A geometrical representation for departures from randomness of the inter-galactic void probablity function*

C.T.J. Dodson,
Department of Mathematics,
University of Manchester Institute of Science and Technology,
Manchester M60 1QD, UK dodson@umist.ac.uk

1999

Abstract

A number of recent studies have estimated the inter-galactic void probability function and investigated its departure from various random models. We study a family of parametric statistical models based on gamma distributions, which do give realistic descriptions for other stochastic porous media. Gamma distributions contain as a special case the exponential distributions, which correspond to the 'random' void size probability arising from Poisson processes. The random case corresponds to the information-theoretic maximum entropy or maximum uncertainty model. Lower entropy models correspond on the one hand to more 'clustered' structures or 'more dispersed' structures than expected at random. The space of parameters is a surface with a natural Riemannian structure, the Fisher information metric. This surface contains the Poisson processes as an isometric embedding and provides the geometric setting for quantifying departures from randomness and perhaps on which may be written evolutionary dynamics for the void size distribution. Estimates are obtained for the two parameters of the void diameter distribution for an illustrative example of data published by Fairall.

Introduction

A number of studies over the past ten years have estimated the inter-galactic void probability function and investigated its departure from randomness. The basic random model is that arising from a Poisson process of mean density n galaxies per unit volume in a large box. Then, in a *given region of volume* V, the probability of finding exactly m galaxies is

$$P_m = \frac{(nV)^m}{m!} e^{-nV} \tag{1}$$

So the probability that the given region is devoid of galaxies is $P_0 = e^{-nV}$. Then it follows that the probability density function for the continuous random variable V in the Poisson case is

$$p_{random}(V) = n\, e^{-nV} \tag{2}$$

For comparison with observations, the approximation fails for very large V since a finite volume box is involved in any catalogue.

A hierachy of N-point correlation functions needed to represent clustering of galaxies in a complete sense was devised by White [26] and he provided explicit formulae, including their continuous limit. In particular, he made a detailed study of the probability that a sphere of radius r is empty and showed that formally it is symmetrically dependent on the whole hierarchy of correlation functions. However, White concentrated his applications on the case when the underlying galaxy distribution was a Poisson process, the starting point for the present approach which is concerned with geometrizing the parameter space of departures from a Poisson process.

*Presented at Workshop on Statistics of Cosmological Data Sets NATO-ASI Isaac Newton Institute 8-13 August 1999

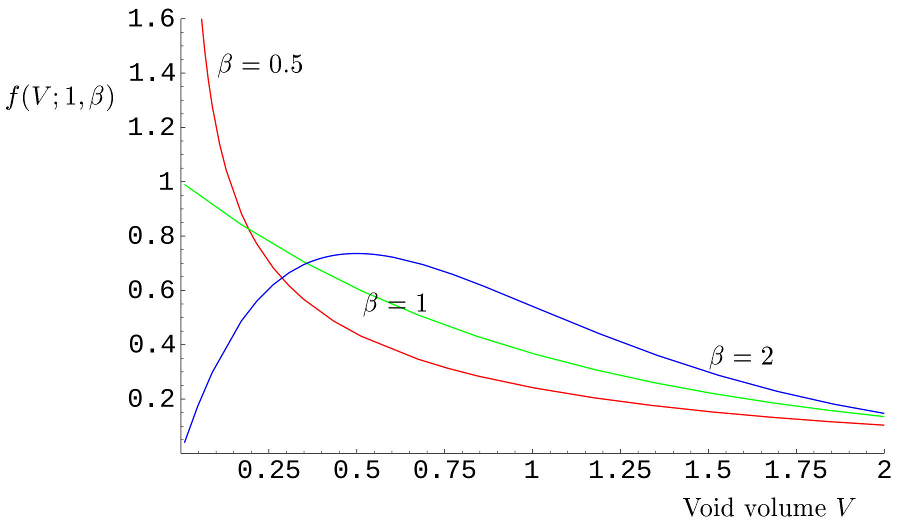

Figure 1: *Probability density functions, $f(V;\mu,\beta)$, for gamma distributions of void volumes with unit mean $\mu = 1$, and $\beta = 0.5,\ 1,\ 2$. The case $\beta = 1$ corresponds to a 'random' distribution from an underlying Poisson process.*

Geometry of gamma models for void volume statistics

We choose a family of parametric statistical models that includes (2) as a special case. There are of course many such families, but we take one that has been successful in modelling void size distributions in terrestrial stochastic porous media [8] and has been used in the representation of clustering of galaxies [6]. The family of gamma distributions has event space $\Omega = \mathbb{R}^+$, parameters $\mu, \beta \in \mathbb{R}^+$ and probability density functions given by

$$f(V;\mu,\beta) = \left(\frac{\beta}{\mu}\right)^{\beta} \frac{V^{\beta-1}}{\Gamma(\beta)}\, e^{-V\beta/\mu} \tag{3}$$

Then $\bar{V} = \mu$ and $Var(V) = \mu^2/\beta$ and we see that μ controls the mean of the distribution while the spread and shape is controlled by, $1/\beta$, the square of the coefficient of variation.

The special case $\beta = 1$ corresponds to the situation when V represents the random or Poisson process in (2) with $\mu = 1/n$. Thus, the family of gamma distributions can model a range of stochastic processes corresponding to non-independent 'clumped' events, for $\beta < 1$, and dispersed events, for $\beta > 1$, as well as the random case (cf. [5, 8]). Thus, if we think of this range of processes as corresponding to the possible distributions of centroids of extended objects such as galaxies that are initially distributed according to a Poisson process with $\beta = 1$, then the three possibilities are:

Chaotic or random structure with no interactions among constituents, $\beta = 1$;

Clustered structure arising from mutually attractive interactions, $\beta < 1$;

Dispersed structure arising from mutually repulsive interactions, $\beta > 1$.

Figure 1 shows a family of gamma distributions, all of unit mean, with $\beta = 0.5,\ 1,\ 2$.

Shannon's information theoretic 'entropy' or 'uncertainty' for such stochastic processes (cf. eg. Jaynes [14]) is given, up to a factor, by the negative of the expectation of the logarithm of the probability density function (3), that is

$$S_f(\mu,\beta) = -\int_0^{\infty} \log(f(V;\mu,\beta)\, f(V;\mu,\beta)\, dV \tag{4}$$

$$= \beta + (1-\beta)\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \log\frac{\mu\,\Gamma(\beta)}{\beta} \tag{5}$$

In particular, at unit mean, the maximum entropy (or maximum uncertainty) occurs at $\beta = 1$, which is the random case, and then $S_f(\mu, 1) = 1 + \log \mu$.

The 'maximum likelihood' estimates $\hat{\mu}, \hat{\beta}$ of μ, β can be expressed in terms of the mean and mean logarithm of a set of independent observations $X = \{X_1, X_2, \ldots, X_n\}$. These estimates are obtained in terms of the properties of X by maximizing the 'log-likelihood' function

$$l_X(\mu, \beta) = \log lik_X(\mu, \beta) = \log \left(\prod_{i=1}^{n} p(X_i; \mu, \beta) \right)$$

with the following result

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad (6)$$

$$\log \hat{\beta} - \psi(\hat{\beta}) = \overline{\log X} - \log \bar{X} \qquad (7)$$

where $\overline{\log X} = \frac{1}{n} \sum_{i=1}^{n} \log X_i$ and $\psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$ is the digamma function, the logarithmic derivative of the gamma function

The usual Riemannian information metric on the parameter space $\mathcal{S} = \{(\mu, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+\}$ is given by

$$ds_{\mathcal{S}}^2 = \frac{\beta}{\mu^2} \, d\mu^2 + \left(\psi'(\beta) - \frac{1}{\beta} \right) d\beta^2 \quad \text{for } \mu, \beta \in \mathbb{R}^+. \qquad (8)$$

For more details about the geometry see [18, 5, 7]. The 1-dimensional subspace parametrized by $\beta = 1$ corresponds to the available 'random' processes. A path through the parameter space $\mathcal{S}$ of gamma models determines a curve

$$c : [a, b] \to \mathcal{S} : t \mapsto (c_1(t), c_2(t)) \qquad (9)$$

with tangent vector $\dot{c}(t) = (\dot{c}_1(t), \dot{c}_2(t))$ and norm $||\dot{c}||$ given via (8) by

$$||\dot{c}(t)||^2 = \frac{c_2(t)}{c_1(t)^2} \, \dot{c}_1(t)^2 + \left(\psi'(c_2(t)) - \frac{1}{c_2(t)} \right) \dot{c}_2(t)^2. \qquad (10)$$

The information length of the curve is

$$L_c(a, b) = \int_a^b ||\dot{c}(t)|| \, dt \qquad (11)$$

and the curve corresponding to an underlying Poisson process has $c(t) = (t, 1)$, so $t = \mu$ and $\beta = 1 = constant$, and the information length is $\log \frac{b}{a}$.

As we know from elementary geometry, arc length is often difficult to evaluate analytically because it contains the square root of the sum of squares of derivatives. Accordingly, we sometimes use the 'energy' of the curve instead of length for comparison between nearby curves. Energy is given by integrating the *square* of the norm of $\dot{c}$

$$E_c(a, b) = \int_a^b ||\dot{c}(t)||^2 \, dt. \qquad (12)$$

so in the case of the curve $c(t) = (t, 1)$, the energy is $\frac{b-a}{ab}$. It is easily shown that a curve of constant μ has $c(t) = (constant, t)$ where $t = \beta$ and $\dot{c}(t) = (0, 1)$; this has energy $\log \frac{a}{b} + \psi'(b) - \psi'(a)$.

Locally, minimal paths joining nearby pairs of points in $\mathcal{S}$ are given by the autoparallel curves or geodesics [7] defined by (8). Some typical sprays of geodesics emanating from various points are provided in [6]. The Gaussian curvature of the surface $\mathcal{S}$ [7] actually controls all of the geometry of geodesics and it is given by

$$K(\mu, \beta) = \frac{\psi'(\beta) + \beta \, \psi''(\beta)}{4(\beta \, \psi'(\beta) - 1)^2} \quad \text{for } \mu, \beta \in \mathbb{R}^+ \qquad (13)$$

$$K(\mu, \beta) \to -\frac{1}{4} \text{ as } \nu \to 0 \qquad (14)$$

$$K(\mu, \beta) \to -\frac{1}{2} \text{ as } \nu \to \infty. \qquad (15)$$

Void diameter statistics

For a general account of large-scale structures in the universe, see Fairall [9]. Kauffmann and Fairall [15] developed a catalogue search algorithm for nearly spherical regions devoid of bright galaxies and obtained a spectrum for diameters of significant voids. This had a peak in the range 8-11 $h^{-1}Mpc$, a long tail stretching at least to 64 $h^{-1}Mpc$, and is compatible with the recent extrapolation models of Baccigalupi et al [1] which yield an upper bound on void diameters of about 100 $h^{-1}Mpc$. We shall return to the data of Kauffmann and Fairall later in this section.

Simulations of Sahni et al. [21] found strong correlation between void sizes and primordial gravitational potential at void centres; void topologies tended to simplify with time. Ghigna et al. [12] found in their simulations that void statistics are sensitive to the passage from CDM to CHDM models. This suggested that the void distribution is sensitive to the type of dark matter but not to the transfer function between types. CHDM simulations gave a void probability in excess of observations, CDM simulations being somewhat better. Vogeley et al. [22] compared void statistics with CDM simulations of a range of cosmological models; good agreement was achieved for samples of very bright galaxies ($M < -19.2$) but for samples containing fainter galaxies the predicted voids were reported to be 'too empty'. Ghigna et al. [11] compared observational data with Gaussian-initiated N-body simulations in a $50\,h^{-1}Mpc$ box and found that at the $2-8\,h^{-1}Mpc$ scales the void probability for CHDM was significantly larger than observed. Ghigna et al. [13] compared simulated galaxy samples with the Perseus-Pisces redshift survey. The void probability function did discriminate between DM and CDM models, the former giving particularly good agreement with the survey.

Little and Weinberg [19] used similar N-body simulations, and found that the void probability was insensitive to the shape of the initial power spectrum. Watson and Rowan-Robinson [23] found that standard CDM predictors do yield reasonably good void probability function estimates whereas Voronoi foam models performed less well.

Lachiéze-Rey and daCosta were of the opinion that the available samples of galaxies were insufficiently representative, because of greater apparent frequency of larger voids in the southern hemisphere. Bernardeau [2] started from a Gaussian field and derived an expression for the void probability function, obtaining

$$\frac{\log P_0}{nV} \approx -(nV\,\sigma^2)^{-3/7} \quad \text{for large } nV\,\sigma^2,$$

with σ^2 the variance of the number of galaxies in volume V. This distribution has an extended large tail because it is asymptotically more like the exponential of $-(nV)^{\frac{4}{7}}$ than the Poisson case which decays like the exponential of $-nV$. Cappi et al. [3] examined the dependence of the void probability function on scale for a range of galaxy cluster samples, finding a general scaling to occur up to void diameters of about $100\,h^{-1}Mpc$. Kerscher et al. [16] used the void probability function to obtain spatial statistics of clusters on scales $10-60\,h^{-1}Mpc$, obtaining satisfactory agreement in a model with a cosmological constant and in a model with breaking of scale invariance of perturbations.

For our model we consider the diameter D of a spherical void with volume $V = \frac{\pi}{6}D^3$ having distribution (3). Something close to the random variable D has direct representation in some theoretical models, for example as polyhedral diameters in Voronoi tesselations [24, 25, 4].

The probability density function for D is given by

$$h(D;\mu,\beta) = f(\frac{\pi}{6}D^3;\mu,\beta)\,\frac{dV}{dD} = \frac{1}{\Gamma(\beta)}\left(\frac{\beta}{\mu}\right)^{\beta}\left(\frac{\pi D^3}{6}\right)^{\beta-1} e^{\frac{-\pi\beta D^3}{6\mu}} \tag{16}$$

Then the mean $\bar{D}$, variance $Var(D)$ and coefficient of variation $CV(D)$ of D are given, respectively,

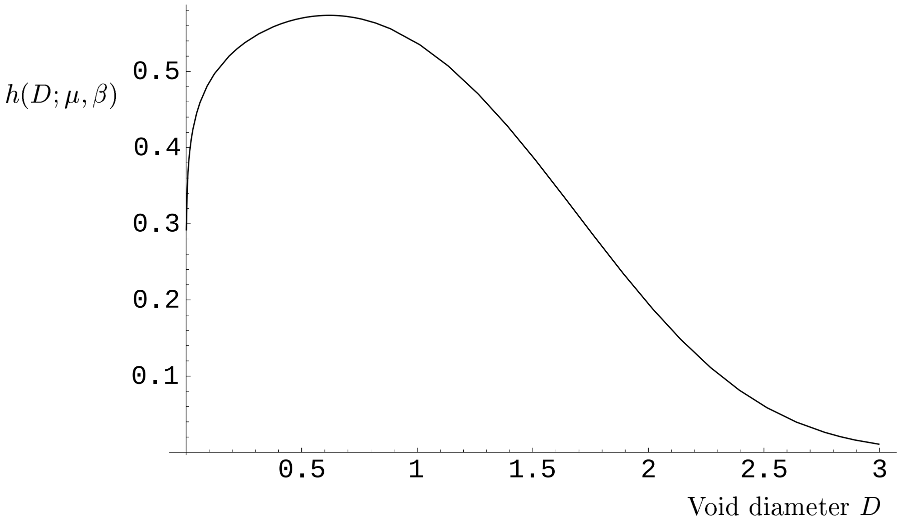

Figure 2: *Probability density function Equation (16), for distributions of void diameters with unit mean, $\mu = 1.244$ and $\beta = 0.370$. These parameter values are the best fit for the data of Kauffmann and Fairall [15], also used in Figure 3.*

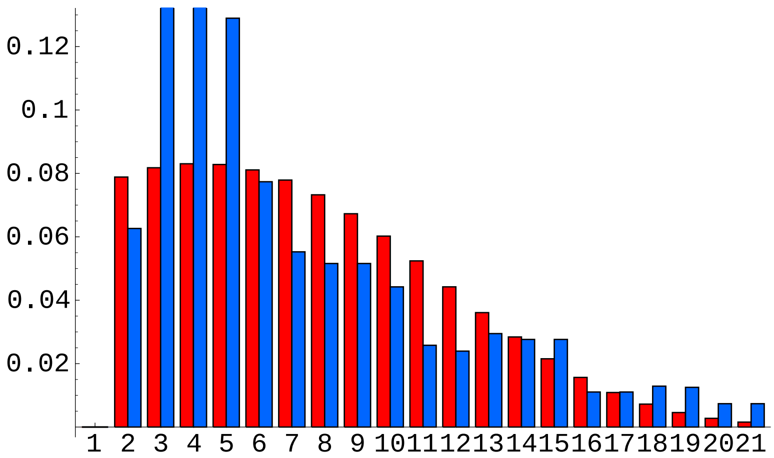

Figure 3: *Histograms of fractions of space occupied by different classes of void diameters, from Kauffmann and Fairall [15] (right hand columns) and predicted from the void diameter distribution Equation (16) (left hand columns) fitted to the same coefficient of variation and mean; the parameters found were $\mu = 1.244$ and $\beta = 0.370$. The class centres are in units of $200\, km/s$ with a mean close to 7 units.*

by

$$\bar{D} = \frac{\Gamma(\beta+\frac{1}{3})}{\left(\frac{\pi\beta}{6\mu}\right)^{\frac{1}{3}}\Gamma(\beta)} \tag{17}$$

$$Var(D) = \frac{\Gamma(\beta)\,\Gamma(\beta+\frac{2}{3}) - \Gamma(\beta+\frac{1}{3})^2}{\left(\frac{\pi\beta}{6\mu}\right)^{\frac{2}{3}}\Gamma(\beta)^2} \tag{18}$$

$$CV(D) = \sqrt{\frac{\Gamma(\beta)\,\Gamma(\beta+\frac{2}{3})}{\Gamma(\beta+\frac{1}{3})^2} - 1} \tag{19}$$

The fact that the coefficient of variation (19) depends only on β gives a rapid fitting of data to (16). Numerical fitting to (19) gives β; this substituted in (17) yields an estimate of μ to fit a given observational mean. By way of illustration, this has been done in Figure 3 for the ZCAT/SRC data from Kauffmann and Fairall [15, 9] Figure 8a (cf also [9] Figure 6.5), both set to unit mean diameter—the true mean for that catalogue was about $7\,h^{-1}Mpc$. The fitted values were $\mu = 1.244$ and $\beta = 0.370$. This fit is not particularly good if the reported peaks are not an artifact but, qualitatively, we observe that the fitted value $\beta = 0.370$ is apparently significantly less than 1, which would correspond to the random model. We conclude tentatively that, for the ZCAT/SRC data subjected to the Kauffmann and Fairall search algorithm, the new model suggests clumping rather than dispersion in the underlying stochastic process. A *Mathematica* NoteBook for performing the fitting procedure is available from the author, who would welcome more sets of data.

Suppose that the parameter space $\mathcal{S}$ of gamma-based models is a meaningful representation of the evolutionary process and that some coordinates such as ($\mu = 1.244$, $\beta = 0.370$) in $\mathcal{S}$ represent current data. Then geodesics in $\mathcal{S}$ through this point represent some kind of extremal path. Moreover, we may consider a vector field U on $\mathcal{S}$ such that ($\mu = 1.244$, $\beta = 0.370$) is the present endpoint of an integral curve of U, the initial point of this curve being presumably in an epoch when less clustering ($\beta > 0.370$), a random state ($\beta = 1$) or even dispersion ($\beta > 1$) was present at higher density. It would be interesting to investigate the various candidate cosmologies for their appropriate vector fields, via the statistics of matter and voids they predict. The present gamma-related parametric statistical models provide the means to convert the catalogue statistics into the coordinate parameters and a background geometrization of the statistics on which dynamical processes may be formulated.

Acknowledgement

The author wishes to thank A.P. Fairall, F. Labini and M.B. Ribeiro for helpful comments during the preparation of this article.

References

[1] C. Baccigalupi, L. Amendola and F. Occhionero. Imprints of primordial voids on the cosmic microwave background *Mon. Not. R. Astr. Soc.* 288, 2 (1997) 387-96.

[2] F. Bernardeau. The gravity-induced quasi-Gaussian correlation hierarchy *Astrophys. J.* 392, no.1, pt.1; 10 June 1992; p.1-14.

[3] A. Cappi, S. Maurogordato and M. Lachiéze-Rey A scaling law in the distribution of galaxy clusters. *Astron. Astrophys.* 243, 1 (1991) 28-32.

[4] P. Coles. Understanding recent observations of the large-scale structure of the universe. *Nature* 346 (1990) 446-.

[5] C.T.J. Dodson. Gamma manifolds and stochastic geometry. In: **Proceedings of the Workshop on Recent Topics in Differential Geometry**, Santiago de Compostela 16-19 July 1997. *Public. Depto. Geometría y Topología* 89 (1998) 85-92.

[6] C.T.J. Dodson. Spatial statistics and information geometry for parametric statistical models of galaxy clustering. *Int. J. Theor. Phys.* 38, 10 (1999) 2585-2597.

[7] C.T.J. Dodson and T. Poston. **Tensor Geometry** Graduate Texts in Mathematics 130, Second edition, Springer-Verlag, New York, 1991.

[8] C.T.J. Dodson and W.W. Sampson. Modeling a class of stochastic porous media. *Appl. Math. Lett.* 10, 2 (1997) 87-89.

[9] A.P. Fairall. **Large-scale structure in the universe** Wiley-Praxis, Chichester 1998.

[10] S. Ghigna.Galaxy surveys and dark matter models: the statistics of voids. In, **Birth of the universe and fundamental physics** Springer-Verlag, Berlin, Germany, 1995, 287-90.

[11] S. Ghigna, S. Borgani, S.A. Bonometto, L. Guzzo, A. Klypin, J.R. Primack, R. Giovanelli and M.P. Haynes. Sizes of voids as a test for dark matter models. *Astrophys. J. Lett.* 437, 2, 2 (1994) L71-4.

[12] S. Ghigna, S.A. Bonometto, J. Retzlaff, S. Gottlober and G. Murante. Void analysis as a test for dark matter composition? *Astrophys. J.* 469, 1, 1 (1996) 40-7.

[13] S. Ghigna, S. Borgani, M. Tucci, S.A. Bonometto, A. Klypin and J.R. Primack. Statistical tests for CHDM and Lambda CDM cosmologies. *Astrophys. J.* 479, 2, 1 (1997) 580-91.

[14] E.T. Jaynes. Information theory and statistical inference. *The Physical Review* 106 (1957) 620-630 and 108 (1957) 171-190.

[15] G. Kauffmann and A.P. Fairall. Voids in the distribution of galaxies: an assessment of their significance and derivation of a void spectrum. *Mon. Not. R. Astr. Soc.* 248 (1990) 313-324.

[16] M. Kerscher, J. Schmalzing, J. Retzlaff, S. Borgani, T. Buchert, S. Gottlober, V. Muller, M. Plionis and H. Wagner. Minkowski functionals of Abell/ACO clusters. *Mon. Not. R. Astr. Soc.* 284, 1 (1997) 73-84.

[17] M. Lachiéze-Rey, L.N. Da-Costa and S. Maurogordato. Void probability function in the Southern Sky Redshift Survey. *Astrophys. J.* 399 (1992) 10-15.

[18] S.L. Lauritzen. Statistical Manifolds. In **Differential Geometry in Statistical Inference**, Institute of Mathematical Statistics Lecture Notes, Volume 10, Berkeley 1987, pp 163-218.

[19] B. Little and D.H. Weinberg. Cosmic voids and biased galaxy formation. *Mon. Not. R. Astr. Soc.* 267, 3 (1994) 605-28.

[20] T. Piran, M.Lecar, D.S. Goldwirth, L. Nicolaci da Costa and G.R. Blumenthal. Limits on the primordial fluctuation spectrum: void sizes and anisotropy of the cosmic microwave background radiation. *Mon. Not. R. Astr. Soc.* 265, 3 (1993) 681-8.

[21] V. Sahni, B.S. Sathyaprakash and S.F. Shandarin. The evolution of voids in the adhesion approximation *Astrophys. J.* 431, 1, 1 (1994) p.20-40.

[22] M.S. Vogeley, M.J. Geller, C. Park and J.P. Huchra. Voids and constraints on nonlinear clustering of galaxies. *Astron. J.* vol.108, 3; (1994) 745-58.

[23] J.M. Watson and M. Rowan-Robinson. The void probability function for flux-limited samples. *Mon. Not. R. Astr. Soc.* 265, 4 (1993) 1027-34.

[24] R. van der Weygaert. Quasi-periodicity in deep redshift surveys. *Mon. Not. R. Astr. Soc.* 249 (1991) 159-.

[25] R. van der Weygaert and V. Icke. Fragmenting the universe II. Voronoi vertices as Abell clusters. *Astron. Astrophys.* 213 (1989) 1-9.

[26] S.D.M. White. The hierarchy of correlation functions and its relation to other measures of galaxy clustering. *Mon. Not. R. Astr. Soc.* 186, (1979) 145-154.