Data Analysis - Galactic Surveys

In Cosmology, observations of galaxy groups is very important for the study and verification of complex theory models, Universe expansion, and its behaviour analysis. There are numerous telescopes used only for galactic surveys and making special statistical analysis. The product of all this are the catalogues of galaxies, which are the result of this observations and complex computational analysis.
Usually, galaxies form in groups inside very large dark matter Halos (that have a spherical shape). The most matter is concentrated in the center, thus having a central large-mass galaxy there is with high probability. In these problem, we will assume that all Halos have one central galaxy and several satellite galaxies.
As one can guess, all of these requires a large amount of computational power: To distinguish galaxies, process corrected coordinates with redshifts and other methods, etc. The complex statistical analysis involves calculating correlation functions for galaxy groups, which is making a histogram for distances between all possible two galaxy pairs. The main goal of this problem is to do theoretical statistical analysis of the observational data and figure out what kind of galaxy groups have been observed. Having a theoretical model that can deduce the characteristics of a galaxy groups with few observations is very useful for scientists and saves lots of computational time and possible errors following those kind of data analysis.
In Figure 1, you can see the probability densities of different galaxy groups. On X axis is the distance, and on Y axis is probability density. There are six galaxy cluster types (A-F) with corresponding densities. The goal of this problem will be to determine what kind of Dark matter Halo groups we are observing.

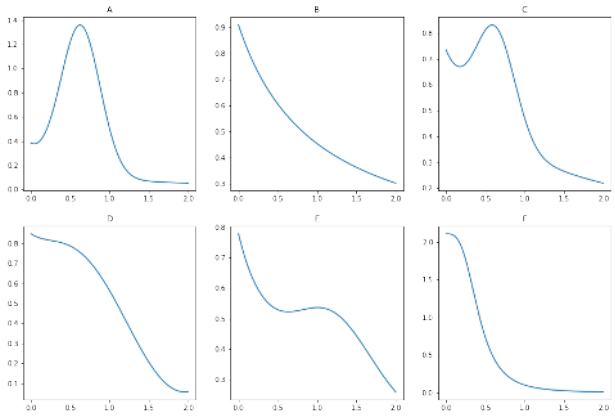


Figure 1: Probability density of all possible galaxy pairs (from different/same dark matter Halos)

Part 1 (35p)

You are given Table 1. that represents the data from observations. From other data and observations it was concluded that all these galaxies belong to the same dark matter halo. You can see the coordinates of satellite galaxies. The coordinate for the central galaxy is $8.300; 6.200; 1.100$. Additional column represents the number of galaxies spotted very near to that points. For simplicity, you can assume that the coordinates are the same for them. These coordinates are given in Cartesian coordinate system center of which coincides to the observer on Earth.

	x (MPC)	y (MPC)	z (MPC)	N (Number of galaxies)
1	8.401	6.309	1.394	291
2	9.883	7.189	1.506	170
3	8.883	6.413	2.226	253
4	8.444	6.569	2.439	8
5	7.782	6.048	-0.358	72
6	8.780	6.305	2.463	120
7	7.990	5.881	0.532	302
8	8.540	6.388	0.369	52
9	7.975	6.030	1.216	28
10	7.072	5.431	0.488	40
11	8.037	5.965	0.615	72
12	8.681	6.483	0.734	379
13	7.115	5.259	1.403	82
14	9.587	7.130	0.322	62
15	8.193	6.104	0.915	305
16	7.613	5.847	1.666	293

Before we start converting Cartesian coordinates to spherical we need to find appropriate local coordinate system with origin in the center of the halo.

(a) Plot all the pairs of Cartesian coordinates on the separate graphs. Given the fact that this halo is shaped like a disk find the direction of its normal. (4p)

(b) Write down the unit vectors of local coordinate system. Choose appropriate orientation and take X' axis such that coincides the line of sight. (3p)

(c) Convert the given Cartesian coordinates to the local $X'Y'Z'$ coordinates. Fill in the values of Table 1 (a separate sheet is given in the answer sheets): X', Y', Z'. (8p)

(d) Convert the local Cartesian coordinates into spherical (take the central galaxy as the center of the coordinate system, ϕ is calculated relative to X' axis and θ - relative to Z' axis). Fill in the other values of Table 1 (a separate sheet is given in the answer sheets): r, θ, ϕ. (8p)

(e) Plot the histogram (parameter dependence on total number of data points that fall in a given limits) with 8 bins of Satellite galaxy distances from the central Galaxy. Take 0.2-2 MPC as the histogram range. The histogram has to be also converted into probability density function. Note: The probability density function is the probability at a given point. You can avoid plotting two separate graphs by just noting other limits of Y axis in case of probability function. (6p)

(f) Theoretically the histogram is given by the following equation

$$H(r) = \frac{1}{A + B \cdot r} \tag{1}$$

Linearize the given dependence and find values A and B from the graph (6p)

Part 2 (20p)

It is well-known that while measuring distance to an object with redshift method (large relative velocity due to the expansion of the universe) errors might occur because of the motion of the galaxies. The aim of this chapter is to correct these errors.
Assume that the central galaxy has a mass of $10^{10} M_{Sun}$, all satellite galaxies move in circular orbits and interactions between them is negligible.

(a) Use the data filled with you in Table 1. to plot the histogram of polar angle (take the width of histogram bins as 60 degrees). The angle distribution should be uniform, but it is not. Calculate the square deviation from the uniform value. (6p)

(b) The galaxies originally in 60-120 and 240-300 bins would need most correction since they are the ones which move along the line of sight. Assume that the angular coordinates coincide with the center of the corresponding bin (90 and 270 in our case) and write the expression as a function of distance from the center which describes change of polar angles caused by circular motion of the galaxies. Hint: You can assume that the distance from the centre doesn't change significantly because of this correction. (6p)

(c) Using the result of previous task correct the angles used in the last histogram. Plot the new histogram (probability dependence on distance) with the corrected data. Calculate the square deviation from the uniform value. (8p)

Part 3 (20p)

We have only calculated the central galaxy - satellite (from the same galaxy) histogram. It is time to move on to calculating other relations.
The same observations and data analysis has been carried out for different Halos and their central galaxies. The data for Figure 2. histogram only includes the distribution probability density function of the distances between central galaxies in different dark matter Halos.
Suppose we have 500 Halos, thus 500 central galaxies ($N_C = 500$). Take that the average number of satellite in one Halo is 1000 ($N_S = 1000$).

(a) Calculate the following quantities: (3 p)

1. N_{CC}, total number of central galaxy pairs
2. N^*_{CS}, total number of central galaxy - satellite pairs in the same Halo
3. N_{CS}, total number of central galaxy - satellite pairs in different Halos
4. N^*_{SS}, total number of satellite - satellite galaxy pairs in the same Halo
5. N_{SS}, total number of satellite - satellite galaxy pairs in different Halos

Below is given the table for probability density values for different distances between central galaxies in different dark matter Halos.

r (MPC)	ρ
0.3125	0.406
0.5375	0.808
0.7625	1.341
0.9875	1.134
1.2125	0.479
1.4375	0.143
1.6625	0.073
1.8875	0.060

We can calculate all probability densities using the given Center-Center data and Center-Satellite data which you calculated in part 1. The relations are following:

$$\rho^*_{SS}(r) = c_1 \rho^*_{CS}(r)^2 r \tag{2}$$

$$\rho_{CS}(r) = c_2 \rho^*_{CS}(r)^2 r \sqrt{\rho_{CC}(r) + 1} \tag{3}$$

$$\rho_{SS}(r) = c_3 \rho^*_{CS}(r)^2 r(\rho_{CC}(r) - 5)^2 \tag{4}$$

Where c_1, c_2 and c_3 are normalization constant and should be found for each density.

(b) Calculate and plot $\rho^*_{SS}(r)$, $\rho_{CS}(r)$ and $\rho_{SS}(r)$. (10p)

(c) Plot the final probability density of all galaxy pairs. Which one of 6 graphs on Figure 1 is most similar to you plot? (7p)

Solutions - Part 1

Part (a)

If we plot given coordinates in XY, XZ and YZ planes we'll see the following pictures. See Figures 2, 3 and 4 (0.5p for each graph). By observing Figure 2 we can deduce that the group is edge-on and its normal lies in XY plane. X' axis is along line of sight so it can be calculated by normalizing $8.3\hat{i} + 6.2\hat{j} + 1.1\hat{k}$. So

$$\hat{x'} = 0.79667866\hat{i} + 0.59510936\hat{j} + 0.10558392\hat{k} \tag{5}$$

(0.5p) Z' axis lays in XY plane and is perpendicular to X' so it can be calculated by normalizing $\hat{x'} \times \hat{k}$

$$\hat{z'} = 0.59845448\hat{i} - 0.80115681\hat{j} \tag{6}$$

(2p)

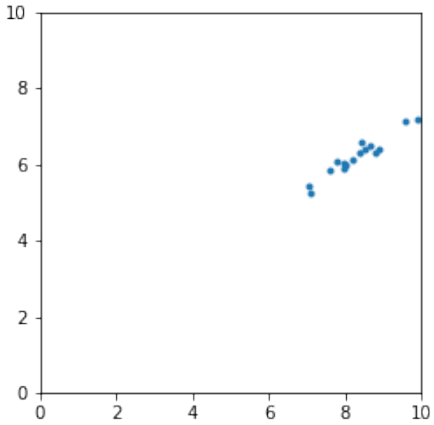


Figure 2: XY viewpoint of a cluster from observational site

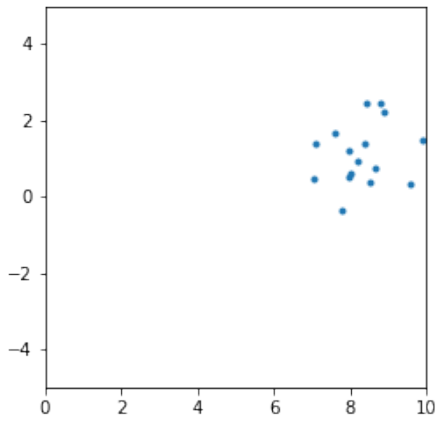


Figure 3: XZ viewpoint of a cluster from observational site

Part (b)

One can choose x' as line of sight (found in part(a)) (1p):

$$\hat{x'} = 0.79667866\hat{i} + 0.59510936\hat{j} + 0.10558392\hat{k} \tag{7}$$

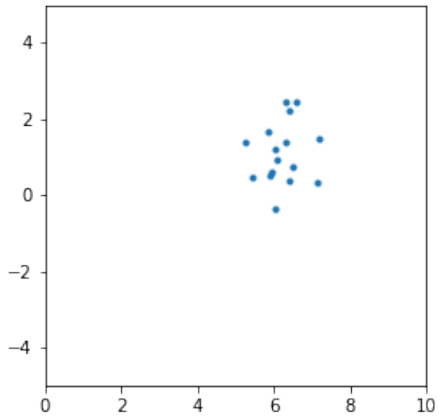


Figure 4: YZ viewpoint of a cluster from observational site

z' can be chosen as halo disk normal (found in part(a))(1p):

$$\hat{z'} = 0.59845448\hat{i} - 0.80115681\hat{j} \tag{8}$$

And finally for y' (1p):

$$\hat{y'} = \hat{x'} \times \hat{z'} = 0.08458928\hat{i} + 0.06318717\hat{j} - 0.9944104\hat{k} \tag{9}$$

Part (c)

Each coordinate in $X'Y'Z'$ system can be found by following equations

$$x' = \hat{x'} \cdot (\vec{r} - \vec{r_c}) \tag{10}$$

$$y' = \hat{y'} \cdot (\vec{r} - \vec{r_c}) \tag{11}$$

$$z' = \hat{z'} \cdot (\vec{r} - \vec{r_c}) \tag{12}$$

Where $\hat{x'}$, $\hat{y'}$ and $\hat{z'}$ are unit vectors of local $X'Y'Z'$ system, $\vec{r}$ is coordinate of a galaxy and $\vec{r_c}$ is coordinate of the center.

This values are calculated in the following table (0.5p for each correct x, y', z' tuple):

	$x - x_c$	$y - y_c$	$z - z_c$	x'	y'	z'
1	0.101124	0.109271	0.294476	0.176684	-0.277372	-0.027025
2	1.582909	0.989392	0.405895	1.892722	-0.207212	0.154641
3	0.582735	0.213285	1.126426	0.710112	-1.057359	0.177866
4	0.144328	0.369381	1.338567	0.476137	-1.295537	-0.209558
5	-0.517770	-0.151687	-1.458202	-0.656729	1.396668	-0.188336
6	0.480059	0.105110	1.362674	0.588881	-1.307808	0.203084
7	-0.309855	-0.319116	-0.568033	-0.496739	0.518483	0.070227
8	0.239541	0.188218	-0.730736	0.225694	0.758807	-0.007438
9	-0.324800	-0.169551	0.116378	-0.347375	-0.153915	-0.058541
10	-1.227659	-0.768684	-0.612379	-1.500158	0.456538	-0.118862
11	-0.263209	-0.234869	-0.485227	-0.400698	0.445410	0.030649
12	0.381200	0.282515	-0.365581	0.433222	0.413634	0.001792
13	-1.185021	-0.941303	0.302593	-1.472310	-0.460620	0.044950
14	1.286640	0.930147	-0.777889	1.496445	0.941150	0.024802
15	-0.106760	-0.095804	-0.184997	-0.161600	0.168879	0.012863
16	-0.686865	-0.352750	0.566244	-0.697350	-0.643470	-0.128449

Part (d)

Spherical coordinates can by calculated by following relations:

$$r = \sqrt{x'^2 + y'^2 + z'2} \tag{13}$$

$$\phi = \arctan \frac{y'}{x'} \tag{14}$$

and signs of x' and y'

$$\theta = \arccos \frac{z}{r} \tag{15}$$

Calculated values are given in the following table (0.5p for each correct r, ϕ, θ tuple):

	x'	y'	z'	r	ϕ	θ
1	0.176684	-0.277372	-0.027025	0.329974	5.279566	1.652789
2	1.892722	-0.207212	0.154641	1.910301	6.174141	1.489757
3	0.710112	-1.057359	0.177866	1.286042	5.303793	1.432047
4	0.476137	-1.295537	-0.209558	1.396079	5.064586	1.721471
5	-0.656729	1.396668	-0.188336	1.554814	2.010330	1.692226
6	0.588881	-1.307808	0.203084	1.448581	5.135477	1.430138
7	-0.496739	0.518483	0.070227	0.721461	2.334780	1.473302
8	0.225694	0.758807	-0.007438	0.791695	1.281697	1.580191
9	-0.347375	-0.153915	-0.058541	0.384430	3.558678	1.723672
10	-1.500158	0.456538	-0.118862	1.572587	2.846171	1.646452
11	-0.400698	0.445410	0.030649	0.599906	2.303400	1.519685
12	0.433222	0.413634	0.001792	0.598981	0.762273	1.567805
13	-1.472310	-0.460620	0.044950	1.543336	3.444801	1.541667
14	1.496445	0.941150	0.024802	1.767973	0.561416	1.556768
15	-0.161600	0.168879	0.012863	0.234094	2.334171	1.515821
16	-0.697350	-0.643470	-0.128449	0.957522	3.886829	1.705349

Part (e)

Using radii calculated above and number of galaxies for each data (rewritten below) one can calculate number of galaxies in each bin. Range of bin 1 is (0.2, 0.425), bin 2 - (0.425, 0.65) and so on bin 8 - (1.775, 2.0).

	r	n
1	0.329974	291
2	1.910301	170
3	1.286042	253
4	1.396079	8
5	1.554814	72
6	1.448581	120
7	0.721461	302
8	0.791695	52
9	0.384430	28
10	1.572587	40
11	0.599906	72
12	0.598981	379
13	1.543336	82
14	1.767973	62
15	0.234094	305
16	0.957522	293

Correct values for each bin are: (0.5p for each correct n)

	Center of the bin	Number of the galaxies
1	0.312500	624
2	0.537500	451
3	0.762500	354
4	0.987500	293
5	1.212500	253
6	1.437500	210
7	1.662500	174
8	1.887500	170

Corresponding histogram is given on Figure 5. (1p)

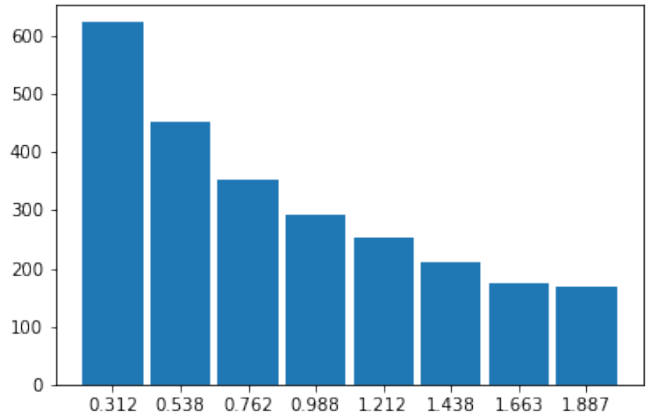


Figure 5: Histogram: Number of galaxies dependence on distance (MPC), with a bin size equal to 0.225 MPC

Histogram data as a table:

In order to get probability density function each value of histogram must be divided by it integral on the whole region. The integral can be approximated by finding total area of the rectangles:

$$I = \sum_{n=1}^{8} H(r_i)d = 569.025 \tag{16}$$

Where $d = 0.225$ is size of a bin. After dividing each data by I we get (0.25p for each correct probability density):

	Center of the bin	Number of the galaxies	Probability density
1	0.312500	624	1.096613
2	0.537500	451	0.792584
3	0.762500	354	0.622117
4	0.987500	293	0.514916
5	1.212500	253	0.444620
6	1.437500	210	0.369052
7	1.662500	174	0.305786
8	1.887500	170	0.298757

Plotting this data gives the graph on Figure 6 (1p)

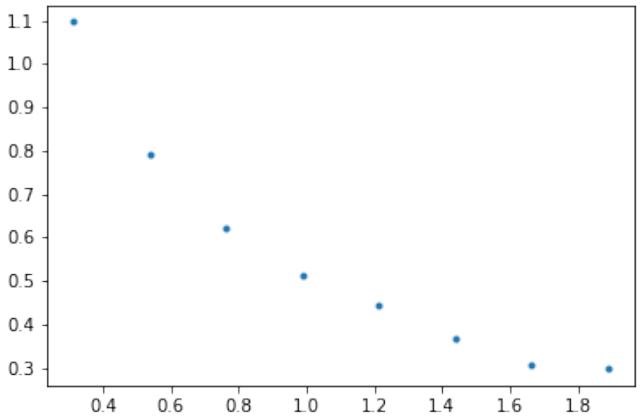


Figure 6: Probability density of center-satellite distances in the same halo

Part (f)

Since histogram equation has following form:

$$H\left(r\right) = \frac{1}{A + B \cdot r} \tag{17}$$

We should calculate $1/H(r)$ which gives the following values (0.125 for each correct 1/H) :

	r	H(r)	1/H(r)
1	0.312500	624	0.001603
2	0.537500	451	0.002217
3	0.762500	354	0.002825
4	0.987500	293	0.003413
5	1.212500	253	0.003953
6	1.437500	210	0.004762
7	1.662500	174	0.005747
8	1.887500	170	0.005882

Plotting this data gives Figure 7

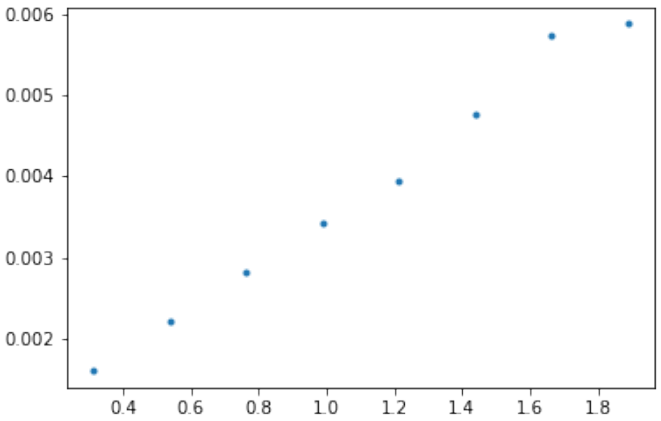


Figure 7: Linearization for regression

Finding the slope and the intercept of this line (by graph or any other method) gives

Following values:

$$A = 0.00065977; B = 0.00285494 \tag{18}$$

2.5p for B in range 0.002-0.003 2.5p for A in range 0.0006-0.0008

Solutions - Part 2

Part (a)

We can ϕ values from Part 1.c to distribute them in 60 degree bins. In bin 1 we have number of galaxies which fall in (0, 60) degree range, in bin 2 - (60, 120) and so on bin 6 - (300, 360).

Part 1.c data in degrees:

	Angle	Number of galaxies
1	302.496846	291
2	353.752245	170
3	303.884926	253
4	290.179429	8
5	115.183442	72
6	294.241133	120
7	133.773013	302
8	73.435830	52
9	203.897219	28
10	163.073590	40
11	131.975076	72
12	43.675004	379
13	197.372555	82
14	32.166754	62
15	133.738158	305
16	222.698869	293

We get the following data for the histogram (0.5p for each correct n):

	Center of the bin	Number of galaxies
1	30	441
2	90	124
3	150	719
4	210	403
5	270	128
6	330	714

Uniform value is total number of galaxies divides by 6 which is 421.5. Square deviation is 241° (1p)

Graph is given on Figure 8 (2p)

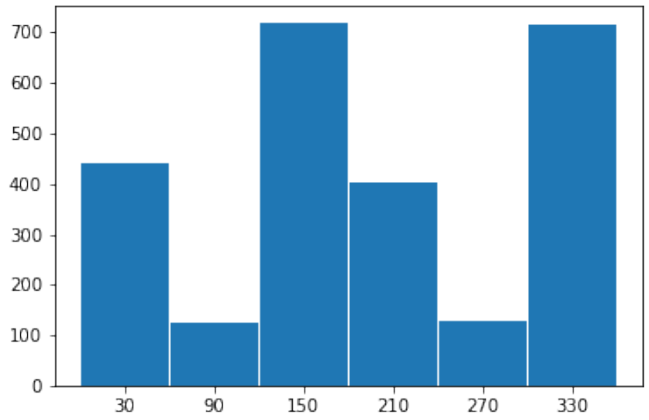


Figure 8: Histogram: Number of galaxies dependence on ϕ angle, with a bin size equal to 60°

Part (b)

A galaxy with the orbital radius r around central mass M has velocity (2p):

$$v = \sqrt{\frac{GM}{r}} \tag{19}$$

This will cause to give the following error in distance along line of sight (2p):

$$d = \frac{v}{H} = \frac{1}{H}\sqrt{\frac{GM}{r}} \tag{20}$$

Angle difference for galaxies at 90 and 270 degrees will be (2p):

$$\alpha = \arctan\frac{d}{r} = \arctan\frac{1}{H}\sqrt{\frac{GM}{r^3}} \tag{21}$$

For detailed drawing see Figure 9

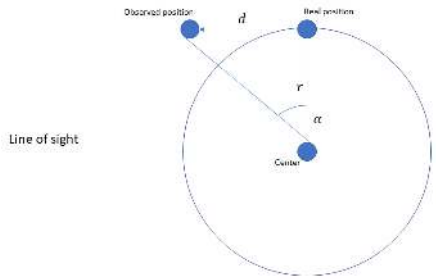


Figure 9: Drawing

Part (c)

Galaxies originally in bins of range (60, 120) and (240, 300) might move to (120, 180) and (300, 360) respectively. So we should subtract correction value found in previous task to the galaxies in bins 2 and 5 to see if they move to bins 1 and 4 respectively. (0.25p for each correct corrected angle(

	Original angle	Corrected angle	Number of galaxies
1	302.494665	276.136451	291
2	353.752069	351.714594	170
3	303.884728	300.199512	253
4	290.178642	290.178642	8
5	115.184577	115.184577	72
6	294.241835	294.241835	120
7	133.776280	125.061576	302
8	73.437747	73.437747	52
9	203.899065	203.899065	28
10	163.073490	160.346665	40
11	131.977481	120.548626	72
12	43.676971	43.676971	379
13	197.372956	197.372956	82
14	32.167948	32.167948	62
15	133.743596	94.077612	305
16	222.694636	222.694636	293

New corrected distribution looks like this (0.3p for each correct n):

	Center of the bin	Number of galaxies	
1	30	441	
2	90	429	
3	150	414	In this case square deviation equals to 12°
4	210	403	
5	270	419	
6	330	423	

(1p) And relevant graph is given on Figure 10 (1.2p)

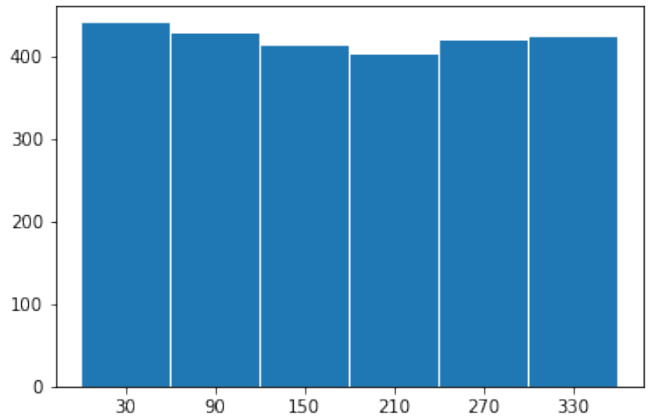


Figure 10: Histogram: Corrected Number of galaxies dependence on ϕ angle, with a bin size equal to 60°

Solutions - Part 3

Part (a)

With simple combinatorics one can get (0.6p for each correct N):

$$N_{CC} = N_C(N_C - 1) = 249500 \tag{22}$$

$$N^*_{CS} = N_C N_S = 500000 \tag{23}$$

$$N_{CS} = N_C(N_C - 1)N_S = 249500000 \tag{24}$$

$$N^*_{SS} = N_C N_S(N_S - 1) = 499500000 \tag{25}$$

$$N_{SS} = N_C N_S(N_C - 1)N_S = 249500000000 \tag{26}$$

Part (b)

We can calculate all probability densities using center-center and center-satellite densities. Non-normalized values (with $c_1 = c_2 = c_3 = 1$) are calculated in the following table (0.5p for each correct tuple):

	r	ρ^*_{CS}	ρ_{CC}	ρ^*_{SS}	ρ_{SS}	ρ_{CS}
1	0.312500	1.096613	0.406141	0.375800	17.365137	0.975747
2	0.537500	0.792584	0.807863	0.337652	12.992898	0.994072
3	0.762500	0.622117	1.340510	0.295110	8.653487	0.988565
4	0.987500	0.514916	1.134011	0.261824	8.568351	0.837480
5	1.212500	0.444620	0.478630	0.239696	10.729180	0.638199
6	1.437500	0.369052	0.143154	0.195787	10.112507	0.458355
7	1.662500	0.305786	0.073165	0.155452	8.262279	0.352612
8	1.887500	0.298757	0.060971	0.168470	8.998531	0.379962

For normalization (same way as in Part 1.d) we get $c_1 = 2.18592; c_2 = 0.79304; c_3 = 0.0516$. Multiplying relevant densities by this values gives (0.5p for each correct normalized tuple):

	r	ρ^*_{SS}	ρ_{SS}	ρ_{CS}
1	0.312500	0.822854	0.900750	0.770962
2	0.537500	0.739325	0.673957	0.785441
3	0.762500	0.646175	0.448866	0.781089
4	0.987500	0.573292	0.444450	0.661713
5	1.212500	0.524839	0.556535	0.504257
6	1.437500	0.428697	0.524548	0.362158
7	1.662500	0.340380	0.428574	0.278607
8	1.887500	0.368883	0.466764	0.300217

Plots are given on Figures 11 - 14 (0.5p for each plot).

Part (c)

Total probability density is weighted sum of all densities (4p):

$$\rho \sim \frac{N_{CC}\rho_{CC} + N^*_{CS}\rho^*_{CS} + N_{CS}\rho_{CS} + N^*_{SS}\rho^*_{SS} + N_{SS}\rho_{SS}}{N_{CC} + N^*_{CS} + N_{CS} + N^*_{SS} + N_{SS}} \tag{27}$$

Calculated values and normalized densities are given in the following table:

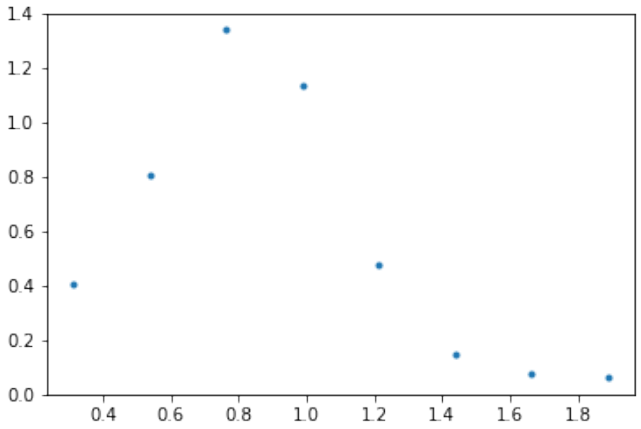


Figure 11: Probability density of center-center distances in for different halos

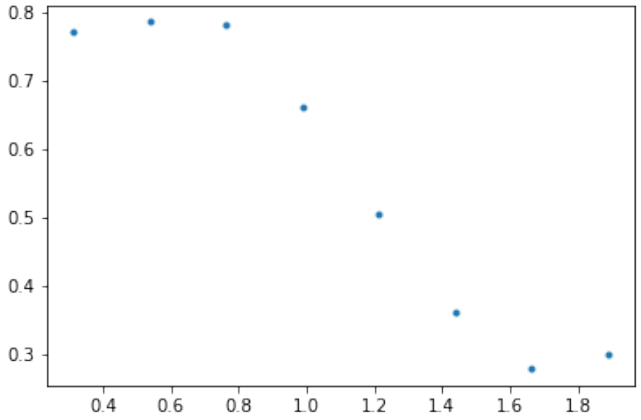


Figure 12: Probability density of center-satellite distances in the different halos

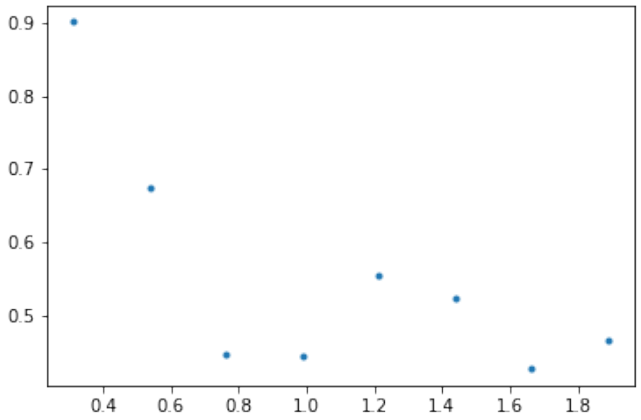


Figure 13: Probability density of satellite-satellite distances in the different halos

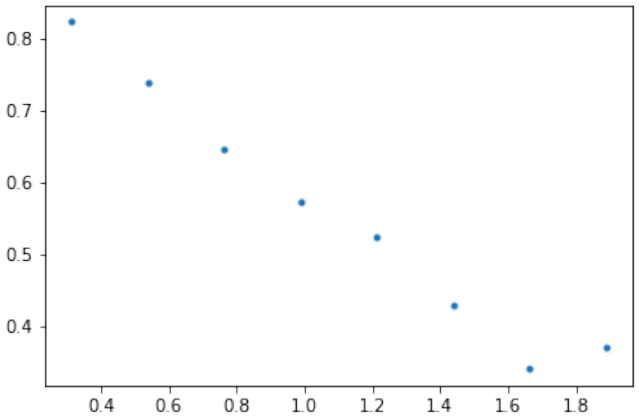


Figure 14: Probability density of satellite-satellite distances in the same halo

	r	Calculated ρ	Normalized ρ
1	0.312500	0.004212	0.843662
2	0.537500	0.003606	0.722433
3	0.762500	0.002967	0.594357
4	0.987500	0.002693	0.539524
5	1.212500	0.002663	0.533361
6	1.437500	0.002265	0.453652
7	1.662500	0.001813	0.363239
8	1.887500	0.001968	0.394216

And final plot looks like this (Figure 15) (2p)

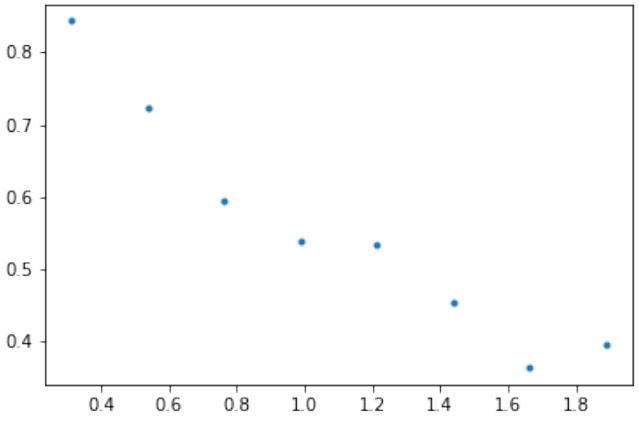


Figure 15: Final probability density

Most similar to this density is plot E on Figure 1 (1p).