*Proceeding Paper*

# SuperNest: Accelerated Nested Sampling Applied to Astrophysics and Cosmology †

Aleksandr Petrosyan [1,2,3,*] and Will Handley [1,2,4]

1   Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK
2   Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK
3   Queens' College, University of Cambridge, Silver Street, Cambridge CB3 9ET, UK
4   Gonville & Caius College, University of Cambridge, Trinity Street, Cambridge CB2 1TA, UK
*   Correspondence: ap886@cam.ac.uk
†   Presented at the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, 18–22 July 2022.

**Abstract:** We present a method for improving the performance of nested sampling as well as its accuracy. Building on previous work we show that posterior repartitioning may be used to reduce the amount of time nested sampling spends in compressing from prior to posterior if a suitable "proposal" distribution is supplied. We showcase this on a cosmological example with a Gaussian posterior, and release the code as an LGPL licensed, extensible Python package supernest.

**Keywords:** Bayesian inference

## 1. Introduction

Nested sampling is a multi-purpose tool invented by John Skilling [1] for simultaneous probabilistic integration of high-dimensional distributions and drawing weighted samples. These are essential tasks for the performance of Bayesian model comparison and parameter estimation [2]. It has become widely used across the physical sciences from astrophysics and cosmology through biochemistry and materials science to high energy physics [3].

The performance of nested sampling algorithms, however, suffer in comparison with alternative posterior samplers such as Metropolis–Hastings [4], in particular, when the latter are provided with a pre-trained proposal distribution and start point. This paper allows an apples-to-apples comparison, by providing a mechanism for accelerating nested sampling using proposal distributions. This is achieved using an extension of the posterior repartitioning techniques developed by Chen et al. [5].

In Section 2, we establish notation and definitions. In Section 3, we review and extend posterior repartitioning. In Section 4, we present our chosen approach and in Section 5, we showcase this on a cosmology example before concluding in Section 6.

## 2. Background

### 2.1. Notation

In this section, we establish notation and recall the key terms in nested sampling and Bayesian inference. A probabilistic inference model $\mathcal{M}$ with a vector of tunable parameters $\theta$ for data $D$ is specified by *likelihood* $\mathcal{L}(\theta) \equiv P(D|\theta, \mathcal{M})$ and *prior* $\pi(\theta) = P(\theta|\mathcal{M})$. The two-step challenge of inference is to compute the *posterior* $\mathcal{P}(\theta) \equiv P(\theta|D, \mathcal{M})$ and *evidence* $\mathcal{Z} = P(D|\mathcal{M})$. Constructing a representation of the posterior is termed *parameter estimation*, and evaluating an estimate of the evidence comprises *model comparison*. The four distributions are linked via Bayes theorem:

$$\mathcal{L}(\theta) \times \pi(\theta) = \mathcal{Z} \times \mathcal{P}(\theta), \tag{1}$$

where both sides of the equation equal to the joint distribution $P(D, \theta)$.

It is useful to define the *Kullback-Leibler divergence* $\mathcal{D}$ [6] between the prior $\pi$ and the posterior $\mathcal{P}$:

$$\mathcal{D}_\pi\{\mathcal{P}\} = \int \mathcal{P}(\theta) \ln \frac{\mathcal{P}(\theta)}{\pi(\theta)} d\theta. \tag{2}$$

One can understand the properties of the KL divergence as a logarithmic ratio of the volume contained in the prior distribution to the volume contained in the posterior distribution:

$$\mathcal{D}_\pi\{\mathcal{P}\} \approx \ln \frac{V_\pi}{V_\mathcal{P}}. \tag{3}$$

When both the prior and posterior are uniform (i.e., constant value where they are non-zero), this relation is exact (one can think of the posterior averaging process in Equation (2) as an information theory inspired "smoothing" for non top-hat distributions).

### 2.2. Nested Sampling

In this section, we will provide a brief primer on the nested sampling algorithm. For a survey of established nested sampling packages and applications, readers are recommended the Nature primer [3]. We begin by initialising $n_{\text{live}}$ initial "live points" drawn from the prior $\pi$. Then at each iteration, the lowest likelihood live point is marked "dead" and replaced with a new live point drawn from the prior, that is at a higher likelihood. The compression procedure allows one to estimate the density of states, which given the recorded locations of live and dead points leads to estimates of posterior and evidence with respective errors.

Crucially, the algorithm terminates once the running estimate of the error in the evidence is greater than a set fraction of the total estimated evidence. Thus, the time complexity of nested sampling is:

$$T \propto n_{\text{live}} \times \langle \mathcal{T}\{\mathcal{L}(\theta)\} \rangle \times \langle \mathcal{T}\{\text{Impl.}\} \rangle \times \mathcal{D}_\pi\{\mathcal{P}\}, \tag{4}$$

where $\mathcal{T}\{\mathcal{L}(\theta)\}$ is the time complexity of evaluating $\mathcal{L}(\theta)$ and $\mathcal{T}\{\text{Impl.}\}$ is the implementation-specific factor representing the time complexity of replacing a dead point with a live point with a higher likelihood, and is the dominant factor for nested sampling over parameter spaces with many dimensions [3]. For rejection nested sampling (e.g., `MultiNest` [7]) this factor is exponential in the number of tunable parameters $d$ (with a turnover constant in tens of dimensions), and linear in $d$ for chain-based approaches (e.g., `PolyChord` [8]). The Kullback–Leibler divergence $\mathcal{D}_\pi\{\mathcal{P}\}$ is linear in the $d$ for parameters that are well constrained, as can be seen from Equation (3) by noticing that volumes in parameter space scale with a power of $d$.

From Equation (4), it follows that the run time of nested sampling, given a resolution fixed by $n_{\text{live}}$ can be reduced by (i) speeding up the likelihood evaluation; (ii) improving the replacement efficiency; (iii) speeding up the prior to posterior compression. The final approach is the one adopted for the *repartitioning techniques* in this paper.

The accuracy of the evidence is associated with the accumulated Poisson noise, such that the error in $\ln \mathcal{Z}$ is [1]:

$$\sigma \propto \sqrt{\mathcal{D}_\pi\{\mathcal{P}\}/n_{\text{live}}}. \tag{5}$$

Thus, faster compression leads to better accuracy and precision. For a fixed accuracy $\sigma$, we can adjust $n_{\text{live}} \propto \mathcal{D}/\sigma^2$, which means $T \propto \mathcal{D}^2/\sigma^2$, given Equation (4). In precision-normalised cases, reducing the effect of the KL divergence term results in quadratic improvements to the speed of nested sampling. Note that the number of live points can be also adjusted. The minimum number is dictated by the quantity of modes in the multi-modal posterior, and implementation dependent considerations: for example, `PolyChord` needs $n_{\text{live}} \sim \mathcal{O}(d)$ to generate enough dead point phantoms for an accurate covariance computation. `MultiNest` needs at least $n_{\text{live}} \sim \mathcal{O}(100)$ to compute ellipsoidal decompositions.

### 2.3. Historical Overview of Acceleration Attempts

One approach to partially achieve this form of speed-up is to perform two passes of nested sampling. First, a low resolution pass (few live points) over the entire original prior volume $V_{\pi,1}$ finds the location of the highest concentration of posterior volume, and an evidence estimate $\mathcal{Z}_1$. Next, one constructs a much tighter bounding prior $\pi_2$ with volume $V_{\pi,2}$ around the largest concentration of the posterior. Trivially:

$$\mathcal{D}_{\pi,1}\{\mathcal{P}\} < \mathcal{D}_{\pi,2}\{\mathcal{P}\} \tag{6}$$

hence, latter inference over the tighter box will finish faster finding the evidence $\mathcal{Z}_2$. With a changed prior, this evidence will not be the same as the original. However, for box-priors we can compensate the volumes geometrically, using the fact that the overall evidence satisfies:

$$\mathcal{Z} = \mathcal{Z}_1 = \mathcal{Z}_2 \frac{V_{\pi,2}}{V_{\pi,1}} \tag{7}$$

See Anstey et al. [9] for examples of this method's application in real astronomical cases.

## 3. Posterior Repartitioning

Nested sampling is unusual amongst Bayesian numerical algorithms, in that it distinguishes between the prior $\pi(\theta)$ and likelihood $\mathcal{L}(\theta)$. Metropolis–Hastings, and Hamiltonian Monte Carlo, for example, are techniques which are only sensitive; $\mathcal{L} \times \pi$ (i.e., the joint or unnormalised posterior) is considered for the purposes of acceptance/rejection. Nested sampling, on the other hand, by "sampling from the prior $\pi$, subject to a hard likelihood constraint $\mathcal{L}$", separates the two functions.

Chen et al. [5] pioneered an approach which exploits this split, noting that we have the freedom to define a new prior $\tilde{\pi}$ and likelihood $\tilde{\mathcal{L}}$, which providing they obey the relation:

$$\tilde{\mathcal{L}}(\theta)\tilde{\pi}(\theta) = \mathcal{L}(\theta)\pi(\theta) \tag{8}$$

will recover the same posterior and evidence. Conceptually, one can think of this as transferring portions of the likelihood into the prior, or vice versa. A cosmological example of this subtle rearranging of portions of likelihood into prior can be found in many codes which implement nested sampling such as `cobaya` [10], `cosmomc` [11,12], `MontePython` [13] or `cosmosis` [14] in the manner in which they consider Gaussian priors. Codes can either make the choice to incorporate this as part of the prior (typically as an inverse-error-function-based unit hypercube transformation), or as a density term added to the rest of the log-likelihood calculations.

This rearranging affects the Kullback–Leibler divergence. The difference can be found using a posterior average:

$$\mathcal{D}_{\tilde{\pi}}\{\mathcal{P}\} = \mathcal{D}_{\pi}\{\mathcal{P}\} + \left\langle \ln \frac{\tilde{\pi}}{\pi} \right\rangle_{\mathcal{P}} \tag{9}$$

If the final term in Equation (9) is negative, repartitioning provides a mechanism for reducing the time to convergence by transferring a reasonable "guess": the proposal distribution, from likelihood to prior, reducing the Kullback–Leibler divergence between the prior $\tilde{\pi}$ and the posterior $\tilde{\mathcal{P}} = \mathcal{P}$. As we have seen before, this both improves performance and increases precision.

### 3.1. Example 1: Power Posterior Repartitioning

The scheme considered in Chen et al. [5] is *power posterior repartitioning* (PPR):

$$\tilde{\pi}(\theta|\beta) = \frac{\pi(\theta)^{\beta}}{z_{\beta}}, \qquad \tilde{\mathcal{L}}(\theta|\beta) = \mathcal{L}(\theta)\pi(\theta)^{1-\beta}z_{\beta}, \qquad z_{\beta} = \int \pi(\theta)^{\beta} d\theta. \tag{10}$$

For $\beta = 0$, any distribution $\pi$ becomes uniform. If a distribution $\pi$ is not uniform, raising it to a power of $\beta$ serves to sharpen or spread the peaks, depending on whether $\beta$ is greater or smaller than 1, in analogy with the thermodynamic inverse temperature of a Boltzmann distribution [3]. For any given value of $\beta$, a nested sampling run with likelihood $\tilde{\mathcal{L}}$ and prior $\tilde{\pi}$ recovers the same posterior samples and evidence as a run with $\mathcal{L}$ and prior $\pi$.

Chen et al. [5] considered this in the context of a geophysical example. The original intention behind repartitioning was not to improve performance, but rather as a way of making nested sampling more resilient to over-zealously specified priors (e.g., if a user specifies a Gaussian prior which is narrow and/or distant from the peak of the likelihood). Still, the performance implications are valuable. Substituting Equation (10) into Equation (9) shows that varying $\beta$ will change the time to convergence as:

$$T(\beta) \propto \mathcal{D}_{\tilde{\pi}}\{\mathcal{P}\} = \mathcal{D}_{\pi}\{\mathcal{P}\} + \left\langle \ln \frac{\pi^{\beta-1}}{z_\beta} \right\rangle_{\mathcal{P}} \tag{11}$$

Whether or not this increases or decreases the speed of nested sampling depends on the choice of $\beta$ and to what extent the sharpened prior overlaps with the posterior. Equation (10) is valid for any choice of $\beta$, and while any specific choice can be useful, Chen et al. [5] demonstrated that still greater power comes from considering it as an additional parameter to sample over, effectively extending the parameter space to $\tilde{\theta} = (\theta, \beta)$ with a prior on $\beta$. A posterior value of the parameter $\beta$ close to 0 therefore indicates that the prior was poorly chosen, i.e., that there is significant *tension* between prior and posterior.

One could consider this for the purposes of acceleration by allowing $\beta$ to take a range wider than $[0, 1]$, for example, by using an exponential prior $P(\beta) = e^{-\beta}$ $[0 \le \beta < \infty]$. There is, however, no need to restrict ourselves to the case of power posterior repartitioning, or indeed to a priors $\tilde{\pi}$, which are direct transformations of the original $\pi$.

### 3.2. Example 0: Replacement Repartitioning

The simplest repartitioning scheme is to simply use a different prior $\pi_*$, but compensate for this with the adjusted likelihood:

$$\tilde{\pi}(\theta) = \pi_*(\theta), \qquad \tilde{\mathcal{L}}(\theta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{\pi_*(\theta)}. \tag{12}$$

From Equation (9), we can see that if $\pi_*$ is close to $\mathcal{P}$ as quantified by the KL divergence, nested sampling terminates sooner (with $\mathcal{D}_{\tilde{\pi}}\{\mathcal{P}\} = 0$ in the limiting case $\tilde{\pi} = \mathcal{P}$). However, a poorly specified proposal $\pi_*$ will pessimise the choices of new live points, thus making the problem harder. Indeed, a poorly chosen proposal can prevent us gaining any usable information.

### 3.3. Example 2: Additive Superpositional Repartitioning

Another repartitioning scheme is to consider an additive superposition of the original prior $\pi$ with another normalised distribution $\pi_*$:

$$\tilde{\pi}(\theta|\beta) = (1-\beta)\pi(\theta) + \beta\pi_*(\theta), \qquad \tilde{\mathcal{L}}(\theta|\beta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{(1-\beta)\pi(\theta) + \beta\pi_*(\theta)} \tag{13}$$

Here, $\pi_*$ does not need to be related to the original prior, and could, for example, be a Gaussian proposal distribution (often supplied with cosmological sampling packages for common likelihood combinations), or a tighter box prior as suggested in Section 2.3.

This scheme can be generalised to superpositions of $N$ alternative distributions $\pi_1, \ldots \pi_N$. Consider:

$$\tilde{\pi}(\theta|\beta) = \left(1 - \sum_{i=1}^{N} \beta_i\right)\pi(\theta) + \sum_{i=1}^{N} \beta_i \pi_i(\theta), \qquad \tilde{\mathcal{L}}(\theta|\beta) = \frac{\mathcal{L}(\theta)\pi(\theta)}{(1-\sum_{i=1}^{N}\beta_i)\pi(\theta) + \sum_{i=1}^{N}\beta_i\pi_i(\theta)} \tag{14}$$

However, additive superpositional repartitioning in practice has the issue of poor generalisation to higher dimensions and/or complex priors due to the difficulty of implementing an analytic unit hypercube transformation. Further work is required to fully explore this using nested samplers that do not rely on prior point–percent functions which produce said transformations. We thank Xy Wang for exploring the possibility of these models in the first instance as part of her University of Cambridge master's dissertation.

### 4. Supernest: **Stochastic Superpositional Repartitioning**

The above examples lead us to the proposed repartitioning scheme. Just as with additive repartitioning, we allow, arbitrarily, many unrelated priors to be superimposed. However, we side-step the multivariate point-percent function correction problem by sampling from each of the priors with some probability. We achieve this by adding an integer selection parameter $\beta$, such that:

$$\tilde{\pi}(\theta|\beta) = \begin{cases} \pi(\theta) & : & \beta = 0 \\ \pi_1(\theta) & : & \beta = 1 \\ \vdots & & \vdots \\ \pi_N(\theta) & : & \beta = N \end{cases} \qquad \tilde{\mathcal{L}}(\theta|\beta) = \begin{cases} \mathcal{L}(\theta) & : & \beta = 0 \\ \mathcal{L}(\theta)\pi(\theta)/\pi_1(\theta) & : & \beta = 1 \\ \vdots & & \vdots \\ \mathcal{L}(\theta)\pi(\theta)/\pi_N(\theta) & : & \beta = N \end{cases} \qquad (15)$$

and then $\beta$ is sampled over with a discrete (typically uniform) prior $P(\beta)$. The choice of said prior depends on the confidence in each of the proposals. When applied to nested sampling, at the end of the run, one will recover a (joint) posterior on $\beta$, which will indicate which proposal was chosen more frequently, as the usual Bayesian balance between the corresponding modified likelihood $\tilde{\mathcal{L}}$ and KL divergence $\mathcal{D}_{\tilde{\pi}}\{\mathcal{P}\}$ of the favoured bin.

In the additive case from Section 3.3, the primary difficulty is in obtaining the point-percent function of the repartitioned prior. The more priors are introduced, the harder the calculation. For the stochastic case, the prior of the superposition is the superposition of the individual priors. Thus, if one knows the point percent functions of the constituent priors, simply copying them is the "calculation".

Additionally, only one of the prior quantile/likelihood pairs is evaluated for each point. Thus, the time complexity of the likelihood evaluation of a superposition of multiple likelihoods is $\mathcal{O}(\langle \mathcal{T}(\mathcal{L}) \rangle)$. By contrast, the equivalent time complexity for the additive superposition is the sum of the time complexities: $\mathcal{O}(\sum \mathcal{T}(\mathcal{L}))$. For comparison, if we fed 100 identical models to the stochastic mixture, we would not see a performance regression, while the additive superposition would take 100 times longer to compute.

### 5. Cosmological Example

For the purposes of demonstration, we choose a likelihood $\mathcal{L}$ which is Gaussian in the parameters, with parameter means and parameter covariances as given by the Planck legacy archive chains [15], and prior given by the default prior widths as specified in those chains. This has six cosmological parameters and 21 nuisance parameters associated with galactic foregrounds and satellite calibration, giving 27 parameters to constrain. In its default mode, this therefore gives $n_{\text{live}} = 675$ and $n_{\text{repeats}} = 135$, and for this example an $n_{\text{dead}} \sim 72{,}000$. This provides a proxy for a nested sampling problem of interests to cosmologists, whilst being fast enough numerically to explore a variety of configurations.

We run PolyChord wrapped with supernest, both with their default settings. PolyChord, in addition, is set with a precision criterion of 10%, i.e., stopping when the live points contain 10% of the remaining evidence. As a representative example, we first choose a Gaussian proposal with the same covariance as the parameters, inflated by a factor of $3^d$ in volume, and offset by a random vector drawn from the posterior. This gives a proposal which is conservative (much wider than the posterior), and approximately correctly centered.

The performance of this example is shown in Figure 1. We can see that the algorithm terminates approximately three times faster, with an error bar that is approximately two

times smaller despite the perturbation. Usually to reduce the error bar by a factor of 2, one would need to increase the number of live points by a factor of 4. Thus, when precision-normalised `supernest` achieved an order-of-magnitude $12\times$ improvement over the traditional nested sampling run. Further improvements to this speed are achievable if a more compact (and possibly larger number of) proposal distributions are used.
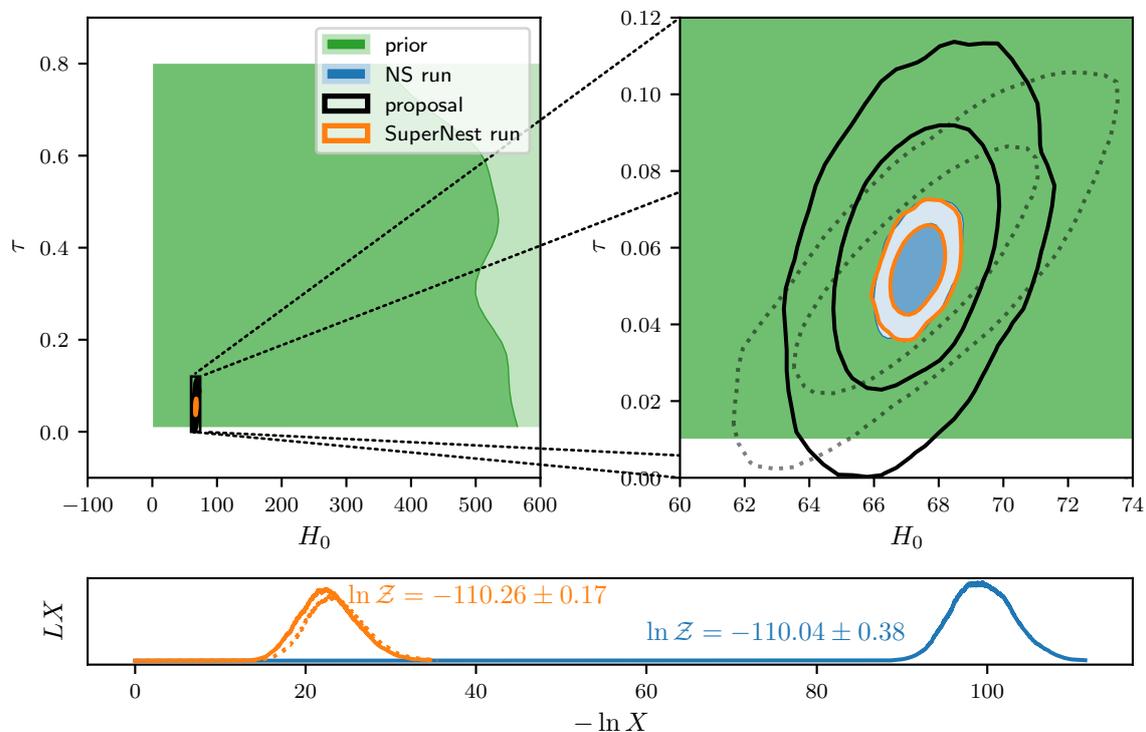


**Figure 1.** `supernest` in action. The top panels show the parameter spaces of the Hubble constant and optical depth to reionisation $(H_0, \tau)$ [15] with all other $\Lambda$CDM and nuisance parameters marginalised out. Contours show the 66% and 95% credibility regions of the prior (green) and nested sampling posterior (blue), alongside the proposal distribution (black) provided to `supernest` which recovers the same posterior (orange) as the original NS run. The left hand plot visualised the prior, whilst the right hand plot zooms in on the posterior for clarity. The lower panel shows the Higson plot [16] of the two runs. Here, the evidences recovered are consistent, and the `supernest` run in orange has a more accurate evidence inference, associated with the fact that the compression in prior volume $\ln X$ is substantially lower, meaning the nested sampling run terminates in approximately a third the time. An alternative proposal distribution with dashed lines is also plotted, which recovers an evidence of $\ln \mathcal{Z} = -110.36 \pm 0.18$, consistent to within sampling error.

As a second example, where the covariance of the proposal does not align with the covariance of the posterior, we draw an inverse Wishart distributed covariance matrix, with $d = 27$ degrees of freedom and scale matrix equal to the covariance, once again inflating the result in volume by $3^d$, and offsetting the result by a random vector drawn from the posterior. The performance is unaffected to within sampling fluctuation, with the recovered evidence equal to within error.

## 6. Conclusions

We presented a method for improving the performance and accuracy of nested sampling using the recent innovations in posterior repartitioning of Chen et al. [5]. The approach functions by reducing the effective prior-to-posterior compression that nested sampling "sees", and hence reducing both the runtime and accumulated Poisson error in the evidence estimate. We discussed a variety of alternative posterior repartitioning strategies before settling on a stochastic superpositional scheme, and demonstrated that on a cosmological

example even conservative proposals have the capacity to speed up nested sampling runs by an order of magnitude.

This performance improvement is achieved by trimming off the long prior-to-posterior compression phase seen in the blue but not orange curve of the lower panel of Figure 1. The remaining majority of the time is spent in traversing the typical set. Accelerating the pace at which nested sampling does this is a matter of ongoing research.

This work will be followed by a re-examination of the repartitioning, in particular, by considering more general cases than the ones considered here. In a follow up study we shall provide a more mathematical look into how further performance improvements can be achieved via modifications to nested samplers, hysteretic priors on the choice parameter, as well as exploring the connection between prior spaces and Hilbert spaces. Finally, we shall consider a reformulation of Bayesian methods using functionals, and explore the implications of *consistent partitioning* (a generalisation of posterior repartitioning) applied to Bayesian inference in functional space.

**Author Contributions:** Conceptualization, A.P. and W.H.; methodology A.P.; software, A.P.; validation, A.P. and W.H.; formal analysis A.P. and W.H.; investigation A.P.; resources W.H.; data curation A.P. and W.H.; writing—original draft preparation A.P.; writing—review and editing W.H.; visualization A.P. and W.H.; supervision W.H.; project demonstration A.P.; funding acquisition W.H. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is available at https://doi.org/10.5281/zenodo.7703906 or at https://gitlab.com/a-p-petrosyan/sspr/.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Anal.* **2006**, *1*, 833–859. [CrossRef]
2. MacKay, D.J.C. Monte-Carlo methods. In *Information Theory, Inference, and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003; Chapter 29, pp. 379–380.
3. Ashton, G.; Bernstein, N.; Buchner, J.; Chen, X.; Csányi, G.; Fowlie, A.; Feroz, F.; Griffiths, M.; Handley, W.; Habeck, M.; et al. Author Correction: Nested sampling for physical scientists. *Nat. Rev. Methods Prim.* **2022**, *2*, 44. [CrossRef]
4. Arminger, G.; Muthén, B.O. A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the metropolis-hastings algorithm. *Psychometrika* **1998**, *63*, 271–300. [CrossRef]
5. Chen, X.; Feroz, F.; Hobson, M. Bayesian automated posterior repartitioning for nested sampling. *arXiv* **2019**, arXiv:1908.04655.
6. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
7. Feroz, F.; Hobson, M.P.; Bridges, M. MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **2009**, *398*, 1601–1614. [CrossRef]
8. Handley, W.J.; Hobson, M.P.; Lasenby, A.N. polychord: Next-generation nested sampling. *Mon. Not. R. Astron. Soc.* **2015**, *453*, 4384–4398. [CrossRef]
9. Anstey, D.; de Lera Acedo, E.; Handley, W. A general Bayesian framework for foreground modelling and chromaticity correction for global 21 cm experiments. *Mon. Not. R. Astron. Soc.* **2021**, *506*, 2041–2058. [CrossRef]
10. Torrado, J.; Lewis, A. Cobaya: Code for Bayesian analysis of hierarchical physical models. *J. Cosmol. Astropart. Phys.* **2021**, *2021*, 57. [CrossRef]
11. Lewis, A.; Bridle, S. Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D* **2002**, *66*, 103511. [CrossRef]
12. Lewis, A. Efficient sampling of fast and slow cosmological parameters. *Phys. Rev. D* **2013**, *87*, 103529. [CrossRef]
13. Brinckmann, T.; Lesgourgues, J. MontePython 3: Boosted MCMC sampler and other features. *Phys. Dark Universe* **2019**, *24*, 100260. [CrossRef]
14. Zuntz, J.; Paterno, M.; Jennings, E.; Rudd, D.; Manzotti, A.; Dodelson, S.; Bridle, S.; Sehrish, S.; Kowalkowski, J. CosmoSIS: Modular cosmological parameter estimation. *Astron. Comput.* **2015**, *12*, 45–59. [CrossRef]

15.　Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* **2020**, *641*, A6. [CrossRef]
16.　Edward Higson and Will Handley and Mike Hobson and Anthony N. Lasenby. Dynamic nested sampling: An improved algorithm for parameter estimation and evidence calculation. *Stat. Comput.* **2018**, *29*, 891–913. [CrossRef]