STATISTICAL STUDIES OF THE CLUSTERING OF GALAXIES

Edward J. Groth
Physics Department, Princeton University

An investigation of the clustering of galaxies making use of correlation function techniques is discussed. Correlation functions are defined and the relations between the angular and spatial functions are presented. The results obtained from application of the method to three samples of galaxies are described and some conclusions are drawn. Finally, some directions for future research are suggested.

1. INTRODUCTION

The study of the clustering of galaxies is an exciting and active field of research. Many complementary techniques from qualitative examination of photographs to quantitative analyses of measured clustering parameters are being brought to bear on the problems. In the limited space available only a small part of this research can be reviewed. In particular, this paper concentrates on the work done at Princeton in recent years. Clustering is studied with statistical methods making use of correlation function techniques. P.J.E. Peebles is the "prime mover" of this work and, building on the earlier work of Limber (1953) and Neyman, Scott, and Shane (1953), has summarized the mathematical details and assumptions used in the application of correlation function techniques (Peebles 1973). These techniques have been refined through many applications (see Groth and Peebles 1977 and earlier references therein) and can be applied to study the clustering of objects among themselves or the "cross clustering" of different classes of objects (e.g. the clustering of galaxies with radio sources has been investigated by Seldner and Peebles 1978). The clustering of galaxies with galaxies is the subject of this paper.

G.O. Abell and P. J. E. Peebles (eds.), Objects of High Redshift, 31–38.

https://doi.org/10.1017/S0074180900068285 Published online by Cambridge University Press

2. CORRELATION FUNCTIONS

2.1. Definition

The angular two-point correlation function for a sample of objects distributed on the sky may be defined through the joint probability of simultaneously finding an object within solid angle $d\Omega_1$ and a second object within solid angle $d\Omega_2$ at angular distance θ from $d\Omega_1$,

$$dP = N^2[1+w(\theta)]d\Omega_1 d\Omega_2, \tag{1}$$

where N is the surface number density of objects in the sample and $w(\theta)$ is the angular two-point correlation function for separation θ. Given a sample of positions on the sky, one can estimate w; however the quantity of interest is the spatial two-point correlation function, $\xi(r)$, which is similarly defined through the joint probability of finding objects within volume elements dV_1 and dV_2 separated by distance r,

$$dP = n^2[1+\xi(r)]dV_1 dV_2, \tag{2}$$

where n is the volume number density of objects. Note that,

$$dP = n[1+\xi(r)]dV, \tag{3}$$

is the conditional probability of finding an object within dV at distance r from a given object.

The angular and spatial three-point correlation functions are defined through the joint probability of finding a triplet of objects,

$$dP = N^3[1+w(\theta_{12})+w(\theta_{23})+w(\theta_{31})+z(\theta_{12},\theta_{23},\theta_{31})]d\Omega_1 d\Omega_2 d\Omega_3, \tag{4}$$

$$dP = n^3[1+\xi(r_{12})+\xi(r_{23})+\xi(r_{31})+\zeta(r_{12},r_{23},r_{31})]dV_1 dV_2 dV_3, \tag{5}$$

where z and ζ are the angular and spatial three-point correlation functions, respectively.

The above definitions are rather dry and do not provide much insight into what the coorelation functions really measure. The following qualitative interpretations of equations (2) and (3) may help to provide a "feel" for the two-point correlation function. First, equation (2) is equivalent to the statement that ξ is the normalized auto-correlation function of the galaxy density distribution, ρ,

$$\xi = (\rho-\bar{\rho})*(\rho-\bar{\rho})/\bar{\rho}^2. \tag{6}$$

With equation (3), $\xi(r)$ may be interpreted as the average correlated or clustered density (in units of the mean density) of galaxies at

https://doi.org/10.1017/S0074180900068285 Published online by Cambridge University Press

distance r from an "average" galaxy. Finally, $\xi(r) \propto \delta\rho/\bar{\rho}$, the average density contrast in a cluster of size r around an "average" galaxy.

2.2. Relations Between the Angular and Spatial Functions

The relation between the angular and spatial two-point correlation functions was derived by Limber (1953); extensions to the relativistic case and relations for the three point functions are given by Groth and Peebles (1977). In brief, the angular functions can be written as integrals of the spatial functions along the line-of-sight,

$$N d\Omega = \int(\text{Volume Element})\times(\text{Comoving Density}) \times(\text{Probability of Seeing an Object}), \quad (7)$$

$$N^2 w(\theta) d\Omega_1 d\Omega_2 = \iint(\text{Volume 1})\times(\text{Volume 2}) \times(\text{Density 1})\times(\text{Density 2}) \times(\text{Probability 1})\times(\text{Probability 2}) \times\xi(r_{12}), \quad (8)$$

with a similar expression for the three-point functions. Note that the probability of seeing an object depends on the luminosity function as well as the method of construction of the sample. The usual procedure for estimating the spatial function given the angular function is to guess a spatial function, choose a luminosity function, plug them into the integral, and iterate until the resulting angular function agrees with that desired. In special cases, the evaluation of the integrals is straightforward; for example, a power law spatial function yields (at small angles) a power law angular function with the power law exponent increased by 1. Fall and Tremaine (1977) have presented a method of solving the integral equation (8).

2.3. Scaling of the Angular Functions with the Depth of the Sample

A characteristic depth, D, may be associated with a magnitude limited sample. For example, D may be the distance at which a galaxy of absolute magnitude M appears at the limiting magnitude of the sample. In the absence of relativistic effects, k corrections, and evolutionary effects, $D \propto \text{dex}(0.2m)$, where m is the limiting magnitude of the sample. A key check that the correlation functions are measuring properties of the galaxy distribution and not local effects is provided by the scaling law relating angular two-point functions for two samples of depths D_1 and D_2,

$$w_2(\theta) = (D_1/D_2) w_1[(D_2/D_1)\theta]. \quad (9)$$

A similar scaling law applies to the three-point function, The factor D_1/D_2 accounts for the greater overlap of clusters along the line of sight as the depth of the sample increases while the

https://doi.org/10.1017/S0074180900068285 Published online by Cambridge University Press

factor D_2/D_1 accounts for the smaller angular size of a given physical structure as the depth of the sample increases. The small relativistic, k, and evolutionary corrections to the simple scaling law are discussed by Groth and Peebles (1977).

2.4. Advantages and Disadvantages of Correlation Function Techniques

Among the advantages of correlation function techniques are the following:

1. The method is straightforward to apply.
2. Only positional information is required.
3. The method is quantitative and in some cases fairly precise.
4. There are well defined relations connecting the angular and spatial functions.
5. The absence of significant local effects (e.g. variable obscuration) can be verified.
6. Correlation functions are easy to compare with theoretical predictions.

Some disadvantages are:

1. A large sample is usually required.
2. The method is insensitive to some features of the galaxy distribution (e.g. holes, filaments, sheets).
3. No specific information concerning any particular cluster or galaxy is obtained.

3. GALAXY CLUSTERING

3.1. Catalogs

Angular correlation functions have been estimated for three magnitude limited samples of galaxies: the Zwicky catalog (Zwicky *et al.* 1961-68), the Lick Catalog (Shane and Wirtanan 1967, Seldner *et al.* 1977), and the Jagellonian field (Rudnicki *et al.* 1973). Properties of these catalogs are summarized in the following table.

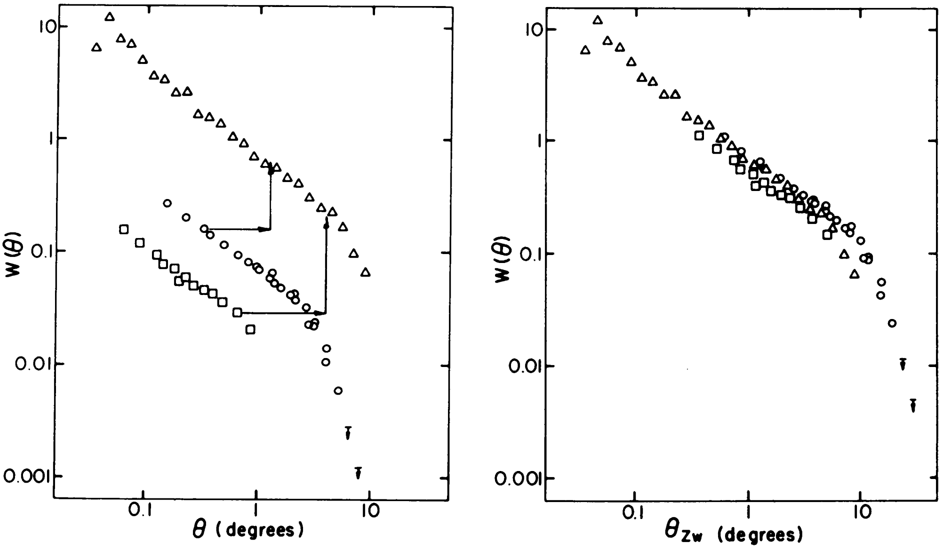
Table 1 Catalogs Analysed

Sample	Limiting Magnitude	Resolution Arcmin	Solid Angle Square Degrees	Typical Redshift	Galaxies
Zwicky	14.9	~2	~6000	0.016	~3700
Lick	18.6	10	~11000	0.07	~590000
Jagellonian	20.3	~3	36	0.13	~12000

https://doi.org/10.1017/S0074180900068285 Published online by Cambridge University Press

3.2. Results

The angular two-point correlation functions for the three samples are shown in Figure 1a. When the Lick and Jagellonian samples are scaled (eq. [9]) to the depth of the Zwicky sample, the curves appear as in Figure 1b. From these data and other data on the



(a) (b)

Figure 1. (a). The angular two-point correlation functions for the Zwicky (triangles), Lick (circles), and Jagellonian (squares) samples. (b). Same as (a) except the Lick and Jagellonian samples have been scaled to the depth of the Zwicky sample.

three-point function, the following results have been obtained:

1. The agreement between the scaled functions indicates that spurious correlations due to variable obscuration are not a problem with these data.

2. Over the range 0.05 Mpc $\lesssim$ hr $\lesssim$ 9Mpc the two-point correlation function is well represented by the power law

$$\xi(r) = (r_o/r)^{\gamma}, \quad \gamma = 1.77 \pm 0.04, \quad hr_o = 4.7\text{Mpc}, \tag{10}$$

where h is Hubble's constant in units of 100 km s^{-1} Mpc^{-1} and r_o is uncertain by about 50% due to uncertainties in the luminosity function.

3. At hr ~ 9Mpc, the correlation function has a "break". For larger distances, it falls much more rapidly than the power law in equation (10).

4. The three-point correlation function is well represented by the model

$$\zeta(r_1,r_2,r_3) = Q[\xi(r_1)\xi(r_2)+\xi(r_2)\xi(r_3)+\xi(r_3)\xi(r_1)], \tag{11}$$

where $Q = 1.29 \pm 0.21$.

3.3. Inferences

From the results presented above, a number of (sometimes controversial) inferences have been drawn:

1. The power law (eq. [10]) may indicate that the mechanism of cluster formation has no preferred scales. One such mechanism is the "gravitational instability picture" (Peebles 1974).
2. The form of the three-point function (eq. [11]) suggests that galaxies are distributed in a continuous clustering hierarchy (Soneira and Peebles 1978, Groth *et al.* 1977).
3. The gravitational instability picture, assuming white noise density fluctuations at recombination, predicts a power law exponent of 1.8, in good agreement with the observed index (Peebles 1974).
4. The break in the two-point function can be interpreted as the transition between linear (longward of the break) and non-linear growth of density fluctuations. The fact that $\xi \sim 0.3$ at the break may indicate that the density is close to the closure density (Davis, Groth, and Peebles 1977).

4. FUTURE DIRECTIONS

Although much has been learned via correlation function techniques, future applications should continue to add substantially to our understanding of galaxy clustering. An exciting development will be the construction of a redshift sample complete to ~15th magnitude and covering a substantial portion of the galactic polar cap. Such a catalog will allow the determination of the galaxy two-point function in both position *and velocity* space.

However, if the discussion is restricted to what can be learned from positional information only, then deeper samples are required. Questions which might be studied with deeper samples are:

1. Is the Universe statistically homogeneous and isotropic?
2. Does the power law (eq. [10]) extend all the way down to galactic dimensions, or is there a discontinuity or other feature at small scales (Peebles 1974)?
3. Is the break (seen mainly in the Lick sample) real or only an artifact of the analysis?

4. How has galaxy clustering (as measured by the correlation functions) evolved? Are clusters still growing?

Since the angular correlation functions decrease with the depth of the sample due to overlapping of clusters along the line of sight, the correlation functions for a sample significantly deeper than those listed in Table 1 will be very small. This means that extension of correlation function techniques to deeper samples will be difficult and will require great care in the construction of the sample. Nevertheless, the questions to be addressed are so interesting that several investigators have already begun to construct the required samples. Kron (1978) and Tyson and Jarvis (1979) have generated several deep samples using 4m prime focus plates. Phillipps *et al.* (1978, see also Ellis *et al.* 1977) are using 1.2m Schmidt plates. In the past year, I have obtained 4m prime focus plates for two fields of approximately $1^\circ \times 10^\circ$. The sample to be constructed from these plates will be especially useful for verifying the existence of the break.

With ground based photography (to obtain the large field of view required) one may perhaps construct a sample with typical redshift ~0.4. With the Space Telescope, the typical redshift will be ~1. At this redshift the angular size of the break will be $\sim 0\overset{\circ}{.}25$, so the small field of view of the ST should not be a major limitation. Thus, the statistical study of galaxy clustering at cosmologically significant distances promises to be an exciting and active field of research in the coming decade.

ACKNOWLEDGMENT

This research was supported in part by the National Science Foundation.

REFERENCES

Davis, M., Groth, E. J., and Peebles, P. J. E.: 1977, *Ap. J. Letters*, 212, p. L107.
Ellis, R. S., Fong, R., and Phillipps, S.: 1977, *M.N.R.A.S.*, 181, p. 163.
Fall, S. M., and Tremaine, S.: 1977, *Ap. J.*, 216, p. 682.
Groth, E. J., and Peebles, P. J. E.: 1977, *Ap. J.*, 217, p. 385.
Groth, E. J., Peebles, P. J. E., Seldner, M., and Soneira, R. M.: 1977, *Scientific American*, 237, No. 5, p. 76.
Kron, R. G.: 1978, Ph.D. Thesis, University of California.
Limber, D. N.: 1953, *Ap. J.*, 117, p. 134.
Neyman, J., Scott, E. L., and Shane, C. D.: 1953, *Ap. J.*, 117, p. 92.
Peebles, P. J. E.: 1973, *Ap. J.*, 185, p. 413.

Peebles, P. J. E.: 1974, *Ap. J. Letters*, 189, p. L51.
Phillipps, S., Fong, R., Ellis, R. S., Fall, S. M., and MacGillivray, H. T.: 1978, *M.N.R.A.S.*, 182, p. 673.
Rudnicki, K., Dworak, T. Z., Flin, P., Baranowski, B., and Sendrakowski, A.: 1973, *Acta Cosmologica*, 1, p. 7.
Seldner, M., and Peebles, P. J. E.: 1978, *Ap. J.*, 225, p. 7.
Seldner, M., Siebers, B., Groth, E. J., and Peebles, P. J. E.: 1977, *A. J.*, 82, p. 249.
Shane, C. D., and Wirtanen, C. A.: 1967, *Pub. Lick Obs.*, vol. 22, Part 1.
Soneira, R. M., and Peebles, P. J. E.: 1978, *A. J.*, 83, p. 845.
Tyson, J. A., and Jarvis, J. F.: 1979, *Ap. J.*, 230, p. L153.
Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., and Kowal, C. T.: 1961-1968, *Catalogue of Galaxies and Clusters of Galaxies*, in 6 vols. (Pasadena: California Institute of Technology).

DISCUSSION

Rees: Do you think that galaxy correlation studies will eventually be able to detect "linear" fluctuations ($\lesssim$ 10%, say) on scales greater than 100 Mpc; or can this be done better by studying the distribution of rarer classes of objects such as X-ray sources or radio sources?

Groth: I can't give a definite answer to this question. However, to set the scale, note that the typical depth of the Lick sample is ~ 200 h^{-1} Mpc. My guess is that it will be very difficult to get at clustering on scales of 100 Mpc with correlation studies of galaxies.

Tyson: Recently Kirshner, Oemler, and Schechter (Astron. J. 83, 1978) have obtained data on magnitudes, positions, and redshifts for a complete sample of galaxies to 16th magnitude and find no evidence for a drop-off at 9 Mpc. Would you please comment on this?

Groth: The KOS sample includes positions and redshifts. The redshifts were used to obtain positions along the line of sight. Thus, KOS had a sample of three-dimensional positions from which they could estimate ξ directly, rather than going through the calculation of the angular function. This has the advantage that ξ, having a much larger amplitude than w, is much easier to estimate from a small sample. On the other hand, this method requires that peculiar velocities be negligible. For example, if "field" galaxies, such as those in the KOS sample, have peculiar velocities of the order of a few hundred km s^{-1}, then the estimated ξ would be smeared out over several Mpc.

I believe that the proper way to treat a position-redshift sample is not to convert velocities to distance, but to try to estimate the correlation function in both position and velocity space, ξ (r, v). This will allow the study of dynamics of clusters as well as the shapes of clusters.