Report of Statistics for Data Science

Name: Jingjing Gao
ID Number: jg2098

10 March 2024

1 i

Upon careful examination of Figure 1, combined with fundamental trigonometric principles, I am able to derive a trigonometric connection among α, β, θ, and x:

$$\beta \tan(\theta_k) = x_k - \alpha \tag{1}$$

2 ii

In general, given a random variable x distributed according to Q and an invertible transformation $f : \mathbb{R}^d \to \mathbb{R}^d$, the random variable $y = f(x)$ is distributed as $y \sim P$, where

$$P(y) = \left|\frac{\partial f}{\partial x}\right|^{-1} P(x) \tag{2}$$

The $d \times d$ matrix $\frac{\partial f}{\partial x}$ is the Jacobian of the transformation.
In our case, the uniform probability density function of θ is given by:

$$P(\theta|\alpha, \beta) = \frac{1}{\pi}, \quad -\frac{\pi}{2} < \theta < \frac{\pi}{2} \tag{3}$$

In equation(1), I also have the connection between the angle θ and the position along the coast x. Differentiating both sides of equation(1) with respect to x, I have

$$\beta \sec^2(\theta) \times \frac{d\theta}{dx} \ = 1 \tag{4}$$

Using the trigonometric identity $\tan^2(\theta) + 1 \equiv \sec^2(\theta)$ and substituting for $\tan(\theta)$ from equation (1), the Jacobian can be written as:

$$\frac{d\theta}{dx} = \left[\beta \sec^2(\theta)\right]^{-1} = \left[\beta\left(1 + \tan^2(\theta)\right)\right]^{-1} = \left[\beta\left(1 + \left(\frac{x-\alpha}{\beta}\right)^2\right)\right]^{-1} \tag{5}$$

Finally, we can use equation (2) to transform the PDF for θ in equation (3) to its equivalent form in terms of x; after a little algebraic rearrangement, we obtain the Cauchy distribution of equation (6):

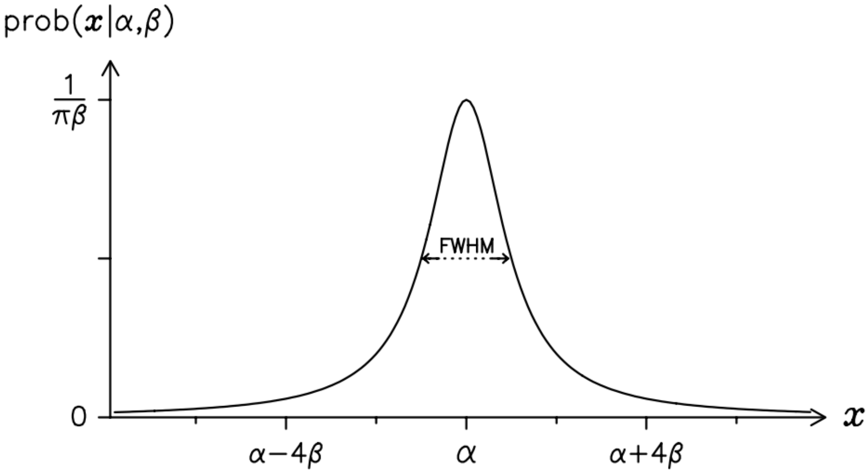
$$\mathcal{L}_x(x|\alpha, \beta) = P(\theta|\alpha, \beta) \times \left|\frac{d\theta}{dx}\right| = \frac{\beta}{\pi\left(\beta^2 + (x-\alpha)^2\right)} \tag{6}$$

3 iii

Equation (6) tells us that the probability of the kth flash being recorded at position x_k, given the coordinates of the lighthouse (α, β), follows a Cauchy distribution. The Cauchy distribution, or Lorentzian distribution, is symmetric about its peak. The most likely location is the point where the probability density function reaches its maximum value.

For this distribution, the maximum probability (or most likely location) for the kth flash to be recorded is at $x_k = \alpha$. This observation is further illustrated in Fig. 1, which plots the Cauchy distribution.

Therefore, the frequentist colleague's point that the most likely location for any flash to be received is $\hat{x} = \alpha$ is correct.



Figuur 1: The Cauchy, or Lorentzian, distribution.

To estimate the lighthouse position, denoted by α, we aim to derive the posterior probability density function (pdf) given the observations $\{x_k\}$ and the parameter β, i.e., $\text{prob}(\alpha|\{x_k\}, \beta)$. Employing Bayes' theorem, the posterior pdf is proportional to the product of the likelihood and the prior:

$$\text{prob}(\alpha|\{x_k\}, \beta) \propto \text{prob}(\{x_k\}|\alpha, \beta) \times \text{prob}(\alpha|\beta) \tag{7}$$

Considering our knowledge of β, which, without the data, provides no new information about the lighthouse position along the shore, we can simplify the prior term:

$$\text{prob}(\alpha|\beta) = \text{prob}(\alpha) \tag{8}$$

Now, let's assume a straightforward uniform pdf for the prior of α:

$$\text{prob}(\alpha|\beta) = \text{prob}(\alpha) = \begin{cases} A & \text{if } \alpha_{\text{min}} < \alpha < \alpha_{\text{max}}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The limits α_{min} and α_{max} may signify known boundaries of the coastline or be set to arbitrarily large values (even up to the size of the Earth) if our knowledge is highly uncertain. The normalization constant A is simply the reciprocal of the prior range.

As each flash is emitted at an independent angle, the likelihood function for a set of flash locations is expressed as the product:

$$\mathcal{L}_x(\{x_k\}|\alpha, \beta) = \prod_k \mathcal{L}_x(x_k|\alpha, \beta) = \prod_k \frac{\beta}{\pi \left(\beta^2 + (x_k - \alpha)^2\right)} \tag{10}$$

Substituting the prior of α from equation (9) and the likelihood function for a set of flash locations resulting from equations (6) and (10) into Bayes' theorem of equation (7) allows us to express the logarithm of the posterior pdf as follows:

$$log_e\left[\text{prob}(\alpha|\{x_k\}, \beta)\right] = constant - \sum_{k=1}^{N} log_e\left[\beta^2 + (x_k - \alpha)^2\right] \tag{11}$$

where the constant includes all terms not involving α.

The best estimate of the most likely location for any flash α_0 is given by the maximum of the posterior pdf; differentiating euation (11) once, with respect to α, we obtain the condition

$$\frac{dlog_e\left[\text{prob}(\alpha|\{x_k\},\beta)\right]}{d\alpha}\bigg|_{\alpha_0} = 2\sum_{k=1}^{N}\frac{x_k-\alpha_0}{\beta^2+(x_k-\alpha_0)^2} = 0 \tag{12}$$

Unfortunately, expressing α_0 in terms of $\{x_k\}$ and β is challenging. The maximum likelihood estimator for α is effective and distinct from the sample mean $\frac{1}{N}\sum_k x_k$. In our case, drawn from a Cauchy pdf, the symmetry around α might suggest the sample mean as a suitable estimate. However, it doesn't align with the solution to the equation (12) for the best estimate α_0. This discrepancy arises because the Cauchy distribution violates a crucial assumption of the central limit theorem. Unlike other distributions, the variability of the mean doesn't decrease with more measurements, persistently remaining 'wrong' regardless of the dataset size. This exception stems from the additive nature of means (μ) and variances (σ^2). In the case of the Cauchy distribution, with its extremely wide wings, σ^2 becomes infinite, and μ is undefined. Consequently, the variability of the mean doesn't decrease with an increasing number of measurements, remaining as 'wrong' after a thousand or a million data points as it is after just one.

4 iv

I assign a simple uniform pdf for the prior of α from equation(9):

$$\text{prob}(\alpha) = \begin{cases} A & \text{if } \alpha_{\min} < \alpha < \alpha_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

I also assign a simple uniform pdf for the prior of β:

$$\text{prob}(\beta) = \begin{cases} B & \text{if } \beta_{\min} < \beta < \beta_{\max}, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

I choose to assign a uniform distribution for the location parameter α (representing the position of the lighthouse along the coastline) and the distance parameter β (representing the distance out to sea for the lighthouse). The range for α, denoted by $\alpha_{\min}$ and $\alpha_{\max}$, can either correspond to known limits of the coastline or be extended arbitrarily large, even up to the size of the Earth if our knowledge is very limited. In either case, the normalization constant A is simply the reciprocal of this prior range.

Similarly, the range for β, denoted by $\beta_{\min}$ and $\beta_{\max}$, can either correspond to known limits of the distance out to sea for the lighthouse or be extended arbitrarily large, again up to the size of the Earth if our knowledge is very limited. The normalization constant B is likewise the reciprocal of this prior range.

The decision to utilize a uniform prior distribution stems from several key considerations. Firstly, it signifies the absence of prior knowledge or bias towards specific values within the parameter space. In our scenario, lacking information about the lighthouse's location leads to the absence of prior knowledge for the β and α. Treating all values as equally probable ensures objectivity in the analysis, particularly in situations where data is scarce or limited. Moreover, the simplicity and ease of interpretation of uniform priors are advantageous. Unlike more intricate distributions, they solely necessitate specifying a range of plausible values, eliminating the need for subjective judgments or complex parameterization. This uncomplicated approach enhances transparency and facilitates clear communication throughout the modeling process. Furthermore, uniform priors offer practical advantages in computational efficiency, notably in sampling techniques like Markov Chain Monte Carlo (MCMC). Sampling from a uniform distribution is computationally efficient and straightforward to implement, rendering uniform priors well-suited for Bayesian inference frameworks. Ultimately, the decision to adopt a uniform prior distribution promotes objectivity, simplicity, and computational tractability in Bayesian analysis, thus bolstering robust and efficient inference procedures.

5 v

I utilized nested sampling to sample from the posterior distribution $P(\alpha, \beta|\{x_k\})$. This approach was inspired by the solution to the lighthouse problem outlined in the textbook by D. S. Sivia, titled "Data Analysis: A Bayesian Tutorial". While the author employed the C programming language to address the problem, I opted for Python.

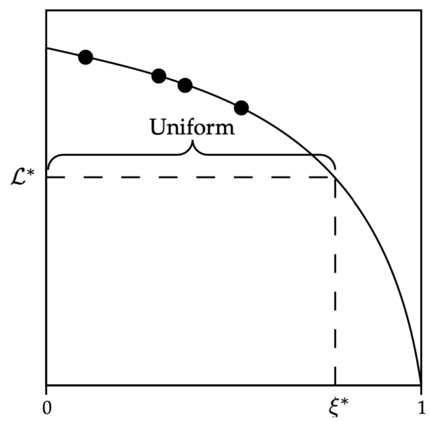
5.1 the basic idea of nested sampling

Nested sampling is a computational technique primarily used in Bayesian inference and statistical estimation, especially in contexts where traditional methods like Markov Chain Monte Carlo (MCMC) are impractical or inefficient. It was introduced by John Skilling in 2004 as a method for calculating the Bayesian evidence and sampling from the posterior distribution.

The core idea behind nested sampling is to transform the problem of computing the evidence integral (which often involves high-dimensional spaces and complex likelihood functions) into a problem of sampling from a sequence of nested sets with progressively smaller prior mass. This is achieved by iteratively sampling from within a series of "shellsör "contours"that enclose regions of higher likelihood values. Each iteration of nested sampling progressively shrinks the prior mass while focusing on regions of higher likelihood, ultimately providing samples from the posterior distribution along with an estimate of the Bayesian evidence.

The new technique of nested sampling (Skilling, 2004) tabulates the sorted likelihood function $\mathcal{L}(x)$ in a way that itself uses Monte Carlo methods. The technique uses a collection of n objects x, randomly sampled with respect to the prior π, but also subject to an evolving constraint $\mathcal{L}(x) > \mathcal{L}^*$ preventing the likelihood from exceeding the current limiting value $\mathcal{L}^*$.

In terms of ξ, the objects are uniformly sampled subject to the constraint $\xi < \xi^*$, where ξ^* corresponds to $\mathcal{L}^*$; this is illustrated in Fig. 2. At the outset, sampling is uniform over the entire prior, meaning that $\xi^* = 1$ and $\mathcal{L}^* = 0$. The idea is then to iterate inwards in ξ and correspondingly upwards in $\mathcal{L}$, in order to locate and quantify the tiny region of high likelihood where most of the joint distribution is to be found.

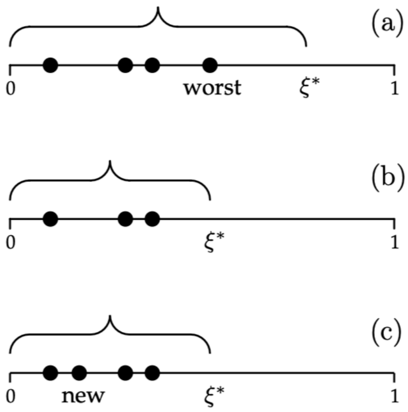


Figuur 2: Four objects ($n = 4$) sampled uniformly in $\xi < \xi^*$, or equivalently in $\mathcal{L} > \mathcal{L}^*$.

On entry, an iteration holds n objects restricted to $\xi < \xi^*$, as shown in Fig. 3(a). The worst of these, being the one with the smallest likelihood and hence the largest ξ, is selected. Located at the largest of n numbers uniformly distributed in $(0, \xi^*)$, it will lie about one part in n less than ξ^*.

Iteration proceeds by using the worst object's $(\xi, \mathcal{L})$ as the new $(\xi^*, \mathcal{L}^*)$. Meanwhile, the worst object, no longer obeying the constraint, is discarded. There are now $n-1$ surviving objects, still distributed uniformly over ξ but confined to a shrunken domain bounded by the new constraint ξ^*; this is illustrated in Fig. 3(b). The new domain is nested within the old, hence the name 'nested sampling'.

The next step is to generate a replacement object, sampled uniformly over the prior but constrained within this reduced domain. For now, we assume that we are able to do this. Having done it, the iteration again holds n objects restricted to $\xi < \xi^*$, as in Fig. 3(c), just like on entry except for the 1-part-in-n shrinkage. The loop is complete, and the next iteration can be started.



Figuur 3: An iteration replaces the worst object with a new one inside the shrunken domain.

The shrinkage ratio $t = \xi/\xi^*$ is distributed as

$$\text{prob}(t) = nt^{n-1}, \tag{15}$$

with mean and standard deviation

$$\log t = \frac{(-1 \pm 1)}{n}. \tag{16}$$

Successive iterations generate a sequence of discarded objects on the edges of progressively smaller nested domains. At iterate k,

$$\mathcal{L}_k = \mathcal{L}^* \quad \text{and} \quad \xi_k = \xi^* = \prod_{j=1}^{k} t_j \tag{17}$$

In which each shrinkage ratio t_j is independently distributed with the pdf of equation (15) with the statistics of equation (16). It follows that

$$\log \xi_k = \frac{(-k \pm \sqrt{k})}{n}. \tag{18}$$

Ignoring uncertainty for a moment, we can proclaim each $\log t$ to be $-1/n$ so that $\xi_k = \exp(-k/n)$, and the sequence then tabulates $\mathcal{L}(\xi)$ just as we require. The evidence is

$$Z = \int_0^1 d\xi \mathcal{L}(\xi) \tag{19}$$

The evidence is evaluated by associating with each object in the sequence a width $h = \Delta\xi$, and hence a vertical strip of area $A = h\mathcal{L}$, whence

$$Z \approx \sum_k A_k, \tag{20}$$

where $A_k = h_k \mathcal{L}_k$. Here, the simplest assignment of the width is $h_k = \xi_{k-1} - \xi_k$. One can try to be more accurate by using the trapezoid or such rule instead, but the uncertainties in ξ tend to overwhelm these minor variations of implementation.

We can generate quantities from the posterior distribution:

- Each sequence in the parameter space $\{x_k\}$ has an associated weight $w_k = h_k \mathcal{L}_k$, where $h_k = \Delta\xi_k$ and $Z = \sum_k h_k \mathcal{L}_k$.
- The weights define the posterior PDF. Any quantity $f(x)$ can be generated from the posterior in the usual way:

$$\langle f \rangle = \sum_k w_k f(x_k) \tag{21}$$

$$\langle f \rangle = \sum_k w_k f^2(x_k) \tag{22}$$

$$\mathrm{var}(f) = \langle f^2 \rangle - \langle f \rangle \tag{23}$$

5.2 programming the lighthouse problem in python

5.2.1 Nested Sampling Algorithm Design

The 'lighthouse' problem involves determining the location of a lighthouse based on observations of flashes along the coastline, emitted in random directions. The lighthouse is assumed to be located within the rectangle $-2 < \alpha < 2$ and $0 < \beta < 4$, with a uniform prior distribution.

Problem: The lighthouse at position (α, β) emits n flashes observed at locations $\{x_k\}$ along the coast.

Inputs:

- **Prior Distribution**: α uniformly distributed in the interval $(-2, 2)$ and β uniformly distributed in the interval $(0, 4)$.
- **Likelihood**: The likelihood function $\mathcal{L}_x(\{x_k\}|\alpha, \beta)$ is given by:

$$\mathcal{L}_x(\{x_k\}|\alpha, \beta) = \prod_k \mathcal{L}_x(x_k|\alpha, \beta) = \prod_k \frac{\beta}{\pi\left(\beta^2 + (x_k - \alpha)^2\right)} \tag{24}$$

$$ln\mathcal{L}_x(\{x_k\}|\alpha, \beta) = \sum_k ln\mathcal{L}_x(x_k|\alpha, \beta) = \sum_k ln\frac{\beta}{\pi\left(\beta^2 + (x_k - \alpha)^2\right)} \tag{25}$$

Outputs:

- The evidence Z is obtained by integrating $\mathcal{L}_x(\{x_k\}|\alpha, \beta)$ multiplied by the prior distribution $\mathrm{Prior}(\alpha, \beta)$ over the region of interest.
- The posterior distribution $P(\alpha, \beta|\{x_k\})$ represents the estimated probability distribution of the lighthouse position, given by $\frac{\mathcal{L}_x(\{x_k\}|\alpha,\beta)}{Z}$.

The algorithm we apply is:

1. Generate N values of α and β from the uniform priors.
2. Calculate $\mathcal{L}_x(\{x_k\}|\alpha, \beta)$ (or $\ln \mathcal{L}_x$ from equation(25)) using the N points and the $\{x_k\}$.
3. Select the value with the lowest $\mathcal{L}_x$ and set it to $\mathcal{L}_x^*$.
4. Use $\mathcal{L}_x^*$ to estimate new limits α^* and β^* and generate new values of α and β subject to these limits. Proceed until termination.

5.2.2 Nested Sampling Algorithm Implementation

In the main function, the nested sampling algorithm is implemented with the following steps:

Initialization:

- Initialize parameters such as the number N of objects (live points) (α, β), and maximum number of iterations MAX.
- Create lists to store objects and samples.

Reading Data:

- Read data from a text file containing arrival positions of flashes observed along the coastline.

Setting Prior Objects:

- Assign uniform prior distributions to the objects (α, β).

Outermost Interval:

- Calculate the outermost interval of prior mass (logwidth) based on the number of live points.

Nested Sampling Loop:

- Iterate through the maximum number of iterations MAX.
- Identify the worst object with the lowest log likelihood.
- Update the evidence based on the worst object's log likelihood and width.
- Store posterior samples.
- Replace the worst object with a copy of another survivor and evolve it within the likelihood constraint.
- Shrink the interval.

Output:

- Print evidence and display posterior results.

The **Object** class that encodes a possible solution needs to contain a trial location (α, β). In accordance with the recommendation to compute with a uniform prior on a unit square, α and β are slaved to controlling variables u and v, also contained in Object and each assigned uniform prior on $(0, 1)$. The transformation is simply:

$$\alpha = 4u - 2, \quad \beta = 4v \tag{26}$$

Object also contains the corresponding likelihood value $\mathcal{L}_x(\{x_k\}|\alpha, \beta)$, as its logarithm $\log \mathcal{L}_x$. Finally, Object contains the weight $A = h\mathcal{L}_x$, as calculated by the main program and stored in $\log Wt$.

Procedure **Prior** merely assigns a random (u, v) within the unit square and transforms that to (α, β) by equation (26), before calculating the corresponding likelihood.

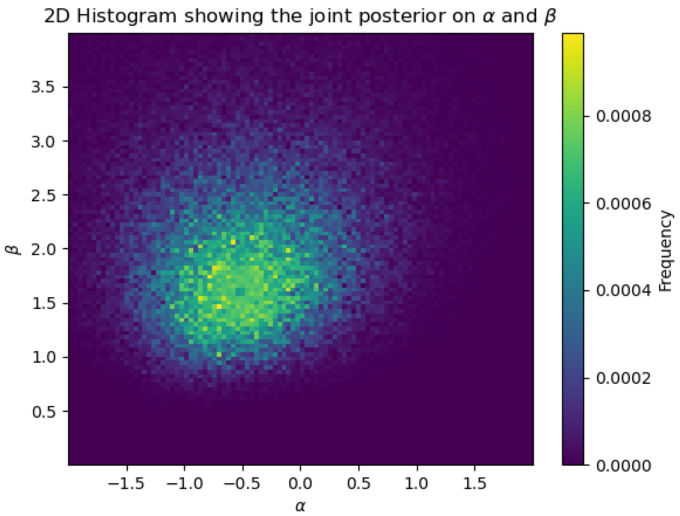
Procedure **Explore** takes a starting position (u, v) — in fact a copy of one of the other positions — and generates a new and supposedly independent position from it, subject to likelihood $\mathcal{L}_x$ exceeding the current limit $\mathcal{L}_x^*$. It does this by adding (or subtracting) a suitable increment to each of u and v, which are then mapped back within $(0, 1)$ if they had escaped, before transforming to the desired (α, β). The increment is chosen uniformly within some range $(-\text{step}, \text{step})$. Any trial position obeying the likelihood constraint is accepted, resulting in movement. Otherwise, the trial is rejected and there is no movement. Thus step should be reasonably large so that movement is reasonably fast, without being so large that the likelihood constraint stops all movement. It is difficult to pre-judge the appropriate size of step, which in any case tends to diminish as iterations proceed and the likelihood constraint becomes tighter. Accordingly, Explore includes a toy learning procedure which adjusts step to balance the number of accepted and rejected trials. Any such procedure in which the step-size almost certainly tends to some long-term limiting value is acceptable, and will allow the eventual position to be correctly distributed. In fact, Explore allows 20 of these MCMC steps of adjustable size, which seems to be adequate.

Finally, procedure **Results** computes whatever results are needed from the sequence of weighted objects found by nested sampling. The particular results chosen are the means and standard deviations of the lighthouse coordinates (α, β).

5.2.3 Results and Discussion

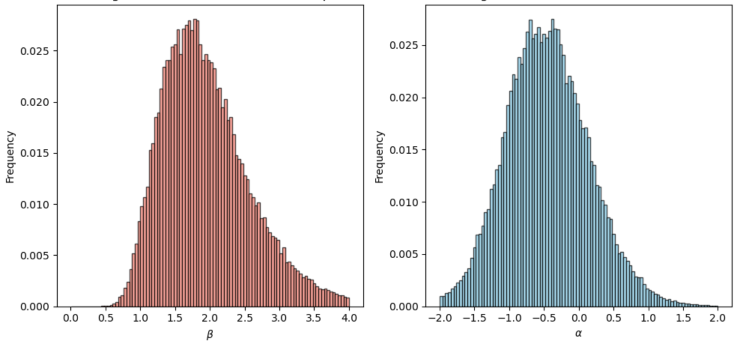
I set 10000 objects (live points) and choose 100000 iterations. The results are as follows:

In Fig.4, a 2-dimensional histogram shows the joint posterior on α and β. This plot visualizes the probability distribution of the parameters α and β based on the collected data.



Figuur 4: 2-dimensional histogram showing the joint posterior on α and β.

Fig.5 presents 1-dimensional histograms of the marginalised posteriors on both parameters. These plots show the probability distributions of individual parameters α and β separately, providing insights into their respective uncertainties.



Figuur 5: 1-dimensional histograms of the marginalised posteriors on both parameters.

Table 1 presents the measurements of both parameters quoted in the form mean $\pm$ standard deviation from equation(23).

Tabel 1: Measurements of both parameters

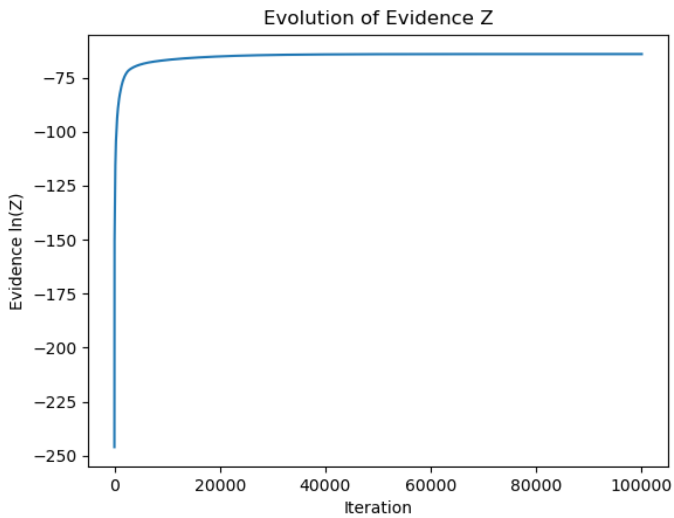
Parameter	Mean	Standard Deviation
α	-0.4477	0.5831
β	1.9392	0.6171

A suitable convergence diagnostic for the nested sampling algorithm is the evolution of the evidence Z as a function of the number of iterations because it provides insight into how well the algorithm is performing and whether it is converging towards a stable value of the evidence.

The evidence Z represents the integral of the likelihood function over the parameter space, and it quantifies the probability of the observed data given the model. In the context of nested sampling, the evidence is a crucial quantity as it allows us to compare different models and assess their relative goodness of fit to the data.

By tracking the evolution of the evidence Z over the course of iterations, we can observe how it changes and whether it stabilizes or converges to a certain value. If the algorithm is converging properly, we would expect to see the evidence Z approach a stable value as the number of iterations increases. On the other hand, if the algorithm is not converging, we may observe erratic behavior or a lack of convergence in the evolution of Z.

The evolution of evidence Z for nested sampling of the lighthouse problem is shown in Figure 6. From the plot, we can observe that the evidence Z converges very well as the number of iterations increases. This indicates that the nested sampling algorithm is performing effectively and converging towards a stable value of the evidence. Hence, we can conclude that our nested sampling approach is successful in efficiently exploring the parameter space and accurately estimating the evidence for the lighthouse problem. Additionally, this convergence suggests that the samples obtained from the posterior distribution are representative and reliable for further analysis.



Figuur 6: Evolution of Evidence Z for Nested Sampling.

6 vi

A suitable prior for the unknown parameter I_0 can be chosen as a log-uniform distribution. This choice is justified for several reasons:

1. **Positive Constraint:** The parameter I_0 represents the absolute intensity of the lighthouse, which is inherently a positive quantity. A log-uniform distribution ensures that I_0 remains strictly positive, which is consistent with the physical interpretation of intensity.

2. **Flexibility:** Log-uniform distributions are flexible and accommodate a wide range of possible values for I_0. This is particularly useful when the true value of I_0 is uncertain and may span several orders of magnitude.

3. **Non-Informative Nature:** Log-uniform priors are often considered non-informative or weakly informative, as they assign equal prior probability density to each order of magnitude. This allows the data to drive the inference process, making the prior less influential in the final results.

4. **Handling Logarithmic Terms:** In the likelihood function, I_0 appears in logarithmic form ($\log I_0$), which can complicate computations with traditional uniform priors. Using a log-uniform prior directly aligns with the logarithmic nature of the likelihood, simplifying calculations and interpretations.

Overall, the log-uniform prior provides a convenient and appropriate choice for the unknown parameter I_0, ensuring positivity, flexibility, and compatibility with the likelihood function.

7 vii

7.1 Problem Analysis

This time, the input for nested sampling consists of priors for α, β, and $\log I_0$, all of which follow a uniform distribution. Specifically, α is uniformly distributed in the interval $(-2, 2)$, β is uniformly distributed in the interval $(0, 4)$, and $\log I_0$ is uniformly distributed in the interval $(-5, 5)$.

Additionally, the likelihood for a combined location and intensity measurement is calculated. Since the location and intensity measurements are independent, the likelihood for this combined measurement is the product of the likelihoods for each individual measurement. The likelihood for a combined location and intensity measurement is given by:

$$\begin{aligned}\mathcal{L}_x(\{x_k\}, \log\{I_k\}|\alpha,\beta,I_0) &= \mathcal{L}_x(\{x_k\}|\alpha,\beta)\mathcal{L}_I(\log\{I_k\}|\alpha,\beta,I_0) \\ &= \prod_k \mathcal{L}_x(x_k|\alpha,\beta)\mathcal{L}_I(\log I_k|\alpha,\beta,I_0) \\ &= \prod_k \frac{\beta}{\pi\left(\beta^2+(x_k-\alpha)^2\right)} \frac{\exp\left(-\frac{(\log I-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}\end{aligned} \tag{27}$$

$$\log\mathcal{L}_x(\{x_k\}, \log I_k|\alpha,\beta,I_0) = \sum_k -\log\pi - 0.5\log(2\pi) + \log\beta - \log d_k^2 - 0.5(\log I_k-\mu)^2 \tag{28}$$

$$= \sum_k -\log\pi - 0.5\log(2\pi) + \log\beta - \log d_k^2 - 0.5(\log I_k - \log I_0 + \log d_k^2)^2 \tag{29}$$

where

$$d_k^2 = \beta^2 + (x_k-\alpha)^2 \tag{30}$$

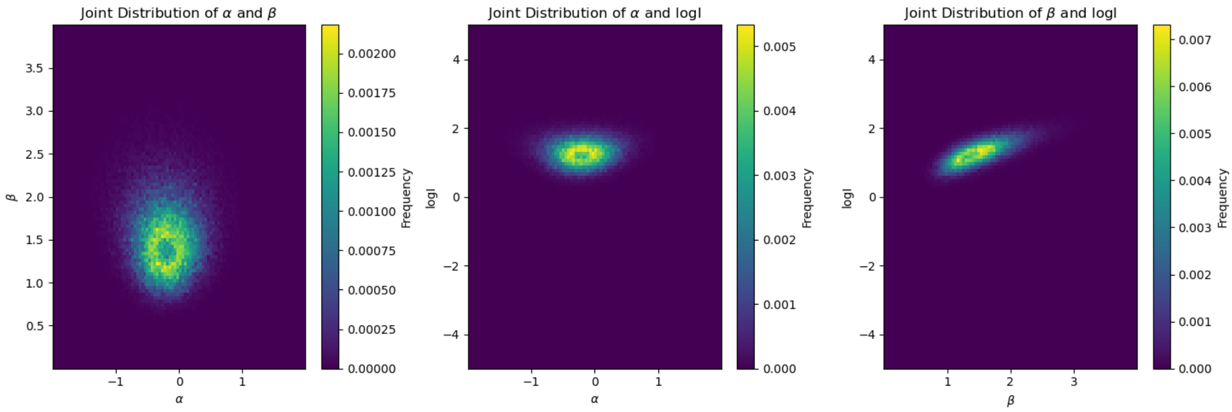
$$\mu = \log(I_0/d^2) \tag{31}$$

$$\sigma = 1 \tag{32}$$

7.2 Results and Discussion

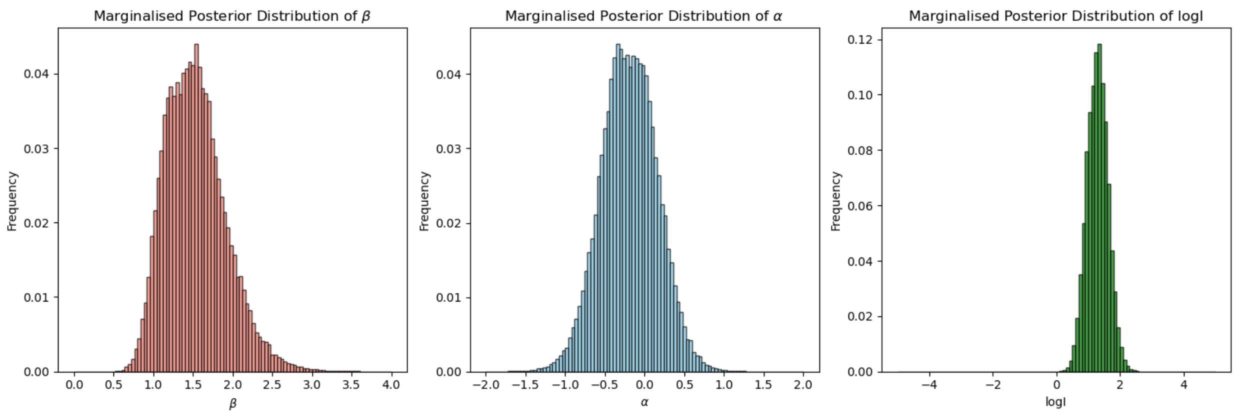
I set 10000 objects (live points) and choose 100000 iterations. The results are as follows:

In Fig.7, three 2-dimensional histograms depict the joint posterior distribution on α, β, and $\log I_0$. These plots visually represent the probability distribution of the parameters α, β, and $\log I_0$ based on the collected data. The color intensity in each plot indicates the frequency of occurrence of parameter combinations, with darker regions representing lower frequencies. Color bars alongside each plot provide a scale for interpreting the frequency of occurrences.



Figuur 7: 2-dimensional histogram showing the joint posterior on α, β, and $\log I_0$.

Fig.8 illustrates 1-dimensional histograms depicting the marginalised posteriors on all parameters. These plots elucidate the probability distributions of individual parameters α, β, and $\log I_0$ separately, offering insights into their respective uncertainties. Each histogram showcases the distribution of one parameter, allowing for a focused examination of its probabilistic characteristics.



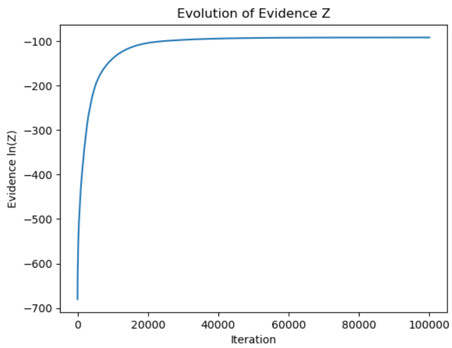
Figuur 8: 1-dimensional histograms of the marginalised posteriors on all parameters.

Table 2 displays the measurements of parameters α, β, and $\log I_0$ quoted in the form of mean $\pm$ standard deviation calculated using equation (23). This table provides a summary of the estimated values for each parameter along with their corresponding uncertainties, aiding in the characterization of the lighthouse system.

Tabel 2: Measurements of both parameters

Parameter	Mean	Standard Deviation
α	-0.1894	0.3286
β	1.5203	0.3830
$\log I_0$	1.2756	0.3299

The evolution of evidence Z for nested sampling of the lighthouse problem is shown in Figure 9. From the plot, we can observe that the evidence Z converges very well as the number of iterations increases. This indicates that the nested sampling algorithm is performing effectively and converging towards a stable value of the evidence. Hence, we can conclude that our nested sampling approach is successful in efficiently exploring the parameter space and accurately estimating the evidence for the lighthouse problem. Additionally, this convergence suggests that the samples obtained from the posterior distribution are representative and reliable for further analysis.



Figuur 9: Evolution of Evidence Z for Nested Sampling.

8 viii

When considering only the flash location in part (v), the measurement for α is $\alpha = -0.4477 \pm 0.5831$. However, in part (vii), when both flash location and intensity data are considered, the measurement for α becomes $\alpha = -0.1894 \pm 0.3286$.

Incorporating intensity data results in a more accurate measurement of α with a lower standard deviation and reduced uncertainty. This improvement is evident when comparing the marginal distributions of α in Fig. 5 and Fig. 8. The distribution becomes more symmetric and focused, resembling a Gaussian distribution more closely.

Therefore, including intensity data has led to an enhanced measurement of α.