Physci 120, Winter 2009

Lab 3

Galaxy environments, groups and clusters of galaxies, statistical clustering

Introduction

Although the distribution of galaxies on large scales is quite uniform, on smaller scales, comparable to the mean separation between bright galaxies (about one million parsecs), galaxy distribution is far from being homogeneous. Galaxies are arranged in a variety of large *structures* from *groups* of a few galaxies (the nearest example is *the Local Group* formed by our Milky Way, Andromeda Galaxy (M31) and a few dozen faint satellite galaxies, including the Magellan Clouds) and *clusters of galaxies* containing dozens and hundreds of bright galaxies, to galaxy superclusters and filaments which span tens of millions of parsecs in size.

Although these structures can be clearly seen in the diagrams showing three-dimensional distribution of galaxies, in most cases they are not as obvious in the distribution of galaxies on the sky. This is because on the sky we see galaxies at different distances projected onto the same area, which tends to "wash out" the structures. Nevertheless, the basic tendency of galaxies to cluster can be estimated by counting numbers of galaxies in different patches of the sky and comparing the counts to the number expected for a uniform distribution. This is a simplified version of the clustering analysis done using correlation function, which has been a very powerful tool in studying clustering of galaxies and is still actively used in astronomy.

In the first part, in Lab 3, you will perform a simple counts exercise to see the tendency of galaxies to cluster for yourself. You will then explore environments of typical galaxies in random fields (called *field galaxies*) and galaxies in groups and clusters. The key lesson of this exercise will be that galaxies in clusters are systematically different than galaxies in random fields: namely, clusters contain unusual concentrations of bright galaxies and most galaxies in a cluster have similar colors and are redder than typical galaxies in the field. In the following Lab (lab 4) you will see that this fact can be used to effectively search for clusters on the sky using galaxy colors and you will carry out such search based on what you learn in the lab 3.

In the final part of the lab 4 you will estimate total gravitational mass of a cluster using the virial theorem and compare it to the total mass associated with its galaxies. You will see that the two masses are vastly different, thereby recovering one of the oldest and strongest observational evidence for the existence of *dark matter* in the universe, which was originally obtained by US astrophysicists Fritz Zwicky and Sinclair Smith in the mid-1930s using very similar method to what you will use.

Lab 3. Galaxy clustering

As noted above, galaxies tend to *cluster* on the scales of the order of the mean separation between galaxies. Astronomers quantify this tendency *statistically* using so-called correlation function. The full exercise would be too complicated for this lab. We will do something simpler but nevertheless illustrative and will quantify clustering tendency of galaxies by comparing the numbers of galaxies in a given magnitude range in the actual SDSS fields to the number expected if the distribution was random.

First, you will need to download the Google Sky script named `SDSS Galaxy Query by R-Magnitude Bins` from the same site as before and load it into the Google Sky (turn off all other scripts you have used before in the previous labs). This script identifies all galaxies within a given range of apparent r-filter magnitudes within your field of you of the SDSS survey. Uncheck all magnitude ranges except for 17-18 (you should see only yellow circles then). Now zoom-in onto an SDSS field to the point, when the angular scale is close to $0^\circ\,20'$. Try to resize the left menu area such that the sky viewing area is as close as possible to a square shape. Wait until the script queries the SDSS database. After it is done, you will see galaxies in the SDSS catalog in the magnitude range $m_r = [17 - 18]$ in your field marked with yellow circles. Count them and record the number.

Now without changing the zoom level (so that the angular extent of the viewing area stays the same — this is important!) shift to some other area fairly distant from the first one. Try to choose an area devoid of very bright stars which can block some areas and result in missed galaxies. Count galaxies marked by yellow circles in the new field again and record the count. Do this for a total of 20 fields (make sure that you do not reach the boundary of the SDSS survey when you do this — there should be no sign of the DSS Consortium imagery near the bottom of the viewing field, only the SDSS survey; the quality of the DSS images is much worse than the SDSS so it should be easy to notice). Record the number of galaxies within each field that you examine. This should be fairly quick, as there will be a dozen of objects marked in each field with some fluctuations. It is these fluctuations that we are after.

Once the counts are ready, compute the mean number of galaxies per field, by summing all counts and dividing the sum by the number of fields. This will be the average count number $\bar{N}$. Sort the counts by the number of galaxies (i.e. first the smallest number, then the next smallest, and so on until the field with the largest number of galaxies). How do we know whether a given number of galaxies in a field is consistent with a uniform distribution or not? We can figure this out as follows. If the distribution was purely uniform, the distribution of counts among fields would be consistent with the Poisson distribution, named after famous french mathematician and physicist Siméon Poisson who discovered it in 1838 when he was studying statistics of verdicts in criminal court cases (trying to determine how random they were). We will use an applet available at: `http://stattrek.com/Tables/Poisson.aspx` to calculate how likely it is that a given count would be observed according to the Poisson distribution with the mean counts $\bar{N}$.

In the site enter the average number $\bar{N}$ in the field named "Average rate of success," while in the field "Poisson random variable" enter the smallest count among your fields. Press the button "Calculate" and examine the number in the field "Cumulative Poisson probability." For counts that are smaller than the mean count $\bar{N}$, mark the count if the

cumulative probability number is smaller than 0.05. Once you find a number for which this probability is larger than 0.05, go to the largest count and compute the probability for it. In this case, for counts larger than the mean, mark count if the cumulative probability is larger than 0.95. Then do this for the next largest count, and so on, until you find the count for which the cumulative probability is ≤ 0.95. How many counts with cumulative probability ≤ 0.05 or ≥ 0.95 did you find? If considerably more than two (assuming you obtained counts in twenty fields), then this is a strong evidence that distribution of galaxies is not uniform, because for the Poisson distribution expected in this case only 10% of fields (i.e. two) should have such values.

Lab tasks.

- *Present the list of counts, the average count $\bar{N}$, and mark the counts for which the cumulative probability is ≤ 0.05 or ≥ 0.95.*
- *State your conclusions about uniformity of galaxy distribution in these fields based on the number of small probability counts.*
- *The typical redshift of galaxies of magnitudes 17-18 is ~ 0.1. Calculate the distance to a typical galaxy at such redshift using the Hubble expansion law $d = cz/H_0$ where $c = 300000$ km/s is the speed of light in vacuum and H_0 is the Hubble constant, which you can assume to be equal to 70 km/s/Mpc. Using the small angle approximation, calculate the physical scale corresponding to this distance, as $r = \theta d$, where θ is the angular size of the field in radians. For the 20 arcminute field that you used it is $\theta = 20\pi/(60 * 180) \approx 0.0058$ radians. Compute and present the scale r on which you quantified the clustering in Megaparsecs.*

Lab 3. Galaxy luminosities and colors in different environments

We will first explore galaxies in the typical random field (called "field galaxies"). Make sure the `SDSS Galaxy Query plugin` used in the previous part is still activated and it identifies galaxies with circles, and activate the `Color-Magnitude plugin`, which plots the r-band apparent magnitude (x-axis) and colors (difference in apparent magnitude in g and r filters, in the y-axis) of galaxies in the field in the graph in the top left corner of the field of view. This plot is called *the color-magnitude diagram* (CMD). You can use one of the random fields used before, or choose your own. Zoom on the field to the scale of about 15 arcminutes or so (number in the lower right corner).

Lab tasks.

- *Describe how galaxies in the random fields are distributed in the CMD.*
- *Load* `Groups` *location file and examine CMD around galaxy groups (small groupings of galaxies. Describe any difference you notice from the CMD distribution of galaxies in the random fields.*

- *Load* `Clusters` *location file and describe the distribution of galaxies in the CMDs and the differences from random and group locations.*
- *Zoom-in onto the region around the brightest cluster galaxies (say central 5 arcminutes) and examine the CMD of galaxies in this region. Now zoom-out, center on a region in the outskirts of the cluster (say 20-30 arcminutes away from the BCG), zoom-in to 5 arcminute level again and examine CMD. How does the distribution of galaxies changes when you move to the outskirts of clusters?*
- *Examine clusters at different redshifts (you can get the redshift by clicking on the brightest galaxy near the center of each cluster). How does the distribution of galaxies change in the CMD for clusters at higher redshifts?*
- *Discuss and interpret your results in terms of what you learned about properties of galaxies in nearby clusters and what you know about spectra of red elliptical galaxies.*

Lab 4. Hunting for clusters

In the previous section you should have learned that cluster fields contain unusual concentration of bright galaxies which look more yellow ("redder") than most field galaxies. Galaxies in clusters form a horizonthal "ridge" or "sequence" in the color-magnitude diagram. In this portion of the lab you will use this fact to carry out a "blind" search for galaxy clusters in the sky. This method of searching for clusters, done in an automated and systematic way using a computer analysis of images, was pioneered just ten years ago by the University of Chicago Prof. Michael Gladders (among others) and is the most efficient way of finding new clusters using the optical images and multi-band photometry (measurements of apparent magnitudes of galaxies through different filters).

To perform the search, go to a random field on the sky (you can go to one of the random fields used above or choose a random area on the sky yourself, just make sure you are well within the SDSS coverage). Zoom-in (or zoom-out depending on the inital zoom of your view) to the level at which the marker in the lower right hand corner of the Sky view window shows approximately 15 arcminutes (i.e., $0°15'$...).

Once you choose the zoom level, keep it the same throughout the exercise (you can play with it once you find a cluster). Now turn off (uncheck) the `imagery layer` in the bottom list on the panel on the left hand side of the Sky window. This should get rid of the SDSS image of the sky, leaving a gray uniform background with the CMD diagram remaining. Now scan the sky making $\approx 15'$ steps. You can do this, for example, by clicking on the left most edge of the field of view and dragging it all the way to the right (or, similarly, from the bottom to top). The order or direction of your steps is not very important, but it may make sense to do them in one direction once you chose it. Please count the number of steps you take (thereby roughly estimating the area you scan to find a cluster).

If you stumble onto an area which may credibly look like it contains a cluster, you can try to make smaller shifts about the center of the field of view to see if it makes cluster signature in the CMD more pronounced. If you believe you may have found a cluster, turn the `imagery` layer back on and scan the field of view visually. If it looks like there is a cluster

in particular area, zoom on it and examine to make sure (hint: look for large, bright elliptical galaxies in the field of view similar to the brightest cluster galaxies you've examined in the Hubble lab; these are candidates for cluster centers which you can examine by zooming in on their locations). If you are not sure there is a cluster after visual inspection, consult with the TA, or simply continue searching for a more obvious case. Once you find the cluster, you are ready for

Lab tasks.

- *Estimate the rough area you surveyed to find a cluster. How many such clusters per square degree of the sky would you then expect? How rare are clusters based on your measurements (how many would there by in the entire sky; the entire sky has the area of approximately* 41253 *square degrees)?*
- *Turn on the* `SDSS Galaxy Query` *layer (uncheck the* `Photometric Layer`*, make sure the* `Spectroscopic layer` *is turned on). Find galaxies with SDSS spectra (identified with circles with the SDSS Spectroscopic layer), click on each bright galaxy with spectrum in the cluster area and record their redshift given in the pop-up bubble. Find the approximate redshift of the cluster.*
- *Do all of the galaxies with spectra have similar redshifts. If not all, why not?*
- *Turn of the CMD layer, zoom-out to view larger than* $15'$ *accross. Examine redshifts of galaxies with SDSS at larger distances from the brightest cluster galaxies. Try to find the most distant galaxy from the BCG which still has redshift similar to that of the BCG. Measure the angular distance between BCG and the most distant galaxy using the ruler icon. Quote the number — this will be your rough estimate of the cluster extent.*
- *Identify several (the more, the better but not more than 10) galaxies within extent of the cluster with redshifts close to that of the BCG. Record their redshifts and calculate the mean redshift of the cluster using their individual redshifts.*
- *Using the angular extent measured above and the estimated redshift of the cluster, compute the actual physical extent of the cluster in Mpc corresponding to the angular extent assuming the Hubble constant value of* $H_0 = 70$ *km/s/Mpc.*

Lab 4. Measuring cluster mass using Virial Theorem

If, in the previous section of the lab, you found a nearby cluster ($z \sim 0.0.05 - 0.15$) with at least five member galaxies that had spectra and similar redshifts to that of the BCG, you can use the cluster you found to measure its mass using the virial theorem in this section. If the cluster you found is at higher redshift and/or does not have enough galaxies with measured spectra, you can use one of the systems in the `Clusters` list used above to examine cluster environments. In the latter case, you have to redo the exercise of estimating cluster extent on the sky and converting it into physical extent in megaparsecs.

Once the cluster is chosen, center the field of view on the BCG and adjust zoom so that the field of view encompasses the entire extent of the cluster. Identify cluster members (i.e., galaxies with SDSS spectra and measured redshifts similar to that of the BCG) among galaxies identified with a circle with the SDSS Galaxy Spectroscopic plugin. Record the redshifts of the member galaxies and calculate the mean redshift of the cluster.

To estimate cluster mass, you will use the so-called *virial theorem*. This theorem relates the kinetic, K, and potential, W, energies of the system and applies to any self-gravitating system in equilibrium (e.g. it also applies to the Sun, Earth, stellar clusters, etc.):

$$W = -2K. \tag{1}$$

The kinetic energy of a cluster is

$$K \approx \frac{1}{2} M_{\rm tot} \langle v_{\rm gal}^2 \rangle, \tag{2}$$

where $M_{\rm tot}$ is the *total* gravitating mass of a cluster (which we aim to measure) and $\langle v_{\rm gal}^2 \rangle$ is characteristic velocity variance of cluster material. We will estimate the latter by using variance of velocities of the cluster member galaxies about the mean cluster velocity, which can be calculated as

$$\langle v_{\rm gal}^2 \rangle \equiv 3\,\frac{\sum_{i=1}^{N_{\rm gal}} (cz_i - c\bar{z})^2}{N-1} = 3c^2 \times \frac{\sum_{i=1}^{N_{\rm gal}} (z_i)^2 - N_{\rm gal}\bar{z}^2}{N_{\rm gal}-1}. \tag{3}$$

The factor of 3 in front is to account for the fact that redshifts measure only velocity along the line of sight, while virial theorem applies to velocities along all three dimensions of space; $\bar{z}$ is the mean redshift of the cluster you measured using the $N_{\rm gal}$ galaxies with spectroscopic redshifts z_i, which you have recorded in the previous step.

The potential energy of the self-gravitating system of mass $M_{\rm tot}$ is negative by definition and is given by

$$W = -\alpha \frac{G M_{\rm tot}^2}{R}, \tag{4}$$

where α is a numeric factor of order unity which depends on how mass is distributed within the system, R is the physical extent of the system, and $G = 6.6726 \times 10^{-8}$ cm^3 s^{-2} g^{-1} (in cgs units) is the universal gravitational constant. For the mass distribution expected in clusters $\alpha \approx 0.5$ (good assumption for the accuracy of your measurement which will be to a factor of two or so), and we will use this number to compute the mass.

Combining equations 1, 2, and 4 we get:

$$M_{\rm tot} = \frac{R \langle v_{\rm gal}^2 \rangle}{\alpha G}. \tag{5}$$

Lab tasks:

- *Using your tabulated redshifts z_i of galaxy members calculate the mean redshift $\bar{z}$ and $\langle v_{\rm gal}^2 \rangle$ in km^2/s^2. The square root of this number will give you velocity dispersion, which is the typical velocity of a galaxy in this cluster with respect to the BCG in km/s — calculate and present this number as well (is this a high velocity? Compare it to ≈ 30 km/s with which the Earth moves around the Sun).*

- *Use equation 5 to estimate total mass of the cluster within the extent you measured. R is half the cluster diameter that you measured and take $\alpha = 0.5$. Be careful with units and give the result in solar masses $M_\odot$ (where $M_\odot = 1.989 \times 10^{33}$ g).*

- *Count the number of red bright galaxies around the cluster within R (with or without spectra) to get a rough number of cluster galaxies. Each of such bright galaxies has total mass in stars of about $10^{11}\,\mathrm{M}_\odot$ (give or take a factor of 2-3). This number is determined from luminosities of these galaxies and average luminosity and mass of old stars which these galaxies consist of. Estimate roughly the total mass in stars within R using the number of galaxies you counted and the average stellar mass of $10^{11}\,\mathrm{M}_\odot$. How does this mass compare to the total mass $M_{\rm tot}$ you calculated using the virial theorem? How big is the difference?*

The difference between mass in stars and total mass is one of the pieces of observational evidence we have for existence of large amounts of dark gravitating matter in the universe on large scales. Actually, in clusters there are large amounts of hot plasma mostly consisting of hydrogen and helium nuclei and electrons, but the mass of this plasma measured from its X-ray emission still falls far short of explaining the measurements of total cluster masses.