Mon. Not. R. Astron. Soc. **387,** 735–740 (2008) doi:10.1111/j.1365-2966.2008.13259.x

Unbiased estimates of galaxy scaling relations from photometric redshift surveys

Graziano Rossi★ and Ravi K. Sheth★

Department of Physics and Astronomy, University of Pennsylvania, 209 S. 33rd Street, Philadelphia, PA 19104, USA

Accepted 2008 March 26. Received 2008 March 4; in original form 2007 October 5

ABSTRACT

Many physical properties of galaxies correlate with one another, and these correlations are often used to constrain galaxy formation models. Such correlations include the colour–magnitude relation, the luminosity–size relation, the fundamental plane, etc. However, the transformation from observable (e.g. angular size, apparent brightness) to physical quantity (physical size, luminosity) is often distance dependent. Noise in the distance estimate will lead to biased estimates of these correlations, thus compromising the ability of photometric redshift surveys to constrain galaxy formation models. We describe two methods which can remove this bias. One is a generalization of the $V_{\rm max}$ method, and the other is a maximum-likelihood approach. We illustrate their effectiveness by studying the size–luminosity relation in a mock catalogue, although both methods can be applied to other scaling relations as well. We show that if one simply uses photometric redshifts one obtains a biased relation; our methods correct for this bias and recover the true relation.

Key words: methods: analytical – methods: statistical – galaxies: formation – cosmology: observations.

1 INTRODUCTION

The 'configuration space' we use to describe galaxies is large, but galaxies do not fill it. The luminosity, colour, size, surface brightness, stellar velocity dispersion, morphology, stellar mass, star formation history and spectral energy distribution of a galaxy are all correlated with one another. These correlations encode important information about galaxy formation, and so quantifying them provides important constraints on models.

Current (e.g. SDSS, Combo-17, MUSYC, Cosmos) and planned surveys (e.g. DES, LSST, SNAP) go considerably deeper in multicolour photometry than in spectroscopy, or are entirely photometric. For such surveys, reasonably accurate photometric redshift estimates are or will be made. The question then arises as to which galaxy observables and correlations are affected by the noisy distance estimate associated with a photometric rather than spectroscopic redshift. The most widely studied property is luminosity – clearly, errors in the distance result in incorrect luminosity estimates. If not accounted for, this leads to a biased estimate of the luminosity function (e.g. Subbarao et al. 1996). Hence, there has been considerable effort devoted to the question of how to correct for this bias (e.g. Chen et al. 2003), and the problem has now been solved (Sheth 2007).

★E-mail: grossi@sas.upenn.edu (GR); shethrk@physics.upenn.edu (RKS)

The next step is to recover an unbiased estimate of not just the luminosity function, but the joint distribution of luminosity, colour, size, etc., from photometric redshift data sets. The main goal of this work is to provide an algorithm which does this for a magnitude-limited photometric redshift survey. Because the same distance error which leads to a misestimate of the luminosity will produce a correlated misestimate of the size, we have chosen to phrase the discussion in terms of the size–luminosity relation – it exhibits all the features of interest.

Section 2 illustrates the nature of the problem by showing the bias in the size–luminosity relation which results from treating photometric redshifts as though they were spectroscopic redshifts. This is done by constructing a mock galaxy sample and then perturbing the true redshifts to mimic photometric redshift errors. Section 3 places this problem in the more general context of inverse problems in statistical astronomy, and argues that a deconvolution algorithm, such as that due to Lucy (1974), is well suited to removing the bias. It shows the result of applying this non-parametric deconvolution technique to a mock galaxy sample. Section 4 provides a maximum-likelihood formulation and solution of the problem. A final section summarizes our findings and discusses possible further studies and applications.

Where necessary, we write the Hubble constant as $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, and we assume a spatially flat cosmological model with $(\Omega_{\rm M}, \Omega_\Lambda, h) = (0.3, 0.7, 0.7)$, where $\Omega_{\rm M}$ and Ω_Λ are the present-day densities of matter and cosmological constant scaled to

the critical density. We use $D_{\rm L}(z)$ to denote the luminosity distance; the angular diameter distance is $D_{\rm A}(z) = D_{\rm L}(z)/(1+z)^2$.

2 CORRELATIONS WITH OBSERVABLES: EFFECT OF DISTANCE ERRORS

In what follows, we use the luminosity–size relation to illustrate how photo-z errors lead to biases.

We begin by generating a magnitude-limited mock galaxy catalogue with parameters chosen to mimic those of Sloan Digital Sky Survey (SDSS) early-type galaxies in the g band, following the method given by Bernardi et al. (2003). The redshift range is restricted to the interval $0.01 \leqslant z \leqslant 0.3$. We ignored passive evolution of the luminosities and colours, as well as K-corrections. The simulated magnitude-limited catalogue has a similar $\mathrm{d}N/\mathrm{d}z$ distribution to that observed, and the distribution of apparent magnitudes, angular sizes and velocity dispersions are very similar to those in the real data.

We then model photometric redshifts ζ by assuming that

$$p[D_{\rm L}(\zeta)\,|\,D_{\rm L}(z)]\,\mathrm{d}D_{\rm L}(\zeta) = \frac{\mathrm{d}x}{x}\,(\gamma x)^{\gamma}\,\frac{\exp(-\gamma x)}{\Gamma(\gamma)}, \qquad (1)$$

where $x = D_{\rm L}(\zeta)/D_{\rm L}(z)$ is the ratio of the photo-z based luminosity distance to the true one and $\gamma = 5$. This distribution has $\langle x\rangle = 1$ and $\sigma_x^2 = 1/\gamma$. With $\gamma = 5$, this error distribution is substantially worse than typical photometric redshift errors. Fig. 1 compares ζ and z. Note that the analysis which follows is not tied to this functional form for the photo-z error distribution; we are simply using it to illustrate our methods.

In what follows, we use M to denote the true absolute magnitude and $\mathcal{M}$ to denote that estimated using ζ rather than z, and, with some abuse of notation, we use R to denote $\log_{10}$ of the physical size $\theta\, D_{\rm A}(z)$, where θ is the measured angular size and $D_{\rm A}(z)$ is the angular diameter distance defined earlier. The estimated size based on the photometric redshift ζ is

$$\begin{aligned}\mathcal{R} &\equiv \log_{10}[\theta D_{\rm A}(\zeta)] = R + \log_{10}[D_{\rm A}(\zeta)/D_{\rm A}(z)] \\ &= R - (\mathcal{M} - M)/5 - 2\log_{10}[(1+\zeta)/(1+z)]; \end{aligned} \qquad (2)$$

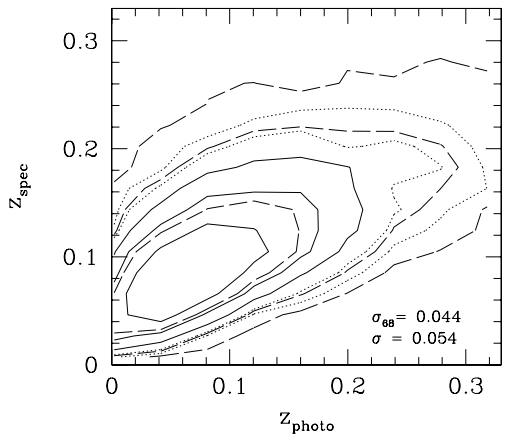

Figure 1. Distribution of spectroscopic and photometric redshifts in our mock catalogue which was set-up to mimic the SDSS early-type galaxy sample. The solid and dotted contours show levels which are $1/2^n$ times the height of the maximum value of the density of sources, with n running from 1 to 5. The dashed contours are the 1, 2 and 3σ levels. Photo-z accuracy, as defined in Oyaizu et al. (2007), is specified in the panel. The photo-z error distribution was assumed to follow equation (1).

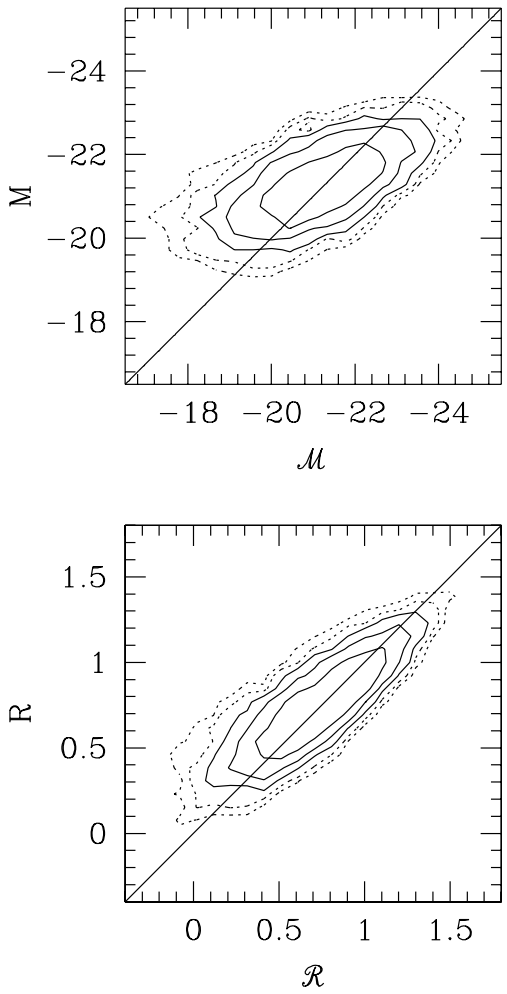

Figure 2. Distributions of intrinsic and estimated absolute magnitudes (top panel) and sizes (bottom panel) which result from the differences between true and photo-z shown in the previous figure.

in principle, there are also evolution and K-correction terms which we set to zero. Fig. 2 compares $\mathcal{M}$ with M and $\mathcal{R}$ with R.

The qualitative nature of the distributions in the two panels is easy to understand. The distribution in $\mathcal{M}$ is broader than that in M, as is the distribution of $\mathcal{R}$ compared to R: photometric redshift errors have broadened both distributions. However, the changes to the estimated absolute magnitude and size are not independent. Assuming an object is closer than it really is makes one infer a fainter luminosity and smaller size than it really has. These correlated changes can have a dramatic effect on the size–luminosity relation, since photo-z errors move each galaxy in the R–M plane left and down or right and up. In general, these motions are not parallel to the principal axis of the true relation, so the mean relation in the photo-z catalogue, $\langle\mathcal{R}\,|\,\mathcal{M}\rangle$, need not be the same as the true relation $\langle R\,|\,M\rangle$. In our mock catalogue, $\langle\mathcal{R}\,|\,\mathcal{M}\rangle \propto -0.20\,\mathcal{M}$, whereas $\langle R\,|\,M\rangle \propto -0.27\,M$ (see Fig. 3).

3 A NON-PARAMETRIC DECONVOLUTION-LIKE METHOD

Sheth (2007) shows that to extend Schmidt's (1968) $V_{\rm max}$ estimator of the luminosity function $\phi(M)$ so that it produces unbiased results in photo-z surveys, it is helpful to think of $\mathcal{N}(\mathcal{M})$, the number of

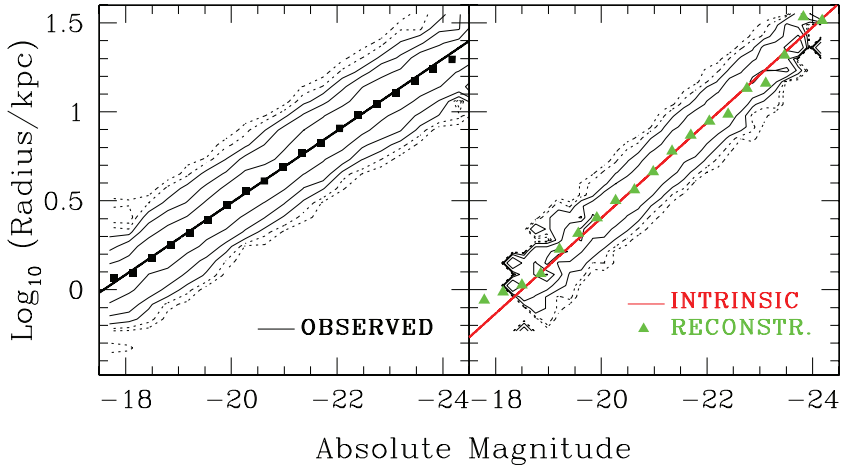

Figure 3. Effect of photo-z on the size–luminosity correlation in our mock catalogue. In the left-hand panel, contours and solid line show the R–M relation associated with photo-zs, whereas the right-hand panel shows the intrinsic R–M relation. Note the strong bias (shallower slope in left-hand panel) which results from the fact that the photo-z distance error moves points down and left or up and right on this plot. The squares in the left-hand panel show the binned starting guess for the 2D deconvolution algorithm, and the triangles in the right-hand panel show the result after 20 iterations (equation 13). Convergence to the correct solution is clearly seen.

observed objects with estimated $\mathcal{M}$, as being a convolution of the true number which have $M, N(M)$, with the probability that an object with magnitude M is thought to have magnitude $\mathcal{M}$. The luminosity function is then estimated by first deconvolving the distribution of $\mathcal{N}(\mathcal{M})$ to obtain $N(M)$, and then using the fact that $N(M) \equiv \phi(M)\, V_{\rm max}(M)$.

For the same reason, the present problem may be thought of as a 2D deconvolution problem (or an n-dimensional deconvolution if we were interested in the full manifold, rather than just 2D projections of it). Specifically, let

$$N(M, R) = N(M)\, p(R \mid M) = V_{\rm max}(M)\, \phi(M)\, p(R \mid M) \qquad (3)$$

denote the (true) number of galaxies with absolute magnitude M and size R in a magnitude-limited catalogue. Here, $p(R \mid M)$ is the probability of having size R when the magnitude is M. Similarly, set $\mathcal{N}(\mathcal{M}, \mathcal{R}) = \mathcal{N}(\mathcal{M})\, p(\mathcal{R} \mid \mathcal{M})$. Then,

$$\mathcal{N}(\mathcal{M}, \mathcal{R}) = \int {\rm d}M \int {\rm d}R\; N(M, R)\; p(\mathcal{M}, \mathcal{R} \mid M, R)$$
$$= \int {\rm d}M\; N[M, \mathcal{R} + F(M, \mathcal{M})]\; p(\mathcal{M} \mid M), \qquad (4)$$

where $F = (\mathcal{M} - M)/5 + 2\log[(1+\zeta)/(1+z)]$. Our problem is to obtain a reliable estimate of the intrinsic $N(M, R)$ given the photo-z biased $\mathcal{N}(\mathcal{M}, \mathcal{R})$ and the error distribution $p(\mathcal{M} \mid M)$. We do this using the deconvolution algorithm proposed by Lucy (1974).

Before we present our algorithm and results, it is worth noting that we could have attempted to invert equation (4) in other ways. Classical naive 'exact' inversion methods include matrix-quadrature techniques (reduction of the integral equation to a linear matrix system), polynomial expansion methods, singular function expansion and product integration methods. These typically run into difficulties because the measured data function usually cannot provide sufficient information on the high-frequency components of the solution. A standard non-classical technique (Phillips 1962; Tikhonov 1963; Twomey 1963) is the method of regularization. A 'regularization' parameter is introduced, which balances the size of the residual against the smoothness of the solution, and the problem is turned into one of minimization. However, there is no general strategy for choosing the optimum regularization parameter; this led Lucy (1974) to formulate his algorithm, and is what has led us to choose his algorithm over these others. One might argue that we are likely to know a fair amount about the expected form of the intrinsic distribution (e.g. luminosity functions are rather smooth, and conditional distributions tend to be bell shaped), so it may be that these other methods are worth pursuing further. This is the subject of work in progress.

3.1 The deconvolution algorithm

The general 2D problem is that of estimating the frequency distribution $\Psi(\xi', \eta')$ of the quantities ξ' and η' when the available measures $x'_1, y'_1; x'_2, y'_2; \ldots, x'_N, y'_N$ are a finite sample drawn from an infinite population characterized by

$$\Phi(x, y) = \int {\rm d}\xi \int {\rm d}\eta\; \Psi(\xi, \eta)\; p(x, y \mid \xi, \eta). \qquad (5)$$

Here, $\Phi(x, y)$ is the function accessible to measurement and $p(x, y \mid \xi, \eta)$ is the conditional probability that x' will fall in $[x, x + {\rm d}x]$ when it is known that $\xi' \equiv \xi$, and that y' will fall in $[y, y + {\rm d}y]$ when $\eta' \equiv \eta$. In many cases, Φ and Ψ represent probability density functions, so they obey normalization and non-negativity constraints.

The iterative procedure for generating estimates to Ψ presented in Lucy (1974) is

$$\Psi^{r+1}(\xi, \eta) = \Psi^{r}(\xi, \eta) \int {\rm d}x \int {\rm d}y\; \frac{\tilde{\Phi}(x, y)}{\Phi^{r}(x, y)}\; p(x, y \mid \xi, \eta), \qquad (6)$$

where

$$\Phi^{r}(x, y) = \int {\rm d}\xi \int {\rm d}\eta\; \Psi^{r}(\xi, \eta)\; p(x, y \mid \xi, \eta). \qquad (7)$$

The index r indicates the rth iteration in the sequence of estimates, and $\tilde{\Phi}$ is an approximation to Φ obtained from the observed sample. Convergence is achieved if $\Phi^{r} = \tilde{\Phi}$. The starting approximation $\Psi^{0}(\xi, \eta)$ should be a smooth, non-negative function having the same integrated density as the observed distribution. The extension to n-dimensions is obvious, although, if the number of dimensions is large, then performing the multidimensional integrations efficiently may become challenging.

The 2D problem simplifies if, as happens in our problem,

$$p(x, y \mid \xi, \eta) = p(x \mid \xi)\, p(y \mid \xi, \eta, x) = p(x \mid \xi)\, \delta_{\rm D}[y = \eta - F(x, \xi)] \quad (8)$$

because the delta function simplifies one of the integrals. The iterative scheme becomes

$$\Psi^{r+1}(\xi, \eta) = \Psi^{r}(\xi, \eta) \int {\rm d}x\, \frac{\tilde{\Phi}[x, \eta - F(x, \xi)]}{\Phi^{r}[x, \eta - F(x, \xi)]}\, p(x \mid \xi) \quad (9)$$

with

$$\Phi^{r}(x, y) = \int {\rm d}\xi\, \Psi^{r}[\xi, y + F(x, \xi)]\, p(x \mid \xi). \quad (10)$$

Partial distributions can be easily computed via marginalization from Ψ^{r+1}:

$$\Psi^{r+1}(\xi) = \int {\rm d}\eta\, \Psi^{r+1}(\xi, \eta), \quad (11)$$

$$\Psi^{r+1}(\eta) = \int {\rm d}\xi\, \Psi^{r+1}(\xi, \eta) \quad (12)$$

and

$$\langle \eta \mid \xi \rangle^{r+1} = \int {\rm d}\eta\, \eta\, \frac{\Psi^{r+1}(\eta, \xi)}{\Psi^{r+1}(\xi)}. \quad (13)$$

The generalization to n-dimensions is a straightforward extension of the expressions above, so we do not present explicit expressions. Note that, just as happens in the 2D case presented here, delta functions will reduce the n-dimensional problem to a simple 1D integral, because the same distance error affects all n quantities.

3.2 2D results

The formalism outlined above is readily applicable to the size–luminosity correlation if we interpret (x, y) as the estimated absolute magnitudes and sizes $(\mathcal{M}, \mathcal{R})$, and (ξ, η) as the true intrinsic ones, (M, R).

Figs 4 and 5 show how well this method recovers the intrinsic distribution of absolute magnitudes and sizes (solid histograms). The broad dashed histograms show the photo-z derived $\mathcal{M}$ and $\mathcal{R}$

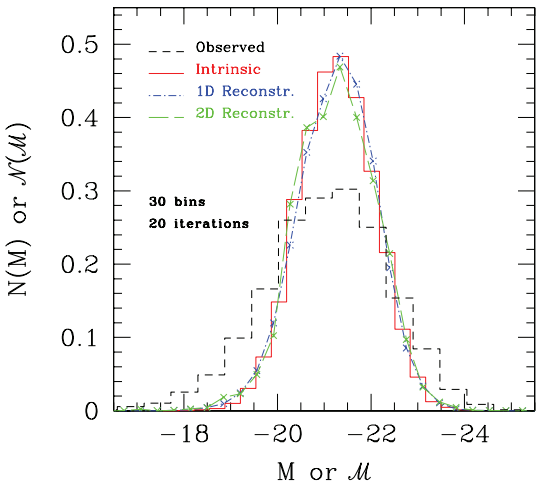

Figure 4. Reconstruction of the intrinsic $N(M)$ distribution from the distribution of estimated redshifts. The dashed histogram shows the observed absolute magnitude distribution, used as a starting guess. The jagged lines show the reconstructed intrinsic distribution after 20 iterations, using the simpler 1D algorithm (blue) or a 2D iterative scheme (green).

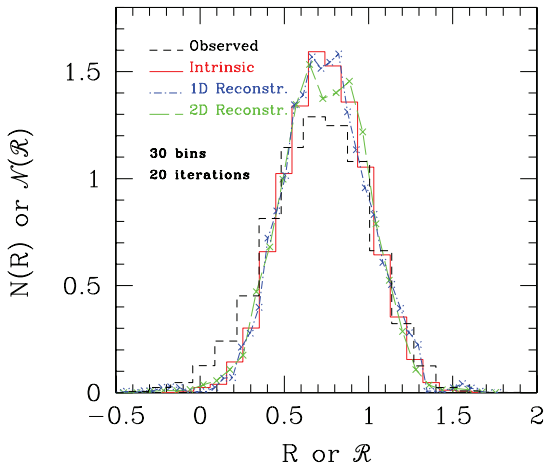

Figure 5. Reconstruction of the intrinsic $N(R)$ distribution from the distribution of estimated redshifts. The line styles are same as in the previous figure.

distributions. These were used as a convenient starting guess in the 2D deconvolution algorithm, although prior knowledge about the expected intrinsic shapes could have been used instead. The reconstruction after 20 iterations is shown by the jagged green lines. Note how well the deconvolved distributions match the intrinsic ones.

A more stringent test is to check if the conditional distributions, $p(R \mid M)$, are also well recovered. Fig. 6 shows $p(R \mid M)$ for three bins of width $\Delta M = 0.355$ centred at -22.046, -21.691 and -20.982. Clearly, the method works well.

The means of these recovered distributions can be used as an estimate of the recovered size–luminosity relation: $\langle R \mid M \rangle$. Recall we had noted that this relation was rather strongly biased because use of the photo-z distance estimate means that the error in the size is correlated with that on the luminosity. The squares in the left-hand panel of Fig. 3 show the starting guess for our algorithm, and the triangles in the right-hand panel show the reconstructed relation obtained from the deconvolution procedure – it is in an excellent agreement with the true one.

3.3 1D results

Of course, we could have chosen to reconstruct $N(M)$ directly from $\mathcal{N}(\mathcal{M})$, using the 1D deconvolution algorithm outlined in Sheth (2007). This procedure converges in about five iterations to the distribution shown by the blue crosses in Fig. 4.

Similarly, one can reconstruct the distribution of sizes with a 1D deconvolution as follows. The true number of objects with size R is

$$N(R) = \int {\rm d}M\, N(M, R) = \int {\rm d}M\, N(M)\, p(R \mid M) = \int {\rm d}M\, \phi(M)\, V_{\rm max}(M)\, p(R \mid M), \quad (14)$$

whereas the number of observed objects with estimated size $\mathcal{R}$ is

$$\mathcal{N}(\mathcal{R}) = \int {\rm d}\mathcal{M}\, \mathcal{N}(\mathcal{M}, \mathcal{R}) = \int {\rm d}\mathcal{M} \int {\rm d}M \int {\rm d}R\, N(M, R)\, p(\mathcal{M}, \mathcal{R} \mid M, R)$$

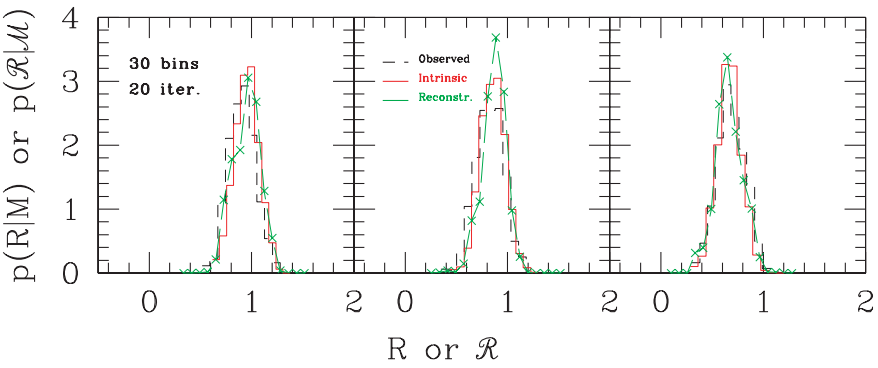

Figure 6. Examples of reconstructed conditional distributions $p(R|M)$ bins in magnitude of width $\Delta M = 0.355$ centred on $M = -22.046$, -21.691 and -20.982. The jagged lines show the distributions recovered by the 2D deconvolution algorithm after 20 iterations.

$$
\begin{aligned}
&= \int \mathrm{d}\mathcal{M} \int \mathrm{d}M \int \mathrm{d}R\; N(R)\; p(M \mid R) \\
&\quad \times p(\mathcal{R} \mid R)\; p(\mathcal{M} \mid M, R, \mathcal{R}) \\
&= \int \mathrm{d}R\; N(R)\; p(\mathcal{R} \mid R) \int \mathrm{d}\mathcal{M} \int \mathrm{d}M\; p(M \mid R) \\
&\quad \times p(\mathcal{M} \mid M, R, \mathcal{R}) \\
&= \int \mathrm{d}R\; N(R)\; p(\mathcal{R} \mid R) \int \mathrm{d}M\; p(M \mid R) \\
&\quad \times \int \mathrm{d}\mathcal{M}\; p(\mathcal{M} \mid M, R, \mathcal{R}) \\
&\equiv \int \mathrm{d}R\; N(R)\; p(\mathcal{R} \mid R). \qquad (15)
\end{aligned}
$$

This algebra shows that one can think of $\mathcal{N}(\mathcal{R})$ as being a convolution of the true number of objects with size R. Hence, by measuring the conditional probability $p(\mathcal{R}|R)$ directly from the catalogue, one can reconstruct the intrinsic distribution $N(R)$ from the observed $\mathcal{N}(\mathcal{R})$ using a simple 1D deconvolution algorithm. The result of doing this is shown by the blue crosses in Fig. 5. Of course, the relation between $N(R)$ and $\phi(R)$ is more complicated than that between $N(M)$ and $\phi(M)$.

4 A MAXIMUM-LIKELIHOOD METHOD

Sheth (2007) describes an algorithm which produces a maximum-likelihood estimate of the luminosity function from magnitude-limited photo-z data sets. It is straightforward to extend that analysis to the present case, in which the quantity of interest is not just the distribution of luminosities, but the joint distribution of luminosity and other observed physical parameters.

Let $\boldsymbol{\mathcal{M}}_i$ denote the vector of physical quantities for galaxy i estimated using the photometric redshift ζ_i when computing distances, and let $\mathcal{N}(\boldsymbol{\mathcal{M}}_i, \zeta_i|\boldsymbol{a})$ denote the number of galaxies in a magnitude-limited catalogue that have estimated redshifts ζ_i and estimated luminosities, sizes, etc., $\boldsymbol{\mathcal{M}}_i$, when the model for the intrinsic joint distribution of physical quantities is specified by the parameters $\boldsymbol{a}$. Then, dropping the understood i-index dependence,

$$
\mathcal{N}(\boldsymbol{\mathcal{M}}, \zeta \mid \boldsymbol{a}) = \int \mathrm{d}z \frac{\mathrm{d}V_\mathrm{c}}{\mathrm{d}z}\, \phi(\boldsymbol{M} \mid \boldsymbol{a})\; p[\zeta \mid z, m(\boldsymbol{\mathcal{M}}, \zeta)], \qquad (16)
$$

where $p(\zeta|z,\ m)$ is the photo-z redshift-error distribution and $\phi(\boldsymbol{M}|\boldsymbol{a})$ denotes the true joint distribution of physical quantities, but evaluated at values which account for the fact that the same photo-z error which affects the absolute magnitudes affects the other observables. For example, in the size–luminosity relation we have been considering,

$$
\boldsymbol{\mathcal{M}} = (\mathcal{M}, \mathcal{R}) \qquad \text{and} \qquad \boldsymbol{M} = (M, R), \qquad (17)
$$

where

$$
M = \mathcal{M} - 5 \log_{10}[D_\mathrm{L}(z)/D_\mathrm{L}(\zeta)], \qquad (18)
$$

$$
R = \mathcal{R} + \log_{10}[D_\mathrm{A}(z)/D_\mathrm{A}(\zeta)]. \qquad (19)
$$

The factor of m is the apparent magnitude associated with photo-z redshift ζ and absolute magnitude $\mathcal{M}$; of course, m is the same for true redshift z and absolute magnitude M.

The predicted number of objects with photometric redshift ζ when the model has parameters $\boldsymbol{a}$ is

$$
\mathcal{N}(\zeta \mid \boldsymbol{a}) = \int \mathrm{d}\boldsymbol{\mathcal{M}}\, \mathcal{N}(\boldsymbol{\mathcal{M}}, \zeta \mid \boldsymbol{a}). \qquad (20)
$$

If the redshift-error distribution is independent of m, then

$$
\begin{aligned}
\mathcal{N}(\zeta \mid \boldsymbol{a}) &= \int \mathrm{d}z\; (\mathrm{d}V_\mathrm{c}/\mathrm{d}z)\; S(z, \boldsymbol{a})\; p(\zeta \mid z) \\
&\equiv \int \mathrm{d}z\; N(z|\boldsymbol{a})\; p(\zeta \mid z), \qquad (21)
\end{aligned}
$$

where

$$
S(z, \boldsymbol{a}) = \int_{M_{\max}(z)}^{M_{\min}(z)} \mathrm{d}M\; \phi(M|z, \boldsymbol{a}). \qquad (22)
$$

This shows that $\mathcal{N}(\zeta \mid \boldsymbol{a})$ is just the convolution of the intrinsic redshift distribution (in a flux-limited catalogue) with the redshift-error distribution.

The expressions above generalize those given in Sheth (2007), where $\boldsymbol{\mathcal{M}} = \mathcal{M}$ and $\boldsymbol{M} = M$, with M given by equation (18). Hence, by analogy to when the distances are known accurately, the likelihood to be maximized is (reintroducing the index i)

$$
\mathcal{L}(\boldsymbol{a}) = \prod_i p_i, \qquad \text{where} \qquad p_i = \frac{\mathcal{N}(\boldsymbol{\mathcal{M}}_i, \zeta_i \mid \boldsymbol{a})}{\mathcal{N}(\zeta_i \mid \boldsymbol{a})}. \qquad (23)
$$

The analysis in Sheth (2007) can now be followed to show analytically that this is indeed the appropriate expression for the likelihood, so we do not reproduce it here.

5 DISCUSSION AND FUTURE WORK

We presented two algorithms for reconstructing the intrinsic correlations between distance-dependent quantities in apparent magnitude-limited photometric redshift data sets. One was a generalization of the non-parametric $V_{\rm max}$ method (Section 3), and the other used a maximum-likelihood approach (Section 4).

Both our reconstruction methods assume that the distribution of photo-z errors is known accurately. In practice, this means that spectroscopic redshifts are available for a subset of the data. The question then arises as to whether or not the number of spectra which must be taken to specify the error distribution reliably is sufficient to also provide a reliable (spectroscopic) estimate of these scaling relations. If so, what is the basis for deciding that it is worth reconstructing these relations from the photo-z data? This is the subject of work in progress, although the methods presented in this paper assume that such reconstructions will indeed be necessary. For example, if the spectra are not simply a random subset of the magnitude-limited photometric sample, then it may be difficult to quantify and so correct for the selection effects associated with the spectroscopic subset.

We used the size–luminosity relation in a mock catalogue which had realistic choices for the correlation to illustrate the biases which are present and must be corrected if photometric redshift data sets are to provide reliable estimates of galaxy scaling relations (Figs 2 and 3). We showed that our iterative deconvolution scheme provides a simple and reliable correction of this bias (Figs 3–6). Note that although we have illustrated our methods using a 2D distribution, the extension to n-correlated variables is trivial.

Because our algorithm permits the accurate measurement of many scaling relations for which spectra were previously thought to be necessary (e.g. the colour–magnitude relation, the size–surface brightness relation, the photometric fundamental plane), we hope that our work will permit photometric redshift surveys to provide more stringent constraints on galaxy formation models at a fraction of the cost of spectroscopic surveys.

Our results may have other applications. For example, Bernardi (2007) has highlighted a bias associated with the correlation between stellar velocity dispersion σ and luminosity L which arises if the distance indicator used to estimate L is correlated with σ (as may happen in the local Universe, where peculiar velocities make spectroscopic redshifts unreliable distance estimators). It may be that the methods presented here would allow an accurate reconstruction of the true relation from the biased one. This is the subject of on-going work.

ACKNOWLEDGMENTS

This work was supported by NSF 0520677. RKS thanks the Aspen Centre for Physics for hospitality during the Summer of 2007, during which time he had interesting conversations with N. Padmanabhan about the pros and cons (but mainly the pros!) of regularized inversion techniques. Shortly thereafter, Paul Schechter taught us that Better is the enemy of Good.

REFERENCES

Bernardi M., 2007, ApJ, 660, 267
Bernardi M. et al., 2003, AJ, 125, 1849
Chen H.-W. et al., 2003, ApJ, 586, 745
Lucy L. B., 1974, AJ, 79, 745
Oyaizu H., Lima M., Cunha C. E., Lin H., Frieman J., 2007, ApJ, submitted (arXiv:0711.0962)
Phillips D. L., 1962, J. Assoc. Comput. Mach., 9, 84
Schmidt M., 1968, ApJ, 151, 393
Sheth R. K., 2007, MNRAS, 378, 709
Subbarao M. U., Connolly A. J., Szalay A. S., Koo D. C., 1996, AJ, 112, 929
Tikhonov A. N., 1963, Sov. Math. Dokl., 4, 1035
Twomey S., 1963, J. Assoc. Comput. Mach. 10, 97

This paper has been typeset from a TEX/LATEX file prepared by the author.