

Astronomical Image Processing: Galaxy Count from a KPNO Deep Optical Image

Dakshina Scott

5th November 2012

Abstract

Galaxy counts were conducted on a deep optical image taken using the 4m telescope at KPNO with a Sloan r-band filter. This was done to test the relationship between galaxy number counts and magnitude, as predicted by the equation $\log_{10}N(m) \propto 0.6m$ for an Euclidean universe.

MatLab was used to write a program to detect galaxies within the image and catalogue their brightnesses. A plot of galaxy counts against magnitude was found to differ significantly from similar surveys, and from the above equation. The steepest gradient was found at the lower magnitudes as 0.35 ± 0.02 up to magnitude 13. This is thought to be due to incompleteness in the image rather than evidence of strong galaxy evolution or about the structure of the universe.

Contents

1	Introduction	2
2	Method	4
2.1	Noise Removal	4
2.2	Saturated Pixels	4
2.3	Finding an Upper Limit for Background Brightness	5
2.4	Fixed Aperture	6
2.4.1	Finding Sources	6
2.4.2	Variable Aperture Method	7
3	Results & Analysis	8
3.1	Fixed Aperture	9
3.2	Variable Aperture	10
4	Conclusion	11
A	Calculating Errors in Magnitude	11

1 Introduction

Galaxy surveys allow us to probe the history and structure of the universe. Number counts and photometric and spectroscopic analysis from such surveys have transformed our understanding of the universe - when in 1917 Einstein's theory of general relativity suggested that the universe must be either expanding or contracting, he brushed the idea aside. It was only after Silpher and Hubble produced data from galaxy surveys relating redshift and distance that the evidence in support of an expanding universe became impossible to ignore[2].

In this experiment we looked at a deep optical image taken using a Sloan r-band filter (central wavelength 620nm) using the CCD mosaic camera at Kitt Peak National Observatory. The aim was to find the number counts of galaxies at different magnitudes, and compare our results to the prediction for an Euclidean universe of galaxies, with no evolution, described by equation 8. We also compared our counts to similar surveys in order to evaluate our cataloguing procedure.

By assuming an Euclidean universe of uniformly distributed galaxies, with a given distribution of luminosities, a simple relationship between magnitude and number of galaxies brighter than that magnitude can be derived. For a given luminosity distribution $f \propto \frac{1}{r^2}$ and so

$$r \propto \frac{1}{f^2} \quad (1)$$

and for a given density in Euclidean space

$$N \propto r^3 \quad (2)$$

Substituting equation 1 into equation 2 we get

$$N(f) \propto f^{-\frac{3}{2}} \quad (3)$$

where $N(f)$ is the number of sources with flux greater than f .

Using

$$m_1 - m_2 = -2.5 \log_{10} \left(\frac{f_1}{f_2} \right) \quad (4)$$

which is the standard equation relating the difference in magnitude between two objects and the flux received from them, it can be seen that

$$m \propto -2.5 \log_{10}(f). \quad (5)$$

Rearranging for f

$$f \propto 10^{-2.5m} \quad (6)$$

and substituting equation 6 into equation 2

$$N(m) \propto 10^{0.6m} \quad (7)$$

Finally, taking the log of both sides

$$\log_{10}(N(m)) = 0.6m + C \quad (8)$$

where $N(m)$ is the number of sources brighter than magnitude m and C is a constant.

Based on this result, the difference in the number of galaxies between each magnitude is given by

$$N(m+1) = 4.0N(m) \quad (9)$$

where m was the lower magnitude.

However, the Friedmann equations tell us that the curvature of the universe is dependent on its density - it's not necessarily Euclidean, so there may be a distance beyond which the structure of the universe influences galaxy counts.

In 1961 Sandage [8] found that galaxy counts from the Hale telescope (the world's largest telescope at the time[9]) were unlikely to reveal information about the structure of the universe because the differences between flat, open and closed models are too small compared to known variations in galaxy distribution. Advances have since been made in astronomical image acquisition, with CCDs allowing much greater quantum efficiency than photographic plates. Eisenstein et al (2001)[6] found that the Sloan Sky Digital Survey, which uses CCDs, would be well suited to studies of large-scale structure, but that makes use of spectroscopic data that was unavailable here.

One of the key results looked at in galaxy surveys (eg Maddox et al. 1990)[4] was how the data deviates from the counts expected by a no-evolution model of galaxies - the further away the galaxy, the further in the past we are looking - so if for example galaxies were brighter/dimmer or more/less numerous, then distant galaxies may give us results inconsistent with equation 8.

Thus it was possible that comparing number counts against magnitude with the prediction of equation 8 will show evidence of galaxy evolution, but it was unlikely that effects of large-scale structure will be seen.

2 Method

Although the image had been processed to adjust for certain effects before it was received, there were still some unwanted effects remaining. Image processing and analysis was done in Matlab, with the image processing toolbox installed.

2.1 Noise Removal

The image consists of a number of sub-exposures, combined to create one image. This smooths out random fluctuating noise, as inconsistent sources will be faded in the final image. However, it can be seen that these sub-exposures don't overlap exactly - as a result there are blank areas in the corners of the image, and around the edges a sharp increase in noise is visible (see figure reffig:edges).

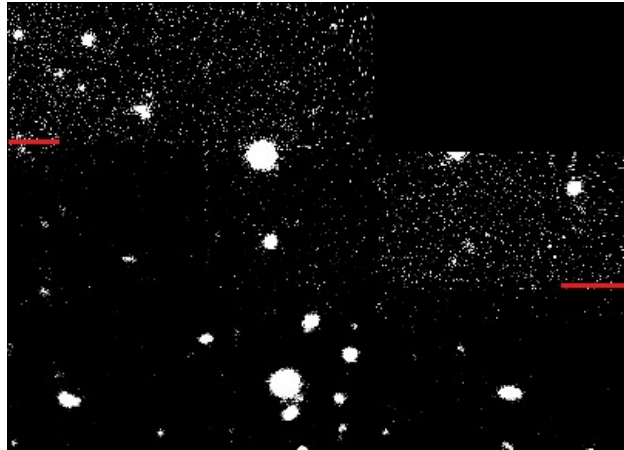


Figure 1: The red lines highlight sharp increases in noise where sub-exposures don't overlap. The top right corner is blank.

The image was cropped before analysis to remove the noisiest areas.

2.2 Saturated Pixels

'Blooming' occurs where pixels in the CCD are saturated and overflow to nearby pixels, something which tends to happen mostly in the vertical and horizontal directions, as is seen in figure 2.

The maximum number of electrons each pixel of the CCD can hold is 65,535. For values approaching this the relationship between number of photons received and the number of electrons stored moves further and further from linearity. By using multiple shorter exposures any given CCD pixel receives fewer photons, so the number of sources bright enough to cause overflowing is reduced. However this effect is still visible in some sources (highlighted in figure 2). These sources and their bloom were blacked out before analysis.

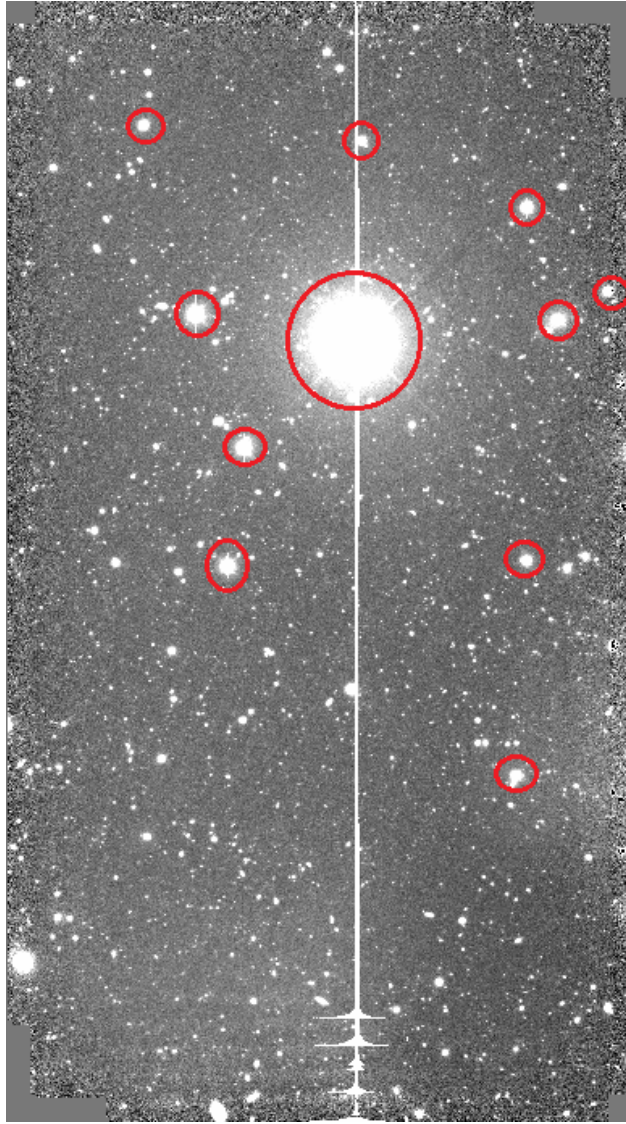


Figure 2: A bright central source which has 'bloomed' vertically across the image. There are also smaller sources showing the same effect to a lesser extent. Red circles indicate sources which were blacked out prior to analysis. The lines of bloom were also blacked out.

A filter was also applied, so that any remaining sources with a maximum pixel value exceeding 50,000 counts are excluded from the survey.

On this fully prepared image, two different algorithms for source detection and photometry are applied, each with different limitations. One method uses a fixed aperture, while the other uses the shape and size of each individual source to analyse the correct areas of the image. Both methods are dependent upon finding a suitable value for the source/background threshold.

2.3 Finding an Upper Limit for Background Brightness

In order to analyse each source in the image, some criterion must be used to determine how high a photon count must be in order for it to be considered to be coming from a source, rather than a background fluctuation. The image was composed mostly of background, with a scattering of sources of varying brightness. Thus finding the mode pixel intensity value of the image would be one way to estimate the background value - doing so produces a value of 3421 counts. However, as the background isn't a constant value, it's clear that much of the background will have values higher than this, and some upper limit of counts must be chosen as the threshold. Inspecting the

distribution of photon counts within the image shows that the frequency of counts per pixel in the region of this modal value was many orders of magnitude greater than that for other values found in the image, and the distribution within this area was approximately Gaussian (see figure 3). Thus by finding the mean value within just this bell-shaped peak, and the standard deviation, a threshold value was chosen as three standard deviations above the mean (at a value of 3461 counts).

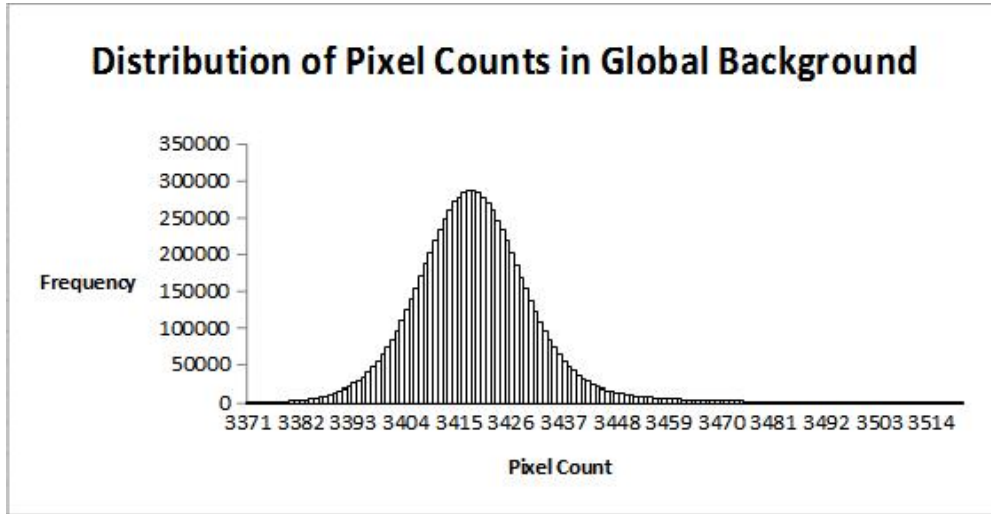


Figure 3: Histogram showing the Gaussian-like distribution of pixel counts in the region of the global background.

2.4 Fixed Aperture

2.4.1 Finding Sources

Initially sources are identified and analysed using a fixed aperture size. The maximum pixel value in the image was found, and the surrounding area analysed. This area was then blacked out so that a search for the new maximum will find the second-brightest source from the original image, which was then analysed, and so on. This continues looping around until the maximum remaining value equals the threshold value.

A number of different aperture sizes are tried in an attempt to find a compromise between a very large aperture and a very small aperture - any source larger than the chosen aperture size will end up with a lower count attributed to it than it should have, reduced even further when the 'local background' (which in this case will at least partially be obscured by the source) was subtracted. Also, the smaller the aperture the more opportunity there was for each source to be counted as multiple sources (although measures are taken within the code to reduce this - see figure 4). On the other hand, sources which are smaller than the aperture will have a reduced mean intensity value, although the total number of counts should be about right once the local background was subtracted.

As the aperture size was increased, more than one source was likely to be contained within a single aperture. Thus an appropriate size would be a compromise. The mean value of the source areas was 51 pixels, suggesting an aperture radius of 4 pixels may be a reasonable compromise. However, as some of the sources are much larger than this, the number counts will be skewed - if lots of apertures fit inside one large source, this will give rise to a jump in the frequency distribution in the range of brightnesses contained in the source.

For each source, an aperture of recommended radius 6 pixels[1] was used as a starting point (3 arc seconds - do calculation?) [reference (arc seconds per pixel)], centred on the maximum value. A larger-radius annular aperture was then placed around this to calculate the background level

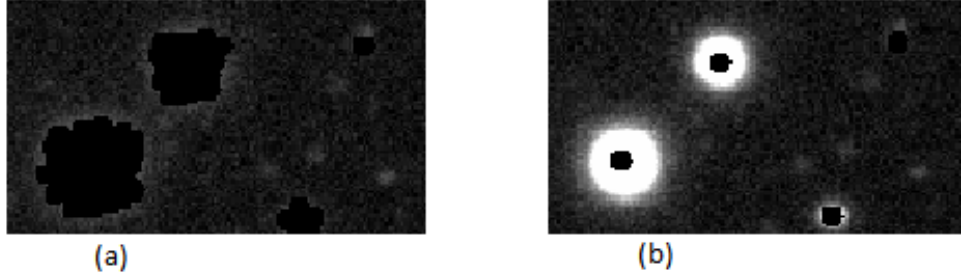


Figure 4: Black circles represent areas that have been treated as sources by the detection program. (a) shows those detected before code was added telling the program to ignore aperture areas which already contained a zero value. (b) shows areas of source detected after this section of code was added. As can be seen, there are some sources which aren't contained by the aperture.

local to each source, which was then subtracted from the source aperture value. Different radii are compared, including a radius of 4 pixels.

2.4.2 Variable Aperture Method

Two methods were developed using a variable aperture size to analyse the sources. Of the two, the first has some significant issues - in cases when the multiple sources are within a few pixels of each other, it's possible that the same source will be analysed multiple times, or it may be skipped altogether. The second version achieves more reliable results. By making better use of inbuilt Matlab functions, the new program was much more effective at quickly detecting and analysing sources (it runs in about 3 seconds). It turns out that both methods appear to produce quite similar results, but only the latter was discussed from here on, as it most accurately represents the data available.

'Thresholding' was used to segment the image based on pixel counts - the image was converted to a binary image, in which any pixels with a count higher than the source/background threshold (as determined above) become ones, the rest zeros. By assigning distinct labels to each separate region of ones, Matlab was then able to run through them sequentially, analysing each corresponding source & local background as it goes. An issue with this method was that sometimes part of the source was 'disconnected' from the main section (as can be seen in the top left and bottom right sides of the binary image in figure 5).

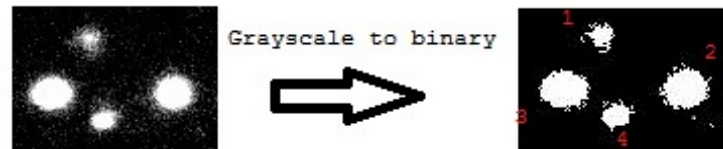


Figure 5: A section of the image in grayscale, and the same section in binary, showing that a good approximation of the source shape and area is obtained. However, notice the small white areas near sources 1 and 2. These are counted as separate sources, regardless of whether they are legitimate sources or disconnections from nearby sources.

These disconnections will lead to an increase in counts of faint sources.

Another issue arises with finding the local background for each source. To do so an area surrounding each source was looked at. If multiple sources are close together and their 'local background areas' overlap then the same local background, averaged over the regions local to each source, was used for all of these sources. (image to illustrate?)

Despite this, the variable aperture method was the preferred method as it always analyses all of the sources in the image (so long as the threshold allows them all to be detected), and unlike the fixed aperture it analyses the actual shape and area of each source (again limited by the accuracy of the threshold used).

3 Results & Analysis

First the results of the fixed aperture method are looked at, with different aperture sizes compared. Then the variable aperture method was looked at in more detail and finally the two sets of results are briefly compared. The raw photon counts are converted to magnitudes using the equation,

where ZP_{inst} was the instrumental zero point, as determined in the AB magnitude system (where a flux density of 3631 Jy was defined to be zero magnitudes). According to the fits header for the image the maximum reliable value for photon counts per CCD pixel was 36000 - for a one-pixel source this translates to a magnitude of 14. Although we have excluded single-pixel sources, this suggests that for magnitude 14 and brighter, more sources will have unreliable results, as determined by the limitations of the equipment used. This should be kept in mind when looking at the results.

3.1 Fixed Aperture

Number count against magnitude was plotted for various aperture radii. The 4-pixel radius appears to produce the straightest graph. This could imply that the 'wobbles' in the other graphs are due to the limits of a fixed aperture rather than a physical phenomenon.

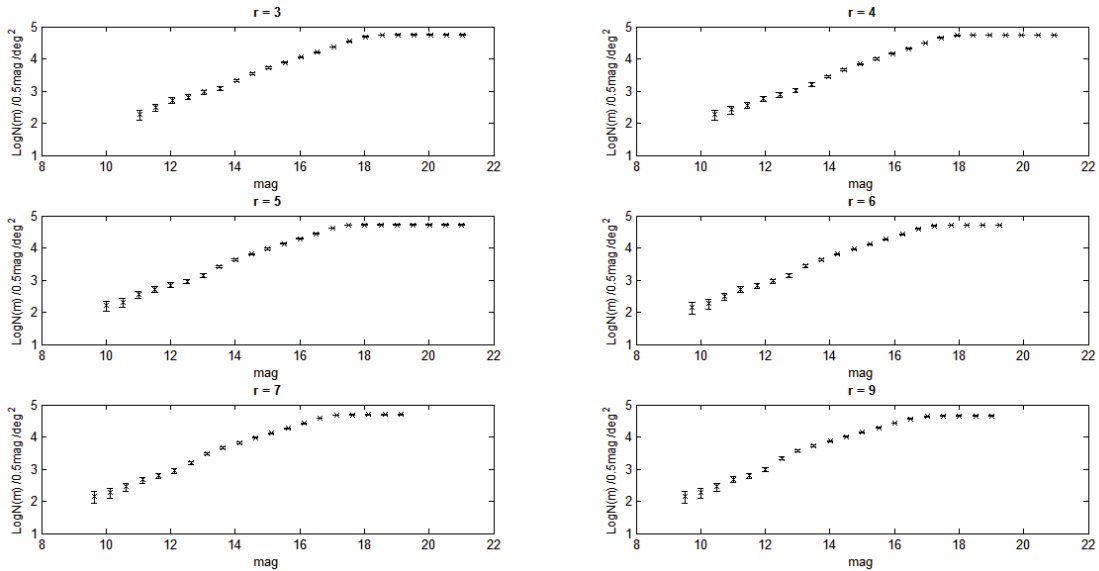


Figure 6: An aperture radius of 4 pixels has the 'straightest' graph. As the radius diverges from this the data points become more and more 'wobbly'. In all cases, just below magnitude 14 there was a dip in the graph.

3.2 Variable Aperture

A variable aperture method is able to much more accurately collect data on each source in the image.

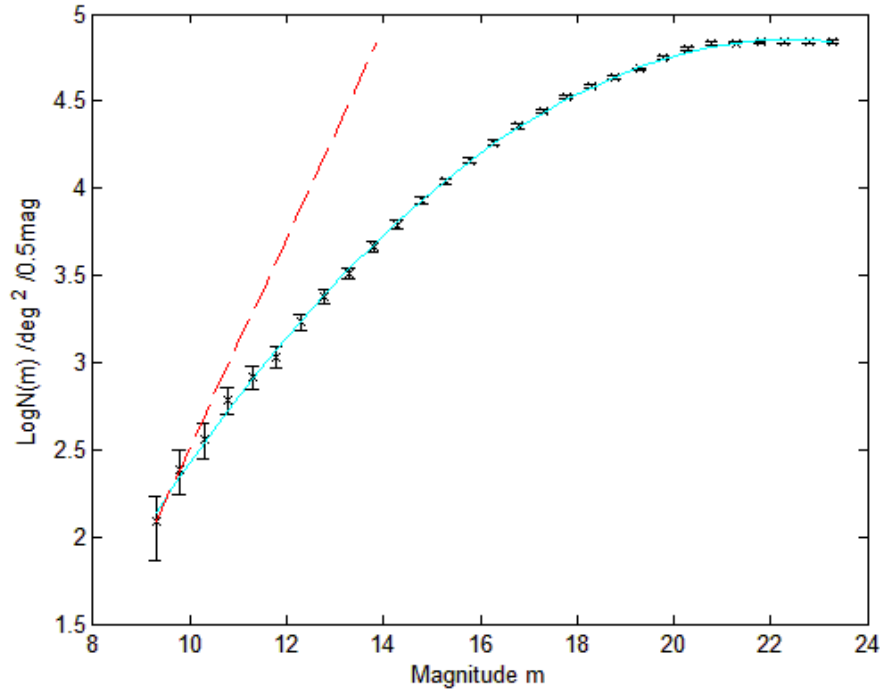


Figure 7: The errors shown are the errors in the number counts of galaxies following a poisson distribution. The largest error in the x-axis is 0.023. This is a combination of the calibration error on the apparatus used and the error in photon counts following a poisson distribution. The solid line shows a quadratic best fit curve. The dashed line shows a gradient of 0.6. The errors in magnitude aren't shown as they are too small to show up on the graph (see appendix for more details).

This graph shows much more curvature than the fixed aperture graphs. This was expected as galaxies become gradually more difficult to detect at higher magnitudes, so the gradient represents the gradual increase of incompleteness in the survey. Excluding the sources which contain a pixel value of over 36000 (the maximum good data value in the fits header) has very little effect - only 14 sources are excluded and the resulting graph looks the same.

The gradient of the graph up to magnitude 13 was 0.34 plus/minus 0.02. As can be seen by the curvature of the graph this gradually decreases, reaching about 0.1 between 18 and 20, and it was finally zero (to one decimal place) between 21 and 23. On the other hand a quadratic line of best fit, as shown in the figure, fits the data very closely.

Even the maximum gradient of about 0.4, at the very start of the graph was much lower than the gradient of 0.6 predicted by equation (number). A number of factors will likely have combined to result in this low value. If the survey was complete, there would be an increase in the galaxy counts at higher magnitudes.

Yasuda[5] et al (2001) present galaxy surveys with very high levels of completeness, in a variety of bands including r-band. At lower magnitudes, where effects of evolution and large-scale structure aren't significant, their number counts fit equation (number) well. This suggests that our data may suffer from incompleteness - and that this effect worsens at increasing faintnesses. The gently decreasing gradient fits with this suggestion, as incompleteness would be expected to increase gradually with magnitude. Removing the brightest sources (which exhibited a blooming effect) also may have affected the gradient - the brightest sources may have decreased the starting value on x-axis.

4 Conclusion

As discussed the variable aperture is believed to be much more reliable and accurate in terms of detecting and collecting data on each source in the image. While the program using a fixed aperture gives results that are closer to those predicted by theory and found by other surveys, it is suspected that this is in fact misleading - that perhaps the unquantified systematic errors in the fixed aperture actually caused the data to appear closer to the predictions than the true data should be. This fits the results found for the variable aperture, as it suggests high levels of incompleteness starting even at the brighter magnitudes. Using a larger number of shorter exposures would improve the results for the brightest sources as less would need to be blacked out before analysis. However, this would reduce completeness at the fainter magnitudes. Alternatively using a better CCD (e.g. with a higher bit counter) would vastly improve the results - long exposure times could be used allowing fainter sources to be detected more reliably, without causing so many pixels to saturate and bloom (so the results would also be more reliable for brighter sources). Another issue was that the detecting programs can't distinguish between a) stars and galaxies, and b) two overlapping sources. Both issues can be addressed by looking at more detailed photometry for each source - for example, Yasuda et al (2001)[5] use a program called 'PHOTO' which looks for multiple peaks in each object, and 'deblends' them into multiple 'child' objects. This program also able to determine whether a source was likely to be a star or a galaxy based on its surface brightness profile (as described by Blanton et al, 2001[7]).

A Calculating Errors in Magnitude

The errors in magnitude are excluded from the graphs as they are very small - and in fact, none are large enough to affect the number of galaxies in each magnitude bin. They are approximated by treating the photon counts for each galaxy as a poisson distribution, with the error as the square root of the counts. These errors in counts are then converted into errors in magnitude, and finally added in quadrature with the zeropoint error included in the fits header file, to give the overall errors in magnitude for each galaxy. The largest error was 0.0227. As we are working with magnitude bins of 0.5, it's meaningless to talk about errors to 4 decimal places. To two decimal places, all of the errors are equal to 0.02, the value given for the zeropoint error.

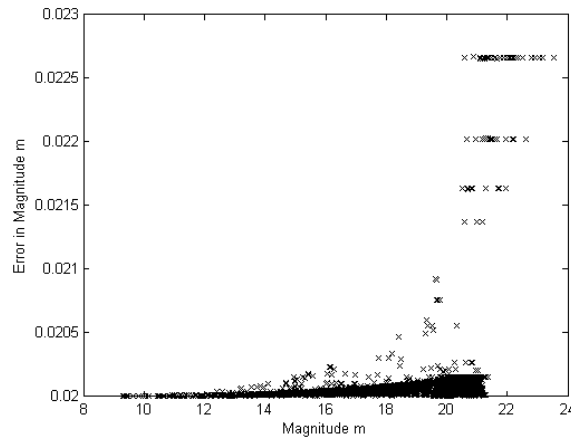


Figure 8: Distribution of errors with magnitude shows that the errors are greatest for the faintest galaxies, but all of the errors are very small.

References

- [1] Clements, D. L., 2012, Computational Image Processing: A Deep Galaxy Survey [Lab script], 2012 ed., Imperial College London
- [2] Raine, D. J., Thomas, E. G., 2002, An Introduction to The Science of Cosmology, 2nd ed., Bristol: Institute of Physics Publishing, p. 60
- [3] Peacock, J. A., 1999, Cosmological Physics, Cambridge: Cambridge University Press, p. 73
- [4] Maddox, S. J. et al, 1990, Galaxy Evolution at Low Redshift, Mon. Not. R. astr. Soc., 247, pp.1-5
- [5] Yasuda, N., et al. 2001, Galaxy Number Counts From The Sloan Digital Sky Survey Commissioning Data, A.J., 122, pp. 1104-1124
- [6] Eisenstein, D. J., et al. 2001, Spectroscopic Target Selection For The Sloan Digital Sky Survey: The Luminous Red Galaxy Sample, A.J., 122, pp. 2267-2280
- [7] Blanton, M. R., et al. 2001, The Luminosity Function Of Galaxies In SDSS Commissioning Data, A.J., 121, pp. 2358-2380
- [8] Sandage, A., 1961, The Ability of the 200-Inch Telescope to Discriminate Between World Models, Amer. Astr. Soc., 133, pp.335-392
- [9] Caltech Astronomy, 2012. The 200 Inch Hale Telescope. [online] Available at <http://www.astro.caltech.edu/palomar/hale.html> [Accessed 4 November 2012].