Ellipsoidal nested sampling, expression of the model uncertainty and measurement

C Palmisano[1], G Mana[2] and G Gervino[1,3]

[1] Dipartimento di Fisica, Università di Torino, Via P. Giuria, 1 - 10125 Torino, Italy
[2] INRIM - Istituto Nazionale di Ricerca Metrologica, Str. delle Cacce, 91 - 10135 Torino, Italy
[3] Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Torino, Via P. Giuria, 1 - 10125 Torino, Italy

E-mail: `palmisan@to.infn.it`, `g.mana@inrim.it`, `gervino@to.infn.it`

Abstract. The measurand value, the conclusions, and the decisions inferred from measurements may depend on the models used to explain and to analyze the results. In this paper, the problems of identifying the most appropriate model and of assessing the model contribution to the uncertainty are formulated and solved in terms of Bayesian model selection and model averaging. As computational cost of this approach increases with the dimensionality of the problem, a numerical strategy, based on multimodal ellipsoidal nested sampling, to integrate over the nuisance parameters and to compute the measurand post-data distribution is outlined. In order to illustrate the numerical strategy, by use of *MATHEMATICA* an elementary example concerning a bimodal, two-dimensional distribution has also been studied.

1. Introduction

Use of Bayesian methods as ways to incorporate the model uncertainty into the data analysis and the uncertainty budget is often computationally expensive, but advances in computing technology have allowed to take these methods into account. Consequently, the choice between two or more hypotheses, for example when we have to the fit the shape of an experimental curve, can be made by considering different models, each indexed by one or more parameter, where the Bayesian model selection and model averaging provides the probabilistic framework to simultaneously treat both the model and data uncertainties.

Let $\boldsymbol{x}$ the list measurement data, $(A, B, \ldots, M, \ldots)$ the list of different possible models, θ the measurand, and $(\theta_A, \theta_B, \ldots, \theta_M \ldots)$ the relevant nuisance parameters. The joint distribution of the data, measurand, parameters, and model can be written in terms of conditional probabilities as

$$P(\boldsymbol{x}, \theta, \theta_M, M) = L(\theta, \theta_M, M|\boldsymbol{x})\pi(\theta, \theta_M|M)\Pi(M), \qquad (1)$$

where we have introduced the likelihood $L(\theta, \theta_M, M|\boldsymbol{x}) = P(\boldsymbol{x}|\theta, \theta_M, M)$, the conditional prior probability distribution of the measurand and parameters $\pi(\theta, \theta_M|M)$, and the prior probability of the model $\Pi(M)$. By conditioning the left hand side of Eq. (1) on the data and the model, and applying the chain rule, we can write (1) as

$$P(\theta, \theta_M|\boldsymbol{x}, M) = \frac{L(\theta, \theta_M, M|\boldsymbol{x})\pi(\theta, \theta_M|M)}{Z(\boldsymbol{x}|M)}, \qquad (2)$$

where we introduced the so-called "evidence" $Z(\boldsymbol{x}|M) = P(\boldsymbol{x}|M)$. It is the normalizing factor of the numerator in (2), i.e. it satisfies the relationship

$$Z(\boldsymbol{x}|M) = \int_{\Theta \oplus \Theta_M} L(\theta, \theta_M, M|\boldsymbol{x})\pi(\theta, \theta_M|M)\, \mathrm{d}\theta\, \mathrm{d}\theta_M \,. \tag{3}$$

By applying the Bayes theorem to $P(\boldsymbol{x}, M)$ we obtain the discrete probability $\mathrm{Prob}(M|\boldsymbol{x})$ for the model M provided by the data

$$\mathrm{Prob}(M|\boldsymbol{x}) = \frac{Z(\boldsymbol{x}|M)\Pi(M)}{P(\boldsymbol{x})}, \tag{4}$$

which, by use of the law of total probability, can be rewritten in terms of evidence $Z(\boldsymbol{x}|M)$ of each model as

$$\mathrm{Prob}(M|\boldsymbol{x}) = \frac{Z(\boldsymbol{x}|M)\Pi(M)}{\sum_M Z(\boldsymbol{x}|M)\Pi(M)} \,. \tag{5}$$

Eq. (5) gives the probability of a model given the observed data, and a simple way to select a model is to choose the most probable. However, when the relevant probabilities are approximately equal and no single model stands out, it is necessary to average over the considered models.

2. Multimodal, ellipsoidal, nested sampling

Since the evaluation of evidence Z in (3) becomes impractical when the integration space has $20 \div 30$ dimensions, and, as in the most practical case, the likelihood is different from zero in a small fraction of the integration domain, we have investigated a nested sampling technique relating the likelihood values to the prior volume [1, 2].

Firstly, p likelihood samples $L_{\theta_1}, \ldots, L_{\theta_p}$ are sampled in $\Theta \oplus \Theta_M$ according to $\pi(\theta, \theta_M|M)$. Next, the smallest, indicated as L_1, is removed and replaced by a new sample, L_{new}, subject to the constraint $L_{\mathrm{new}} > L_1$. The prior volume enclosed by the iso-likelihood surface $L = L_1$ is estimated as $p/(p+1)$, where $p/(p+1)$ is the mean value of the largest of p uniform samples in $[0,1]$, the total volume of $\Theta \oplus \Theta_M$ being normalised to 1. The discharge of the lowest likelihood L_n, sampling of a replacement constrained to $L_{\mathrm{new}} > L_n$, and shrinking of the prior volume of the associated iso-likelihood surface to $V_n = p^n/(p+1)^n$ are repeated until some stopping criterion is satisfied, for example when the contribution to (3) of the surviving likelihood samples, i.e. $L_{\max}p^n/(p+1)^n$, where $L_{\max}$ is the maximum sample, is less than some pre-defined value.

By using the sequence of the discarded likelihoods $0 < L_1 < L_2 \ldots < L_N$, and the differences $\Delta V_n = V_{n-1} - V_n$ of the associated prior volumes $V_0 = 1 > V_1 > V_2 > \ldots V_N > 0$, Eq. (3) can be approximated as (rectangle method)

$$Z(\boldsymbol{x}|M) \approx \sum_{n=1}^{N} L_n \Delta V_n + L_{\max}\left(\frac{p}{p+1}\right)^N \,. \tag{6}$$

The challenge in implementing (6) is sampling within the iso-likelihood surfaces $L > L_n$. As the Monte Carlo Markov Chain algorithm may be inefficient, improvements have been proposed in [3]-[6], which we are presently investigating for application to metrology.

Ellipsoidal sampling [3] replaces the iso-likelihood surface $L = L_n$ by a hyper-ellipsoid given by the covariance matrix of the living samples and centered in their mean value, and the $L_{\mathrm{new}} > L_n$ is sampled within the intersection of the domain of integration and this hyper-ellipsoid [7].

When the integrand in (3) is a multimodal function, the hyper-ellipsoid will be often bonded to the maxima of the integrand and the sampling will result in an unacceptable decrease in the acceptance rate of L_{new}.

In this case, a possible strategy is partitioning the set of all living points in clusters, and then enclose each cluster in a "small" hyper-ellipsoid, centered in the mean value of the cluster points and defined by the covariance matrix calculated with the cluster points.

In [4] is described a method to avoid hyper-ellipsoids to overlap, but we will consider a method to manage the overlapping according to [5, 6, 8]. At each step of the method the set of all living points is partitioned in clusters. If the number of hyper-ellipsoids is greater than 1, a single hyper-ellipsoid is sampled with probability proportional to its volume; then a $L_{\text{new}} > L_n$ is sampled and accepted with probability $1/(\text{number of hyper-ellipsoids containing } L_{\text{new}})$. If the point is rejected, a new sampling of a hyper-ellipsoid takes place.

2.1. Bimodal example in $\mathbb{R}^2$
By use of *MATHEMATICA* (Wolfram Inc.), we applied the ellipsoidal nested sampling to this elementary integral

$$I = V_{\text{D}} \int_{30}^{45} \left(\int_{31}^{44} f(x,y)\,\mathrm{d}y \right) \frac{1}{V_{\text{D}}}\,\mathrm{d}x, \tag{7}$$

where $V_{\text{D}} = (45 - 30) \times (31 - 44) = 195$ is the volume of the integration domain and (see Fig. 1)
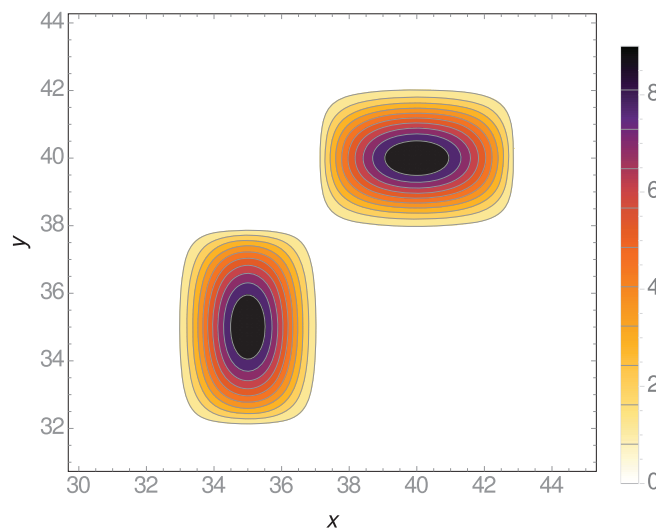

Figure 1. Contour plot of the function $f(x,y)$ defined in Eq. 8.

$$f(x,y) = \cos^2\left(\frac{2\pi}{10}(x - 35)\right)\mathtt{UnitBox}\left(\frac{x - 35}{5}\right)\left(-(y - 35)^2 + 9\right)\mathtt{UnitBox}\left(\frac{y - 35}{6}\right)$$
$$+ \left(-(x - 40)^2 + 9\right)\mathtt{UnitBox}\left(\frac{x - 40}{6}\right)\cos^2\left(\frac{2\pi}{10}(y - 40)\right)\mathtt{UnitBox}\left(\frac{y - 40}{5}\right). \tag{8}$$

The bimodal function $f(x,y)$, which is centered in $(x,y) = (35, 35)$ and $(x,y) = (40, 40)$, plays the role of a likelihood; $1/V_{\text{D}}$ is the prior distribution and $\mathtt{UnitBox}(x)$ is by definition equal to 1 for $|x| \leqslant 1/2$ and 0 elsewhere. Our goal is to reobtain the value of the analytical calculation, that is $I = 180$. Fig. 2 shows an ellipsoid (red dashed curve) calculated after sampling uniformly 200 points from the domain of integration. The external ellipsoid, red continuous curve corresponding to $h\,k_1$, where $h = 1.1$, enlarge the sampling volume and is useful to better study the contour lines (iso-likelihood surfaces) of the integrand. It is also shown the minimum point of the integrand (green large disk). Strictly speaking, the integrand
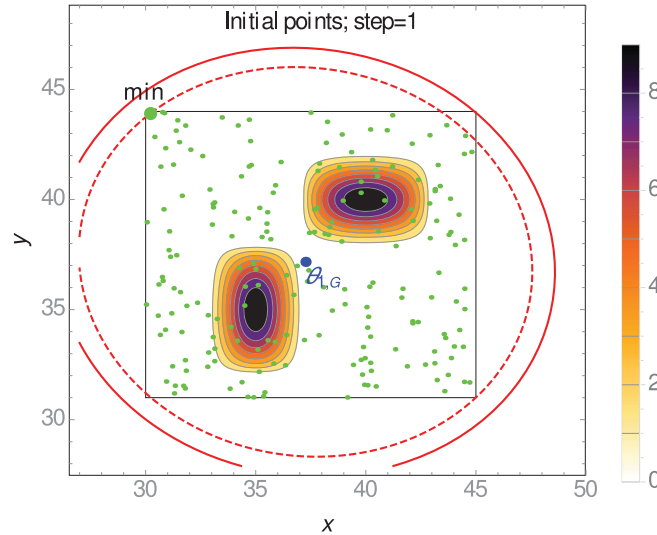
Figure 2. Contour plot of the function $f(x,y) + l(x,y)$; are also shown 200 sampled points (green), a couple of ellipsoids (red curves, corresponding to k_1 and $h\,k_1$, with $h = 1.1$) containing the sampled points, the mean value $\theta_{1,\mathrm{G}}$ (blue) and the minimum point (green disk).
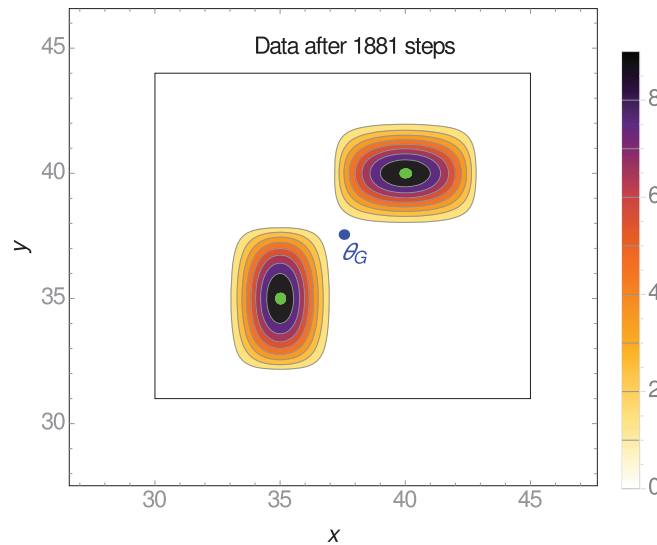

Figure 3. Contour plot of the function $f(x,y) + l(x,y)$; after 1881 steps the 200 living points (green) shrunk around the maximum points of $f(x,y)$. The internal rectangle indicates the integration domain $[30, 45] \times [31, 44]$.

should be equal to zero outside the support of $f(x,y)$, but in order to implement the nested sampling it is necessary to add to $f(x,y)$ a little appropriate function $l(x,y)$. In our example with function (8), after 1881 steps, the set of the 200 living points is totally shrunk in the maximum points of the integrand $f(x,y)$, see Figs. 3; moreover, a file containing a list of 1881 "minimal likelihood" values L_n is available to calculate the evidence Z according to Eq. (6). With the trapezoid rule, the final result is integral= 148.7 with uncertainty= 14.3. The uncertainty has been assessed by use of one of the equivalent formulae suggested in [9]. The set of the lowest likelihood points is shown in Fig. 4 and the list of the number of clusters at each step is shown in Fig. 5 A realization of 100 independent calculations of the integral with 200 points, that is 100
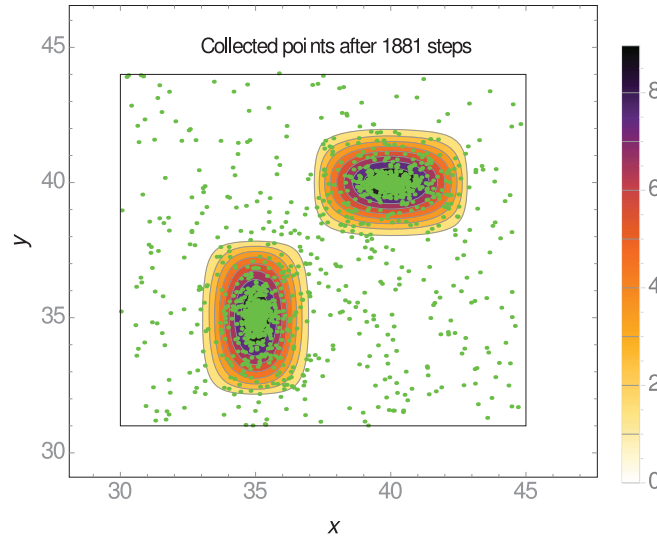
Figure 4. Set of the lowest likelihood points (green), corresponding to the discharged likelihood values L_n by which to calculate the integral in Eq. (6). The internal rectangle indicates the integration domain $[30, 45] \times [31, 44]$.
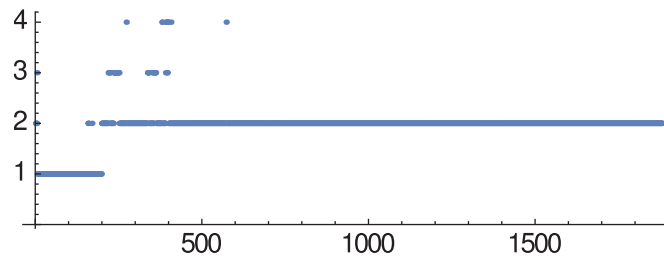

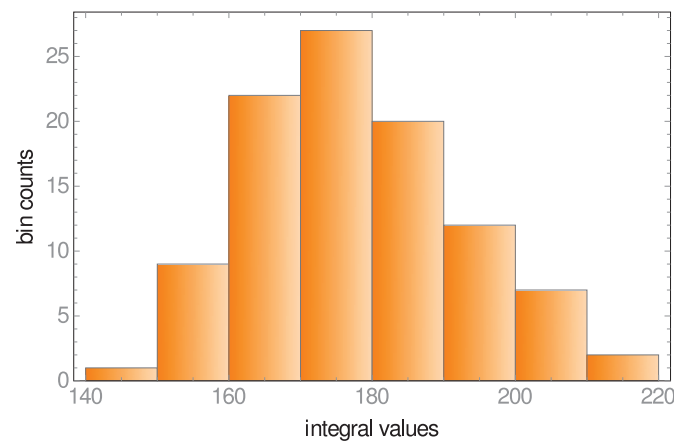**Figure 5.** Number of clusters at each step.

Figure 6. Histogram obtained by 100 independent calculations of (7). Mean value $= 178.1$, standard deviation of the mean $= 1.5$, exact value $= 180$.

different sets of 200 initial points, has generated the histogram in Fig. 6 having a mean value $= 178.1$ and a standard deviation of the mean $= 1.5$, the trapezoid rule having been exploited again. Obviously, the mean value of the histogram is very close to the analytical calculation, while the value obtained by a single calculation can be very different from the real value.

5

3. Conclusion

The role of the Bayesian approach to the expression of the model uncertainty in measurements consists of the quantification of both the sampling and model contributions to the post-data uncertainty. The multimodal nested sampling has been fruitful in astrophysics, but at authors knowledge has not yet been exploited in metrology. In any case, to develop an algorithm and the relevant software to tackle numerical integration of multimodal functions over high-dimensional spaces can be fruitful from a more general point of view.

Acknowledgments

References

[1] Skilling J 2004 *AIP Conf. Proc.* vol 735, ed R Fischer, R Preuss and U von Toussaint pp 395-405
[2] Skilling J 2006 *Bayesian Analysis* **1** 833
[3] Mukherjee P, Parkinson D and Liddle A R 2006 *The Astr. J.* **638** L51
[4] Shaw R, Bridges M and Hobson M P 2007 *Mon. Not. R. Astron. Soc.* **378** 1365
[5] Feroz F and Hobson M P 2008 *Mon. Not. R. Astron. Soc.* **384** 449
[6] Feroz F, Hobson M P and Bridges M 2009 *Mon. Not. R. Astron. Soc.* **398** 1601
[7] Allison R and Dunkley J 2014 *Mon. Not. R. Astron. Soc.* **437** 3918
[8] Feroz F, Hobson M P, Cameron E and Pettitt A N 2013 Importance Nested Sampling and the MultiNest Algorithm *Preprint* astro-ph/1306.2144v2
[9] Keeton C R 2011 *Mon. Not. R. Astron. Soc.* **414** 1418