Astronomy & Astrophysics

On the shape of the mass-function of dense clumps in the Hi-GAL fields

II. Using Bayesian inference to study the clump mass function

L. Olmi[1,2], D. Anglés-Alcázar[3], D. Elia[4], S. Molinari[4], M. Pestalozzi[4], S. Pezzuto[4], E. Schisano[5,4], L. Testi[6,1], and M. Thompson[7]

[1] INAF, Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
e-mail: olmi.luca@gmail.com
[2] University of Puerto Rico, Rio Piedras Campus, Physics Dept., Box 23343, UPR station, San Juan, Puerto Rico, USA
[3] Department of Physics, University of Arizona, 1118 E. 4th Street, Tucson AZ 85721, USA
[4] INAF, Istituto di Astrofisica e Planetologia Spaziali, via Fosso del Cavaliere 100, 00133 Roma, Italy
[5] Infrared Processing and Analysis Center, California Institute of Technology, Pasadena CA 91125, USA
[6] ESO, Karl-Schwarzschild-Str. 2, 85748 Garching bei Munchen, Germany
[7] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK

ABSTRACT

Context. Stars form in dense, dusty clumps of molecular clouds, but little is known about their origin, their evolution, and their detailed physical properties. In particular, the relationship between the mass distribution of these clumps (also known as the "clump mass function", or CMF) and the stellar initial mass function (IMF) is still poorly understood.
Aims. To better understand how the CMF evolve toward the IMF and to discern the "true" shape of the CMF, large samples of bona-fide pre- and proto-stellar clumps are required. Two such datasets obtained from the *Herschel* infrared GALactic Plane Survey (Hi-GAL) have been described in Paper I. Robust statistical methods are needed to infer the parameters describing the models used to fit the CMF and to compare the competing models themselves.
Methods. In this paper, we apply Bayesian inference to the analysis of the CMF of the two regions discussed in Paper I. First, we determine the posterior probability distribution for each of the fitted parameters. Then, we carry out a quantitative comparison of the models used to fit the CMF.
Results. We have compared several methods of sampling posterior distributions and calculating global likelihoods, and we have also analyzed the impact of the choice of priors and the influence of various constraints on the statistical conclusions for the values of model parameters. We find that both parameter estimation and model comparison depend on the choice of parameter priors.
Conclusions. Our results confirm our earlier conclusion that the CMFs of the two Hi-GAL regions studied here have very similar shapes but different mass scales. Furthermore, the lognormal model appears to better describe the CMF measured in the two Hi-GAL regions studied here. However, this preliminary conclusion is dependent on the choice of parameter priors.

Key words. stars: formation – ISM: clouds – methods: data analysis – methods: statistical

1. Introduction

Stars form in dense, dusty cores, or clumps[1], of molecular clouds, but the physical processes that regulate the transition from molecular clouds/clumps to (proto)stars are still being debated. In particular, the relationship between the mass distribution of molecular clumps (also known as the clump mass function, or CMF[2]) and the stellar initial mass function (IMF) is poorly understood (McKee & Ostriker 2007). To improve our understanding of this relationship, it is necessary to undertake the study of statistically significant samples of pre- and proto-stellar clumps.

The *Herschel* infrared GALactic Plane Survey (Hi-GAL), a key program of the *Herschel* Space Observatory (HSO) that has been carrying out a 5-band photometric imaging survey at 70, 160, 250, 350, and 500 μm of a $|b| \leq 1°$-wide strip of the Milky Way Galactic plane (Molinari et al. 2010), is now providing us with large samples of starless and proto-stellar clumps in a variety of star-forming environments. In the first paper (Olmi et al. 2013, Paper I, hereafter), we gave a general description of the Hi-GAL data and described the source extraction and photometry techniques. We also determined the spectral energy distributions and performed a statistical analysis of the CMF in the two regions mapped by HSO during its science demonstration phase (SDP). The two SDP fields were centered at $\ell = 59°$ and $\ell = 30°$, and the final maps spanned $\simeq 2°$ in both Galactic longitude and latitude.

The goal of this second paper is twofold. On one side, we build on the premises of Paper I, and apply a Bayesian analysis to the CMF of the two SDP fields. First, we determine the posterior probability distribution of the parameters specific to each of the

---

[1] Here, the term "clump" refers to any compact density enhancement (generally $\gtrsim 0.1$–1 pc in diameter) which is identified by the source-extraction algorithm and may be composed of sub-structures, or "cores", when observed at higher angular resolution.
[2] Throughout this paper, CMF is referred to "clump mass function" instead of "core mass function".

CMF models analyzed in this work, i.e., powerlaw and lognormal. Next, we carry out a quantitative comparison of these models by using the given data and an explicit set of assumptions.

The other major aim of our paper is to compare the results of several popular methods, which sample posterior probability distributions and compute global likelihoods, to highlight the effects that different algorithms may have on the results. In particular, we are interested in analyzing the impact of the choice of *priors* and the influence of various constraints on the statistical conclusions for the values of model parameters.

The outline of the paper is thus as follows: in Sect. 2, we summarize the models used to describe the mass distribution. In Sect. 3, we give a general description of Bayesian inference and how it is applied to the analysis of the CMF. We describe the algorithms used in Sect. 4 and discuss our results in Sect. 5. We finally draw our conclusions in Sect. 6.

2. Description of models used to fit the CMF

In the following sub-sections we give a short description of the mathematical functions used in this analysis, but the reader should refer to Paper I for more details.

2.1. Definitions

We start by defining the CMF, $\xi(M)$, in the general case of a *continuous* distribution. If dN represents the number of objects of mass M lying between M and $M + \mathrm{d}M$, then we can define the number density distribution per mass interval, d$N/\mathrm{d}M$ with the relation (Chabrier 2003):

$$\xi(M) = \frac{\mathrm{d}N}{\mathrm{d}M} = \frac{\xi(\log M)}{M \ln 10} = \left(\frac{1}{M \ln 10}\right) \frac{\mathrm{d}N}{\mathrm{d}\log M}; \tag{1}$$

thus, $\xi(M)\mathrm{d}M$ represents the number of objects with mass M lying in the interval $[M, M + \mathrm{d}M]$. The *probability* of a mass falling in the interval $[M, M + \mathrm{d}M]$ can be written for a continuous distribution as $p(M)\mathrm{d}M$, where $p(M)$ represents the mass probability density function (or distribution, PDF). For the case of *discrete* data, $p(M)$ can be written as

$$p(M) = \frac{\xi(M)}{N_{\mathrm{tot}}}, \tag{2}$$

where N_{tot} represents the total number of objects being considered in the sample. The PDF and CMF must obey the following normalization conditions (which we write here for continuous data):

$$\int_{M_{\mathrm{inf}}}^{M_{\mathrm{sup}}} p(M)\mathrm{d}M = 1 \text{ and}$$
$$\int_{M_{\mathrm{inf}}}^{M_{\mathrm{sup}}} \xi(M)\mathrm{d}M = N_{\mathrm{tot}} \tag{3}$$

where M_{inf} and M_{sup} denote, respectively, the inferior and superior limits of the mass range for the objects in the sample beyond which the distribution does not follow the specified behavior.

2.2. Powerlaw form

The most widely used functional form for the CMF is the powerlaw:

$$\xi_{\mathrm{pw}}(\log M) = A_{\mathrm{pw}} M^{-\alpha}, \text{ or} \tag{4}$$

$$\xi_{\mathrm{pw}}(M) = \frac{A_{\mathrm{pw}}}{\ln 10} M^{-\alpha-1}, \tag{5}$$

where A_{pw} is the normalization constant. The original Salpeter value for the IMF is $\alpha = 1.35$ (Salpeter 1955).

The PDF of a powerlaw (continuous) distribution is given by (Clauset et al. 2009):

$$p_{\mathrm{pw}}(M) = C_{\mathrm{pw}} M^{-\alpha-1}, \tag{6}$$

where the normalization constant can be approximated as $C_{\mathrm{pw}} \simeq \alpha M_{\mathrm{inf}}^{\alpha}$, if $\alpha > 0$ and $M_{\mathrm{sup}} \gg M_{\mathrm{inf}}$ (see Paper I and references therein). As described later in Sect. 3, Bayesian inference provides a technique to estimate the probability distribution of the model parameters α and M_{inf}.

2.3. Lognormal form

The continuous lognormal CMF can be written (e.g., Chabrier 2003) as

$$\xi_{\mathrm{ln}}(\ln M) = \frac{A_{\mathrm{ln}}}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(\ln M - \mu)^2}{2\sigma^2}\right], \tag{7}$$

where μ and $\sigma^2 = \langle(\ln M - \langle\ln M\rangle)^2\rangle$ denote, respectively, the mean mass and the variance in units of $\ln M$ and A_{ln} represents a normalization constant (see Paper I).

The PDF of a continuous lognormal distribution can be written as (e.g., Clauset et al. 2009)

$$p_{\mathrm{ln}}(M) = \frac{C_{\mathrm{ln}}}{M} \exp\left[-x^2\right], \tag{8}$$

where we have defined the variable $x(M) = (\ln M - \mu)/(\sqrt{2}\sigma)$. If the condition $M_{\mathrm{sup}} \gg M_{\mathrm{inf}}$ holds, the normalization constant, C_{ln}, can be approximated as (see Paper I)

$$C_{\mathrm{ln}} \simeq \sqrt{\frac{2}{\pi\sigma^2}} \times [\mathrm{erfc}(x_{\mathrm{inf}})]^{-1}, \tag{9}$$

where $x_{\mathrm{inf}} = x(M_{\mathrm{inf}})$. As already mentioned for the powerlaw case, the Bayesian inference allows us to estimate the probability distribution of the three model parameters μ, σ and M_{inf}.

3. Bayesian Inference

3.1. Overview of Bayesian methodology and prior information

Our main goal is to confront theories for the origin of the IMF using an analysis of the CMF data that provides information on the processes responsible for cloud fragmentation and clump formation. Bayesian inference allows the quantitative comparison of alternative models by using the data and an explicit set of assumptions. This last topic is known as *model selection*, or the problem of distinguishing competing models, which generally feature different numbers of free parameters.

Bayesian statistics also provides a mathematically well-defined framework that allows one to determine the posterior probability distribution of the parameters of a given model. As we have already seen in Sect. 2, the powerlaw model for the

CMF depends on two parameters, α and $M_{\inf}$, while the lognormal model contains three parameters, μ, σ, and $M_{\inf}$ (unless $M_{\inf}$ is considered a *fixed* parameter, see Sect. 5.1.2).

A very distinctive feature of Bayesian inference is that it deals quite naturally with *prior information* (for example, on the parameters of a given model), which is highly relevant in many cases, such as when the parameters of interest have a physical meaning that restricts their possible values (e.g., masses or positive quantities in general). The prior choice in Bayesian statistics has been regarded both as a weakness and as a strength. In principle, prior assignment becomes less important as better and better data make the posterior distribution of the parameters dominated by the likelihood of the data (see, e.g., Trotta 2008). However, more often the data are not strong enough to override the prior, in which case the final inference may depend on the prior choice. If different prior choices lead to different posteriors one should conclude that the data are not *informative* enough to completely override our prior state of knowledge. An analysis of the role of priors in cosmological parameter extraction and Bayesian cosmological model building has already been presented by Trotta et al. (2008).

The situation is even more critical in model selection. In this case, the impact of the prior choice is much stronger, and care should be exercised in assessing how much the outcome would change for physically reasonable changes in the prior (see Berger & Pericchi 2001; Pericchi 2005). In addition to being nonrobust with respect to the choice of parameter priors, Bayesian model selection also suffers from another deep difficulty, specifically with the computation of a quantity, known as the *global likelihood*, which is difficult to calculate to the required accuracy.

In this section, we first give a short introduction to Bayesian inference by reviewing the basic terminology and describing the most common prior types. We also briefly discuss model comparison and define global likelihood. The mathematical tools required to efficiently evaluate the global likelihood and their limitations are then discussed in the subsequent sections.

3.2. Definitions

Here we give a short list of defintions that are used later.

1. We denote a particular model by the letter $\mathcal{M}$. This particular model is characterized by Q parameters, which we denote by θ_q, $q = 1, ..., Q$ (with size Q dependent on the model). The set of θ_q constitutes the parameter vector $\boldsymbol{\theta}$. In this paper we consider two models for the CMF: the powerlaw ($Q = 2$, $\boldsymbol{\theta} = [\alpha, M_{\inf}]$) and the lognormal ($Q = 3$, $\boldsymbol{\theta} = [\mu, \sigma, M_{\inf}]$) models.

2. We denote the data by the letter D. In this work, the data consist of the observed CMF for the $\ell = 59°$ and $\ell = 30°$ Hi-GAL SDP fields (see Paper I). More restrictive selection criteria have been applied to ensure that all methods described in Sect. 4 did converge. Individual clump masses are denoted by M_i (not to be confused with the model, $\mathcal{M}$).

3. In the following, we use the *likelihood* of a given set of data D, or the combined probability that D would be obtained from model $\mathcal{M}$ and its set of parameters θ_q, which we denote by $P(D|\boldsymbol{\theta}, \mathcal{M})$ or $\mathcal{L}$. For the present case and for a set of data $D = \{M_i\}$, the likelihood can be written as

$$P(D|\boldsymbol{\theta}, \mathcal{M}) = \mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{N_{tot}} p(M_i|\boldsymbol{\theta}), \tag{10}$$

where it is assumed that the data are drawn from the PDF associated with model $\mathcal{M}$, as denoted by $p(M_i|\boldsymbol{\theta})$ (see Sect. 2).

4. In principle, the model parameters θ_q can take any value, unless we have some information limiting their range. We can constrain the expected ranges of parameter values by assigning the probability distributions of the unknown parameters θ_q. These are called the *parameters prior probability distribution* (often called simply the *parameters prior* and are denoted by $P(\theta_q|\mathcal{M})$. Hence:

$$P(\boldsymbol{\theta}|\mathcal{M}) = \prod_{q=1}^{Q} P(\theta_q|\mathcal{M}). \tag{11}$$

5. In contrast to traditional point estimation methods (e.g., maximum-likelihood estimation, or MLE) Bayesian inference does not provide specific estimates for the parameters. Rather, it provides a technique to estimate the *probability distribution* (assumed to be continuous) of each model parameter θ_q, also known as the *posterior* PDF, or simply the posterior distribution, $P(\boldsymbol{\theta}|D, \mathcal{M})$. In Bayesian statistics, the posterior distribution encodes the full information coming from the data and the prior, and it is given by the *Bayes theorem*:

$$P(\boldsymbol{\theta}|D, \mathcal{M}) = \frac{P(\boldsymbol{\theta}|\mathcal{M}) \, P(D|\boldsymbol{\theta}, \mathcal{M})}{\mathcal{P}(D|\mathcal{M})} = \frac{P(\boldsymbol{\theta}|\mathcal{M}) \, \mathcal{L}(\boldsymbol{\theta})}{\mathcal{P}(D|\mathcal{M})}, \tag{12}$$

where $\mathcal{P}(D|\mathcal{M})$ is a normalization factor and is often called *global likelihood* or *evidence* for the model,

$$\mathcal{P}(D|\mathcal{M}) = \int P(\boldsymbol{\theta}|\mathcal{M}) \, \mathcal{L}(\boldsymbol{\theta}) \, d\boldsymbol{\theta}. \tag{13}$$

Thus, the global likelihood of a model is equal to the weighted (by the parameters prior, $P(\boldsymbol{\theta}|\mathcal{M})$; also known as *marginalization* over the parameter space) average likelihood for its parameters, which is why it is also often called the *total marginalized likelihood*. We work mostly with logarithmic probabilities; thus Eq. (12) becomes

$$\ln P(\boldsymbol{\theta}|D, \mathcal{M}) = \text{const.} + \ln P(\boldsymbol{\theta}|\mathcal{M}) + \ln \mathcal{L}(\boldsymbol{\theta}) \tag{14}$$

where

$$\ln \mathcal{L}(\boldsymbol{\theta}) = \ln \prod_{i=1}^{N_{tot}} p(M_i|\boldsymbol{\theta}) = \sum_{i=1}^{N_{tot}} \ln p(M_i|\boldsymbol{\theta}). \tag{15}$$

3.3. Specifying the parameter priors

If we have some expectation of the ranges in which the parameter values lie, then we can incorporate this information in the parameter priors $P(\boldsymbol{\theta}|\mathcal{M})$. Even when parameters in a given range of values are equally probable, we can specify plausible bounds on parameters.

Here, we briefly introduce the two forms of priors most commonly used, the *uniform* and *Jeffreys'* priors. In Sect. 4, we then discuss the priors actually used in our computations.

1. When dealing with *scale parameters*[3], the preferred form is the Jeffreys' priors, which assign an equal probability per decade interval (appropriate for quantities that are scale invariant), and are given by (see Gregory 2005):

$$P(\theta_q|\mathcal{M}) = \begin{cases} \frac{1}{\theta_q \ln(\theta_q^{\max}/\theta_q^{\min})}, & \text{for } \theta_q^{\min} \le \theta_q \le \theta_q^{\max} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

where $[\theta_q^{\min}, \theta_q^{\max}]$ represents the range allowed for parameter θ_q to vary.

---

[3] A *scale parameter* is typically associated with the magnitude (scale) of an event and is always a positive quantity.

2. On the other hand, when dealing with *location parameters*[4], the preferred form is the uniform priors, which give uniform probability per arithmetic interval:

$$P(\theta_q|\mathcal{M}) = \begin{cases} \frac{1}{\theta_q^{\max}-\theta_q^{\min}}, & \text{for } \theta_q^{\min} \le \theta_q \le \theta_q^{\max} \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

3.4. Model comparison

In many cases, such as the present one, more than one parameterized model is available to explain a given set of data, and it is thus of interest to compare them. The models may differ in form and/or in number of parameters. Use of Bayes' theorem allows one to compare competing models by calculating the probability of each model as a whole. The equivalent form of Eq. (12) to calculate the posterior probability of a model $\mathcal{M}$, $P(\mathcal{M}|D, I)$, which represents the probability that model $\mathcal{M}$ has actually generated the data D, is the following (Gregory 2005):

$$P(\mathcal{M}|D, I) = \frac{P(\mathcal{M}|I)\,P(D|\mathcal{M}, I)}{P(D|I)}, \quad (18)$$

where I represents our prior information that one of the models under consideration is true. One can recognize $\mathcal{P}(D|\mathcal{M}, I)$ as the global likelihood for model $\mathcal{M}$, which can be calculated according to Eq. (13). The function $P(\mathcal{M}|I)$ represents the *prior probability* for model $\mathcal{M}$, while the term at the denominator $P(D|I)$ is again a normalization constant, which is obtained by summing the products of the priors and the global likelihoods of all models being considered.

Then, the plausibility of two different models $\mathcal{M}_1$ and $\mathcal{M}_2$, parameterized by the model parameters vectors θ_1 and θ_2, can be assessed by the *odds ratio*:

$$\frac{P(\mathcal{M}_2|D, I)}{P(\mathcal{M}_1|D, I)} = BF_{21}\,\frac{P(\mathcal{M}_2, I)}{P(\mathcal{M}_1, I)}, \quad (19)$$

which can be interpreted as the "odds provided by the data for model $\mathcal{M}_2$ versus $\mathcal{M}_1$". In Eq. (19) we have also introduced the *Bayes factor* (e.g., Gelfand & Dey 1994; Gregory 2005):

$$BF_{21} = \frac{\mathcal{P}(D|\mathcal{M}_2, I)}{\mathcal{P}(D|\mathcal{M}_1, I)} = \frac{\int P(\theta_2|\mathcal{M}_2)\,\mathcal{L}(\theta_2)\,d\theta_2}{\int P(\theta_1|\mathcal{M}_1)\,\mathcal{L}(\theta_1)\,d\theta_1}. \quad (20)$$

In general, it is also assumed that $P(\mathcal{M}_2, I) = P(\mathcal{M}_1, I)$ (i.e., no model is favored over the other); thus, the odds ratio becomes equal to the Bayes factor. A value of $BF_{21} > 1$ would indicate that the data provide evidence in favor of model $\mathcal{M}_2$ vs. the alternative model $\mathcal{M}_1$. Usually, the Bayes factor is quoted in ln units; that is, we change to $\ln(BF)$:

$$\mathcal{BF}_{21} = \ln(BF)_{21} = \ln\mathcal{P}(D|\mathcal{M}_2, I) - \ln\mathcal{P}(D|\mathcal{M}_1, I). \quad (21)$$

Then, $sgn(\mathcal{BF})$ indicates the most probable model with positive values of $\mathcal{BF}$ favoring model $\mathcal{M}_2$ and negative values favoring model $\mathcal{M}_1$.

---

[4] A *location parameter* is typically associated with the location of an event in space, which depends on our choice of origin. The *uniform prior* is invariant to a shift in location.

Thus, for each model, we must compute the global likelihoods, which means evaluating the integrals in Eq. (20), which can be written in general as[5]

$$\mathcal{P}(D|\mathcal{M}_j, I) = \int \mathcal{L}(\theta_j)P(\theta_j|\mathcal{M}_j)\,d\theta_j$$

$$= \int \left[\prod_{i=1}^{N_{\text{tot}}} p_j(M_i|\theta_j)\right]\left[\prod_{q=1}^{Q} P(\theta_{jq}|\mathcal{M}_j)\right] d\theta_j \quad (22)$$

$$j = 1, 2 \quad q = 1, ..., Q_j.$$

The above equation can be written more explicitly, for example, in the powerlaw case and using Jeffreys priors for both parameters as

$$\mathcal{P}(D|\mathcal{M}_{\text{pw}}, I) =$$

$$\int \left[\prod_{i=1}^{N_{\text{tot}}} C_{\text{pw}} M_i^{-\alpha}\right] P(\theta_{\text{pw}}|\mathcal{M}_{\text{pw}})\,d\alpha\,dM_{\text{inf}}, \quad \text{with} \quad (23)$$

$$P(\theta_{\text{pw}}|\mathcal{M}_{\text{pw}}) = \frac{1}{M_{\text{inf}}\,\ln(M_{\text{inf}}^{\max}/M_{\text{inf}}^{\min})} \times \frac{1}{\alpha\,\ln(\alpha_{\max}/\alpha_{\min})},$$

where we have used Eq. (6) and C_{pw} has been given in Sect. 2.2.

4. Computation of model parameters and global likelihood

4.1. General constraints

We now turn to the description of various methods that allow the computation of the posterior distributions of the model parameters which, in some cases, also allow to estimate the global likelihood as a by-product. Several open software resources exist that can perform the computation of model parameters. However, estimating the multidimensional integrals in equations of the type (22) and (23) is impossible when done analytically in most cases, and is otherwise computationally very intensive when done numerically. Therefore, popular statistical packages (e.g., WinBUGS[6], $\mathcal{R}$[7]) are usually not able to compute the global likelihood directly, and more specialized programs or dedicated procedures must be used.

As discussed in greater detail in Sect. 5, we note that all methods described below have proved to be sensitive, to various degrees, to both the priors type and range, as well as to other parameters specific to each algorithm. In addition, some of these algorithms would either crash, if, for example, the prior range was too wide, or some of the model parameters (e.g., α, μ) would converge toward one end of the prior range. Therefore, to have meaningful comparisons of various methods, and to avoid software problems, we selected the same type of priors and range, as well as an even more restrictive subset of our data, as compared to those used in Paper I. As shown in Fig. 1, both datasets have been imposed $M \ge 1\,M_\odot$, which resulted in masses below the completeness limit in the $\ell = 30°$ region. This allowed us to test how this may affect the final posterior distributions. For the case of the $\ell = 59°$ field, Fig. 1 shows that ignoring sources with mass below the completeness limit may lead to deviations at the low-mass end.

---

[5] Here and in the following, we use the symbol of simple integration, $\int$, instead of that for multiple integration over the parameters space, $\int ... \int$.
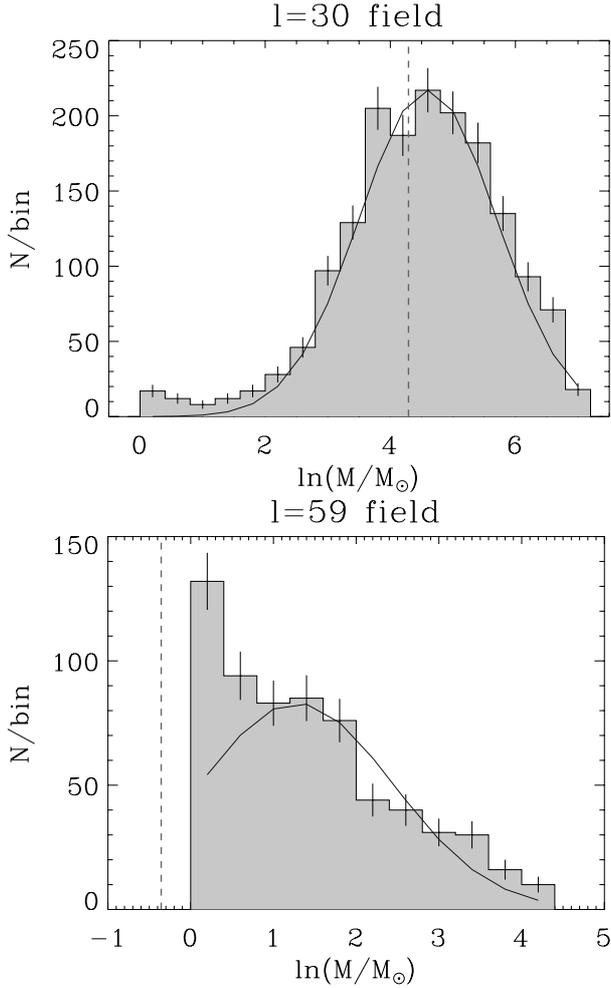
[6] http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

[7] http://www.r-project.org/

Fig. 1. Distribution of $\ln(M)$ for the selected subset of clumps in the $\ell = 30°$ (*top*) and $\ell = 59°$ (*bottom*) fields. Vertical bars show Poisson errors. The solid lines show one realization of the posterior distributions (corresponding to the mean values obtained with the Metropolis-Hastings method, see Sect. 4.3) discussed in Sect. 5. The vertical dashed lines represent the completeness limits as described in Paper I.

Table 1. Ranges for uniform priors used toward the $\ell = 30°$ and $\ell = 59°$ fields.

Region	Powerlaw		Lognormal		
	α	M_{inf} $[M_\odot]$	μ $[\ln(M_\odot)]$	σ $[\ln(M_\odot)]$	M_{inf} $[M_\odot]$
$\ell = 30°$	[0, 3]	[10, 30]	[0.8, 20]	[0.4, 10]	[10, 30]
$\ell = 59°$	[0, 3]	[0.3, 0.7]	[0.2, 5]	[0.2, 4]	[0.3, 0.7]

Since WinBUGS only supports a small set of proper[8] priors, we have selected *uniform* priors (see Table 1) as they are *non-informative* (which supposedly provide "minimal" influence on the inference) compared to, for example, Gaussian priors. The ranges for all parameters, except M_{inf}, have been selected around the values originally estimated in Paper I and have then been somewhat enlarged while maintaining convergence of all methods used here. The parameter M_{inf} has proved to be more critical and its range has been selected to achieve convergence with all methods (see Sect. 5).

---

[8]   The prior is called a proper prior if the sum or integral of the prior values is equal to 1.

4.2. Laplace approximation and harmonic mean estimator

In this section we describe two methods to implement the computation of the global likelihood that use open software resources (e.g., WinBUGS in this work) and, thus, do not require users to develop their own software.

4.2.1. Laplace approximation

One of the most popular approximations of the global likelihood is the so-called *Laplace approximation*. In this method, the product of prior and likelihood in Eq. (13) is approximated by a multivariate Gaussian. The integral then becomes the volume of a Gaussian distribution, and thus the accuracy of this method depends on how well the posterior is approximated by a Gaussian. The Laplace approximation for Eq. (13) results in (e.g., Gregory 2005, Ntzoufras 2009):

$$\mathcal{P}(D|\mathcal{M}) \approx (2\pi)^{Q/2} |\boldsymbol{H}(\hat{\theta})|^{-1/2} P(D|\hat{\theta}, \mathcal{M}) P(\hat{\theta}|\mathcal{M}), \tag{24}$$

where Q represents the number of parameters (see Sect. 3.2, item 1), $\hat{\theta}$ is the posterior mode of the parameters of model $\mathcal{M}$, and $\boldsymbol{H}$ is equal to the minus of the second derivative matrix (with respect to the parameters, or the *Hessian*) of the posterior PDF, which is $P(\theta|D, \mathcal{M})$ (see Sect. 3.2, item 5), evaluated at the posterior mode $\hat{\theta}$.

As described by Ntzoufras (2009), the Laplace-Metropolis estimator can be used to evaluate Eq. (24), where $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{H}$ can be estimated from the output of a Markov chain Monte Carlo (MCMC) algorithm (see Sect. 4.3.1). Thus, Eq. (24) becomes

$$\ln \mathcal{P}(D|\mathcal{M}) \approx \frac{1}{2} Q \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{R}_\theta| + \sum_{q=1}^{Q} \ln s_q$$
$$+ \sum_{i=1}^{N_{\text{tot}}} \ln p(M_i|\bar{\boldsymbol{\theta}}) + \ln P(\bar{\boldsymbol{\theta}}|\mathcal{M}), \tag{25}$$

where the posterior means (replacing the posterior modes) of the parameters of interest are denoted by $\bar{\boldsymbol{\theta}}$, $\boldsymbol{R}_\theta$ represents the posterior correlation between the parameters of interest, s_q are the posterior standard deviations of the θ_q parameters of model $\mathcal{M}$, and $p(M_i|\bar{\boldsymbol{\theta}})$ is the PDF associated to model $\mathcal{M}$ and evaluated at data point i (see Sect. 3.2, item 3).

Following Ntzoufras (2009), we estimate the posterior means, standard deviations, and correlation matrix from a MCMC run in WinBUGS, and then we calculate the global likelihood from Eq. (25). The results are listed in Tables 2 and 3, while the Bayes factors are listed in Table 4 and are discussed later in Sect. 5.

4.2.2. Harmonic mean estimator

The harmonic mean estimator (HME) is also based on a MCMC run and provides the following estimate for the global likelihood (Ntzoufras 2009):

$$\mathcal{P}(D|\mathcal{M}) \approx \left\{ \frac{1}{T} \sum_{t=1}^{T} \mathcal{L}(\boldsymbol{\theta}_t)^{-1} \right\}^{-1}, \tag{26}$$

where $\mathcal{L}(\boldsymbol{\theta}_t)$ represents the likelihood of the data corresponding to the t-th sample of the MCMC simulation, which has a total of T samples. Although very simple, this estimator is quite unstable and sensitive to small likelihood values; hence, it is not recommended. However, we present the HME values obtained by us with WinBUGS in Table 4 as a comparison for the Laplace-Metropolis method.

Table 2. Mean values and standard deviations estimated from the posterior distributions of the parameters, obtained using the methods described in the text (WinBUGS, MHPT, and MultiNest), for the powerlaw distribution.

Method	$\ell = 30°$		$\ell = 59°$	
	α	$\ln \mathcal{P}(D\|\mathcal{M})$	α	$\ln \mathcal{P}(D\|\mathcal{M})$
HME	0.73 ± 0.03	-9008	0.56 ± 0.02	-1924
Laplace	0.73 ± 0.03	-9454	0.56 ± 0.02	-2051
MHPT	0.7 ± 0.1	-4753	0.55 ± 0.01	-4985
MultiNest	0.73 ± 0.02	-9018	0.56 ± 0.02	-1932

Notes. Parameter M_{inf} is undetermined by all methods considered. Therefore, in the case of the Laplace method (requiring the posterior means), the mid-value of the priors range was used. Only results obtained using *uniform* priors are shown (see Table 1).

Table 3. Same as Table 2 for the lognormal distribution.

Method	$\ell = 30°$			$\ell = 59°$		
	$\mu\,[\ln M_\odot]$	$\sigma\,[\ln M_\odot]$	$\ln \mathcal{P}(D\|\mathcal{M})$	$\mu\,[\ln M_\odot]$	$\sigma\,[\ln M_\odot]$	$\ln \mathcal{P}(D\|\mathcal{M})$
HME	4.76 ± 0.02	0.96 ± 0.02	-6424	1.43 ± 0.04	1.07 ± 0.03	-543
Laplace	4.76 ± 0.02	0.96 ± 0.02	-9130	1.43 ± 0.04	1.07 ± 0.03	-1892
MHPT	4.6 ± 0.1	1.1 ± 0.1	-4610	1.28 ± 0.01	1.17 ± 0.01	-4622
MultiNest	4.0 ± 0.1	1.4 ± 0.1	-8904	1.1 ± 0.1	1.3 ± 0.1	-1823

Table 4. Estimated Bayes factor using the results listed in Tables 2 and 3 for the case where M_{inf} is a free parameter.

Method	$\ell = 30°$ **field** $\ln(BF)_{ln/pw}$	$\ell = 59°$ **field** $\ln(BF)_{ln/pw}$
HME	2584	1381
Laplace-Metropolis	324	159
MHPT	143	363
MultiNest	114	109

Notes. Positive values of $\ln(BF)_{ln/pw}$ favor the lognormal model over the powerlaw one.

4.3. Computation of the global likelihood with the Metropolis-Hastings algorithm

In this section, we describe how the global likelihood and the Bayes factor can also be estimated by implementing our own MCMC procedure.

4.3.1. Markov chain Monte Carlo

As we previously mentioned in Sect. 3.2, the Bayesian inference parameter estimation consists of calculating the posterior PDF, or density, $P(\theta|D, \mathcal{M})$, given by Eq. (12). However, since we must vary all θ_q parameters, we need a method for exploring the parameter space because gridding in each parameter direction would lead to an unmanageably large number of sampling points. The Q-dimensional parameter space can be explored with the aid of MCMC techniques, which are able to draw samples from the unknown posterior density (also called the *target distribution*) by constructing a pseudo-random walk in model parameter space, such that the number of samples drawn from a particular region is proportional to its posterior density. Such a pseudo-random walk is achieved by generating a *Markov chain*, which we create using the *Metropolis-Hastings* (MH) algorithm (Metropolis et al. 1953; Hastings 1970).

Briefly, the MH algorithm proceeds as follows. Given the posterior density $P(\theta|D, \mathcal{M})$ and any starting position θ_t in the parameter space, the step to the next position θ_{t+1} in the random walk is obtained from a *proposal distribution* (for example, a normal distribution) $g(\theta_{t+1}|\theta_t)$. Assuming g to be symmetric in θ_t and θ_{t+1}, the requirement of detailed balance leads to the following rule: accept the proposed move to θ_{t+1} if the *Metropolis ratio* $r \geq u$, where $r = P(\theta_{t+1}|D, \mathcal{M})/P(\theta_t|D, \mathcal{M})$, and u is a random variable drawn from a uniform distribution in the interval 0 to 1. If $r < u$, remain at θ_t. This sequence of proposing new steps and accepting or rejecting these steps is then iterated until the samples (after a burn-in phase) have converged to the target distribution. Since the factor at the denominator of Eq. (12) in the Metropolis ratio cancels out, the evaluation of r requires only the calculation of the parameter priors and of the likelihoods but *not* of the global likelihood, $\mathcal{P}(D|\mathcal{M})$.

A modified version of the MH algorithm to fully explore all regions in the parameter space with significant probability employs the so-called *parallel tempering* (MHPT; Gregory 2005, see also Handberg & Campante 2011). In the MHPT method, several versions, or *chains* (n_β in total), of the MH algorithm are launched in parallel. Each of these n_β chains is characterized by a different tempering parameter, β, and the new target distribution can be written by modifying Eq. (12):

$$P(\theta|D, \mathcal{M}, \beta) \propto P(\theta|\mathcal{M})\,\mathcal{L}(\theta)^\beta, \qquad 0 < \beta \leq 1. \qquad (27)$$

Thus, by modifying the chains according to Eq. (27) a discrete set of progressively flatter versions of the target distribution, also known as *tempered distributions*, are generated. For $\beta = 1$, we recover the target distribution. The MHPT method allows one to visit regions of parameter space containing significant probability, which are not accessible to the basic algorithm. The main steps in the MHPT algorithm are described in Appendix A.

4.3.2. Application of MHPT method to model comparison

Going back, now, to the issue of model comparison, an important property of the MHPT method is that the samples drawn from the tempered distributions can be used to compute the global likelihood, $\mathcal{P}(D|\mathcal{M})$, of a given model $\mathcal{M}$. It can be shown that the global log-likelihood of a model is given by (for a derivation see Gregory 2005)

$$\ln \mathcal{P}(D|\mathcal{M}) = \int_0^1 \langle \ln \mathcal{L}(\theta) \rangle_\beta \, d\beta, \qquad (28)$$

where

$$\langle \ln \mathcal{L}(\boldsymbol{\theta}) \rangle_\beta = \frac{1}{T} \sum_{t=1}^{T} \ln \mathcal{L}(\boldsymbol{\theta}_{t\beta}), \tag{29}$$

where T is the number of samples in each set after the burn-in period. The log-likelihoods in Eq. (29), $\ln \mathcal{L}(\boldsymbol{\theta}_{t\beta})$, can be evaluated from Eq. (10) and MHPT results, which consist of sets of $\{\boldsymbol{\theta}_t\}$ samples with one set (i.e., Markov chain, $\boldsymbol{\theta}_1 \rightarrow \boldsymbol{\theta}_2 \rightarrow ... \rightarrow \boldsymbol{\theta}_t \rightarrow ...$) for each value of the tempering parameter β.

As a by-product of the computation of the global likelihood, the MHPT method can also be used to determine the posteriors of the parameters. With both the posterior summaries and the global likelihoods, the results are also listed in Tables 2 and 3.

4.4. Computation of the global likelihood using the nested sampling method

The main problem of the methods outlined in Sect. 4.2 is the approximations involved, whereas the MCMC sampling methods, such as the MHPT technique described in Sect. 4.3, may have problems in estimating the parameters of some model, if the resulting posterior distribution is, for example, multimodal. In addition, calculation of the Bayesian evidence for each model is still computationally expensive using MCMC sampling methods.

The nested sampling method introduced by Skilling (2004) is supposed to greatly reduce the computational expense of calculating evidence and also produces posterior inferences as a by-product. This method replaces the multidimensional integral in Eq. (13) with a one-dimensional integral over an unit range:

$$\mathcal{P}(D|\mathcal{M}) = \int_0^1 \mathcal{L}(X) \mathrm{d}X \tag{30}$$

where $\mathrm{d}X = P(\boldsymbol{\theta}|\mathcal{M}) \mathrm{d}\boldsymbol{\theta}$ is the element of "prior volume". A sequence of values X_j can then be generated and the evidence is approximated numerically using standard quadrature methods (see Skilling 2004; and Feroz & Hobson 2008). Here, we use the "multimodal nested sampling algorithm" (MultiNest[9], Feroz & Hobson 2008; Feroz et al. 2009) to calculate both global likelihood and posterior distributions. The results are summarized in Tables 2 and 3.

5. Discussion

We now turn to the discussion of the effects of priors and different algorithms on the CMF parameter inference and model comparison using Bayesian statistics. As it was mentioned in Sect. 4, our comparison is limited to uniform priors only, because they are non-informative and also because of other software constraints.

Our purpose is not to perform a general analysis of the impact of priors type and range on Bayesian inference, since, as previously discussed in Sect. 3.1, this is a very complex topic that goes well beyond the scope of the present work. Instead, we are interested in the two distributions considered here for analyzing the sensitivity of our results to variations in the bounds of our uniform priors. In particular, we are interested in the role of the parameter M_{inf}, which is clearly critical for both powerlaw and lognormal distributions, as discussed below.

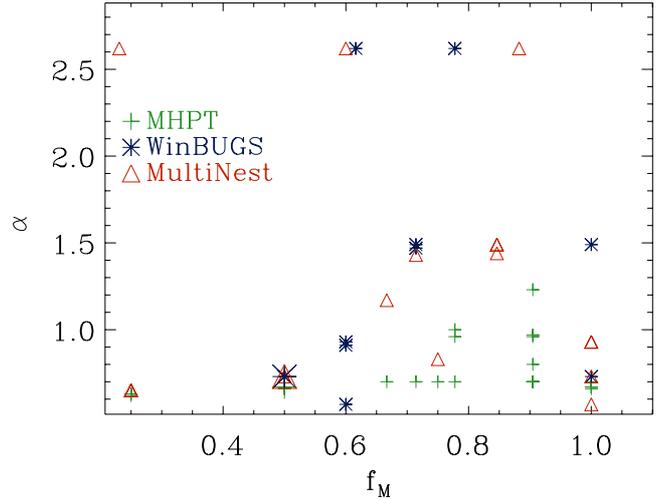[9] http://ccpforge.cse.rl.ac.uk/gf/project/multinest/

Fig. 2. Comparison of results obtained for the mean value of the parameter α (powerlaw case) in the $\ell = 30°$ field using the MHPT (green plus signs), MultiNest (red triangles), and WinBUGS (blue asterisks) methods, as a function of $f_M = (M_{\mathrm{inf2}} - M_{\mathrm{inf1}})/(M_{\mathrm{inf1}} + M_{\mathrm{inf2}})$, where M_{inf1} and M_{inf2} are the extremes of the prior range for M_{inf} (see text). The larger symbols correspond to the values listed in Table 2.

5.1. Powerlaw results

5.1.1. Non-regular likelihood: Consideration of M_{inf} as a free parameter

We start our discussion by analyzing the results of the posterior distributions for the powerlaw model, when the parameter M_{inf} is free to vary. Then, Table 2 shows that the three methods described in Sect. 4 deliver remarkably similar values of the α parameter, which are separate for the two SDP fields and with the prior ranges shown in Table 1. However, the powerlaw slope estimated for the two fields is different and is also different from the values quoted in Paper I ($\alpha \simeq 1.1$ and $\simeq 1.2$, for the $\ell = 30°$ and $\ell = 59°$ fields, respectively).

This discrepancy, however, is less significant compared to the sensitivity of the posteriors on the priors range and, in particular, on the range $[M_{\mathrm{inf1}}, M_{\mathrm{inf2}}]$ for the uniform prior on the parameter M_{inf}. We checked this sensitivity toward one of the two SDP fields, the $\ell = 30°$ region. Thus, in Fig. 2, we plot the values of α obtained with the three methods discussed above, as a function of the parameter $f_M = (M_{\mathrm{inf2}} - M_{\mathrm{inf1}})/(M_{\mathrm{inf1}} + M_{\mathrm{inf2}})$. The parameter f_M, thus, represents a measure of the amplitude of the prior range, and the scatter in Fig. 2 is due either to different M_{inf} ranges (which we have varied between 0 and 200 $M_\odot$), which may have the same value of f_M, or to the variation of other parameters specific to the method used. Despite their sensitivity to the parameter f_M, the values of α are much less sensitive to the range of its own uniform prior, that is, $[\alpha_1, \alpha_2]$ (which we have varied between 0 and 4; see also Sect. 5.1.2).

Looking at Fig. 2 it is not surprising that the values listed in Table 2 are somewhat different from those quoted in Paper I (Table 4). The values listed in Paper I could be easily reproduced with the proper choice of f_M. In addition, it should be noted that the values of α and M_{inf} listed in Paper I were determined using the PLFIT method (Clauset et al. 2009). Even with this method, the result for these parameters depends on whether an upper limit for M_{inf} is selected or not.

We also note that all of the methods used were unable to deliver a well-defined value for the M_{inf} parameter, unless *Gaussian* (i.e., *informative*) priors were used. In all cases
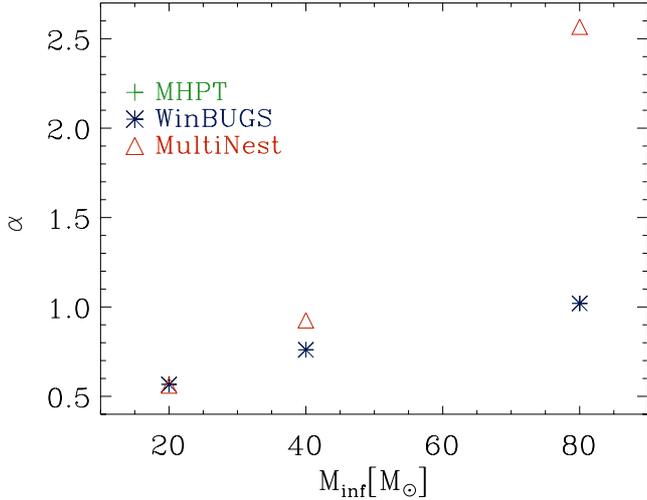
Fig. 3. Comparison of results obtained for the mean value of the parameter α (powerlaw case) in the $\ell = 30°$ field keeping $M_{\rm inf}$ fixed. Symbols and colors are as in Fig. 2. The points representing the MHPT and WinBUGS results overlap almost exactly, and the error bars on α are not shown because they are typically contained within the symbol size (see text).

considered, this parameter tends to converge toward the higher end of the prior range. However, this is not an effect caused by the specific data samples used. To test this issue, we generated a set of power-law distributed data, using the method described in Clauset et al. (2009) and applied to it the MHPT and MultiNest methods. In both cases, $M_{\rm inf}$ tended to converge toward the higher end of the prior range. Therefore, it is more likely that the convergence problems of $M_{\rm inf}$ arise because it is this unknown parameter that determines the range of the distribution. The likelihoods associated to such probability distributions are known as *non-regular* (see Smith 1985) and both likelihood and Bayesian estimators may be affected, requiring alternative techniques (see Atkinson et al. 1991; Nadal & Pericchi 1998) whose discussion is outside the scope of the present work.

In conclusion, the bayesian estimators considered here cannot constrain the value of the $M_{\rm inf}$ parameter, and it also appears that our data are not yet strong enough to override the prior of the powerlaw slope. Therefore, the value of α is sensitive to the choice of priors, and their range (mostly on $M_{\rm inf}$) and the present data do not allow us to draw statistically robust conclusions on possible differences between the two SDP fields, if we allow $M_{\rm inf}$ to be a free parameter. The uncertainty on the estimated value of α should be that derived from scatter plots like the one shown in Fig. 2 rather than the formal errors estimated by a specific method.

5.1.2. Regular likelihood: Keeping the value of $M_{\rm inf}$ fixed

To remove the potential problems associated with non-regular likelihoods, we carried out some tests to determine whether the sensitivity to the prior range would be the same even when the value of the $M_{\rm inf}$ parameter is kept fixed. We therefore modified the MHPT, MultiNest, and WinBUGS procedures to have only one free parameter, or α in the case of the powerlaw model. $M_{\rm inf}$ was kept fixed, but the range of the uniform prior on α was allowed to vary.

In Fig. 3, we show the results obtained for the $\ell = 30°$ field. We selected a series of values for $M_{\rm inf}$, and then we ran the three methods discussed above for each of these values,

repeating each procedure several times using a different range $[\alpha_1, \alpha_2]$ for the uniform prior on α. The figure shows three main features: (i) the two MCMC methods (MHPT and WinBUGS) yield almost identical results for α (In Fig. 3 their corresponding symbols almost exactly overlap.), which is independent of the selected value of $M_{\rm inf}$, whereas MultiNest progressively diverges from the other two methods. (ii) Then, the estimated values of α become larger for all methods, when $M_{\rm inf}$ is increased. (iii) Finally, for each specific value of $M_{\rm inf}$ all three methods are rather *insensitive* to variations of the prior range $[\alpha_1, \alpha_2]$. (In Fig. 3, the error bars representing the variations of α when using a different prior range are not shown because they are typically contained within the symbol size.)

Therefore, the sensitivity to the uniform priors range, which has been discussed in Sect. 5.1.1 and graphically shown in Fig. 2, disappears when $M_{\rm inf}$ is fixed, and the only free parameter left is α. This result would appear to confirm that the effects discussed in Sect. 5.1.1 are indeed a special consequence of the non-regularity of the likelihood.

5.2. Lognormal results

5.2.1. Non-regular likelihood: Consideration of $M_{\rm inf}$ as a free parameter

The results from the posterior distribution of the parameters are listed in Table 3, and they are also shown graphically in Figs. 4 and 5. Although we do not show a scatter plot similar to Fig. 2 for the lognormal case, we have equally noted a high sensitivity of all methods to the range of the uniform priors on $M_{\rm inf}$. This must be taken into account when comparing the results of Table 3 to those quoted in Paper I (Table 5). Further differences with respect to Paper I may also be caused by the different samples selected, particularly for the $\ell = 59°$ field, where the selection of sources with mass above the completeness limit may lead, for example, to overestimate parameter μ compared to the value reported in Paper I (see Fig. 1)

Similar to what happens for the powerlaw PDF (Sect. 5.1.1), even in the lognormal case, all methods considered here are unable to constrain the $M_{\rm inf}$ parameter, which makes the likelihood non-regular. Our comparison is thus limited to the μ and σ parameters. By comparing their values in Table 3 then, we note again that the three methods discussed in Sect. 4 deliver similar values for the μ and σ parameters. This is also clearly visible in Fig. 4, where it can be seen that the two MCMC-based methods deliver somewhat higher values of the μ parameter, compared to MultiNest. In both SDP fields, we also note the different shape of the posteriors distribution with the MHPT and MultiNest distribution having a similar shape, that is, with comparable widths in μ and σ (although with a different scale), while WinBUGS tends to have a more symmetrical distribution.

Therefore, if the parameter $M_{\rm inf}$ is allowed to vary even for the lognormal case, the data are not constraining enough to allow one to reliably predict the *absolute* values of some key parameters discussed here. However, our estimates are still good enough to allow a relative comparison between the two SDP regions. Even accounting for the different posterior distributions obtained with the various methods and the different choice of priors, the parameter μ is substantially higher in the $\ell = 30°$ field than in the $\ell = 59°$ region. On the other hand, the values of the parameter σ are much more alike between the two SDP fields. This result appears to confirm our earlier conclusion from Paper I. That is, the CMFs of the two SDP fields have very similar shapes but different mass scales, which, according to the
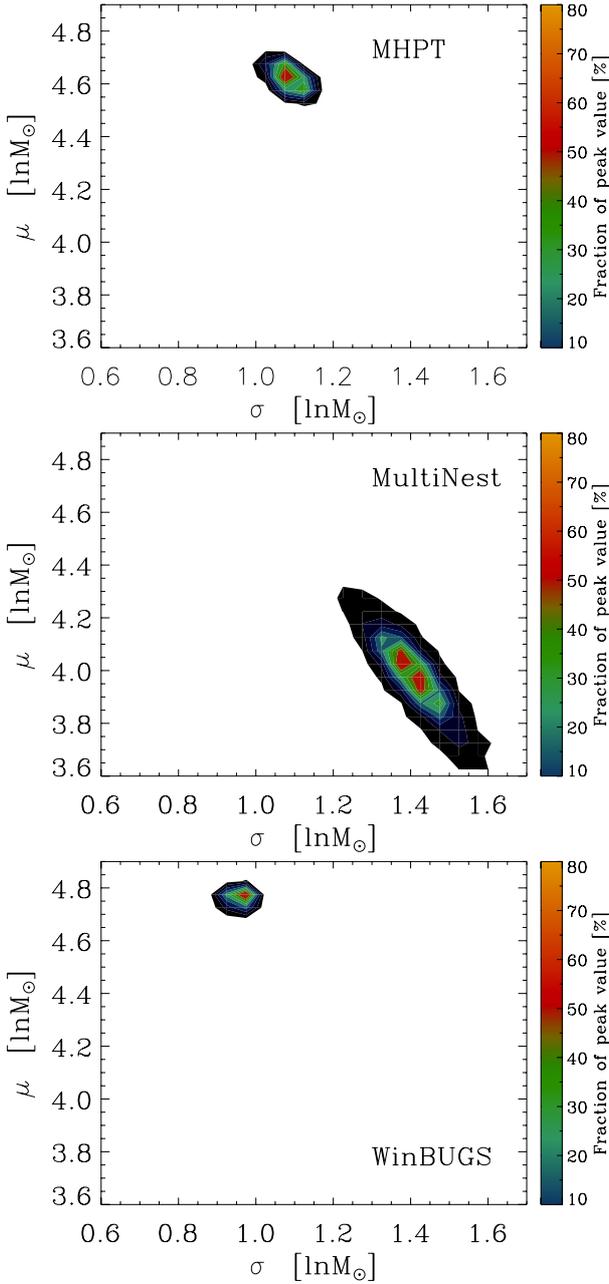
Fig. 4. Results on the posterior distributions of the μ and σ parameters for the lognormal PDF shown simultaneously in the $\mu - \sigma$ plane ($M_{\rm inf}$ is the third model parameter, see text) for the MHPT (*top panel*), MultiNest (*middle panel*), and WinBUGS (*bottom panel*) methods. The results refer to the $\ell = 30°$ field.

Fig. 5. Same as Fig. 4 for the $\ell = 59°$ field.

simulations discussed in Paper I, cannot be explained by distance effects alone.

5.2.2. Regular likelihood: Keeping the value of $M_{\rm inf}$ fixed

As already done in Sect. 5.1.2 for the powerlaw model, we have also run similar tests in the case of the lognormal distribution. Thus, we have selected some specific values of the $M_{\rm inf}$ parameter, and then we ran the three methods discussed above for each of these values, repeating each procedure several times using different ranges, $[\mu_1, \mu_2]$ and $[\sigma_1, \sigma_2]$, for the uniform priors on the μ and σ parameters. The results for the $\ell = 30°$ field are shown in Fig. 6. As with the powerlaw case, the two MCMC methods
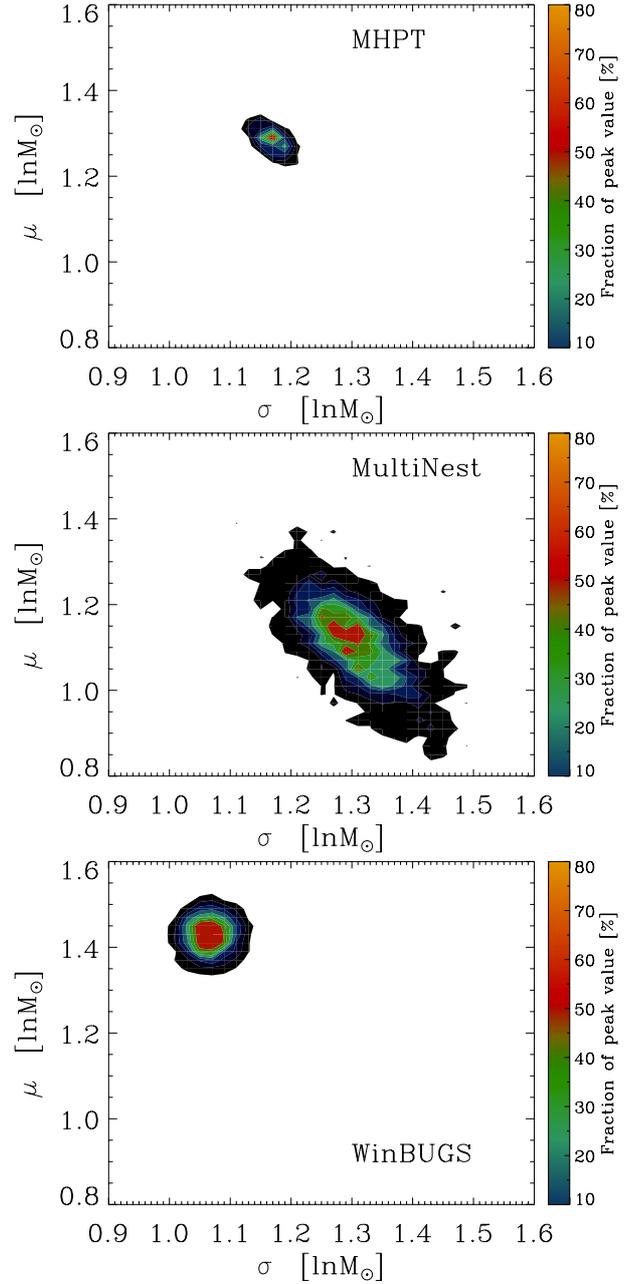
yield similar values, although to a lesser extent when compared to Fig. 3. The MultiNest algorithm converges toward one end of the $[\mu_1, \mu_2]$ prior range when $M_{\rm inf} \geq 40 \, M_\odot$, and this may also be the reason for the fluctuations seen in the parameter σ. We also note that the parameter μ tends to increase with larger values of $M_{\rm inf}$, while σ appears to be more stable at least in the case of the MCMC methods.

As with the powerlaw model, when the parameter $M_{\rm inf}$ is fixed, the posteriors of the remaining parameters, μ and σ, are *not* very sensitive to the ranges of the uniform priors for all three methods. For the distributions analyzed here, it would thus appear that the extreme sensitivity to the range of the uniform prior for the parameter $M_{\rm inf}$ is directly linked to the non-regularity of the likelihood, as described in Sect. 5.1.1, rather than being related to the more general sensitivity of Bayesian inference to the choice of priors type and range. As an additional comparison, we plot the posterior distributions in the $\mu - \sigma$ plane when $M_{\rm inf}$

Table 5. Estimated Bayes factor for the case where M_{inf} is fixed (M_{inf} = 20 $M_\odot$ and M_{inf} = 0.5 $M_\odot$ for the ℓ = 30° and ℓ = 59° fields, respectively; see Figs. 3 and 6).

Method	ℓ = 30° field $\ln(BF)_{ln/pw}$	ℓ = 59° field $\ln(BF)_{ln/pw}$
HME	2675	1418
Laplace-Metropolis	$-^a$	145
MHPT	189	586
MultiNest	356	171

Notes. [a] No convergence obtained.
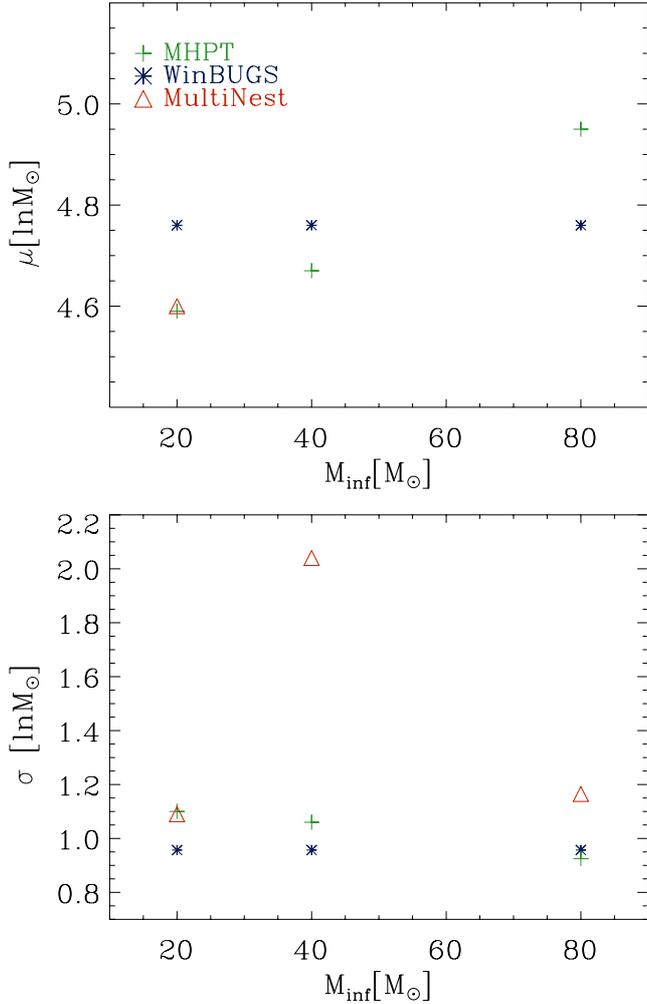
Fig. 6. Comparison of results obtained for the mean value of the parameters μ and σ (lognormal case) in the ℓ = 30° field, which keeps M_{inf} fixed. For $M_{inf} \geq 40\ M_\odot$, no convergence is obtained for parameter μ with MultiNest. Symbols and colors are as in Fig. 2.

is held fixed in Figs. 7 and 8. Compared to Figs. 4 and 5 one can note a better agreement among all algorithms. One can also note that the WinBUGS distributions (with a variable and fixed parameter M_{inf}) look very much the same.

5.3. Model comparison

We have previously seen how the posterior distributions can depend on the type of the parameter priors used and on their range, as well as depend on the specific algorithm and software used to
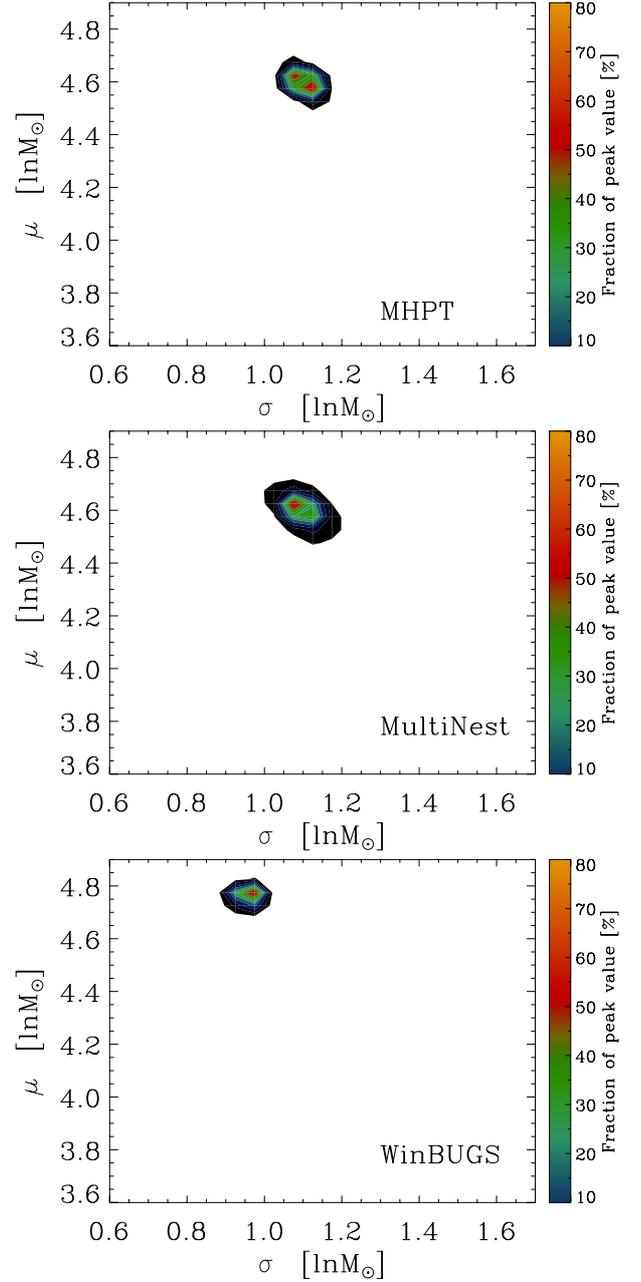
Fig. 7. Same as Fig. 4 for the case M_{inf} fixed (M_{inf} = 20 $M_\odot$).

estimate the posteriors. As we mentioned in Sect. 3.1, the situation is even more critical in model comparison, where the global likelihood, a priors-weighted average likelihood, and the Bayes factor also depend on the choice of priors. Even more so, given we usually wish to compare models with different number of parameters.

Therefore, the results listed in Table 4 should be regarded with some caution. In the table, we show the resulting Bayes factors (in logarithmic units) as estimated using the global likelihoods reported in Tables 2 and 3. According to our discussion in Sect. 3.4, the Bayes factors estimated by our analysis support the lognormal vs. the powerlaw model. In Table 4, we can note several features. (i) All methods result in the same conclusion for both SDP regions. (ii) However, the values of $\ln(BF)_{ln/pw}$ are quite large: In the so-called "Jeffreys scale", a value of $\ln(BF)_{21} > 5$ should be interpreted as "strong" support in favor of model 2 over model 1. The values that we find are
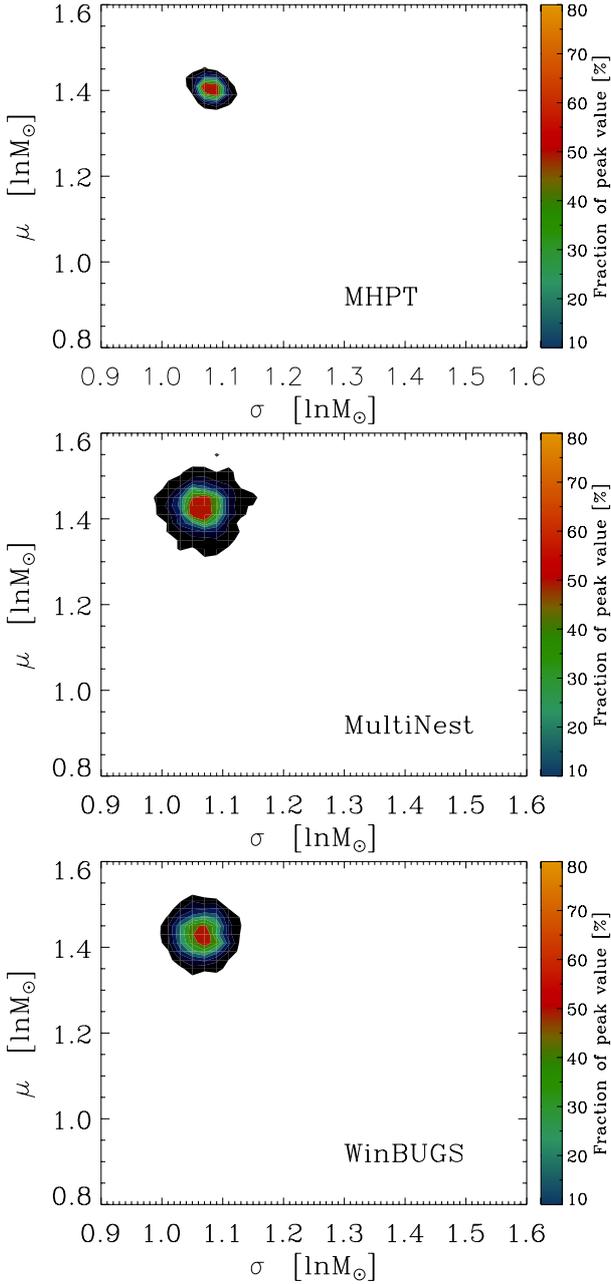
Fig. 8. Same as Fig. 5 for the case M_{inf} fixed ($M_{inf} = 0.5\,M_\odot$).

suspiciously large and suggest that they may be a consequence of the choice of priors. (iii) With the exception of MultiNest, all other methods show a substantial difference in the Bayes factors estimated for the two SDP fields.

We have also estimated the Bayes factors for the case of the parameter M_{inf} fixed to check for any significant difference. The results are shown in Table 5, and we can note that the values of $\ln(BF)_{ln/pw}$ are still positive and still quite large. The values $\ln(BF)_{ln/pw}$ are also strongly dependent on the selected value of the M_{inf} parameter and are typically lower for higher values of M_{inf}. Therefore, our preliminary conclusion is that the *lognormal* model appears to better describe the CMF measured in the two SDP regions. However, we caution that this conclusion may be affected by different choices of priors and their ranges.

6. Conclusions

Following our study (Paper I) of the two Hi-GAL, SDP fields centered at $\ell = 59°$ and $\ell = 30°$, we have applied a Bayesian analysis to the CMF of these two regions to determine how well powerlaw and lognormal models describe the data. First, we have determined the posterior probability distribution of the model parameters. Next, we have carried out a quantitative comparison of these models by using the data and an explicit set of assumptions. This analysis has highlighted the peculiarities of Bayesian inference compared to more commonly used MLE methods. In parameter estimation, the Bayesian inference allows one to estimate the probability distribution of each parameter, making it easier in principle to obtain realistic error bars on the results and, in addition, to include prior information on the parameters. However, Bayesian inference may be computationally intensive, and we have also shown that the results may be quite sensitive to the priors type and range, particularly if the parameters limit the range of the distribution (such as the M_{inf} parameter).

In terms of the powerlaw model, we found that the three Bayesian methods described here deliver remarkably similar values of the powerlaw slope for both SDP fields. Likewise, for the lognormal model of the CMF, we have found that the three computational methods deliver similar values for the μ (center of the lognormal distribution) and σ (width of the lognormal distribution) parameters, which are separate for the two SDP fields. In addition, the parameter μ is substantially higher in the $\ell = 30°$ field than in the $\ell = 59°$ region, while the values of the parameter σ are much more alike between the two SDP fields. This result confirms our earlier conclusion from Paper I, that is, the CMFs of the two SDP fields have very similar shapes but different mass scales. We have also shown that the difference with respect to the values of the parameters determined in Paper I may be due to the sensitivity of the posterior distributions to the specific choice of the parameter priors, and, in particular, of the M_{inf} parameter.

As far as model comparison is concerned, we have discussed and compared several methods to compute the global likelihood, which, in general, cannot be calculated analytically and is fundamental to estimate the Bayes factor. All methods tested here show that the lognormal model appears to better describe the CMF measured in the two SDP regions. However, this preliminary conclusion is dependent on the choice of parameter priors and needs to be confirmed using more constraining data.

Appendix A: Procedure to implement the Metropolis-Hastings algorithm with parallel tempering (MHPT)

We briefly list here the main steps of the Metropolis-Hastings algorithm with the inclusion of parallel tempering (see Sects. 2.1 and 4.3 for defintions).

1: Initialize the parameters vector for all tempered distributions
$\boldsymbol{\theta}_{0,i} = \boldsymbol{\theta}_0, \quad 1 \le i \le n_\beta$
2: Start MCMC loop
for $t = 1, ..., (T - 1)$
3:    Start *parallel tempering* loop
for $i = 0, 1, ..., (n_\beta - 1)$
4:       Propose a new sample drawn from, e.g., a Normal distribution with a mean equal to current parameters values and standard deviation fixed
$\boldsymbol{\theta}_{prop} \sim N(\boldsymbol{\theta}_{t,i}; \boldsymbol{\sigma})$

5:      Compute the *Metropolis ratio* using Eq. (27)
        $\ln r = \ln P(\theta_{\text{prop}}|D, \mathcal{M}, \beta_i) - \ln P(\theta_{t,i}|D, \mathcal{M}, \beta_i)$
6:      Sample a uniform random variable
        $u_1 \sim Uniform(0, 1)$
7:      if $\ln u_1 \leq \ln r$ then
            $\theta_{t+1,i} = \theta_{prop}$
        else
            $\theta_{t+1,i} = \theta_{t,i}$
        end if
8:      end for   End parallel tempering loop
9:      Sample another uniform random variable
        $u_2 \sim Uniform(0, 1)$
10:     Do swap between chains?
        (n_{swap} = N. of swaps between chains)
11:     if $u_2 \leq 1/n_{\text{swap}}$ then
12:         Select random chain:
            $j \sim UniformInt(1, n_\beta - 1)$
13:         Compute r_{swap}
            $\ln r_{\text{swap}} = \ln P(\theta_{t,j+1}|D, \mathcal{M}, \beta_j) + \ln P(\theta_{t,j}|D, \mathcal{M}, \beta_{j+1})$
            $- \ln P(\theta_{t,j}|D, \mathcal{M}, \beta_j) - \ln P(\theta_{t,j+1}|D, \mathcal{M}, \beta_{j+1})$
14:         $u_3 \sim Uniform(0, 1)$
15:         if $\ln u_3 \leq \ln r_{\text{swap}}$ then
                Swap parameters states of chains j and $j + 1$
                $\theta_{t,j} \leftrightarrow \theta_{t,j+1}$
15:         end if
16:     end if
17: end for   End MCMC loop

References

Berger, J. O., & Pericchi, L. R. 2001, in Model selection, ed. P. Lahiri, IMS-Lecture Notes, Institute of Mathematical Statistics, Beachwood (OH), 38, 135

Chabrier, G. 2003, PASP, 115, 763

Clauset, A., Shalizi, C. R., & Newman, M. E. J. 2009, SIAM Rev., 51, 661

Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Gelfand, A. E., & Dey, D. K. 1994, J. Roy. Stat. Soc. Ser. B Methodol., 56, 501

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support (Cambridge University Press)

Handberg, R., & Campante, T. L. 2011, A&A, 527, A56

Hastings, W. K. 1970, Biometrika, 57, 97

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, J. Chem. Phys., 21, 1087

Molinari, S., Swinyard, B., Bally, J., et al. 2010, PASP, 122, 314

Ntzoufras, I. 2009, Bayesian Modeling Using WinBUGS (Wiley)

Olmi, L., Anglés-Alcázar, D., Elia, D., et al. 2013, A&A, 551, A111

Pericchi, L. R. 2005, in Bayesian thinking: modeling and computation, eds. D.K. Dey, & C.R. Rao, Handbook of Statistics (Elsevier), 25, 115

Salpeter, E. E. 1955, ApJ, 121, 161

Skilling, J. 2004, in AIP Conf. Ser. 735, eds. R. Fischer, R. Preuss, & U. V. Toussaint, 395

Trotta, R. 2008, Contemp. Phys., 49, 71

Trotta, R., Feroz, F., Hobson, M., Roszkowski, L., & Ruiz de Austri, R. 2008, J. High Energy Phys., 12, 24