Very large-scale correlations in the galaxy distribution

F. Sylos Labini(a)

Museo Storico della Fisica e Centro Studi e Ricerche Enrico Fermi
Piazzale del Viminale 1, 00184 Rome, Italy, EU
Istituto dei Sistemi Complessi CNR - Via dei Taurini 19, 00185 Rome, Italy, EU

received 6 September 2011; accepted in final form 6 October 2011
published online 16 November 2011

PACS 98.80.-k – Cosmology
PACS 05.40.-a – Fluctuations phenomena, random processes, noise, and Brownian motion
PACS 02.50.-r – Probability theory, stochastic processes, and statistics

Abstract – We characterize galaxy correlations in the Sloan Digital Sky Survey by measuring several moments of galaxy counts in spheres. We firstly find that the average counts grows as a power-law function of the distance with an exponent \(D = 2.1 \pm 0.05\) for \(r \in [0.5, 20]\) Mpc/h and \(D = 2.8 \pm 0.05\) for \(r \in [30, 150]\) Mpc/h. In order to estimate the systematic errors in these measurements we consider the counts variance finding that it shows systematic finite-size effects which depend on the samples sizes. We clarify, by making specific tests, that these are due to the galaxy long-range correlations extending up to the largest scales of the sample. The analysis of mock galaxy catalogs, generated from cosmological N-body simulations of the standard LCDM model, shows that for \(r < 20\) Mpc/h the counts exponent is \(D \approx 2.0\), weakly dependent on the galaxy luminosity, while \(D = 3\) at larger scales. In addition, contrary to the case of the observed galaxy samples, no systematic finite-size effects in the counts variance are found at large scales, a result that agrees with the absence of large-scale (\(r \approx 100\) Mpc/h) correlations in the mock catalogs. We thus conclude that the observed galaxy distribution is characterized by correlations, fluctuations and hence structures, which are larger, both in amplitude and in spatial extension, than those predicted by the standard model LCDM of galaxy formation.

Copyright © EPLA, 2011

Introduction. – The quantification of large-scales galaxy correlations and fluctuations is a central problem in cosmology. The advent of massive redshift surveys has made possible precise measurements of the galaxy two-point correlation function on ten Mpc scales, where power-law correlations (in the redshift space) have been well established [1–3]. On larger scales the situation is less clear as large density fluctuations on 150 Mpc/h scales were observed [4–11]: these are not obviously compatible with the absence of strong clustering on those scales as predicted by standard models of galaxy formation [12,13]. In addition, in a deep sample of very bright galaxies [14–16], on 200 Mpc/h scales, an unexpected strong signal with respect to the predictions of the standard model was observed; recently an excess of clustering was also found by [17,18].

In order to characterize the large-scale galaxy correlations, the average density \(\overline{n(r)}\) was measured in the Sloan Digital Sky Survey (SDSS) [19,20]: this was shown to present a scaling behavior as \(\overline{n(r)} \sim r^{-1}\) for \(r < 20\) Mpc/h [3,21]. However, while some authors [21] noticed that around 70 Mpc/h there occurs a transition toward uniformity (i.e. \(\overline{n(r)} \sim \text{const}\)), others concluded that \(\overline{n(r)} \sim r^{-0.2}\) for 20 Mpc/h < \(r < 80\) Mpc/h [22]. Furthermore this latter behavior was shown to be associated to a scaling of the variance slower than for a uniform distribution and to a probability density function (PDF) of galaxy counts in spheres well fitted by a Gumbel distribution (i.e., different from a simple Gaussian PDF) [22]. It was then found that the PDF, in different spatial volumes, presents systematic differences at large enough scales which, by making specific tests, were interpreted as due to large-scale structures [23,24]. However, the considered tests used non-overlapping subsamples covering different redshift ranges, so that a physical or observational scale-dependent systematic effect can, at least partially, affect the observed behaviors.

For instance, it was noticed that the number density of bright galaxies increases by a factor \(\approx 3\) as the redshift increases from \(z = 0\) to \(z = 0.3\) [25], and to explain these observations a significant evolution in the luminosity and/or number density of galaxies at redshifts \(z < 0.3\)
was proposed. In this context evolution can be parameterized as a redshift-dependent correction [26]. However, by performing several tests to determine the possible effect of evolution on the SDSS data, it was concluded [24] that, while the evolution may change the galaxy redshift counts behavior as a function of redshift, it is unlikely that it can produce large amplitude fluctuations of large spatial extension.

In this letter, by considering SDSS equal volume samples covering the same distance scales, we are able to disentangle finite-size effects due to large-scale correlations from redshift-dependent systematic effects. This allows us to confirm the results by [22,24], to extend them to larger scales, i.e. \( r \approx 150 \text{ Mpc}/h \), and to clarify several other properties of the large-scale galaxy distribution.

**The data.** – We have constructed several subsamples of the main-galaxy (MG) sample of the spectroscopic catalog SDSS-DR7 [27,28]. The selection criteria used to construct the volume-limited (VL) samples are (see [24] for more details): we selected galaxies from the MG sample with redshift \( z \in [10^{-3}, 0.3] \) with redshift confidence \( \varepsilon_{\text{conf}} \geq 0.35 \), without significant redshift determination errors and with apparent magnitude \( m_r < 17.77 \). In order to avoid the irregular edges of the survey boundaries, and to consider a simple sky area, we have constructed a sample \( (R_1) \) limited, in the internal angular coordinates of the survey [24], by \( \eta \in [-33.5程度, 36.0程度] \) and \( \lambda \in [-48°, 51.5°] \). We also consider in what follows that two half samples of equal volume are limited by \( (R_2) \) \( \lambda \in [-48°, 0] \) and \( (R_3) \lambda \in [0°, 51.5°] \).

The survey is known to have a small angular incompleteness; indeed, in average, about the 5% of the target galaxies do not have a measured redshift [28]. In general, there is not a procedure free of a priori assumptions, to correct for such an incompleteness. Indeed, given that a detailed information on the real galaxy distribution is unknown, one has to make some assumptions on the statistical properties of such a distribution (see, e.g., [1]). Since we do not want to use such **ad hoc** assumptions, as we employ statistical methods that aim to tests some of the most common assumptions in the analysis of galaxy correlations [29], we have adopted the following strategy. From the one hand, we have considered an angular region which does not include the survey edges where completeness varies mostly. On the other hand, we have done a few tests to control the effect of completeness on the correlation analysis, and in particular, as we discuss below, we have focused on testing the stability the results. Note that the incompleteness due to fiber collisions can be neglected in measurements of large-scale (i.e., \( r > 10 \text{ Mpc}/h \)) galaxy correlations as this

**Table 1:** Main properties of the SDSS VL samples: \( R_{\text{min}}, R_{\text{max}} \) (in Mpc/h) are the chosen limits for the metric distance; \( M_{\text{min}}, M_{\text{max}} \) define the interval for the absolute magnitude in each sample and \( N_p \) is the number of galaxies in the sample.

| VL sample | \( R_{\text{min}} \) | \( R_{\text{max}} \) | \( M_{\text{min}} \) | \( M_{\text{max}} \) | \( N_p \) |
|-----------|-----------------|-----------------|-----------------|-----------------|-------|
| VL1       | 50              | 200             | -18.9           | -21.1           | 73811 |
| VL2       | 125             | 400             | -20.5           | -22.2           | 129975|
| VL3       | 200             | 600             | -21.6           | -22.8           | 51698 |

effects is only relevant at very small scales (see [24] and references therein).

In order to construct VL samples (see table 1) we computed the metric distances \( R \) using the standard cosmological parameters, i.e., \( \Omega_M = 0.3 \) and \( \Omega_\Lambda = 0.7 \). Absolute magnitudes \( M \) are computed using Petrosian apparent magnitudes in the \( m_r \) filter corrected for Galactic absorption and applying standard \( K \)-corrections [30].

**Statistical methods.** – We consider the number of galaxies \( N_i(r) \) within radius \( r \) around the \( i \)-th galaxy: this quantity differs for each galaxy and hence we consider it as a random variable characterizing its statistical properties. The average over an ensemble of realizations, \( \langle N(r) \rangle \), can be estimated by the volume average (assuming ergodicity)\(^2\)

\[
\langle N(r) \rangle = \frac{1}{M(r)} \sum_{i=1}^{M(r)} N_i(r).
\]

Note that the number of points contributing to the average eq. (1), \( M(r) \), depends on the scale \( r \) as an effect of the requirement that the spheres must be fully included in the sample’s boundaries [24]. To estimate the typical fluctuations of the random variable \( N_i(r) \) we may consider the conditional variance \( \Sigma^2(r) = \langle N^2(r) \rangle - \langle N(r) \rangle^2 \). In general [29], this can be written as the sum of two contributions: \( \Sigma^2(r) = \Sigma^2_i(r) + \Sigma^2_p(r) \), where \( \Sigma_i(r) \) is the intrinsic part, due to correlations, and \( \Sigma_p^2(r) = \langle N(r) \rangle \), is due to the Poisson noise. The normalized variance is defined as \( \sigma^2(r) = \Sigma^2(r)/\langle N(r) \rangle^2 = \sigma_i^2(r) + \sigma_p^2(r) \), where its intrinsic part can be estimated by

\[
\sigma_i^2(r) = \frac{1}{N_i(r)^2} \sum_{i=1}^{M(r)} (N_i - \langle N_i \rangle)^2 \quad \frac{1}{\langle N(r) \rangle}.
\]

**Results.** – We have measured the conditional average density, \( \bar{n}(r) = \langle N(r) \rangle/V(r) \) where \( V(r) = 4/3 \pi r^3 \), in the different SDSS VL samples (see fig. 1). We find that: i) at small scales, \( r \in [0.5, 20] \text{ Mpc}/h \) \( \bar{n}(r) \propto r^{-\gamma} \) with \( \gamma = 0.88 \pm 0.05 \) (i.e. \( \bar{n}(r) \propto r^D \)), with \( D = 3 - \gamma = 2.12 \pm 0.05 \). The error of the exponent has been derived by measuring the scattering in the different samples.

\(^2\)The symbols \( \langle \ldots \rangle \) \( (\ldots) \) stands for the ensemble (volume) average performed with the condition that the sphere center coincides with a point of the distribution, i.e. it is a conditional average [29].
ii) At larger scales, i.e., $r > 30 \text{Mpc/h}$ and up to $\approx 150 \text{Mpc/h}$ in the deepest sample, the exponent is $\gamma = 0.2 \pm 0.05$ (i.e., $D = 2.8 \pm 0.05$). The amplitude of the conditional average density in the different VL samples is in principle fixed by a luminosity factor that depends on the absolute luminosity of the galaxies therein included [24]. However, in a finite sample, as long as correlators extend on scales of the order of the sample size, the amplitude of $\sigma(r)$ also depends on the particular structures present in that specific volume. Therefore, the galaxy counts normalization in samples of different sizes, covering different space volumes and including galaxies of different luminosity depends on systematic effects which become negligible only for very large sample sizes. For this reason in fig. 1, by choosing an arbitrary normalization, we have plotted $n(r)/n(r_0)$ where $r_0 = 30 \text{Mpc/h}$. On the other hand, note that the exponent of $n(r)$ does not show variations larger than the estimated error bar (see below) in the different samples.

The main result from this analysis is that the slopes of the galaxy average density, both at small and large scales, show a very good agreement in the different samples. When the radius $r$ approaches the size of the largest sphere included in each sample there are systematic effects, as shown by the fact that the large-scale tail of $n(r)$ (where $M(r)$ in eq. (1) becomes rapidly very small) grows at different scales in the different samples.

In order to quantify fluctuations affecting the large-scale (i.e., $r > 10 \text{Mpc/h}$) behavior of $n(r)$ it is necessary to estimate statistical errors. For instance, one may simply compute them as $\sqrt{\Sigma^2(r)V^{-2}(r)M^{-1}(r)}$: however, in this way one underestimates the true errors because large-scale correlations can break the Central-Limit Theorem and thus the different determinations are not independent [29] (note that such error bars would not be even visible in the plot in fig. 1). A possible evaluation of the scattering of $n(r)$ can be performed by measuring sample-to-sample variations [15]. To this aim, we calculate thus eq. (2) in the different samples (see fig. 2) finding that at small scales results are very similar while at large scales a clear difference is detected.

Interestingly, the differences in $\sigma^2(r)$ occur at a scale that grows with sample’s sizes. The determination in the smallest sample, i.e., VL1, shows a marked large-scale difference with respect to those in VL2 and VL3, which instead present a more similar behavior everywhere but a very large-scale tail (i.e., $r \approx 100 \text{Mpc/h}$). Therefore a possible explanation of this large-scale behavior is that it is due to a finite-size effect, i.e., that the scale at which the abrupt decay of $\sigma^2(r)$ occurs in the different samples, depends on the their sizes. To show that this is the case we have considered the behavior of $\sigma^2(r)$ in subvolumes of the VL2 and VL3 samples limited to a depth (i.e., $R_{\text{max}}$) smaller than the ones reported in table 1. In this way we are able to clearly single out the finite-size effect (see fig. 3).

We performed another test by cutting the sample into two subsamples of equal volume (i.e., limited in angles by the regions $R_2$ and $R_3$) and determining $\bar{n}(r)$ and $\sigma^2(r)$ in each of them (see fig. 4). In the case of VL3 the determinations of both the average density and the variance in two subsamples agree better than for VL1: this corroborates the result, obtained also with the previous tests, that systematic finite-size effects become weaker in samples with larger volumes [22,24,31]. This is in agreement with the larger dimension observed at large-scale, which also implies less wild fluctuations [29].

From these behaviors we may conclude that both $\bar{n}(r)$ and $\sigma^2(r)$ converge to a well-defined value only for scales $r < r_s$ smaller than the radius of the largest sphere included in the sample. We estimate that for VL1...
Fig. 3: (Colour on-line) Normalized conditional variance in subsamples of VL2 (upper panel), respectively with $R_{\text{max}} = 250, 300, 400$ Mpc/h and VL3 (bottom panel) with $R_{\text{max}} = 300, 400, 500, 600$ Mpc/h.

$r_s \approx 50$ Mpc/h, for VL2 $r_s \approx 100$ Mpc/h and for VL3 $r_s \approx 150$ Mpc/h. Beyond these scales the behaviors in the two half samples systematically differ: this is in agreement with the results by [8,9,22–24,31] where the whole behavior of the PDF of $N_i(r)$ was considered.

As mentioned above it is interesting to consider the effect of the small (i.e. $\sim 5\%$) angular incompleteness on the measurements of galaxy correlations. For instance, there are some small sky regions where, due to the presence of bright stars, galaxies have not been observed. Each bright star corresponds to a circular hole where galaxies cannot be observed. The radius of the circle around each bright star depends on the star apparent magnitude [30]. An upper limit to such a size is $\theta = 3$ arcsec, corresponding to very bright stars [30]. A simple way to test for effect of such holes is the following.

We distribute randomly holes of size $\theta$ with a surface density of 50 per square degree (roughly corresponding the surface density of bright stars [32]). Galaxies which are placed inside one of the holes are cut from the resulting sample. In such a way we artificially have taken away from the sample, in a correlated way at small angles, about $\sim 10\%$ of the galaxies (i.e. about the double than the angular incompleteness of the catalog). The question is whether the large-scales (i.e., $r > 1$ Mpc/h) correlation properties are affected by such an incompleteness. Results are shown in fig. 5: one may note that, apart from an obvious $10\%$ shift in the amplitude of the average conditional density, no scale-dependent changes are manifested. Thus we can confidently conclude that incompleteness effects do not play a major role in the results of the correlation analysis. We refer to a forthcoming work for a more complete series of tests on the angular incompleteness of this survey.

Mock galaxy catalogs. – We have repeated the same analysis in some mock galaxy samples. These are constructed from cosmological N-body simulations of the standard LCDM model [33,34], by applying the same cuts in absolute magnitude and distance as those reported in table 1 and by computing the redshift positions (i.e., simply applying the corrections to the redshift due to the peculiar velocities along the line of sight). The conditional average density (see fig. 6) shows i) a slope $\gamma \approx 1$ for $r \in [0.5, 20]$ Mpc/h that is weakly dependent on the average luminosity of the galaxies, and ii) a well-defined crossover to uniformity, i.e. $\gamma = 0$, at $\sim 30$ Mpc/h. Both features are
Large-scale galaxy correlations

Fig. 5: (Colour on-line) Conditional average density for the full sample VL2 and for a modified version of it (VL2h) where $\sim 10\%$ of the galaxies have been cut in a correlated way (see text): apart from an obvious 10\% shift in the amplitude of the average conditional density, no scale-dependent changes are manifested.

Fig. 6: (Colour on-line) Conditional average density, normalized to its value at 30 Mpc/h, for the different samples of the mock galaxy catalog. The exponent $\gamma$ reported in the labels is the best fit in the range [0.5, 20] Mpc/h.

Fig. 7: (Colour on-line) Normalized conditional variance for the different samples of the mock galaxy catalog. No finite-size effects are present in this case.

different from the real SDSS samples. Correspondingly to the crossover toward uniformity (absence of large-scales strong correlations), the conditional variance (fig. 7) does not show a finite-size dependence, contrary to what occurs for the real samples (fig. 2). In this case, the stability of the normalized variance in the different samples is shown by diving the samples into two parts and by comparing the behaviors: because of the lack of large-scales correlations, no systematic differences are found between the two subsamples.

Note that a given mock galaxy sample is constructed from the underlying dark-matter particles distribution, by assuming a certain prescription to identify galaxies. Such a prescription is generally based on a physical mechanism which links the local density of dark-matter particles to the probability to form a mock galaxy [34]. In principle, one can introduce different physically motivated prescriptions to relate dark- to visible-matter distributions in cosmological simulations. The question is then whether a different prescription can substantially change the resulting mock galaxies correlation properties, possibly giving rise to a better agreement with observations. We do not explore this question here and we cannot exclude that a better agreement could be obtained with a different bias prescription. However we note that, given that large-scale correlations are not present in the underlying dark-matter distribution, as this is what theoretical models predicts on scales of the order of $\sim 100$ Mpc/h [13,35], the only way to introduce such correlations is via a bias mechanism: biasing should correspond to a large-scale correlated selection. This implies the biasing mechanism to be non-local, i.e. a completely different one from what is generally considered to be a physically plausible prescription.

Discussion. – In summary, we have analyzed the scaling properties of the galaxy distribution in the SDSS-DR7 samples. We found that at small scales $r \in [0.5, 20]$ Mpc/h the galaxy (conditional) average density decays as a power-law function of distance, $n(r) \propto r^{-\gamma}$ with an exponent $\gamma \approx 0.9$, while at large scales, i.e. $r > 30$ Mpc/h, it decays more slowly: $\gamma \approx 0.2$. The analysis of the variance (eq. (2)) allows to clearly single out a finite-size effect, that can be simply understood as due to galaxy correlations extending up to the largest scales of the considered samples. These behaviors differ from those found in mock galaxy samples where: i) at small scales the exponent, $\gamma \approx 1$, slightly depends on galaxy luminosity, and ii) at large scales, i.e. $r > 30$ Mpc/h, a well-defined crossover to uniformity (i.e. $\gamma = 0$) is found. In addition, no finite-size effects are detected in the large-scale behavior of the normalized counts variance.

The detection of very large-scale galaxy correlations and of finite-size effects, allow us to conclude that the observed
galaxy distribution is more correlated, fluctuating and thus clustered, than the one predicted by the standard LCDM model of galaxy formation through cosmological $N$-body simulations [33,34]. These findings are in agreement with the results obtained in the same samples, although at smaller scales $r < 100\,\text{Mpc}/\text{h}$, through different tests [10,22–24,31], with the results obtained in the 2-degree-field redshift survey [4–9] and other surveys [14–18]. Note that the results obtained by [21] are also compatible with a slow decay of $n(r)$ at large scales.

Because of the large-scale scaling of the galaxy average density, we conclude that any statistical quantity normalized to the estimation of the sample average density (e.g., the standard two-point correlation function) is biased by finite-size effects [29]. This implies that the volume of current galaxy samples is not yet large enough to measure the standard two-point correlation function at $\sim 100\,\text{Mpc}/\text{h}$. From a theoretical point of view, the main challenge of our results concerns the way in which the deviations from a pure statistically homogeneous and isotropic, and spatially uniform density field [31] imply the consideration of the inhomogeneities effects on the dynamics of the universe and the deviations that can possibly be introduced with respect to the simple FRW models [36–40].

***

I thank Y. V. Baryshev and D. Tekhanovich for useful discussions and comments. I acknowledge the use of the Sloan Digital Sky Survey data (http://www.sdss.org), of the NYU Value-Added Galaxy Catalog (http://ssds.physics.nyu.edu/) and of the Millennium run semi-analytic galaxy catalog (http://www.mpa-garching.mpg.de/galform/agnpaper).

REFERENCES

[1] Zehavi I. et al., Astrophys. J., 571 (2002) 172.

[2] Hawkins E. et al., Mon. Not. R. Astron. Soc., 346 (2003) 78.

[3] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., Astron. Astrophys., 465 (2007) 23.

[4] Frith W. J. et al., Mon. Not. R. Astron. Soc., 345 (2003) 1049.

[5] Busswell G. S. et al., Mon. Not. R. Astron. Soc., 354 (2004) 991.

[6] Frith W. J. et al., Mon. Not. R. Astron. Soc., 371 (2006) 1601.

[7] Einasto J. et al., Astron. Astrophys., 450 (2006) 1.

[8] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., EPL, 85 (2009) 29002.

[9] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., Astron. Astrophys., 496 (2009) 7.

[10] Yang A. and Laszlo W. C., Astrophys. J., 729 (2011) 1.

[11] Einasto J. et al., Astrophys. J., 736 (2011) 51.

[12] Peebles P. J. E., Principles Of Physical Cosmology (Princeton University Press, Princeton, NJ) 1993.

[13] Gabrielli A. et al., Phys. Rev. D, 65 (2002) 083523.

[14] Martínez V. J. et al., Astrophys. J., 696 (2009) L93.

[15] Sylos Labini F. et al., Astron. Astrophys., 505 (2009) 981.

[16] Kazin et al., Astrophys. J., 710 (2010) 1444.

[17] Thomas S. A., Abdalla F. B. and Lahav O., Phys. Rev. Lett., 106 (2011) 241301.

[18] Sawangwit U. et al., Mon. Not. R. Astron. Soc., 416 (2011) 3033.

[19] York D. et al., Astron. J., 120 (2000) 1579.

[20] Adelman-McCarthy J. K. et al., Astrophys. J. Suppl., 175 (2008) 297.

[21] Hogg D. W. et al., Astrophys. J., 624 (2005) 54.

[22] Antal T., Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., EPL, 88 (2009) 59001.

[23] Sylos Labini F. et al., EPL, 86 (2009) 49001.

[24] Sylos Labini F., Vasilyev N. L. and Baryshev Yu. V., Astron. Astrophys., 508 (2009) 17.

[25] Loveday J., Mon. Not. R. Astron. Soc., 347 (2004) 601.

[26] Blanton M. R. et al., Astrophys. J., 592 (2003) 819.

[27] Abazajian K. et al., Astrophys. J. Suppl., 182 (2009) 543.

[28] Strauss M. A. et al., Astron. J., 124 (2002) 1810.

[29] Gabrielli A., Sylos Labini F., Joyce M. and Pietronero L., Statistical Physics for Cosmic Structures (Springer Verlag, Berlin) 2005.

[30] Blanton M. R. et al., Astron. J., 129 (2005) 2562.

[31] Sylos Labini F. and Baryshev Yu. V., JCAP, 06 (2010) 021.

[32] Hod E. et al., Astron. Astrophys., 355 (2000) L27.

[33] Springel V. et al., Nature, 435 (2005) 629.

[34] Croton D. J. et al., Mon. Not. R. Astron. Soc., 365 (2006) 11.

[35] Sylos Labini F. and Vasilyev N. L., Astron. Astrophys., 477 (2008) 381.

[36] Rasanen S., Int. J. Mod. Phys., D17 (2008) 2543.

[37] Buchert T., Gen. Relativ. Gravit., 40 (2008) 467.

[38] Wiltshire D. L., Int. J. Mod. Phys., 17 (2008) 641.

[39] Clifton T. and Ferreira P. G., Phys. Rev. D, 80 (2009) 103503.

[40] Maartens R., arXiv:1104.1300 (2011).