The Distribution Mapping Functions

Jiřina, Marcel
2015

Institute of Computer Science
Academy of Sciences of the Czech Republic

THE DISTRIBUTION MAPPING FUNCTION

Marcel Jiřina

Technical report No. 1222

April 2015

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 053 350, fax: +420 286 585 789,
e-mail:marcel@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

THE DISTRIBUTION MAPPING FUNCTION[1]

Marcel Jiřina

Technical report No. 1222

April 2015

Abstract:

The target of this study is to make clear the difference of the distribution mapping function introduced in 2003 and the classical notion of point processes theory, the counting function N, Ripley's K-function, and other two distance functions, F and G-functions.

We summarize here necessary starting points from the point process theory using the famous work by Baddeley, a two-volume book of Daley and Vere-Jones and short paper by Dixon. When dealing with the distribution mapping function we use up-to-date formulations used in various papers since 2003.

Keywords:
Multidimensional data, Correlation dimension, Distance, Metrics, Distribution Mapping Function.

[1]The work was done with institutional support RVO:67985807.

(30 April 2015)

THE MULTIDIMENSIONAL POINT PROCESSES AND THE DISTRIBUTION MAPPING FUNCTION

MARCEL JIŘINA,* *Institute of Computer Science AS CR*

Abstract

The target of this study is to make clear the difference of the distribution mapping function introduced in 2003 and the classical notion of point processes theory, the counting function N, Ripley's K-function, and other two distance functions, F and G-functions.

We summarize here necessary starting points from the point process theory using the famous work by Baddeley, a two-volume book of Daley and Vere-Jones and short paper by Dixon. When dealing with the distribution mapping function we use up-to-date formulations used in various papers since 2003.

Keywords: Multidimensional data, Correlation dimension, Distance, Metrics.

1. INTRODUCTION

The aim here is to formulate the problem of similarity and dissimilarity of the distribution mapping function to the counting function, Ripley's K-function and other similar functions as F- and G-function. For this purpose this work summarizes some important notions from the point processes theory.

2. DATA AND POINT PROCESS

2.1. Multivariate data in $\mathbb{R}^d$

Let the data set U of total $N_U = N(U) < \infty$ samples be given. Each sample $x_t = \{x_{t1}, x_{t2}, \dots x_{td}\}; t = 1, 2, ...N_U, x_{tk} \in \mathbb{R}; k = 1, 2, ..., d$ corresponds to a point in d-dimensional metric space $\mathbb{M}^d$, where d is the sample space dimension. For each $x_t \in U$ a mark, a class function $T : \mathbb{R}^d \to 1, 2, ...\mathcal{C} : T(x_t) = c; c \in 1, 2, ..., \mathcal{C}$ is introduced. With the class function the set U is decomposed into disjoint classes

* Postal address: Pod Vodarenskou vezi 2, 182 07 Praha 8, Czech Republic

$U_c = \{x_t \in U \mid T(x_t) = c\}; U = \cup_{c=1}^{C} U_c, U_c \cap U_b = \emptyset; c, b \in 1, 2, ..., \mathcal{C}; c \neq b$. Let the cardinality of set U_c be $N_c; \sum_{c=1}^{C} N_c = N_U$.

2.2. Point Process

2.2.1. *Point process N* (Acording to Pawlas [17] and Baudin [3] modified so that d-dimensional metric space $\mathbb{M}^{\,d}$ instead of d-dimensional Euclidean space $\mathbb{R}^d$ is considered.) Let B^d be the Borel σ-algebra (generated by open sets) in $\mathbb{M}^{\,d}$ and $B_0^d \subseteq \mathbb{B}^{\,d}$ be the system of all bounded Borel sets. We define the space of locally finite subsets of $\mathbb{M}^{\,d}$ as

$$N_{lf} = \{\, x \subseteq \mathbb{R}^d : N(x_B) \;<\; \infty \;\forall\; B \;\in \mathbb{B}_0^{\,d}\}$$

where $x_B = x \cap B$ and $N(y)$ denotes the cardinality of the set y. Elements of N_{lf} are called *locally finite point configurations*. We equip N_{lf} with σ-algebra

$$\mathcal{N}_{lf} = \sigma\ \{\ \{x \in N_{lf} : N(x_{B0}) = m\ \},\ m \in \mathbb{N}_0, B \in B_0^d\ \},$$

where $\mathbb{N}_0 = \mathbb{N} \cup 0 = 0, 1, 2, 3, ...$ A point process is defined as a random locally finite point configuration.

Definition 1. A point process X is a measurable mapping $X : (\omega, \mathcal{F}, \mathbb{P}) \to (N_{lf}, \mathcal{N}_{1f})$, where $(\omega, \mathcal{F}, \mathbb{P})$ is an abstract probability space. Distribution of the point process is a measure P_x on $(N_{lf}, \mathcal{N}_{1f})$ defined by the relation $P_x(F) = \mathbb{P}(X \in F) = \mathbb{P}(\ \{\ \omega \in \Omega : X(\omega) \in F\ \}\ ),\ F\ \in \mathcal{N}_{1f}$.

Definition 2. We say that the point process is finite if $n(X) < \infty$ almost surely.

Note 1. In these definitions only so-called *simple* point processes are considered, i.e. point processes where the points of X are mutually distinct. A point process can be defined more generally as a random locally finite integer-valued measure (allowing multiple points). Also note that the term "process" does not imply a dynamic evolution over time. Spatio-temporal point processes, where both time evolution and spatial dispersion of points are present, form the topic which is outside the scope of this study.

2.3. Multidimensional point process

Proposition 1. *Multidimensional point processes are a special case of random closed sets* [3].

In higher dimensions, there is no natural ordering of the points, so that there is no natural analogue of the inter-arrival times nor of the counting process. Instead, the most useful way to handle a spatial point process is to generalize the interval counts $N(a, b]$ to the region counts $N(B)$ = number of points falling in B defined for each bounded closed set $B \in \mathbb{M}^{d}$ [2].

2.4. Counting variables and vacancy indicators

For a point process X we will denote the number of points falling in the set B by $N(B) = n(X_B) = n(X \cap B)$ and refer to the function N as a *count function*.

Proposition 2. *X is a points process if and only if $N(B)$ is a random variable for any $B \in \mathcal{B}_0^d$.*

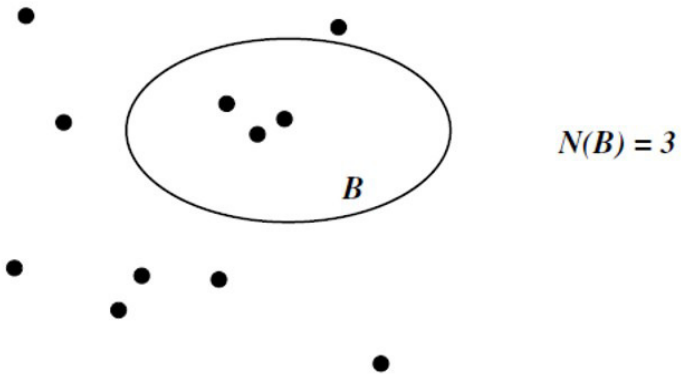

FIGURE 1: Counting variables $N(B)$ for a spatial point process.

It is often sufficient to study a point process using only the vacancy indicators $V(B) = 1\{N(B) = 0\} = 1$ (there are no points falling in B); 1 is the indicator function (1 if its operand is true, 0 otherwise).

The counting variables $N(B)$ are natural for exploring additive properties of a point process. For example, suppose we have two point processes, of 'red' and 'blue' points respectively, and we superimpose them (forming a single point process by discarding

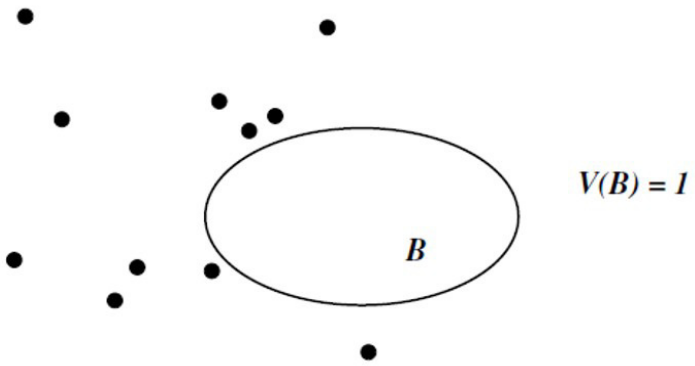

FIGURE 2: Vacancy indicators $V(B)$ for a spatial point process.

the colors). If $N_{red}(B)$ and $N_{blue}(B)$ are the counting variables for red and blue points respectively, then the counting variable for the superimposed process is $N(B) = N_{red}(B) + N_{blue}(B)$.

The vacancy indicators $V(B)$ are also natural for exploring geometric and 'multiplicative' properties of a point process. If $V_{red}(B)$ and $V_{blue}(B)$ are the vacancy indicators for two point processes, then the vacancy indicator for the superimposed process is $V(B) = V_{red}(B).V_{blue}(B)$. The values of the counting variables $N(B)$ for all subsets B give us sufficient information to reconstruct completely the positions of all the points in the process. Indeed, the points of the process are those locations x such that $N(x) > 0$. Hence we may as well define a point process as a collection of random variables $N(B)$ indexed by subsets B.

The counting variables $N(B)$ for different sets B satisfy certain relationships, including additivity $N(A \cup B) = N(A) + N(B)$ whenever A and B are disjoint sets $(A \cap B = \phi)$ and of course $N(\phi) = 0$. Furthermore, they are continuous in the sense that, if A_n is a decreasing sequence of closed, bounded sets $(A_n \subseteq A_{n+1})$ with limit $\bigcap_n \mathcal{A}_n = \mathcal{A}$, then we must have $N(A_n) \to N(A)$. These properties must hold for each realization (see Sect. 2.8) of the point process, or at least, with probability 1. They amount to the requirement that N is a measure (or at least, that with

probability 1, the values $N(B)$ can be extended to a measure). Formally, then, a point process may be defined as a random measure in which the values $N(B)$ are nonnegative integers. Usually it is assumed that the point process is locally finite: $N(B) < \infty$ with probability 1 for all bounded $B \subset \mathbb{R}^d$. That is, any bounded region contains only a finite number of points, with probability 1.

Definition 3. The point process is *simple* if

$$N(\{x\}) \leq 1 \; \forall \; x \; \in \mathbb{R}^d$$

with probability 1. That is, with probability 1, no two points of the process are coincident.

A simple point process can be regarded as a random set of points.

The vacancy indicators must satisfy

$$V(A \cup B) = min \; \{ \; V(A), V(B) \; \}$$

for any sets A, B, and have other properties analogous to those of the count variables $N(B)$. Thus, we could alternatively define a simple point process as a random function V satisfying these properties almost surely.

We use notation X (for a point process when it is considered as a random set) or N_x (for the counting variables associated with the same point process).

2.5. Finite dimensional distributions

Definition 4. The **finite-dimensional distributions** or **fidis** of a point process are the joint probability distributions of

$$(N(B_1), ..., N(B_m))$$

for all finite integers $m > 0$ and all compact $B_1, B_2, ...$

Equivalently, the fidis specify the probabilities of all events of the form

$$N(B_1) = k_1, ..., N(B_m) = k_m$$

involving finitely many regions [2].

Definition 5. The **capacity functional** of a simple point process X is the functional

$$T(K) = \mathbb{P}(N(K) > 0), \; K \; compact.$$

2.5.1. *Stationarity* The concept of a **stationary** point process plays an important role.

Definition 6. A point process X in $\mathbb{R}^d$ is called **stationary** if, for any fixed vector $v \in \mathbb{R}^d$, the distribution of the shifted point process $X + v$ (obtained by shifting each point $x \in X$ *to* $x + v$) is identical to the distribution of X.

Lemma 1. *A point process is stationary if and only if its capacity functional is invariant under translations,* $T(K) = T(K + v)$ *for all compact sets* $K \subset \mathbb{R}^d$ *and all* $v \in \mathbb{R}^d$.

2.5.2. *Randomness* According to [4] verbatim: "In a random distribution of a set of points on a given area, it is assumed that any point has had the same chance of occurring on any sub-area as any other point, that any sub-area of specified size has had the same chance of receiving a point as any other sub-area of that size, and that the placement of each point has not been influenced by that of any other point. Thus, randomness as here employed is a spatial concept, intimately dependent upon the boundaries of the space chosen by the investigator. A set of points may be random with respect to a specified area but decidedly non-random with respect to a larger space which includes the specified area. For meaningful results, therefore, the areas selected for investigation should be chosen with care."

The Poisson point process serves as a canonical model for no interaction between points (complete spatial randomness) [17].

Proposition 3. *A homogeneous Poisson point process is both stationary and isotropic.*

2.5.3. *The General Finite Point Process* Here we suppose only that the following conditions hold concerning a finite point process.

Conditions "5.3.1" [5]. (a) The points are located in a complete separable metric space (c.s.m.s.) χ, as, for example, $\chi = \mathbb{R}^d$.

(b) A distribution $\{p_n\}$ $(n = 0, 1, ...)$ is given determining the total number of points in the population, with $\sum_{n=0}^{\infty} p_n = 1$.

(c) For each integer $n \geq 1$, a probability distribution $\prod_n(.)$ is given on the Borel sets of $\chi^n \equiv \chi \times ... \times \chi$, and it determines the joint distribution of the positions of the points of the process, given that their total number is n.

Such a formulation provides a constructive definition that could be used to simulate

the process: first, generate a random number N_U according to the distribution p_n (and note that $Pr\{0 \leq N_U < \infty\} = 1$), and then, supposing $N_U = n$ and excepting the case $n = 0$ in which case there is nothing else to do, generate a random vector $(x_1, ..., x_n)$ according to the distribution $\Pi_n(.)$.

2.6. Marked Point Processes

There is the idea that the points of a point process might be labeled with extra information called marks. For example, in a map of the locations of emergency calls, each point might carry a label stating the time of the call and the nature of the emergency. A marked point can be formalized as a pair (x, m) where x is the point location and m is the mark attached to it [1].

Definition 7. A marked point process on a space S with marks in a space M is a point process Y on $S \times M$ such that $N_y(K \times M) < \infty$ $a.s.$ for all compact $K \subset S$. That is, the corresponding projected process (of points without marks) is locally finite.

Note 2. The space of marks M can be very general. It may be a finite set, a continuous interval of real numbers, or a more complicated space.

2.7. Transforming a Point Process

One pragmatic way to construct a new point process is by transforming or changing an existing point process. Convenient transformations include mapping, thinning, superposition, and clustering.

2.8. Repeated realizations of the point process

For point process X let in a region W there be a finite set of points $X_k = \{x_1, x_2, ...x_n\}$, where $x_i \in \mathbb{M}^{\,d}, i = 1, 2, ...$, i.e. $X_k \in W, k = 1, 2, ...$ and call it a realization of point process X in W.

There can appear many realizations $X_k, k = 1, 2, ...$ of process X mutually with different points and also with a different number of points lying in region W. We call processes W_k repeated realizations of point process X. As the X_k has arisen from the same data generating process X, we suppose that this process is statistically stable, stationary. What we are interested in are some (mostly statistical) chararacteristics of point process X. Some characteristics can be estimated from a single realization,

some not. For example, the distance to the nearest point of point x, the G function, the nearest neighbor function, can be estimated if X is a homogenous process.

3. DISTANCES IN POINT PROCESS

3.1. Empty Space Function F (The Contact Distribution)

One simple way to analyze a point process is in terms of the distances between points. If X is a point process, let $dist(u, X)$ for $u \in \mathbb{M}^d$ denote the shortest distance from the given location u to the nearest point of X. This is sometimes called the *contact distance*. Note the key fact that

$$dist(u, X) \leq r \; if \; and \; only \; if \; N(b(u, r)) > 0,$$

where $b(u, r)$ is the ball of radius r centred at x. Since $N(b(u, r))$ is a random variable for fixed u and r, the event $N(b(u, r)) > 0$ is measurable, so the event $dist(u, X) \leq r$ is measurable for all r, which implies that the contact distance $dist(u, X)$ is a well-defined random variable.

Definition 8. Let X be a stationary point process in $\mathbb{M}^d$. The contact distribution function or empty space function F is the cumulative distribution function of the distance

$$R = dist(u, X)$$

from a fixed point u to the nearest point of X. That is

$$\begin{aligned} F(r) &= P(dist(u, X) \leq r) \quad (1) \\ &= P(N(b(u, r)) > 0). \quad (2) \end{aligned}$$

Note 3. By stationarity this does not depend on u.

More generally:

Definition 9. Take a convex compact set $B \ni 0$. The contact distribution function F_B is given by

$$F_B(r) = \mathcal{P} \{ N(rB) > 0 \},$$

irrespective of $\mathcal{P}$ being stationary or not.

Note 4. a) the "ball" can be ball in the sense of different distance measures, eventually metrics; b) the last says the "local" nature of the F-function dependent on the "spatial origin". Here it is explicitly said that this case is a nonstationarity. Simply, the F-function is the distribution function of the distance of the (first) nearest neighbor (of the origin).

3.1.1. *Estimation of F-function from data* In applications, spatial point pattern data usually take the form of a finite configuration of points $x = x_1, ..., x_n$ in a region (**window**) W, where $X_i \in W$ and where $n = n(x) \geq 0$ is not fixed. The data would often be treated as a realization of a stationary point process X inside W. It is then important to estimate properties of the process X.

An estimator of F is

$$\hat{F}(r) = \frac{1}{\lambda_d(W)} \int_W 1\{dist(u, X) \leq r\} du. \tag{3}$$

[It means that for given r the integral goes over all points u in W, and for nearest neighbor, say p, of u, the case when $dist(u, p) \leq r$ is counted. The largest value found can be used as estimate of $\lambda_d(W)$. To keep estimate of F truly "local", the window W should be "small".]

A practical problem is that, if we only observe $X \cap W$, the integrand in (3) is not observable. When u is a point close to the boundary of the window W, the point of X nearest to u may lie outside W. More precisely, we have $dist(u, X) \leq r$ if and only if $n(X \cap b(u, r)) > 0$. But data are a realization of $X \cap W$, so we can only evaluate $n(X \cap W \cap b(u, r))$.

It was once a common mistake to ignore this, and simply to replace X by $X \cap W$ in (3). But this results in a negatively biased estimator of F. Call the estimator $\hat{F}_W(r)$. Since $n(X \cap W \cap b(u, r)) \leq n(X \cap b(u, r))$, we have

$$1 \{ \ n(X \cap W \cap b(u, r)) > 0 \ \} \leq 1 \ \{n(X \cap b(u, r)) > 0 \ \} \tag{4}$$

so that $E\hat{F}_W(r) \leq F(r)$. This is called a bias due to edge effects.

One simple strategy for eliminating the edge effect bias is the **border method**. When estimating $F(r)$, we replace W in equation (3) by the erosion

$$W_{-r} = W$$

so that

$$b(0,r) = x \in W : dist(X, \partial W) \geq r$$

consisting of all points of W that are at least r units away from the boundary ∂W. Clearly, $u \in W_{-r}$ if and only if $b(u,r) \subset W$. Thus, $n(x \cap b(u,r))$ is observable when $u \in W_{-r}$. Thus, we estimate $F(r)$ by

$$\hat{F}_b(r) = \frac{1}{\lambda_2(W_{-r})} \int W_{-r} 1\{dist(u,x) \leq r\} du.$$

This is observable, and by the previous argument, it is an unbiased estimator of $F(r)$. For a survey of corrections for edge effects, see [1].

Note 5. If W is large enough to contain all points of a realization of X and all nearest neighbors of points of X then no corrections are needed.

In short - the function F can be estimated from data, and provides a simple summary of the process. It can be useful in statistical analysis of point patterns. It is the distribution function of the distance from an arbitrary point of the process, selected as origin, to the nearest other point, of the process, i.e. the probability of existing nonzero points in the ball of radius r.

3.2. Nearest Neighbor Function; the G-function

A related concept to the F function is the nearest neighbor distance distribution G.

Definition 10. Let X be a stationary point process in $\mathbb{R}^d$. The **nearest neighbor function** G is the cumulative distribution function of the distance

$$R' = dist(x, X)$$

from a **typical point** (it means a chosen **fixed** point) $x \in X$ to the nearest other point of X. That is

$$G(r) = \mathbb{P}^x(dist(x, X \backslash x) \leq r) \tag{5}$$

$$= \mathbb{P}^x(N(b(x,r) \backslash x) > 0). \tag{6}$$

Note 6. This function depends heavily on *typical point* x. By stationarity, this does not depend on x.

3.2.1. *G-function estimation for stationary Poisson process.* For a stationary Poisson process in $\mathbb{R}^d$, since $X^x \equiv X \cup \{x\}$, we have

$$G(r) = \mathbb{P}^x(dist(x, X\backslash x) \leq r) \quad (7)$$

$$= \mathbb{P}(dist(x, X) \leq r) \quad (8)$$

In this case $G(r) \equiv F(r)$. If we write for each $x_i \in \mathbf{x}$

$$d_i = dist(x_i, \mathbf{x}\backslash x_i) \quad (9)$$

$$b_i = dist(x_i, \partial W) \quad (10)$$

so that d_i is the observed nearest-neighbor distance and b_i is the distance to the boundary of the observation window, then the estimator can be rewritten

$$\hat{G}_b(r) = \frac{\sum_i 1\ \{d_i \leq r, b_i \geq r\ \}}{\sum_i 1\ \{b_i \geq r\ \}}$$

G-function estimation. To find the G-function, suppose a given point of the process at the origin, and consider separately the distance to the nearest point from the same cluster, and to the nearest point from a different cluster. For any given cluster structure, there will be a well-defined distribution function tail, $Q_{c1}(r)$ say, for the probability that within a distance r of some given point of a cluster there is no other point of the same cluster. The distance to the nearest point in a different cluster, however, has the same distribution $F(r)$ as in (1). This implies [6] that

$$1 - G(r) = Q_{c1}(r)[1 - F(r)],$$

and hence that $J(r) = Q_{cl}(r)$. Thus, for a stationary Poisson cluster process, $J(r)$ is equal to the probability that no two points from the same cluster lie within a distance r of each other, and therefore satisfies $1 \geq J(r) \downarrow (0 \leq r \uparrow)$.

Estimation of $G(r)$ and edge effect corrections. Practical estimation of the F- and G-functions raises the usual problems of allowing for edge effects and possible biases arising from nonhomogeneity.

Here we mention only the edge correction for estimates of the nearest-neighbor distribution proposed by Hanisch in 1984 [5]. This has the advantage of preserving the monotonicity of the estimate as a function of r. It replaces the naive estimate

$$\hat{G}(r) = \frac{1}{N(W)} \sum_{k=1}^{N(W)} 1\ \{\ N[(S_r x_k) = 0]\ \}$$

with the form

$$\hat{G} + h(r) = \frac{\ell(W)}{N(W)} \sum_{k=1}^{N(W)} \frac{1\ \{\ N[(S_r x_k) \cap W = 0]\ \}}{\ell(W^{-d(x_k, \partial W)})} \tag{11}$$

where $d(x, \partial W)$ is the distance from the point x to the boundary of the observation region W.

The interpretation is that when a point x is too close to the boundary of W for the ball $S_r(x)$ to be wholly contained in W, the count from $S_r(x) \cap W$ is inflated by the weight factor $\ell(W)/\ell W^{-d(x_k, \partial W)}$.

See also Note 5.

3.3. J-function

An interesting combination of the empty space function F and the nearest neighbor function G is the following.

Definition 11. Let X be a *stationary* point process in R^d. The J-function of X is

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

for all $r \geq 0$ such that $F(r) < 1$.

For a uniform Poisson process, we know that $F(r) \equiv G(r)$ and hence $J(r) \equiv 1$. The J-function of a stationary process can be written explicitly in terms of the conditional intensity:

$$\begin{aligned}
J(r) &= \frac{\mathbb{P}^0(dist(0, X \backslash 0) > r)}{\mathbb{P}^0(dist(0, X) > r)} && (12)\\
&= \frac{\mathbb{P}^{!0}(dist(0, X) > r)}{\mathbb{P}(dist(0, X) > r)} && (13)\\
&= \frac{\mathbb{E}\left[\frac{\beta^*(0,X)}{\beta(0)} 1\{dist(0, X) > r\}\right]}{\mathbb{P}(dist(0, X) > r)} && (14)\\
&= \mathbb{E}\left[\frac{\beta^*(0, X)}{\beta(0)} \mid dist(0, X) > r\right]. && (15)
\end{aligned}$$

This representation can often be evaluated, while F and G often cannot be evaluated explicitly.

3.3.1. *Nearest neighbor distances in marked point processes* For nearest-neighbor distances there are in principle four different options to consider: the distance from a point of the process with arbitrary mark to the nearest point with arbitrary mark (giving the nearest-neighbor distribution function $C_g(x)$ for the ground process); the distance from a point with arbitrary mark at the origin to the nearest neighbor with mark in a specified set B [giving the distribution $G_{(g,B)}(x)$]; the distance from a point at the origin with specified mark κ to the nearest point of the process regardless of its mark, [giving $G_{(\kappa,g)}(x)$]; and the distance from a point with mark κ at the origin to the nearest point with mark in the subset $B \in B_\kappa$ [giving $G_{\kappa,B}(x)$].

3.4. Ripley's K-function

3.4.1. *Introduction.* Ripley's K and L functions [7] are closely related descriptive statistics for detecting deviations from spatial homogeneity. The K-function is informally defined as

$\lambda K(r) = E(number\ of\ points\ within\ r\ of\ the\ origin \mid point\ at\ the\ origin)$

The K-function sample-based estimate is defined as

$$\hat{K}(t) = \lambda^{-1} \sum_{i \neq j} 1\{d_{ij} < t\}/n,$$

where d_{ij} is the distance between the ith and jth points in a data set of n points, t is the search radius, λ is the average density of points (generally estimated as n/A, where A is the area (volume) of the region containing all points).

For data analysis, the variance stabilized Ripley K function called the L function is generally used. The sample version of the L function is defined as

$$\hat{L}(t) = \left(\hat{K}(t)/\pi\right)^{1/2}.$$

For approximately homogeneous data, the L function has expected value t and its variance is approximately constant in t. A common plot is a graph of $t - \hat{L}(t)$ against t, which will approximately follow the horizontal zero-axis with constant dispersion if the data follow a homogeneous Poisson process.

The function $K(t)$ does not uniquely define the point processes in the sense that two different processes can have the same $K(t)$-function. Also, while $K(t)$ is related to the nearest-neighbor distribution function $G(.)$, the two functions describe different

aspects of a point process. In particular, processes with the same $K(t)$-function may have different nearest-neighbor distribution functions, $G(t)$, and vice versa. $K(t)$ is also closely related to the pair correlation function, $g(t)$ [18]. Although it is usual to assume stationarity, $K(t)$ is interpretable for nonstationary processes because $K(t)$ is defined in terms of a randomly chosen event. It is also customary to assume isotropy, i.e. that one unit of distance in the y direction has the same effect as one unit of distance in the x direction.

3.4.2. *The second order measures*

Definition 12. *For a point process X we define n-th order moment measure by*

$$\mu^{(n)}(A) = \mathbb{E} \sum_{\xi_1,...,\xi_n,\in X} 1\{(\xi_1, ..., \xi_n) \in A\}, \quad A \in (\mathbb{B}^{\,d})^n.$$

Definition 13. Let X be a point process with intensity measure μ. We define the Campbell measure as

$$C(A) = \mathbb{E} \sum_{\xi \in X} 1\{(\xi, X) \in A\}, \quad \mathcal{A} \in \mathbb{B}^{\,d} \times \mathcal{N}_{1f}.$$

The reduced Campbell measure is given by

$$\mathcal{C}^!(A) = \mathbb{E} \sum_{\xi \in X} 1\{(\xi, X\ \xi) \in A\}, \quad \mathcal{A} \in \mathbb{B}^{\,d} \times \mathcal{N}_{1f}.$$

Note 7. Campbell measure is also determined by the relation

$$\mathcal{C}(B \times F) = \mathbb{E}1\ \{\ X \in F\ \}\ N(B), B \in \mathcal{B}^d,\ F \in \mathcal{N}_{1f}.$$

We will assume that intensity measure μ is σ-finite (it holds, for example, when it is locally finite). Then there exists (μ-a.s unique) Radon-Nikodym density $\xi \mapsto P_\xi(F)$, i.e.

$$\mathcal{C}(B \times F) = \int_B P_\xi(F)\mu(d\xi).$$

It can be shown that there exists a regular version $P_\xi(F)$, i.e. a Markov kernel:

(i) for any $F \in \mathcal{N}_{1f} \xi \mapsto P_\xi(F)$ is a nonnegative measurable function on $\mathbb{R}^d$,

(ii) for any $\xi \in \mathbb{R}^d, P_\xi(\cdot)$ is a probability measure.

Definition 14. The distribution P_ξ is called Palm distribution of a point process X at a point ξ. Analogously we can define the distribution $P_\xi^!$ called reduced Palm distribution. It satisfies the relation

$$\mathcal{C}^!(B \times F) = \int_B P_\xi^!(F)\mu(d\xi).$$

Note 8. Palm distribution P_ξ can be interpreted as the conditional distribution of a point process given that ξ is a point of the process. For $\varepsilon > 0$ small we have:

$$\begin{aligned}\mathbb{P}(X \in F | N(b(\xi,\varepsilon)) > 0) &= \frac{\mathbb{P}(X \in F, N(b(\xi,\varepsilon)) > 0)}{\mathbb{P}(N(b(\xi,\varepsilon)) > 0)} \approx \frac{\mathbb{E}\ 1\{X \in F\}\ N(b(\xi,\varepsilon))}{\mathbb{E}\ N(b(\xi,\varepsilon))} \\ &= \frac{\mathcal{C}(b(\xi,\varepsilon) \times F)}{\mu(b(\xi,\varepsilon))} \approx P_\xi(F).\end{aligned}$$

(I would write $\approx P_\xi\ (X)$ here.)

Probability that in point process X [that belongs to F (that is a part of point configurations algebra)] there is $N(b(\xi,\varepsilon)) > 0$ [the number of points in distance ε from point ξ] is approximately equal to Palm distribution of process X at point ξ. Palm distribution of process X at point ξ is approximately equal to the probability that in point process X [that belongs to F (that is a part of the algebra of point configurations)] there is $N(b(\xi,\ \varepsilon)) > 0$ [the number of points in distance ε from point ξ is positive].

3.4.3. *Ripley's K-function*

Definition 15. Let X be a point process with intensity function ρ. Suppose that the measure

$$\mathcal{K}(B) = \frac{1}{\mid A \mid}\mathbb{E} \sum_{\xi,\eta \in X, \xi \neq \eta} \frac{1\ \{\ \xi \in A, \eta - \xi \in B\ \}}{\rho(\xi)\rho(\eta)}, B \in \mathcal{B}^d,$$

where ξ is a point, and $\rho(.)$ is intensity at $(.)$, does not depend on the choice of $A \in \mathcal{B}^d$ with positive and finite Lebesgue measure $(0 < |A| < \infty)$. Then, X is said to be the second order intensity re-weighted stationary and $\mathcal{K}$ is called the second order reduced moment measure.

Proposition 4. Every stationary point process X is the second order intensity reweighted stationary.

The reduced second moment measure $\mathcal{K}$ carries important information about the dependence or **interaction** between different points of the process. For practical data

analysis, we need some simplification of the measure $\mathcal{K}$. Ripley [39] suggested the function

$$K(t) = \frac{1}{\beta}\mathcal{K}(b(o,t)),\ t \geq 0.$$

where $b(o,t)$ is a ball with center at o (origin) and radius t, and β is an intensity. Here it is supposed that any point $x \in X$ can be set (moved) to be the origin.

Definition 16. Let X be a second order intensity reweighted stationary point process. We define the second order reduced moment function or shortly **K-function** as

$$K(r) = \mathcal{K}(b(o,r)), r \geq 0,$$

and L-function as

$$L(r) = (K(r)/\omega_d)^{1/d}, r \geq 0,$$

where $\omega_d = |b(o,1)|$ is the volume of the d-dimensional unit ball.

Proposition 5. *For stationary Poisson point process,* $K(r) = \omega_d r^d$ *and* $L(r) = r$.

It was shown [7] that $\beta K(t)$ is the expected number of points y of the process that satisfy $0 < ||yx|| \leq t$ for a given point x of the process. In other words, $\beta K(t)$ is the expected number of points close to a given point of the process, where close means within a distance t.

Example 1. For a uniform Poisson process in $\mathbb{R}^d$,

$$K(t) = \omega_d t^d,\ t \geq 0,$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$.

Lemma 2. (Invariance of K under thinning). Suppose X is a stationary point process, and Y is obtained from X by random thinning (each point of X is deleted or retained, independently of other points, with retention probability p). Then, the K-functions of X and Y are identical.

3.4.4. *Relation to near neighbors distribution.* There is an important interpretation when X is a point process rather than a general random measure. We assume that the process is orderly. Then, the second order measure

$$M_2(A) = E(number\ point - pairs(x_i, x_j) : x_i \in U^d\ and\ x_j \in x_1 + A) \qquad (16)$$

$$E(rate\ of\ occurrence\ of\ point-pairs(x_i,x_j):x_j-x_i\in A) \tag{17}$$

Dividing by the mean density m (= intensity = average rate of occurrence) yields an interpretation of M_2 in terms of the expectation measure of the *Palm* process obtained by conditioning on the presence of a point at the origin:

$$E[number\ of\ poits\ x_i\in A\mid point\ at\ x=0]=M_2(A)/m \tag{18}$$

In considering the reduced measures $M_2(A)$ and related functions, spheres $S_r(0)$ constitute a natural class of sets to use for A in dimension $d\geq 2$; define

$$K_2(r)=M_2(S_r(0)\backslash 0)=M_{[2]}(S_r(0)) \tag{19}$$

the equivalent formulation here being a consequence of orderliness. Ripley [6] introduced this function, though what is now commonly called *Ripleys* K-function (including Ripley, 1981 [19]) is the density-free version

$$K(r)=\frac{M_2(S_r(0)\backslash\ 0)}{m^2}=\frac{K_2(r)}{m^2} \tag{20}$$

so, since $\lambda=m$ because of orderliness.

$$\lambda K(r)=E(number\ of\ points\ within\ r\ of\ the\ origin|point\ at\ the\ origin) \tag{21}$$

where on the right-hand side the origin itself is excluded from the count. The function $K(r)$ is monotonically nondecreasing on its range of definition $r>0$ and converges to 0 as $r\to 0$. This function is particularly useful in studying stationary isotropic point processes because it then provides a succinct summary of the second-order properties of the process. For a Poisson process, $K(r)=\ell(S_r(0))$.

Recall the definition of $K(r)$ in terms of the sphere $S_r(o)$. Noting the interpretation in (21), we see that the derivative $(d/dr)K_2(r)=K'(r)$ gives the conditional probability of a point on the surface of a spherical shell of radius r, conditional on a point at the center of the shell. Consequently, for an isotropic process in $\mathbb{R}^2$, the probability density that a point is located at distance r from a given point of the process and in the direction 0 equals $K'(r)/(2\pi r)$, independent of 0 because of isotropy. In dimension $d\geq 3$, the same quality holds on replacing the denominator $2\pi r$ by the surface area of $S_r(0)$.

3.4.5. *Estimating K-function*

K-function estimation from data – eliminating the edge effect. One simple strategy for eliminating the edge effect bias is the border method, When estimating $K(t)$, we replace W by the erosion

$$W_{-t} = W \ominus b(0,t) = x \in W : dist(x, \partial W) \geq t$$

consisting of all points of W that are at least t units away from the boundary ∂W. Clearly, $u \in W_{-t}$ if and only if $b(u,t) \subset W$. Thus, $n(x \cap b(x_i,t) \backslash x_i)$ is observable when $x_i \in W_{-t}$. Thus, we estimate $K(t)$ by

$$\hat{K}(t) = \frac{\sum_{x \in W_{-t}} n(x \cap (b(x,t) \backslash x)}{\hat{\beta} n(x \cap W_{-t})} \tag{22}$$

$$= \frac{\sum_{i=1}^{n} \sum_{j \neq i} 1\{\| x_i \| \leq t \}}{\hat{\beta} n(x \cap W_{-t})} \tag{23}$$

where $\hat{\beta}$ is usually $n(x)/\lambda_2(W)$. This is called the **border method** of edge correction. There are more sophisticated edge corrections with better performance.

Estimation of K-Functions for Multivariate Spatial Patterns The generalization of $K(t)$ to more than one type of point (a multivariate spatial point process) is

$K_{ij}(t) = \lambda_j^{-1} E$ [number of type j events within distance t of a randomly chosen type i event].

When there are g types of events, there are g^2 K functions, $K_{11}(t)$, $K_{12}(t)$, ...,$K_{1g}(t)$, $K_{21}(t)$, ..., $K_{2g}(t)$, ..., $K_{gg}(t)$. It is helpful to distinguish the cross-K functions $K_{ij}(t)$, where $i \neq j$ from the self-K functions, $K_{ii}(t)$. Analytical expressions for $K_{ij}(t)$ are known for various multivariate points. Various edge corrections have been suggested; one common example is the extension of Ripley's estimator:

$$\hat{K}_{ij}(t) = (\hat{\lambda_i} \hat{\lambda_j} A)^{-1} \sum_k \sum_l w(i_k, j_l) I(d_{i_k, j_l} < t)$$

where $w(i_k, j_l)$ is the fraction of the circumference of a circle centered at the kth location of process i with radius d_{i_k, j_l} that lies inside the study region, and A is the area of the study region.

If the spatial process is stationary, then corresponding pairs of cross-K functions are equal, i.e. $K_{12}(t) = K_{21}(t)$ and $K_{ij}(t) = K_{ji}(t)$. When edge corrections are used, then $\hat{K}_{ij}(t)$ and $\hat{K}_{ji}(t)$ are positively correlated but not equal. This suggests the use of a more efficient estimator, $K *_{ij} (t) = [\hat{\lambda}_j \hat{K}_{ij}(t) + \hat{\lambda}_i \hat{K}_{ji}(t)]/(\hat{\lambda}_i + \hat{\lambda}_j)$, although other linear combinations of $\hat{K}_{ij}(t)$ and $\hat{K}_{ji}(t)$ may have even smaller variance.

4. RESUME

4.1. Implicit assumptions

1. Where not given explicitly, the center is at point $0 = (0, 0, ..., 0)$ and this point is not a point of the point process X. More generally, any fixed point outside the point process X (eventually inside the point process but not counted) is considered.

2. We consider also a finite point process of points $\xi = \{\xi_1, \xi_2, \ldots, \xi_n\} \in X_\xi \subset \mathbb{M}^{\,d}$.

3. Distance does not necessarily mean the Euclidean distance, but some assertions above follow from implicit Euclidian metric assumption.

4.2. Summary of main properties

4.2.1. *Counting function* The **counting variable** $N(B)$ is the number of points in set B. $N(S_r)$ is the number of points in the ball of radius r (implicitly with the center at the origin, or at an *arbitrary fixed* point).

The vacancy indicator $V(B) = 1\{N(B) = 0\} = 1$, there are no points falling in B [else it is zero].

4.2.2. *F-function* The Empty Space Function, The Contact Distribution is the cumulative distribution function of the distance from a randomly selected point $u \notin X$ to the *first* nearest point of X.

$$1 - F(r) = \mathcal{P}N \{ (S_r) = 0 \}.$$

4.2.3. *G-function* The Nearest Neighbor Function, the probability of existing nonzero points in the ball of radius r and with the center in an arbitrary fixed point, the

distribution function of the distance from an arbitrary fixed point x (selected as origin) to the nearest other point of X.

$$\begin{aligned} G(r) &= P\ \{\ N(S_r \backslash 0) > 0\ \} \\ G(r) &= P^x(dist(x, X) \leq r) \\ G(r) &= P^x(N(b(x, r) \backslash x) > 0) \end{aligned}$$

$$G(r) \to 1\ for\ r\ \to \infty\ .$$

In X_ξ get for many (m) points the distances to nearest neighbors of $X_\xi, r_i, i = 1, 2, ...m$. Then

$$\hat{G}(r) = \frac{1}{m} \sum 1\ \{\ r > r_i\ \}.$$

Empirically one cannot get G from points of X only because for one sample of process X and one fixed point x_i we have one nearest point with one distance that cannot yield any statistics.

4.2.4. *J-function* It is the ratio of the two survivor functions

$$\begin{aligned} J(r) &= \frac{1 - G(r)}{1 - F(r)}\ if\ F(r) < 1, \\ J(r) &= 1\ if\ F(r) = 1. \end{aligned}$$

Alternatively

$$J(r) = \frac{\mathcal{P}^0\ \{\ N(s_r, 0) = 0\ \}}{\mathcal{P}\ \{\ N(S_r) = 0\ \}}.$$

For Poisson process $J(r)\ =\ 1$.

4.2.5. *K-function* This function can be defined as

$$\lambda K(r) = E(number\ of\ points\ within\ r\ of\ the\ origin \mid point\ at\ the\ origin).$$

Note that any *randomly selected* point can become the origin. In X_ξ get for many (m) points the distances to nearest neighbors of X_ξ, $\quad r_i, i = 1, 2, \ldots \mathrm{m}$. Then

$$\lambda \hat{K}(r) = \lambda \hat{K}(r) = \mathrm{E}_{m\,points} \sum_{all\ neighbors} 1\ \{\ r > r_i\ \}.$$

More concretely

$$\lambda \hat{K}(r) = \lambda \hat{K}(r) = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} 1\ \{\ r > r_i\ \}. \tag{24}$$

4.2.6. *Correlation integral, pair correlation function* In X_ξ get the distances of all pairs of points $X_\xi,\ r_k, k = 1, 2, ..., n(n-1)/2$. Then [9]

$$\hat{C_I}\,(r) = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} 1\ \{\ r >\ |r_i - r_j|\ \}.$$

The correlation integral is also called the "pair correlation function" and it is shown [18] that it is the conditional probability density of finding a particle at r, given that there is a particle at the origin. Thus, $\mathcal{C}_I(r)$ provides a measure of local spatial ordering.

5. DISTRIBUTION MAPPING FUNCTION

In this part we consider point process X_ξ.

5.1. Density

Let $W\ \subset\ R^d$ and

$$a(r, W) = \sum 1\{r_i < r\}. \tag{25}$$

An average of this quantity represents the relative frequency, or probability density of some particle being situated near r [18]. In other words, $\mathrm{E}a(r) = \rho(r)$ is simply the mean density at position r:

$$p\,(r) = \mathrm{E} \sum_i 1\{r_i < r\}.$$

5.2. Definitions

5.2.1. *Mapping the distribution* Let us have an example of a ball in an n-dimensional space containing uniformly distributed points over its volume. Let us divide the ball on concentric "peels" of the same volume. Using the formula $r_i = \sqrt[d]{V_i/S_d}$, which is, in fact, inverted formula for volume V_i of d-dimensional ball of radius r_i, we obtain a quite interesting succession of radii corresponding to the individual volumes - peels. The symbol S_d denotes the volume of a ball with unit radius in E^d; note $S_3\ =\ 4/3\pi$. A mapping between the mean density ρ_i in an i-th peel and its radius r_i is $\rho_i\ =\ p(r_i); p(r_i)$ is the mean probability density in the i-th ball peel with radius r_i. The probability distribution of points in the neighborhood of a given point x is thus

simplified to a function $p(r_i)$ of a scalar variable r_i. We call this function a probability distribution mapping function $D(x,\ r)$ and its partial differentiation with respect to r the distribution density mapping function $d(x,\ r)$. Functions $D(x,\ r)$ and $d(x,\ r)$ for x fixed are, in fact, the probability distribution function and the probability density function of variable r, i.e. of distances of all points from the given point x. More exact definitions follow [15].

Definition 17. The probability distribution mapping function $D(x,\ r)$ of the given point x is the function $D(x,r) = \int\limits_{B(x,r)} p(z)dz$, where r is the distance from the given point and $B(x,\ r)$ is ball with center x and radius r.

Definition 18. Let there be a finite point process X with total n points. The *empirical probability distribution mapping function* of the given point x is function

$$\hat{D}(x,r) = N(B(x,r))/n, \tag{26}$$

where r is the distance from the given point and $B(x,\ r)$ is ball with center x and radius r.

Note 9. This definition transforms the counting variable N over the ball $B(x,r)$ to a distribution function of its radius r.

Definition 19. The distribution density mapping function $d(x,\ r)$ of the given point x is function $d(x,r) = \frac{\partial}{\partial r}D(x,r)$, where $D(x,\ r)$ is a probability distribution mapping function of the given point x and radius r.

Note 10. When it is necessary (in the marked point processes) to differentiate class of point in distance r from point x, we write $D(x,\ r,\ c)$ or $d(x,\ r,\ c)$.

In X_ξ get the distances from a fixed point to all other points of $X_\xi, r_i, i = 1, 2, ...n-1$. Then

$$\hat{D}\left(r\right) = \frac{1}{n-1}\sum 1\ \{\ r > r_i\}\,.$$

Theorem 1. *The correlation integral is the mean of distribution mapping functions,*

$$\hat{C}_I\left(r\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\sum_{j=1;j\neq i}^{n} 1\ \{\ r > r_i\ \} = \frac{1}{n}\sum_{i=1}^{n} D(r_i, r).$$

Proof. For proof see [14].

Corollary. In point process X_ξ the λK is the mean of m distribution mapping functions $D(r_i,\ r)$

$$\lambda\hat{K}\,(r) = \lambda\hat{K}\,(r) = \frac{1}{m}\sum_{i=1}^{m} nD(r_i, r)\,.$$

Proof. Compare this relation with (24).

5.3. Comparisons of the D-function with F, G, and K-functions in X_ξ

1. The $D(x,\ r)$ is closely related to the counting function $N(B)$ over balls with center at x, see (26).

2. The $D(x,\ r)$ is reminiscent of the F-function (the contact distribution) in that both (D and F) go farther than to the first nearest neighbor. The F-function does not proceed from a fixed point but from an arbitrary point, i.e. probability is given by probabilities from ALL points of X_ξ.

3. The $D(x,\ r)$ is also suggestive of the G-function (The Nearest Neighbor Function, $\mathcal{F}(r) = \mathbb{P}^x(dist(x, X) \leq r)$) in the sense that it goes from a fixed point; but goes farther than to the first nearest neighbor only as the G-function does.

4. Both (D and K) go farther than to the first nearest neighbor but the K-function is constructed from all points (so that each point can be moved to the origin); the $D(x,\ r)$ relates to a single fixed point x. This comparison is analogous to comparison functions F and G.

5. Moreover, $D(x,\ r)$ is a component of the K-function according to the corollary above.

It follows from this that $D(.,.)$ differs from the N, F, G, and K-functions common in theory of multidimensional point processes.

Using notation above, one can write

$$D\,(x, r) = P(dist(x, X_\xi \setminus x) \leq r).$$

Note 11. The *dist* means any distance measure or metrics that can be used for measuring distances in R^d. There is no distance measure or metrics assumed in advance or implicitly, and thus generally when building the D/function we can use distance

measures as we like. This "as we like" is limited by the necessity not to get bizzare results. If the distance measure was semimetrics and continuous, the "embedding space" is a topological space [8] with all its advantages but also with disdvatages and limitations not being the Euclidean metric space as commonly assumed in the point processes theory.

5.4. ACKNOWLEDGEMENTS

This work was supported by the Institute of Computer Science of the Czech Academy of Sciences RVO: 67985807.

References

[1] Baddeley, A.J. (1999). Spatial sampling and censoring. In: Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. (eds.) *Stochastic Geometry. Likelihood and Computation.* Monographs on Statistics and Applied Probability, 80, Chapman and Hall/CRC, Boca Raton, FL (1999)

[2] Baddeley, A. (2007). Spatial point processes and their applications. In A. Baddeley, I. Bárány, R. Schneider, and W. Weil, editors, *Stochastic Geometry: Lectures given at the C.I.M.E.* Summer School held in Martina Franca, Italy, September 13-18, 2004, *Lecture Notes in Mathematics* 1892, Springer-Verlag, Berlin, 2007. ISBN 3-540-38174-0, pp. 1-75. MR2327290 (2008c:60045)

[3] Baudin, M. (1984). Multidimensional point processes and random closed sets. *Jour. of Applied Probability*, Vol. 21, pp. 173–178.

[4] Clark, P.J. and Evans, F.C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, Vol. 35, No. 4 (Oct., 1954), pp. 445–453.

[5] Daley, D.J., Vere-Jones, D. *An Introduction to the Theory of Point Processes I and II.* Springer 2002.

[6] Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Processes.* Arnold Publishers, London, 1983, 160 pp.

[7] Dixon, P.M. (2002). Ripley's K-function. In: El-Straavani, A.H.-Piegorsch, W.W. (Eds.) *Encyklopedia of Environmentrics.* John Wiley, Chichester, UK.

[8] Fischer, L. (1972). *A Survey of the Mathematical Theory of Multidimensional Point Processes.* In: Stochastic point processes: statistical analysis, theory, and applications, Peter A. W. Lewis (ed.), Wiley-Interscience, 1972, pp. 468-512.

[9] Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors, *Physica*, Vol. 9D, pp. 189–208.

[10] Hanisch, K.H. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Mathematische Operationsforschung und Statistik, series Statistics* 15, pp. 409–412.

[11] Jiřina, M. (2003). Dimensionality Reduction and Classification Using the Distribution Mapping Exponent. *Proceedings of the International Conference in Memoriam John von Neumann.* Budapest Muszaki Foiskola (Budapest Polytechnic), ISBN 963-7154-21-3. pp. 155-168. Held: International Conference in Memorial John von Neumann, Budapest, 12.12.2003, HU.

[12] Jiřina, M. (2004). Dimensionality Reduction and Classification Using the Distribution Mapping Exponent. *12th European Symposium on Artificial Neural Networks*, ESANN'2004, Bruges, Belgium, April 28-29-30,2004, Proceedings, Ed. M. Verleysen published by d-side, 24 av. L. Mommaerts - B-1140 Evere, Belgium ISBN 2-930307-04- 8, pp. 169–174.

[13] Jiřina, M. and Jiřina, Jr., M. (2009). "Classification by the Use of Decomposition of Correlation Integral" In: Abraham, A.; Hassanien, A.; Snášel, V. (Eds.) *Foundations of Computational Intelligence (Studies in Computational Intelligence 205)*, Springer, Berlin, pp. 39–55.

[14] Jiřina, M. and Jiřina Jr., M. (2013). "Utilization of singularity exponent in nearest neighbor based classifier", *Journal of Classification*, Vol. 30, No. 1, pp. 3–29.

[15] Jiřina, M. and Jiřina, Jr., M. (2014). "Correlation Dimension-Based Classifer", *IEEE Transactions on Cybernetics*, Vol. 44, in print.

[16] Jiřina, M. and Jiřina Jr., M. (2014b). "Classification Using Zipfian Kernel", *Journal of Classification* (Springer), Vol. 32, 2014, in print.

[17] Pawlas, Z. (2008). Spatial modelling and spatial statistics. *Course Notes*, Autumn 2008. Available at: http://www.math.ku.dk/ (Feb. 2015)

[18] Vesely, F. (2003). Statistical Physics and Simulation. *Course material. Chap. 5.2.* Pair Correlation Function. Available at: http://homepage.univie.ac.at/franz.vesely/simsp/dx/node22.html (Feb. 2015)

[19] Ripley, B.D. (1976). The Second Order Analysis of Stationary Point Processes. *J. Appl. Prob* Vol. 13, pp. 255-266.

6. APPENDIX – MOMENT MEASURES, CAMPBELL MEASURE, AND PALM DISTRIBUTION

(According to [17])

6.1. Moment measures

Definition 20. *For a point process X we define n-th order moment measure by*

$$\mu^{(n)}(A) = \mathbb{E} \sum_{\xi_1,...,\xi_n \in X} 1_[(\xi_1, ..., \xi_n) \in A] A \in (\mathcal{B}^d)^n$$

and n-th order factorial moment measure by

$$\alpha^{(n)}(A) = \mathbb{E} \sum_{\neq \xi_1,...,\xi_n \in X} 1_[(\xi_1, ..., \xi_n) \in A] A \in (\mathcal{B}^d)^n,$$

where $\sum^{\neq}_{\xi_1,...,\xi_n \in X}$ means that the summation goes over the n-tuples of mutually distinct points $\xi_1, ..., \xi_n$.

Remark 1. First order moment measures coincide with intensity measure: $\mu^{(1)}(A) = \alpha^{(1)}(A) = \mu(A)$ for any $A \in \mathcal{B}^d$. For $n = 2$ and $A = B_1 \times B_2 \in \mathcal{B}^d \times \mathcal{B}^d$ we have

$$\begin{aligned} \mu^{(2)}(B_1 \times B_2) &= E \sum_{\xi,\eta \in X} 1_{[\xi \in B_1 \eta \in B_2]} && (27) \\ &= EN(B_1)N(B_2) && (28) \\ &= E \sum_{\xi,\eta \in X}^{\neq} 1_{[\xi \in B_1 \eta \in B_2]} + E \sum_{\xi, \in X} 1_{[\xi \in B_1 \cap B_2]} && (29) \\ &= \alpha^{(2)}(B_1 \times B_2) + \mu(B_1 \cap B_2). && (30) \end{aligned}$$

Hence,

$$\alpha^{(2)}(B_1 \times B_2) = \mathbb{E}N(B_i)N(B_2) - \mathbb{E}N(B_1 \cap B_2).$$

Similarly, factorial moment measure is related to the factorial moment of the number of points in a given region,

$$\alpha^{(n)}(B \times ...B) = \mathbb{E}[N(B)(N(B) - 1)...(N(B) - n + 1)]. \quad (31)$$

6.2. Campbell measure

Theorem 2. (Campbell theorem) *For a point process X and arbitrary nonnegative measurable function h we have*

$$\mathbb{E}\sum_{\xi_1,...,\xi_n\in X} h(\xi_1,...,\xi_n) = \int ... \int h(\xi_1,...,\xi_n)\mu^{(n)}(d\xi_1,...,d\xi_n)$$

and

$$\mathbb{E}\sum_{\xi_1,...,\xi_n\in X, \xi_k\neq\xi_j, k\neq j} h(\xi_1,...,\xi_n) = \int ... \int h(\xi_1,...,\xi_n)\alpha^{(n)}(d\xi_1,...,d\xi_n)$$

Remark 2. If intensity function ρ and second order product density $\rho^{(2)}$ exist, we get from Theorem 7 the following useful formulas

$$\mathbb{E}\sum_{\xi\in X} h(\xi) = \int h(\xi)\rho(\xi)d\xi$$

for any measurable function $h : \mathbb{R}^d \to \mathbb{R}+$ and

$$\mathbb{E}\sum_{\xi,\eta\in X,\xi\neq\eta} h(\xi,\eta) = \int\int h(\xi,\eta)\rho^{(2)}(\xi,\eta)d\xi d\eta$$

for any measurable function $h : \mathbb{R}^d \times \mathbb{R}^d + \mathbb{R}+$.

Definition 21. *Let X be a point process with intensity measure μ. We define the Campbell measure as*

$$\mathcal{C}(A) = \mathbb{E}\sum_{\xi\in X} 1\ \{\ (\xi, X) \in A\ \}, A \in\ \mathcal{B}^d \times \mathcal{N}_{1f}.$$

The reduced Campbell measure is given by

$$\mathcal{C}^d(A) = \mathbb{E}\sum_{\xi} \in 1\ \{\ (\xi, X\ \xi)\ \in\ A\ \}\ , A \in \mathbb{B}^d \times \mathcal{N}_{1f}.$$

Remark 3. Campbell measure is also determined by the relation

$$\mathcal{C}(B \times F) = \mathbb{E}\ 1\ \{X \in F\ \}\ N(B), B \in \mathcal{B}^d, F \in \mathcal{N}_{1f}.$$

Note 12. $\mu(\cdot) = \mathcal{C}(\cdot \times N_{1f})$. For every $B \in \mathcal{B}^d$ and $F \in \mathcal{N}_{1f}, \mu(B) = 0$ implies $\mathcal{C}(B \times F) = 0$. It means that the measure $\mathcal{C}(\cdot \times F)$ is absolutely continuous with respect to μ. We will assume that intensity measure μ is σ-finite (it holds, for example, when it is locally finite). Then there exists (μ-a.s. unique) Radon-Nikodym density $\xi \mapsto P_{\xi(F)}$, i.e.

$$\mathcal{C}(B \times F) = \int_B P_\xi(F)\mu(d\xi).$$

It can be shown that there exists a regular version $P_\xi(F)$, i.e. a Markov kernel:
(i) for any $\mathcal{F} \in \mathcal{N}_{1f}, \xi \in\mapsto P_\xi(F)$ is a nonnegative measurable function on $\mathbb{R}^d$,
(ii) for any $\xi \in \mathbb{R}^d, P_\xi(\cdot)$ is a probability measure.

6.3. Palm distribution

In the study of stochastic processes, Palm calculus, named after Swedish teletrafficist Conny Palm, is the study of the relationship between probabilities conditioned on a specified event and time average probabilities. A Palm probability, or Palm expectation, is a probability or expectation conditioned on a specified event occurring at time 0.

Definition 22. *The distribution P_ξ is called Palm distribution of a point process X at a point ξ. Analogously, we can define the distribution $P^!_\xi$ called reduced Palm distribution. It satisfies the relation*

$$\mathcal{C}^!(B \times F) = \int_B P^!_\xi(F)\mu(d\xi).$$

Remark 4. Palm distribution P_ξ can be interpreted as the conditional distribution of a point process given that ξ is a point of the process. For $\varepsilon 0$ small we have:

$$\mathbb{P}(X \in F \mid N(b(\xi,\varepsilon)) > 0) = \frac{\mathbb{P}(X \in F, N(b(\xi,\varepsilon)) > 0)}{\mathbb{P}(N(b(\xi,\varepsilon)) > 0)} \approx \frac{E1\{X \in F\}N(b(\xi,\varepsilon))}{\mathbb{E}N(b(\xi,\varepsilon))} \quad (32)$$

$$= \frac{\mathcal{C}(b(\xi,\varepsilon))}{\mu(b(\xi,\varepsilon)) \approx P_\xi(F)}\ , \quad (33)$$

where $b(\xi\varepsilon)$ denotes a ball with centre ξ and radius ϵ. Similarly $P^!_\xi$ can be interpreted as the conditional distribution of a point process given that ξ is a point of the process that is not itself counted.

Theorem 3. (Slivnyak/Mecke) *Let X be a Poisson point process with intensity measure μ. Then*

$$\mathbb{E}\sum_{\xi\in X} h(\xi, X\ \{\ \xi\ \}\ ) = \int_{R^d} \mathbb{E}h(\xi, X)\mu(d\xi)$$

for any nonnegative measurable function h on $\mathbb{R}^d \times N_{1f}$.

Theorem 4. *For a stationary point process X with intensity ρ and for any nonnegative measurable function h,*

$$\mathbb{E}\sum_{\xi\in X} h(\xi, X) = \rho\int_{R^d}\int_{N_{1r}} h(\xi, x)P_\xi(dx)d\xi = \rho\int_{R^d}\int_{N_{1f}} h(\xi, x+\xi)P_0(dx)d\xi$$

and

$$\mathbb{E}\sum_{\xi\in X} h(\xi, X(\ xi)) = \rho\int_{R^d}\int_{N_{1r}} h(\xi, x)P_\xi(dx)d\xi = \rho\int_{R^d}\int_{N_{1f}} h(\xi, x+\xi)P_0^!(dx)d\xi$$

where $x+\xi = \eta+\xi : \eta \in x$.

Definition 23. *If both intensity function ρ and second order product density $\rho^{(2)}$ exist, we define the pair correlation function by*

$$g(\xi,\eta) = \frac{\rho^{(2)}}{(\xi,\eta)\rho(\xi)\rho(\eta)}, \xi,\eta \in \mathbb{R}^d : \rho(\xi) > 0.$$

If a point process is stationary, then $\xi^{(2)}(\xi,\eta) = \rho^{(2)}(\xi\eta, o) = \rho^{(2)}(\xi\eta)$ and $g(\xi,\eta) = \frac{\rho^{(2)}(\xi-\eta)}{\rho^2} = g(\xi-\eta)$ are functions of difference $\xi-\eta$. If X is moreover inotropic, then $\rho^{(2)}(\xi,\eta) = (||\xi-\eta||)$ and $g(\xi,\eta) = g(\|\ \xi-\eta\ \|)$ are functions of distance between ξ and η. Note that we abuse notation and denote by $\rho^{(2)}$ and g also the corresponding functions of $\xi-\eta$ or $\|\ \xi-\eta\ \|$.

Remark 5. Pair correlation function can take values in $[0,\infty)$. Hence, the name "correlation function" is slightly misleading.

ooo

o