Fermi National Accelerator Laboratory

FERMILAB-Pub-97/022-A

Cumulant Correlators from the APM

Istvan Szapudi

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

Alexander S. Szalay

Johns Hopkins University
Baltimore, Maryland 21218

February 1997

Submitted to *Astrophysical Journal Letters*

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CH03000 with the United States Department of Energy

FERMILAB-Pub-97/022-A

Cumulant Correlators from the APM

István Szapudi[1] and Alexander S. Szalay[2]

[1]NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510-0500

[2] Johns Hopkins University, Baltimore, MD 21218

Abstract

This work presents a set of new statistics, the cumulant correlators, aimed at high precision analysis of the galaxy distribution. They form a symmetric matrix, Q_{NM}, related to moment correlators the same way as cumulants are related to the moments of the distribution. They encode more information than the usual cumulants, S_N's, and their extraction from data is similar to the calculation of the two-point correlation function. Perturbation theory (PT), its generalization, the extended perturbation theory (EPT), and the hierarchical assumption (HA) have simple predictions for these statistics. As an example, the factorial moment correlators measured by Szapudi, Dalton, Efstathiou & Szalay (1996, hereafter SDES) in the APM catalog are reanalyzed using this technique. While the previous analysis assumed hierarchical structure constants, this method can directly investigate the validity of HA, along with PT, and EPT. The results in agreement with previous findings indicate that, at the small scales used for this analysis, the APM data supports HA. When all non-linear corrections are taken into account it is a good approximation at the 20 percent level. It appears that PT, and a natural generalization of EPT for cumulant correlators does not provide such a good fit for the APM. Once the validity the HA is approximately established, cumulant correlators can separate the amplitudes of different tree-types in the hierarchy up to fifth order. As an example, the weights for the fourth order tree topologies are calculated including all non-linear corrections.

keywords large scale structure of the universe – galaxies: clustering – methods: numerical – methods: statistical

1. Introduction

Direct determination of higher order correlation functions (Fry & Peebles 1978, Peebles 1980, and references therein) is burdened with the combinatorial explosion of terms, which severely complicates their measurement and interpretation. Thus in the recent years indirect methods became increasingly popular for high precision measurements of higher order correlations. The simplest of these methods consists of calculating the (factorial) moments of the distribution of counts in cells, and from that, the cumulants, S_N's, of the underlying distribution (see e.g. Peebles 1980, Gaztañaga 1992, Bouchet *et al.* 1993, Gaztañaga 1994, Colombi *et al.* 1995, Szapudi, Meiksin, & Nichol 1996). For a point process, these quantities measure the amplitude of the N-point correlation function averaged in a particular window. The advantages of this technique lie in its simplicity, and its direct relation to the predictions of PT (Peebles 1980, Juszkiewicz, Bouchet, & Colombi 1993, Bernardeau 1992, Bernardeau 1994, EPT (Colombi *et al.* 1996) and the HA (Peebles 1980). Since the averaging causes a significant loss of information, alternative methods based on moment correlators use a pair of cells (Szapudi, Szalay & Boschán 1992, Meiksin, Szapudi, & Szalay 1992, SDES). In the past such methods were used mainly to estimate the average amplitude of the different N-point correlation functions in the HA, the Q_N's, motivated by the theory of the BBKGY equations in the strong clustering regime. This work presents an alternative analysis of the factorial moment correlators which is free of assumptions, except for the widely accepted infinitesimal Poisson model to relate the continuum limit quantities to the measured discrete process. Instead of fitting for the Q_N, a matrix Q_{NM} is defined: the cumulant correlators. Both HA and PT have specific predictions for these possibly scale dependent quantities. After elaborating these predictions, the method is illustrated by reanalyzing the factorial moment correlators obtained from the APM catalog by SDES. Once the HA is established, cumulant correlators contain enough information to separate the weights of different tree topologies up to fifth order. The next section outlines the basic theory, section §3 presents the predictions of PT, EPT, and HA. The measurements of the 4th order coefficients of the hierarchy from the APM catalog are described in section §4.

2. Cumulant Correlators

Following SDES we define the factorial moment correlators for a pair of cells separated by a distance r_{12} as

$$w_{kl}(r_{12}) = \frac{\langle (N_1)_k (N_2)_l \rangle - \langle (N)_k \rangle \langle (N)_l \rangle}{\langle N \rangle^{k+l}}, \quad k \neq 0, l \neq 0, \tag{1}$$

and the normalized factorial moments for a single cell

$$w_{k0} = \frac{\langle (N)_k \rangle}{\langle N \rangle^k}. \tag{2}$$

The notation $(N)_k = N(N-1)..(N-1+k)$ is introduced for the factorial moments of the counts in cells, $\langle\rangle$ denotes averaging over all cell positions in the survey. The generating function for the factorial moments in terms of the cumulants Q_N is

$$W(x) = \exp \sum_{N=1}^{\infty} \Gamma_N x^N Q_N, \tag{3}$$

with

$$\Gamma_N = \frac{N^{N-2}\xi_s^{N-1}}{N!}, \tag{4}$$

where ξ_s is the average of the two point correlation function over the cell. The generating function can be written in the above form for any distribution that has cumulants. Generally, the Q_N's can have a scale dependence, while for the HA $Q_N = \mathrm{const}$ is expected. Similarly, the generating function of the factorial moment correlators can be written as

$$W(x,y) = W(x)W(y)\left(\exp Q(x,y) - 1\right), \tag{5}$$

with

$$Q(x,y) = \xi_l \sum_{M=1,N=1}^{\infty} x^M y^N Q_{NM} \Gamma_M \Gamma_N NM. \tag{6}$$

This latter equation defines the cumulant correlators, Q_{NM}, with $\xi_l = w_{11}$, the two-point correlation function. Note, that the linear dependence is factored out, however, Q_{NM} is not necessarily a constant.

Cumulants and cumulant correlators are related to the continuum limit connected moments because of the continuum properties of the factorial moments

$$\frac{\left\langle \delta_1^N \right\rangle}{N!} = Q_N \Gamma_N \tag{7}$$

$$\frac{\left\langle \delta_1^N \delta_2^M \right\rangle_c}{N!M!} = Q_{NM} \Gamma_M \Gamma_N NM. \tag{8}$$

Although the above equations are formally identical to SDES, there are two subtle differences: there is no reference to the hierarchical assumption, therefore Q_{NM} becomes a matrix, and it is understood as an *exact* equation, i.e. the non-linearities are included. It is convenient to define also linear order cumulant correlators, denoted by $\tilde{Q}_{NM}$, which are obtained from the generating function with the approximation of $\exp Q(x,y) - 1 \simeq Q(x,y) + \mathcal{O}(\xi_l^2)$. The $\tilde{Q}_{NM}$'s coincide up to linear order and normalization with the C_{NM}'s calculated from PT by Bernardeau 1995 (see next section).

The quantities Q_N and Q_{NM} can be calculated for any well behaved point process. The calculation simply involves expanding $W(x,y)/[W(x)W(y)]$ according to equation 5. For instance the third and fourth order moments are

$$Q_{12}\Gamma_1\Gamma_2 2\xi_l = w_{12}/2 - \xi_l \tag{9}$$

$$Q_{13}\Gamma_1\Gamma_3 3\xi_l = w_{13}/6 - w_{12}/2 - w_{20}/2 + \xi_l \tag{10}$$

$$Q_{22}\Gamma_2^2 4\xi_l = w_{22}/4 - w_{12} + \xi_l - \xi_l^2/2, \tag{11}$$

where ξ_l is the usual two-point correlation function between the two cells, w_{11}. The higher order moments can be obtained from the generating function analogously. Note, that $Q_{22} = \tilde{Q}_{22} - \xi_l/2\xi_s^2$.

3. Predictions

3.1. Hierarchical Assumption

In the highly nonlinear regime $\xi_2 \gg 1$, an ansatz for the structure of the N-point correlation functions is the HA (e.g., Peebles 1980; BS) stating that the N-point correlation functions can be written as a sum of products of $N-1$ two-point correlation functions. Each product corresponds to a tree spanning the N-points, and there is a summation over all possible trees. The different tree topologies, labeled with k, are weighted with a constant Q_{Nk}. Our notation in detail can be found in Boschán, Szapudi, & Szalay 1994, Szapudi & Colombi 1996. One of the goals of this paper is validate this assumption to an unprecedented accuracy.

Comparing Equation 6 with the corresponding equation in SDES, and Szapudi & Szalay 1993, and using $\exp Q(x,y) - 1 \simeq Q(x,y)$ yields a simple linear order prediction for the HA

$$Q_{N+M} \simeq \tilde{Q}_{NM} \simeq const. \tag{12}$$

For instance the 4th order cumulant Q_4 is approximately equal to the linear cumulant correlators $\tilde{Q}_{13}$, $\tilde{Q}_{22}$, and constant, etc. The weights for different tree topologies were implicitly equated with the average, and form factors of the two point window and non-linearities were neglected. While the form factors were shown to be negligible (Boschán, Szapudi, & Szalay 1994), different tree topologies and non-linear corrections will be taken into account next for a more accurate prediction.

The only 3rd order cumulant correlator is Q_{12}. Tree graphs spanning three points have only one possible topology (its weight denoted by Q_3 with form factors neglected), giving altogether three possible graphs.

$$\left\langle \delta_1^2 \delta_2 \right\rangle = 2Q_{12}\xi_s\xi_l = Q_3(2\xi_l\xi_s + \xi_l^2), \tag{13}$$

reproducing $Q_{12} = Q_3$ at linear order.

At fourth order there are two cumulant correlators Q_{13}, and Q_{22}. The sixteen possible trees spanning four points come in two distinct topologies: four "snake" graphs and twelve "star" graphs. Their respective amplitudes are denoted with R_a and R_b in the HA. Summing all possible graphs with the appropriate statistical weights gives

$$\left\langle \delta_1^3 \delta_2 \right\rangle_c = 9Q_{13}\xi_l\xi_s^2 = 6\xi_l\xi_s^2 R_a + 3\xi_l\xi_s^2 R_b + 6\xi_l^2\xi_s R_a + \xi_l^3 R_b, \tag{14}$$

and

$$\left\langle \delta_1^2 \delta_2^2 \right\rangle_c = 4Q_{22}\xi_l\xi_s^2 = 4\xi_l\xi_s^2 R_a + 4\xi_l^2\xi_s R_a + 4\xi_l^2\xi_s R_b + 4\xi_l^3 R_a. \tag{15}$$

These two equations are linear in R_a and R_b, therefore they can be solved yielding equations (with non-linear coefficients in terms of ξ) in terms of Q_{13} and Q_{22}. Note, that the linear solution is $R_a = \tilde{Q}_{22}$ and $R_b = 3\tilde{Q}_{13} - 2\tilde{Q}_{22}$.

3.2. Perturbation Theory

Direct comparison of equation 6 with the coefficients C_{NM} in Bernardeau 1995 reveals that they are identical to the linear order cumulant correlators up to normalization

$$C_{NM} = \tilde{Q}_{NM} N^{N-1} M^{M-1} + \mathcal{O}(\xi_l^2). \tag{16}$$

Perturbation theory predicts that the coefficients factorize such that

$$C_{NM} = C_{N1} C_{M1}, \tag{17}$$

and the series C_{N1} was calculated up to first non-trivial order. For instance

$$C_{2,1} = \frac{68}{21} - \frac{n+3}{3} \tag{18}$$

for a power spectrum with effective index n at the scale of the cells. The interested reader is referred to Bernardeau 1995 for detailed predictions in the weakly non-linear regime, for the present work only equation 17 is needed.

4. Measurements from the APM Catalog

4.1. Linear Cumulant correlators

For an initial assessment, the linear cumulant correlators were first calculated from the factorial moment correlators measured in the APM survey (Maddox et al. 1990a, Maddox et al. 1990a, Maddox et al. 1990c by SDES). Details on the properties of the density maps are found there. In what follows, a density map of cell size $0.23°$ and magnitude cut of $b_J = 17 - 20$ was used. The bottom panel of the Figure shows the measured $\tilde{q}_{NM}$'s (the linear projected cumulant correlators; lower case symbols refer to projected quantities) up to fifth order. The degeneracy and the approximate parallel nature of the curves immediately suggest that the HA is a reasonable approximation. At larger scales the cumulant correlators appear to roll off, while the prediction stays flat, and the degeneracy of the curves is slightly broken. This is mainly due to the approximations used: different tree topologies and non-linearities were not taken into account. Possibly, the effect is partly caused by the fact that the cumulants are not exactly constant at all scales as shown by Gaztañaga 1994, Szapudi, Meiksin, & Nichol 1996 (i.e. HA is slightly broken). In the next subsection it will be shown by improving on the approximations, that HA is a good approximation at the 20 percent level up to fourth order.

The middle panel of the Figure. illustrates equation 17 predicted by linear PT. The solid lines are the cumulant correlators $\tilde{q}_{NM}, N, M > 1$, while the dotted lines show the corresponding $\tilde{q}_{1N}\tilde{q}_{1M}$. Only the fourth and fifth order are shown. The degree of validity of PT can be judged from how well the dotted and solid lines match. Since the dotted lines appear to be consistently smaller than the solid ones this model provides a less accurate description of the data than HA. Possibly, second order PT could improve the representation of the data; it is left for future work.

Although perturbation theory is not expected to be valid at the scales of the cells, it was found in N-body simulations (Colombi *et al.* 1996), and galaxy data (Szapudi, Meiksin, & Nichol 1996), that the higher order correlation amplitudes, S_N, measured from counts in cells are similar to the one prescribed by PT, but with a steeper power spectrum. This phenomenological extension of PT is the essence of EPT. The previous exercise taken at face value would suggest that EPT cannot be generalized for moment correlators. Since moment correlators contain extra information compared to the moments, they might provide a means to differentiate between weakly non-linear distribution with low power index, and a truly non-linear distribution. However, the data are not conclusive, therefore the validity of a possible extension of EPT for two-point quantities in N-body simulations, and high quality data is left for future work.

4.2. Non-linear Cumulant Correlators

The approximate validity of the hierarchy established previously warrants a closer examination by relaxing the assumptions on the uniform weighting of the different tree topologies, and using the full non-linear description. The form factors resulting from the pair of cells are expected to be smaller than the measurement errors and will be neglected. Counting the number of degrees of freedom reveals that from the cumulants and cumulant correlators it is possible to separate the different tree topologies up to fifth order. The errors are a limiting factor for the fifth order, therefore a calculation for the third and fourth order is presented here. The fifth order calculation is analogous, although somewhat tedious. At higher than fifth order additional information is needed to separate the different graph types.

The long dashed line on the top panel of the Figure shows the non-linear measurement of q_3 as calculated from q_{21} of the APM according to Eq. 13. The dotted lines show the linear solution r_a, and r_b as computed from $\tilde{q}_{22}$, and $\tilde{q}_{31}$. The hierarchy predicts two horizontal lines, with the constraint that $16q_3 = 12r_a + 4r_b$. The linear approximations on the other hand show a strong scale dependence, increasing and even crossing over at the smallest scales: a possible sign of non-linear effects. The full non-linear equations (14, 15) yield the result plotted with solid lines: the non-linear corrections remove most of the scale dependence, as expected if HA is satisfied. The residuals are probably due to the neglected form factors, measurement errors. On the left side of the panel several amplitudes are plotted for comparison; for these points the angular scale is irrelevant. The three sided symbols refer to third order quantities, the four sided to fourth order. The filled triangles and squares shows the value of $q_3 = 1.15$ and $q_4 = 4.7$ calculated from the averaged value of $q_{21} = 1.15$, and $r_a = 5.3$, and $r_b = 1.15$, respectively. The open symbols correspond to the values of $q_3 = 1.7$, and $q_4 = 4.17$ measured from the factorial moments alone, w_{k0}, at the scale of the cell. For a comparison, the two stars show the respective measurements of SDES $q_3 = 1.16$, and $q_4 = 2.96$. The reason that SDES measured a somewhat lower q_4 is that they used linear approximations (dotted lines) only. The measurements of $q_3 = 1.7$ by Gaztañaga 1994 in the APM and $q_3 = 1.6$ Szapudi, Meiksin, & Nichol 1996 at similar scales are in excellent agreement with the same results from w_{k0}. The values for the fourth order in the same sources, $q_4 = 3.7$ and $q_4 = 3.2$, are slightly lower than above, but the agreement is still within $20 - 30$ percent.

The above numbers suggest that, while the different measurements using the same method are consistent with each other even in different catalogs, there is some disagreement between the results based on moment correlators and moments. The error distribution studied by Szapudi & Colombi 1996 provides useful clues to resolve this apparent discrepancy. Since the distribution of errors is positively skewed and increasingly so for higher order moments, an upward fluctuation is more likely than a downward. This effect is increasing with the order of the moments measured. In the method proposed by this work q_3 is estimated from the value of q_{21}. The behavior of the errors is similar to the multiple of a second and first order quantity, thus the variance is reduced. Note, that this is possible, only after the hierarchy is established, i.e. a prior information is used to reduce the scatter from cosmic errors.

An accurate error estimation in this case would involve a tedious calculation, a non-trivial generalization of Szapudi & Colombi 1996. The fluctuations on the Figure indicate roughly 20 percent errors. Although finite volume effects could introduce an overall shift as well, this effect is expected to be much less (Gaztañaga 1994).

The de-projection using the coefficients in SDES yields $Q_3 = 1$, $R_a = 3.7$, and $R_b = 0.8$, giving $Q_4 = 3$. This is to be compared with with Fry & Peebles 1978, where the direct determination of the four-point correlation function from the Lick catalog yielded $R_a = 2.5 \pm 0.6$ and $R_b = 4.3 \pm 1.2$. These results could give a clue for solving the BBKGY equations in the highly non-linear regime. The assumption of Hamilton 1988, that only the snake graphs have a contribution, appears to be close to our results: although both graph types have a contribution, the average is closer to the snake coefficient. The ansatz of Bernardeau & Schaeffer 1992, $\sqrt{R_a} \simeq Q_3$, is not a particularly good approximation. In conclusion the statistics of the cumulant correlators is in excellent agreement with HA. The method outlined here in conjunction with future data and N-body simulations will be able to pin down the amplitudes of the higher order correlations with unprecedented accuracy.

It is a pleasure to acknowledge discussions with F. Bernardeau, S. Colombi, and J. Frieman. The original measurement of the factorial moment correlators was carried out in collaboration with G. Dalton, and G. Efstathiou. I.S. was supported by DOE and NASA through grant NAG-5-2788 at Fermilab. A.S.Z. was supported by a NASA LTSA grant.

5. Figure Caption

Lower Panel. The linear cumulant correlators, $\tilde{q}_{nm}$, the main raw results of the paper are displayed up to fifth order. The parallel degenerate lines suggest the HA.

Middle Panel. The linear cumulant correlators are shown on a linear scale (solid lines) together with the prediction from PT (dotted line). The agreement is improving towards the higher scales.

Upper Panel. The hierarchical amplitudes as calculated from the fully non-linear cumulant correlators are displayed. The long dashed line corresponds to the estimator of q_3, the solid lines to the estimator of r_a, and r_b, the amplitudes of the fourth order snake, and star graphs, respectively. The dotted lines show the linear approximation, which breaks down at smaller scales at this level of precision. The filled symbols mark q_3 (triangle), and q_4 (square) as calculated from the moment correlators. The open symbols are the same as measured from the moments of counts in cells only. Finally, the crosses show the measurements of q_3 (triangular), and q_4 (square) by SDES for comparison.

REFERENCES

Bernardeau, F. 1992, ApJ, 292, 1

Bernardeau, F. 1994, ApJ, 433, 1

Bernardeau, F. 1995, A&A, 301, 309

Bernardeau, F. 1995, & Schaeffer, R. 1992, A&A, 255, 1

Boschán, P., Szapudi, I., & Szalay, A. 1994, ApJS, 93, 65

Bouchet, F.R., Strauss, M.A., Davis, M., Fisher, K.B., Yahil, A., & Huchra, J.P. 1993, ApJ, 417, 36

Colombi, S., Bouchet, F.R., & Hernquist, L. 1995, A&A, 281, 301

Colombi, S., Bernardeau, F., Bouchet, F.R., & Hernquist, L. 1996, (astro-ph 9610253)

Fry, J.N., & Peebles, P.J.E. 1978, ApJ, 221, 19

Gaztañaga, E. 1992, ApJ, 319, L17

Gaztañaga, E. 1994, MNRAS, 268, 913

Hamilton, A.J.S. 1988, ApJ, 332,67

Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, ApJ, 412, L9

Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, L. 1990a, MNRAS, 242, 43P

Maddox, S. J., Sutherland, W. J., Efstathiou, G., & Loveday, L. 1990b, MNRAS, 243, 692

Maddox, S. J., Sutherland, W. J., Efstathiou, G., & Loveday, L. 1990b, MNRAS, 246, 433

Meiksin, A., Szapudi, I., & Szalay, A., 1992, ApJ, 394, 87

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

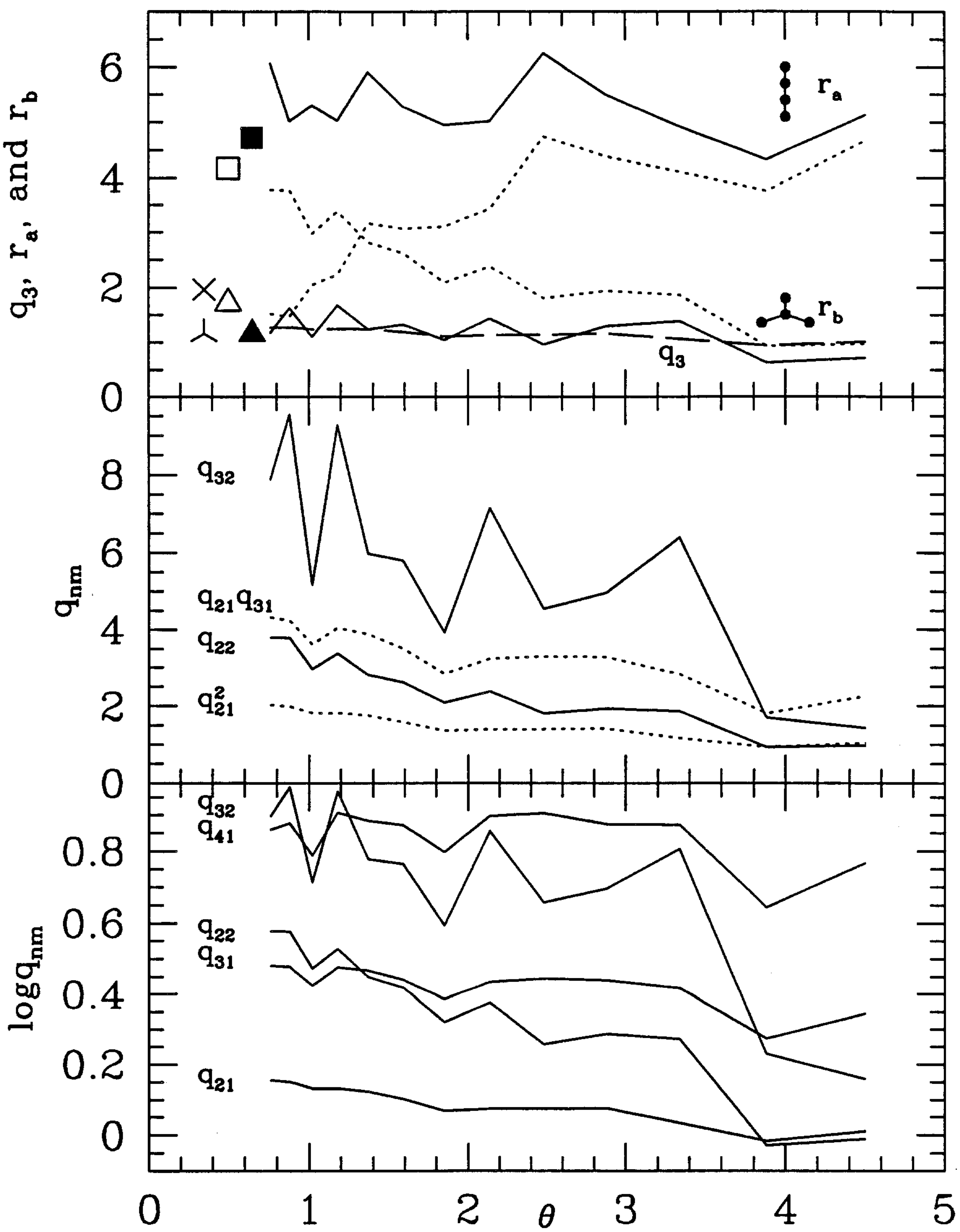
Szapudi, I., & Colombi, S. 1996, ApJ, 470, 131

Szapudi, I., Dalton, G., Efstathiou, G.P., & Szalay, A. 1995, ApJ, 444, 520

Szapudi, I., Meiksin, A., & Nichol, R.C. 1996, ApJ, 473, 15

Szapudi, I. & Szalay, A. 1993, ApJ, 408, 43

Szapudi, I., Szalay, A., & Boschán, P. 1992, ApJ, 390, 350



q_3, r_a, and r_b
r_a
r_b
q_3
q_{nm}
q_{32}
$q_{21}q_{31}$
q_{22}
q_{21}^2
$\log q_{nm}$
q_{32}
q_{41}
q_{22}
q_{31}
q_{21}
θ