ACTA STEREOL 1996; 15/1: 91-96
ORIGINAL SCIENTIFIC PAPER

SPHERICAL CONTACT AND NEAREST NEIGHBOUR DISTANCES IN BOOLEAN CLUSTER FIELDS

Ivan Saxl, Jan Rataj *

Mathematical Institute, Acad. Sci. of the Czech Republic, Žitná 25, CS-115 67 Praha 1, Czech Republic
*Charles University, Sokolovská 83, CS-186 00 Praha 8, Czech Republic

ABSTRACT.

First, Boolean cluster fields in $\mathbb{R}^d$ with grains being point clusters of several types (random points within a d–ball, random or regular points on a d–sphere, fixed or Poisson distributed number of points) are introduced. Then the formulae for spherical contact distance distribution function, its hazard rate, recluse probability of cluster and the nearest neighbour distance distribution function in such point fields are given. Finally, the statistical testing of cluster fields is briefly covered.

Key Words: Boolean cluster, hazard rate, nearest neighbour distance, recluse probability, spherical contact distance.

INTRODUCTION.

The recently renewed interest in the first contact distributions (Hansen *et al.*, 1994) invoked this paper, in which several cluster fields differing in the arrangement of daughter points and in the distribution of their number are compared on the basis of the derived formulae for important field characteristics. The possibility to recognize reliably cluster fields can be examined by means of test statistics for Poisson point process hypothesis (Cressie, 1992), some of which are sensible measures of the degree of clustering.

THEORY.

Let the random cluster $Z = \{z_1, \ldots, z_m\}$ be a random element of the space $\mathcal{Z}$ of all finite subsets of $\mathbb{R}^d$. Given the distribution of Z, its main characteristics are the mean number of points (called the daughters) in a cluster $N = \mathbf{E}\mathrm{card}\, Z$ and the cluster size, *e.g.* the mean content $\mathbf{E}\nu_k(\mathrm{conv}\, Z)$ of its convex hull (ν_k is the k–dimensional measure in $\mathbb{R}^d$, $0 \leq k \leq d$; hence ν_d is the volume content, ν_{d-1} is the surface content, ν_1 is the length *etc.*). Other useful quantities are interdaughter distances $\sigma_{ij} = \|z_i - z_j\|$, in particular their extreme value $\sigma_Z = \min(\sigma_{ij}; z_i, z_j \in Z, z_i \neq z_j)$. A *Boolean cluster model* X can be represented by $X = \bigcup_{i=0}^{\infty}(x_i + Z_i)$, where $\{x_i\}$ is a stationary Poisson point process (PPP) with the intensity λ in $\mathbb{R}^d$ (germ process) and Z_i are independent copies of a random cluster Z (grain process). X is a random closed set.

Let B_r be the d–ball of radius r centred in O, $B_r(x)$ its translate by x and κ_d the volume of B_1 . The important characteristics of cluster fields are the *mean volume of*

Research supported by Grant Agency of the Czech Republic (grant 201/93/2172).

the spherical dilation $\psi_Z(r) = \mathbf{E}\nu_d(Z \oplus B_r)$, the *spherical contact distribution function* $F_X(r) = Pr\{B_r \cap X \neq \emptyset\} = 1 - \exp(-\lambda\psi_Z)$, $r > 0$, and its *hazard rate* (defined in analogy to the survival analysis - Baddeley & Gill, 1993) $g_X(r) = -\mathrm{d}(\ln[1 - F_X(r)])/\mathrm{d}(\kappa_d r^d)$ with the interpretation that $g_X(r)\mathrm{d}\nu_d$ is the risk that a ball B_r missing X will hit it if its volume increases by $\mathrm{d}\nu_d = \mathrm{d}(\kappa_d r^d)$). Important is also the *nearest neighbour distance distribution function* $D_X(r) = 1 - [1 - F_X(r)]R_Z(r)$, where the *recluse probability* $R_Z(r)$ is the probability that there is no other point of the same cluster Z in a ball of radius r centred in a point $z_j \in Z$ (in the present case, $R(z)$ is equal to the *J-function* $J(r) = (1 - D(r))/(1 - F(r))$ introduced by Lieshout and Baddeley, 1995). It follows from the formulae given above that the knowledge of ψ_Z and R_Z permits to calculate all the remaining quantities.

A. *Mean dilation, spherical contact distribution function and hazard rate*

The formulae for $\psi_Z(r)$ in three particular cases of Boolean clusters are derived in Rataj & Saxl (1995):

i) deterministic clusters - Z is a fixed m-tuple of points (*e.g.* vertices of an $(m+1)$–simplex $\mathcal{S}_{m+1}$, $m + 1 \leq d$, or of a k-cube $\mathcal{C}_m$, $m = 2^k$, $k \leq d$, *etc.*), R is the radius of the circumball of Z. The mean dilation $\psi_Z(r) = \mathbf{E}\nu_d(Z \oplus B_r)$ in $\mathbb{R}^d$ can be given iteratively, for $\mathcal{C}_m, \mathcal{S}_{m+1}$ in $\mathbb{R}^2, \mathbb{R}^3$ explicitly (*cf.* Saxl, 1993).

ii) Poisson clusters - $Z \subset B_R$ (*globular* or *Matérn cluster*) or $Z \subset \partial B_R$ (*spherical cluster*) and m is Poisson distributed with the mean N:

(1)
$$\psi_Z(r) = \int_{\mathbb{R}^d} [1 - \exp(-\Lambda(B_r(x)))]\,\mathrm{d}x,$$

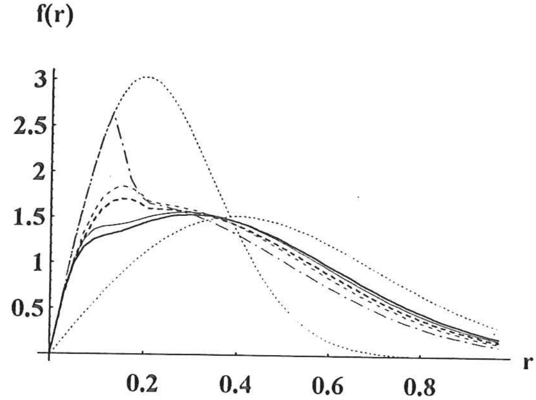
where $\Lambda(B_r(x)) = N\nu_d(B_R \cap B_r(x))/\nu_d(B_R)$ for globular clusters (Stoyan, 1992) and $\Lambda(B_r(x)) = N\nu_{d-1}(\partial B_R \cap B_r(x))/\nu_{d-1}(\partial B_R)$ for spherical clusters. The intersection contents $\nu_d(B_R \cap B_r(x))$ and $\nu_{d-1}(\partial B_R \cap B_r(x))$ are given in Rataj & Saxl (1995), for $d = 2$ in Stoyan & Stoyan (1994). Only a numerical calculation is possible here as well as in the following case.

iii) Binomial (globular and spherical) clusters - as *ii) with fixed* $m = N$ (a binomial cluster is a Poisson cluster with fixed number of daughters). Then

(2)
$$\psi_Z(r) = \int_{\mathbb{R}^d} (1 - (1 - \mu(B_r(x)))^N)\,\mathrm{d}x$$

and $\mu(B_r(x)) = \Lambda(B_r(x))/N$. The following notation will be used: P for Poisson, B for binomial, G for globular and S for spherical - hence PG is the Matérn cluster *etc.* Deterministic $\mathcal{C}_N$ clusters can be considered as regular spherical clusters on their circumball B_R, they will be denoted by RG and their size is characterized by the radius R. Finally, P is the parent Poisson point process (intensity λ) and D is the daughter Poisson point process (intensity λN) corresponding to the cluster size $R \to \infty$. The lengths (r, R *etc.*) will be measured in the units $(1/\sqrt{\lambda})^{1/d}$, or equivalently, we set $\lambda = 1$.

A rough description of the behaviour of clusters of equal size parameter R and equal mean daughter number N is as follows: at the values of $R < 0.1$ ($R > 1$), the Poisson point process of intensity λ ($N\lambda$) is approached. In the intermediate range,



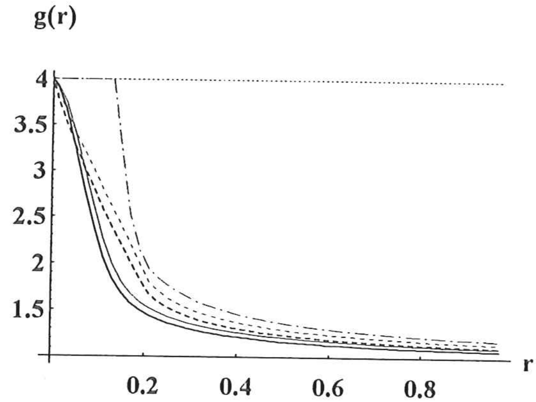


FIG.1.. The p.d.f. $f_Z(r)$ and the hazard rate $g_Z(r)$ for Boolean clusters of size $R = 0.2$ with $N = 4$. Z=RG (dash-dotted), PG (thick full), BG (thin full), PS (thick dashed), BS (thin dashed), P (lower dotted curve), D (upper dotted).

the characteristics of cluster fields of different types fulfill a sequence of inequalities similar to that of the expected volumes $\mathrm{E}\nu_d(\mathrm{conv}\ Z)$ (Affentranger, 1988), *e.g.*

$$(3)\quad \psi_D(r) \equiv N\psi_P(r) \geq \psi_{RG}(r) \geq \psi_{BS}(r) \geq \psi_{PS}(r) \geq \psi_{BG}(r) \geq \psi_{PG}(r) \geq \psi_P(r),$$

where $\psi_P(r) = \kappa_d r^d$. The remaining functions are calculated using the formulae given above. Two effects disturb this regular behaviour. First, at low values of N ($1 \leq N \leq 2$) and R, the non-zero Poissonian probability of a void cluster decreases the apparent intensity of Poisson clusters to $\lambda' < \lambda$. Further, at higher values of N and R, the Boolean spherical clusters approach the Poisson point process of intensity $N\lambda$ more slowly than the other cluster models.

Examples of shapes of the probability density functions $f_Z(r)$ and hazard rates $g_Z(r)$ are shown in Fig. 1 for clusters with $N = 4$, $R = 0.2$. Note that $f_{RG}(r) = f_D(r)$ and $g_{RG}(r) = g_D(r)$ for $r \leq R/\sqrt{d}$ (0.14 in the present case), which is a typical feature of deterministic clusters because of $\sigma_{RG} > 0$. The local maxima or inflections

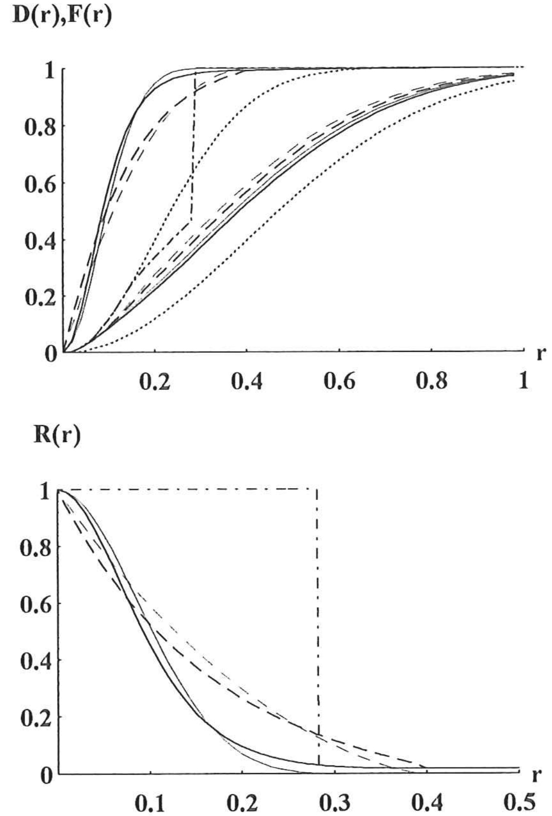


FIG.2.. The spherical contact distribution $F_Z(r)$, the nearest neighbour distance distribution $D_Z(r)$ and the recluse probability $R_Z(r)$ for Boolean clusters of size $R = 0.2$ with $N = 4$. Z=RG (dash-dotted), PG (thick full), BG (thin full), PS (thick dashed), BS (thin dashed), P (lower dotted), D (upper dotted).

in the shape of $f_Z(r)$ are observable only below $r \simeq 0.3$. The hazard rate of Poisson clusters is monotonously decreasing, a pronounced decrease ends at about $r \simeq R$. The difference in the shape occurs between spherical and globular clusters, mainly because of the difference in size (for moderate N, $\mathbf{E}\nu_d($ conv $Z)$ of Poisson globular cluster is roughly one half of the value corresponding to the similar Poisson spherical cluster). Especially at higher size of cluster, $R \simeq 0.7$ say, the hazard rate is more sensitive to clustering than other characteristics as it takes on the values from $(1, N)$ independently of R, whereas f_Z , F_Z as well as their moments approaches the characteristics of the daughter process D.

B.*Recluse probability and nearest neighbour distance distribution function*

The recluse probability R_Z in the above enumerated cases is as follows:

ad i) Clearly $R_Z(r) = 1$ for $r < R/\sqrt{d}$ (2^d–clusters) or $r < R[(1+d^{-1})/2]^{1/2}$ (regular simplices) and $R_Z(r) = 0$ otherwise.

ad ii) For Poisson clusters, $R_Z(r) = \exp[-N\pi(r)]$, where $\pi(r)$ is the probability that the ball of radius r centred at a random point of a cluster contains an independent random point with the same distance distribution from the parent point (see Stoyan & Stoyan, 1994 for the 2D version). Then for Poisson globular clusters,

$$(4) \quad \pi(r) = \frac{d}{\kappa_d R^{2d}} \int_0^R t^{d-1} \nu_d(B_R \cap B_r(t))dt \text{ for } r \le 2R \text{ and } \pi(r) = 1 \text{ otherwise.}$$

For Poisson spherical clusters, $\pi(r) = \nu_{d-1}(\partial B_R \cap B_r(R))/\nu_{d-1}(\partial B_R)$. Then for $r \le 2R$, we obtain $\pi(r) = \pi^{-1}\arccos(1 - 0.5r^2/R^2)$ for $d = 2$ and $\pi(r) = 0.25.r^2/R^2$ for $d = 3$).

ad iii) For binomial clusters, $R_Z(r) = (1 - \pi(r))^{N-1}$ with $\pi(r)$ as above.

Examples of shapes of the distributions $F_Z(r)$, $D_Z(r)$ and of the recluse probability $R_Z(r)$ are given in Fig.2, again for clusters with $N = 4$, $R = 0.2$ ($F_{RG}(r)$ is omitted in order to simplify the figure). Note that $D_{RG}(r) = D_D(r)$ for $r \le R/\sqrt{d}$ (0.14 in the present case), then it increases more slowly and jumps to 1 at $r = 2R/\sqrt{d}$. The values of $D_Z(r)$ for all other cluster types lie above $D_D(r) = F_D(r)$, but with increasing cluster size, $D_Z(r)$ and $F_Z(r)$ approach $F_D(r)$. The shapes of recluse probabilities are similar to those ones of hazard rates; the binomial and Poisson clusters of the same type behave likewise at $r < 2R$, only their asymptotic behaviour is different: for $r \ge 2R$, $R_{B\bullet}(r) = 0$ and $R_{P\bullet}(r) = \exp(-N)$.

TESTS OF POINT PATTERNS BASED ON DISTANCE METHODS.

Let $\{X_i\}$ and $\{W_i\}$ be samples of size n of spherical contact and nearest neighbour distances, respectively. As the moments of these distances are

$$\mu'_k = (\lambda\pi)^{-k/2}\Gamma(1 + k/2)$$

for PPP, the test statistics (and "measures" of clustering)

$$A(F) = 2\sqrt{\lambda_d}\sum X_i/n \,(\text{Clark and Evans}), C(F) = \pi\lambda_d \sum X_i^2/n \,(\text{Pielou})$$

have the asymptotic normal distributions $N(1, (4-\pi)/n\pi)$ and $N(1, 1/n)$, resp., for samples from PPP of intensity λ_d . Passing to squared distances, $I = \sum X_i^2 / \sum W_i^2$ (Hopkins) has, under the assumption of independence of $\{X_i\}$, $\{W_i\}$, Fisher $F_{2n,2n}$ distribution for PPP (Cressie, 1991). The first and second moments of $F(r)$, $D(r)$ for clusters of different size and type in the units of the corresponding moments of PPP, hence the asymptotic mean values of $A(\bullet)$, $C(\bullet)$ for clusters of various types, are summarized in the Tbl. 1. In the last column are the moment ratios ($\mu'_2(F)/\mu'_2(D)$ is the asymptotic mean of I).

Clearly, the Hopkins statistics is the most sensitive one and a reliable recognition of clustering can be expected within the whole range $R \in (0, 0.7)$ for all cluster types with the exception of RG at $R > 0.3$. The differentiation between various cluster types on the basis of statistics A, C would be difficult, but I should be sensible sufficiently. If, moreover, also $g_Z(r)$ and $R_Z(r)$ are estimated, more detailed information concerning clustering will be obtained. For example, if the examined cluster field is a mixture of clusters of several sizes, a stepwise decrease of $g_Z(r)$, $R_Z(r)$ is obligatory.

Table 1. Moments of spherical contact and nearest neighbour distances.

$N=4$ in $\mathbb{R}^2$	R	$\mu'_\bullet(F)$ PG	BG	PS	BS	RG	$\mu'_\bullet(D)$ PG	BG	PS	BS	RG	$\mu'_\bullet(F)$ / $\mu'_\bullet(D)$ PG	BG	PS	BS	RG
$\mu'_1(\bullet)$	0.1	1.8	1.8	1.8	1.7	1.7	0.23	0.21	0.28	0.28	0.49	7.9	8.4	6.3	6.2	3.3
	0.2	1.7	1.6	1.6	1.5	1.4	0.41	0.40	0.49	0.50	0.90	4.1	4.0	3.2	3.0	1.5
	0.4	1.4	1.4	1.3	1.3	1.1	0.68	0.70	0.72	0.75	1.05	2.1	2.0	1.8	1.7	1.0
	0.7	1.2	1.2	1.2	1.1	1.0	0.87	0.89	0.84	0.87	1.01	1.4	1.3	1.4	1.3	1.0
$\mu'_2(\bullet)$	0.1	3.5	3.4	3.3	3.1	3.0	0.11	0.05	0.14	0.09	0.19	33	73	23	34	15
	0.2	3.0	2.9	2.7	2.6	2.2	0.21	0.17	0.32	0.31	0.72	14	17	8.4	8.4	3.1
	0.4	2.3	2.1	1.9	1.8	1.4	0.51	0.50	0.65	0.67	1.13	4.5	4.3	3.0	2.6	1.2
	0.7	1.6	1.5	1.4	1.3	1.0	0.82	0.84	0.82	0.85	1.02	2.0	1.8	1.7	1.5	1.0

REFERENCES.

Affentranger F. Expected volume of a random polytope in a ball. J Microsc 1988; 151: 277-87.

Baddeley A, Gill RD. Kaplan-Meier estimators of interpoint distance distributions for spatial point processes. Res Rep BS-R9315, Amsterdam: CWI, 1993.

Cressie N. Statistics for spatial data. New York: J Wiley & Sons, Inc., 1991.

Hansen MB, Gill RD, Baddeley A. Some regularity properties for first contact distributions. Preprint nr 889, Dept of Mathematics. Utrecht: University Utrecht, 1994.

Lieshout MC, Baddeley A.: Spatial statistics. Lecture at the seminar "Mathematische Stochastik", Mathematisches Forschunginstitut, Oberwolfach, March 3.-5. 1995 (unpublished).

Rataj J, Saxl I. Boolean cluster models. Math Bohemica 1995; 120 (in press).

Saxl I. Spherical contact distances in Neyman-Scott process of regular clusters. Acta Stereol 1993; 12: 115-22.

Stoyan D. Statistical estimation of model parameters of planar Neyman-Scott cluster process. Metrika 1992; 39: 67-74.

Stoyan D, Kendall WS, Mecke J. Stochastic geometry and its applications. Berlin: Akademie-Verlag, 1987: 142-5.

Stoyan D, Stoyan H. Fractals, random shapes and point fields. New York: J.Wiley & Sons, 1994: 309-17.

Presented at the 9th International Congress for Stereology, Copenhagen, August 20th to 25th, 1995.