NISS

Measures of Dependence Structure for Random Sets

Jeffrey D. Picka

Technical Report Number 88
August, 1998

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Measures Of Dependence Structure For Random Sets

Jeffrey D Picka

National Institute of Statistical Sciences

Abstract

A set of measures of dependence are proposed for the analysis of the structure of random sets. These measures are similar to cumulants in some aspects, but are far more suited to the structure of random sets than cumulants. Differences were found between cumulants and the new functions in terms of probabilistic, statistical and geometrical criteria. In each case, the new measures were found to be superior to ordinary cumulants in the context of this particular type of stochastic process.

Keywords: cumulants, random sets, germ-grain model, Boolean Model, Dead Leaves Model

AMS Subject Classifications: 60D05

1 Introduction

In the study of random processes, cumulants take on a central role in describing dependence in multivariate distributions. A random set process is not a standard example of such a distribution, and so the ordinary cumulants may not be the measures of the dependence structure for a process consisting of dependent 0-1 random variables located at every point in $\mathbb{R}^d$. Instead, an alternative measure of dependence is proposed which yields more interpretable information about the higher order dependence structure of these and other binary random structures. In the study of germ-grain models with non-intersecting grains, these will provide information about the global structure of a realization, in contrast to the local information that would be provided by tessellation-based statistics.

Initially, a series of definitions for the first four moments of a random set process will be given. Two different standardizations of the moments will follow, as well as examples of their form for a Boolean model and for a Dead Leaves model. A discussion of the relative merits of each standardization will be undertaken, which will also involve the meaning of standardization in the random set context.

2 Notation and Basic Definitions

In the most general case, a realization of a random set Φ in $\mathbb{R}^d$ can be represented by the indicator functions defined at all $x \in \mathbb{R}^d$ by

$$I_\Phi(x) = \begin{cases} 1 & \text{if} \quad x \in \Phi \\ 0 & \text{if} \quad x \notin \Phi. \end{cases}$$

Matheron [1] demonstrated the existence of a probability measure that allows such processes to be considered as stochastic processes in the conventional sense, but the laws of such processes have no convenient and succinct analytical expressions. Instead, random sets in common use are defined in terms of the algorithm that produces them and so are generally referred to as *models.* Classic examples of such models are the Boolean Model [2] and the Dead Leaves Model [3], both of which belong to the more general category of germ-grain models [4]. Much of this paper is concerned with these models, which use the points in the realization of a point process (know as *germs*) as locations at which geometric objects of possibly random shape, size and orientation (known as *grains*) can be attached.

While the probability distributions associated with random sets are complex, the moments of these distributions can be easily expressed and estimated, particularly if the process is stationary or isotropic.

The first four moments of a general random set process are defined for any $x, r, s, t \in \mathbb{R}^d$ as

$$M_1(x) = \Pr[x \in \Phi]$$

$$M_2(x, r) = \Pr[x, r \in \Phi]$$

$$M_3(x, r, s) = \Pr[x, r, s \in \Phi]$$

$$M_4(x, r, s, t) = \Pr[x, r, s, t \in \Phi].$$

If the process is stationary, the *reduced moments* of the process are defined as

$$m_1 = \Pr[0 \in \Phi]$$

$$m_2(r) = \Pr[0, r \in \Phi]$$

$$m_3(r, s) = \Pr[0, r, s \in \Phi]$$

$$m_4(r, s, t) = \Pr[0, r, s, t \in \Phi].$$

Given a window A onto the process, these reduced moments may be estimated for stationary processes by estimators such as

$$\widehat{m}_1 = \frac{\nu_d(\Phi \cap A)}{\nu_d(A)}$$

where ν_d is Lebesgue measure on $\mathbb{R}^d$. While it must be noted that discretized versions of such estimators are used in practice, this form is more useful for analytical and expository purposes.

From this point, the random set process will be assumed to be stationary and isotropic, while all definitions will be stated in terms of reduced moments for greater clarity. It should be noted that this is not the conventional notation for random set moments [2, 4], but that system of notation is incapable of coping with moments higher than the second. Lower case letters will be used to refer to position vector quantities, while upper case letters will be used to refer to lengths of vectors whose orientation is irrelevant.

While the complement of a random set Φ is not a random set because it is not closed, in the context of certain models this is not practically relevant. If one were to restrict attention entirely to germ-grain models in which the grains could not interpenetrate and had simple, non-pathological boundaries, then the difference between the complement Φ^c of such a random set and its closure $cl(\Phi^c)$ would be geometrically trivial. In this case, if the Φ were stationary, then a set of reduced moments $m_k^c(r_1, \ldots, r_{k-1})$ could be defined for $cl(\Phi^c)$ and they would be related to the moments of the original random set as follows:

$$
\begin{aligned}
m_1^c &= 1 - m_1 \\
m_2^c(r) &= 1 - 2m_1 + m_2(r) \\
m_3^c(r,s) &= 1 - 3m_1 + m_2(r) + m_2(s) + m_2(r-s) - m_3(r,s) \\
m_4^c(r,s,t) &= 1 - 4m_1 + m_2(r) + m_2(s) + m_2(t) + m_2(r-s) + m_2(r-t) + m_2(s-t) \\
&\quad - m_3(r,s) - m_3(s,t) - m_3(r,t) - m_3(r-t, s-t) + m_4(r,s,t)
\end{aligned}
$$

These relationships are also obeyed by the estimators of these moments.

3 Standardized Moment Measures For Random Sets

Standardized moment measures are of use in describing the dependence structure of the set. For a stationary random set, this structure is what produces dependence between the events $0 \in \Phi$, $x_1 \in \Phi, \ldots, x_{k-1} \in \Phi$ for arbitrary configurations $\{0, x_1, \ldots, x_{k-1}\}$. For random sets not possessed of any ordered periodic structure, it is expected that this dependence will vanish as the points in a configuration grow further apart, causing any moment defined on k mutually distant points to revert to a value of m_1^k. By standardizing this value to 1 or 0, configurations for which the dependence structure is not present may be identified.

In the case of a stationary random set process Φ on $\mathbb{R}^d$, ordinary cumulants are defined for configurations of points defined by the vectors $0, r, s, t \in \mathbb{R}^d$.

$$
\begin{aligned}
\kappa_1 &= m_1 \\
\kappa_2(r) &= m_2(r) - m_1^2 \\
\kappa_3(r,s) &= m_3(r,s) - m_1\left(m_2(r) + m_2(s) + m_2(r-s)\right) + 2m_1^3 \\
\kappa_4(r,s,t) &= m_4(r,s,t) \\
&\quad - m_1\left(m_3(r,s) + m_3(s,t) + m_3(t,r) + m_3(s-r, t-r)\right) \\
&\quad - \left(m_2(r)m_2(s-t) + m_2(s)m_2(r-t) + m_2(t)m_2(r-s)\right) \\
&\quad + 2m_1^2\left(m_2(r) + m_2(s) + m_2(t) + m_2(r-s) + m_2(s-t) + m_2(t-r)\right) \\
&\quad - 6m_1^4.
\end{aligned}
$$

Cumulants are of interest as measures of the dependence structure since they might possibly be able to identify configurations of points for which the random set exhibits *inherent k^{th} order dependence.* For

$k = 2$ and the configuration $\{0, x\}$, inherent 2^{nd}-order dependence would occur if the event that $0 \in \Phi$ was dependent on whether or not $x \in \Phi$. For a configuration of points $\{0, x_1, \dots, x_{k-1}\}$, this type of dependence occurs if the event that any one of those points is in Φ is dependent on whether or not *all* of the other $(k-1)$ points in the configuration are in the set. Given the complexity of the geometry of germ grain models, it may be difficult to determine if such dependence is possible for a given event. However, for the configuration of points {0,r,s }, if 0 and r are close but s is far from both of them, then the event $s \in \Phi$ should be independent of whether or not $0 \in \Phi$ or $r \in \Phi$, and thus

$$\kappa_3(r, s) \approx m_1 m_2(s) - m_1(m_2(s) - 2m_1^2) + 2m_1^3 = 0$$

For configurations of three or more points, the complexity of the relationship between the geometry of the model and the question of independence between events does not make it possible to determine if a condition such as

$$\kappa_3(r, s) \neq 0$$

is necessary when inherent 3^{rd}-order dependence exists. Instead, configurations of p points which satisfy the condition $\kappa_p \neq 0$ will be said to display *additive k^{th}-order dependence.* As in the distinction between lack of correlation and independence, the cumulant being 0 does not necessary imply the absence of k^{th} order dependence. This can be seen in Figure 4 below, where the third cumulant is 0 at distances where inherent third-order dependence must exist.

An alternative potential measure of inherent k^{th}-order dependence, which will be called the *k^{th} spatial cumulant*, is defined for a stationary random set process as follows:

$$\begin{aligned}
\Sigma_1 &= m_1 \\
\Sigma_2(t) &= \frac{m_2(t)}{m_1^2} \\
\Sigma_3(s,t) &= \frac{m_3(s,t)/m_1^3}{\Sigma_2(s)\Sigma_2(t)\Sigma_2(s-t)} \\
&= \frac{m_3(s,t)m_1^3}{m_2(s)m_2(t)m_2(s-t)} \\
\Sigma_4(r,s,t) &= \frac{(m_4(r,s,t)/m_1^4)}{\Sigma_2(r)\Sigma_2(s)\Sigma_2(t)\Sigma_2(r-s)\Sigma_2(s-t)\Sigma_2(r-t)\Sigma_3(r,s)\Sigma_3(s,t)\Sigma_3(r,t)\Sigma_3(r-s,t-s)} \\
&= \frac{m_4(r,s,t)m_2(r)m_2(s)m_2(t)m_2(r-s)m_2(s-t)m_2(t-r)}{m_3(r,s)m_3(s,t)m_3(r,t)m_3(r-s,t-s)m_1^4}.
\end{aligned}$$

These definitions result in a multiplicative decomposition of the moments instead of an additive one. That is,

$$
\begin{aligned}
m_1 &= \Sigma_1 \\
m_2(t) &= m_1^2 \Sigma_2(t) \\
m_3(s,t) &= m_1^3 \Sigma_2(s)\Sigma_2(t)\Sigma_2(s-t)\Sigma(s,t) \\
m_4(r,s,t) &= m_1^4 \Sigma_2(r)\Sigma_2(s)\Sigma_2(t)\Sigma_2(r-s)\Sigma_2(s-t)\Sigma_2(t-r) \\
&\quad \times \Sigma_3(r,s)\Sigma_3(s,t)\Sigma_3(r,t)\Sigma_3(r-s,t-s)\Sigma_4(r,s,t).
\end{aligned}
$$

For spatial cumulants, these too may be indices of inherent k^{th}-order dependence. In the case of a configuration $\{0, r, s\}$ where 0 and r are close but s is far away, $\Sigma_3(r,s) \approx 1$. This suggests that configurations of k points for which $\Sigma_k \neq 1$ exhibit *multiplicative inherent k^{th}-order dependence*, which, as shown below, is distinct from the additive version. The logarithms of the spatial cumulants can be used to decompose $\log(m_k)$ additively, but this decomposition is distinct from that of the ordinary cumulants.

The pattern of the decomposition is a simple one as well, at least as far as the first four moments are concerned. If Γ_j is the set of all subsets of j points from among the k points involved in a configuration, then the decomposition has the general form

$$
m_k(x_1, \ldots, x_{k-1}) = m_1^k \prod_{j=2}^{k} \prod_{\gamma \in \Gamma_j} \Sigma_j(\gamma).
$$

The inversion of this leads to

$$
\Sigma_k(x_1, \ldots, x_{k-1}) = \prod_{j=1}^{k} \prod_{\gamma \in \Gamma_j} m_j(\gamma)^{(-1)^{k-j}}
$$

In all of the above cases, the decomposition isolates the effects of each combination of inter-point dependencies in a single term.

4 Examples

Expressions for the moments of random set processes may only be found for a very small number of models, most of which are Boolean models with simple grains of non-random shape and size. For many models that are simple to simulate, there are no analytical expressions for the moments, and so simulations must be used to estimate the moment functions. Two processes are considered here:

(1) A Boolean model in R^2 with grains that are discs of diameter $\mu = 1$ located at germs generated by a Poisson process of intensity λ. The grains in this model are free to interpenetrate.

(2) A stationary Dead Leaves model whose grains are discs of diameter $\mu = 1$. In this case, 150 points were drawn from a uniform distribution on a square window slightly larger than 10 units by 10 units, augmented to avoid edge effects. Each new point is compared to all the points that came before it, and any of those points that come within μ of the new point are removed from the list. The final list of points becomes a list of the germs in a realization of the model. The grains in this model will not interpenetrate.

For a Boolean model, the spatial cumulants of its complement have a very simple form. In the most general case, if a random set is produced from germs that arise from a Poisson process with intensity λ and from grains that have random shape and size, then the spatial cumulants of the complementary process are directly related to the process which produces the grains. In the case of the first spatial cumulant, if the process producing the grains is designated Z, then

$$\log(\Sigma_1^c) = \log(m_1^c) = -\lambda E[|Z|] \tag{1}$$

where $E[|Z|]$ is the average grain volume. For the higher dimensional spatial cumulants, the relationships are:

$$\log(\Sigma_2^c(r)) = \lambda E[|Z \cap (Z \oplus r)|] \tag{2}$$

$$\log(\Sigma_3^c(r,s)) = -\lambda E[|Z \cap (Z \oplus r) \cap (Z \oplus s)|] \tag{3}$$

$$\log(\Sigma_4^c(r,s,t)) = \lambda E[|Z \cap (Z \oplus r) \cap (Z \oplus s) \cap (Z \oplus t)|]. \tag{4}$$

If the grains are convex, then $E[|Z \cap (Z \oplus r)|]$ corresponds to the average volume of the subset of a random grain formed by intersecting that grain with a subset of itself translated by r, and similarly for the higher cumulants. If the grains are discs of fixed diameter, then

$$\log(\Sigma_1^c) = -\lambda\left(\frac{\pi\mu^2}{4}\right)$$

while

$$\log(\Sigma_2^c(r)) = \begin{cases} \frac{\lambda}{4}\left(\mu^2 \arccos\left(\frac{|r|}{\mu}\right) - \sqrt{|r|^2\left(\mu^2 - |r|^2\right)}\right) & \text{iff} \quad |r| \le \mu \\ 0 & \text{otherwise.} \end{cases}$$

The expression for a general third spatial cumulant would be analytically complicated and difficult to represent graphically. If attention is restricted to the configurations $\triangle R$ whose three points form the vertices of an equilateral triangle with sides of length R, then the third spatial cumulant simplifies greatly:

$$\log(\Sigma_3^c(\triangle R)) = \begin{cases} -\frac{\lambda}{4}\left(\sqrt{3}s^2 + 3\left(\mu^2 \arccos\left(\frac{s}{\mu}\right) - \sqrt{s^2\left(\mu^2 - s^2\right)}\right)\right) & \text{iff} \quad R \leq \frac{\sqrt{3}}{2}\mu \\ 0 & \text{otherwise.} \end{cases}$$

where

$$s = \frac{1}{2}\left(\sqrt{3(\mu^2 - R^2)} - R\right) \quad R \in (0, \frac{\sqrt{3}}{2}\mu).$$

Plots of $\Sigma_3(\triangle R)$, $\Sigma_2(R)$, $\kappa_3(\triangle R)$, and $m_3(\triangle R)$ are given for models with grains of size $\mu = 1$ and first moments $m_1 = 0.1$ and $m_1 = 0.75$ in Figures 1 and 2.

No analytical expressions exist for the moments or spatial cumulants of the Dead Leaves model, or most models whose grains are not allowed to interpenetrate. One hundred realizations of the Dead Leaves model were generated, and the moments were estimated using a Monte Carlo method that examined points in randomly chosen triangular configurations for membership in the random set. The estimates of $\Sigma_3(\triangle R)$, $\Sigma_2(R)$, $\kappa_3(\triangle R)$, and $m_3(\triangle R)$ are plotted in Figure 3, along with the equivalent functions for a Boolean model of the same mean volume fraction.

5 Discussion

The attributes of spatial and ordinary cumulants can be broken down into three basic classes. First, there are a number of general probabilistic observations that can be made outside of any reference to a particular model. If examples of particular models are considered, certain geometrical aspects of the cumulants can be examined that are not obvious from functional considerations alone. By comparing methods by which the various cumulants can be estimated, a statistical comparison may also be made.

5.1 Probabilistic Comparisons

From the expressions (1)-(4), the spatial cumulants of the complement of a Boolean model are completely determined by the process that generates the grains of that model and the intensity of the Poisson process that generates the germs. There is no such direct interpretation of the ordinary cumulant in terms of the grain and germ processes, but this is not necessarily a strong argument in favour of the spatial cumulant. While much data from nature could be modeled by a germ-grain model, in almost every case that model would not have germs that arose from a Poisson process. Instead, these data would be often better represented by germ-grain models in which the grains do not interpenetrate, and for these models the relationship between the grain process and the spatial cumulants is unclear. While ordinary cumulants are invariant under linear transformations of Φ, it is not clear that such transformations are meaningful in the context of random sets, particularly when there is only a large single observation of a stationary process.

For real-valued multivariate random variables, their joint distributions are often determined by the characteristic function of that distribution, from which the ordinary cumulants can be derived. Random sets possess characteristic functionals rather than characteristic functions, which are far more complicated structures whose nature will not be determinable from cumulant functions. Characterization of random closed sets may be achieved through knowledge of the hitting function [1] , defined as

$$T(K) = Pr[K \cap \Phi^c]$$

where K is any closed subset of $\mathbb{R}^d$. Both ordinary and spatial cumulants can be expressed as functions of hitting functions, when K is taken to be the sets of points in a configuration. While point sets alone cannot completely determine an arbitrary random set processes from its hitting function, for certain Boolean models Nagel [5] and Rataj [6] have shown that m_1, m_2, and m_3 can completely determine the hitting function.

The spatial cumulants may be used to decompose conditional probabilities in a very similar manner to that used in the decomposition of the moments of the process. For example,

$$\Pr[0, t \in \Phi | s \in \Phi] = m_1^2 \Sigma_2(s)\Sigma_2(t)\Sigma_2(s-t)\Sigma_3(s,t).$$

$$\Pr[t \in \Phi | 0, s \in \Phi] = m_1 \Sigma_2(t)\Sigma_2(s-t)\Sigma_3(s,t).$$

For ordinary cumulants, the additive nature of the definition of the cumulant lead to more complicated decompositions that yield little insight into the dependence structure of the set. For example,

$$\Pr[t \in \Phi | 0, s \in \Phi] = \frac{\kappa_3(r,s) + \kappa_1\left(\kappa_2(r) + \kappa_2(s) + \kappa_2(r-s)\right) + \kappa_1^3}{\kappa_2(s) + \kappa_1^2}.$$

The decomposition of conditional probabilities involving both Φ and Φ^c does not produce a multiplicative decomposition, regardless of the standardized moments used.

If the complement of a random set can also be thought of as a random set, then there is a further major difference between the spatial and ordinary cumulants. The relationship between each of the first four ordinary cumulants of a random set and its inverse is as follows:

$$\kappa_1^c = 1 - \kappa_1$$

$$\kappa_2^c(r) = \kappa_2(r)$$

$$\kappa_3^c(r,s) = -\kappa_3(r,s)$$

$$\kappa_4^c(r,s,t) = \kappa_4(r,s,t).$$

The sign of the third cumulant at $r = s = 0$ is determined by κ_1, in that if $\kappa_1 = m_1 < 0.5$, then $\kappa_3(0,0) > 0$, with a sign reversal for $m_1 > 0.5$. Since the third and fourth moments of the random set process are affected by shape, this suggests that the shape information is lost when this form of standardization is used. The relationships between the spatial cumulants of a random set and their counterparts for its complement are considerably more complex:

$$\Sigma_2^c(r) = 1 + \lambda^2 \left(\Sigma_2(r) - 1\right)$$
$$\Sigma_3^c(r,s) = \frac{1 + \frac{\lambda^2}{1-m_1}\left(\Sigma_2(r) + \Sigma_2(s) + \Sigma_2(r-s)\right) - \lambda^3 \left(\Sigma_2(r)\Sigma_2(s)\Sigma_2(r-s)\Sigma_3(r,s) - 1\right)}{\left(1 + \lambda^2\left(\Sigma_2(r) - 1\right)\right)\left(1 + \lambda^2\left(\Sigma_2(s) - 1\right)\right)\left(1 + \lambda^2\left(\Sigma_2(r-s) - 1\right)\right)}$$

where $\lambda = m_1/(1 - m_1)$ is the odds ratio. The relationship for the fourth cumulant is omitted, as it is more complicated than that of the third. These relationships illustrate that the influence of the first moment is not fully removed from the standardized higher order spatial cumulants, particularly in the region where multiplicative inherent k^{th}-order dependence occurs.

From the relationships between the spatial cumulants of the random set process and its complement, this standardization process does not remove the dependence of the higher order spatial cumulants on the size of the first moment. In the case of the Boolean model, the appearance of a realization will vary drastically with increasing m_1, changing from almost isolated discs to a large region with discrete and irregularly shaped holes. The standardization affects the functions primarily at distances where the dependence structure breaks down entirely, but the remainder of the standardized moments are not fully cleared of the influence of lower order moments. In the case of Boolean models, an alternative set of dependence measures can be constructed. If Z is the process that generates the germs, then (1)-(4) can be re-arranged to yield

$$\Gamma_2(r) = -\frac{\log\left(\Sigma_2^c(r)\right)}{\log\left(\Sigma_1^c\right)} = \frac{E[|Z \cap (Z \oplus r)|]}{E[|Z|]}$$
$$\Gamma_3(r,s) = \frac{\log\left(\Sigma_3^c(r,s)\right)}{\log\left(\Sigma_1^c\right)} = \frac{[|Z \cap (Z \oplus r) \cap (Z \oplus s)|]}{E[|Z|]}$$
$$\Gamma_4(r,s,t) = -\frac{\log\left(\Sigma_4^c(r,s,t)\right)}{\log\left(\Sigma_1^c\right)} = \frac{E[|Z \cap (Z \oplus r) \cap (Z \oplus s) \cap (Z \oplus t)|]}{E[|Z|]}$$

which do not depend at all on the intensity of the underlying Poisson germ process. It is not clear that these functions have any meaning for general germ-grain models, but Γ_k is identically 0 for any configuration on which multiplicative inherent k^{th}-order dependence for Φ^c is absent.

5.2 Geometrical Comparisons

The ordinary and spatial cumulants also differ in the manner in which they describe the geometry of the random set realization. Both types of cumulants and the moments describe average properties of any realization rather than the local properties that are inherited from the generating algorithms, but the connections between the algorithms and these average properties is not obvious. As a result, it is not possible to say if configurations that exhibit inherent k^{th}-order dependence also exhibit the additive or multiplicative inherent dependence. Nonetheless, both ordinary and spatial cumulants describe the average properties of a random set in different ways, and careful observation of their properties can allow some conjectures to be made concerning which of these functions is more informative.

Figures 1 and 2 illustrate a selection of second and third moment-based dependence measures for the Boolean model, when $m_1 = 0.1$ and $m_1 = 0.75$, respectively. The second moment declines monotonically to m_1^2, showing no sign of the slight oscillation associated with models in which the germs must be at least a fixed distance apart. Since $\Sigma_2(R)$ is a rescaling of this function, there is no real difference between the shapes of these functions.

For the third moment-based functions, striking differences appear. In both cases shown, and for all cases examined, $m_3(\triangle R)$ declines monotonically to m_1^3 as $|r|$ approaches μ, and has roughly the same shape from $m_1 = 0.00001$ to $m_1 = 0.9$, after which it becomes markedly more concave. In contrast, the spatial cumulants exhibit a number of interesting properties which may be related to the geometry of the model. For $m_1 = 0.00001$ and $m_1 = 0.0001$, $\Sigma_3(\triangle R)$ looks like a step function with the step at $R = \mu$. In such a model, the probability that two grains overlap or are anywhere close to each other is very small. Consequently, it is almost certain that an event associated with inherent 3^{rd}-order dependence will be one in which all three points in $\triangle R$ fall into a single grain. As m_1 increases, $\triangle R$ can intersect with two grains which are close, so that one vertex of $\triangle R$ lies outside one of the grains and another vertex lies outside the other disc. This arrangement would be possible when $R \in [0, \mu]$, while the event with all three vertices appearing in a single grain is only possible when $R \in [0, (\sqrt{3}/2)\mu]$. The perturbation in $\Sigma_3(\triangle R)$ at $R = (\sqrt{3}/2)\mu$ that grows more extreme as m_1 becomes smaller may be explained by this. It is also interesting to note that the second derivative of $\Sigma_3(\triangle R)$ is generally positive for $R < \mu/2$ when $m_1 < 0.5$, generally negative when $m_1 > 0.5$, and almost constant when $m_1 = 0.5$. This dependence of the second derivative on m_1 is reversed when $\Sigma_3^c(\triangle R)$ is considered, although for large values of m_1, $\Sigma_3^c(\triangle R)$ does not look like a step function. At values of m_1 as high as 0.99999, it slopes gradually upward from $R = \mu/2$ to $R = \mu$. This difference may be related to the fact that isolated grains in the Boolean model have a constant size and shape, while the areas left uncovered by a model with large m_1 are irregular in size and shape. Finally, multiplicative inherent 3^{rd}-order dependence for the Boolean model disappears when $R > \mu$, but disappears when $R > (\sqrt{3}/2)\mu$ on its complement. Since the complement is not a grain-model, it is not obvious why these should be different, but the fact that they are illustrates a fundamental difference between additive and multiplicative inherent 3^{rd}-order dependence.

The third cumulant function for the Boolean model is not all that different in shape from $m_3(\triangle R)$ for $m_1 < 0.5$, but as m_1 approaches 0.5, the function develops a slight wobble that eventually becomes something like a cycle of a sine curve (Figure 4). As m_1 continues to increase, the cumulant takes on the shape of $m_3(\triangle R)$ reflected in the line $m_3(\triangle R) = m_1^3$. In addition, $\kappa_3^c(\triangle R) = -\kappa_3(\triangle R)$, and both functions become identically 0 only when $R > \mu$, even though it appears to converge to 0 at $R = (\sqrt{3}/2)\mu$ when m_1 is small.

Examination of the moment and cumulant functions for the Dead Leaves Model and a Boolean model having the same m_1 value as shown in Figure 3 demonstrates further differences between the descriptive power of spatial and ordinary cumulants. The third moment function for the Dead Leaves model is not dissimilar in shape to the second, dropping slightly below m_1^3 in the vicinity of $R = \mu$. The third spatial cumulants for the Dead Leaves model and its complement are radically different in shape, even though the curves for the Boolean model are quite similar. For the Dead Leaves model, it is possible to catalog all of the different types of event in which inherent 3^{rd}-order depenence might occur:

G_1: all three vertices of $\triangle R$ lie in the same grain which can occur for $R \in [0, (\sqrt{3}/2)\mu]$,

G_2: two vertices of $\triangle R$ lie in the one grain, while the third vertex lies in a second grain which can occur for $R \in [0, \mu]$, and

G_3: all three vertices of $\triangle R$ lie in different grains, which can occur for $R > \mu(1 - \sqrt{3}/2)$.

The local minima and maxima of $\Sigma_3(\triangle R)$ for the Dead Leaves model are close to the locations where events of type G_1 and type G_2 can no longer occur, which suggests that $\Sigma_3(\triangle R)$ is strongly related to inherent 3^{rd}-order dependence. Unfortunately, no such classification has yet been found for events in which inherent 3^{rd} order dependence is present in the complement of the dead leaves model, but the smoother shape of $\Sigma_3^c(\triangle R)$ may be related to the greater randomness in shape of the components of the complement.

The shape of $\kappa_3(\triangle R)$ for the Dead Leaves model resembles $\Sigma_3(\triangle R)$ reflected in a horizontal line. This might suggest that $\kappa_3(\triangle R)$ is somehow averaging what is seen in $\Sigma_3(\triangle R)$ and in $\Sigma_3^c(\triangle R)$, since the sharpness of the shape of $\Sigma_3(\triangle R)$ might be related to the fixed shape and size of the grains.

To summarize, there appears to be a significant difference between the aspects of a model described by the spatial and ordinary cumulants. The tentative connections between the behaviour of the spatial cumulants and aspects of inherent 3^{rd}-order dependence together with the manner in which the spatial cumulants appear to capture more details of this structure suggests that $\Sigma_3(\triangle R)$ is a better descriptor of structure than $\kappa_3(\triangle R)$. As for the ordindary cumulants, it is not clear at all what $\kappa_3(\triangle R)$ actually represents in terms of the dependence structure of the random set. When $m_1 = 0.5$, the spatial cumulants may reflect the shift in dominance from a germ-grain model to its complement through a gradual change in convexity of the function, while the spatial cumulant marks this transition by radically changing its shape to that seen in Figure 4 for m_1 within a small neighbourhood of 0.5. This lack of interpretability further suggests that ordinary cumulants may not be appropriate for the analysis of random sets.

5.3 Statistical Comparisons

It can be shown that the asymptotic variance of estimators of both types of cumulants can be greatly reduced if the estimators are balanced in their use of sample information [7]. While the most obvious estimates of the second spatial and ordinary cumulants are

$$\widehat{\Sigma}_2(r) = \frac{\widehat{m}_2(r)}{\widehat{m}_1^2} \qquad \text{and} \quad \widehat{\kappa}_2(r) = \widehat{m}_2(r) - \widehat{m}_1^2$$

where

$$\widehat{m}_2(r) = \frac{\nu_d\,(\Phi \cap \Phi_r \cap A \cap A_r)}{\nu_d\,(A \cap A_r)} \qquad \widehat{m}_1 = \frac{\nu_d\,(\Phi \cap A)}{\nu_d\,(A)}$$

and $A_r = \{x : x - r \in A\}$, these estimates overuse information from parts of the window A that are not used to estimate the second moment. By restricting attention to estimators of the first moment that use only information from those portions of the window that were used to estimate the second moment, significant reductions in estimator variance can be obtained. The intrinsically balanced second spatial and ordinary cumulant estimators are

$$\widehat{\Sigma}_2^{\circledast}(r) = \frac{\widehat{m}_2(r)}{\widehat{m}_1[\check{0}, r]\widehat{m}_1[0, \check{r}]} \qquad \widehat{\kappa}_2^{\circledast}(r) = \widehat{m}_2(r) - \widehat{m}_1[\check{0}, r]\widehat{m}_1[0, \check{r}]$$

where

$$\widehat{m}_1[\check{0}, r] = \frac{\nu_d\,(\Phi \cap A \cap A_r)}{\nu_d\,(A \cap A_r)} \qquad \widehat{m}_1[0, \check{r}] = \frac{\nu_d\,(\Phi_r \cap A \cap A_r)}{\nu_d\,(A \cap A_r)}$$

Similar modifications have been found for estimators of $\Sigma_3(r, s)$ and $\kappa_3(r, s)$, but the modified estimate of the spatial cumulant is much simpler than that of the ordinary cumulant.

6 Summary

From the discussion given above, the spatial cumulant appears to be a more useful tool than the ordinary cumulant in studying the dependence structure of a random set. It is potentially interpretable at distances where the dependence structure still exists, it can be used to construct conditional probabilities, and it reflects the differences in structure between the original random set process and its complement. The ordinary cumulant fails to achieve any of these things, and seems much better suited to the analysis of real-valued multivariate data, where the spatial cumulant would be meaningless. While a connection between the law of the random set process more fundamental than the connection through its moments is lacking, the existence and form of such a connection would yield useful general insights into the behaviour of germ-grain models, and perhaps more general random sets.

7 Acknowledgements

The problem that stimulated this research was proposed by Michael Stein, who also supervised the Ph.D. dissertation that emerged from it. Funding for this research was provided by the University of Chicago Louis Block Fund and NSF Grants DMS 89-05292 and DMS 92-08758.

References

[1] G. Matheron. *Random Sets and Integral Geometry.* Wiley, New York, 1975.

[2] N.A.C. Cressie. *Statistics For Spatial Data.* Wiley, New York, 1993.

[3] J. Serra. *Mathematical Morphology and Image Analysis.* Academic Press, London, 1982.

[4] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications.* John Wiley & Sons, New York, 2nd edition, 1995.

[5] W. Nagel. Orientation-dependent chord length distributions characterize convex polygons. *Journal of Applied Probability*, 30:730–736, 1993.

[6] J. Rataj. Characterization of compact sets by their dilation volume. *Math. Nachr.*, 173:287–295, 1995.

[7] J. D. Picka. Variance reducing modifications for estimated standardized moments of random sets. *Advances in Applied Probability*, 1999? submitted.

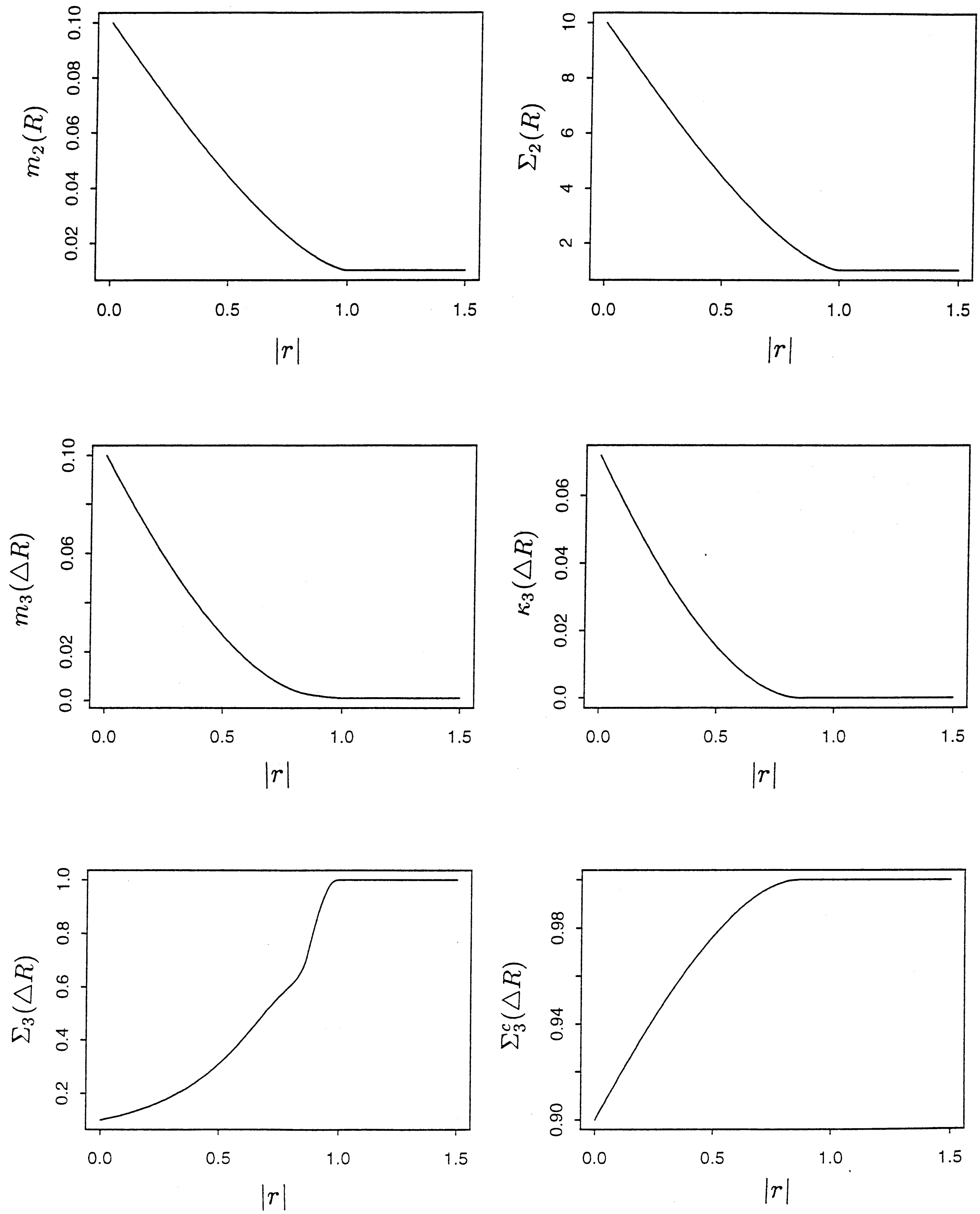

Figure 1: Moment and cumulant functions for the Boolean Model with $m_1 = 0.1$.

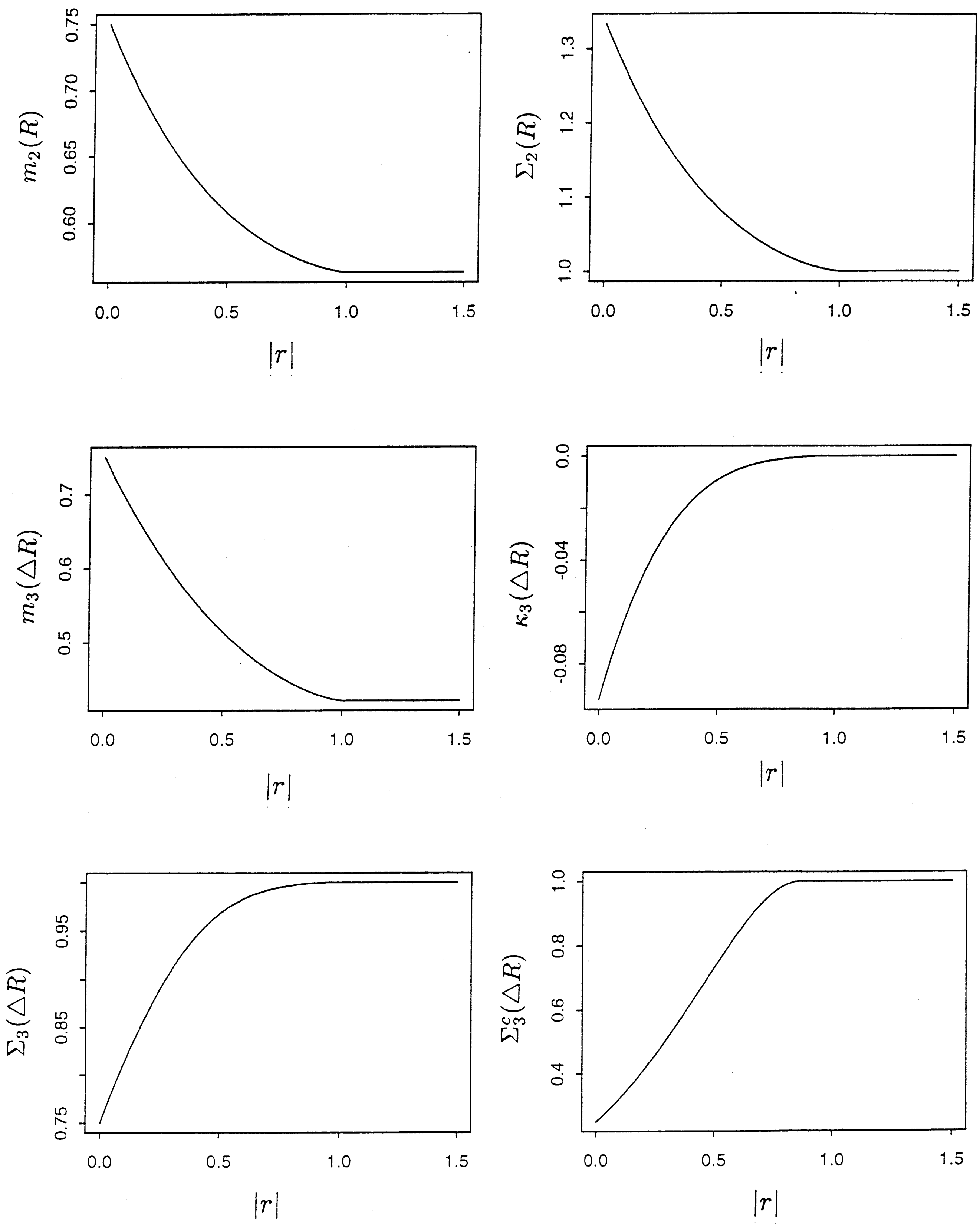

Figure 2: Moment and cumulant functions for the Boolean Model with $m_1 = 0.75$.

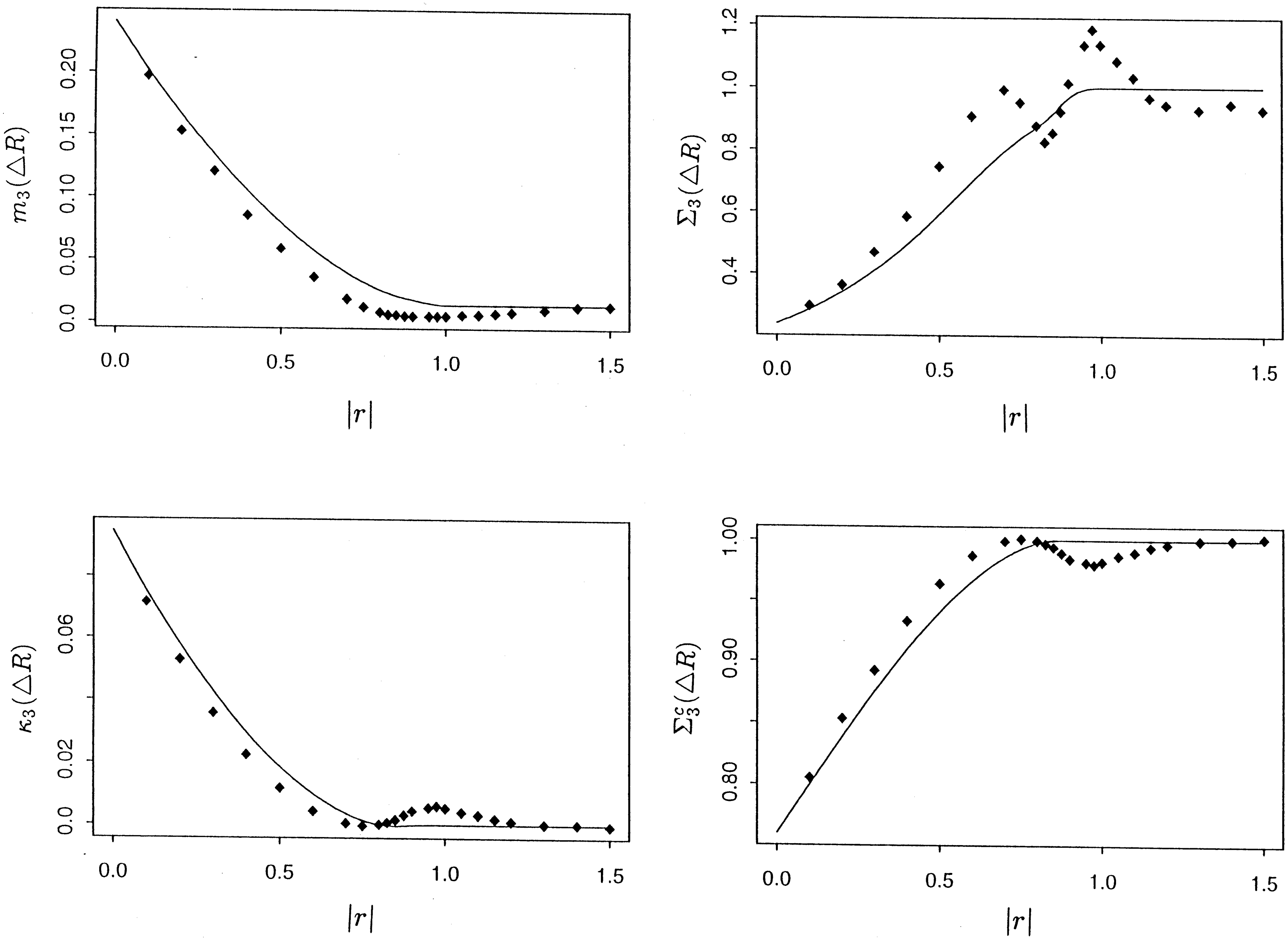

Figure 3: Third moment and cumulant functions for the Dead Leaves Model and a Boolean Model with the same value of m_1.

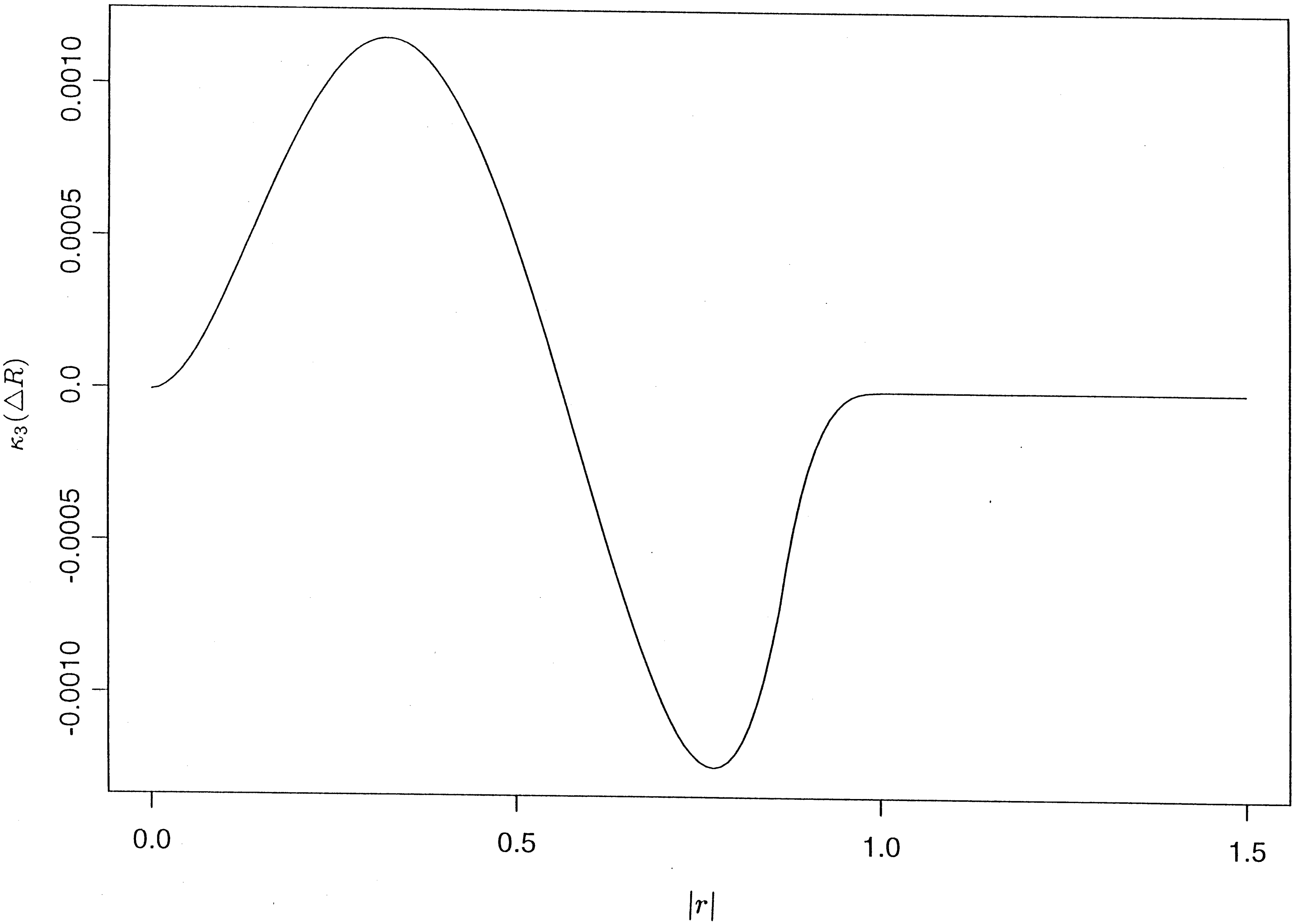

Figure 4: Third ordinary cumulant function for the Boolean Model with $m_1 = 0.5$.