Fermi National Accelerator Laboratory

FERMILAB-Pub-92/71-A
February 1992

Probing the statistics of primordial fluctuations and its evolution

ENRIQUE GAZTAÑAGA AND JUN'ICHI YOKOYAMA[1]

NASA/Fermilab Astrophysics Center,

Fermi National Accelerator Laboratory, Batavia, IL60510-0500, USA

ABSTRACT

We propose a new measure of statistical distribution of fluctuations on various scales, namely, counts-in-cells of smoothed density field. Using volume-limited samples of galaxy redshift catalogs, it is shown that the distribution on large scales, with volume-average of the two-point correlation function of the smoothed field $\lesssim 0.05$, is consistent with Gaussian. Statistics on smaller scale are shown to agree with Fry's BBGKY hierarchical model and the negative binomial distribution remarkably well, which may suggest that our universe started with Gaussian fluctuation and evolved keeping hierarchical form.

Subject headings: cosmology – galaxies:clustering – galaxies:redshift

[1]On leave of absence from Department of Physics, Faculty of Sciences, The University of Tokyo, Tokyo 113, Japan

1 Introduction

Increasing number of possible scenarios of galaxy formation have been proposed, as high energy physics theories are considered more deeply, which should describe the early universe when the seeds of primordial fluctuations are provided. It is an important role of observational cosmology to provide statistical tests which help us to single out the correct one. Primary tests, so far used, are the amplitude and shape of the two-point correlation function of galaxies as well as the isotropy of microwave background radiation. Both of which can in principle be calculated from the amplitude and spectrum of the primordial fluctuations once the matter content and the thermal history of the universe are specified. Scenarios yet surviving today, of course, satisfy minimal requirements of these constraints. In order to go a step further, we should apply other tests, examining different statistical properties such as the statistical distribution of density fluctuations, on which two-point correlation function does not provide any information.

Although standard inflationary models predict Gaussian distribution of primordial density fluctuations, some other scenarios, such as cosmic strings (*e.g.* Zel'dovich 1980; Vilenkin 1981; Perivolaropoulos, Brandenberger and Stebbins 1990) global textures (*e.g.* Turok and Spergel 1991), or solitons (Frieman *et al.* 1989; Griest and Kolb 1989) might have different, non-Gaussian, distributions. One expects that the distribution of fluctuations at earlier epochs of the evolution should still be present and observable at large scales, where neither gravity nor any other physical interaction has had enough time to deform it. At smaller scales, gravity has changed the original distribution to a different one which we know, from the observed galaxy distribution, it is non-Gaussian. Hence it is important to examine the shape of the statistical distribution of density fluctuations on various scales.

Unfortunately, however, it is not easy to find a good measure for this purpose. It is quite likely to find apparently Gaussian distribution due to the central limit theorem even though fluctuations are intrinsically non-Gaussian. For example, Baumgart and Fry (1991) have recently investigated statistical distribution of the power spectrum at several wave numbers and found that it is well approximated by Gaussian even on small scales, where we know statistics can not be Gaussian (see below). The apparent Gaussian distribution they obtained would be a result of averaging over the direction of the wave number vector, as pointed out by Suginohara and Suto (1991). Other tests for the statistical distribution include analyses of higher order correlation functions (*e.g.* Lahav *et al.* 1991), statistics of genus numbers (*e.g.* Gott *et al.* 1989) and those of contour crossing numbers (Ryden *et al.* 1989), none of which directly measures the distribution itself (see also Scherrer and Bertschinger 1991). In this paper we propose a different statistical test by modifying counts-in-cells method, which measures the statistical distribution of fluctuations on various scales directly.

Counts in cells, *i.e.*, the probability, $P_N(V)$, to have N galaxies within a randomly chosen cell of volume V, reflects the statistics of fluctuations because it depends not only on the power spectrum or two-point correlation function, but also on arbitrary higher-order connected correlation functions or cumulants (White 1979). In particular if fluctuations obey Gaussian statistics the void probability is given by

$$P_0(V) = \exp\left[-\overline{n}V + \frac{1}{2}\overline{n}^2V^2\overline{\xi}(V)\right], \tag{1}$$

where $\overline{n}$ is the mean number density of galaxies, and $\overline{\xi}(V)$ is the volume average of two-point correlation function:

$$\overline{\xi}(V) = \int_V \frac{d^3r_1 d^3r_2}{V^2}\xi(|\mathbf{r}_1 - \mathbf{r}_2|). \tag{2}$$

One may naturally think that if primordial fluctuations had Gaussian statistics the above expression

would fit observational data at large enough volume V, where scale fluctuations are still in the linear regime, keeping the memory of initial condition (Peebles 1980 §39). In fact the above expression has been tested by Fry (1986 and 1989), who compared predictions of various models with observational data and found that the above Gaussian model was the worst even on large scales among the six different models considered. (See also Figure 7 below).

This does not necessarily imply that the primordial fluctuations were highly non-Gaussian for two reasons. One is that no matter how large the volume is, $\overline{\xi}(V)$ depends on correlation function at small scales as well where fluctuation is non-Gaussian due to nonlinear evolution. The other, more generic reason, is related to the positive semidefiniteness of number-counts or density.

In order to realize Gaussian distribution, the dispersion, $\overline{\sigma}$, must be much smaller than the average number, $\overline{N}$, to avoid negative counts. For a continuous density field $\overline{\sigma} = \overline{N}\sqrt{\overline{\xi}}$, so that $\sigma \ll \overline{N}$ implies $\overline{\xi} \ll 1$. For discrete counts $\overline{\sigma} = \sqrt{\overline{N} + \overline{\xi}\overline{N}^2}$, so that $\overline{\sigma} \ll \overline{N}$ implies $\overline{\xi} \ll 1 - \overline{N}^{-1}$. Thus the distribution cannot be Gaussian on scales $\lesssim$ 8Mpc where observed galaxy distribution has $\overline{\xi} \gtrsim 1$ (Davis and Peebles 1983). Fry (1986) has further argued that it is impossible to obtain Gaussian distribution of number-counts in a volume containing $\overline{N}$ galaxies with $\overline{\xi} \gtrsim 1/\overline{N}$ on average. Since $\overline{N}\overline{\xi}$ is an increasing function of length scale in the interested range, we cannot make use of counts-in-cells statistics as it is in order to examine if primordial fluctuations obey Gaussian statistics.

In the present paper we consider a statistical measure which is free from the difficulties mentioned above, that is, counts-in-cells of a smoothed density field. By smoothing the data we can focus on the information on large scales, with smaller values of $\overline{\xi}$, and examine the statistics of fluctuations on various scales, so that we can not only probe the primordial statistical distribution but also trace its evolution in the presence of gravity.

The rest of the paper is organized as follows. In §2 we consider properties of discrete and smoothed distribution in relation to the original continuous density distribution function. §3 is devoted to explanation of the algorithm and discussion of the significance of our method through the analysis of artificially-generated data. In §4 we report the results of our analyses of four different galaxy catalogs both on large and small scales and their implication is discussed. Finally §5 contains the conclusion. Throughout the paper we take the Hubble parameter $H_0 = 100h$km/sec/Mpc $= 100$km/sec/Mpc. Dependence on h is readily recovered by putting h^{-1} before the unit of Mpc.

2 Discrete and smooth density fields

2.1 Effect of discreteness of galaxy distribution

As stated in the introduction, our primary purpose is to extract the statistics of original continuous density field, $\rho(\mathbf{r})$, out of the discrete distribution of galaxies. It is evident that the smoothing procedure of a point distribution alone does not recover the original distribution, since galaxy formation is considered to be a sort of random process which may depend on $\rho(\mathbf{r})$ nontrivially. Here, in order to remove uncertainty, we assume the simplest possibility that each galaxy is a faithful tracer of background density distribution and that galaxy formation is a Poisson process out of it. Possible effects of biasing (Kaiser 1984, Bardeen *et al.* 1986) will be considered elsewhere.

In the Poisson model the probability that a galaxy emerges in a volume δV at $\mathbf{r}$ may be calculated according to the Poisson statistics with local mean number $\langle \delta N(\mathbf{r}) \rangle = \rho(\mathbf{r})\delta V$. The moment generating function yields

$$\mathcal{M}_{\delta V}(t) = 1 - \rho(\mathbf{r})\delta V + \rho(\mathbf{r})\delta V e^t = \exp[(e^t - 1)\rho(\mathbf{r})\delta V]. \tag{3}$$

The generating function of moments of counts in a volume $V_R = \frac{4\pi}{3}R^3$ is found by multiplying equation (3):

$$\mathcal{M}_D(t) = \left\langle \exp\left[\int_{V_R} \rho(\mathbf{r})dV(e^t - 1)\right]\right\rangle \equiv e^{\mathcal{K}_D(t)}, \tag{4}$$

where $\overline{N}(R)$ is the mean number of galaxies found in a volume V_R and $\mathcal{K}_D(t)$ is the corresponding generating function of cumulants. It is related with the cumulant generating function of the original continuous field, $\mathcal{K}_C(t)$, as

$$\mathcal{K}_D(t) = \mathcal{K}_C(e^t - 1), \tag{5}$$

by which we find the following equalities.

$$\begin{aligned}
\overline{N}^2\overline{\kappa}_{D2} &= \overline{N}^2\overline{\kappa}_{C2} + \overline{N}, \\
\overline{N}^3\overline{\kappa}_{D3} &= \overline{N}^3\overline{\kappa}_{C3} + 3\overline{N}^2\overline{\kappa}_{C2} + \overline{N}, \\
\overline{N}^4\overline{\kappa}_{D4} &= \overline{N}^4\overline{\kappa}_{C4} + 6\overline{N}^3\overline{\kappa}_{C3} + 7\overline{N}^2\overline{\kappa}_{C2} + \overline{N}, \\
&\cdots\cdots\cdots\cdots, \\
\overline{N}^J\overline{\kappa}_{DJ} &= \overline{N}^J\overline{\kappa}_{CJ} + \frac{1}{2}J(J-1)\overline{N}^{J-1}\overline{\kappa}_{C(J-1)} + \ldots + (2^{J-1} - 1)\overline{N}^2\overline{\kappa}_{C2} + \overline{N},
\end{aligned} \tag{6}$$

where $\overline{\kappa}_{DJ}$ and $\overline{\kappa}_{CJ}$ are volume-averaged cumulants of discrete and continuous fields which are calculated through differentiating generating functions as $\mathcal{K}_D^{(J)}(t=0) = \overline{N}^J\overline{\kappa}_{DJ}$ and $\mathcal{K}_C^{(J)}(t=0) = \overline{N}^J\overline{\kappa}_{CJ}$, respectively. Thus the J-th order cumulant of observable discrete field depends lower order cumulants of the corresponding continuous field as well.

As is seen above, if the cell is so small that it contains only a small number of galaxies on average, the last term on the right hand side of each equality, the shot noise, dominates and the distribution does not reflect the background density field at all. On the other hand, we may avoid effects of shot noise by employing large enough cells with a large value of $\overline{N}$, so that cumulants of discrete counts are well approximated by those of the original continuous field. This is a merit of counts-in-cells analysis which is absent in the statistics of density at each point (Saunders *et al.* 1991). Strictly speaking, one must have

$$\overline{N}\overline{\kappa}_{CJ} > \frac{1}{2}J(J-1)\overline{\kappa}_{C(J-1)}, \tag{7}$$

in order to guarantee $\overline{N}^J\overline{\kappa}_{DJ} \cong \overline{N}^J\overline{\kappa}_{CJ}$ for all J. This inequality, however, should merely be regarded as a *sufficient* condition for the probability distribution based on $\overline{\kappa}_{DJ}$'s to coincide with the original distribution. In practice, the measured distribution of density fluctuations might only be sensitive up to a finite number of J, in particular in the linear regime, and one would like to check if inequality (7) is satisfied up to a large enough J for each particular value of $\overline{N}$. We return to this point later when we consider specific models of $\overline{\kappa}_{CJ}$.

2.2 Properties of smoothed distribution

In this paper we treat each galaxy in the catalogs not as a point but as if it has a continuous Gaussian density distribution peaked at its original point with dispersion σ. Thus, the fluctuation of the smoothed field at $\mathbf{x}$, $\widehat{\delta_\sigma n}(\mathbf{x})$, is given by

$$\begin{aligned}
\widehat{\delta_\sigma n}(\mathbf{x}) &= \int \delta n(\mathbf{x}')\frac{1}{(\sqrt{2\pi}\sigma)^3}e^{-\frac{(\mathbf{x}-\mathbf{x}')^2}{2\sigma^2}}d^3x' \\
&= \int \left[\sum_i \delta_D(\mathbf{x}' - \mathbf{x}_i) - \overline{n}\right]\frac{1}{(\sqrt{2\pi}\sigma)^3}e^{-\frac{(\mathbf{x}-\mathbf{x}')^2}{2\sigma^2}}d^3x' = \sum_i \frac{1}{(\sqrt{2\pi}\sigma)^3}e^{-\frac{(\mathbf{x}-\mathbf{x}_i)^2}{2\sigma^2}},
\end{aligned} \tag{8}$$

where δ_D is the Dirac delta function, $\mathbf{x}_i$'s are original positions of galaxies, and $\delta n(\mathbf{x}')$ represents their fluctuation. We denote quantities of smoothed field with overhat and suffix σ as above.

As is well known, physical meaning of this smoothing is prominent seen in Fourier space. Define Fourier transform of each fluctuation as

$$\delta(\mathbf{k}) = \int \delta n(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}} d^3x, \tag{9}$$

and

$$\widehat{\delta_\sigma}(\mathbf{k}) = \int \widehat{\delta_\sigma n}(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}} d^3x, \tag{10}$$

respectively. Then the latter is related with former as

$$\widehat{\delta_\sigma}(\mathbf{k}) = e^{-\frac{k^2\sigma^2}{2}} \delta(\mathbf{k}). \tag{11}$$

Thus the smoothed field contains full information on large scale with $k \lesssim \sigma^{-1}$ but small-scale information is exponentially suppressed. This means that one may concentrate on large scale which keeps memory of initial data without being contaminated by effects of nonlinear evolution. In particular, if initial spectrum of fluctuation obeys Gaussian statistics with random phase and if this holds for $k \lesssim \sigma^{-1}$ today, higher-order moments of fluctuation in a volume V, $\widehat{\delta_\sigma N}(V)$, are given by

$$\begin{aligned} \langle \widehat{\delta_\sigma N}(V)^{2s} \rangle &\approx \frac{(2s)!}{2^s s!} \left[\int_{V^2} d^3r_1 d^3r_2 \langle \widehat{\delta_\sigma n}(r_1) \widehat{\delta_\sigma n}(r_2) \rangle \right]^s \\ &= \frac{(2s)!}{2^s s!} \langle \widehat{\delta_\sigma N}(V)^2 \rangle^s \equiv \frac{(2s)!}{2^s s!} \left[\widehat{\bar{\xi}_\sigma}(V) V^2 \widehat{\overline{n_\sigma}}^2 \right]^s, \end{aligned} \tag{12}$$

and $\langle \widehat{\delta_\sigma N}(V)^{2s+1} \rangle \approx 0$, where $\widehat{\overline{n_\sigma}}$ and $\widehat{\bar{\xi}_\sigma}(V)$ are mean number density and averaged correlation function of the smoothed field, respectively. Thus the probability distribution function of $\widehat{\delta_\sigma N}(V)$ will be Gaussian,

$$P[\widehat{\delta_\sigma N}(V)] = \frac{1}{\sqrt{2\pi \langle \widehat{\delta_\sigma N}(V)^2 \rangle}} \exp\left[-\frac{\widehat{\delta_\sigma N}(V)^2}{2\langle \widehat{\delta_\sigma N}(V)^2 \rangle} \right], \tag{13}$$

if the average value, $\widehat{\overline{N_\sigma}}(V)$, is much larger than $\sqrt{\widehat{\delta_\sigma N}(V)^2}$ to ensure positivity of the distribution.

In this paper we choose the sampling cell to be a sphere with varying radius R. (See Elizalde and Gaztañaga (1992) for dependence of discrete counts on cell's shape.) Then $\langle \widehat{\delta_\sigma N}(V)^2 \rangle \equiv \langle \widehat{\delta_\sigma N}(R)^2 \rangle$ above is given by

$$\langle \widehat{\delta_\sigma N}(R)^2 \rangle = 8 \int \langle |\widehat{\delta_\sigma}(\mathbf{k})|^2 \rangle \left(\frac{\sin kR - kR\cos kR}{k^2} \right)^2 dk. \tag{14}$$

If the correlation function of the original distribution is approximated by a power law,

$$\xi(r) = \left(\frac{r_0}{r} \right)^\gamma, \tag{15}$$

then $\langle |\delta(\mathbf{k})|^2 \rangle$ is given by

$$\langle |\delta(\mathbf{k})|^2 \rangle = \frac{2(\gamma-1)\pi^2 r_0^\gamma}{\Gamma(\gamma) \sin\left[\frac{(\gamma-1)\pi}{2}\right]} k^{\gamma-3}, \tag{16}$$

(Peebles 1980) so that $\langle\widehat{\delta_\sigma N(R)^2}\rangle$ is theoretically calculated:

$$\langle\widehat{\delta_\sigma N(R)^2}\rangle = \frac{16(\gamma-1)\pi^2 r_0^\gamma}{\Gamma(\gamma)\sin\left[\frac{(\gamma-1)\pi}{2}\right]}\int k^{\gamma-7}e^{-\sigma^2k^2}(\sin kR - kR\cos kR)^2 dk, \tag{17}$$

which will provide a consistency test of our analysis below.

Somehow we have smoothed the field twice, first with a Gaussian window of size σ and them with a top hat window of size R. We do this for practical reasons and one should keep in mind that there is only one significant parameter that quantifies the smoothing. A good way to characterize the amount of smoothing is to follow the values of the smoothed dispersion $\widehat{\bar{\xi}_\sigma}$.

We can also consider how smoothing affects shot noise by calculating the cumulant generating functions of the smoothed field,

$$\widehat{\mathcal{K}_{C\sigma}}(t) = \ln\left\langle\exp\left[\int_{V_T} d^3r\rho(\mathbf{r})W_\sigma(R,\mathbf{r})t\right]\right\rangle, \tag{18}$$

and

$$\widehat{\mathcal{K}_{D\sigma}}(t) = \ln\left\langle\exp\left[\int_{V_T} d^3r\rho(\mathbf{r})\left(e^{W_\sigma(R,\mathbf{r})t}-1\right)\right]\right\rangle, \tag{19}$$

for continuous and originally discrete distributions, respectively. Here the integral is over the total volume, V_T, of the entire sample and $W_\sigma(R,\mathbf{r})$ represents integrated smoothing function defined by

$$W_\sigma(R,\mathbf{r}) \equiv 4\pi\int_0^R r'^2dr'\frac{1}{(\sqrt{2\pi}\sigma)^3}\exp\left[-\frac{(\mathbf{r}'-\mathbf{r})^2}{2\sigma^2}\right].$$

Differentiating equations (18) and (19) with respect to t, one can readily find that redundant terms on the right hand side of equalities (6) decreases considerably through smoothing, which is natural since shot noise is most prominent on small scales.

3 Method to measure statistical distribution of fluctuation on various scales

3.1 Algorithm

The algorithm of our analysis is simple. First choose a set of parameters, σ and R. We always take $R > \sigma$, since information on scales smaller than σ is suppressed. Next choose a center of a sphere randomly in the survey sample. We adopt and compare two different prescriptions to deal with the boundary problem associated with it. One is to take a center so that the entire sphere is contained in the surveyed region. The other is to adopt a periodic boundary condition. If some part of the cell falls outside the sample, we choose an antipodes appropriately and place another cell there and add up counts in both cells so that the deficit is compensated. Either prescription, however, is by no means free from problems; the former tend to overweight the central region of the sample, while the latter corresponds to forbidding fluctuations on scales larger than the dimension of the sample. Since observational data are available only for a tiny portion of the entire sky today, all we may do is to compare the results of these prescriptions and guess the magnitude of errors caused by the smallness of the data. Both prescription should give the same results if R is adequately smaller than the size of the sample $\sim$ 100Mpc. Below we find that

periodic boundary condition gives a better results in general, since it mimics the contribution of galaxies outside the sample which is absent in the nonperiodic boundary prescription.

Once a sphere is placed in the sample, we sum up the contribution of each galaxy to the cell. The galaxy at $\mathbf{x}$ contributes to the count in the sphere centered at the origin with the amount of

$$\begin{aligned} N(\mathbf{0},\mathbf{r},\sigma,R) &= \frac{1}{(\sqrt{2\pi}\sigma)^3}\int_{|\mathbf{x}|\le R}\exp\left[-\frac{(\mathbf{x}-\mathbf{r})^2}{2\sigma^2}\right]d^3x \qquad (20)\\ &= \frac{\sigma}{\sqrt{2\pi}r}\left\{\exp\left[-\frac{(R+r)^2}{2\sigma^2}\right]-\exp\left[-\frac{(R-r)^2}{2\sigma^2}\right]\right\}+\frac{1}{2}\mathrm{erf}\left(\frac{R+r}{\sqrt{2}\sigma}\right)+\frac{1}{2}\mathrm{erf}\left(\frac{R-r}{\sqrt{2}\sigma}\right), \end{aligned}$$

where

$$\mathrm{erf}(z)\equiv\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$$

is the error function.

We repeat the above procedure many times so that the sampling spheres overlap with each other to a considerable extent, which we can quantify in terms of the frequency each point is covered by a cell on average, which we shall call the coverage, c. Although the value of c does not affect our estimate of probability distribution $P[\widehat{\delta_\sigma N}]$, it certainly does control the magnitude of estimated error bars. In fact if we took c too large, statistical error would be reduced incorrectly. Hence we should impose some upper bound on c, that is, one must know which is the total number of statistically independent cells (see Otto *et al.* 1986).

One may wonder that it should be $c = 1$ to avoid overcounting. However, this is not the case, because what we accumulate is not the position of each galaxy within each cell but its effective total number, a coarse-grained quantity. Hence c can be much larger than unity in general. The information we can obtain, after performing the analysis with coverage c, is the effective total number of galaxies within a volume the of order of the "resolution volume" $V_R \equiv V_c/c \ll V_c$ all over the sample, although its size and accuracy may fluctuate due to the randomness of our sampling procedure.

Now let us ask a question: with the above information alone, to which extent could we predict the effective total number of galaxies within another randomly placed sphere with volume V_c ? One can write it as

$$N = \sum_{V_{Ri}\in V_c} n_i V_{Ri} + \sum_{V_{Rj}\in \partial V_c} f_j n_j V_{Rj}, \qquad (21)$$

where the first summation is over the resolution cells, V_{Ri}, which are fully contained in V_c and the second runs over those crossing the boundary with f_j being the effective number-fraction of galaxies contained in V_c. The former involves $\sim V_c/V_R = c$ terms while the latter has $\sim (V_c/V_R)^{2/3} = c^{2/3}$ terms. It is the second part that causes error in the estimate of N. Thus we can predict N up to the accuracy of

$$\frac{\Delta N}{N} \approx \frac{\sum_{V_{Rj}\in\partial V_c} f_j n_j V_{Rj}}{\sum_{V_{Ri}\in V_c} n_i V_{Ri}} \sim \left(\frac{1}{c}\right)^{1/3}. \qquad (22)$$

If this quantity was found too small, this would imply that there had already been a cell in our analysis whose center was very close to the new one's, which in turn would mean that we had placed too many cells in the original procedure. In order to keep independence of each cell in our analysis, the above error should remain *large* enough. Its reasonable lower bound would be given by the statistical fluctuation in the sample. Thus we take the coverage to satisfy

$$\frac{\Delta N}{N} \gtrsim \frac{\widehat{\delta_\sigma N}}{\widehat{\overline{n}_\sigma} V_c} = \widehat{\xi}_\sigma^{1/2}(V_c), \qquad (23)$$

or

$$c \lesssim \widehat{\overline{\xi_\sigma}}^{-3/2}(V_c). \tag{24}$$

This constraint becomes weaker as we increase σ or V_c, since it is more difficult to verify that two different cells are actually dependent with each other in smoother distribution. However, it is not meaningful to have the resolution volume V_R much smaller than the smoothing volume $V_\sigma \equiv \frac{3\pi}{4}\sigma^3$. We have chosen the value of c keeping this point in mind in the case inequality (24) does not give us a strong constraint. In practice the number of cells placed in each calculation ranged from 210 to 770 depending on samples and values of parameters.

Once the number of independent cells, N_c, is calculated, one would like to set the size of the bins, $\Delta\widehat{\delta_\sigma N}$, into which we accumulate the calculated results in order to reconstruct probability distribution. When N_c is small we have to take a larger value of $\Delta\widehat{\delta_\sigma N}$ so that each bin contains large enough events to keep the relative error small enough. Thus we end up with less number of histogram bars.

The error in the estimation of the probability of a given bin, $p \equiv P[\widehat{\delta_\sigma N}]\Delta\widehat{\delta_\sigma N}$, is calculated in terms of the binomial distribution, which gives,

$$\frac{\delta p}{p} = \sqrt{\frac{1}{pN_c}}, \tag{25}$$

for $p \ll 1$. By imposing an upper bound on this error one may find a lower bound on $\Delta\widehat{\delta_\sigma N}$ when $P[\widehat{\delta_\sigma N}]$ is known. We fix it so that $pN_c > 10$ at the bins corresponding to $\widehat{\delta_\sigma N} = \pm\sqrt{\langle\widehat{\delta_\sigma N}^2\rangle}$. That is, each probability bin near the peak is estimated at least using ten cells, which is adequate for a χ^2 test.

In practice, since we do not know the probability distribution *a priori*, we first assume Gaussian distribution with dispersion given by equation (17) using the observed values of γ and s_0 (see Table 1 below) for the first estimate of N_c and $\Delta\widehat{\delta_\sigma N}$. Then we run the program once to estimate $P[\widehat{\delta_\sigma N}]$, from which we compute the final values of N_c and $\Delta\widehat{\delta_\sigma N}$. With these values of parameters we repeat the calculation for ten times to compute the final result of $P[\widehat{\delta_\sigma N}]$ and its errors. In the figures the probability distribution is depicted in terms of points, indicating the hight of a histogram, with errorbars. The probability distribution of density fluctuation contains all the statistical information as it is related with the characteristic function by equation (32).

3.2 Significance of the method

As we are proposing a new method to probe statistical distribution of fluctuations on various scales, we should make sure that it is not affected by some bias, such as shot noise or the central limit theorem, which artificially reproduces a Gaussian distribution. Although it would be interesting to generate various data artificially using N-body simulations to test our algorithm, we have generated some simple distributions with extreme statistical properties to check this point for the moment. We have restricted our simulated samples to have the same total volume, shape, and two-point correlation function as the galaxy samples. We also change the density (with fixed correlation and volume) to check how the distribution is affected.

First, in order to test the sensitivity to the intrinsically non-Gaussian samples, we have generated a non-Gaussian sample by introducing large-scale coherence *i.e.* a wall in which points are distributed randomly with higher density than the background. One could choose the size of the wall in a given sample to obtain the desired two-point correlation function. In Figures 1 and 2 we present the results for an artificial distribution with points located in the same region as the galaxy sample CfAN80 (see Table 1 below) and with the same number of points. The artificial sample, WA80, has all the points in a spherical

wall of 10 Mpc thick (between redshift 6500 km/s and 7500 km/s) with a density contrast of ~ 3. This is a simple simulation of the so called 'Great Wall' (Geller and Huchra 1989). As is seen in Figures 2, the statistical distribution is observed to be non-Gaussian even for very large cells and smoothing parameter σ. Deviation from Gaussian distribution is more important at the characteristic scale of the wall, that is, around 10Mpc. We have also tried various walls with different thickness and density contrast, in all of which we have obtained non-Gaussian distributions at some characteristics scales. These preliminary results prove that our algorithm is sensitive to non-Gaussian fluctuations and not affected by the central limit theorem.

Next we have changed the number of points in the samples considerably, without changing the correlation function, to test if our method is affected by shot noise. As a result, no change in the statistical distribution was observed between the simulations with the same number of points as the galaxy sample and those with much more points.

As mentioned in section 2.2 we smooth the the density field twice, so that the same amount of smoothing can be achieved with different combinations of the values R and σ which we use to characterize the fluctuations. An independent way to characterize the amount of smoothing is to measure how much we dump the density fluctuations which is measured by $\widehat{\overline{\xi_\sigma}}(R)$. A good test for our algorithm would be to check that the measured distribution of fluctuations depends only on $\widehat{\overline{\xi_\sigma}}(R)$ and not on the combination of R and σ. Although limited by the actual size of the sample, we have performed this test both for simulated samples and for real samples, and a good agreement has been found. For example, in the sample CfAN80 the measured value of the dispersion for $\sigma = 12$Mpc and $R = 18$Mpc is $\widehat{\overline{\xi_\sigma}} = 2.1$, the same as the value obtained for $\sigma = 14$Mpc and $R = 16$Mpc, and the shape of the distribution of density fluctuations was found very similar to each other. This is another good proof that our analysis is free from the shot noise because it shows that the result is independent of the average density in a cell, $\widehat{\overline{N_\sigma}}$.

As a further test, we compare the first and second moments of the probability distribution in each case with the expected values from the density and the two-point correlation function of the discrete sample. In principle the average density in a cell $\widehat{\overline{N_\sigma}}$ should be equal to the sample density times the volume on the cell, $4\pi R^3/3$, which is independent of σ. We have found, however, that $\widehat{\overline{N_\sigma}}$ decreases with increasing σ for very large cells. This is because the Gaussian smoothing spreads the points outside the boundaries of the samples. When we use the prescription of periodic boundaries most of this effect is compensated provided $\sigma \ll R$. But there are still some differences for $\sigma \sim R$. This discrepancy does not seem to affect much the shape of the distribution of fluctuations, as has been tested by comparing the resulting distribution with the distribution obtained when we used different values of R and σ with larger $(R - \sigma)$ but with similar $\widehat{\overline{\xi_\sigma}}$.

A further check is obtained from the agreement between periodic and non-periodic boundaries strategy. We find both prescriptions agrees with each other reasonably well except for the fact that there is a characteristic cut-off in the tail of large density fluctuations for large cells in the non-periodic boundary prescription. This comes from the Gaussian spread of each galaxy outside the boundaries.

4 Analysis of observational data

4.1 Galaxy catalogs

We use two different catalogs of galaxies for our analysis. One is North Zwicky Center for Astrophysics (CfA) catalog with $m_B < 14.5$, $\delta \geq 0$, and $b^{II} \geq 40°$ which has a solid angle of 1.83sr (Huchra *et al.*

1983). The other is the Southern Sky Redshift Survey (SSRS) of diameter-selected galaxies from the ESO catalog with $\delta \leq -17.5°$, $b^{II} \leq -30°$, and the solid angle of 1.75sr (Da Costa *et al.* 1988). Heliocentric redshifts are corrected only from our motion with respect to the rest frame of the Cosmic Microwave Background; $v = 365$km/sec and the direction $(\alpha, \delta) \sim (11.2^h, -7°)$ (Smoot *et al.* 1991).

We have taken two different volume-limited samples out of each redshift catalogy as shown in Table 1. In addition to the limiting radius, $r_{\rm lim}$, we have also put a cutoff at $r = 25$Mpc. Galaxies with smaller redshift are not taken into account since their typical peculiar velocity is as large as recession velocity in general. Samples CfAN80 and CfAN92 include galaxies brighter than $M_B = m_B - 25 - 5\log r_{\rm lim}$, where $m_B = 14.5$ is the limiting apparent magnitude. Samples SSRS80 and SSRS115 include galaxies with physical diameter greater than $d_{\rm lim} = r_{\rm lim}\theta_{\rm cut}$, where $\theta_{\rm cut} = 1'.26$ is the 'face-on' diameter cut-off. We use volume-limited samples because they provide an unbiassed trace of the galaxy distribution of a given range of absolute magnitude. This is not the case for apparent-magnitude-limited samples, where a selection function must be introduced. The price we pay is that we end up with a small number of bright galaxies. As we are interested in the smoothed field at large scales, having more galaxies does not necessarily give a better statistics as shown in the previous section, whereas using a selection function would introduce big uncertainties even in the value of the density. Properties of similar volume-limited samples have been studied by several authors (*e.g.* Einasto, Klypin and Saar, 1986; Davis *et al.*, 1988; Hamilton 1988; Pellegrini *et al.* 1990; Maurogordato and Lachièze-Rey 1991; Zucca *et al.*, 1991).

The two-point redshift correlation function is estimated in the usual way, that is, counting pairs at a given redshift separation interval, $(s, s + ds)$. To avoid problems with edge effects, the counts are normalized with pair counts between the galaxies and a simulated catalog of randomly distributed points within the same sample boundary (the control sample), *i.e.*,

$$1 + \xi(s) = \frac{N_{DD}(s)}{N_{DR}(s)} \left(\frac{N_R}{N_D}\right) . \tag{26}$$

Here, $N_{DD}(s)$ is the number of real galaxy pairs with redshift space separation $(s, s + ds)$, and $N_{DR}(s)$ is the number of pairs at the same separation with one point in the data and the other in the random control sample. The total number of galaxies in the random and galaxy catalogs are N_R and N_D. To obtain a better trace of the boundaries, the control sample has been chosen to have three times as many points as the corresponding galaxy sample, $N_R = 3N_D$. Each correlation function is an average of the correlation functions computed over different realizations of the control sample.

For the shake of clarity, we use for Figure 3 the volume-average of two-point correlation, that is,

$$\frac{1}{V}\int_V \xi(s')d^3s' = VJ_3(s) \sim \bar{\xi}(s) \tag{27}$$

with V the volume of a sphere of radius s. For a power law, $VJ_3(s)$ is just proportional to $\xi(s)$ and also to $\bar{\xi}(s)$. The correlation length, s_0, and power index γ of each correlation function are also shown in Table 1. There we have fit the data in the range of $s = 5 - 20$Mpc, since we are interested in these large scales. The best fit values of these parameters, however, change considerably depending on the range of scales we try to fit. Thus one should keep in mind that s_0 and γ typically suffer from the errors of, at least, ± 1Mpc and ± 0.1, respectively.

The correlation in all of the samples is very similar, in agreement with Davis *et al.* (1988) and Pellegrini *et al.* (1990), but different from the 'standard' value $r_0 \sim 5$Mpc, $\gamma \sim 1.8$ (Davis and Peebles 1983). This might be caused, partially, by redshift distortions, but it is not clear yet whether there are other selection effects (see Einasto, Klypin and Saar 1986; Hamilton 1988; Davis *et al.* 1988; Maurogordato

and Lachièze-Rey 1991; Frieman and Gaztañaga 1992). Using the average values of our samples, $\gamma = 1.55$ and $s_0 = 7.5$Mpc, we find the scale of nonlinearity with $\bar{\xi} \simeq 1$ to be $\simeq 10$Mpc in contrast to $\simeq 8$Mpc found by Davis and Peebles (1983).

4.2 Gaussian fluctuations on large scale

In terms of the algorithm presented in section 3 to measure the distribution of density fluctuations, we have analyzed the samples with various values of R and σ using a number of tests to measure how Gaussian the distribution is. The result is summarized in figures 4 and Tables 2, where we showed the dispersion $\widehat{\xi_\sigma}$, the skewness, $\widehat{\eta_\sigma} \equiv \langle (N - \widehat{N_\sigma})^3 \rangle / \widehat{N_\sigma}^3$, the and consistency with a Gaussian distribution in terms of χ^2-test. N_χ represents the number of data bins used for fitting, each of which contains ten or more events. As is seen in Tables 2 the distribution becomes less skewed as we increase σ with fixed R. Statistical errors are 1-3% in $\widehat{N_\sigma}$ and 4-6% in $\langle \widehat{\delta_\sigma N}^2 \rangle$, so that the error in $\widehat{\xi_\sigma}$ is around 10%. Errors for $\langle \widehat{\delta_\sigma N}^3 \rangle$ increase with the cell size from 5-10% to up to 60%. For the smallest values of the skewness, the error is comparable to the observed value.

The χ^2 fit does not necessarily exhibit monotonic behavior, because the test is sensitive only to those bins around the peak. We have also tried Kolmogorov-Smirnov test using the binned data but it always resulted in high consistency with more than 99%.

Our results imply that it is not until smoothing length is as large as $\sigma \simeq 14$Mpc or $\widehat{\xi_\sigma} \sim .05$, which is considerably larger than the scale of nonliniarity, that the distribution becomes practically Gaussian. This is consistent with previous analyses using different methods (*e.g.* Ryden *et al.* 1989). We also note that the calculated values of $\langle \widehat{\delta_\sigma N}(R)^2 \rangle$ were in agreement with the theoretical estimate (e.q. [17]). Figures 4 depicts the case $\sigma = 16$Mpc for all the samples with periodic boundary condition.

4.3 Gravitational effects on the statistics

As is mentioned above, the statistics deviate from Gaussian distribution even in the linear regime with $\widehat{\xi_\sigma} \ll 1$. We now consider theoretical modeling of this deviation which is a consequence of gravitational instability. Although a lot of models have been tested in the discrete version of counts-in-cells (*e.g.* Fry 1986; Balian and Schaeffer 1989b; Fry *et al.* 1989; Maurogordato and Lachièche-Rey 1991; Cappi *et al.* 1991), few of them are theoretically well-motivated. Here we consider two specific models and compare them with observation.

4.3.1 Hierarchical BBGKY distribution

This model has been formulated by Fry (1985) based on the hierarchy ansatz of irreducible J-point correlation function, κ_J, which is expressed as

$$\kappa_J = Q_J \sum_{\{ij\}} \xi_{ij}^{J-1}. \tag{28}$$

(see also Balian and Schaeffer 1989a, 1989b and references therein). Here the product of two-point function, ξ_{ij}, is over $J-1$ independent pairs of relative separations and the sum, consisting of J^{J-2} terms, is over equivalent reassignments of labels $i, j = 1, 2, \ldots, J$. The above equality is supported by some observational evidence at least up to some finite order (Groth and Peebles 1977; Davis and Peebles 1977). Extending the earlier work of Davis and Peebles (1977), Fry (1984) has found that BBGKY

equations in the strong clustering limit can be consistently solved with constant Q_J's, independent of the form of $\xi(r)$, provided we iteratively choose Q_J's to be

$$Q_J = \left(\frac{4Q}{J}\right)^{J-2} \frac{J}{2(J-1)}, \tag{29}$$

where Q is the three-point amplitude.

Thus J-th order cumulant of fluctuation in the volume V, $\langle \delta N^J(V) \rangle_C$, is given by

$$\begin{aligned} \langle \delta N^J(V) \rangle_C &= (\overline{n}V)^J \int_{V^J} d^3r_1 \ldots d^3r_J \kappa_J \\ &= (\overline{n}V)^J J^{J-2} Q_J \bar{\xi}^{J-1}(V) \equiv (\overline{n}V)^J k_J(V). \end{aligned} \tag{30}$$

Hence the characteristic function of $\Delta \equiv \delta N(V)/(\overline{n}V)$ is calculated as

$$\begin{aligned} \langle e^{it\Delta} \rangle &= \exp\left[\sum_{J=2}^{\infty} \frac{(it)^J}{J!} k_J\right] \\ &= \exp\left[\frac{1}{8Q} \int_0^{4Q\bar{\xi}it} \frac{dx}{x}(e^x - 1)\right] \equiv f(t). \end{aligned} \tag{31}$$

Finally, the probability distribution of Δ is derived in terms of the inverse Fourier transform,

$$P(\Delta) = \int \frac{dt}{2\pi} f(t) e^{-it\Delta}, \tag{32}$$

which depends on $\bar{\xi}$ and Q as well. This distribution approaches Gaussian as $\bar{\xi}$ decreases. The non-Gaussian behavior is prominent when $\bar{\xi} > (4Q)^{-2} \simeq .05$ with $Q \simeq 1$ (Fry 1985). From equations (28) and (29) we find that the inequality (7) is valid up to $J = 5$ for $\overline{N\xi} = 5$ in this model. To compare with the observational data we have performed a numerical fast Fourier transform to evaluate equation (32).

4.3.2 Negative binomial distribution

The other model we consider is the negative binomial or modified Bose-Einstein distribution. It has been used in a number of fields with different physical backgrounds such as quantum optics (Klauder and Sudarshan 1968), hadronic multiplicity (Carruthers and Shih 1983), galaxy counts in a Zwicky cluster (Carruthers and Minh 1983) or large scale galaxy counts (Fry 1986). The distribution has been theoretically re-derived by Elizalde and Gaztañaga (1992) in an appropriate manner to our problem by introducing a correlation to uniform Poisson distribution.

What they call quasi-Poisson distribution is derived as follows. First divide the sample of N points to m identical cells of volume $V_c = V_T/m$ and consider the probability, $P_N(k_1, \ldots, k_m)$, of having k_j particles in the j-th cell. We assume that each cell is independent with each other and the distribution is characterized by volume-averaged two-point correlation function $\bar{\xi}(V_c) \ll 1$ with negligible higher-order correlations. Now we consider adding one particle to the system without disturbing its statistical properties. The probability that this particle is placed in the j-th cell, which contains k_j particles, is proportional to

$$P_{E_j(k_j)}[E_j(k_j+1)] = \frac{P[E_j(k_j) \cap E_j(k_j+1)]}{P[E_j(k_j)]} = \frac{P[E_j(k_j+1)]}{P[E_j(k_j)]}$$

$$= \frac{1+\binom{k_j+1}{2}\bar{\xi}+O(k_j^4\bar{\xi}^2)}{1+\binom{k_j}{2}\bar{\xi}+O(k_j^4\bar{\xi}^2)} \cong 1+k_j\bar{\xi}, \tag{33}$$

where $E_j(l)$ denotes the event that the j-th cell contains l particles and we have assumed $k_j^2\bar{\xi} \ll 1$. Hence we find the following recurrence formula.

$$P_N(k_1,\ldots,k_m) = \sum_{l=1}^{m} P_{N-1}(k_1,\ldots,k_l-1,\ldots,k_m)\frac{1+(k_l-1)\bar{\xi}}{m+(N-1)\bar{\xi}}, \tag{34}$$

which yields,

$$P_N(k_1,\ldots,k_m) = \frac{N!}{k_1!\cdots k_m!}\prod_{j=1}^{m}\prod_{l=1}^{k_j}\frac{1+(l-1)\bar{\xi}}{m+\left(\sum_{i=1}^{j-1}k_i+l-1\right)\bar{\xi}}. \tag{35}$$

We can also express the probability in terms of a new variable x_i, number of cells containing i particles inside, as

$$P(x_0,\ldots,x_N) = \frac{N!}{(0!)^{x_0}\cdots(N!)^{x_N}}\frac{\prod_{l=1}^{N-1}(1+l\bar{\xi})^{\sum_{j=l+1}^{N}x_j}}{\prod_{l=0}^{N-1}(m+l\bar{\xi})}, \tag{36}$$

from which we find the expectation value of x_i:

$$\langle x_i\rangle = \frac{mN!\Gamma\left(\bar{\xi}^{-1}+i\right)\Gamma\left(m\bar{\xi}^{-1}\right)\Gamma\left((m-1)\bar{\xi}^{-1}+N-i\right)}{(N-i)!i!\bar{\xi}\Gamma\left(\bar{\xi}^{-1}+1\right)\Gamma\left((m-1)\bar{\xi}^{-1}\right)}. \tag{37}$$

Finally the desired quantity, the probability to find i particles in a cell, is obtained by taking $m \to \infty$ and $N \to \infty$ with $N/m = \bar{n}V_c$ =constant:

$$P_i(V_c) = \lim\frac{\langle x_i\rangle}{m} = \frac{(\bar{n}V_c)^i}{i!}(1+\bar{\xi}\bar{n}V_c)^{-i-\bar{\xi}^{-1}}\prod_{j=1}^{i-1}(1+j\bar{\xi}), \tag{38}$$

which is nothing but the negative binomial distribution with volume-averaged two-point correlation function $\bar{\xi}(V_c)$.

In the limit $\bar{\xi} \to 0$ equation (38) reproduces the Poisson distribution. This is why it was named quasi-Poisson. Note, however, that it also mimics Gaussian distribution in that higher-order correlation has been set to zero in each cell and that cells are assumed to be independent with each other so that its power spectrum has random phase. In this sense one may well call it quasi-Gaussian as well. In the continuum limit $\bar{n}V_c \to \infty$ with $x \equiv i/(\bar{n}V_c)$ constant, the probability distribution yields

$$P[x] = \lim_{\bar{n}\to\infty}\bar{n}V_cP_i[V_c] = \frac{1}{\Gamma(\bar{\xi}^{-1})\bar{\xi}}(\bar{\xi}^{-1}x)^{\bar{\xi}^{-1}-1}e^{-\bar{\xi}^{-1}x}. \tag{39}$$

If we further take the limit $\bar{\xi} \to 0$, it approaches Gaussian:

$$P[x] = \frac{1}{\sqrt{2\pi\bar{\xi}}}\exp\left[-\frac{(x-1)^2}{2\bar{\xi}}\right], \qquad \bar{\xi} \ll 1. \tag{40}$$

As is seen above, this model has no free parameter unlike Fry's hierarchical model. We also note that the negative binomial distribution, too, has the hierarchical property (e.q. [28]) with Q_J given by

$$Q_J = \frac{(J-1)!}{J^{J-2}}. \tag{41}$$

In this model the inequality (7) is satisfied up to as large J as $J = 10$ for $\overline{N\xi} = 5$, so that, in practice, the discrete distribution still contains information of the continuous field at large scales.

An interesting feature of this distribution is that for all scales the density is always positive which, as pointed in the introduction, is not the case with the Gaussian distribution or the hierarchical BBGKY model.

4.4 Comparison with observation

Figures 5a-d show the statistical distribution of the smoothed field in a sphere of radius $R = 16$ or 18Mpc with different smoothing scales with periodic boundary condition. Those with nonperiodic boundary condition are depicted in Figures 6 for CfAN92 sample for comparison. In these figures we have depicted Fry's distribution with $Q = 0.6$ and negative binomial distribution. As seen there, both of the curves fit remarkably well. We note, however, the best fit value of Fry's parameter $Q = 0.6$ is significantly lower than the observed value $Q \simeq 1$, although there is much uncertainty in it (Groth and Peebles 1977). In fact $Q \simeq 1$ does not fit the data at all for $\sigma \lesssim$ 10Mpc where fluctuation is non-Gaussian. Tables 3 summarize the results of fitting to Fry's model with $Q = 0.6$ and negative binomial distribution based on the χ^2 test.

Finally, for completeness, we depict the result of counts-in-cells of discrete galaxies in Figure 7. The scaling function, $\chi \equiv -\ln(P_0/\overline{n}V_c)$ is plotted for different values of $\overline{n}V_c\overline{\xi}$, where $\overline{n}V_c$ and $\overline{\xi}$ are the average density and dispersion from the measured counts-in-cells in a spherical cell of volume V_c. We have used a similar algorithm as in §3.1 with periodic boundary condition, which is necessary to reproduce the expected values of $\overline{n}V_c$ and $\overline{\xi}$.

5 Conclusion

In the present paper we proposed a new method to examine the statistical distribution of fluctuations on various scales. We have been able to prove that the method can significantly distinguish the deviations from a Gaussian distribution, and is not affected by the central limit theorem. Besides we have shown that the analysis is free from shot noise in the particular situation we inspected, as explained in section 3.2.

Although it is usually assumed that density fluctuations on very large scale should trace primordial density fluctuations, it is not clear how large this scale should be. We have found that the statistical distribution of density fluctuations on large scales becomes practically Gaussian if the value of the smoothed correlation function drops to $\widehat{\overline{\xi_\sigma}} \sim 0.05$. In our analysis this value correspond to scales of $\sim$ 14Mpc as we use a Gaussian window function. For the standard top-hat window function, the corresponding scale should be larger. If the correlation function is a power law with power $\gamma \sim 2$, and amplitude one at 8 Mpc, the corresponding scale for a top-hat window would be $\sim$ 30 Mpc.

Fluctuations on smaller scales evolve according to Fry's model or negative binomial model. We suppose the latter is the better, since it is a natural extension of Gaussian distribution as explained

above[2]. However, this explanation is valid only when correlation is very small. On the other hand, Fry's model is theoretically well-motivated in the strongly nonlinear regime. This agreement, together with the fact that both models approach a Gaussian distribution for small correlations (large scales) confirm our point that the distribution on large scale is Gaussian. Furthermore one can evaluate how large the deviation from a Gaussian distribution is by simply calculating $\bar{\xi}$.

With either of the models our results confirm the hypothesis that we live in a hierarchical universe. Independent evidence comes from the scale invariance properties of counts in cells, both in angular catalogs (*e.g.* Sharp 1981; Sharp, Bonometto and Lucchin 1984) and redshift surveys (*e.g.* Fry 1986; Balian and Schaeffer 1989b; Fry *et al.* 1989; Maurogordato and Lachièche-Rey 1991; Cappi *et al.* 1991; Elizalde and Gaztañaga 1992).

Although what we have measured is the fluctuations in redshift space, we expect them to be a good tracer of real space at very large scales. The analyses have been done using redshift samples from two very different catalogs. Not only from different regions of the sky, (the CfA from the North galactic plane, SSRS from the South) but also selected with different criteria (CfA by luminosity, SSRS by size). Because of the last point, galaxies in CfA samples have about 50% of spiral galaxies whereas in the SSRS there are around 80% (Table 1). There are indications of segregation of the statistical properties of galaxies with different morphologies at small scales: ellipticals and S0 galaxies seems to be more clustered than spirals (Davis and Geller 1976; Giovanelli, Haynes and Chincarini 1986) and are found preferentially in the cores of rich clusters (Dressler 1980; Postman and Geller 1984). We have not detected any difference in the distribution of density fluctuations at large scales.

It would also be interesting to apply our analysis to redshift surveys complete to a larger apparent magnitude, like the CfA2 (Geller and Huchra 1989), where there seems to be apparent large-scale coherent structures, such as the 'Great Wall', in order to verify whether they correspond to non-Gaussian statistic. Although our samples CfAN80 and CfAN92 include, at least partially, the so called 'Great Wall' region, its statistical distribution, corresponding to very large absolute magnitude, seems completely Gaussian.

Finally we remind the reader that all our analyses have been done assuming the Poisson model is a reasonable approximation. If this is indeed the case, our result imply that the universe was imprinted Gaussian density fluctuations and has evolved keeping hierarchical form as described by Fry's model or the negative binomial distribution. We plan to clarify if it is a generic feature in a gravitating system using N-body simulations.

Acknowledgements

We thank Josh Frieman and Albert Stebbins for useful comments. This work was supported in part by DOE and NASA (grant NAGW-2381) at Fermilab. E.G. acknowledges support from Fulbright/MEC (grant FU91 14594408) and J.Y. acknowledges support from JSPS.

[2]Discrete counts in cells also seems to follow the discrete version of the negative binomial, see Figure 7 and Elizalde and Gaztañaga (1992), and references therein.

References

Balian, R. and Schaeffer, R. 1989a, *Astr. Astrophys.*, **220**, 1.

Balian, R. and Schaeffer, R. 1989b, *Astr. Astrophys.*, **226**, 373.

Bardeen, J.M., Bond, J.R., Kaiser, N. & Szalay, A.S. 1986, *Ap.J.*, **304**, 15.

Baumgart, D.J. and Fry, J.N. 1991, *Ap.J.*, **375**, 25.

Cappi, A., Maurogordato, S., and Lachièze-Rey, M. 1991, *Astr.Astrophys.*, **243**, 28.

Carruthers, P. and Shih, C.C. 1983, *Phys. Lett. B*, **127**, 242.

Carruthers, P. and Minh, D.-V. 1983, *Phys. Lett. B*, **131**, 116.

Da Costa, L.N. *et al.* 1988, *Ap.J.*, **327**, 544.

Davis, M. and Geller, M.J. 1976, *Ap.J.*, **208**, 13.

Davis, M. and Peebles, P.J.E. 1977, *Ap.J. Suppl.*, **35**, 425.

Davis, M. and Peebles, P.J.E. 1983, *Ap.J.*, **267**, 465.

Davis, M., Efstathiou, G., Frenk, C.S. and White, S.D.M. 1985, *Ap.J.*, **292**, 371.

Davis, M., Meiksin, A., Strauss, A. da Costa, L.N. and Yahil,A. 1988, *Ap.J.(Letters.)*, **333**, L9.

Dressler, A. 1980, *Ap.J.*, **236**, 351.

Elizalde, E. and Gaztañaga, E. 1992, *MNRAS*, **254**, 247.

Einasto, J., Klypin, A.A. and Saar, E., 1986, *MNRAS*, **219**, 457.

Fry, J.N. 1984, *Ap.J.(Letters)*, **227**, L5.

Fry, J.N. 1985, *Ap.J.*, **289**, 10.

Fry, J.N. 1986, *Ap.J.*, **306**, 358.

Fry, J.N., Giovanelli, R., Haynes, M.P., Mellot, A.L. and Sherrer, R.J. 1989, *Ap.J.*, **340**, 11.

Frieman, J.A., Olinto, A.V., Gleiser, M., and Alcock, C. 1989, *Phys. Rev. D*, **40**, 3241.

Frieman, J.A., and Gaztañaga, E. 1992, in preparation.

Geller, M.J. and Huchra, J.P. 1989, *Science*, **246**, 897.

Giovanelli, R., Haynes, M.P., Chincarini G.L. 1986, *Ap.J.*, **300**, 77.

Gott III, J.R. *et al.* 1989, *Ap.J.*, **340**, 625.

Griest, K. and Kolb, E.W. 1989, *Phys. Rev. D*, **40**, 3231.

Groth, E.J. and Peebles, P.J.E. 1977, *Ap.J.*, **217**, 385.

Hamilton, A.J.S. 1988, *Ap.J.(Letters)*, **331**, L59.

Huchra, J., Davis, M., Latham, D., and Tonry, J. 1983, *Ap.J. Suppl*, **52**, 89.

Kaiser, N. 1984, *Ap.J.(Letters)*, **284**, L9.

Klauder, J.R. and Sudarshan, E.C.G. 1968, *Fundamentals of Quantum Optics* (Benjamin: New York).

Lahav, O., Itoh, M., Inagaki, S., and Suto, Y. 1992, Kyoto University preprint YITP/U-92-01.

Maurogordato, S., and Lachièze-Rey, M. 1991, *Ap.J.*, **369**, 30.

Otto, S., Politzer, D.H., Preskill, J., and Wise, M.B. 1986, *Ap.J.*, **304**, 62.

Peebles, P.J.E. 1980, *The Large-Scale Structure of the Universe* (Princeton University Press: Princeton).

Pellegrini, P.S., Willmer, C.N.A., Da Costa, L.N., Santiago, B.X. 1990, *Ap.J.*, **350**, 95.

Perivolaropoulos, L., Brandenberger, R., and Stebbins, A. 1990, *Phys.Rev D*, **41**, 1764.

Postman, M., and Geller, M.J. 1984, *Ap.J.*, **281**, 95.

Ryden, B.S. *et al.* 1989, *Ap.J.*, **340**, 647.

Saunders, W. *et al.* 1991, *Nature*, **349**, 32.

Scherrer, R.J., and Bertschinger, E. 1991, Ohio State University OSU-TA-3/91..

Sharp, N.A. 1981, *MNRAS*, **195**, 857.

Sharp, N.A., Bonometto, S.A. and Lucchin F. 1984, *Astr. Astrophys.*, **130**, 79.
Smoot, G.F. *et al.* 1991, *Ap.J.(Letters)*, **371**, L1.
Suginohara, T. and Suto, Y. 1991, *Ap.J.*, **371**, 470.
Turok, N., and Spergel, D.N. 1991, *Phys.Rev.Lett.*, **66**, 3093.
Vilenkin, A. 1981, *Phys.Rev.Lett.*, **46**, 1169.
White, S.D. 1979, *MNRAS*, **186**, 145.
Zel'dovich, Y.B. 1980, *MNRAS*, **192**, 663.
Zucca, E., Bardelli, S., Cappi, A., and Moscardini, L. 1991, *MNRAS*, **253**, 401.

Table 1 Samples of galaxies used for the analyses. $r_{\rm lim}$, M_B, $d_{\rm lim}$ and $N_{\rm tot}$ denote, respectively, maximum distance or depth, limiting absolute magnitude (CfA), limiting physical diameter (SSRS) and total number of each sample. Also shown are percentage of spiral galaxies and values of $\overline{N}$ and $\overline{\xi}$ within the sphere of $R = 18$Mpc based on discrete-count. Samples CfAN80 and CfAN92 include galaxies brighter than $M_B = m_B - 25 - 5\log r_{\rm lim}$, where $m_B = 14.5$ is the limiting apparent magnitude. Samples SSRS80 and SSRS115 include galaxies with physical diameter greater than $d_{\rm lim} = r_{\rm lim}\theta_{\rm cut}$, where $\theta_{\rm cut} = 1'.26$ is the 'face-on' diameter cut-off.

Sample	CfAN80	CfAN92	SSRS80	SSRS115
$r_{\rm lim}$ (Mpc)	80	92	80	115
$M_B/d_{\rm lim}$(kpc)	-20.0	-20.3	29.2	42.0
$N_{\rm tot}$	214	146	220	134
Volume (Mpc^3)	3.0×10^5	4.7×10^5	2.9×10^5	8.8×10^5
$\overline{n}$ (Mpc^{-3})	7.1×10^{-4}	3.1×10^{-4}	7.6×10^{-4}	1.5×10^{-4}
Spiral (%)	52	55	80	83
s_0(Mpc)	8	7	7	8
γ	1.5	1.7	1.6	1.4
$\overline{N}$(18Mpc)	17	7.1	18	3.3
$\overline{\xi}$(18Mpc)	0.32	0.23	0.24	0.25

Tables 2: Comparison of data with Gaussian distribution. $\widehat{\overline{\xi_\sigma}}$, $\widehat{\overline{\eta_\sigma}}$, and consistency with Gaussian based on χ^2-test are shown. N_χ denotes numbers of binns used for the fitting. Periodic (non-periodic) boundary conditions are used.

CfAN80	R (Mpc)	16	16	16	16	20
	σ (Mpc)	6	8	10	12	16
	$10^2\widehat{\overline{\xi_\sigma}}$	35 (28)	25 (21)	20 (17)	16 (15)	9.3
	$10^2\widehat{\overline{\eta_\sigma}}$	15 (15)	7.0 (8.2)	4.3 (4.9)	1.7 (1.1)	.25
	$\chi^2_{\rm Gauss}$ (%)	22 (4.0)	31 (4.7)	31 (.64)	74 (71)	92
	N_χ	7 (8)	11 (12)	15 (13)	8 (9)	7

CfAN92	R (Mpc)	18	18	18	18	24
	σ (Mpc)	5	8	11	14	16
	$10^2\widehat{\overline{\xi_\sigma}}$	28 (25)	19 (18)	12 (12)	8.4 (8.6)	5.3
	$10^2\widehat{\overline{\eta_\sigma}}$	15 (11)	7.3 (6.1)	3.0 (2.5)	1.0 (.94)	.01
	$\chi^2_{\rm Gauss}$ (%)	88 (1.5)	68 (11)	34 (11)	98 (91)	99
	N_χ	14 (14)	18 (14)	18 (20)	13 (12)	9

SSRS80	R (Mpc)	16	16	16	16	20
	σ (Mpc)	6	8	12	14	16
	$10^2\widehat{\overline{\xi_\sigma}}$	24 (28)	17 (19	7.2 (7.4)	5.3 (4.8)	2.6
	$10^2\widehat{\overline{\eta_\sigma}}$	10 (12)	5.1 (6.2)	.90 (.89)	.36 (.22)	.11
	$\chi^2_{\rm Gauss}$ (%)	7.2 (.15)	1.5 (0.0)	97 (2.0)	99 (98)	97
	N_χ	13 (12)	20 (21)	14 (19)	8 (6)	6

SSRS115	R (Mpc)	18	18	18	18	28
	σ (Mpc)	7	10	13	16	16
	$10^2\widehat{\overline{\xi_\sigma}}$	30 (27)	19 (16)	11 (9.7)	7.2 (6.3)	3.7
	$10^2\widehat{\overline{\eta_\sigma}}$	17 (11)	7.3 (3.8)	2.1 (8.9)	5.5 (-.04)	.24
	$\chi^2_{\rm Gauss}$ (%)	17 (79)	45 (91)	79 (80)	100 (72)	99
	N_χ	15 (12)	18 (18)	19 (21)	16 (18)	14

Tables 3: Consistency of the data with Fry's hierarchical model with $Q = 0.6$ and with negative binomial model based on the χ^2-test using periodic (non-periodic) boundary conditions.

CfAN80	R (Mpc)	16	16	16	16
	σ (Mpc)	6	8	10	12
	$\chi^2_{\rm Fry}$ (%)	77 (87)	99 (96)	98 (99)	98 (95)
	$\chi^2_{\rm NB}$ (%)	91 (99)	99 (100)	86 (100)	93 (96)
	N_χ	7 (8)	11 (12)	15 (13)	8 (9)

CfAN92	R (Mpc)	18	18	18	18
	σ (Mpc)	5	8	11	14
	$\chi^2_{\rm Fry}$ (%)	73 (98)	49 (92)	74 (98)	99 (100)
	$\chi^2_{\rm NB}$ (%)	87 (99)	63 (99)	78 (99)	98 (100)
	N_χ	14 (14)	18 (14)	18 (20)	13 (12)

SSRS80	R (Mpc)	16	16	16	16
	σ (Mpc)	6	8	12	14
	$\chi^2_{\rm Fry}$ (%)	98 (90)	97 (1.9)	100 (37)	98 (98)
	$\chi^2_{\rm NB}$ (%)	100 (97)	98 (4.2)	100 (34)	97 (98)
	N_χ	13 (12)	20 (21)	14 (19)	8 (6)

SSRS115	R (Mpc)	18	18	18	18
	σ (Mpc)	7	10	13	16
	$\chi^2_{\rm Fry}$ (%)	75 (96)	100 (67)	98 (86)	97 (83)
	$\chi^2_{\rm NB}$ (%)	98 (82)	100 (57)	86 (40)	79 (33)
	N_χ	15 (12)	18 (18)	19 (21)	16 (18)

Figure Captions

Figure 1. Comparison of discrete two-point correlation function, $\xi(s)$, of redshift sample CfAN80 (in Table 1) and simulated sample WA80.

Figure 2. Probability distribution of density fluctuations, estimated from counts-in-cells of the smoothed density field, for a simulated non-Gaussian sample WA80. Each graph shows the distribution for a fixed spherical cell of radius $R = 16$Mpc and increasing value of σ, the dispersion of the Gaussian smoothing function. The short-dashed line is the negative binomial model and the long-dashed the Fry's BBGKY hierarchical model (See §4 below). Also shown is the value of the dispersion, $\widehat{\overline{\xi}_\sigma}$, in units of $\frac{1}{16}$.

Figure 3. Volume-average of the two-point correlation function, $\frac{1}{V(s)}\int_{V(s)}\xi(s')d^3s' \sim \bar{\xi}(s)$, with $V(s)$ being the volume of a sphere of radius s, as a function of the redshift separation, s, in Mpc. The corresponding amplitude for the correlation function, $\xi(s)$, is one at $s \sim 7.5$Mpc. The dashed line shows a power law $\sim s^{-1.8}$.

Figure 4. Probability distribution of density fluctuations, estimated from counts-in-cells of the smoothed density field for each of the samples in Table 1. Each graph shows the distribution for a spherical cell of radius R in Mpc and a fixed value for the dispersion of the Gaussian smoothing function, $\sigma = 16$Mpc. The continuous line is a Gaussian distribution, the dashed one is the Fry's BBGKY hierarchical model. Also shown is the value of the dispersion, $\widehat{\overline{\xi}_\sigma}$, in units of $\frac{1}{16}$.

Figures 5. Probability distribution of density fluctuations, estimated from counts-in-cells of the smoothed density field, with periodic boundary conditions. Figures a,b,c and d show four different distributions for samples CfAN80, CfAN92, SSRS80 and SSRS92 in Table 1, respectively. Each figure displays four distribution for a fixed spherical cell of radius $R = 16$Mpc (Fig.5a and Fig.5c) or $R = 18$Mpc (Fig.5b or Fig.5d) and increasing value of σ, the dispersion of the Gaussian smoothing function. The short-dashed line is the negative binomial model and the long-dashed the Fry's BBGKY hierarchical model. Also shown is the value of the dispersion, $\widehat{\overline{\xi}_\sigma}$, in units of $\frac{1}{16}$.

Figure 6. Same as Fig.5b using non-periodic boundary conditions.

Figure 7. Void probabilities, from discrete counts-in-cells, for samples in Table 1. The scaling function, $\chi \equiv -\ln(P_0/\overline{n}V_c)$ is plotted for different values of $\overline{n}V_c\overline{\xi}$, where $\overline{n}$ and $\overline{\xi}$ are the average density and dispersion from the measured counts-in-cells in a spherical cell of volume V_c, respectively. The continuous line is the discrete negative binomial model, the dashed one is the discrete Gaussian distribution.

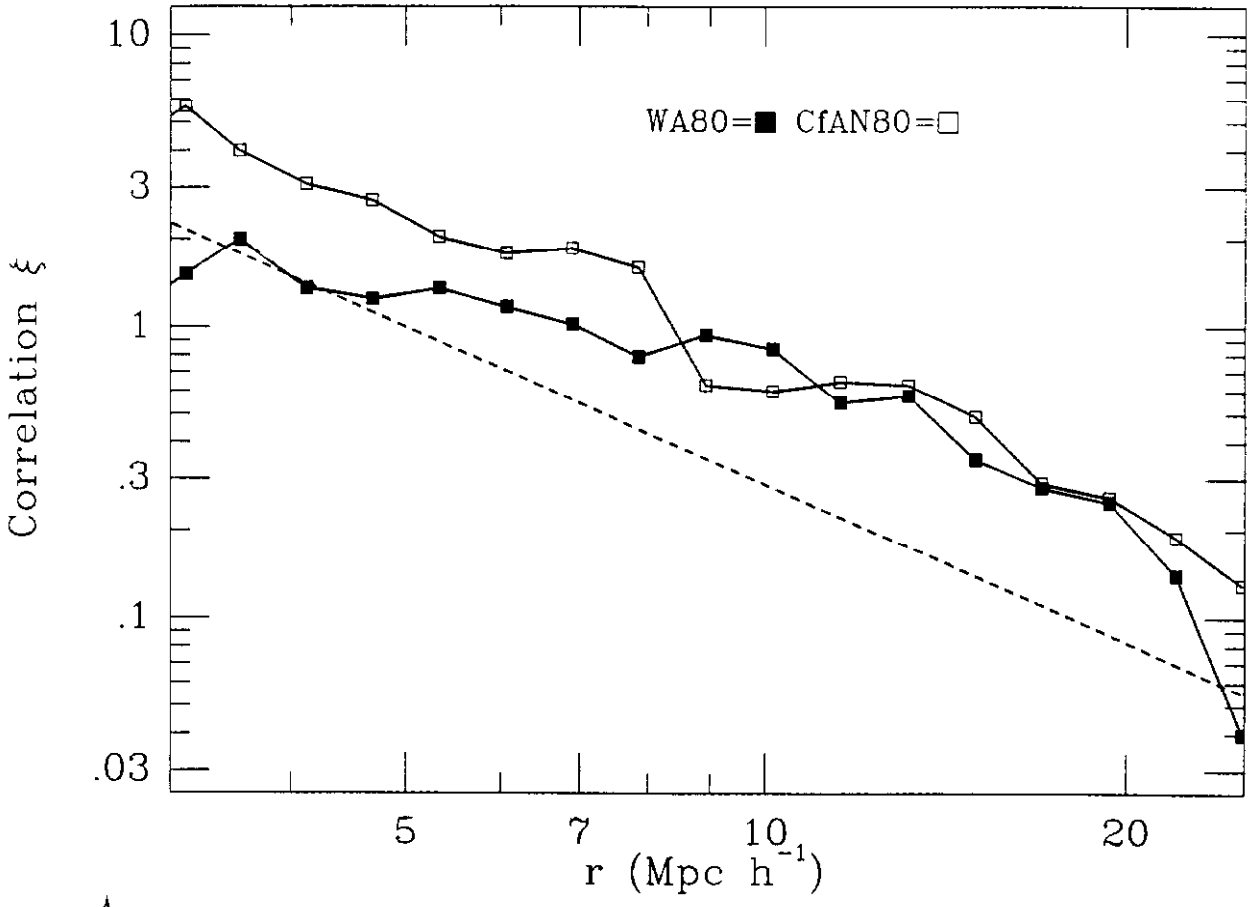
WA80=■ CfAN80=□
Correlation ξ
r (Mpc h^{-1})
10
3
1
.3
.1
.03
5
7
10
20

Fig 1.

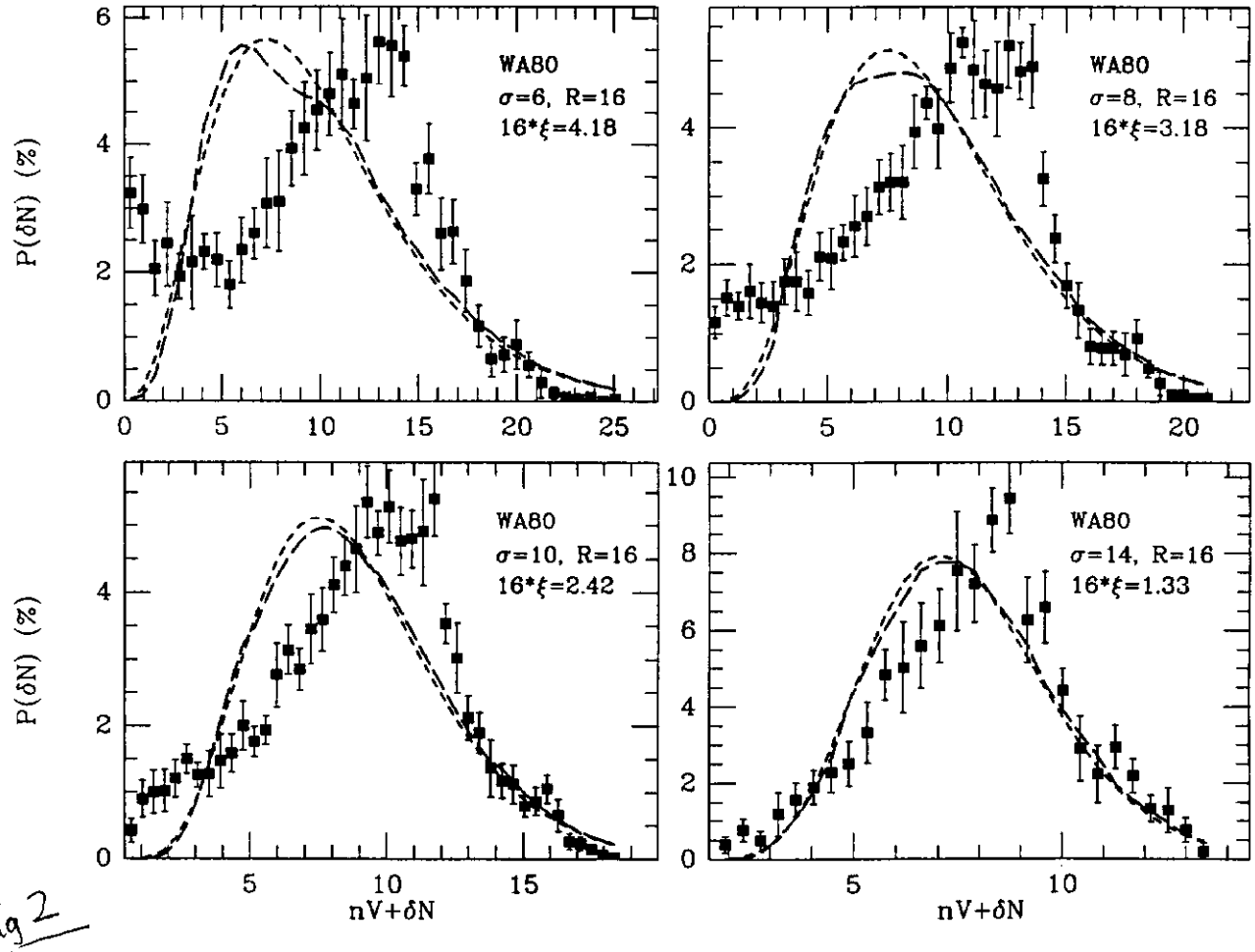
WA80
σ=6, R=16
16*ξ=4.18
WA80
σ=8, R=16
16*ξ=3.18
WA80
σ=10, R=16
16*ξ=2.42
WA80
σ=14, R=16
16*ξ=1.33
P(δN) (%)
nV+δN

Fig 2

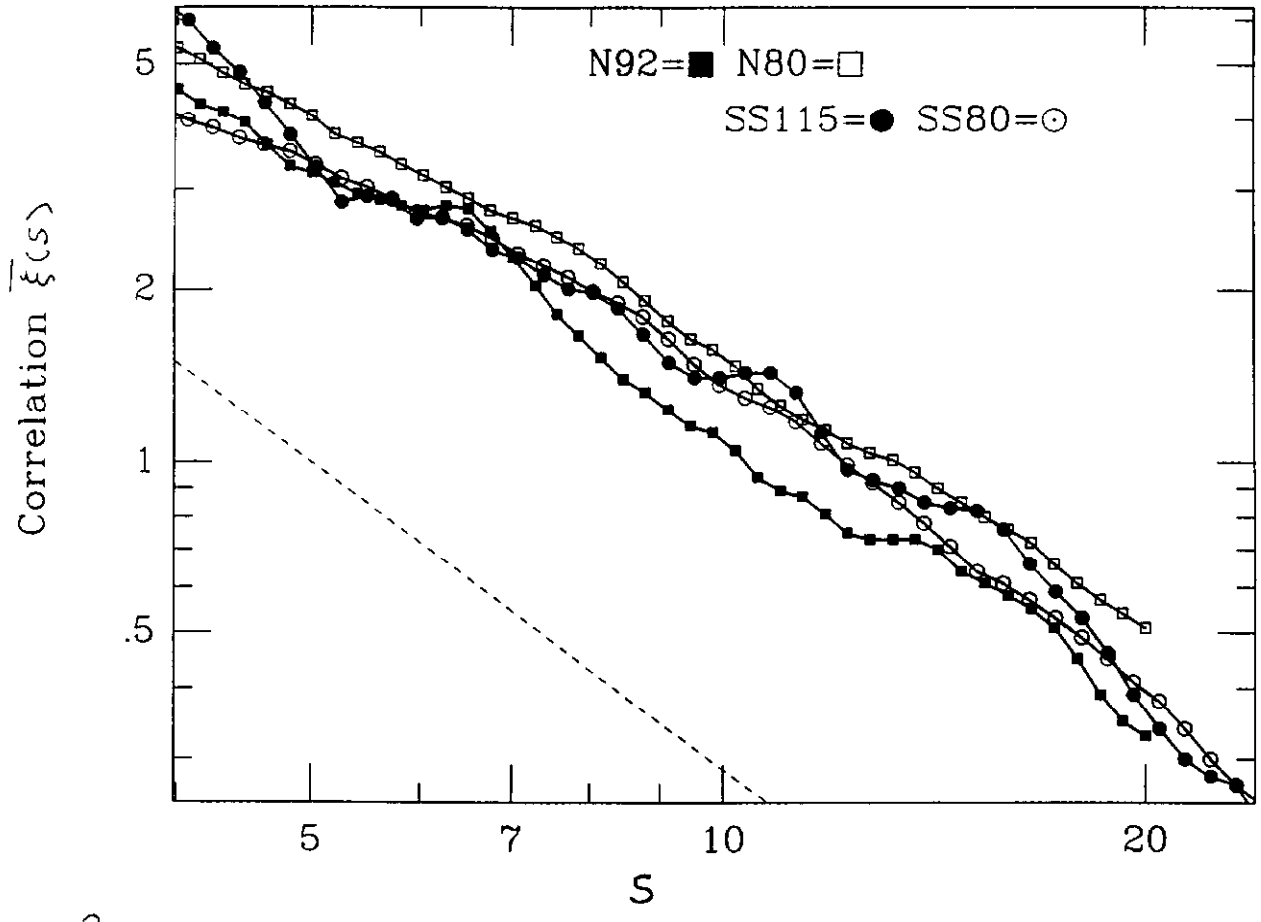
N92=■ N80=□
SS115=● SS80=⊙
Correlation $\bar{\xi}(s)$
5
2
1
.5
5
7
10
20
s

Fig. 3

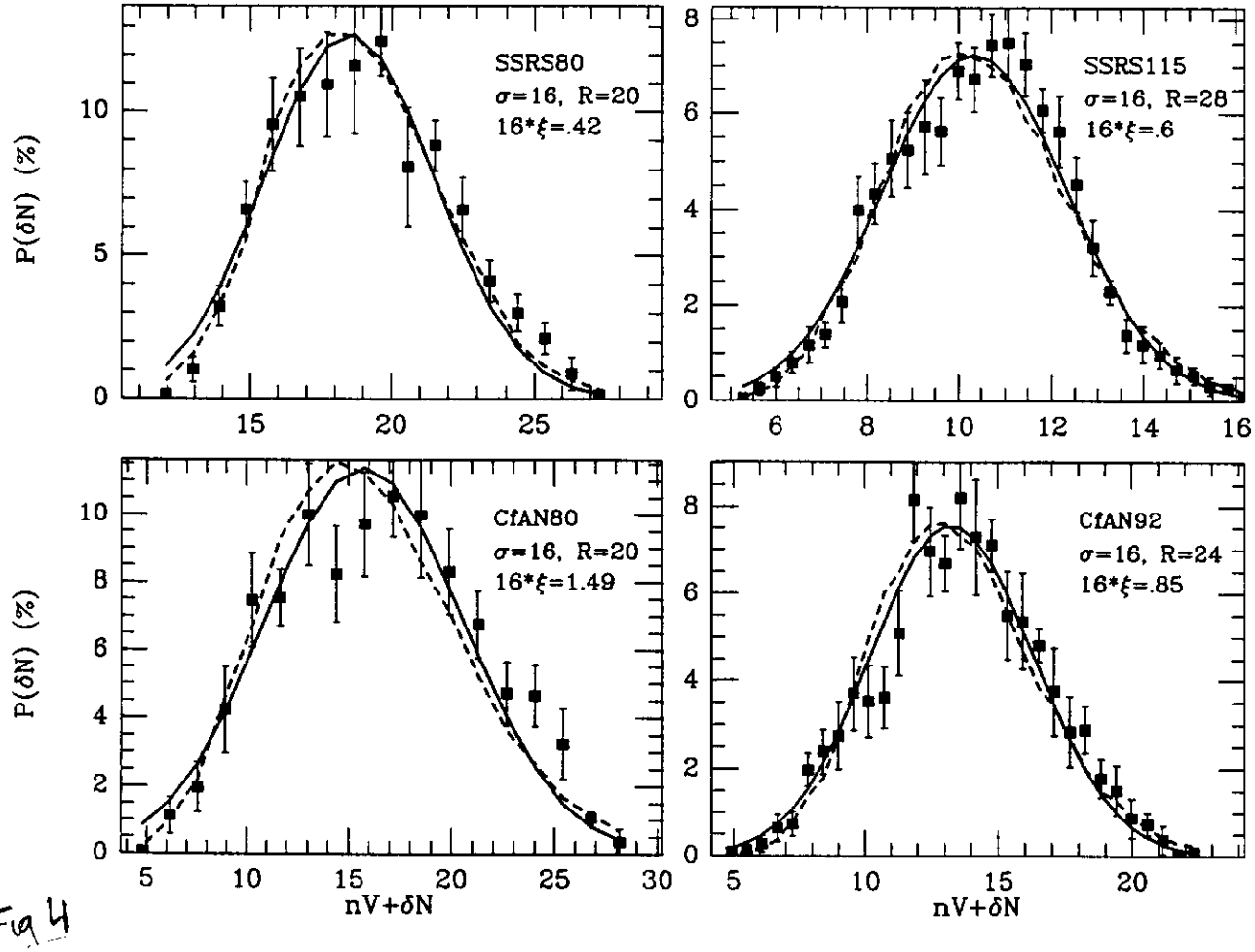
SSRS80
σ=16, R=20
16*ξ=.42
SSRS115
σ=16, R=28
16*ξ=.6
CfAN80
σ=16, R=20
16*ξ=1.49
CfAN92
σ=16, R=24
16*ξ=.85
P(δN) (%)
nV+δN

Fig 4

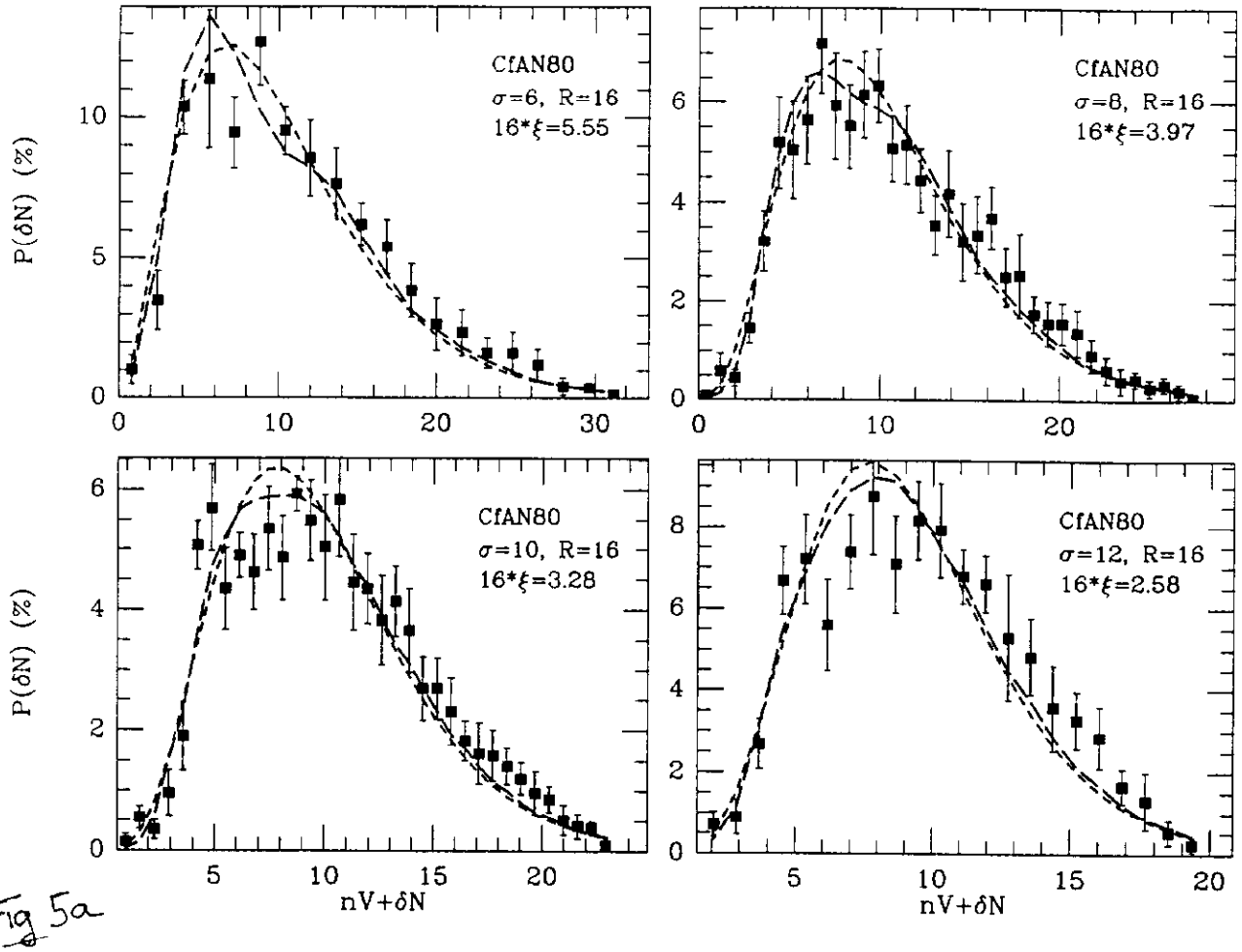
CfAN80
σ=6, R=16
16*ξ=5.55
CfAN80
σ=8, R=16
16*ξ=3.97
CfAN80
σ=10, R=16
16*ξ=3.28
CfAN80
σ=12, R=16
16*ξ=2.58
P(δN) (%)
nV+δN

Fig 5a

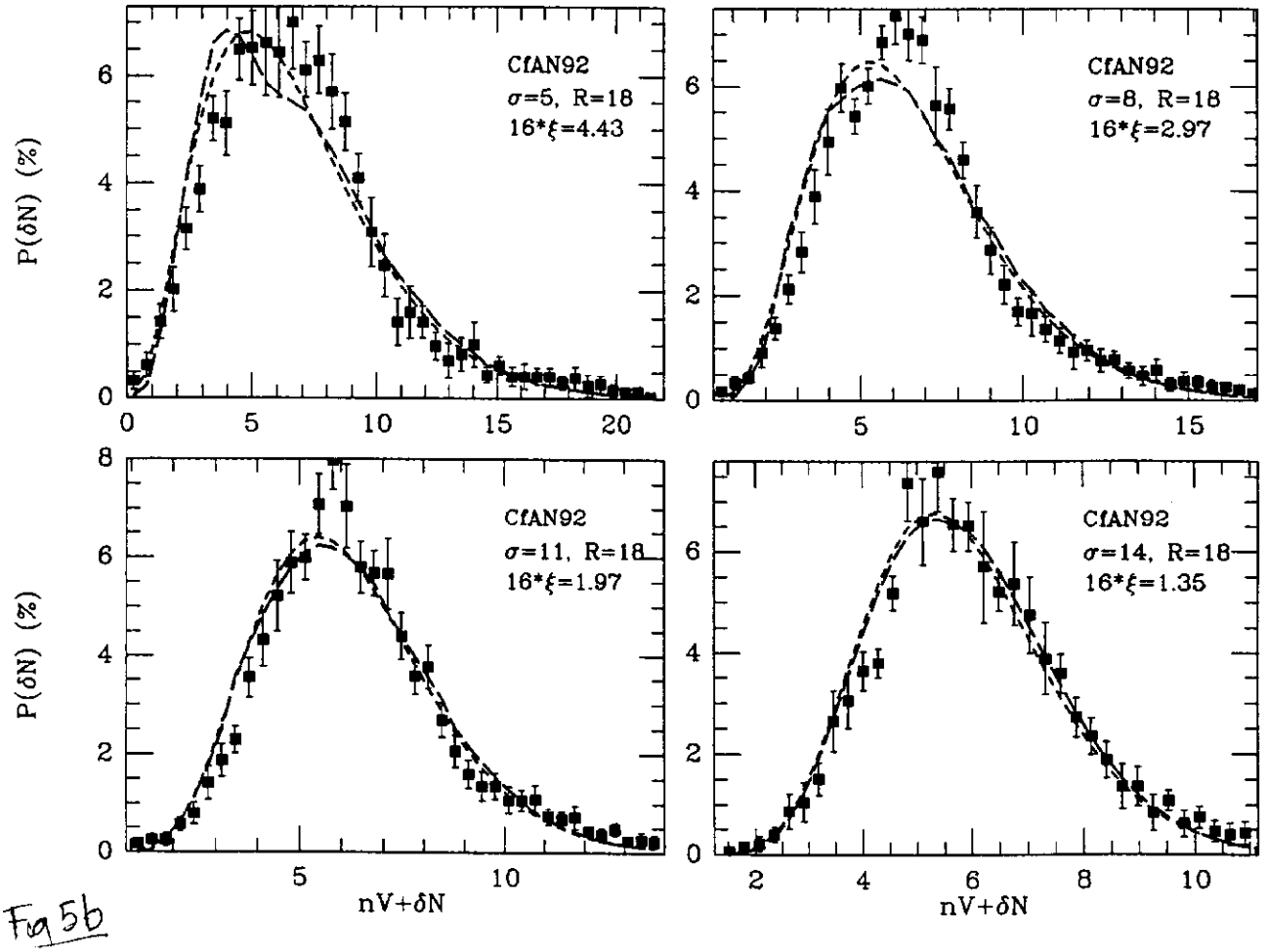
CfAN92
σ=5, R=18
16*ξ=4.43
CfAN92
σ=8, R=18
16*ξ=2.97
CfAN92
σ=11, R=18
16*ξ=1.97
CfAN92
σ=14, R=18
16*ξ=1.35
P(δN) (%)
nV+δN

Fig 5b

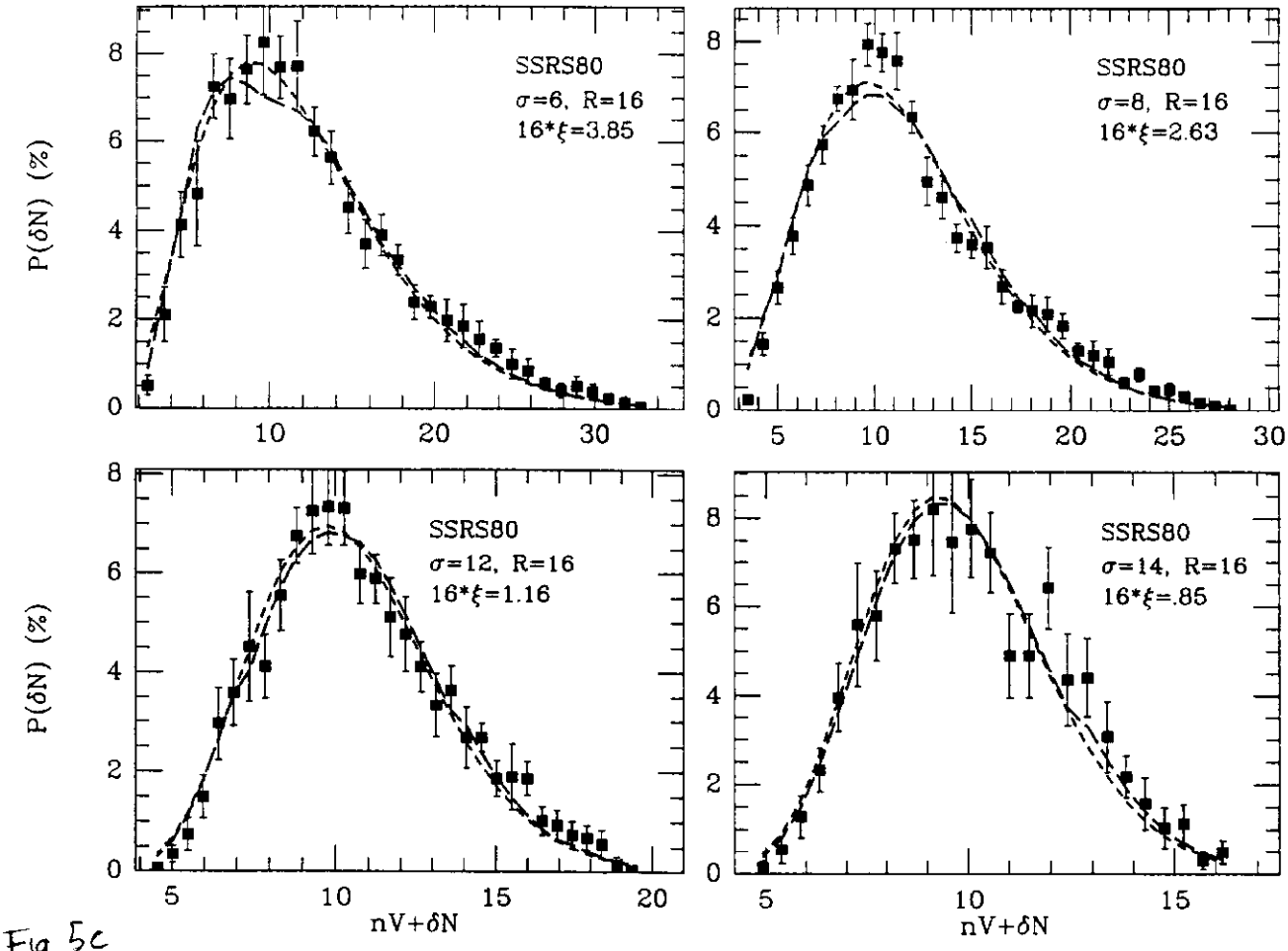
SSRS80
σ=6, R=16
16*ξ=3.85
SSRS80
σ=8, R=16
16*ξ=2.63
SSRS80
σ=12, R=16
16*ξ=1.16
SSRS80
σ=14, R=16
16*ξ=.85
P(δN) (%)
nV+δN

Fig 5c

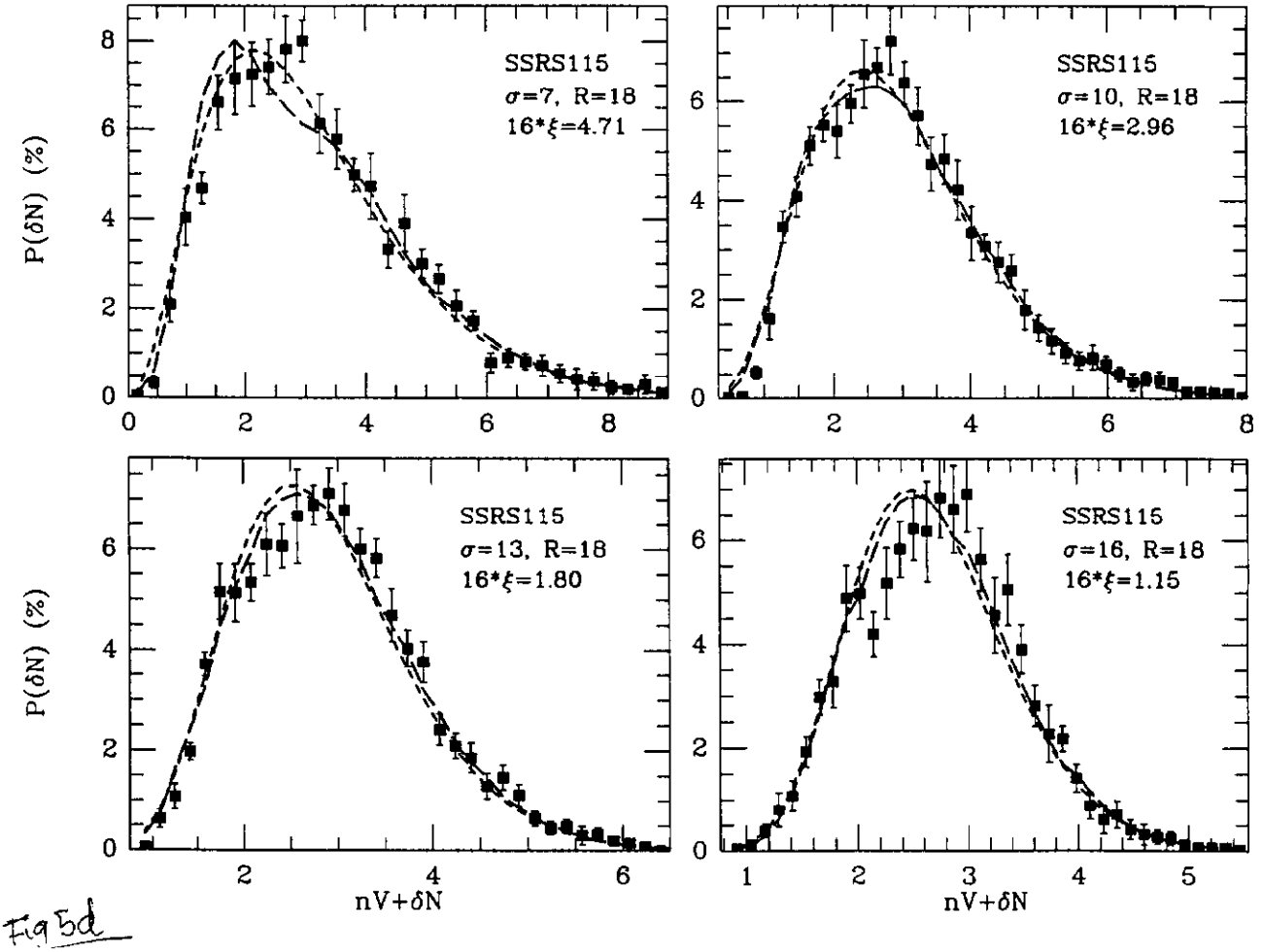
SSRS115
σ=7, R=18
16*ξ=4.71
SSRS115
σ=10, R=18
16*ξ=2.96
SSRS115
σ=13, R=18
16*ξ=1.80
SSRS115
σ=16, R=18
16*ξ=1.15
P(δN) (%)
nV+δN

Fig 5d

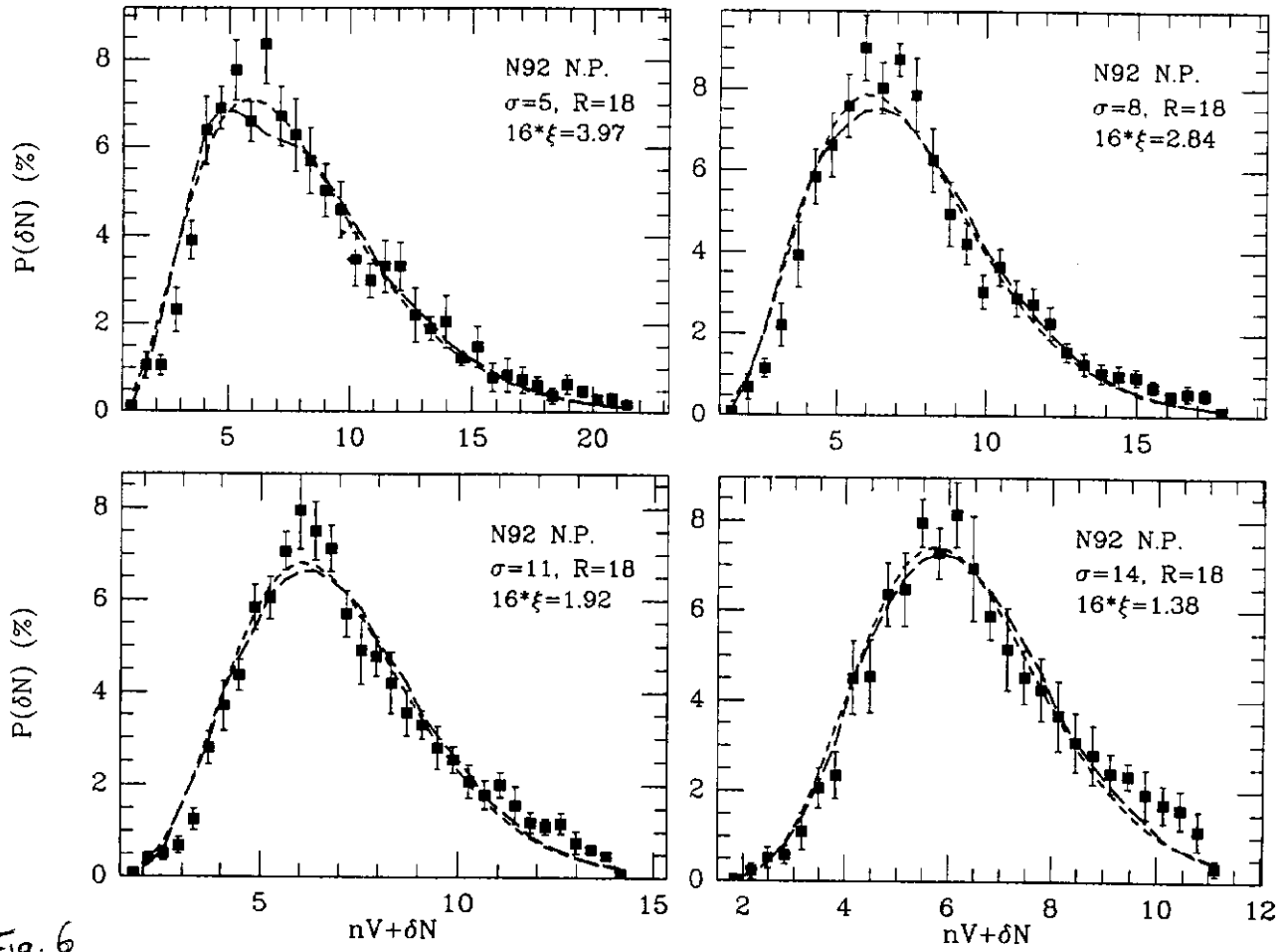
N92 N.P.
σ=5, R=18
16*ξ=3.97
N92 N.P.
σ=8, R=18
16*ξ=2.84
N92 N.P.
σ=11, R=18
16*ξ=1.92
N92 N.P.
σ=14, R=18
16*ξ=1.38
P(δN) (%)
P(δN) (%)
nV+δN
nV+δN

Fig. 6

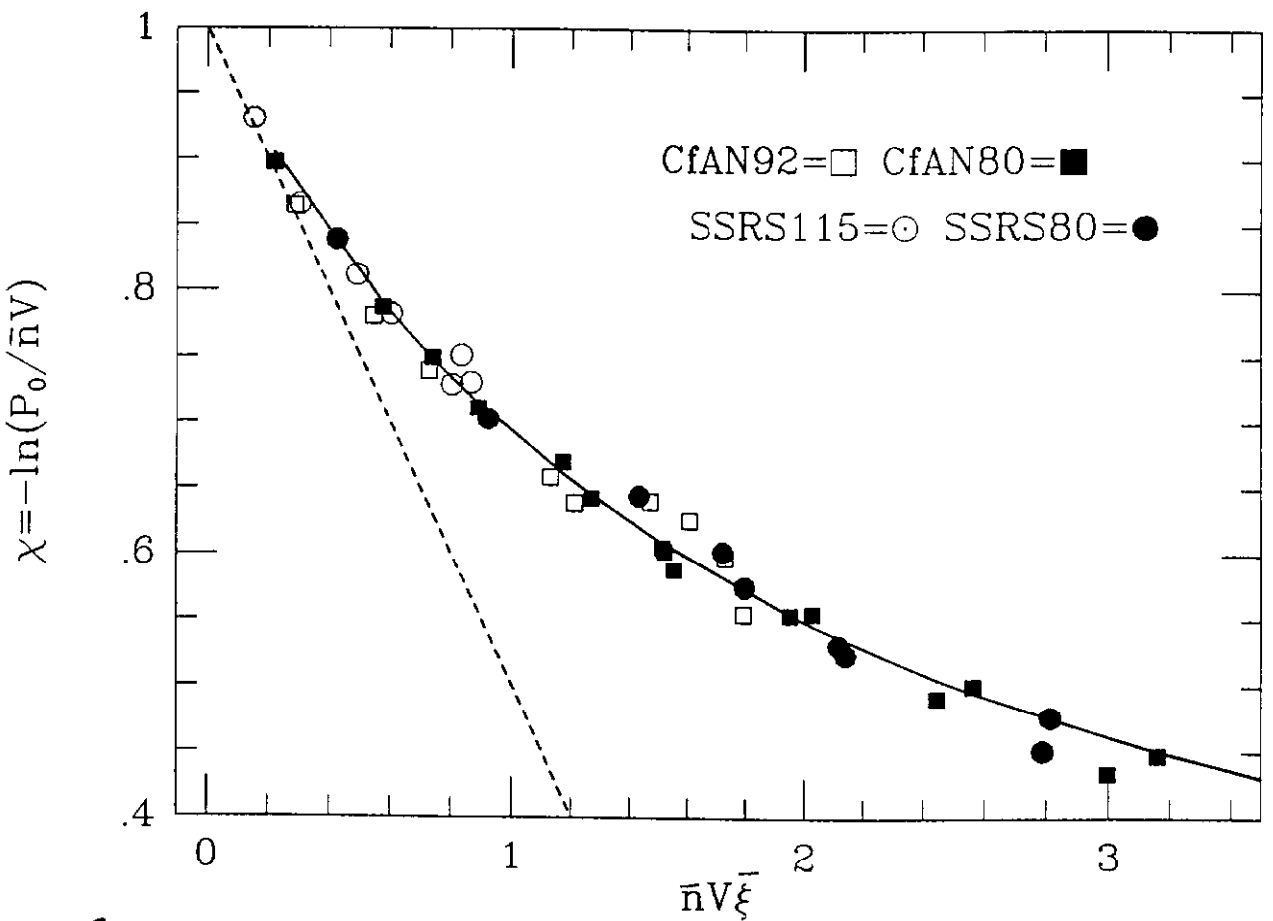
CfAN92=□ CfAN80=■
SSRS115=⊙ SSRS80=●
χ=−ln(P₀/n̄V)
n̄Vξ̄

Fig. 7