(n_{\text{live}}^{-1/2})$ ).
- ▶ set of *reliability* parameters [2101.04525], which don't improve results if set arbitrarily high, but introduce systematic errors if set too low.
- ▶ e.g. Multinest efficiency  $\text{eff}$  or PolyChord chain length  $n_{\text{repeats}}$ .

## Chain-based samplers

- ▶ e.g. PolyChord, ProxNest.
- ▶ Run Markov chain starting at a live point, generating many valid (correlated) points.
- ▶ Linear  $\sim \mathcal{O}(d)$  penalty in decorrelating new live point from the original seed point.

# Applications: 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  $\Lambda$ CDM 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.  $\Lambda$ CDM vs a dynamic dark energy cosmology

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

$$\frac{\mathcal{Z}_M \Pi_M}{\sum_m \mathcal{Z}_m \Pi_m}$$

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

## Tension quantification

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

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

$$\begin{aligned} \log \mathcal{S} &= \langle \log \mathcal{L}_{AB} \rangle_{\mathcal{P}_{AB}} \\ &\quad - \langle \log \mathcal{L}_A \rangle_{\mathcal{P}_A} \\ &\quad - \langle \log \mathcal{L}_B \rangle_{\mathcal{P}_B} \end{aligned}$$

# Applications of nested sampling

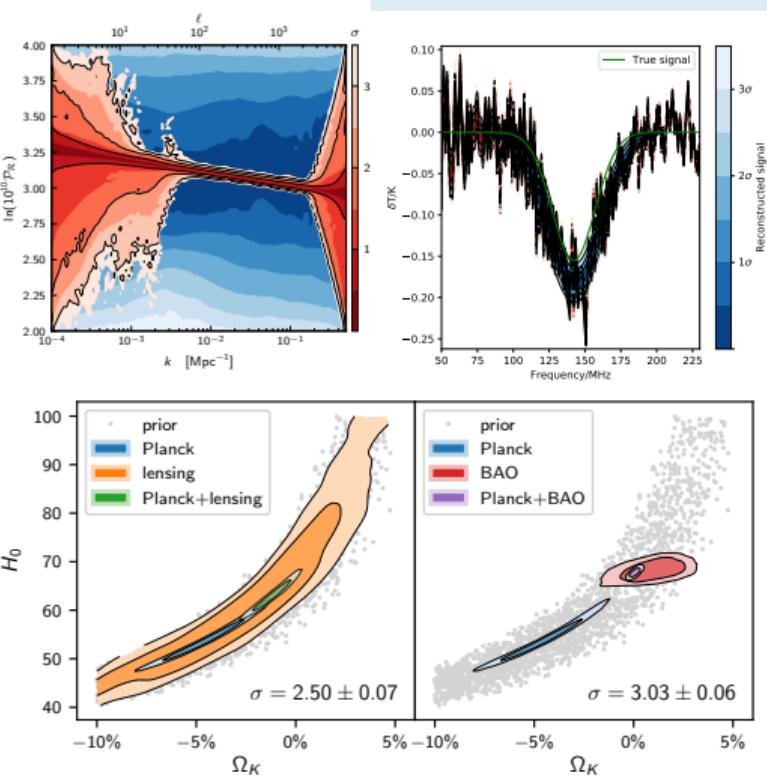
Adam Ormondroyd



PhD

## Cosmology

- ▶ 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  $\Lambda$ CDM 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.



1

# Applications of nested sampling

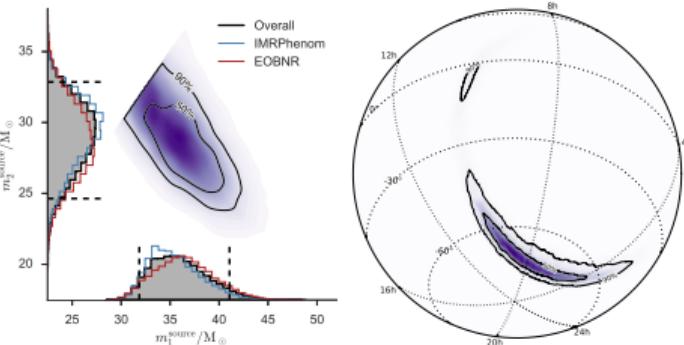
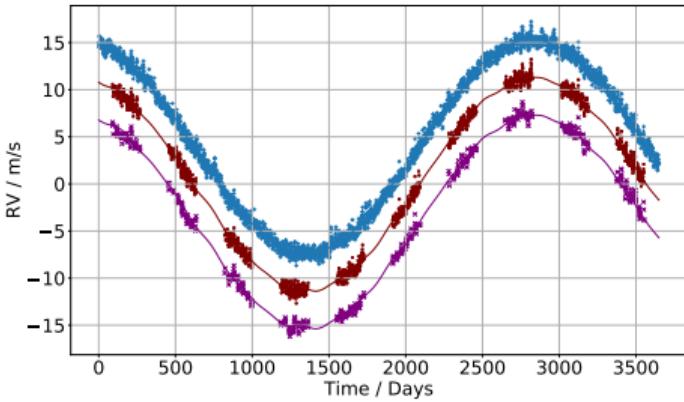
Metha Prathaban

PhD



## Astrophysics

- ▶ In exoplanets [1806.00518]
  - ▶ Parameter estimation: determining properties of planets.
  - ▶ Model comparison: how many planets? Stellar modelling [2007.07278].
  - ▶ exoplanet problems regularly have posterior phase transitions [2102.03387]
- ▶ In gravitational waves
  - ▶ Parameter estimation: Binary merger properties
  - ▶ Model comparison: Modified theories of gravity, selecting phenomenological parameterisations [1803.10210]
  - ▶ Likelihood reweighting: fast slow properties



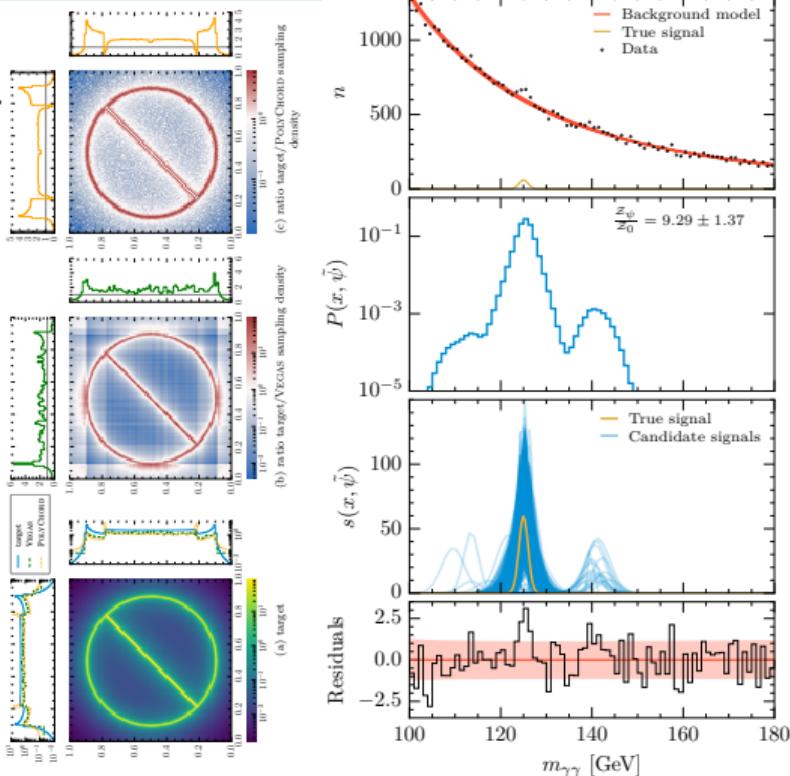
# Applications of nested sampling

## Particle physics

- ▶ Nested sampling for cross section computation/event generation  $\sigma = \int_{\Omega} d\Phi |\mathcal{M}|^2$ .
- ▶ Nested sampling can explore the phase space  $\Omega$  and compute integral blind with comparable efficiency to HAAG/RAMBO [2205.02030].
- ▶ Bayesian sparse reconstruction [1809.04598] applied to bump hunting allows evidence-based detection of signals in phenomenological backgrounds [2211.10391].
- ▶ Fine tuning quantification
- ▶ Fast estimation of small  $p$ -values [2106.02056](PRL), just make switch:  
 $X \leftrightarrow p, \mathcal{L} \leftrightarrow \lambda, \theta \leftrightarrow x.$

David Yallup

PDRA



# Applications of nested sampling

David Yallup

PDRA

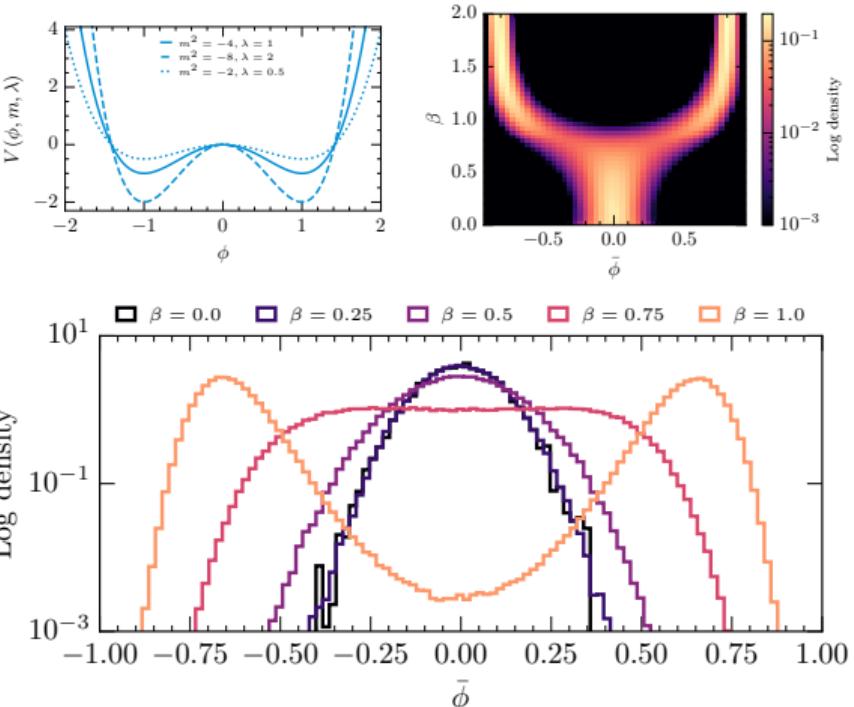


## Lattice field theory

- Consider standard field theory Lagrangian:

$$Z(\beta) = \int D\phi e^{-\beta S(\phi)}, \quad S(\phi) = \int dx^\mu \mathcal{L}(\phi)$$

- Discretize onto spacetime grid.
- Compute partition function
- NS unique traits:
  - Get full partition function for free
  - allows for critical tuning
  - avoids critical slowing down
- Applications in lattice gravity, QCD, condensed matter physics
- Publication imminent (next week)

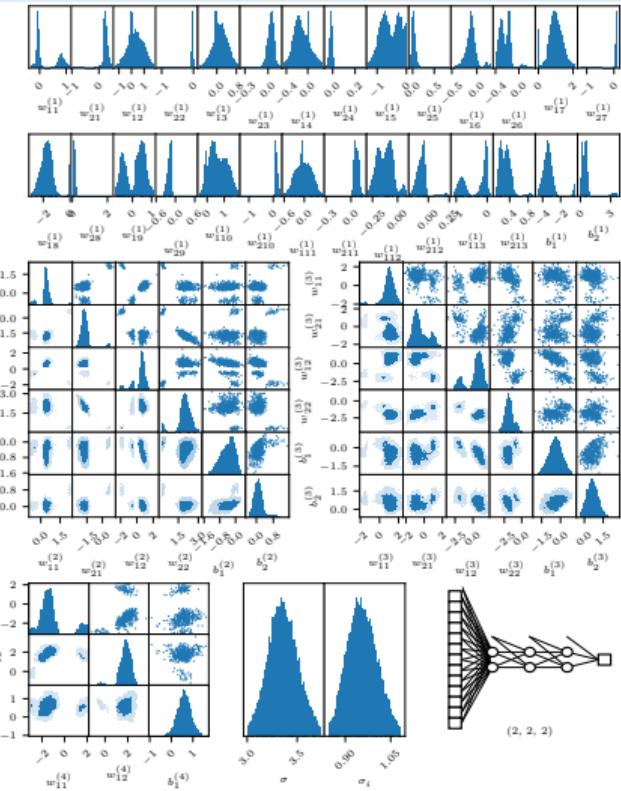




# Applications of nested sampling

## Machine learning

- ▶ Machine learning requires:
  - ▶ Training to find weights
  - ▶ Choice of architecture/topology/hyperparameters
- ▶ Bayesian NNs treat training as a model fitting problem
- ▶ Compute posterior of weights (parameter estimation), rather than optimisation (gradient descent)
- ▶ Use evidence to determine best architecture (model comparison), correlates with out-of-sample performance!
- ▶ Solving the full “shallow learning” problem without compromise [2004.12211][2211.10391].
  - ▶ Promising work ongoing to extend this to transfer learning and deep nets.
- ▶ More generally, dead points are optimally spaced for training traditional ML approaches e.g. [2309.05697]

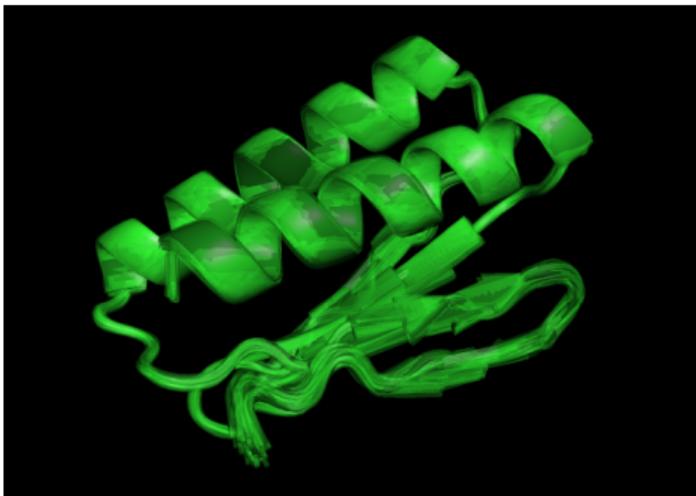




# Applications of nested sampling

And beyond...

- ▶ Techniques have been spun-out (PolyChord Ltd) to:
- ▶ Protein folding
  - ▶ Navigating free energy surface.
  - ▶ Computing misfolds.
  - ▶ Thermal motion.
- ▶ Nuclear fusion reactor optimisation
  - ▶ multi-objective.
  - ▶ uncertainty propagation.
- ▶ Telecoms & DSTL research (MIDAS)
  - ▶ Optimising placement of transmitters/sensors.
  - ▶ Maximum information data acquisition strategies.



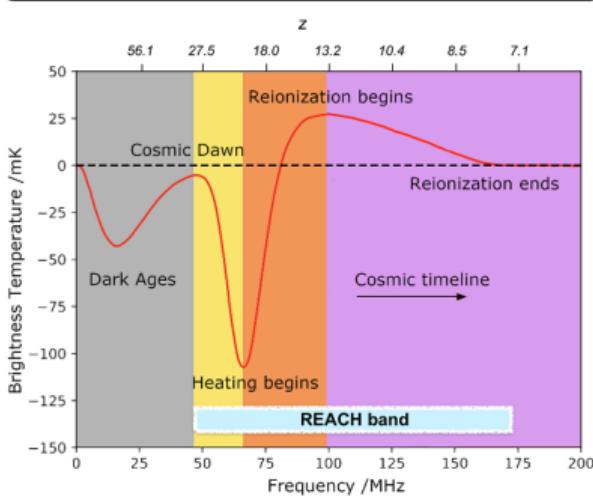
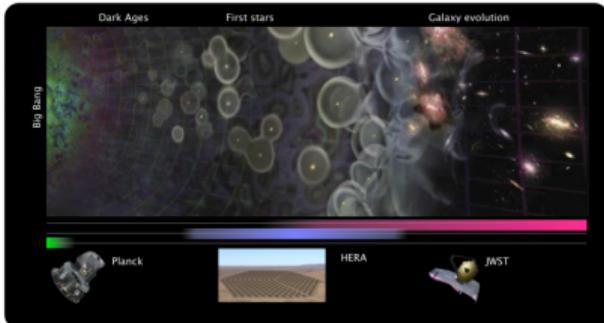
# REACH: Global 21cm cosmology [2210.07409](NatAstro)

Ian Roque

PhD

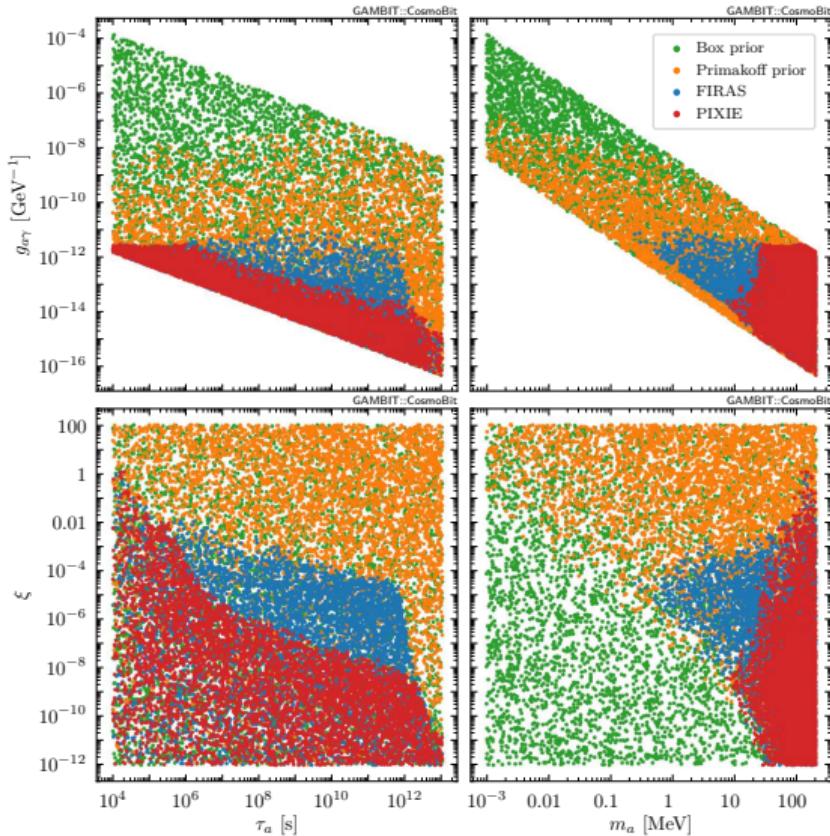


- ▶ Imaging the universal dark ages using CMB backlight.
- ▶ 21cm hyperfine line emission from neutral hydrogen.
- ▶ Global experiments measure monopole across frequency.
- ▶ Challenge: science hidden in foregrounds  $\sim 10^4 \times$  signal.
- ▶ Lead data analysis team (REACH first light in January)
- ▶ Nested sampling woven in from the ground up (calibrator, beam modelling, signal fitting, likelihood selection).
- ▶ All treated as parameterised model comparison problems.



# GAMBIT: combining particle physics & cosmological data

- ▶ Multinational team of particle physicists, cosmologists and statisticians.
- ▶ Combine cosmological data, particle colliders, direct detection, & neutrino detectors in a statistically principled manner [2205.13549].
- ▶ Lead Cosmo/Dark Matter working group [2009.03286].
- ▶ Nested sampling used for global fitting, and fine-tuning quantification [2101.00428]



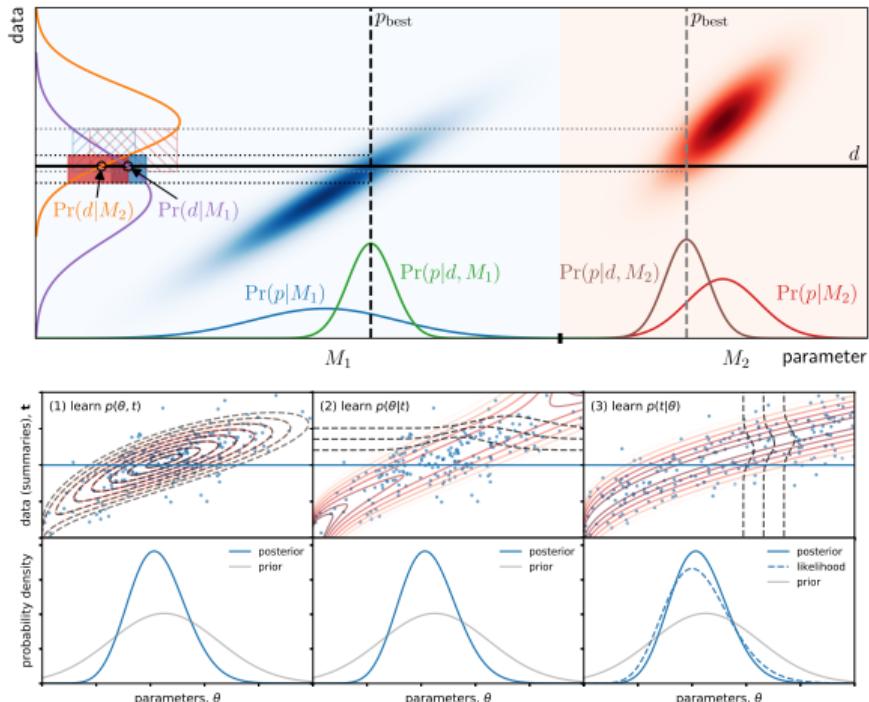
# Likelihood-free inference (aka SBI)

Kilian Scheutwinkel

PhD



- ▶ How do you do inference if you don't know the likelihood  $P(D|\theta)$ ?
  - ▶ e.g. if you can simulate a disease outbreak, how can you infer a posterior on  $R_0$ , or select the most predictive model?
- ▶ If you can forward simulate/model  $\theta \rightarrow D$ , then you have an implicit likelihood.
- ▶ LFI aims to (machine-)learn the likelihood from forward simulations  $\{(\theta, D)\}$ .
- ▶ Nested sampling has much to offer
  - ▶ truncation strategies (PolySwyft)
  - ▶ evidence driven compression
  - ▶ marginalised machine learning
- ▶ In my view, LFI represents the future of inference – in twenty years time this will be as well-used as MCMC techniques are today.



# unimpeded: PLA for the next generation

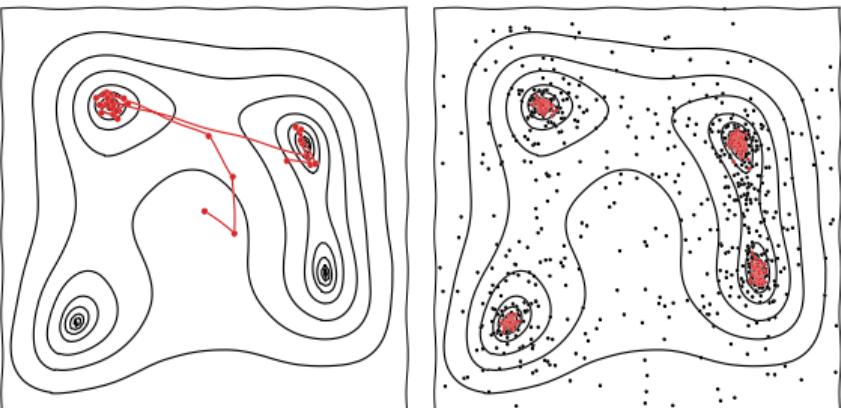
Harry Bevins



PhD→JRF

- ▶ DiRAC 2020 RAC allocation of 30MCPUh
- ▶ Main goal: Planck Legacy Archive equivalent
- ▶ Parameter estimation → Model comparison
- ▶ MCMC → Nested sampling
- ▶ Planck → {Planck, DESY1, BAO, ...}
- ▶ Pairwise combinations
- ▶ Suite of tools for processing these
  - ▶ anesthetic 2.0
  - ▶ unimpeded 1.0
  - ▶ zenodo archive
  - ▶ margarine
- ▶ MCMC chains also available.
- ▶ Library of bijectors emulators for fast re-use

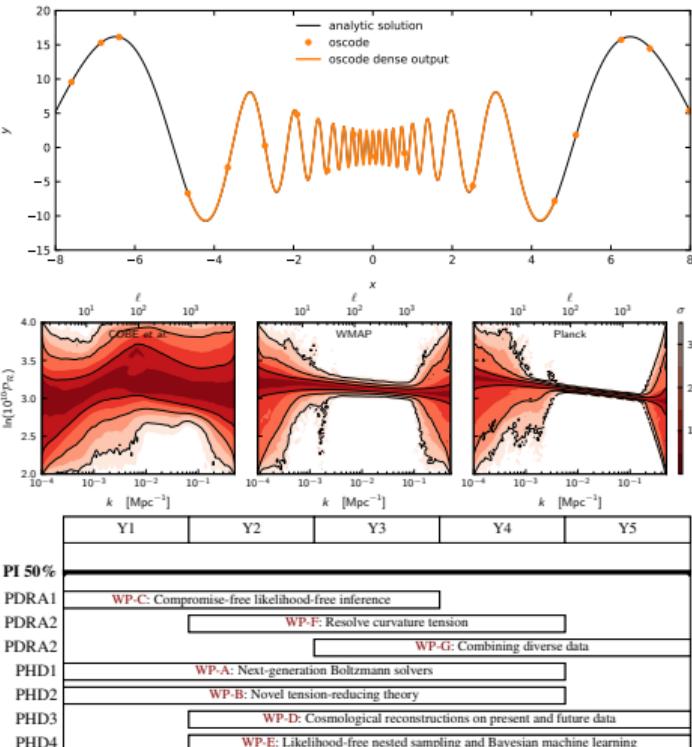
# DiRAC





Resolving cosmological tensions with diverse data, novel theories and Bayesian machine learning

- ▶ ERC grant ⇒ UKRI Frontier, commencing 2023.
- ▶ Funds 3 PDRAs and 4 PhDs over 5 years.
- ▶ Research programme centered around combining novel theories of gravity, Boltzmann solvers [1906.01421], reconstruction [1908.00906], nested sampling & likelihood free inference.
- ▶ Aims to disentangle cosmological tensions  $H_0$ ,  $\sigma_8$ ,  $\Omega_K$  with next-generation data analysis techniques.

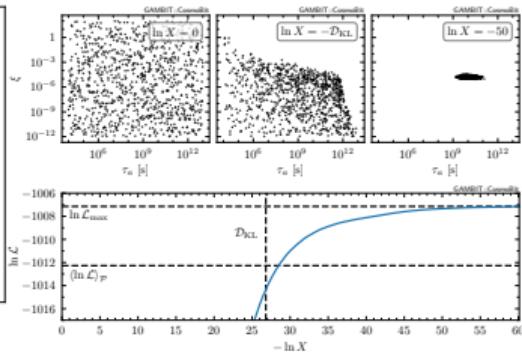
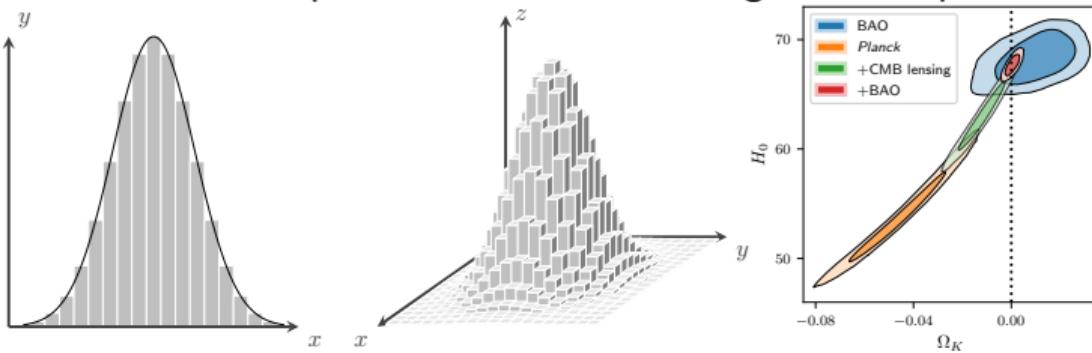
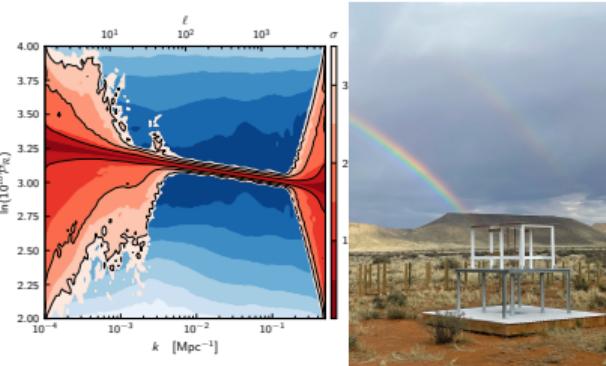


# Conclusions



[github.com/handley-lab](https://github.com/handley-lab)

- ▶ Nested sampling is a multi-purpose numerical tool for:
  - ▶ Numerical integration  $\int f(x)dV$ ,
  - ▶ Exploring/scanning/optimising *a priori* unknown functions,
  - ▶ Performing Bayesian inference and model comparison.
- ▶ It is applied widely across cosmology, particle physics & machine learning.
- ▶ It's unique traits as the only numerical Lebesgue integrator mean with compute it will continue to grow in importance.



# How does Nested Sampling compare to other approaches?

- ▶ In all cases:
    - + NS can handle multimodal functions
    - + NS computes evidences, partition functions and integrals
    - + NS is self-tuning/black-box
- Modern Nested Sampling algorithms can do this in  $\sim \mathcal{O}(100s)$  dimensions

## Optimisation

- ▶ Gradient descent
  - NS cannot use gradients
  - + NS does not require gradients
- ▶ Genetic algorithms
  - + NS discarded points have statistical meaning

## Sampling

- ▶ Metropolis-Hastings?
  - Nothing beats well-tuned customised MH
  - + NS is self tuning
- ▶ Hamiltonian Monte Carlo?
  - In millions of dimensions, HMC is king
  - + NS does not require gradients

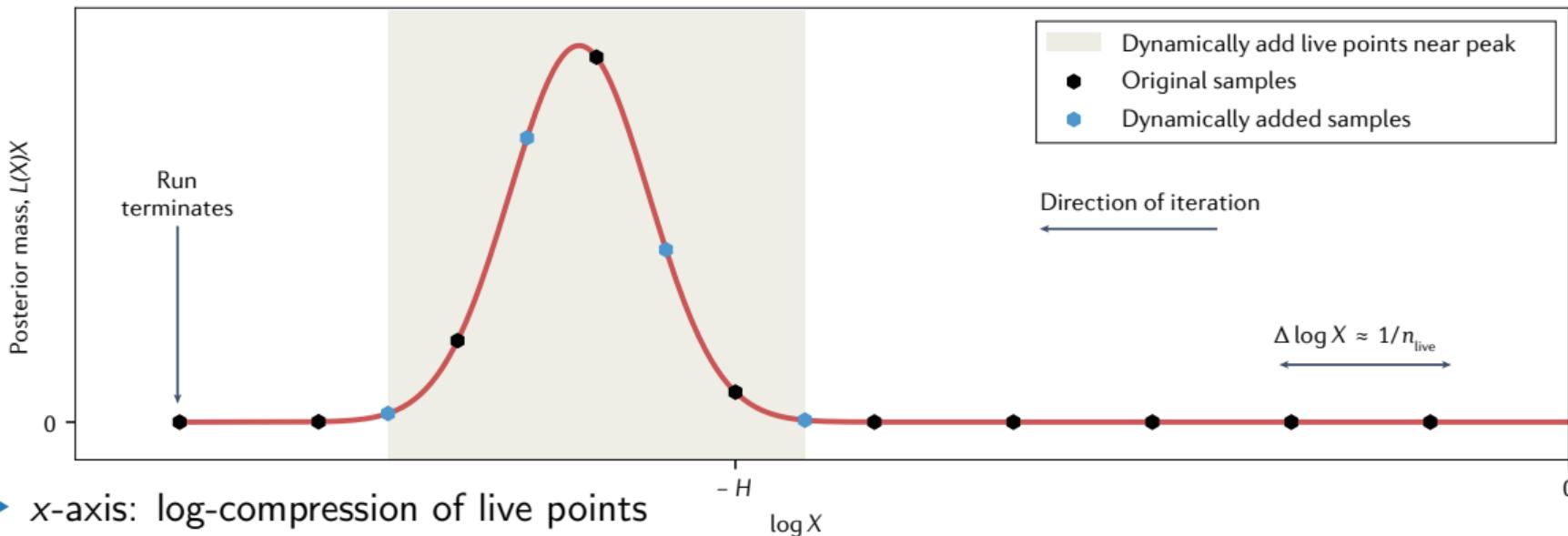
## Integration

- ▶ Thermodynamic integration
  - protective against phase transitions
  - + No annealing schedule tuning
- ▶ Sequential Monte Carlo
  - SMC experts classify NS as a kind of SMC
  - + NS is athermal

# Nested Sampling: a user's guide

1. Nested sampling is a likelihood scanner, rather than posterior explorer.
  - ▶ This means typically most of its time is spent on burn-in rather than posterior sampling.
  - ▶ Changing the stopping criterion from  $10^{-3}$  to 0.5 does little to speed up the run, but can make results very unreliable.
2. The number of live points  $n_{\text{live}}$  is a resolution parameter.
  - ▶ Run time is linear in  $n_{\text{live}}$ , posterior and evidence accuracy goes as  $\frac{1}{\sqrt{n_{\text{live}}}}$ .
  - ▶ Set low for exploratory runs  $\sim \mathcal{O}(10)$  and increased to  $\sim \mathcal{O}(1000)$  for production standard.
3. Most algorithms come with additional reliability parameter(s).
  - ▶ e.g. MultiNest:  $\text{eff}$ , PolyChord:  $n_{\text{repeats}}$ .
  - ▶ These are parameters which have no gain if set too conservatively, but increase the reliability.
  - ▶ Check that results do not degrade if you reduce them from defaults, otherwise increase.

# Time complexity of nested sampling



- ▶ x-axis: log-compression of live points
- ▶ Area  $\propto$  posterior mass
- ▶ Shows Bayesian balance of likelihood vs prior
- ▶ Run proceeds right to left
- ▶ Run finishes after bump (typical set)

## Time complexity

$$T = n_{\text{live}} \times T_{\mathcal{L}} \times T_{\text{sampler}} \times D_{\text{KL}}(\mathcal{P} \parallel \pi)$$

## Error complexity

$$\sigma \propto \sqrt{D_{\text{KL}}(\mathcal{P} \parallel \pi) / n_{\text{live}}}$$

# 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

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

- ▶ Evidence is composed of a “goodness of fit” term and “Occam Penalty”.
- ▶ RHS true for all  $\theta$ . Take max likelihood value  $\theta_*$ :
- ▶ Be more Bayesian and take posterior average to get the “Occam's razor equation”

$$\log \mathcal{Z} = -\chi_{\min}^2 - \text{Mackay penalty.}$$

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

- ▶ Natural regularisation which penalises models with too many parameters.

# Kullback Liebler divergence

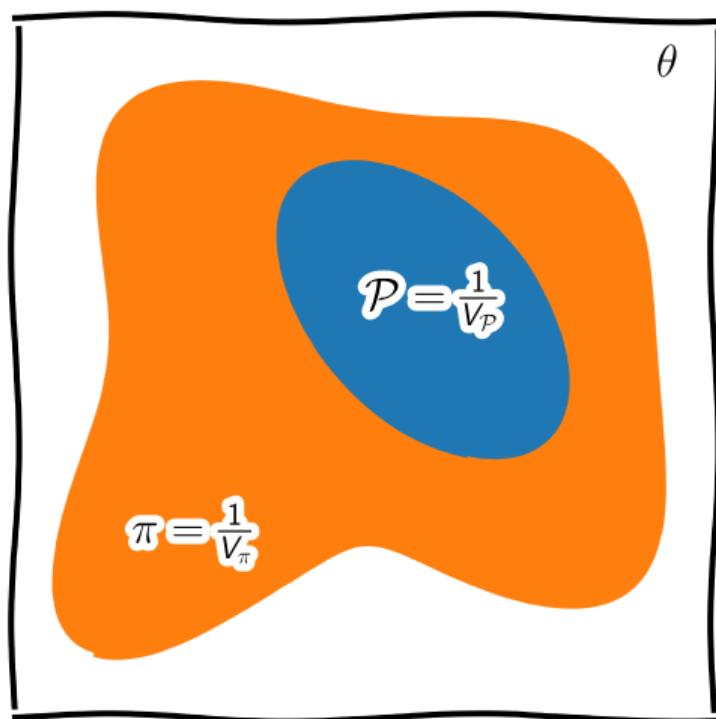
- The KL divergence between prior  $\pi$  and posterior  $\mathcal{P}$  is defined as:

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

- 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”:

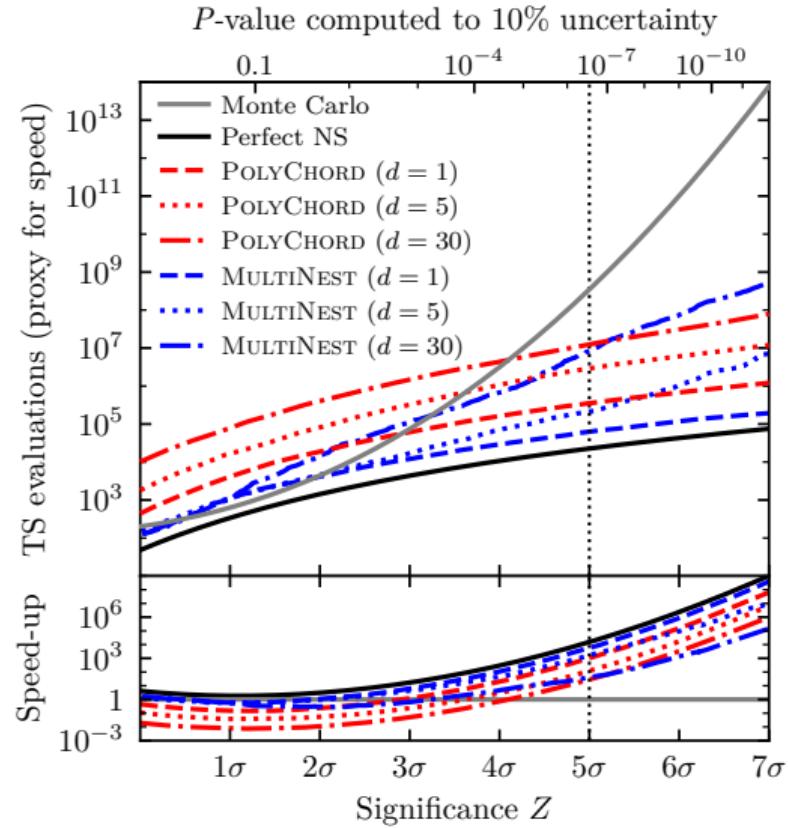
$$\mathcal{D}_{\text{KL}} \approx \log \frac{V_\pi}{V_{\mathcal{P}}}.$$

(this is exact for top-hat distributions).



# Statistics: fast estimation of small $p$ -values [2106.02056](PRL)

- ▶ Nested sampling for frequentist computation!?
- ▶  $p$ -value:  $P(\lambda > \lambda^* | H_0)$  – probability that test statistic  $\lambda$  is at least as great as observed  $\lambda^*$ .
- ▶ Computation of a tail probability from sampling distribution of  $\lambda$  under  $H_0$ .
- ▶ For gold-standard  $5\sigma$ , this is very expensive to simulate directly ( $\sim 10^9$  by definition).
- ▶ Need insight/approximation to make efficient.
- ▶ Nested sampling is tailor-made for this, just make switch:  $X \leftrightarrow p$ ,  $\mathcal{L} \leftrightarrow \lambda$ ,  $\theta \leftrightarrow x$ .
- ▶ The only real conceptual shift is switching the integrator from parameter- to data-space.



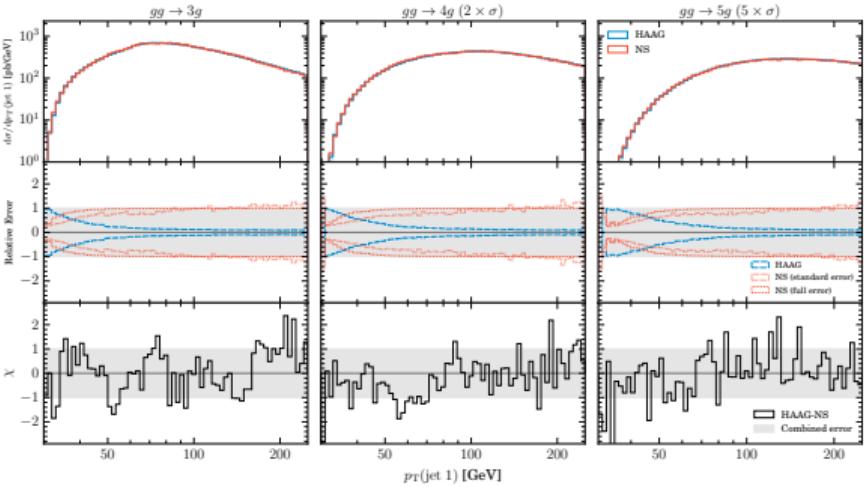
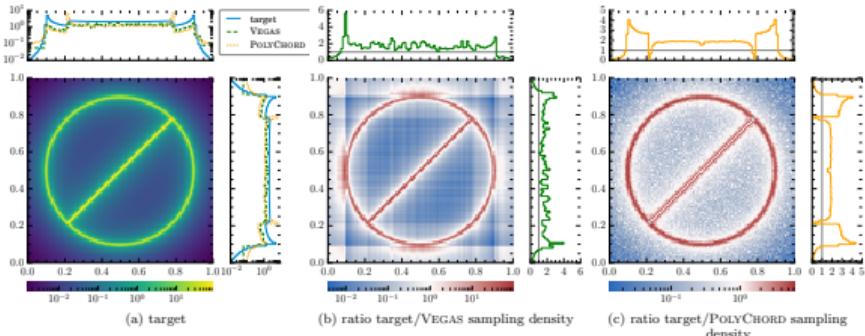
# Exploration of phase space [2106.02056]

- ▶ Nested sampling for cross section computation/event generation.
- ▶ Numerically compute collisional cross section

$$\sigma = \int_{\Omega} d\Phi |\mathcal{M}|^2,$$

$\Omega$  phase space of kinematic configurations  $\Phi$ , each with matrix element  $\mathcal{M}(\Phi)$ .

- ▶ Current state of the art e.g. HAAG (improvement on RAMBO) requires knowledge of  $\mathcal{M}(\Phi)$ .
- ▶ Nested sampling can explore the phase space and compute integral blind with comparable efficiency.



# Quantification of fine tuning [2101.00428] [2205.13549]

- ▶ Example: Cosmological constraints on decaying axion-like particles [2205.13549].
- ▶ Subset of parameters  $\xi, m_a, \tau, g_{a\gamma}$ : ALP fraction, mass, lifetime and photon coupling. (Also vary cosmology,  $\tau_n$  and nuisance params)
- ▶ Data: CMB, BBN, FIRAS, SMM, BAO.
- ▶ Standard profile likelihood fit shows ruled out regions and best-fit point.
- ▶ Nested sampling scan:
  - ▶ Quantifies amount of parameter space ruled out with Kullback-Liebler divergence  $\mathcal{D}_{KL}$ .
  - ▶ Identifies best fit region as statistically irrelevant from information theory/Bayesian.
  - ▶ No evidence for decaying ALPs. Fit the data equally well: but more constrained parameters create Occam penalty.

