Recent advances in Bayesian inference in cosmology and astroparticle physics thanks to the MultiNest algorithm

Trotta, Roberto
Imperial College London, Blackett Laboratory, Prince Consort Road, London, SW7 2AZ, UK
r.trotta@imperial.ac.uk

Feroz, Farhan
University of Cambridge, Cavendish Laboratory JJ Thomson Avenue, Cambridge CB3 0HE, UK
f.feroz@mrao.cam.ac.uk

Hobson, Mike
University of Cambridge, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK
mph@mrao.cam.ac.uk

Ruiz de Austri, Roberto
University of Valencia, Instituto de Física Corpuscular, IFIC-UV/CSIC, Valencia, Spain
rruiz@ific.uv.es

1 Introduction

Cosmology (the study of the large-scale properties of the Universe) and astroparticle physics (the intersection of particle physics and astrophysics, studying high-energy particles emitted by astrophysical bodies) present many interesting inference challenges for statisticians and physicists alike. Both disciplines have seen a great surge in the quality and quantity of data in the last decade, and this has in turn spurred the development of more sophisticated statistical modeling and analysis methods. The launch of the ISI Astrostatistics Section, which is present for the first time at this ISI World Congress, is a testament to the growing importance of astrostatistics, the application of advanced statistical techniques to astrophysical and cosmological problems.

Among the many different statistical challenges that large and complex astrophysical data sets pose, three classes of problems arise often in very disparate contexts. They can be broadly described as follows:

1. *Bayesian inference for multi-dimensional, multi-modal likelihoods.* Problems in as disparate fields as Supersymmetry phenomenology and gravitational wave detection have likelihood functions that are multi-modal, with modes presenting widely different characteristic scales. Numerical exploration of the full posterior distribution is a challenging problem for traditional Markov Chain Monte Carlo (MCMC) methods in this context, which risk getting stuck in local modes.

2. *Bayesian model selection aimed at model building or object detection.* Bayesian model selection is used to decide whether one or several extra parameters are needed in a model to explain the data. This has applications to questions such as deciding whether the Universe is flat, whether dark energy is a cosmological constant, and astronomical object detection (is there a galaxy cluster in a weak lensing image? Does this star have one or several exo-planets orbiting it?). This requires a reliable and efficient algorithm to compute the Bayesian evidence (or model likelihood).

3. *Profile likelihood evaluation for multi-modal likelihoods.* An alternative to posterior-based inference is represented by the (frequentist) construction of confidence regions based on the profile likelihood ratio. For highly non-Gaussian, multi-modal problems, Bayesian credible regions can

differ very considerably from confidence regions (especially in problems in high dimensions and/or when the prior is informative). Evaluation of the profile likelihood is in general a much more challenging task than exploring the posterior distribution, and the question of the accuracy of the recovered credible intervals is a relevant one in domains such as supersymmetry phenomenology (i.e., the reconstruction of properties of particles beyond the Standard Model from data from accelerators, cosmology and other astrophysical probes).

In this paper, we present the MULTINEST algorithm (section 2), an implementation of nested sampling, and we give examples of its successfull application to problems in cosmology and astroparticle physics in the above three classes (section 3).

2 Nested sampling and the MULTINEST algorithm

Bayesian inference is often the statistical framework of choice in cosmology (see e.g. [1, 2]), and, increasingly so, in astroparticle physics. The posterior pdf $p(\Theta|d, \mathcal{M})$ for the n-dimensional parameters vector Θ of a model $\mathcal{M}$ is given by

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

Here, $p(\Theta|\mathcal{M})$ is the prior, $p(d|\Theta, \mathcal{M})$ the likelihood and $p(d|\mathcal{M})$ the model likelihood, or marginal likelihood (usually called "Bayesian evidence" by physicists). The Bayesian evaluation of a model's performance in the light of the data is based on the Bayesian evidence, the normalization integral on the right–hand–side of Bayes' theorem, Eq. (1):

$$p(d|\mathcal{M}) \equiv \int p(d|\Theta, \mathcal{M})p(\Theta|\mathcal{M})\mathrm{d}^n\Theta. \tag{2}$$

Thus the Bayesian evidence is the average of the likelihood under the prior for a specific model choice. From the evidence, the model posterior probability given the data is obtained by using Bayes' Theorem to invert the order of conditioning:

$$p(\mathcal{M}|d) \propto p(\mathcal{M})p(d|\mathcal{M}), \tag{3}$$

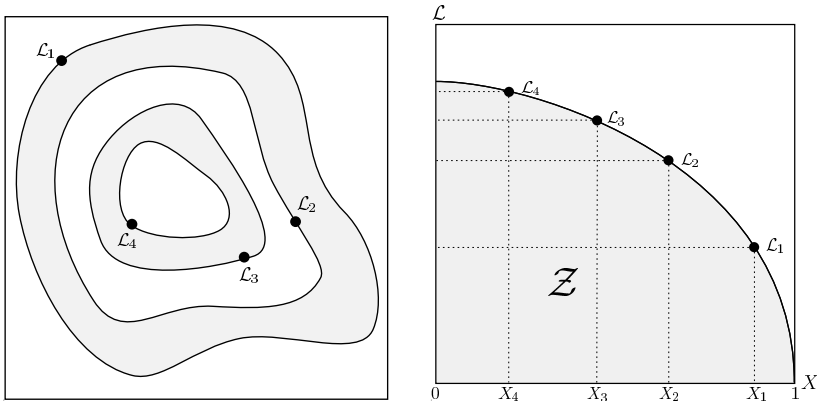
where $p(\mathcal{M})$ is the prior probability assigned to the model itself. Usually this is taken to be non–committal and equal to $1/N_m$ if one considers N_m different models. When comparing two models, $\mathcal{M}_0$ versus $\mathcal{M}_1$, one is interested in the ratio of the posterior probabilities, or *posterior odds*, given by

$$\frac{p(\mathcal{M}_0|d)}{p(\mathcal{M}_1|d)} = B_{01}\frac{p(\mathcal{M}_0)}{p(\mathcal{M}_1)} \tag{4}$$

and the *Bayes factor* B_{01} (see e.g. [3] for details) is the ratio of the models' evidences:

$$B_{01} \equiv \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)} \quad \text{(Bayes factor)}. \tag{5}$$

A value $B_{01} > (<)\ 1$ represents an increase (decrease) of the support in favour of model 0 versus model 1 given the observed data. From Eq. (4) it follows that the Bayes factor gives the factor by which the relative odds between the two models have changed after the arrival of the data, regardless of what we thought of the relative plausibility of the models before the data, given by the ratio of the prior models' probabilities. Bayes factors are usually interpreted against the Jeffreys' scale [4] for the strength of evidence. This is an empirically calibrated scale, with thresholds at values of the odds of about $3:1$, $12:1$ and $150:1$, representing weak, moderate and strong evidence, respectively.



(a) (b)

Figure 1: Cartoon illustrating (a) the likelihood of a two-dimensional problem; and (b) the transformed $\mathcal{L}(X)$ function where the prior volumes X_i are associated with each likelihood $\mathcal{L}_i$.

Nested sampling [5, 6] is a Monte Carlo technique aimed at an efficient evaluation of the Bayesian evidence, which also produces posterior inferences as a by-product. It calculates the evidence by transforming the multi-dimensional evidence integral of Eq. (2) into a one–dimensional integral that is easy to evaluate numerically. This is accomplished by defining the prior volume X as $\mathrm{d}X = p(\Theta)\mathrm{d}^n\Theta$, so that

$$X(\lambda) = \int_{\mathcal{L}(\Theta)>\lambda} p(\Theta)\mathrm{d}^n\Theta, \tag{6}$$

where $\mathcal{L}(\Theta) \equiv p(d|\Theta)$ is the likelihood function and the integral extends over the region(s) of parameter space contained within the iso-likelihood contour $\mathcal{L}(\Theta) = \lambda$ (in this section we drop the explicit conditioning on model $\mathcal{M}$, as this is understood). Assuming that $\mathcal{L}(X)$, i.e. the inverse of (6), is a monotonically decreasing function of X (which is trivially satisfied for most posteriors), the evidence integral (2) can then be written as

$$\mathcal{Z} \equiv p(d) = \int_0^1 \mathcal{L}(X)\mathrm{d}X, \tag{7}$$

Thus, if one can evaluate the likelihoods $\mathcal{L}_j = \mathcal{L}(X_j)$, where X_j is a sequence of decreasing values,

$$0 < X_M < \cdots < X_2 < X_1 < X_0 = 1, \tag{8}$$

as shown schematically in Fig. 1, the evidence can be approximated numerically using standard quadrature methods as a weighted sum

$$\mathcal{Z} = \sum_{i=1}^{M} \mathcal{L}_i w_i. \tag{9}$$

If one uses a simple trapezium rule, the weights are given by $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$. An example of a posterior in two dimensions and its associated function $\mathcal{L}(X)$ is shown in Fig. 1.

2.1 Evaluation of the Bayesian evidence

The nested sampling algorithm performs the summation (9) as follows. To begin, the iteration counter is set to $i = 0$ and N "live" (or "active") samples are drawn from the full prior $p(\Theta)$ (which is often simply the uniform distribution over the prior range), so the initial prior volume is $X_0 = 1$. The

samples are then sorted in order of their likelihood and the smallest (with likelihood $\mathcal{L}_0$) is removed from the live set and replaced by a point drawn from the prior subject to the constraint that the point has a likelihood $\mathcal{L} > \mathcal{L}_0$. The corresponding prior volume contained within this iso-likelihood contour will be a random variable given by $X_1 = t_1 X_0$, where t_1 follows the distribution $p(t) = Nt^{N-1}$ (i.e. the probability distribution for the largest of N samples drawn uniformly from the interval $[0, 1]$). At each subsequent iteration i, the discarding of the lowest likelihood point $\mathcal{L}_i$ in the live set, the drawing of a replacement with $\mathcal{L} > \mathcal{L}_i$ and the reduction of the corresponding prior volume $X_i = t_i X_{i-1}$ are repeated, until the entire prior volume has been traversed. The algorithm thus travels through nested shells of likelihood as the prior volume is reduced.

The expectation value and standard deviation of $\log t$, which dominates the geometrical exploration, are:

$$E[\log t] = -\frac{1}{N}, \qquad \sigma[\log t] = \frac{1}{N}. \tag{10}$$

Since each value of $\log t$ is independent, after i iterations the prior volume will shrink down such that $\log X_i \approx -(i \pm \sqrt{i})/N$. Thus, one takes $X_i = \exp(-i/N)$.

The nested sampling algorithm is terminated on determining the evidence to some specified precision. An adequate and robust condition [5] is given by the upper limit on the evidence that can be determined from the remaining set of current active points. By selecting the maximum-likelihood value $\mathcal{L}_{\max}$ in the set of active points, one can safely assume that an upper bound to the evidence contribution from the remaining portion of the posterior is $\Delta\mathcal{Z}_{\rm i} = \mathcal{L}_{\max} X_{\rm i}$, i.e. the product of the remaining prior volume and maximum likelihood value. We choose to stop when this quantity would no longer change the final evidence estimate by some user-defined value, described by a tolerance parameter, tol. As described in section 3.3 below, it is important to adjust this tolerance value appropriately if one wants to use MultiNest for profile likelihood evaluation.

2.2 Posterior inferences

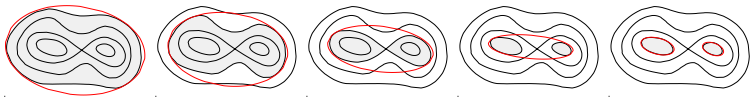
Once the evidence $\mathcal{Z}$ is found, posterior inferences are obtained from the full sequence of discarded samples from the nested sampling process, as well as the live set at termination. Each such samples is assigned the probability weight

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}}. \tag{11}$$

These samples can then be used to calculate inferences of posterior parameters such as means, standard deviations, covariances and so on, or to construct marginalised posterior distributions. The use of the posterior sample to approximate the profile likelihood is further discussed below.

2.3 The MultiNest algorithm

The most challenging task in implementing the nested sampling algorithm is drawing samples from the prior within the hard constraint $\mathcal{L} > \mathcal{L}_i$ at each iteration i. Employing a naive approach that draws blindly from the prior would result in a steady decrease in the acceptance rate of new samples with decreasing prior volume (and increasing likelihood). Ellipsoidal nested sampling [7] tries to overcome this problem by approximating the iso-likelihood contour of the point to be replaced by an n–dimensional ellipsoid determined from the covariance matrix of the current set of live points. New points are then selected from the prior within this (enlarged) ellipsoidal bound until one is obtained that has a likelihood exceeding that of the discarded lowest-likelihood point. In the limit that the ellipsoid coincides with the true iso-likelihood contour, the acceptance rate tends to unity.



(a) (b) (c) (d) (e)

Figure 2: Cartoon of ellipsoidal nested sampling from a simple bimodal distribution. In (a), the red ellipsoid represents a good bound to the active region. From (a) to (d), as we nest inward we can see that the acceptance rate will rapidly decrease as the bound steadily worsens. Figure (e) illustrates the increase in efficiency obtained by sampling from each clustered region separately. From [10].

(a) (b)

Figure 3: Illustration of the ellipsoidal decompositions used by MULTINEST to deal with degeneracies in the posterior distribution: the points given as input are overlaid on the resulting ellipsoids. 1000 points were sampled uniformly from: (a) two non-intersecting ellipsoids; and (b) a torus. From [10].

Ellipsoidal nested sampling is efficient for simple uni-modal posterior distributions without pronounced degeneracies, but is not well suited to multi-modal distributions. As advocated by [8, 9, 10] and shown in Fig. 2, the sampling efficiency can be substantially improved by identifying distinct *clusters* of live points that are well separated and constructing an individual ellipsoid for each cluster. In some problems, however, some modes of the posterior might possess very pronounced curving degeneracies, which are problematic to explore efficiently. To sample with maximum efficiency from such distributions, the MULTINEST algorithm of [10] divides the live point set into sub-clusters which are then enclosed in ellipsoids and a new point is then drawn uniformly from the region enclosed by these 'overlapping' ellipsoids, see Fig. 3. The number of points in an individual sub-cluster and the total number of sub-clusters is decided by a an 'expectation-maximization' algorithm so that the total sampling volume, which is equal to the sum of volumes of the ellipsoids enclosing the sub-clusters, is minimized. This allows maximum flexibility and efficiency by breaking up a mode resembling a Gaussian into relatively fewer number of sub-clusters, while using a larger number of small overlapping ellipsoid in the presence of severe, curving degeneracies.

3 Applications of MULTINEST

The MULTINEST algorithm sketched above has been succesfully applied to a broad range of challenging inference problems, including cosmological model selection [12], exo-planets detection and characterization [13], astronomical object detection [14], inference on Supersymmetric parameters [11, 15, 16], gravitational waves analysis [17], cosmic ray propagation models [18], and others. As an illustration, we briefly present one example for each of the three classes of problems mentioned in the Introduction.

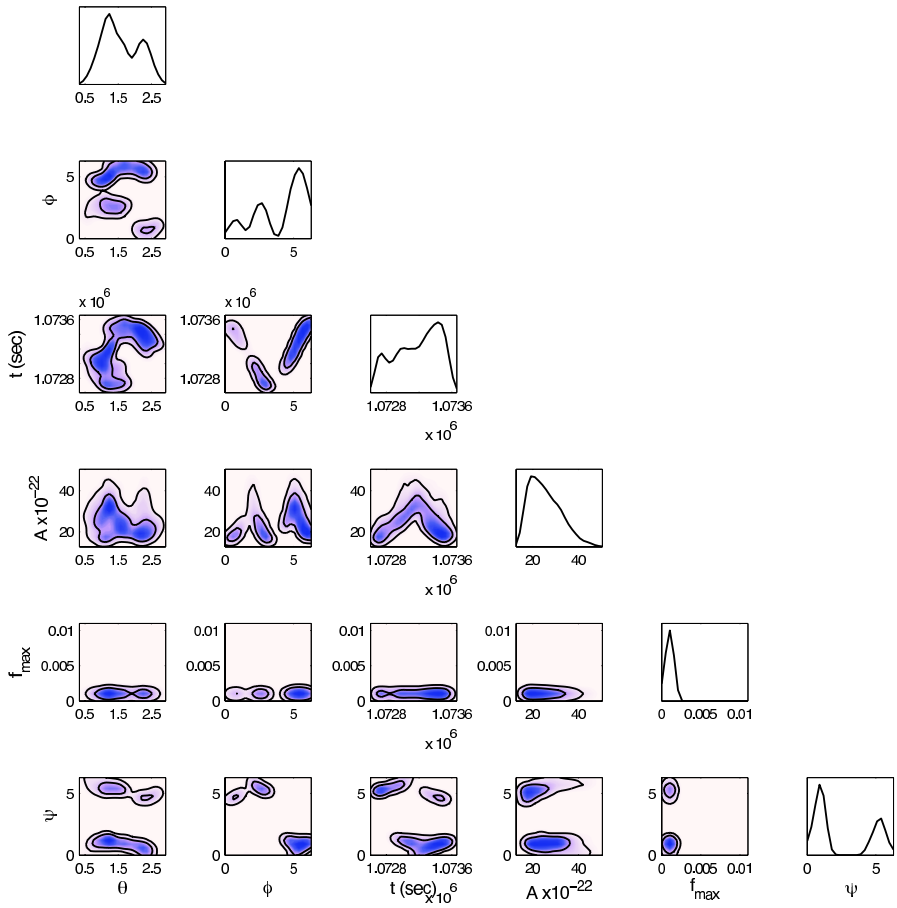


Figure 4: 2D and 1D marginalised posteriors as recovered by MultiNest in the search for the second of the cosmic string bursts in the LISA gravitational wave detector training data set. The parameters, from top-to-bottom and left-to-right, are colatitude, longitude, burst time, burst amplitude, burst break frequency and waveform polarization. From [17].

3.1 Sampling of multi-modal posteriors in gravitational waves astronomy

The proposed space-based gravitational wave detector, the Laser Interferometer Space Antenna (LISA) is expected to observe thousands of gravitational wave signals from many different types of sources, including galactic compact binaries, inspiral and mergers of supermassive black holes (SMBH) and extreme mass ratio inspirals. Detecting and characterizing this many sources presents a significant data analysis challenge. In order to encourage the research in this area, a program of Mock LISA Data Challenges (MLDC) (e.g., [19]) has been taking place, with each round consisting of one or more datasets containing simulated instrumental noise and gravitational waves from sources with undisclosed parameters. Several of the LISA sources exhibit degeneracies in their parameter space resulting in multiple modes in the posterior pdf, which proved to be a very challenging problem for traditional MCMC methods. With its ability to explore highly multi-modal distributions efficiently, MultiNest is well poised to tackle data analysis problems in gravitational wave astronomy, as demonstrated in [17].

MultiNest was used for the detection and characterisation of cosmic string burst sources in mock LISA data of increasing realism (including for example instrumental noise and partially subtracted galactic background). As a search tool, the algorithm was successful in finding the three cosmic string bursts that were present in the MLDC challenge data set. These sources, and the five sources in the MLDC training data, were correctly identified in the sense that the full signal-to-noise

ratio of the injected source was recovered, and a posterior distribution for the parameters obtained. The maximum likelihood and maximum a-posteriori parameters were not particularly close to the true parameters of the injected signals, but this was a consequence of the intrinsic degeneracies in the cosmic string model parameter space (shown in Fig. 4) and in all cases the true parameters were consistent with the recovered posterior distributions.

3.2 Inflationary Bayesian model comparison

A second example is the inflationary model comparison carried out in Ref. [12]. Although the technical details are fairly involved, the underlying idea can be sketched as follows.

The term "inflation" describes a period of exponential expansion of the Universe in the very first instants of its life, some 10^{-32} seconds after the Big Bang, during which the size of the Universe increased by at least 25 orders of magnitude. This huge and extremely fast expansion is required to explain the observed isotropy of the cosmic microwave background on large scales. It is believed that inflation was powered by one or more "scalar fields". The behaviour of the scalar field during inflation is determined by the shape of its potential, which is a real-valued function $V(\phi)$ (where ϕ denotes the value of the scalar field). The detailed shape of $V(\phi)$ controls the duration of inflation, but also the spatial distribution of inhomogeneities (perturbations) in the distribution of matter and radiation which emerge from inflation. It is from those perturbations that galaxies and cluster form out of gravitational collapse. Hence the shape of the scalar field can be constrained by observations of the large scale structures of the Universe and of cosmic microwave background anistropies.

Theories of physics beyond the Standard Model motivate certain functional forms of $V(\phi)$, which however typically have a number of free parameters, Θ. The fundamental model selection question is to use cosmological observations to discriminate between alternative models for $V(\phi)$ (and hence alternative fundamental theories). The major obstacle to this programme is that very little if anything at all is known *a priori* about the free parameters Θ describing the inflationary potential. What is worse, such parameters can assume values across several orders of magnitude, according to the theory. Hence the Occam's razor effect of Bayesian model comparison can vary in a very significant way depending on the prior choices for Θ. Furthermore, a non-linear reparameterization of the problem (which leaves the physics invariant) does in general change the Occam's razor factor, and hence the model comparison result.

In Ref. [12] a first attempt was made to tackle inflationary model selection from a principled point of view. The main result of the analysis is shown in Fig. 5, which presents the Bayes factors between models, obtained using MultiNest (suitably normalized w.r.t. a reference model, here the so-called LFI_2 model). Two classes of models for $V(\phi)$ have been considered, namely so-called Small Field Inflation (SFI) models and Large Field Inflation (LFI) models. The two classes of model differ in the parameterized form of $V(\phi)$, and have different sets of parameters, differing in dimensionality, as well. Within each class of models, sub-classes are defined (denoted by subscripts in Fig. 5) based on theoretical considerations, e.g. by fixing some of the parameters to certain values. The priors on the models' parameters have been chosen based on theoretical considerations of possible values achievable under each class of models. Typical priors are uniform on the log of the parameter (to reflect indifference w.r.t. the characteristic scale of the quantity), within a range chosen as a reflection of physical model building. The models' priors are chosen in such a way to lead to non-committal priors for the two classes as a whole, i.e. $p(\mathrm{SFI}) = p(\mathrm{LFI}) = 1/2$.

Fig. 5 shows that some models in the LFI class are fairly strongly disfavoured by the data (e.g., LFI_3 and LFI_4), while the model comparison is inconclusive in most other cases. One finds that the posterior probability for the SFI model class evaluates to $p(\mathrm{SFI}|d) \approx 0.77$. Therefore, the probability of the SFI class has increased from 50% in the prior to about 77% in the posterior, signalling a weak

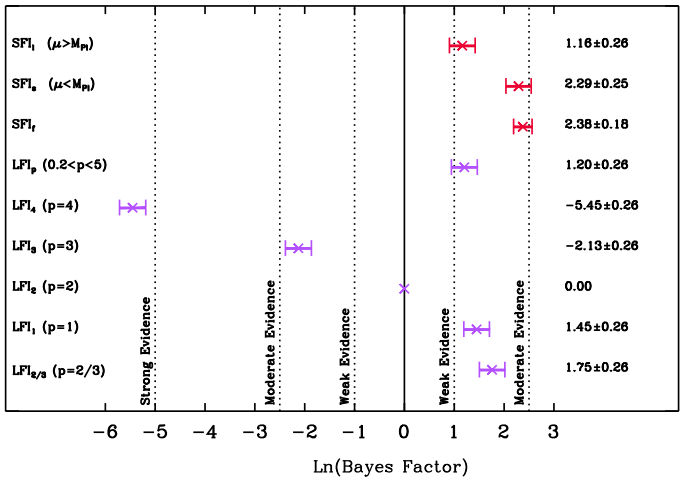


Figure 5: Results of Bayesian model comparison between 9 inflationary models (vertical axis), subdivided in two categories (SFI models and LFI models), from Ref. [12]. Errorbars reflect the 68% uncertainty on the value of the Bayes factor from the numerical evaluation performed with MultiNest.

preference for this type of models in the light of the data.

3.3 Challenges of profile likelihood evaluation

For highly non-Gaussian problems like supersymmetric (SUSY) parameter determination, inference can depend strongly on whether one chooses to work with the posterior distribution (Bayesian) or profile likelihood (frequentist) [11, 20, 21]. There is a growing consensus that both the posterior and the profile likelihood ought to be explored in order to obtain a fuller picture of the statistical constraints from present-day and future data. This begs the question of the algorithmic solutions available to reliably explore both the posterior and the profile likelihood in the context of SUSY phenomenology.

When Θ is composed of parameters of interest, θ, and nuisance parameters, ψ, the profile likelihood ratio is defined as

$$\lambda(\theta) \equiv \frac{\mathcal{L}(\theta, \hat{\hat{\psi}})}{\mathcal{L}(\hat{\theta}, \hat{\psi})}. \tag{12}$$

where $\hat{\hat{\psi}}$ is the conditional maximum likelihood estimate (MLE) of ψ with θ fixed and $\hat{\theta}, \hat{\psi}$ are the unconditional MLEs. The profile likelihood ratio defined in Eq. (12) is an attractive choice as a test statistics, for under certain regularity conditions, Wilks [22] showed that the distribution of $-2\ln\lambda(\theta)$ converges to a chi-square distribution with a number of degrees of freedom given by the dimensionality of θ. Clearly, for any given value of θ, evaluation of the profile likelihood requires solving a maximisation problem in many dimensions to determine the conditional MLE $\hat{\hat{\psi}}$. While posterior samples obtained with MultiNest have been used to estimate the profile likelihood, the accuracy of such an estimate has been questioned [23]. As mentioned above, evaluating profile likelihoods is much more challenging than evaluating posterior distributions. Therefore, one should not expect that a vanilla setup for MultiNest (which is adequate for an accurate exploration of the posterior distribution) will automatically be optimal for profile likelihoods evaluation. In Ref. [24] the question of the accuracy of profile likelihood evaluation from MultiNest was investigated in detail. We report below the main results.

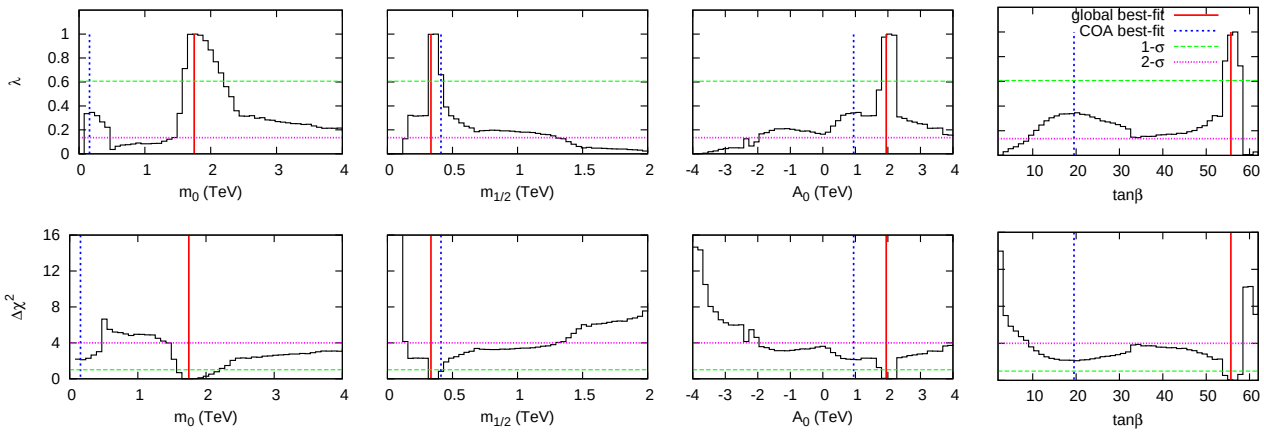


Figure 6: 1-D profile likelihoods from present-day data for the parameters of interest of a Supersymmetric model (so-called CMSSM), normalized to the global best-fit point. The red solid and blue dotted vertical lines represent the global best-fit point ($\chi^2 = 9.26$, located in the focus point region) and the best-fit point found in the stau co-annihilation region ($\chi^2 = 11.38$) respectively. The upper and lower panel show the profile likelihood and $\Delta\chi^2$ values, respectively. Green (magenta) horizontal lines represent the 1σ (2σ) approximate confidence intervals. MultiNest was run with 20,000 live points and tol $= 1 \times 10^{-4}$ (a configuration deemed appropriate for profile likelihood estimation), requiring approximately 11 million likelihood evaluations. From [24].

The two most important parameters that control the parameter space exploration in MultiNest are the number of live points N – which determines the resolution at which the parameter space is explored – and a tolerance parameter tol, which defines the termination criterion based on the accuracy of the evidence. Generally, a larger number of live points is necessary to explore profile likelihoods accurately. Moreover, setting tol to a smaller value results in MultiNest gathering a larger number of samples in the high likelihood regions (as termination is delayed). This is usually not necessary for the posterior distributions, as the prior volume occupied by high likelihood regions is usually very small and therefore these regions have relatively small probability mass. For profile likelihoods, however, getting as close as possible to the true global maximum is crucial and therefore one should set tol to a relatively smaller value. In Ref. [24] it was found that $N = 20,000$ and tol $= 1 \times 10^{-4}$ produce a sufficiently accurate exploration of the profile likelihood in toy models that reproduce the most important features of the parameter space of a popular supersymmetric scenario.

In principle, the profile likelihood does not depend on the choice of priors. However, in order to explore the parameter space using any Monte Carlo technique, a set of priors needs to be defined. Different choices of priors will generally lead to different regions of the parameter space to be explored in greater or lesser detail, according to their posterior density. As a consequence, the resulting profile likelihoods might be slightly different, purely on numerical grounds. We can obtain more robust profile likelihoods by simply merging samples obtained from scans with different choices of Bayesian priors. This does not come at a greater computational cost, given that a responsible Bayesian analysis would estimate sensitivity to the choice of prior as well. The results of such a scan are shown in Fig. 6, which was obtained by tuning MultiNest with the above configuration, appropriate for an accurate profile likelihood exploration, and by merging the posterior samples from two different choices of priors (see [24] for details). This high-resolution profile likelihood scan using MultiNest compares favourably with the results obtained by adopting a dedicated Genetic Algorithm technique [23], although at a slightly higher computational cost (a factor of ~ 4). In general, an accurate profile

likelihood evaluation was about an order of magnitude more computationally expensive than mapping out the Bayesian posterior.

4 Conclusions

We have given a short overview of the capabilities of the MultiNest algorithm, which has been used very successfully in a variety of problems to (a) reconstruct multi-modal posterior distributions; (b) evaluate the model likelihood for Bayesian model comparison and (c) approximate the profile likelihood in the challenging case of multi-modal likelihoods in many dimensions. Current statistical challenges in cosmology and astroparticle physics still present many open questions, and no doubt this will provide fertile ground for an increased collaboration between physicists and statisticians in the burgeoning field of astrostatistics.

Acknowledgments. We would like to thank Joseph Hilbe for organizing the first Astrostatistics session at an ISI World Congress and for his excellent efforts at getting the Astrostatistics section organized within ISI. We would also like to thank our colleagues for many stimulating discussions over the years: J. Berger., R. Cousin, K. Cranmer, A. Jaffe, M. Kunz, O. Lahav, A. Liddle, T. Loredo, L. Lyons and D. van Dyk.

References

[1] *Bayesian Methods in Cosmology*, M. Hobson *et al.*, (Eds), Cambridge University Press (2010).

[2] R. Trotta, Contemp. Phys. **49**, 71 (2008)

[3] Sellke, T., Bayarri, M.J. and Berger, J. (2001) The American Statistician (55), 62–71.

[4] H. Jeffreys, *Theory of probability*, 3rd edn, Oxford Classics series (reprinted 1998) (Oxford University Press, Oxford, UK, 1961).

[5] J. Skilling, Nested sampling, in R. Fischer, R. Preuss and U. von Toussaint (Eds) *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 735 (Amer. Inst. Phys. conf. proc. 2004), pp. 395–405.

[6] J. Skilling, Bayesian Analysis **1** 833–861 (2006).

[7] P. Mukherjee, D. Parkinson and A. R. Liddle, Astrophys. J. **638** L51–L54 (2006).

[8] R. Shaw, M. Bridges and M. P. Hobson, Mon. Not. Roy. Astron. Soc. **378** 1365–1370 (2007).

[9] F. Feroz and M. P. Hobson, Mon. Not. Roy. Astron. Soc., 384, 2, 449-463 (2008).

[10] F. Feroz, M. P. Hobson, M. Bridges, [arXiv:0809.3437 [astro-ph]].

[11] R. Trotta, F. Feroz, M. P. Hobson, L. Roszkowski and R. Ruiz de Austri, JHEP **0812**, 024 (2008)

[12] J. Martin, C. Ringeval, R. Trotta, Phys. Rev. D83, 063524 (2011).

[13] F. Feroz, S. T. Balan, M. P. Hobson, Mon. Not. Roy. Astron. Soc., in press (2011).

[14] F. Feroz, M. P. Hobson, J. T. L. Zwart, R. D. E. Saunders, K. J. B. Grainge, [arXiv:0811.1199 [astro-ph]].

[15] F. Feroz, B. C. Allanach, M. Hobson, S. S. AbdusSalam, R. Trotta, A. M. Weber, JHEP **0810**, 064 (2008).

[16] G. Bertone, D. G. Cerdeno, M. Fornasa, R. R. de Austri, R. Trotta, Phys. Rev. **D82**, 055008 (2010).

[17] Feroz F., Gair J. R., Graff P., Hobson M. P., Lasenby A., 2010, Classical and Quantum Gravity, 27, 075010

[18] R. Trotta, G. Johannesson, I. V. Moskalenko, T. A. Porter, R. R. de Austri, A. W. Strong, Astrophys. J. **729**, 106 (2011).

[19] Babak S. et al., 2010, Classical and Quantum Gravity, 27, 084009

[20] B. C. Allanach, K. Cranmer, C. G. Lester, and A. M. Weber, *JHEP* **0708** (2007) 023.

[21] P. Scott, J. Conrad, J. Edsjö, L. Bergström, C. Farnier, and Y. Akrami, JCAP **1001**, 031 (2010).

[22] S. Wilks, *Ann. Math. Statist.* **9** (1938) 60–2.

[23] Y. Akrami, P. Scott, J. Edsjo, J. Conrad, and L. Bergstrom, *JHEP* **04** (2010) 057.

[24] F. Feroz, K. Cranmer, M. Hobson, R. Ruiz de Austri, R. Trotta, (2011), pre-print: [http://xxx.lanl.gov/abs/1101.3296].