Bayesian Object Detection in Astrophysics

Farhan Feroz[1,2] and Mike Hobson[1]
[1]Astrophysics Group, Cavendish Laboratory, Cambridge CB3 0HE, UK
[2]Corresponding author: Farhan Feroz, e-mail: f.feroz@mrao.cam.ac.uk

Abstract

Detecting objects from noisy data-sets is common practice in astrophysics. Object detection presents a particular challenge in terms of statistical inference, not only because of its multi-modal nature but also because it combines both the parameter estimation (for characterizing objects) and model selection problems (in order to quantify the detection). Bayesian inference provides a mathematically rigorous solution to this problem by calculating marginal posterior probabilities of models with different number of objects, but the use of this method in astrophysics has been hampered by the computational cost of evaluating the Bayesian evidence. Nonetheless, Bayesian model selection has the potential to improve the interpretation of existing observational data. In this work we discuss several Bayesian approaches to object detection problems and describe how the statistical inference on them can be done in an efficient and robust manner. We also describe some recent applications of Bayesian object detection to problems like galaxy cluster and extra-solar planet detection. These approaches are generic in nature and may therefore be applied beyond astrophysics.

Key Words: statistical inference, data analysis, nested sampling

1 Introduction

Identification and characterisation of discrete objects immersed in some general background of diffuse signal, instrumental noise or systematic effects, is a long-standing problem in astrophysics. Bayesian inference provides an elegant solution for detecting and characterising all the objects in an image simultaneously by exploring the joint posterior distribution of all the parameters in the model used to describe them. Bayesian inference also provides a rigorous way of performing model selection required to determine the number of objects favoured by the data. The main problem in applying such Bayesian model selection techniques is the computational cost involved in calculating the Bayesian evidence. However, recent advances in Markov-Chain Monte Carlo (MCMC) techniques have made it possible for Bayesian techniques to be applied to astrophysical object detection. In this paper, we present several approaches for performing objection detection and review a few applications of these approaches in astrophysics.

2 Bayesian inference

Bayesian inference provides a consistent approach to the estimation of a set of parameters Θ in a model (or hypothesis) H for the data D. Bayes' theorem states that

$$\Pr(\Theta|D,H) = \frac{\Pr(D|\Theta,H)\Pr(\Theta|H)}{\Pr(D|H)}, \tag{1}$$

where $\Pr(\Theta|D,H) \equiv P(\Theta|D)$ is the posterior probability distribution of the parameters, $\Pr(D|\Theta,H) \equiv \mathcal{L}(\Theta)$ is the likelihood, $\Pr(\Theta|H) \equiv \pi(\Theta)$ is the prior, and $\Pr(D|H) \equiv \mathcal{Z}$ is the Bayesian evidence given by:

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d^N\Theta, \tag{2}$$

where N is the dimensionality of the parameter space. Bayesian evidence being independent of the parameters, can be ignored in parameter estimation problems and inferences can be obtained by taking samples from the (unnormalized) posterior distribution using standard MCMC methods.

Model selection between two competing models H_0 and H_1 can be done by comparing their respective posterior probabilities given the observed data-set D, as follows

$$R = \frac{\Pr(H_1|D)}{\Pr(H_0|D)} = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D|H_0)\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)}, \tag{3}$$

where $\Pr(H_1)/\Pr(H_0)$ is the prior probability ratio for the two models, which can often be set to unity in situations where there is not a prior reason for preferring one model over the other, but occasionally requires further consideration. It can be seen from Eq. (3) that the Bayesian evidence plays a central role in Bayesian model selection.

As the average of the likelihood over the prior, the evidence is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values, even if the likelihood function is very highly peaked. Thus, the evidence automatically implements Occam's razor.

Evaluation of the multidimensional integral in Eq. (2) is a challenging numerical task. Standard techniques like thermodynamic integration are extremely computationally expensive which makes evidence evaluation at least an order of magnitude more costly than parameter estimation. Various alternative information criteria for astrophysical model selection are discussed by Liddle (2007), but the evidence remains the preferred method.

The nested sampling approach, introduced by Skilling (2004), is a Monte Carlo method targeted at the efficient calculation of the evidence, but also produces posterior inferences as a by-product. Feroz & Hobson (2008) and Feroz et al. (2009) built on this nested sampling framework and have introduced the `MultiNest` algorithm which is very efficient in sampling from posteriors that may contain multiple modes and/or large (curving) degeneracies and also calculates the evidence. This technique has greatly reduces the computational cost of Bayesian parameter estimation and model selection and has already been applied to several object detection problems in astrophysics (see e.g. Feroz et al. 2008, 2009a,b).

3 Bayesian Object Detection

To detect and characterise an unknown number of objects in a dataset, one would ideally like to infer simultaneously the full set of parameters $\Theta = \{N_{\rm obj}, \Theta_1, \Theta_2, \cdots, \Theta_{N_{\rm obj}}, \Theta_{\rm n}\}$, where $N_{\rm obj}$ is the (unknown) number of objects, $\Theta_{\rm i}$ are the parameters values associated with the ith object, and $\Theta_{\rm n}$ is the set of (nuisance) parameters common to all the objects. This however requires any sampling based approach to move between spaces of different dimensionality as the length of the parameter vector depends on the unknown value of N_{obj}. Such techniques are discussed in Hobson & McLachlan (2003) and Brewer et al. (2012). Nevertheless, due to this additional complexity of variable dimensionality, these techniques are generally extremely computationally intensive.

An alternative approach for achieving virtually the same result is the 'multiple source model'. By considering a *series* of models $H_{N_{\rm obj}}$, each with a *fixed* number of objects, i.e. with $N_{\rm obj} = 0, 1, 2, \ldots$. One then infers $N_{\rm obs}$ by identifying the model with the largest marginal posterior probability $\Pr(H_{N_{\rm obj}}|\mathbf{D})$. Assuming that there are $n_{\rm p}$ parameters per object and $n_{\rm n}$ (nuisance) parameters common to all the objects, for $N_{\rm obj}$ objects, there would be $N_{\rm obj} n_{\rm p} + n_{\rm n}$ parameters to be inferred, Along with this increase in dimensionality, the complexity of the problem also increases with $N_{\rm obj}$ due to exponential increase

in the number of modes as a result of counting degeneracy (there are $n!$ more modes for $N_{\rm obj} = n$ than for $N_{\rm obj} = 1$).

If the contributions to the data from each object are reasonably well separated and the correlations between parameters across objects is minimal, one can use the alternative approach of 'single source model' by setting $N_{\rm obj} = 1$ and therefore the model for the data consists of only a single object. This does not, however, restrict us to detecting only one object in the data. By modelling the data in such a way, we would expect the posterior distribution to possess numerous peaks, each corresponding to the location of one of the objects. Consequently the high dimensionality of the problem is traded with high multi-modality in this approach, which, depending on the statistical method employed for exploring the parameter space, could potentially simplify the problem enormously. For an application of this approach in detecting galaxy cluster from weak lensing data-sets see Feroz et al. (2008).

Calculating Bayesian evidence accurately for large number of objects is extremely difficult, due to increase in dimensionality and severe complexity of the posterior, however, parameter estimation can still be done accurately. In order to circumvent this problem, Feroz et al. (2011b) proposed a new general approach to Bayesian object detection called 'residual data model' that is applicable even for systems with a large number of planets. This method is based on the analysis of residual data after detection of $N_{\rm obj}$ objects. We summarize this method as follows:

Let $H_{N_{\rm obj}}$ denote a model with $N_{\rm obj}$ objects. The observed (fixed) data is denoted by $\mathbf{D} = \{d_1, d_2, \cdots, d_{\rm M}\}$, with the associated uncertainties being $\{\sigma_1, \sigma_2, \cdots, \sigma_{\rm M}\}$. In the general case that $N_{\rm obj} = n$, the random variable $\mathbf{D}_n$ is defined as the data that would be collected if the model H_n were correct, and the random variable $\mathbf{R}_n \equiv \mathbf{D} - \mathbf{D}_n$, as the data residuals in this case. If we set $N_{\rm obj} = n$ and analyse $\mathbf{D}$ to obtain samples from the posterior distribution of the model parameters Θ from which it is straightforward to obtain samples from the posterior distribution of the data residuals $\mathbf{R}_n$. This is given by:

$$P(\mathbf{R}_n|\mathbf{D}, H_n) = \int P(\mathbf{R}_n|\Theta, H_n) P(\Theta|\mathbf{D}, H_n)\, d\Theta, \tag{4}$$

where

$$P(\mathbf{R}_n|\Theta, H_n) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{[D_i - R_i - D_{{\rm p},i}(\Theta)]^2}{2\sigma_i^2} \right\}, \tag{5}$$

and $\mathbf{D}_{\rm p}(\Theta)$ is the (noiseless) predicted data-set corresponding to the parameter values Θ. Assuming that the residuals are independently Gaussian distributed with mean $\langle \mathbf{R}_n \rangle = \{r_1, r_2, \cdots, r_{\rm M}\}$ and standard deviations $\{\sigma'_1, \sigma'_2, \cdots, \sigma'_{\rm M}\}$ obtained from the posterior samples, $\langle \mathbf{R}_n \rangle$ can then be analysed with $N_{\rm obj} = 0$, giving the 'residual null evidence' $Z_{{\rm r},0}$, which is compared with the evidence value $Z_{{\rm r},1}$ obtained by analysing $\langle \mathbf{R}_n \rangle$ with $N_{\rm obj} = 1$. The comparison is thus being made between the model H_0 that the residual data does not contain an additional object and the model H_1 in which an additional object is favoured.

With no prior information about the number of objects in a data-set, the original data-set $\mathbf{D}$ is analysed with $N_{\rm obj} = 1$. If, in the analysis of the corresponding residuals data, H_1 is favoured over H_0, then the original data $\mathbf{D}$ are analysed with $N_{\rm obj} = 2$ and the same process is repeated. In this way, $N_{\rm obj}$ is increased in the analysis of the original data $\mathbf{D}$, until H_0 is favoured over H_1 in the analysis of the corresponding residual data. The resulting value for $N_{\rm obj}$ gives the number of objects favoured by the data. This approach thus requires the detection and estimation of orbital parameters for $N_{\rm obj} = n$ model but Bayesian evidence only needs to be calculated for $N_{\rm obj} = 1$ model (and the $N_{\rm obj} = 0$ model, which is trivial); this reduces the computational cost of the problem significantly.

4 Applications of Bayesian Object Detection in Astrophysics

4.1 Modelling of Galaxy Clusters

Clusters of galaxies are the most massive gravitationally bound objects in the universe and, as such, are critical tracers of the formation of large-scale structure. The number counts of clusters as a function of their mass and redshift have been predicted both analytically (see e.g. Press & Schechter 1974) and from large scale numerical simulations (see e.g. Jenkins et al. 2001), and are particularly sensitive to the cosmological parameters σ_8 (amplitude of power spectrum on the scale $8h^{-1}$Mpc) and $\Omega_{\rm m}$ (matter density normalized to critical density) (see e.g. Battye & Weller 2003).

4.1.1 SZ Effect

The SZ effect (Sunyaev & Zeldovich 1970, 1972) produces secondary anisotropies in the cosmic microwave background (CMB) radiation through inverse-Compton scattering from the electrons in the hot intracluster gas (which radiates via thermal Bremsstrahlung in the X-ray waveband), and the subsequent transfer of some of the energy of the electrons to the low-energy photons. Apart from the receiver noise, there are numerous other sources of noise in SZ observations, including the primordial CMB which can mimic galaxy clusters, resolved and unresolved radio sources. The presence of these noise components make analysis of SZ observations quite challenging as one is not only required to characterize any galaxy clusters present, but also to quantify its detection.

Feroz et al. (2009a) presented a Bayesian approach to modelling galaxy clusters using multi-frequency observations from telescopes that exploit the SZ effect using `MultiNest` to explore the high-dimensional parameter spaces and also to calculate the Bayesian evidence. By performing tests on simulated Arcminute Microkelvin Imager (AMI; AMI Consortium: Zwart et al. 2008) observations of a cluster in the presence of primary CMB signal, radio point sources (detected as well as an unresolved background) and receiver noise, they showed that the algorithm is able to analyse jointly the data from six frequency channels, sample the posterior space of the model and calculate the Bayesian evidence very efficiently. This technique has since been used in several studies of real SZ observations (see e.g. Zwart et al. 2011; AMI Consortium: Rodriguez-Gonzalvez et al. 2011) and has also resulted in the discovery of previously unknown galaxy clusters (see e.g. AMI Consortium: Shimwell et al. 2012).

4.1.2 Weak Gravitational Lensing

Observations of galaxy clusters through gravitational lensing exploit the fact that the spacetime around a massive object is curved, and as a result light rays from a background source (e.g. galaxies), propagating through the spacetime are bent. This results in magnification and distortion of the images of background sources. Gravitational lensing is classified as 'strong' when these distortions are easily visible, and 'weak' when they are much smaller and can only be studied by averaging over large number of background sources. A cluster mass distribution is investigated using weak gravitational lensing through the relationship $\langle \epsilon(x) \rangle = g(x)$, that is, at any point x on the sky, the local average of the complex ellipticities $\epsilon = \epsilon_1 + i\epsilon_2$ of a collection of background galaxy images is an unbiased estimator of the local complex reduced shear field, $g = g_1 + ig_2$, due to the cluster.

The quantification of cluster detection is extremely important in weak lensing analyses as despite the advances in data quality, the weak lensing data remains very sparse and noisy. By adopting the 'single source model' (see Sec. 3), and using `MultiNest` to explore the highly multi-modal parameter space, Feroz et al. (2008) analysed simulated

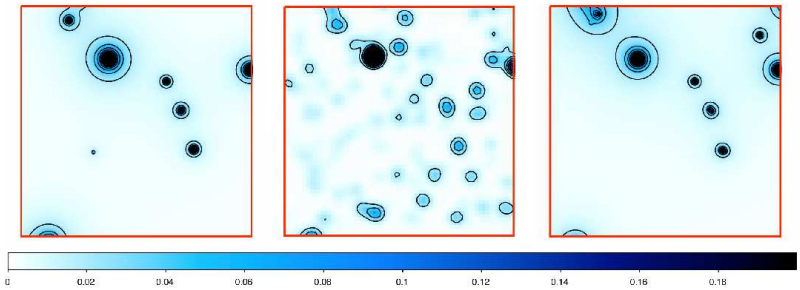

Figure 1: Galaxy cluster detection from weak gravitational lensing observations. Left panel shows the true convergence (projected mass density) map of the simulated clusters discussed in Section 3.1 of (Feroz et al., 2008). Middle panel shows the denoised convergence map made with the LenEnt2 algorithm. Right panel shows the inferred convergence map made with the `MultiNest` algorithm.

wide field weak lensing data with hundreds of clusters and extracted many potential galaxy clusters as well as calculated the probability odds for each detected cluster being 'true'. An example of reconstructed mass map from simulated weak lensing observations using this technique is shown in Fig. 1.

An important feature of using the Bayesian model selection for quantifying cluster detection is that it gives the probability distribution of detected clusters being 'true'. Once the ratio $R_{\rm i}$ of the probabilities of the $i^{\rm th}$ detected clusters being 'true' and 'false' has been calculated as given in Eq. (3), the probability $p_{\rm i}$, of this cluster being 'true' can then be calculated as $p_{\rm i} = R_{\rm i}/(1+R_{\rm i})$. Given a 'threshold probability' $p_{\rm th}$, defined such that detected clusters with $p_{\rm i} \geq p_{\rm th}$ are identified as candidate clusters, the expected number of *false positives*, $\hat{n}_{\rm FP}$ can then be calculated as,

$$\hat{n}_{\rm FP} = \sum_{\rm i=1,p_i \geq p_{th}}^{\rm k} (1-p_{\rm i}), \tag{6}$$

where k is total number of detected clusters. The expected 'purity', defined as the fraction of the cluster candidates that are expected to be 'true' can be similarly calculated. The choice of $p_{\rm th}$ would depend on the application. See Feroz et al. (2008) and Karpenka et al. (2013) for examples of such analyses.

4.2 Exoplanetary Searches

Exoplanetary research has made great advances in the last decade as a result of the data gathered by several ground and space based telescopes and so far more than 800 exoplanets have been discovered. One of the most successful methods for detecting exoplanets is the so-called 'Radial Velocity' (RV) method. The gravitational force between the planets and their host star results in the planets and star revolving around their common centre of mass. This produces doppler shifts in the spectrum of the host star according to its RV, the velocity along the line-of-sight to the observer. Several such measurements, usually over an extended period of time, can then be used to detect extrasolar planets.

Feroz et al. (2011a) adopted the 'multiple source model' (see Sec. 3) to determine the number of companions orbiting star HIP 5158. By analysing high-precision RV measurements of HIP 5158, they found conclusive evidence for the presence of two companions and estimated their orbital parameters. Owing to the large uncertainty on the mass of the second companion, they were unable to determine whether it is a planet or a brown dwarf. They also analysed a three-companion model, but found it to be e^8 times less probable than the two-companion model.

For systems with 4 or more planets, calculating Bayesian evidence accurately is extremely difficult, however, parameter estimation can still be done accurately. Adopting the 'residual data model' (see Sec. 3), Feroz et al. (2011b) analysed the RV measurements of HD 10180 using `MultiNest`. RV analysis of HD 10180 by Lovis et al. (2011) reported at least five and as many as seven planets orbiting this star. Feroz et al. (2011b) found strong evidence for the presence of six planets orbiting HD 10180 and although their analysis of the residual data of the 6-planet model did reveal several peaks in the posterior distribution with periods around 6.51 and 1 days, they were not found to be sufficiently significant (residual data after detection of 6 planets favouring an additional planet over no additional planets only by a factor $e^{-0.7}$). They therefore ruled out the presence of a 'detectable' seventh planet in the data.

5 Conclusions

With the availability of vast amounts of high quality data, statistical inference is increasingly playing an important role in astrophysics. MCMC techniques and algorithms based on nested sampling have been employed successfully in the detection of different classes of astrophysical objects. We have given a short introduction to Bayesian object detection and reviewed some of its applications in astrophysics. With the forthcoming data from future missions, there will be no doubt many more exciting areas of research which will foster further development of this inter-disciplinary field.

References

AMI Consortium: Rodriguez-Gonzalvez C., et al., 2011, MNRAS, 414, 3751
AMI Consortium: Shimwell A. C. T. W., et al., 2012, MNRAS, 423, 1463
AMI Consortium: Zwart J. T. L., et al., 2008, MNRAS, 391, 1545
Battye R. A., Weller J., 2003, Phys.Rev.D, 68, 083506
Brewer B. J., Foreman-Mackey D., Hogg D. W., 2012, arXiv e-prints [arXiv:1211.5805]
Feroz F., Balan S. T., Hobson M. P., 2011a, MNRAS, pp L295+
Feroz F., Balan S. T., Hobson M. P., 2011b, MNRAS, pp 1007–+
Feroz F., Gair J. R., Hobson M. P., Porter E. K., 2009b, Classical and Quantum Gravity, 26, 215003
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009c, MNRAS, 398, 1601
Feroz F., Hobson M. P., Zwart J. T. L., Saunders R. D. E., Grainge K. J. B., 2009a, MNRAS, 398, 2049
Feroz F., Marshall P. J., Hobson M. P., 2008, arXiv e-prints [arXiv:0810.0781]
Hobson M. P., McLachlan C., 2003, MNRAS, 338, 765
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Karpenka N. V., Feroz F., Hobson M. P., 2013, MNRAS, 429, 1278
Liddle A. R., 2007, MNRAS, 377, L74
Lovis C., et al., 2011, A&A, 528, A112+
Press W. H., Schechter P., 1974, ApJ, 187, 425
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, American Institute of Physics Conference Series Nested Sampling. pp 395–405
Sunyaev R. A., Zeldovich Y. B., 1970, CoASP, 2, 66
Sunyaev R. A., Zeldovich Y. B., 1972, CoASP, 4, 173
Zwart J. T. L., et al., 2011, MNRAS, 418, 2754