MDPI

Proceeding Paper

Inferring Evidence from Nested Sampling Data via Information Field Theory [†]

Margret Westerkamp [1,2,*], **Jakob Roth** [1,2,3], **Philipp Frank** [1], **Will Handley** [4,5] **and Torsten Enßlin** [1,2]

1 Max Planck Institute for Astrophysics, 85748 Garching, Germany
2 Faculty for Physics, Ludwig-Maximilians-Universität, 80539 Munich, Germany
3 Department for Computer Science, Technical University Munich, 85748 Munich, Germany
4 Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK
5 Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0EZ, UK
* Correspondence: margret@mpa-garching.mpg.de
† Presented at the 42nd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 3–7 July 2023.

Abstract: Nested sampling provides an estimate of the evidence of a Bayesian inference problem via probing the likelihood as a function of the enclosed prior volume. However, the lack of precise values of the enclosed prior mass of the samples introduces probing noise, which can hamper high-accuracy determinations of the evidence values as estimated from the likelihood-prior-volume function. We introduce an approach based on information field theory, a framework for non-parametric function reconstruction from data, that infers the likelihood-prior-volume function by exploiting its smoothness and thereby aims to improve the evidence calculation. Our method provides posterior samples of the likelihood-prior-volume function that translate into a quantification of the remaining sampling noise for the evidence estimate, or for any other quantity derived from the likelihood-prior-volume function.

Keywords: nested sampling; Bayesian evidence; information field theory

1. Introduction

Nested sampling is a computational technique for Bayesian inference developed by [1]. Whereas previous statistical sampling algorithms were primarily designed to sample the posterior, the nested sampling algorithm focuses on computing the evidence by estimating how the likelihood function relates to the prior. As discussed in [2], Bayesian inference consists of parameter estimation and model comparison. In Bayesian parameter estimation, the model parameters $\theta_{\mathcal{M}}$ for a given model $\mathcal{M}$ and data d are inferred via Bayes' theorem,

$$\mathcal{P}(\theta_{\mathcal{M}}|d,\mathcal{M}) = \frac{\mathcal{P}(d|\theta_{\mathcal{M}},\mathcal{M})\mathcal{P}(\theta_{\mathcal{M}}|\mathcal{M})}{\mathcal{P}(d|\mathcal{M})}. \tag{1}$$

Here, $\mathcal{P}(\theta_{\mathcal{M}}|d,\mathcal{M})$ is the posterior probability for the model parameters $\theta_{\mathcal{M}}$ given the data d. The likelihood $\mathcal{P}(d|\theta_{\mathcal{M}},\mathcal{M})$ describes the measurement process, which generated the data d, and the prior $\mathcal{P}(\theta_{\mathcal{M}}|\mathcal{M})$ encodes our prior knowledge of the parameters within the given model. The normalization of the posterior,

$$Z = \mathcal{P}(d|\mathcal{M}) = \int d\theta_{\mathcal{M}} \mathcal{P}(\theta_{\mathcal{M}},d|\mathcal{M}), \tag{2}$$

is called the evidence and is the focus of this study. In Bayesian parameter estimation, it is common to work with not normalized posteriors. Thus, in this scenario, the computation of the evidence is less critical. In contrast, when comparing different Bayesian models,

estimating the evidence for different models is very important. In this case, the aim is to find the most probable model $\mathcal{M}_i$ given the data,

$$P(\mathcal{M}_i|d) = \frac{P(d|\mathcal{M}_i)P(\mathcal{M}_i)}{P(d)}. \tag{3}$$

Assuming a uniform prior for all arbitrary models $P(\mathcal{M}_i) = \text{const}$, this turns out to be equivalent to choosing the model with the highest evidence.

In nested sampling, the possibly multidimensional integral of the posterior in Equation (2) is transformed into a one-dimensional integral by directly using the prior mass X. In particular, by transforming the problem into a series of nested spaces, nested sampling provides an elegant way to compute the evidence. The algorithm starts by drawing N samples from the prior, called the live points. For each of these points, the likelihood is calculated and the live point with the lowest likelihood is removed from the set of live points and added to another set, called the dead points. A new live point is then sampled that has a higher likelihood value than the last added dead point. This type of sampling is commonly referred to as likelihood-restricted sampling. However, the specific methods associated with likelihood-restricted sampling are not discussed further in this paper. As a consequence of the procedure, the prior volume shrinks from one to zero, contracting around the peak of the posterior. The prior mass X contained in the parameter space volume with likelihood values larger than L can be computed by

$$X(L) = \int_{P(d|\theta_\mathcal{M},\mathcal{M})>L} P(\theta_\mathcal{M}|\mathcal{M}) \, d\theta_\mathcal{M}. \tag{4}$$

Thus, Equation (2) simplifies to a one-dimensional integral,

$$Z = \int_0^1 L(X)dX, \tag{5}$$

where $L(X)$ is the inverse of Equation (4). Accordingly, this integral can be approximated by the weighted sum over all m dead points

$$Z \approx \sum_{i=1}^{m} \omega_i L_i. \tag{6}$$

As proposed in [1], we calculate the weights via $\omega_i = \frac{1}{2}(X_{i-1} - X_{i+1})$ assuming $X_0 = 1$ and $X_{m+1} = 0$. Adding dead points to their set and adjusting the evidence accordingly continues until the remaining live points occupy a tiny prior volume that would contribute little to the weighted sum in Equation (6).

For the calculation in Equation (6) not only the known live and dead contours of the likelihood are needed but also the corresponding prior volumes encoded in ω_i, which are not precisely known. According to [1] there are two different approaches to approximate the prior volumes X_i, a stochastic scheme and a deterministic scheme. In the stochastic scheme the change of volume due to each removed shell i is a stochastic process characterised by a Beta distributed random variable t_i,

$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad P(t_i) = \text{Beta}(t_i|1, N) = Nt_i^{N-1}, \tag{7}$$

where we assume a constant number of live points N. Approaches with a varying number of live points were i.a. introduced in dynamic nested sampling by [3,4] and extend beyond the boundaries of this research until the present moment. This probabilistic description of the prior volume evolution allows to draw several samples of prior volumes X, according to

the likelihood values L, and to thereby get uncertainty estimates on the evidence calculation (Equation (6)). In the deterministic scheme the logarithmic prior volume is estimated via,

$$\ln(X_i) = -\frac{i}{N},$$ (8)

at the ith iteration. This estimate is derived from the fact that the expectation value of the logarithmic volume changes is $\langle \ln t_i \rangle = -1/N$. However, this estimate does not take the uncertainties in the evidence calculation [5] into account and differs from unbiased approaches introduced and analysed in [6–8]. In any case, the imprecise knowledge of the prior volume introduces probing noise that can potentially hinder the accurate calculation of the evidence. In order to improve the accuracy of the posterior integration, we aim to reconstruct the likelihood-prior-volume function given certain a priori assumptions on the function itself using Bayesian inference. Here, we introduce a prior and likelihood model for the reconstruction of the likelihood-prior-volume function, which we will call the reconstruction prior and the reconstruction likelihood to avoid confusion with likelihood contour and prior volume information obtained from nested sampling.

The left side of Figure 1 illustrates the nested sampling likelihood dead contours generated by the software package anesthetic [9] for the simple Gaussian example discussed in Section 2 and two live points (N = 2) as a function of prior volume. In the following, we call the likelihood values of the dead points the likelihood data $\vec{d}_L$ and the prior volume, approximated by Equation (8), the prior volume data $\vec{d}_X$. Additionally, the analytical solution of the likelihood-prior-volume function, which we call the ground truth, is plotted.
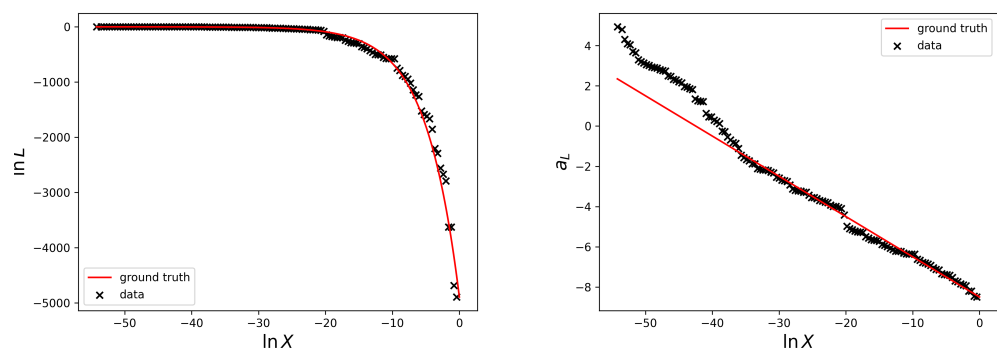

Figure 1. (**Left**): Visualisation of the nested sampling dead point logarithmic likelihoods, $\vec{d}_L$, as a function of logarithmic prior mass data, $\vec{d}_X$, for the normalized simple Gaussian in Equation (14) ($\sigma_X = 0.01$, $D = 10$). The corresponding data was generated by the software package anesthetic [9]. (**Right**): Visualisation of the reparametrised nested sampling dead point logarithmic likelihoods according to Equation (13) as a function of logarithmic prior mass for the same case as shown left.

In accordance with the here considered example, we assume the likelihood-prior-volume function to be smooth for most real-world applications of nested sampling. In this study, we propose an approach that incorporates this assumption of a-priori-smoothness and enforces monotonicity. In particular, we use Information Field Theory (IFT) [10] as a versatile mathematical tool to reconstruct a continuous likelihood-prior-volume function from a discrete dataset of likelihood contours and to impose the prior knowledge on the function.

As noted in [11], the time complexity of the nested sampling algorithm depends on several factors. First, the time complexity depends on the information gain of the posterior over the prior, which is equal to the shrinkage of the prior required to reach the bulk of the posterior. This is described mathematically by the Kullback-Leibler divergence (KL) [12],

$$\mathcal{D}_{\text{KL}} = \int d\theta_{\mathcal{M}} \mathcal{P}(\theta_{\mathcal{M}}|d, \mathcal{M}) \ln \frac{\mathcal{P}(\theta_{\mathcal{M}}|d, \mathcal{M})}{\mathcal{P}(\theta_{\mathcal{M}}|\mathcal{M})}.$$ (9)

Second, the time complexity increases with the number of live points N, which defines the shrinkage per iteration. Furthermore, the time for evaluating the likelihood $L(\theta)$, $\mathcal{T}_L$, and the time for sampling a new live point in the likelihood restricted volume, $\mathcal{T}_{\text{samp}}$, contribute to the time complexity. Accordingly, in [13] the time complexity of the nested sampling algorithm T and the error σ_Z have been characterised via,

$$T \propto N \times \langle \mathcal{T}_L \rangle \times \langle \mathcal{T}_{\text{samp}} \rangle \times \mathcal{D}_{\text{KL}} \tag{10}$$

$$\sigma_Z \propto \sqrt{\mathcal{D}_{\text{KL}}/N}. \tag{11}$$

Upon examining the error, σ_Z, it becomes evident that reducing the error by increasing the number of live points leads to significantly longer execution times. Accordingly, by inferring the likelihood-prior-volume function, we aim to reduce the error in the log-evidence for a given $\mathcal{D}_{\text{KL}}$ and a fixed number of live points, N, avoiding a significant increase in time complexity.

The rest of the paper is structured as follows. In Section 2, the description of the reconstruction prior of the likelihood-prior-volume curve is discussed. The model for the reconstruction likelihood and the inference of the likelihood-prior-volume function and the prior volumes using IFT is described in Section 3. The corresponding results for a Gaussian example and the impact on the evidence calculation are shown in Section 4. And eventually, the conclusion and outlook for future work are given in Section 5.

2. The Reconstruction Prior Model for the Likelihood-Prior-Volume Function

A priori we assume that the likelihood-prior-volume function is smooth and monotonically decreasing. This is achieved by representing the negative rate of change of the logarithmic prior volume, $\ln X$, with a monotonic function of the likelihood a_L as a log-normal process,

$$-\frac{d \ln X}{da_L} = e^{\tau(\ln X)}. \tag{12}$$

In the words of IFT, we model the one-dimensional field τ, which assigns to each logarithmic prior volume a value, as a Gaussian process with $\mathcal{P}(\tau) = \mathcal{G}(\tau, T)$. Thereby, we do not assume a fixed power spectrum for the Gaussian process, but reconstruct it simultaneously with τ itself. An overview of this Gaussian process model is given in Appendix A. The details can be found in [14].

In the most relevant volume for the evidence, the peak region of the posterior is expected to be similar to a Gaussian in a first order approximation. Therefore, the function a_L is chosen such that τ is a constant for the Gaussian case. Deviations from the Gaussian are reflected in deviations of τ from the constant. Accordingly, we define,

$$a_L = -\ln\left(-\ln\left(\frac{L}{L_{\text{max}}}\right)\right), \tag{13}$$

with L_{max} being the maximal likelihood. We consider the simple Gaussian example proposed by [1],

$$L_{\text{gauss}}(X) = \frac{1}{C(\sigma_X)} \exp\left(-\frac{X^{2/D}}{2\sigma_X^2}\right), \tag{14}$$

where D is the dimension and σ_X is the standard deviation. We find that the function $a_L(\ln X)$, defined in Equation (13), becomes linear in this case,

$$a_{\text{gauss}}(\ln X) = \ln(2\sigma_X^2) - \frac{2}{D} \ln X. \tag{15}$$

Figure 1 illustrates the data and the ground truth on log-log-scale on the left and the linear relation $a_L(\ln X)$ on the right. According to the log-normal process defined in

Equation (12), we define the function $a_L(\ln X)$, for arbitrary likelihoods, which is able to account for deviations from the Gaussian case,

$$a_L(\ln X) = a_0 - \int_0^{\ln X} e^{-\tau(z)} dz. \tag{16}$$

By inverting Equation (13) we then get the desired likelihood-prior-volume function. The logarithmic prior volume values given the likelihood contours are obtained by inversion of Equation (16). In Figure 2 several prior samples given the model for the reconstruction prior according to Equation (16) are shown.
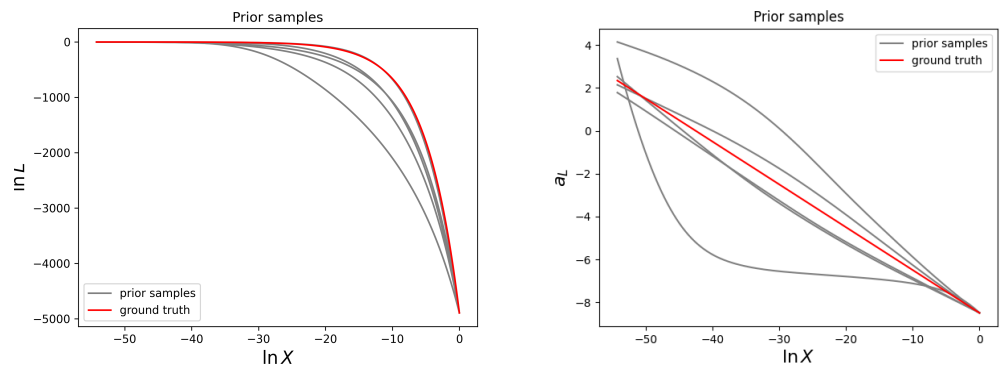

Figure 2. Reconstruction prior samples of the likelihood-prior-volume function plotted together with the ground truth. (**Left**): Log-log-scale. (**Right**): Parametrisation according to Equation (13).

However, often the maximum log-likelihood, $\ln L_{\max}$, is not known. In [15], the calculation of the maximum Shannon entropy $\mathcal{I} = \ln\left(\frac{\mathcal{P}(\theta|d)}{\mathcal{P}(\theta)}\right)$ is given. Using this approach, we can calculate the logarithmic maximum likelihood $\ln L_{\max}$ and thus calculate a_L for unknown likelihoods.

$$\mathcal{I}_{\max} = \mathcal{D}_{\mathrm{KL}} + \frac{D}{2} \tag{17}$$

$$\rightarrow \ln L_{\max} = \langle \ln L \rangle_{\mathcal{P}(\theta_{\mathcal{M}}|d,\mathcal{M})} + \frac{D}{2} \tag{18}$$

Hence, based on the likelihood contours obtained from the nested sampling run, we calculate the data based evidence, Z_d, using the approximated prior volumes according to Equation (8). This allows us to obtain an estimate of the maximum log-likelihood, $\ln L_{\max}$, of the model for reparametrisation,

$$Z_d \approx \sum_{i=1}^{m} d_{L_i} \times (d_{X_{i-1}} - d_{X_i}) \tag{19}$$

$$\rightarrow \ln L_{\max} \approx \sum_{i=1}^{m} \frac{d_{L_i}}{Z_d} (d_{X_{i-1}} - d_{X_i}) \ln d_{L_i} + \frac{D}{2}. \tag{20}$$

3. The Reconstruction Likelihood Model and Joint Inference

In this section we will derive a model for the reconstruction likelihood for the joint inference of the likelihood-prior-volume function and the changes of logarithmic prior volume according to the likelihood data $\vec{d}_L$. Here, IFT and the software package NIFTy [16], which facilitates the implementation of IFT algorithms, allow us to infer the reparametrised likelihood-prior-volume function in Equation (16) from the data,

$$\vec{d}_a = -\ln\left(-\ln\left(\frac{\vec{d}_L}{L_{\max}}\right)\right), \tag{21}$$

given the reconstruction prior and reconstruction likelihood model. For the inference of the likelihood-prior-volume function we first define the likelihood function a_L as a function of the Beta distributed t_i (Equation (7)),

$$a_L(z_j = \sum_{i=1}^{j} t_i) = a_0 - \int_0^{\sum_{i=1}^{j} t_i} e^{-\tau(z)} dz,$$

(22)

where a_0 is the likelihood for $X_0 = 1$. We perform a joint reconstruction of the function a_L and the vector $\vec{t}_d$ representing changes in prior volume according to the likelihood data $\vec{d}_a$.

$$\mathcal{P}(a, \vec{t}_d | \vec{d}_a) \propto \mathcal{P}(\vec{d}_a | \vec{t}_d, a) \, \mathcal{P}(\vec{t}_d, a)$$

(23)

$$= \mathcal{P}(\vec{d}_a | \vec{t}_d, a) \, \mathcal{P}(a | \vec{t}_d) \, \mathcal{P}(\vec{t}_d)$$

$$= \delta(\vec{d}_a - a_L(\vec{t}_d)) \, \mathcal{G}(\tau | T) \prod_{i=1}^{m} \text{Beta}(t_i | 1, N)$$

$$\approx \mathcal{G}(\vec{d}_a - a_L(\vec{t}_d), \sigma_\delta) \, \mathcal{G}(\tau | T) \prod_{i=1}^{m} \text{Beta}(t_i | 1, N)$$

Here, the Gaussian uncertainty σ_δ is supposed to be chosen small in order to approximately represent the δ-function. So far, we have managed to obtain a probabilistic model for the non-normalized reconstruction posterior. To this end, we use variational inference, in particular the geoVI algorithm supposed by [17], to get an approximate of the true reconstruction posterior distribution. In the end, this statistical approach allows us to get an estimate of the likelihood-prior-volume function and the prior volumes via the posterior mean, its uncertainty and any other quantity of interest, which can be derived from the posterior samples. The results of the here developed method are shown in Section 4.

4. Results

To test the presented method we perform a reconstruction for the simple Gaussian example discussed in Section 2 and introduced in Figure 1. The according results for the likelihood-prior-volume function are shown in Figure 3.
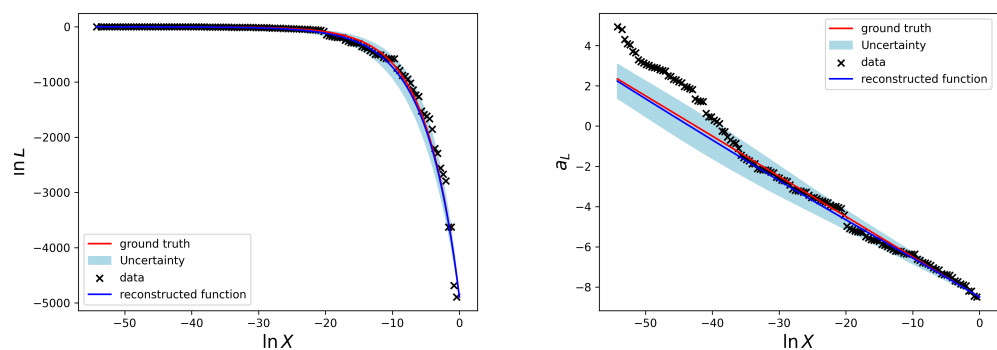

Figure 3. Reconstruction results for the likelihood-prior-volume function for the simple Gaussian example in Equation (14). The plots show the data, the ground truth and the reconstruction as well as its uncertainty. (**Left**): Reconstruction results on log-log-scale. (**Right**): Reconstruction results in reparametrised coordinates according to Equation (13).

Moreover, the posterior estimates of the logarithmic prior volumes to the according likelihood data $\vec{d}_L$ are shown in Figure 4.
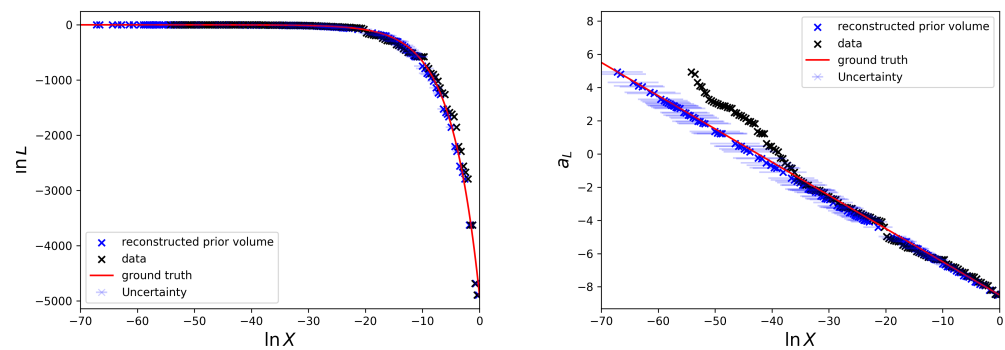
Figure 4. Reconstruction results for the prior volumes given the likelihood data $\vec{d}_L$ for the simple Gaussian example in Equation (14). The plots show the data, the ground truth and the reconstruction as well as its uncertainty. (**Left**): Reconstruction results on log-log-scale. (**Right**): Reconstruction results in reparametrised coordinates according to Equation (13).

Since the main goal of nested sampling is to compute the evidence, we want to quantify the impact of the proposed method on the evidence calculation. To do this, we use $n_{\text{samp}} = 200$ posterior samples for the prior volumes X^* and calculate the evidence given the likelihood contours $\vec{d}_L$ for each of these samples according to Equation (6),

$$Z^* = \sum_{i=1}^{m} (\vec{d}_L)_i \frac{1}{2} (X_{i-1}^* - X_{i+1}^*). \tag{24}$$

Similarly, we generate by means of anesthetic [9] $n_{\text{samp}} = 200$ samples of the prior volume via the probabilistic nested sampling approach described in Equation (7). Also for these samples we calculate the evidence according to Equation (24). A comparison of the histograms of evidences for both sample sets (classical nested sampling and reconstructed prior volumes) is shown in Figure 5.
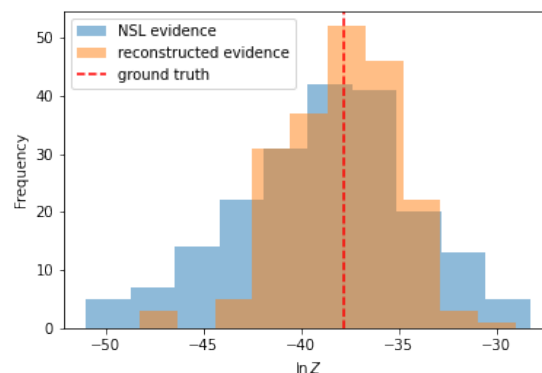

Figure 5. Comparison of histograms for logarithmic evidences for $n_{\text{samp}} = 200$ samples for the classical nested sampling (NSL) approach and the reconstructed prior volumes.

From the comparison of the histograms, one can already see that the standard deviation for the posterior sample evidences for the reconstructed prior volumes got smaller. This is also mirrored as soon as we look at numbers: The ground truth logarithmic evidence for this Gaussian case is $\ln Z_{\text{gauss}} = -37.798$. The result for the evidence for the classical nested sampling approach given $n_{\text{samp}} = 200$ is $\ln Z_d = -38.92 \pm 4.50$. And finally, the result for the evidence inferred with the here presented approach from the likelihood contours assuming smoothness and enforcing monotonicity is $\ln Z = -37.97 \pm 2.89$.

5. Conclusions & Outlook

In our search for a more accurate estimate of the evidence, we set out to reconstruct the likelihood-prior-volume function. In particular, a Bayesian method was developed to infer jointly the likelihood-prior-volume function and the vector of prior volumes from the dead-point likelihood contours given by the nested sampling algorithm. For the reconstruction we enforce monotonicity and assume smoothness. The test of the reconstruction algorithm on a Gaussian example shows a significant improvement in the accuracy of the computed logarithmic evidence.

In general, the approach presented here will only show notable improvements if the assumption of smoothness for the likelihood- prior-volume-curve holds. Fortunately, we can reasonably expect this assumption to hold in the majority of cases. Future work will apply the reconstruction algorithm to further likelihoods where the ground truth likelihood-prior-volume function is known for testing, with the ultimate goal of applying it to actual nested sampling outputs. In particular, the results on non-Gaussian likelihoods shall be tested.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Appendix A. Gaussian Process Model in NIFTy

In NIFTy [16] we represent our reconstruction priors via generative models as described in [18]. More precisely we use the reparametrisation trick by [19] according to [20] to describe the field τ with correlation structure T in Equation (12) as a generative process,

$$\tau = A\xi, \ \ \mathcal{P}(\xi) = \mathcal{G}(\xi, \Xi), \ \ T = AA^{\dagger}. \tag{A1}$$

Under the a priori assumption of statistical homogeneity and isotropy, A becomes diagonal in Fourier space and can be fully represented by the square root of the power spectrum $p_T(|k|)$,

$$A_{kk'} = (FAF^{\dagger})_{kk'} = 2\pi\delta(k - k')\sqrt{p_T(|k|)}, \tag{A2}$$

where F is the corresponding Fourier transformation. Here, we model the logarithmic amplitude spectrum, $\sqrt{p_T(|k|)}$, as a power-law with non-parametric deviations represented by an integrated Wiener process on logarithmic coordinates $l = \ln(|k|)$ according to [14],

$$\sqrt{p_T(l)} \propto e^{\gamma(l)}, \ \ \frac{d^2\gamma}{dl^2} = \eta\xi_W(l), \ \ \mathcal{P}(\xi_W) = \mathcal{G}(\xi_W|\mathbb{1}). \tag{A3}$$

The resulting shape of the power spectrum is encoded in ξ_W, η and additional integration and normalization parameters. These additional parameters are represented by Gaussian and log-normal processes themselves and such their generative prior models are defined by hyperparameters characterising their mean and variance.

References

1. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Anal.* **2006**, *1*, 833–859. [CrossRef]
2. Handley, W.J.; Hobson, M.P.; Lasenby, A.N. polychord: Next-generation nested sampling. *Mon. Not. R. Astron. Soc.* **2015**, *453*, 4385–4399. [CrossRef]
3. Speagle, J.S. dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **2020**, *493*, 3132–3158. [CrossRef]
4. Higson, E.; Handley, W.; Hobson, M.; Lasenby, A. Dynamic nested sampling: An improved algorithm for parameter estimation and evidence calculation. *Stat. Comput.* **2018**, *29*, 891–913. [CrossRef]
5. Keeton, C.R. On statistical uncertainty in nested sampling. *Mon. Not. R. Astron. Soc.* **2011**, *414*, 1418–1426. [CrossRef]
6. Walter, C. Rare Event Simulation and Splitting for Discontinuous Random Variables. *ESAIM Probab. Stat.* **2015**, *19*, 794–811.
7. Chopin, N.; Robert, C.P. Properties of nested sampling. *Biometrika* **2010**, *97*, 741–755. [CrossRef]
8. Salomone, R.; South, L.F.; Drovandi, C.C.; Kroese, D.P. Unbiased and Consistent Nested Sampling via Sequential Monte Carlo. *arXiv* **2018**, arXiv:1805.03924.
9. Handley, W. Anesthetic: Nested sampling visualisation. *J. Open Source Softw.* **2019**, *4*, 1414. [CrossRef]
10. Enßlin, T.A. Information Theory for Fields. *Ann. Phys.* **2019**, *531*, 1800127. [CrossRef]
11. Buchner, J. Nested Sampling Methods. *Stat. Surv.* **2023**, *17*, 169–215.
12. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
13. Petrosyan, A.; Handley, W.J. SuperNest: Accelerated nested sampling applied to astrophysics and cosmology. *Phys. Sci. Forum* **2022**, *5*, 51.
14. Arras, P.; Frank, P.; Haim, P.; Knollmüller, J.; Leike, R.; Reinecke, M.; Enßlin, T. Variable structures in M87 from space, time and frequency resolved interferometry. *Nat. Astron.* **2022**, *6*, 259–269. [CrossRef]
15. Handley, W.; Lemos, P. Quantifying dimensionality: Bayesian cosmological model complexities. *Phys. Rev. D* **2019**, *100*, 023512. [CrossRef]
16. Arras, P.; Baltac, M.; Enßlin, T.A.; Frank, P.; Hutschenreuter, S.; Knollmueller, J.; Leike, R.; Newrzella, M.N.; Platz, L.; Reinecke, M.; et al. *Nifty5: Numerical Information Field Theory v5*; Astrophysics Source Code Library: College Park, MD, USA, 2019; record ascl:1903.008.
17. Frank, P.; Leike, R.; Enßlin, T.A. Geometric Variational Inference. *Entropy* **2021**, *23*, 853. [CrossRef]
18. Enßlin, T. Information Field Theory and Artificial Intelligence. *Entropy* **2022**, *24*, 374. [CrossRef]
19. Kingma, D.P.; Salimans, T.; Welling, M. Variational Dropout and the Local Reparameterization Trick. In Proceedings of the Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, Montreal, QC, Canada, 7–12 December 2015; pp. 2575–2583.
20. Knollmüller, J.; Enßlin T.A. Metric Gaussian Variational Inference. *arXiv* **2020**, arXiv:1901.11033.