

The VIMOS Public Extragalactic Redshift Survey (VIPERS) [★]

On the correct recovery of the count-in-cell probability distribution function.

J. Bel², E. Branchini^{10,28,29}, C. Di Porto⁹, O. Cucciati^{17,9}, B. R. Granett², A. Iovino², S. de la Torre⁴, C. Marinoni^{7,30,31}, L. Guzzo^{2,27}, L. Moscardini^{17,18,9}, A. Cappi^{9,21}, U. Abbas⁵, C. Adami⁴, S. Arnouts⁶, M. Bolzonella⁹, D. Bottini³, J. Coupon³², I. Davidzon^{9,17}, G. De Lucia¹³, A. Fritz³, P. Franzetti³, M. Fumana³, B. Garilli^{3,4}, O. Ilbert⁴, J. Krywult¹⁵, V. Le Brun⁴, O. Le Fèvre⁴, D. Maccagni³, K. Małek²³, F. Marulli^{17,18,9}, H. J. McCracken¹⁹, L. Paioro³, M. Polletta³, A. Pollo^{22,23}, H. Schlegelhauser^{24,20}, M. Scodreggio³, L. A. .M. Tasca⁴, R. Tojeiro¹¹, D. Vergani^{25,9}, A. Zanichelli²⁶, A. Burden¹¹, A. Marchetti^{1,2}, Y. Mellier¹⁹, R. C. Nichol¹¹, J. A. Peacock¹⁴, W. J. Percival¹¹, S. Phleps²⁰, and M. Wolk¹⁹

(Affiliations can be found after the references)

ABSTRACT

We compare three methods to measure the count-in-cell probability density function of galaxies in a spectroscopic redshift survey. From this comparison we found that when the sampling is low (the average number of object per cell is around unity) it is necessary to use a parametric method to model the galaxy distribution. We used a set of mock catalogues of VIPERS, in order to verify if we were able to reconstruct the cell-count probability distribution once the observational strategy is applied. We find that in the simulated catalogues, the probability distribution of galaxies is better represented by a Gamma expansion than a Skewed Log-Normal. Finally, we correct the cell-count probability distribution function from the angular selection effect of the VIMOS instrument and study the redshift and absolute magnitude dependency of the underlying galaxy density function in VIPERS from redshift 0.5 to 1.1. We found very weak evolution of the probability density distribution function and that it is well approximated, independently from the chosen tracers, by a Gamma distribution.

Key words. Cosmology: cosmological parameters – cosmology: large scale structure of the Universe – Galaxies: high-redshift – Galaxies: statistics

1. Introduction

The galaxy clustering offers a formidable playground to try to understand how structures have been growing during the evolution of the universe. A number of statistical tools have been developed and used over the past thirty years (see Bernardeau et al. 2002, for a review). In general, these statistical methods use the fact that the clustering of galaxies is due to the gravitational pull of the underlying matter distribution. Hence, the study of the spatial distribution of galaxies in the universe allows us to get information about the statistical properties of its matter content. As a result, it is of paramount importance to be able to measure the statistical quantities describing the galaxy distribution from a redshift survey.

Send offprint requests to: Bel., J.
e-mail: julien.bel@brera.inaf.it

* based on observations collected at the European Southern Observatory, Cerro Paranal, Chile, using the Very Large Telescope under programmes 182.A-0886 and partly 070.A-9007. Also based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. The VIPERS web site is <http://www.vipers.inaf.it/>.

The development of multi-object spectrographs on 8-m class telescopes during the 1990s triggered a number of deep redshift surveys with measured distances beyond $z \sim 0.5$ over areas of $1-2 \text{ deg}^2$ (e.g. VVDS Le Fevre et al. 2005, DEEP2 Newman et al. 2012 and zCOSMOS Lilly et al. 2009). Even so, it was not until the wide extension of VVDS was produced (Garilli et al. 2008), that a survey existed with sufficient volume to attempt cosmologically meaningful computations at $z \sim 1$ (Guzzo et al. 2008). In general, clustering measurements at $z \approx 1$ from these samples remained dominated by cosmic variance, as dramatically shown by the discrepancy observed between the VVDS and zCOSMOS correlation functions at $z \approx 0.8$ (de la Torre et al. 2010).

The VIMOS Public Extragalactic Redshift Survey (VIPERS) is part of this global attempt to take cosmological measurements at $z \sim 1$ to a new level in terms of statistical significance. In contrast to the BOSS and WiggleZ surveys, which use large-field-of-view ($\sim 1 \text{ deg}^2$) fibre optic positioners to probe huge volumes at low sampling density, VIPERS exploits the features of VIMOS at the ESO VLT to yield a dense galaxy sampling over a moderately large field of view ($\sim 0.08 \text{ deg}^2$). It reaches a volume at $0.5 < z < 1.2$ comparable to that of the 2dFGRS (Colless et al. 2001) at $z \sim 0.1$, allowing the cosmological evolution to be tested with small statistical errors.

The VIPERS redshifts are being collected by tiling the selected sky areas with a uniform mosaic of VIMOS fields. The area covered is not contiguous, but presents regular gaps due to the specific footprint of the instrument field of view, in addition to intrinsic unobserved areas due to bright stars

or defects in the original photometric catalogue. The VIMOS field of view has four rectangular regions of about 8×7 square arcminutes each, separated by an unobserved cross (Guzzo et al. 2014; de la Torre et al. 2013). This creates a regular pattern of gaps in the angular distribution of the measured galaxies. Additionally, the Target Sampling Rate and the Survey Success Rate vary among the quadrants, and a few of the latter were lost because of mechanical problems within VIMOS (Garilli et al. 2014). Finally, the slit-positioning algorithm, SPOC (see Bottini et al. 2005), also introduces some small-scale angular selection effects, with different constraints along the dispersion and spatial directions of the spectra, as thoroughly discussed in de la Torre et al. (2013). Clearly, this combination of angular selection effects has to be taken properly into account when estimating any clustering statistics.

In this paper we measure the probability distribution function from the VIPERS Public Data Release 1 (PDR-1) redshift catalogue, including $\sim 64\%$ of the final number of redshifts expected at completion (see Guzzo et al. 2014; Garilli et al. 2014 for a detailed description of the survey data set). The paper is organized as follows. In §2 we introduce the VIPERS survey and the features of the PDR-1 sample. In §3 we review the basics of the three methods we compared. In §4 we present a null test of the three method on a synthetic galaxy catalogue. In §5 we use galaxy mock catalogues to assess performances of two of the methods. Magnitude and redshift dependence of the probability distribution function of VIPERS PDR-1 galaxies are presented in §6 and conclusions are drawn in §7.

Throughout, the Hubble constant is parameterized via $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, all magnitudes in this paper are in the AB system (Oke & Gunn 1983) and we will not give an explicit AB suffix. In order to convert redshifts into comoving distances we assume that the matter density parameter is $\Omega_m = 0.27$ and that the universe is spatially flat with a Λ CDM cosmology without radiations.

2. Data

The VIMOS Public Extragalactic Redshift Survey (VIPERS) is a spectroscopic redshift survey being built using the VIMOS spectrograph at the ESO VLT. The survey target sample is selected from the Canada-France-Hawaii Telescope Legacy Survey Wide (CFHTLS-Wide) optical photometric catalogues (Mellier et al. 2009). The final VIPERS will cover $\sim 24 \text{ deg}^2$ on the sky, divided over two areas within the W1 and W4 CFHTLS fields. Galaxies are selected to a limit of $i_{AB} < 22.5$, further applying a simple and robust *gri* colour pre-selection, as to effectively remove galaxies at $z < 0.5$. Coupled to an aggressive observing strategy (Scodreggio et al. 2009), this allows us to double the galaxy sampling rate in the redshift range of interest, with respect to a pure magnitude-limited sample ($\sim 40\%$). At the same time, the area and depth of the survey result in a fairly large volume, $\sim 5 \times 10^7 \text{ h}^{-3} \text{ Mpc}^3$, analogous to that of the 2dFGRS at $z \sim 0.1$. Such combination of sampling and depth is quite unique over current redshift surveys at $z > 0.5$. The VIPERS spectra are collected with the VIMOS multi-object spectrograph (Le Fevre et al. 2003) at moderate resolution ($R = 210$), using the LR Red grism, providing a wavelength coverage of 5500-9500Å and a typical redshift error of $141(1+z) \text{ km sec}^{-1}$. The full VIPERS area is covered through a mosaic of 288 VIMOS pointings (192 in the W1 area, and 96 in the W4 area). A discussion of the survey data reduction and management infrastructure is presented in Garilli et al. (2012). An early subset of the

Table 1. List of the magnitude selected objects (in B-band) in the VIPERS PDR-1

z_{\min}	z_{\max}	luminosity $M_B - 5 \log(h) <$	$\bar{\rho}$ (Eq. 1) $10^{-3} h^3 \text{ Mpc}^{-3}$
0.5	0.7	$-18.6 - z$	4.49
0.5	0.7	$-19.1 - z$	2.96
0.5	0.7	$-19.5 - z$	1.88
0.5	0.7	$-19.7 - z$	1.43
0.5	0.7	$-19.9 - z$	1.04
0.7	0.9	$-19.1 - z$	2.47
0.7	0.9	$-19.5 - z$	1.66
0.7	0.9	$-19.7 - z$	1.25
0.7	0.9	$-19.9 - z$	0.912
0.9	1.1	$-19.5 - z$	0.622
0.9	1.1	$-19.7 - z$	0.535
0.9	1.1	$-19.9 - z$	0.425

spectra used here is analyzed and classified through a Principal Component Analysis (PCA) in Marchetti et al. (2012).

A quality flag is assigned to each measured redshift, based on the quality of the corresponding spectrum. Here and in all parallel VIPERS science analyses we use only galaxies with flags 2 to 9 inclusive, corresponding to a global redshift confidence level of 98%. The redshift confirmation rate and redshift accuracy have been estimated using repeated spectroscopic observations in the VIPERS fields. A more complete description of the survey construction, from the definition of the target sample to the actual spectra and redshift measurements, is given in the parallel survey description paper (Guzzo et al. 2014).

The data set used in this paper and the other papers of this early science release is the VIPERS Public Data Release 1 (PDR-1) catalogue, which have been made publicly available in September 2013. This includes 55,359 objects, spread over a global area of $8.6 \times 1.0 \text{ deg}^2$ and $5.3 \times 1.5 \text{ deg}^2$ respectively in W1 and W4. It corresponds to the data frozen in the VIPERS database at the end of the 2011/2012 observing campaign, i.e. 64% of the final expected survey. For the specific analysis presented here, the sample has been further limited to its higher-redshift part, selecting only galaxies with $0.55 < z < 1.1$. The reason for this selection is related to minimizing the shot noise and maximizing the volume. This reduces the usable sample to 18135 and 16879 galaxies in W1 and W4 respectively (always with quality flags between 2 and 9). The corresponding effective volume of the two samples are 6.57 and $6.14 \times 10^6 \text{ h}^{-3} \text{ Mpc}^3$. At redshift $z = 1.1$ they span respectively the angular comoving distances ~ 370 and $230 \text{ h}^{-1} \text{ Mpc}$. We divide the W1 and W4 fields in three redshift bins and we build magnitude limited subsamples in each of them. For convenience, we use the magnitude limits listed in Table (1) of di Porto et al. (2014), which we recall in Tab. (2).

The VIMOS footprint has an important impact on the observed probability of finding N galaxies in a randomly placed spherical cell in the survey volume. As a matter of fact, a direct appreciation of the masked area can be shown on the first moment of the probability distribution, i.e. the expectation value of the number count $\bar{N} \equiv \sum_{N=0}^{\infty} NP_N$. On one hand, we can predict the mean number of objects per cells from the knowledge of the number density in each considered redshift bins and on the other hand we can estimate it by placing a regular grid of spherical cells of radius R into the volume surveyed by VIPERS. In fact, given the solid angle of W1 and W4 and the corresponding num-

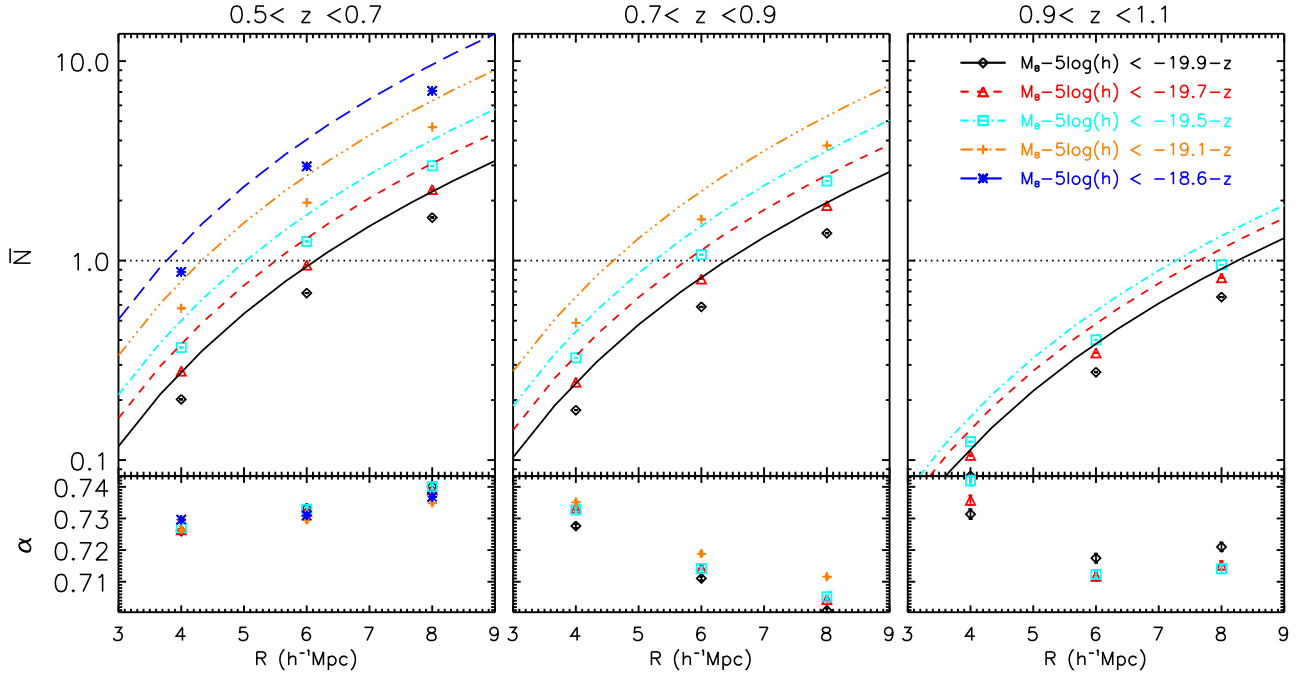


Fig. 1. *Upper:* Expected mean number count in spheres (solid line, from Eq. 2) with respect to the observed one (symbols) for the various luminosity cuts and for the three redshift bins [0.5, 0.7] (left panel), [0.7, 0.9] (central panel), and [0.9, 1.1] (right panel). The selection in absolute magnitude M_B in B-band corresponding to each symbols/lines and colors are indicated in the inset. The dotted line displays the $\bar{N} = 1$. *Lower:* Displays the deviation α (see Eq. 3) between the expected mean number \bar{N}_R and the observed one \bar{N} with respect to the radius R of the cells.

ber of galaxies N_1 and N_4 contained in a redshift bin extracted from each field, one can estimate the total number density as

$$\bar{\rho} = \frac{N_1 + N_4}{\Omega_1 + \Omega_4} \frac{1}{V_k}, \quad (1)$$

where V_k is defined as the volume corresponding to a sector of a spherical shell with solid angle equal to unity. In the case of VIPERS PDR-1 the effective solid angles corresponding to W1 and W4 are respectively $\Omega_1 = 1.6651683 \times 10^{-3}$ and $\Omega_4 = 1.5573021 \times 10^{-3}$ (in square radians). One can therefore predict the corresponding expected number of objects in each cell by multiplying the averaged number density by the volume of a cell. It reads

$$\bar{N}_R = \frac{4}{3} \pi R^3 \bar{\rho}, \quad (2)$$

in the case of the spherical cells of radius R considered in this work. The expectation value \bar{N}_R with respect to the radius of the cells corresponding to each luminosity sub-sample extracted from VIPERS-PDR1 is represented by lines in Fig. (1). On the same figure we display the measured mean number of object \bar{N} in each redshift bins. Note that to perform this measurement we place a grid of equally separated ($4h^{-1}\text{Mpc}$) spheres of radius $R = 4, 6, 8h^{-1}\text{Mpc}$ and we reject spheres with more than 40% of their volume outside the observed region (see Bel et al. 2014). We quantify the effect of the mask using the quantity

$$\alpha \equiv \frac{\bar{N}}{\bar{N}_R}, \quad (3)$$

in fact the bottom panels of Fig. (1) shows that for all sub samples and at all redshifts the neat effect of the masks is to under-sample the galaxy field by roughly 72%. It also shows that the correction factor α depends on the considered redshift, on the

luminosity and on the cell-size. The scale dependency can be explained by the fact that the correction parameter α depends on how the cells overlap with the masked regions. The left panel of Fig. (1) suggests that at low redshift the mask effect behaves in the same way for all the luminosity samples while the middle panel shows a clear dependency with respect to luminosity. The correction factor α depends on the redshift distribution, as a result the apparent dependency with respect to the luminosity is due to the dependence of the number density with respect to the luminosity of the considered objects.

The mask not only modifies the mean number of object but it also modifies the higher order moments of the distribution, such that the measured P_N will be systematically altered. In the present paper we show that this systematic effect can be taken into account by measuring the underlying probability density function of the galaxy density contrast δ . It has been shown (see Fig. (8) of Bel et al. 2014) that after rejecting spheres with more than 40% of their volume outside the survey, the local poisson process approximation holds. In particular, it allows to use the “wrong” probability distribution function in order to get reliable information on the underlying probability density function $p(\delta)$. Then applying the Poisson sampling one can recover the unaltered P_N using that $\bar{N} = \bar{N}(\text{masked})/\alpha$. For the sake of completeness we provide the reader with the measured probability function obtained after rejecting the cells with more than 40% of their volume outside the survey (see Fig. 8).

In particular, let P_M and P_N , respectively, be the observed and the true Counting Probability Distribution Function (CPDF). Assuming that from the knowledge of P_M there exists a process to get the underlying probability density function of the stochastic field Λ , which is associated to the random variable N , one can compute the true CPDF applying

$$P_N = \int_0^\infty P[N|\Lambda]p(\Lambda)d\Lambda, \quad (4)$$

where $P[N|\Lambda]$ is called the sampling conditional probability; it determines the sampling process from which the discrete cell-count arises. In the following we assume that this sampling conditional probability follows a Poisson law (Layzer 1956), as a result in Eq. (4) we substitute

$$P[N|\Lambda] = K[N, \Lambda] \equiv \frac{\Lambda^N}{N!} e^{-\Lambda}. \quad (5)$$

It is also convenient to express Eq. (4) in terms of the density contrast of the stochastic field Λ , $\delta \equiv \Lambda/\bar{\Lambda} - 1$, it follows that

$$P_N = \int_{-1}^{\infty} K[N|\bar{N}(1+\delta)]p(\delta)d\delta, \quad (6)$$

where we used that $\bar{\Lambda} = \bar{N}$, which is a property of the Poisson sampling.

Continuing along this direction that we propose to compare three methods which aim at extracting the underlying probability density function (PDF) in order to correct the observed CPDF from the angular selection effects of VIPERS.

3. Methods

In this section we review the PDF estimators that we use and compare with each others in this paper. The purpose is to select the method which will be more adapted to the VIPERS characteristics.

3.1. The Richardson-Lucy deconvolution

This is an iterative method which aims at inverting Eq. (6) without parametrizing the underlying PDF, it has been investigated by Szapudi & Pan (2004). This method starts with an initial guess p_0 for the probability density function p which is used to compute the corresponding expected observed $P_{N,0}$ via

$$P_{N,0} = \int_{-1}^{\infty} \hat{K}[N, \bar{N}(1+\delta)]p_0(\delta)d\delta,$$

where $\hat{K}[N, \bar{N}(1+\delta)] \equiv K/\sum_N K$. The probability density function used at the next step is obtained using

$$\hat{p}_{i+1}(\delta) = \hat{p}_i(\delta) \sum_{N=0}^{N_{max}} \frac{P_N}{P_{N,i}} \hat{K}[N, \bar{N}(1+\delta)],$$

where $\hat{p} \equiv p \sum_N K$. For each step the agreement between the expected observed probability distribution $P_{N,i}$ and the true one P_N is quantified by

$$\chi_i^2 \equiv \sum_{N=0}^{N_{max}} \left(\frac{P_N}{P_{N,i}} - 1 \right)^2.$$

It is therefore possible to know the evolution of the cost function χ^2 with respect to the steps i .

In fact it has been shown by Szapudi & Pan (2004) that it converges toward a constant value which corresponds to the best evaluation of the probability density function p given the observed probability distribution P_N . Since these authors have shown that this convergence occurs after around 30 iterations. We did our own convergence tests which have shown that adopting a value of 30 iterations is enough. However, it happens that the evolution of the χ^2 is not always monotonic. In practice, we store the χ^2 result of each step and we look for the step for which the χ^2 is minimum, i.e. $p(\delta) = p_{i_{min}}(\delta)$. As an initial guess we set that the discret CPDF is equal to the continuous one ($p_0(\delta) = p$).

3.2. The Skewed Log-Normal

This is a parametric method where the shape of the probability density depends on a given number of parameters, in this case the probability density function is assumed to be well described by a Skewed Log-Normal (Colombi 1994) distribution. It is derived from the Log-Normal distribution (Coles & Jones 1991) but it is more flexible. It is indeed built upon an Edgeworth expansion; be the stochastic field $\Phi \equiv \ln(1+\delta)$, following a Normal distribution then the density contrast δ follows instead a Log-Normal distribution. In the case of the Skewed Log-Normal (SLN) density function, the field Φ follows an Edgeworth expanded Normal distribution

$$P_\Phi(\Phi) \equiv \left\{ 1 + \frac{\langle v^3 \rangle_c}{6} H_3(v) + \frac{\langle v^4 \rangle_c}{24} H_4(v) + \frac{5}{72} \langle v^3 \rangle_c^2 H_6(v) \right\} \frac{G(v)}{\sigma_\Phi}, \quad (7)$$

where $v \equiv \frac{\Phi - \mu_\Phi}{\sigma_\Phi}$, G is the central reduced Normal distribution

$G(v) \equiv \frac{e^{-v^2/2}}{\sqrt{2\pi}}$ and $\langle v^n \rangle_c$ denotes the cumulant expectation value of v . As a result, the SLN is parameterized by the four parameters μ_Φ , σ_Φ , $\langle v^3 \rangle_c$, and $\langle v^4 \rangle_c$ which are related, respectively to the mean, the dispersion, the skewness and the kurtosis of the stochastic variable Φ . They can all be expressed in terms of cumulants $\langle \Phi^n \rangle_c$ of order n of the weakly non-Gaussian field Φ . In Szapudi & Pan (2004) they use a best fit approach and determine these parameters by minimizing the difference between the measured counting probability P_N and the one obtained from

$$P_N^{th} = \int_{-1}^{\infty} K[N, \bar{N}(1+\delta)] P_\Phi[\ln(1+\delta), \mu_\Phi, \sigma_\Phi^2, \langle \Phi^3 \rangle_c, \langle \Phi^4 \rangle_c] \times d \ln(1+\delta). \quad (8)$$

However, this requires us to perform the integral (Eq. 8) in a four dimensional parameter space which is numerically expensive.

In the present paper we use an alternative implementation which is computationally more efficient. Instead of trying to maximize the likelihood of the model given the observations, we rather use the observations to predict the parameters of the SLN. To do so we use the property of the local Poisson sampling (Bel & Marinoni 2012); the factorial moments $\langle (N)_f \rangle$ of the discrete counts are equal to the moments of the underlying continuous distribution $\langle \Lambda^n \rangle$. Since the transformation between the density contrast δ and the Edgeworth expanded field Φ is local and deterministic, it is possible to find a relation between the moments $\langle \Lambda^n \rangle$ and the cumulants $\langle \Phi^n \rangle_c$.

By definition, the moments of the positive continuous field Λ are given by

$$\langle \Lambda^n \rangle \equiv \int_0^{\infty} \Lambda^n P(\Lambda) d\Lambda,$$

then for a local deterministic transformation the conservation of probability imposes $P(\Lambda)d\Lambda = P_\Phi(\Phi)d\Phi$, it follows that the moments of Λ can be recast in terms of Φ

$$\langle \Lambda^n \rangle = \bar{\Lambda}^n \int_0^{\infty} e^{n\Phi} P_\Phi(\Phi) d\Phi.$$

In the right hand side one can recognize the definition of the moment generating function $\mathcal{M}_\Phi(t) \equiv \langle e^{t\Phi} \rangle$ we therefore obtain that

$$\mathcal{M}_\Phi(t = n) = \frac{\langle \Lambda^n \rangle}{\bar{\Lambda}^n} \equiv A_n. \quad (9)$$

This equation allows us to link the moment of Λ to the cumulants of Φ via the moment generating function \mathcal{M}_Φ .

Moreover, since the probability density P_ϕ is the product of a sum of Hermite polynomials with a Gaussian function it is straightforward to compute the explicit expression of the moment generating function we obtain

$$\mathcal{M}_\Phi(t) = \left\{ 1 + \langle \Phi^3 \rangle_c \frac{t^3}{6} + \langle \Phi^4 \rangle_c \frac{t^4}{24} + \langle \Phi^3 \rangle_c^2 \frac{5}{72} t^6 \right\} e^{t\mu_\Phi + t^2 \frac{\sigma_\Phi^2}{2}}. \quad (10)$$

As a matter of fact, Eq. (10) and Eq. (9) together allow to set up a system of four equations, for $n = 1, 2, 3, 4$ it reads

$$Y^{n^2} X^n B_n = A_n, \quad (11)$$

where $Y \equiv e^{\frac{\sigma_\Phi^2}{2}}$, $X \equiv e^{\mu_\Phi}$ and $B_n \equiv \mathcal{M}_\Phi(t = n, \mu_\Phi = 0, \sigma_\Phi = 0)$. In the system of equations (Eq. 11) the right hand side is given by observations and the left hand side depends on the cumulants μ_Φ , σ_Φ^2 , $\langle \Phi \rangle_c^3$ and $\langle \Phi \rangle_c^4$ parameterized in terms of X , Y , $x \equiv \langle \Phi \rangle_c^3$ and $y \equiv \langle \Phi \rangle_c^4$. In appendix (A), we detail the procedure to solve this non-linear system of equations.

We therefore, get the values of the four parameters of the SLN by simply measuring the moments of the counting variable N up to the fourth order.

3.3. The Gamma expansion

The Gamma expansion method follows the same idea as described in §3.2 but it uses a Gamma distribution instead of a Gaussian one. It uses the orthogonality properties of the Laguerre polynomials in order to modify the moments of the Gamma PDF. Such an expansion has been investigated in Gaztañaga, Fosalba & Elizalde (2000) where they compared it to the Edgeworth expansion in order to model the one-point PDF of the matter density field. Then it has been further extended, in a more general context, to multi-point distributions by Mustapha & Dimitrakopoulos (2010).

As mentioned above the Gamma expansion requires the use of the Gamma distribution ϕ_G defined as

$$\phi_G(u) \equiv \frac{u^{k-1}}{\theta \Gamma(k)} e^{-u}, \quad (12)$$

where Γ is the Gamma function (for an integer n , $\Gamma(n+1) = n!$), θ and k are two parameters which are related to the two first moments of the PDF. If the galaxy probability density function is well described by a Gamma expansion at order n then it can be formally written as

$$P(\Lambda) = \phi_G(u) f_n^{(k-1)}(u), \quad (13)$$

where by definition $u \equiv \frac{\Lambda}{\theta}$, $k = \frac{\bar{\Lambda}^2}{\sigma_\Lambda^2}$, $\theta \equiv \frac{\bar{\Lambda}}{k} = \frac{\sigma_\Lambda^2}{\bar{\Lambda}}$. The function $f_n^{(k-1)}$ represents the expansion aiming at tuning the moments of the Gamma distribution; note that the exponent $(k-1)$ is not the derivative of order $k-1$. Since this expansion is built upon the orthogonal properties of products of Laguerre polynomials with the Gamma distribution, the function $f_n^{(k-1)}$ is given by the sum

$$f_n^{(k-1)}(x) \equiv \sum_{i=0}^n c_i L_i^{(k-1)}(x), \quad (14)$$

where $L_i^{(k-1)}$ are the generalized Laguerre polynomials of order i and the coefficients c_i represent the coefficients of the Gamma

expansion and therefore depend on the moments of the galaxy field Λ

$$c_n \equiv \sum_{i=0}^n \binom{n}{i} \frac{\Gamma(k)}{\Gamma(k+i)} (-1)^i \frac{\langle \Lambda^i \rangle}{\theta^i}. \quad (15)$$

The main interest of the Gamma expansion with respect to the SLN is that the coefficients of the expansion are directly related to the moments of the distribution we want to model, i.e. it is not necessary to solve a complicated non-linear system of equations nor to perform a Likelihood estimation of the coefficients. Moreover, it can be easily performed at higher order to describe as best as possible the underlying probability density function of galaxies.

Another advantage of describing the galaxy field Λ by a Gamma expansion probability density function is that the corresponding observed P_N can be expressed analytically, which is not the case for the SLN which must be integrated numerically.

In Appendix (B) we demonstrate the previous statement, it follows that the CPDF P_N can be calculated from

$$P_N = \frac{(-\theta)^N}{N!} \sum_{i=0}^n c_i \frac{\Gamma(i+k)}{\Gamma(k)} h_i^{(N)}(\theta), \quad (16)$$

where $h_i \equiv \frac{1}{i!} \frac{\theta^i}{(1+\theta)^{i+k}}$ and in this case we use the notation $h_i^{(N)} \equiv \frac{d^N h_i}{dy^N}$. The successive derivatives of h_i can be obtained from the recursive relation

$$h_i^{(N)}(\theta) = \frac{(i)_f^N}{\theta^N} h_i(\theta) - \sum_{m=1}^N \binom{N}{m} \frac{(i+k)_f^m}{(1+\theta)^m} h_i^{(N-m)}(\theta).$$

In addition to the fact that having the possibility of computing the corresponding observed P_N without requiring an infinite integral for each number N is computationally more efficient, it is also practical to have the analytical calculation for some peculiar values of the k parameter of the distribution. In fact, when k is lower than 1 which occurs on small scales ($4h^{-1}$ Mpc), the probability density function goes to infinity when Λ goes to 0 (although the distribution is still well defined). In particular, this numerical divergence would induce large numerical uncertainties in the computation of the void probability P_0 . In addition, one can see that for the void probability we have the simple relation

$$P_0 = \sum_{i=0}^n c_i \frac{\Gamma(k+i)}{\Gamma(k)} h_i(\theta), \quad (17)$$

which can be used to recover the true void probability in VIPERS.

4. Application of the methods on a synthetic galaxy distribution

In this section we analyse a suite of synthetic galaxy distributions generated from 20 realizations of a Gaussian stochastic field. The full process involved in generating these benchmark catalogues is detailed in Appendix C. Each comoving volume has a cubical geometry of size $500h^{-1}$ Mpc. We generate the galaxies by discretizing the density field according to the sampling conditional probability $P[N|\Lambda]$ which we assume to be a Poisson distribution with mean Λ . In this way we know the true

underlying galaxy density contrast δ . We can therefore perform a fair comparison between the methods introduced in §3.

In order to avoid the effect of the grid ($0.95h^{-1}\text{Mpc}$) we smooth both the density field and the discrete field using a spherical Top-Hat filter of radius $R = 8h^{-1}\text{Mpc}$. We apply the three methods mentioned in §3 and compare the reconstructed probability density function to the expected one obtained directly from the density field δ .

The discrete distribution of points contains an average number of object per cell $\bar{N} = 8$ which is the one expected according to our sampling process. The corresponding P_N is given by the black histogram in the lower panel of Fig. (2), from this measurement we apply the three methods R-L, SLN and Γ_e and obtain an estimation of the probability density function corresponding to each method. In the upper panel of Fig. (2) we compare the performance of the three methods in recovering the true probability density function (black histogram referred to as reference in the inset). Note that, for this test case, we use a Gamma expansion at order 4 in order to be coherent with the order of the expansion of the Skewed Log-Normal. We have also represented the probability density function estimated when neglecting the shot noise (red dotted line), which is used as the initial guess in the case of the R-L method.

From the top panel of Fig. (2) we can conclude that the three methods perform reasonably well. It seems that the Γ_e method reproduces better the density distribution of under-dense regions ($\delta \sim -1$) but this is expected in the sense that the distribution used to generate the synthetic catalogues is a Gamma distribution (see Appendix C). Although, it is not obvious because the scale on which the density field has been set up is one order of magnitude smaller than the scale of the reconstruction $R = 8h^{-1}\text{Mpc}$.

The performance of the three methods is also represented in the bottom panel of Fig. (2), in which we compare the expected observed P_N from each method to the true one. One can see that they all agree at the 15% level, hence it is not possible to conclude that one is better than another. This was actually expected, from the comparison on the underlying density field (Fig. 2). On the contrary if one of the methods would not agree with the PDF then we would expect also a disagreement on the observed CPDF (see §6).

In the following part we investigate the sensitivity of the three methods with respect to the shot noise. In fact, as shown in Fig. (1), in most of the sub-samples of VIPERS PDR-1 we will work with a high shot noise level ($\bar{N} \leq 1$). We therefore randomly under-sample the fake galaxy distribution by keeping only 10% of the total number of object contained in each comoving volume. This process gives an average number per cell of 0.8, which is more representative in the context of the application of the reconstruction method. We perform the same comparison as in the ideal case ($\bar{N} \simeq 8$) and found that the R-L method appears to be highly sensitive to the shot noise. In fact if the mean number of object per cell is too few then the output of the method depends too much on the initial guess. It follows that, if it is too far from the true PDF the process does not converge (see top panel of Fig. 3) and the corresponding P_N does not match the observed one (see bottom panel of Fig. 3). Note that we explicitly checked this effect by increasing the number of iterations from 30 to 200. While in the case of both, the SLN and the Gamma expansion, one can see in Fig. (3), the output probability density function is in agreement (with a larger scatter) with the one obtained in the $\bar{N} \simeq 8$ case. This means that the sensitivity regarding to the shot noise is much smaller when considering parametric methods.

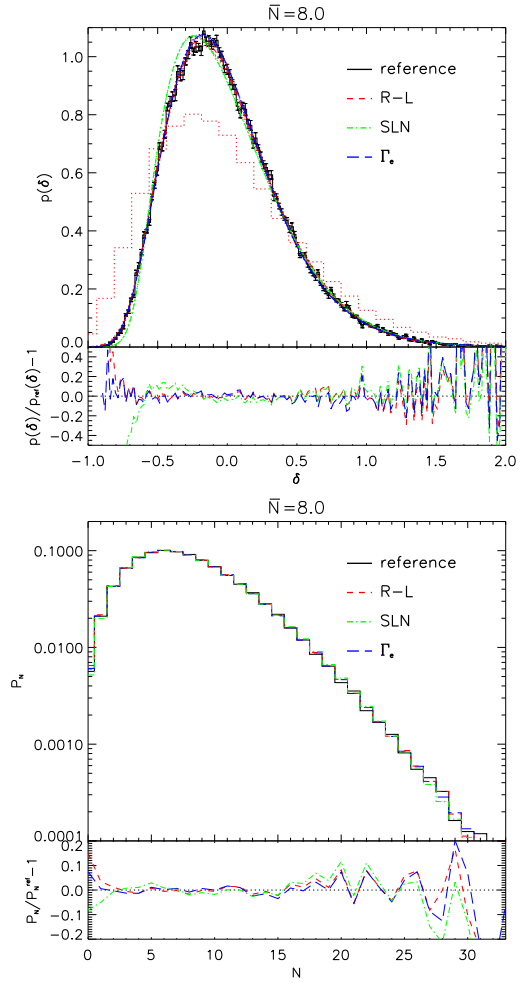


Fig. 2. *Upper:* The black histogram with error bars shows the true underlying probability density function (referred to as reference in the inset) compared to the reconstruction obtained with the R-L (red dashed line), the SLN (green dot-dashed line), and the Γ_e (blue long dashed line) methods. The red dotted histogram shows the PDF used as the initial guess for the R-L method and the colored dotted lines around each method line represent the dispersion of the reconstruction among the 20 fake galaxy catalogues. We also display the relative difference of the result obtained from each method with respect to the true PDF. *Lower:* The black histogram with error bars shows the observed probability density function (referred to as reference in the inset) compared to the reconstruction obtained with the R-L (red dashed line), the SLN (green dot-dashed line), and the Γ_e (blue long dashed line) methods. We also display the relative difference of the result obtained from each method with respect to the observed P_N .

Considering the sensitivity of the R-L method to the initial guess, knowing that the average number of galaxies per cell can be lower than unity and finally taking into account computational time, we shall continue our analysis only using the two parametric methods SLN and Γ_e . In the following, we will compare them using more realistic mock catalogues but for which we don't know a priori the true underlying PDF.

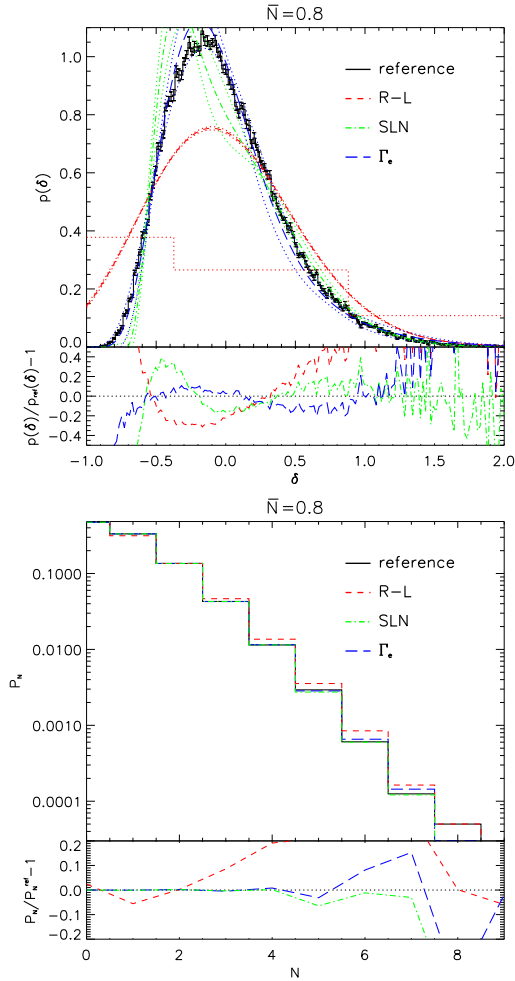


Fig. 3. Same as in Fig. (2) but we use only 10% of the galaxies contained in the fake galaxy catalogues as a result the average number of galaxy per cell drops from $\bar{N} = 8$ to $\bar{N} = 0.8$.

5. Performances in realistic conditions

In this section we discuss how observational effects have been accounted for in our analysis and test the robustness of the reconstruction methods SLN and Gamma expansion. For this purpose we use a suite of mock catalogues created from the Millenium simulation, they are also used in the analysis performed by di Porto et al. (2014).

We shall compare the reconstruction methods between two catalogues, namely REFERENCE and MOCK. The reference is a galaxy catalogue obtained from semi-analytical models. We simulate the redshift errors of VIPERS PDR-1 by perturbing the redshift (including distortions due to peculiar motions) with a Normally distributed error with rms $0.00047(1+z)$. Each MOCK catalogue is built from the corresponding REFERENCE catalogue by applying the same observational strategy (de la Torre et al. 2013) which is applied on VIPERS PDR-1; spectroscopic targets are selected from the REFERENCE catalogue by applying the slit-positioning algorithm (SPOC, Bottini et al. 2005) with the same setting as for the PDR-1. This allows us to reproduce the VIPERS footprint on the sky, the small-scale angular incompleteness due to spectra collisions and the variation of target sampling rate across the fields. Finally, we deplete each quadrant to reproduce the effect of the survey success rate (SSR, see de la Torre et al. 2013). In

Table 2. List of the magnitude selected objects (in B-band) in the mock catalogues

z_{min}	z_{max}	luminosity $M_B - 5 \log(h) <$
0.5	0.7	$-18.42 - z$
0.5	0.7	$-19.12 - z$
0.5	0.7	$-19.72 - z$
0.7	0.9	$-19.12 - z$
0.7	0.9	$-19.72 - z$
0.9	1.1	$-19.72 - z$

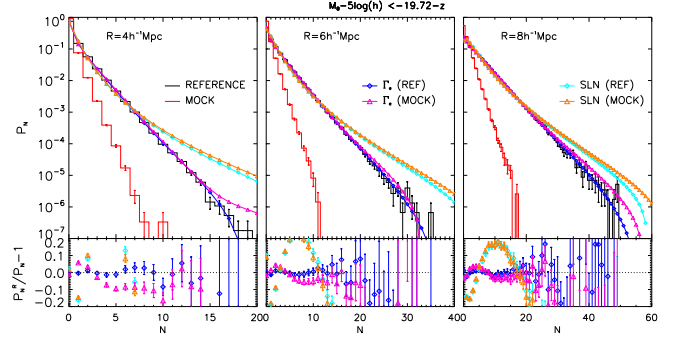


Fig. 4. Comparison between the SLN and Γ_e methods at $0.9 < z < 1.1$. Each panel corresponds to a cell radius R of 4, 6 and $8h^{-1}$ Mpc from the left to the right. *Top*: The red histogram shows the observed PDF in the MOCK catalogues while the black histogram displays the PDF extracted from the REFERENCE catalogues. The blue diamonds with lines and the magenta triangles show, respectively, the Γ_e expansion performed in the REFERENCE and MOCK catalogues. On the other hand, the cyan diamonds with lines and the orange triangles show, respectively, the SLN expansion performed in the REFERENCE and MOCK catalogues. *Bottom*: Relative deviation of the Γ_e and SLN expansions applied both on the REFERENCE and MOCK catalogues with respect to the PDF of the REFERENCE catalogues.

this way, we end up with 50 realistic mock catalogues (named MOCK hereafter), which simulate the detailed survey completeness function and observational biases of VIPERS in the W1 and W4 fields.

In order to perform a similar analysis as the one we aim at doing for VIPERS PDR-1, we construct sub-samples of galaxies selected according to their absolute magnitude M_B in B-band; we take all objects brighter than a given luminosity. We list those samples in Tab. (5), we have in total 6 galaxy samples. The highest luminosity cut ($M_B - 5 \log(h) < -19.72 - z$) allows us to follow a single population of galaxies at three cosmic epochs.

In Fig. 4, 5 and 6 we show the reconstruction performances for the SLN and the Γ_e method. We consider the same population ($M_b - 5 \log h + z < -19.72$) but in three redshift bins, $0.9 < z < 1.1$, $0.7 < z < 0.9$ and $0.5 < z < 0.7$. In order to test the stability of the methods we perform the reconstruction at three smoothing scales, $R = 4, 6$ and $8h^{-1}$ Mpc. The comparison is done as follows, on one hand we estimate the true P_N from the REFERENCE catalogue (before applying the observational selection) and we perform the reconstruction on it, in this way we can test the intrinsic biases due to the assumed parametric method (SLN or Γ_e). On the other hand, we estimate the observed P_M in the MOCK catalogues, from which we perform

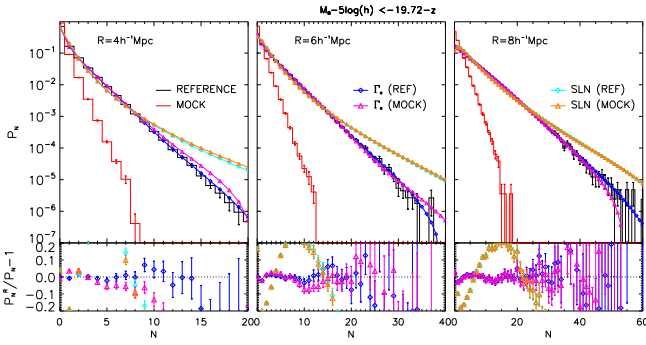


Fig. 5. Comparison between the SLN and Γ_e methods at $0.7 < z < 0.9$. Each panel corresponds to a cell radius R of 4 , 6 and $8h^{-1}\text{Mpc}$ from the left to the right.

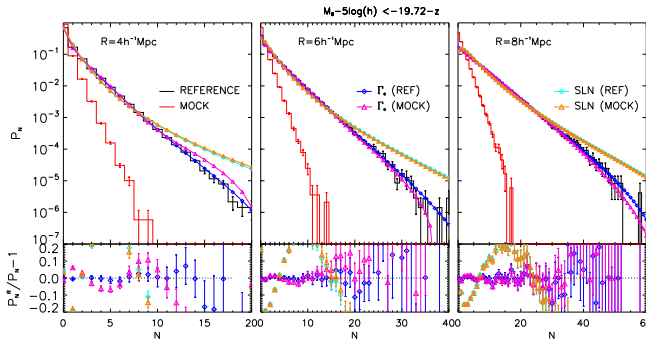


Fig. 6. Comparison between the SLN and Γ_e methods at $0.5 < z < 0.7$. Each panel corresponds to a cell radius R of 4 , 6 and $8h^{-1}\text{Mpc}$ from the left to the right.

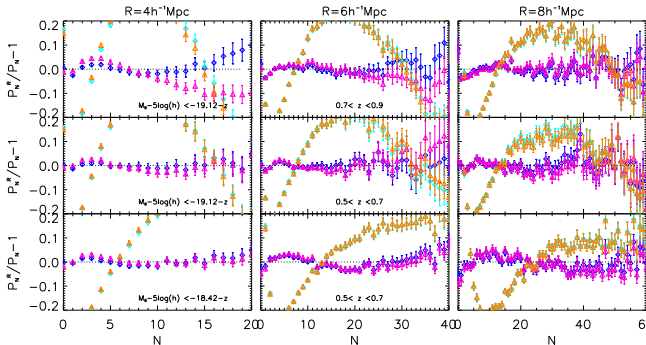


Fig. 7. Comparison between the SLN and Γ_e methods. Each column corresponds to a cell radius R of 4 , 6 and $8h^{-1}\text{Mpc}$ from the left to the right, and each row corresponds to a combination of redshift and magnitude cut.

the reconstruction to verify if we recover the expected P_N from the REFERENCE catalogue.

Inspecting Fig. (4) we can first see that the intrinsic error due to the specific modeling of the methods is much larger for the SLN (cyan diamonds compared to the black histogram) than for the Γ_e (magenta diamonds compared to the black histogram). From the top panels we see that the SLN does not reproduce the tail of the CPDF and from the bottom panel we see that even for low counts it is showing deviations as large as 20%. This intrinsic limitation is propagating when performing the reconstruction on the MOCK catalogue (orange triangles compared to the black histogram) while for the Γ_e we see that the agreement is better than 10% (magenta triangles compared to the black his-

toграм) in the low count regime and the tail is fairly well reproduced. In the second place, comparing the Γ_e performed on the REFERENCE and the MOCK catalogues (blue diamonds with respect to magenta triangles) one can see the loss of information due to the observational strategy has at most an impact of 10% on the reconstructed CPDF which reduces when considering larger cells (less shot noise).

In general, examination of Fig. (5) and (6) confirms that for the considered galaxy population the same results hold at lower redshifts. However, in particular the reconstruction at $R = 4h^{-1}\text{Mpc}$ can exhibit deviations larger than 20%, this is at odds with the fact that the shot noise contribution is expected to be the same for the three redshift bins (magnitude limited). We attribute this larger instability to the fact that not only the shot noise contribution is higher for $R = 4h^{-1}\text{Mpc}$ but also the volume probed is smaller when decreasing the redshift.

The performances of the reconstruction for the last three galaxy samples are shown in Fig.(7) where each row corresponds to a galaxy sample (we only show the residual with respect to the REFERENCE). This last comparison allows to say that the reconstruction instability at $4h^{-1}\text{Mpc}$ was indeed due to the high level of shot noise. We can conclude that in the HOD galaxy mock catalogues, the galaxy distribution is more likely to be modelled by a Γ_e instead of an SLN. Finally, for a chosen reconstruction method, the information contained in the MOCK catalogues is enough to be able to reconstruct the CPDF of the REFERENCE catalogue to better than 10%.

6. VIPERS PDR-1 data

In this section we apply the reconstruction method to the VIPERS PDR-1. We saw in the previous sections that the SLN and Γ_e methods are sensitive to the assumption we make about the underlying PDF. In fact, we saw in §4 that if the underlying PDF is close to the chosen model then the reconstruction works. We found in §5 that the galaxy distribution arising from semi-analytic models is better described by a Γ_e than an SLN distribution. However, in the following we will not take for granted that the same property holds for the galaxies in the PDR-1.

We want to choose which one of the two distributions (Log-Normal or Gamma) best describes the observed galaxy distribution in VIPERS PDR-1, when no expansion is applied. Thus, we compare the observed PDF to the one expected from the Poisson sampling of the Log-Normal probability density function (PS-LN) and to the one expected from the Poisson sampling of the Gamma distribution (the so-called Negative Binomial). Error bars are obtained by performing a Jack-knife resampling of 3×7 subregions in each fields W1 and W4.

The SP-LN distribution does not have an analytic expression and must be obtained by numerically integrating Eq. (6) while the Poisson sampling of the Gamma distribution leads to the Negative Binomial distribution defined as

$$P_N = \frac{\theta^N r(r+1)\dots(r+N-1)}{N! (1+\theta)^{N+r}}, \quad (18)$$

where $\theta = \frac{\bar{N}}{r}$ and $r = \frac{\bar{N}^2}{\sigma_N^2 - \bar{N}}$ to ensure that the first two moments of the Negative Binomial match those of the observed distribution. We show in Fig. (8) the outcome of this comparison, it follows that the Negative Binomial is much closer to the observed PDF than the PS-LN. As a result, the underlying galaxy distribution is more likely to be described by a Gamma distribution than by a Log-Normal. Hence, we only use the Gamma expansion to model the galaxy distribution of VIPERS PDR-1.

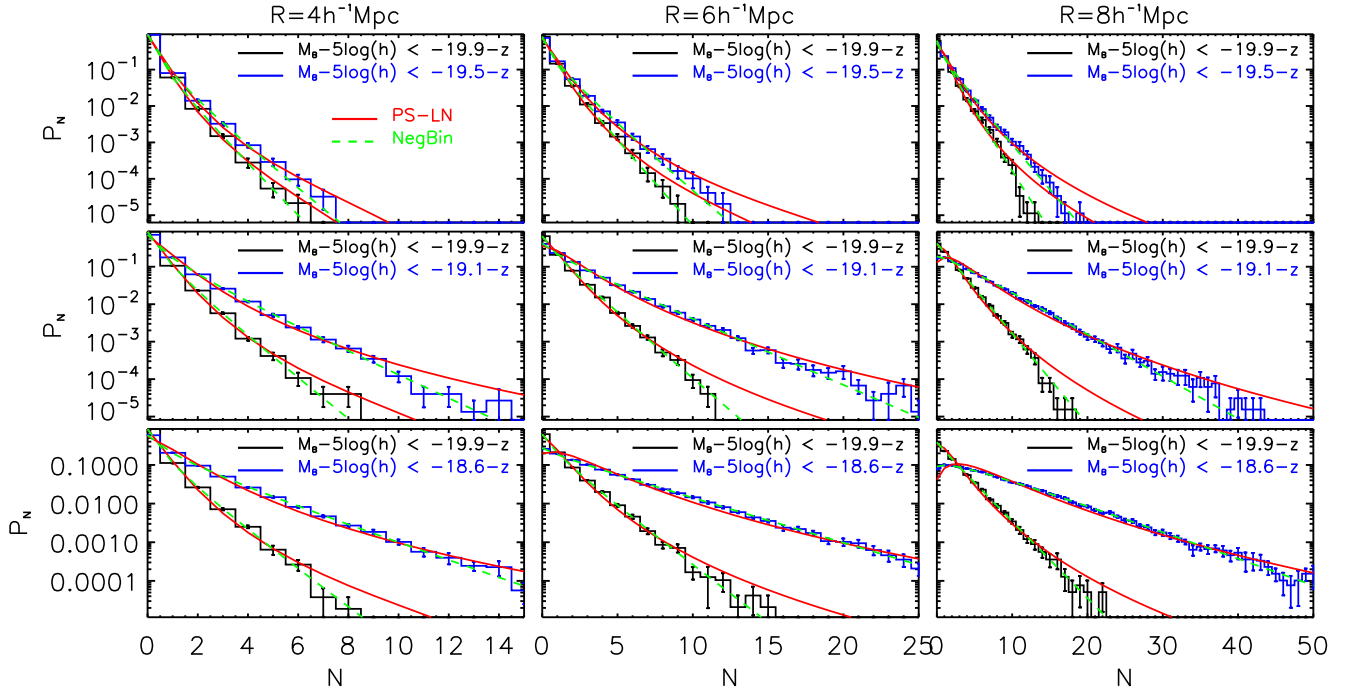


Fig. 8. Observed count-in-cell probability distribution function P_N (histograms) from VIPERS PDR-1 for various luminosity cuts (indicated in the inset). Each row corresponds to a redshift bin, from the bottom to the top $0.5 < z < 0.7$, $0.7 < z < 0.9$, and $0.9 < z < 1.1$. Each column corresponds to a cell radius $R = 4, 6, 8h^{-1}\text{Mpc}$ from the left to the right. Moreover we added the expected PDF from two models which match the two first moment of the observed distribution; the red solid line shows the prediction for a Poisson sampled Log-Normal (PS-LN) CPDF while the green dashed line displays the Negative Binomial model for the CPDF.

Moreover, the use of the Gamma expansion instead of the SLN simplifies substantially the analysis. In Fig. (9) we provide the reconstructed probability distribution function of VIPERS PDR-1 together with the corresponding underlying probability density function for each redshift bin and luminosity cut. Each panel of Fig. (9) shows how the choice of a particular class of tracers (selected according to their absolute magnitude in B-band) influence the PDF of galaxies. When measuring specific properties of the intrinsic galaxy distribution for each luminosity cut, it is enough to look at the CPDF however, when comparing the distributions with each other it is necessary to take care about the averaged number of objects per cell which varies from sample to sample. As a result it appears more useful to compare the properties of the different galaxy samples using their underlying probability density function which, assuming Poisson sampling, is free from sampling rate variation between different type of tracers.

For the two first redshift bins, we can see that the probability density function is broadening when selecting more luminous galaxies, this goes in the direction of increasing the linear bias with respect to the matter distribution. However, despite a less significant trend, for the highest redshift bin it seems that it goes in the opposite direction. This trend might be an artifact; indeed by analyzing Fig. (1) we see that for all these samples the averaged number of object per cell is between 0.2-0.4 which shows that these samples could be highly affected by shot noise effects. As a result, specific care should be taken when interpreting those three high redshift samples.

In the following we focus on the evolution of the underlying PDF for a particular class of objects on the wide redshift range probed by VIPERS PDR-1. The Fig. (10) displays the outcome of this study, it shows how the PDF, for three populations (the

three highest magnitude cuts), evolves regarding to the redshift at which it is measured. The three populations (top, middle and bottom panels) exhibit non-monotonic evolution with respect to the redshift. In particular, the more luminous population is showing that the PDF at $0.9 < z < 1.1$ appears to be systematically different than in the two lower redshift bins. However, we see also that some instabilities are appearing in the reconstruction (see wiggles at high $1 + \delta$). This might be due to the fact that we have fewer galaxies in this sample giving rise to a large shot-noise contribution ($\bar{N} < 0.3$). We indeed verified that for the high mass bin and the two other galaxy populations we vary the order of the expansion from 6 to 4 the resulting PDF changes by less than $1-\sigma$ while for the most luminous population, truncating the expansion at order 4 only removes the instability without changing significantly the overall behavior of the PDF. This consistency test shows that the radical change in the measured PDF for the highest redshift bin appears to be the true feature. Probably only the final VIPERS data set will be able to give a robust conclusion.

Finally, in Tab. (6), we list the relevant coefficients of the Gamma expansion which we measured from the VIPERS PDR-1 at the scale $R = 6h^{-1}\text{Mpc}$. They can be used in order to model both the CPDF (Eq. 16) and the PDF (Eq. 13).

7. Summary

The main goal of the present paper is to measure the probability of finding N galaxies falling into a spherical cell randomly placed inside a sparse sampled (i.e. with masked areas or with low sampling rate) spectroscopic survey. Our general approach to this problem has been to use the underlying probability density distribution of the density contrast of galaxies in order to

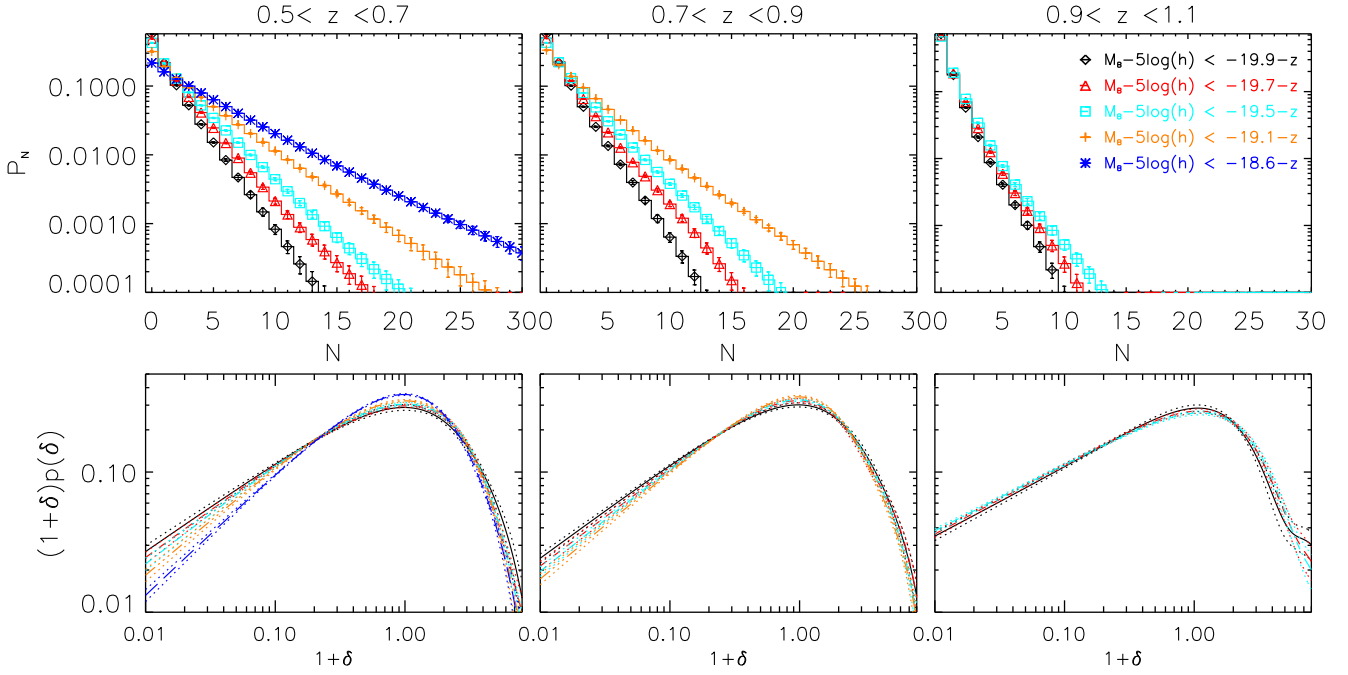


Fig. 9. *Top:* Reconstructed PDF applying the Γ_e method in three redshift bins (from left to right) at the intermediate smoothing scale $R = 6h^{-1}\text{Mpc}$. *Bottom:* Underlying PDF corresponding to the CPDF in the top panel, for each luminosity cut the 1-sigma uncertainty is represented by the dotted lines.

Table 3. Coefficients of the Γ_e expansion which describe the VIPERS PDR-1 data for $R = 6h^{-1}\text{Mpc}$

z	$M_B - 5 \log(h)$	k	θ	c_3	c_4	c_5	c_6
0.5 – 0.7	$-18.6 - z$	0.87961819	4.5053822	-0.027583435	-0.030026522	-0.018218867	-0.019292756
	$-19.1 - z$	0.78883961	3.2677238	-0.011759548	-0.0041201299	0.0076149367	-0.0010233871
	$-19.5 - z$	0.72531432	2.2643581	-0.020667396	0.00070338969	0.021056193	-0.00061403852
	$-19.7 - z$	0.64267892	1.4068744	-0.034276861	-0.022797814	0.022229339	0.023963984
	$-19.9 - z$	0.64267892	1.4068744	-0.0071341640	-0.0072444524	-0.0030038079	-0.045733910
0.7 – 0.9	$-19.1 - z$	0.76911853	2.9737929	-0.063844766	-0.046627985	-0.032441385	-0.067589757
	$-19.5 - z$	0.73969794	2.0841542	-0.032831012	-0.032693436	-0.028383261	-0.064019117
	$-19.7 - z$	0.70270085	1.6638888	-0.019063352	-0.048572844	-0.061832661	-0.078445546
	$-19.9 - z$	0.67984433	1.2608492	0.013646925	-0.028325455	-0.042087256	-0.021113201
0.9 – 1.1	$-19.5 - z$	0.47473429	1.3138704	-0.10794135	-0.17074978	-0.10267837	-0.0089188521
	$-19.7 - z$	0.49470455	1.0926144	-0.075805086	-0.16739016	-0.13623398	-0.019540367
	$-19.9 - z$	0.48382041	0.90259279	-0.076620326	-0.20604275	-0.23060122	-0.14506575

recover the counting probability corrected from sparseness effects. We therefore compared three ways (R-L, SLN and Γ_e) of measuring the probability density of galaxies classified in two categories; direct and parametric. We found that when the sampling is high ($\bar{N} \approx 10$) the direct method (Rycharson-Lucy deconvolution) performs well and avoids putting any prior on the shape of the distribution. On the other hand, we saw that when the sampling is low ($\bar{N} \approx 1$) the direct method fails to converge to the true underlying distribution. We thus concluded that, in such cases, the only alternative is to use a parametric method.

We presented two parametric forms aiming at describing the galaxy density distribution, the SLN which is often used in the literature to model the matter distribution and the Γ_e . Despite the fact that the two distributions used in this paper have been already investigated in previous works, the approach we propose to estimate their parameters is completely new. Previously, fitting procedures were used in order to estimate them. Here we propose to measure directly the parameters of the distributions from the observations. The method can be applied to both dis-

tributions SLN and Γ_e and decreases considerably the computational time of the process.

Relying on simulated galaxy catalogues of VIPERS PDR1, we tested the reconstruction scheme of the counting probability (P_N) under realistic conditions in the case of the SLN and Γ_e expansions. We found, that the reconstruction depends on the choice of the model for the galaxy distribution. However, we have also shown that it is possible to test which distribution better describes the observations.

Using VIPERS PDR1, on the relevant scales investigated in this paper ($R = 4, 6, 8h^{-1}\text{Mpc}$), we found that the Γ distribution gives a better description of the observed P_N than the one provided by the Log-Normal (see Fig. 8). We therefore adopted the Γ_e parametric form in order to reconstruct the probability density functions of galaxies. From these reconstruction we studied how their PDF changes according to their absolute luminosity in B-band and we also studied their redshift evolution. We found that little evolution has been detected in the two first redshift bins while it seems that the density distribution of the galaxy field is strongly evolving in the last redshift bin.

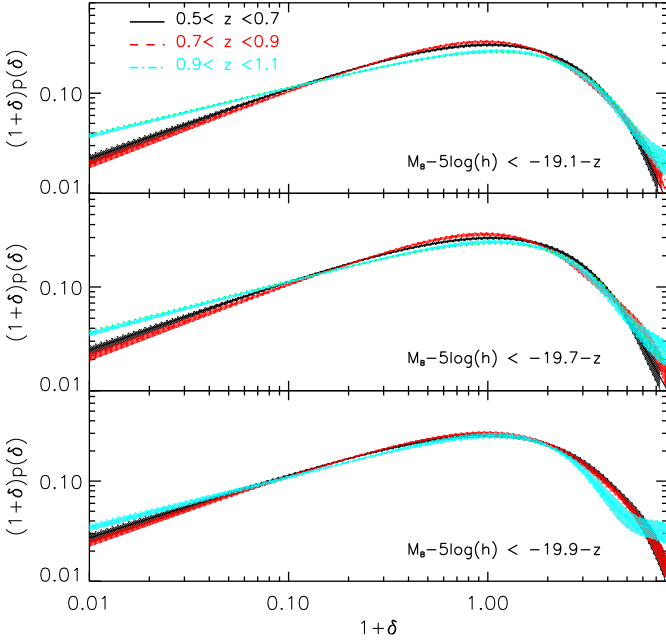


Fig. 10. Evolution of three galaxy populations selected according to their luminosity (from bottom to top). On each panel, the black solid, red dashed and, cyan dot-dashed lines represent, respectively, the three redshift bins $0.5 < z < 0.7$, $0.7 < z < 0.9$ and, $0.9 < z < 1.1$.

We finally used, the measured pdf in order to reconstruct the counting probability (CPDF) one would observe if VIPERS was not masked by gaps between the VIMOS quadrants.

Acknowledgements. JB acknowledges useful discussions with E. Gaztañaga. We acknowledge the crucial contribution of the ESO staff for the management of service observations. In particular, we are deeply grateful to M. Hilker for his constant help and support of this programme. Italian participation to VIPERS has been funded by INAF through PRIN 2008 and 2010 programmes. JB, LG and BJG acknowledge support of the European Research Council through the Darklight ERC Advanced Research Grant (# 291521). OLF acknowledges support of the European Research Council through the EARLY ERC Advanced Research Grant (# 268107). AP, KM, and JK have been supported by the National Science Centre (grants UMO-2012/07/B/ST9/04425 and UMO-2013/09/D/ST9/04030), the Polish-Swiss Astro Project (co-financed by a grant from Switzerland, through the Swiss Contribution to the enlarged European Union), the European Associated Laboratory Astrophysics Poland-France HECOLS and a Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Foreign Researchers (P11802). GDL acknowledges financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n. 202781. WJP and RT acknowledge financial support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n. 202686. WJP is also grateful for support from the UK Science and Technology Facilities Council through the grant ST/I001204/1. EB, FM and LM acknowledge the support from grants ASI-INAF I/023/12/0 and PRIN MIUR 2010-2011. CM is grateful for support from specific project funding of the *Institut Universitaire de France* and the LABEX OCEVU.

References

- Bel, J. & Marinoni, C. 2012, *MNRAS*, 424, 971
 Bel, J., et al. (the VIPERS Team) 2014, *A&A*, 563, A37
 Bernardeau, F., Colombi, S., Gaztañaga, E. & Scoccimarro, R. 2002, *PR*, 367, 1
 Bottini D., Garilli, B., Maccagni, D., et al. 2005, *PASP*, 117, 996
 Coles, P., Jones, B. 1991, *MNRAS*, 248, 1
 Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, 328, 1039
 Colombi, S. 1994, *ApJ*, 435, 536

- de la Torre, S., Guzzo, L., Kovac, K., et al. (the ZCOSMOS collaboration) 2010, *MNRAS*, 409, 867
 de la Torre, S., Guzzo, L., Peacock, J.A., et al. (VIPERS team) 2013, *A&A*, 557A, 54
 di Porto, C., et al. (VIPERS team) 2014, submitted, arXiv:1406.6692D
 Eisenstein, D. J. & Hu, W. 1998, *ApJ*, 496, 605
 Garilli, B., Le Fèvre, O., Guzzo, L., et al. (the VVDS collaboration) 2008, *A&A*, 486, 683
 Garilli, B., Paioro, L., Scodreggio, M. et al. 2012, *PASP*, 124, 1232
 Garilli, B., Guzzo, L., Scodreggio, M., et al. (the VIPERS team) 2014, *A&A*, 562, 23
 Gaztañaga, E., Fosalba, P. & Elizalde, E. 2000, *ApJ*, 539, 522
 Greiner, M., Enßlin, T. A. 2015, *A&A*, 574, 86
 Guzzo, L., Pierleoni, M., Meneux, B., et al. (the VVDS team) 2008, *Nature*, 451, 541
 Guzzo, L., Scodreggio, M., Garilli, B., et al. (the VIPERS team) 2014, *A&A*, 566, 108
 Layzer, D. 1956, *AJ*, 61, 383
 Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, *Proc. SPIE*, 4841, 1670
 Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, 439, 845
 Lilly, S. J., Le Brun, V., Maier, C., et al. (the ZCOSMOS collaboration) 2009, *ApJS*, 184, 218
 Marchetti, A., Granett, B.R., Guzzo, L., et al. (the VIPERS team) 2012, *MNRAS*, in press, arXiv:1207.4374
 Mellier, Y., Bertin, E., Hudelot, P., et al. 2008, The CFHTLS T0005 Release, <http://terapix.iap.fr/cpl/oldSite/Descart/CFHTLS-T0005-Release.pdf>
 Mustapha, H. & Dimitrakopoulos, R. 2010, *C&M*, 60, 2178
 Newman, J. A., Cooper, M.C., Davis, M., et al. (the DEEP2 collaboration) 2012, arXiv:1203.3192
 Oke, J. B. & Gunn, J. E. 1983, *ApJ*, 266, 713
 Scodreggio M., Franzetti P., Garilli B., et al. 2009, *Msngr*, 135, 13
 Szapudi, I. & Pan, J. 2004, *ApJ*, 602, 26

Appendix A: Non-linear system

The problem of this system of equations is that it is non-linear, it is therefore difficult to solve however it can be reduced to a one dimensional equation which can be solved numerically.

The two first equations ($n = 1$ and $n = 2$) can be used to express the two first cumulants with respect to the third and fourth order ones

$$\sigma_{\Phi}^2 = \ln(A_2) + \ln\left(\frac{B_1^2}{B_2}\right) \quad (\text{A.1})$$

$$\mu_{\Phi} = -\frac{1}{2} \left[\ln(A_2) + \ln\left(\frac{B_1^4}{B_2}\right) \right] \quad (\text{A.2})$$

where B_1 and B_2 are both functions of x and y . Then using other combinations of equation one can express a system of two equations for x and y alone

$$B_3^2 = a_1 B_1^2 B_4 \quad (\text{A.3})$$

$$B_3 B_1^3 = a_2 B_2^3, \quad (\text{A.4})$$

where $a_1 \equiv \frac{A_3^2}{A_4}$ and $a_2 \equiv \frac{A_3}{A_2^2}$. In order to solve properly the system we prefer to express it in term of one parameter $\eta \equiv B_2/B_1$, moreover one can see that polynomials B_1 to B_4 are not independent, as a result

$$B_4 = d + aB_1 + bB_2 + cB_3,$$

where $a = 96, b = -32, c = \frac{224}{27}, d = -\frac{1925}{27}$ and which can be substituted in Eq. (A.3). Combining Eq. (A.3) and Eq. (A.4) one obtains a parametric equation for B_1

$$(a + b\eta)B_1^3 + (d + c f(\eta))B_1^2 - g(\eta) = 0, \quad (\text{A.5})$$

which can be solved for each value of the parameter η and an independent parametric equation for B_3

$$B_3 = f(\eta).$$

As a result we can find a couple B_1, B_3 for each value of the parameter η , it follows that one can express x and y with respect to η and given the definition of η the possible solution x and y must satisfy the condition

$$B_2[x(\eta), y(\eta)] - \eta B_1[x(\eta), y(\eta)] = 0,$$

which gives the possible values of η from which one can recover x and y . Finally, from Eq. (A.1) and Eq. (A.2) we can compute the values of σ_Φ and μ_Φ corresponding to each couple (x, y) of solutions. This allows us to select the solution which provides a value of A_5 closer to the observed one.

Once the values of the cumulants $\mu_\Phi, \sigma_\Phi^2, \langle \Phi^3 \rangle_c$ and $\langle \Phi^4 \rangle_c$ are known from the process detailed above, we know that the moments of the corresponding P_N^h will match those of the observed on up to order 4. At the end, one can check whether the SLN distribution provides a good match to data by integrating numerically the probability density function convolved with the Poisson kernel K (see Eq. 5).

Appendix B: Generating function

We show that the CPDF associated to a Gamma expanded PDF can be calculated analytically from an expression which depends explicitly on the coefficients c_i of the Gamma expansion.

Be \mathcal{G}_N the generating function associated to the probability distribution P_N , it is defined as

$$\mathcal{G}_N(\lambda) \equiv \sum_{i=0}^{\infty} \lambda^i P_N. \quad (\text{B.1})$$

In case of the Poisson sampling of a Gamma distribution, after some algebra, one can show that it can be expressed with respect to the coefficients of the Gamma expansion as

$$\mathcal{G}_N(\lambda) = \frac{1}{\Gamma(k)} \sum_{i=0}^n c_i F_i(\gamma), \quad (\text{B.2})$$

where $\gamma \equiv (1 - \lambda)\theta$ and

$$F_i(\gamma) \equiv \int_0^{\infty} x^{k-1} e^{-x} L_i^{(k-1)}(x) e^{-\gamma x} dx.$$

Nevertheless, this integral can be computed using the Laguerre expansion of the exponential

$$e^{-\gamma x} = \sum_{i=0}^{\infty} \frac{\gamma^i}{(1 + \gamma)^{i+\alpha+1}} L_i^{(\alpha)}(x),$$

it reads to

$$F_i(\gamma) = \frac{\gamma^i}{(1 + \gamma)^{i+k}} \frac{\Gamma(i+k)}{i!}. \quad (\text{B.3})$$

The formal expression of the generating function is therefore given by

$$\mathcal{G}_N(\lambda) = \frac{(1 + \gamma)^{-k}}{\Gamma(k)} \sum_{i=0}^n c_i \frac{\Gamma(i+k)}{i!} \left(\frac{\gamma}{1 + \gamma} \right), \quad (\text{B.4})$$

where we still use $\gamma = (1 - \lambda)\theta$. From the explicit expression of the moment generating function (Eq. B.4) one can get the probability distribution P_N by iteratively deriving the generating function with respect to γ

$$P_N \equiv \frac{1}{N!} \left. \frac{d^N \mathcal{G}_N(\lambda)}{d\lambda^N} \right|_{\lambda=0} = \frac{(-\theta)^N}{N!} \left. \frac{d^N \mathcal{G}_N(\gamma)}{d\gamma^N} \right|_{\gamma=\theta}.$$

These derivatives can be calculated explicitly.

Appendix C: Synthetic galaxy catalogues

In this Appendix we describe how we generate synthetic galaxy catalogues from Gaussian realizations. The first requirement of these catalogues is that they must be characterized by a known power spectrum and 1-point probability distribution function. The second requirement is that the probability distribution function must be measurable.

The basic idea is simple, we generate a Gaussian random field in Fourier space (assuming a power spectrum), we inverse Fourier transform it to get its analog in configuration space. We further apply a local transform in order to map the Gaussian field into a stochastic field characterized by the target PDF. The two crucial step of this process are the choice of the input power spectrum and the choice of the local transform.

Be ν a stochastic field following a centered ($\langle \nu \rangle = 0$) reduced ($\sigma_\nu^2 \equiv \langle \nu^2 \rangle_c = 1$) Gaussian distribution. From a realization of this field, one can generate a non-Gaussian density field δ by applying a local mapping L between the two, hence

$$\delta = L(\nu). \quad (\text{C.1})$$

The local transform L must be chosen in order to match some target PDF P_δ for the density contrast δ . Assuming that the local transform is a monotonic function which maps the ensemble $]-\infty, +\infty[$ into $]-1, +\infty[$ then, due to the probability conservation $P_\delta(\delta)d\delta = P_\nu(\nu)d\nu$, the local transform must verify the following matching

$$C_\delta[\delta] = C_\nu[\nu], \quad (\text{C.2})$$

where C_x stands for the cumulative probability distribution function. Be $[a, b]$ the definition assemble of the variable x then its cumulative probability distribution function is defined as $C_x[x] \equiv \int_a^x P_x(x') dx'$, where P_x is the PDF of x . By definition a probability density function is positive, it follows that its cumulative is a monotonic function and therefore Eq. (C.2) can always be inverted, it reads

$$\delta = C_\delta^{-1} [C_\nu(\nu)],$$

where the exponent -1 stands for the reciprocal function such that $F^{-1} [F(x)] = x$. For example, by definition the local mapping L which allows transform a Normal distribution into a Log-Normal distribution is $\delta = e^\nu - 1$. Note that depending on the PDF that must be matched this inversion can require a numerical evaluation which can be tabulated.

Once a local transform is chosen, we need to adress the question of finding the appropriate power spectrum of the Gaussian field ν which, once locally mapped into the density field δ , will match the expected power spectrum. Following Greiner & Enßlin (2015), who considered a log-transform we generalized their result to a generic local transformation. This

mapping is not direct in Fourier space while it is in configuration. Writing the two point moment of order two of the density field δ and assuming the probability conservation leads to

$$\xi_\delta \equiv \langle \delta_1 \delta_2 \rangle = \int L(v_1)L(v_2)\mathcal{B}(v_1, v_2, \xi_v)dv_1dv_2, \quad (\text{C.3})$$

where \mathcal{B} is a bivariate Gaussian defined as

$$\mathcal{B}(v_1, v_2, \xi_v) \equiv \frac{1}{2\pi|C_v|^{1/2}} \exp\left\{-\frac{1}{2}v^T C_v^{-1}v\right\}. \quad (\text{C.4})$$

One can notice that in our case (central reduced Gaussian) the covariance matrix C_v takes the simple form $C_v = \begin{bmatrix} 1 & \xi_v \\ \xi_v & 1 \end{bmatrix}$. Once integrated over the definition domain of v_1 and v_2 , Eq. (C.3) provides a mapping between the 2-point correlation function of the Gaussian field v and the 2-point correlation function of the density field δ . However, we prefer to rotate the coordinate system before performing the integral (C.3) because in case of high correlation (~ 1) then the gaussian will be comparable with a straight line; most of the sampling of this function will be useless. That's why we look for the rotation allowing to diagonalize the matrix C_v and therefore convert v into a new variable \mathbf{x} . It follows that

$$C_x = \begin{bmatrix} 1 - \xi_v & 0 \\ 0 & 1 + \xi_v \end{bmatrix}$$

and the integral becomes

$$\xi_\delta = \frac{1}{2\pi\sqrt{1-\xi_v^2}} \int L\left(\frac{x_2 - x_1}{2}\right)L\left(\frac{x_2 + x_1}{2}\right)e^{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2}\right)} dx_1 dx_2, \quad (\text{C.5})$$

where $\sigma_1^2 = 1 - \xi_v$ and $\sigma_2^2 = 1 + \xi_v$ we can therefore integrate over a bounded domain corresponding to the $-8\sigma_1, 8\sigma_1$ along the x_1 axis and $-8\sigma_2, 8\sigma_2$ along the x_2 axis. An other possibility to perform the integral C.3 is to use the Mehler's formula, doing so, one can show that the 2-point correlation of the density field can be expressed as a Taylor expansion on the 2-point correlation function of the v field. It reads,

$$\xi_\delta = \lambda(\xi_v) \equiv \sum_{n=0}^{\infty} n! c_n^2 \xi_v^n, \quad (\text{C.6})$$

where the c_n are the coefficients of the Hermit transform of the local mapping $L(v) = \sum_{n=0}^{\infty} c_n H_n(v)$ and they can be calculated using the orthogonal properties of Hermit polynomials

$$c_n = \frac{1}{n!} \int_{-\infty}^{+\infty} L(v)H_n(v)P_v(v)dv. \quad (\text{C.7})$$

The latter approach considerably speed up the numerical evaluation of Eq. (C.5), it allows to compute the 2-D integral as a finite sum of 1-D integrals. It also allows to verify that when the 2-point function of the field v is positive then the derivative of ξ_δ with respect to ξ_v is positive. Moreover, from Eq.(C.3) one can see that $\xi_v = 0$ implies $\xi_\delta = 0$. This means that the function which transforms ξ_v into ξ_δ is invertible as long as ξ_δ is positive. On the other hand we know that the zero-crossing of the 2-point correlation function occurs at very large scales at which one can safely assume that $|\xi_\delta| \ll 1$ thus by continuity one can truncate the Eq.(C.6) at order one providing a linear relation between ξ_δ and ξ_v . As a result, one can take the reciprocal of the function λ such that $\xi_v = \lambda^{-1}(\xi_\delta)$.

Once the local transform L and the 2-point correlation mapping λ are known, then the input power spectrum of the Gaussian field v can be obtained as follow. We choose a power spectrum $P(k)$, in the present case Eisenstein & Hu (1998), for the density field δ , we calculate its corresponding 2-point correlation function

$$\xi_\delta = \int P(k)e^{ik \cdot r} d^3 k. \quad (\text{C.8})$$

At each scale r , one can deduce the 2-point correlation function of the Gaussian field $\xi_v = \lambda^{-1}(\xi_\delta)$ and finally using a Fourier transform we obtain the input power spectrum

$$P_{in}(k) = \frac{1}{(2\pi)^3} \int \xi_v(r)e^{-ik \cdot r} d^3 r. \quad (\text{C.9})$$

Finally, in order to make sure that the PDF target will be reproduced, it is necessary to verify that, once the input power spectrum $P_{in}(k)$ have been set up on regular k-space grid which will be used to generate the Gaussian field, its integral is indeed equal to the expected variance on the size of the mesh. $\hat{\sigma}_a^2 = \left(\frac{2\pi}{L}\right)^3 \sum_n P(k_n)$ should be equal to $\sigma_a^2 = \int P(k)d^3 k$. In general, σ_a and $\hat{\sigma}_a$ are not equal, thus we renormalize the target power spectrum by the quantity $S = \hat{\sigma}_a^2/\sigma_a^2$, $\hat{P}_{in}(k) = S P_{lin}(k)$.

We generate a Gaussian field (with a flat power spectrum), on a regular mesh of $a = 0.95h^{-1}\text{Mpc}$ and a comoving box of $500^3 h^{-3}\text{Mpc}^3$. We then Fourier transform with an FFT and keep only the phases of the field $v_k = e^{i\theta(k)}$. We generate at each position \mathbf{k}_n the value of the modulus of $v_k = \sqrt{X_k}e^{i\theta(k)}$, where $X_k = -\hat{P}_{in}(k) \ln(1 - \epsilon)$ and ϵ is a random number with a uniform probability distribution between 0 and 1. We then inverse Fourier transform the field to get a centered reduced Gaussian field. In Fig. (C.1) we show the input power spectrum of the Gaussian field v compared to the one measured using a FFT, and to the one expected from the local transformation applied to the v field in order to generate the density field δ .

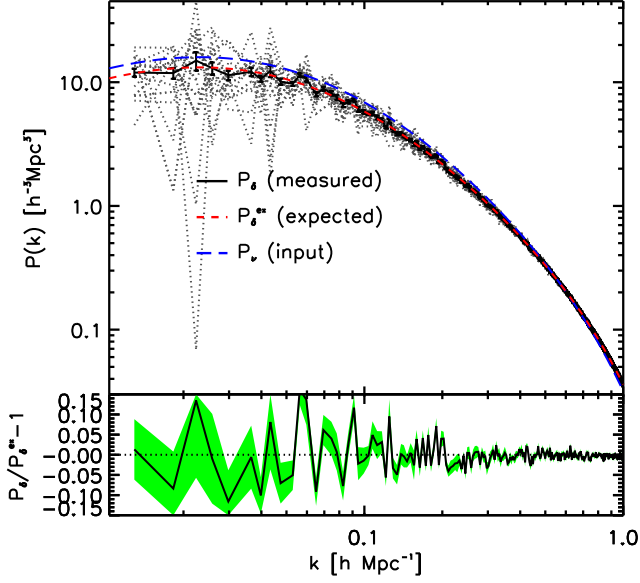


Fig. C.1. *Upper:* Grey dotted lines show the power spectrum measured in each of the 20 fake galaxy distributions, the black solid line represent their average and the errors display the dispersion of the measurements. The blue long dashed line displays the input power spectrum used too generate the Gaussian stochastic field δ and the red dashed line shows the corresponding expectation value for the power spectrum of the density contrast δ . *Lower:* Shows the deviation between the measured power spectrum of the δ -field and the expected one.

¹ Università degli Studi di Milano, via G. Celoria 16, 20130 Milano, Italy

² INAF - Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, via E. Bianchi 46, 23807 Merate, Italy

³ INAF - Istituto di Astrofisica Spaziale e Fisica Cosmica Milano, via Bassini 15, 20133 Milano, Italy

⁴ Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388, Marseille, France

⁵ INAF - Osservatorio Astronomico di Torino, 10025 Pino Torinese, Italy

⁶ Canada-France-Hawaii Telescope, 65–1238 Mamalahoa Highway, Kamuela, HI 96743, USA

⁷ Aix Marseille Université, CNRS, CPT, UMR 7332, 13288 Marseille, France

⁸ Université de Lyon, F-69003 Lyon, France

⁹ INAF - Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127, Bologna, Italy

¹⁰ Dipartimento di Matematica e Fisica, Università degli Studi Roma Tre, via della Vasca Navale 84, 00146 Roma, Italy

¹¹ Institute of Cosmology and Gravitation, Dennis Sciama Building, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3FX

¹² Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan

¹³ INAF - Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, 34143 Trieste, Italy

¹⁴ SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

¹⁵ Institute of Physics, Jan Kochanowski University, ul. Swietokrzyska 15, 25-406 Kielce, Poland

¹⁶ Department of Particle and Astrophysical Science, Nagoya University, Furo-cho, Chikusa-ku, 464-8602 Nagoya, Japan

¹⁷ Dipartimento di Fisica e Astronomia - Alma Mater Studiorum Università di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy

¹⁸ INFN, Sezione di Bologna, viale Berti Pichat 6/2, I-40127 Bologna, Italy

¹⁹ Institute d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France

²⁰ Max-Planck-Institut für Extraterrestrische Physik, D-84571 Garching b. München, Germany

²¹ Laboratoire Lagrange, UMR7293, Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, 06300 Nice, France