Astron. Astrophys. 328, 1–4 (1997)

ASTRONOMY AND ASTROPHYSICS

Research Note

On the role of Bernoulli distribution in cosmology

Attila Mészáros

Department of Astronomy, Charles University, 150 00 Prague 5, Švédská 8, Czech Republic (e-mail: meszaros@mbox.cesnet.cz)

Received 5 February 1997 / Accepted 10 July 1997

Abstract. Cosmological observations suggest that the transition scale to homogeneity is not smaller than $\simeq (200-300)h^{-1}$ Mpc (h is the Hubble parameter in units 100 km s^{-1} Mpc^{-1}), and that - in addition - this transition scale seems to decrease with increasing redshift. Nevertheless, these claims are highly uncertain, and call for further observational tests allowing to verify the homogeneity even on scales $\simeq (400-1000)h^{-1}$ Mpc for different redshifts. The key ideas of such a test are presented. Simultaneous pencil beam surveys are proposed to be done, and then the randomness of observed objects should be verified by the so called "Bernoulli test". The method is in fact a generalization of the usual counts-in-cells analysis, where the Poisson distribution is substituted by the Bernoulli distribution. Hence, it is shown that this distribution may have a great importance in cosmology.

Key words: cosmology: large-scale structure of Universe – methods: statistical

1. Introduction

In the Friedmann model of Universe the so called "transition scale to the homogeneous Universe" (denoted here as r_t) is the maximal scale of structures still existing in Universe. Statistically this means that on scales $\leq r_t$ ($\geq r_t$) the objects are not (are) distributed randomly; for more details see, e.g., Peebles (1993) or Einasto & Gramann (1993). It must be $r_t \ll c/H = 3000h^{-1}$ Mpc, where $H = 100\ h$ km s^{-1}Mpc^{-1} is the Hubble parameter, and c is the velocity of light.

Several observations supported the existence of structures even with sizes $\simeq (50-300)h^{-1}$Mpc (cf. Batuski & Burns 1985; Tully 1987; Efstathiou, Sutherland & Maddox 1990; Scaramella, Vettolani & Zamorani 1991; Carruthers 1991; Loveday et al. 1992; Bahcall & West 1992; Einasto & Gramann 1993; Loveday et al. 1995, Cohen et al. 1996). As far as it is known the largest structure with size $\simeq 300h^{-1}$ Mpc was discovered by Batuski & Burns (1985). Generally, the so called complexes of superclusters with sizes $\simeq (200-300)h^{-1}$ Mpc are thought to be the biggest observed structures (Tully 1987).

r_t may be depending on z (z is the redshift), and hence $r_t = r_t(z)$. Different redshift surveys of galaxies (cf. Broadhurst, Ellis & Shanks 1988; Longo 1991; West 1991; Loveday et al. 1992; Bernstein at al. 1994; Infante & Pritchet 1995; Loveday et al. 1995; Cohen et al. 1996; Loveday et al 1996; Shectman et al. 1996; Le Fèvre et al. 1996) suggest that on higher redshifts the transition scale tends to decrease; e.g., for $z \simeq (0.1-1.0)$ it should be $r_t(z) \simeq (10-100)h^{-1}$ Mpc. This tendency is supported also by the studies of the clustering of quasars (cf. Bahcall & Chokshi 1991; Ho & Fang 1993) claiming departures from the homogeneity only on scale $\sim 10h^{-1}$ Mpc. Hence, it seems that one may take $r_t(z) \simeq (0.07-0.1) \times 3000h^{-1}$ Mpc for $z \simeq (0.0-0.2)$; in addition, $r_t(z)$ seems to be decreasing with increasing z.

Nevertheless, some structures with sizes $\simeq (400-1000)h^{-1}$ Mpc cannot be excluded yet from the observational point of view. In addition, recently (Quashnock et al. 1996) clustering on scales $\sim 100h^{-1}$ Mpc was detected even at high redshifts ($1.2 \leq z \leq 4.5$). Hence, the decrease of transition scale by increasing redshift is not sure yet. Simply, further observational efforts are needed to test the randomness of distribution of objects even on scales $\simeq (400-1000)h^{-1}$ Mpc for different z.

The aim of this note is to present the basic ideas of such observational test. In this test the Bernoulli distribution will play the key role; hence the name "Bernoulli test". This distribution is a standard matter of statistics, and its *special case* is the Poisson distribution (cf. Trumpler & Weaver 1953). As far as it is known, the more general Bernoulli distribution was never used yet in cosmology, because the Poisson distribution was usually also enough.

2. Bernoulli distribution

Be given N randomly distributed points in a volume V. Then the mean number density of points in a volume unit is $n = N/V$. Be given a part of V having the volume V_1 ($V_1 < V$). The probability to find k points in V_1 (it may be $k = 0, 1, 2, ..., N$)

is given by the Bernoulli distribution (called also as "binomial distribution") having the form (cf. Trumpler & Weaver 1953)

$$B_{(V_1/V)}(N,k) = \frac{N!}{(N-k)!k!}\left(\frac{V_1}{V}\right)^k\left(\frac{V-V_1}{V}\right)^{N-k}. \quad (1)$$

It is essential to note that the forms of both volumes V_1 and V are fully arbitrary; it is enough that the volume V_1 be fully in the interior of volume V. In other words, it is not necessary that $V_1 \ll V$ be fulfilled; $V_1 < V$ is also enough. For our purpose it is also essential to remark that volume V need not be a "compact" single volume; it may be a sum of separated volumes. For example, it may also be that $V = MV_1$, where $M \geq 2$ is an integer number, and volume V is a sum of M separated identical volumes V_1. The mean number of points in volume V_1 is given by $\bar{n} = nV_1$, and the variance is defined by $\sigma^2 = \overline{(k-\bar{n})^2} = \bar{n}(1-(V_1/V))$ (Trumpler & Weaver 1953). Note yet that k is a positive integer, but $\bar{n}$ need not be integer. Of course, $\bar{n}$ is different for different V_1 ($\bar{n} = nV_1$).

Poisson distribution is *a special case*, when $V_1/V \rightarrow 0$ for fixed $n = N/V$ (for details see, cf., Trumpler & Weaver 1953; p. 161). Then one has $\bar{n} = \overline{(k-\bar{n})^2}$. The Bernoulli distribution may be substituted by the Poissonian one, when $V_1/V \ll 1$; the accuracy of this substitution is $\simeq V_1/V$.

Consider now the "Bernoulli test". Be given a redshift interval ("redshift bin") defined by $[(z_b - \Delta z_b/2), (z_b + \Delta z_b/2)]$. This is a spherical layer with thickness $(c/H)\Delta z_b$, if $\Delta z_b \ll z_b$. (Note that $\Delta z_b \ll z_b$ is not necessary; the condition $\Delta z_b \leq 2z_b$ is also enough. But for the purpose of this paper one may take the special case $\Delta z_b \ll z_b$.) Consider $M \geq 2$ solid angles in different directions, having the same size ω steradians. The intersections of these M solid angles and the redshift bin define M identical "cells" with volume $V_1 = \omega(c/H)^3[1-(1+z_b)^{-1/2}]^2\Delta z_b$ for spatially flat Friedmann model (it is standard cosmology to write down V_1 also for the other two Friedmann models). To do the Bernoulli test one may simply identify V_1 with these cells, and to take $V = MV_1$. (It is essential to precise that $M\omega$ may but does not need to cover the whole sky; one has $M\omega \leq 4\pi$, and $M\omega \ll 4\pi$ is also possible.) If there are N points in V, then Eq. (1) takes the form

$$B_{1/M}(N,k) = \frac{N!}{(N-k)!k!}M^{-k}(1-M^{-1})^{N-k}. \quad (2)$$

Obviously, $\bar{n} = N/M$ is the mean number of points in a cell, and the variance is $\sigma^2 = \bar{n}(M-1)/M$. The choices of ω, M, z_b and Δz_b are *arbitrary*.

(2) may be substituted by the Poisson distribution only for $M \gg 1$ with $\simeq 1/M$ inaccuracy. Then this Bernoulli test may be substituted by the well-known Poisson test, which is often used in cosmology (Hubble 1934; Trumpler & Weaver 1953; Abell 1958; Efstathiou et al. 1990; Press et al. 1992; Oliver et al. 1996; Ueda & Yokoyama 1996). This standard "counts-in-cells" method considers practically always the Poisson distribution as the theoretical expectation; any remarks about a more general distribution are *rare* (cf. Carruthers & Minh 1983; Bouchet et al. 1993; Ueda & Yokoyama 1996), and *no use* of Bernoulli distribution is known yet. We see that Bernoulli test is more general, because any $M \geq 2$ is allowed, and one does not need $M \gg 1$.

All this means that one may test the randomness of distribution of objects for any redshift interval defined by z_b and Δz_b. Of course, in order to avoid the problems with the evolutionary effects (see cf. Butcher & Oemler 1984) and also to study the dependence of transition scale on redshift one should take thin redshift intervals (say, $\Delta z_b \lesssim 0.1$, or even much smaller). For simplicity, one should consider the objects having well defined point-like positions instead of their structures (clusters, superclusters, voids, complexes, filaments,...). In addition, one may consider either a concrete type of objects (special type of galaxies, quasars, AGNs, ...) or all objects together.

Concretely, one may proceed as follows. Obviously, ω should be as small as possible in order to reduce the huge number of observed objects in the considered solid angle. This number should not be bigger than $\sim 10^5$, because the present-day instruments are able to measure such amount of redshifts (cf. Loveday et al. 1996; Shectman et al. 1996). In the so called pencil beam survey (Broadhurst et al. 1990) one used the solid angle ~ 1 square degree. Similar or even smaller solid angles should also be considered here. Then one should take arbitrary $M \geq 2$ such solid angles. The angular distances among these solid angles should be roughly the same (θ), and this θ is again arbitrary. (Note that - trivially - both θ and M are not fully arbitrary simultaneously. First, one has $\sqrt{\omega} < \theta \leq \pi$, of course. Second, it is an elementary mathematics to calculate the maximal possible θ for $M \geq 2$ objects; for example, for $M = 2$; $M = 3$; $M = 8$; $M \gg 1$, this maximal value is $\pi/2$; $2\pi/3$; $\pi/4$; $\simeq \sqrt{4\pi/M}$, respectively.) Having this arrangement one has to measure the redshifts of objects in M solid angles; either of all objects or of a special type of objects. Then one has to choose a redshift interval $[(z_b - \Delta z_b/2), (z_b + \Delta z_b/2)]$ with $\Delta z_b \ll z_b$, and to compare the observed number of objects in M cells with the theoretical Bernoulli distribution defined by (2). The best is to begin with the smallest possible M and with biggest possible angular scales (say, $2 \leq M < 8$ and $\pi \geq \theta > \pi/2$). If for these angular scales the comparison with Bernoulli distribution gives wrong fit, then in this redshift interval the Friedmann model is not fulfilled yet. If on the contrary one obtains acceptable fit, then one may increase M with the simultaneous decreasing of θ for the same z_b up to the angular scale $\theta(z_b)$ for which the fit will already be wrong. To transform $\theta(z_b)$ into the transverse proper distance for a given z_b in order to obtain $r_t(z_b)$ is a standard matter (cf. Weinberg 1972). Of course, one has to take several different values of z_b, and to repeat the previous procedure. If there were no good fit for any possible redshift and for any angular scales, then the Friedmann model would be in a serious doubt.

Probably the simplest and the fastest comparison of theoretically expected Bernoulli distribution and the observations may be done as follows. Assume that there are n_j observed objects at j-th cell ($j = 1, 2, ...M$). Then the maximum-likehood estimates (cf. Feller 1957; Morrison 1967) of mean and variance are

$$n_{est.} = \frac{1}{M}\sum_{j=1}^{M} n_j, \quad \sigma^2_{est.} = \frac{1}{M-1}\sum_{j=1}^{M}(n_j - n_{est.})^2. \qquad (3)$$

If the Bernoulli distribution is correct, then it must be

$$\sum_{j=1}^{M} n_j = [M/(M-1)]^2 \sum_{j=1}^{M}(n_j - n_{est.})^2 \qquad (4)$$

in the statistical sense. The acceptance of this equality may be verified, e.g., as follows. Assume that $n_{est.} = n$. Then the theoretically expected variance is $\sigma^2 = [(M-1)/M]n_{est.}$, and the observed variance is given by Eq. (3). Then the indentity of these two variances should be tested by the standard F-test (Fisher-test) for $M-1$ degrees of freedom (Press et. al 1992).

Note that negative F-test rejects the Bernoulli distribution, but once F-test gives indentity in statistical sense, this does not mean yet that one surely has a Bernoulli distribution. (The identity $n_{est.} = [M/(M-1)]\sigma^2_{est.}$ may occur for a fully different distribution, too.) Then further tests are needed. One may use, e.g., the standard Kolmogorov-Smirnov-test needing the integral distribution function $\sum_{k'=0}^{k} B(N,k')$ $(0 \leq k \leq N)$. Remark yet that for $N \gg 1$ the Bernoulli distribution may well be approximated by the normal one, and hence this integral distribution may well be approximated for $0 \ll k \ll N$ by standard error function. Both this approximation and the Kolmogorov-Smirnov test are standard procedures of statistics (cf. Trumpler & Weaver 1953; Press et al. 1992); no further details are needed here.

3. Remarks and conclusion

In the previous section a new statistical method allowing to determine the transition scale for different redshift bins was outlined. Of course, several questions are emerging. To keep the brevity only two short remarks are presented here.

First, one has to note that tremendous technical problems can occur in the observational application of this test, which are not discussed in detail here. It is only noted that recently several redshift surveys are in progress (see, cf. Loveday et al. 1996; Shectman et al. 1996, and references therein). Nevertheless, these surveys are done for great solid angles and for relatively near objects. To search for transition scale one should proceed oppositely, and to use small solid angles but large redshift ranges similarly to pencil beam surveys (Broadhurst et al. 1990).

Second, it is not alleged that the test presented here is the only possible statistical procedure leading to the determination of $r_t(z)$ for a given redshift bin. For example, the method based on the usual angular correlation function (Chandrasekhar & Münch 1952; Chandrasekhar 1954; Limber 1953; Peebles 1973; 1980; 1993; Börner 1993) may immediately be used for a given redshift bin, too. Or, as a fully different new method, the so called nearest neighbor analysis may also be applied for a redshift interval (Scott & Tout 1989). Trivially, the best is the combination of several statistical tests. In addition, because searching for $r_t(z)$ one in fact tests the validity of Friedmann model itself, the galaxy (quasars, AGNs,...) counting methods should also be compared with fully different tests of the Friedmann model (cf. Mészáros & Vanýsek 1997).

Summing up: As it was already mentioned, the use of any generalized version of Poisson distribution is rare in cosmology (see, cf., Sect. 2. of Ueda & Yokoyama (1996) for a brief survey), and - as far as it is known - concretely the Bernoulli distribution was never used yet. This is remarkable, because this paper shows its importance. This is the key conclusion of this paper.

The author hopes that all this will encourage the observers to consider seriously this Bernoulli test for different thin redshift intervals. Such surveys would have a great importance: searching for the transition scale for different redshifts they would in essence test the fulfilment of the Friedmann model itself.

Acknowledgements. The author would like to thank P. Mayer and anonymous referee for useful notes.

References

Abell, G.O. 1958, ApJS 3, 211
Bahcall, N.A. & Chokshi, A. 1991, ApJ 380, L9
Bahcall, N.A. & West, M.J. 1992, ApJ 392, 419
Batuski, D.J. & Burns, J.O. 1985, ApJ 299, 5
Bernstein, G.M., Tyson, J.A., Brown, W.R. & Jarvis, J.F. 1994, ApJ 426, 524
Bouchet, F.R., Strauss, M.A., Davis, M., Fisher, K.B., Yahil, A. & Huchra, J.P. 1993, ApJ 417, 36
Börner, G. 1993, The Early Universe, Springer, Berlin
Broadhurst, T.J, Ellis, R.S. & Shanks, T. 1988, MNRAS 235, 827
Broadhurst, T.J, Ellis, R.S., Koo, D.C. & Szalay, A.S. 1990, Nature, 343, 726
Butcher, H. & Oemler, A. 1984, Nature 310, 31
Carruthers, P. & Minh D.-V. 1983, Phys. Lett. B, 131, 116
Carruthers, P. 1991, ApJ 380, 24
Chandrasekhar, S. 1954, in Selected Papers on Noise and Stochastic Processes, ed. N. Wax, Dover, New York
Chandrasekhar, S. & Münch, G. 1952, ApJ 115, 103
Cohen, J.G., Hogg, D.W., Pahre, M.A. & Blandford, R. 1996, ApJ 462, L9
Einasto, J. & Gramann, M. 1993, ApJ 407, 443
Efstathiou, G., Sutherland, W.J. & Maddox, S.J. 1990, Nature 348, 705
Efstathiou, G. et al. 1990, MNRAS 247, 10P
Feller, W. 1957, An Introduction to Probability Theory and its Applications, J.Wiley, New York
Ho, H.J. & Fang, L.Z. 1993, ApJ 410, 493
Hubble, E. 1934, ApJ 79, 8
Infante, L. & Pritchet, C.J. 1995, ApJ 439, 565
Le Fèvre, O., Hudson, D., Lilly, S.J., Crampton, D., Hammer, F. & Tresse, L. 1996, ApJ 461, 534
Limber, D.N. 1953, ApJ 117, 134
Longo, M.J. 1991, ApJ 372, L59
Loveday, J., Efstathiou, G., Peterson, B.A. & Maddox, S.J. 1992, ApJ 400, L43
Loveday, J., Maddox, S.J., Efstathiou, G. & Peterson, B.A. 1995, ApJ 442, 457

Loveday, J., Efstathiou, G., Maddox, S.J. & Peterson, B.A. 1996, ApJ 468, 1
Mészáros, A. & Vanýsek, V. 1997, A&A 319, 371
Morrison, D.F. 1967, Multivariate statistical methods, McGraw-Hill, New York
Oliver, S.J. et al. 1996, MNRAS 280, 673
Peebles, P.J.E. 1973, ApJ 185, 413
Peebles, P.J.E. 1980, The Large-Scale Structure of the Universe, Princeton University Press, Princeton
Peebles, P.J.E. 1993, Principles of Physical Cosmology, Princeton University Press, Princeton
Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1992, Numerical Recipes, Cambridge University Press, Cambridge
Quashnock, J.M., Vanden Berk, D.E. & York, D.G. 1996, ApJ 472, L69
Scaramella, R., Vettolani, G. & Zamorani, G. 1991, ApJ 376, L1
Scott, D. & Tout, C.A. 1989, MNRAS 241, 109
Shectman, S.A. et al. 1996, ApJ, 470, 142
Trumpler, R.J. & Weaver, H.F. 1953, Statistical Astronomy, University of California Press, Berkeley
Tully, R.B. 1987, ApJ 323, 1
Ueda, H. & Yokoyama, J. 1996, MNRAS 280, 754
Weinberg, S. 1972, Gravitation and Cosmology, J.Wiley, New York
West, M.J. 1991, ApJ 379, 19

This article was processed by the author using Springer-Verlag LaTeX A&A style file *L-AA* version 3.