Random Galaxy Catalogues

Shaun Cole, Institute for Computational Cosmology

The dependence of galaxy clustering on galaxy properties is playing an increasingly strong role in constraining the physics and modelling of galaxy formation. For instance, simple Halo Occupation Distribution (HOD) modelling can reveal the typical masses of host dark matter haloes if one has a good measurement of the spatial two-point correlation function of a galaxy population.

The traditional method of estimating a galaxy correlation function, $\xi(r)$, is by counting pairs as a function of separation, r, and comparing to an equivalent count for a random unclustered catalogue. To determine $\xi(r)$ for galaxies with different, luminosities, colours, star formation rates or other properties one, ideally, needs a random galaxy catalogue in which the random galaxies carry all the same properties (luminosity, colour, star formation rate) as those of the genuine catalogue. Here and in astro-ph:1104.0009 we describe a procedure in which galaxies from an observational flux limited catalogue can be cloned and randomly redistributed in redshift to produce such a random catalogue. (The angular position of each galaxy can be independently randomly chosen within the angular footprint of the survey.)

A very simple approach would be to randomly distribute $N_{\rm clones}$ of each galaxy within the volume, $V_{\rm max}$, accessible to that galaxy. That is, to assign each clone a redshift distributed uniformly in the cumulative volume, $V(<z)$, between the minimum redshift limit of the survey and the maximum redshift $z_{\rm max}$ at which this galaxy would still satisfy the survey selection criteria.

As illustrated in Figure 1, this process is biased by any density perturbations in the original survey. This occurs as in a flux limited survey the luminosities of objects are correlated with their redshifts. Hence if there is an overdensity at a particular redshift then the catalogue will have an excess of galaxies of a particular range of luminosities that, in turn, will bias the redshift distribution of the random catalogue. To avoid this one must vary the number of clones produced according to the inverse of the overdensity, $\Delta(z)$, at their redshift. This requires an estimator of the overdensity in a redshift shell, $\Delta(z)$.

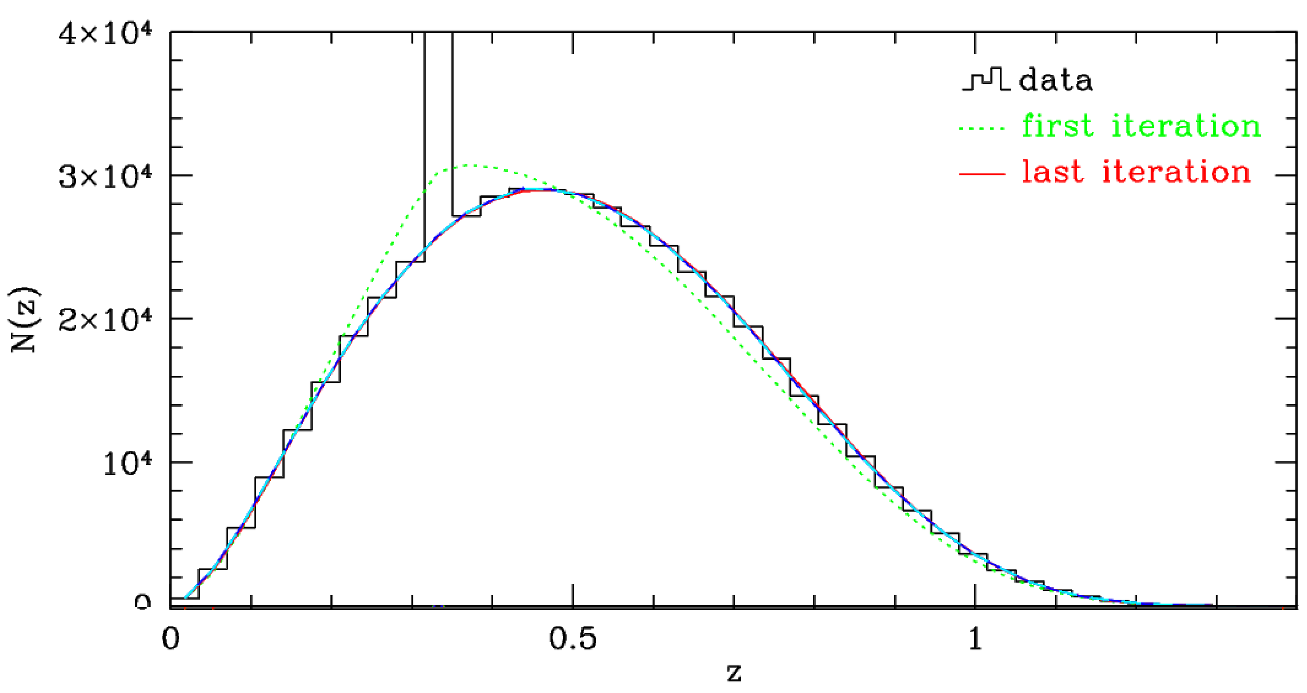

Figure 1: The black histogram shows the redshift distribution of a mock galaxy catalogue drawn an analytic distribution (blue/cyan curve) derived from Schechter LF except for an imposed overdensity in one redshift bin. The green curve shows the (biased) redshift distribution of a random catalogue produced by cloning galaxies and randomly placing them within the accessible $V^{\rm max}$ volume. The red curve, which is practically coincident with the blue/cyan, curve is the redshift distribution of the random catalogue resulting from the new algorithm.

If one first ignores redshift evolution and makes the usual assumption that the LF is independent of the large scale environment, then a simple maximum likelihood estimator of both the overdensity, $\Delta(z)$, and the luminosity function (LF), $\phi(L)$, can be obtained by maximising the likelihood

$$\mathcal{L} = \Pi_\alpha p_\alpha, \qquad \text{where} \qquad p_\alpha = \frac{\Delta(z_\alpha)\frac{dV(z_\alpha)}{dz}\phi(L_\alpha)}{\int \Delta(z)\frac{dV}{dz}\int_{L^{\rm min}(z)}^{\infty}\phi(L)\,dL\,dz} \qquad (1)$$

is the joint probability of finding galaxy, α, at redshift z_α with luminosity L_α.

The result is quite intuitive. The estimator of the LF is simply the standard "$1/V_{\rm max}$" estimate,

$$\phi(L) = \sum_\alpha \frac{1}{V^{\rm dc,max}(L_\alpha)}, \qquad (2)$$

but with the normal $V^{\rm max}$ replaced by a **density corrected**

$$V_\alpha^{\rm dc,max} = \sum_p \Delta_p V_p\, S(L_p^{\rm min}|L_\alpha) \equiv \int_{z_{\rm min}}^{z_{\rm max}} \Delta(z)\,\frac{dV}{dz}\,dz, \qquad (3)$$

which is simply the integral/sum of the accessible volume weighted by the overdensity in redshift bins ($S(L_p^{\rm min}|L_\alpha)$ is a step function which is unity if the galaxy luminosity L_α is brighter than the luminosity $L_p^{\rm min}$ corresponding to the survey flux limit at redshift bin p). The corresponding estimator of the overdensity

$$\Delta_p = \frac{N_p}{V_p \hat{\bar{n}}_p} \qquad (4)$$

is simply the actual number of galaxies in the redshift bin divided by the number one would expect, $\hat{\bar{n}}_p V_p$, based on the LF. Here V_p is the bin volume and

$$\hat{\bar{n}}_p = \sum_i \phi(L_i)\, S(L_p^{\rm min}|L_i) \equiv \int_{z_{\rm min}}^{z_{\rm max}(L_\alpha)} \Delta(z)\,\frac{dV}{dz}\,dz \qquad (5)$$

is the expected galaxy number density in redshift bin z_p computed by integrating/summing the LF over luminosities sufficiently bright to be selected at that redshift. The two estimators are coupled but are easily solved by a simple iterative scheme starting with $\Delta(z) = 1$, such that the estimate from the first iteration is the standard $1/V^{\rm max}$ result.

The red (coincident with the blue) line in Figure 1 shows the result of estimating the overdensity, $\Delta(z)$, and then generating a random catalogue by cloning each galaxy with a rate proportional to $1/\Delta(z)$ and redistributing them uniformly within their accessible volume, $V^{\rm max}$. This removes the bias and generates an random catalogue that is consistent with the known underlying distribution in this test case. As each galaxy carries the properties of the original from which it was cloned the random catalogue can be split by any of these observational properties. (For redshift dependent properties one needs to have models, e.g. k-corrections, for their redshift dependence.)

Figure 2 shows a more realistic example of the algorithm applied to a mock catalogue constructed from a semi-analytic model applied to a large N-body simulation.

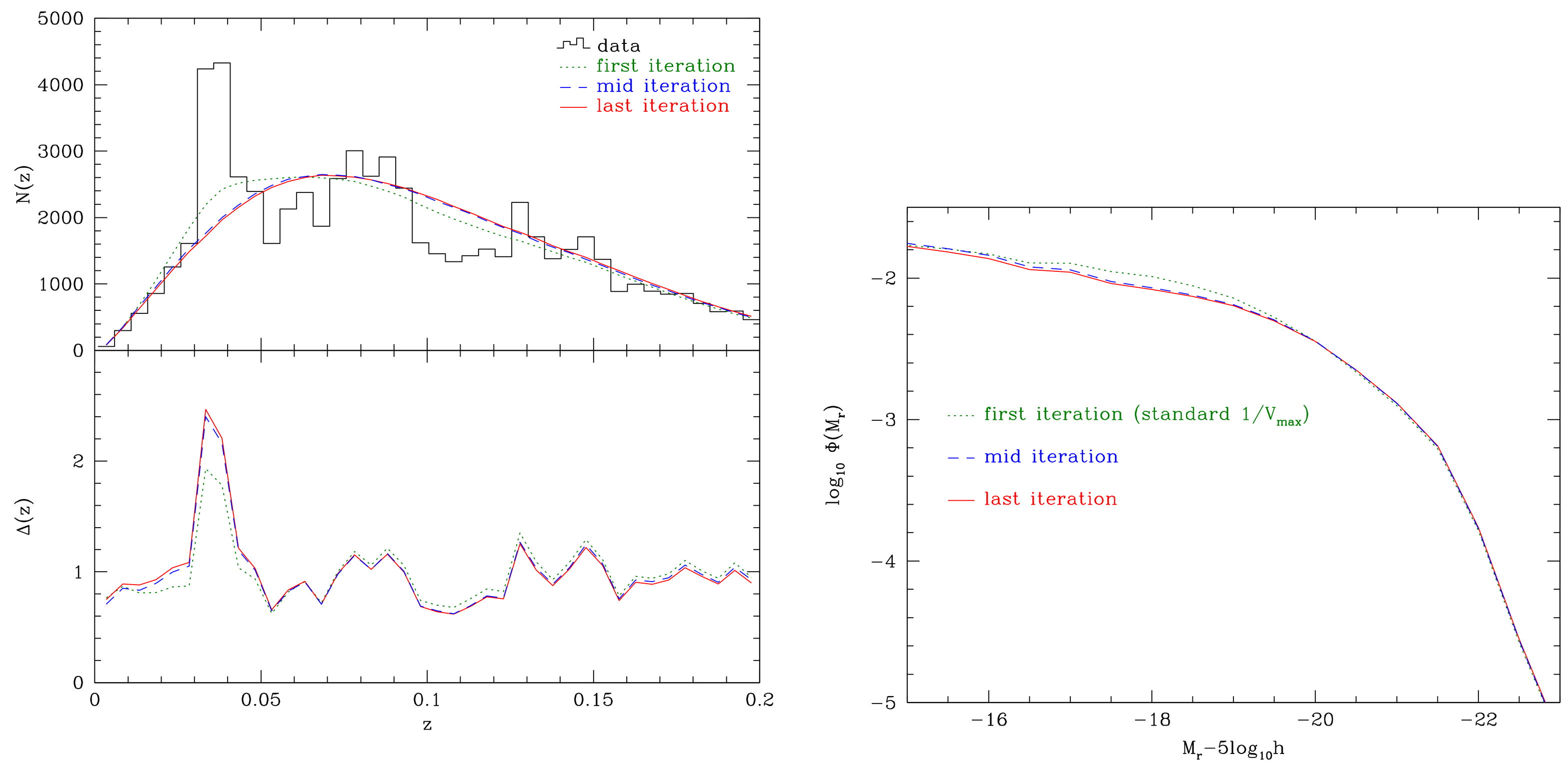

Figure 2: **Left (upper):** Comparison of the redshift distribution of the mock catalogue and that of the resulting random catalogue after several iterations. **Left (lower):** The estimated overdensity $\Delta(z)$. **Right:** The corresponding estimated LF. In all cases the green lines are the first iteration based simply on $V^{\rm max}$.

It can be seen that the method quickly iterates to a stable smooth solution which avoids biases due to the large scale over- and under-densities that are present in the original mock dataset, which if ignored (green curves) bias both the redshift distribution and the luminosity function.

To make the method applicable to deeper redshift surveys where one cannot ignore evolution one can extend the likelihood analysis to include a parametric model of the evolution of the galaxy LF. We have implemented and tested a model in which both the characteristic luminosity and number density of the LF are allowed to evolve with redshift. Evolution of the galaxy number density with redshift is degenerate with a systematic trend of the overdensity, $\Delta(z)$, with redshift. However, this degeneracy can be broken by using prior information regarding the rms magnitude of the expected density fluctuations. If the redshift bins are sufficiently large in volume we can make a simple estimate of the expected fluctuations in the galaxy overdensity using the integral $J_3 = \int \xi(r) r^2 dr$ (assumed to be a constant when integrated to scales $\gtrsim 10h^{-1}$ Mpc) of the galaxy correlation function, $\xi(r)$ (Peebles 1980). The resulting expected variance in Δ for redshift bin p is

$$\sigma_p^2 = \frac{1 + 4\pi\hat{\bar{n}}_p J_3}{\hat{\bar{n}}_p V_p}, \qquad (6)$$

with the second term enhancing the variance above the Poisson value because galaxy positions are correlated and tend to come in clumps of $4\pi\hat{\bar{n}}J_3$ galaxies at a time.

The Likelihood analysis results in a coupled set of equations which constrain the evolution parameters as well as the LF and $\Delta(z)$. These equations can again be solved by a simple iterative scheme and we provide code that implements this algorithm (http://astro.dur.ac.uk/~cole/publications.html#Software).

As a final test we show, in Figure 3, the results of applying this procedure to a deep pencil beam redshift survey. The data was constructed by sampling from a known evolving Schechter function LF and imposing known density fluctuations so that the accuracy of the reconstruction could be tested.

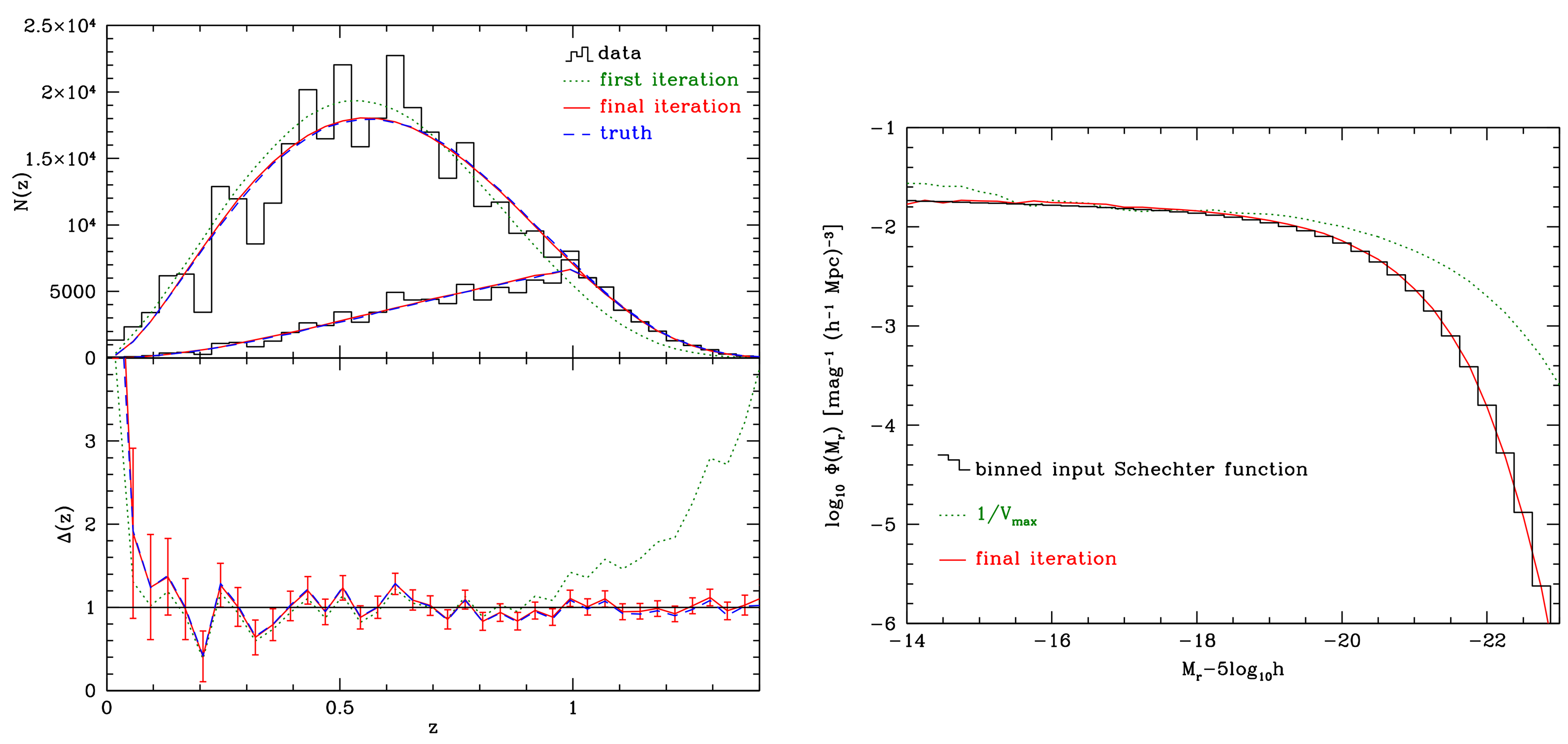

Figure 3: **Left (upper):** Comparison of the redshift distribution of a deep pencil beam mock catalogue (black) and that of the resulting random catalogue after several iterations (red). Also shown is the redshift distribution of a bright galaxy subset. **Left (lower):** The estimated overdensity $\Delta(z)$ (red) compared with the known input (blue). **Right:** The corresponding estimated LF (red) compared with the known input LF (black). In all cases the green lines are the first iteration based simply on $V^{\rm max}$ and ignoring evolution.

Summary

We have presented a maximum likelihood method of generating a random catalogue with a smooth redshift distribution that corresponds to an observed flux limited galaxy catalogue. The approach is superior to simply fitting a parametric model to the observed redshift distribution as it makes use of additional information, namely the distribution of galaxy luminosities. The algorithm works by cloning galaxies from the original catalogue and consequently produces a random catalogue in which the random galaxies have all the attributes of the galaxies in the observed catalogue. This makes the catalogue particularly well suited for use in estimating the dependence of galaxy clustering on galaxy properties. This technique should be particularly applicable to multi-wavelength surveys such as GAMA (Driver et al 2011) and its overlap with H-ATLAS (Eales et al 2010), 6dF (Jones et al 2009), zCOSMOS (Lilly et al 2007) and future redshift surveys designed to probe galaxy evolution.

References

Driver S. P. et al., 2011, MNRAS, 413, 9

Eales S. et al., 2010, PASP, 122, 49971

Jones D. H. et al.. 2009. MNRAS. 399. 683

Institute for Computational Cosmology