Observatoire de Paris
Institut d'Astrophysique de Paris

THESIS

Toward a new level of modeling of environmental effects on galaxies

École doctorale
Astronomie et Astrophysique d'Île-de-France

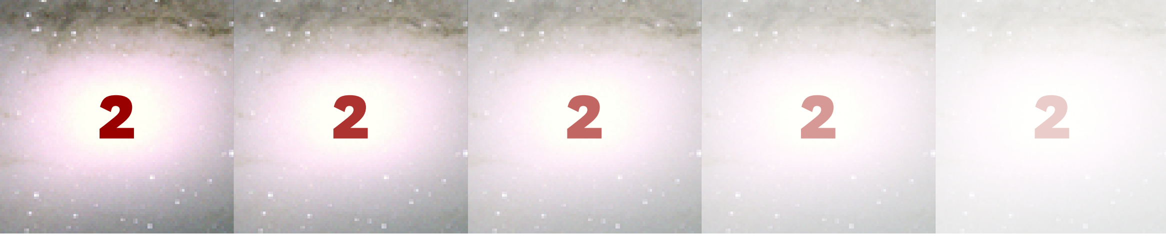
PhD thesis
Astronomie et Astrophysique

DUARTE Manuel

Director:
MAMON Gary

Jury:

Sponge Bob,
Bender,
Fry,
Pr. Farnsworth



THESIS

Contents

THESIS

CONTENTS

Chapter 1.

Luminosity function

There is very lot of words in this note but I can't stop to write, else the world will end!

1.1 Determine the LF

When we want to compare the results from our galaxy group finder to other existing algorithms, we have to compare a flux limited catalogue with our algorithm. But as said before, our algorithm works on a double complete sample of galaxies. So, we need to develop a flux limited version of our algorithm.

The problem when working with a flux limited sample of galaxies is that we must correct for missing galaxies. The galaxies observed by the survey can be seen just when they are brighter than a luminosity limit which depends on the redshift (the distance of the galaxy). We can determine this luminosity easily by the theory:

$$L_{\text{lim}}(z) = \left(\frac{d_{\text{lum}}(z)}{10pc}\right)^2 10^{0.4(M_\odot - m_{\text{lim}})} \tag{1.1}$$

with m_{lim} the magnitude limit of the survey and $M_\odot$ the absolute magnitude of the sun, all in the same band filter. This luminosity is in unit of sun luminosity. In ours groups, when they are at a distance above the distance limit to see galaxies with the luminosity threshold of the catalogue, some galaxies are missing because they can't be seen. In order to correct for the number of missing galaxies, we have to know the distribution of galaxy luminosities. With this luminosity function (LF), we can calculate the "fraction" of galaxy luminosities in mean that we can see:

$$f\left(L_{\text{lim}}(z)\right) = \frac{\int_{L_{\text{lim}}(z)}^{\infty} L\phi(L)\mathrm{d}L}{\int_{L_{thres}}^{\infty} L\phi(L)\mathrm{d}L} \tag{1.2}$$

where $\phi(L)$ is the LF. So determining this LF is useful to correct for missing galaxies.

But, as we expect with the goal of this thesis, it's clear for us that properties of galaxies depend on the environment. So the luminosity function may probably depend on the host halo. The LF have to depend on different characteristics of the halo. The unique "observable" property is the virial mass so we want that the LF depends on it. This is the better way to correct for the incompleteness of groups, with a particular correction for each group. **Manuel: Put something in order to justify physically this dependence on the halo mass!** Resulting from this idea, the LF used in (1.2) becomes a conditional LF, which is the LF in groups of a given halo mass:

$$\phi(L) \to \phi(L|M) \tag{1.3}$$

In galaxy groups, we separate galaxies in two classes: centrals and satellites. Centrals are expected to be the most massives galaxies in groups, and consequently, it's probable that the central is the brighter galaxy. A consequence is that if we can't see the central galaxy, we can't see other galaxies in the group and the correction is not needed because we don't know how to correct for incompleteness. So for the correction we just need to constrain the distribution of luminosities in satellite galaxies.

In practice, we have to choose a functional for this conditional luminosity function (CLF) which can be easily fitted and integrated to determine the correction factor in our group luminosities. In studies of the galaxy sample from the SDSS survey as in Blanton et al. [3], the LF has been well fitted by a double Schechter functional form which can be written:

$$\phi(L) = \left(\phi_1^*\left(\frac{L}{L_*}\right)^{\alpha_1} + \phi_2^*\left(\frac{L}{L_*}\right)^{\alpha_2}\right)\exp\left(-\left(\frac{L}{L_*}\right)\right) \tag{1.4}$$

Now we assume that the CLF have the same form that (1.4). The dependence on the halo mass M is taken with the parameters of the double Schechter (DS). For example $\alpha_1 \to \alpha_1(M|\theta)$, where the functional form of this dependence is not given explicitly here, and θ is a set of parameters relative to the function used to describe the dependence with halo mass. The number of parameters in θ can vary greatly, depending on the function used.

The form of this dependence can't be determine in advance when we want to fit the CLF on the data. For example in the SDSS, we have to know in advance the properties of the groups in order to choose a certain dependence for the parameters of the DS with the virial mass. So, for testing the viability of this method, we have to select a functional that describes correctly the modulation of the parameters with the halo, and samples of galaxies that can give us this information are present in outputs of semi-analytical models (SAM). In such samples, we know in which group a galaxy is, and the virial mass of the host halo is known too. To validate this method of correction for incompleteness, we can test it in mock galaxy catalogues.

1.1.1 Estimating parameters

We need to use a method for estimating the parameters that fit well the data (real or simulated). When working with distribution function, it is common and better to use the maximum likelihood estimation defined as:

$$\mathcal{L}(\theta|X) = \prod_i p_i(X_i|\theta) \tag{1.5}$$

where X is the set of data (in our case the luminosity L) and θ the set of parameters of the model that have to be estimated.

If we consider Bayesian statistics, the likelihood is defined as $p(X|\theta)$ and it seems to be incoherent. But using the Bayes's theorem, we can see that *our* likelihood is in reality the posterior distribution which is proportional to the likelihood in the definition of Bayesian statistics, multiply by a prior. But we don't have any prior on the parameters distribution (which can be discussed...). So, if we take a constant for the prior (probability equal for each parameter), we get same results.

There
very l
words
this r
but l
stop
writ
the
will

It's more convenient to use the logarithm of the likelihood in order to prevent numerical problems when calculating the likelihood. The product in (1.5) becomes at this moment a sum, and the computation is simplified. It's easier too to minimize a function numerically so we rewrite it.

$$-\log \mathcal{L}\left(\theta|X\right) = -\sum_i \log\left(p_i\left(X_i|\theta\right)\right) \tag{1.6}$$

We define $p_i\left(X_i|\theta\right)$ as the probability to get the value X_i given the parameters θ, so it's the probability density. To determine this density, we need to calculate the number of "points" in the sample which are between X_i and $X_i + \mathrm{d}X_i$ compare to the total number of points in the set X:

$$p_i\left(X_i|\theta\right)\mathrm{d}X_i = \frac{\mathrm{d}N_i}{N_{\mathrm{tot}}} \tag{1.7}$$

By definition of the CLF, which is the number of galaxies by unit of volume comprised between L and $L + \mathrm{d}L$ at a given halo mass M, we can write:

$$\mathrm{d}^2N = \phi\left(L|M\right)\mathrm{d}L\mathrm{d}V \tag{1.8}$$

Summing on all the volume in which we are working, we get:

$$\mathrm{d}N = \phi\left(L|M\right)\mathrm{d}L \tag{1.9}$$

Manuel: Because we are working in limited volume region, we can rewrite the probability density as:

$$p_i\left(L_i|\theta\right)\mathrm{d}L_i\mathrm{d}V = \frac{\mathrm{d}^2N_i}{N_{\mathrm{tot}}} \tag{1.10}$$

So we can write:

$$p_i\left(L_i|\theta\right)\mathrm{d}L_i = \frac{\phi\left(L_i|M\right)\mathrm{d}L_i}{N_{\mathrm{tot}}} \tag{1.11}$$

and the total number of galaxies is just:

$$N_{\mathrm{tot}} = \int_{L_{\mathrm{thres}}}^{\infty} \phi\left(L|M\right)\mathrm{d}L \tag{1.12}$$

In this way, the density probability for the DS can be written:

$$p_i\left(L_i \middle| \alpha_1, \alpha_2, M_*, \frac{\phi_2^*}{\phi_1^*}\right) = \frac{\left(\left(\frac{L}{L_*}\right)^{\alpha_1} + \frac{\phi_2^*}{\phi_1^*}\left(\frac{L}{L_*}\right)^{\alpha_2}\right)\exp\left(-\left(\frac{L}{L_*}\right)\right)}{\left(\Gamma\left(1+\alpha_1, \frac{L_{\mathrm{thres}}}{L_*}\right) + \frac{\phi_2^*}{\phi_1^*}\Gamma\left(1+\alpha_2, \frac{L_{\mathrm{thres}}}{L_*}\right)\right)} \tag{1.13}$$

where $\Gamma\left(a, x\right) = \int_x^\infty e^{-t}t^{a-1}\mathrm{d}t$ is the incomplete gamma function.

The principle of the estimation by the method of the maximum likelihood is that when we maximize the likelihood relatively to the parameters θ, we get the maximum of probability, likelihood, of having the parameters that correspond denoted $\hat{\theta}$. The parameters $\hat{\theta}$ are the parameters that best fit the data according to the functional form assumed for the CLF. Numerically we minimize the equation (1.6).

There are many ways of doing such a minimization. When the probability density isn't too complex, $\hat{\theta}$ can be determined analytically. But in this case, with the DS, the incomplete gamma function prevent us to do it in this way. So we are constrained to use numerical methods in order to minimize the likelihood. Many algorithms exist to do this job like Powell's method, Newton-Raphson's method, etc..., but they share the same problem: when they find a minimum, we can't know if it is the global minimum or if it is a local minimum. The result depends on the initial starting point of the algorithm in the parameter space. Some other methods try, using Monte-Carlo methods, to do a better exploration of this parameter space, allowing some "jumps" to other regions in order to see if there isn't a best minimum near. An example of such an algorithm is the simulated annealing method which implement the cooling of a material where the function to minimize becomes the energy of the system and a fictive temperature T is introduce to allow some temperature jumps. But it is not always sure that we get the global minimum. Moreover, we can't easily determine errors on the estimation of the parameters, except using bootstraps or jackknife techniques which need many estimation of the parameters varying the sample which may be expensive in calculation time.

We have chosen to use the Markov Chains Monte Carlo method (MCMC) to minimize our function. This method is the better in all the universe. **Manuel: Explain why!**. We can estimate easily with results of the algorithm the errors on the parameters.

We will now resume the result of works on mock catalogues of our method of estimating the CLF.

1.1.2 Tests on mock catalogues

There is two steps in order to determine the dependence on the halo mass of the parameters of the DS model. First, we have to determine what is the best functional form to fit this dependence which can be done on a complete sample of galaxy. Secondly, see if we can recover this parametrisation and modulation with a flux limited sample of this galaxies to know if the method works well when applied in a real survey.

Complete sample

In order to determine the dependence on the halo mass of the parameters, we use a complete sample of galaxies taken from the outputs of the SAM of Guo et al. [4] applied on dark matter halos from the Millennium II run. We limit our sample of galaxies from this catalogue to galaxies with a luminosity such that the absolute magnitude in the r band is $M_r < -12$. For each galaxy, we have the virial mass of the halo which contains this galaxy. Our complete sample is defined just as this galaxy catalogues from Guo et al. [4] with the truncation on the data through the r band magnitude.

First, we determine what is the best model for the "total" CLF, *i.e.* the LF when we don't do a segregation with the halo mass. We have tried to adjust a simple Schechter and a double Schechter. Results are shown on figure (**??**). We can see that the minimization works well because the fit seems to be good enough in the figure. The double Schechter fits better the data than the simple Schechter because we can constrain with this form the two populations of galaxies in the sample from the Guo et al. [4] SAM. We see that there is a low population with high slope and a brighter population with a slope more little. Differences with the data at luminous galaxies is due to the fact that the number

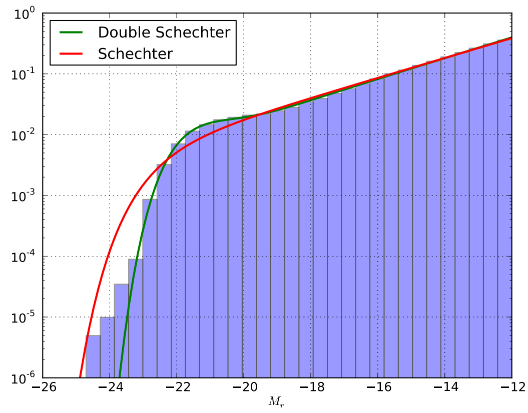


Figure 1.1: Results of the fit on the data in blue with a Schechter distribution in red and a double Schechter distribution in green. The double Schechter fit is better because this form allow to constrain the two galaxy populations we can see in our sample: a low population with high slope and a brighter population with a more little slope in absolute value.

of galaxies with $M_r < -24$ is very low, in some bins there is just one galaxy. But we need a quantitative proof of this fit. We use for that the Kolmogorov-Smirnov test ans the P-Value associated.**Manuel: TODO: KS test on the fit in order to get a good idea of the robustness of the fit.**

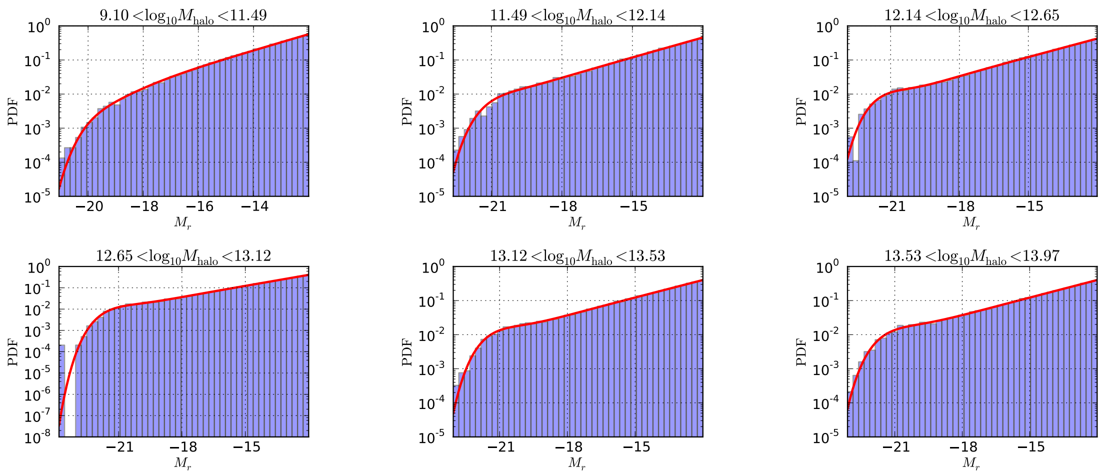
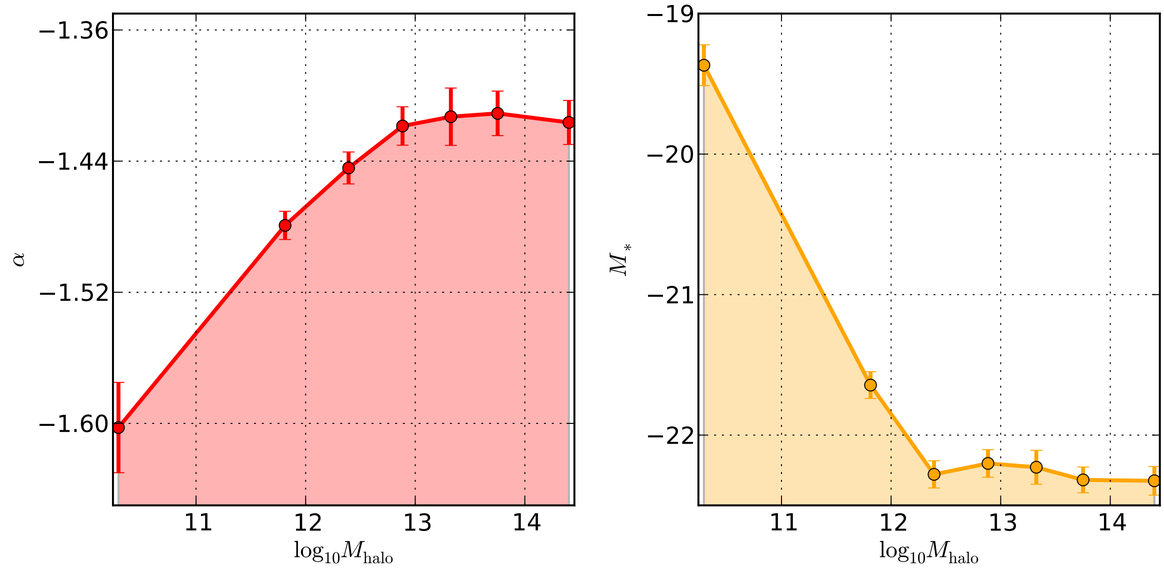


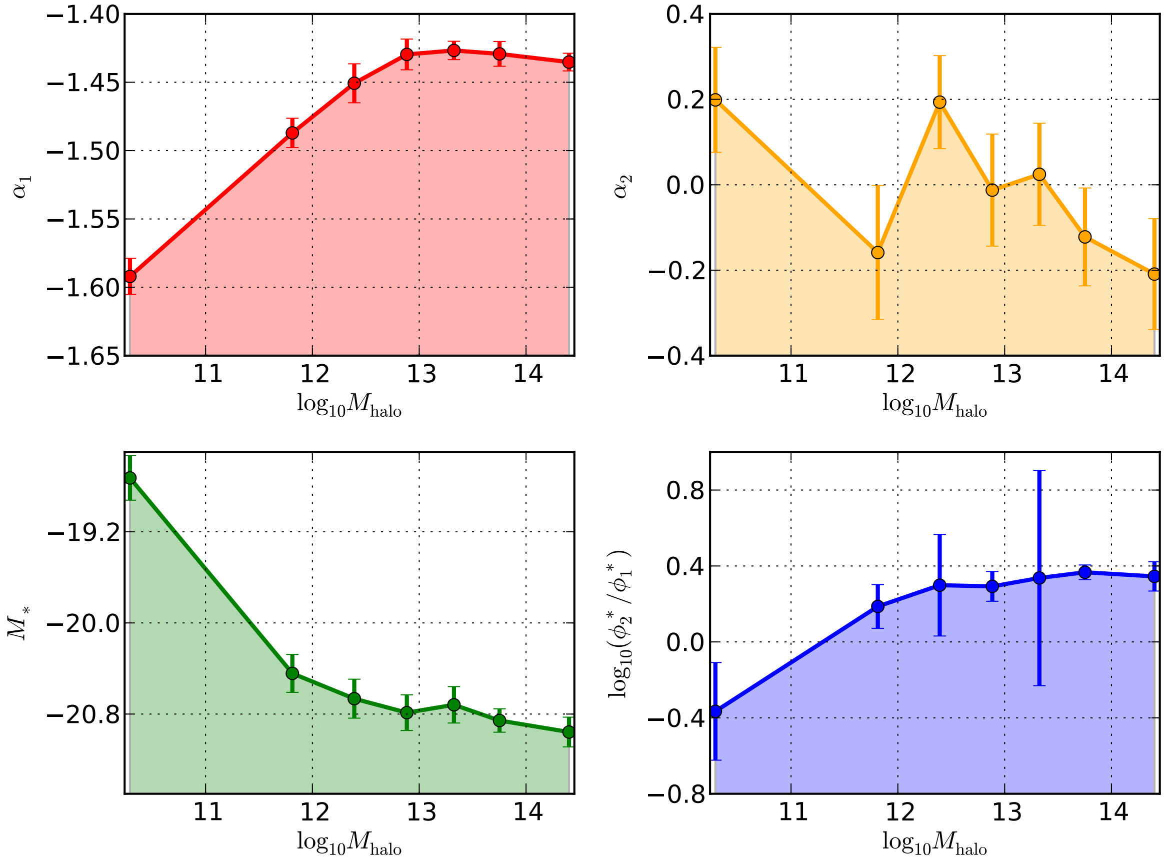
Figure 1.2: Results of the fit for the double Schechter distribution in different bins of halo mass. In blue the PDF of the data, and in red the fit with parameters obtained from the MCMC algorithm.

Happy to see that the minimization is good, we want to see the modulation of the parameters of the DS with the halo mass. For doing that we take galaxies in bins of the certain width in halo mass, and we compute the parameters that fit well the data in each, as previously. This modulation is represented in the figure (**??**). The resulting fit of the probability density function is shown for the double Schechter in the figure (**??**).

As we expect, we can see some physics process putted in the SAM in this fig-



(a) For a Schechter distribution



(b) For a double Schechter distribution

Figure 1.3: Modulation of the parameters of both Schechter and double Schechter luminosity distributions with the halo mass.

ure.**Manuel: To detail.**

Therefore, **Manuel: Verify Cependant→Therefore.**, we don't see a particular modulation, *i.e.* a functional form to use for each parameter of the DS. So we have decided to use polynomials functions to adjust this modulation. For a given parameter θ_i we write:

$$\theta_i(M) = \sum_{j=0}^{N} a_{ij} M^j \tag{1.14}$$

where N is the order of the polynomial.

We now integrate this form of parametrization of the parameters of the DS into the minimization of the likelihood using directly the halo masses of the galaxies.

But it doesn't work!!!**Manuel: Some blabla!**

Flux limited sample

Working with a flux limited sample, it's like having gaps into the set of galaxies. But we know why we are missing this galaxies: we can't see them at a given distance. Determining the luminosity limit which we can see at redshift z can be done analytically. So for that we use the STY method.**Manuel: Put Bibtex of STY and description related to this.** We need to modify the probability density used in the likelihood to take into account missing data. The procedure is the following: in order to estimate the likelihood, we have to calculate the probability density for a galaxy i at the redshift z_i of having a magnitude between M_i and $M_i + \mathrm{d}M_i$. The probability that a galaxy have a magnitude $\mathcal{M}$ superior to M is given by:

$$P(\mathcal{M} > M|z) = \frac{\int_{-\infty}^{M} \phi(M')\,\rho(z)\,f(M')\mathrm{d}M'}{\int_{-\infty}^{\infty} \phi(M')\,\rho(z)\,f(M')\mathrm{d}M'} \tag{1.15}$$

where f is the completeness function and $\rho(z)$ is the redshift distribution. We can express it like this because the number of galaxies at a given redshift z with a given magnitude M is $\mathrm{d}N = \phi(M)\,\mathrm{d}M$ in a given volume. So to avoid the volume dependence, we multiply by the number of galaxies in a given volume $\rho(z) = \mathrm{d}N/\mathrm{d}V$. So the number of galaxies with a magnitude between M and $M+\mathrm{d}M$ is $\mathrm{d}^2N = \phi(M)\,\rho(z)\,\mathrm{d}M\mathrm{d}V$. But with the problem of the completeness, we can see just galaxies with a certain magnitude defined by f so $\mathrm{d}^2N = \phi(M)\,\rho(z)\,f(M)\,\mathrm{d}M\mathrm{d}V$. The probability (1.15) results from this. Calculating the probability density is straightforward because we have:

$$P(\mathcal{M} > M|z) = \int_{-\infty}^{M} p(M'|z)\mathrm{d}M' \tag{1.16}$$

and so:

$$p(M|z) = \frac{\partial P(\mathcal{M} > M|z)}{\partial M} \tag{1.17}$$

Finally:

$$p(M_i|z_i) = \frac{\phi(M_i)}{\int_{M_{\mathrm{bright}}(z_i)}^{M_{\mathrm{faint}}(z_i)} \phi(M')\,\mathrm{d}M'} \tag{1.18}$$

and this defines the new likelihood in the case of a flux limited sample.

We apply this to the incomplete sample generated with the mock algorithm we have created. Firstly, we have tried to recover the parameters in the mock with the apparent magnitudes calculated applying a "K-decorrection". The results are very bad. Parameters can't be recovered correctly and with the two models chosen (simple Schechter and DS). We don't have the same estimations as in the complete sample, although the flux limited sample is the same as the complete sample but "truncated". It's not clear why this parameters can't be found. The procedure isn't in cause, but it's possible that the data are the problem. We think that our method for estimating the K-decorrection have some troubles when the redshift is higher than 0.2.

For verifying this assumption, we have made more simple flux limited samples. We know how to generate random variables following a Schechter distribution and a DS distribution two. We have placed galaxies in a homogeneous universe, up to few hundred Mpc. We have assigned redshifts to this galaxies according to their distances D using simply $z = v/c = H_0D/c$. We apply a Schechter distribution to the galaxies generated, without taking clusters effects into account. Calculating apparent magnitudes is done subtracting the distance modulus using galaxies redshifts. We have done the same applying a DS. Results are the followings:

- We are enable to recover the Schechter parameters used to generate the distribution in the flux limited sample. When data are perfect like in this situation, there are no troubles.
- Using a DS distribution, we have more difficulties in finding those parameters used to generate the flux limited sample. The slopes of both the low galaxies and brighter galaxies aren't near the true values.

We think that the number of low galaxies in the flux limited sample isn't not sufficient to well constrain the slopes in the DS. As a result, the maximum likelihood can't find real parameters. It's like a degeneracy is present and the algorithm can't decide to the true parameters.

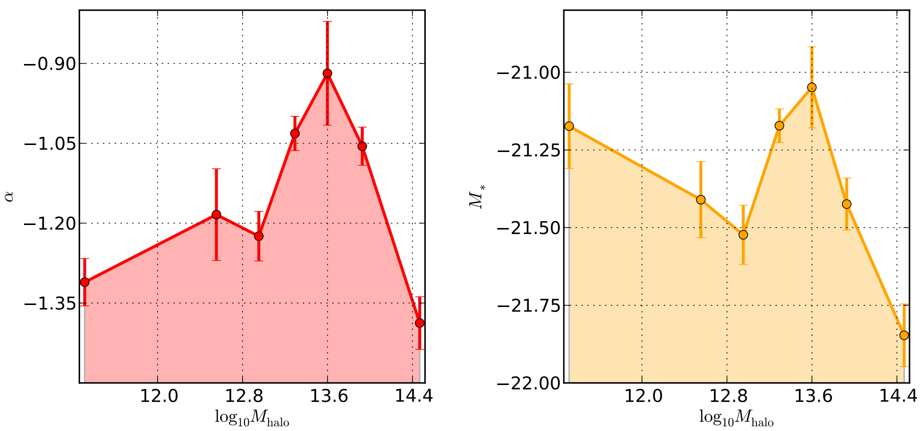


Figure 1.4: Modulation of the parameters for a Schechter distribution with the halo mass in a mock catalogue constructed with the galaxy catalogue from the Guo2010a, without taking a K-decorrection for apparent magnitudes of galaxies.

So, it's a clue that the K-decorrection isn't good at high redshift and underestimate the number of galaxies when we apply a flux limit using apparent magnitudes. We have made an other mock catalogue without this K-decorrection in order to show that is the real problem. We have extended too the upper magnitude limit of galaxies in the sample in order to improve the number of low galaxies to estimate better slopes in the case of the DS. The former limit was -15 in r band magnitude and now we go to -12. So we expect that the number of low luminosity galaxies increases and improves the parametrization. All results hereafter and before are for this new limit.

Unfortunately, the increasing number of low mass galaxies and taking no K-decorrection doesn't improve the estimation of the parameters. Results are shown in figure (**??**).

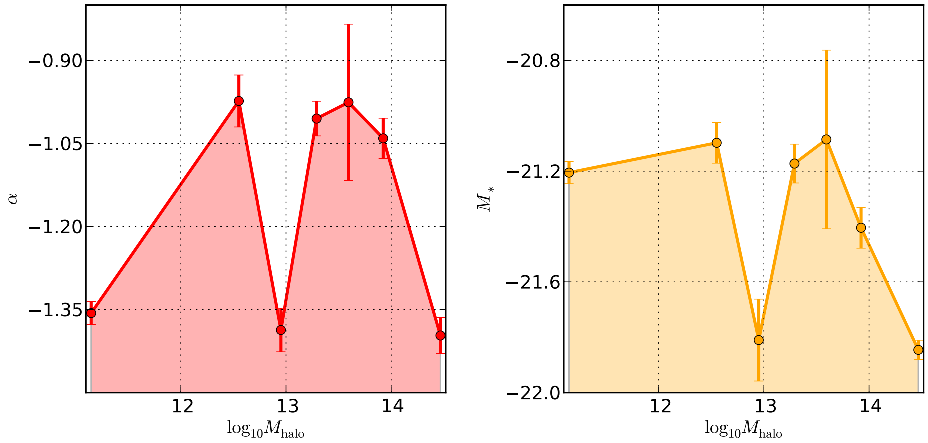


Figure 1.5: Modulation of the parameters for a Schechter distribution with the halo mass in a mock catalogue constructed with the galaxy catalogue from the Guo2010a, without taking a K-decorrection for apparent magnitudes of galaxies and removing galaxies in groups that are closer to the border of cube simulation on the mock's cube.

We can see that the modulation of this parameters with the halo mass isn't the same as in the figure (**??**) for the Schechter distribution. It's not very clear why when we construct the mock catalogue, we can't recover the same parameters as for the data used to build this mock catalogue. It's maybe a problem of spatial truncation of data in flux and redshift that may cause some variations on the intrinsic LF of the data from the galaxy catalogue used for the mock catalogue. The mock catalogue have a problem: periodic conditions in the box used in order to build this mock can move away from each other galaxies belonging to the same group. So in groups of a certain halo mass, we have more missing data than expected and the estimation of parameters is affected too. To see if this assumption is correct, we have removed from the sample of galaxies in the mock catalogue these ones that are in groups too close to the border of the box in the mock catalogue. The modulation is shown on the figure (**??**).

We can't find parameters as in the galaxy catalogue from Guo et al. [4] both with modulation of the halo mass and for global data. The behaviour is the same as the latter situation.

The last test to understand why we can't find the same parameters is described in what follows. We have used the algorithm described in the appendix **Manuel: add the description in the appendix to generate a galaxy population from a simulation** in order to create a sample of galaxy following a NFW profil in halos, with a velocity dispersion calculated using the Mamon and Łokas [2] model for the anisotropy factor.

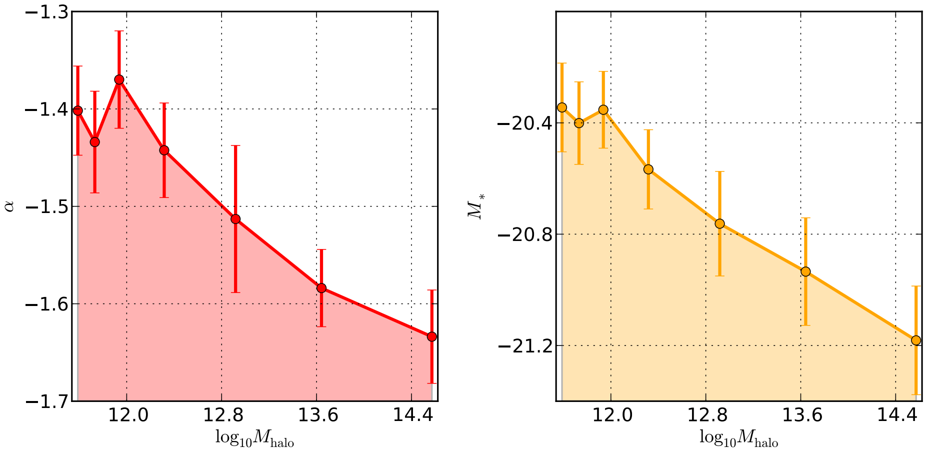


Figure 1.6: Modulation of the parameters for a Schechter distribution with the halo mass in the sample of galaxy constructed using a HOD model, and imposing a linear evolution of the parameters with halo mass. We have imposed that α goes from -1.2 to -1.7 and M_* from -19.5 and -21.5 between the extreme halo mass of the simulation.

Galaxy luminosities in the halo are generated in order to have a linear modulation of the parameters of a Schechter distribution with the halo mass. We imposed this modulation in the data of the galaxy sample generated with the HOD model of [5]. The magnitude sample has $-23 < M_r < -12$. The result of this modulation in the complete sample is shown in figure (**??**).

We now construct an other mock catalogue using this galaxy sample in order to see what happened when we fix the modulation in the parameters with the halo mass in a flux limited sample. Results are shown on the figure (**??**)

The mock catalogue we have created goes to a redshift of 0.1 while the mock with the galaxy sample from Guo et al. [4] is limited to a redshift of 0.3. The modulation of the parameters we have imposed in the galaxy sample is well recovered. In the case of M_*, the estimation is always good, same with a DS. But there are more uncertainties in finding the slope of the LF with both Schechter and DS. Slopes are less constrained by the data when with have a flux limited sample.

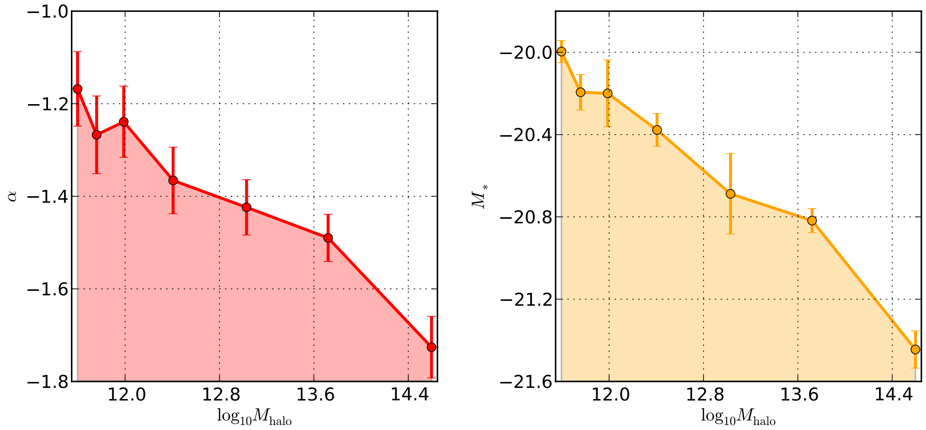


Figure 1.7: Modulation of the parameters for a Schechter distribution with the halo mass in the mock catalogue constructed as described in the previous figure. Parameters are recovered with a given uncertainties.

Appendix A.

Analysing SDSS-DR8

A.1 Introduction

In order to realize the mock catalogue for the group finder algorithm, we need to mimic the SDSS. This mock have to be realistic. But our algorithm needs to have the redshift for *all* galaxies in the volume selected and the SDSS provides spectroscopic redshifts just for galaxies that could have been targeted due to the problem of fibre collision. So for galaxies in this situation we use the photometric redshift. The problem is that dense regions on the survey are more susceptible to don't have a spectroscopic redshift than a galaxy in a lower dense region. So we have to determine how the fraction of photometric galaxies depends on the density of galaxies in the sky in order to apply this in our mock catalogue.

In the SDSS, there are different ways to estimate photometric so we list the methods here. **Manuel: Add the list of method available in the SDSS**.

We can select galaxies in the sample with an SQL query. All queries used will be summary here.

A.2 Analysis

A.2.1 Definitions

In the SDSS there is something called "stripes" which is a band of observations in the sample. Those *bands* can overlap contrary to "chunks" which are similar bands but don't overlapping (they make a complete partition of the survey in their union). We can use this stripes in order to select galaxies in regions of interest for our studies. Data on the SDSS provide limits of this stripes, so we can use it to fix borders of the survey. In reality, in the region of the survey we consider, we don't see overlapping of the stripes. So it is more useful to

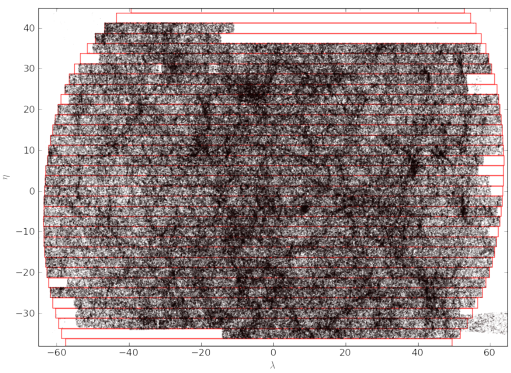


Figure A.1: Spectroscoped galaxies in the SDSS DR8 with stripes limits as planned. Coordinates are in degrees.

use them in order to define limits of survey in region of our interest.

Following definitions given in the SDSS website, we can define two other coordinate systems in the survey which we can use to select galaxies.

Great Circle: This coordinates system is define with two angle (μ, ν). Coordinates are relatives to one stripe so it can be use when working with galaxies in the region of the stripe we consider. **Manuel: More definitions of this**.

Survey Coordinates: It's an other system similar to celestial coordinates but "centred" on the "block" of galaxies of the survey that we can see in maps. Coordinates are written (η, λ). If we use celestial coordinates, we have:

$$(0, 90°)_{(\eta,\lambda)} = (275°, 0)_{(\alpha,\delta)} \quad (57.5°, 90°)_{(\eta,\lambda)} = (0, 90°)_{(\alpha,\delta)} \tag{A.1}$$

It results from this that $-\frac{\pi}{2} < \eta < \frac{\pi}{2}$ and $-\pi < \lambda < \pi$.

Wee this informations we can write the transformations between the different coordinate systems.

Survey coordinates to celestial coordinates

From previous definitions, we see that the relation between those systems is just a rotation. So:

$$\begin{aligned} \delta &= \arcsin\left(\cos\lambda \sin\left(\eta + 32.5°\right)\right) \\ \alpha &= \operatorname{atan2}\left(\sin\lambda, \cos\lambda\cos\left(\eta + 32.5°\right)\right) + 185° \end{aligned} \tag{A.2}$$

Celestial coordinates to survey coordinates

The inverse transformation is in consequence:

$$\begin{aligned} \eta &= \operatorname{atan2}\left(\sin\delta, \cos\delta\cos\left(\alpha - \alpha_0\right)\right) - \delta_0 \\ \lambda &= \arcsin\left(\cos\delta \sin\left(\alpha - \alpha_0\right)\right) \end{aligned} \tag{A.3}$$

with $(\alpha_0, \delta_0)_{(\alpha,\delta)} = (0, 0)_{(\eta,\lambda)}$. We have to apply too periodic conditions in the angles founded by the latter equation in order to have values in the correct range. So conditions are:

$$\begin{aligned} \text{Where} \qquad & \eta < -90° \text{ or } \eta > 90° : \\ & \eta \rightarrow \eta + 180° \\ & \lambda \rightarrow 180° - \lambda \end{aligned} \tag{A.4}$$

$$\text{Where} \quad \begin{aligned} &\eta > 180^\circ : \\ &\eta \to \eta - 360^\circ \end{aligned} \tag{A.5}$$

$$\text{Where} \quad \begin{aligned} &\lambda > 180^\circ : \\ &\lambda \to \lambda - 360^\circ \end{aligned} \tag{A.6}$$

Determining the number of a stripe to which a galaxy pertains is easy too because stripes are organized such they have a constant width along the η coordinate, with a width of 2.5°. The number of the stripe n of a galaxy with η position is:

$$n = \text{floor}\left(\frac{(\eta + 58.75^\circ)}{2.5^\circ}\right) \tag{A.7}$$

A.2.2 Galaxies selection

There are many tables in the SDSS saving galaxies and other objects properties extracted from images of the survey. Those tables are the results of different selections in objects detected in images. When crossing objects between images of the survey that overlap, there are some differences of positions between the same object in the two images. So there are possibilities that an object is observed twice or more. In many of those tables, there is no "double objects".

The `Galaxy` view is a selection from the `PhotoPrimary` for objects flagged as *galaxy*. The `Galaxy` view contains the photometric parameters (no redshifts or spectroscopic parameters) measured for resolved primary objects. But we have other useful informations to link with tables that give us photometric and spectroscopic redshifts. There is the `specobjid` to link with spectroscopic redshifts in the table `SpecObj` which doesn't contain duplicates (it's a clean table of `SpecObjAll` with clean redshifts). If `specobjid=0`, the galaxy doesn't have a spectroscopic redshift. The `objid` is a link to the `Photoz` table which contains all photometric redshifts for galaxies in the `Galaxy` table. Estimation is based on a robust fit on spectroscopically observed objects with similar colors and inclination angle. There is also the `PhotozRF` where estimates are based on the Random Forest technique. Galaxies in the `SpecObj` are limited to $m_r < 17.77$ and a surface brightness selection **Manuel: Add This!!**. So we need to do the same flux limitations when selecting galaxies on the `Galaxy` table. A possible SQL query for selecting galaxies in this table and link them with redshifts tables could be for spectroscoped galaxies:

```
select GG.ra, GG.dec, GG.petroMag_u, GG.petroMag_g, GG.petroMag_r,
GG.petroMag_i, GG.petroMag_z, GG.specobjid, GG.objid, Z.z, Z.Zerr
from Galaxy as GG, SpecObj as Z
where Z.specobjid=GG.specobjid and GG.specobjid!=0 and GG.petroMag_r<17.77
and GG.ra<275 and GG.ra>100 and GG.dec>-10 and GG.dec<75
```

and for galaxies which couldn't be spectroscoped:

```sql
select GG.ra, GG.dec, GG.petroMag_u, GG.petroMag_g, GG.petroMag_r,
GG.petroMag_i, GG.petroMag_z, GG.specobjid, GG.objid, Z.z, Z.Zerr
from Galaxy as GG, Photoz as Z
where GG.specobjid=0 and GG.objid=Z.objid and GG.petroMag_r<17.77
and GG.ra<275 and GG.ra>100 and GG.dec>-10 and GG.dec<75
```

Limits of stripes are given in the SDSS table `StripeDefs` but this limits aren't actual limits, they are planned limits when survey started. We can see it on the figure (**??**) where planned limits are shown in red and spectroscoped galaxies are the points.

We see that some planned regions aren't still observed (spectroscopically speaking). So we need to define other limits in λ coordinates for that stripes that aren't completes. We find by hand the new limits of stripes which contains spectroscoped galaxies. Now, the survey mask is like in figure (**??**). We will consider just galaxies in this mask in order to find groups in the SDSS. Other galaxies aren't easy in order to define borders of the survey and find groups.

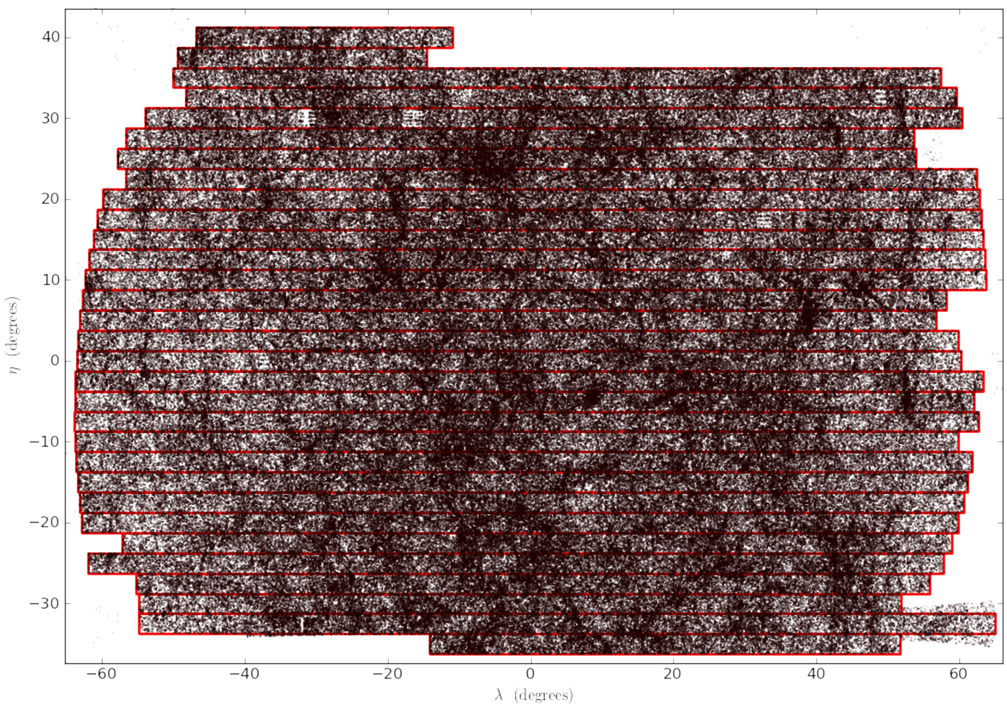


Figure A.2: Spectroscoped galaxies in the SDSS DR8 with stripes limits chosen in order to find easily groups at the border of the survey.

For fibre collisions galaxies, we use galaxies selected in the table of the photometric redshifts and keep galaxies that are in the mask defined previously. Now we have a sample of galaxies in a region of the SDSS for which we can easily characterize borders and where all galaxies, given the flux limit of the SDSS, are presents. There is just the problem of fibre collisions galaxies for which the redshift in our possession is photometric, in consequence less precise than spectroscopic redshifts. But our algorithm is tested on a mock catalogue which is "perfect" if we don't take in account this problem of less robust photometric redshifts. In order to know the behaviour of the algorithm with those problematic redshifts, we need to implement this in our mock catalogue.

Flags in the SDSS

Galaxies can have some troubles with photometry due to fit and estimations in the SDSS. in the general case, those objects are flagged with the `clean` property which indicates by 1 that the photometry is OK and by 0 when there is a problem. Details of the problems are in the bit flag. But for groups, we need to select all galaxies, whether there are not clean.

`Galaxy` table is a selection from `PhotoPrimary` view for objects with `type` = 3 (galaxy). I think that we don't have to care of the "good" photometry of galaxies in the `Galaxy` view, but we can leave a flag in the group finder algorithm to say if a galaxy is in this case.

However, we have to take into account the error on the redshift estimation using the `zErr`. For photometric redshift I think that if the `zErr` is too high, we can use the `nnAvgZ` which is the average redshift of galaxies in the neighbourhood of the considered galaxy. It can be better too if the photometric redshift is too different from it.

The `SpecObjAll` contains duplicates and bad datas. But the `SpecObj` contains just clean spectras. We use `zWarning` to decide if we keep the redshift (`zWarning`=0) or not. In the latter case, we use the photometric redshift instead.

A.2.3 Fibre collision estimation

In the SDSS, obtaining spectroscopic redshifts of galaxies is done using a plate of 1.5° diameter, in which there is a certain number of fibres in order to get spectrum of the galaxy. But in the field of the plate, the number of fibres is limited, and the number of coverings of a portion of the sky is limited too because of the time needed to obtain a spectrum. Although runs may overlap, there is sometime galaxies that can't be spectroscoped. Moreover, fibres have a dimension of 55", so when galaxies are closer than this size, one (or more) of those galaxies aren't spectroscoped. We can see that in the figure (**??**) where we have taken the nearest neighbour of a galaxy and determined the differences in angular size and redshift between the two galaxies. As expected, the number of galaxies which are closer than 55" decreases dramatically. There are still some galaxies because the overlapping of runs can permit to get redshifts for galaxies behind this limit.

A consequence of those problems is that in denser regions, the number of fibre collision increases, affecting more our groups analysis because the number of photometric redshifts is higher in those dense regions.

We need to implement this selection effect in our mock catalogue. For that we compute the local density in the field, taking all galaxies in the neighbourhood of 1.5° of a galaxy, and in the same time, we determine the fraction of galaxies that don't have a spectroscopic redshift. We deduce of this a relation between the density field and the fraction of fibre collisions. In the mock catalogue, we compute the same density field and we apply the

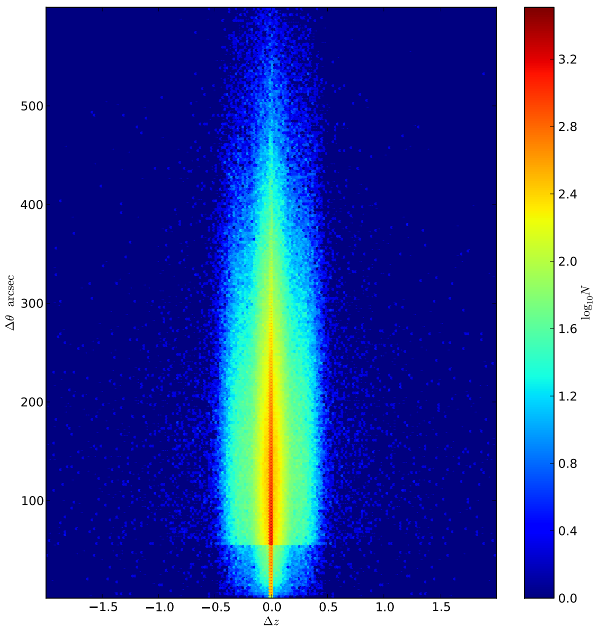


Figure A.3: Distribution of spectroscoped galaxies in the SDSS DR8 in angular size and redshift differences with the nearest neighbour galaxy.

relation estimated in the SDSS sample to the mock.**Manuel: Do it!!** We need for each galaxy to count the fraction of non spectroscoped galaxies in a region of 1.5° radius around. We have to remove galaxies that are to close of the border of the survey, because if we don't remove those galaxies, there are some regions with missing galaxies and the fraction estimation will be affected. The way of selecting those galaxies is to compute a circle of 1.5° around a galaxy, and if a generated point is out of the survey, the galaxy is defined as to be closer to the limits.

We can generate samples of points at an angular distance d to a point of coordinate (α_0, δ_0) using formulas of the spherical triangle. If we define a triangle by the pole, the point (α_0, δ_0) and the point whose we want coordinates (α, δ) denoted M, we can write the following relations:

$$\begin{aligned} \sin(\alpha - \alpha_0) &= \frac{\sin d \sin\gamma}{\cos\delta} \\ \sin\delta_0 \cos\gamma &= \cos\delta_0 \cot d - \sin\gamma \cot(\alpha - \alpha_0) \end{aligned} \tag{A.8}$$

where γ is like a polar angle, which have all the values between 0 and 2π. So we have now:

$$\begin{aligned} \alpha - \alpha_0 &= \arctan\left(\frac{\sin\gamma}{\cos\delta_0 \cot d - \sin\delta_0 \cos\gamma}\right) \\ \delta &= \arccos\left(\frac{\sin d \sin\gamma}{\sin(\alpha - \alpha_0)}\right) \end{aligned} \tag{A.9}$$

There are problems in poles and equator with those formulas. For a γ limit, angles can't be recovered with those formulas. We have in those cases:
with $\gamma_0 = \arccos\left(\frac{-\sin\delta_0 \cos d}{\cos\delta_0 \sin d}\right)$

$$\text{Where} \quad \begin{aligned} &\delta_0 - d < 0 \text{ and } \delta_0 > 0 \text{ and } \gamma_0 < \gamma < 2\pi - \gamma_0 : \\ &\delta \to -\delta \end{aligned} \tag{A.10}$$

$$\text{Where} \quad \begin{aligned} &\delta_0 + d > 0 \text{ and } \delta_0 < 0 \text{ and } \gamma_0 > \gamma \text{ or } \gamma > 2\pi - \gamma_0 : \\ &\delta \to -\delta \end{aligned} \tag{A.11}$$

with $\gamma_0 = \arccos\left(\frac{\cos\delta_0 \cot d}{\sin\delta_0}\right)$

$$\text{Where} \quad \begin{aligned} &\delta_0 + d > \frac{\pi}{2} \text{ and } \gamma_0 > \gamma : \\ &\delta \to \alpha + \pi \\ &\alpha \to \pi - \delta \end{aligned} \tag{A.12}$$

$$\text{Where} \quad \begin{aligned} &\delta_0 + d > \frac{\pi}{2} \text{ and } \gamma > 2\pi - \gamma_0 : \\ &\delta \to \alpha - \pi \\ &\alpha \to \pi - \delta \end{aligned} \tag{A.13}$$

$$\text{Where} \quad \begin{aligned} &\delta_0 - d < -\frac{\pi}{2} \text{ and } \gamma_0 < \gamma < 2\pi - \gamma_0 : \\ &\delta \to \alpha + \pi \\ &\alpha \to -\pi - \delta \end{aligned} \tag{A.14}$$

An other way to draw circles in the sphere is to consider the point for which we want to know celestial coordinates around a given angular distance as the pole of a new coordinate system. In this system, points at given distance of our central point are just points with $\pi/2 - \delta$ and α running between 0 and 2π. We now can determine cartesian coordinates of those points in this system and apply a rotation to go from the "real" system and the system where the central point is the pole. In the new system we have:

$$\begin{aligned} X' &= r\cos\alpha'\cos\delta' \\ Y' &= -r\sin\alpha'\cos\delta' \\ Z' &= r\sin\delta' \end{aligned} \tag{A.15}$$

Then the rotation matrix to go from the "real" system to the new is:

$$R = \begin{pmatrix} \cos\left(\frac{\pi}{2}-\delta_0\right)\cos\alpha_0 & \sin\alpha_0 & \sin\left(\frac{\pi}{2}-\delta_0\right)\cos\alpha_0 \\ -\cos\left(\frac{\pi}{2}-\delta_0\right)\sin\alpha_0 & \cos\alpha_0 & -\sin\left(\frac{\pi}{2}-\delta_0\right)\sin\alpha_0 \\ -\sin\left(\frac{\pi}{2}-\delta_0\right) & 0 & \cos\left(\frac{\pi}{2}-\delta_0\right) \end{pmatrix} \tag{A.16}$$

with $\vec{X} = R\vec{X}'$ where $\vec{X} = (X, Y, Z)$. Then we have just to convert those coordinates in celestial angles using:

$$\alpha = \begin{cases} -\arctan2(Y,X) + 2\pi & \text{if} \quad Y > 0 \\ -\arctan2(Y,X) & \text{else} \end{cases}$$

$$\delta = \operatorname{sign}(Z)\arccos\left(\frac{\sqrt{X^2+Y^2}}{\sqrt{X^2+Y^2+Z^2}}\right) \tag{A.17}$$

Fibre collisions are more probable in dense region of the sky in projection, but for the mock catalogue we need to quantify this. In order to do that, we have selected for all galaxies in the SDSS survey as defined previously galaxies that are closer than 1.5° in angular size, which is the radius of a plate used for spectroscopy in the SDSS. With that, we can estimate the local density field $\Sigma_{1.5°}$ in unit of number of galaxies per degree^2. In this selection, we can determine which galaxies had been spectroscoped or not, and so we can estimate the fraction of non-spectroscoped galaxies. We remove for computing it galaxies that are to close of the border of the survey, and so galaxies which are closer than the radius selected can't be used to search neighbours because some galaxies may be missed and can affect our estimations. Results are shown in the figure (**??**). We can't see the trend we have expected with the density field, so we thought that it can be due to the large region in which we consider galaxies and we ran the same with a radius of 0.3°. Results are in figure (**??**).

In order to decide which redshift to assign to a galaxy in the mock catalogue which has been chosen to have a photometric redshift, we have estimated the distribution of photometric redshifts versus the spectroscopic redshift in the SDSS sample of spectroscoped

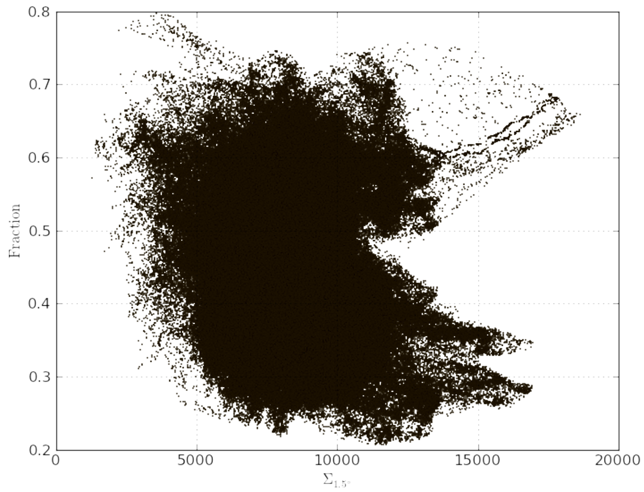


Figure A.4: Fraction of galaxies non-spectroscoped in the SDSS versus the local density field computed in a 1.5° radius region around galaxies not to close than this radius to the border of the survey. Density is in unit of galaxies per degree^2.

galaxies. Results show that a normal distribution is a well fit for those distributions, so we get for this parameters in figure (**??**).

In the mock catalogue, we have interpolated this parameters and we assign a photometric redshift for a galaxy chosen to be in fibre collision.

A.3 Coverage of the SDSS

For many computations in this thesis, we need to determine the surface covered on the sky by the data in our selection. The way we have selected galaxies allows us to easily determine if a point in the sky is in our area, so we can use a Monte Carlo process to compute this area.

First, we generate a number N of points around a point of coordinates (α_0, δ_0) with a maximal angular separation $\theta_{\max}$ which is larger than the maximal angular separation in our sample. The fraction of those points which reside in our selection area gives the area of the selection in fraction of the area of points generation. This area is just $2\pi\,(1 - \cos\theta_{\max})$. . I have made this calculation for different cone angle $\theta_{\max}$ and for different number of points to see if we have a convergence in the value of the area. Results are shown on figure (**??**).

I think I don't have squared angles in the expression of the solid angle.

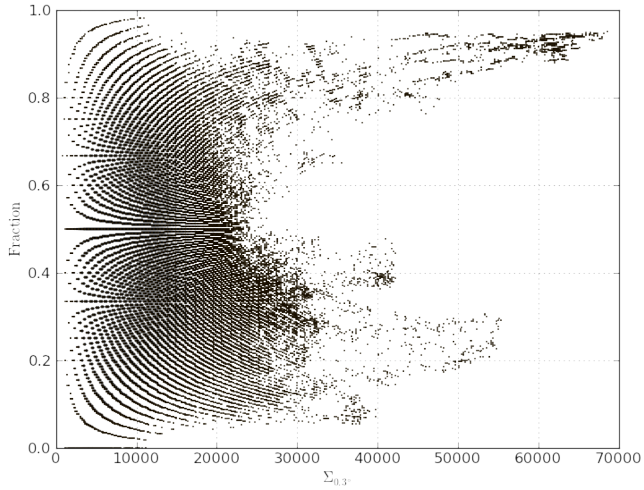


Figure A.5: Fraction of galaxies non-spectroscoped in the SDSS versus the local density field computed in a 0.3° radius region around galaxies not to close than this radius to the border of the survey. Density is in unit of galaxies per degree^2.

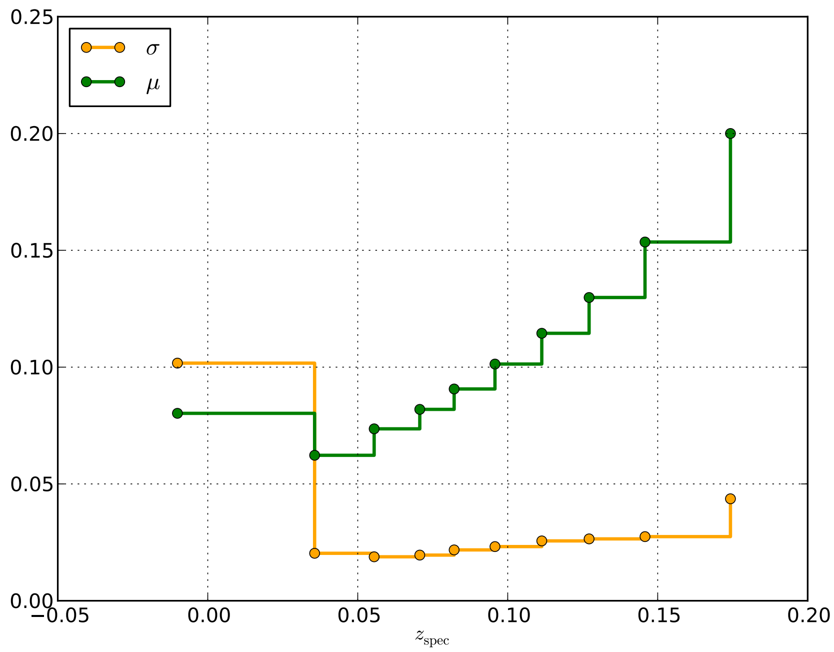


Figure A.6: Parameters of a normal distribution for photometric redshifts versus spectroscopic redshifts in the SDSS.

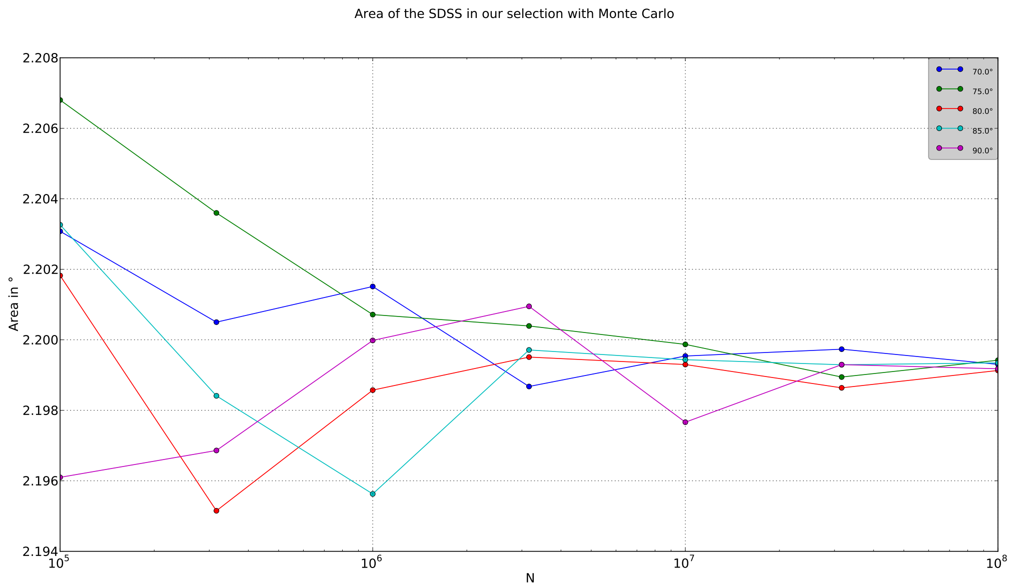
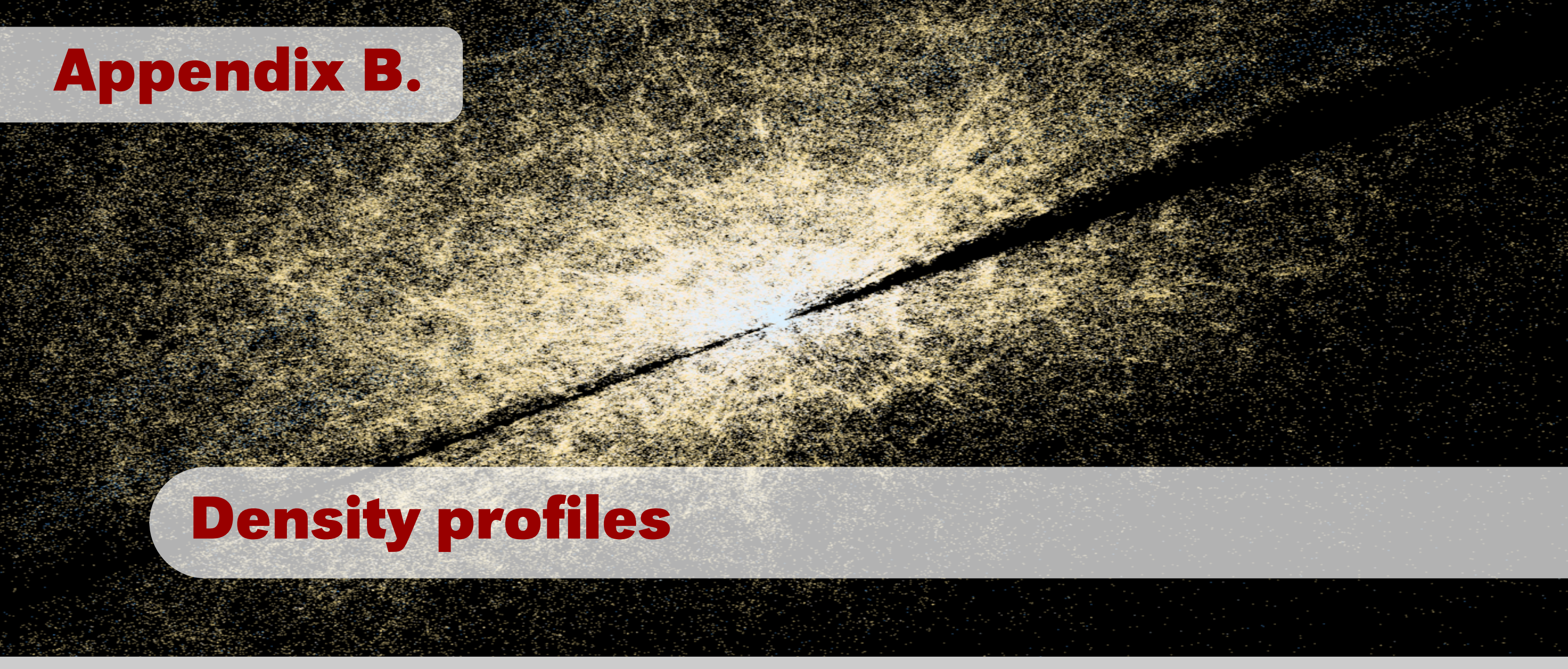


Figure A.7: Determination of the area of the SDSS for our selection with a Monte Carlo process. Results seem to converge on a value of 2.1993 steradians.

THESIS



Appendix B.

Density profiles

B.1 Introduction

In this chapter, there is some details on the computation of the density profiles and there derived quantities. We define here the different normalization used along the thesis for some useful current density profiles.

B.1.1 Definitions

The number of galaxies in a sphere of radius r with a density profile in number $\nu(r)$ is the case of a spherical symmetry:

$$N(r) = \int_0^r 4\pi r'^2 \nu(r') \mathrm{d}r' \tag{B.1}$$

B.2 Navarro et al. [1]

To start, we define some functions with no dimension used to make easier some computations.

$$\begin{aligned} N(r) &= N(a)\widetilde{N}(r/a) \\ \nu(r) &= \frac{N(a)}{4\pi a^3}\widetilde{\nu}(r/a) \end{aligned}$$

with a the radius at which the slope of the density profile is equal to -2.

B.2.1 Definitions

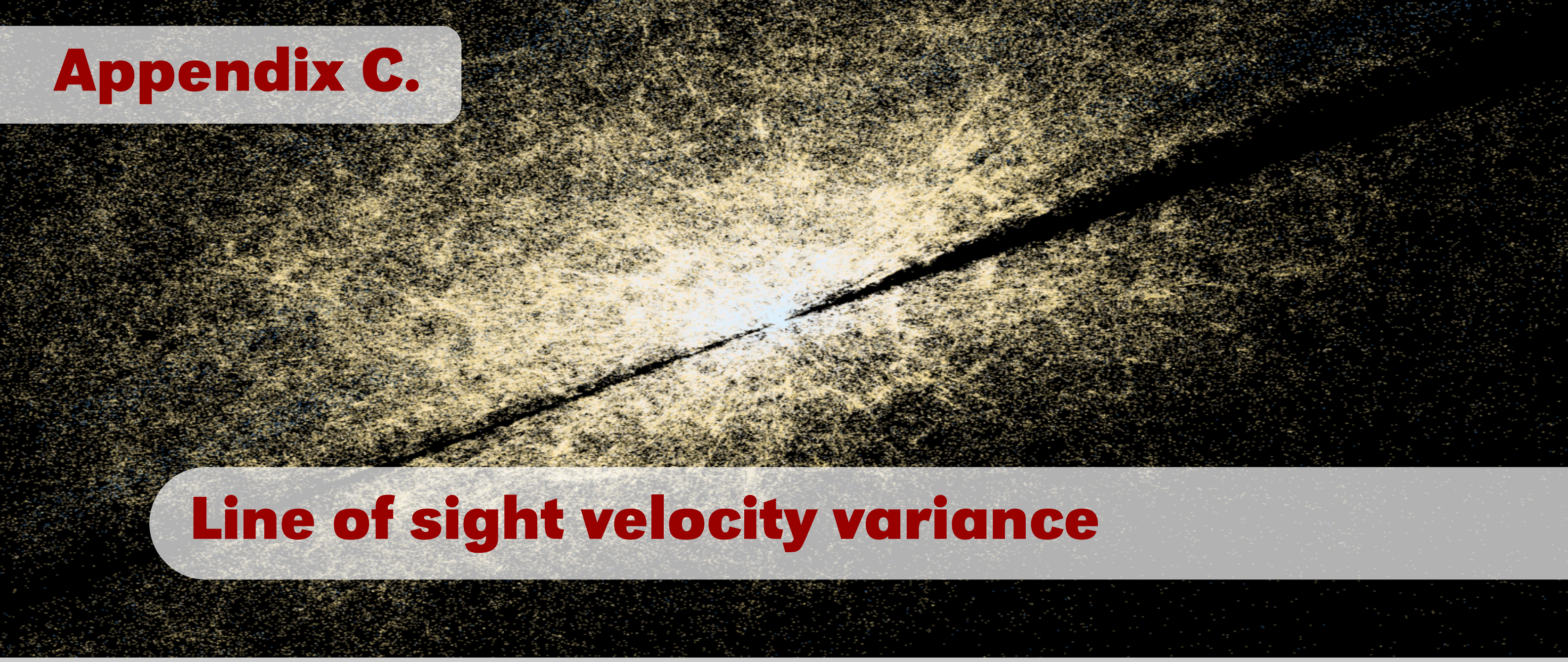
In the case of an NFW profile, we can write:

$$\widetilde{\nu}(x) = \frac{1}{\ln 2 - 1/2} \quad \frac{1}{x(1+x)^2}$$

and by integration:

$$\widetilde{N}(x) = \frac{1}{\ln 2 - 1/2} \quad \left(\ln\left(1+x\right) - \frac{x}{x+1} \right)$$

B.2. Navarro et al. [1]



Appendix C.

Line of sight velocity variance

C.1 Introduction

We will compute in this section the line of sight velocity dispersion of galaxies in a general spherical density profile, and then compute it specifically for an NFW profile.

This useful to make cuts at some sigma in the velocity profile to check where is the most important part of a group.

C.2 Calculus

By definition, the variance is the mean of the squared quantity under the assumption of a distribution function. We use a general density profile which is invariant under rotations $\nu(r)$. In our case, we make this mean on the line of sight, so:

$$\sigma^2_{LOS}\left(R\right) = \frac{\int_{-\infty}^{\infty} v^2_{LOS}\nu(r)\mathrm{d}z}{\int_{-\infty}^{\infty} \nu(r)\mathrm{d}z} \tag{C.1}$$

But in the group $r^2 = R^2 + z^2$ so:

$$\sigma^2_{LOS}\left(R\right) = \frac{2\int_R^{r_{\max}} v^2_{LOS}\frac{\nu(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r}{2\int_R^{r_{\max}} \frac{\nu(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r} \tag{C.2}$$

The denominator is by definition the projected density surface along the line of sight and we denote it

$$\Sigma(R) = 2\int_R^{r_{\max}} \frac{\nu(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r \tag{C.3}$$

Normally the integration is for $r_{\max} \to \infty$ but in our case we want to restrict to a limited region in the group (to virial sphere precisely).

In the same coordinate system as previously, the line of sight velocity can be expressed in spherical coordinates as:

$$v_{\mathrm{LOS}} = v_r \cos\theta - v_\theta \sin\theta \tag{C.4}$$

We suppose that we are at the equilibrium and so that there is no flow in the group in consequence we can neglect means of velocities. In terms of velocity variance we have now:

$$\Sigma(r)\sigma^2_{\mathrm{LOS}}(R) = 2\int_R^{r_{\max}} \left(\sigma^2_r(r)\cos^2\theta + \sigma^2_\theta \sin^2\theta\right) \frac{\nu(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r \tag{C.5}$$

Check word sphericity!

If we want to use the anisotropy parameter $\beta(r) = 1 - \sigma^2_\theta(r)/\sigma^2_r(r)$ in case of sphericity , we can write:

$$\Sigma(r)\sigma^2_{\mathrm{LOS}}(R) = 2\int_R^{r_{\max}} \left(1-\beta(r)\frac{R^2}{r^2}\right) \frac{\nu(r)\sigma^2_r(r)r}{\sqrt{r^2-R^2}}\mathrm{d}r \tag{C.6}$$

We can compute the radial velocity dispersion using the Jeans equation for a spherical system at equilibrium.

$$\frac{\partial\left(\nu(r)\sigma^2_r(r)\right)}{\partial r} + \frac{2\beta(r)}{r}\left(\nu(r)\sigma^2_r(r)\right) = -\nu(r)\frac{GM(r)}{r^2} \tag{C.7}$$

The solution to this equation is given by:

$$\nu(r)\sigma^2_r(r) = \int_r^{r_{\max}} K_r(r,s)\nu(s)\frac{GM(s)}{s^2}\mathrm{d}s \tag{C.8}$$

with $K_r(r,s)$ the kernel of the integral defined as:

$$K_r(r,s) = \exp\left[2\int_r^s \beta(r)\frac{\mathrm{d}t}{t}\right] \tag{C.9}$$

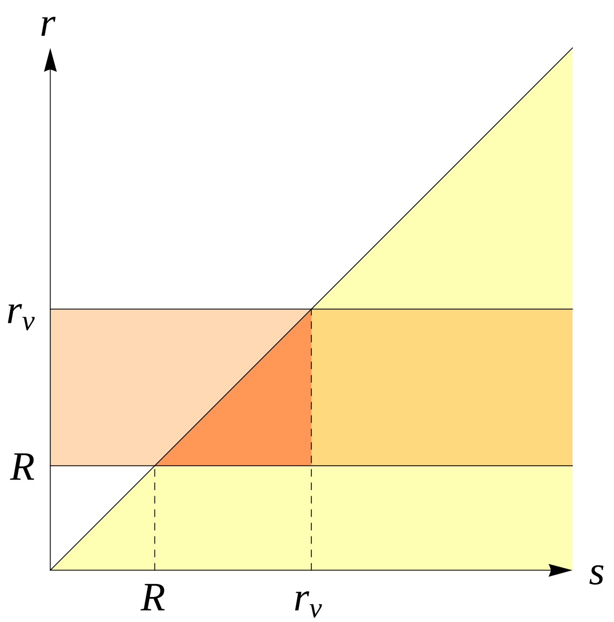


Figure C.1: Integration domain for the line of sight radial dispersion.

C.2.1 Supposing Mamon and Łokas [2] anisotropy

With the decomposition of the integral over the domain of integration, we can write

$$\begin{aligned}\Sigma(R){\sigma_{LOS}}^2(R) &= 2\int_R^{r_v}\frac{(s+a)}{s^2}\nu(s)GM(s)\mathrm{d}s\\ &\times\left(\int_R^s\left(\frac{r}{r+a}-\frac{1}{2}\left(\frac{R}{r+a}\right)^2\right)\frac{1}{\sqrt{r^2-R^2}}\mathrm{d}r\right)\\ &+\ 2\int_{r_v}^{\infty}\frac{(s+a)}{s^2}\nu(s)GM(s)\mathrm{d}s\\ &\times\left(\int_R^{r_v}\left(\frac{r}{r+a}-\frac{1}{2}\left(\frac{R}{r+a}\right)^2\right)\frac{1}{\sqrt{r^2-R^2}}\mathrm{d}r\right)\end{aligned} \tag{C.10}$$

where we are setting $r_{\max}$ to r_v. So now we can write using the fact that $4a\nu(a)\widetilde{\Sigma}(R/a,c)$:

$$\begin{aligned}{\sigma_{LOS}}^2(R) &= {v_v}^2\frac{c/2}{\widetilde{M}(c)\widetilde{\Sigma}(R/a,c)}\\ &\times\left(\int_{R/a}^{c}K\left(x\frac{a}{R},\frac{a}{R}\right)\widetilde{\nu}(x)\frac{\widetilde{M}(x)}{x}\mathrm{d}x+I\left(c\frac{a}{R},\frac{a}{R}\right)J(c)\right)\end{aligned} \tag{C.11}$$

$$I(u,u_a)=\begin{cases}-u_a\mathrm{sign}(u_a-1)\frac{{u_a}^2-1/2}{|{u_a}^2-1|^{3/2}}C^{-1}\left(\frac{1+uu_a}{u+u_a}\right) & \\ \qquad +\mathrm{acosh}u+\frac{1/2}{u_a+u}\frac{\sqrt{u^2-1}}{{u_a}^2-1}, & u_a\neq 1\\ \mathrm{acosh}u-\sqrt{\frac{u-1}{u+1}}\left(\frac{8+7u}{6(1+u)}\right), & u_a=1\end{cases} \tag{C.12}$$

with:

$$K(u,u_a)=\left(1+\frac{u_a}{u}\right)I(u,u_a) \tag{C.13}$$

and:

$$C^{-1}(X)=\begin{cases}\mathrm{acosh}X & u_a>1\\ \mathrm{acos}X & u_a<1\end{cases} \tag{C.14}$$

We have too an other integral:

$$J(y)=\int_y^{\infty}\frac{x+1}{x^2}\widetilde{\nu}(x)\widetilde{M}(x)\mathrm{d}x \tag{C.15}$$

In the case of an NFW profile, this can be expressed in an analytical way:

$$\begin{aligned}J(y) &= \frac{2}{3y^2(1+y)\left(\ln 4-1\right)^2}\left(y\left(-3+y\left(-9+\pi^2\left(1+y\right)\right)\right)\right.\\ &+3y^3\ln\left(1+\frac{1}{y}\right)+3\ln\left(1+y\right)\left(1-y+y^2\left(1+y\right)\ln\left(1+y\right)\right)\\ &\left.-3y^2\ln\left(y\left(1+y\right)\right)+6y^2\left(1+y\right)\mathrm{Li}_2\left(-y\right)\right)\end{aligned}$$

where the dilogarithm function is defined in our case as:

$$\mathrm{Li}_2\left(z\right) = -\int_0^1 \frac{\ln\left(1 - zt\right)}{z}\mathrm{d}t \tag{C.16}$$

Still in the case of the NFW profil, in Mamon et al. [6] there is the expression of $\widetilde{\Sigma}$:

$$\begin{aligned}\widetilde{\Sigma}(X,c) &= \frac{1}{2\ln 2 - 1}\int_X^c \frac{\mathrm{d}x}{(1+x)^2\sqrt{x^2 - X^2}} \\ &= \frac{1}{2\ln 2 - 1}\begin{cases} \frac{1}{(1-X^2)^{3/2}}\cosh^{-1}\left[\frac{c+X^2}{(c+1)X}\right] - \frac{1}{(c+1)}\frac{\sqrt{c^2-X^2}}{1-X^2} & \text{if } 0 < X < 1 \\ \frac{\sqrt{c^2-1}(c+2)}{3(c+1)^2} & \text{if } X = 1 < c \\ \frac{1}{(c+1)}\frac{\sqrt{c^2-X^2}}{X^2-1} - \frac{1}{(X^2-1)^{3/2}}\cos^{-1}\left[\frac{c+X^2}{(c+1)X}\right] & \text{if } 1 < X < c \\ 0 & \text{if } X = 0 \text{or} X > c \end{cases}\end{aligned} \tag{C.17}$$

Appendix D.

Generate mock catalogues

D.1 Introduction

A mock catalogue is a useful tool to test algorithms involving galaxies in order to see if it is operational in a realistic situation. Many of the properties of galaxy surveys can be simulated: the spatial clustering of galaxies, luminosity function, incompleteness and measures errors are some examples of them. There are different methods to obtain such a mock catalogue. All of them involves cosmological simulations and there halos of dark matter. According to the model of galaxy formation, we can use halo occupation distribution (HOD) to populate dark matter haloes with galaxy and putting some LF as constraint. We can too follow galaxies in semi analytical models (SAM) in those cosmological simulations outputs in order to have statistical properties of galaxies which agree with observational results. Then with such realistic galaxies we can use those simulation boxes to place an observer into it and create a mock survey. But to have a realistic mock catalogue, it's necessary to take care of many things which will be described in the next section.

Add things to this section!

D.2 Mock structure

In all this section, we will assume that we have already in our possession a dark matter simulation box which has been populated with galaxies with one of the methods described below (SAM, HOD...). At this step, physical properties of those galaxies aren't interesting.

D.2.1 Placing boxes

The first step to make a mock catalogue is to get galaxies positions like in a survey, to get an (α, δ) frame to simulate the sky coverage of survey and, at the same time, project galaxies on the sky, masking to us some spatial modulations of galaxy properties.

We want that a false observer see the same volume extension that a true survey. For example for the SDSS survey, we can measure redshift to a value of 0.3 (and more!). But the problem is that the majority of the simulation boxes have a size of around $L_{\rm box} = 100 - 300h^{-1}$ Mpc, letting us with a maximal redshift in our false survey of around $H_0 L_{\rm box}/c \approx 0.025$ in the case of a box of $100h^{-1}$ Mpc sized. Bigger simulations exist, and maybe can allow us to access to bigger redshifts, but this increasing size reduces the resolution of the simulation in particle mass and therefore we can't have low mass halos in the simulations.

The solution is to take a "little" simulation box and to replicate it and to make some bigger "Tetris" cube until we reach the maximal redshift we want. An example of the resulting "mock cube" is shown on figure (??).

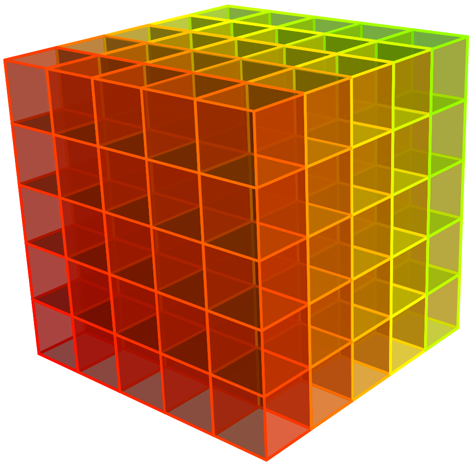


Figure D.1: The structure of the mock catalog once we have replicated the simulation box chosen to populate dark matter halos.

Maybe add details if I use it later...

Now if we take an observer at some position into this big box, we can have different sky coverage for the observer. The simplest is to place the observer at a corner which we give a solid angle of $\pi/2$ steradians. At the centre, we have a full sky coverage but we reduce the redshift extension by 2. At this time we don't care about a redshift evolution of galaxies for the observer, but if we want to care about it, we need to use other snapshots at different redshifts. Box sizes are similar in comobile coordinates, and different in physical coordinates due to the variation of the Hubble constant with redshift (h depends en z). So juxtaposing cubes isn't as easy as the case without redshift evolution.Many simulations give coordinates in units of h^{-1}Mpc but we place galaxies in the mock in physical positions which we can measure. So coordinates in the mock catalog are scaled to get positions in units of Mpc.

Placing boxes as described previously can create a perspective effect from the point of view of an observer, and the consequences aren't predictable in a statistical sense when we try to use the mock catalog. To avoid this, we apply some transformations on galaxies in the initial cube like inversions, rotations and periodic translations. Rotations are multiples of $\pi/2$ around the three principal coordinates axes, because if other rotations are allowed, it can create some over-densities in some regions of the final mock which aren't physical. An example of such a problem is illustrated in figure (??). Translations are performed on the three principal axes and when galaxies are out of the initial cube, periodic conditions are applied. All of those transformations are randomly generated for each cube in the final mock catalog.

D.2.2 Physics

We have now galaxies in a realistic coverage of the Universe, restricted to a given volume. But an observer can't use it because galaxies are seen in projection and he doesn't have any idea of the real distance of galaxies from him in a statistical sense. So we need to add physic and errors to those galaxies.

Survey mask

When studying galaxies in a survey, we are confronted to the problem of the area of this survey. Limits are not defined in a easy analytical way in many cases which can create some problems to correct for observers. The first step to simulate this is to transform cartesian coordinates in the 3D space to celestial coordinates ((α, δ) frame). Get these coordinates is the same as to compute spherical coordinates.

$$\alpha = \begin{cases} \arctan2(Y,X) + 2\pi & \text{if} \quad Y > 0 \\ \arctan2(Y,X) & \text{else} \end{cases}$$

$$\delta = \operatorname{sign}(Z) \arccos\left(\frac{\sqrt{X^2+Y^2}}{\sqrt{X^2+Y^2+Z^2}}\right) \tag{D.1}$$

In our case, the origin of coordinates is the observer. If we keep the distance as calculated previously, the observer can still have precise determination of the distance of a galaxy. In reality, we observe it in redshift space so the redshift as distance indicator is biased by peculiar velocities. Our initial galaxy catalog allow us to get the velocity of a galaxy, so we can compute the line of sight (los) velocity of this galaxy relatively to the observer.

$$v_{\text{los}} = \frac{\vec{OG}.\vec{v_{\text{pec}}}}{||\vec{OG}||} \tag{D.2}$$

where O is the observer and G the galaxy, $\vec{v_{\text{pec}}}$ its peculiar velocity. This velocity has a sign. The redshift is just the expression a shift in wavelength due to a velocity. The observed wavelength λ is linked to the original (emitted) wavelength λ_0 by:

$$\lambda = (1+z)\lambda_0 \tag{D.3}$$

The shift caused by Universe expansion is $\lambda_{\text{cos}} = (1+z_{\text{cos}})\lambda_0$ where the subscript cos refer to the cosmological expansion. The shift caused by the peculiar velocity is $\lambda = (1+z_{\text{pec}})\lambda_{\text{cos}}$. So the observed wavelength is $\lambda = (1+z_{\text{pec}})(1+z_{\text{cos}})\lambda_0$. The resulting observed redshift is just:

$$(1+z) = (1+z_{\text{pec}})(1+z_{\text{cos}}) \tag{D.4}$$

expansion. The peculiar redshift is the just due to the relativist Doppler effect:

$$(1+z_{\text{pec}}) = \sqrt{\frac{1+\beta}{1-\beta}} \tag{D.5}$$

with $\beta = v_{\text{los}}/c$. The cosmological redshift is approximated by $z_{\text{cos}} = H_0 D/c$ where D is the physical distance of the galaxy to the observer and H_0 the Hubble constant. Applying this method to the galaxies in the mock catalog, we can have galaxies whose distance is biased by peculiar velocities in redshift space. With such a treatment, the velocity dispersion of galaxies in groups leads to the apparition of "fingers of God" as seen in observations in redshift space.

I think we need to add the velocity of the Local Group in the redshift, because in our case the observer has a null velocity. Maybe corrected in SDSS

With our frame in redshift space relative to the observer, we can apply different masks on angular coordinates according to the survey we want to mimic. An example of such a mask is in appendix (A).

K-corrections

In reality, an observer study galaxies in a given bandwidth in wavelength and can't use the bolometric flux of the object. With the expanding Universe, all the spectral energy distribution (SED) of galaxy is shifted. All wavelengths are shifted by the same value for a given redshift. So, knowing the luminosity L of a galaxy in a given band in reality (using the true SED), computing its apparent magnitude for an observer isn't as easy as correcting for the distance modulus. The observer in the same band sees a different part of the true SED. The flux observed in the same band as the true flux is maybe higher or lower. A correction for this effect is needed and must be taken into account in our mock catalogue.

As explained before, this correction depends on the SED of galaxies and the band used in the survey. The common way of correcting it when we have a multi-band photometry is to fit the observed SED in those bands with theoretical templates of SEDs. Such templates can be obtained with existing programs as PEGASE , which give us SEDs with some assumptions on the galaxy. But those programs are a little time consuming, which can be a problem for mock when we want to run several of them. A good solution is given by Chilingarian et al. [7], where the K-correction is fitted on templates for SED as given by PEGASE in terms of a polynomial of the redshift of the galaxy and its colour. The corresponding K-correction is precise for redshifts until 0.3 in different survey bands (including *ugriz* for the SDSS). This work reduces the computation of K-corrections to the use of simple polynomial relations and make easier our task.

Add references

By definition, the K-correction K for a galaxy of apparent magnitude m_X in a given band X and absolute magnitude M_X in the same band is:

$$m_X = M_X + 5\log_{10}\left(d_{\text{lum}}\,[pc]\right) - 5 + K \tag{D.6}$$

In our case, the K-correction depends on the redshift of the galaxy and its colour in apparent magnitude given two bands. So we can rewrite:

$$m_X = M_X + 5\log_{10}\left(d_{\text{lum}}\,[pc]\right) - 5 + K(z, m_X - m_{X'}) \tag{D.7}$$

where:

$$K(z, m_X - m_{X'}) = \sum_{i=0}^{N_i}\sum_{j=0}^{N_j} a_{ij} z^i (m_X - m_{X'})^j \tag{D.8}$$

and a_{ij} is a $N_i \times N_j$ matrix containing the coefficients of the two dimensional polynomial. These coefficients depend on the bands of the survey used for the colour computation.

The observer in the mock can just, in theory, access to apparent magnitude of the survey. But we don't know in advance these magnitudes, and as we can see in the expression of equation (D.7), we need apparent magnitudes to compute apparent magnitudes. If we use the other bands of the survey, with a_{ij} coefficients, we can always write a set of equations for a galaxy which involves all apparent magnitudes of the survey. So we can write a set of non linear equations with polynomial of order N_j (redshift of the galaxy is supposed to be known). Numerically it's easy to solve this set of equations, and relatively fast with equations solvers or by iterations. In practice, the first is faster than the second method, unless both methods give similar results in apparent magnitudes.

Flux limit

We have seen in appendix (A) that spectroscoped galaxies are just defined for galaxies whose apparent magnitude is less than 17.77 in r band. So, in all the redshift sample, we miss some galaxies not sufficiently bright. To take into account this effect, we remove galaxies which don't reach the limit apparent magnitude of the survey.

Some surveys have variation of this flux limit with the sky position. Maybe something like this to check for SDSS?

Spectroscopic and photometric redshifts

Sometimes, we can't access to spectroscopic redshifts which are more precise than photometric redshifts. In the SDSS, for example, this is due to tiling process. Fibers analysing the spectrum of galaxies can't be closer from each other than 55", so if for a target galaxy (selected to get a spectrum) there is an other galaxy closer than those 55", the tile containing all fibers doesn't have the possibility to measure the redshift of this galaxy. This problem is more significant for dense regions in the celestial plane. A very good algorithm to placing tiles in order to limit the number of missed galaxies (*ie* the number of fiber collision) has been applied in the galaxy sample of the SDSS. But there is still some galaxies without spectroscoped redshifts. If we remove those galaxies from our sample, there will be a spectroscopic incompleteness with unknown effects on our results.

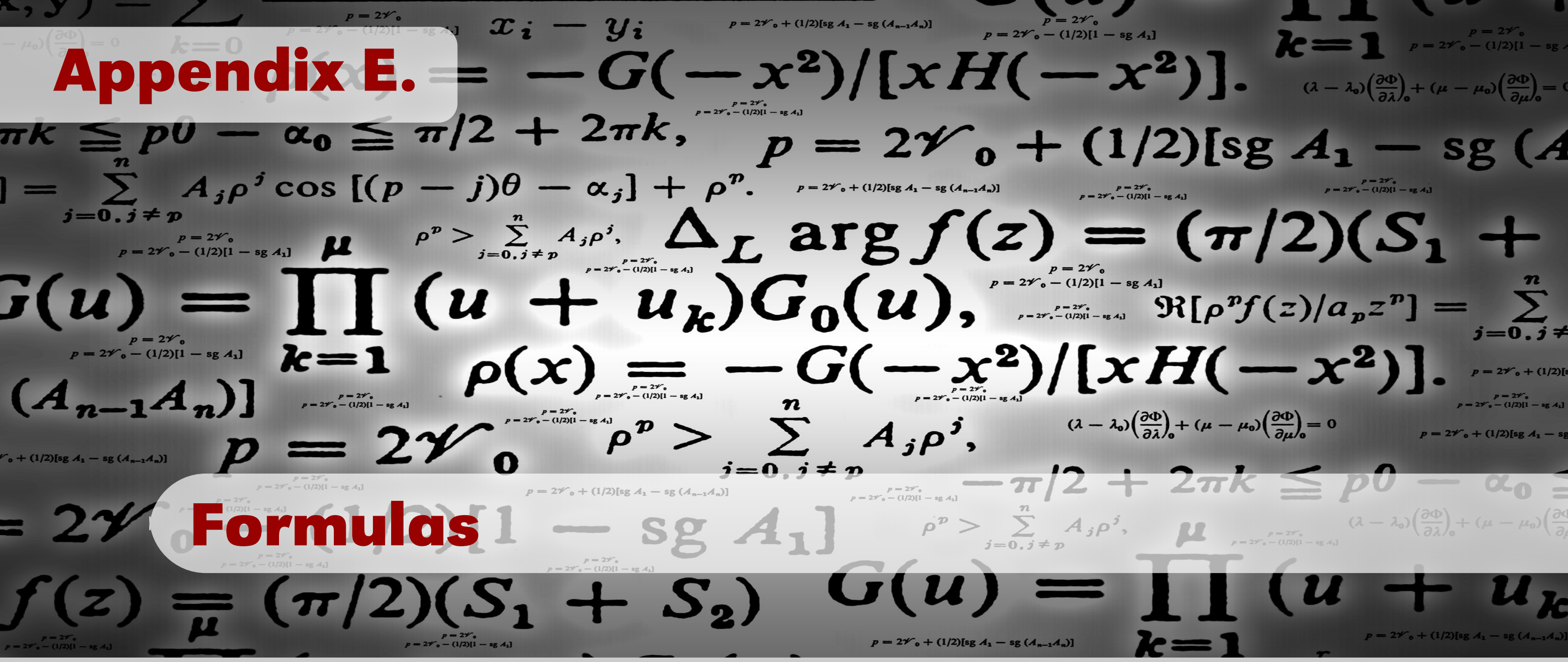
To "correct" this problem, non-spectroscoped galaxies will be affected a photometric redshift in our mock catalogue in the sense that we will affect a redshift following a normal distribution modulate mean and dispersion depending the distance of the nearest spectroscoped neighbour and the magnitude difference between them. It is applied randomly in a given number of galaxies in the mock catalogue according to the fraction of non-spectroscoped galaxies in the survey we want to mimic.

Observational errors

Errors exist in redshift measurement, photometry, astrometry... We add them in our mock catalogue by simply assuming a distribution for errors around the true value, and then applying it to the galaxy in our mock catalogue..

Think in a way of adding the surface brightness selection criterion on the mock catalogue. It will be great!

THESIS



Appendix E. Formulas

E.1 Introduction

In this appendix are described the formulas used in all computations realized during my thesis. Its just a simple way to share and verify that the job id done correctly. References to those formulas are indicated too, in order to improve search when some doubts are presents.

Add a little more in the introduction.

E.2 Formulas

E.2.1 Cosmology

Formulas which are related to the cosmology.

The luminosity distance is defined as the relation between the galaxy flux S and its absolute luminosity L by:

$$d_{\text{lum}} = \sqrt{\frac{L}{4\pi S}} \tag{E.1}$$

An analytical precise computations isn't possible, but numerical computations exist. Although precise, numerical recipes aren't sufficiently fast in practice. Some other analytical approximations of this distance was created.

For example, in Wickramasinghe and Ukwatta [8], an approximation good to 0.3% is available for a range of values in Ω_Λ compatible with WMAP and Planck results.In this approximation we have:

$$d_{\text{lum}}(z) = \frac{c}{3H_0} \frac{1+z}{\Omega_\Lambda^{1/6}(1-\Omega_\Lambda)^{1/3}} \left[\Psi\left(x\left(0, \Omega_\Lambda\right)\right) - \Psi\left(x\left(z, \Omega_\Lambda\right)\right)\right] \tag{E.2}$$

with:

$$\Psi(x) = 3x^{1/3}2^{2/3}\left[1 - \frac{x^2}{252} - \frac{x^4}{21060}\right] \tag{E.3}$$

$$x(\alpha) = \ln\left(\alpha + \sqrt{\alpha^2+1}\right) \tag{E.4}$$

$$\alpha\left(z,\Omega_\Lambda\right)=1+2\frac{\Omega_\Lambda}{1-\Omega_\Lambda}\frac{1}{\left(1+z\right)^3} \tag{E.5}$$

The other distances are simply linked to this luminosity distance. The angular distance d_{ang} and the proper distance d_{pm} are $d_{\text{lum}}(z)=(1+z)^2 d_{\text{ang}}(z)=(1+z)d_{\text{pm}}(z)$.

The element of comoving volume is expressed using the Robertson-Walker metric as:

$$\mathrm{d}V=\frac{c}{H\left(z\right)}d_{\text{pm}}\left(z\right)^2\mathrm{d}\Omega\mathrm{d}z \tag{E.6}$$

The evolution of the fraction of matter, and dark energy is the following:

$$\Omega_m\left(z\right)=\Omega_{m,0}\frac{\left(1+z\right)^3}{E\left(z\right)^2} \tag{E.7}$$

$$\Omega_\Lambda\left(z\right)=\frac{\Omega_{\Lambda,0}}{E\left(z\right)^2} \tag{E.8}$$

where z is the redshift and the subscript 0 refers to the actual value of the parameter.

The distance modulus represents the magnitude difference betweenthe observed flux of the galaxy and it would be if the galaxy was at a distance of $10pc$. So it's:

$$DM\left(z\right)=5\log_{10}\left(\frac{d_{\text{lum}}\left(z\right)}{10pc}\right) \tag{E.9}$$

where z is the redshift of the galaxy and d_{lum} is the luminosity distance.

The apparent magnitude m of galaxy in the perfect case where isn't K-correction, extinction…, is just:

$$m=M+DM\left(z\right) \tag{E.10}$$

where M is the absolute magnitude of this galaxy in the same band of m and $DM(z)$ is the distance modulus at redshift z.

Magnitudes are defined at a given constant which is the same for each object so:

$$M-M_\odot=-2.5\log_{10}\left(\frac{L}{L_\odot}\right) \tag{E.11}$$

where M is absolute magnitude, L the luminosity of the object and $\odot$ refers to Sun's quantities. We can determined the luminosity by this relation which gives:

$$\frac{L}{L_\odot}=10^{0.4(M_\odot-M)} \tag{E.12}$$

For galaxies at a given redshift z, we can see all galaxies with an absolute magnitude lower than (using equation (E.10)):

$$m_{\text{lim}}=M+DM\left(z\right) \tag{E.13}$$

where m_{lim} is the apparent magnitude limit for a survey, and M is the absolute magnitude threshold to be seen at this redshift.

The virial radius r_Δ is defined as the radius at which the density is Δ times the critical density of the Universe. So we have:

$$\rho(r_\Delta) = \Delta\rho_c \tag{E.14}$$

with $\rho_c = \dfrac{3H(z)^2}{8\pi G}$.

If we suppose that the density is constant in this radius, we have:

$$\Delta\frac{3H(z)^2}{8\pi G} = \frac{M_\Delta}{4\pi {r_\Delta}^3/3} \tag{E.15}$$

where M_Δ is the virial mass. We can now defined three quantities, the virial mass as:

$$M_\Delta = \frac{\Delta H(z)^2 {r_\Delta}^3}{2G} \tag{E.16}$$

the virial radius as:

$$r_\Delta = \left(\frac{2GM_\Delta}{\Delta H(z)^2}\right)^{1/3} \tag{E.17}$$

and the virial velocity as:

$$v_\Delta = \sqrt{\frac{GM_\Delta}{r_\Delta}} = \sqrt{\frac{\Delta}{2}}H(z)\, r_\Delta \tag{E.18}$$

Sometimes, the density at the virial radius isn't defined in relation with the critical density but instead with mean density of the Universe. So the equation (E.14) becomes:

$$\rho(r_\Delta) = \Delta\rho_m = \Delta\Omega_m\rho_c \tag{E.19}$$

We can treat this situation in the same way as previously, but formally with $\Delta \to \Delta\Omega_m$.

Appendix F.

Halo mass functions

A description of how to compute halo mass functions given simple models in some articles.

F.1 Theory

F.1.1 Definition

By definition, the halo mass function by unit of comobile volume is the number of halos with mass M comprise between M and $M + \mathrm{d}M$. If N is the number of halos, the halo mass function $\phi(M)$ can be written:

$$\phi(M) = \frac{\mathrm{d}N^2}{\mathrm{d}M\mathrm{d}V} = \frac{\mathrm{d}n}{\mathrm{d}M} \tag{F.1}$$

In this case, n can be the comobile density of halos, or the CDF of the density. In the latter case, we have:

$$n(M,z) = \int_0^M \phi(M,z)\,\mathrm{d}M \tag{F.2}$$

and so:

$$\frac{\mathrm{d}n}{\mathrm{d}M} = \frac{\mathrm{d}}{\mathrm{d}M}\int_0^M \phi(M,z)\mathrm{d}M = \frac{\mathrm{d}}{\mathrm{d}M}\left(\Phi(M,z) - \Phi(0,z)\right) = \phi(M,z) \tag{F.3}$$

where Φ is a primitive of ϕ.

F.1.2 In practice

Cosmological simulations give results with $f(\sigma)$ a fitted function on simulations. $\sigma(M)$ is the variance in mass of the smoothed density fields. We can link this function to the halo mass function by:

$$\phi(M,z) = \frac{\mathrm{d}\ln\sigma^{-1}}{\mathrm{d}M}\frac{\rho_m(z)}{M} f(\sigma) = \frac{\rho_m(z)}{M^2}\left|M\frac{\mathrm{d}\ln\sigma}{\mathrm{d}M}\right| f(\sigma) \tag{F.4}$$

where the computation of σ involves the power spectrum $P(k)$ and the filter for spectrum $\tilde{W}(k)$:

$$\sigma^2(M) = \frac{1}{2\pi^2}\int_0^\infty P(k)\,\tilde{W}(k)^2\,k^2\mathrm{d}k \tag{F.5}$$

This form is time consuming for the computation of the halo mass function and model dependent. In [9], there is a good approximation for this formula which is resumed to:

$$\sigma(M) = \sigma_8 \frac{f(u)}{f(u_8)} \tag{F.6}$$

with the function f:

$$f(u) = 64,087(1 + 1,074u^{0,3} - 1,581u^{0,4} + 0.954u^{0.5} - 0.185u^{0.6})^{-10} \tag{F.7}$$

and u, u_8 which are:

$$\begin{aligned} u &= 3.804e-4\Gamma\left(\frac{Mh}{\Omega_{m,0}}\right)^{1/3} \\ u_8 &= 32\Gamma \\ \Gamma &= \Omega_{m,0}h\exp\left[-\Omega_b(1+\sqrt{2h}/\Omega_{m,0})\right] \end{aligned} \tag{F.8}$$

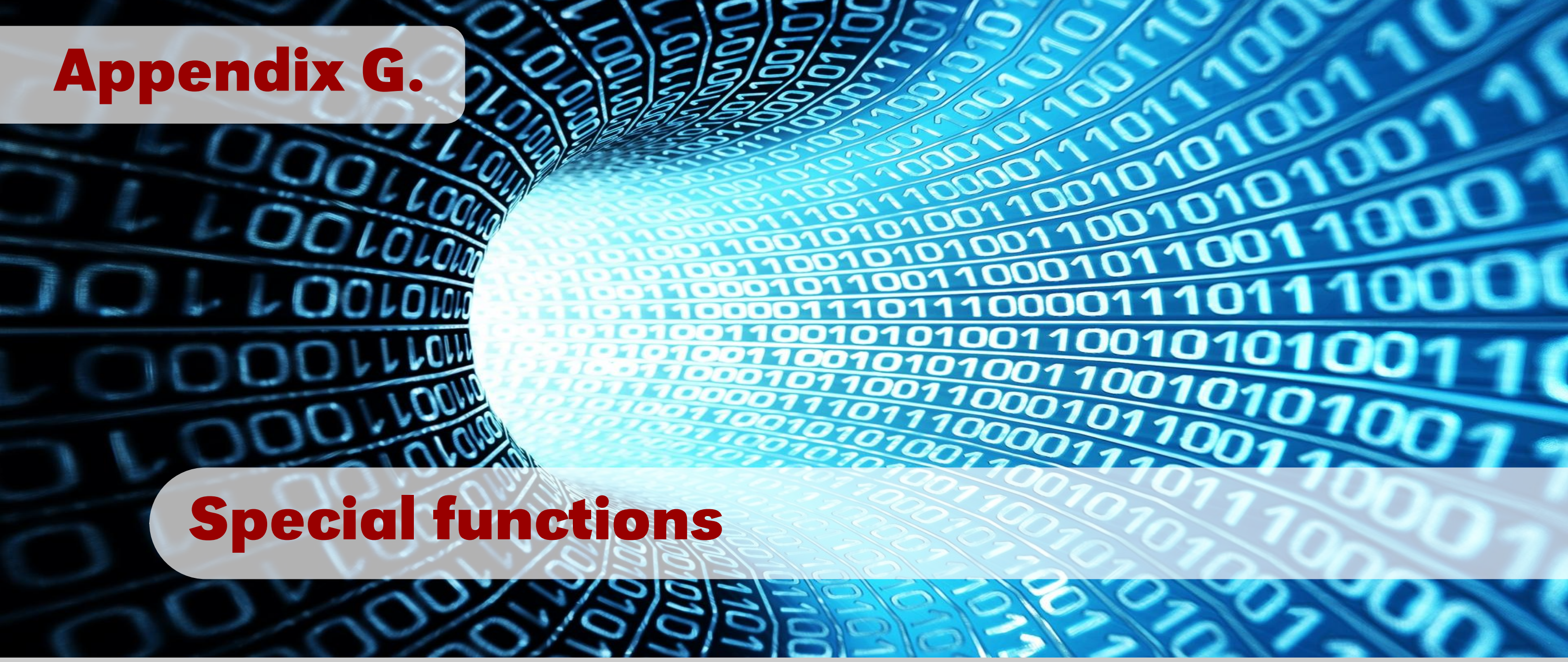
Now, with this approximation, we can compute easily the derivative of σ and:

$$\left(M\frac{\mathrm{d}\ln\sigma}{\mathrm{d}M}\right)^{-1} + \frac{1}{2} = \frac{(-0.000310111X^{1.7} + 0.00225895X^{1.6} - 0.00505879X^{1.5} - 0.1X^{1.2})}{(-0.000328357X^{1.8} + 0.00310111X^{1.7} - 0.0090358X^{1.6} + 0.0101176X^{1.5})} \tag{F.9}$$

Check if we can used directly the power spectrum in the calculation without too many CPU time consuming...

with:

$$X = \left(h\Omega_{m,0}e^{-\Omega_b\left(\frac{\sqrt{2}\sqrt{h}}{\Omega_{m,0}}+1\right)}\sqrt[3]{\frac{hM}{\Omega_{m,0}}}\right) \tag{F.10}$$



Appendix G. Special functions

G.1 Legendre elliptic integral function

G.1.1 Introduction

The Legendre elliptic integral function appears naturally when evaluationg distances like the luminosity distance in a flat Universe. But the most of the time, this function isn't used directly because of the difficulty of implementation, and when it's already adapted, this is not for all kinds of value. In following sections, we described how to use the NSWC implementation of elliptic integrals.

G.1.2 Luminosity distance

Our goal is to easily and precisely compute the luminosity distance given a cosmology according to the formula:

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{\mathrm{d}t}{\sqrt{\Omega_m (1+t)^3 + \Omega_\Lambda}} \tag{G.1}$$

Making the change of variable $u = 1/t$ and defining $s = \sqrt{3}(1-\Omega_m)/\Omega_m$ gives us:

$$d_L(z) = \frac{c(1+z)}{H_0\sqrt{s\Omega_m}} \left[T(s) - T\left(\frac{s}{1+z}\right) \right] \tag{G.2}$$

with

$$T(x) = \int_0^x \frac{\mathrm{d}u}{\sqrt{u^4+u}} \tag{G.3}$$

As described in **?**], this integral is an elliptic integral, and such integrals can be expressed in terms of Carlson symmetric forms $R_F(x_1, x_2, x_3)$:

$$R_F(x_1, x_2, x_3) = \frac{1}{2} \int_0^\infty \frac{\mathrm{d}t}{\sqrt{(t+x_1)(t+x_2)(t+x_3)}} \tag{G.4}$$

With help of the reduction theorem, we can write:

I can't remember how to do the computation. Some boring day, do it again for fun!

$$T(x) = 4R_F(m, m+3+2\sqrt{3}, m+3-2\sqrt{3}) \tag{G.5}$$

where

$$m(x) = \frac{2\sqrt{x^2 - x + 1}}{x} + \frac{2}{x} - 1 \tag{G.6}$$

G.1.3 Algorithm

G.2 Incomplete gamma function

G.2.1 Introduction

By default, many algorithm used to compute incomplete gamma function doesn't allowed to have negative parameters. By definition, the incomplete gamma function $\Gamma(a, x)$ is:

$$\Gamma(a, x) = \int_x^{\infty} e^{-t} t^{a-1} \mathrm{d}t \tag{G.7}$$

when $a \leq 0$, we can't compute this function with usual algorithms. Moreover, we need to used an algorithm which doesn't use the "simple" gamma function $\Gamma(a)$:

$$\Gamma(a) = \int_0^{\infty} e^{-t} t^{a-1} \mathrm{d}t \tag{G.8}$$

Indeed this function have singularities for negative values of a where a is an integer, as we can see in figure (**??**). So we need an algorithm which not involves to use the gamma

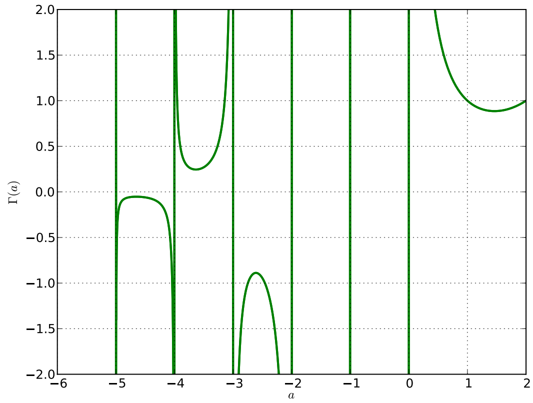


Figure G.1: The gamma function.

function for negative values. Here is described such algorithm.

G.2.2 Algorithm

Theory

The best way to compute the incomplete gamma function for a negative values is to use recurrence relations. Let us define:

$$\Gamma(a+1, x) = \int_x^{\infty} e^{-t} t^{a} \mathrm{d}t \tag{G.9}$$

Defining $u' = e^{-t}$ and $v = t^a$, we can use integration by parts:

$$\Gamma(a+1,x) = \left[-e^{-t}t^a\right]_x^\infty + a\int_x^\infty e^{-t}t^{a-1}\mathrm{d}t \tag{G.10}$$

The all integrated part is always zero for all values of a at infinity, and the second member of the right hand side of the previous equation lets appear the definition of the incomplete gamma function. So the recurrence relation for the incomplete gamma function is:

$$\Gamma(a+1,x) = e^{-x}x^a + a\Gamma(a,x) \tag{G.11}$$

We can see that computing the incomplete gamma function for $a <= 0$ can be done with a recursive function using the function at higher values of a.

$$\Gamma(a,x) = \frac{\Gamma(a+1,x) - e^{-x}x^a}{a} \tag{G.12}$$

The previous equation shows that there is still a problem for integer values of a because if $a = -2$ for example, at a moment in the recursion, we have a value of 0 for a which create problems. If we refer to Abramowitz and Stegun [10], the definition of the elliptical integral is:

$$E_n(z) = \int_1^\infty e^{-zt}t^{-n}\mathrm{d}t \tag{G.13}$$

for integer values of n. If we change the variable in the integral to $t' = zt$, we can rewrite the equation to have:

$$E_n(z) = z^{n-1}\Gamma(1-n,z) \tag{G.14}$$

so:

$$\Gamma(a,x) = x^a E_{1-a}(x) \tag{G.15}$$

for $a \leq 0$ and a integer. Now we have a good computation for the incomplete gamma function. But numerically, there is still a problem near integer negative values of a. If a is very close to an integer value, at a moment in the recursion, a is very small. So $1/a$ can be bigger than the overflow value for the machine. To avoid this, we add a condition for a when it is near zero.

An other definition of the incomplete gamma function is:

$$\Gamma(a,x) = \Gamma(a) - \gamma(a,x) = \Gamma(a)\left(1 - P(a,x)\right) \tag{G.16}$$

with:

$$\gamma(a,x) = \int_0^x e^{-t}t^a\mathrm{d}t \tag{G.17}$$

In Press et al. [11] exists a precise computation of the function $P(a,x)$. We can remark that this function isn't needed in the recursion if we have already access to a function which compute the incomplete gamma function for positive values of a.

Following is described the algorithm for computing incomplete gamma function without loss of precision and without numericals problems for negative values of a.

Numerical

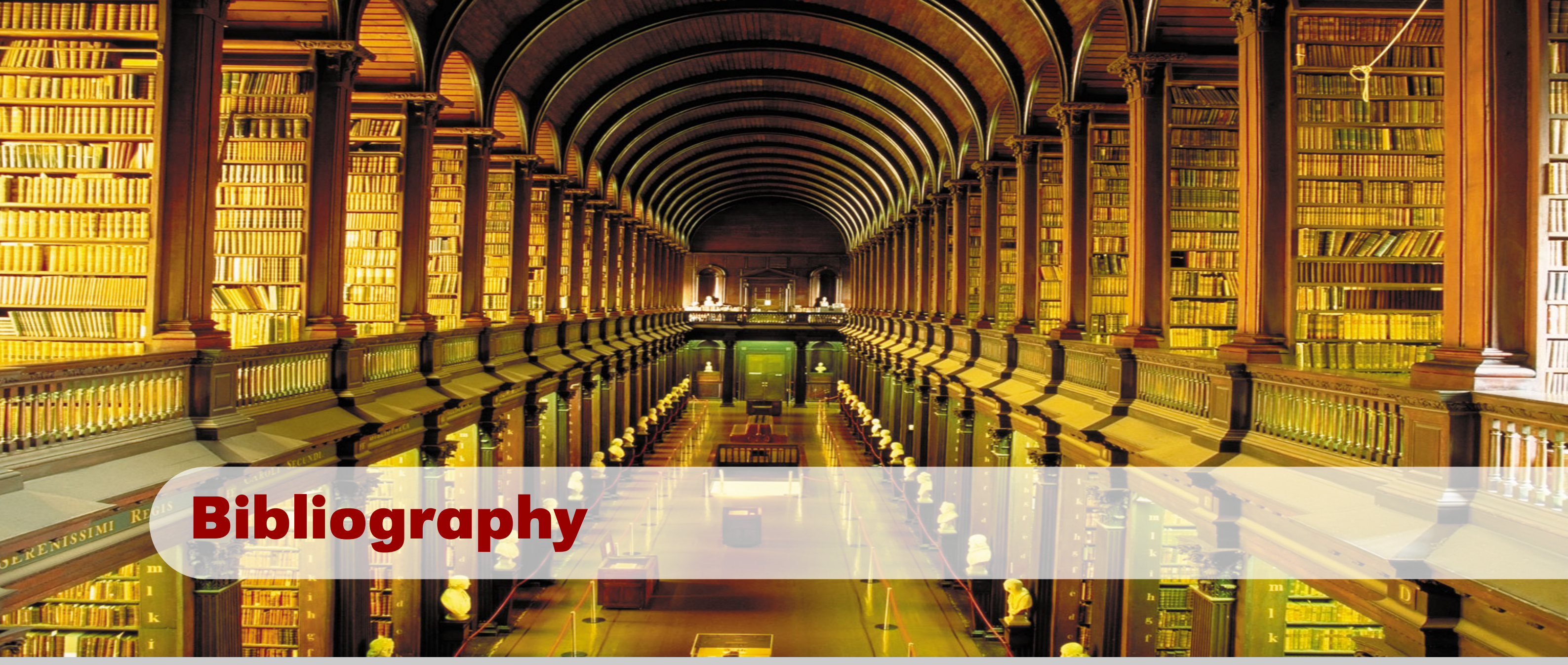
```
def gammainc(a, x):

    """
    To compute the incomplete gamma function
    without loss of precision or without numerical
    problems. OF is the value of the overflow for
    the machine and expint( n, x ) the function
    which computes the integral function for n and x.
    """

    import numpy as np

    if x >= 0.:
        if a <= 1.:
            if a == int(a) or OF * abs(a) < 1 :
                return x * int(a) * expint(1 - int(a), x)
            else :
                return (gammainc(a + 1, x) -
                    np.exp(-x) * (x ** a)) / a

        else :
            return gamma(a) (1 - P(a, x))
            # or call the function which computes
            # the incomplete gamma function for
            # positive values of a
```



Bibliography

[1] J. F. Navarro, C. S. Frenk, and S. D. M. White. A universal density profile from hierarchical clustering. ApJ, 490:493–+, December 1997.

[2] G. A. Mamon and E. L. Łokas. Dark matter in elliptical galaxies - ii. estimating the mass within the virial radius. MNRAS, 363:705–722, November 2005.

[3] M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, M. Fukugita, and J. Loveday. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. ApJ, 631:208–230, September 2005.

[4] Q. Guo, S. White, M. Boylan-Kolchin, G. De Lucia, G. Kauffmann, G. Lemson, C. Li, V. Springel, and S. Weinmann. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. MNRAS, pages 164–+, February 2011.

[5] I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, J. E. Gunn, R. H. Lupton, R. C. Nichol, W. J. Percival, D. P. Schneider, R. A. Skibba, M. A. Strauss, M. Tegmark, and D. G. York. Galaxy clustering in the completed sdss redshift survey: The dependence on color and luminosity. ApJ, 736:59, July 2011.

[6] G. A. Mamon, A. Biviano, and G. Murante. The universal distribution of halo interlopers in projected phase space. Bias in galaxy cluster concentration and velocity anisotropy? ApJ, 520:A30+, September 2010.

[7] I. V. Chilingarian, A.-L. Melchior, and I. Y. Zolotukhin. Analytical approximations of K-corrections in optical and near-infr ared bands. MNRAS, 405:1409–1420, July 2010.

[8] T. Wickramasinghe and T. N. Ukwatta. An analytical approach for the determination of the luminosity distance in a flat universe with dark energy. MNRAS, 406:548–550, July 2010.

[9] F. C. van den Bosch. The universal mass accretion history of cold dark matter haloes. MNRAS, 331:98–110, March 2002. doi: 10.1046/j.1365-8711.2002.05171.x.

[10] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth Dover printing, tenth GPO printing edition, 1964.

[11] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, volume 1. Cambridge University Press, 2 edition, September 1992. ISBN 9780521430647.