Conditional cumulants in a weakly non-linear regime

Jun Pan*† and István Szapudi

Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA

Accepted 2005 April 29. Received 2005 April 7; in original form 2004 May 28

ABSTRACT

We introduce conditional cumulants as a set of unique statistics closely related to \( N \)-point correlation functions and to cumulants of moments of counts in cells. We show that they can be viewed in three equivalent ways: (i) as particular integrals of the \( N \)-point correlation functions, (ii) as integrated monopole moments of the bispectrum, and (iii) as statistics associated with neighbour counts. As monopole statistics, they carry similar information to the cumulants \( S_N \), the most widely spread higher-order statistics usually measured from counts in cells. While it has been proved that counts in cells can only be approximately corrected for edge effects, we show that well-tested, edge-corrected estimators can be successfully adapted for conditional cumulants. Since edge-effect errors typically dominate large scales, it is expected that it will be possible to measure conditional cumulants with higher accuracy in the interesting large-scale regime. To lay the theoretical ground work for future applications, we compute the predictions of weakly non-linear perturbation theory for conditional cumulants. We demonstrate the use of edge-corrected estimators in a set of simulations and measure conditional cumulants, and compare the results with our theoretical predictions in real and redshift space. We find excellent agreement, especially on scales \( \gtrsim 20 \, h^{-1} \) Mpc. Owing to their advantageous statistical properties and well-understood dynamics, we propose conditional cumulants as tools for high-precision cosmology. Potential applications include constraining bias and redshift distortions from galaxy redshift surveys.

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

1 INTRODUCTION

Large-scale structure statistics of higher than second order contain a wealth of information on cosmological parameters, gravitational amplification of initial fluctuations, and on structure formation in general. In particular, higher-order statistics have the potential to provide some of the best constraints on the phenomenon of ‘biasing’ (Kaiser 1984). While the core ideas have been known for over two decades (cf. Peebles 1980), the latest wide-field galaxy surveys, such as the Sloan Digital Sky Survey (e.g. York et al. 2000, SDSS), and the 2-degree Filed Galaxy Redshift Survey (Colless et al. 2001, 2dF), have motivated a concerted effort to enhance theories and techniques of higher-order statistics to the level of ‘high-precision cosmology’ (for a summary, see Bernardeau et al. 2002).

Counts in cells (CIC) and related statistics have the best-understood theoretical background and consequently have yielded some of the most successful applications on galaxy surveys both in two dimensions (e.g. Gaztañaga & Frieman 1994; Szapudi, Meiksin & Nichol 1996; Magliocchetti et al. 1998; Szapudi et al. 2002) and in three dimensions (e.g. Hoyle, Szapudi & Baugh 2000; Szapudi et al. 2000; Croton et al. 2004). The estimation methods and the corresponding errors have also been worked out in detail (e.g. Szapudi & Colombi 1996; Szapudi, Colombi & Bernardeau 1999). When CIC are measured to constrain theories at the level of a few per cent, edge effects present a major problem as a result of the complex geometry and cut-out holes of realistic surveys. As shown by Szapudi & Colombi (1998a), edge effects cannot be corrected for exactly, while an approximate estimator exists if the shape dependence of counts is weak (Szapudi 1998a).

Edge-effect correction is possible, though, when using the class of estimators introduced by Szapudi & Szalay (1998) for the \( N \)-point correlation functions. \( N \)-point correlation functions are, however, inherently more complicated objects than cumulants of CIC. They depend on a large number of parameters, their measurement is computationally intensive (Moore et al. 2001), and consequently their interpretation is difficult. Much of the theoretical effort has been concentrated on the three-point function, yet weakly non-linear perturbation theory and halo models have only limited success when contrasted with simulations (Barriga & Gaztañaga 2002; Takada & Jain 2003). Redshift distortions present an even more formidable challenge.

The goal of the present paper is to explore a set of statistics, the conditional cumulants, a goal that is inspired by our effort to
combine the simplicity and transparency of cumulants with the estimator of the N-point correlation. They can be viewed as integrated N-point correlation functions, or integrals of the monopole moment of the bispectrum (Szapudi 2004). They are also closely related to (factorial) moments of neighbour counts (Peebles 1980). While they have been known for some time, neighbour counts have been used fairly infrequently in comparison to cumulants (e.g. Borgani 1995). The terminology ‘conditional cumulants’ was introduced by Bonometto et al. (1995). They used an estimator based on moments of neighbour counts and developed a theory under assumptions of stable clustering and scale invariance.

In Section 2, we present a formal definition of the conditional cumulants and their relevant properties. Predictions of the third-order conditional cumulant in the weakly non-linear regime are described in Section 3. In Section 4, we adapt the edge-corrected estimator derived by Szapudi & Szalay (1998) to measure conditional cumulants, and compare predictions with measurements in simulations. In Section 5 we summarize results, present the theory in redshift space together with simulation results, and discuss implications for estimating bias.

2 CONDITIONAL CUMULANTS

Conditional cumulants are defined as the joint connected moment of one unsmoothed and N − 1 smoothed density fluctuation fields. They are realized by integrals of the N-point correlation function through N − 1 spherical top-hat windows,

\[ U_N(r_1, \ldots, r_{N-1}) = \int \xi_N(s_1, \ldots, s_{N-1}, 0) \prod_{i=1}^{N-1} d^3 s_i \frac{W_r(s_i)}{V_i}, \]

(1)

where \( V_i = 4\pi r_i^3 / 3 \) is the volume of the window function \( W_r \). In the most general case, each top-hat window might have a different radius. Further simplification arises if all the top hats are the same, i.e. we define \( U_N(r) \) with \( r_1 = \ldots = r_{N-1} = r \) as the conditional cumulant (cf. Bonometto et al. 1995). The \( U_N \) is subtly different from the usual cumulant of smoothed-field \( \xi_N \) by one less integral over the window function.

The second-order cumulant, \( U_2 \), is equivalent to the confusingly named \( J_3 \) integral (e.g. Peebles 1980):

\[ U_2(r) = \frac{3}{r^3} J_3(r) = \frac{1}{(2\pi)^3} \int P(k) w(kr) 4\pi k^2 dk. \]

(2)

where \( w(kr) = 3(\sin kr - kr \cos kr) / (kr)^3 \) is the Fourier transform of \( W_r \), and \( P(k) \) is the power spectrum.

For higher orders, we can construct reduced conditional cumulants as

\[ R_N(r) = \frac{U_N(r)}{U_2^{N-1}(r)}. \]

(3)

\( U_N \) and \( R_N \) have a close connection with moments of neighbour counts (e.g. Peebles 1980). Let us define the partition function \( Z[J] = \langle \exp \int iJ \rho \rangle \) (cf. Szapudi & Szalay 1993), where \( \rho = [1 + \delta] \) is the smoothed density field. Then we can use the special source function \( iJ(x) = W(x) s + \delta \rho(x) \) to obtain the generating function \( G(s, t) \). This is related to the generating function of neighbour-count factorial moments as \( G(s) = \partial_t G(s, t) |_{t=0} \). The final result is

\[ G(s) = \sum_{M \geq 0} \frac{(snV)^M}{M!} U_{M+1} \exp \sum_{N \geq 1} \frac{(snV)^N}{N!} \xi_N. \]

(4)

where \( nV = \bar{N} \) is the average count of galaxies, and \( \xi_1 = U_1 = 1 \) by definition. This generating function can be used to obtain \( U_N \) and/or \( R_N \) from neighbour-count factorial moments in a way analogous to the way in which the generating functions in Szapudi & Szalay (1993) are used to obtain \( S_N \)s from factorial moments of CIC.

For completeness, the generating function for the neighbour-count distribution is obtained by the substitution \( s \rightarrow s - 1 \), while the ordinary moment generating function by \( s \rightarrow e^s - 1 \). We have checked that we recover the formulae of Peebles (1980), section 36, from \( G(e^s - 1) \). The above generating function facilitates the extraction of \( U_N \) from neighbour-count statistics. Further details can be found in Szapudi & Szalay (1993): the entire theory for CIC can be adapted to neighbour counts. So far our discussion has been general; in what follows we will focus on the first non-trivial conditional cumulant, \( U_3 \).

\[ U_3(r_1, r_2) = \frac{1}{\bar{V}^2} \int B(k_1, k_2, k_3) \delta_0(k_1 + k_2 + k_3) \]

\[ w(k_1 r_1) w(k_2 r_2) d^3 k_1 d^3 k_2 d^3 k_3, \]

(5)

where \( \delta_0 \) is the Dirac delta function. To further elucidate the above relation, we use the multipole expansion of the bispectrum and the three-point correlation function proposed by Szapudi (2004):

\[ B(k_1, k_2, \theta) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} U_l(k_1, k_2) P_l(\cos \theta); \]

\[ \zeta(r_1, r_2, \theta) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \zeta_l(r_1, r_2) P_l(\cos \theta), \]

where \( \cos \theta = k_1 k_2 / (k_1 k_2) \) or \( r_1 r_2 / r_1 r_2 \), and \( P_l \) are Legendre polynomials. The multipole moments can be obtained as \( B_l = 2\pi \int B P_l d \cos \theta \), \( \zeta_l = 2\pi \int \zeta P_l d \cos \theta \). Substituting into the general equation, we find

\[ U_3(r_1, r_2) = \frac{4\pi}{V_1 V_2} \int_0^{r_1} \int_0^{r_2} \delta_V(\xi (r_1, r_2) r_1^2 r_2^2) d^9 \Xi \]

\[ = \frac{4\pi}{(2\pi)^6} \int d^3 k_1 d^3 k_2 \frac{3k_1^3}{r_1} j_3(k_1 r_1) j_3(k_2 r_2) B_0(k_1, k_2), \]

(7)

in which \( j_3 \) is the first-order spherical Bessel function. It can be seen that \( U_3 \) depends only on the monopole moment of the bispectrum/three-point correlation function. This property significantly simplifies the transformation of the statistics under redshift distortions.

3 U_3 IN A WEAKLY NON-LINEAR REGIME

On large scales, where the fluctuations are reasonably small, clustering of cosmic structures can be understood in Eulerian weakly non-linear perturbation theory (EPT) (Bernardeau et al. 2002, and references therein). To predict the behaviour of \( U_3 \) from Gaussian initial conditions, we assume an expansion of the density field into first, second, etc. order: \( \delta = \delta^{(1)} + \delta^{(2)} + \ldots \) EPT can then be used to calculate the leading-order contribution to \( U_3 = \langle \delta(0) \delta_{1c}^2 \rangle \), where \( \delta_1 \) is the density field filtered at the scale \( r \), and \( \langle \rangle_1 \) means connected moment. We use the second-order EPT kernel (Fry 1984; Goroff et al. 1986),

\[ F_2(k, k') = \frac{10}{7} + k \cdot k' k \cdot k' k + \frac{4}{7} \left( \frac{k \cdot k}{kk'} \right)^2, \]

(8)

the linear power spectrum \( P(k) \), and integrals of the kernel multiplied with the top-hat window function (Bernardeau 1994a) to finally obtain

\[ R_3(r_1, r_2) = \frac{U_3(r_1, r_2)}{U_3^2(r_1) U_3^2(r_2)} \]

\[ = \frac{34}{21} \left[ 1 + \frac{2}{3} \frac{\xi_{(r_1, r_2)}}{U_3(r_1) U_3(r_2)} + \frac{1}{3} \frac{\xi_{(r_1, r_2)}}{U_3(r_1) U_3(r_2)} \right] \]

\[ + \frac{1}{3} \frac{\xi_{(r_1, r_2)}}{U_3(r_1) U_3(r_2)} \left[ \frac{d \ln U_3(r_1)}{d \ln r_1} + \frac{\partial \ln U_3(r_1)}{\partial \ln r_1} \right], \]

(9)
in which $\bar{E}(r_1, r_2) = (1/2\pi^2) \int k^2 P(k)u(kr_1)u(kr_2) \, dk$. The special case when $r_1 = r_2 = r$ reads

$$R_3 = \frac{34}{21} \left[ 1 + \frac{\sigma^2}{U_2} \right] + \frac{\sigma^2}{3U_2} \left[ 2 \frac{d\ln U_2}{d\ln r} + \frac{d\ln \sigma^2}{d\ln r} \right],$$

(10)

where $\sigma^2 = (1/2\pi^2) \int k^2 P(k)u^2(k) \, dk$. The above equations constitute the main results of this paper. Note the similarity of $R_3$ with the skewness, which is calculated in weakly non-linear perturbation theory as $S_3 = 34/7 + d\ln \sigma^2/d\ln r$ (Juszkiewicz, Bouchet & Colombi 1993; Bernardeau 1994b).

4 MEASUREMENTS

Intuitively, conditional cumulants can be estimated from moments and/or factorial moments of neighbour counts, a fact that is easily seen from the generating function. However, in order to calculate $U_3$, for example, we are forced to measure $\sigma^2$ of a spherical top-hat window function (Peebles 1980). We have therefore not gained any efficiency or advantages over CIC in this way.

We realized that $U_n(r)$ can be measured in a similar way to $N$-point correlation functions. For instance, $U_2$ can be thought of as a two-point correlation function in a bin $[r_0, r_0] = [0, r]$. Taking the lower limit to be a very small number instead of 0, one can correct for discreteness effects arising from self-counting (this is equivalent to using factorial moments when neighbour counts are calculated directly). Given a set of data and random points, the class of estimators of Szapudi & Szalay (1998) will provide an edge-corrected technique to measure conditional cumulants:

$$\hat{U}_n = \frac{(D - R)^n}{R^n};$$

(11)

for example,

$$U_3(r_1, r_2) = (DDD - 3DDR + 3DRR)/RRR - 1,$$

in which the bin configuration is $[\epsilon, r_1], [\epsilon, r_2], [\epsilon, r_1 + r_2]$, where $\epsilon$ is a very small positive number. Obviously, existing $N$-point correlation function codes can be used for the estimation; for higher than third order, one also has to take connected moments in the usual way. Note that the above estimator is corrected both for ‘sharp’ edges (cut-out hole and survey geometry), and for gradual changes, such as incompleteness. In the latter case, randoms need to be generated according to the varying density, and the variance of the estimator can be improved by minimum-variance (for example $J_3$ etc.) weighting.

While the above suggests a scaling similar to $N$-point correlation functions, the relation to neighbour-count factorial moments outlined in the previous section can be used to realize the estimator using two-point correlation function codes. To develop such an estimator, neighbour-count factorial moments need to be collected for each possible combination in which data and random points play the role of centre and neighbour.

Note that the edge correction of equation (11) is expected to be less accurate for conditional cumulants than for $N$-point correlation functions; however, the estimator will be more accurate than CIC estimators. Several ways of correcting edge effects are known that are directly applicable to conditional cumulants (Ripley 1988; Kerscher 1999; Pan & Coles 2002). In what follows, we use equation (11) for all results presented. Future high-precision measurements could benefit from a detailed comparison of possible estimators, as in Kerscher (1999).

To test our theory, we performed measurements in the $\Lambda$CDM simulations of the Virgo Supercomputing Consortium. We used outputs of the Virgo simulation (Jenkins et al. 1998) and the Very Large Simulation (Macfarland et al. 1998, VLS). These two simulations have identical cosmology parameters: $\Omega_m = 0.3$, $\Omega_v = 0.7$, $\Gamma = 0.21$, $h = 0.7$, and $\sigma_8 = 0.9$; the Virgo simulation is in a box of $239.5 \, h^{-1} \text{Mpc}$ with $256^3$ particles, and the VLS is in a box of $479 \, h^{-1} \text{Mpc}$ with $512^3$ particles. In order to estimate measurement errors, we divide the VLS simulation into eight independent subsets, each with the same size and geometry as the original Virgo simulation. We used the resulting nine realizations to estimate errors. The periodic boundary condition is not used in the process of estimation. Note that we corrected for cosmic bias by always taking the average before ratio statistics were formed.

Our measurements of the second- and third-order conditional cumulants in full scale range are displayed in Fig. 1 to give an overall impression. Results from EPT (equation 10) are denoted with solid lines. To clarify the difference between conditional cumulants and cumulants, we plot $S_3$ both in simulation and perturbation theory along with $R_3$ in Fig. 2. It is very clear from the ratio of $R_3$ in simulations over theory (subpanel in Fig. 2) that measurements in simulations are in excellent agreement with EPT, especially on large

### Figure 1

- $U_2(r)$ (left) and $U_3(r)$ (right) in real space measured in $\Lambda$CDM $N$-body simulations (triangles with error bars) compared with predictions from perturbation theory (solid line).

© 2005 RAS, MNRAS 361, 357–361
scales $\gtrsim 20 \, h^{-1} \, \text{Mpc}$. It can also be seen that the error bar of $R_3$ is roughly 50 per cent smaller than that of $S_3$ at $r \sim 40 \, h^{-1} \, \text{Mpc}$.

5 SUMMARY AND DISCUSSION

We have presented the theory of conditional cumulants in a weakly non-linear regime. This unique set of statistics can be thought of as integrated $N$-point correlation functions, or as an integrated monopole. We have derived the generating function of neighbour-count factorial moments, revealing a deep connection to conditional cumulants. We introduced the reduced quantity $R_N$, which is analogous to the cumulant $S_N$. We calculated leading-order perturbation theory predictions, and showed that results are similar to those of the $S_N$s. While edge correction for CIC is not feasible, however, we have proposed an edge-corrected estimation method for the conditional cumulants. This was applied to a set of measurements in simulations, which yielded results in excellent agreement with the theory, especially on large scales $\gtrsim 20 \, h^{-1} \, \text{Mpc}$. The agreement with theory should encourage further development of this statistic for high-precision cosmological applications, such as constraining bias. Before we start measuring conditional cumulants, however, we need to further calibrate the estimator in order to achieve better edge correction with real survey geometry.

As three-dimensional galaxy catalogues are produced inherently in redshift space, it is crucial to understand the effects of redshift distortions on these statistics before practical applications can follow. In the distant-observer approximation, the formula by Kaiser (1987) and Lilje & Efstathiou (1989) is expected to provide an excellent approximation for $U_2(s)$. According to Section 2, we only need to consider the monopole enhancement

$$U_2(s) = \left(1 + \frac{2}{3} f + \frac{1}{5} f^2\right) U_2(r)$$

(12)

where $f \approx \Omega_0^{1/3}$. This formula essentially predicts a uniform shift of the real-space results. To test it, we repeated our measurements in redshift space, and found that the above is indeed an excellent approximation in redshift space (Fig. 3).

Considering the relatively simple, monopole nature of the statistics, we expect that the overall effect on $U_3$ should also be a simple shift, similar to the Lagrangian calculations of Hivon et al. (1995) and the more general Eulerian results of Scoccimarro, Couchman & Frieman (1999). Specifically, we propose that the ratio of $R_3$ in redshift space to that in real space can be approximated by

$$\frac{5(2520 + 3360 f + 1260 f^2 + 9 f^3 - 14 f^4)}{98(15 + 10 f + 3 f^2)^2}.$$  

(13)

This is motivated by the notion that the shift from redshift distortions of equilateral triangles should be similar to the corresponding shift for our monopole statistic. Our simulation results (see Fig. 3) show that this simple idea is indeed a surprisingly good approximation, although the phenomenological theory based on the above formula appears to have $\pm 5$ per cent bias on scales $\gtrsim 20 \, h^{-1} \, \text{Mpc}$, where we expect that weakly non-linear perturbation theory is a good approximation. For practical applications, this bias can be calibrated by $N$-body, or second-order Lagrangian perturbation theory (2LPT) (Scoccimarro 2000) simulations.

In addition to the above simple formula, we have calculated the shift due to redshift distortions by angular averaging the bispectrum monopole term in Scoccimarro et al. (1999). We found that the

![Figure 2](https://academic.oup.com/mnras/article-abstract/361/1/357/1023789)

**Figure 2.** $R_3$ in real space. $S_3$ in simulations and in perturbation theory are also given as reference; ratios of $R_3$ in simulations against perturbation theory are plotted in the subpanel as an indication of their agreement on large scales.

![Figure 3](https://academic.oup.com/mnras/article-abstract/361/1/357/1023789)

**Figure 3.** $U_2$ and $R_3$. The solid line in the left panel comes from equation (12). In the right panel the solid line shows a phenomenological model based on equation (13). The theory appears to be a reasonable approximation at the 5 per cent level, as shown by the ratio of $R_3$ in simulations to that in theory in the subpanel of the right panel.
results overpredict redshift distortions; however, they would agree with simulations at the 1–2 per cent level if we halved the terms classified as FOG (Finger Of God). At the moment there is no justification for such a fudge factor, and therefore we opt to use the above phenomenology, which is about 5 per cent accurate. While redshift distortions of third-order statistics are still not fully understood because of the non-perturbative nature of the redshift-space mapping (R. Scoccimarro, private communication), detailed calculations taking into account velocity dispersion effects will improve the accuracy of the redshift-space theory \( U_3 \).

For applications to constrain bias, it is necessary to keep in mind that redshift distortions and non-linear bias do not commute. At the level of the above simple theory, however, it is clear that one can understand the important effects at least for the third-order statistic. There are several ways to apply conditional cumulants for bias determination, either in combination with one other statistic (CIC or cumulant correlators, cf. Szapudi 1998b), or using the configuration dependence of the more general \( R_3(r_1, r_2) \). Care must be taken because in practical applications ratio statistics will contain cosmic bias (Hui & Gaztaña 1999; Szapudi et al. 1999). We propose that joint estimation with \( U_2 \) and \( U_3 \) will be more fruitful, even if \( R_1 \) is better for plotting purposes. Details of the techniques to constrain bias from these statistics, as well as determination of the bias from wide-field redshift surveys, are left for future work.

An alternative way to get around redshift distortions is to adapt conditional cumulants for projected and angular quantities. Such calculations are straightforward, and entirely analogous to those performed for \( S_3 \) in the past. Another possible generalization of our theory would be to use halo models (Cooray & Sheth 2002) to extend the range of applicability of the theory to well below 20 h\(^{-1}\) Mpc. These generalizations are left for future research.

ACKNOWLEDGMENTS

This work was supported by NASA through grants AISR NAG5-11996, ATP NAG5-12101, and by NSF through grants AST02-06243 and ITR 1120201-128440. We thank Pablo Fosalba for stimulating discussions. The simulations in this paper were carried out by the Virgo Supercomputing Consortium using computers based at the Computing Centre of the Max-Planck Society in Garching and at the Edinburgh Parallel Computing Centre.

REFERENCES

Borgani S., 1995, Phys. Rep., 251, 1

This paper has been typeset from a \TeX/\LaTeX file prepared by the author.