

Galaxy Survey

Aidan Boxall

1st May 2015

Abstract

The aim of this investigation was to produce a catalogue of galaxies from an image taken by the Spitzer Space Telescope. From this catalogue the number of galaxies were plotted against apparent magnitude to see if the relationship matched that expected if the galaxies were uniformly distributed with a given flux distribution. The relationship between galaxy population and flux density was found to be similar but not entirely consistent with the predicted relationship. A number of possible explanations are considered but the most likely explanation is that the discrepancies are due to the small sample size of galaxies investigated.

Introduction

The aim of this investigation was to produce a catalogue of galaxies from the image shown in Figure 1 taken by the Spitzer Space Telescope. The catalogue produced includes the position, count and magnitude of the galaxies in the image. The catalogue was produced using a python script. A number of different algorithms were investigated to determine the number of galaxies and their counts.

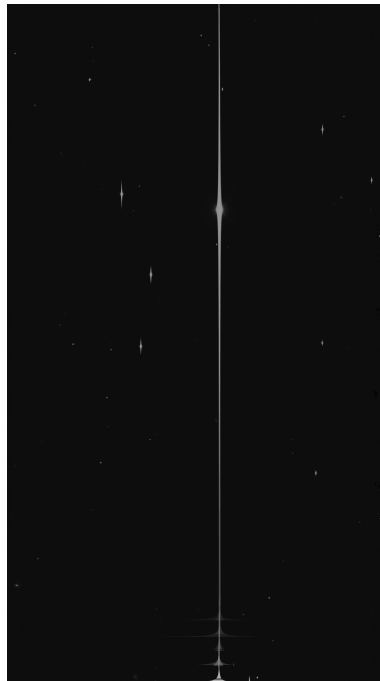


Figure 1: The image used as the data source for this investigation.

The galaxy magnitudes were used to investigate the relationship:

$$\log(N(m)) = 0.6m + \text{constant} \quad (1)$$

where $N(m)$ is the number of galaxies below an apparent magnitude m . This relationship was derived by making several assumptions about galaxies. It assumes a uniform distribution of galaxies with a given distribution of flux and uniform density. A full derivation can be found in the appendix.

The image was taken using a charge coupled device (CCD) camera. A CCD produces pixel counts which are linearly related to the number of photons absorbed by the device [1]. This makes it straightforward to locate the brightest points in the image and find an average background count.

Experimental Method

Background Count

The first aim of the experiment was to find a measure of the background count. The background count was needed for two reasons. Firstly it was necessary as it gave a lower limit to the pixel count that could be considered a galaxy. The minimum pixel count of a galaxy was determined by the average background count plus a number of standard deviations. The number of standard deviations to use was determined experimentally to be five. This gave the best compromise between not discarding too many galaxies with low counts and not counting background noise as galaxies. The background count was found by producing a histogram of all the pixel counts in the image, which is shown in Figure 2. From this histogram a peak could be seen in the low count range corresponding to the background. A second histogram shown in Figure 3 was produced focusing on this region. The mean and standard deviation of the background count were calculated and a Gaussian distribution with the same mean and standard deviation is shown in Figure 3. This shows that the background noise is approximately Gaussian distributed.

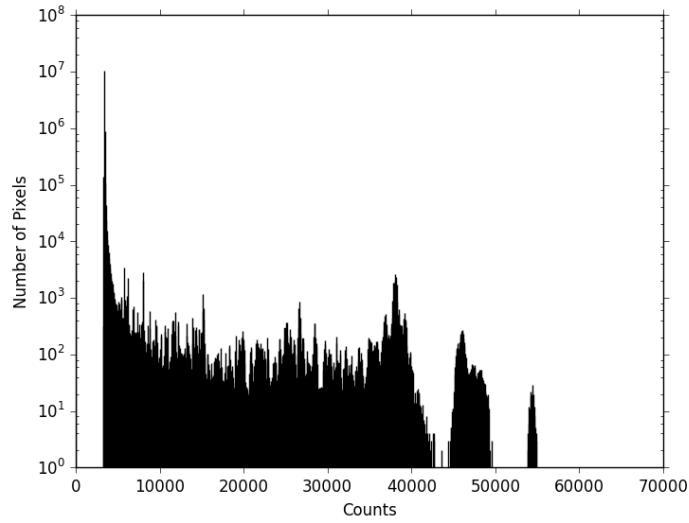


Figure 2: A histogram of the pixel values of the entire image. There is a clear peak at around 3500 counts which corresponds to the background noise in the image.

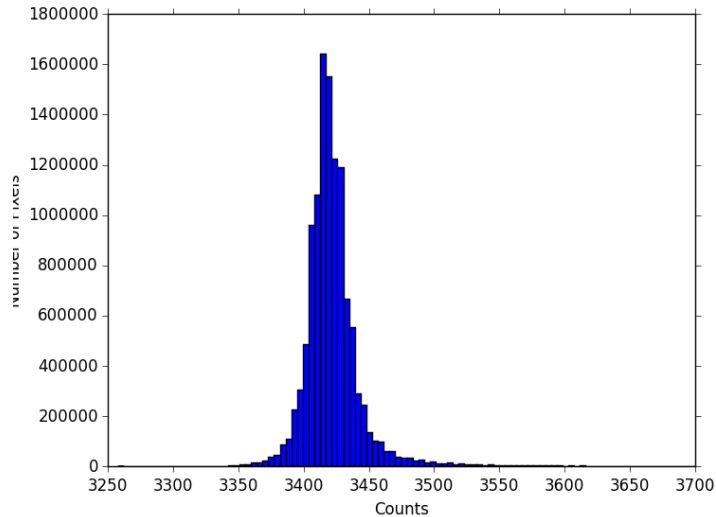


Figure 3: A histogram of the background noise pixel counts.

Masking Stars and Fixing Dead Pixels

Once the background count had been determined the next stage was to determine an algorithm for cataloging galaxies. Galaxies were located by successively finding the brightest pixel in the image, taking counts within a circular aperture around the pixel and then marking that galaxy as counted in a mask image. This method required the stars and bleed pixels to be masked out so that they would not be counted as galaxies. Bleed pixels occur

when a bright source causes a cell in the CCD to overflow with electrons into neighboring pixels [1]. The stars and surrounding bleed pixels were found by eye and masked as shown in Figure 4. The image also contained a number of dead pixels, pixels with a count of zero. These were potentially problematic as if they were included in a galaxy count they would make the count lower than it should be. These pixels were dealt with by replacing them with the average count of the neighboring pixels.



Figure 4: A copy of the original image with the masked "bad" regions removed.

Algorithms for Finding Galaxies

Several methods were investigated to determine the aperture radius. The first method was to simply have a fixed aperture radius. The problem with this method was that if the aperture was too large multiple galaxies were counted as a single source. If the aperture was too small then a single large galaxy was counted as several galaxies. Therefore there was no ideal fixed aperture size and it was determined that a variable aperture was required. The first variable aperture method investigated was as follows. After the bright pixel at the center of the galaxy had been found counts of the pixels in eight directions were measured until the counts went below the minimum count. This gave eight possible radii. The largest was chosen to be the aperture radius. This method was quite successful apart from when there were two or more galaxies close together. When this happened the count did not go below the minimum until the radius covered several galaxies and they were counted as a single galaxy. A diagram showing a group of galaxies which caused this problem is shown in Figure 5. The first attempt to fix this was to limit the aperture radius to a maximum size. The problem with this is unless there isn't a radius that can be chosen that fixes all instances of this error and is still bigger than the biggest galaxy, If the maximum aperture size is smaller than the biggest galaxy then then it takes several apertures to cover a single galaxy as shown in Figure 6.



Figure 5: Figure shows how two galaxies close together could be counted as one if the algorithm allows for large apertures. The aperture shown was not an aperture produced by the algorithm but is similar to apertures produced by the algorithm.

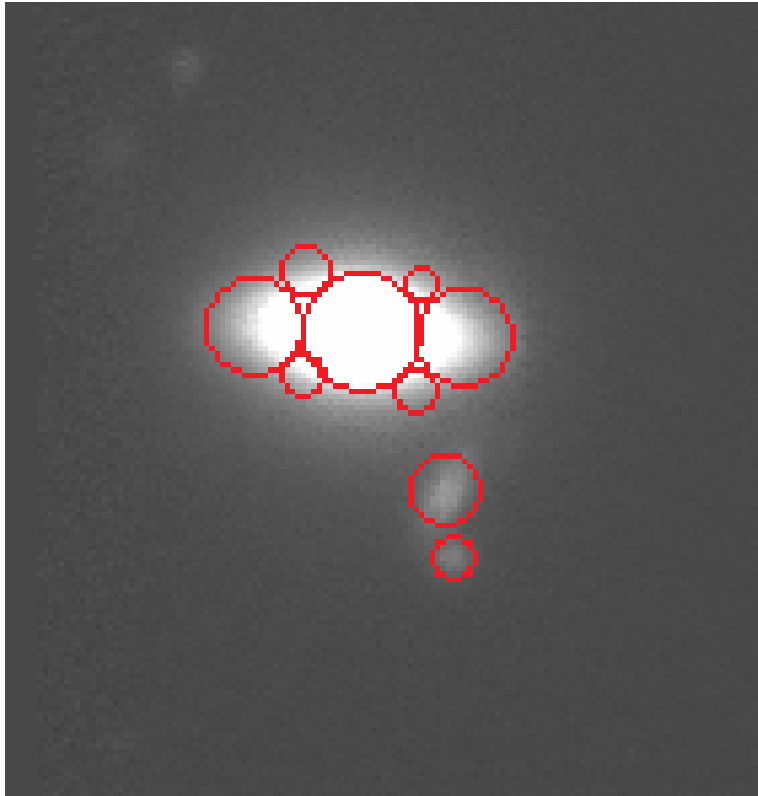


Figure 6: Figure shows how a large galaxy could be counted as many if the algorithm is designed to not miss the smaller galaxy next to it. The apertures shown are not apertures produced by the algorithm but is similar to apertures produced by the algorithm.

To fix these problems a new algorithm was designed. Around the central bright pixel disks of counts were taken and averaged. The radius was chosen to be either when the average count of the disk went below the minimum count or when the count stopped decreasing at a rate which was determined experimentally.

The average background count was taken from the galactic counts to give the true count of the galaxies. The use of local background counts was considered to take into account the varying background counts across the image but in the final data only a universal background was used for two main reasons. Firstly as shown by Figure 3 the local background count is normally distributed with a small standard deviation. The background count is 3423 ± 26 , this means the variation is about the mean is not significant compared to the galactic count as the variation is at most around 1% of the count even for the faintest detectable galaxies. Further variation in the background flux could be caused by faint galaxies being detected next to bright sources. Taking local background counts was considered for this reason but when investigated it caused further issues. Local backgrounds were counted by counting pixels in a disk around the galaxy. Often this disk included other galaxies or stars. This made it possible for the disk to have a similar and sometimes greater average count than the galaxy itself. This made it possible for the algorithm to return galaxies with negative counts.

Results, Errors and Discussion

The aim of the experiment was to measure the number and magnitude of galaxies within an image to catalogue them and see if they obeyed equation (1). Figure 7 shows that the galaxy count does have a linear relationship on a logarithmic plot. The gradient is slightly different though to the gradient predicted by equation (1). The measured gradient is 0.650 ± 0.002 which is slightly different to 0.6. This could be due to a number of factors. The most significant factor is in this experiment only a small area of sky containing a relatively small number of galaxies was measured. For comparison there were 1455 galaxies found in the image used but there were 900,000 galaxies measured in the Sloan Digital Sky Survey [2]. The relationship described by equation (1) requires a large homogeneous sample of the sky. This is because on smaller local scales such as this one the homogeneousness of this area of the sky may not be complete. Further difference may be due to galaxy evolution. Due the finite speed of light looking at galaxies very far away is equivalent to looking at them in the past. In the distant past the flux produced by the galaxies may have been different which would contradict the assumption that the flux distribution of the galaxies is constant. The sample of galaxies in this survey is too small to determine if variations in the results are due to galaxy evolution or statistical variations.

The error in galaxy count is the square root of the count as the count obeys Poisson statistics. This means there is a more significant chance of there being less bright (low magnitude) sources than expected. This is shown in the graph by the non-linear green part below a magnitude of around 10. The error on the galaxy count is the square root of the count as the counts are Poisson distributed.

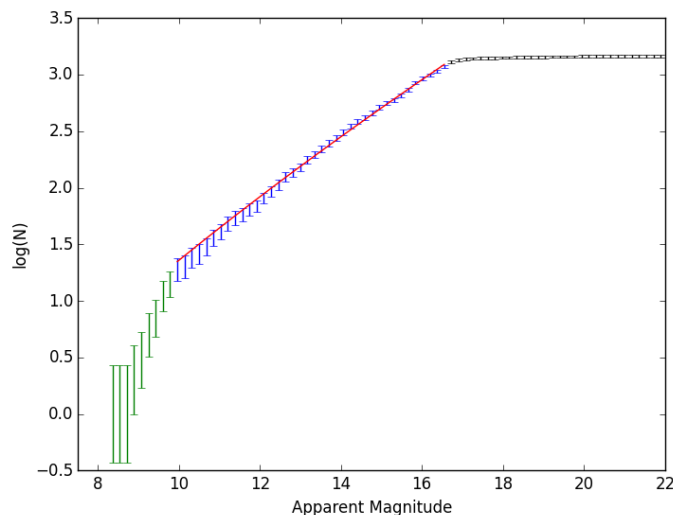


Figure 7: A plot of the log of the number of galaxies with an apparent magnitude m below m vs. apparent magnitude. The gradient of the straight section is 0.650 ± 0.002 . The apparent magnitudes of the galaxies have been calculated relative to the AB magnitude system.

The flattening of the graph above a magnitude of around 16 is due to incompleteness of the survey. As the galactic flux gets lower the chance of it being detected falls as it becomes indistinguishable from the background flux in the image. Eventually above a certain magnitude no further galaxies can be detected and the graph remains flat.

The catalogue produced contains among other information the location and magnitude of all 1455 galaxies found in the image. An extract from the catalogue is shown in Figure 8 in the appendix.

Conclusion

A catalogue of 1455 galaxies was successfully produced. From the catalogue a graph of the number of galaxies vs absolute magnitude was produced. The gradient of the graph was similar but not in agreement within error to the predicted value. This is most likely to be due to the small sample of galaxies in this image. To achieve more consistent results a larger sample size of galaxies would be required such as in the Sloan Digital Sky Survey [2]. The results could be further improved by improving the algorithm used to detect galaxies, for example by using elliptical apertures.

References

- [1] Dr. D.L. Clements and N. Skrzypek. *Computational Image Processing: A Deep Galaxy Survey*. Imperial College London, 2013.
- [2] Masataka; Narayanan Vijay K.; Lupton Robert H.; Strateva Iskra; Strauss Michael A.; Ivezi eljko; Kim Rita S. J.; Hogg David W.; Weinberg David H.; Shimasaku Kazuhiro; Loveday Jon; Annis James; Bahcall Neta A.; Blanton Michael; Brinkmann Jon; Brunner Robert J.; Connolly Andrew J.; Csabai Istvn; Doi Mamoru; Hamabe Masaru; Ichikawa Shin-Ichi; Ichikawa Takashi; Johnston David E.; Knapp G. R.; Kunszt Peter Z.; Lamb D. Q.; McKay Timothy A.; Munn Jeffrey A.; Nichol Robert C.; Okamura Sadanori; Schneider Donald P.; Szokoly Gyula P.; Vogeley Michael S.; Watanabe Masaru; York Donald G. Yasuda, Naoki; Fukugita. Galaxy number counts from the sloan digital sky survey commissioning data. *The Astronomical Journal*, 112:1104–1124, 2001.

Appendix

Derivation of Equation (1)

The derivation assumes the galaxies are uniformly distributed in space so the number of galaxies $N(r)$ in a given radius is:

$$N(r) = \frac{4}{3}\pi r^3 \rho \quad (2)$$

where ρ is the density of galaxies. Magnitudes are always measured against a reference flux f_R . Magnitudes are given by:

$$m = -2.5 \log\left(\frac{f}{f_R}\right) \quad (3)$$

where f is the flux received from the galaxy. As the flux distribution is assumed to be fixed the expectation value the total flux produced by a galaxy f_G is constant. The flux f is given by:

$$f = \frac{f_G}{r^2} \quad (4)$$

Therefore by subbing (4) into (3):

$$m = -2.5 \log\left(\frac{f_G}{f_R r^2}\right) \quad (5)$$

$$m = 5 \log(r) - 2.5 \log\left(\frac{f_G}{f_R}\right) \quad (6)$$

$$m = 5 \log(r) + c' \quad (7)$$

where $c' = -2.5 \log\left(\frac{f_G}{f_R}\right)$ is a constant. Rearranging (7):

$$r = 10^{\frac{m}{5} - c'} \quad (8)$$

Then subbing (8) into (2):

$$N(m) = \frac{4}{3}\pi 10^{\frac{3m}{5} - 3c'} \rho \quad (9)$$

Then taking the log:

$$\log(N(m)) = \log\left(\frac{4}{3}\pi 10^{\frac{3m}{5} - 3c'} \rho\right) \quad (10)$$

$$\log(N(m)) = \log(10^{\frac{3m}{5} - 3c'}) + \log\left(\frac{4}{3}\pi \rho\right) \quad (11)$$

$$\log(N(m)) = \frac{3m}{5} - 3c' + \log\left(\frac{4}{3}\pi \rho\right) \quad (12)$$

$$\log(N(m)) = 0.6m + c \quad (13)$$

where $c = -3c' + \log\left(\frac{4}{3}\pi \rho\right)$ is a constant. Equation (13) is identical to equation (1) as required.

Centre	Radius	Count	Count Error	Magnitude	Magnitude Error
(657, 2530)	3	61408	247	13.33	0.02
(3443, 2562)	3	61943	248	13.32	0.02
(3600, 2535)	3	61689	248	13.32	0.02
(4548, 1054)	3	61988	248	13.32	0.02
(30, 1692)	5	1036273	1017	10.26	0.02
(1495, 634)	16	2219740	1489	9.43	0.02
(1938, 675)	14	1839211	1356	9.64	0.02
(2300, 446)	16	2389080	1545	9.35	0.02
(578, 1772)	16	2392490	1546	9.35	0.02
(4273, 55)	16	1846498	1358	9.63	0.02
(3923, 188)	14	1287143	1134	10.03	0.02
(2263, 704)	14	1548444	1244	9.83	0.02
(662, 65)	31	5998917	2449	8.35	0.02
(1115, 2522)	14	1566317	1251	9.81	0.02
(1654, 966)	14	1267580	1125	10.04	0.02
(4400, 1312)	15	2168704	1472	9.46	0.02
(4330, 1365)	14	1534698	1238	9.83	0.02
(3373, 894)	14	1295880	1138	10.02	0.02
(3034, 2519)	16	3016912	1736	9.1	0.02
(2529, 2534)	13	1735335	1317	9.7	0.02

Figure 8: A sample of the galaxy catalogue produced. Radii are in units of pixels.