PROBABILITY CATALOGS FOR CROWDED STELLAR FIELDS

BRENDON J. BREWER\textsuperscript{1}, DANIEL FOREMAN-MACKEY\textsuperscript{2}, AND DAVID W. HOGG\textsuperscript{2}

\textsuperscript{1}Department of Statistics, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand; bj.brewer@auckland.ac.nz
\textsuperscript{2}Center for Cosmology and Particle Physics, Department of Physics, New York University, Washington Place, New York, NY 10003, USA

Received 2012 November 22; accepted 2013 April 20; published 2013 June 5

ABSTRACT

We present and implement a probabilistic (Bayesian) method for producing catalogs from images of stellar fields. The method is capable of inferring the number of sources $N$ in the image and can also handle the challenges introduced by noise, overlapping sources, and an unknown point-spread function. The luminosity function of the stars can also be inferred, even when the precise luminosity of each star is uncertain, via the use of a hierarchical Bayesian model. The computational feasibility of the method is demonstrated on two simulated images with different numbers of stars. We find that our method successfully recovers the input parameter values along with principled uncertainties even when the field is crowded. We also compare our results with those obtained from the SExtractor software. While the two approaches largely agree about the fluxes of the bright stars, the Bayesian approach provides more accurate inferences about the faint stars and the number of stars, particularly in the crowded case.

Key words: catalogs – methods: data analysis – methods: statistical – stars: luminosity function, mass function

Online-only material: color figures

1. INTRODUCTION

Traditional practice in astronomy is to take images of the sky, detect or enumerate sources visible in those images, and create catalogs. These catalogs are then used to perform fundamental astronomical measurements, for example, reconstructing the three-dimensional structure of the Galaxy or the two-point correlation function of galaxies. Indeed, the process of catalog construction is so “baked in” to our ideas about what astronomy is, that we sometimes forget that the catalog is not the fundamental data product of astronomy; catalogs are produced from imaging; their production involves many decisions and ideas that go beyond the information provided to the telescope by the incident intensity field. In addition, catalogs are not usually the final goal of any imaging project or survey. Typically, they are produced in order to facilitate the scientific study of populations of objects (e.g., the initial mass function of a population of stars) or to provide a sky-search capability to the community who might be interested in only a small subset of objects. Standard tools for generating catalogs from astronomical imaging include SExtractor (Bertin & Arnouts 1996), DAOPHOT (Stetson 1987), DOLPHOT (Dolphin 2000), and SDSS Photo (Lupton et al. 2001).

Telescopes do not make catalogs (Hogg & Lang 2011), they measure the intensity field. Viewed through the lens of probabilistic inference, the goals of astronomy are to take the information in the telescope-generated records of the intensity field and use this information to obtain quantities of astronomical interest with as little loss as possible. Insertion of a catalog-generation step in the inference pipeline between the raw imaging and the final astrophysical analyses is potentially lossy. The hard decisions of catalog making destroy information, at least in principle. Probability theory suggests that it may be less lossy to pass forward not a catalog but a probabilistic description of all the sources in the image (and that should therefore be listed in the catalog) is itself unknown. Second, if $N$ is large, then the parameter space of positions and properties (flux, size, etc.) of the objects is also large. This can cause MCMC algorithms’ difficulties—they may take a long time to converge to the target posterior distribution over the space of catalogs. Third, this problem is subject to the so-called label-switching problem that is commonly encountered in mixture modeling (e.g., Jasra et al. 2005). Given any proposed catalog, another catalog that is equally plausible is the catalog obtained by shuffling the entries of the first catalog. This leads to a posterior distribution with $N!$ identical peaks in parameter space. This can lead to difficulties with certain (otherwise very effective) MCMC algorithms such as the affine-invariant stretch move (Goodman & Weare 2010; Foreman-Mackey et al. 2013).

This article represents an attempt at implementing this ambitious goal in the specific situation where the only objects in the field are stars or other point sources.

Beyond these philosophical concerns, there are practical issues; standard methods for constructing catalogs can have difficulty in some challenging situations. For example, when multiple sources overlap partially or completely, it can be difficult to determine how many sources are present, and how much flux belongs to each source. In principle, the uncertainty about the existence and properties of the objects can be significant and should be propagated into any inferences about the stellar population. A Bayesian approach that obtains the posterior distribution over catalog space (rather than a single catalog estimate) has the potential to overcome these problems by deblending objects when it is possible, and clearly indicate the uncertainty remaining when it is not possible.

In practice, Bayesian inferences are often implemented using Markov Chain Monte Carlo (MCMC) methods (Mackay 2003) to sample from the posterior distribution. Sampling a posterior probability distribution for catalogs is a challenging numerical task for a number of reasons. First, the number $N$ of objects in the image (and that should therefore be listed in the catalog) is itself unknown. Second, if $N$ is large, then the parameter space of positions and properties (flux, size, etc.) of the objects is also large. This can cause MCMC algorithms’ difficulties—they may take a long time to converge to the target posterior distribution over the space of catalogs. Third, this problem is subject to the so-called label-switching problem that is commonly encountered in mixture modeling (e.g., Jasra et al. 2005). Given any proposed catalog, another catalog that is equally plausible is the catalog obtained by shuffling the entries of the first catalog. This leads to a posterior distribution with $N!$ identical peaks in parameter space. This can lead to difficulties with certain (otherwise very effective) MCMC algorithms such as the affine-invariant stretch move (Goodman & Weare 2010; Foreman-Mackey et al. 2013).

Bayesian object detection (as this problem is sometimes called) has been implemented both inside and outside of astronomy (e.g., Harkness & Green 2000; Hobson & McLachlan...
the parameters and the data, otherwise it would be impossible to learn about parameters by obtaining data.

When specific data \( x^* \) are taken into account, our state of knowledge about \( \theta \) gets updated from the prior distribution to the posterior distribution via Bayes’ rule:

\[
p(\theta | x = x^*) \propto p(\theta) p(x | \theta)|_{x=x^*},
\]

\[
= p(\theta) \mathcal{L}(\theta; x).
\]

The term \( p(x | \theta)|_{x=x^*} = \mathcal{L}(\theta; x) \) is the likelihood function, which is the probability of obtaining the actual data set \( x^* \) as a function of the parameters. In the case that the sampling distribution is a probability density function (PDF), the likelihood is the PDF evaluated at the observed data. This usually causes no problems, although one should be aware of the Borel–Kolmogorov paradox (Jaynes 2003). As suggested by the above notation, the likelihood function is obtained from the sampling distribution with the actual data substituted in and is therefore a function of the parameters only.

To proceed with the model for inferring catalogs from image data, we must specify a definite hypothesis space and choices for the prior distribution and the sampling distribution. These choices are presented and discussed in Section 3.

3. THE SPECIFIC MODEL FOR STELLAR FIELDS

3.1. The Hypothesis Space

The hypothesis space is the set of possible catalogs, or the set of possible answers to the question: “What objects are present in the field and what are their properties?” We shall assume that there are an unknown number of stars \( N \) in the field. Each star has an unknown position \((x, y)\) in the plane of the sky and an unknown flux \( f \). We also describe the distribution of fluxes (commonly known as the luminosity function) of the stars by some parameters denoted collectively by \( \beta \). In summary, the unknown parameters are

\[
\theta = \{N, \beta, \{x_i, y_i\}_{i=1}^N, \{f_i\}_{i=1}^N\}.
\]

We note that models similar to this have been implemented for general image modeling and deconvolution (e.g., Skilling 1998); however, in this case it is more justified as we are actually searching for point fluxes.

3.2. The Prior

The prior probability distribution for the unknown parameters can be factorized using the product rule of probability theory. With a variety of independence assumptions, the prior can be factorized as

\[
p(\theta) = p(\beta) p(N | \beta) \prod_{i=1}^N p(x_i, y_i) p(f_i | \beta).
\]

Here, we have assumed that the luminosity function does not depend on position. Finally, the fluxes of the stars come independently from a common distribution. If we knew the luminosity function of the stars, then the location and flux of a particular star would not tell us anything about the location and flux of another star. Really, this is just a way of implementing exchangeability of the stars and is often called a hierarchical model.
the mock image in that pixel. This can be modeled by assuming
the following distribution:
\[ \epsilon_{ij} \sim \mathcal{N}(0, \sigma_0^2 + \eta M(X_{ij}, Y_{ij})) \],  \hspace{1cm} (13)
where \( \sigma_0 \) is a constant noise level and \( \eta \) is an unknown coefficient that allows for the possibility that the noise level is higher in brighter regions of the image. This dependence of the noise variances on the model intensity arises as a result of the Poissonian nature of photon counts, but allows for the fact that a “sky” background may have already been subtracted from the image in the reduction process. This parameterization has been used by Brewer et al. (2011a) and is an alternative to the common practice of producing a “variance map” from the image data that are then assumed to be known.

3.4. The Prior Distribution

The prior distribution for the number of stars \( N \) is assigned to be uniform between 0 and some maximum number \( N_{\text{max}} \). The extent of the image is assumed to be from \( x = x_{\text{min}} \) to \( x = x_{\text{max}} = x_{\text{min}} + x_{\text{range}} \) and from \( y = y_{\text{min}} \) to \( y = y_{\text{max}} = y_{\text{min}} + y_{\text{range}} \) in arbitrary units, and the positions of the stars are assigned independent uniform priors:
\[ x_i \sim \text{Uniform}(x_{\text{min}} - 0.1x_{\text{range}}, x_{\text{max}} + 0.1x_{\text{range}}) \]
\[ y_i \sim \text{Uniform}(y_{\text{min}} - 0.1y_{\text{range}}, y_{\text{max}} + 0.1y_{\text{range}}). \] (14, 15)

The stars are allowed to be slightly outside of the observed image because the PSF can scatter light from these stars into the image.

For the purposes of this paper, we model the luminosity function as a broken power-law distribution, which has four free parameters:
\[ \beta = [h_1, h_2, \alpha_1, \alpha_2], \] (16)
where \( h_1 \) is a lower flux limit, \( h_2 \) is a break point, \( \alpha_1 \) is the slope of the distribution between \( h_1 \) and \( h_2 \), and \( \alpha_2 \) is the slope of the distribution above \( h_2 \). For mathematical details on the broken power-law model, see Appendix A. While the broken power law is likely to be unrealistic in many cases, it is a reasonably flexible distribution and this is sufficient for demonstrating the properties of our method.

The prior distribution on \( h_1, h_2, \alpha_1, \) and \( \alpha_2 \) is assigned to be
\[ \ln h_1 \sim \text{Uniform}(-10^{-3}, 10^3) \] (17)
\[ \ln h_2 \sim \text{Uniform}(\ln(h_1), \ln(h_2) + 2.3) \] (18)
\[ \alpha_1 \sim \text{Uniform}(1, 5) \] (19)
\[ \alpha_2 \sim \text{Uniform}(1, 5). \] (20)

These priors express vague prior knowledge about \( \alpha_1 \) and \( \alpha_2 \) in addition to vague prior knowledge about \( h_1 \) and \( h_2 \) apart from the fact that the flux units are not extreme and that \( h_2 \) should be no more than an order of magnitude greater than \( h_1 \).

This simply parameterized model for the luminosity function can be criticized on the basis that information from bright stars can be used to infer the parameters of the luminosity function which then still apply at lower flux levels. In principle, this can
be resolved by using a more flexible distribution (e.g., Kelly et al. 2008) where each star’s measured brightness affects the inference of the luminosity function locally but not globally.

The priors for the PSF parameters and the noise parameters were assigned to be

\[
\ln s_1 \sim \text{Uniform}(\ln(0.3L), \ln(30L)) \quad (21)
\]

\[
\ln s_2 \sim \text{Uniform}(\ln(s_1), \ln(s_1) + 2.3) \quad (22)
\]

\[
w \sim \text{Uniform}(0, 1) \quad (23)
\]

\[
\ln \sigma_0 \sim \text{Uniform}(\ln(10^{-3}), \ln(10^3)) \quad (24)
\]

\[
\ln \eta \sim \text{Uniform}(\ln(10^{-3}), \ln(10^3)), \quad (25)
\]

where \( L = \frac{x_{\text{range}}}{n} \) is the width of a pixel. These priors describe vague prior knowledge about the overall scale of the PSF except that the wider component is less than 10 times as wide as the narrow component, as well as the knowledge that the noise variance is not extreme relative to the fluxes of the stars.

4. MCMC IMPLEMENTATION

The MCMC sampling was implemented using the Diffusive Nested Sampling (DNS; Brewer et al. 2011b) method. DNS is a variant of the Nested Sampling (Skilling 2006) algorithm that uses Metropolis–Hastings updates, and is very generally applicable. The main difference between DNS and the standard Metropolis–Hastings algorithm is that the target distribution is modified. Rather than simply exploring the posterior distribution over catalog space, DNS constructs an alternative target distribution which is a mixture of the prior distribution with more constrained versions of the prior distribution. The modified target distribution assists the sampling in several ways. First, the target distribution shrinks at a constant rate with time during the initial phase of the exploration. This is similar to the popular “simulated annealing” method (Kirkpatrick et al. 1983; Neal 2001) but with an optimal annealing schedule. Second, communication with the prior is maintained: once a catalog is found that fits the data, the catalog can “disintegrate” back to the prior distribution and re-fit, allowing different peaks in the parameter space to be explored (if they exist). This all happens naturally within the context of a valid MCMC sampler. The MCMC may also be run using the standard Metropolis algorithm targeting the posterior distribution.

5. SIMULATED DATA

In order to test our approach, we applied the method to two illustrative simulated images generated from the above model (Figure 1). The purpose of this experiment was to test the computational feasibility of the model, as well as to compare the inferences from the model with those from more standard techniques.

The true parameter values for the two simulated data sets are listed in Table 1. The broken power-law parameter values were chosen so that roughly half of the stars’ fluxes were below and above the break point, respectively. Figure 7 in Appendix A also shows the true flux distribution used for the simulated images. Each of the images is 100 × 100 pixels in extent and covers a range from −1 to 1 in arbitrary units for both the x and y axes. The first image contains 100 stars (including stars just outside of the image); there are 63 stars whose central positions lie within the image) and the second image contains ∼1000 stars (699 of which are positioned within the image).

5.1. Test Case 1

Test Case 1 was run with the DNS algorithm and usable results were obtained within about an hour on a modern desktop PC. The inferences on the parameters \( N, h_1, h_2, \alpha_1, \) and \( \alpha_2 \) are shown in Figure 2. The number of stars is correctly inferred, and the posterior distributions for the other parameters comfortably contain the true input values. As \( N \) is a parameter of our model, there is no need for Bayesian “Model Comparison” calculations to be done between different values of \( N \). The DNS method does compute the “evidence” value that is required for model comparison, but this is useful only to test completely separate models; it is not needed to infer the value of \( N \).

Note that the uncertainty in \( h_2, \alpha_1, \) and \( \alpha_2 \) is quite large. This is because the broken power-law model (Figure 7) does not change drastically in shape as the parameters are varied. Therefore, a large number of stars would be required to tightly constrain these parameters.

The PSF parameters \( \{s_1, s_2, w\} \) and the noise parameters \( \{\sigma_0, \eta\} \) were also inferred accurately with small uncertainties.

5.2. Test Case 2

Test Case 2 is more challenging than Test Case 1 because the image contains more stars. This increases the size of the

![Figure 1](image-url)
Figure 2. Inference about the parameters for Test Case 1. The left panel shows the posterior distribution for the number of stars \( N \), and the right panels show the joint posterior distributions for the flux distribution parameters. Note that there is considerable uncertainty (particularly about \( h_2 \)), which occurs because the shape of the broken power law does not depend strongly on the parameters. The true input values are plotted as filled squares.

(A color version of this figure is available in the online journal.)

Figure 3. Fundamentally, the output from our method is samples from the posterior distribution over the catalog space. Nine example catalogs are shown, sampled from the posterior distribution for Test Case 2. Features in common represent features with high probability, and differences between the catalogs represent conclusions that are uncertain. The area of each circle is proportional to the flux of the star. The posterior samples may be used to compute summary images; some of these are presented in Figure 4.

Computational task in two ways: First, there will be more unknown parameters to infer, so any MCMC algorithm will require more iterations in order to converge to the posterior distribution. Second, the time taken to compute the predicted image from a proposed catalog (in order to evaluate the likelihood) is longer because of the larger number of stars. Hence, each MCMC step also takes more time. Using DNS, some samples from the posterior distribution can be obtained in about a day on a modern multi-core PC.

Each catalog in the posterior sample represents a scenario for the true underlying image that we would observe if we had a hypothetical noise-free, infinite resolution telescope. Figure 3 shows nine possible catalogs sampled from the posterior distribution. Features that are common to these nine samples are plausible, and features that differ are uncertain.

Figure 4. Summary images produced from the posterior distribution for Test Case 2. The upper left panel shows the posterior mean high-resolution scene. The upper right panel shows the posterior mean scene when observed at the resolution of the data, and the bottom panels show the standardized model residuals.

(A color version of this figure is available in the online journal.)

From these samples, we can construct the posterior expected true scene and other summaries. Summary images are shown in Figure 4. The residuals provide a check on the validity of the model assumptions, and the posterior expected true scene provides a useful visual guide to the uncertainties present in the catalogs. In this example, the residuals show only noise because the simulated images were actually generated from the model.

The inferences on the number of stars \( N \) and the luminosity function parameters are shown in Figure 5. The uncertainty about the luminosity function parameters is still considerable despite the larger number of stars, as the fluxes of the fainter stars are not well constrained by the data. The true values are still well within the range of plausible values in the posterior distribution. As with Test Case 1, the PSF parameters \([s_1, s_2, w]\) and the noise parameters \([\sigma_0, \eta]\) were also inferred accurately with small uncertainties.
Figure 5. Inference about the parameters for Test Case 2. Note that the flux distribution parameters are still not very well constrained even with the larger number of stars. This occurs because the fluxes of faint stars are not accurately measured and because the shape of the broken power-law distribution does not vary rapidly as a function of its parameters. The true input values are plotted as filled squares.

(A color version of this figure is available in the online journal.)

Figure 6. Cumulative luminosity functions (number of stars above a given flux, as a function of flux) produced by the Bayesian method (several posterior samples shown) and SExtractor (for various values of the threshold parameters), compared with the actual cumulative luminosity function. Both methods correctly identify the fluxes at the bright end, with some uncertainty due to overlapping sources. However, at the lower end SExtractor is unable to detect all of the stars, whereas the true CLF is typical of the posterior distribution.

(A color version of this figure is available in the online journal.)

6. COMPARISON TO SExtractor

In the previous section, we established that the inference of the catalogs from the data is computationally feasible and that the number of stars and the luminosity function can be inferred from the image data, albeit with moderate uncertainties. We now compare this approach to an alternative analysis that makes use of the standard tool SExtractor (Bertin & Arnouts 1996). To achieve this, we executed SExtractor on the two test images, for various values of the detection threshold parameters DETECT_THRESH and ANALYSIS_THRESH ranging from 0.5 to 6.5. This results in a set of catalogs for each image, with more conservative thresholds resulting in less stars detected as compared to more aggressive thresholds. To compute flux estimates that are directly comparable to the fluxes in our input catalogs, we configured SExtractor to compute object fluxes within fixed circular apertures that were known to contain 70% of the mass of the PSF. The SExtractor flux estimates were then scaled up to account for this finite aperture.

In Figure 6, we present the cumulative luminosity function (CLF) of the stars in the two fields, defined as the number of stars above a given flux. The true CLF is plotted along with several posterior samples from the Bayesian method and catalogs produced by SExtractor for various values of the detection and analysis thresholds. We note that the true CLF is typical of the posterior samples, as expected. For both test cases, the SExtractor catalog is also consistent with the posterior distribution at the bright end. However, the inferences from SExtractor and the Bayesian method differ at the faint end, with the former significantly underestimating the number of faint stars.

This result may be attributable to the fact that the Bayesian method knows about the existence and form of the luminosity function, even though it does not know the values of the parameters. To test this, we ran the inference on the data using an incorrect exponential distribution for the luminosity function. The resulting CLF from this run did undershoot the true CLF at the faint end. However, the Bayesian evidence for the exponential model was significantly lower (by a factor of approximately $10^7$) than for the (correct) broken power-law model.

In practice, we note that the wings of the PSF might become degenerate with a nonzero flat background level in the data. To test whether this was influencing the inferences (particularly about the faint end of the CLF), we also ran a model that included an unknown constant background. This had only a minor effect on the resulting inferences.
7. DISCUSSION AND CONCLUSIONS

In this paper, we have developed and demonstrated a Bayesian approach to making catalogs from astronomical images in the case where the image contains only stars (or other point sources). The key idea is that instead of computing a single catalog, the method creates a posterior probability distribution on the space of possible catalogs that represents our state of knowledge about the presence and properties of objects in the image. When this is done, the uncertainties in the imaging are accurately propagated through to scientific conclusions, for example, about the luminosity function of the stars. This approach was contrasted with the results from the standard SExtractor software. For the bright sources, the results were essentially consistent; however, the Bayesian approach was more successful at modeling the distribution of faint stars. Of course, the Bayesian method is much more computationally intensive, which is a significant issue in practice. However, the great value of upcoming imaging data sets and the irreproducibility of astronomical imaging data in areas of time-domain astrophysics (for example, in observations of rare events) make it important to extract as much information as possible from every patch of imaging. Our view is that the additional CPU time and the non-triviality of the outputs from our method will be worth the effort in the next generation of astronomical experiments. To build a more complete picture of when our approach is necessary, it will need to be tested against a wider variety of alternative methods such as DAOPHOT (Stetson 1987) and DOLPHOT (Dolphin 2000).

We note that there are many limitations to the model presented in this paper, some of which will be important to relax when it is applied to real data. In principle, our model should be a model of the physical state of the universe, and not a simple model where the only stellar properties are a two-dimensional position and a flux. Another limitation is that we have not considered multi-epoch or multi-band imaging. In the former case, PSF variations and stellar motions may be relevant (Lang et al. 2009), and in the latter, a model for the spectral energy distributions of the stars will need to be considered: essentially, the luminosity function will need to be a probability distribution over more than one dimension.

In practice, it may also be necessary to improve the model for the prior distribution of stellar positions and fluxes. One area where this is clearly needed is the application of this approach to images of stellar clusters. The model would need to be revised to take into account the fact that we expect the stars’ positions to be clustered together, whereas the current model implies a large prior probability for the stars being scattered evenly across the image. In this and other applications, the luminosity function would also require multiple components, for example, consisting of stars that are associated with a cluster or a stream and those that are not.

Throughout this paper, we have also assumed that the pixel-convolved PSF can be adequately modeled using simple components and that there are no PSF variations across the field. Relaxing these assumptions provides a significant challenge for the future.

It is a pleasure to thank Anna Nierenberg (UCSB), Jonathan Goodman (NYU), Fengji Hou (NYU), Dustin Lang (CMU), Andrew Dolphin (Raytheon), Juliane Dalcanton (Washington), Geraint Lewis (Sydney), and Phil Marshall (Oxford) for their comments and discussions. The referee is thanked for thoughtful comments which helped us to improve the paper. B.J.B. thanks Tommaso Treu (UCSB) for his support, and Wayne Stewart and Arden Miller (Auckland) for their encouragement and advice. D.F.M. and D.W.H. were partially supported by the U.S. National Science Foundation (grant IIS-1124794) and the U.S. National Aeronautics and Space Administration (grant NNX12AI50G).

APPENDIX A

BROKEN POWER-LAW DISTRIBUTION

The broken power-law distribution is based on a straightforward extension to a simple power-law distribution (also known as a Pareto distribution, particularly in the statistics literature). The power-law distribution for a variable \( x \) (given a lower cutoff \( x = h_1 \) and a slope \( \alpha \)) is defined by

\[
p(x) \propto \begin{cases} 
0, & x < h_1 \\
x^{-\alpha-1}, & x \geq h_1.
\end{cases}
\]

(A1)

In contrast, the broken power-law distribution for a variable \( x \) is defined by a lower cutoff \( x = h_1 \), two slopes \( \{\alpha_1, \alpha_2\} \), and a break point \( x = h_2 \):

\[
p(x) \propto \begin{cases} 
0, & x < h_1 \\
x^{-\alpha_1-1}, & h_1 \leq x \leq h_2 \\
x^{-\alpha_2-1}, & x > h_2.
\end{cases}
\]

(A2)

The free parameters of the broken power law are

\[
\beta = \{h_1, h_2, \alpha_1, \alpha_2\}.
\]

(A3)

With normalizing terms included, the proportionality becomes an equality:

\[
p(x) = \begin{cases} 
Z^{-1}_1 x^{-\alpha_1-1}, & h_1 \leq x \leq h_2 \\
Z^{-1}_2 x^{-\alpha_2-1}, & x > h_2.
\end{cases}
\]

(A4)

Two conditions will be used to determine the normalizers \( Z_1 \) and \( Z_2 \). First, the PDF should be continuous at \( x = h_2 \):

\[
Z^{-1}_1 h_2^{-\alpha_1-1} = Z^{-1}_2 h_2^{-\alpha_2-1}
\]

\[
\Rightarrow Z_2 = Z_1 h_2^{\alpha_1-\alpha_2}.
\]

(A5)

The second condition is that the total probability must be 1:

\[
\int_{h_1}^{h_2} Z^{-1}_1 x^{-\alpha_1-1} \, dx + \int_{h_2}^{\infty} Z^{-1}_2 x^{-\alpha_2-1} \, dx = 1
\]

\[
Z^{-1}_1 \alpha_1 [h_1^{-\alpha_1} - h_2^{-\alpha_1}] + Z^{-1}_2 \alpha_2 h_2^{-\alpha_2} = 1
\]

\[
Z^{-1}_1 \alpha_1 [h_1^{-\alpha_1} - h_2^{-\alpha_1}] + Z^{-1}_1 \alpha_2 h_2^{-\alpha_2} = 1
\]

\[
\Rightarrow Z_1 = \alpha_1 [h_1^{-\alpha_1} - h_2^{-\alpha_1}] + h_2^{-\alpha_2} \alpha_2^{-1}.
\]

(A6)

The cumulative distribution function (CDF) is a useful property of a probability distribution and is given by the antiderivative of the PDF:

\[
P(X \leq x) = F(x)
\]

\[
= \begin{cases} 
0, & x < h_1 \\
(Z_1 \alpha_1)^{-1} (h_1^{-\alpha_1} - x^{-\alpha_1}), & h_1 \leq x \leq h_2 \\
1 - (Z_2 \alpha_2)^{-1} x^{-\alpha_2}, & x > h_2.
\end{cases}
\]

(A11)
The inverse of the CDF is also useful and is given by

\[ F^{-1}(u) = \begin{cases} h_1 - uZ_1^{\alpha_1} \alpha_1^{-1}, & 0 < u < 1 - (Z_2 \alpha_2)^{-1} h_2^{-\alpha_2}; \\ Z_2(1 - u)^{-1/\alpha_2}, & 1 - (Z_2 \alpha_2)^{-1} h_2^{-\alpha_2} < u < 1. \end{cases} \]  

(A12)

An example of a broken power-law distribution is shown in Figure 7.

**APPENDIX B**

**PROPOSAL DISTRIBUTIONS**

To implement Metropolis–Hastings moves for the space of possible catalogs, proposal distributions are required. See Table 2 for a list of proposal distributions used in this study.

**REFERENCES**

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Brewer, B. J., Lewis, G. F., Belokurov, V., et al. 2011a, MNRAS, 412, 2521
Brewer, B. J., Pártay, L. B., & Csányi, G. 2011b, Stat. Comput., 21, 649
Cattani, A. 2009, in AIP Conf. Proc. 1193, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. P. M. Goggans & C.-Y. Chan (Melville, NY: AIP), 60
Cox, R. T. 1946, Am. J. Phys., 14, 1
Dolphin, A. E. 2000, PASP, 112, 1383
Feroz, F., Balan, S. T., & Hobson, M. P. 2011, MNRAS, 415, 3462
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Green, P. J. 1995, Biometrika, 82, 711
Harkness, M. A., & Green, P. J. 2000, in Proc. British Machine Vision Conf., Parallel Chains, Delayed Rejection and Reversible Jump MCMC for Object Recognition, ed. M. Mirmeendi & B. Thomas (Manchester: British Machine Vision Association Press), 73–1
Hobson, M. P., & McLachlan, C. 2003, MNRAS, 338, 765
Hogg, D. W. 2012, Data Analysis Recipes: Probability Calculus for Inference, arXiv:1205.4446
Hogg, D. W., & Lang, D. 2011, Telescopes Don’t Make Catalogues!, ed. C. Turon, F. Meynadier, & F. Arenou (EAS Publ. Ser., Vol. 45; Cambridge: Cambridge Univ. Press), 351
Irwin, M. J. 1985, MNRAS, 214, 575
Jasra, A., Holmes, C. C., & Stephens, D. A. 2005, StAsC, 20, 50
Jaynes, E. T. 2003, Probability Theory: The Logic of Science (Cambridge: Cambridge Univ. Press)
Kelly, B. C., Fan, X., & Vestergaard, M. 2008, ApJ, 682, 874
Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. 1983, Sci., 220, 671
Lang, D., Hogg, D. W., Jester, S., & Rix, H.-W. 2009, AJ, 137, 4400
Lupton, R., Gunn, J. E., Ivezić, Z., et al. 2001, in ASP Conf. Proc. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco, CA: ASP), 269
Mackay, D. J. C. 2003, Information Theory, Inference and Learning Algorithms (Cambridge: Cambridge University Press)
Magain, P., Courbin, F., Gillon, M., et al. 2007, A&A, 461, 373
Metcalf, S. A., & Grindlay, J. E. 2002, MNRAS, 335, 73
Neal, R. M. 2001, Stat. Comput., 11, 125
Skilling, J. 1998, Massive Inference and Maximum Entropy, in Maximum Entropy and Bayesian Methods (Dordrecht: Kluwer), 1
Skilling, J. 2006, Bay. A, 1, 833
Stetson, P. B. 1987, PASP, 99, 191
Zhang, Z. W., Kim, D. W., Wang, J. H., et al. 2009, PASP, 121, 1429