

Nested Sampling for physical scientists

Greg Ashton ^{1,2} Noam Bernstein ³ Johannes Buchner ⁴ Xi Chen ⁵ Gábor Csányi ⁶ Farhan Feroz ⁷ Andrew Fowlie ⁸ Matthew Griffiths ⁹ Michael Habeck ¹⁰ Will Handley ^{11,12} Edward Higson ¹³ Michael Hobson ¹² Anthony Lasenby ^{11,12} David Parkinson ¹⁴ Livia B. Pártay ¹⁵ Matthew Pitkin ¹⁶ Doris Schneider ¹⁷ Leah South ¹⁸ Joshua S. Speagle  (沈佳士)^{19,20,21} John Veitch ²² Philipp Wacker ¹⁷ David J Wales ²³ David Yallup ^{11,12}

We review Skilling’s nested sampling (NS) algorithm for Bayesian inference and more broadly multi-dimensional integration. After recapitulating the principles of NS, we survey developments in implementing efficient NS algorithms in practice in high-dimensions, including methods for sampling from the so-called **constrained prior**. We outline the ways in which NS may be applied and describe the application of NS in three scientific fields in which the algorithm has proved to be useful: cosmology, gravitational-wave astronomy, and materials science. We close by making recommendations for best practice when using NS and by summarizing potential limitations and optimizations of NS.

Author addresses

¹School of Physics and Astronomy, Monash University, VIC 3800, Australia

²Department of Physics, Royal Holloway, University of London, TW20 0EX, United Kingdom

³Center for Materials Physics and Technology, U. S. Naval Research Laboratory, Washington, DC 20375, USA

⁴Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85741 Garching, Germany

⁵Department of Computer Science, University of Bath, Bath, BA2 7PB, United Kingdom

⁶Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

⁷Independent researcher

⁸Department of Physics and Institute of Theoretical Physics, Nanjing Normal University, Nanjing, Jiangsu 210023, China. andrew.j.fowlie@NJNU.edu.cn

⁹Concr Ltd, Babraham Hall House, Cambridge, CB22 3AT

¹⁰Microscopic Image Analysis Group, Jena University Hospital, Jena, Germany

¹¹Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge, CB3 0HE, United Kingdom

¹²Kavli Institute for Cosmology, Cambridge, CB3 0HA, United Kingdom

¹³The D. E. Shaw Group, 1166 Avenue of the Americas, New York, NY 10036, USA

¹⁴Korea Astronomy and Space Science Institute, Yuseong-gu, Daedeok-daero 776, Daejeon 34055, Korea

¹⁵Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

¹⁶Department of Physics, Lancaster University, Lancaster, LA1 4YB, United Kingdom

¹⁷Department of Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

¹⁸School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia

¹⁹Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

²⁰David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto ON M5S 3H4, Canada

²¹Department of Statistical Sciences, University of Toronto, 100 St. George St, Toronto, ON M5S 3G3, Canada

²²School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, United Kingdom

²³Yusuf Hamied Department of Chemistry, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

of live points through the parameter space. The algorithm was immediately adopted in cosmology, owing to the fact that it partially overcomes three major difficulties in the traditional algorithm for Bayesian computation, **Markov chain Monte Carlo (MCMC; see for example ref. [3, 4])**. First, it simultaneously returns results for model comparison and parameter inference. Second, it is successful in **multi-modal** problems. Third, it is naturally self-tuning, permitting it to be applied immediately to new problems. In the 15 years since, the theoretical properties of the algorithm and connections to other computational methods have been partially clarified, and efficient implementations, variants and cross-checks of NS have been developed. The range of applications now extends beyond cosmology and into many other branches of science.

Here the review of those developments and applications is structured as follows. Below in **Introduction**, we recapitulate the origins and principles of NS. We summarize implementations and variants of the NS algorithm, including the developments since its inception, in **Experimentation** and results from NS in **Results**. We describe scientific applications from cosmology, gravitational-wave astronomy, particle physics and materials science in **Applications**. We outline best practices when using NS, including **Reproducibility and data deposition** and discuss issues with the technique in **Limitations and optimizations**. We close by looking forward to the future of NS and Bayesian computation in **Outlook**. Further details are presented in **Supplementary Information**, including a glossary in **Supplementary Information A** and a simple numerical example in **Supplementary Information G**.

1.1 Multi-dimensional integrals

Since NS is primarily an algorithm for integration, let us write a general multi-dimensional integral of a function L over parameters Θ as

$$Z = \int L(\Theta) d\mu(\Theta). \quad (1)$$

In many scientific problems we need to be able to integrate in high dimensions and for challenging **integrands**. We assume that the **integrand** is positive, $L(\Theta) \geq 0$.

1 Introduction

The nested sampling algorithm (NS; [1, 2]) was introduced by Skilling in 2004 in the context of Bayesian inference and computation (described in box 1). The NS algorithm solves otherwise challenging high-dimensional integrals by evolving a collection

Box 1 | Bayesian inference

Although NS is a general purpose algorithm for integration, its major application has been integrals in Bayesian inference, and we describe NS using that language. In Bayesian inference [5–11] our state of knowledge is quantified by probability and we learn from data by updating probabilities using Bayes' theorem,

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$

To use this to learn from data about a model and its parameters, we write it as

$$P(\Theta) = \frac{L(\Theta)\pi(\Theta)}{Z},$$

where the **prior**, $\pi(\Theta) \equiv \Pr(\Theta)$ represents what was known about a model's parameters before seeing the data and the **posterior**, $P(\Theta) \equiv \Pr(\Theta|D)$, represents what is known after learning from the data. The observed data, D , was encoded into the **likelihood** function, $L(\Theta) \equiv \Pr(D|\Theta)$.

The denominator, $Z \equiv \Pr(D)$, is the **evidence** value that appears in **Bayesian model comparison** [12]. It may be written,

$$Z = \int L(\Theta)\pi(\Theta) d\Theta,$$

and so is also known as the marginal **likelihood** and as the normalizing constant, since it normalizes the **posterior** such that $\int P(\Theta) d\Theta = 1$. The ratio of evidences computed for different models is known as a Bayes factor,

$$B_{10} = \frac{Z_1}{Z_0}.$$

The Bayes factor tells us how we must update the relative plausibility of two models in light of data.

Often, Z may be a physical quantity such as the total mass of an object distributed with density L across volumes $d\mu(\Theta)$. Whilst NS is a general method for integration, for concreteness, we view all such applications through the lens of Bayesian inference (see box 1), with $d\mu(\Theta) \equiv \pi(\Theta)d\Theta$ seen as an element of *prior probability* with π the **prior**, normalized by its nature to $\int \pi(\Theta)d\Theta = 1$. The **integrand**, L , is the modulating **likelihood** function (hence the symbol) and Z is the **evidence**. In scientific inference problems, the integral could be over tens if not hundreds of parameters, required to model fundamental effects as well as the calibration and systematics of complicated experimental measurements [13].

We may rewrite the **integrand** through the elementary factorization known as Bayes' theorem,

$$L(\Theta) \times \pi(\Theta) = Z \times P(\Theta), \quad (2)$$

where

$$P(\Theta) = \frac{L(\Theta)\pi(\Theta)}{Z}, \quad (3)$$

is the **posterior**, normalized to $\int P(\Theta) d\Theta = 1$. Notation apart, however, all we are doing here is decomposing the **integrand** into a magnitude Z and a shape $P(\Theta)$.

Historically, Bayesian computation focused on only the shape $P(\Theta)$, partly owing to controversies around **Bayesian model comparison** such as its sensitivity to the choices of **prior** [12], and partly due to computational difficulties [14]. However, shapes and magnitudes both matter, especially in the general setting of multi-dimensional integration beyond **Bayesian model comparison**. NS [1, 2] surmounts the challenge by computing shapes and magnitudes simultaneously.

1.2 Simplifying multi-dimensional integrals

Before introducing NS, let us attempt to simplify the general integral in eq. (1). Consider traditional Riemann-style integration. This decomposes the space into volume elements $\Delta\Theta$, typically small cubes, and performs a sum over them. Small cubes, however, rapidly become infeasible in multi-dimensional integration because their cost grows exponentially with dimension — this is the “**curse of dimensionality**.”

We don't, however, need to decompose our space into little cubes; our cells can be any shape we want. The integrals needed for quantification,

$$Z = \int L(\Theta)\pi(\Theta) d\Theta = \lim_{|\Delta\Theta| \rightarrow 0} \sum L(\Theta)\pi(\Theta) \Delta\Theta, \quad (4)$$

are defined as limiting sums over volume elements which should be small enough to keep $P(\Theta)$ almost constant *regardless of shape*. We may therefore combine the cells in which the **integrand** is almost constant. Schematically, we may write

$$Z = \sum L(X)\Delta X, \quad (5)$$

where ΔX is the volume of cells that share **likelihood** $L(X)$ weighted by the **prior** $\pi(\Theta)$. This is illustrated schematically in fig. 1a and works whether the **integrand** is **uni-** or **multi-modal**.

We can reach eq. (5) more concretely by noting that the **evidence** is the expectation of a non-negative random variable, such that it may be written as

$$Z = \int X(L) dL, \quad (6)$$

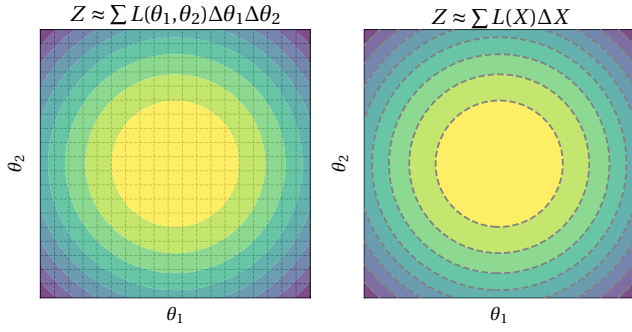
where the volume variable X ,

$$X(L^*) = \int_{L>L^*} \pi(\Theta) d\Theta, \quad (7)$$

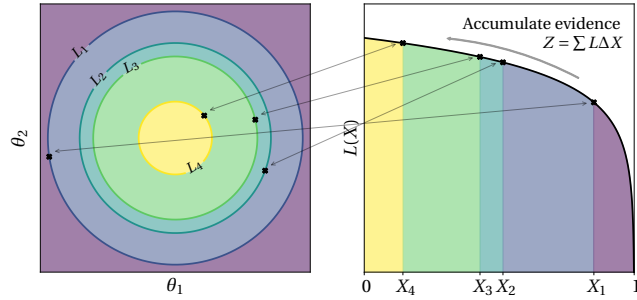
is the volume enclosed by contour L^* . This result can be readily proven by integration by parts (see also sec. 21 in ref. [17]). Applying integration by parts again to eq. (6), we obtain the familiar NS **evidence** identity,

$$Z = \int_0^1 L(X) dX, \quad (8)$$

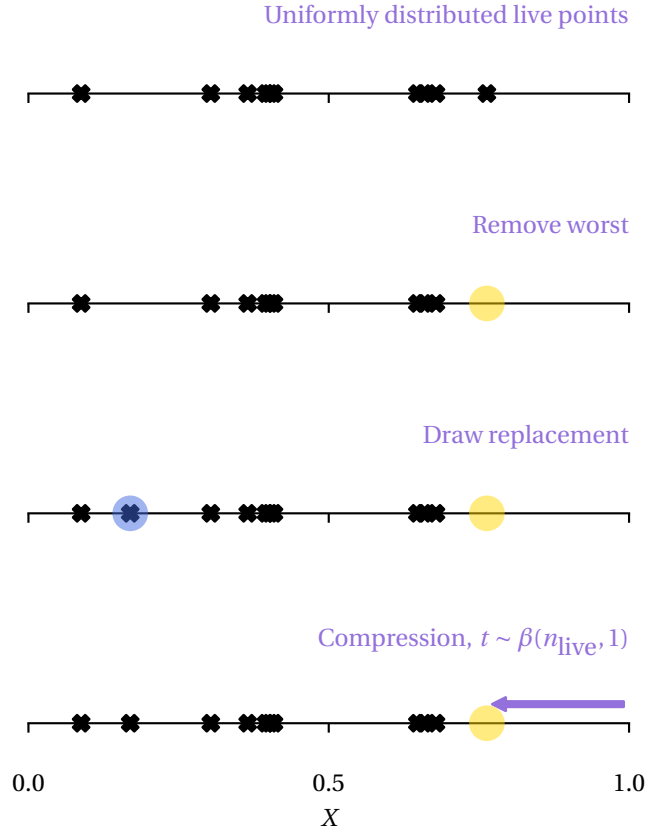
providing that $L(X)$, the inverse of $X(\lambda)$, indeed exists and that the **evidence** is finite. This formalizes the schematic eq. (5). We discuss this result more formally in box 2.



a | **The NS evidence identity.** The colours represent contours of a two-dimensional likelihood function. Rather than summing over little cubes (left), we combine cubes of similar likelihood together and sum over them (right).



b | **NS on a two dimensional problem.** We show the dead points and their iso-likelihood contours (left) and the corresponding contributions to the evidence integral (right). The volumes X_i are estimated statistically in NS.



c | **Compression in one iterate of NS.**

Figure 1 | **Illustrations of NS algorithm.**

1.3 Nested sampling

As the multi-dimensional integral in eq. (7) is impractical in high dimension, some sort of statistical estimation is inevitable. NS starts with an ensemble of n_{live} random locations Θ drawn from the prior, $\pi(\Theta)$, each of which has its likelihood $L(\Theta)$ which we can place in ascending order. Crudely, if we discarded the lowest half of the values, the survivors would be random samples taken within the restricted volume $L > \text{Median}[L]$, which would statistically be roughly half the original volume. This allows us to make a statistical estimate of the volume variable in eq. (7). Repeating that n_{iter} times would yield compression by a factor of about $2^{n_{\text{iter}}}$. This is the exponential behaviour required to overcome the curse of dimensionality.

Thus NS works by statistical estimates of the compression, which is a general and fundamental operation that can be used in various ways not limited to those in table 1. The evidence identity in eq. (6) isn't required in every application; many applications only use the compression in eq. (7). We present scientific applications in Applications.

1.4 Formulation

We now present the NS algorithm in more detail. We assume that there are no regions of constant likelihood resulting in like-

likelihood plateaus (see Limitations and optimizations for further discussion). The NS algorithm begins by drawing an ensemble of n_{live} samples from the prior. We compute the likelihood for each sample. We denote the smallest likelihood by L^* and we discard that point. The remaining live points are now distributed over a compressed volume; we denote the factor by which the volume compressed by t . Finally, a replacement point is drawn from the prior subject to $L > L^*$, that is, from the constrained prior,

$$\pi^*(\Theta) \propto \begin{cases} \pi(\Theta) & \text{if } L(\Theta) > L^* \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

This leaves a new ensemble with n_{live} samples obeying a likelihood constraint $L > L^*$.

As they are drawn from the constrained prior, the volumes X associated with the live points are uniformly distributed. Thus the compression associated with the discarded outermost sample, t , corresponds to the smallest of n_{live} uniform random variables. This follows a Beta($n_{\text{live}}, 1$) distribution,

$$P(t) = n_{\text{live}} t^{n_{\text{live}}-1}. \quad (10)$$

The first factor accounts for the fact that any live point could be the outermost and the second factor for the fact that $n_{\text{live}} - 1$

NS application	Details
Integration	Perform the general multi-dimensional integral eq. (1) for positive integrands.
Global optimization	Maximize the likelihood , L , by compressing to $\hat{\Theta}$, the maximum of $L(\Theta)$, with no restriction to uni-modal distributions. This may require strict settings and be more computationally expensive than integration; see ref. [148, 239] for further discussion.
Bayesian inference	NS simultaneously computes the posterior and the Bayesian evidence , allowing parameter inference and model comparison.
Approximate Bayesian Computation	Perform efficient Approximate Bayesian Computation by applying NS to the joint space of parameters and data [241]
Statistical thermodynamics	If we use the Boltzmann factor as the likelihood , i.e., $L(\Theta) = e^{-\beta E(\Theta)}$, where E is the energy of the state Θ among N , we may use NS to compute the partition function . By accumulating for several temperatures T in parallel, one can plot thermodynamic functions such as the specific heat $C_V = T dS/dT$ as functions of temperature without needing multiple runs.
Rare event sampling	The volume variable X may be interpreted as the probability of a rare event [24, 25] or used to compute a p -value in frequentist statistics [161].

Table 1 | **Applications of NS.**

uniformly distributed samples lie above the outermost sample at t (the remaining sample lies at t).

We started from unconstrained samples drawn from the full original volume $X_0 = 1$. As we repeat this process of replacing the outermost points, the successive compressions by factors t_1, t_2, t_3, \dots lead to exponentially decreasing inferred volumes X_1, X_2, X_3, \dots

$$X_0 = 1. \quad X_{i+1} = t_{i+1} X_i. \quad (11)$$

We illustrate a single iteration of NS in fig. 1c. If we are most interested in the magnitude of the **evidence**, $\log Z$, we should consider,

$$\log t \approx \langle \log t \rangle = -\frac{1}{n_{\text{live}}}, \quad (12)$$

as under repeated multiplication $X_k = t_1 t_2 \dots t_k$ it's the *logarithms* that add. For a more complete inference, the compression factors t may be sampled directly from Beta($n_{\text{live}}, 1$). See Supplementary Information B for further discussion. NS therefore estimates volumes through probability not geometry, topology, or even dimension. We do not get a definite compression value, only a distribution of what it might have been.

Having obtained estimates of the volume $X(L^*)$ at each of n_{iter} iterations, we may accumulate the **evidence** via eq. (6) or eq. (8), for example by the trapezium rule,

$$Z = \sum_{i=1}^{n_{\text{iter}}} w_i L_i^*, \quad (13)$$

where the weights are

$$w_i = \frac{1}{2} (X_{i-1} - X_{i+1}). \quad (14)$$

The sum in eq. (13) converges to the desired integral [18–20]. This is the magnitude; we obtain shape by assigning weights to

each sample (see ref. [18] for discussion),

$$P_i = \frac{w_i L_i^*}{Z}, \quad (15)$$

normalized such that $\sum P_i = 1$. These are the **posterior** weights of the dead points; the shape may be recovered by for example a weighted histogram or other density estimation methods. We show the whole algorithm schematically in algorithm 1 and the summation for a two-dimensional problem in fig. 1b.

Compared to the sketch of NS in section 1.3, we replace a single live point per iteration, rather than half of the live points. Whilst the number of replacements can be varied, one replacement is optimal (though see considerations in **Parallelization**). This is because for $r \ll n_{\text{live}}$, replacing r points per iteration would reach the **posterior bulk** in about r times fewer iterations. While the computational expense wouldn't change as r replacements are required per iteration, the error estimates would scale as \sqrt{r} because reducing the number of iterations increases the relative Poisson noise in the number of iterations.

In the last decade or so, analogies between NS and statistical mechanics (see box 3) and other statistical methods, including sequential Monte Carlo (SMC; [21]) and subset simulation in rare event sampling [22–25], among others [26–30], have been recognized. The connections to SMC and an SMC variant of NS are discussed in box 4. There are, of course, other strategies for computing the evidence; see ref. [31–34] for reviews. Notable examples include approximating the integrand by a tractable function [35], Chib's method using density estimation and Gibbs' sampling [36], importance sampling [37], and techniques that re-use MCMC draws [38]. Broadly speaking, NS lies in a class of algorithms that form a path of bridging distributions, and evolves samples along that path [39, 40]. NS stands out because the path is automatic and smooth — we compress in $\log X$ by on average $1/n_{\text{live}}$ at each iteration — and

Algorithm 1 | **Schematic of the NS algorithm for the general multi-dimensional integral in eq. (1).** Techniques for drawing replacements and stopping criteria are discussed in [Exploration strategies](#) and [Stopping conditions](#) respectively.

- 1 Choose an estimate of the compression factor, e.g., $t = e^{-1/n_{\text{live}}}$
 - 2 Initialize volume, $X = 1$ and integral, $Z = 0$
 - 3 Sample n_{live} points from the [prior](#) — the live points
 - 4 **repeat**
 - 5 Let L^* be the minimum L of the live points
 - 6 Replace live point corresponding to L^* by one drawn from the [prior](#) subject to $L > L^*$
 - 7 Increment the estimate of the integral, $Z = Z + L^* \Delta X$, with e.g., $\Delta X = (1 - t)X$
 - 8 Contract volume, $X = tX$
 - 9 **until** *stopping criteria satisfied*
 - 10 Add estimate of remaining [evidence](#), e.g., $Z = Z + \bar{L}X$ where \bar{L} is the average [likelihood](#) among the live points
 - 11 **return** *Estimate of integral, Z*
-

because along the path we compress through constrained [priors](#), rather than from the [prior](#) to the [posterior](#). This was in fact a motivation for NS as it avoids phase transitions — abrupt changes in the bridging distributions — that cause problems for other methods including path samplers such as annealing. We further discuss NS’s historical background and contrast it with annealing in Supplementary Information C.

1.5 Uncertainties

Our estimates of the magnitude and shape in eq. (1) must be accompanied by a discussion of the bias and statistical uncertainty. The latter originates from our noisy estimates of the compression factors. We may estimate the resulting statistical uncertainty in the [evidence](#) by considering the compression required to reach the [bulk](#) of the [posterior](#). This may be quantified by the information content [41, 42]

$$H = \int P(\Theta) \log \left(\frac{P(\Theta)}{\pi(\Theta)} \right) d\Theta \quad (16)$$

known in statistics as the [Kullback-Leibler \(KL\) divergence](#). We can write it using the volume variable as

$$H = \int P(X) \log P(X) dX \quad (17)$$

$$= - \int P(X) \log X dX + \int P(\log X) \log P(\log X) d \log X \quad (18)$$

where $P(X) \equiv L(X)/Z$ is the posterior density of the volume. In eq. (18) we see that the KL divergence equals minus the posterior expectation of $\log X$ minus the differential entropy associated with the posterior of $\log X$. As the first term typically dominates, the KL divergence provides a measure of compression.

Thus from eq. (12) it’s likely to take about

$$n_H \simeq n_{\text{live}} H \quad (19)$$

iterations to compress to the [bulk](#) of the [posterior](#) at $\log X = -H$, and this count is likely to be subject to $\sqrt{n_H}$ Poisson variability. Neglecting contributions from outside the [bulk](#), we may write

the evidence sum in eq. (13) as

$$Z \simeq e^{-n_H/n_{\text{live}}} \sum_{i=n_H}^{n_{\text{iter}}} (1-t)^{i-n_H} L_i^*. \quad (20)$$

We see that the final estimate of $\log Z$ will be plausibly subject to an approximately Gaussian uncertainty from the first factor,

$$\Delta \log Z \approx \sqrt{\frac{H}{n_{\text{live}}}}. \quad (21)$$

Thus NS statistical uncertainties scale as $1/\sqrt{n_{\text{live}}}$ as usual for statistical uncertainties (see ref. [18] for an alternative proof and discussion).

Although NS was first introduced with this estimate [1], it can be unreliable. Authority rests with repeated simulation through eq. (10) of what the compressions might actually have been. See ref. [43] for further discussion of the statistical uncertainty in the NS estimates and Supplementary Information E for discussion of uncertainties in NS estimates of the posterior. As well as this statistical uncertainty, there are four potential sources of bias: bias originating from failure to faithfully sample from the [constrained prior](#), bias originating from the choice of estimator for the compression factor, the generally negligible quadrature error, and the potentially important truncation error in eq. (13). The latter occurs as we stop after a finite number of iterations and is further discussed in [Stopping conditions](#). Provided that NS is appropriately configured, the statistical uncertainty usually dominates.

We see that difficulty in NS does not in fact lie in dimension but in compression from the [prior](#) to the [posterior](#): the compression and the resolution n_{live} alone determine the uncertainty and the run-time. To maintain a given uncertainty, by eq. (21) we require $n_{\text{live}} \propto H$ live points and by eq. (19) $n_{\text{iter}} \propto H^2$ iterations. If the [prior](#) and [posterior](#) are factorizable into a term for each dimension, by eq. (16) the [KL divergence](#) is additive, and so scales linearly with dimension, D , and so run-time goes like $\mathcal{O}(D^2)$. NS beats the exponential scaling with dimension expected from the [curse of dimensionality](#).

Box 2 | Mathematical details

The idea of NS is to transform eq. (1) into eq. (8) which can be approximated more efficiently using the described Monte Carlo approach with active samples:

$$Z = \int_{\Omega} L(\Theta) d\mu(\Theta) = \int_0^1 \tilde{L}(X) dX,$$

where Ω is the **parameter domain** and \tilde{L} is an **overloaded** form of L , as described next. To achieve this transformation, we define the **survival function** $X: \mathbb{R} \rightarrow [0, 1]$, $X(\lambda) = \mu(\{z \in \Omega: L(z) > \lambda\})$, that is the μ -measure of the λ -super-level-sets. Then, we introduce a mapping $\Phi: \Omega \rightarrow [0, 1]$ with

$$\Phi(\Theta) = X(L(\Theta)) = \mu(\{z \in \Omega: L(z) > L(\Theta)\})$$

This mapping Φ is the transformation which allows us to shift the integration from Ω to $[0, 1]$, by virtue of the **push-forward measure** $\mu \circ \Phi^{-1}$ on $[0, 1]$. Now, we can define

$$\tilde{L}: [0, 1] \rightarrow \mathbb{R}, \quad \tilde{L}(\xi) = \sup\{\lambda \in \text{Im } L: X(\lambda) > \xi\}.$$

Thus the integral transformation above which provides an alternative characterization of Z is true because

$$\begin{aligned} \int_{\Omega} L(\Theta) d\mu(\Theta) &= \int_{\Omega} \tilde{L}(X(L(\Theta))) d\mu(\Theta) \\ &= \int_{\Omega} (\tilde{L} \circ \Phi)(\Theta) d\mu(\Theta) = \int_{[0,1]} \tilde{L}(x) d(\mu \circ \Phi^{-1})(x) \\ &= \int_0^1 \tilde{L}(x) dx \end{aligned}$$

Here, \tilde{L} is a generalized inverse of X under suitable assumptions on L , see ref. [15, 16]. This integral transformation holds at least in the case that L has no plateaus of positive **prior measure**, owing to the fact that in this case $\mu \circ \Phi^{-1}$ is indeed the uniform **measure** on $[0, 1]$, that is $d\mu \circ \Phi^{-1}(x) = dx$.

If L has a plateau of non-negligible mass, i.e., there exists a level λ^* such that $\mu(\{z \in \Omega: L(z) = \lambda^*\}) > 0$, then the derivation is more challenging, see also **Limitations and optimizations**.

2 Experimentation

In this section we will discuss how to implement NS, including considerations such as choosing the number of live points, drawing new live points from the **constrained prior**, parallelization and deciding when to stop.

2.1 Choice of the number of live points

As discussed in **Formulation**, the number of live points controls the *rate* of exponential compression during an NS run; we compress inwards by about $\Delta \log X \approx 1/n_{\text{live}}$ per iteration. This means that run-time scales as $\mathcal{O}(n_{\text{live}})$ (eq. (19)) and that the dominant uncertainties on the **evidence** integral (eq. (21)) and **posterior** scale as $\mathcal{O}(1/\sqrt{n_{\text{live}}})$. The above considerations are

Box 3 | Statistical mechanics analogy

There is a strong analogy between Bayesian inference and statistical mechanics. This suggests that NS might be useful in exploring problems that are typically the subject of statistical mechanical analysis. Consider Θ a **microstate** and $E(\Theta) = -\log L(\Theta)$ its energy. Then the **microcanonical ensemble** includes all states with $E(\Theta) = \epsilon$ where ϵ is a constant energy level. The volume of state space corresponding to a given energy ϵ is given by the density of states,

$$g(\epsilon) = \int \delta(E(\Theta) - \epsilon) \pi(\Theta) d\Theta$$

and taking the **prior** to be uniform corresponds to the **ergodic hypothesis**, namely that each **microstate** that is allowed by applicable conservation laws is equally likely to be observed. The Laplace transform of the density of states is the canonical **partition function** $Z(\beta) = \int e^{-\beta E} g(E) dE$, which corresponds to the generalized **evidence** (eq. (38)). The canonical ensemble is an alternative description of thermodynamic states that is based on the **inverse temperature** β rather than the energy level ϵ . In the canonical ensemble, states follow the Boltzmann distribution $p(\Theta | \beta) = \exp\{-\beta E(\Theta)\} / Z(\beta)$. NS in essence tracks the cumulative density of states via:

$$X(L) = \int_{E \leq -\log L} g(E) dE,$$

The practical approximation corresponding to eqs. (13) and (14) is

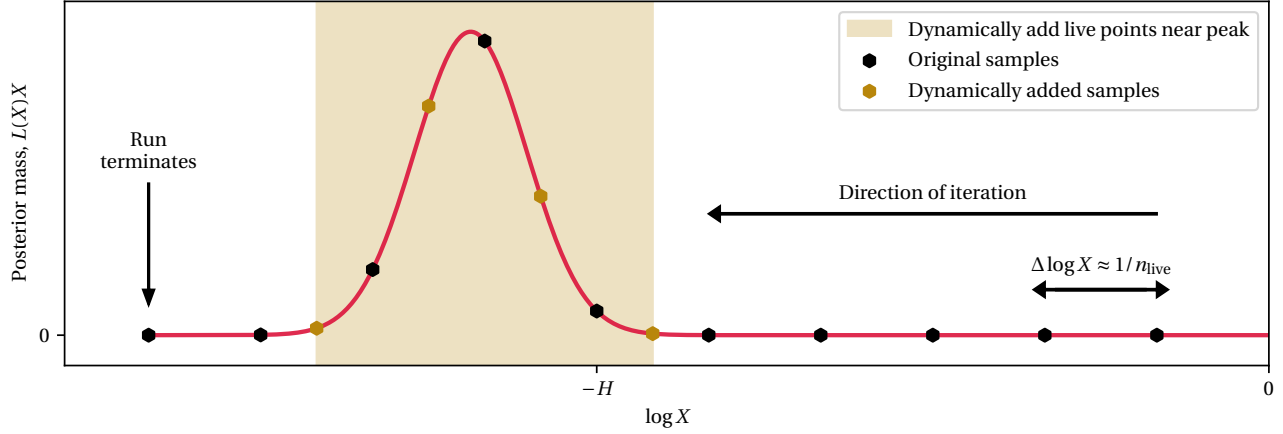
$$Z(\beta) = \sum e^{-\beta E_i} (X_{i-1} - X_{i+1}) / 2.$$

During NS, states are generated from the **prior** constrained by an upper energy limit $\epsilon = -\log L^*$, which can be achieved with any number of techniques, such as simple rejection sampling, Galilean Monte Carlo and Hamiltonian dynamics [44] or Creutz' microcanonical demon algorithm [45, 46]. The information entropy of the **constrained prior** is the **volume entropy** $\log X(\epsilon)$ (also known as Gibbs entropy). As NS progresses from one energy limit ϵ to the next $\epsilon' < \epsilon$, the volume entropy changes by $\Delta H = \log[X(\epsilon) / X(\epsilon')]$ at a rate that is constant on average: $\langle \Delta H \rangle = 1/n_{\text{live}}$.

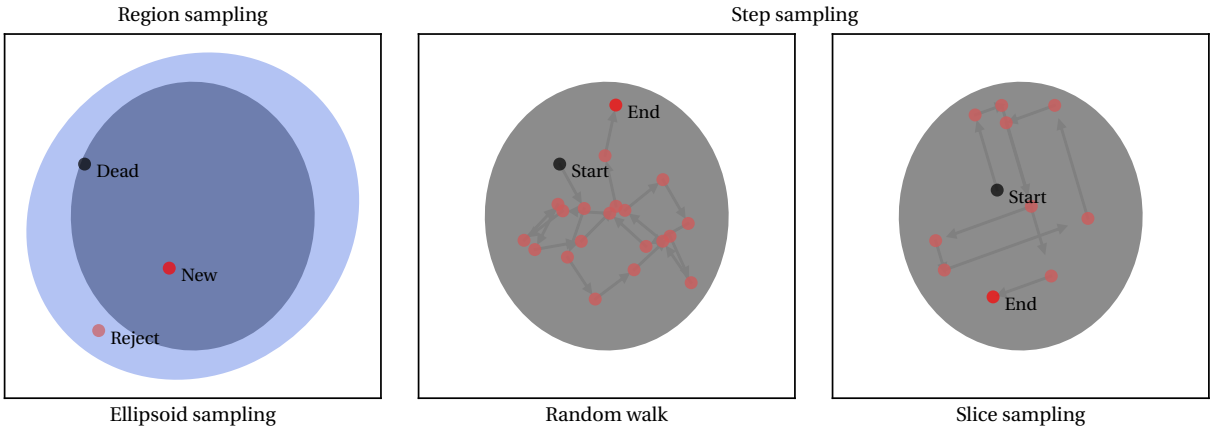
Widely-used tempering methods such as simulated annealing work in the canonical ensemble and use the inverse temperature β as a control parameter to weight each **microstates**. Alternatives include the Wang-Landau method [47] and NS, both of which use energy as a control parameter. In contrast to the ensemble property β , a key advantage of $E(\Theta)$ is that it can be evaluated for a single **microstate**. The Wang-Landau method uses a fixed, predefined set of energy bins, while NS constructs a sequence of energy levels at runtime. The sequence is optimal in that it achieves constant thermodynamic speed because changes in volume entropy are constant on average. Therefore, NS elegantly avoids a major problem of canonical annealing methods: designing a good temperature schedule.

represented graphically in fig. 2a.

There is thus a trade-off between run-time and uncertainty. Despite that, there isn't a straight-forward method for choos-



a | **Schematic representation of an NS run.** The curve $L(X)X$ shows the relative posterior mass, the bulk of which lies in a tiny fraction e^{-H} of the volume. Most of the original samples lie in regions with negligible posterior mass. In dynamic NS, we add samples near the peak.



b | **Examples of strategies for sampling from the constrained prior.** We must sample from the true iso-likelihood contour (grey ellipse). In region sampling (left) we bound the existing live points (blue ellipse) and draw a new sample from within that bound; some proposals may be rejected. In step sampling (right) we select a live point and perform a sequence of steps inside the contour to obtain an independent draw.

Figure 2 | **Experimentation in NS.** (a) Illustration of an NS run and (b) strategies for sampling from the constrained prior.

ing the number of live points if the required compression isn't known ahead of time. The number should be chosen, furthermore, bearing in mind the alternative role that it plays as a resolution parameter for NS [48, 49], especially in multi-modal problems. In particular, n_{live} should be large enough that at any time the constrained prior splits into disjoint modes, at least one live point lies inside the footprint of each mode. As a rough rule of thumb, if the constrained prior occupies a total volume X , only modes with a footprint greater than about X/n_{live} may be reliably found [8, 49]. This defines a resolution down to which the posterior is reliably sampled. Modes with smaller footprints are typically not located and correctly sampled, and hence will also not contribute to the evidence estimate. Moreover, to sample reliably and efficiently from the constrained prior, it is usually advisable that n_{live} exceeds the dimensionality of the parameter space.

2.2 Dynamic nested sampling

We have so far only considered NS with a fixed number of live points, and noted that the uncertainties in both posterior distributions and evidence estimates are reduced by increasing this number. However, the evidence depends on an accurate estimate of the total compression when we reach the posterior bulk. The posterior, on the other hand, depends only on an accurate estimate of the relative compression once inside the posterior bulk. The former uncertainty cancels in the posterior weights in eq. (15), as they are invariant under rescaling the estimates of the volume variable. This reflects the fact that whereas the evidence may depend strongly on the size of the prior, the posterior usually depends only weakly on its shape.

We are thus motivated to consider a *dynamic* number of live points to efficiently reduce uncertainties in parameter inference. As shown in fig. 2a, we may quickly compress to the posterior bulk using few live points. Upon reaching it, we may increase the number of live points, reducing uncertainty in the bulk of the posterior mass. We denote schemes that vary the number

Box 4 | Sequential Monte Carlo

Connections between NS and a form of rare-event SMC sampler [50] can be exploited to develop an NS-SMC algorithm with desirable theoretical properties [21].

SMC samplers evolve n samples (referred to as “particles”) through a series of distributions π_t for $t = 0, \dots, T$ using reweighting, resampling and mutation steps. Of particular interest is a form of rare-event SMC [50] that involves sampling from the sequentially **constrained prior** distribution. This SMC sampler has the sequence of distributions

$$\eta_t \propto \mathbb{1}(L(\Theta) \geq L_t^*)\pi(\Theta),$$

where $\mathbb{1}(\cdot)$ denotes an **indicator function** and $0 = L_0^* < \dots < L_t^* < \dots < L_{T+1}^* = \infty$. Each iteration t of the sampler involves sampling n particles with replacement from the current set of points satisfying $L(\Theta) \geq L_t^*$ and then diversifying those particles through several iterations of an η_t -invariant **MCMC** kernel.

Noticing the similarities between the two algorithms, ref. [21] built upon these rare-event SMC samplers to create the NS-SMC algorithm for evidence and posterior estimation. Approximation of Z and P in NS-SMC is done by weighting the particles from iteration t to target a shell of the **posterior**,

$$P_t \propto \mathbb{1}(L_t^* < L(\Theta) \leq L_{t+1}^*)L(\Theta)\pi(\Theta).$$

While original NS and NS-SMC both sample from the sequentially **constrained prior** distribution, there are several key differences. Crucially, NS-SMC uses weights based on importance sampling rather than numerical quadrature, it naturally handles the region beyond the largest threshold with no truncation error and it uses a different sampling mechanism. As a result, NS-SMC avoids the issues with bias in NS. Under mild conditions, including the situation where a finite number of **MCMC** steps are used, NS-SMC produces unbiased and consistent estimates of Z and consistent estimates of P as $n \rightarrow \infty$ [21]. Practically, NS can result in noticeably biased estimates of the **evidence** compared to NS-SMC for the same computational cost when the **MCMC** kernel is inefficient.

of live points as *dynamic* NS [51]. Open-source dynamic NS software packages include *dynesty* [52] and *dyPolyChord* [53]; see table 2. The gains are greatest in problems with substantial compression to the **posterior bulk**.

In dynamic NS schemes, the number of live points can be automatically adjusted to maximize a user-specified objective for a fixed computational budget. Usually, the run starts with an exploratory NS run using a constant number of live points. We spend the remaining computational budget by repeatedly increasing the number of live points in the most important regions of volume judged according to the objective (for example the shaded region in fig. 2a). Because dynamic NS re-winds a run and adds extra samples anywhere, running for longer reduces uncertainties and increases the effective sample size, unlike in ordinary NS.

In the original dynamic NS algorithm [51], a user specifies

their objective by assigning a relative importance to reducing uncertainties in the **posterior** and the **evidence**. When focusing on **posterior** inferences, dynamic NS can achieve orders of magnitude reductions in computational cost for a fixed uncertainty. The approach can also improve **evidence** calculation accuracy for a fixed number of samples, and can improve **posterior** inferences and **evidence** calculations simultaneously. This objective was generalised by the reactive NS [54] variant of dynamic NS. This considers the computation as a graph, with the nodes being the live and dead points, and edges indicating the replacement of a point by another. Multiple agents can then add live points (edges) where needed, and optimize towards additional goals, such as the effective sample size, or the number of samples per cluster. Lastly, we note a significant variant of NS, *diffusive* NS [55], in which the number of live points at a given **likelihood** threshold can change. In this variant, random walks starting from the existing live points are permitted to step down as well as up in **likelihood**, refining the typical **likelihood** in a volume range X .

2.3 Exploration strategies

NS progresses by replacing live points by independent samples drawn from the **constrained prior** in eq. (9), that is the prior restricted to regions in which the **likelihood** exceeds a threshold. This is the major difficulty in efficiently and reliably implementing NS, especially in **multi-modal** problems. Provided one successfully samples from the constrained prior, however, NS works identically in **uni-modal** and **multi-modal** settings.

Whilst we could simply sample from the entire prior until we find a sample for which the **likelihood** exceeds the threshold, this rapidly becomes incredibly inefficient due to the exponential reduction in the volume contained within the **constrained prior** at every iteration. Fortunately, the current set of live points and the estimate of the volume enclosed by the contour may guide our search for new live points. There are two main classes of methods for sampling from the **constrained prior** [56]: region samplers and step samplers. They are illustrated in fig. 2b. Analogous to the choices of **transition kernels** in **MCMC**, the choices here lead to various flavors of NS with different performance characteristics and different behaviour as dimension grows. Although a priori they require just as much tuning as **MCMC**, the live points allow them to build proposal structures and apply clustering algorithms (analogously to ensemble samplers [57]), such that NS is naturally amenable to being robustly self-tuning. As they are guided by them, their reliability and efficiency usually improve when the number of live points is increased; see ref. [54, 58] for numerical investigations. For both region and step samplers, if no current live points lie inside a mode in a **multi-modal** problem, that region of the constrained prior almost certainly won't be sampled, i.e., they miss that mode (see section 2.1).

Both samplers usually operate in the hypercube, a parameterization in which the prior is uniform over a unit hypercube (though this is not a requirement in NS). This slightly simplifies the problem to that of sampling uniformly from within a contour defined by the threshold. This is usually achieved by the

inverse transformation method. In this case, users specify their **priors** by an inverse-cumulative density function rather than a density function. For example, suppose we desired a Gaussian prior for a parameter, $\theta \sim \mathcal{N}(\mu, \sigma^2)$. We could transform a unit hypercube parameter, $u \sim \mathcal{U}(0, 1)$, using the standard normal distribution’s inverse-cumulative density function, Φ^{-1} ,

$$\theta = \mu + \Phi^{-1}(u) \sigma. \quad (22)$$

This in principle allows all manner of **priors**, though see [Limitations and optimizations](#) for further discussion.

Region sampling

In region sampling, we try to construct a region that bounds the **iso-likelihood contour** defined by the threshold. To find the region, we construct a geometric shape around the current distribution of live points. The shape must contain at least the currently estimated volume. We then draw independent and identically distributed (iid) samples from within that region until we obtain a sample that passes the current **likelihood** threshold. To be confident that the region did not encroach the contour, implementations of region sampling often expand the region by a user-specified factor or by a factor found through cross-validation. For example, dividing the live points into a training and test set and ensuring that the region found from the training set includes points in the test set [56, 59]. The expansion factor improves reliability at the expense of efficiency.

The simplest region sampler would be sampling from the entire unit hypercube, that is the entire prior, which rapidly becomes prohibitively inefficient. Instead, most region samplers attempt to estimate the **constrained prior** by wrapping the live points with one [60, 61] or more overlapping ellipsoids. Using more than one ellipsoid allows complicated and **multi-modal iso-likelihood contours** to be efficiently bounded. An appropriate number of ellipsoids can be found by applying clustering algorithms to the live points to identify distinct **modes**, such as x-means (ref. [62], see [Cosmology](#)). The shape and location of the ellipsoids may be approximately found from the mean and covariance of the live points they contain, and their volumes may be judiciously expanded by a tuning parameter. This has successfully been implemented in [MultiNest](#) [48]. Alternatively, [MLFriends](#) [56, 59] places an ellipsoid around every live point, and determines the ellipsoid scale by bootstrapping (similar to kernel density estimation with a uniform kernel).

Region samplers suffer from two major limitations when the dimension or complexity of the contour grow. First, the ability to accurately bound a complicated contour strongly depends on the number of live points. For instance, we may fail to identify a substructure with too few points, resulting in overly large regions that bound complicated substructures and in poor efficiency. Alternately, since live points are distributed uniformly within the **constrained prior** rather than near the edge, wrapping a small number of live points can result in overly small estimates of the **constrained prior**. Second, the accuracy and efficiency of region samplers suffer from the **curse of dimensionality**. To see how it strikes, consider region sampling with a single ellipsoid. Suppose that the true contour is a unit hyper-cube.

The smallest ellipsoid that we could construct that enclosed the contour would be a sphere of diameter \sqrt{D} . The volume of such a sphere blows up exponentially as dimension increases, leading to $\mathcal{O}(e^{-D})$ efficiency. This follows the general result that an exponentially increasing fraction of volume lies near the boundaries of high-dimensional shapes. As a result, region samplers are efficient and practical only for moderate-to-low dimensionality ($D \lesssim 20$).

Step sampling

Step samplers, by contrast, do not attempt to construct a region that bounds the **iso-likelihood contour** and so avoid some of the issues described above. Instead, they evolve a randomly chosen existing live point through a sequence of steps to a new approximately independent position. The acceptance rule is simply that we accept a transition to Θ' if

$$L(\Theta') > L^* \quad (23)$$

i.e., each step must stay inside the contour (cf. eq. (36)). Such step samplers are akin to running constrained **MCMC** inside NS, and were in fact the originally proposed solution. Strategies for generating new positions vary widely, and currently include:

- random-walk Metropolis [8], where new positions are proposed based on a local target distribution (e.g., a multivariate Gaussian),
- ensemble proposals [63], which use the distribution of all live points to propose new positions using strategies such as differential evolution [64]
- slice sampling variants, where new positions within the **constrained prior** are proposed along a randomly chosen principal axis (slice sampling; [65, 66]), or randomly chosen direction (hit-and-run; [67, 68]), and
- gradient-based trajectories [46, 52, 69–71] that ‘reflect’ off the current **likelihood** constraint.

See ref. [54, 72, 73] for further discussion. In step samplers with a step size parameter, such as random walk Metropolis, the step size is often tuned to ensure a substantial fraction (> 20%) of proposed positions are accepted. This avoids an unacceptable overall sampling efficiency [3, 8]. This tuning may be performed continuously over the course of an NS run [8, 52, 66], although this can introduce biases [71].

It can be challenging to judge the number of steps required to ensure that the live points are independent draws from the constrained prior (see Supplementary Information D for further discussion). While mild violations of this requirement might be inconsequential [52], strong violations lead to unreliable NS **evidence** estimates [74]. The number of iterations required for new samples to approximately de-correlate scales as $\mathcal{O}(D^2)$ for random-walk proposals with tuned step sizes, and $\mathcal{O}(D^1 - D^{5/4})$ for slice sampling or gradient-based trajectories [3]. In practice, the number of steps is often chosen to be aD^b , where a is of order one and b is the anticipated dimensional scaling.

The computational cost scales linearly with the number of steps. Thus unlike region samplers, step samplers escape the **curse of dimensionality** as their cost shows only polynomial $\mathcal{O}(D^b)$ scaling with dimensionality. Nevertheless, region samplers are often more efficient in low dimensions. As a result, step samplers are more often used when applying NS in high-dimensions ($D \gtrsim 20$).

Table 2 compares the approaches in several publicly available NS implementations that originated in different research fields. For a multitude of programming languages, there are well-documented, free and open source codes. Support for parallelization to computing clusters and checkpointing are also common features. They usually work with the logarithms of the **evidence** and **likelihood**, as the latter may be numerically tiny such that it cannot be represented as a floating point value. In general, sensible defaults for numerous region and step samplers have been found to work across a large variety of problems [52]. We present a simple numerical example in Supplementary Information G.

2.4 Parallelization

To utilize computing resources, we could parallelize the computation of the likelihood function. If that is impossible or impractical, we may wish to design an efficient scheme for parallelizing an NS run itself. As discussed in ref. [51, 52], statistically independent NS runs may be combined into an equivalent NS run with $\sum n_{\text{live}}$ live points. To achieve this, pool the initial live points from the independent runs a, b, \dots together, giving $\sum n_{\text{live}}$ live points drawn from the prior. Remove the worst, supposing it lay at L^* and originated from run a . To draw a replacement from the prior subject to $L > L^*$, simply take the replacement used in run a , since it is already a draw from the prior subject to $L > L^*$. We may continue in this fashion, weaving together independent NS runs to build a new NS run with $\sum n_{\text{live}}$ live points. This allows parallelization of an NS run with n_{live} live points into n_{CPU} independent runs with about $n_{\text{live}}/n_{\text{CPU}}$ live points each. The independent runs themselves proceed linearly and may be ultimately combined, resulting in a speedup of about n_{CPU} . Even simpler, estimates of the **evidence** integrals may be combined by weighted averaging.

However, the reduction in the number of live points per run, $n_{\text{live}}/n_{\text{CPU}}$, impacts the exploration schemes discussed in section 2.3, especially for ellipsoidal rejection sampling where it could lead to inefficient or faulty bounding ellipsoids. It may therefore be desirable to utilize parallelization within individual NS runs. For example, by drawing n_{CPU} candidate replacement points and evaluating their **likelihoods** in parallel. We could subsequently replace the worst n_{CPU} live points at each iteration [75, 76], replace a single live point and consider the other evaluated points at subsequent iterations [66], or replace a single live point and discard as many as $n_{\text{CPU}} - 1$ acceptable candidate points [48]. The latter is wasteful if more than one viable point are likely to be found among the n_{CPU} candidates,

$$\text{Speed-up} \approx \min[n_{\text{CPU}}, 1/\epsilon] \quad (24)$$

If points are considered at subsequent iterations,

$$\text{Speed-up} \approx n_{\text{live}} \log \left(1 + \frac{n_{\text{CPU}}}{n_{\text{live}}} \right) \quad (25)$$

and so a speedup of about n_{CPU} if $n_{\text{CPU}} \ll n_{\text{live}}$. The expression originates from the fact that the threshold increases as the run progresses, meaning that points drawn from the **constrained prior** might not be valid at a subsequent iteration. Lastly, replacing n_{CPU} points in parallel per iteration results in a speedup of about n_{CPU} but increases the variance in the **evidence** integral by about $\sqrt{n_{\text{CPU}}}$. Increasing n_{live} by a factor $\sqrt{n_{\text{CPU}}}$ to maintain the same uncertainty decreases the speedup to about $\sqrt{n_{\text{CPU}}}$.

2.5 Stopping conditions

We must decide when to stop an NS run. The fact that we can only perform a finite number of iterations introduces a truncation error in eq. (13) that we wish to be negligible. Skilling originally proposed to stop NS once we reached the **posterior bulk** at $X \simeq e^{-H}$ at iteration $n_{\text{iter}} \simeq n_{\text{live}}H$, or using an estimate of the remaining **evidence**. Popular NS software later adopted the latter. In **MultiNest**, this was based on the maximum **likelihood** found so far, $\max LX$, whereas **PolyChord** chose the mean **likelihood**, \bar{LX} . They stopped once $\Delta Z/Z \leq \text{tol}$, where tol is a user-specified parameter. Upon deciding to stop, the truncation error in the **evidence** was corrected by either adding an estimate of the remaining **evidence** or by killing the live points one by one without replacement and incrementing the **evidence** in the usual manner until no live points remain. The latter is in keeping with the NS approach. For the former the remaining **evidence** may be estimated by \bar{LX} . See ref. [43] for further discussion of the statistical properties of this estimate of the remainder. When NS is used to calculate the **partition function** of a material system (box 3), physically motivated stopping conditions can be based on the expected minimum energy (negative \log **likelihood**) or the sampled temperature, which is proportional to the derivative of the limiting energy with respect to NS iteration [77]

None of these approaches guarantees that summation hasn't been terminated too early; there could well be a spike of enormous **likelihood** lurking inward. The computational budget, furthermore, cannot be easily anticipated ahead of time. However, runs that are terminated prematurely may still be used to illustrate what was learned so far about the **posterior**. Proposals to construct termination criteria for a fixed computational budget that result in unbiased estimates of the **evidence** have been suggested [24].

3 Results

NS results in an estimate of the integral in eq. (1) and, in the context of Bayesian statistics, a weighted set of draws from the **posterior** distribution. We discuss some examples in **Applications** and how to report the results in **Reproducibility and data deposition**. The error estimate in eq. (21) depends on the compression and cannot be known ahead of time. If the

Code	Methods	Dynamic	Languages	Field	Pub. Year
CosmoNest [60, 61]	ellipsoid	fixed	Fortran	Cosmology	2006
MultiNest [48, 84]	multi-ellipsoid	fixed	Fortran, C/C++, Python	Cosmology	2008
DIAMONDS [249]	multi-ellipsoid	fixed	C++	Astrophysics	2015
nestle [250]	ellipsoid, multi-ellipsoid	fixed	Python	Astrophysics	2015
nessai [90, 91]	normalising flow ellipsoid	fixed	Python	Gravitational waves	2021
(dy)PolyChord [53, 66]	slice	dynamic	Fortran, C/C++, Python	Cosmology	2015
LALInferenceNest [180]	random walk, ensemble, differential evolution	fixed	C	Gravitational waves	2015
Nested_fit [105, 257, 258]	random walk	fixed	Fortran	Atomic physics	2016
cpnest [259]	slice, differential evolution, Gauss, Hamiltonian, ensemble	fixed	Python	Gravitational waves	2017
pymatnest [44]	random walk, Galilean, symplectic Hamiltonian	fixed	Python	Materials	2017
NNest [261]	normalising flow random walk	fixed	Python	Cosmology	2019
DNest5 [55]	user-defined, random walk	diffusive	C++	Astrophysics	2020
BayesicFitting [263]	random walk, slice, Galilean, Gibbs	fixed	Python	Astronomy	2021
dynesty [52]	ellipsoid, multi-ellipsoid, MLFriends & Gauss, slice, Hamiltonian	dynamic	Python	Astrophysics	2020
UltraNest [92]	MLFriends + ellipsoid & Gauss, hit-and-run, slice	reactive	Python, Julia, R, C/C++, Fortran	Astrophysics	2020
jaxns [266]	multi-ellipsoid & slice	fixed	jax	Astronomy	2021

Table 2 | **Comparison of NS codes.** The first two groups are region samplers and step samplers, respectively, whereas the third group offers both. Dynamic implementations allow the number of live points to be changed during a run. We show the language in which the NS code was written followed by any additional languages for which interfaces exist, and the field from which the code originated (though most are general purpose codes).

achieved error is unacceptable, the NS run can be repeated with more live points or combined with a new run. The shape of the posterior can be found from the posterior weights by density estimation. There are dedicated software packages for making publication quality figures of marginalized posterior densities from weighted samples, including *anesthetic* [78], *superplot* [79], *pippi* [80], *dynesty* [52], *getdist* [81], *corner* [82] and *pygtc* [83].

There are a few ways to check the results. First, we can check the NS implementation, rather than the particular run. To do so, we can compute the evidence integral in eq. (1) for problems with known analytic solutions; see ref. [66, 84–86] for examples including a multi-dimensional Gaussian, an eggbox function, the Rosenbrock function [87], Gaussian shells and a mixture of a Gaussian and a log-gamma distribution. Similarly, in some cases, the NS estimates of the volume variable at each iteration, $X(L^*)$, may be checked against analytic results [56]. If discrepancies are found, the implementation is suspect. Alternatively,

we can repeat calculations and check whether the distribution of results is consistent with what would be expected if the software was working correctly. Ref. [74] describes procedures for doing this, including tests requiring only two NS runs; these are implemented in *nestcheck* [88].

Second, the particular NS run of interest may also be checked using a test of the insertion indexes of new live points [89]. If NS draws new live points independently from the constrained prior, as it should, the ranks in likelihood of each new live point compared to the current live points should be uniformly distributed. This test is implemented in the *anesthetic* [78] NS analysis software, which is compatible with *PolyChord* and *MultiNest*, and used on the fly in the *nessai* [90, 91] and *UltraNest* [92] NS implementations. If this test fails, but other implementation checks pass, the choices of exploration strategy for the problem at hand may be inadequate. Similarly, ref. [93, 94] discuss testing whether live points are uniformly distributed in the unit hypercube in two-dimensional problems. Last, in the context

of parameter inference we can compare the posterior samples obtained from different NS implementations or from MCMC and NS (see for example ref. [95]) or perform **Simulation-Based Calibration** to check expected properties of the posterior.

As discussed in **Stopping conditions**, most NS implementations stop once the estimate of the remaining **evidence** appears negligible. This could omit spikes in **likelihood** lying inside the remaining unexplored volume. If that is a concern, it may be beneficial to optimize the **likelihood** using a local optimization algorithm starting from the live point with the greatest **likelihood**. The run may be restarted with stricter stopping conditions if local optimization finds a maximum **likelihood** that is orders of magnitude greater than that found during the NS run.

Lastly, in many settings we may be concerned about the possibility that **modes** were missed. While the tests described above may indicate whether implementation errors are present, it is in general impossible to know for certain whether all **modes** were identified. We can inspect the **posterior** by eye, or perform a **mode** identification algorithm on the **posterior** samples. If the expected number of **modes** was known and **modes** are evidently missing, the number of live points should be increased.

4 Applications

Here we present the most established NS applications, highlighting its advantages in each case. Besides these established applications, NS is beginning to be utilized in many other areas, including signal processing [96], phylogenetics [97], systems biology [98, 99], acoustics [100, 101], nuclear physics [102, 103], atomic physics [104–108], exoplanet searches [109–113], and geology [114].

4.1 Cosmology

The rapid spread of Bayesian methods in cosmology [9] in the early 2000s was generated by the growth of data, specifically the new cosmic microwave background (CMB) temperature power spectrum measured by the Wilkinson Microwave Anisotropy Probe (WMAP) [115], and the type-Ia supernovae distance measurements [116, 117]. However, these new and powerful cosmological datasets had raised difficult questions regarding the cosmological model. The favoured model of the Universe included a mysterious accelerating force, the dark energy, which accounted for 70% of the energy density today. And the initial spectrum of density fluctuations in the early Universe, which were shown to be Gaussian and adiabatic by WMAP, indicated a period of accelerated expansion at very early times (about 10^{-23} seconds after the Big Bang), known as cosmic inflation [118]. Finally, there was still the question of the missing mass of the Universe, which generates the gravitational fields required for cosmic structure, known as cold dark matter (CDM). All three of these phenomena (cold dark matter, dark energy, and cosmic inflation) had proposed explanations from the field of high-energy theoretical physics, and these model predictions could be combined with the new wealth of cosmological data to be evaluated with respect to each other, model by

model. Thus model selection in general, and the computation of the Bayesian **evidence** using NS specifically, became a tool of choice.

The simplest model of cosmic inflation is one driven by a single scalar field, a particle physics object similar (but not identical) to the Higgs boson. The behaviour of the scalar field driving inflation is determined by its potential $V(\phi)$ (ϕ being the value of the scalar field) [118, 119]. The formulation of $V(\phi)$ (as a function of some set of parameters Θ) determines the duration of inflation along with the distribution of anisotropies found in the CMB. Ref. [120] performed a detailed **Bayesian model comparison** between 193 inflationary models using the region sampler **MultiNest**. They used the cosmological observations (including CMB data from WMAP mission) to discriminate between alternative models for $V(\phi)$ and found slight preference for so-called Small Field Inflation (SFI) models over Large Field Inflation (LFI) models.

Another area where NS has been used extensively in cosmology is the modelling of galaxy clusters [121, 122]. Clusters of galaxies are the most massive gravitationally bound objects in the Universe and therefore, can be used to trace the formation of large scale structure in the Universe. Galaxy clusters can be observed through several methods including X-ray observations, weak gravitational lensing and by exploiting the Sunyaev–Zel’dovich effect (SZ). Weak gravitational lensing involves the distortion of the images of the background galaxies by the presence of a large mass lying along the line-of-sight. Weak lensing allows to probe the total mass distribution, including the dark matter, of the galaxy cluster. Ref. [123] presented a Bayesian approach using NS for the detection of galaxy clusters in SZ data, including estimation of parameters associated with the physical model assumed for the cluster and quantification of the detection using Bayesian model selection, making use of the statistics of the CMB anisotropies in the **likelihood** function. Ref. [124] presented a joint Bayesian analysis of weak lensing and SZ observations of several galaxy clusters, allowing for the estimation of gas fraction of individual clusters. In modern weak lensing surveys [125] such as DES [126] and KiDS [127], NS forms a critical part of their parameter estimation, model comparison and tension quantification pipelines [128] (fig. 3).

The expansion history of the Universe can also be measured through the use of distance measurements, such as observations of type Ia supernovae (SNIa), which can be used as ‘standard candles’. The important results from the Supernova Cosmology Project and the High-Z Supernova Search Team [116, 117] presented **evidence** for the accelerated expansion of the Universe, requiring the existence of a mysterious dark energy acting against gravity. Estimation of cosmological parameters from the observations of SNIa light curves has historically been done using a χ^2 approach (see e.g. [129]) which lacks a rigorous approach for the determination of systematic uncertainties. Ref. [130] introduced a Bayesian hierarchical model for the determination of cosmological parameters. By using NS, they showed that their principled Bayesian approach delivered tighter statistical constraints on the cosmological parameters over 90% of the time, reduced statistical bias by a factor ~ 2 -3

times and that it had better coverage properties than the standard χ^2 approach.

The measurement of the CMB anisotropies by Planck [132] was an increase in the statistical power over the previous experiment (WMAP), but required more sophisticated modeling of galactic foregrounds and instrumental calibration, introducing $\mathcal{O}(20)$ “nuisance parameters”. Metropolis-Hastings techniques for parameter estimation were able to accommodate these by exploiting the fact that these parameters were “fast” in comparison to the cosmological ones; namely, by caching results from previous calculations, these parameters can be changed with negligible computational cost providing the cosmological parameters remain fixed.

Region-based samplers such as **MultiNest** are not able to exploit this hierarchy of parameter speeds, and could not navigate the now high-dimensional cosmological + nuisance parameter space reliably. To address this, cosmologists turned to step sampling based strategies as instantiated in the **PolyChord** algorithm [66, 133] which uses slice sampling.¹ These were successfully applied throughout the 2015 Planck inflation paper [135] to non-parametric reconstructions [136] (fig. 3) and general inflationary model comparison, in particular to the challenging example of axion monodromy models. The non-parametric reconstruction approach, which only became possible with the ability to reliably (and fully) explore the parameter space of a large number of parameters, was an important model-independent demonstration of the simple power-law behavior of the primordial power spectrum of density perturbations.

Although step sampling was originally introduced in cosmology to exploit a fast-slow hierarchy of parameters, the better dimensional scaling opened up a new range of cosmological analysis possibilities that were inaccessible to region samplers. It has been applied to constraining kinetically dominated inflation models [137], model comparison for the quantum mechanical initial conditions for the Universe [138], additional reconstructions of the dark energy equation of state [139, 140], astronomical sparse reconstruction [141], and played a critical role in the GAMBIT combined cosmology and particle analyses [142, 143]. Step sampling takes a leading role in the REACH 21cm global cosmology analysis [144], and at the other end of the astrophysical scale it has also been applied to high-dimensional exoplanet analyses [111, 113].

4.2 Particle Physics

Particle physics is a field related to cosmology that has also seen various applications of NS. Around 2010, when NS tools such as **MultiNest** were reaching maturity, the particle physics community was focused on the first results from the Large Hadron Collider (LHC). A particularly favoured theoretical framework was supersymmetry (SUSY; see e.g., ref. [145]). SUSY introduces

¹Axial slice sampling had been applied in the then near past to systems biology [134] although unsurprisingly the cosmological authors were not aware of this work. **PolyChord** improved upon the existing slice-sampling method by implementing covariance-based non-axial steps and **mode** clustering in addition to the ability to exploit a hierarchy of parameter speeds.

an array of new particles and unknown parameters that can be fitted to collider data as well as observations from other experiments in a so-called global fit. The physics goals of a global fit are to understand what the model predicts in future experiments and how best to discover the new particles.

These global fits of $\mathcal{O}(10)$ free parameters in SUSY models presented a problem for which NS tools were naturally well-suited, as they were typically **multi-modal**. A package named **SuperBayeS** [146–148] utilized the **MultiNest** implementation of NS and was used to make a number of early LHC predictions and fits [149–154]. As the LHC results have poured in throughout the 2010s, the theoretical landscape shifted and a wider set of models have been considered using NS (see e.g., ref. [155–158]). More recently, the GAMBIT collaboration has driven many such global fits making use of **MultiNest** and **PolyChord** to sample parameter spaces and compute Bayesian evidences, and benchmarked NS against a number of other MC based and gradient-free methods [159, 160]. Related problems such as tuning phenomenological parameters in event generators have seen preliminary work, and could be rich avenues of NS application in the field. Lastly, we note that NS was recently applied to the sampling space, rather than the parameter space, of a statistical model. This enables efficient computation of small p -values that are used in the discovery of new particles at the LHC [161].

4.3 Gravitational waves

Gravitational-wave (GW) astronomy is a field that has blossomed since the first observation in 2015 of two colliding black holes [162] by the LIGO [163] and Virgo [164] interferometers. The signals are produced by non-axisymmetric changes in the gravitational field, typically sourced by the rapid motion of neutron stars and black holes in binary systems [165]. The binary orbits decay through GW emission, increasing the orbital frequency and the rate of energy loss until the objects merge. Such events, known as compact binary coalescences (CBCs), produce signals at frequencies of 10–1000 Hz which are recorded in the detectors as a time-series.

Early development of Bayesian methods for GW analysis was contemporaneous with the first publication of NS, and it was found that NS provided an efficient means to robustly sample the **posterior** distributions of GW signals [166]. This is important as the **posteriors** provide a rich new astrophysical view: from measuring the masses of neutron stars to the expansion rate of the Universe itself. The small signal-to-noise ratios of observed signals and degeneracies in the model parameter space produce **posterior** distributions that are often **multi-modal** and highly correlated. To date, only NS and **MCMC** approaches have been able to robustly sample the **posteriors** of all observed events. While each has its own advantages, agreement between the NS and **MCMC** approaches has been critical in building confidence in results. However, the NS approach is better-suited to providing robust **evidence** estimates for model comparisons [167]; the efficiency of the NS approach does not depend so much on the use of problem-specific proposals; and massively parallel approaches using the **dynesty** code have

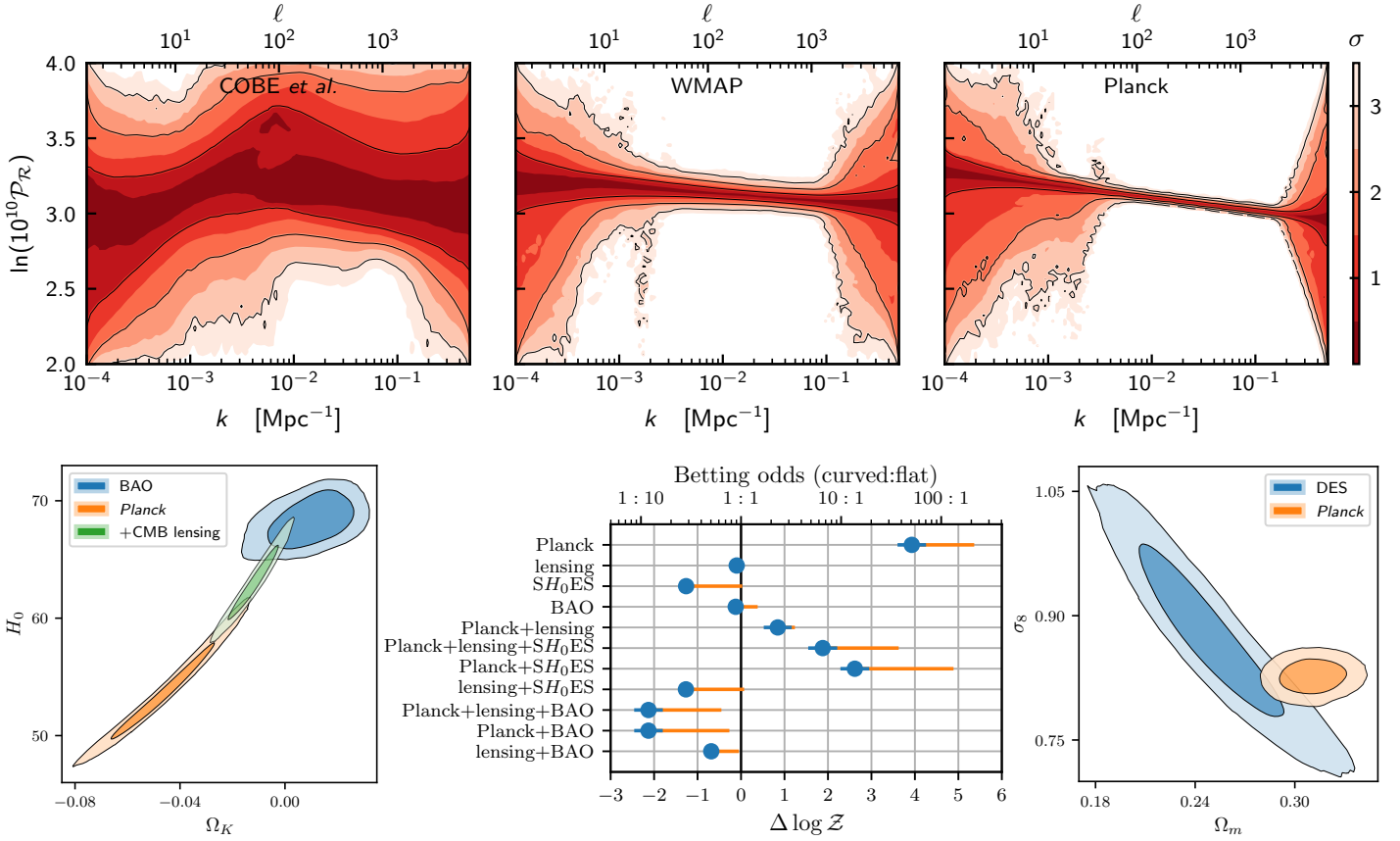


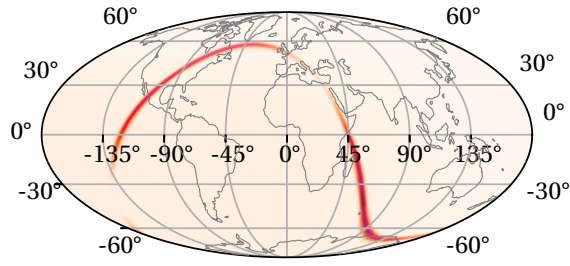
Figure 3 | **Cosmological applications of NS.** *Top row:* Cosmological non-parametric reconstruction of the power spectrum of primordial cosmological fluctuations, as measured by cosmic microwave background satellites across human history. The reconstruction uses a linear spline-based procedure with N movable knots. Each of the locations of the knots along with the cosmological and nuisance parameters are varied in a full NS fit. The **evidence** is then used to marginalize over N to produce the final plots. The standard model of cosmology predicts a featureless tilted power spectrum, so such non-parametric reconstructions are of great interest for astrophysicists searching for **evidence** of beyond standard model physics. *Bottom row:* NS in cosmological tension quantification. Bayesian **evidences** computed by NS can be used to quantify the level of disagreement which may be hidden by marginalization of high-dimensional parameter spaces. *Bottom left:* Planck CMB data is in tension with CMB lensing and BAO in the context of curved cosmologies, so caution should be exercised in combining them [131]. If one includes only CMB data, a **Bayesian model comparison** shows preference for a closed Universe relative to a flat one (*Bottom middle*) in spite of the Occam penalty associated with the additional parameter (orange bar). *Bottom right:* There is also tension between weak lensing data (DES) and the CMB (Planck).

made analyses with more advanced signal models [168] computationally tractable. The success of NS in analysing merging binary signals has inspired many other efforts such as the analysis of continuous signals from individual rapidly-rotating non-axisymmetric neutron stars [169, 170]; detection of unmodelled sources [171]; model selection between different physical mechanisms of core-collapse supernova encoded in the GW signal [172]; and detection of a stochastic superposition of weak merger sources [173].

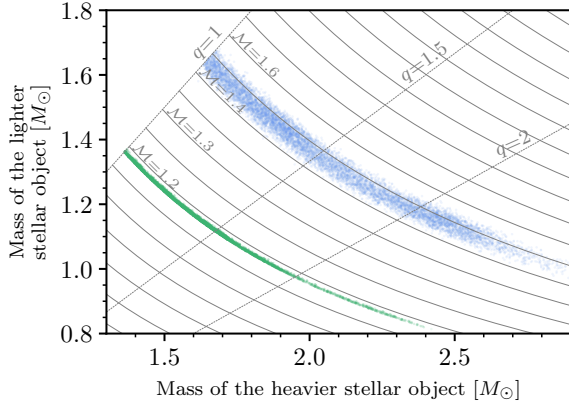
CBCs consisting of two black holes have been the dominant sources of GW signals seen by the LIGO and Virgo detectors [165]. The signals are modelled through a combination of post-Newtonian approximations to General Relativity and relativistic numerical modelling [174–177]. The signal model, as observed in a detector, is parameterized by 8 parameters that

are intrinsic to the binary system (the individual masses, m_1 and m_2 , and their 3D spin vectors) and 7 parameters related to the relative orientation and position of the system with respect to the detector (including the source’s luminosity distance D_L and location on the sky). For systems that include at least one neutron star, the signal model’s phase evolution requires additional parameters related to the neutron star equation-of-state [178].

Signals from black hole mergers last only a few seconds within the sensitive frequency regime of the detectors. On these timescales the detector data can be approximated as having noise drawn from a stationary Gaussian process described by a known power spectral density, which can either be estimated from data surrounding the signal [180], or inferred directly using a parameterized model of its shape [181]. For inference, a



a | Sky location of the source of the binary black hole merger signal GW151226. The posterior probability distribution from NS is shown in orange [179].



b | Masses of stellar objects involved in two binary neutron star mergers. The posterior samples from NS are shown for the neutron star mergers GW170817 (green) and GW190524 (blue). Solid grey lines mark curves of constant chirp mass \mathcal{M} , while dashed lines mark curves of constant mass ratio q . When viewed in \mathcal{M} - q space, the posteriors show only weak correlation.

Figure 4 | Applications of NS in GW astronomy.

Gaussian likelihood function of the form given by Whittle [182] can be used [180]. For multiple detectors, which have independent noise, the likelihoods can be coherently combined using the product rule of probability. The prior distributions used are discussed in ref. [180], but are generally set to be uninformative (e.g., uniform over a sphere for sky coordinates), or constrained to be within a physically reasonable range. In addition to the parameters related to the source, the likelihood can also contain ~ 60 unknown parameters that relate the frequency-dependent uncertainties in the phase and amplitude calibration of the detectors [183].

For an observed signal, using the likelihood and priors discussed above, the joint posterior of all these 15 (or more) parameters can be extracted after application of NS. The posterior distribution of the typical events observed so far are not unimodal or Gaussian. Due to the relatively small signal-to-noise ratio, they exhibit significant degeneracies and correlations. For example, the posterior of the source sky location is largely determined by differences in arrival time of the signal at different detectors. This produces ring-shaped degeneracies as demonstrated in fig. 4a.

The many degeneracies encountered for typical CBC inference have led to significant work to identify optimal parameterization and, where possible, utilized a marginalized likelihood (see ref. [184] for a review). An optimal parameterization involves identifying a mapping between the physical model parameters and combinations of these which reduce the complexity of the target density. Usually, a good reparameterization involves identifying the combinations of physical parameters which are ‘best measured’. As an example, the physical mass of the two component stellar objects are m_1 and m_2 . However, there is a strong banana-like correlation between the two masses (see, e.g. fig. 4b) and an exact degeneracy under exchange of m_1 and m_2 . To enable efficient sampling, Veitch *et al* [180] proposed sampling in the chirp-mass, \mathcal{M} , and mass ratio q : two algebraic combinations of the component masses. In fig. 4b, we show that the posterior, as viewed in $m_1 - m_2$ space, follows contours of the chirp mass. Physically, this is because the chirp-mass is the most well measured mass parameter (i.e. has the smallest posterior width) followed by the mass ratio.

The chirp-mass and mass ratio parameterization, along with numerous others have greatly improved sampling efficiencies. We note that the choice of sampling parameters, which we choose for computational efficiency, is separate from the choice of prior distributions. If required, a Jacobian transformation [185] can be made to enable sampling in the optimal sampling parameters while setting priors on the physical parameters.

4.4 Materials science

As hinted at in box 3 above, NS can be used to study the thermodynamic properties of molecules and materials, which ultimately derive from the partition function

$$Z(\beta) = \int e^{-\beta H(q,p)} dq dp, \quad (26)$$

where $q \in \mathbb{R}^{3N}$ specifies the spatial coordinates of the N particles in the system, $p \in \mathbb{R}^{3N}$ is the corresponding momentum vector, $H(q,p)$ is the Hamiltonian, and $\beta = 1/k_B T$ is the inverse temperature.

The classical Hamiltonian can be separated into the configuration dependent potential energy, $U(q)$, and momentum dependent kinetic energy, $K(p) = \sum p^2/2m$ (where m is the mass of each particle), giving

$$Z(\beta) = \int e^{-\beta K(p)} dp \int e^{-\beta U(q)} dq. \quad (27)$$

(For the fully quantum treatment using NS, see ref. [186].) The first factor is a Gaussian that may be computed analytically,

$$Z_p \equiv \int e^{-\beta K(p)} dp = \frac{1}{N!} \left(\frac{2\pi m}{\beta \hbar^2} \right)^{3N/2}, \quad (28)$$

and we tackle the second, configuration dependent factor by NS. From the partition function all thermodynamic quantities of relevance can be calculated, for example, the average energy,

$$\langle H(q,p) \rangle = -\frac{\partial \ln Z(\beta)}{\partial \beta} \quad (29)$$

and the heat capacity

$$C_V = \frac{\partial \langle H(q, p) \rangle}{\partial T}, \quad (30)$$

and these may be interpreted as **posterior** expectations to be computed after the NS run. In this application,

- the factor $e^{-\beta U(q)}$ plays the role of the **likelihood** function, parameterized by the inverse temperature, β ;
- the prior, according to the ergodic hypothesis, is uniform;
- the dimensionality is commonly on the order of 10^2 to 10^3 or more.

The striking difference from standard Bayesian inference is the inverse temperature parameter β in the **likelihood**. Rather than having a single inference problem for some fixed value of β , almost always we want to know the behaviour of observables *as a function of temperature*, so effectively we have a continuous family of inference problems. And for this, NS (and other density-of-states methods, see box 3) have a remarkable feature: the **likelihood** is a monotonic function of β and hence the entire NS algorithm is invariant to changes in β . In practice this means that a single NS run can be used to calculate observables at all temperatures. For large β values (corresponding to low temperatures), the **partition function** is dominated by the lowest energy minima (highest **likelihood modes**), and for molecules and materials these configurations correspond to the globally stable structures. On the other hand, when β is small (corresponding to high temperatures), the **partition function** is dominated by the large volume associated with high energy states — or in the parlance of materials science, “entropic effects” [187, 188].

Some of the most interesting phenomena in molecular and materials science (and more generally in statistical mechanics) are associated with the above change of regime. Collectively known as *phase transitions*, they are characterized by a dramatic change of where the **bulk** of the **posterior** mass lies, as the temperature (or other system parameters such as pressure) is varied — and this is what makes it very challenging to study them using numerical sampling schemes. Experimentally, phase transitions are often observed indirectly as changes in the expectation value of observables (e.g. discontinuously, in the case of first order phase transitions such as melting and evaporation), or more directly as sharp peaks in response functions, such as the heat capacity or magnetic susceptibility. The promise of NS is to enable the calculation of such response functions in general with high reliability and the minimum of fuss.

There are some aspects of the materials application of NS that turn out to be favourable in comparison with the general inference problem. The first is a rather easy stopping criterion for the NS iterations. For any given model of the potential energy, it is generally not hard to come up with a good global lower bound on the energy, which translates into an upper bound in the **likelihood**. As the **likelihood** values sampled by NS appear to converge, if this is close to the known bound, the

iterations can be stopped without the risk of missing the highest **likelihood** mode.

Secondly, convergence of NS with the number of live points and other sampling parameters is desired and observed in terms of convergence of the heat capacity peak locations, and this typically occurs far earlier than the decorrelation of the sampler chains that are used to explore the **constrained prior**.

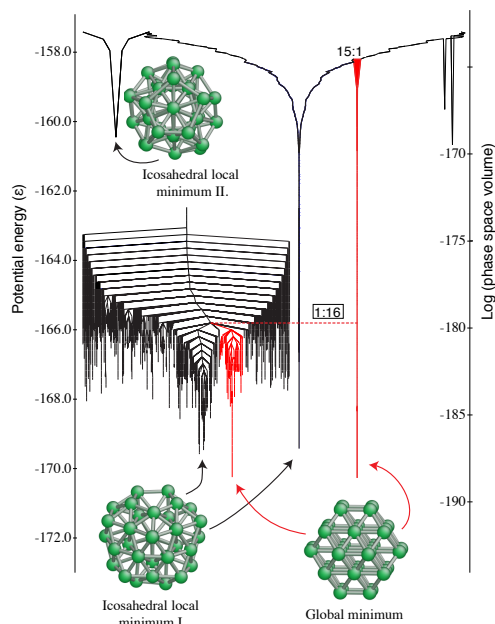
Thermodynamics of LJ Clusters

The Lennard-Jones (LJ) potential is a simple model for describing the pairwise interactions between atoms and it provides the basis for benchmarking algorithms for modelling materials. In particular, specific sized clusters of LJ atoms exhibit complex thermodynamic properties due to solid-solid transitions (the finite system analogue of first order phase transitions mentioned above) caused by existence of competing low energy minima, in addition to solid-fluid melting [188–190].

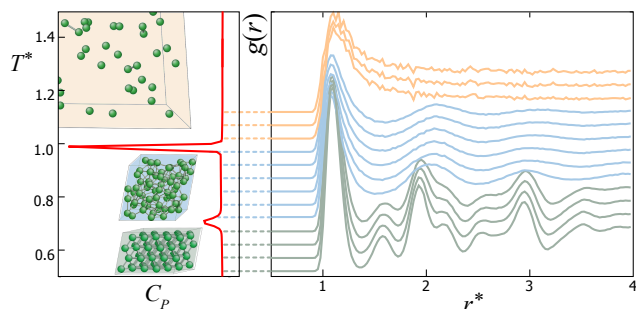
We illustrate some features of atomistic energy landscapes in fig. 5a using the example of the cluster of 38 particles (LJ₃₈) along with the corresponding *disconnectivity graph* [191, 192] which help to convey the relationships between the large number of local energy minima (**likelihood modes**). Each leaf of the tree structure corresponds to a local minimum of the energy (or to a closely related set of them), and junctions represent saddle points connecting the minima. As the number of particles in the system is increased, the number of distinct local minima grows exponentially. According to one estimate, LJ₃₁ for example has 10^{15} distinct minima, not counting permutation isomers. [188]

Potential energy disconnectivity graphs are good at showing the topology of the energy minima, but to reflect the volumes’ free energy, disconnectivity graphs are required [194, 195]. Using NS, *energy landscape charts* [49] can be generated, and we show one for LJ₃₈ in fig. 5a. The vertical axis is the potential energy, and the curve shows a series of (possibly nested) basins, whose width at each potential energy level is proportional to the volume of the corresponding **posterior mode** slice. In order to fit the chart in one diagram, the volumes are also scaled by an exponential factor whose logarithm is shown on the right hand axis. Each basin, encompassing a range of closely related configurations, corresponds to a macroscopic “state” of the system — illustrative structures are shown. The relative volume of the funnel associated with the global minimum (shown in red) and the entropically dominant minimum (in the centre) is 1 : 15 at the energy level where the two separate (at the resolution of this particular NS run), whereas the relative volume reverses to 16 : 1 at the energy level where the known lowest barrier path actually connects the two states [196–198].

It is through these landscape charts that the challenges of thermodynamic sampling can be best understood. If there are not enough live points, even though a few may make it into the global minimum basin where that splits off, there is a danger of that population of walkers “dying out” much before the global minimum is reached. This extinction can happen even while the different basins are still connected in principle, but the volume of states that connect them is so small (or rather, the path connecting them is so narrow) that in practice there is



a | Potential energy landscape of LJ_{38} , showing the energy landscape chart (see text) obtained by NS, and the disconnectivity graph calculated using discrete path sampling, showing the 200 lowest energy minima [193]. The global minimum basin is shown in red, and the relative phase space volumes of the basins at the separation point and the lowest known connecting path are indicated by the ratios.



b | Constant pressure heat capacity curve (red) peaks are used to locate transitions between the gas (top), liquid (middle) and solid (bottom) phases. Sampled configurations and weights enable calculations of temperature dependent observables such as the radial distribution function (right panel) [77] (note that starred variables correspond to quantities in reduced units).

Figure 5 | **Illustrations of materials science applications.**

no communication between the basins. This phenomenon is generally referred to as *broken ergodicity* [188].

Note that while small LJ clusters serve as illustration and allow one to develop, benchmark and parameterize various details of sampling algorithms, NS is not the most efficient method for calculating observables in this case. Specialized “bottom-up”

algorithms such as Basin-Sampling [188] are significantly more efficient, because they start from a low lying minimum (not necessarily the global minimum) and build a database of all the neighbouring minima by a series of perturbations of the particles and subsequent relaxation. The performance of NS for LJ clusters can be somewhat improved if either (i) the sampler is allowed the use of the pre-generated database of local minima, as in the superposition-enhanced NS (SENS) [199], or (ii) a large number of independent NS runs each with a single live point are suitably combined, as in Nested Basin-Sampling [71], but neither of these enhancements makes NS competitive for studying small and moderate sized particle clusters.

There is one case for which NS appears to be an effective tool for studying even the smallest clusters, and this is the determination of thermodynamically favourable *transformation paths* between different states at moderately high temperatures where harmonic transition state theory is no longer applicable. [200]

Phase diagrams of materials

It is in the study of condensed phase systems that NS really comes into its own. There are two reasons for this: one is that it is generally difficult to make efficient sampling moves due to the geometric constraints, the other is that as the system size grows the phase transitions become sharper such that for a hundred or more particles, thermal methods become essentially ineffective.

At high energies and moderate pressures entropy wins and all materials are gases, but as the energy decreases they typically condense into a liquid, then freeze into a solid, and sometimes even undergo solid-phase structural transitions. A specific heat curve is shown as an example in fig. 5b for the periodic Lennard-Jones system. In the thermodynamic limit, the discontinuity of the potential energy across the phase transition would result in a divergence of the heat capacity, but in a numerical simulation they are broadened by finite size effects and appear as sharp peaks. Performing the sampling at a range of pressure values, the loci of the peaks define the boundaries between stability regions of different phases in the pressure-temperature phase diagram. Notice the structural similarity of the solid phase, a closed packed face centred cubic (fcc) crystal, to the narrow global minimum of the LJ_{38} cluster. As the cluster size grows, the repeated occurrence of these fcc-like clusters hint at the ground state of the infinite crystal. There is no such periodic analogue for the broad icosahedral basin of the clusters because the five-fold symmetry is incompatible with periodicity (periodicity is further discussed in Supplementary Information F).

NS has been successful in characterizing the behaviour of a wide range of materials. These include model systems such as the Potts model [201, 202], hard spheres [203], Lennard-Jones [77, 204] and the Jagla potential [205] as well as more chemically realistic potentials for aluminium and the shape memory alloy NiTi [44], as well as lithium [206], and iron [207]. Recently, machine-learning interatomic potentials [208] are now being applied [209] to increase the predictive power of calculated phase diagrams, leveraging the increased accuracy of the potentials. A detailed overview of materials applications

can be found in ref. [210].

5 Reproducibility and data deposition

We recommend a set of minimum considerations and reporting standards for NS computations, shown in box 5. You should state clearly the number of live points and stopping conditions. In addition, you should report implementation specific settings for example the number of repeats if using slice sampling, or enlargement factor if using ellipsoidal sampling. If you are using a public software package, report the version number. Since NS is an MC algorithm, for reproducibility we suggest fixing a random seed so that identical results can be replicated.

Since NS computes an integral, make sure that the integrand is explained clearly, including the choices of likelihood and prior, and, for example, any overall constant factors are sometimes omitted from the likelihood. To help achieve these goals, consider publishing your computer code along with your research. Similarly, consider depositing the NS output files publicly. This would permit further scrutiny and re-use of the NS results. The output data should be accompanied by sufficient metadata (e.g., column labels; see ref. [211]).

Although the ultimate result might be for example, a ratio of NS results (a Bayes factor), we recommend reporting the results of all individual NS calculations. To do so, we suggest the triplet of $\log Z$, the estimated uncertainty, and the KL divergence, H . The first two are most relevant for inference, whereas the KL divergence indicates the numerical challenge, as it impacts runtime (eq. (19)) and uncertainty (eq. (21)). You may also wish to report the effective dimensionality or an Occam factor [212].

The NS error estimates are usually reliable, such that if we wish to reproduce NS computations, we should find agreement within uncertainties with the original calculation. As discussed previously, though, it is $\log Z$ rather than Z alone that is distributed with a roughly symmetric Gaussian error.

6 Limitations and optimizations

6.1 Limitations

Although NS is broadly applicable, there are several potential limitations. The first limitation relates to the prior: we sample from the constrained prior and thus require a proper prior. For many NS implementations, this proper prior must be transformed from the unit hypercube. Normalizing flows were recently proposed [213] for cases where this is inconvenient or impossible analytically, including using the posterior from an NS run as a prior.

There are, furthermore, limitations related to the likelihood. Integration using NS requires a non-negative integrand (though the compression itself makes no such restriction). Whilst this condition is always fulfilled in statistical applications, it could be violated when NS is used as a general purpose integrator. We wrote eq. (6) assuming that $L \geq 0$ whereas eq. (8) makes no restriction. NS, however, compresses upwards in likelihood,

Box 5 | Check-list for reliability and reproducibility

1. Pick an appropriate technique for sampling from the **constrained prior**; some choices perform more reliably and efficiently in high-dimensions
2. Pick appropriate settings and tailor them to your goals and problem. Relaxed exploration settings (for example fewer steps [66]) may be adequate if you are only concerned about parameter inference.
3. Report software version numbers and settings
4. Describe **priors** and **likelihood** in adequate detail
5. Ideally, publish the code used for the computation, permitting the calculation to be replicated and scrutinized
6. Perform cross-checks on NS run [89], as implemented in e.g., anesthetic. If practical, consider simulation-based calibration to check results
7. Report triplet of $\log Z$, the associated uncertainty, and H for each NS computation
8. Consider publishing the NS output, allowing further re-use and checks of the NS run

such that positive and negative likelihood regions of the integral would be treated differently, with the latter explored at inadequate resolution. NS, furthermore, requires a tractable likelihood. For cases in which the likelihood is intractable, ref. [99] proposed a likelihood-free NS in the context of systems biology, assuming that an unbiased estimator of the likelihood was available.

Lastly, plateaus in the likelihood function spoil the estimates of the compression [2, 214–216]. That is, sets A with non-negligible prior mass, $\mu(A) > 0$, and constant likelihood, which means $L(\Theta) \equiv c$ for all $\Theta \in A$. We may modify the likelihood function to remove plateaus by adding a negligible unique iid tie-breaking random draw to the likelihood or promoting that draw to a parameter and increasing the dimension of the problem. We may alternatively modify the algorithm itself such that it sums plateaus correctly but reduces to the ordinary NS algorithm in their absence. In the presence of plateaus, the minimum likelihood among the live points, $\min L$, may be shared by several live points.

For example, we could modify NS by removing all q points in the plateau and then replacing them all [217]. In this case, the compression factors follow Beta($n_{\text{live}} + 1 - q, q$). This may be applied retrospectively to any NS runs, though suffers from increased uncertainty as the number of live points is temporarily reduced to as few as $n_{\text{live}} - q$. Alternatively, we could first increase the live points by sampling subject to $L > L^*$ until $n_{\text{live}} - 1$ lie at $L > \min L$. Note that $\min L$ may decrease as we add live points at $\min L > L > L^*$. We then remove q points in the outermost contour. The compression factor follows Beta(n_{live}, q). This cannot be applied retrospectively but the uncertainties are

reduced (at the cost of greater run-time) as the number of live points is temporarily increased beyond n_{live} . This may in fact be seen as a dynamic version of the first modification and we could increase the number of live points according to a different criterion. Lastly, we could split the integral into an integral over plateaus and an integral without plateaus, and perform only the latter with NS [216].

In addition, there are computational limitations. In ordinary NS running for longer doesn't reduce uncertainties or increase the number of **posterior** samples. If one is only interested in using NS for optimization, this isn't a drawback. In any case, as discussed in **Dynamic nested sampling**, dynamic NS overcomes this problem by rewinding and resuming an NS run. Besides that, NS requires $\mathcal{O}(D^2)$ iterations (eq. (19)), the memory required to store the co-ordinates of every dead point scales as $\mathcal{O}(D^3)$ (with potentially nasty scaling factors for clustering in particular exploration strategies). This means that NS implementations that store every dead point become memory bound at $\mathcal{O}(500)$ dimensions.

We stress, furthermore, that users should be aware of ways in which NS may fail:

1. NS may fail to successfully draw independent samples from the **constrained prior**. Consequently NS results including error estimates may be faulty and anticipated properties of NS, such as convergence, won't hold. In **Results**, we discussed cross-checks on NS results, such as ref. [89], that may identify this issue.
2. Owing to item 1, NS may miss **modes**. **multi-modal** problems pose challenges in Bayesian computation, particularly in **MCMC**, where the chain must make a sequence of unlikely steps between **modes**. Unfortunately, it is in general impossible to know whether all **modes** have been found. Broadly speaking, as NS does not depend on slow transitions between **modes**, it is well-suited to **multi-modal** problems. Once a **mode** is established, furthermore, it won't be abandoned until the **likelihood** threshold exceeds that of the points that were in the mode.
3. NS may sample inefficiently from the **constrained prior**. As discussed in **Exploration strategies**, this may occur in high-dimensional problems with rejection sampling strategies.

Our check-list in box 5 includes checks of NS failures.

6.2 Optimizations

There are several ways NS runs may be optimized to make best use of computing resources (see also **Parallelization**). First, one may utilize fast and slow parameters by breaking the **likelihood** function into fast and slow operations that involve subsets of the D parameters. The parameters associated with fast and slow operations are referred to as fast and slow parameters, respectively; for example, if the **likelihood** function may be written as,

$$L(x, y, z, \dots) = \text{slow}(x) \times \text{fast}(y, z, \dots), \quad (31)$$

then y, z, \dots are fast parameters and x is a slow parameter. When selecting a new point, to minimize runtime we should where possible change the fast parameters and keep the slow parameters constant, allowing caching of the slow operation [218]. This is natural in the slice sampling exploration strategy discussed in **Exploration strategies**, as we may pick slices along fast directions [66]. If D_s parameters are slow, we require D_s rather than D slow operations per iteration. Similarly, it may be exploited in modified Metropolis algorithms [22, 219] in which the proposals change blocks of parameters at a time.

In addition, one may repartition the **posterior**. NS must compress exponentially through the entire prior volume, which could be slow for a diffuse prior with substantial compression to the **posterior**. This in practice limits NS to $\mathcal{O}(100s - 1000s)$ of dimensions. This prohibits fitting complex hierarchical Bayesian models and deep neural networks, or reaching HMC like dimensionalities of $\mathcal{O}(10^6)$. In some cases, it may be possible to use a narrower prior and correct the **evidence** estimates from NS post-hoc.

This issue may occur in problems in which the prior is unrepresentative [220], i.e., where the observed data lies in the tail of the prior predictive distribution. This may be mitigated by Bayesian automatic prior repartitioning [221] in which one redefines the prior and **likelihood** while leaving their product unchanged. This allows one to keep the **evidence** integral in eq. (1) unchanged, but reduce the prior to **posterior** compression in eq. (16). For example, consider a Gaussian **likelihood**, $\mathcal{N}(x, \sigma^2)$, for a parameter $x \sim \mathcal{U}(-L/2, L/2)$ for $\sigma \ll L$. The compression would be approximately $H \approx \log(\sigma/L)$. Repartitioning the prior and **likelihood** by swapping them, we obtain $H \approx 0$.

The possibility of improving the robustness and efficiency of NS by exploiting the intrinsic degeneracy between the 'effective' **likelihood** and prior in the formulation of Bayesian inference problems was also discussed in ref. [123], and posterior repartitioning can further be viewed as the vanilla case (when the importance weight function equals to 1) of nested importance sampling proposed in ref. [18].

Lastly, rather than performing new NS runs, one may reuse runs for similar **likelihoods** and **priors**. If the prior or **likelihood** are modified, it may be possible to reweight the **posterior** weights and **evidence** integral [40],

$$P'_i = P_i \times \frac{L'_i \pi'_i}{Z'} \frac{Z}{L_i \pi_i} \quad (32)$$

$$Z' = Z \times \sum P_i \times \frac{L'_i \pi'_i}{L_i \pi_i} \quad (33)$$

where $P_i = w_i L_i$. This re-weighting may be interpreted as a **pseudo-importance sampling** in which the original estimated **posterior** plays the role of the kernel. This is particularly useful for investigating prior sensitivity in the context of Bayesian inference. The effective sample size of draws from the new **posterior** may be used to judge the reliability of this procedure; if the new and original **posterior** distributions differ substantially, a fresh NS run needs to be performed.

There may, furthermore, be cases in which we wish to in-

investigate several similar **likelihood** functions at once; ref. [59] presents a collaborative version of NS that operates on more than one **likelihood** function at once and in which parts of the **likelihood** evaluation may be recycled.

7 Outlook

The **MCMC** computational revolution of the 1990s solved the problem of computing shapes. Skilling’s remarkable NS algorithm solved the outstanding problem of computing magnitudes at the same time. Although it is naturally expressed in the language of Bayesian inference, NS is a powerful and general-purpose integration algorithm. For that reason, we expect NS to remain relevant long into the future. We are now in the midst of a revolution in data science, and high-dimensional spaces and integration are more important than ever. NS rises to the challenge.

As we reviewed, there have been many theoretical developments in understanding NS, including its convergence, errors, diagnostics and techniques for sampling from the **constrained prior**. We expect them to continue, especially as connections to other statistical methods and machine learning are explored. Indeed, theoretical analysis of NS combined with constrained **MCMC** exploration may be helped by the connections to SMC discussed in box 4, and we anticipate further developments in understanding uncertainties in NS, especially for parameter inference and in the case of a dynamic number of live points. Lastly, NS may be considered a meta-algorithm, as it doesn’t specify an algorithm for sampling from the **constrained prior**. We are already seeing that this opening allows NS to dovetail with developments in machine learning, such as normalising flows that are beginning to be used to sample from the **constrained prior**.

The successes and breadth of applications of NS stem from the fact that it is fundamentally simple. Whilst in the future NS may be understood and improved in the context of more sophisticated computational methods, NS will retain advantage and appeal due to its simplicity.

Acknowledgements

We thank John for his wonderful algorithm. The success of nested sampling may be that simple beats clever; but the beauty of nested sampling is that it is both simple and clever.

We thank Kyle Barbary for discussions. AF was supported by an NSFC Research Fund for International Young Scientists grant 11950410509. LBP acknowledges support from the EPSRC through an Early Career Fellowship (EP/T000163/1). MH acknowledges support from the Carl Zeiss Foundation. NB was funded by the U. S. Naval Research Laboratory’s base 6.1 research program, and CPU time from the U. S. DoD’s HPCMPO at the AFRL and ARL DSRCs. MP acknowledges support from the STFC (ST/V001213/1 and ST/V005707/1). WH was supported by a Royal Society University Research Fellowship.

Author Contributions

Introduction (A. F., D. S., L. S., P. W.); Experimentation (J. B., E. H., J. S. S.); Results (A. F.); Applications (G. A., N. B., X. C., G. C., F. F., M. G., W. H., M. H., A. L., D. P., L. B. P., M. P., J. V., D. W., D. Y.); Reproducibility and data deposition (A. F.); Limitations and optimizations (A. F.); Outlook (A. F.); Overview of the Primer (A. F.).

Competing Interests

There are no competing interests to declare.

Supplementary Information

A Glossary

Bayesian model comparison A method for comparing models based on computing the change in their relative plausibility in light of data using Bayes’ theorem. Pages 2, 12, 14, 20.

bootstrap Techniques that estimate statistical variation by repeatedly drawing samples from the true dataset with replacement. Pages 23.

bulk Region with size of order e^{-H} that contains the overwhelming majority of the posterior mass. Closely related to typical sets. Usually won’t lie near the mode of the posterior, especially in high-dimensions [222, 223]. Pages 4, 5, 7, 8, 10, 16.

constrained prior The **prior** for the parameters restricted to the region in which the **likelihood** exceeds a threshold. Pages 1, 3, 5–11, 16, 18–20, 22, 23.

curse of dimensionality The phenomenon that the difficulty of a problem often increases dramatically with dimension [224]. Pages 2, 3, 5, 9, 10.

evidence A factor in Bayes’ theorem that may be written as an integral and that plays an important role in **Bayesian model comparison**. In NS this is the multi-dimensional integral we wish to compute. Pages 2–14, 19, 22.

indicator function A function $\mathbb{1}(\cdot)$ that takes the value 1 if the condition \cdot holds and 0 otherwise. Pages 8.

insertion indexes The indexes at which the elements of a list must be inserted into an ordered list to maintain ordering. Pages 11.

integrand A function that is being integrated, for example, $f(x)$ is the integrand in $\int f(x)dx$. Pages 1, 2, 18, 21, 22, 24.

iso-likelihood contour The set of points for which the likelihood is equal to a particular constant; in two-dimensions, this set forms a contour line. Pages 3, 7, 9, 23.

Kullback-Leibler divergence A measure of difference between two distributions that may be interpreted as the information gained by switching from one to the other. See ref. [222, 225] for pedagogical discussions. Pages 5.

likelihood The probability of the observed data as a function of a model's parameters. Pages 2–5, 8–12, 15, 16, 18–20, 22, 24.

Markov chain A sequence of random states for which the probability of a state depends only on the previous state. Pages 21, 22.

Markov chain Monte Carlo (MCMC) A class of algorithms for drawing a correlated sequence of samples from a distribution using the states of a **Markov chain**. Pages 1, 4, 8, 9, 12, 13, 19, 20, 22.

measure A probability measure is a function assigning probabilities to events. See ref. [225] for a pedagogical discussion. Pages 6.

microcanonical ensemble Assigns equal probability to states Θ with $E(\Theta) = \epsilon$ and zero probability if $E(\Theta) \neq \epsilon$ such that the energy level ϵ rather than the inverse temperature β characterizes a thermodynamic state. Pages 6.

microstate The state of all degrees of freedom in a physical system, for example, the microstate of a multi-particle system includes the positions and momenta of all particles. Pages 6.

mode A peak in a probability distribution. Pages 7, 9, 12, 13, 16, 19, 21–23.

multi-modal Problems in which the **integrand** contains more than one **mode**. In inference problems, **modes** may correspond to distinct ways in which the model could predict the observed data. Pages 1, 2, 7–9, 13, 19, 21–23.

overloaded A function with a definition that depends on the type of its argument. In NS the likelihood L is an overloaded function since we consider separate functions $L(\theta)$ and $L(X)$. Pages 6.

parameter domain The set of a priori possible parameters, usually the reals \mathbb{R}^n or a subset thereof. Pages 6.

partition function Normalizing constant $Z(\beta) = \int e^{-\beta E(\Theta)} d\Theta$ that fully characterizes a physical system, because many important thermodynamic variables can be derived from it. Pages 4, 6, 10, 15, 16, 22, 24.

posterior A probability density conditioned on the observed data found by updating a **prior** using Bayes' theorem. In NS it describes the shape of our **integrand**. Pages 2, 4–8, 10–13, 15, 16, 18, 19, 22, 23.

prior A probability density that isn't yet conditioned on observed data. In NS it is a measure in our integral. Pages 2, 3, 5, 6, 9, 15, 18–21, 24.

pseudo-importance sampling Algorithms in which an importance sampling density is defined a posteriori. Pages 19.

push-forward measure The push-forward measure (or image measure) is the distribution of a random variable under a probability measure. Pages 6.

Simulation-Based Calibration Techniques that use simulations from the model to check the correctness of Bayesian computation [226–228]. Pages 12.

super-level-sets A λ -super-level-set of any function contains all points for which the function value exceeds λ . Pages 6.

survival function A function $F(x)$ associated with a distribution that returns the probability of obtaining a sample greater than x . Pages 6.

tractable An adjective used to describe problems that are feasible to solve under computational, monetary or time constraints. Pages 18.

transition kernel A function that describes the likely steps of a **Markov chain**. Pages 8.

uni-modal Problems in which the **integrand** contains only one **mode** (cf. **multi-modal**). Pages 2, 4, 8, 15, 23.

B Estimators for compression factor

Besides the estimator $\langle \log t \rangle = -n_{\text{iter}}/n_{\text{live}}$ in eq. (12), we may instead consider,

$$\langle t \rangle = \frac{n_{\text{live}}}{n_{\text{live}} + 1}. \quad (34)$$

Similarly, ref. [24] suggests the estimator

$$\hat{t} = 1 - \frac{1}{n_{\text{live}}}. \quad (35)$$

In any case, the relative differences in $\log t$ are of order $\mathcal{O}(1/n_{\text{live}})$.

C Historical background

In the 1950s, when computational algorithms were beginning to be developed, the Bayesian revolution lay decades in the future and inquiry was focused at best on **posterior** distributions. The first practical computer codes for this were **MCMC** [229]; first, the Metropolis algorithm [230] and later Metropolis-Hastings [231]. They build a **Markov chain** of correlated points drawn from an unnormalized **posterior**. The Metropolis algorithm accepted a transition $\Theta \rightarrow \Theta'$ only if

$$L(\Theta') > uL(\Theta) \quad (36)$$

where $u \sim \mathcal{U}(0, 1)$. This conditional acceptance of random trial transitions satisfies detailed balance and converges towards an equilibrium. This, however, only gives the shape of the distribution, not the normalizing constant, Z .

Later, in the 1970s, a connection between the **likelihood** and the temperature-dependent energy,

$$L(\Theta) \propto e^{-\beta E(\Theta)} \quad (37)$$

where $\beta = 1/T$, suggested generalizing the **likelihood** to L^β . Upon which, the **evidence** generalized to

$$Z(\beta) = \int L^\beta(\Theta) \pi(\Theta) d\Theta, \quad (38)$$

with $Z = Z(1)$ being the desired **evidence** and $Z(0)$ being 1. Whereas the **posterior** generalized to

$$P^\beta(\Theta) = \frac{L^\beta(\Theta) \pi(\Theta)}{Z(\beta)}, \quad (39)$$

with $P^0(\Theta)$ the prior and $P^1(\Theta)$ the desired **posterior**. These values are connected by the computable differential

$$\frac{d \log Z}{d\beta} = \int \log L(\Theta) P^\beta(\Theta) d\Theta = \langle \log L \rangle_\beta \quad (40)$$

which is simply the log-likelihood averaged over the distribution $P^\beta(\Theta)$ at inverse temperature β . Add that up as the distribution is slowly cooled from $\beta = 0$ to 1, i.e., from the prior to the **posterior**, and the **evidence** arrives as [39, 232, 233]

$$\log Z = \int_0^1 \langle \log L \rangle_\beta d\beta. \quad (41)$$

All that seems to be needed to compute the **evidence** integral is to fix the cooling schedule to track β suitably slowly (hence annealing) from 0 to 1 such that the **posterior** distribution $P^\beta(\Theta)$, represented by a collection of samples drawn from it, changes slowly. This annealing schedule may, however, be problematic. Maintaining the thermodynamics analogy, which is further expanded in box 3, the trouble occurs with changes of state as the system cools. The **evidence** plays the role of the thermodynamic **partition function** and at each inverse temperature the samples are expected to coalesce closely around the local

maximum of the **posterior** distribution $P^\beta(\log X)$. This implies that

$$\frac{d \log L(X)}{d \log X} = -\frac{1}{\beta} \quad (42)$$

In smooth cross-over transitions, volume and energy shrink in tandem as the system cools from $\beta = 0$ (vertical slope) to $\beta = 1$ (diagonal slope) in fig. 6-a, which is what happens in stable physical systems.

If there is a second-order phase transition, however, a tiny cooling around the critical β rapidly compresses the ensemble exponentially all the way from disorder to order. We see this in the iso-thermal region in fig. 6-b. The sensitivity of the system to temperature means that it becomes numerically impossible to use β as the control parameter. Lastly, if there's a first-order phase transition, two phases coexist at the same temperature or in a temperature interval, e.g., steam and water in fig. 6c. Starting from the steam phase, to reach the water phase requires an abrupt first-order phase transition. It thus becomes impossible even in principle to use inverse temperature to change smoothly from the steam phase to the water phase.

The first-order phase transition was connected to the fact that there were multiple solutions to eq. (42) for some β . This implies that the **integrand** in the **evidence** integral expressed as

$$Z(\beta) = \int L^\beta(X) X d \log X, \quad (43)$$

was **multi-modal** for at least some temperatures. Physically, this may occur in a problem with distinct solutions at different **likelihood** levels. In our analogy, we may interpret the **modes** as phases. Traditional annealing methods struggle to move particles between them during a phase transition but with NS, the system's evolution can be controlled through compression instead.

D Step sampling

For detailed balance, the transition kernel in a step sampler must satisfy

$$\int P(\Theta_f | \Theta) \pi^*(\Theta) d\Theta = \pi^*(\Theta_f). \quad (44)$$

We are thus drawing a new live point Θ_f from

$$P(\Theta_f | \{\Theta_{\text{live}}\}) = \frac{1}{n_{\text{live}}} \sum_{i=1}^{n_{\text{live}}} P(\Theta_f | \Theta_{\text{live}}^i) \quad (45)$$

where $\{\Theta_{\text{live}}\}$ denotes the set of live points and we average over the possible choices of initial live point indexed by i , Θ_{live}^i . For a draw from eq. (45) to be approximately equivalent to an independent draw from the **constrained prior**, we thus require

$$P(\Theta_f | \{\Theta_{\text{live}}\}) \approx \pi^*(\Theta_f). \quad (46)$$

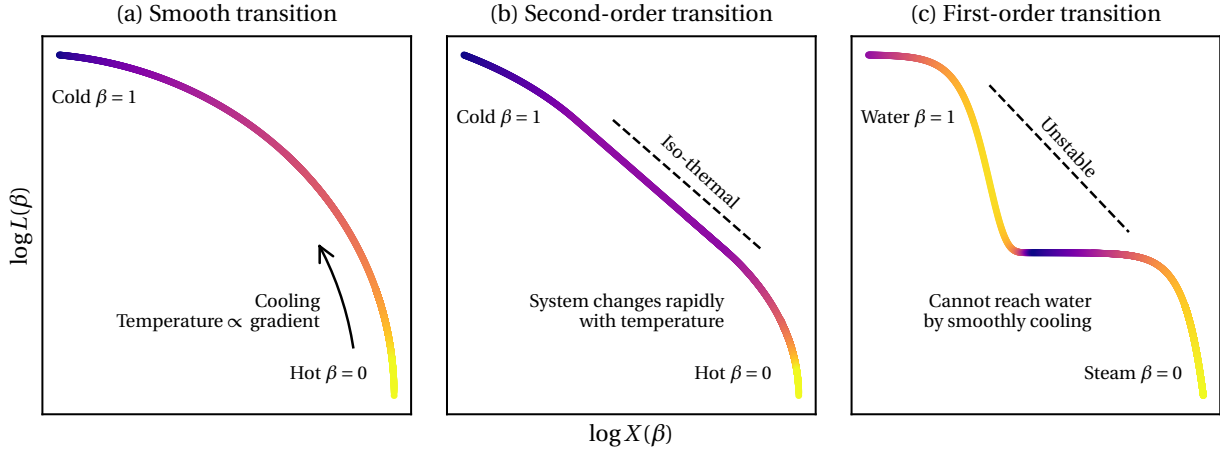


Figure 6 | **Phase transitions in simulated annealing.** We show $\log L(\beta)$ and $\log X(\beta)$ as we cool from $\beta = 0$ to 1 in an ordinary application, a second-order transition and a first-order transition. The colour scale indicates temperature.

We anticipate that this might approximately hold because the live points should be independent draws from the **constrained prior** such that eq. (45) is a Monte Carlo estimate of the **constrained prior** through eq. (44). In any case, to satisfy eq. (46), we may attempt to satisfy a condition of decorrelation between the chosen initial point and the final point,

$$P(\Theta_f | \Theta_{\text{live}}^i) \approx \pi^*(\Theta_f). \quad (47)$$

In **uni-modal** problems, this can be achieved by performing a sufficient number of steps. In **multi-modal** problems, it may be impossible as walks between **modes** are all but impossible. Since we only require eq. (46), however, correlation between the **modes** in which the initial and final points lie may be acceptable. The number of steps required to satisfy eq. (46) may decrease when n_{live} is increased, as Monte Carlo errors in the approximation in eq. (46) shrink as $1/\sqrt{n}$.

Lastly, these considerations show that starting from the last dead point instead of a randomly chosen live point isn't sensible [201], even though it avoids correlation with the existing live points. In **uni-modal** problems, we could require many steps to avoid biasing the draw towards the lowest likelihood levels near the last dead point. In **multi-modal** problems, it would lead to catastrophe, as the fractions of live points in each mode could not change, since a point would always be replaced by a point in the same mode. Consequently, modes would never die and NS would become stuck at the likelihood threshold of the lowest lying mode.

E Uncertainty in the posterior

There is a further source of uncertainty in **posterior** expectations. From the dead points Θ_i , with importance weights P_i , a

posterior mean for some quantity of interest $f : \Omega \rightarrow \mathbb{R}$ may be estimated by [18]

$$\langle f(\Theta) \rangle = \int f(\Theta) P(\Theta) d\Theta = \int f(X) P(X) dX \quad (48)$$

$$\approx \sum_i P_i f(X_i) \approx \sum_i P_i f(\Theta_i), \quad (49)$$

where $f(X)$ is the mean of f over an **iso-likelihood** contour,

$$f(X) = \frac{\int f(\Theta) \delta(L(\theta) - L(X)) \pi(\Theta) d\Theta}{\int \delta(L(\theta) - L(X)) \pi(\Theta) d\Theta}. \quad (50)$$

There are thus two sources of uncertainty in eq. (49): the statistical uncertainty in the estimates of the importance weights P_i , as before, and the fact that eq. (49) replaces the mean around the contour by a single draw from around the contour, $f(X_i) \approx f(\theta_i)$. As before, the first source can be accounted for by simulating the compression factors through eq. (10). Whilst both sources are reduced by increasing the number of live points, the second source usually dominates.

Ref. [234] demonstrated that it can be assessed by decomposing an NS run with n_{live} live points into n_{live} NS runs with one live point each called *threads*. The threads can be recombined in many different combinations using resampling methods such as **bootstrap** to generate simulated runs with the same number of live points as the original run. The variance in **posterior** inferences in these simulated runs captures both sources of uncertainty (see UltraNest [92] for an NS implementation of these simulations).

F Periodic boundary conditions, degrees of freedom and Monte Carlo moves

In any condensed phase, a computationally tractable small finite system would be dominated by surface effects, and eliminating these requires a periodic supercell description. The typical volume per atom changes by orders of magnitude between the gas and condensed phases, and must be allowed to vary for NS to sample the relevant structures. This can be achieved by fixing the pressure P , replacing the potential energy U in the sampling algorithm with the enthalpy $U + PV$, where V is the system volume, and sampling the cell degrees of freedom [44, 77]. The corresponding [partition function](#) is

$$Z(N, P, \beta) = Z_p \beta^P \int d\mathbf{h}_0 \delta(\det \mathbf{h}_0 - 1) \times \int_0^\infty dV V^N \int_{(0,1)^{3N}} d\mathbf{s} e^{-\beta(U(V^{1/3}\mathbf{h}_0\mathbf{s}) + PV)},$$

where \mathbf{h}_0 is the reduced cell shape matrix (with unit determinant) and \mathbf{s} are the scaled (often called fractional) atomic positions. When generating uniformly distributed configurations V and \mathbf{h}_0 must be sampled in addition to the scaled positions \mathbf{s} . The expectation value of $V(\beta)$ must be calculated using the same weights as the [partition function](#). To compute an entire pressure-temperature phase diagram the NS run is repeated for each pressure value. After a completed NS run, the expectation value of any observable can be calculated as a function of temperature, e.g. the radial distribution function plotted in fig. 5b, which can be compared to the results of scattering experiments and used to identify each equilibrium phase.

Condensed phase atomic position Monte Carlo moves with a reasonable acceptance probability are also challenging to generate. Single atom moves become inefficient if the resulting energy change cannot be computed in $\mathcal{O}(1)$ time, and the probability of accepting naive collective moves decreases as $1/\sqrt{N}$. A more efficient alternative is Galilean Monte Carlo [44, 69, 70], where a random direction in $3N$ -dimensional configuration space is proposed, the entire system is propagated for a fixed number of steps along straight lines that reflect specularly from the allowed $U(q) < U^*$ boundary. Another is total-energy Hamiltonian Monte Carlo [44, 235], where the kinetic energy is added to U , so the NS iteration constrains *total* energy, and atomic moves are proposed by carrying out short time constant energy molecular dynamics trajectories that are accepted or rejected in their entirety. In both cases, Monte Carlo moves that propose to change the volume and the cell shape are also used.

In condensed phase systems with multiple types of particles the probability for particles to switch places is low, especially in the solid phase, and explicit particle swap proposals are required to ensure mixing. Further, the *composition* of different phases may change discontinuously across phase transitions. The semi-grand-canonical ensemble, with fixed total number of particles but variable composition, describes this situation.

Monte Carlo proposals include changes to particle type, and the energy is augmented by $\sum_i \mu_i N_i$, where i runs over the types, μ_i is a specified chemical potential, and N_i is the number of particles of that type i . Like $V(\beta)$ above, $N_i(\beta)$ is an output of the simulation.

G Example

Let us demonstrate NS using standard NS software on a toy problem using Python. The libraries used in this example may be installed using pip, for example,

```
$ pip install numpy scipy matplotlib pypolychord anesthetic
```

To record our software version numbers,

```
>>> import platform
>>> print(platform.python_version())
3.8.10
>>> from importlib.metadata import version
>>> print(version('pypolychord'))
1.20.1
>>> print(version('anesthetic'))
1.3.6
```

Let us attempt to compute the two-dimensional Gaussian integral,

$$Z = \frac{1}{(2a)^2} \int_{-a}^a \int_{-a}^a e^{-x^2 - y^2} dx dy. \quad (51)$$

We may write the [integrand](#) in the form eq. (1) by defining the [likelihood](#)

$$L(x, y) = e^{-x^2 - y^2} \quad (52)$$

and the [prior](#),

$$\pi(x, y) = \frac{1}{(2a)^2} \quad (53)$$

for $-a < x < a$ and $-a < y < a$ and vanishing elsewhere. So long as a is greater than about 1, $Z \approx \pi/(2a^2)$ and $H \approx -1 - \log \pi - \log 2a^2$. For concreteness, we take $a = 5$ such that $\log Z \approx -3.46$ and $H \approx 2.46$

To run NS on this problem we first implement the logarithm of the [likelihood](#) in eq. (52),

```
>>> def loglike(theta):
...     return -(theta**2).sum(), [] # [] is anything else to be
...     saved
```

and a transformation of the unit hypercube representing the prior in eq. (53),

```
>>> def prior(unit_hypercube):
...     a = 5.
...     return 2. * a * unit_hypercube - a
```

Here we map from the unit hypercube $\mathcal{U}(0, 1)^2$ to our $\mathcal{U}(-a, a)^2$ prior.

We use the `PolyChord` [66, 133] NS implementation — this uses slice sampling to draw replacement live points from the constrained prior — see table 2 for alternative NS software. We import it by

```
>>> from pypolychord import run_polychord
>>> from pypolychord.settings import PolyChordSettings
```

We specify that our integral is two-dimensional ($\text{ndim} = 2$), that there are no derived quantities to save to disk ($\text{nderived} = 0$), and our NS settings. We wish to use $n_{\text{live}} = 1000$ ($\text{settings.nlive} = 1000$), perform 5 slice sampling steps ($\text{settings.num_repeats} = 5$), and fix our random seed for reproducibility ($\text{settings.seed} = 67$).

```
>>> ndim = 2
>>> nderived = 0
>>> settings = PolyChordSettings(ndim, nderived)
>>> settings.nlive = 1000 # this is n_live
>>> settings.num_repeats = 5
>>> settings.seed = 67
```

Finally, we run the NS algorithm,

```
>>> run_polychord(loglike, ndim, nderived, settings, prior)
```

This by default writes our results to files named `chains/test*` and information to the screen. We may further inspect the results using, for example, `anesthetic` [78]. First, we load the data from disk and for reproducibility fix the random seed using `numpy` [236],

```
>>> from anesthetic import NestedSamples
>>> import numpy as np

>>> samples = NestedSamples(root='chains/test')
>>> np.random.seed(71)
```

We may compute and print the triplet $\log Z$, its uncertainty, and the KL divergence,

```
>>> H = samples.D()
>>> logZ = samples.logZ()
>>> uncertainty = (H / settings.nlive)**0.5 # this is eq. (21)
>>> print(logZ, uncertainty)
-3.5109005855438395 0.04994705730822035
>>> print(H)
2.494708533750648
```

Thus in this problem NS estimates $\log Z = -3.51 \pm 0.05$ and $H = 2.5$ in agreement with the analytic results. We may compare these estimates to those found from 1000 simulations ($\text{nsamples}=1000$) of the compression factors,

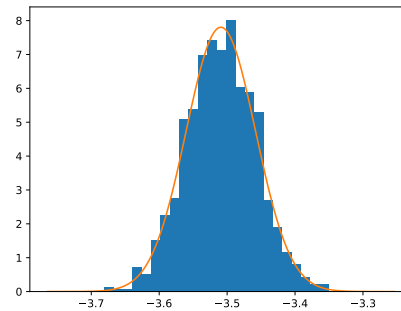
```
>>> draws = samples.logZ(nsamples=1000)
>>> mean = np.mean(draws)
>>> std = np.std(draws)
>>> print(mean, std)
-3.510624723981491 0.05232090274229939
```

showing that the standard NS estimators are reliable in this

case. We may check the assumption that $\log Z$ is approximately Gaussian distributed by histogramming the draws of $\log Z$. Here we use the `matplotlib` [237] histogramming functionality and the `scipy` [238] implementation of the normal distribution,

```
>>> import matplotlib.pyplot as plt
>>> from scipy.stats import norm

>>> plt.hist(draws, bins='auto', density=True)
>>> x = np.linspace(mean - 5. * std, mean + 5. * std, 1000)
>>> plt.plot(x, norm(mean, std).pdf(x))
>>> plt.show()
```



We see that $\log Z$ approximately follows a Gaussian distribution.

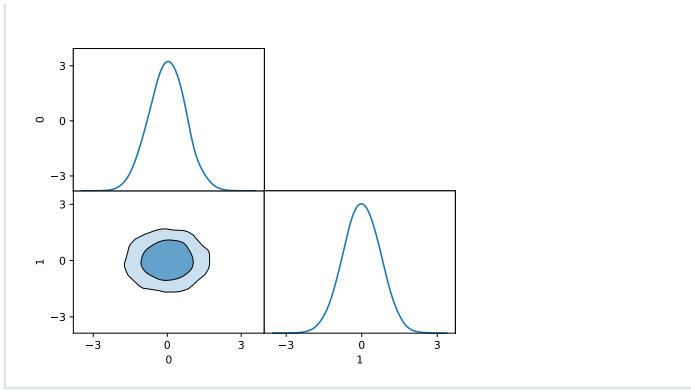
As a cross-check, we can test whether the insertion indexes of new live points are uniformly distributed using `anesthetic`,

```
>>> from anesthetic.utils import insertion_p_value
>>> samples._compute_insertion_indexes()
>>> ks = insertion_p_value(samples.insertion, settings.nlive)
>>> print(ks['p-value'])
0.4965573665241407
```

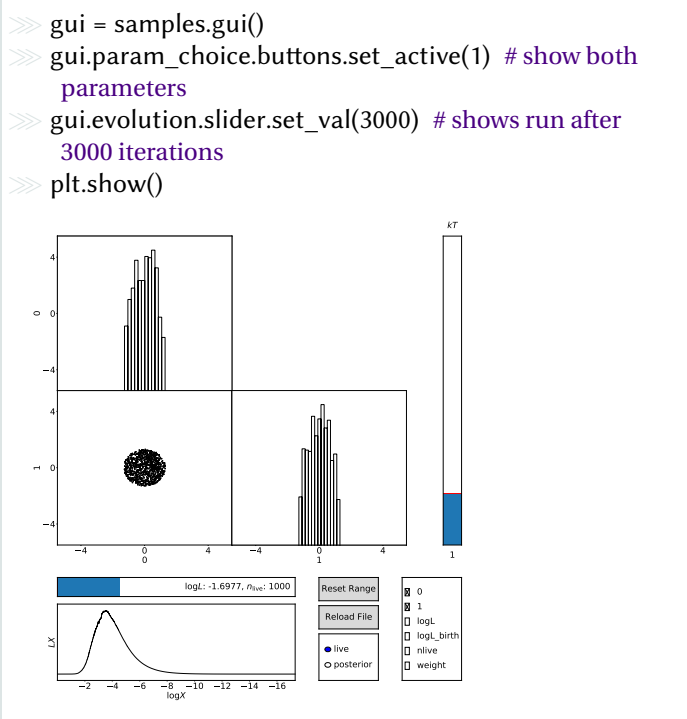
In this case we find p -value of about 0.5, which does not indicate any discrepancy from uniform.

Lastly, we may wish to compute and plot posterior distributions. Here we plot one- and two-dimensional posterior distributions using `anesthetic` and kernel density estimation ('kde'),

```
>>> samples.plot_2d(samples.columns[:ndim], types={'
lower': 'kde', 'diagonal': 'kde'})
>>> plt.show()
```



We may examine the evolution of live points interactively using a graphical user interface (GUI)



This shows, among other things, the distribution of live points at iteration 3000, and we may use the slider to see how it evolved.

References

[1] Skilling, J. Nested Sampling. In Fischer, R., Dose, V., Preuss, R. & von Toussaint, U. (eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2004)*, 735, 395–405, DOI: [10.1063/1.1835238](https://doi.org/10.1063/1.1835238). American Institute of Physics (AIP, New York, 2004).
NS was first presented at MAXENT 2004 and appeared in the subsequent proceedings

[2] Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Analysis* **1**, 833–859, DOI: [10.1214/06-ba127](https://doi.org/10.1214/06-ba127) (2006).
Skilling presented NS and explained it in detail in this landmark publication

[3] Brooks, S., Gelman, A., Jones, G. & Meng, X. L. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods (CRC Press, 2011).

[4] Hogg, D. W. & Foreman-Mackey, D. Data analysis recipes: Using Markov Chain Monte Carlo. *Astrophys. J. Suppl.* **236**, 11, DOI: [10.3847/1538-4365/aab76e](https://doi.org/10.3847/1538-4365/aab76e) (2018). [[arXiv:1710.06068](https://arxiv.org/abs/1710.06068)].

[5] van de Schoot, R. *et al.* Bayesian statistics and modelling. *Nature Reviews Methods Primers* **1**, DOI: [10.1038/s43586-020-00001-2](https://doi.org/10.1038/s43586-020-00001-2) (2021).

[6] D’Agostini, G. *Bayesian Reasoning In Data Analysis: A Critical Introduction* (World Scientific Publishing Company, 2003).

[7] Gregory, P. *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, 2005).

[8] Sivia, D. & Skilling, J. *Data Analysis: A Bayesian Tutorial* (Oxford University Press, 2006).
Sivia & Skilling’s popular textbook includes a chapter on NS

[9] Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.* **49**, 71–104, DOI: [10.1080/00107510802066753](https://doi.org/10.1080/00107510802066753) (2008). [[arXiv:0803.4089](https://arxiv.org/abs/0803.4089)].

[10] von der Linden, W., Dose, V. & von Toussaint, U. *Bayesian Probability Theory: Applications in the Physical Sciences* (Cambridge University Press, 2014).

[11] Bailer-Jones, C. A. L. *Practical Bayesian Inference: A Primer for Physical Scientists* (Cambridge University Press, 2017).

[12] Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795, DOI: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572) (1995).
A classic modern reference for Bayes factors

[13] AbdusSalam, S. S. *et al.* Simple and statistically sound strategies for analysing physical theories. *arXiv e-prints* (2020). [[arXiv:2012.09874](https://arxiv.org/abs/2012.09874)].

[14] Martin, G. M., Frazier, D. T. & Robert, C. P. Computing Bayes: Bayesian Computation from 1763 to the 21st Century. *arXiv e-prints* (2020). [[arXiv:2004.06425](https://arxiv.org/abs/2004.06425)].

[15] Embrechts, P. & Hofert, M. A note on generalized inverses. *Math. Methods Oper. Res.* **77**, 423–432, DOI: [10.1007/s00186-013-0436-7](https://doi.org/10.1007/s00186-013-0436-7) (2013).

- [16] de la Fortelle, A. A study on generalized inverses and increasing functions Part I: generalized inverses. *hal-01255512* (2015).
- [17] Billingsley, P. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics (Wiley, 2013), third edn.
- [18] Chopin, N. & Robert, C. P. Properties of nested sampling. *Biometrika* **97**, 741–755, DOI: [10.1093/biomet/asq021](https://doi.org/10.1093/biomet/asq021) (2010). [arXiv:0801.3887].
- [19] Skilling, J. Nested Sampling’s Convergence. In Goggans, P. M. & Chan, C.-Y. (eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAX-ENT 2009)*, vol. 1193, 277–291, DOI: [10.1063/1.3275625](https://doi.org/10.1063/1.3275625). American Institute of Physics (AIP, New York, 2009).
- [20] Evans, M. Discussion of nested sampling for Bayesian computations by John Skilling. In Bernardo, J. M. *et al.* (eds.) *Bayesian Statistics 8*, vol. 8, 491–524 (Oxford University Press, 2007).
- [21] Salomone, R., South, L. F., Drovandi, C. C. & Kroese, D. P. Unbiased and consistent nested sampling via sequential Monte Carlo. *arXiv e-prints* (2018). [arXiv:1805.03924].
Introduces connections between NS and Sequential Monte Carlo
- [22] Au, S.-K. & Beck, J. L. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics* **16**, 263–277, DOI: [10.1016/s0266-8920\(01\)00019-4](https://doi.org/10.1016/s0266-8920(01)00019-4) (2001).
- [23] Beck, J. L. & Zuev, K. M. Rare-event simulation. In Ghanem, R., Higdon, D. & Owhadi, H. (eds.) *Handbook of Uncertainty Quantification*, 1–26, DOI: [10.1007/978-3-319-11259-6_24-1](https://doi.org/10.1007/978-3-319-11259-6_24-1) (Springer International Publishing, Cham, 2016). [arXiv:1508.05047].
- [24] Walter, C. Point process-based Monte Carlo estimation. *Statistics and Computing* **27**, 219–236, DOI: [10.1007/s11222-015-9617-y](https://doi.org/10.1007/s11222-015-9617-y) (2015). [arXiv:1412.6368].
- [25] Birge, J. R., Chang, C. & Polson, N. G. Split Sampling: Expectations, Normalisation and Rare Events. *arXiv e-prints* (2013). [arXiv:1212.0534].
- [26] Burrows, B. L. A New Approach to Numerical Integration. *IMA Journal of Applied Mathematics* **26**, 151–173, DOI: [10.1093/imamat/26.2.151](https://doi.org/10.1093/imamat/26.2.151) (1980).
- [27] McDonald, I. R. & Singer, K. Machine calculation of thermodynamic properties of a simple fluid at supercritical temperatures. *The Journal of Chemical Physics* **47**, 4766–4772, DOI: [10.1063/1.1701695](https://doi.org/10.1063/1.1701695) (1967).
- [28] Thin, A. *et al.* NEO: Non Equilibrium Sampling on the Orbits of a Deterministic Transform. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W. (eds.) *Advances in Neural Information Processing Systems*, vol. 34, 17060–17071 (Curran Associates, Inc., 2021). [arXiv:2103.10943].
- [29] Rotskoff, G. M. & Vanden-Eijnden, E. Dynamical computation of the density of states and bayes factors using nonequilibrium importance sampling. *Phys. Rev. Lett.* **122**, 150602, DOI: [10.1103/physrevlett.122.150602](https://doi.org/10.1103/physrevlett.122.150602) (2019). [arXiv:1809.11132].
- [30] Polson, N. G. & Scott, J. G. Vertical-likelihood Monte Carlo. *arXiv e-prints* (2015). [arXiv:1409.3601].
- [31] Robert, C. P. & Wraith, D. Computational methods for Bayesian model choice. In *AIP Conference Proceedings*, DOI: [10.1063/1.3275622](https://doi.org/10.1063/1.3275622) (AIP, 2009). [arXiv:0907.5123].
- [32] Knuth, K. H., Habeck, M., Malakar, N. K., Mubeen, A. M. & Placek, B. Bayesian evidence and model selection. *Digital Signal Processing* **47**, 50–67, DOI: [10.1016/j.dsp.2015.06.012](https://doi.org/10.1016/j.dsp.2015.06.012) (2015). [arXiv:1411.3013].
- [33] Zhao, Z. & Severini, T. A. Integrated likelihood computation methods. *Computational Statistics* **32**, 281–313, DOI: [10.1007/s00180-016-0677-z](https://doi.org/10.1007/s00180-016-0677-z) (2016).
- [34] Llorente, F., Martino, L., Delgado, D. & Lopez-Santiago, J. Marginal likelihood computation for model selection and hypothesis testing: an extensive review. *arXiv e-prints* (2020). [arXiv:2005.08334].
- [35] Tierney, L. & Kadane, J. B. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82–86, DOI: [10.1080/01621459.1986.10478240](https://doi.org/10.1080/01621459.1986.10478240) (1986).
- [36] Chib, S. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321, DOI: [10.1080/01621459.1995.10476635](https://doi.org/10.1080/01621459.1995.10476635) (1995).
- [37] Kloek, T. & van Dijk, H. K. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* **46**, 1–19 (1978).
- [38] Newton, M. A. & Raftery, A. E. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)* **56**, 3–26, DOI: [10.1111/j.2517-6161.1994.tb01956.x](https://doi.org/10.1111/j.2517-6161.1994.tb01956.x) (1994).
- [39] Gelman, A. & Meng, X.-L. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185, DOI: [10.1214/ss/1028905934](https://doi.org/10.1214/ss/1028905934) (1998).
Classic reference that introduces the notion of path sampling
- [40] Cameron, E. & Pettitt, A. N. Recursive Pathways to Marginal Likelihood Estimation with Prior-Sensitivity Analysis. *Statistical Science* **29**, 397–419, DOI: [10.1214/13-sts465](https://doi.org/10.1214/13-sts465) (2014). [arXiv:1301.6450].

- [41] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423, DOI: 10.1002/j.1538-7305.1948.tb01338.x (1948).
- [42] Jaynes, E. T. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* **4**, 227–241, DOI: 10.1109/tssc.1968.300117 (1968).
- [43] Keeton, C. R. On statistical uncertainty in nested sampling. *MNRAS* **414**, 1418–1426, DOI: 10.1111/j.1365-2966.2011.18474.x (2011). [arXiv:1102.0996].
- [44] Baldock, R. J. N., Bernstein, N., Salerno, K. M., Pártay, L. B. & Csányi, G. Constant-pressure nested sampling with atomistic dynamics. *Phys. Rev. E* **96**, 43311–43324, DOI: 10.1103/physreve.96.043311 (2017). [arXiv:1710.11085].
- [45] Creutz, M. Microcanonical Monte Carlo simulation. *Phys. Rev. Lett.* **50**, 1411, DOI: 10.1103/physrevlett.50.1411 (1983).
- [46] Habeck, M. Nested sampling with demons. In Mohammad-Djafari, A. & Barbaresco, F. (eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2014)*, vol. 1641, 121–129, DOI: 10.1063/1.4905971. American Institute of Physics (AIP, New York, 2015).
- [47] Wang, F. & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**, 2050–2053, DOI: 10.1103/physrevlett.86.2050 (2001). [arXiv:cond-mat/0011174].
- [48] Feroz, F. & Hobson, M. P. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS* **384**, 449–463, DOI: 10.1111/j.1365-2966.2007.12353.x (2008). [arXiv:0704.3704].
- Popularized NS in astrophysics and cosmology by introducing the MultiNest implementation**
- [49] Pártay, L. B., Bartók, A. P. & Csányi, G. Efficient sampling of atomic configurational spaces. *The Journal of Physical Chemistry B* **114**, 10502–10512, DOI: 10.1021/jp1012973 (2010). [arXiv:0906.3544].
- Introduced NS for atomistic modelling**
- [50] Cérou, E., Moral, P., Furon, T. & Guyader, A. Sequential Monte Carlo for rare event estimation. *Statistics and Computing* **22**, 795–808, DOI: 10.1007/s11222-011-9231-6 (2012).
- [51] Higson, E., Handley, W., Hobson, M. & Lasenby, A. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing* **29**, 891–913, DOI: 10.1007/s11222-018-9844-0 (2018). [arXiv:1704.03459].
- Introduced an important dynamic variant of NS that speeds up parameter inference**
- [52] Speagle, J. S. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS* **493**, 3132–3158, DOI: 10.1093/mnras/staa278 (2020). [arXiv:1904.02180].
- [53] Higson, E. dyPolyChord: dynamic nested sampling with PolyChord. *The Journal of Open Source Software* **3**, 965, DOI: 10.21105/joss.00965 (2018).
- [54] Buchner, J. Nested Sampling Methods. *arXiv e-prints* (2021). [arXiv:2101.09675].
- [55] Brewer, B. J., Pártay, L. B. & Csányi, G. Diffusive nested sampling. *Statistics and Computing* **21**, 649–656, DOI: 10.1007/s11222-010-9198-8 (2010). [arXiv:0912.2380].
- [56] Buchner, J. A statistical test for Nested Sampling algorithms. *Statistics and Computing* **26**, 383–392, DOI: 10.1007/s11222-014-9512-y (2016). [arXiv:1407.5459].
- [57] Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* **5**, 65–80, DOI: 10.2140/camcos.2010.5.65 (2010).
- [58] Allison, R. & Dunkley, J. Comparison of sampling techniques for Bayesian parameter estimation. *MNRAS* **437**, 3918–3928, DOI: 10.1093/mnras/stt2190 (2014). [arXiv:1308.2675].
- [59] Buchner, J. Collaborative Nested Sampling: Big Data versus Complex Physical Models. *PASP* **131**, 108005, DOI: 10.1088/1538-3873/aae7fc (2019). [arXiv:1707.04476].
- [60] Mukherjee, P., Parkinson, D. & Liddle, A. R. A Nested Sampling Algorithm for Cosmological Model Selection. *Ap. J.* **638**, L51–L54, DOI: 10.1086/501068 (2006). [arXiv:astro-ph/0508461].
- [61] Parkinson, D., Mukherjee, P. & Liddle, A. R. Bayesian model selection analysis of WMAP3. *Phys. Rev. D* **73**, 123523, DOI: 10.1103/physrevd.73.123523 (2006). [arXiv:astro-ph/0605003].
- [62] Shaw, J. R., Bridges, M. & Hobson, M. P. Efficient Bayesian inference for multimodal problems in cosmology. *MNRAS* **378**, 1365–1370, DOI: 10.1111/j.1365-2966.2007.11871.x (2007). [arXiv:astro-ph/0701867].
- [63] Veitch, J. & Vecchio, A. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Phys. Rev. D* **81**, 062003, DOI: 10.1103/physrevd.81.062003 (2010). [arXiv:0911.3820].
- [64] Ter Braak, C. J. F. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing* **16**, 239–249, DOI: 10.1007/s11222-006-8769-1 (2006).

- [65] Jasa, T. & Xiang, N. Nested sampling applied in Bayesian room-acoustics decay analysis. *Acoustical Society of America Journal* **132**, 3251, DOI: [10.1121/1.4754550](https://doi.org/10.1121/1.4754550) (2012).
- [66] Handley, W. J., Hobson, M. P. & Lasenby, A. N. POLYCHORD: next-generation nested sampling. *MNRAS* **453**, 4384–4398, DOI: [10.1093/mnras/stv1911](https://doi.org/10.1093/mnras/stv1911) (2015). [[arXiv:1506.00171](https://arxiv.org/abs/1506.00171)].
- [67] Smith, R. L. Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research* **32**, 1296–1308 (1984).
- [68] Zabinsky, Z. B. & Smith, R. L. *Hit-and-Run Methods*, 721–729 (Springer US, Boston, MA, 2013).
- [69] Betancourt, M. Nested Sampling with Constrained Hamiltonian Monte Carlo. In Mohammad-Djafari, J. B. A. & Bessière, P. (eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2010)*, vol. 1305, 165–172, DOI: [10.1063/1.3573613](https://doi.org/10.1063/1.3573613). American Institute of Physics (AIP, New York, 2011).
- [70] Skilling, J. Bayesian computation in big spaces-nested sampling and Galilean Monte Carlo. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2011)*, vol. 1443, 145–156, DOI: [10.1063/1.3703630](https://doi.org/10.1063/1.3703630). American Institute of Physics (AIP, New York, 2012).
- [71] Griffiths, M. & Wales, D. J. Nested basin-sampling. *J. Chem. Theory Comput.* **15**, 6865, DOI: [10.1021/acs.jctc.9b00567](https://doi.org/10.1021/acs.jctc.9b00567) (2019).
- [72] Olander, J. *Constrained space MCMC methods for nested sampling Bayesian computations*. Ph.D. thesis, Chalmers tekniska högskola, Institutionen för fysik (2020).
- [73] Stokes, B., Tuyl, F. & Hudson, I. New prior sampling methods for nested sampling — Development and testing. In Verdoolaege, G. (ed.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2016)*, vol. 1853, 110003, DOI: [10.1063/1.4985378](https://doi.org/10.1063/1.4985378). American Institute of Physics (AIP, New York, 2017).
- [74] Higson, E., Handley, W., Hobson, M. & Lasenby, A. NESTCHECK: diagnostic tests for nested sampling calculations. *MNRAS* **483**, 2044–2056, DOI: [10.1093/mnras/sty3090](https://doi.org/10.1093/mnras/sty3090) (2019). [[arXiv:1804.06406](https://arxiv.org/abs/1804.06406)].
- [75] Burkoff, N. S., Várnai, C., Wells, S. A. & Wild, D. L. Exploring the energy landscapes of protein folding simulations with Bayesian computation. *Biophysical Journal* **102**, 878–886, DOI: [10.1016/j.bpj.2011.12.053](https://doi.org/10.1016/j.bpj.2011.12.053) (2012).
- [76] Henderson, R. W. & Goggans, P. M. Parallelized nested sampling. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2013)*, vol. 1636, 100–105, DOI: [10.1063/1.4903717](https://doi.org/10.1063/1.4903717). American Institute of Physics (AIP, New York, 2014).
- [77] Baldock, R. J. N., Pártay, L. B., Bartók, A. P., Payne, M. C. & Csányi, G. Determining the pressure-temperature phase diagrams of materials. *Phys. Rev. B* **93**, 174108, DOI: [10.1103/physrevb.93.174108](https://doi.org/10.1103/physrevb.93.174108) (2016). [[arXiv:1503.03404](https://arxiv.org/abs/1503.03404)].
Adapted NS for materials simulation with periodic boundary conditions
- [78] Handley, W. anesthetic: nested sampling visualisation. *The Journal of Open Source Software* **4**, 1414, DOI: [10.21105/joss.01414](https://doi.org/10.21105/joss.01414) (2019). [[arXiv:1905.04768](https://arxiv.org/abs/1905.04768)].
- [79] Fowlie, A. & Bardsley, M. H. Superplot: a graphical interface for plotting and analysing MultiNest output. *Eur. Phys. J. Plus* **131**, 391, DOI: [10.1140/epjp/i2016-16391-0](https://doi.org/10.1140/epjp/i2016-16391-0) (2016). [[arXiv:1603.00555](https://arxiv.org/abs/1603.00555)].
- [80] Scott, P. Pippi - painless parsing, post-processing and plotting of posterior and likelihood samples. *Eur. Phys. J. Plus* **127**, 138, DOI: [10.1140/epjp/i2012-12138-3](https://doi.org/10.1140/epjp/i2012-12138-3) (2012). [[arXiv:1206.2245](https://arxiv.org/abs/1206.2245)].
- [81] Lewis, A. GetDist: a Python package for analysing Monte Carlo samples. *arXiv e-prints* (2019). [[arXiv:1910.13970](https://arxiv.org/abs/1910.13970)].
- [82] Foreman-Mackey, D. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software* **1**, 24, DOI: [10.21105/joss.00024](https://doi.org/10.21105/joss.00024) (2016).
- [83] Bocquet, S. & Carter, F. W. pygtc: beautiful parameter covariance plots (aka. giant triangle confusograms). *The Journal of Open Source Software* **1**, DOI: [10.21105/joss.00046](https://doi.org/10.21105/joss.00046) (2016).
- [84] Feroz, F., Hobson, M. P. & Bridges, M. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS* **398**, 1601–1614, DOI: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x) (2009). [[arXiv:0809.3437](https://arxiv.org/abs/0809.3437)].
- [85] Feroz, F., Hobson, M. P., Cameron, E. & Pettitt, A. N. Importance Nested Sampling and the MultiNest Algorithm. *Open J. Astrophys.* **2**, 10, DOI: [10.21105/astro.1306.2144](https://doi.org/10.21105/astro.1306.2144) (2019). [[arXiv:1306.2144](https://arxiv.org/abs/1306.2144)].
- [86] Beaujean, F. & Caldwell, A. Initializing adaptive importance sampling with Markov chains. *arXiv e-prints* (2013). [[arXiv:1304.7808](https://arxiv.org/abs/1304.7808)].
- [87] Rosenbrock, H. H. An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal* **3**, 175–184, DOI: [10.1093/comjnl/3.3.175](https://doi.org/10.1093/comjnl/3.3.175) (1960).
- [88] Higson, E. nestcheck: error analysis, diagnostic tests and plots for nested sampling calculations. *The Journal of Open Source Software* **3**, 916, DOI: [10.21105/joss.00916](https://doi.org/10.21105/joss.00916) (2018).
- [89] Fowlie, A., Handley, W. & Su, L. Nested sampling cross-checks using order statistics. *MNRAS* **497**, 5256–5263, DOI: [10.1093/mnras/staa2345](https://doi.org/10.1093/mnras/staa2345) (2020). [[arXiv:2006.03371](https://arxiv.org/abs/2006.03371)].

Identified a previously unused property of NS and showed how it could be used to test individual NS runs

- [90] Williams, M. J. *nessai*: Nested Sampling with Artificial Intelligence, DOI: [10.5281/zenodo.4550693](https://doi.org/10.5281/zenodo.4550693) (2021).
- [91] Williams, M. J., Veitch, J. & Messenger, C. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D* **103**, 103006, DOI: [10.1103/physrevd.103.103006](https://doi.org/10.1103/physrevd.103.103006) (2021). [arXiv:2102.11056].
- [92] Buchner, J. UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software* **6**, 3001, DOI: [10.21105/joss.03001](https://doi.org/10.21105/joss.03001) (2021). [arXiv:2101.09604].
- [93] Stokes, B., Tuyl, F. & Hudson, I. Equidistribution testing with Bayes factors and the ECT. In Giffin, A. & Knuth, K. H. (eds.) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MAXENT 2015)*, vol. 1757, 040001, DOI: [10.1063/1.4959055](https://doi.org/10.1063/1.4959055). American Institute of Physics (AIP, New York, 2016).
- [94] Stokes, B. J. *New prior sampling methods and equidistribution testing for nested sampling*. Ph.D. thesis, University of Newcastle (2018).
- [95] Romero-Shaw, I. M. *et al.* Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO-Virgo gravitational-wave transient catalogue. *MNRAS* **499**, 3295–3319, DOI: [10.1093/mnras/staa2850](https://doi.org/10.1093/mnras/staa2850) (2020). [arXiv:2006.00714].
- [96] Henderson, R. W., Goggans, P. M. & Cao, L. Combined-chain nested sampling for efficient Bayesian model comparison. *Digit. Signal Process.* **70**, 84–93, DOI: [10.1016/j.dsp.2017.07.021](https://doi.org/10.1016/j.dsp.2017.07.021) (2017).
- [97] Russel, P. M., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model selection and parameter inference in phylogenetics using nested sampling. *Syst. Biol.* **68**, 219–233, DOI: [10.1093/sysbio/syy050](https://doi.org/10.1093/sysbio/syy050) (2019). [arXiv:1703.05471].
- [98] Pullen, N. & Morris, R. J. Bayesian model comparison and parameter inference in systems biology using nested sampling. *PLOS ONE* **9**, e88419, DOI: [10.1371/journal.pone.0088419](https://doi.org/10.1371/journal.pone.0088419) (2014).
- [99] Mikelson, J. & Khammash, M. Likelihood-free nested sampling for parameter inference of biochemical reaction networks. *PLOS Computational Biology* **16**, e1008264, DOI: [10.1371/journal.pcbi.1008264](https://doi.org/10.1371/journal.pcbi.1008264) (2020).
- [100] Beaton, D. & Xiang, N. Room acoustic modal analysis using Bayesian inference. *J. Acoust. Soc. Am.* **141**, 4480–4493, DOI: [10.1121/1.4983301](https://doi.org/10.1121/1.4983301) (2017).
- [101] Van Soom, M. & de Boer, B. Detrending the waveforms of steady-state vowels. *Entropy* **22**, 331, DOI: [10.3390/e22030331](https://doi.org/10.3390/e22030331) (2020).
- [102] Lewis, S., Ireland, D. & Vanderbauwhede, W. Development of Bayesian analysis program for extraction of polarisation observables at CLAS. *Journal of Physics: Conference Series* **513**, 022020, DOI: [10.1088/1742-6596/513/2/022020](https://doi.org/10.1088/1742-6596/513/2/022020) (2014).
- [103] Ozturk, F. C. *et al.* New test of modulated electron capture decay of hydrogen-like ^{142}Pm ions: precision measurement of purely exponential decay. *Phys. Lett. B* **797**, 134800, DOI: [10.1016/j.physletb.2019.134800](https://doi.org/10.1016/j.physletb.2019.134800) (2019). [arXiv:1907.06920].
- [104] Trassinelli, M. *et al.* Measurement of the charged pion mass using X-ray spectroscopy of exotic atoms. *Phys. Lett. B* **759**, 583–588, DOI: [10.1016/j.physletb.2016.06.025](https://doi.org/10.1016/j.physletb.2016.06.025) (2016). [arXiv:1605.03300].
- [105] Trassinelli, M. Bayesian data analysis tools for atomic physics. *Nucl. Instrum. Meth. B* **408**, 301–312, DOI: [10.1016/j.nimb.2017.05.030](https://doi.org/10.1016/j.nimb.2017.05.030) (2017). [arXiv:1611.10189].
- [106] Covita, D. S. *et al.* Line shape analysis of the $K\beta$ transition in muonic hydrogen. *Eur. Phys. J. D* **72**, 72, DOI: [10.1140/epjd/e2018-80593-1](https://doi.org/10.1140/epjd/e2018-80593-1) (2018). [arXiv:1709.05950].
- [107] De Anda Villa, M. *et al.* Assessing the surface oxidation state of free-standing gold nanoparticles produced by laser ablation. *Langmuir* **35**, 11859–11871, DOI: [10.1021/acs.langmuir.9b02159](https://doi.org/10.1021/acs.langmuir.9b02159) (2019).
- [108] Machado, J. *et al.* High-precision measurements of $n = 2 \rightarrow n = 1$ transition energies and level widths in He- and Be-like argon ions. *Phys. Rev. A* **97**, 032517, DOI: [10.1103/physreva.97.032517](https://doi.org/10.1103/physreva.97.032517) (2018). [arXiv:1802.05970].
- [109] Brewer, B. J. & Donovan, C. P. Fast Bayesian inference for exoplanet discovery in radial velocity data. *MNRAS* **448**, 3206–3214, DOI: [10.1093/mnras/stv199](https://doi.org/10.1093/mnras/stv199) (2015). [arXiv:1501.06952].
- [110] Lavie, B. *et al.* HELIOS-RETRIEVAL: An Open-source, Nested Sampling Atmospheric Retrieval Code; Application to the HR 8799 Exoplanets and Inferred Constraints for Planet Formation. *AJ* **154**, 91, DOI: [10.3847/1538-3881/aa7ed8](https://doi.org/10.3847/1538-3881/aa7ed8) (2017). [arXiv:1610.03216].
- [111] Hall, R. D., Thompson, S. J., Handley, W. & Queloz, D. On the Feasibility of Intense Radial Velocity Surveys for Earth-Twin Discoveries. *MNRAS* **479**, 2968–2987, DOI: [10.1093/mnras/sty1464](https://doi.org/10.1093/mnras/sty1464) (2018). [arXiv:1806.00518].
- [112] Kitzmann, D. *et al.* Helios-r2: A New Bayesian, Open-source Retrieval Model for Brown Dwarfs and Exoplanet Atmospheres. *ApJ* **890**, 174, DOI: [10.3847/1538-4357/ab6d71](https://doi.org/10.3847/1538-4357/ab6d71) (2020). [arXiv:1910.01070].
- [113] Ahrer, E. *et al.* The HARPS search for southern extrasolar planets - XLV. Two Neptune mass planets orbiting HD 13808: a study of stellar activity modelling's impact on planet detection. *MNRAS* **503**, 1248–1263, DOI: [10.1093/mnras/stab373](https://doi.org/10.1093/mnras/stab373) (2021). [arXiv:2102.03387].

- [114] Elsheikh, A. H., Wheeler, M. F. & Hoteit, I. Nested sampling algorithm for subsurface flow model selection, uncertainty quantification, and nonlinear calibration. *Water Resources Research* **49**, 8383–8399, DOI: 10.1002/2012wr013406 (2013).
- [115] Spergel, D. N. *et al.* First-year wilkinson microwave anisotropy probe (WMAP) observations: Determination of cosmological parameters. *The Astrophysical Journal Supplement Series* **148**, 175–194, DOI: 10.1086/377226 (2003). [arXiv:astro-ph/0302209].
- [116] Riess, A. G. *et al.* Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Ap. J.* **116**, 1009–1038, DOI: 10.1086/300499 (1998). [arXiv:astro-ph/9805201].
- [117] Perlmutter, S. *et al.* Measurements of Ω and λ from 42 High-Redshift Supernovae. *Ap. J.* **517**, 565–586, DOI: 10.1086/307221 (1999). [arXiv:astro-ph/9812133].
- [118] Guth, A. H. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D* **23**, 347–356, DOI: 10.1103/physrevd.23.347 (1981).
- [119] Liddle, A. R., Parsons, P. & Barrow, J. D. Formalizing the slow-roll approximation in inflation. *Phys. Rev. D* **50**, 7222–7232, DOI: 10.1103/physrevd.50.7222 (1994). [arXiv:astro-ph/9408015].
- [120] Martin, J., Ringeval, C. & Trota, R. Hunting down the best model of inflation with Bayesian evidence. *Phys. Rev. D* **83**, DOI: 10.1103/physrevd.83.063524 (2011). [arXiv:1009.4157].
- [121] Allen, S. W., Schmidt, R. W. & Fabian, A. C. Cosmological constraints from the X-ray gas mass fraction in relaxed lensing clusters observed with chandra. *MNRAS* **334**, L11–L15, DOI: 10.1046/j.1365-8711.2002.05601.x (2002). [arXiv:astro-ph/0205007].
- [122] Allen, S. W., Evrard, A. E. & Mantz, A. B. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics* **49**, 409–470, DOI: 10.1146/annurev-astro-081710-102514 (2011). [arXiv:1103.4829].
- [123] Feroz, F., Hobson, M. P., Zwart, J. T. L., Saunders, R. D. E. & Grainge, K. J. B. Bayesian modelling of clusters of galaxies from multifrequency-pointed Sunyaev-Zel’dovich observations. *MNRAS* **398**, 2049–2060, DOI: 10.1111/j.1365-2966.2009.15247.x (2009).
- [124] Hurley-Walker, N. *et al.* Bayesian analysis of weak gravitational lensing and Sunyaev-Zel’dovich data for six galaxy clusters. *MNRAS* **419**, 2921–2942, DOI: 10.1111/j.1365-2966.2011.19937.x (2011).
- [125] Joudaki, S. *et al.* KiDS+VIKING-450 and DES-Y1 combined: Cosmology with cosmic shear. *A&A* **638**, L1, DOI: 10.1051/0004-6361/201936154 (2020). [arXiv:1906.09262].
- [126] DES Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D* **98**, 043526, DOI: 10.1103/physrevd.98.043526 (2018). [arXiv:1708.01530].
- [127] Asgari, M. *et al.* KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics. *A&A* **645**, A104, DOI: 10.1051/0004-6361/202039070 (2021). [arXiv:2007.15633].
- [128] Handley, W. & Lemos, P. Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio. *Phys. Rev. D* **100**, 043504, DOI: 10.1103/physrevd.100.043504 (2019). [arXiv:1902.04029].
- [129] Conley, A. *et al.* Supernova Constraints and Systematic Uncertainties from the First 3 Years of the Supernova Legacy Survey. *Astrophys. J. Suppl.* **192**, 1, DOI: 10.1088/0067-0049/192/1/1 (2011). [arXiv:1104.1443].
- [130] March, M. C., Trota, R., Berkes, P., Starkman, G. D. & Vaudrevange, P. M. Improved constraints on cosmological parameters from Type Ia supernova data. *MNRAS* **418**, 2308–2329, DOI: 10.1111/j.1365-2966.2011.19584.x (2011).
- [131] Handley, W. Curvature tension: evidence for a closed universe. *Phys. Rev. D* **103**, L041301, DOI: 10.1103/physrevd.103.L041301 (2021). [arXiv:1908.09139].
- [132] Planck Collaboration. Planck 2013 results. I. Overview of products and scientific results. *A&A* **571**, A1, DOI: 10.1051/0004-6361/201321529 (2014). [arXiv:1303.5062].
- [133] Handley, W. J., Hobson, M. P. & Lasenby, A. N. polychord: nested sampling for cosmology. *MNRAS* **450**, L61–L65, DOI: 10.1093/mnras/slv047 (2015). [arXiv:1502.01856].
- Opened up high-dimensional problems with slice-sampling implementation of NS**
- [134] Aitken, S. & Akman, O. E. Nested sampling for parameter inference in systems biology: application to an exemplar circadian model. *BMC systems biology* **7**, 1–12, DOI: 10.1186/1752-0509-7-72 (2013).
- [135] Planck Collaboration. Planck 2015 results. XX. Constraints on inflation. *A&A* **594**, A20, DOI: 10.1051/0004-6361/201525898 (2016). [arXiv:1502.02114].
- [136] Handley, W. J., Lasenby, A. N., Peiris, H. V. & Hobson, M. P. Bayesian inflationary reconstructions from Planck 2018 data. *Phys. Rev. D* **100**, 103511, DOI: 10.1103/physrevd.100.103511 (2019). [arXiv:1908.00906].
- [137] Hergt, L. T., Handley, W. J., Hobson, M. P. & Lasenby, A. N. Constraining the kinetically dominated universe. *Phys. Rev. D* **100**, 023501, DOI: 10.1103/physrevd.100.023501 (2019). [arXiv:1809.07737].

- [138] Gessey-Jones, T. & Handley, W. J. Constraining quantum initial conditions before inflation. *Phys. Rev. D* **104**, 063532, DOI: [10.1103/physrevd.104.063532](https://doi.org/10.1103/physrevd.104.063532) (2021). [arXiv:2104.03016].
- [139] Zhao, G.-B. *et al.* Dynamical dark energy in light of the latest observations. *Nature Astronomy* **1**, 627–632, DOI: [10.1038/s41550-017-0216-z](https://doi.org/10.1038/s41550-017-0216-z) (2017). [arXiv:1701.08165].
- [140] Hee, S., Vázquez, J. A., Handley, W. J., Hobson, M. P. & Lasenby, A. N. Constraining the dark energy equation of state using Bayes theorem and the Kullback-Leibler divergence. *MNRAS* **466**, 369–377, DOI: [10.1093/mnras/stw3102](https://doi.org/10.1093/mnras/stw3102) (2017). [arXiv:1607.00270].
- [141] Higson, E., Handley, W., Hobson, M. & Lasenby, A. Bayesian sparse reconstruction: a brute-force approach to astronomical imaging and machine learning. *MNRAS* **483**, 4828–4846, DOI: [10.1093/mnras/sty3307](https://doi.org/10.1093/mnras/sty3307) (2019). [arXiv:1809.04598].
- [142] Renk, J. J. *et al.* CosmoBit: a GAMBIT module for computing cosmological observables and likelihoods. *J. Cosmology Astropart. Phys.* **2021**, 022, DOI: [10.1088/1475-7516/2021/02/022](https://doi.org/10.1088/1475-7516/2021/02/022) (2021). [arXiv:2009.03286].
- [143] Stöcker, P. *et al.* Strengthening the bound on the mass of the lightest neutrino with terrestrial and cosmological experiments. *Phys. Rev. D* **103**, 123508, DOI: [10.1103/physrevd.103.123508](https://doi.org/10.1103/physrevd.103.123508) (2021). [arXiv:2009.03287].
- [144] Anstey, D., de Lera Acedo, E. & Handley, W. A general Bayesian framework for foreground modelling and chromaticity correction for global 21 cm experiments. *MNRAS* **506**, 2041–2058, DOI: [10.1093/mnras/stab1765](https://doi.org/10.1093/mnras/stab1765) (2021). [arXiv:2010.09644].
- [145] Martin, S. P. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.* **18**, 1–98, DOI: [10.1142/9789812839657_0001](https://doi.org/10.1142/9789812839657_0001) (1998). [arXiv:hep-ph/9709356].
- [146] de Austri, R. R., Trotta, R. & Roszkowski, L. A Markov chain Monte Carlo analysis of the CMSSM. *JHEP* **05**, 002, DOI: [10.1088/1126-6708/2006/05/002](https://doi.org/10.1088/1126-6708/2006/05/002) (2006). [arXiv:hep-ph/0602028].
- [147] Trotta, R., Feroz, F., Hobson, M. P., Roszkowski, L. & de Austri, R. R. The Impact of priors and observables on parameter inferences in the Constrained MSSM. *JHEP* **12**, 024, DOI: [10.1088/1126-6708/2008/12/024](https://doi.org/10.1088/1126-6708/2008/12/024) (2008). [arXiv:0809.3792].
- [148] Feroz, F., Cranmer, K., Hobson, M., de Austri, R. R. & Trotta, R. Challenges of Profile Likelihood Evaluation in Multi-Dimensional SUSY Scans. *JHEP* **06**, 042, DOI: [10.1007/jhep06\(2011\)042](https://doi.org/10.1007/jhep06(2011)042) (2011). [arXiv:1101.3296].
- [149] Trotta, R. *et al.* Constraints on cosmic-ray propagation models from a global Bayesian analysis. *Astrophys. J.* **729**, 106, DOI: [10.1088/0004-637x/729/2/106](https://doi.org/10.1088/0004-637x/729/2/106) (2011). [arXiv:1011.0037].
- [150] AbdusSalam, S. S., Allanach, B. C., Quevedo, F., Feroz, F. & Hobson, M. Fitting the Phenomenological MSSM. *Phys. Rev. D* **81**, 095012, DOI: [10.1103/physrevd.81.095012](https://doi.org/10.1103/physrevd.81.095012) (2010). [arXiv:0904.2548].
- [151] Strece, C. *et al.* Updated global fits of the cMSSM including the latest LHC SUSY and Higgs searches and XENON100 data. *J. Cosmology Astropart. Phys.* **03**, 030, DOI: [10.1088/1475-7516/2012/03/030](https://doi.org/10.1088/1475-7516/2012/03/030) (2012). [arXiv:1112.4192].
- [152] Buchmueller, O. *et al.* The CMSSM and NUHM1 after LHC Run 1. *Eur. Phys. J. C* **74**, 2922, DOI: [10.1140/epjc/s10052-014-2922-3](https://doi.org/10.1140/epjc/s10052-014-2922-3) (2014). [arXiv:1312.5250].
- [153] Fowlie, A. *et al.* The CMSSM Favoring New Territories: The Impact of New LHC Limits and a 125 GeV Higgs. *Phys. Rev. D* **86**, 075010, DOI: [10.1103/physrevd.86.075010](https://doi.org/10.1103/physrevd.86.075010) (2012). [arXiv:1206.0264].
- [154] Fowlie, A., Kowalska, K., Roszkowski, L., Sessolo, E. M. & Tsai, Y.-L. S. Dark matter and collider signatures of the MSSM. *Phys. Rev. D* **88**, 055012, DOI: [10.1103/physrevd.88.055012](https://doi.org/10.1103/physrevd.88.055012) (2013). [arXiv:1306.1567].
- [155] Catena, R. & Gondolo, P. Global fits of the dark matter-nucleon effective interactions. *J. Cosmology Astropart. Phys.* **09**, 045, DOI: [10.1088/1475-7516/2014/09/045](https://doi.org/10.1088/1475-7516/2014/09/045) (2014). [arXiv:1405.2637].
- [156] de Vries, K. J. *et al.* The pMSSM10 after LHC Run 1. *Eur. Phys. J. C* **75**, 422, DOI: [10.1140/epjc/s10052-015-3599-y](https://doi.org/10.1140/epjc/s10052-015-3599-y) (2015). [arXiv:1504.03260].
- [157] Hernández, P., Kekic, M., López-Pavón, J., Racker, J. & Salvado, J. Testable Baryogenesis in Seesaw Models. *JHEP* **08**, 157, DOI: [10.1007/jhep08\(2016\)157](https://doi.org/10.1007/jhep08(2016)157) (2016). [arXiv:1606.06719].
- [158] Kreisch, C. D., Cyr-Racine, F.-Y. & Doré, O. Neutrino puzzle: Anomalies, interactions, and cosmological tensions. *Phys. Rev. D* **101**, 123505, DOI: [10.1103/physrevd.101.123505](https://doi.org/10.1103/physrevd.101.123505) (2020). [arXiv:1902.00534].
- [159] Martinez, G. *et al.* Comparison of statistical sampling methods with scannerbit, the GAMBIT scanning module. *Eur. Phys. J. C* **77**, DOI: [10.1140/epjc/s10052-017-5274-y](https://doi.org/10.1140/epjc/s10052-017-5274-y) (2017). [arXiv:1705.07959].
- [160] Balázs, C. *et al.* A comparison of optimisation algorithms for high-dimensional particle and astrophysics applications. *JHEP* **05**, 108, DOI: [10.1007/jhep05\(2021\)108](https://doi.org/10.1007/jhep05(2021)108) (2021). [arXiv:2101.04525].
- [161] Fowlie, A., Hoof, S. & Handley, W. Nested Sampling for Frequentist Computation: Fast Estimation of Small p-Values. *Phys. Rev. Lett.* **128**, 021801, DOI: [10.1103/physrevlett.128.021801](https://doi.org/10.1103/physrevlett.128.021801) (2022). [arXiv:2105.13923].

- [162] Abbott, B. P. *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* **116**, 061102, DOI: [10.1103/physrevlett.116.061102](https://doi.org/10.1103/physrevlett.116.061102) (2016). [arXiv:1602.03837].
- [163] Aasi, J. *et al.* Advanced LIGO. *Class. Quant. Grav.* **32**, 074001, DOI: [10.1088/0264-9381/32/7/074001](https://doi.org/10.1088/0264-9381/32/7/074001) (2015). [arXiv:1411.4547].
- [164] Acernese, F. *et al.* Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* **32**, 024001, DOI: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001) (2015). [arXiv:1408.3978].
- [165] Abbott, R. *et al.* GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv e-prints* (2021). [arXiv:2111.03606].
- [166] Veitch, J. & Vecchio, A. A Bayesian approach to the follow-up of candidate gravitational wave signals. *Phys. Rev. D* **78**, 022001, DOI: [10.1103/physrevd.78.022001](https://doi.org/10.1103/physrevd.78.022001) (2008). [arXiv:0801.4313].
- [167] Abbott, B. P. *et al.* Model comparison from LIGO–Virgo data on GW170817’s binary components and consequences for the merger remnant. *Class. Quant. Grav.* **37**, 045006, DOI: [10.1088/1361-6382/ab5f7c](https://doi.org/10.1088/1361-6382/ab5f7c) (2020). [arXiv:1908.01012].
- [168] Smith, R. J. E., Ashton, G., Vajpeyi, A. & Talbot, C. Massively parallel Bayesian inference for transient gravitational-wave astronomy. *MNRAS* **498**, 4492–4502, DOI: [10.1093/mnras/staa2483](https://doi.org/10.1093/mnras/staa2483) (2020). [arXiv:1909.11873].
- [169] Pitkin, M., Isi, M., Veitch, J. & Woan, G. A nested sampling code for targeted searches for continuous gravitational waves from pulsars. *arXiv e-prints* (2017). [arXiv:1705.08978].
- [170] Abbott, B. P. *et al.* Searches for Gravitational Waves from Known Pulsars at Two Harmonics in 2015–2017 LIGO Data. *Ap. J.* **879**, 10, DOI: [10.3847/1538-4357/ab20cb](https://doi.org/10.3847/1538-4357/ab20cb) (2019). [arXiv:1902.08507].
- [171] Lynch, R., Vitale, S., Essick, R., Katsavounidis, E. & Robinet, F. Information-theoretic approach to the gravitational-wave burst detection problem. *Phys. Rev. D* **95**, 104046, DOI: [10.1103/physrevd.95.104046](https://doi.org/10.1103/physrevd.95.104046) (2017). [arXiv:1511.05955].
- [172] Powell, J., Gossan, S. E., Logue, J. & Heng, I. S. Inferring the core-collapse supernova explosion mechanism with gravitational waves. *Phys. Rev. D* **94**, 123012, DOI: [10.1103/physrevd.94.123012](https://doi.org/10.1103/physrevd.94.123012) (2016). [arXiv:1610.05573].
- [173] Smith, R. & Thrane, E. Optimal search for an astrophysical gravitational-wave background. *Phys. Rev. X* **8**, 021019, DOI: [10.1103/physrevx.8.021019](https://doi.org/10.1103/physrevx.8.021019) (2018). [arXiv:1712.00688].
- [174] Futamase, T. & Itoh, Y. The Post-Newtonian Approximation for Relativistic Compact Binaries. *Living Rev. Relativ.* **10**, 2, DOI: [10.12942/lrr-2007-2](https://doi.org/10.12942/lrr-2007-2) (2007).
- [175] Blanchet, L. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Rev. Relativ.* **17**, 2, DOI: [10.12942/lrr-2014-2](https://doi.org/10.12942/lrr-2014-2) (2014). [arXiv:1310.1528].
- [176] Hannam, M. Modelling gravitational waves from precessing black-hole binaries: progress, challenges and prospects. *General Relativity and Gravitation* **46**, 1767, DOI: [10.1007/s10714-014-1767-2](https://doi.org/10.1007/s10714-014-1767-2) (2014). [arXiv:1312.3641].
- [177] Bishop, N. T. & Rezzolla, L. Extraction of gravitational waves in numerical relativity. *Living Rev. Relativ.* **19**, 2, DOI: [10.1007/s41114-016-0001-9](https://doi.org/10.1007/s41114-016-0001-9) (2016). [arXiv:1606.02532].
- [178] Baiotti, L. Gravitational waves from neutron star mergers and their relation to the nuclear equation of state. *Progress in Particle and Nuclear Physics* **109**, 103714, DOI: <https://doi.org/10.1016/j.pnpnp.2019.103714> (2019). [arXiv:1907.08534].
- [179] Abbott, B. P. *et al.* GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.* **116**, 241103, DOI: [10.1103/physrevlett.116.241103](https://doi.org/10.1103/physrevlett.116.241103) (2016). [arXiv:1606.04855].
- [180] Veitch, J. *et al.* Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D* **91**, 042003, DOI: [10.1103/physrevd.91.042003](https://doi.org/10.1103/physrevd.91.042003) (2015). [arXiv:1409.7215].
- [181] Cornish, N. J. *et al.* BayesWave analysis pipeline in the era of gravitational wave observations. *Phys. Rev. D* **103**, 044006, DOI: [10.1103/physrevd.103.044006](https://doi.org/10.1103/physrevd.103.044006) (2021). [arXiv:2011.09494].
- [182] Whittle, P. Curve and periodogram smoothing. *Journal of the Royal Statistical Society. Series B (Methodological)* **19**, 38–63 (1957).
- [183] Vitale, S. *et al.* Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the advanced detectors era. *Phys. Rev. D* **85**, 064034, DOI: [10.1103/physrevd.85.064034](https://doi.org/10.1103/physrevd.85.064034) (2012). [arXiv:1111.3044].
- [184] Thrane, E. & Talbot, C. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *PASA* **36**, e010, DOI: [10.1017/pasa.2019.2](https://doi.org/10.1017/pasa.2019.2) (2019). [arXiv:1809.02293].
- [185] Callister, T. A. A Thesaurus for Common Priors in Gravitational-Wave Astronomy. *arXiv e-prints* (2021). [arXiv:2104.09508].

- [186] Szekeres, B., Pártay, L. B. & Mátyus, E. Direct Computation of the Quantum Partition Function by Path-Integral Nested Sampling. *J. Chem. Theory Comput.* **14**, 4353–4359, DOI: [10.1021/acs.jctc.8b00368](https://doi.org/10.1021/acs.jctc.8b00368) (2018). [arXiv:1804.05987].
- [187] Sciortino, F., Kob, W. & Tartaglia, P. Thermodynamics of supercooled liquids in the inherent-structure formalism: a case study. *J. Phys.: Condens. Matt.* **12**, 6525–6534, DOI: [10.1088/0953-8984/12/29/324](https://doi.org/10.1088/0953-8984/12/29/324) (2000). [arXiv:cond-mat/9911062].
- [188] Wales, D. J. Surveying a complex potential energy landscape: Overcoming broken ergodicity using basin-sampling. *Chemical Physics Letters* **584**, 1–9, DOI: [10.1016/j.cplett.2013.07.066](https://doi.org/10.1016/j.cplett.2013.07.066) (2013).
- [189] Wales, D. J. *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- [190] Wales, D. J. & Bogdan, T. V. Potential energy and free energy landscapes. *J. Phys. Chem. B* **110**, 20765–20776, DOI: [10.1021/jp0680544](https://doi.org/10.1021/jp0680544) (2006).
- [191] Becker, O. M. & Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **106**, 1495–1517, DOI: [10.1063/1.473299](https://doi.org/10.1063/1.473299) (1997).
- [192] Wales, D. J., Miller, M. A. & Walsh, T. R. Archetypal energy landscapes. *Nature* **394**, 758–760, DOI: [10.1038/29487](https://doi.org/10.1038/29487) (1998).
- [193] Wales, D. J. The energy landscape as a unifying theme in molecular science. *Phil. Trans. R. Soc. A* **363**, 357–377, DOI: [10.1098/rsta.2004.1497](https://doi.org/10.1098/rsta.2004.1497) (2005).
- [194] Krivov, S. V. & Karplus, M. Free energy disconnectivity graphs: Application to peptide models. *J. Chem. Phys.* **117**, 10894–10903, DOI: [10.1063/1.1517606](https://doi.org/10.1063/1.1517606) (2002).
- [195] Evans, D. A. & Wales, D. J. Free energy landscapes of model peptides and proteins. *J. Chem. Phys.* **118**, 3891–3897, DOI: [10.1063/1.1540099](https://doi.org/10.1063/1.1540099) (2003).
- [196] Li, Z. & Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615, DOI: [10.1073/pnas.84.19.6611](https://doi.org/10.1073/pnas.84.19.6611) (1987).
- [197] Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116, DOI: [10.1021/jp970984n](https://doi.org/10.1021/jp970984n) (1997). [arXiv:cond-mat/9803344].
- [198] Wales, D. J. & Scheraga, H. A. Global optimization of clusters, crystals and biomolecules. *Science* **285**, 1368–1372, DOI: [10.1126/science.285.5432.1368](https://doi.org/10.1126/science.285.5432.1368) (1999).
- [199] Martiniani, S., Stevenson, J. D., Wales, D. J. & Frenkel, D. Superposition enhanced nested sampling. *Phys. Rev. X* **4**, 031034, DOI: [10.1103/physrevx.4.031034](https://doi.org/10.1103/physrevx.4.031034) (2014). [arXiv:1402.6306].
- [200] Bolhuis, P. G. & Csányi, G. Nested transition path sampling. *Phys. Rev. Lett.* **120**, 250601, DOI: [10.1103/physrevlett.120.250601](https://doi.org/10.1103/physrevlett.120.250601) (2018).
- [201] Murray, I., MacKay, D. J. C., Ghahramani, Z. & Skilling, J. Nested sampling for Potts models. In Weiss, Y., Schölkopf, B. & Platt, J. (eds.) *Advances in Neural Information Processing Systems*, vol. 18 (MIT Press, 2005).
- [202] Pfeifenberger, M. J., Rumetshofer, M. & von der Linden, W. Nested sampling, statistical physics and the Potts model. *Journal of Computational Physics* **375**, 368–392, DOI: [10.1016/j.jcp.2018.08.049](https://doi.org/10.1016/j.jcp.2018.08.049) (2018).
- [203] Pártay, L. B., Bartók, A. P. & Csányi, G. Nested sampling for materials: The case of hard spheres. *Phys. Rev. E* **89**, 022302, DOI: [10.1103/physreve.89.022302](https://doi.org/10.1103/physreve.89.022302) (2014). [arXiv:1208.1721].
- [204] Wilson, B. A., Gelb, L. D. & Nielsen, S. O. Nested sampling of isobaric phase space for the direct evaluation of the isothermal-isobaric partition function of atomic systems. *J. Chem. Phys.* **143**, 154108, DOI: [10.1063/1.4933309](https://doi.org/10.1063/1.4933309) (2015).
- [205] Bartók, A. P., Hantal, G. & Pártay, L. B. Insight into liquid polymorphism from the complex phase behavior of a simple model. *Phys. Rev. Lett.* **127**, 015701, DOI: [10.1103/physrevlett.127.015701](https://doi.org/10.1103/physrevlett.127.015701) (2021). [arXiv:2103.03406].
- [206] Dorrell, J. & Pártay, L. B. Pressure–Temperature Phase Diagram of Lithium, Predicted by Embedded Atom Model Potentials. *J. Phys. Chem. B* **124**, 6015–6023, DOI: [10.1021/acs.jpcc.0c03882](https://doi.org/10.1021/acs.jpcc.0c03882) (2020).
- [207] Pártay, L. B. On the performance of interatomic potential models of iron: comparison of the phase diagrams. *Comput. Mater. Sci* **149**, 153–157, DOI: [10.1016/j.commatsci.2018.03.026](https://doi.org/10.1016/j.commatsci.2018.03.026) (2018). [arXiv:1803.08400].
- [208] Deringer, V. L. *et al.* Gaussian process regression for materials and molecules. *Chemical Reviews* **121**, 10073–10141, DOI: [10.1021/acs.chemrev.1c00022](https://doi.org/10.1021/acs.chemrev.1c00022) (2021). PMID: 34398616.
- [209] Rosenbrock, C. W. *et al.* Machine-learned interatomic potentials for alloys and alloy phase diagrams. *npj Comp. Mat.* **7**, 24, DOI: [10.1038/s41524-020-00477-2](https://doi.org/10.1038/s41524-020-00477-2) (2021). [arXiv:1906.07816].
- [210] Pártay, L. B., Csányi, G. & Bernstein, N. Nested sampling for materials. *Eur. Phys. J. B* **94**, 159, DOI: [10.1140/epjb/s10051-021-00172-1](https://doi.org/10.1140/epjb/s10051-021-00172-1) (2021).

- [211] Wilkinson, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3**, DOI: 10.1038/sdata.2016.18 (2016).
- [212] Hergt, L. T., Handley, W. J., Hobson, M. P. & Lasenby, A. N. Bayesian evidence for the tensor-to-scalar ratio r and neutrino masses m_ν : Effects of uniform vs logarithmic priors. *Phys. Rev. D* **103**, 123511, DOI: 10.1103/physrevd.103.123511 (2021). [arXiv:2102.11511].
- [213] Alsing, J. & Handley, W. Nested sampling with any prior you like. *MNRAS* **505**, L95–L99, DOI: 10.1093/mnras/505/lab057 (2021). [arXiv:2102.12478].
- [214] Murray, I. *Advances in Markov chain Monte Carlo methods*. Ph.D. thesis, University College London (2007).
- [215] Riley, T. E. *Neutron star parameter estimation from a NICER perspective*. Ph.D. thesis, Anton Pannekoek Institute for Astronomy (2019).
- [216] Schittenhelm, D. & Wacker, P. Nested Sampling And Likelihood Plateaus. *arXiv e-prints* (2020). [arXiv:2005.08602].
- [217] Fowlie, A., Handley, W. & Su, L. Nested sampling with plateaus. *MNRAS* **503**, 1199–1205, DOI: 10.1093/mnras/stab590 (2021). [arXiv:2010.13884].
- [218] Lewis, A. Efficient sampling of fast and slow cosmological parameters. *Phys. Rev. D* **87**, 103529, DOI: 10.1103/physrevd.87.103529 (2013). [arXiv:1304.4473].
- [219] Lewis, A. & Bridle, S. Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D* **66**, 103511, DOI: 10.1103/physrevd.66.103511 (2002). [arXiv:astro-ph/0205436].
- [220] Chen, X., Hobson, M., Das, S. & Gelderblom, P. Improving the efficiency and robustness of nested sampling using posterior repartitioning. *Statistics and Computing* **29**, 835–850, DOI: 10.1007/s11222-018-9841-3 (2019). [arXiv:1803.06387].
- [221] Chen, X., Feroz, F. & Hobson, M. Bayesian automated posterior repartitioning for nested sampling. *arXiv e-prints* (2019). [arXiv:1908.04655].
- [222] MacKay, D. J. C. *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, 2003).
- [223] Betancourt, M. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv e-prints* (2018). [arXiv:1701.02434].
- [224] Bellman, R. *Adaptive Control Processes: A Guided Tour*. Princeton Legacy Library (Princeton University Press, 1961).
- [225] Applebaum, D. *Probability and Information: An Integrated Approach* (Cambridge University Press, 2008).
- [226] Dawid, A. P. The Well-Calibrated Bayesian. *Journal of the American Statistical Association* **77**, 605–610, DOI: 10.1080/01621459.1982.10477856 (1982).
- [227] Cook, S. R., Gelman, A. & Rubin, D. B. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* **15**, 675–692, DOI: 10.1198/106186006x136976 (2006).
- [228] Talts, S., Betancourt, M., Simpson, D., Vehtari, A. & Gelman, A. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv e-prints* (2018). [arXiv:1804.06788].
- [229] Robert, C. & Casella, G. A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science* **26**, 102–115, DOI: 10.1214/10-sts351 (2011). [arXiv:0808.2902].
- [230] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092, DOI: 10.1063/1.1699114 (1953).
- [231] Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 97–109, DOI: 10.1093/biomet/57.1.97 (1970).
- [232] Kirkwood, J. G. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* **3**, 300–313, DOI: 10.1063/1.1749657 (1935).
- [233] Kirkpatrick, S., Gelatt, C. D. G. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680, DOI: 10.1126/science.220.4598.671 (1983).
- [234] Higson, E., Handley, W., Hobson, M. & Lasenby, A. Sampling errors in nested sampling parameter estimation. *Bayesian Analysis* **13**, 873–896, DOI: 10.1214/17-ba1075 (2018). [arXiv:1703.09701].
- [235] Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216, DOI: 10.1016/0370-2693(87)91197-x (1987).
- [236] Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362, DOI: 10.1038/s41586-020-2649-2 (2020).
- [237] Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* **9**, 90–95, DOI: 10.1109/mcse.2007.55 (2007).
- [238] Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272, DOI: 10.1038/s41592-019-0686-2 (2020).

References in tables

- [239] Akrami, Y., Scott, P., Edsjo, J., Conrad, J. & Bergstrom, L. A Profile Likelihood Analysis of the Constrained MSSM with Genetic Algorithms. *JHEP* **04**, 057, DOI: [10.1007/jhep04\(2010\)057](https://doi.org/10.1007/jhep04(2010)057) (2010). [arXiv:0910.3950].
- [240] Feroz, F., Cranmer, K., Hobson, M., de Austri, R. R. & Trotta, R. Challenges of Profile Likelihood Evaluation in Multi-Dimensional SUSY Scans. *JHEP* **06**, 042, DOI: [10.1007/jhep06\(2011\)042](https://doi.org/10.1007/jhep06(2011)042) (2011). [arXiv:1101.3296].
- [241] Brewer, B. J. & Foreman-Mackey, D. DNest4: Diffusive Nested Sampling in C++ and Python. *Journal of Statistical Software* **86**, 1–33, DOI: [10.18637/jss.v086.i07](https://doi.org/10.18637/jss.v086.i07) (2018). [arXiv:1606.03757].
- [242] Walter, C. Point process-based Monte Carlo estimation. *Statistics and Computing* **27**, 219–236, DOI: [10.1007/s11222-015-9617-y](https://doi.org/10.1007/s11222-015-9617-y) (2015). [arXiv:1412.6368].
- [243] Birge, J. R., Chang, C. & Polson, N. G. Split Sampling: Expectations, Normalisation and Rare Events. *arXiv e-prints* (2013). [arXiv:1212.0534].
- [244] Fowlie, A., Hoof, S. & Handley, W. Nested Sampling for Frequentist Computation: Fast Estimation of Small p-Values. *Phys. Rev. Lett.* **128**, 021801, DOI: [10.1103/physrevlett.128.021801](https://doi.org/10.1103/physrevlett.128.021801) (2022). [arXiv:2105.13923].
- [245] Mukherjee, P., Parkinson, D. & Liddle, A. R. A Nested Sampling Algorithm for Cosmological Model Selection. *Ap. J.* **638**, L51–L54, DOI: [10.1086/501068](https://doi.org/10.1086/501068) (2006). [arXiv:astro-ph/0508461].
- [246] Parkinson, D., Mukherjee, P. & Liddle, A. R. Bayesian model selection analysis of WMAP3. *Phys. Rev. D* **73**, 123523, DOI: [10.1103/physrevd.73.123523](https://doi.org/10.1103/physrevd.73.123523) (2006). [arXiv:astro-ph/0605003].
- [247] Feroz, F. & Hobson, M. P. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS* **384**, 449–463, DOI: [10.1111/j.1365-2966.2007.12353.x](https://doi.org/10.1111/j.1365-2966.2007.12353.x) (2008). [arXiv:0704.3704].
- Popularized NS in astrophysics and cosmology by introducing the MultiNest implementation**
- [248] Feroz, F., Hobson, M. P. & Bridges, M. MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS* **398**, 1601–1614, DOI: [10.1111/j.1365-2966.2009.14548.x](https://doi.org/10.1111/j.1365-2966.2009.14548.x) (2009). [arXiv:0809.3437].
- [249] Corsaro, E. & De Ridder, J. DIAMONDS: a new Bayesian nested sampling tool. In *European Physical Journal Web of Conferences*, vol. 101 of *European Physical Journal Web of Conferences*, 06019, DOI: [10.1051/epj-conf/201510106019](https://doi.org/10.1051/epj-conf/201510106019) (2015). [arXiv:1509.08311].
- [250] Barbary, K. nestle: Pure Python, MIT-licensed implementation of nested sampling algorithms for evaluating Bayesian evidence. <https://github.com/kbarbary/nestle>.
- [251] Williams, M. J. nessai: Nested Sampling with Artificial Intelligence, DOI: [10.5281/zenodo.4550693](https://doi.org/10.5281/zenodo.4550693) (2021).
- [252] Williams, M. J., Veitch, J. & Messenger, C. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D* **103**, 103006, DOI: [10.1103/physrevd.103.103006](https://doi.org/10.1103/physrevd.103.103006) (2021). [arXiv:2102.11056].
- [253] Handley, W. J., Hobson, M. P. & Lasenby, A. N. POLYCHORD: next-generation nested sampling. *MNRAS* **453**, 4384–4398, DOI: [10.1093/mnras/stv1911](https://doi.org/10.1093/mnras/stv1911) (2015). [arXiv:1506.00171].
- [254] Higson, E. dyPolyChord: dynamic nested sampling with PolyChord. *The Journal of Open Source Software* **3**, 965, DOI: [10.21105/joss.00965](https://doi.org/10.21105/joss.00965) (2018).
- [255] Veitch, J. *et al.* Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D* **91**, 042003, DOI: [10.1103/physrevd.91.042003](https://doi.org/10.1103/physrevd.91.042003) (2015). [arXiv:1409.7215].
- [256] Trassinelli, M. Bayesian data analysis tools for atomic physics. *Nucl. Instrum. Meth. B* **408**, 301–312, DOI: [10.1016/j.nimb.2017.05.030](https://doi.org/10.1016/j.nimb.2017.05.030) (2017). [arXiv:1611.10189].
- [257] Trassinelli, M. The Nested_fit Data Analysis Program. *Proceedings* **33**, DOI: [10.3390/proceedings2019033014](https://doi.org/10.3390/proceedings2019033014) (2019).
- [258] Trassinelli, M. & Ciccodicola, P. Mean shift cluster recognition method implementation in the nested sampling algorithm. *Entropy* **22**, DOI: [10.3390/e22020185](https://doi.org/10.3390/e22020185) (2020).
- [259] Veitch, J. *et al.* cpnest, DOI: [10.5281/zenodo.592884](https://doi.org/10.5281/zenodo.592884) (2017).
- [260] Baldock, R. J. N., Bernstein, N., Salerno, K. M., Pártay, L. B. & Csányi, G. Constant-pressure nested sampling with atomistic dynamics. *Phys. Rev. E* **96**, 43311–43324, DOI: [10.1103/physreve.96.043311](https://doi.org/10.1103/physreve.96.043311) (2017). [arXiv:1710.11085].
- [261] Moss, A. Accelerated Bayesian inference using deep learning. *MNRAS* **496**, 328–338, DOI: [10.1093/mnras/staa1469](https://doi.org/10.1093/mnras/staa1469) (2020). [arXiv:1903.10860].
- [262] Brewer, B. J., Pártay, L. B. & Csányi, G. Diffusive nested sampling. *Statistics and Computing* **21**, 649–656, DOI: [10.1007/s11222-010-9198-8](https://doi.org/10.1007/s11222-010-9198-8) (2010). [arXiv:0912.2380].
- [263] Kester, D. & Mueller, M. BayesicFitting, a PYTHON Toolbox for Bayesian Fitting and Evidence Calculation. *Astronomy and Computing* **37**, 100503, DOI: [10.1016/j.ascom.2021.100503](https://doi.org/10.1016/j.ascom.2021.100503) (2021). [arXiv:2109.11976].

- [264] Speagle, J. S. DYNesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS* **493**, 3132–3158, DOI: [10.1093/mnras/staa278](https://doi.org/10.1093/mnras/staa278) (2020). [[arXiv:1904.02180](https://arxiv.org/abs/1904.02180)].
- [265] Buchner, J. UltraNest - a robust, general purpose Bayesian inference engine. *The Journal of Open Source Software* **6**, 3001, DOI: [10.21105/joss.03001](https://doi.org/10.21105/joss.03001) (2021). [[arXiv:2101.09604](https://arxiv.org/abs/2101.09604)].
- [266] Albert, J. G. JAXNS: a high-performance nested sampling package based on JAX. *arXiv e-prints* (2020). [[arXiv:2012.15286](https://arxiv.org/abs/2012.15286)].