Fermi National Accelerator Laboratory

FERMILAB-Pub-91/175-A
June 1991

Void probability as a function of the void's shape and scale-invariant models

E. ELIZALDE
Department E.C.M., Faculty of Physics,
University of Barcelona, Diagonal 647,
E-08028 Barcelona, Spain

E. GAZTANAGA
NASA/Fermilab Astrophysics Center
Fermi National Accelerator Laboratory,
P.O. Box 500, Batavia, IL 60510, USA

May 1991

Abstract

The dependence of counts in cells on the shape of the cell for the large-scale galaxy distribution is studied. A very concrete prediction can be done concerning the void distribution for scale invariant models. The prediction is tested on a sample of the CfA catalog, and good agreement is found. It is observed that the probability of voids is bigger for spherical cells than for elongated ones, whereas the probability of a cell to be occupied is bigger for some elongated cells. A phenomenological scale-invariant model for the observed distribution of the counts in cells —an extension of the negative binomial distribution— is presented in order to illustrate how this dependence can be quantitatively determined. An original, intuitive derivation of this model is presented.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

1 Introduction

One of the standard approaches to the study of the 3-D space distributions of galaxies over distances of tens of Mpc has been the consideration of the two-point correlation function, $\xi(r)$, which can be derived both from observational data, angular (e.g. Groth and Peebles 1977) or redshift (e.g. Davis and Peebles 1983) catalogs of galaxy positions, and from specific cosmological scenarios: ξ is directly related to the power spectrum of fluctuations (see e.g. Peebles 1980). Some striking large-scale observational results concerning big voids (Kirschner 1981, Bothun et al. 1986), 'bubbles' (de Lapparent, Geller and Huchra 1986), 'great walls' (Geller and Huchra 1989) or 'sponge-like' topologies (Gott et al. 1989), have also proven to be relevant differential criteria to study this structure and do not seem to be trivially related with ξ —but rather with a complicated integration of all N-point correlation functions. It is then important to address the question of extracting statistical information from the higher order correlation functions.

Counts in cells, the probabilities P_i to have i galaxies inside a randomly chosen cell of volume V, and in particular the void probability P_0, are related with higher order correlations —as is clear e.g. from (6) below— and have received increasing attention as a tool to study 3-D redshift catalogs from data, and to compare them with simulations or theoretical predictions (see e.g. Fall et. al 1976, White 1979, Sharp 1981, Fry 1984, Ryden and Turner 1984, Saslaw and Hamilton 1984, Fry 1985, Hamilton 1985, Schaeffer 1985, Fry 1986, Otto et al. 1986, Bouchet and Lachièze-Rey 1986, White et. al 1987, Maurogordato and Lachièze-Rey 1987, Mellot 1987, Balian and Schaeffer 1988, Elizalde and Gaztanaga 1988, Fry et. al. 1989, Balian and Schaeffer 1989, Elizalde and Gaztanaga 1990, Maurogordato and Lachièze-Rey 1991, and references therein). The void probability P_0 is of special interest, since it generates all the other counts in cells P_i (see White 1979), although is not trivially related to the observed big voids, because they are not statistically significant (see Otto et al. 1986). The void probability contains higher-order correlation functions, generates counts in cells and provides an easy way to test scale-invariant models. We would like to study in this paper the dependence of the void probability and of the other counts in cells on the shape of the cell. This dependence might be appropriate to address questions such as whether elongated or spherical cell occupancy is statistically more significant. Furthermore, the analysis of the shape dependence might be useful when cell counts must be binned in non-spherical cells for practical purposes.

We shall concentrate on a phenomenological scale invariant model for counts in

cells: an extension of the negative binomial distribution. It is very simple and seems to reproduce fairly well the observed redshift distribution. We first present (in Section 2) an original and intuitive derivation of this distribution. In Sections 3 and 4 we compare the model with a sample from the CfA redshift catalog. We will argue in the discussion (Section 5) that the analysis presented below can be applied to any scale invariant model. For scale invariant models the shape dependence will be proven to be fixed by the volume dependence.

2 The negative binomial distribution

A continuous generalization of the negative binomial distribution, see (17) below, which was originally applied to hadron multiplicities at high-energy colliders (Carruthers and Shih 1983), has been used by Fry (1986) and Fry et al. (1989) as a phenomenological illustration of a model with a scaling relation on N-point correlation functions. This scaling property seems to be related with the so-called hierarchical universes. Although no physical or intuitive explanation have yet been given, the negative binomial distribution does provide a fair agreement with the observational distributions (Fry et. al 1989) of the Giovanelli-Haynes (1985) catalogue over Pisces-Perseus. As pointed out by Fry et al. (1989), the agreement —as shown in the scaling function χ— is not perfectly accurate but can be used as a first analytical approximation to the observed galaxy distribution. It is not clear to us whether the small, but systematic, discrepancies between the model and the observational counts correspond to real differences, since systematic errors coming from peculiar velocities have not yet been estimated. It is well known that peculiar velocities produce spurious effects on the estimate of the two point correlation function ξ from redshift catalogs (e.g. Davis and Peebles 1983, de Lapparent, Geller and Huchra 1988). This, by itself, affects the analysis by Fry et al. (1989), because the scaling function χ is studied as a function of ξ, which is directly extracted from data, but is not corrected from peculiar velocities. One might also expect further distortions, coming from peculiar velocities, on higher order correlations, and thus on P_0. These effects are by no means small, specially when samples are close to a big cluster (like Pisces-Perseus), since they can distort small scale into large scale statistics. In fact, the anisotropies presented in Section 4, below, might be caused by peculiar velocity distortions.

Balian and Schaeffer (1989) compiled different values for P_0 obtained from different catalogs, including the ones from Fry et al. (1989) of Pisces-Perseus, and the ones

from Maurogordato and Lachièze-Rey 1987 of the CfA redshift catalog (Huchra et al 1983), showing that the scaling function, $\chi = -(\log P_0)/nV$, with n the density and V the cell volume, can be fitted to a power law: $\chi \sim (nV\widehat{\xi})^{-\omega}$, being $\widehat{\xi}$ an average over ξ [see (19) below] and ω a parameter fitted about $\omega \sim .7$. This same compilation also fits the negative binomial expression: $\chi = [\log(1 + nv\widehat{\xi})]/nv\widehat{\xi}$ which has no free parameter. As pointed out by Balian and Schaeffer (1988 and 1989) the agreement[1] obtained for different catalogs and samples, with different luminosities and densities, is highly non-trivial and provides a good verification of scale-invariant models.

We have independently found a good agreement with the negative binomial distribution, which we called the quasi-Poisson model, both in an analysis over the CfA2 slice sample from the Lapparent, Geller and Huchra (1986) and Huchra et al. (1990) redshift catalog (Elizalde and Gaztanaga 1988) and over a sample from the Huchra et al. (1983) CfA1 redshift catalog (Elizalde and Gaztanaga 1990).

The original discrete negative binomial distribution is well known from standard statistics (see e.g. Eadie et al. 1971). It accounts for the probability, P_i, of the number i of trials necessary for m successes to occur, the events being independent and having each a probability p of success:

$$P_i = \binom{i-1}{m-1} p^m (1-p)^{i-m}, \tag{1}$$

with $i > m$. If we use a new variable, defined as $s \equiv i - m$, we have the (also common) expression

$$P_s = \binom{m+s-1}{s} p^m (1-p)^s. \tag{2}$$

with $s = 0, 1, ...$ It is this last form the one that leads to the distribution under study when generalized to continuous values of m. Indeed for $m = 1/\widehat{\xi}$ and $p = 1/(1 + nV\widehat{\xi})$ it is easy to reproduce formula (28) of Fry (1986), or the equivalent expression (17) below. After performing these changes one completely looses the original simple interpretation of the discrete negative binomial distribution. In order to recover such an interpretation we summarize below our original and independent derivation of the binomial distribution, which was constructed from a simple and intuitive statistical model (we called it quasi-Poisson) for the box distribution of a sample of galaxies. It has not been until recently that we have realized that our quasi-Poisson model has the same distribution of counts in cells as the binomial distribution (2). The quasi-Poisson model is, however, much more concrete, because we also have explicit expressions for the distribution in boxes, the variance and the higher-order momentums corresponding to the cell counts. In Elizalde and Gaztanaga (1990) a different derivation of the

[1] In particular, the fact that all data yield χ as a function of $nV\widehat{\xi}$ only.

model was presented. There, the quasi-Poisson distribution was obtained directly from an initial guess for the configurational probability, $F_N \sim \prod[1+\xi(i,j)]$, the same as expression (15) of Politzer and Wise (1984) for bias statistics (Bardeen et al. 1986). This alternative derivation was only valid under a very restrictive, low-density condition, where the cell occupancy is strictly less or equal than one galaxy per cell.

2.1 Counts in cells

Let us first introduce some concepts and notations. The problem of characterizing the large-scale structure of the universe can be formally solved, statistically, from the knowledge of the (configurational) probability of having a certain spatial point distribution $r_1,...,r_N$. This can be shown to be equivalent to the knowledge of *all* the N-point correlation functions, $\xi_N(r_1,...,r_N)$ (Peebles 1980), which are direct observational quantities, although impossible to compute in practice for $N > 3$. In general, evaluations are restricted to $N = 2$, where there are already big uncertainties and, eventually, they reach $N = 3$ (Groth and Peebles 1977) or N=4 (Fry and Peebles 1978), which are known to be very elusive correlation functions, specially for redshift surveys. The correlation functions can be also related with the counts in cells $P_i(V_c)$, the probabilities of having i galaxies inside an arbitrary cell of volume V_c. These probabilities are easy-to-measure observable quantities that contain statistical information about the higher-order correlation functions. The connection is performed by introducing a generating function

$$G(\lambda) \equiv \sum_i P_i \lambda^i, \tag{3}$$

which can be shown to be (see e.g. White 1979, Fry 1985, Fry 1986, Otto et al. 1986, Balian and Schaeffer 1988):

$$G(\lambda) = \exp\left[\sum_{N=1}^{\infty} \frac{n^N(\lambda-1)^N}{N!} \int_{V_c} dr_1 ... \int_{V_c} dr_N \xi_N(r_1,...,r_N)\right], \tag{4}$$

where n is the system density, dr is the differential of volume and V_c is the cell volume. Moreover, $\xi_1 = 1$. By taking convenient derivatives of $G(\lambda)$ and by evaluating them at $\lambda = 1$, one can easily find the relations between the P_i and the ξ_N. For instance, the second derivate gives

$$\sum_i i^2 P_i = (nV_c)^2 + nV_c + n^2 \int_{V_c}\int_{V_c} \xi_2(r_1,r_2) dr_1 dr_2, \tag{5}$$

which is the well known result that the two-point correlation function, ξ for brief, gives the system number density fluctuations. It is also interesting to notice that for $\lambda = 0$ we have

$$P_0 = \exp\left[\sum_{N=1}^{\infty} \frac{n^N(-1)^N}{N!} \int_{V_c} dr_1 ... \int_{V_c} dr_N \xi_N(r_1, ..., r_N)\right], \tag{6}$$

so that P_0 depends, in principle, on all order correlation functions. Notice in particular that for an uncorrelated system all the ξ_N are zero (except for ξ_1), and we get the Poisson result $\exp(-nV_c)$. We see also that P_0 is obtained in terms of a complicated integration of all N-point correlation functions.

2.2 Distribution in boxes

To introduce the binomial model let us consider a more restricted characterization of a point distribution. Divide the sample of μ points into m identical cells of volume $V_c = V/m$ and of a given shape, and consider the probability $P_\mu(k_1, ..., k_m)$ of having k_1 particles in cell 1, k_2 particles in cell 2, ..., and k_m particles in cell m. For the uniform distribution, where any point has the same probability of being in any cell, the probability of the configuration $k_1, ..., k_l - 1, ..., k_m$ is related with the previous one by

$$P_\mu(k_1, ..., k_m; N) = \frac{1}{m} \sum_{l=1}^{m} P_{\mu-1}(k_1, ..., k_l - 1, ..., k_m; N). \tag{7}$$

This gives the uniform distribution; for $\mu = N$,

$$P_N(k_1, ..., k_m; N) = \frac{m!}{k_1!...k_m!} \prod_{l=1}^{m} \frac{1}{m^{k_l}} = \frac{m!}{k_1!...k_m!} \left(\frac{1}{m}\right)^N. \tag{8}$$

What we have called the quasi-Poisson distribution is defined by considering the probability that a point belongs to a particular cell to be proportional to the number of points which are already occupying this cell. The proportionality constant will be called g. It may depend on the volume and on the shape of the cell, and can be interpreted as a kind of attraction/repulsion parameter. The value $g = 0$ represents no interaction and reproduces the uniform distribution. With these considerations, the recurrence relation corresponding to (7) is in this case

$$P_\mu(k_1, ..., k_m; N) = C \sum_{l=1}^{m} P_{\mu-1}(k_1, ..., k_l - 1, ..., k_m; N)\,[1 + (k_l - 1)g], \tag{9}$$

with C a normalization constant that can be easily computed, what leads to

$$P_\mu(k_1,...,k_m;N) = \sum_{l=1}^{m} P_{\mu-1}(k_1,...,k_l-1,...,k_m;N)\frac{1+(k_l-1)g}{m+(N-m)g}. \qquad (10)$$

2.3 Explicit solution for the distribution

We have been able to solve the recurrence (10) analytically. In Appendix A.1 we show that

$$P_N(k_1,...,k_m;N) = \frac{m!}{k_1!...k_m!}\prod_{l=1}^{m}\prod_{j=1}^{k_l}\frac{1+(l-1)g}{m+g(\sum_{i=1}^{j-1}k_i+j-1)}. \qquad (11)$$

This result reproduces the uniform distribution (8) when $g = 0$. Once g is known, all the information about the system in the configuration space is obtained. We can compute, in particular, the expectation value for the number of particles in a cell and also its fluctuation. This gives a directly observable magnitude which can therefore be checked experimentally: the counts in cells. With the aim of computing this average, let us first elaborate on (11) in order to obtain from it the corresponding distributions of the usual random variables x_i, *the number of cells with i points inside.* In Appendix A.1 it is found that

$$P_N(x_0,...,x_N) = \frac{m!\ N!\ \Gamma(m/g)\ \prod_{i=2}^{N}[\Gamma(1/g+i)]^{x_i}}{\Gamma(m/g+N)g^{m-x_0}[\Gamma(1/g+1)]^{m-(x_0+x_1)}\prod_{i=1}^{N}x_i!\ \prod_{i=1}^{N}(i!)^{x_i}}. \qquad (12)$$

Out of this expression it is possible to calculate the mean value $E(x_i)$ and the variance $V(x_i)$ associated with this random variable. This is done by adding up the corresponding contributions from (12)

$$E(x_i) = \sum_{(\sum x_j=1)\ (\sum jx_j=N)} x_j P_N(x_0,...,x_j,...,x_N). \qquad (13)$$

In Appendix A.2 this is seen to give

$$E(x_i) = A(m,N,i), \qquad (14)$$

where

$$A(m,N,i) \equiv \binom{N}{i}\frac{N\Gamma(1/g+i)\Gamma(m/g)\Gamma(\frac{m-1}{g}+N-i)}{g\Gamma(1/g+1)\Gamma(m/g+N)\Gamma(\frac{N-1}{g})} \qquad (15)$$

is a convenient expression introduced in order to express the different characteristics of the probability distribution in a simple way. For instance, its variance can be written as

$$V(x_i) = A(m,N,i)[1 + A(m-1,N-i,i) - A(m,N,i)]. \tag{16}$$

These results can be tested in the limit $g \to 0$, where they actually reproduce the results for the binomial distribution —since $g = 0$ means no interaction.

If we now take the *continuum limit* for the probability of finding i points in a certain cell, $P_i = E(x_i)/m$, we get

$$P_i(V_c) = \frac{(nV_c)^i}{i!}(1 + gnV_c)^{-1/g-i}\prod_{j=1}^{i-1}(1 + gj), \tag{17}$$

which also reproduces the Poisson distribution for $g \to 0$. This is precisely the so-called negative binomial distribution, presented in Section 2.

The distribution (17) above can also be tested by calculating $\sum_i i^2 P_i$ directly, which in Appendix A.3 is shown to yield

$$\sum_i i^2 P_i = (nV_c)^2 + nV_c + g(nV_c)^2. \tag{18}$$

By comparing with (5), the value of g is completely fixed:

$$g(V_c) = \frac{1}{V_c^2}\int_{V_c}\int_{V_c}\xi(r_1,r_2)dr_1dr_2, \tag{19}$$

which happens to be a well-known quantity, usually referred to in the literature as $\widehat{\xi}$ (cf. e.g. Peebles 1980, Fry 1986, Balian and Schaeffer 1988) and is also related to the known J_3: $\widehat{\xi} \sim J_3/V_c$. Notice that one of the integrals in (19) can be readily performed using the fact that $\xi(r_1,r_2) = \xi(r_1 - r_2)$, but the second one is restricted to some different contour and will depend on the shape of the cell.

It is important to observe that, in addition to the new interpretation of the binomial distribution, we have been able to derive a very concrete model for the box distribution, $P_N(k_1,k_2,...,k_m;N)$, eq. (11), a discrete version of the counts in cells for a finite sample, eqs. (14) and (15), and its variance, eq. (16).

3 Comparison with data

We now turn to the crucial point of comparing this information with real data. In doing so, we face a rather strong imposition from the theoretical model, namely that

the expression of counts in cells (17) represents a very concrete prediction *without any free parameter to be fitted*, provided we compute the fluctuation $\widehat{\xi}$ directly from the data.

We have checked counts in cells in a redshift survey using a volume-limited sample from the completed *North Zwicky Forty* CfA catalog, $m_B < 14.5$ (Huchra et al. 1983). We will focus on a sample with $M_B < -20$ and $v < 8000$ Km/s. Cell counts P_i have been computed by placing a random cell on the sample and by counting the number of objects inside the cell. This has been repeated a number of times for 40,000 to 4,000,000 cells, depending on the statistical uncertainties, and then we have recounted the number of cells having a given number, i, of galaxies inside. This provides our result for P_i. Errors are estimated by going through different independent samples.

Boundary problems have been faced up by choosing only cells which fell completely inside the sample. This has introduced a bias, because some zones of the sample (the central ones) are overweighted. The effect is bigger for bigger cells (for spherical 10 Mpc cells the contour zone which is underweighted represents 80% of the sample) and thus values for big cells contain inescapable uncertainties.

We have used the *standard* known value for the correlation function: $\xi(r) \simeq (a/r)^\gamma$, with $\gamma \simeq 1.8$ and $a \simeq 5$ Mpc ($H_0 = 100$ Km s^{-1}Mpc^{-1}) for the 14.5 m_B CfA catalog (Davis and Peebles 1983), to estimate $g(V_c)$ (or $\widehat{\xi}$) in (19). The two-variable probability observational distribution of counts in cells $P_i(R)$ is automatically fitted with the negative binomial distribution (17) for all significant values of R and i. The fitted curve is shown in Fig. 1, where for the sake of clarity we have plotted observational values only for voids, P_0, but for different cell sizes, compared with theoretical values for different γ and a corresponding to different uncertainties in the knowledge of ξ. The average fitted values, that is $a = 5$ Mpc and $\gamma = 1.8$, perfectly match with the observations.

4 The shape dependence and anisotropies

An interesting characteristic of counts in cells is their ability to provide a good measure of some statistical shape features. When the shape of the cell is no longer an sphere but rather e.g. an ellipsoid of a given fixed volume but of varying shape, information about the way galaxies choose to cluster can be easily extracted. We can, for simplicity, characterize an ellipsoid by its three half-axes: a, b, c with the restriction: $abc = R^3$, and take one of them fixed and the other two: $\sqrt{s}R$ and $R/\sqrt{s}$, so that the excentricity, s, and the radius, R, characterize the shape of the cell $V_c(R, s)$. If clusters or groups

tend to appear with elongated shapes, counts in cells $P_i(R,s)$ will be increasing with s ($s>1$) —up to a maximum value of s— and if voids are statistically more probable to be spherical, $P_0(R,s)$ will be decreasing with s. Since the probability, P_i, must be normalized when summed over all values of i, there is a relation between these last effects: if the void probability decreases with increasing s, the occupancy probability must increase.

For the negative-binomial model the shape dependence comes from the integration of ξ in (19) since, for a given volume, different shapes will contribute to $\widehat{\xi}$ with different numerical factors. Formally, $\widehat{\xi}$ is then a function of R and s, $\widehat{\xi}(R,s)$. We can introduce a shape factor $\eta(R,s)$,

$$\widehat{\xi}(R,s)=\eta(R,s)\widehat{\xi}(R,1) \tag{20}$$

so that, $\eta(R,1)=1$. For ellipsoids this factor is a decreasing function of the excentricity, s, as can be seen in Fig. 2, where the values of $\widehat{\xi}$ are obtained as a function of s, by numerical integration of (19), using a power law for ξ. The corresponding effect on P_i is shown in Fig. 3, where the values of P_i coming from eq. (17) are shown as a function of s, for a given V_c. We see that there is a bigger probability to find spherical voids and elongated clusters than elongated voids and spherical clusters.

Let us now compare, as in the previous section, these predictions with observational data, just by performing the same counts for elongated cells. We find the striking result that there is a big anisotropy in the catalog, because $P_i(R,s)$ depends strongly on which one of the axes is chosen to be fixed, or, in other words, on the precise plane where we perform the elongation, and also on the orientation of the ellipsoid's section inside this plane. In Fig. 4 we have plotted different probabilities for P_0, extracted from the catalog, varying with s and for six differently oriented ellipsoids —corresponding to three different frames of orthogonal directions for the fixed axes or plane. Inside each plane we choose two distinct perpendicular orientations.

Very different values are found for the various orientations; the differences between them increase with excentricity, that is, with the main scale involved. However, some warnings should be made concerning this anisotropy. As pointed out before, there exist inescapable uncertainties due to the contour, and this effect also increases with the scale, just as in our anisotropy. The contour forces us to center the cells in a certain inner region which represents only a 10% of the total volume, for R=5 Mpc and s=5 (the main distance involved is 25 Mpc). Anisotropies in this central region, where the Coma cluster is located, highly contribute to the effect. There is another point to be taken into account, which is the fact that the contour is not symmetric and, thus, the mentioned central region is not the same for every direction. Peculiar

velocity distortions coming from the Coma cluster (the "finger of God effect") are, very possibly, responsible for this effect. Actually this anisotropy had already been observed (Elizalde and Gaztanaga 1988) in a slice catalog, almost included in the one we are dealing with: the first CfA2 $m_B = 15.5$ slice (de Lapparent, Geller and Huchra 1986, Huchra et al. 1990) which containts the Coma cluster.

We have performed estimations of counts in cells with different corrections in the redshifts, to take into account the local motion of our galaxy. No significant differences in the results have been found when data is corrected either from the MWB inertial frame, the local group frame or galactocentric velocities.

The above is a new result that can be perfectly measured quantitatively via the $P_i(R, s, \alpha)$, where α stands for the orientation angles, but interferes somehow with our original purpose of checking the dependence with the shape. Although a detailed analysis of the anisotropy dependence has not yet been performed, we can average over all the orientations in order to obtain an estimation for the $P_i(R, s)$. The real average would probably have some weight which we do not know and thus, we can only expect to find some fit within the standard deviations.

In Fig. 6 error bars are compared with predictions from the model using the corresponding average on (19) and good overall agreement is found, within the error bars, *again without having to adjust any parameter.*

5 Discussion and conclusions

As mentioned in the introduction, our analysis of the shape dependence can be extended to very different models, being the negative binomial distribution an appropriate phenomenological tool that can be used to illustrate it.

In general, the dependence of P_0 on the shape of the cell can be easily obtained for any scale-invariant model, for which we already know the volume dependence. For the scale invariant models of the galaxy distribution, it is already known (Fry 1986, Balian and Schaeffer 1988) that $\chi = -(\log P_0)/nV$ has to be a function of the combined variable $nV\widehat{\xi}$ alone. This statement is true for a cell of any shape, so that the only way the shape can affect the probability P_0 is through $\widehat{\xi}$. We can use, for example, elipsoids as cells and characterize their shape, for a fixed volume, by the value of the excentricity s —for spheres $s = 1$. The dependence $\widehat{\xi}(s)$ is obtained, from a known correlation function, ξ, by performing the integration in (19), which will introduce and additional factor, $\eta(s)$, coming just from the shape, so that, $\widehat{\xi}(s) = \eta(s)\widehat{\xi}(1)$ (see

eq. (20)). Thus, given a sample where the correlation function and the functional dependence of χ (or P_0) on nV are known for a given shape, e.g. for $s = 1$, one can easily predict the values corresponding to other shapes, since this just requires the appropriate scaling of nV. Formally

$$\chi[nV\widehat{\xi}(s)] = \chi[nV'\widehat{\xi}(1)], \tag{21}$$

where $nV' = nV\eta(s)$. In the previous sections it has been illustrated how to do this prediction for a concrete analytical distribution, but our whole study is completely general, and can be reproduced, numerically, without referring to the binomial distribution.

It is concluded then, that the shape dependence of P_0 (or χ) can be predicted from the values which are obtained for a given fixed shape, provided that the distribution is scale invariant. Although the analysis presented here is quite restricted (only one sample has been considered) and affected by large anisotropies (coming probably from the Coma cluster), a good qualitative agreement has been observed. We propose this type of analysis as a further way to test the functional dependence of χ on $nV\widehat{\xi}$, $\chi = \chi(nV\widehat{\xi})$ and, thus, to study the scale invariant characteristics of the galaxy distribution.

Acknowledgments

Financial support from Comisión Interministerial de Ciencia y Tecnologia (CICYT, Spain), Project No. AEN 90-0033, is gratefully acknowledged. This work was supported in part by DOE and NASA (grant NAGW-1340) at Fermilab, and by Fulbright/MEC (grant FU90 14594408).

A Appendix

A.1 Configurational probability into boxes

We would like first to show how to solve the recurrent relations in (10):

$$P_\mu(k_1,...,k_m;N) = \sum_{l=1}^{m} P_{\mu-1}(k_1,...,k_l-1,...,k_m;N)\frac{1+(k_l-1)g}{m+(\mu-m)g}. \tag{22}$$

This can be done by applying the recurrence several times, one after the other. It is convenient to realize at this stage that $\mu = \sum_{q=1}^{m} k_q$, since this is the total number of points to be distributed into the m cells, with occupation number $k_q, q = 1,...,M$. If we apply (22) two successive times, and take into account that μ must be replaced by $\mu - 1$ (since one of the μ total points is already fixed in k_l), we get

$$\begin{aligned} P_\mu(k_1,...,k_m;N) &= \sum_{i=1}^{m}\sum_{l=1}^{m} P_{\mu-1}(k_1,...,k_i-1,...,k_l-1,...,k_m;N) \\ &\frac{1+(k_i-1)g}{m-g+g\sum_{i=1}^{m}k_i}\ \frac{1+(k_l-1)g}{m-2g+g\sum_{l=1}^{m}k_l} \end{aligned} \tag{23}$$

Going on with the recurrence and re-expressing the summations in terms of products, we arrive to

$$P_N(k_1,...,k_m;N) = \frac{N!}{k_1!...k_m!}\prod_{j=1}^{m}\prod_{l=1}^{k_j}\frac{1+(l-1)g}{m+g(\sum_{i=1}^{j-1}k_i+l-1)}, \tag{24}$$

which is indeed eq. (11).

In terms of the variables x_i (number of cells which have i points inside), this expression can be further elaborated, in the form,

$$P_\mu(k_1,...,k_m;N) = \frac{N!}{(0!)^{x_0}(1!)^{x_1}...(N!)^{x_n}}\frac{\prod_{l=1}^{N-1}(1+lg)^{\sum_{j=l+1}^{N}x_j}}{\prod_{l=0}^{N-1}(m+lg)} \tag{25}$$

from which we get the probability density in terms of the x_i:

$$\begin{aligned} P(x_0,...,x_m;N) &= \frac{m!}{x_0!x_1!...x_N!}\ \frac{N!}{(0!)^{x_0}(1!)^{x_1}...(N!)^{x_n}} \\ &\frac{\Gamma(m/g)}{\Gamma(m/g+N)}\ \frac{\prod_{l=2}^{N}\Gamma(1/g+l)^{x_l}}{g^{m-x_0}[\Gamma(1/g+1)]^{m-(x_0+x_1)}} \end{aligned} \tag{26}$$

this is the same as eq. (12).

A.2 Expectation value and variance

Let us now calculate the mean value $E(x_i)$ and the variance $V(x_i)$:

$$\begin{aligned} E(x_i) &= \sum_{(\sum x_j=m)\ (\sum jx_j=N)} x_j P_N(x_0,...,x_j,...,x_N), \\ V(x_i) &= E(x_i^2) - E(x_i)^2, \end{aligned} \tag{27}$$

for expression (12). This is computed by noticing that

$$\sum_{(\sum x_j=m)\ (\sum jx_j=N)} P_N(x_0,...,x_j,...,x_N) = 1, \tag{28}$$

and that, for any $x_j \geq 1$,

$$\sum_{(\sum x_j=m-1)\ (\sum jx_j=N-i)} P_N(x_0,...,x_j-1,...,x_N) = 1. \tag{29}$$

Use of these expressions allows us to write

$$\begin{aligned} E(x_i) &= \sum_{(\sum x_j=m)(\sum jx_j=N)(x_i\geq 1)} x_i P(x_0,...,x_i,...,x_N) \\ &= \sum_{(\sum x_j=m)(\sum jx_j=N)(x_i\geq 1)} x_i P(x_0,...,x_i,...,x_{N-i},0,...,0), \end{aligned} \tag{30}$$

and replacing (12), we get

$$E(x_i) = \frac{mN!\Gamma(1/g+i)}{g(N-i)!i!\Gamma(1/g+1)} \sum \frac{(m-1)!}{x_0!...(x_i-1)!..x_{N-i}!}$$
$$\frac{(N-i)!}{(0!)^{x_0}...(i!)^{x_i-1}(N!)^{x_{N-i}}} \frac{\Gamma(m/g)}{\Gamma(m/g+N)} \frac{\prod_{i=2}^{N-i}\Gamma(1/g+i)^{x_i}}{g^{m-1-x_0}[\Gamma(1/g+1)]^{m-1-(x_0+x_1)}}. \tag{31}$$

By calling

$$\overline{x_i} \equiv x_i - 1, \tag{32}$$

and by writing the sum which appears in (31) as

$$\sum = \sum_{(x_0+..+\overline{x_i}+..+x_N=m-1)(x_1+..+i\overline{x_i}+..+x_N=N-i)} P(x_0,...,\overline{x_i},...,x_n) = 1. \tag{33}$$

we arrive at the very simple expression

$$E(x_i) = \frac{mN!\Gamma(1/g+i)\Gamma(m/g)\Gamma(\frac{m-1}{g}+N-i)}{g(N-i)!i!\Gamma(1/g+1)\Gamma(m/g+N)\Gamma(\frac{N-1}{g})}. \tag{34}$$

This can be written in the form

$$E(x_i) = A(m, N, i), \tag{35}$$

where

$$A(m, N, i) \equiv \binom{N}{i} \frac{\Gamma(1/g + i)\Gamma(m/g + 1)\Gamma(\frac{m-1}{g} + N - i)}{\Gamma(1/g + 1)\Gamma(m/g + N)\Gamma(\frac{N-1}{g})} \tag{36}$$

is a convenient expression introduced in order to write in a simple compact way the different characteristics of the probability distribution. A similar calculation for the variance yields

$$V(x_i) = A(m, N, i)[1 + A(m - 1, N - i, i) - A(m, N, i)]. \tag{37}$$

A.3 Fluctuations

We now want to obtain g in (17), by first calculating $\sum_i i^2 P_i$ and by comparing then with (5):

$$\sum_i i^2 P_i = (nV_c)^2 + nV_c + n^2 \int_{V_c} \int_{V_c} \xi_2 dr_1 dr_2. \tag{38}$$

We start from (17)

$$P_i = \frac{(nV_c)^i}{i!}(1 + gnV_c)^{-1/g-i} \prod_{j=1}^{i-1}(1 + g), \tag{39}$$

with

$$\sum_i P_i i = nV_c. \tag{40}$$

Differentiating (40) with respects to n, we get,

$$\sum_i \frac{dP_i}{dn} i = V_c. \tag{41}$$

On the other hand, by computing $\frac{dP_i}{dn}$ from (39)

$$\frac{dP_i}{dn} = i\frac{P_i}{nV_c} - i\frac{P_i g V_c}{(1 + gnV_c)}, \tag{42}$$

and substituting in (41), we get

$$\sum_i i^2 P_i = (nV_c)^2 + nV_c + n^2 V_c^2 g. \tag{43}$$

Finally, comparison of this equality with (38) yields the expression for g we were looking for

$$g = \frac{1}{V_c^2} \int_{V_c} \int_{V_c} \xi(r_1, r_2) dr_1 dr_2. \tag{44}$$

This is eq. (19).

References

[1] Balian, R., and Schaeffer, R.: 1988, Astron. Astrophys., **220**, 1.

[2] Balian, R., and Schaeffer, R.: 1989, Astron. Astrophys., **226**, 373.

[3] Bardeen, J.M., Bond, J.R., Kaiser, N., and Szalay, A.S.: 1986, *Ap.J.* **304**, 15.

[4] Bothun, G.D., Beers, T.C., Mould, J., and Huchra, J.P.: 1986, *Ap.J.* **308**, 510.

[5] Bouchet, F.R., and Lachièze-Rey, M.: 1987, Ap. J. Lett, **302**, L37.

[6] Davis, M., and Peebles P.J.E.: 1983, Ap. J., **267**, 465.

[7] de Lapparent, V., Geller, M.J., and Huchra, J.P.: 1988, Ap. J., **322**, 56.

[8] de Lapparent, V., Geller, M.J., and Huchra, J.P.: 1986, Ap. J. Lett., **302**, L1.

[9] Eadie, W.T., James, F.E., Prijard, D., Roos, M., and Sadoulet, B.: 1971, *Statistical methods in experimental physics*, North-Holland (Amsterdam).

[10] Elizalde, E., and Gaztanaga, E.: 1988, Phys. Lett. **129A**, 295. Addendum: 1989, Phys. Lett., **136A**, 513.

[11] Elizalde, E., and Gaztanaga, E.: 1990, Nucl. Phys. (Proc. Suppl.), **16**, 650. For a more detailed analysis see Gaztanaga, E.: 1989, Ph.D. Thesis, University of Barcelona (Barcelona, Spain).

[12] Fall, S.M., Geller, M.P., Jones, B.J.T., and White, S.D.M.: 1976, Ap. J., **205**, L121.

[13] Fry, J.N., and Peebles, P.J.E.: 1978, Ap. J., **221**, 19.

[14] Fry, J.N.: 1984, Ap. J., **277**, L5. Fry, J.N.: 1984, Ap. J., **279**, 499.

[15] Fry, J.N.: 1985, Ap. J., **289**, 10.

[16] Fry, J.N.: 1986, Ap. J., **306**, 358.

[17] Fry, J.N., Giovanelli, R., Haynes, M.P., Mellot, A.L., and Scherrer, R.J.: 1989, Ap. J., **340**, 11.

[18] Geller, M.J., and Huchra, J.P.: 1989, *Science* **246**, 897.

[19] Giovanelli, R., and Haynes, M.P.: 1985, Ap. J., **292**, 404.

[20] Gott III, J.R., Miller, J., Thuan T.X., Schneider, S.E., Weinberg D.H., Gammie, C., Polk, K., Vogeley, M., Jeffrey, S., Bhavsar, S.P., Mellot, A.L., Giovanelli, R., Haynes, M.P., Tully, R.B., and Hamilton A.J.S.: 1989, *Ap. J.* **340**, 625.

[21] Groth, E.J., and Peebles, P.J.E.: 1977, Ap. J., **217**, 385.

[22] Hamilton, A.J.S.: 1985, Ap. J., **292**, L35.

[23] Huchra, J., Davis, M., Latham, D., and Tonry, J.: 1983, Ap. J. Supplement Series, **52**, 89.

[24] Huchra, J., Geller, M.J., de Lapparent, V., and Corwin, H.G.: 1990, Ap. J. Supplement Series, **72**, 433.

[25] Kirshner, R.P., Oemler, A., Schechter, P.L., and Schectman, S.A.: 1981, Ap. J., **248**, L57.

[26] Maurogordato, S., and Lachièze-Rey, M.: 1987, Ap. J., **320**, 13.

[27] Maurogordato, S., Lachièze-Rey, M.: 1991, Ap. J., **369**, 30.

[28] Mellot, A.L.: 1986, Month. Not. Roy. Astron. Soc., **228**, 1001.

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

[30] Politzer, D.H., Wise, M.B.: 1984, Ap. J., **285**, L1.

[31] Peebles, P.J.E.: 1980, *The large scale structure of the universe*, Princeton University Press (Princeton).

[32] Ryden, B.S., and Turner, E.L.: 1984, Ap. J., **287**, L59.

[33] Saslaw, W.C., and Hamilton, A.J.S.: 1984, Ap. J., **276**, 13.

[34] Schaeffer, R.: 1984, Astron. Astrophys. (Lett.), **134**, L15.

[35] Schaeffer, R.: 1985, Astron. Astrophys. (Lett.), **144**, L1.

[36] Sharp, N.A.: 1981, Month. Not. Roy. Astron. Soc., **191**, 857.

[37] White, S.D.M.: 1979, Month. Not. Roy. Astron. Soc., **186**, 145.

[38] White, S.D.M., Frenk, C.S., Davis, M., and Efstathiou, G.: 1987, Ap.J.,**313**, 505.

Figure captions

Fig. 1: Values of P_0 (in %) for cells of different radii: $R = 1.5 - 7.5$ Mpc. Dashed circles represent observational values, the dashed line is the theoretical value for $a = 5$ Mpc and $\gamma = 1.8$, and the continuous upper and lower lines are the theoretical values for $a = 4$ Mpc and $\gamma = 1.6$, and $a = 8$ Mpc and $\gamma = 1.85$, respectively.

Fig. 2: Values of the averaged two-point correlation function, $\widehat{\xi}$, normalized to the value of $\widehat{\xi}$ for a sphere, are computed numerically for a power law correlation function, ξ, and plotted as a function of the excentricity of the cell, for a fixed volume, over which ξ is averaged.

Fig. 3: Values of counts in cells, P_i, for $i = 0, 1, 2, 3$, are plotted for the negative binomial model as a function of the excentricity of the cell for a fixed volume.

Fig. 4: Variation of P_0 (in %) with the excentricity, $s = 1-5$, with fixed $R = 5$ Mpc, for different orientations. Lines with equal type of dash represent two different orientations in the same plane. Different dashes correspond to three different orthogonal choices of the plane of elongation.

Fig. 5: Values of P_0 (in %) for different shapes of the cell, parametrized by its excentricity $s = 1 - 5$, being $R = 5$ Mpc. The error bars correspond to observed deviations with orientation. The continuous line corresponds to the theoretical prediction, with $a = 5$ Mpc and $\gamma = 1.8$.

Fig. 6: The same as in Fig. 5, for P_2.

Fig. 1

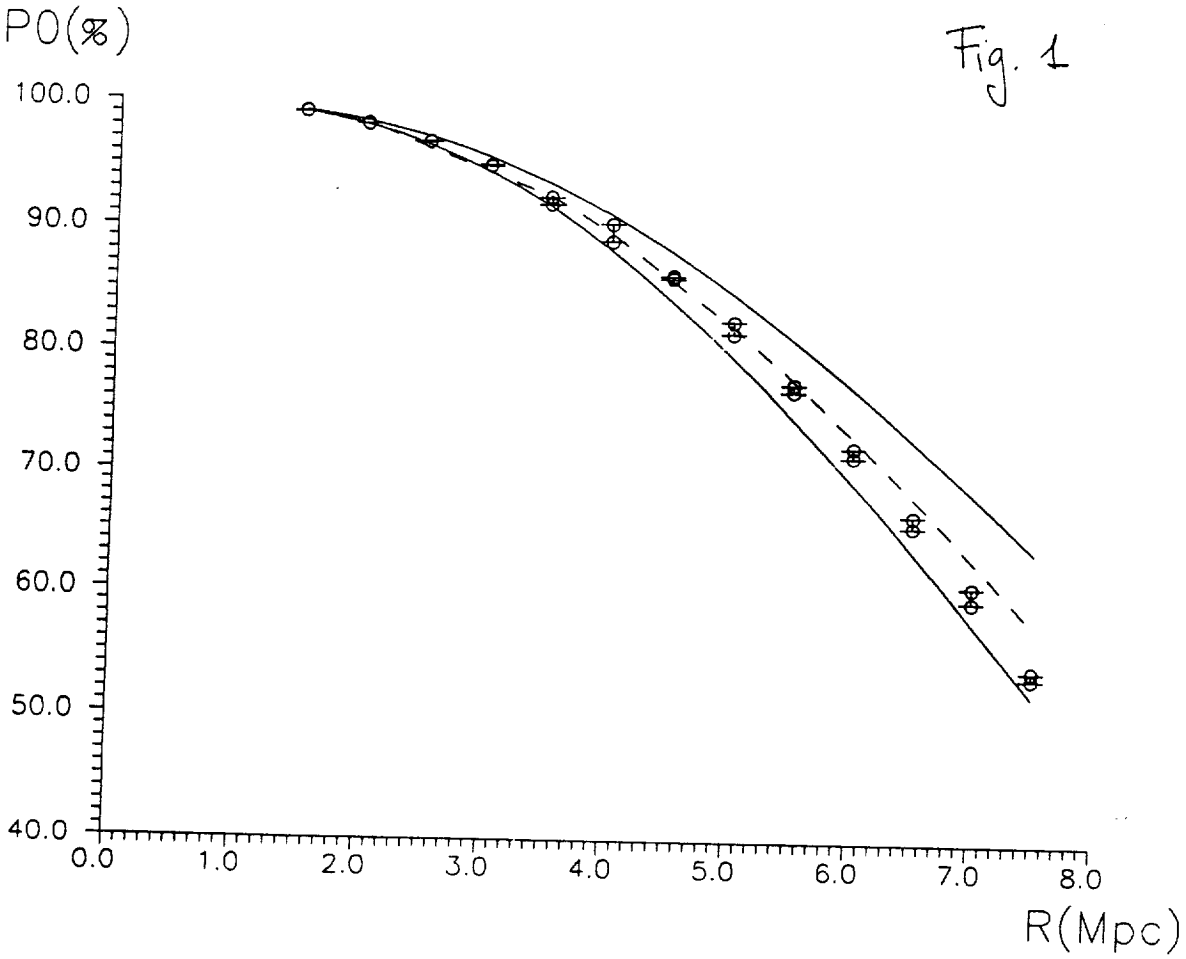

PO(%)
100.0
90.0
80.0
70.0
60.0
50.0
40.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
R(Mpc)

Fig. 2

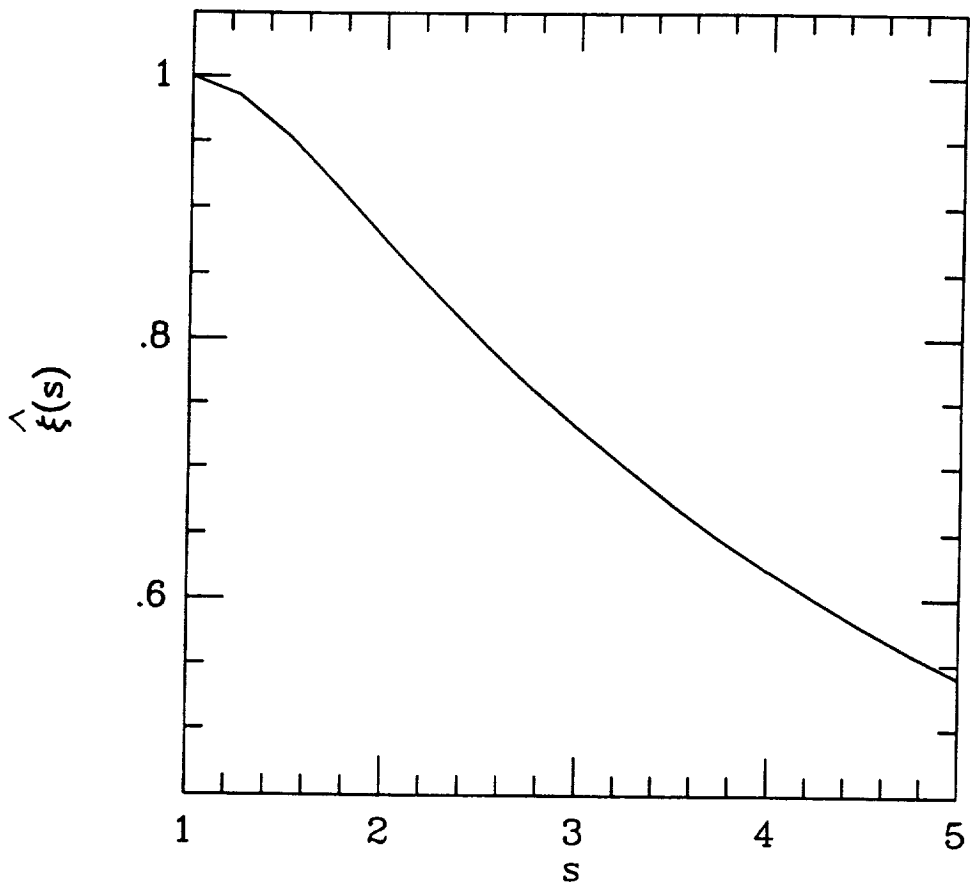

1
.8
.6
$\hat{\xi}(s)$
1
2
3
4
5
s

Fig. 3

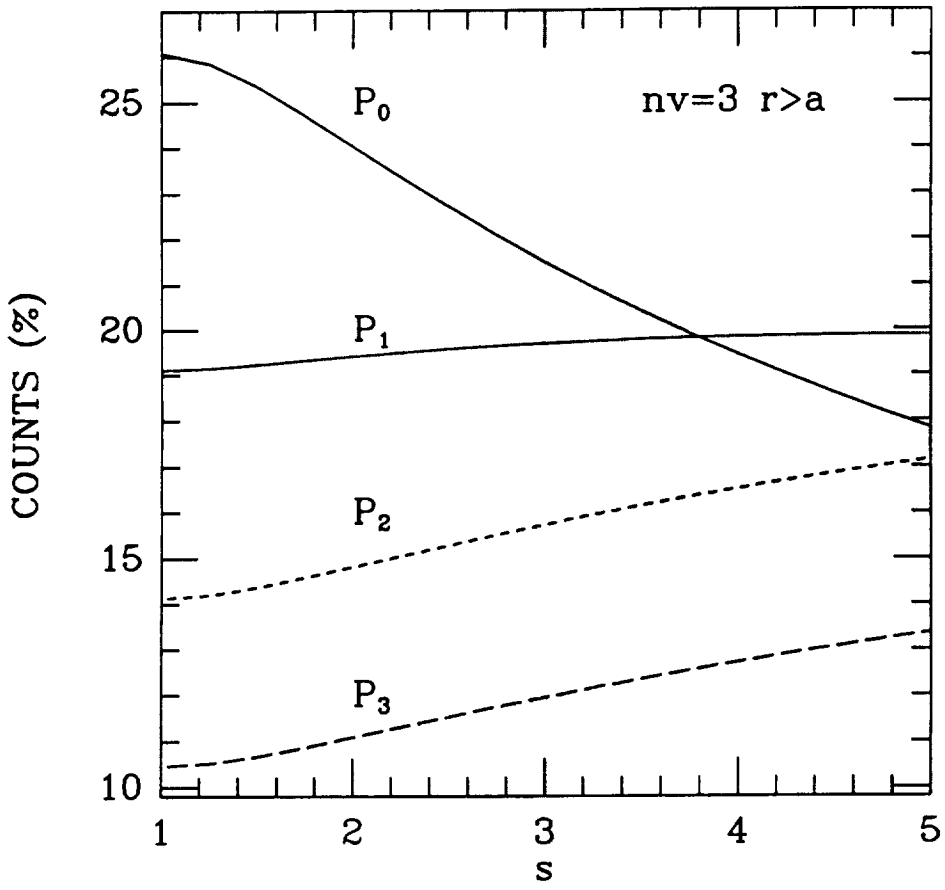

P0
nv=3 r>a
P1
P2
P3
COUNTS (%)
25
20
15
10
1
2
3
4
5
s

Fig. 4

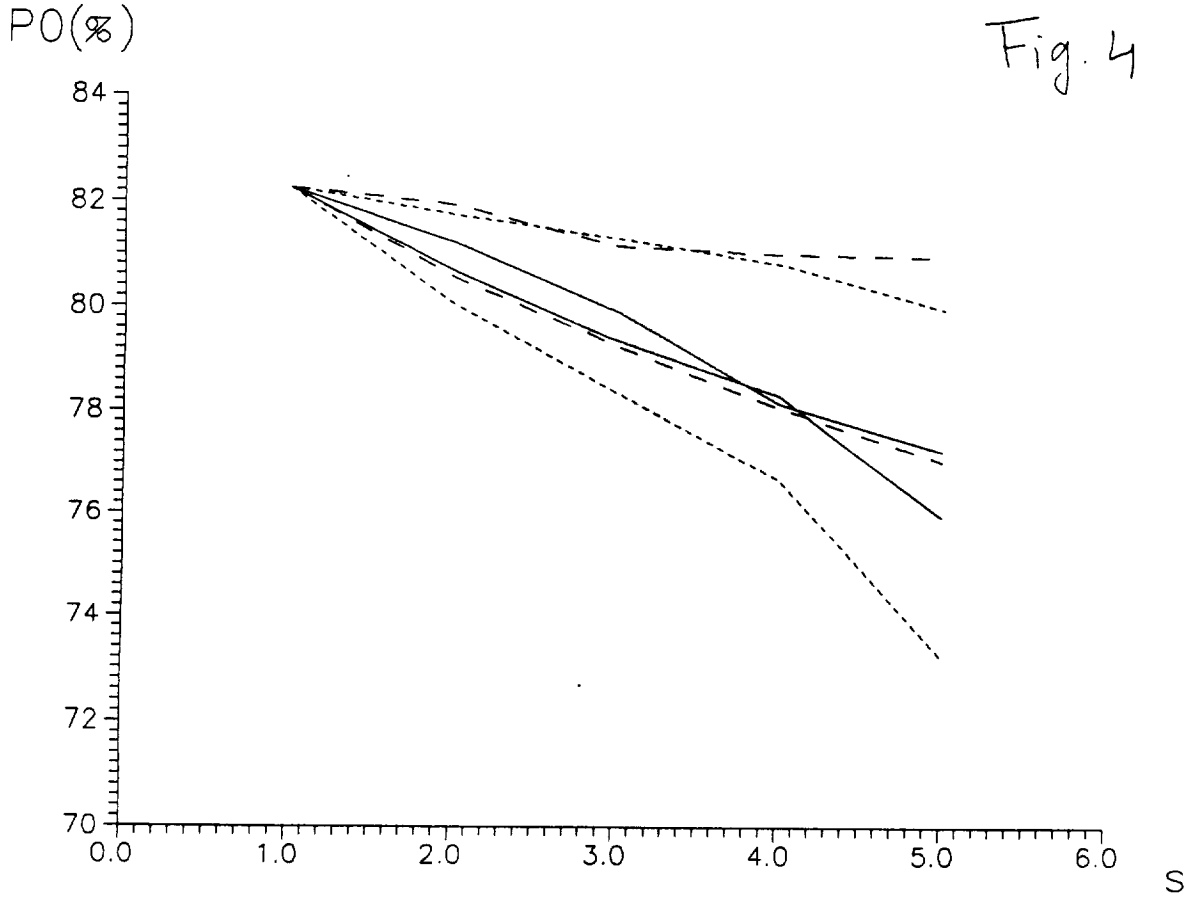

P0(%)
84
82
80
78
76
74
72
70
0.0
1.0
2.0
3.0
4.0
5.0
6.0
S

Fig. 5

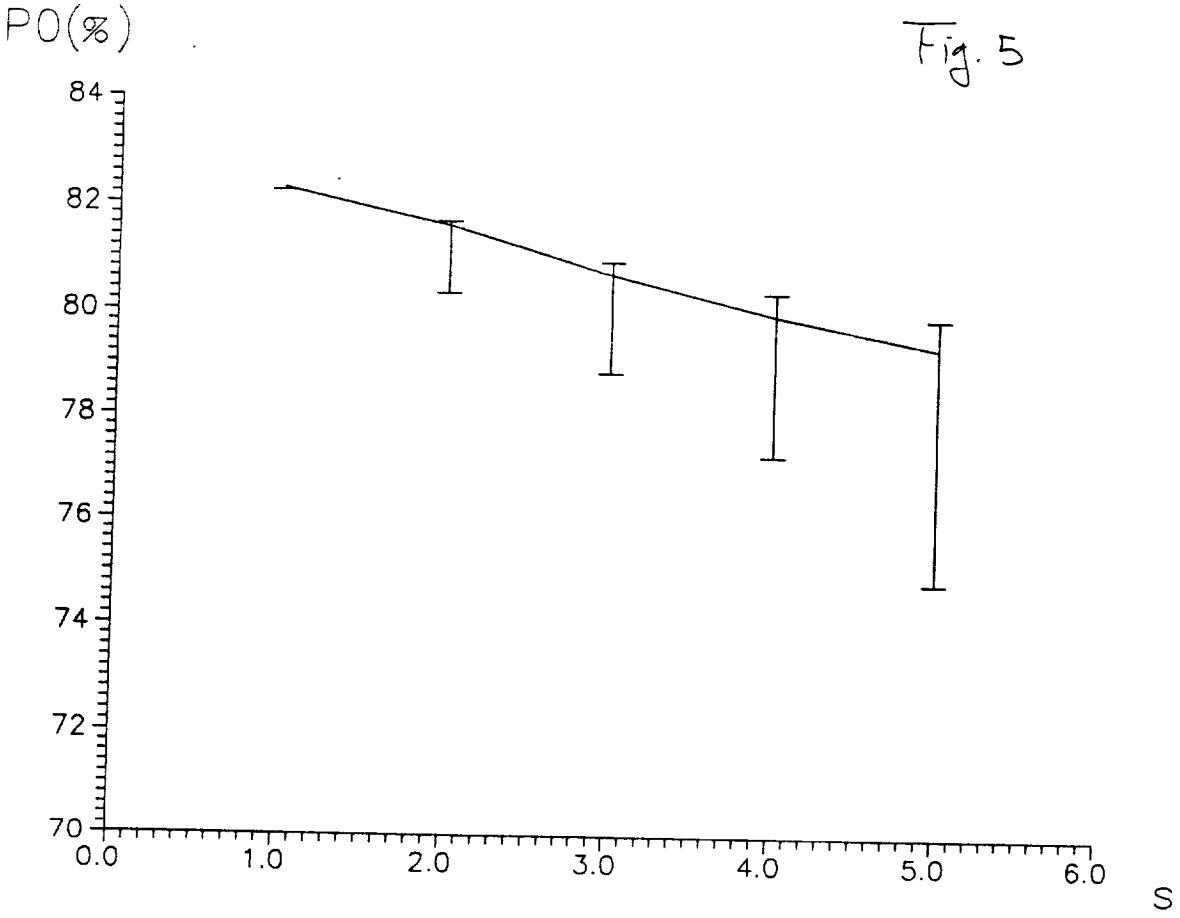
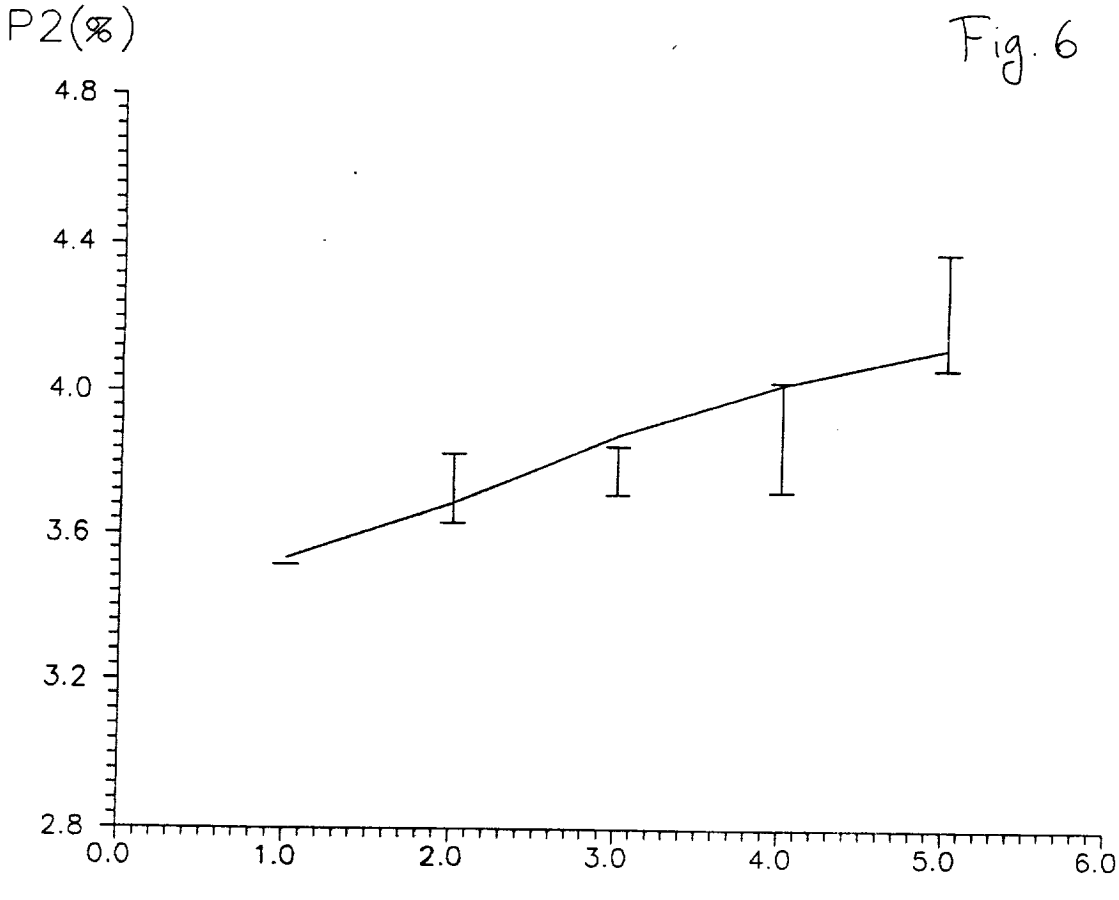

P0(%)
84
82
80
78
76
74
72
70
0.0
1.0
2.0
3.0
4.0
5.0
6.0
S

Fig. 6


P2(%)
4.8
4.4
4.0
3.6
3.2
2.8
0.0
1.0
2.0
3.0
4.0
5.0
6.0
S