Nested sampling with plateaus

Andrew Fowlie¹*, Will Handley²,³†, and Liangliang Su¹‡

¹Department of Physics and Institute of Theoretical Physics, Nanjing Normal University, Nanjing, Jiangsu 210023, China
²Astrophysics Group, Cavendish Laboratory, J.J.Thomson Avenue, Cambridge, CB3 0HE, UK
³Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK

ABSTRACT

It was recently shown by Riley (2019); Schittenhelm & Wacker (2020) that in the presence of plateaus in the likelihood function nested sampling (NS) produces faulty estimates of the evidence and posterior densities. After informally explaining the cause of the problem, we present a modified version of NS that handles plateaus and can be applied retrospectively to NS runs from popular NS software using anesthetic. In the modified NS, live points in a plateau are evicted one by one without replacement, with ordinary NS compression of the prior volume after each eviction but taking into account the dynamic number of live points. The live points are replenished once all points in the plateau are removed. We demonstrate it on a number of examples. Since the modification is simple, we propose that it becomes the canonical version of Skilling’s NS algorithm.

Key words: methods: statistical – methods: data analysis – methods: numerical

1 INTRODUCTION

Nested sampling (NS) (Skilling 2004; Skilling 2006) is a popular algorithm for Bayesian inference in cosmology, astrophysics and particle physics. The algorithm handles multimodal and degenerate problems, and returns weighted samples for parameter inference as well as an estimate of the Bayesian evidence for model comparison. It was recently emphasised independently by Riley (2019); Schittenhelm & Wacker (2020), however, that assumptions in NS are violated by plateaus in the likelihood, that is regions of parameter space that share the same likelihood. Although we should not be surprised by subtleties caused by ties in the likelihood as NS is based on order statistics, this particular problem was not fully appreciated prior to Riley (2019); Schittenhelm & Wacker (2020).

In physics problems, plateaus most often occur when regions of a model’s parameter space make predictions that are in such severe disagreement with observations that they are assigned a likelihood of zero. For example, in particle physics, parameter points that fail to predict electroweak symmetry breaking would be vetoed and in cosmology portions of parameter space may be excluded if they result in unphysically large power spectra for the purposes of applying lensing, or if it is impossible to trace a consistent cosmic history (see section XIII.C of Hergt et al. 2020 for more detail). We generically refer to such points as unphysical and the observation that renders them unphysical as \( U \). In fact, within popular implementations of NS, such as MultiNest (Feroz & Hobson 2008; Feroz et al. 2009, 2013) and PolyChord (Handley et al. 2015a; Handley et al. 2015b), unphysical points can be assigned a likelihood of zero or a prior of zero through the \( \logZero \) setting. Any log likelihoods below it are treated as if the prior were in fact zero. Statistically, this makes a difference to the evidences and Bayes factors, as it changes whether we consider \( U \) to be prior knowledge or information with which we are updating, i.e. whether we wish to compute \( p(D \mid M, U) \) or \( p(D, U \mid M) \), where \( D \) represents experimental data and \( M \) represents a model. The latter is problematic within NS. We consider both cases interesting: on the one hand, taking a basic observation, e.g., electroweak symmetry breaking, as prior knowledge is reasonable, but, on the other, so is judging models by their ability to predict a basic observation. Although the problem with plateaus can in general lead to faulty posterior distributions as well, when plateaus occur only at zero likelihood, they do not impact posterior inferences about the parameters of the model. There are, furthermore, realistic situations in which plateaus could occur at non-zero likelihood, e.g., if in some regions of parameter space, the likelihood function or the physical observables on which it depends are approximated by a constant.

In Riley (2019); Schittenhelm & Wacker (2020), the problem caused by plateaus was formally demonstrated. After reviewing the relevant aspects of NS in section 2, in section 3, we instead make an informal explanation of the problem. We continue in section 4 by proposing a modified NS algorithm that deals with plateaus and reduces to ordinary NS in their absence. We show examples in section 5 before concluding in section 6. An implementation of the modified NS algorithm that can be used to correct evidences and posteriors found from MultiNest and PolyChord runs is implemented in anesthetic starting from version 2.0.0-beta.2 (Handley 2019).
2 NESTED SAMPLING

To establish our notation and the assumptions in ordinary NS, let us briefly review the NS algorithm. NS works by computing evidence integrals,

\[ Z = \int \mathcal{L}(\Theta) \pi(\Theta) \, d\Theta, \]

where \( \mathcal{L}(\Theta) \) is the so-called likelihood function for the relevant experimental data and \( \pi(\Theta) \) is the prior density for the model’s parameters, as Riemann sums of a one-dimensional integral,

\[ Z = \int_0^1 \mathcal{L}(X) \, dX, \]

where

\[ X(\lambda) = \int_{\mathcal{L}(\Theta) > \lambda} \pi(\Theta) \, d\Theta, \]

is the prior volume contained within the iso-likelihood contour at \( \lambda \) and \( \mathcal{L}(X) \) is the inverse of \( X(\lambda) \), i.e., \( \mathcal{L}(X(\lambda)) = \lambda \). This evidently requires that such an inverse exists over the range of integration.

To tackle the one-dimensional integral, we first sample \( n_{\text{live}} \) points from the prior—the live points. Then, at each iteration of NS, we remove the live point with the worst likelihood \( \mathcal{L}^* \) and replace it with one drawn from the prior subject to the constraint that \( \mathcal{L} > \mathcal{L}^* \). Thus we remove a sequence of samples of increasing likelihood \( \mathcal{L}_i \). In NS we estimate \( X_i = X(\mathcal{L}_i) \) by the properties of how that sequence was generated. Indeed, at each iteration the volume contracts by a factor \( t \), where the arithmetic and geometric expectations of \( t \) are

\[ \langle t \rangle = \frac{n_{\text{live}}}{n_{\text{live}} + 1}, \]

\[ \langle \ln t \rangle = -\frac{1}{n_{\text{live}}}. \]

We can then make a statistical estimate of the prior volume at the \( i \)-th iteration, \( X_i = e^{-\langle \ln t \rangle} n_{\text{live}} \). This enables us to compute

\[ Z \approx \sum \mathcal{L}(X_{i-1} - X_i). \]

The estimates in eq. (4) assume that the live points are uniformly distributed in \( X \). In Fowlie et al. (2020) we proposed a technique for testing the veracity of this assumption within the context of a numerical implementation such as MultiNest or PolyChord. In the following section we discuss why plateaus violate that assumption.

3 PLATEAUS

In Skilling (2006), plateaus were recognised as a problem, since they offer no guidance about the existence or location of points of greater likelihood and since they cause ambiguity in the ranking of samples by likelihood. The latter problem was addressed by breaking ties by assigning a rankable second property to each live point, that is expected to be unique, such as a cryptographic hash of the parameter coordinates or just a random variate from some pre-specified distribution. This was implemented by extending the likelihood to include that tie-breaking second property, \( t \), suppressed by a tiny factor, \( \epsilon \), so that it doesn’t impact the evidence estimate,

\[ \mathcal{L} \rightarrow \mathcal{L} + \epsilon t. \]

It was not stated explicitly that plateaus in fact violate an assumption in NS, as shown by Riley (2019); Schittenhelm & Wacker (2020), though it seems likely that it was known.

In the presence of plateaus, the outermost contour could contain more than one point with the same likelihood. When we replace one of the points in the plateau by a point with a greater likelihood, the volume cannot contract at all, since the outermost contour still contains other points at the same likelihood. Once we’ve replaced all the points in the plateau, the volume finally contracts. Crucially, however, it was not possible for any of the replacement points to affect the contraction, as the replacement points could never be the worst point whilst points in the plateau remain in the live points. The latter subtlety changes statistical estimates of the volume. Without plateaus, it is possible that a replacement point is or soon becomes the worst point, slowing the volume contraction. Without that possibility, the volume contracts faster and thus ordinary NS underestimates volume contraction and overestimates the evidence when there are plateaus.

The problem is illustrated by fig. 1. We show a Gaussian likelihood with mean \( \mu = 0.5 \) and standard deviation \( \sigma = 0.25 \) (orange) and a modified Gaussian likelihood with a plateau in its tails at \( L_P = 0.5 \) (blue). In each case we consider a flat prior for the parameter \( x \) from 0 to 1. The upper right panel shows the prior distribution of the likelihood. The plateau manifests as an atom in that distribution at \( L_P \). The lower left panel shows the resulting enclosed prior volume as a function of the likelihood. The plateau causes a jump discontinuity at \( L_P \). Finally, the lower right panel shows the distribution of \( X(L) \). The plateau leads to an inaccessible region between about 0.8 and 1 and an atom of probability mass at \( X(0.5) \approx 0.8 \), and so \( X \) is not uniformly distributed.

Indeed, in ordinary NS, if \( q \) outermost live points were in a plateau, we would compress by

\[ e^{-\frac{q}{n_{\text{live}}}}. \]

We should, however, compress by about

\[ 1 - \frac{q}{n_{\text{live}}}, \]

which is an unbiased estimate of the volume outside the plateau based on binomial statistics (see section 4.1 for further discussion). The compressions are only similar for \( q \leq 0.5 n_{\text{live}} \), since

\[ e^{-\frac{q}{n_{\text{live}}}} \approx 1 - \frac{q}{n_{\text{live}}} + O \left( \frac{q^2}{n_{\text{live}}} \right). \]

The breakdown in the NS compression in eq. (8) is shown in fig. 2. Note that this problem doesn’t impact importance nested sampling (Feroz et al. 2013), since it does not use the estimated volumes in eq. (8).

In fact, the arguments above show that in the presence of plateaus the inverse of \( X(\lambda) \), denoted \( \mathcal{L}(X) \) in overloaded notation, does not exist for all \( 0 \leq X \leq 1 \) (see the lower right panel in fig. 1). As shown in Schittenhelm & Wacker (2020), in this case we should instead consider the generalised inverse

\[ \mathcal{L}(X) \equiv \{ \sup \lambda \in \text{Im} \mathcal{L} : X(\lambda) > X \}, \]

with which the evidence may be written as

\[ Z = \int_0^1 \mathcal{L}(X) \, dX. \]

We now introduce a modified NS algorithm that correctly computes the evidence even in the presence of plateaus via eq. (12).
Figure 1. Infographic showing the impact of plateaus on assumptions in NS. We show an ordinary Gaussian (orange) and a modified Gaussian with plateaus in the tails (blue). In the lower right panel, we see that the distribution of $X$ is no longer uniform, breaking assumptions in NS.

After removing a point, the number of live points drops by one, such that if we were to remove $i = 1, \ldots, q$ such points (i.e., if there were $q$ points in the plateau) we would compress by

$$\prod_{i=1}^{q} \frac{n_{\text{live}}}{n_{\text{live}} + 1} \cdot \frac{n_{\text{live}} - 1}{n_{\text{live}}} \cdot \frac{n_{\text{live}} - (q - 1)}{n_{\text{live}} - (q - 2)} = 1 - \frac{q}{n_{\text{live}} + 1} \tag{13}$$

if using an arithmetic estimate of the compression, and by

$$\sum_{i=1}^{q} (\ln t_i) = -\sum_{i=1}^{q} \frac{1}{n_{\text{live}} - (i - 1)} \tag{14}$$

$$\approx \ln \left(1 - \frac{q}{n_{\text{live}}} \right) \quad \text{for } n_{\text{live}} \gg q \tag{15}$$

if using a geometric one. Thus we find in both cases that the compression from removing $q$ points would be about $1 - q/n_{\text{live}}$, in agreement with eq. (9), the difference being of order $1/n_{\text{live}}$. The difference is noteworthy when $q \approx n_{\text{live}}$, in which case the unbiased estimate of the contraction would be zero, but the above estimates are about $1/n_{\text{live}}$.

In algorithm 1 and algorithm 2 we show the original and our modified NS, respectively. For concreteness we show the geometric estimator of the compression. We highlight the parts of the algorithm that are changed in red. The simple difference is that whereas in the

4 MODIFIED NS ALGORITHM

In our modification to NS, we remove all live points at the contour $L = L^*$ one by one without replacement, contracting the volume after each removal. If there is a plateau, there may be more than one such point; if not, our algorithm reduces to ordinary NS. After removing all such points, we finally replenish the live points by adding points sampled from the constrained prior, as usual.
In the latter, the number of live points, in estimates of the compression factor from these two approaches. The size of the plateau. Let us consider more carefully the differences when dealing with plateaus, Schittenhelm & Wacker (2020) consider estimating the beta distributions for the compression factor. When dealing with taking into account the changing number of live points. This assumes that it is in the spirit of dynamic NS (Higson et al. 2019), since and thatitisinthespiritofdynamicNS(Higsonetal.2019),sincetherebycomputingaposteriordistribution, \( p(t|q) \propto \text{size}^{-q} (1-t)^q \), (18) for \( q \) points in the plateau and \( n_{\text{live}} - q \) points above it. We could make inferences about \( t \) by computing a posterior distribution, \( p(t|q) \propto \text{size}^{-q} (1-t)^q \), (18) where \( p(t) \) is our prior for the size of the non-plateau region. For \( p(t) = \text{const.} \) this is in fact the probability density for a \( t \sim \text{Beta}(n_{\text{live}} + 1 - q, q + 1) \) distribution.

On the other hand, taking the approach in modified NS, the number of points that lie in the plateau \( q \) is no longer treated as a random variable. Instead, the factor \( t \) is assumed to follow a beta distribution when \( q \) points are removed (Henderson & Goggans 2014),

\[
t \sim \text{Beta}(n_{\text{live}} + 1 - q, q).
\]

The density for \( t \) is

\[
p(t|q) \propto \text{size}^{-q} (1-t)^q \]

(20) corresponding to \( q - 1 \) points below \( \mathcal{L}^* \), \( n_{\text{live}} - q \) points above it, and one point at \( \mathcal{L}^* \).

The fact that we must consider \( q - 1 \) points below \( \mathcal{L}^* \) and one point at \( \mathcal{L}^* \), rather than \( q \) points inside a plateau, results in a factor of \( (1-t)q \) difference between eqs. (18) and (20). A further factor of \( p(t) \) originates from any prior knowledge about the size of the plateau. If the factors may be neglected inferences based on ordinary NS compression may be reliable.

With a flat prior for the unknown size of the plateau, the difference is that between a Beta\( (n_{\text{live}} + 1 - q, q) \) and a Beta\( (n_{\text{live}} + 1 - q, q+1) \) distribution. As shown in fig. 4, the first two moments are similar, with only moderate differences of order \( 1/n_{\text{live}} \) even when

### 4.1 Distribution of compression factor

In our modified NS, we apply ordinary NS compression factors taking into account the dynamic number of live points. This assumes beta distributions for the compression factor. When dealing with plateaus, Schittenhelm & Wacker (2020) consider estimating the compression using the fact that the number of points inside the plateau should follow a binomial distribution parameterized by the size of the plateau. Let us consider more carefully the differences in estimates of the compression factor from these two approaches. In the latter, the number of live points, \( q \), that fall in the outermost contour plateau follows a binomial

\[
q \sim \text{Binom}(n_{\text{live}} - 1 - t)
\]

(16) where \( 1-t \) is the size of the plateau and thus \( t \) is the compression factor. The probability mass function is thus

\[
P(q|t) \propto t^{q-1} (1-t)^{(1-q)}
\]

(17) for \( q \) points in the plateau and \( n_{\text{live}} - q \) points above it. We could make inferences about \( t \) by computing a posterior distribution, \( p(t|q) \propto \text{size}^{-q} (1-t)^q \), (18) where \( p(t) \) is our prior for the size of the non-plateau region. For \( p(t) = \text{const.} \) this is in fact the probability density for a \( t \sim \text{Beta}(n_{\text{live}} + 1 - q, q + 1) \) distribution.

On the other hand, taking the approach in modified NS, the number of points that lie in the plateau \( q \) is no longer treated as a random variable. Instead, the factor \( t \) is assumed to follow a beta distribution when \( q \) points are removed (Henderson & Goggans 2014),

\[
t \sim \text{Beta}(n_{\text{live}} + 1 - q, q).
\]

The density for \( t \) is

\[
p(t|q) \propto \text{size}^{-q} (1-t)^q \]

(20) corresponding to \( q - 1 \) points below \( \mathcal{L}^* \), \( n_{\text{live}} - q \) points above it, and one point at \( \mathcal{L}^* \).

The fact that we must consider \( q - 1 \) points below \( \mathcal{L}^* \) and one point at \( \mathcal{L}^* \), rather than \( q \) points inside a plateau, results in a factor of \( (1-t)q \) difference between eqs. (18) and (20). A further factor of \( p(t) \) originates from any prior knowledge about the size of the plateau. If the factors may be neglected inferences based on ordinary NS compression may be reliable.

With a flat prior for the unknown size of the plateau, the difference is that between a Beta\( (n_{\text{live}} + 1 - q, q) \) and a Beta\( (n_{\text{live}} + 1 - q, q+1) \) distribution. As shown in fig. 4, the first two moments are similar, with only moderate differences of order \( 1/n_{\text{live}} \) even when

---

**Algorithm 1:** Original NS.

Sample \( n_{\text{live}} \) points from the prior — the live points;

Let \( P \) be the set of live points;

Set \( X_0 = 1 \);

Initialise evidence, \( Z = 0 \);

Initialise \( i = 0 \);

while \( i \leq \text{number of iterations} \) do

Let \( L^* \) be the minimum \( L \) of the live points;

Let \( r \) be the live point with \( L > L^* \);

Increment iteration, \( i = i + 1 \);

Contract volume, \( X_i = X_{i-1} \cdot \exp (1/\text{size}(P)) \);

Assign importance weight, \( w_i = (X_{i-1} - X_i) \cdot L^* \);

Increment evidence, \( Z = Z + w_i \);

Remove the point \( r \) from the live points;

Add a new live point sampled from the prior subject to \( L > L^* \);

end

return Estimate of evidence, \( Z \)

---

**Algorithm 2:** Modified NS.

Sample \( n_{\text{live}} \) points from the prior — the live points;

Let \( P \) be the set of live points;

Set \( X_0 = 1 \);

Initialise evidence, \( Z = 0 \);

Initialise \( i = 0 \);

while \( i \leq \text{number of iterations} \) do

Let \( L^* \) be the minimum \( L \) of the live points;

Let \( R \) be the set of live points with \( L = L^* \);

foreach point \( r \) in \( R \) do

Increment iteration, \( i = i + 1 \);

Contract volume, \( X_i = X_{i-1} \cdot \exp (1/\text{size}(P)) \);

Assign importance weight, \( w_i = (X_{i-1} - X_i) \cdot L^* \);

Increment evidence, \( Z = Z + w_i \);

Remove the point \( r \) from the live points;

end

Add \( n_{\text{new}} = \text{size}(R) \) new live points sampled from the prior subject to \( L > L^* \);

end

return Estimate of evidence, \( Z \)
\( q \approx 1 \) and \( q \approx n_{\text{live}} \).\(^1\) The factor vanishes entirely for a logarithmic prior for the unknown size of the plateau,

\[
p(t) \propto \frac{1}{1-t} \quad \text{and so} \quad p(f) \propto \frac{1}{f}.
\]

Thus we see that our modified NS treatment is approximately the same as the binomial treatment, the full details of which would depend on a prior for the size of the plateau. If the differences are important, though, we could instead use the binomial statistics, as discussed at the end of section 4, and a more careful treatment of the prior.

### 4.2 Error estimates

The classic NS error estimate

\[
\Delta \log Z \approx \sqrt{\frac{H}{n_{\text{live}}}},
\]

where \( H \) is the Shannon entropy between the posterior and prior, assumes the ordinary NS compression and a constant number of live points, and so is not applicable to our modified NS. The anesthetic error estimates, however, already account for a dynamic number of live points, and so are applicable to the modified NS algorithm. The anesthetic error estimates are found (as initially suggested by Skilling 2006) through simulations; sequences of possible compression factors are drawn from beta-distributions and used to make a set of estimates \( \ln Z \). One could alternatively compute an analytic estimate using the equivalents of the arithmetic expressions in eq. (4) for the variance, which can be found in Appendix B of Handley et al. (2015b). Future versions of anesthetic will also support such analytic estimates.

If the plateau at \( L_P \) makes up a large fraction \( f \) of the current prior volume, the error in the estimated compression could be substantial, as few points lie outside the plateau. Indeed, we show that the fractional error in the compression blows up when \( q \approx n_{\text{live}} \) in fig. 4. In such cases, there may exist efficient schemes for dynamically increasing the number of live points to ensure that sufficient points lie above the plateau. For example, in algorithm 2 we must ultimately replenish the live points such that there are \( n_{\text{live}} \) live points at \( L > L_P \). We could instead dynamically increase the number of live points immediately prior to the plateau by sampling them from \( L \geq L_P \), stopping at \( n_{\text{dynamic}} \) once \( n_{\text{live}} \) of the \( n_{\text{dynamic}} \) points lie at \( L > L_P \). Once we remove points from the plateau, there would be \( n_{\text{live}} \) live points remaining, and no need to top up the live points.

### 5 EXAMPLES

We now consider a few examples. First, in section 5.1 we apply our modified NS to examples considered in Schittenhelm & Wacker (2020). Second, in section 5.2 we construct a ‘wedding-cake’ function that exhibits a series of plateaus and check that our modified NS correctly computes the evidence.

---

\(^1\) The fact that the distributions as functions of \( t \) are approximately identical is enough to ensure that our inferences are approximately identical (Berger & Wolpert 1988). This holds despite the fact that in the binomial case the number of live points in the outermost plateau is a random variable and the size of the plateau \( 1-t \) is fixed, and in the beta case the compression factor \( t \) is a random variable and \( q \) is fixed. This needn’t be true for frequentist estimates of the factor \( t \).
but the remaining evidence isn’t accounted for, we underestimate the evidence by a term
\[
\Delta \log Z = \log (Z + \Delta Z) - \log Z
\]  
where \(\Delta Z = f \max L\)

\[\text{(24)}\]

\[\text{(25)}\]

\[\text{where } f \text{ is the fraction of the prior volume occupied by the plateau.}\]

\[\text{For example, in Scenario 3 of Schittenhelm & Wacker (2020), the plateau occupies } f = 0.161 \text{ of the prior volume at the peak of the likelihood at } \max \log L \approx -2.21, \text{ and we find } \Delta \log Z \approx 1.006, \text{ in good agreement with the difference found numerically in Schittenhelm & Wacker (2020), which was 1.007.}\]

### 5.2 Wedding cake likelihood

We now construct a challenging but semi-analytical example for which we can numerically confirm our approach. Consider an infinite sequence of concentric square plateaus of geometrically decreasing volume \(a^i(1-a)\) for \(i = 0, 1, 2, \ldots\). The edges of the plateaus lie at

\[
r_i = a^i/D/2 \quad \text{and} \quad r = |x - 1/2|_0
\]

\[\text{(26)}\]

for \(i = 0, 1, 2, \ldots\), where \(|x|_0 = \max_j (|x_j|)\) denotes the infinity norm. We define the height of each plateau to have a Gaussian profile:

\[
\log \mathcal{L} = -\sum_{i=0}^{\infty} \frac{r_i^2}{2\sigma^2} I_{r_{i+1} < r < r_i}
\]

\[\text{(27)}\]

where \(I\) is an indicator function that, for any given \(r\), selects a single term in the sum. The resulting likelihood is therefore a set of hypercuboids with a hypercubic base of side length \(r_i\) and height \(\exp(-r_i^2/(2\sigma^2))\). If the base is two-dimensional, this creates a tiered “wedding cake” surface, as can be seen in fig. 6. The \(i\) selected by the indicator function is in fact,

\[
i(r) = \lfloor D \log_2 2r \rfloor
\]

\[\text{(28)}\]

where \(\lfloor y \rfloor\) the floor function (namely the greatest integer less than or equal to \(y\)), enabling us to write

\[
\log \mathcal{L} = -\frac{a^{2i(r)/D}}{8\sigma^2}
\]

\[\text{(29)}\]

Given that the volume of the region \([r_i < r < r_{i-1}]\) is \(a^i(1-a)\), the evidence can be expressed as:

\[
Z = \sum_{i=0}^{\infty} e^{-a^{2i(D)/8\sigma^2}} a^i(1-a)
\]

\[\text{(30)}\]

which as an infinite series converges sufficiently rapidly and stably to be evaluated numerically for any number of dimensions \(D\), but if speed is a requirement then a Laplace approximation shows that one only needs to consider the terms in the series around

\[
i \sim \sqrt{\frac{D \log(4D\sigma^2)^2 - 1}{\log a}}.
\]

\[\text{(31)}\]

For reference, putting all of these equations together, the likelihood can be computed as:

\[
\log \mathcal{L}(x) = -\frac{a^{2i(D \log_2 2|x-1/2|_0)/D}}{8\sigma^2}
\]

\[\text{(32)}\]

\[\text{where } a \text{ is a hyperparameter controlling the depth of the plateaus, and } \sigma \text{ controls the width of the overall Gaussian profile.}\]

This likelihood forms part of the aesthetic test suite, which confirms that the approach suggested in this paper recovers the true likelihood.

The wedding cake likelihood can be very useful for testing nested sampling implementations as unlike a traditional Gaussian it can be trivially sampled from using a simple random number generator, has no unexpected edge effects as the boundaries of the prior are also a likelihood contour.

### 6 CONCLUSIONS

Following from Riley (2019); Schittenhelm & Wacker (2020), which showed formally that NS breaks down if there are plateaus in the likelihood, we first presented the problem of plateaus in an informal but accessible way. We then constructed a modified version of the NS algorithm. The simple modification permits it to remove points in a plateau one by one, without replacement. Once all points in a plateau are removed, the live points are finally replenished. This leads to correct compression in the presence of plateaus and ordinary NS in their absence.

We discussed examples from Schittenhelm & Wacker (2020), shedding light on them by finding the impact of plateaus on ordinary NS in simple analytic formulae. The impact was previously shown only numerically. Lastly, our especially constructed wedding-cake problem showed a particularly pathological case, with multiple plateaus. The modified NS algorithm successfully dealt with them.

Runs from popular NS software such as PolyChord and Multi
Nest may be resummed retrospectively via the modified NS algorithm using anesthetic starting from version 2.0.0-beta.2. Our modified NS makes a minimal change to ordinary NS, to the extent that we recommend it becomes the canonical version of NS.

ACKNOWLEDGEMENTS
AF was supported by an NSFC Research Fund for International Young Scientists grant 11950410509. WH was supported by a George Southgate visiting fellowship grant from the University of Adelaide, Gonville & Caius College, and STFC IPS grant number ST/T001054/1.

DATA AVAILABILITY
There is little data associated with this paper, though any data or code will be shared on reasonable request to the corresponding author.

REFERENCES
Berger J., Wolpert R., 1988, The Likelihood Principle. Lecture notes – monographs series Vol. 6, Institute of Mathematical Statistics
Feroz F., Hobson M. P., 2008, Mon. Not. Roy. Astron. Soc., 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, Mon. Not. Roy. Astron. Soc., 398, 1601
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2013, The Open Journal of Astrophysics
Fowlie A., Handley W., Su L., 2020, Mon. Not. Roy. Astron. Soc., 497, 5256
Handley W., 2019, J. Open Source Softw., 4, 1414
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, Mon. Not. Roy. Astron. Soc., 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, Mon. Not. Roy. Astron. Soc., 453, 4384
Henderson R. W., Goggans P. M., 2014, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering. pp 100–105, doi:10.1063/1.4903717
Hergt L., Agocs F., Handley W., Hobson M., Lasenby A., 2020, (In preparation)
Higson E., Handley W., Hobson M., Lasenby A., 2019, Statistics and Computing, 29, 891
Riley T., 2019, PhD thesis, Neutron star parameter estimation from a NICER perspective, https://hdl.handle.net/11245.1/aa86fcf3-2437-4bc2-810e-cf9f30a98f7a
Schittenhelm D., Wacker P., 2020, arXiv e-prints, p. arXiv:2005.08602
Skilling J., 2004, in Fischer R., Preuss R., Toussant U. V., eds, American Institute of Physics Conference Series Vol. 735, American Institute of Physics Conference Series. pp 395–405, doi:10.1063/1.1835238
Skilling J., 2006, Bayesian Analysis, 1, 833

This paper has been typeset from a TeX/LaTeX file prepared by the author.