Sky simulation

Filippo Tognini

December 2019

Abstract

In the following document there will be a description of the … r it from the PSF and the noise background. The sky is generate… ved with a gaussian filter and finally reconstructed.

1 Sky simulation

The first thing to describe is how the sky is simulated. The stars follow a mass function

$$P \propto M^{-2} \tag{1}$$

and are generated using the custom function "Crea-s". The function needs the dimension of the sky matrix, the maximum and minimum mass and the density of the stars to create the Sky. Firstly the function generates a matrix 500x500 (the matrix dimension can be changed) full of random numbers between 0 and 1 and then searches for all the numbers higher than a threshold, set by default to 0.998, those will be the positions of the stars, so that the star's density will be two per thousand. The density can be increased and I will show the effect in the next sections. After the the function finds a valid position it generates a star with a mass between M_{min} and M_{max}, set in the function to $0.5M_{\odot}$ and $10M_{\odot}$. The generation is done using the cumulative function, integrating the probability function between M_{min} and x, and then inverting it obtaining:

$$M = \frac{M_{min} * a}{(a - M_{min} * P)} \tag{2}$$

Where P is a random number generated between 0 and 1, and a is the normalization of our probability function:

$$a = \frac{1}{{}^{1}/_{Mmin} - {}^{1}/_{M_{max}}} \tag{3}$$

Figure 1 is the histogram of the sky generated with a density of two per thousand, this will be the sky used for all the simulation. Figure 2 represent the fit in log log of the sky histogram, as expected the slope is -2. The fitted line is not perfectly matching the data because of the lack of statistics. Figure 3 and figure 4 show the histogram and the fit with a sample of 250.000 stars, obtained using a a threshold of 0, the fit improved a lot thanks to this, the slope is still -2 as expected.

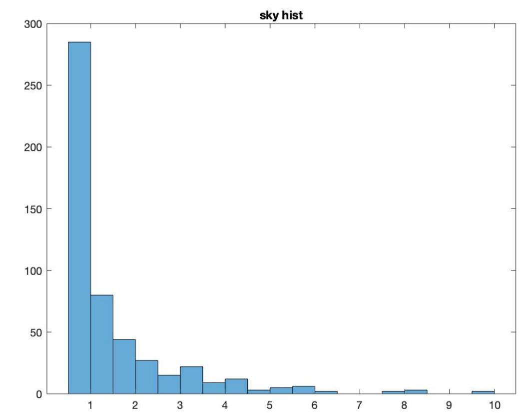

Figura 1: Histogram of the generated sky with a density of 2/1000

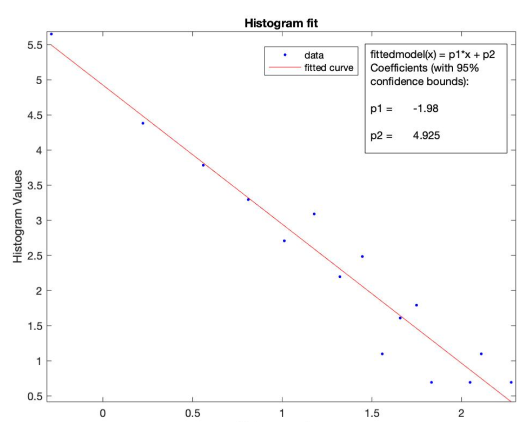

Figura 2: Fit of the histogram in log log a matrix full of stars, histogram edges vs values

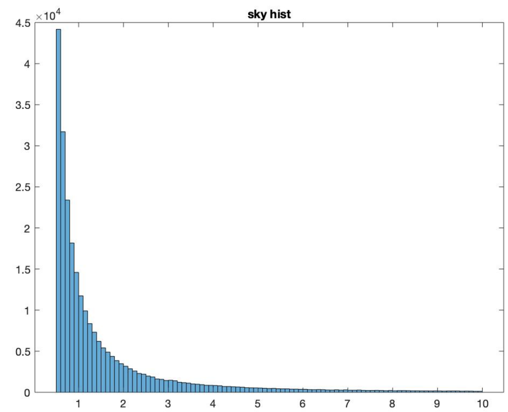

Figura 3: Histogram made with the same distribution, but with 250.000 stars

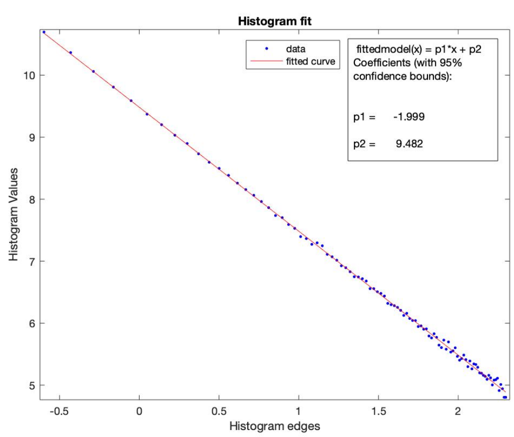

Figura 4: Fit of the histogram in log log, histogram edges vs value

2 Psf convolution and removal

After the sky generation, the next thing to do is to convolve it with the psf. The psf is assumed to be a gaussian and is generated and convolved by the matlab build-in function "imgaussfilt". This function requires as input the image to convolve, the σ and the width of the PSf, then generates a matrix with a PSF with the specified parameters and convolves it. The sky is reconstructed with the homebrew function findmax, which takes as input:

- the matrix of the sky + psf
- the dimension dim of the psf
- a threshold (the importance of this parameter will be clear in the next sections)
- the minimum mass of the stars

Firstly, the function copies the sky matrix in the center of a bigger one with edges width equal to the dimension of the psf and zeros in those positions. This operation is necessary to avoid problems with the edges of the sky matrix during the reconstruction. After this first adjustment the function finds the absolute maximum of the matrix and sums all the pixels in a square of side dim and centered on the maximum coping the result in the reconstructed matrix, only the result of the sum is higher than the minimum mass. Next, the square is set to zero to eliminate the maximum and the function searches for the next maximum until in the copied matrix contains only number lesser than the threshold. The functions requires to know the width of the PSF in pixels and that is symmetric. The dimension of the PSF can be calculated looking at image of the sky and the hypothesis of symmetry, in this case, is well defined. The reconstructed sky and the original one are then confronted with the homemade function "confronta-cielo". The function compare the reconstructed sky with the original to count the number of recognised stars, unrecognised stars and false one. Firstly the function finds all the position (j,k) of the stars in the original sky and then confront a square of size dim centered in (j,k) in the reconstructed sky to find out if there is a star or not. The square is then set to zero, so that all the remaining stars are counted as false ones.

The results that will be described next were produced with a PSF width equal to 5 pixel and a sigma of 1. The number of stars on the generated sky are 517, after the reconstruction the number of recognised stars are 502, of the unrecognised are 15, of false stars is 0. This loss is due to the point spread function and the correlation between pixel. In fact if two stars are near each other only the brightest one is recognised, the other one is lost. As shown in figure 5 the histogram of the reconstructed sky is nearly perfectly matching the original one, except for the loss of some stars because of the PSF. Figure 6 shows the histogram fit, the slope of the line is compatible with the one of the original sky as expected.

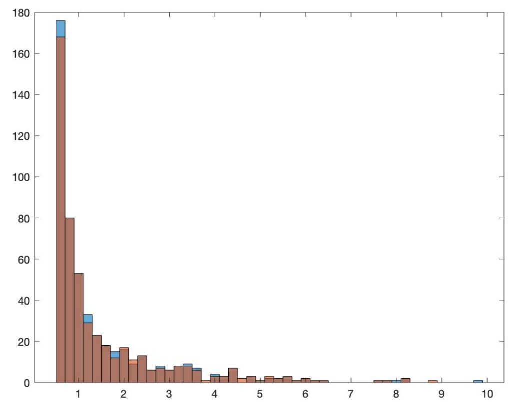

Figura 5: histogram of the reconstructed sky (red) and the original one (blue)

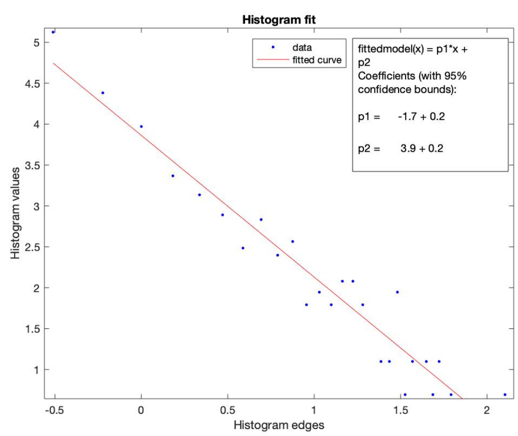

Figura 6: Fit of the histogram of the reconstructed sky in log log, histogram edges vs values, the fit is not perfect because of the lack of stars

Next, as expected, if the dimension of the PSF is brought up to 9 pixels the number of recognised stars decrease to 471, the one of unrecognised stars 46 and the one of false stars remains 0. The same thing happens if the dimension of the PSF is further increased to 13, instead, whereas if the dimension is reduced the recognised stars are 511 and the lost are 6. Increasing the density of stars increases the ratio of lost stars and recognised stars, the ratio with the default threshold is 0.0299. Using a threshold of 0.995, equal to a density of 5 per thousand, the ration is increased to 0.0817, this is because of the fact that more stars are correlated by the PSF and lost during the reconstruction.

3 Effect of the Noise

In this section two type of noise would be discussed, the additive noise and the convolved one. The additive noise is only added after the convolution of the sky with the PSF, the other one is added before the convolution.
The noise is generated in the same way with the matlab function "normrnd". The "normrnd" function generates a matrix filled with normal random generated numbers with mean and a sigma that can be modified. To exclude the negative values before the sum the script takes the absolute value of the noise matrix and then sums it. Figure 7 is the histogram of the noise matrix with mean $\mu = M_{min}/3$ and $\sigma = \mu/10$, those are the default values in the script Simul-s.

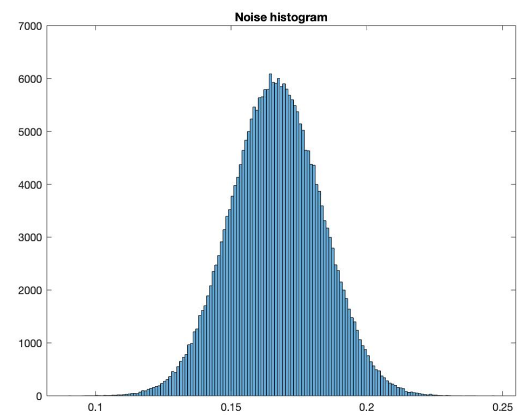

Figura 7: Noise histogram, the Mean is $\mu = M_{min}/3$ and the $\sigma = \mu/10$

3.1 Additive noise

As seen in the last section in this case the noise is summed after the convolution with the PSF. The reconstruction of the sky is made with the "findmax" function but with a threshold greater than zero. The threshold chosen in the end is $\mu + 5 * \sigma$ because is the best to avoid losing too many stars and creating false ones. The number of stars recognised with this set up is 468,the number of unrecognised stars is 49 and the false ones are 0. If the threshold were lowered to $\mu + 4 * \sigma$ the number of recognised stars would increase to 507, but the script would create 7 false stars.
The loss of stars this time depends on the mean and sigma of the noise. Figure 8 contains the overlap of the original sky's histogram and the reconstructed one. As shown in the figure there is a significant loss of low mass stars because of the noise, some star are brought in the higher mass bins and some are lost because of the threshold.

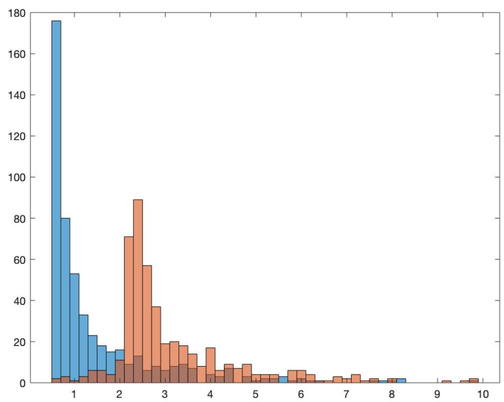

Figura 8: The blue histogram is the original one, the orange one is the reconstructed one

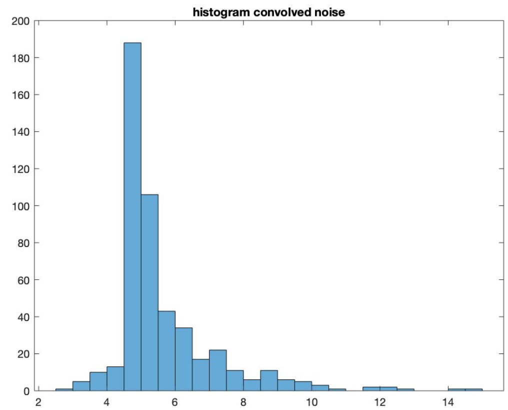

Figura 9: Histogram of the reconstructed sky made with the convolved noise

If the mean is raised the number of lost stars raises as well, as expected. With a mean $\mu = {}^{M_{min}}/_2$ the number of recognised star is 327, of unrecognised is 190 and false is 0. The reconstruction gets way worse if the the mean is set equal to M_{min}, the number of unrecognised stars becomes 362 exceeding the recognised that are 155; the number of false stars is still zero. Obviously decreasing the mean reduces the unrecognised star, with a $\mu = {}^{M_{min}}/_4$ the number of recognised stars are 504, unrecognised stars are 13 and false 0. An interesting thing to note is that the number of recognised stars increases if a small noise is present. This is due to the fact that the "findmax" function reconstructs only the stars that after the sum are over M_{min}. It is possible, in the case of the PSF only, that if two stars are correlated by the PSF only the brighter one is recognised as a stars and the other is lost because it doesn't exceed the control parameter. This changes when adding the noise, because the second star can now be seen by the script. Using a neglectable noise the number of recognised and unrecognised stars return to the original ones of 502 and 17.

3.2 Convolved noise

The effect of the convolved noise on the reconstruction is similar to the additive one. For the same reason as the last section the threshold is set to $\mu + 5 * \sigma$, where μ and σ also remained untouched. The number of recognised stars in this case is 489, unrecognised stars is 28 and false ones is 0. The number of recognised stars drastically reduces when the noise is increased. If $\mu = M_{min}/2$ the number of recognised stars is 304, unrecognised is 213, false is 0. If the mean of the noise is decreased to $M_{min}/5$ the number of recognised stars goes up to 500 and the number of unrecognised stars to 17. Figure 9 is the histogram of the reconstructed sky in the case of $\mu = 3$

As shown in the histogram the low mass stars are either lost because of the noise or brought into an higher luminosity bin. A consequence is that in the reconstruction is lost all the information about the generation of the stars, that can be recovered only knowing the original values of the mass or deconvolving.