

Eddington’s demon: inferring galaxy mass functions and other distributions from uncertain data

D. Obreschkow,^{1,2★} S. G. Murray,^{2,3} A. S. G. Robotham¹ and T. Westmeier¹

¹International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

²ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO), Australia

³International Centre for Radio Astronomy Research (ICRAR), Curtin University, 1 Turner Ave, Bentley, WA 6102, Australia

Accepted 2017 December 2. Received 2017 December 2; in original form 2017 September 25

ABSTRACT

We present a general modified maximum likelihood (MML) method for inferring generative distribution functions from uncertain and biased data. The MML estimator is identical to, but easier and many orders of magnitude faster to compute than the solution of the exact Bayesian hierarchical modelling of all measurement errors. As a key application, this method can accurately recover the mass function (MF) of galaxies, while simultaneously dealing with observational uncertainties (Eddington bias), complex selection functions and unknown cosmic large-scale structure. The MML method is free of binning and natively accounts for small number statistics and non-detections. Its fast implementation in the *R*-package `dftools` is equally applicable to other objects, such as haloes, groups, and clusters, as well as observables other than mass. The formalism readily extends to multidimensional distribution functions, e.g. a Choloniewski function for the galaxy mass–angular momentum distribution, also handled by `dftools`. The code provides uncertainties and covariances for the fitted model parameters and approximate Bayesian evidences. We use numerous mock surveys to illustrate and test the MML method, as well as to emphasize the necessity of accounting for observational uncertainties in MFs of modern galaxy surveys.

Key words: galaxies: luminosity function, mass function – methods: statistical.

1 INTRODUCTION

Few challenges have incited more publications by astrophysicists of several generations than the task of fitting a luminosity function (LF) or mass function (MF) to samples of stars, galaxies, groups, and clusters (e.g. Schmidt 1968; Lynden-Bell 1971; Kirshner, Oemler & Schechter 1979; Sandage, Tammann & Yahil 1979; Turner 1979; Davis & Huchra 1982; Efstathiou, Ellis & Peterson 1988; Zwaan et al. 2003; Teerikorpi 2004; Cole 2011; Loveday et al. 2015; Weigel, Schawinski & Bruderer 2016). LFs and MFs quantify the space density of these objects in the universe as a function of their luminosity and mass, respectively. They are fundamental observables that summarize the statistical outcome and scale-dependence of complex formation processes (e.g. Croton et al. 2006; Murray, Power & Robotham 2013). LFs and MFs thus constitute a crucial bridge between astrophysical observations and theory, having shaped their synergetic progress for more than half a century.

For simplicity, this paper will talk about the ‘galaxy MF’, however the concepts, formalisms and implementations are readily transferable to other objects, e.g. stars, star clusters, galaxy clusters, and dark haloes. Likewise, the galaxy mass M is not further specified

to emphasize its applicability to any choice, e.g. stellar mass, gas mass, or dynamical mass. In fact, M can be substituted for any other observable, including luminosity, absolute magnitude, and even multidimensional observables. This article hence treats galaxy MFs as an example of the general problem of inferring a distribution function (DF) generating empirical data subject to measurement uncertainties and sample biases.

Faint galaxies are far more common than luminous, massive ones. In a fixed cosmic volume, the number of galaxies per unit mass approximately declines as a power law up to a cut-off scale, beyond which the number density declines almost exponentially. In appreciation of the power-law behaviour, it is convenient to define the MF $\phi(x)$ as the the density of galaxies per unit volume and unit of logarithmic mass $x = \log_{10}(M/M_{\odot})$. Formally, in a cosmic volume V , the expected number of galaxies in an interval $[x, x + dx]$ is

$$dN = \phi(x) V dx. \quad (1)$$

The choice of \log_{10} -units is a subjective preference and easily converts to the natural-log MF, $\phi'(x) = \phi(x)/\ln 10$, or linear MF, $\Phi(M) = \phi(x)/(M \ln 10)$.

Many analytical parametric functions have been proposed to fit observed MFs. They can all be written as $\phi(x|\theta)$, where θ is a vector

* E-mail: dana.il.obreschkow@gmail.com

of P scalar model parameters. A common example is the Schechter function (Schechter 1976),

$$\phi(x|\theta)_{\text{Schechter}} = \ln(10)\phi_*\mu^{\alpha+1}e^{-\mu}, \quad (2)$$

where $\mu = M/M_* = 10^x M_\odot/M_*$. This model depends on three parameters: the amplitude ϕ_* , the break mass M_* and the power-law slope α . The Schechter function is the best-known MF model that captures the truncated power-law behaviour. We will use the Schechter function for most illustrations, but the formalism and implementations remain applicable to any DF model, including quasi non-parametric ('stepwise') models, discretized into a custom number of bins or vertices.

The measurement of MFs requires a redshift survey, providing a sample of galaxies with approximate distances and hence a means of converting apparent to absolute magnitudes and intrinsic mass. Fitting an MF model $\phi(x|\theta)$ to these data is not trivial. The most basic and intuitive approach (Schmidt 1968) is to bin the data in mass and estimate an observed space density ϕ_j in each bin j , by dividing the number of detections by the bin width and the maximal volume V_{max} , in which galaxies of that bin could have been detected ($1/V_{\text{max}}$ method). A model function $\phi(x|\theta)$ is then fitted to the bin values ϕ_j . This method suffers from a list of limitations:

(i) The fit depends on the choice of binning (bin centres and spacing).

(ii) It is not clear how $\phi(x|\theta)$ is best fitted to the binned data. Given the varying number of galaxies per bin and Poisson statistics, least-square minimization is inaccurate.

(iii) The inclusion of non-detections, i.e. mass bins which happen to contain no galaxy, is often cumbersome because of the impossibility to assign Poisson errors to such bins.

(iv) Dealing with observational uncertainties in the galaxy masses is difficult.

(v) Complex detection limits with source-dependent completeness and reliability (defined in Section 2.2) make the choice of V_{max} ambiguous.

(vi) Cosmic large-scale structure (LSS) can introduce systematic errors (see Section 2.3).

Most of these challenges cannot be overcome by brute force alone, e.g. by simply observing more galaxies, and hence remain an issue for modern spectroscopic redshift surveys detecting hundreds of thousands (York et al. 2000; Colless et al. 2001; Drinkwater et al. 2010; Grazian et al. 2015; Liske et al. 2015; Davidson et al. 2017) to millions (Amiaux et al. 2012; Dawson et al. 2013) of galaxies. Some challenges, such as small number statistics and LSS, are particularly pronounced in small samples ($\lesssim 10^3$ galaxies) and samples with a strong mass bias. Such samples often arise, by construction, when subsamples are drawn from larger sets to address cutting-edge topics (e.g. highest redshifts, satellite population, rare environments).

All the caveats above have been addressed from different angles in the literature (e.g. references in the first paragraph). The persisting problem is to overcome all of them at once (see Pihajoki 2017 for the related case of non-linear model fitting). In particular, approaches dealing with measurement uncertainties (e.g. Teerikorpi 2004), especially prominent in cluster studies (Mortonson, Hu & Huterer 2011; Evrard et al. 2014), are hard to reconcile with approaches addressing all the other issues. Perhaps for this reason, a surprising number of MF and LF studies in modern galaxy surveys neglect measurement uncertainties or do not deal with them in a statistically accurate way. This bias is none the less significant in spite (or rather because) of the large number of

detections in modern surveys. Another problem is that MF fitting methods are quite tricky or at least time-consuming to implement due to their inevitable mathematical complexity. Thus, most authors of MF papers of redshift surveys have spent considerable time developing/publishing their own techniques prior to processing the actual data.

The objective of this paper is to derive, demonstrate, and implement a method that simultaneously overcomes all the caveats above, building on the extensive literature. This method, derived in Section 2, is based on a variation of the maximum likelihood (ML) method – with the statistical axioms that this approach entails (Appendix A) – and makes only a few assumptions on the nature of the data and model to be fitted. It can be used to infer the most likely model parameters θ of any DF model $\phi(x|\theta)$, while accounting for generic measurement uncertainties and complex selection functions. Section 3 presents a fast algorithm for this method and describes its numerical implementation in the R statistical language. Sections 4 and 5 test and illustrate the method using controlled mock data. For clarity and simplicity, we avoid the use of real observations in this paper. Section 6 concludes with a critical summary and outlook.

2 MATHEMATICAL METHOD

2.1 Generative model search

Let us consider a galaxy survey detecting N galaxies $i = 1, \dots, N$ with masses $M_i = 10^{x_i} M_\odot$. We temporarily assume that these mass measurements are exact and bin them into a finite number of bins j , equally spaced in log-mass x , with bin widths Δx and bin centres x_j . In this case, the measurement can be summarized via the discrete *source counts* $n_j \in \mathbb{N}_0$, where n_j is the number of galaxies with masses in the interval $x \in \{x_j \pm \Delta x/2\}$.

We then assume that the space density of the whole galaxy population is described by an MF $\phi(x|\theta)$, e.g. a Schechter function (equation 2). The objective is to find the most likely model parameters θ generating the observed source counts n_j . To do so, we note that the predicted number of galaxies detected in the mass bin j is

$$\lambda_j(\theta) = \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} \phi(x|\theta)V(x)dx. \quad (3)$$

For the moment, the effective volume $V(x)$ can be thought of as the volume probed by galaxies of log-mass x in terms of a sharp detection limit: every galaxy of log-mass x is detected if and only if it lies inside this volume and there are no false detections. Explicit expressions for $V(x)$ in the presence of general selection functions will be derived in Section 2.2.

Given the expected source counts $\lambda_j(\theta)$, the likelihood $L_j(\theta)$ of detecting n_j galaxies in bin j is assumed to be given by the Poisson distribution function

$$L_j(\theta) = \frac{e^{-\lambda_j(\theta)}\lambda_j(\theta)^{n_j}}{\Gamma(n_j + 1)}. \quad (4)$$

The total likelihood function $L = \prod_j L_j$ is conveniently written as logarithm (akin to the photon statistics of Cash 1979),

$$\ln L(\theta) = \sum_j (n_j \ln \lambda_j(\theta) - \lambda_j(\theta) - \ln \Gamma(n_j + 1)) \quad (5)$$

with the convention $0 \ln 0 = 0$ to account for bins expected to be empty. This likelihood (without the parameter-independent last term) constitutes the core of numerous MF papers since its

introduction for parametric (Sandage et al. 1979) and non-parametric (Efstathiou et al. 1988) MFs.

We now generalize equation (5) to account for statistical measurement errors. To this end, each datum x_i is replaced by a probability distribution function (PDF) $\rho_i(x)$ (with normalization $\int \rho_i(x) dx = 1$). This PDF represents the probability that the galaxy i has a true log-mass x , based solely on the measurement, that is *assuming a flat prior* without using any knowledge on the underlying MF $\phi(x|\theta)$ and selection $V(x)$. A typical example is the case where each object i has a measured value x_i with a normally distributed uncertainty. In this case, $\rho_i(x)$ is a Gaussian centred at x_i . The evaluation of $\rho_i(x)$ can be a subtle task, which depends on whether the underlying error model is conditional on the true or the observed mass. A non-trivial example will be provided in Section 5.4.

Interestingly, in an uncertain measurement, the mode of $\rho_i(x)$ is not a good proxy for the true log-mass if the MF is steep (i.e. varies considerably across the width of $\rho_i(x)$). This feature, known as *Eddington bias*, can be accounted for using Bayes theorem to write the bias-corrected PDFs as

$$\tilde{\rho}_i(x|\hat{\theta}) = \frac{\rho_i(x)\phi(x|\hat{\theta})V(x)}{\int \rho_i(x)\phi(x|\hat{\theta})V(x)dx}, \quad (6)$$

where $\hat{\theta}$ is the vector of the most likely model parameters. The problem with equation (6), of course, is that the most likely model parameters $\hat{\theta}$ are a priori unknown, which will lead to an interesting optimization problem.

Note that equation (6) is an approximation, because the exact posteriors $\tilde{\rho}_i(x)$ would require integrating over the full posterior PDF of θ instead of using the mode $\hat{\theta}$. However, we will show later that this approximation does not in fact change the solution. Another subtle point is that equation (6) assumes that the observational uncertainty is introduced to the data after drawing it from the population source counts $\phi(x|\hat{\theta})V(x)$, that is *after* applying the selection encoded in the effective volume $V(x)$. In astrophysical observations, it is not uncommon that some scatter is introduced to the data already before applying the selection function, or that there are multiple layers of scattering events and selection processes. For the moment, we assume that these cases can be recast into a single selection function $V(x)$, followed by an uncertain measurement. Further discussion and justification, along with an example of fitting scatter followed by selection are provided in Appendix D.

We then split each bias-corrected mass measurement into the mass bins j via

$$n_j(\hat{\theta}) = \sum_i \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} \tilde{\rho}_i(x|\hat{\theta}) dx. \quad (7)$$

In this way, the source counts $n_j \in \mathbb{R}$ become non-integers, but the normalization $\sum_j n_j = N$ persists.

Finally, we let Δx become infinitesimal and rewrite the *predicted* (equation 3) and *observed posterior* (equation 7) source count densities $\lambda_j/\Delta x$ and $n_j/\Delta x$ as

$$\lambda(x|\theta) = \phi(x|\theta)V(x), \quad (8)$$

$$n(x|\hat{\theta}) = \sum_i \tilde{\rho}_i(x|\hat{\theta}). \quad (9)$$

The likelihood function (equation 5) then becomes

$$\ln \mathcal{L}(\theta, \hat{\theta}) = \int (n(x|\hat{\theta}) \ln \lambda(x|\theta) - \lambda(x|\theta)) dx, \quad (10)$$

where we have dropped a constant¹ that does not depend on θ . We refer to $\mathcal{L}(\theta, \hat{\theta})$ as the *modified likelihood* function to emphasize its subtle difference to the true likelihood function $L(\theta, \hat{x}_i)$, i.e. the likelihood of the full two-stage Bayesian hierarchical model (Allenby, Rossi & McCulloch 2005) that treats the true values \hat{x}_i leading to the uncertain measurements x_i as N additional model parameters. In a self-consistent ML solution, the parameter-vector θ maximizing equation (10) (while keeping $\hat{\theta}$ fixed) must be equal to $\hat{\theta}$ used for debiasing the observations. In other words, we are looking for the parameter-vector $\hat{\theta}$ satisfying

$$\hat{\theta} = \mathcal{F}(\hat{\theta}), \quad (11)$$

where

$$\mathcal{F}(\hat{\theta}) = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta, \hat{\theta}). \quad (12)$$

We call this approach the modified maximum likelihood (MML) method and refer to the solution $\hat{\theta}$ of equation (11) as the MML estimator (MMLE).

Importantly, it can be shown analytically (see Appendix A) that if a unique solution exists for the standard ML estimator (MLE) of the full Bayesian hierarchical model, then this same solution exists uniquely also for equation (11). This identity of the MMLE and MLE is the central theorem of this work that ultimately justifies the MML method and puts it on a robust mathematical basis. By virtue of this theorem the MMLE inherits all of the interesting features of the MLE. In particular, it is an asymptotically unbiased, minimum-variance and normally distributed estimator (Kendall & Stuart 1979). These are precisely the properties that one would naturally request from an optimal estimator. Incidentally, if the ML solution is *not* unique, i.e. if the true likelihood has multiple maxima, the MMLE is also not unique. In practice, different solutions $\hat{\theta}$ can then be found by sampling the initial parameters $\hat{\theta}_0$ (see Section 3.1), for instance as part of an MCMC algorithm, and the solution that maximizes $\mathcal{L}(\hat{\theta}, \hat{\theta})$ is the most likely model. However, in the wide range of mock examples considered in this work, such an approach has never been necessary.

Solving equation (11) is much easier than maximizing the full likelihood in the presence of general observational uncertainties, but remains none the less a challenging optimization problem. A fast and stable algorithm is presented in Section 3.1 and implemented in Section 3.2.

2.2 Selection function

This section elaborates on how to evaluate the effective volume $V(x)$, so far assumed to be given.

Most galaxy surveys do not have sharp sensitivity limits. There is no well-defined maximum volume inside which all galaxies of a fixed mass are detected, while outside none are detected. Instead, the fraction of detected sources (true positives) decreases gradually when reaching the detection limit, while the fraction of wrong detections (false positives) increases. These fractions are typically quantified via the *completeness* $C \in [0, 1]$, defined as the probability of a real source to be detected, and the *reliability* $R \in [0, 1]$, defined as the probability of a detection corresponding to a real

¹ Equation (5) diverges as $\Delta x \rightarrow 0$. The trick to avoid this divergence is to subtract the term $\sum_j (n_j \ln \Delta x - \ln \Gamma(n_j + 1))$ before letting $\Delta x \rightarrow 0$, which is possible because this term does not depend on the parameter-vector θ to be fitted.

source. Both fractions generally depend on the mass and the position $\mathbf{r} \in \mathbb{R}^3$ in the survey volume, here defined as the comoving position relative to the observer. Sometimes C and R also depend on known or unknown extra properties (here labelled ...), such as the galaxy inclination. For what follows, it is convenient to introduce the *selection function*

$$f(x, \mathbf{r}) = \left\langle \frac{C(x, \mathbf{r}, \dots)}{R(x, \mathbf{r}, \dots)} \right\rangle, \quad (13)$$

where the expectation $\langle \rangle$ averages over the extra variables. In this way, $f(x, \mathbf{r})$ always equals the expected ratio between the number of detections (including false positives) and the number of true sources (whether detected or not). Hence, the expected number density of detections per unit log-mass x and comoving volume is $\phi(x|\theta)f(x, \mathbf{r})$ and the expected number density over the whole survey volume is

$$\lambda(x|\theta) = \phi(x|\theta) \int f(x, \mathbf{r}) d^3r. \quad (14)$$

Matching this equation to equation (8) implies that the MML formalism (equations 8–11) applies to arbitrary selection functions, upon defining the effective volume as

$$\begin{aligned} V(x) &= \int f(x, \mathbf{r}) d^3r \\ &= \int_0^{2\pi} d\alpha \int_{-\pi/2}^{\pi/2} d\delta \int_0^\infty dr r^2 \cos(\delta) f(x, \alpha, \delta, r), \end{aligned} \quad (15)$$

where the second line is the explicit form in spherical coordinates: right ascension α , declination δ and distance $r = |\mathbf{r}|$.

Two common cases, worth expanding, are those where the selection function only depends on x and r (Section 2.2.1) and where the effective volume is only given on a galaxy-by-galaxy basis (Section 2.2.2).

2.2.1 Isotropic selection function

Often the selection function is independent of the direction (α and δ) and only varies with the comoving distance r . Such *isotropic* selection functions $f(x, r)$ reduce equation (15) to

$$V(x) = \int V'(r) f(x, r) dr, \quad (16)$$

where $V'(r)$ is the derivative of the total observed comoving volume $V(r)$. For instance, a survey of redshift-independent solid angle Ω encompasses a volume $V(r) = \Omega r^3/3$, i.e. $V'(r) = \Omega r^2$ and hence

$$V(x) = \Omega \int r^2 f(x, r) dr. \quad (17)$$

Isotropic selection functions are sometimes expressed as $f(x, z)$, where z is the cosmological redshift corresponding to the distance r . To write equations (16) and (17) in terms of z , it suffices to substitute r for its expression as a function of z . For instance, in the local universe, $r = cz/H_0$ with Hubble constant H_0 and speed of light c , and hence $V(z)' = \Omega c^3 H_0^{-3} z^2$.

2.2.2 Volume given for each galaxy

It is quite common for galaxy surveys to store a single effective volume $V_i = V(x_i)$ for each galaxy i , without specifying the continuous functions $V(x)$ or $f(x, r)$ needed in the MML formalism. In practice, the V_i are often computed from the maximal volume $V_{\max}(x_i)$ in which a galaxy of log-mass x_i could have been detected, corrected for the estimated completeness and reliability of the source,

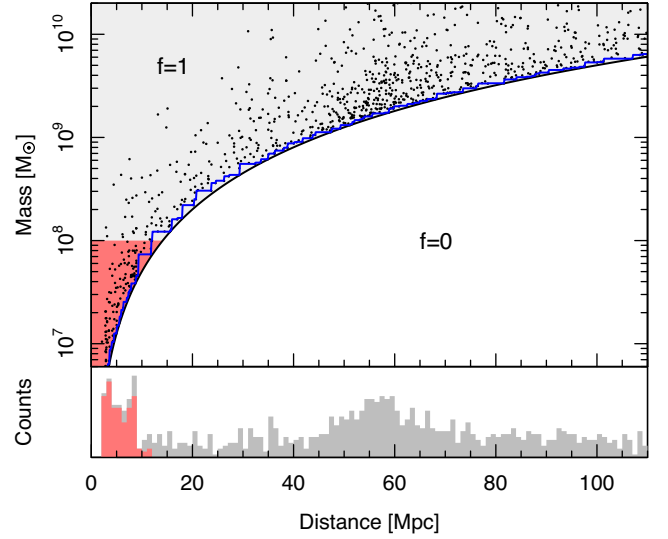


Figure 1. Example of a sensitivity-limited mock galaxy survey, subject to LSS with a notable overdensity at distances smaller than 10 Mpc (red). This overdensity causes an over-average detection rate of galaxies with masses below $10^8 M_\odot$. See Section 2.3 for details.

$V_i = C_i R_i^{-1} V_{\max}(x_i)$. The important point is that the galaxy i contributes a density V_i^{-1} to the galaxy MF at its particular mass.

The challenge consists in reconstructing the continuous function $V(x)$ from the finite set $\{V_i\}$. This challenge also has to face the fact that two (or more) galaxies of identical mass x can have different values V_i . For example, some galaxies might be seen edge-on with strong dust extinction, while others are seen face-on with little extinction. Thus, the first class is harder to detect and hence admits a smaller effective volume than the second one. So what value should be chosen for $V(x)$? The answer comes from the requirement that space densities must add up (following from conservation of mass), which implies that $V(x)$ is the harmonic mean of the individual effective volumes of all detections. Formally,

$$V(x_i) = \left\langle V_i^{-1} \right\rangle^{-1}, \quad (18)$$

where the expectation goes over all sources of the same log-mass x_i and hence marginalizes over all other variables. A simple example demonstrating the applicability of this harmonic mean is given in Appendix C.

The reasoning behind equation (18) implies that $V^{-1}(x)$ can be interpolated linearly from the given values $V^{-1}(x_i)$. For values x outside the range of $\{x_i\}$, the function $V(x)$ can normally be extrapolated based on the survey specifications. For example, in a volume-limited survey where the most massive galaxy $i = N$ could have been detected anywhere in the survey volume, $V(x) = V(x_N)$ for all $x > x_N$.

Numerical tests using the implementation of Section 3 show that this way of interpolating $V(x)$ from the set $\{V_i = V(x_i)\}$ generally leads to MML solutions that are statistically consistent with those obtained using the exact $V(x)$.

2.3 Cosmic large-scale structure

Cosmic large-scale structure (LSS) can introduce systematic errors in the estimation of the MF. How these errors arise is illustrated in Fig. 1. In this example, the average density within the survey volume is the mean density of the universe, hence the counts of

massive galaxies that can be detected to the edge of this volume are insensitive to galaxy clustering. However, due to a nearby overdensity (red), the low-mass galaxies ($M < 10^8 M_\odot$), which can only be detected to a smaller distance (~ 10 Mpc) are overrepresented. If not corrected, this selection bias will result in overestimating the steepness of the MF in the low-mass end.

LSS is suitably quantified by the relative density $g(r)$, defined as the mean density in the survey volume at comoving distance $r = |\mathbf{r}|$, relative to the mean density of the universe. Assuming that the number density of galaxies of any mass scales with $g(r)$, the expected number density of detections per unit log-mass x and comoving volume is $\phi(x|\boldsymbol{\theta})f(x, \mathbf{r})g(r)$ and equation (15) becomes

$$V_{\text{LSS}}(x) = \int f(x, \mathbf{r})g(r)d^3r. \quad (19)$$

In the absence of LSS, we expect $g(r) \equiv 1$ and equation (19) reduces to equation (15). If a model of $g(r)$ is known, for instance from a pre-existing survey, LSS is accounted for using equation (19) in the MML method.

Interestingly, $g(r)$ can be estimated, up to an overall normalization factor, directly from the distance distribution of the galaxies in the sample. Here, we restrict the discussion of this procedure to the case of an isotropic selection function $f(x, r)$ (see Section 2.2.1). In Appendix B, we show that if $g(r)$ is derived from the data, the LSS bias-corrected effective volume of equation (19) becomes

$$V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}}) = \sum_i \frac{f(x, r_i)}{\int \phi(\tilde{x}|\hat{\boldsymbol{\theta}})f(\tilde{x}, r_i)d\tilde{x}}. \quad (20)$$

As with the posterior PDFs of the data, $\tilde{\rho}_i(x|\hat{\boldsymbol{\theta}})$, the effective volume with LSS is determined at the MML solution $\hat{\boldsymbol{\theta}}$. Since this solution is a priori unknown, $V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}})$ is also evaluated iteratively as part of the algorithm introduced in Section 3.1.

Equation (20) requires a selection function, $f(x, r)$. A special situation is the case of a sharp survey limit, $f(x, r) \in \{0, 1\}$, such that $f(x, r) = 0$ if and only if x is smaller than a distance dependent threshold $x_{\min}(r)$, shown as the black line in Fig. 1. In this case, equation (20) reduces to

$$V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}}) = \sum_{i \in \{x \geq x_{\min}(r_i)\}} \left(\int_{x_{\min}(r_i)}^{\infty} \phi(\tilde{x}|\hat{\boldsymbol{\theta}})d\tilde{x} \right)^{-1}. \quad (21)$$

If $x_{\min}(r)$ increases monotonically with r , one can use the approximation

$$x_{\min}(r) = \min_{r_i \geq r} (x_i), \quad (22)$$

as illustrated in Fig. 1 (solid blue line). There is no need to use the approximations of equations (21) and (22) or variations thereof, unless the selection function $f(x, r)$ is unavailable.

It is important to stress that the overall normalization of $g(r)$ and therefore $V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}})$ cannot be derived from the data. It is simply impossible to know purely from a list of galaxies whether the survey volume represents an underdensity or an overdensity, relative to the rest of the universe. We must therefore make a choice of how to normalize $V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}})$, for instance by demanding that

$$\int \phi(x|\hat{\boldsymbol{\theta}})V_{\text{LSS}}(x|\hat{\boldsymbol{\theta}})w(x)dx = \int \phi(x|\hat{\boldsymbol{\theta}})V(x|\hat{\boldsymbol{\theta}})w(x)dx, \quad (23)$$

where $w(x)$ is a weighing function. Choosing $w(x) \equiv 1$ preserves the total number of galaxies. If $w(x) = 10^x$, the total mass of the survey is conserved.

Equation (21) sums over the cumulative MF, representing the total density of galaxies with log-masses above $x_{\min}(r)$. This appearance

of the cumulative MF (or LF) is almost universal to all published approaches accounting for LSS. In our case, the cumulative MF naturally appears through the short analytical derivation in Appendix B and the assumption of a monotonic selection function, but it is instructive to follow the alternative derivations presented in the original works of Lynden-Bell (1971), introducing the so-called C^- method, Turner (1979), introducing the φ/ϕ method, Sandage et al. (1979) and Kirshner et al. (1979). All these and derived ‘density-corrected’ methods (e.g. Baldry et al. 2012; Wright et al. 2017), explicitly or implicitly make the same basic assumptions that led to equations (21) and (22) and account for LSS using the same idea of modelling $g(r)$ directly from the distance- or redshift-distribution of the data. The advantage of the current formalism consists in the MML framework, which simultaneously handles mass uncertainties.

2.4 Multidimensional distributions

The entire formalism presented so far is straightforward to generalize to D -dimensional galaxy properties $\mathbf{x} \equiv (x_1, \dots, x_D)$ of any integer $D \geq 1$. For instance, if we construct the distribution of galaxies in the mass-size plane ($D = 2$), x_1 could represent the mass and x_2 the half-mass radius. Let us write the D -dimensional DF by generalizing equation (1) to

$$dN = \phi(\mathbf{x})Vd^Dx, \quad (24)$$

where dN denotes the expected number of galaxies in the infinitesimal volume d^Dx around \mathbf{x} . Given a model $\phi(\mathbf{x}|\boldsymbol{\theta})$ with model parameters $\boldsymbol{\theta}$, the expected number density of detections per unit of x_1, x_2, \dots and x_D then reads

$$\lambda(\mathbf{x}|\boldsymbol{\theta}) = \phi(\mathbf{x}|\boldsymbol{\theta})V(\mathbf{x}), \quad (25)$$

where $V(\mathbf{x})$ is calculated in analogy to the 1D case. For instance, if we know the multidimensional selection function $f(\mathbf{x}, \mathbf{r})$, then $V(\mathbf{x}) = \int f(\mathbf{x}, \mathbf{r})d^3r$ [see equation (15)]; or, if we only know the effective volume V_i of each galaxy, then $V(\mathbf{x})^{-1}$ can be linearly interpolated between V_i^{-1} (see Section 2.2.2). The bias-corrected observed source counts are (generalization of equations 6 and 9),

$$n(\mathbf{x}|\hat{\boldsymbol{\theta}}) = \sum_i \frac{\rho_i(\mathbf{x})\lambda(\mathbf{x}|\hat{\boldsymbol{\theta}})}{\int \rho_i(\mathbf{x})\lambda(\mathbf{x}|\hat{\boldsymbol{\theta}})d^Dx}. \quad (26)$$

The modified likelihood function (equation 10) then generalizes to

$$\ln \mathcal{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \int (n(\mathbf{x}|\hat{\boldsymbol{\theta}}) \ln \phi(\mathbf{x}|\boldsymbol{\theta}) - \lambda(\mathbf{x}|\boldsymbol{\theta}))d^Dx. \quad (27)$$

All comments on $\ln \mathcal{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ made following equation (10) also apply to equation (27). A numerical example of the MMLE for $D = 2$ is presented in Section 5.6.

2.5 Parameter uncertainties

At second order, the covariance matrix $\text{cov}(\hat{\boldsymbol{\theta}})$ of the best-fitting parameters $\hat{\boldsymbol{\theta}}$ can be approximated as (Laplace approximation)

$$\text{cov}(\hat{\boldsymbol{\theta}}) = -(H(\hat{\boldsymbol{\theta}}))^{-1}, \quad (28)$$

where H is the Hessian matrix (second derivatives) of the log-likelihood. However, the Hessian of our modified log-likelihood $\ln \mathcal{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ (equation 10), defined as

$$H(\hat{\boldsymbol{\theta}})_{ij} = \left. \frac{\partial^2 \ln \mathcal{L}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (29)$$

is not identical to the Hessian of the standard log-likelihood (despite the identity of the MLE and MMLE), as proven in Appendix A. Hence, equation (28) is not necessarily a good approximation. This inaccuracy is negligible if the measurement uncertainties of the data $\{x_i\}$ are small (< 50 per cent) compared to the range (standard deviation) of all data $\{x_i\}$. It is possible to express the correct Hessian of the standard log-likelihood at the MMLE solution [see equation (A7)], but its numerical evaluation is rather laborious, requiring $3P^2ND$ integrals. Moreover, neither of the Hessian approaches normally accounts for parameter uncertainties due to the removal of the LSS bias (Section 2.3), which is itself uncertain. Also, if the model parameters are fully degenerate or if the likelihood is non-Gaussian (e.g. non-linear parameter correlations), the Laplace approximation breaks down.

These limitations of the Hessian approach in MML can all be addressed by estimating $\text{cov}(\hat{\theta})$ via a non-parametric bootstrapping (Efron & Tibshirani 1993) approach that resamples the N data points, treating them as the whole population. Explicitly, one performs $Q > 1$ bootstrap iterations, labelled $q = 1, \dots, Q$. Each iteration includes three steps: (1) choose a random number N_q from a Poisson distribution with expectation N ; (2) draw a new sample of N_q data points (e.g. galaxy masses) from the original sample of N points *with replacement* (i.e. allowing for repetitions); and (3) determine the MMLE $\hat{\theta}_q$ of this new sample. The covariance matrix of the original MMLE $\hat{\theta}$ is then approximated as the covariances of the $\{\hat{\theta}_q\}$. Following Babu & Singh (1983), $Q \approx N(\ln N)^2$ iterations typically suffice for a good estimate. An explicit example of this method is provided in Section 5.5.

If the MML method is performed while correcting for LSS bias (Section 2.3), we can either refit $V_{\text{LSS}}(x|\hat{\theta}_q)$ at each resampling iteration q or fix this function to $V_{\text{LSS}}(x|\hat{\theta})$ across all iterations. Depending on this choice, the bootstrap parameter covariances, respectively, include or exclude the uncertainty of the cosmic LSS itself. Accounting for the limited knowledge of LSS generally increases the uncertainties of the model parameters, sometimes by a significant amount as illustrated in the example of Section 5.1. When fitting real galaxy surveys it is hence advisable to quote both the uncertainties with and without LSS uncertainties.

2.6 Estimator bias correction

The MMLE (or MLE) can be biased, meaning that its expectation $E[\hat{\theta}]$, equal to the average $\hat{\theta}$ for an infinite number of random samples from the same population, differs from the true population model θ_{true} . This is a general property of the MLE: only as the sample size tends to infinity, is the MLE guaranteed to be unbiased (Kendall & Stuart 1979 for theory; Appendix A of Robotham & Obreschkow 2015 for an example).

It is possible to construct a bias-corrected MLE analytically using the higher order derivatives of the likelihood function (Cordeiro & Klein 1994). This approach demonstrates that the leading bias term varies as N^{-1} , but the correction terms are rather cumbersome and depend on the choice of the MF model $\phi(x|\theta)$. Here, we resort to a more generic jackknifing approach to approximately correct the bias to order N^{-1} of any ML estimator. Following Efron & Stein (1981), the bias-corrected estimator reads

$$\tilde{\theta} = N\hat{\theta} - \frac{N-1}{N} \sum_{i=1}^N \hat{\theta}_{(i)}, \quad (30)$$

where $\hat{\theta}_{(i)}$ is the ML estimator of the $N-1$ galaxies, in which object i has been removed from the original sample of N galaxies.

Importantly, since the number of galaxies itself affects the overall normalization of the MF, the parameters $\hat{\theta}_{(i)}$ must be estimated using the renormalized volumes

$$V_{(i)}(x) = \frac{N-1}{N} V(x). \quad (31)$$

As illustrated in Section 4.3, this bias correction performs remarkably well. That said, it is often debatable whether the true MMLE $\hat{\theta}$ or the bias-corrected estimator $\tilde{\theta}$ is the ‘better’ solution. In many ways unbiased estimators exhibit less favourable properties (see Hardy 2002 for an example). The choice ultimately depends on the application. In any case, we will show (e.g. Fig. 7) that the difference between $\hat{\theta}$ and $\tilde{\theta}$ only becomes appreciable for very small samples ($\lesssim 10$ galaxies) and even then their difference is small compared to the overall parameter uncertainties.

3 NUMERICAL IMPLEMENTATION

3.1 Optimization algorithm

Solving the implicit equation (11) is a tricky optimization problem. This is because of the subtlety that its solution $\hat{\theta}$ is *not* obtained by maximizing $\mathcal{L}(\hat{\theta}, \hat{\theta})$. In other words, the solution $\hat{\theta}$ of equation (11) does not generally satisfy $(\partial_{\theta} \mathcal{L})(\hat{\theta}, \hat{\theta}) = 0$ (Appendix A). To solve equation (11), we developed a customised algorithm, referred to as the ‘fit-and-debias’ algorithm: First, evaluate the observed source count function $n_0(x) = \sum_i \rho_i(x)$. Then, repeat the following iteration for $k = 1, 2, 3, \dots$:

Find the parameter-vector $\hat{\theta}_k$ that maximizes

$$\ln \mathcal{L}(\hat{\theta}_k) = \int (n_{k-1}(x) \ln \phi(x|\hat{\theta}_k) - \phi(x|\hat{\theta}_k) V(x)) dx. \quad (32)$$

Use $\hat{\theta}_k$ as new estimator to debias the source counts,

$$n_k(x) = \sum_i \frac{\rho_i(x) V(x) \phi(x|\hat{\theta}_k) dx}{\int \rho_i(x) V(x) \phi(x|\hat{\theta}_k) dx}. \quad (33)$$

The algorithm can be stopped as soon as a certain convergence criterion is reached, for instance if

$$\|\hat{\theta}_k - \hat{\theta}_{k-1}\| \leq \Delta, \quad (34)$$

where $\Delta \in \mathbb{R}_+$ is a predefined tolerance. In this work, we set Δ equal to the relative precision error ($\sim 10^{-15}$) of the double-precision floating-point representation (IEEE 754). If a guess of initial parameters $\hat{\theta}_0$ is available, the fit-and-debias algorithm can be accelerated by evaluating n_0 via equation (33) instead of using $n_0 = \sum_i \rho_i(x)$.

The algorithm can readily account for (unknown) cosmic LSS (theory in Section 2.3). It suffices to substitute $V(x)$ in equations (32) and (33) for $V_{\text{LSS}}^{k-1}(x)$ and add a third step.

Use $\hat{\theta}_k$ to update the effective volume,

$$V_{\text{LSS}}^k(x) = A \sum_i \frac{f(x, r_i)}{\int \phi(\tilde{x}|\hat{\theta}_k) f(\tilde{x}, r_i) d\tilde{x}}, \quad (35)$$

with a normalization factor A computed via equation (23) at $\hat{\theta} = \hat{\theta}_k$. Computing $V_{\text{LSS}}^0(x)$ requires an initial guess $\hat{\theta}_0$.

Appendix A proves analytically that the fit-and-debias algorithm always converges towards the solution of equation (11), which is itself unique and identical to the MLE of the true likelihood. To illustrate the typical convergence, we draw a random mock sample of 10^3 galaxies from a fixed Schechter function [with parameters

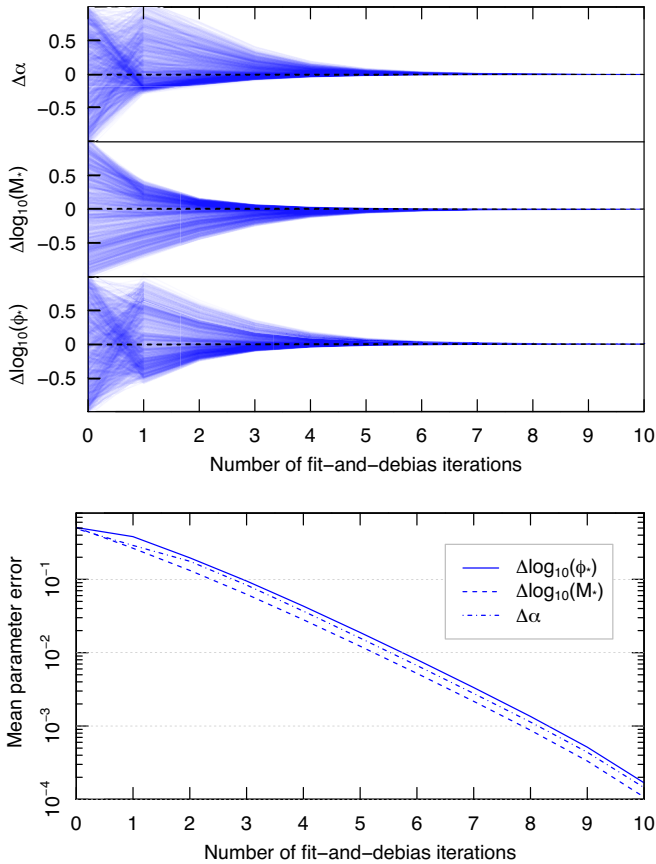


Figure 2. Test of the ‘fit-and-debias’ algorithm. Good (excellent) convergence is reached after 5 (10) iterations through equations (32) and (33). See Section 3.1 for details.

θ_{true} in equation (36)] using a sensitivity-limited effective volume – example discussed later in Section 4.2. The Schechter function parameters are then inferred from the mock data using the fit-and-debias algorithm, starting with initial parameters $\hat{\theta}_0$ that have been perturbed from θ_{true} by uniform random numbers in the interval $[-1, +1]$. This numerical experiment is repeated 10^3 times. As shown in Fig. 2 the algorithm quickly converges towards a stable solution in every single run. The top panel shows the evolution of the parameter errors $\Delta\theta$, defined as the absolute difference between the parameters at a given iteration and their final value after 20 iterations. Only the first 10 iterations are shown, since, for all practical purposes, the parameters are sufficiently converged after 10 iterations. The bottom panel shows the evolution of the average parameter errors $\Delta\theta$, revealing a monotonic decrease by a factor of ~ 2.5 per iteration. In Section 4.2, we will demonstrate that the solutions of the 10^3 random experiments are indeed consistent with the true parameters θ_{true} .

3.2 The R-package `dftools`

The fit-and-debias algorithm for the MML method has been implemented in the package `dftools` for the R-language, freely available for most operating systems (including Windows, MacOS, Linux). We refer the reader to the detailed documentation that comes with this package. In this documentation, all routines are explained alongside many examples. Here, we summarize the core functionality with some selected examples.

The `dftools` package is distributed via GitHub. To install the package in R use

```
install.packages('devtools')
library(devtools)
install_github('obreschkow/dftools')
```

The package is then activated by calling

```
library('dftools')
```

To view the inbuilt documentation, type

```
?dftools
```

The package includes a routine to generate galaxy data given an arbitrary MF and selection function. For instance, to draw a sample of 10^3 galaxies with Gaussian measurement errors of $\sigma = 0.5$ (in $\log_{10}x$) from a Schechter function (equation 2) with the default parameters $\theta_{\text{true}} = (\log_{10}\phi_*, \log_{10}M_*, \alpha) = (-2, 11, -1.3)$ and a built-in sensitivity-limited selection function, use

```
dat = dfmockdata(n=1e3, sigma=0.5)
```

All mock galaxies and survey specifications are stored in the list `dat`. For instance, the observed log-masses and their Gaussian uncertainties are stored in the vectors `dat$x` and `dat$x.err`, respectively, while the effective survey volume function $V(x)$ is stored in `dat$veff`. Given this mock survey, the most likely generative MF can be fitted via

```
survey = dffit(dat$x, dat$veff, dat$x.err)
```

This function executes the fit-and-debias algorithm (Section 3.1), using the in-built `optim` function with the default algorithm (Nelder & Mead 1965) to maximize equation (32). The output argument `survey` is a list of several sub-lists, such as `survey$data`, keeping track of the fitted galaxy data, and `survey$fit`, containing the fitted parameters and their covariances. To visualize the fit and mock data, type

```
ptrue = dfmodel(output='initial')
mfplot(survey, xlim=c(1e7, 2e12), p=ptrue)
```

This command produces a plot similar to Fig. 3 (without legend and true counts). The fit and its 68 per cent-uncertainty (light blue) is consistent with the input model. This can also be seen in the parameter covariance plot, obtained via

```
dfplotcov(list(survey, ptrue))
```

Unless other graphical parameters are specified, this plot (Fig. 4) shows the best-fitting parameters (blue dots) with their 68 per cent and 95 per cent confidence regions (ellipses) in the Gaussian approximation, as well as the input parameters (black crosses). The numerical values of the fitted parameters and their uncertainties can be displayed by calling

```
dfwrite(survey)
```

For an extended discussion of the physics and mathematics conveyed by Figs 3 and 4, we refer to the detailed examples in Section 4.

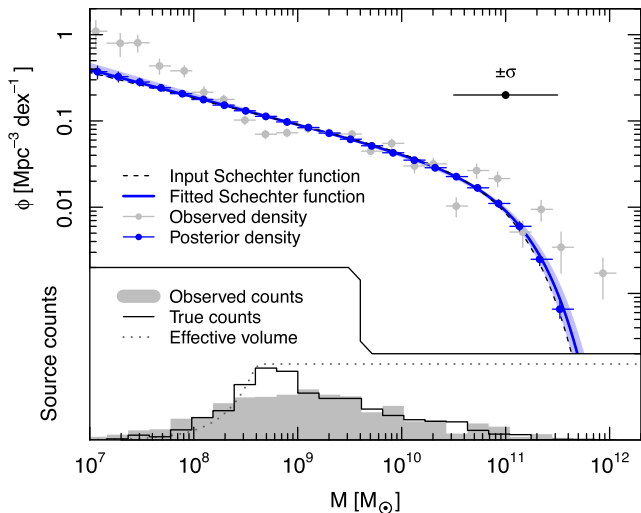


Figure 3. Basic example of `dftools`. 10^3 Mock galaxies are drawn from the input Schechter function, assuming the effective volume shown at the bottom and perturbing the masses by random log-normal errors of $\sigma = 0.5$ dex standard deviation. The resulting source counts are shown as grey histogram and the corresponding raw MF (using the $1/V_{\max}$ method) is shown in bins as purple dots. Horizontal bars are the bin widths and vertical bars are Poisson errors. A Schechter function is then fitted to these mock data using the fit-and-debias algorithm (equations 32 and 33), while accounting for the observational uncertainties. The best-fitting solution (blue solid line) and its 68 per cent confidence regions (blue shading) are consistent with the input MF; and posterior masses (blue dots), computed from the grey data via equation (6), align with the input model.

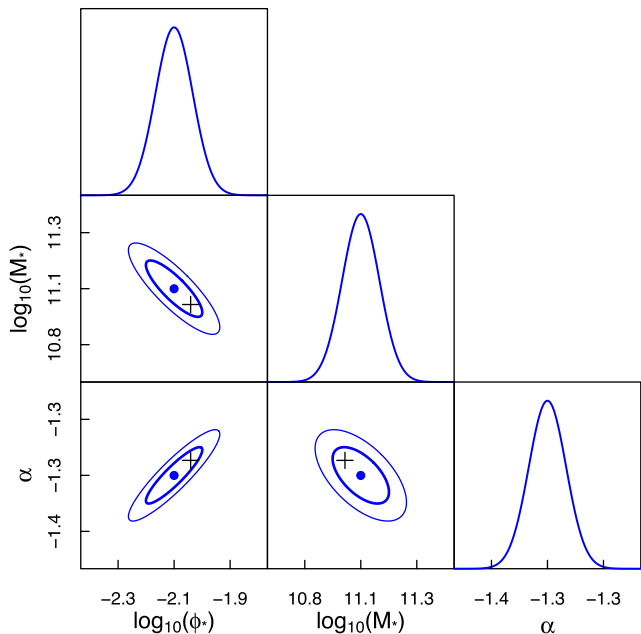


Figure 4. Uncertainties and covariances of the best-fitting parameters for the example in Fig. 3. The best-fitting values are plotted as blue dots with 68 per cent and 95 per cent confidence regions, in the Gaussian approximation, drawn as thick and thin blue lines. The true input parameters are given by the black crosses.

The `dftools` package includes various example routines that can be executed via `dfexample(case)`. `ss:2dexexample` using the integer argument `case` from 1 to 4 produces examples similar to those shown in Sections 3.2, 5.1, 5.2, and 5.6. The open source code of `dfexamples` discloses the implementation of these examples. More pedagogical tutorials, called ‘vignettes’ in *R*, will be included in the code and always updated along with the package.

4 BASIC EXAMPLES AND BENCHMARKS

To benchmark the MML formalism and its implementation in the package `dftools`, we generate and then fit mock surveys with precisely controlled input parameters. All surveys in this section assume that the true galaxy MF is a Schechter function (equation 2) with parameters

$$\theta_{\text{true}} = (\log_{10} \phi_*, \log_{10} M_*, \alpha) = (-2, 11, -1.3). \quad (36)$$

We will often drop the units, as they don’t matter for the examples. However, upon adopting ϕ_* in units of $\text{Mpc}^{-3} \text{dex}^{-1}$ and M_* in units of M_{\odot} , the vector θ_{true} is consistent with the rounded parameters of the observed galaxy MF (Bell et al. 2003; Papastergis et al. 2012) for baryonic matter (stars and cold gas). Other MF models are considered in Section 5.

Using this fixed input Schechter function, the following subsections consider widely different sample sizes to illustrate different effects. The largest sample ($N = 10^5$ galaxies, Section 4.1) serves to isolate the effect of Eddington bias from other effects and demonstrate its removal. The mid-sized samples ($N \approx 10^3$ galaxies, Section 4.2) serve to test the uncertainties of the MMLE and illustrate their dependence on the selection function. Finally, an array of small galaxy samples ($N \approx 10$ galaxies, Section 4.3) is picked to show the effect and removal of intrinsic estimator bias, only noticeable in such small samples.

4.1 Large galaxy samples

Let us model a galaxy survey that is purely sensitivity limited. For a constant mass-to-light ratio, this implies that a galaxy of mass M is detectable to a maximum distance proportional to $M^{1/2}$. Hence, the effective volume scales as

$$V(x) \propto M^{3/2} \propto 10^{1.5x}, \quad (37)$$

shown as the dotted line in the bottom panel of Fig. 5. We randomly pick $N = 10^5$ galaxies (i.e. log-masses x_i) from the expected source count function, i.e. from the PDF $\phi(x|\theta_{\text{true}})V(x)$. Each log-mass x_i is then perturbed by adding a random observing error Δx_i , drawn from a normal distribution with standard deviation $\sigma = 0.5$ (large horizontal error bar in Fig. 5). The increase in $V(x)$ with mass makes the source count distribution of the sample (histogram in Fig. 5) biased towards high masses compared to the underlying MF (short-dashed line in upper panel) – a typical feature known as Malmquist bias.

For purely illustrative purposes we bin the masses of this survey and compute the space density using the $1/V_{\max}$ method, i.e. by dividing the number of galaxies in each bin by the mean volume $V(x)$ associated with this bin. The resulting MF (purple points in Fig. 5) differs significantly and systematically from the input MF: it clearly overestimates the number of low-mass ($M < 10^9 M_{\odot}$) and high-mass ($M > 2 \cdot 10^{10} M_{\odot}$) galaxies, while underestimating the intermediate mass range. This important offset is due to Eddington bias: the declining MF makes it more likely that a galaxy with

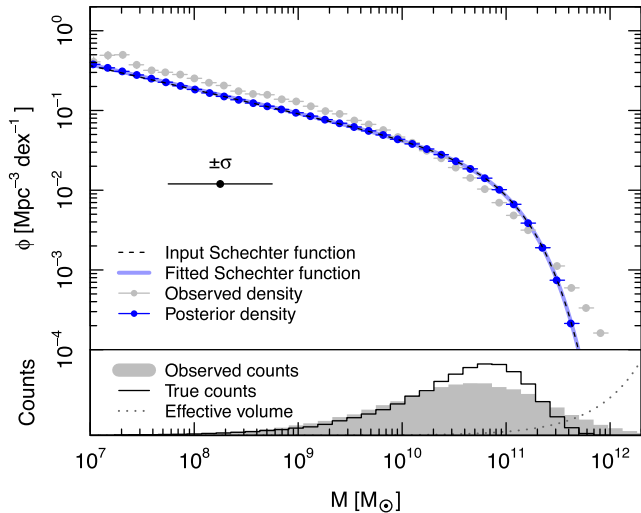


Figure 5. Example of recovering a Schechter function from a mock data set with 10^5 galaxies. Their large observing errors of standard deviation $\sigma = 0.5$ dex give rise to significant Eddington bias, which is fully removed by the fit. Details are given in Section 4.1.

an observed log-mass x is truly a lower-mass galaxy scattered upwards than a higher-mass galaxy scattered downwards. Overall this tends to smooth out the true MF, here by a Gaussian filter in log-mass x . The challenge consists in recovering the true MF, given the Eddington biased mock data – a key purpose of the MML method. The fit-and-debias algorithm converges in seven iterations in this example, taking only about 20s on a 3 GHz Intel Core i7 CPU, which is a very reasonable computation time for fitting 10^5 uncertain measurements. The fit $\hat{\theta} = (-2.000 \pm 0.004, 10.998 \pm 0.002, -1.303 \pm 0.003)$ is statistically consistent with and closely matched by the input parameters θ_{true} . In fact, graphically the fitted MF in Fig. 5 is indistinguishable from the input MF, illustrating the accurate removal of all Eddington bias.

We stress that the uncorrected Eddington bias (as seen in the poor fit of the purple data) is orders of magnitude larger than the uncertainties of the fit. Even if the observational uncertainties were as small as typical high-precision multiwavelength stellar mass errors of $\sigma = 0.1$ (in $\log_{10}M$; Wright et al. 2017), Eddington bias would still dominate over shot noise in a survey with 10^5 (or more) galaxies. We therefore expect the MFs of modern galaxy surveys (references in Section 1) to depend significantly on Eddington bias removal.

Finally, we can visualize the posterior data, computed as part of the MML method: equation (6) yields the posterior PDFs $\tilde{p}_i(x|\hat{\theta})$ for the log-mass of each galaxy i individually, which can be summed up via equation (9) to obtain to posterior for the observed source counts, $n(x|\hat{\theta})$. The posterior observed MF is then computed as

$$\tilde{\phi}(x|\hat{\theta}) = \frac{n(x|\hat{\theta})}{V(x)}. \quad (38)$$

For illustrative purposes, we chose to bin this function into the same mass-bins as the observed data, while defining the bin-values as the mean of $\tilde{\phi}(x|\hat{\theta})$ and the bin-centre as the $\tilde{\phi}$ -weighted mean of x in each bin. These binned posterior data (black points in Fig. 5) admit a similarly excellent agreement with the input MF as the best-fitting model itself.

4.2 Medium galaxy samples

The previous example was so large that the fitting errors vanish on the scale of Fig. 5. To discuss these errors, their covariance and dependence on the selection function, we now transition to smaller samples of $N \approx 10^3$ galaxies, where the expected model uncertainties are much larger. The samples are again drawn from the input Schechter function with parameters in equation (36), but using two selection functions: one that is purely sensitivity-limited ($V \propto M^{3/2}$) and one that is approximately volume-limited, providing a constant effective volume in the mass range $10^8 M_{\odot} \leq M \leq 2 \times 10^{11} M_{\odot}$ with a deliberate exponential cut-off for higher masses. The effective volumes of these two mock surveys are shown as the dotted lines in the ‘selection’ panel of Fig. 6. The respective expected source count densities $\phi(x|\theta_{\text{true}})V(x)$ are shown as grey shading. The galaxy masses are randomly drawn from this source count density and then perturbed with random observing errors from a log-normal distribution with standard deviation σ , where σ itself is different for each galaxy and drawn from a uniform random distribution between 0 and $\sigma_{\text{max}} = 0.5$. Hence, the mean error scale is $\sigma_{\text{mean}} = 0.25$. We chose to vary the uncertainty scale for each source to verify the MML method in this case.

For both selection functions, 10^4 random mock samples of $N \approx 10^3$ galaxies are generated and fitted with a Schechter function using `dftools`, while accounting for the different measurement uncertainties. The actual number of galaxies in each sample is itself drawn from a Poisson distribution with expectation $\langle N \rangle = 10^3$ to mimic the shot noise inherent to any real survey. The fits are plotted as light blue lines in the upper panel of Fig. 6. The distribution of these blue lines are very different in Fig. 6(a) and (b): in the first case, the mock sample is biased towards massive galaxies, leaving the low-mass end of the model poorly constrained; the second case shows the opposite situation. We deliberately chose these two extreme cases to illustrate the robustness of the MML method irrespective of the selection function.

The distributions of the 10^4 fitted Schechter parameters are displayed as blue histograms in the bottom panels of Fig. 6. They are approximately Gaussian, as expected to the extent that the Laplace approximation applies, i.e. that the log-likelihood is described by a second-order Taylor expansion around its maximum. The parameter covariances are shown as blue point-clouds with 68 per cent (thick blue lines) and 95 per cent (thin blue lines) elliptical contours. These contours are centred on the average fitted parameters (blue dots). For comparison, the black crosses show the input parameters θ_{true} and the black lines, centred on these parameters, show the average Gaussian uncertainties and covariances predicted from the averaged inverse Hessians of the log-likelihoods (see Section 2.5).

The covariance figures convey two messages: First, the agreement between mean fitted parameters and the input parameters is excellent in the sense that their difference is small (< 10 per cent) relative to the mean parameter uncertainties. This can also be seen in the visual overlap of the Schechter function associated with the mean parameters (dashed blue line) and the input Schechter function (thick black solid line). Statistically, the differences between the average fits and input parameters are consistent with being equal to zero. In other words, the MMLE behaves nearly like an unbiased estimator (i.e. its expectation matches the true population value) in these examples. Secondly, the agreement between the parameter covariances determined from the 10^4 runs and those predicted from the Hessian is good in the sense that the difference is smaller than the actual variances. Hence, for practical purposes, the Hessian approximation of the parameter uncertainties normally suffices. This

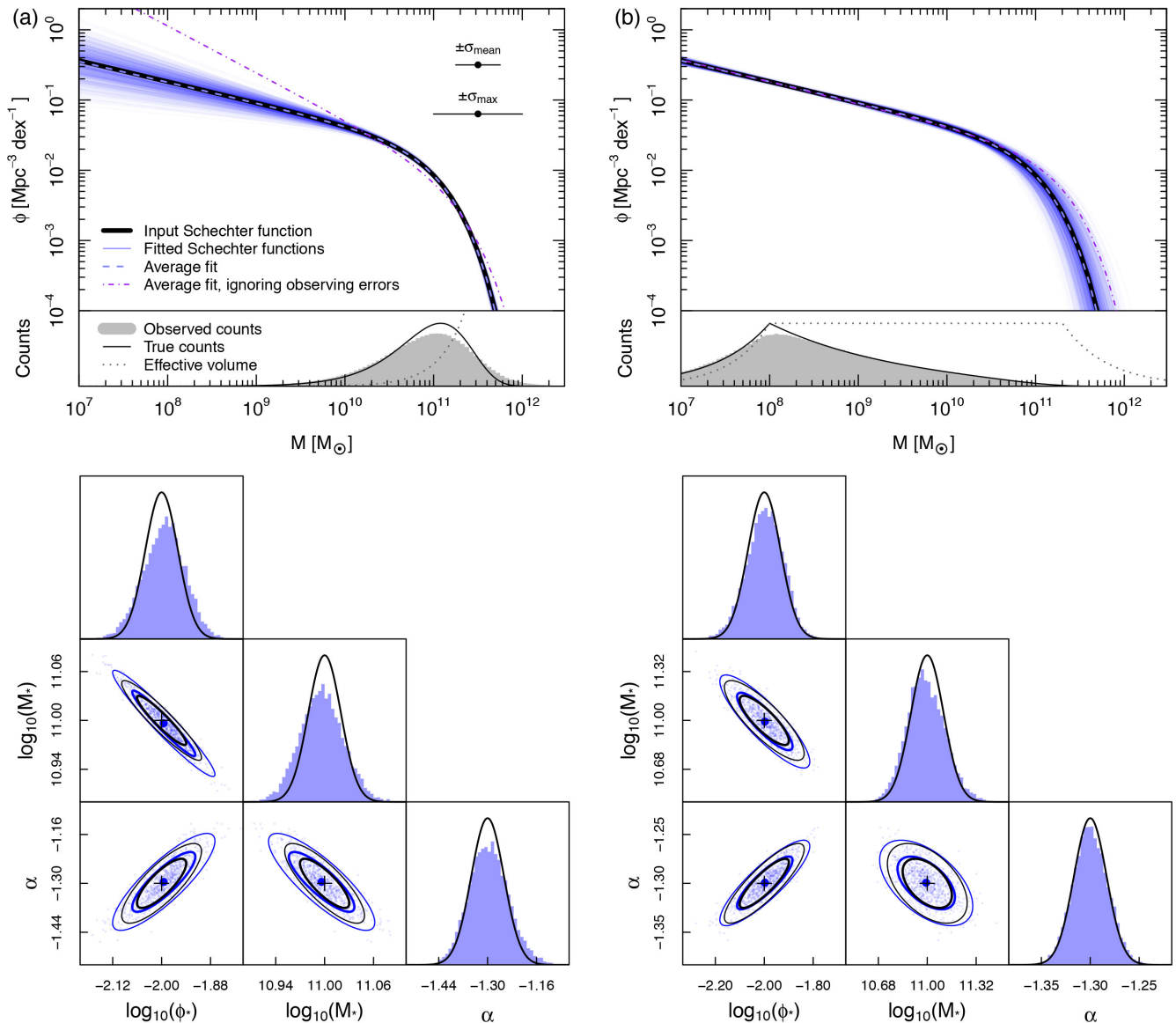


Figure 6. Examples of recovering a Schechter function from mock surveys with 10^3 galaxies, in the presence of widely different selection functions, biased towards high masses (a) and low masses (b). In each case, the MML fit is repeated with 10^4 random mock samples, resulting in a zoo of MFs with parameter distributions shown in the bottom panels. The mean fits are in excellent agreement with the input model and the parameter distributions (blue ellipses) roughly match the Hessian predictions (black ellipses). Details are given in Section 4.2.

statement does not apply in general, but counter-examples are rare and rather unphysical (see Section 5.5), except when accounting for cosmic LSS (see Section 5.1).

Finally, we emphasize that Eddington bias – the focus of Section 4.1 – also affects the examples of Fig. 6, but has been dealt with automatically by the MML method. To show this, Fig. 6 also displays the Schechter functions associated with the average parameters obtained when ignoring observational errors (dash-dotted purple lines). These MFs significantly deviate from the input function, especially in the poorly constrained parts, due to Eddington bias.

4.3 Small galaxy samples

As explained in Section 2.6, the MMLE, like the MLE, is expected to be slightly biased, i.e. its expectation differs from the true population

model. This estimator bias vanishes as N^{-1} as the samples size $N \rightarrow \infty$ (Kendall & Stuart 1979) and is hence most pronounced when fitting small samples. In the examples of Sections 4.1 and 4.2 based on $N = 10^5$ and $N = 10^3$ galaxies, the bias was negligible. So let us consider mock surveys of only $N \approx 5$ to $N \approx 100$ galaxies. These surveys are drawn from the input Schechter function [with parameters in equation (36)] using an effective volume varying as $V(x) \propto M^{0.8}$ (dotted line in Fig. 7), which is similar to a sensitivity-limited survey (equation 37), but less biased towards high masses to ensure that the low-mass end of the MF is at least marginally constrained. The expected source count density is shown as grey shading in Fig. 7. The masses drawn from this distribution are perturbed with observing errors from a log-normal distribution of standard deviation $\sigma = 0.3$ (error bar in Fig. 7).

We generate surveys of five different sizes with expected numbers of objects ($\langle N \rangle = 5, 10, 20, 50, 100$). For each size, 10^4 mock surveys

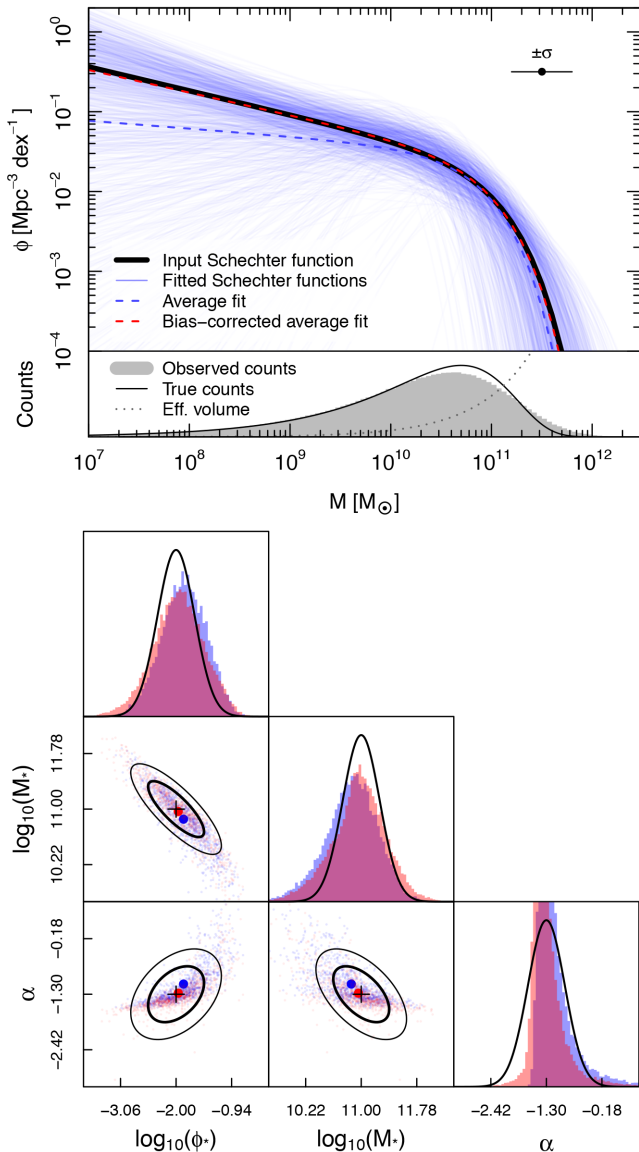


Figure 7. Example of recovering a Schechter function from mock surveys with only 10 galaxies. The MML fit is repeated with 10^4 random mock samples, resulting in a zoo of MFs with parameter distributions shown in the bottom panel. The average fit deviates significantly from the input model, revealing the classical estimator bias of the ML method for small samples. This bias is corrected using jackknifing. The parameter distributions of all 10^4 runs are shown in the bottom panel for the normal (blue) and bias-corrected (red) MML estimator. Details are given in Section 4.3.

are generated, each with an actual number of N galaxies, drawn from a Poisson distribution of mode $\langle N \rangle$. Surveys with only two or less galaxies are excluded (12 per cent for $\langle N \rangle = 5$ and 0.3 per cent for $\langle N \rangle = 10$), as they can admit unbound MLEs for three parameters. The MML method is then used to fit a Schechter function to each survey, while accounting for the observational uncertainties.

For the $\langle N \rangle = 10$ surveys, the Schechter fits are shown as light blue lines in Fig. 7. Naturally, the scatter between these fits is much larger than in the previous example (Fig. 6) due to the 100-times smaller sample size. The distributions and covariances of the fitted parameters are shown in blue in the bottom panel of Fig. 7. As in the previous example, black lines represent the covariances from the Hessian of the log-likelihood. The average fitted param-

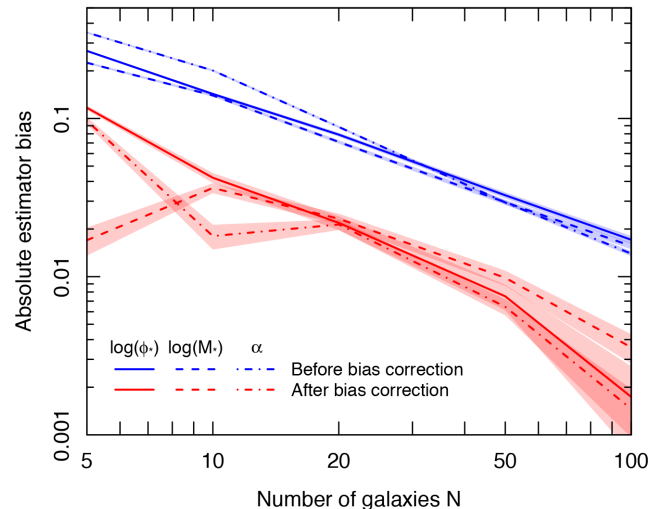


Figure 8. The MML estimator is, like any ML estimator, biased. Plotting this bias as a function of N reveals the typical bias scaling as approximately N^{-1} . The first-order jackknifing method reduces the bias by an order of magnitude. Details are given in Section 4.3.

ters (blue dots) lie significantly off the input values (black crosses). Analogously, in the upper panel, the Schechter function associated with the average parameters (dashed blue line) clearly differs from the input Schechter function (thick black line). This difference between expected MML fits and true parameters is the estimator bias we aimed to evidence. While clearly visible, this bias remains small compared to the parameter uncertainties. Hence, correcting this bias might not be necessary.

If desired, the MMLE bias can be removed to first order in N^{-1} by the jackknifing method presented in Section 2.6. Applying this method results in the red parameter distributions in Fig. 7. The average best-fitting parameters produce the Schechter function shown as the red dashed line in the upper panel. This function is almost identical to the input Schechter function, demonstrating the effectiveness of this approach. The precise biases of the corrected (red) and uncorrected (blue) parameters is shown in Fig. 8 for all considered sample sizes. We find that jackknifing reduces the estimator bias by about an order of magnitude. The somewhat strange behaviour of the parameter errors at $N = 5$ is due to the higher order correction terms becoming important for such low numbers of galaxies.

5 ADVANCED EXAMPLES

So far, all examples focused on a homogenous universe without LSS, where the galaxy population is fully described by a three-parameter Schechter function. This section expands the view towards additional complications encountered when working with real data. All of these complications can be dealt with using `dftools`.

5.1 Cosmic large-scale structure

Galaxy surveys are inevitably subject to cosmic LSS, which can bias the reconstruction of the galaxy MF as explained in Section 2.3. To test the removal of this bias [via equation (20)], we consider a typical, sensitivity-limited survey with a fuzzy detection limit. Isocontours of the selection function $f(x, r)$ are shown in the upper panel of Fig. 9 for $f = 0.1$ (short-dashed line), $f = 0.5$ (solid line) and $f = 0.9$ (long-dashed line). Next, we pick a non-uniform function

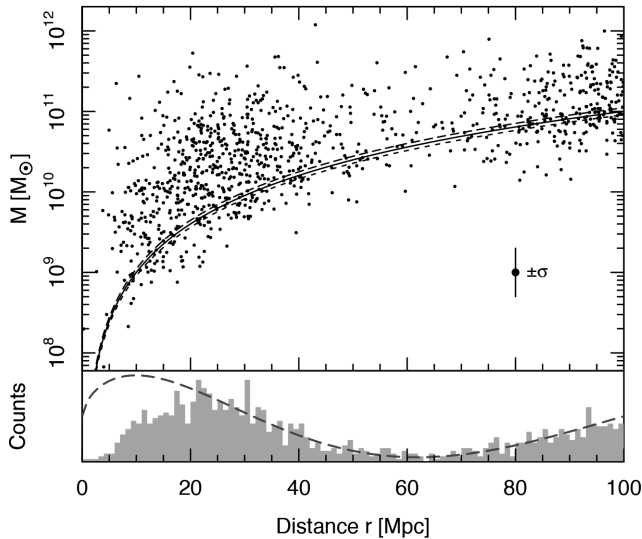


Figure 9. Distance–mass distribution of a mock sample with a sensitivity-limited selection function (solid line) assuming the non-uniform cosmic LSS represented by the function $g(r)$ shown as long dashed line in the bottom panel. Details are given in Section 5.1.

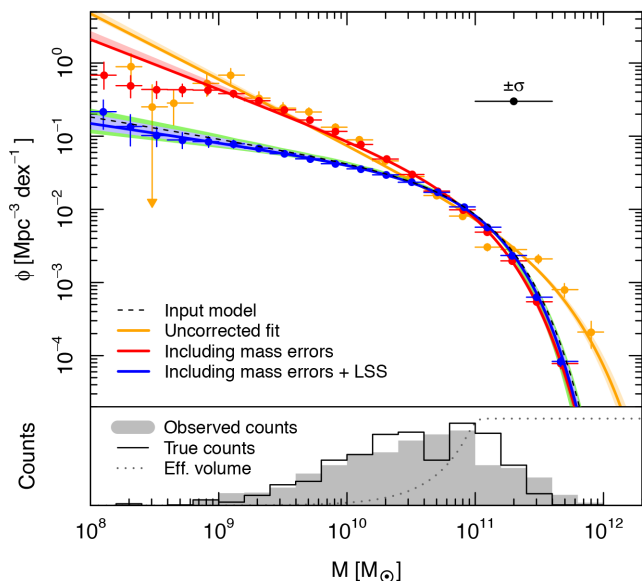


Figure 10. Illustration of fitting an MF to a mock sample (shown in Fig. 9) with non-uniform LSS. If LSS-correction is activated, the input function is accurately recovered. Details are given in Section 5.1.

$g(r)$ (dashed line in the bottom panel of Fig. 9), representing the number density contrast due to cosmic LSS. Using the resulting overall selection function $f(x, r)g(r)$ and our reference Schechter function (with parameters θ_{true} of equation (36)), we draw a sample of $N = 10^3$ galaxies and perturb their masses by random, log-normal observing errors of standard deviation $\sigma = 0.3$ (in $x = \log_{10} M/M_{\odot}$). The resulting mock sample (black points in Fig. 9) then admits the oscillating source counts shown as grey histogram in Fig. 10.

This mock sample is subject to two major biases: Eddington and LSS bias. If neither is accounted for when fitting a Schechter function (using a $1/V_{\text{max}}$ or simplistic ML approach), the best-fitting

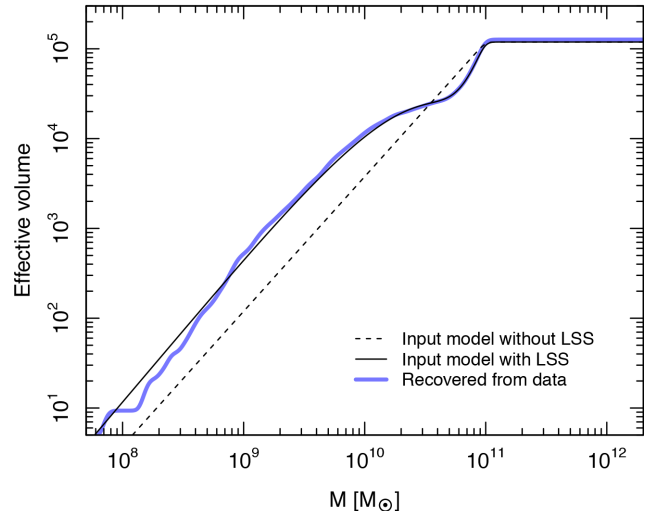


Figure 11. Effective volume functions. The *unknown* effective volume with LSS $V_{\text{LSS}}(x)$, needed to recover the MF in the presence of LSS, is approximately recovered using the known selection function $f(x, r)$ without LSS and the distance distribution of the survey (Fig. 9). Details are given in Section 5.1.

solution (yellow line in Fig. 10) differs widely from the input MF (dashed black line).

Let us first look at the solution, if only Eddington bias is corrected, i.e. mass observing errors are accounted for but not LSS. In this case, the effective survey volume is computed (automatically by `df-tools`) using equation (15). The effective volume is shown as the dashed line in Fig. 11. Solving the MML method using the ‘fit-and-debias’ algorithm in `df-tools`, results in the Schechter function fit shown as the red line in Fig. 10. This fit works well in the high-mass end, but fails on the low-mass side, dominated by LSS bias.

The LSS bias can be approximately removed by using equation (20) instead of equation (15) to compute the effective volume. To do so in `df-tools`, it suffices to set `correct.lss.bias = TRUE` when calling `df-fit`. The resulting effective volume is shown as the blue line in Fig. 11. This effective volume accounts for LSS to the extent that this LSS is imprinted in the distance distribution of the galaxies in the sample. It strongly resembles the ‘true’ effective volume with LSS (black solid line in Fig. 11), given in equation (19), which requires the input function $g(r)$ that is unknown to the observer. In fact, the best-fitting parameters $\hat{\theta} = (\log_{10} \phi_*, \log_{10} M_*, \alpha) = (-1.97 \pm 0.07, 10.96 \pm 0.04, -1.28 \pm 0.06)$ are statistically consistent with the input parameters θ_{true} .

The parameter uncertainties quoted above are standard deviations, i.e. square roots of the diagonal covariance elements, computed from Hessian matrix (see Section 2.5). These covariant parameter uncertainties are represented by the light blue envelope around the solid blue line in Fig. 10. As detailed in Section 2.5, uncertainties can also be computed by resampling the data. In `df-fit`, this is achieved by specifying an integer value for the argument `n.bootstrap` (equal to Q in Section 2.5). The user can choose whether to refit $V_{\text{LSS}}(x)$ at each resampling iteration via the logical argument `lss.errors`. If set to `FALSE` (no refitting of $V_{\text{LSS}}(x)$), the resulting parameter covariances are statistically consistent with the values computed from the Hessian matrix. If set to `TRUE`, $V_{\text{LSS}}(x)$ is refitted at each iteration. This approach results

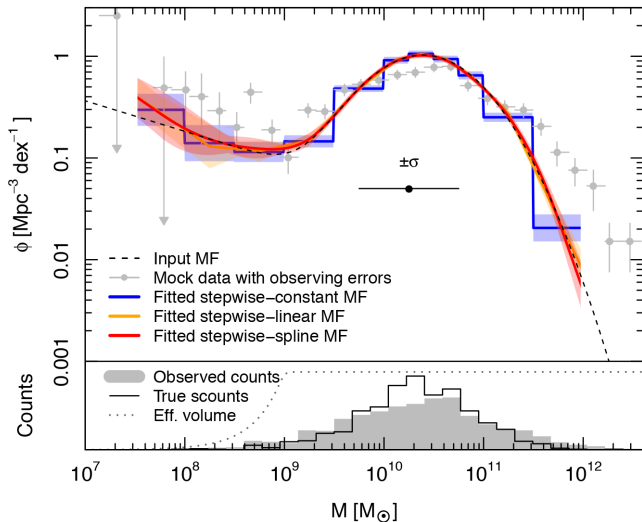


Figure 12. Illustration of fitting quasi-non-parametric MF models to mock data generated from a survey drawn from the dashed input MF. Details are given in Section 5.2.

in ~ 2 times larger standard errors, represented by the light green envelope in Fig. 10. The parameter uncertainties are bound to increase if $V_{\text{LSS}}(x)$ is refitted at each iteration, because the variance of $V_{\text{LSS}}(x)$ implies an additional uncertainty in the model parameters. The bootstrap method allows us to evaluate this additional uncertainty.

This error analysis reveals that the uncertainties of the most likely MF parameters increase significantly if the uncertainty of the LSS is accounted for. As emphasized in Section 2.5, it is therefore advisable to always quote parameter uncertainties with and without LSS uncertainties, when fitting real galaxy data.

5.2 Quasi-non-parametric modelling

So far all examples have used galaxy populations generated and fitted by a Schechter function. The MML formalism can none the less deal with any real MF model $\phi(x|\theta)$, including so-called ‘non-parametric models’, which parametrize the MF in bins rather than a single analytical function. In the literature, the ML formalism for fitting such binned MFs is often called the stepwise ML (SWML) formalism. It was first introduced by Efstathiou et al. (1988). Unlike classical SWML methods, our MML formalism fully accounts for observing errors (Eddington bias).

The `dftools` package includes the function `dftswmodel` to generate stepwise MFs, which can then be fitted using `dffit` in exactly the same way as when fitting other MFs (cf. example in Section 3.2). To test this approach we generate a mock sample of $N = 10^3$ galaxies in the same way as in previous examples. However, this time the input MF (short-dashed black line in Fig. 12) differs from a Schechter function in that it has a second turning point with opposite curvature. Using this arbitrary input MF with the effective volume shown as the dotted line in the bottom panel and random observing errors of $\sigma = 0.5$ (in $x = \log_{10} M/M_{\odot}$) results in the number counts of the grey histogram. The binned MF obtained using the $1/V_{\text{max}}$ method is shown as grey data points. These points differ significantly from the input MF due to Eddington bias.

As in all previous examples, we then fit the mock data using the MML method. As fitting functions, we use three stepwise MFs generated by `dftswmodel`:

(i) A stepwise MF made of N_{bins} constant bins (blue line in Fig. 12) is the simplest quasi-non-parametric form.

(ii) Better fits can be obtained by choosing the model as a power law in each bin (yellow line). Upon requiring this stepwise power law to be continuous, this function is fully specified by its values at the N_{bins} bin centres.

(iii) The MF can also be modelled by a continuous cubic spline (red line) connecting vertices at the N_{bins} bin centres.

The function `dffit` automatically extrapolates these functions linearly outside their domain while finding the MML solution. This extrapolation avoids issues with measurements that lie on the edge of the MF domain. In the present example, the three stepwise MFs have been parametrized on $N_{\text{bins}} = 10$ bins of different sizes between $x = 7.5$ and $x = 12$. Fig. 12 highlights two clear advantages of the stepwise power-law and spline functions relative to the stepwise constant model (blue). First, they are statistically consistent with the input model (black dashed-line) at any point, not just somewhere along each mass bin. Secondly, by construction, they satisfy the continuity (and smoothness for spline) condition, which one would expect for most physically meaningful MFs. Therefore, it seems advisable to use these continuous MF models when performing stepwise fits.

5.3 Model evidence

Which is the right MF model to fit? The MML method can fit (nearly) any MF model $\phi(x|\theta)$, including quasi-non-parametric ones (cf. Section 5.2). So how can we decide, solely from the data, on the best model, at least amongst a finite set of proposals? Bayesian inference offers a powerful tool to answer this question: The conditional probability Z of a model given the data is proportional to the integral of the likelihood function over the full parameter space. We here compute this integral in the Laplace approximation (Daniels 1954), which treats the likelihood in the Gaussian approximation. In other words, the log-likelihood function is approximated at second order around its maximum. In the MML nomenclature, this approximation reads

$$Z = \int \mathcal{L}(\theta, \hat{\theta}) \exp \left[-\frac{1}{2} (\theta - \hat{\theta})^{\dagger} \text{cov}(\hat{\theta})^{-1} (\theta - \hat{\theta}) \right] d^P \theta$$

$$= \mathcal{L}(\hat{\theta}, \hat{\theta}) \sqrt{(2\pi)^P |\text{cov}(\hat{\theta})|}, \quad (39)$$

where $\mathcal{L}(\hat{\theta}, \hat{\theta})$ is the modified likelihood at the MML solution (Section 2.1), P is the number of scalar parameters (i.e. the number of elements in the vector θ) and $\text{cov}(\hat{\theta})$ is the covariance of the optimal parameters (Section 2.5). The value of $\ln Z$ is automatically computed when calling `dffit` and provided in the output list of this routine. If two competing models are a priori (i.e. before using the data) equally likely, then the odds of the first model over the second is given by the ratio $B = Z_1/Z_2$, known as Bayes factor.

As an illustration, we use an extension of the Schechter function to the four-parameter, $\theta = (\log_{10} \phi_*, \log_{10} M_*, \alpha, \beta)$, MRP function (Murray, Robotham & Power 2017),

$$\phi(x|\theta)_{\text{MRP}} = \ln(10) \phi_* \mu^{\alpha+1} e^{-\mu^{\beta}}, \quad (40)$$

where $\mu = M/M_* = 10^x M_{\odot}/M_*$. The only difference to the Schechter function is the additional parameter β , which modulates the steepness of the exponential cut-off at the high-mass end. A Schechter function is recovered if $\beta = 1$.

Fig. 13 shows the reference Schechter function [dotted line, parameters in equation (36)], compared to an MRP function (dashed

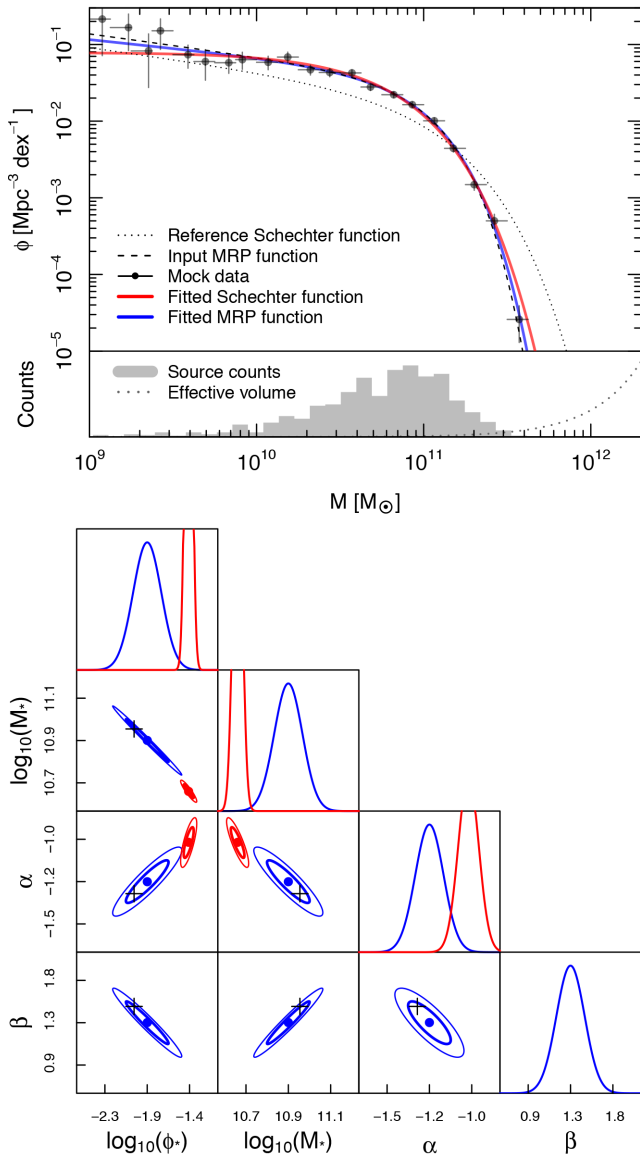


Figure 13. Example of fitting an MRP function and a Schechter function to a mock survey generated from an MRP function. The input MRP function can be approximately mimicked by a Schechter function with parameters ϕ_* , M_* , α differing from those used in the MRP function, as can be seen from the parameter covariances in the bottom panel. Details are given in Section 5.3.

line) with the same Schechter parameters and $\beta = 1.3$. Comparing the dotted to the dashed line, it is clear that the value $\beta > 1$ steepens the high-mass end and lowers the low-mass amplitude. From this MRP function, we draw a mock sample of $N = 10^3$ galaxies, adopting a sensitivity-limited effective volume function (equation 37) without LSS. For illustration, these data are binned by mass and shown as black dots in Fig. 13 – without observing errors, hence matching the input model (dashed line).

Using the MML method, we fit the mock data with both an MRP function (blue line in Fig. 13) and a Schechter function (red line). The respective model parameters and covariances are shown in the bottom panel. While the MRP function is closer to the input model (as expected), the Schechter function provides a surprisingly good fit. This is because values of $\beta \neq 1$ in the MRP function, can be partially compensated by adjusting the three Schechter function

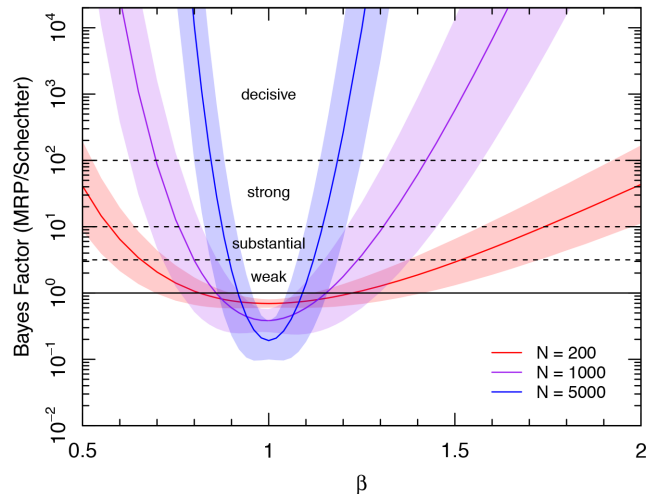


Figure 14. Evidence for the MRP model against the Schechter model as a function of the β -parameter in the input MRP function (where $\beta = 1$ is a Schechter function). Lines and shaded regions show the mean Bayes factors and standard deviations determined from 10^3 realizations for each β , for three different survey sizes. Details are given in Section 5.3.

parameters – a statement that is also obvious from the strong covariances of the Schechter parameters with β . Naturally, the best-fitting Schechter parameters differ significantly from the MRP parameters, because of this compensation of a $\beta \neq 1$.

If we do not know whether the data in Fig. 13 are drawn from an MRP or a Schechter function, finding the true population model is graphically quite tricky. The Bayes factor of the MRP model over the Schechter model is $B = 10.3 > 1$ in this example, meaning that the MRP model is favoured. Note that this factor drops to $B \approx 4$ if the LSS is considered to be unknown, because LSS could also be responsible for a deviation from the Schechter model (see Section 5.1).

One would expect that it becomes easier to distinguish between the MRP and Schechter model, if β deviates more strongly from $\beta = 1$ and if more data is available. This expectation is tested in Fig. 14, which shows the Bayes factor (not accounting for LSS and without mass uncertainties) as a function of β for three different survey sizes N . For each pair $\{\beta, N\}$, we generated and fitted 10^3 random mock surveys. The distribution of their Bayes factors (one standard deviation) is shown as transparent shading. Interestingly, the Bayes factor implicitly penalizes models with more free parameters – a property sometimes referred to as the Bayesian version of Ockham's razor. Therefore, if $\beta = 1$, the Bayes factor favours the three-parameter Schechter model over the four-parameter MRP model, despite the fact that the MRP function is identical to the Schechter function for $\beta = 1$. This penalization also implies that smaller samples require larger deviations from $\beta = 1$ in order to favour the MRP model.

In the terminology of Kass & Raftery (1995), a Bayes factor $B > 10$ (or $B < 0.1$) is called 'strong evidence' (other denominations shown Fig. 14). According to Fig. 14, a sensitivity-limited galaxy survey must detect at least $N = 200$ (10^3 , 5×10^3) galaxies to provide such strong evidence for a value of $\beta \approx 0.6$ (0.8, 0.9) or $\beta \approx 1.7$ (1.3, 1.1). Through explicit calculations, we found that roughly 10-times more galaxies are required if the (unknown) LSS is accounted for and mass uncertainties of $\sigma = 0.3$ (in $\log_{10}M$) are assumed.

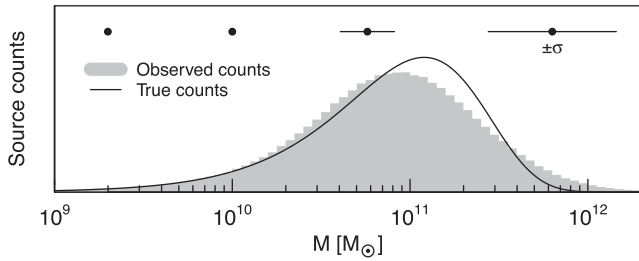


Figure 15. Source counts of the sensitivity-limited mock survey of Fig. 6(a), when subjected to observational uncertainties σ that depend on the mass. Details are given in Section 5.4.

5.4 Mass-dependent measurement uncertainties

The mock data considered so far included statistical errors, whose magnitude was independent of the measured mass. In most real measurements, the statistical uncertainty of a datum none the less depends on its value. Such systematic variations of uncertainties are naturally dealt with by the MML framework, given a correct handling of the prior PDFs $\rho_i(x)$ of the observed data. To illustrate this point, let us assume that a datum of true value x yields a measured value x_{obs} with probability $\varrho(x_{\text{obs}}|x)$. Hence, an observation i with measured value x_i has a true value x with probability

$$\rho_i(x) = \frac{\varrho(x_i|x)}{\int \varrho(x_i|\tilde{x})d\tilde{x}}. \quad (41)$$

As an explicit example, we reconsider the sensitivity-limited survey with $N = 10^3$ galaxies of Fig. 6(a), but assume the Gaussian uncertainty model

$$\varrho(x_{\text{obs}}|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left(-\frac{(x - x_{\text{obs}})^2}{2\sigma(x)^2}\right), \quad (42)$$

with a σ that depends on the true value x via,

$$\sigma(x) = \max(0, 0.2x - 2). \quad (43)$$

With this choice of $\sigma(x)$, the range and mean of σ in the samples is similar to that in the example of Fig. 6(a). The true source counts and perturbed by this mass-dependent error model are shown in Fig. 15.

We then evaluate the observed PDFs $\rho_i(x)$ via equation (41) and apply the MML algorithm to recover the most likely parameters of the Schechter function generating the data. This experiment is repeated with 10^4 random mock samples drawn from the same population, resulting in the fitted parameter distributions shown in blue in Fig. 16. The excellent match between the input parameters (black crosses) and the numerical expectation of the MML solution (big blue dots) demonstrates the applicability of the MML method to such non-trivial error models.

Note that the observed probabilities $\rho_i(x)$, computed via equation (41), are *not* Gaussian in this example, despite the Gaussian form of $\varrho(x_{\text{obs}}|x)$. This is a subtle, but crucial point. If, instead of using equation (41), we incorrectly forced the observed probabilities $\rho_i(x)$ to be Gaussians with standard deviations $\sigma(x_i)$, the most likely fitted parameters (purple in Fig. 16) were no longer consistent with the input model. This comparison emphasizes the general point that correct error models of the data are important for an accurate recovery of the population model – a statement that is not specific to the MML formalism.

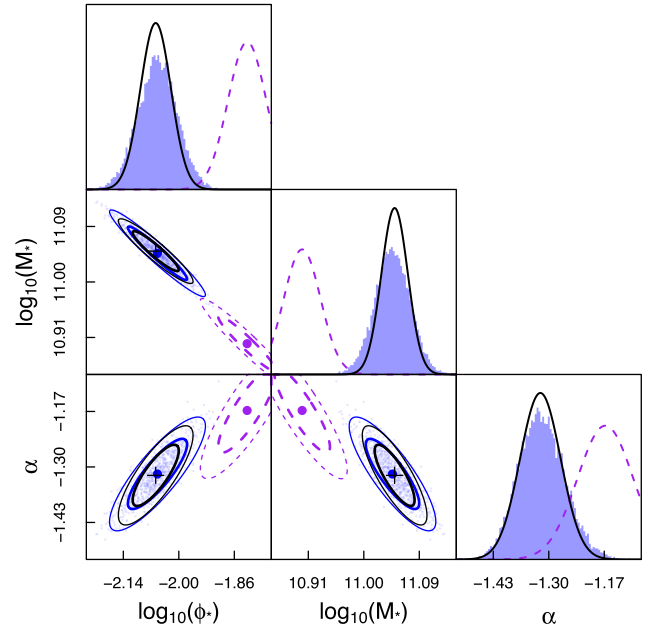


Figure 16. Schechter function parameters of a sensitivity-limited mock survey with mass-dependent uncertainties. Black crosses show the input parameters. Their covariances, computed from the Hessian of the modified likelihood function, are shown as black thick (68 per cent confidence) and thin (95 per cent) ellipses. Blue point-clouds and histograms represent the fitted parameters for 10^4 random mock samples. Their means and covariances are shown as big blue dots and ellipses. Purple dots and dashed ellipses indicate the parameter solutions, if the observational uncertainties are incorrectly assumed to be Gaussian. Details are given in Section 5.4.

5.5 Resampling uncertainties

The computations of parameter (co)variances and Bayes factors presented so far relied on estimating the covariance matrix from the Hessian matrix of the modified likelihood function (Laplace approximation). As noted in Section 2.5, this approach is only valid if the likelihood function is approximately Gaussian and if the uncertainties of the data are smaller than their range (standard deviation). In most realistic examples this is indeed the case, and the straightforward Hessian covariance estimations work well (cf. bottom panel of Fig. 6) – except that they exclude LSS uncertainties, as illustrated in Section 5.1.

The Hessian covariances become inaccurate if the observational uncertainties of the log-masses x are close to or larger than the range (standard deviation) of the log-masses themselves. This is best illustrated by adopting a Gaussian MF with a controlled standard deviation τ ,

$$\phi(x|\theta)_{\text{Gaussian}} = \frac{A}{\sqrt{2\pi}\tau^2} \exp\left(-\frac{(x - \mu)^2}{2\tau^2}\right). \quad (44)$$

The parameters A and μ set the amplitude and mode of the MF. As an example, we pick the parameters $\theta = (\log_{10} A, \tau, \mu) = (-1, 1, 9)$ and sample the MF adopting a constant effective volume without LSS. We generate 10^3 mock surveys, each with an expected number of $N = 10^3$ galaxies. The log-masses x are perturbed by Gaussian random errors of standard deviation $\sigma = 1$. Hence, the data uncertainty is identical to the width of the MF τ .

The distribution of the fitted model parameters is shown as blue histograms and point-clouds in Fig. 17. The blue ellipses are the 1σ (68 per cent) and 2σ (95 per cent) contours fitted directly to these points. For comparison, the black Gaussians and ellipses show the

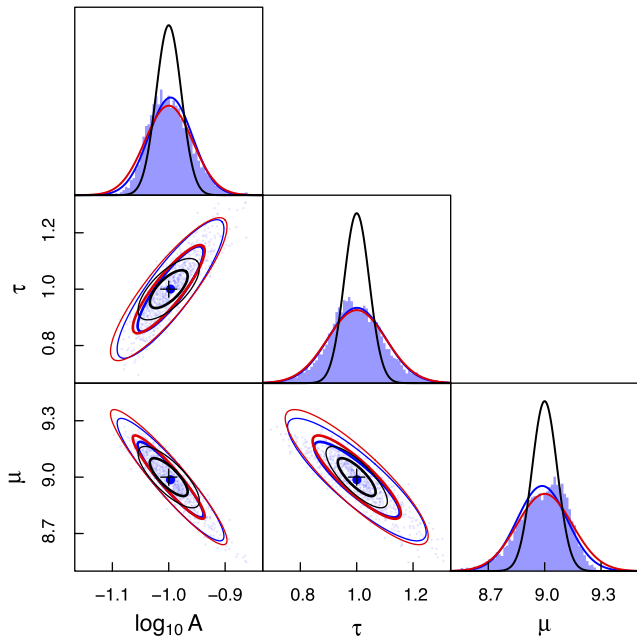


Figure 17. Illustration of parameter covariances in the case of very large observing errors that are comparable to the width of the MF given in equation (44). The adopted input parameters are shown as black crosses. The black ellipses centred on these crosses show the 68 per cent (thick lines) and 95 per cent (thin line) confidence intervals of fitted parameters predicted from the Hessian of the modified likelihood function. These predictions differ significantly from the best-fitting solutions of 10^3 random mock samples (blue). However, the covariance of the fitted parameters is well predicted using a resampling method (red lines). Details are given in Section 5.5.

average (co)variances computed from the Hessian matrices. The mean values of the fits (big blue dots) agree with the input parameters (black crosses), showing that the expectation of the MMLE is correct, i.e. the estimator bias (Section 2.6) is negligible for $N = 10^3$ galaxies. However, the (co)variances estimated from the Hessian are clearly too small, showing the failure of the Hessian approximation in the presence of large observing errors.

It is possible to compute accurate parameter covariances for any data errors via the bootstrapping method described in Section 2.5. This technique is implemented in `dftools` and activated by setting `n.bootstrap` when calling `dffit`. The argument `n.bootstrap` is the integer number of resampling iterations, called Q in Section 2.5.

The average covariances obtained by bootstrapping are shown as the red lines in Fig. 17. They agree with the numerical expectations (blue), within the statistical uncertainties of these expectations. This example demonstrates the power of bootstrapping in computing the covariances. Moreover, bootstrapping allows an accurate sampling of the parameter posterior, even if the parameter correlations are highly non-linear, i.e. if the covariance matrix provides a poor description of the parameter uncertainties. If, in a particular instance, the user does not know whether the Hessian parameter uncertainties are good enough, it suffices to activate the bootstrapping mode and compare the covariance matrices of the two approaches.

5.6 2D distribution

Finally, this section illustrates the MML fitting of a multidimensional DF, theoretically discussed in Section 2.4. We limit the example to the 2D mass–angular momentum distribution of galaxies: each galaxy has two observables, its mass

M and specific angular momentum j . As in all previous cases, we won't further specify the type of matter to which these quantities apply. The two observables are summarized in the vector $\mathbf{x} = (x_1, x_2)$ with components $x_1 = \log_{10} M/M_{\odot}$ and $x_2 = \log_{10} j/[\text{kpc km s}^{-1}]$. We assume that the population is described by the 2D generative DF, first introduced by Choloniewski (1985),

$$\phi(\mathbf{x}|\boldsymbol{\theta})_{\text{Mj}} = \frac{\ln(10)\phi_*\mu^{\alpha+1}}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{[x_2 - a(x_1 - b)]^2}{2\tau^2} - \mu\right), \quad (45)$$

which is a Schechter function for the mass distribution, combined with a power-law M – j relation of slope a , zero-point b and Gaussian scatter τ . The ‘true’ parameters are assumed to be $\boldsymbol{\theta}_{\text{true}} = (\log_{10} \phi_*, \log_{10} M_*, \alpha, a, b, \tau) = (-2, 11, -1.3, \frac{2}{3}, 7, 0.3)$, adopting the same three Schechter parameters as before (equation 36) and the rounded values of the observed M – j distribution of baryons (stars and cold gas) in disc galaxies (Obreschkow & Glazebrook 2014). Isocontours of this DF are shown in Fig. 18.

To draw galaxy samples from this M – j distribution, we adopt an effective volume that depends both on M and j : $V(\mathbf{x})$ scales with M as in a sensitivity-limited survey with constant mass-to-light ratio (equation 37) and it depends on j following an error-function that is roughly constant for $x_2 > 3$, but decreases rapidly for smaller angular momenta, mimicking a natural decrease in low- j detections due to the difficulty of measuring the sizes of these small galaxies. Explicitly,

$$V(\mathbf{x}) \propto 10^{1.5x_1} [\text{erf}(x_2 - 3) + 1]. \quad (46)$$

Three isocontours of $V(\mathbf{x})$ are shown in Fig. 18.

A random mock sample of $N = 200$ galaxies is drawn from $\phi(\mathbf{x}|\boldsymbol{\theta}_{\text{true}})_{\text{Mj}}V(\mathbf{x})$ and perturbed by random errors drawn from a fixed covariant Gaussian distribution (shown as grey ellipses in Fig. 18). The resulting randomized data are shown as black points in Fig. 18. Fitting the six-parameter DF of equation (45) to these mock data results in the model shown in blue. The ‘source count model of the fit’, also shown in Fig. 18 represents the distribution of galaxies expected if no observing errors were made.

To check the statistical accuracy of the fitting solution, we generate and fit 10^4 independent mock samples. The distributions of the best-fitting parameters are shown in blue in the bottom panel of Fig. 18. For comparison, we also show the input parameters and expected Gaussian uncertainties from the average Hessian matrix of the modified likelihood functions (black). As in the case of fitting an MF (e.g. Fig. 6), the expectation of the fit (big blue dots) is statistically consistent with the input parameters (crosses). The expected covariances (blue ellipses) are approximately consistent with the Hessian prediction (black ellipses) – the slight deviation being due to the slight inaccuracy of the Hessian approach, explained in Section 2.5 and corrected in Section 5.5.

Robotham & Obreschkow (2015) derived a general method for fitting D -dimensional data ($D \geq 2$) with a $(D - 1)$ -dimensional linear model, i.e. a straight line if $D = 2$. This method, implemented in the `hyper.fit` package for R , accounts for heteroscedastic errors that are correlated between the different dimensions. Thus, `hyper.fit` normally outperforms standard regression or bisector techniques, frequently used in astronomy. However, nor `hyper.fit` or these other techniques account for the fact that the data were sampled via a known or unknown selection function (for extensions, see Pihajoki 2017). The MML approach overcomes this limitation: the example of Fig. 18 can be regarded as fitting a linear model, $x_2 = a(x_1 - b)$, with intrinsic scatter τ to the M – j relation, while accounting for the *known* effective volume and *unknown* MF. Applying `hyper.fit` to the same data results in the red parameters in Fig. 18. It appears that ignoring the non-uniform distribution

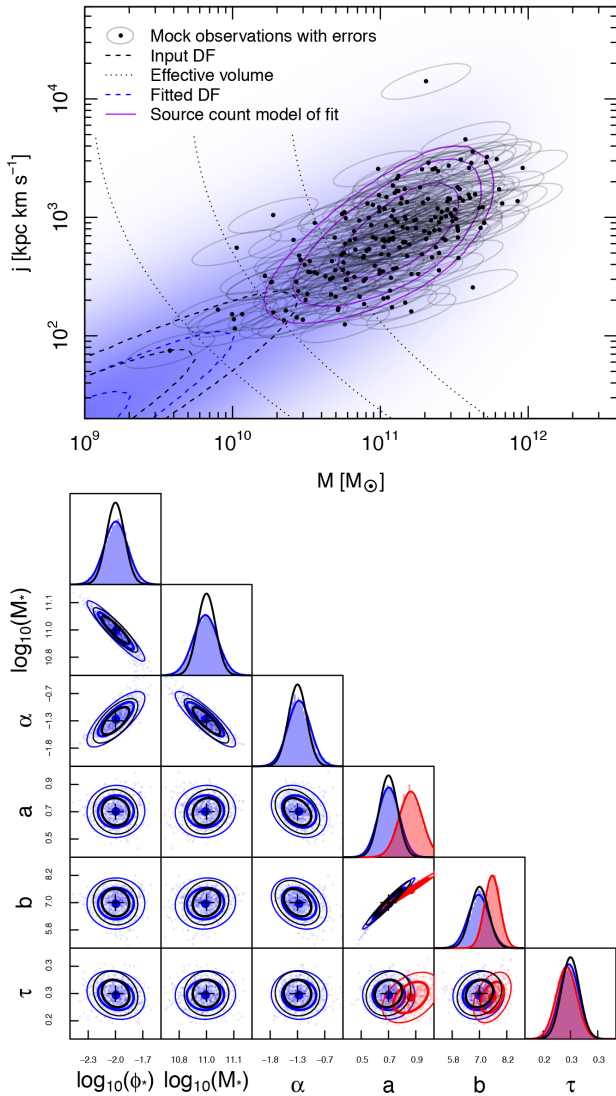


Figure 18. Example of fitting a 2D mass–angular momentum distribution function to a mock survey. The top panel shows the data of a single random survey, drawn from the 2D DF in equation (45) using the effective volume of equation (46), both represented by isocontours spaced by factors of 2. The blue shading and dashed blue isocontours represent the best MML fit recovered from this particular mock survey. The MML fit is repeated with 10^4 random mock samples, resulting in the distribution of best-fitting parameters (blue) shown in the bottom panels. The mean fits (big blue dot) are in excellent agreement with the input model (black crosses), and the parameter distributions (blue ellipses) roughly match the Hessian predictions (black ellipses). The `hyper.fit` solution is shown in red. Details are given in Section 5.6.

of the data leads to overestimating the slope a , which also affects the offset b . Hence, accounting for the MF and effective volume is important when fitting the M – j relation in this example. In conclusion, `dftools` can be used as a generalization of `hyper.fit` to account for non-uniform data.

6 CONCLUSION

The purport of this work reaches beyond astrophysics. The central aim was to develop, implement and test a general method to determine the most likely P -dimensional ($P \geq 1$) parameter θ of a distribution function model $\phi(\mathbf{x}|\theta)$, generating a sample of N objects

with D -dimensional ($D \geq 1$) properties $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. This inference problem is subject to two simultaneous complications: First, the data (i.e. the values of \mathbf{x}_i) can have arbitrary, heteroscedastic measurement uncertainties, known only up to the prior probability distributions $\rho_i(\mathbf{x})$, also called ‘belief functions’ in statistics (Denooux 2013). Secondly, the data sample can be biased in that the probability $V(\mathbf{x})$ of detecting an object of the underlying population depends on its true value \mathbf{x} . This problem is very generic, because data uncertainties and sample biases naturally appear in many applications (e.g. Aggarwal & Yu 2009). In fact, they are almost inevitable whenever the data are gathered using subsamples of larger sets, imperfect sensors or mathematical approximations, such as extrapolation.

We found that the solution to this challenging inference problem is provided by the implicit equation (11), which relies on our modified likelihood function $\mathcal{L}(\theta, \hat{\theta})$. The solution of equation (11) was shown to be unique and identical to the maximum of the ‘true’ likelihood function, i.e. the likelihood of the full Bayesian hierarchical model that treats the uncertain measurements as ND additional model parameters. However, equation (11) can be solved orders of magnitude faster than this hierarchical model using the iterative fit-and-debias algorithm of Section 3.1. Its implementation in the *R*-package `dftools` was tested thoroughly, demonstrating a quick convergence towards the correct solution. Gaussian uncertainties and covariances of the model parameters can be estimated using the Hessian matrix of the modified likelihood function (Section 2.5). However, this approach fails if the uncertainties of the data are larger than their range. In this case, parameter covariances either require evaluating the full Hessian (equation A7) or, more conveniently, bootstrapping the data (Section 5.5).

In astrophysics, the most prominent application of the MML method is the fitting of space DFs, quantifying the number of astrophysical objects per unit cosmic volume as a function of some intrinsic property \mathbf{x} . The detection probability $V(\mathbf{x})$ can then be interpreted as the effective volume (hence the symbol ν) in which the objects can be detected.

By far the most common space DFs are MFs (or LFs), which therefore dominate the terminology of this article. MFs are 1D generative DFs ($D = 1$) with scalar observables $x = \log_{10}(M/M_{\odot})$. The natural steepness of galaxy MFs makes their recovery prone to Eddington bias – an effect that many modern galaxy surveys tend to neglect. However, various mock examples (Sections 4.1 and 4.2) demonstrate that Eddington bias is very significant compared to the otherwise shot noise and LSS limited fitting uncertainties. The same examples prove that the MML method robustly removes Eddington bias, given a model of the observational uncertainties.

With the fast development of galaxy redshift surveys with imaging capabilities or even integral field spectroscopy modes, analyses of higher dimensional DFs ($D \geq 2$) are on the rise. Prominent examples include the 2D mass–size (Lange et al. 2015), mass–angular momentum (Romanowsky & Fall 2012) and spin–ellipticity (Emsellem et al. 2011) distributions, as well as the 3D mass–size–velocity (Koda, Sofue & Wada 2000) and mass–spin–morphology (Obreschkow & Glazebrook 2014) distributions. The implementation of the MML method in `dftools` accurately handles such higher dimensional DFs as illustrated in one example (Section 5.6). In particular, the method can also be used to fit linear models of any dimension, similarly to the `hyper.fit` method (Robotham & Obreschkow 2015), but accounting for arbitrary selection functions.

A specific problem with MFs and other space DFs of astrophysical objects is that the detectability of these objects not only depends on their intrinsic properties \mathbf{x} , but generally also on the distance r

to the observer, sometimes even on the 3D position \mathbf{r} . This addition to the problem is further complicated by the inevitable presence of unknown cosmic LSS. Often, the detectability also depends on hidden properties (e.g. the galaxies' inclination or colour) that are not part of the fitted observables \mathbf{x} . All these effects can be accounted for in the definition of the effective volume (Section 2.2), which, in the case of LSS, depends on the best-fitting model parameters $\hat{\theta}$ (Section 2.3).

Let us now turn to some limitations of the current presentation of the MML method, which might require further investigation. First, there are a number of secondary uncertainties, not yet included in the MML method or any other fitting algorithm to our knowledge:

(i) *Uncertain selection functions*: In principle, the formalism could be extended to include such uncertainties, probably at the cost of slowing down the algorithm considerably. At the moment, we recommend to adopt a bootstrapping technique, i.e. a wrapper around `dftools` that resamples the selection function and refits an MF at each iteration.

(ii) *Uncertainties in the measurement uncertainties*: In practice, the functions $\rho_i(x)$ are themselves subject to both systematic and random errors. The former are hard to address, but the effect of random errors can again be estimated by refitting MFs to different choices of $\{\rho_i(x)\}$.

(iii) *Distance uncertainties for LSS*: We have not included distance uncertainties in estimating and removing LSS bias. (Of course distance uncertainties can be included in the uncertainties of \mathbf{x} , such as in mass uncertainties.) In the case of spectroscopic redshift measurements, these uncertainties are negligible relative to the typical scales of density fluctuations that dominate LSS. Only photometric redshift measurements might require accounting for distance uncertainties in the removal of LSS bias.

Another important aspect not considered in this work is cosmic evolution, which makes DFs, such as galaxy MFs depend on redshift z or comoving distance r . Of course, it is possible to subdivide a galaxy sample into different redshift bins and fit an MF individually to each of them to evidence trends in the parameter evolution. However, sometimes it is (arguably) desirable to fit just one or two additional MF parameters of an analytic evolution model. ML methods dealing with this case have been presented (Lin et al. 1999; Loveday et al. 2012), but they do not simultaneously account for observational errors (Eddington bias). Fitting evolution models in the context of the MML method is possible, but it is not as straightforward as including redshift or distance as an additional observable in \mathbf{x} . Fitting evolution models requires extending the formalism to redshift-dependent DFs $\phi(\mathbf{x}|z, \theta)$, which will lead to a redshift-integral.

The natural next step is to apply the MML method to real galaxy data from existing and future surveys. This brings all the benefits of the standard ML method, while fully accounting for major empirical unknowns, especially mass errors and LSS. Therefore, MML fits allow a robust comparison of different data sets and, within the Laplace approximation, a clean identification of the best DF model. More generally, we hope that the MML estimator and `dftools` will spread to other fields within and outside astrophysics, where measurement uncertainties and sample biases play a significant role in statistical inference.

ACKNOWLEDGEMENTS

Parts of this research were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAAS-

TRO), through project number CE110001020. We thank the anonymous referee for their in-depth feedback.

REFERENCES

- Aggarwal C. C., Yu P. S., 2009, *IEEE Trans. Knowl. Data Eng.*, 21, 609
- Allenby G. M., Rossi P. E., McCulloch R. E., 2005, in *Hierarchical Bayes Models: A Practitioners Guide*
- Amiaux J. et al., 2012, in Clampin M. C., Fazio G. G., MacEwen H. A., Oschmann J. M., eds, *Proc. SPIE Conf. Ser. Vol. 4482, Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*. SPIE, Bellingham, p. 84420Z
- Babu G. J., Singh K., 1983, *Ann. Statist.*, 11, 999
- Baldry I. K. et al., 2012, *MNRAS*, 421, 621
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJ*, 585, L117
- Cash W., 1979, *ApJ*, 228, 939
- Choloniewski J., 1985, *MNRAS*, 214, 197
- Cole S., 2011, *MNRAS*, 416, 739
- Colless M. et al., 2001, *MNRAS*, 328, 1039
- Cordeiro G. M., Klein R., 1994, *Stat. Probab. Lett.*, 19, 169
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Daniels H. E., 1954, *Ann. Math. Stat.*, 25, 631
- Davidzon I. et al., 2017, *A&A*, 605, A70
- Davis M., Huchra J., 1982, *ApJ*, 254, 437
- Dawson K. S. et al., 2013, *AJ*, 145, 10
- Denoeux T., 2013, *IEEE Trans. Knowl. Data Eng.*, 25, 119
- Drinkwater M. J. et al., 2010, *MNRAS*, 401, 1429
- Efron B., Stein C., 1981, *Ann. Statist.*, 9, 586
- Efron B., Tibshirani R., 1993, *An introduction to the bootstrap*. Monographs on statistics and applied probabilities. Chapman & Hall/CRC, CRC Press LLC, Boca Raton, Florida
- Efstathiou G., Ellis R. S., Peterson B. A., 1988, *MNRAS*, 232, 431
- Emmelmeier E. et al., 2011, *MNRAS*, 414, 888
- Evrard A. E., Arnault P., Huterer D., Farahi A., 2014, *MNRAS*, 441, 3562
- Grazian A. et al., 2015, *A&A*, 575, A96
- Hardy M., 2002, eprint ([arXiv:e-prints](https://arxiv.org/abs/200204011))
- Kass E., Raftery E., 1995, *J. Am. Stat. Assoc.*, 90, 773
- Kendall M. G., Stuart A., 1979, *The Advanced Theory of Statistics*. Vol. 2: Inference and Relationship, 4th edn. Macmillan, New York
- Kirshner R. P., Oemler A., Jr, Schechter P. L., 1979, *AJ*, 84, 951
- Koda J., Sofue Y., Wada K., 2000, *ApJ*, 531, L17
- Lange R. et al., 2015, *MNRAS*, 447, 2603
- Lin H., Yee H. K. C., Carlberg R. G., Morris S. L., Sawicki M., Patton D. R., Wirth G., Shepherd C. W., 1999, *ApJ*, 518, 533
- Liske J. et al., 2015, *MNRAS*, 452, 2087
- Loveday J. et al., 2012, *MNRAS*, 420, 1239
- Loveday J. et al., 2015, *MNRAS*, 451, 1540
- Lynden-Bell D., 1971, *MNRAS*, 155, 95
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
- Mortonson M. J., Hu W., Huterer D., 2011, *Phys. Rev. D*, 83, 023015
- Murray S. G., Power C., Robotham A. S. G., 2013, *MNRAS*, 434, L61
- Murray S., Robotham A., Power C., 2017, *MNRAS*, submitted
- Nelder J. A., Mead R., 1965, *Comput. J.*, 7, 308
- Obreschkow D., Glazebrook K., 2014, *ApJ*, 784, 26
- Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, *ApJ*, 759, 138
- Pihajoki P., 2017, *MNRAS*, 472, 3407
- Robotham A. S. G., Obreschkow D., 2015, *Publ. Astron. Soc. Aust.*, 32
- Romanowsky A. J., Fall S. M., 2012, *ApJS*, 203, 17
- Sandage A., Tammann G. A., Yahil A., 1979, *ApJ*, 232, 352
- Schechter P., 1976, *ApJ*, 203, 297
- Schmidt M., 1968, *ApJ*, 151, 393
- Teerikorpi P., 2004, *A&A*, 424, 73
- Turner E. L., 1979, *ApJ*, 231, 645
- Weigel A. K., Schawinski K., Bruderer C., 2016, *MNRAS*, 459, 2150
- Westmeier T. et al., 2017, *MNRAS*, 472, 4832

Wright A. H. et al., 2017, MNRAS, 470, 283
 York D. G. et al., 2000, AJ, 120, 1579
 Zwaan M. A. et al., 2003, AJ, 125, 2842
 Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

APPENDIX A: PROOF OF MML SOLUTION

The MML method presented in this article is able to produce rapid parameter fits due to a subtle approximation, namely that the posterior PDF $\tilde{\rho}_i(x|x_i, \theta)$ – the probability of a given observation x_i to have true properties x – is fully specified by its prior PDF, $\rho(x|x_i)$, and the generative DF model *at the ML solution*. With this approximation, one is able to define the ‘posterior’ PDF of each observation as (equation 6)

$$\tilde{\rho}_i(x|\theta) = \frac{\rho_i(x|x_i)\phi(x|\theta)V(x)}{\int \rho_i(x|x_i)\phi(x|\theta)V(x)dx}. \quad (\text{A1})$$

Effectively, this equation makes the point that if we just knew the true generative distribution (via its parameters θ), then we would know the true probability of each observation having underlying properties of x . Of course, we don’t know the true generative distribution a priori. This suggests splitting the problem into two layers of iteration. First, in a meta-iteration, make a choice of parameters, $\hat{\theta}_0$, calculate the expensive factor $\tilde{\rho}_i$, and then solve (by downhill iteration) for the best set of parameters, θ'_0 using the likelihood equation (10). This inner optimization is cheap, since the expensive calculation is only performed once. Following this, set the ‘guess’, $\hat{\theta}_1$, to θ'_0 , and so on until $\theta'_i = \hat{\theta}_i$, at which point we postulate that $\theta \equiv \hat{\theta}_i$.

It is clear that this is an approximation. The true posterior PDF for each datum must involve marginalizing over the posterior of θ , that is, the true posterior probability of any observation having true value x is a weighted sum over the distribution-corrected prior for all possible values of the parameters, not just their most likely value. It may be surprising then that this approximation ‘works’.

Here, we present the true likelihood, based on full marginalization over the posterior, and show that it analytically produces the same MLE as the modified likelihood. For completeness, we also show that it does *not* produce the same Hessian (and therefore covariance) as the modified likelihood.

Let the underlying data, \hat{x}_i , be drawn from $\lambda(x|\theta) = \phi(x|\theta)V(x)$, and let the observed data, x_i , include some uncertainty in its measurement, $x_i \sim \rho'_i(x|\hat{x}_i)$ (note that $\rho'_i \neq \rho_i$ in general, except for symmetric distributions, but that one can be computed from the other, cf. Appendix D). A Bayesian hierarchical model will require an estimation of both the model parameters θ and the underlying value \hat{x}_i . In particular, for a Poisson-distributed set of infinitesimal bins in x , the likelihood is the product of the expectation, λ , at each estimate \hat{x} , normalized by the expectation of the total number of observations, and then multiplied by the prior probability of measuring the set of x_i given the estimates \hat{x}_i . Succinctly, the log-likelihood is

$$\ln \mathcal{L}(\hat{x}, \theta) = \sum_i \ln \lambda(\hat{x}_i|\theta) + \ln \rho_i(x_i|\hat{x}_i) - \int \lambda(x|\theta)dx, \quad (\text{A2})$$

where $\hat{x} = \{\hat{x}_i\}$.

We are not typically interested in the values of \hat{x}_i , but rather in $\hat{\theta}$, so we may marginalize over the former by integrating over the likelihood (not the log-likelihood!). Doing so involves a highly multidimensional integral (as many dimensions as there are data).

However, so long as the uncertainties on each are independent, the integral is completely factorizable, and the final likelihood is

$$\ln \mathcal{L}(\theta) = - \int \lambda(x|\theta)dx + \sum_i \ln \left(\int \lambda(x|\theta)\rho_i(x)dx \right). \quad (\text{A3})$$

We note that while this equation is very similar to equation (10), the logarithm is outside the integral in this case, and there is no $\tilde{\rho}$ factor involved.

The ML solutions are defined as $\hat{\theta} = \theta$ at which the Jacobian of $\ln \mathcal{L}$ is zero (note that there may in general be more than one solution). We find that the Jacobian of equation (A3) is

$$\mathbf{J}_{\text{true}}(\theta) = \frac{\partial \ln \mathcal{L}}{\partial \theta} = - \int \lambda_{\theta}(x)dx + \sum_i \frac{\int \lambda_{\theta}(x)\rho_i(x)dx}{\int \lambda(x|\theta)\rho_i(x)dx}, \quad (\text{A4})$$

where a subscript denotes partial differentiation. Suppose that an MLE solution exists and further that it is unique (multiple peaks will be a challenge for either likelihood), and let such a solution be denoted $\hat{\theta}$, i.e. $\mathbf{J}_{\text{true}}(\hat{\theta}) = 0$ and $\mathbf{H}_{\text{true}}(\hat{\theta})$ negative definite (with \mathbf{H}_{true} the Hessian). To show that the MML method will yield the same solution, our task is then threefold: (i) show that the MML likelihood admits an identical MLE, (ii) show that this solution is unique, and (iii) show that the iterative procedure always converges towards this solution.

In an iterative solution with n meta-iterations, let $0 \leq j < n$ be the current meta-iteration, and $\hat{\theta}_j$ be the current ‘meta-estimate’ of θ . The Jacobian, with respect to the inner-iteration parameters, θ , of equation (10) is

$$\begin{aligned} \mathbf{J}_{\text{MML}}(\theta, \hat{\theta}_j) &= - \int \lambda_{\theta}(x)dx + \sum_i \int dx \tilde{\rho}_i(x|\hat{\theta}_j) \frac{\lambda_{\theta}(x)}{\lambda(x|\theta)} \\ &= - \int \lambda_{\theta}(x)dx + \sum_i \frac{\int dx \rho_i(x)\lambda_{\theta}(x)\lambda(x|\hat{\theta}_j)/\lambda(x|\theta)}{\int dx \rho_i(x)\lambda(x|\hat{\theta}_j)}. \end{aligned} \quad (\text{A5})$$

The existence of the solution is shown easily by substituting $\theta = \hat{\theta}_j = \hat{\theta}$. In this case, the λ in the numerator and denominator of equation (A5) cancel one another, to leave precisely $\mathbf{J}_{\text{true}}(\hat{\theta})$ which by construction is zero.

The uniqueness of the solution is easily shown by contradiction. Suppose an alternative solution, $\tilde{\theta}$, was found. To be the final solution to the fit-and-debias algorithm, this requires that the solution within the iteration, $\theta = \theta'$, is the same as the ‘meta-estimate’ $\hat{\theta}_j$, i.e. we suppose that $\theta' = \hat{\theta}_j = \tilde{\theta}$, and that $\mathbf{J}_{\text{MML}}(\tilde{\theta}, \tilde{\theta}) = 0$. However, whenever $\theta = \hat{\theta}_j$, we have $\mathbf{J}_{\text{MML}}(\theta', \theta') \equiv \mathbf{J}_{\text{true}}(\theta')$. Thus we imply that $\mathbf{J}_{\text{true}}(\tilde{\theta}) = 0$, which is in contradiction to the original statement that $\hat{\theta}$ is a unique solution to the true likelihood. Conversely, if $\tilde{\theta}$ were *not* a unique solution to the true likelihood, but rather multiple peaks existed, then it follows that $\tilde{\theta}$ must be *one of them*, which shows that the *set* of solutions to the MML likelihood are the same as the true solution (but does not guarantee that the same solution will be reached given the same initial estimate).

Our final task is to show that the iterative scheme always converges towards the solution. One could imagine that even though a true solution, $\hat{\theta}$, exists, the fit-and-debias algorithm nevertheless diverges or eternally oscillates around this solution while the true likelihood converges. Stated mathematically, our task is as follows. Let $\hat{\theta}_j$ and θ'_j retain their meanings from previous arguments, so that $\mathbf{J}_{\text{MML}}(\hat{\theta}_j, \theta'_j) = 0$ (but note that we may not have converged yet, so that we may have $j < n$ and $\hat{\theta}_j \neq \theta'_j$). Then, we require that $|\hat{\theta}_j - \theta'_j| \leq |\hat{\theta} - \hat{\theta}_j| \forall j$. That is, we must show that every meta-iteration returns an estimate which is closer to the true solution. We can simplify this picture by noting that the next meta-iteration

($j + 1$) will begin at $\theta = \theta'_j$, and use the meta-estimate $\hat{\theta}_{j+1} = \theta'_j$. For the solution of this iteration to be closer to $\hat{\theta}$ only requires that the Jacobian at this point be positive if $\tilde{\theta} < \hat{\theta}$ and vice versa (i.e. gradient must roll towards the true solution from this point). Furthermore, it suffices to show that the Jacobian at such a starting point has these properties for just one starting point to the left and right of $\hat{\theta}$, as the uniqueness proved above ensures that the Jacobian cannot cross through zero except at $\hat{\theta}$. Taking these offset points to the infinitesimal limit, we require only the gradient of the Jacobian at the solution itself, i.e. we need only show that

$$\frac{\partial}{\partial \theta} \mathbf{J}_{\text{MML}}(\theta, \theta)|_{\hat{\theta}} < 0. \quad (\text{A6})$$

To this end, we simply note that $\mathbf{J}_{\text{MML}}(\theta, \theta) = \mathbf{J}_{\text{true}}(\theta)$ so that the derivative above is merely the true Hessian. This Hessian, by construction, is negative definite at $\hat{\theta}$, so we have proved the MML method.

Despite the above proof, the Hessians for both the true likelihood and the MML are not in general the same. For completeness, we here give the Hessian at the MLE in each case:

$$\mathbf{H}_{\text{true}}^{\text{lm}} = Q(\hat{\theta}) - \sum_i \frac{\int \lambda_{\theta_i}(x|\hat{\theta}) \rho_i(x) dx \int \lambda_{\theta_m}(x|\hat{\theta}) \rho_i(x) dx}{\left[\int \lambda(x|\hat{\theta}) \rho_i(x) dx \right]^2}, \quad (\text{A7})$$

$$\mathbf{H}_{\text{MML}}^{\text{lm}} = Q(\hat{\theta}) - \sum_i \frac{\int dx \rho_i(x) \lambda_{\theta_i}(x|\hat{\theta}) \lambda_{\theta_m}(x|\hat{\theta}) / \lambda(x|\hat{\theta})}{\int dx \rho_i(x) \lambda(x|\hat{\theta})}, \quad (\text{A8})$$

where

$$Q(\hat{\theta}) = - \int \lambda_{\theta_i \theta_m}(x|\hat{\theta}) + \sum_i \frac{\int \lambda_{\theta_i \theta_m}(x|\hat{\theta}) \rho_i(x) dx}{\int \lambda(x|\hat{\theta}) \rho_i(x) dx}. \quad (\text{A9})$$

APPENDIX B: EFFECTIVE VOLUME FOR LSS

This appendix derives the equation to estimate $g(r)$ directly from the redshift-distribution of the galaxy sample. In most MF work, the galaxy distances r_i are inferred from spectroscopic redshifts. For the purpose of correcting for LSS bias, we assume that these distances are exact, thus neglecting redshift-measurement errors and redshift-space distortions due to peculiar velocities. In principle, both uncertainties can be incorporated in the MML formalism, but in the majority of cases this will likely complicate the formalism without measurable improvement. Note that this assumption of exact distances only applies to the LSS bias correction. Of course, distance uncertainties can be accounted for in the mass uncertainties $\phi_i(x)$, which are a central part of the MML formalism.

To the benefit of intuition, let us temporarily introduce distance bins k of mean redshifts r_k and width Δr . Let n_k be the observed number of galaxies in each bin k . The relative density g_k can then be calculated as the ratio of n_k and the expected number of galaxies in bin k in the absence of LSS,

$$g_k = \frac{n_k}{\int_{r_k - \Delta r/2}^{r_k + \Delta r/2} \phi(x|\hat{\theta}) V'(r) f(x, r) dr dx}. \quad (\text{B1})$$

As in the removal of Eddington bias (equation 6), we here used the most likely model parameters $\hat{\theta}$ to estimate the expected counts. We now let Δr become infinitesimal while introducing the densities $g(r) = g_k / \Delta r$ and $n(r) = n_k / \Delta r$. In the isotropic case ($d^3 r = V'(r) dr$),

equation (19) then becomes

$$V_{\text{LSS}}(x) = \int \frac{f(x, r) n(r)}{\int \phi(x|\hat{\theta}) f(x, r) dx} dr. \quad (\text{B2})$$

Note that the terms $V(r)$ have cancelled out. Since we are assuming precisely known distances r_i , the source density simplifies to $n(r) = \sum_i \delta_D(r_i - r)$, where δ_D is the Dirac delta function. Hence, the integral over r reduces to a sum over the galaxies i , i.e. to equation (20).

APPENDIX C: MEAN EFFECTIVE VOLUMES

This section gives a numerical example for the statement that several types of galaxies of identical mass, but different effective volumes can be combined in the MF by constructing the harmonic mean of their effective volumes.

Let us consider a simplified case (Fig. C1) of two types of galaxies, blue and red ones, all of identical log-mass x . Our hypothetical universe contains exactly one blue and one red galaxy per each unit volume. Hence, the MF at x has a value of $\phi(x) = 2$. We survey a volume of 8 units, which contains 8 blue and 8 red galaxies. The blue galaxies can be detected across the whole survey volume, i.e. their effective volume is $V_{\text{blue}} = 8$ and we count $n_{\text{blue}} = 8$ such galaxies. The red galaxies are much harder to detect, such they can only be detected in the most nearby volume unit, where there is just one such galaxy, i.e. $V_{\text{red}} = 1$ and $n_{\text{red}} = 1$.

In total, there are $n = n_{\text{blue}} + n_{\text{red}} = 9$ galaxies. Since these galaxies must collectively have an MF value of $\phi(x) \equiv n/V = 2$, their combined effective volume V must be equal to 4.5. This value is indeed identical to the *harmonic* mean of the effective volumes of all detected galaxies,

$$V = \left[\frac{1}{n} (n_{\text{blue}} V_{\text{blue}}^{-1} + n_{\text{red}} V_{\text{red}}^{-1}) \right]^{-1} = 4.5. \quad (\text{C1})$$

In other words, we can evaluate the MF directly by dividing the number of detections by the harmonic mean of their volumes. In doing so, the different detectability of different galaxy types is accurately accounted for. This result readily generalizes to any number of galaxies and types of galaxies.

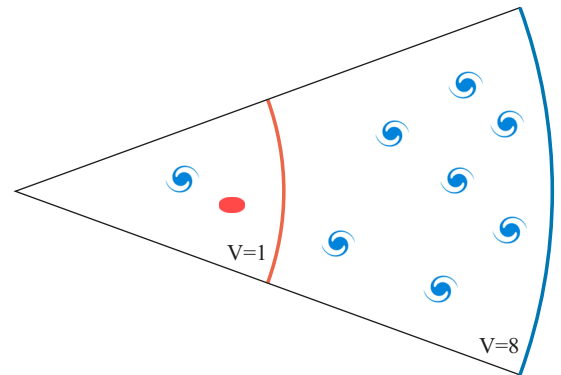


Figure C1. Sketch of surveying a simplified universe with red and blue galaxies. Both types are equally frequent, but red galaxies are only detectable in a smaller effective volume. In this case, the MF is recovered by taking the harmonic mean of the effective volumes of all detections, as explained in Appendix C.

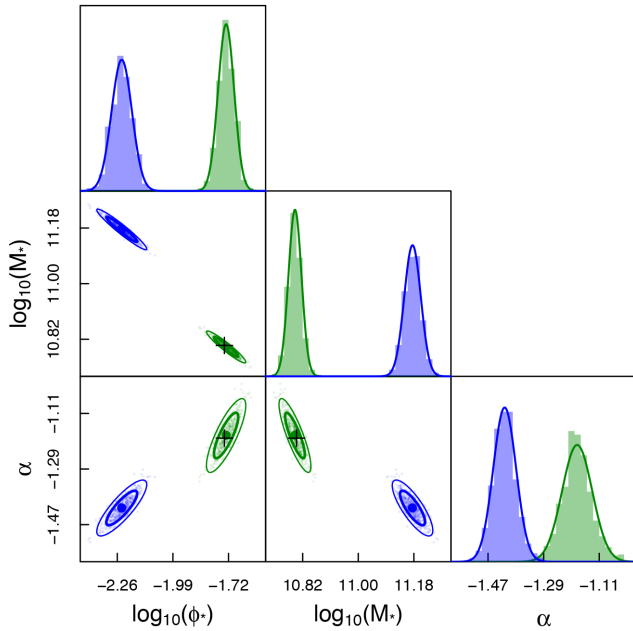


Figure D1. Illustration of recovering the Schechter function parameters from mock data, drawn from a Schechter function, then scattered and then selected with a weight $V(x_{\text{obs}})$. The true model parameters are shown as black crosses. The big blue dots/ellipses show the average and Gaussian 68 per cent (thick) and 95 per cent (thin) contours of 10^3 random mock samples, fitted by incorrectly assuming that the data was first selected and then scattered. The green dots/ellipses show the fits that account for the correct ordering via the substitutions of equations (D1) and (D2).

APPENDIX D: SCATTER, THEN SELECT

This work focused on the case, where data are drawn from a population with weights $V(x)$ and then scattered by measurement uncertainties. There are other cases where this order is (partially) reversed. For instance, astronomical observations at the sensitivity limit of the telescope might only detect a few photons per object, subject to Poisson statistics. If we impose a fixed number of photons as the detection limit, the Poisson scattering is followed by the selection. In this particular situation, it is possible to approximately include the Poisson statistics in the completeness C (Section 2.2), when calculating the effective volume of each galaxy.

There are none the less more complex cases. For instance, the photon counting subject to Poisson noise might be followed by an apparent magnitude cut, conversion to absolute magnitudes using uncertain distances, additional cut by absolute magnitude, conversion to masses using uncertain mass-to-light ratios, etc. Such layered uncertainties are typically handled by a hierarchical scheme, where the model MF is convolved with an uncertainty model and passed through a selection function in a repeated sequence; e.g. first convolution with Poisson noise, then apparent magnitude cut, convolution with distance errors, absolute magnitude cut and convolution with mass-to-light error. Currently, only the last uncertainty can then be accelerated by the MML scheme, using the backward correction of equation (6) instead of a forward convolution.

Let us provide one example of fitting for an MF, first scattered and then subjected to a single selection. Explicitly the data x are drawn from $\phi(x|\theta)$, then scattered ($x \rightarrow x_{\text{obs}}$) and then selected with weight $V(x_{\text{obs}})$. It suffices to note that scattering first and then selecting is equivalent to drawing data from a modified MF, which has been smoothed (convoluted) by the observing error, without

adding further noise. Formally, this corresponds to the following substitution in the formalism of Section 2.1,

$$\phi(x|\theta) \rightarrow \phi'(x|\theta) = \int \phi(s|\theta) \varrho(x|s) ds, \quad (\text{D1})$$

$$\rho_i(x) \rightarrow \rho'_i(x) = \delta_D(x - x_i), \quad (\text{D2})$$

where $\varrho(x|s)$ is the probability of observing a value x given a true value s , solely due to the uncertainty of the measurement process.

As an illustration, we adopt the advanced example of Section 5.4, which exhibits measurement uncertainties that depend on the true value of the log-mass x . The mock data is generated exactly as described in Section 5.4, except that the scattering ($x \rightarrow x_{\text{obs}}$) is performed before the selection with weight $V(x_{\text{obs}})$. We use 10^3 random mock samples. As shown in Fig. D1, their fitted Schechter function parameters are consistent with the true input values, if and only if the substitutions of equations (D1) and (D2) are included.

APPENDIX E: USING REAL OBSERVATIONS

This paper intentionally used mock data instead of real observations to allow for a comparison of the fits against the *known* true solutions. This section illustrates how, explicitly, `dftools` is applied to real data. We use the data recently published in Westmeier et al. (2017), where an early version of `dftools` has already been applied to estimate the MF of neutral atomic hydrogen (H I) in the nearby Sculptor filament, a loosely bound elongated group of galaxies.

In *R*, load the `dftools` library and the data of Westmeier et al. (2017):

```
library(dftools)

data = read.table

('`http://quantumholism.com/
m/dftools/westmeier2017.txt`', header=TRUE)
```

There are 31 galaxies in this sample, hence the dataframe `data` has 31 rows and three columns specifying the H I masses (in M_{\odot}), the mass uncertainties (standard deviations) and the effective volumes (in Mpc^3). We recast these data into separate vectors containing the log-masses x_i ($i = 1, \dots, 31$), log-normal uncertainties (using linear error propagation) and effective volumes of each galaxy,

```
x = log10(data$MHI)

x.err = data$errMHI/data$MHI/log(10)

veff.values = data$Vmax
```

To fit a Schechter function (the default MF of `dftools`) and display the best-fitting parameters, it suffices to call

```
survey = dffit(x,veff.values,x.err)

dfwrite(survey)
```

which outputs the text

```
dN/(dVdx) = log(10)*10^p[1]*mu^(p[3]+1)
```

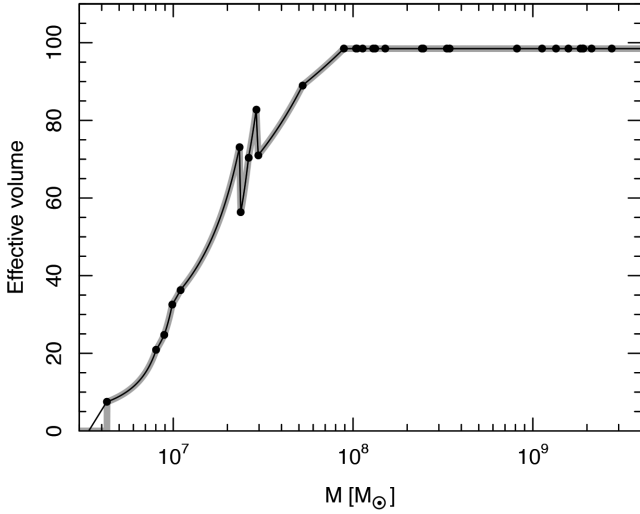


Figure E1. Effective volume function of the Sculptor H I survey. The individual values provided with the sample are shown as black dots. The default fit derived by `dftools` (grey thick line) is slightly corrected manually (black thin line) based on additional knowledge of the survey, not contained in the data itself.

```
*exp(-mu), where mu=10^(x-p[2])
```

```
p[1] = -1.315 (+-0.276)
```

```
p[2] = 9.541 (+-0.308)
```

```
p[3] = -1.103 (+-0.147)
```

The thick line in Fig. E1 shows the effective volume as a function of mass, recovered from linear interpolation of the values $1/V(x_i)$ (see Section 2.2.2). Outside the observed mass range, the effective volume defaults to $V(x) = 0$ for $x < \min\{x_i\}$ and $V(x) = \max\{V(x_i)\}$ for $x > \max\{x_i\}$. If a better model is available from survey-specific considerations, then this information can be exploited to improve the fit. Here, we like to replace the effective volume for $x < \min\{x_i\}$ by $V(x) = \max(0, 75 \cdot (x - 6.53))$, while keeping the rest unchanged (thin line in Fig. E1). To apply this modification to the fit, it suffices to define a list `selection`, composed of the vector `veff.values` and a new function `veff.fn`, specifying the effective volume outside the observed mass range:

```
veff.fn = function(x) {
  veff.max = max(veff.values)
  return(pmax(0, pmin(veff.max,
    (x - 6.53) * 75)))
}
selection = list(veff.values,veff.fn)
```

In addition to using the modified effective volume (dashed line in Fig. E1), we wish to determine the parameter uncertainties via bootstrapping, hence allowing for asymmetric uncertainties. To use 10^3 bootstrap iterations with a fixed seed for the random number generator, call

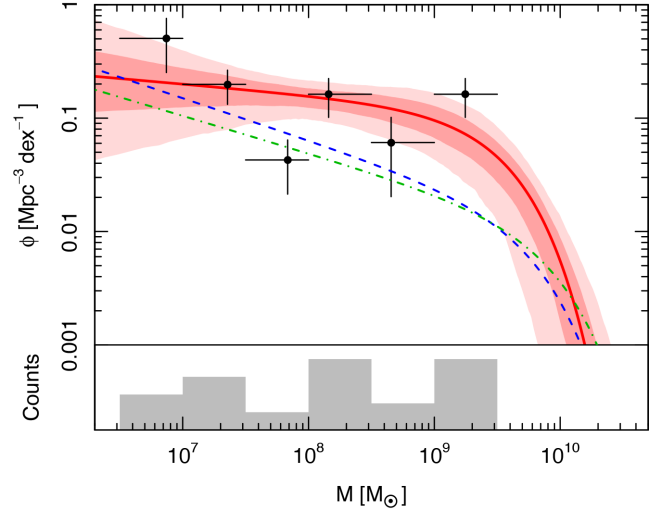


Figure E2. H I MF of the Sculptor group. The data (shown as binned black points and grey histogram) is fitted with a Schechter function, shown as red solid line with 68 per cent and 95 per cent confidence intervals as red shading. The mean H I MF of the local universe as inferred from the HIPASS and ALFALFA surveys are shown as blue dashed and green dash-dotted lines, respectively.

```
set.seed(1)
```

```
survey = dffit(x,selection,x.err,
```

```
n.bootstrap = 1e3)
```

Finally, we produce the MF plot (Fig. E2) with 68 per cent and 95 per cent confidence regions around the best fit. The chosen graphical arguments display the fitting function in red, display the observed data in black, remove posterior data, suppress the effective volume line and adjust the binning of input data.

```
mplot(survey,xlim=c(2e6,5e10),
ylim=c(1e-3,1),uncertainty.type=3,
col.fit='red',col.data.input='black',
show.posterior.data=FALSE,col.veff=NULL,
nbins=6,bin.xmin=6.5,bin.xmax=9.5)
```

To add the reference H I MFs of the HIPASS (Zwaan et al. 2005) and ALFALFA (Martin et al. 2010) surveys, use

```
x = c(survey$grid$x)
y = dfmodel(x,c(log10(6.0e-3),9.80,-1.37))
lines(10^x,y,lty=2,lwd=1.5,col='blue')
y = dfmodel(x,c(log10(4.8e-3),9.96,-1.33))
lines(10^x,y,
lty=4,lwd=1.5,col='#00bb00')
```

In Fig. E2, the Sculptor H I MF clearly differs from the mean H I MF in the local universe, as discussed by Westmeier et al. (2017). The best-fitting parameters are displayed via

```
dfwrite(survey)
```

resulting in the output

```
dN/(dVdx) = log(10)*10^p[1]*mu^(p[3]+1)
```

```
*exp(-mu), where mu=10^(x-p[2])
```

```
p[1] = -1.308 (+0.252 -0.260)
```

```
p[2] = 9.535 (+0.143 -0.197)
```

```
p[3] = -1.097 (+0.174 -0.141)
```

Note that there are marginal differences to the uncertainty ranges quoted by Westmeier et al. (2017) due to a difference in the bootstrapping technique used in the previous (parametric bootstrapping) and the current (non-parametric bootstrapping) version of `dftools`.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.