The Annals of Applied Probability
1999, Vol. 9, No. 2, 465–492

POISSON APPROXIMATION IN CONNECTION WITH CLUSTERING OF RANDOM POINTS[1]

BY MARIANNE MÅNSSON

Chalmers University of Technology

Let n particles be independently and uniformly distributed in a rectangle $\mathbf{A} \subset \mathbb{R}^2$. Each subset consisting of $k \leq n$ particles may possibly aggregate in such a way that it is covered by some translate of a given convex set $C \subset \mathbf{A}$. The number of k-subsets which actually are covered by translates of C is denoted by W. The positions of such subsets constitute a point process on $\mathbf{A}$. Each point of this process can be marked with the smallest necessary "size" of a set, of the same shape and orientation as C, which covers the particles determining the point. This results in a marked point process.

The purpose of this paper is to consider Poisson (process) approximations of W and of the above point processes, by means of Stein's method. To this end, the exact probability for k specific particles to be covered by some translate of C is given.

1. Introduction. Assume n particles are uniformly and independently distributed in a rectangle $\mathbf{A} \subset \mathbb{R}^2$, and let $C \subset \mathbf{A}$ be a convex set, small relative to $\mathbf{A}$. To avoid problems with the boundaries of $\mathbf{A}$, the torus convention will be used throughout the paper. The aim of this paper is to investigate Poisson approximation of certain variables and point processes which concern subsets of particles which are covered by some translate of C.

First, a brief review of related problems will be given. Thereafter, necessary notation and some general Poisson approximation theorems are presented. In Section 3 the probability that a fixed number of particles aggregate in such a way that they are all covered by some translate of C is given, and its connection with integral geometry is discussed.

Section 4 constitutes the bulk of the paper. First, the number of subsets consisting of $k < n$ particles, called *k-subsets* in the following, which are covered by some translate of C is considered. The Stein–Chen method is used to bound the error made in the approximation of this number by a Poisson variable.

To each k-subset, which is covered by some translate of C, can be attached its position on $\mathbf{A}$ and the smallest $s \in \mathbb{R}^+$ for which some translate of sC covers the k particles. Poisson process approximation of three point processes determined by these positions and sizes is also dealt with in Section 4, as is

[1]Parts of this work was done while the author was visiting Universität Zürich, supported by the Swedish Natural Science Research Council (NFR).

AMS 1991 *subject classifications*. Primary 60D05; secondary 52A22, 60G55.

Key words and phrases. Poisson approximation, Stein's method, total variation distance, integral geometry, convex sets, mixed areas, Poisson process.

the joint distribution of the number of covered k-subsets, and the corresponding processes, for different k. The next section deals with asymptotic aspects, such as rates of convergence. The paper concludes with some problems of possible interest for more detailed investigation.

Note that the parameters used in the approximations follow from results in Section 3. For those readers whose main interest is the approximations and not the geometrical considerations leading to the parameters, it is enough to use Section 3 as a reference.

1.1. *A brief historical account.* There are $\binom{n}{k}$ different k-subsets of particles, some of which are covered by translates of C. Order the k-subsets in some way and let W denote the number of k-subsets which actually are covered by some translate of C. Then W can be written as

$$W = \sum_{i=1}^{\binom{n}{k}} I_i,$$

where

$$I_i = \begin{cases} 1, & \text{if the } i\text{th } k\text{-subset is covered by some translate of } C, \\ 0, & \text{otherwise.} \end{cases}$$

A k-subset which is covered by some translate of C will in this introduction be referred to as a k-aggregate in accordance with previous literature. Silberstein (1945), Mak (1948, 1949) and Eggleton and Kormack (1944) are concerned mainly with the calculation of the expected number of k-aggregates, $E[W]$, in one dimension (then C obviously is a line segment) and, for certain shapes of C, in two dimensions. Circular sets are handled by Mack (1948) and Silberstein (1945), while rectangular sets are handled by Mack (1949) and Eggleton and Kormack (1944). Mack (1949) discusses arbitrary shapes in two dimension and also gives some results in higher dimensions, and Mack (1948) argues for a Poisson limit of the number of k-aggregates as n tends to infinity.

The cases of a disc and a square are treated by Aldous (1989). He gives approximate formulas for the distribution of the radius (side length) of the smallest disc (square) which can be translated to a position where it covers at least k particles.

In the special case of $k = 2$ and a circular set, C, the number of k-aggregates equals the number of pairs of particles with interpoint distance less than the diameter of C. Convergence of this number to a Poisson limit is discussed in Silverman and Brown (1978). The results therein are complemented with a bound for the total variation distance in Silverman and Brown (1979). In, for example, Kryscio and Saunders (1983) Poisson convergence of the interpoint distances is considered in the case where the underlying process of particles is not necessarily stationary.

In Silverman and Brown (1978) the number of close pairs is given as an example of the classical U-statistics. Poisson approximation of U-statistics

and sums of dissociated variables (defined in Section 2.2) are treated by Barbour and Eagleson (1984) and Barbour, Holst and Janson (1992). In these two references, the Stein–Chen method is used to bound the total variation distance for Poisson approximations of the sums, in particular, the bound in Silverman and Brown (1979) is improved.

The number of close pairs is sometimes used to test the null hypothesis that a sample is uniformly distributed against some clustering or regular alternative. This is discussed, for example, in Silverman and Brown (1978) and in Barbour, Holst and Janson (1992), page 34.

Another related concept is the so-called scan statistic, that is, the maximal number of particles which are covered by some translate of C. In this context the particles usually constitute a Poisson process rather than being a fixed number. In one dimension, considerable attention has been devoted toward finding approximations of the distribution of the scan statistic; see, for example, Naus (1982), Alm (1983), Janson (1984) and Glaz (1989). The references on scan statistics in higher dimension are more scattered. The two-dimensional case is treated in Alm (1997) and Loader (1991), the latter, however, being restricted to rectangular sets.

In one dimension, the scan statistic is used in the same way as the number of close pairs is used in two dimensions, in testing the null hypothesis of an underlying Poisson process against a clustering or regular alternative.

2. Preliminaries.

2.1. *Notions and notation.* Let $\mathbb{R}^2$ denote the two-dimensional Euclidean space, with a fixed origin, O, and orthogonal coordinate-axes. The *area* of a (measurable) subset of $\mathbb{R}^2$ is its two-dimensional Lebesgue measure, which we denote by μ.

For $B, C \subset \mathbb{R}^2$ and $c \in \mathbb{R}$, the *Minkowski sum* and *scalar multiple* are defined as

$$B + C = \{x + y\colon x \in B,\ y \in C\} \quad \text{and} \quad cB = \{cx\colon x \in B\},$$

respectively. If $c = -1$ we get $\check{B} = \{-x\colon x \in B\}$, which we call the *reflected set* of B. For $x \in \mathbb{R}^2$, $B + \{x\}$ is the *translate* of B by x, which is denoted by $B + x$. If $B = \check{B} + x$ for some $x \in \mathbb{R}^2$, B is said to be *centrally symmetric*. An alternative, and for us more useful, way of writing the Minkowski sum is

$$B + C = \left\{x\colon B \cap (\check{C} + x) \neq \varnothing\right\}. \tag{2.1}$$

For the set $xB + yC$, where $x, y \in \mathbb{R}^+$ and $B, C \subset \mathbb{R}^2$ are nonempty convex sets, the area can be written as

$$\mu(xB + yC) = x^2\mu(B) + 2xy\nu(B, C) + y^2\mu(C), \tag{2.2}$$

where $\nu(B, C)$ is the *mixed area* of B and C, which is actually defined by (2.2) [see, e.g., Bonnesen and Fenchel (1948), page 40]. These mixed areas are usually only defined for nonempty sets. We make the natural extension that

if B or $C = \emptyset$, then $\nu(B, C) = 0$. It can be shown that if C is a convex set in $\mathbb{R}^2$, then

$$\mu(C) \le \nu(C, \check{C}) \le 2\mu(C), \tag{2.3}$$

where the lower bound is attained if and only if C is centrally symmetric, while the upper bound is attained if and only if C is a triangle. For a proof of these facts, see, for example, Bonnesen and Fenchel (1948), page 105.

Assume that the rectangle in which the particles are distributed, $\mathbf{A}$, is centered at the origin. Let $\mathscr{K}$ denote the family of convex sets, $C \subset \mathbb{R}^2$, with the properties that $C + \check{C} \subset \mathbf{A}$ and that O is an interior point of C.

Let $(\mathscr{X}, \mathscr{A})$ be any measurable space. The *total variation distance*, d_{TV}, between two probability measures μ and ν on $\mathscr{X}$ is defined to be

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathscr{A}} |\mu(A) - \nu(A)|.$$

If the state space is discrete, then

$$d_{\mathrm{TV}}(\mu, \nu) = \tfrac{1}{2} \sum_{i \in \mathscr{X}} |\mu\{i\} - \nu\{i\}|,$$

and in this case convergence in total variation distance, $d_{\mathrm{TV}}(\mathscr{L}(X_n), \mathscr{L}(X)) \to 0$, is equivalent to $\{X_n\}$ converging in distribution to X.

2.2. *General Poisson approximation theorems.* Following the notation of Barbour, Holst and Janson (1992), let Γ be an arbitrary finite collection of indices and let

$$W = \sum_{\alpha \in \Gamma} I_\alpha, \qquad \pi_\alpha = E[I_\alpha] \quad \text{and} \quad \lambda = E[W],$$

where I_α, $\alpha \in \Gamma$, are possible dependent, indicator variables. The following method for bounding the total variation distance between the distribution of $\mathscr{L}(W)$ and a Poisson variable with parameter λ is called the local approach of the Stein–Chen method. It is based on Stein's method, which is a general mean for approximating the distribution of random quantities. For Poisson approximations, it was worked out by Chen (1975), and further developed by Barbour and Eagleson (1983), Arratia, Goldstein and Gordon (1989) and by Barbour, Holst and Janson (1992), which hereafter is referred to as BHJ (1992). The Stein–Chen method is suitable to use when there is a natural dependence structure which allows every pair of indicators to be classified as either strongly or weakly dependent. For each $\alpha \in \Gamma$, let $\Gamma \setminus \{\alpha\}$ be divided into two subsets, one consisting of those $\beta \in \Gamma$ for which I_β is weakly dependent on I_α, and one consisting of the indices of the indicators which are strongly dependent on I_α. Denote these subsets by Γ_α^w and Γ_α^s, respectively, and let

$$Y_\alpha = \sum_{\beta \in \Gamma_\alpha^w} I_\beta \quad \text{and} \quad Z_\alpha = \sum_{\beta \in \Gamma_\alpha^s} I_\beta.$$

The function $g\colon \mathbb{Z}^+ \to \mathbb{R}$ in the theorem below, is a solution of the Stein equation $\lambda g(j+1) - jg(j) = I\{j \in A\} - \mathrm{Po}(\lambda)\{A\}$, $A \subset \mathbb{Z}^+$. See BHJ (1992) for details.

THEOREM 2.1 [Theorem 1.A, BHJ (1992).] *Let* Γ *be an arbitrary finite collection of indices. With the above definitions, for any choice of the index sets* Γ_α^w *and* Γ_α^s, $\alpha \in \Gamma$,

$$d(\mathscr{L}(W), \mathrm{Po}(\lambda)) \le \sum_{\alpha\in\Gamma} \left(\pi_\alpha^2 + \pi_\alpha E[Z_\alpha] + E[I_\alpha Z_\alpha]\right)\lambda^{-1}(1-e^{-\lambda}) + \sum_{\alpha\in\Gamma} \eta_\alpha \min(1, \lambda^{-1/2}),$$

where η_α *is any quantity satisfying*

$$\left|E[I_\alpha g(Y_\alpha + 1)] - \pi_\alpha E[g(Y_\alpha + 1)]\right| \le \eta_\alpha \sup_{j\ge 1} |g(j)|,$$

for instance,

$$\eta_\alpha = E\Big[\Big|E\Big[I_\alpha\Big|\big(I_\beta,\ \beta \in \Gamma_\alpha^w\big)\Big] - \pi_\alpha\Big|\Big].$$

The next theorem concerns approximation of marked point processes. Let $\Xi = \sum_{\alpha\in\Gamma} I_\alpha \delta_{Y_\alpha}$, where δ_y denotes the unit point mass at y and the Y_α's are random variables with state space $\mathscr{Y}$, which is assumed to be metric and separable. Then Ξ is a random element on the space $\mathscr{X}$ of configurations of finite point processes over $\mathscr{Y}$. Assume that a measure on $\mathscr{Y}$ is defined for each $\alpha \in \Gamma$ by

$$\Lambda_\alpha(A) = P(I_\alpha = 1, Y_\alpha \in A) \tag{2.4}$$

and let

$$\Lambda = \sum_{\alpha\in\Gamma} \Lambda_\alpha.$$

Furthermore, let

$$\Xi_\alpha^w = \sum_{\beta\in\Gamma_\alpha^w} I_\beta \delta_{Y_\beta} \quad \text{and} \quad \Xi_\alpha^s = \sum_{\beta\in\Gamma_\alpha^s} I_\beta \delta_{Y_\beta}.$$

The following theorem is the "local" version of Theorem 10.E in BHJ (1992). Its proof is a combination of the proofs of Theorem 10.A and Theorem 10.E of that reference, which should be consulted for more details.

THEOREM 2.2. *Let* Ξ *and* Λ *be defined as above. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda)) \le \sum_{\alpha\in\Gamma} \Bigg(\pi_\alpha^2 + \pi_\alpha E[Z_\alpha] + E[I_\alpha Z_\alpha] + \int_{\mathscr{Y}} d_{\mathrm{TV}}\big(\mathscr{L}(\Xi_\alpha^w \mid I_\alpha = 1, Y_\alpha = x), \mathscr{L}(\Xi_\alpha^w)\big)\Lambda_\alpha(dx)\Bigg).$$

PROOF. Let

$$(\mathscr{A}h)(\xi) = \int_{\mathscr{Y}} [h(\xi + \delta_x) - h(\xi)]\Lambda(dx) + \int_{\mathscr{Y}} [h(\xi - \delta_x) - h(\xi)]\,\xi\{dx\},$$

$\xi \in \mathscr{X}$, be the generator of the immigration-death process on $\mathscr{Y}$ with immigration intensity $\Lambda = \sum_{\alpha \in \Gamma} \Lambda_\alpha$ and unit per capita death rate, whose equilibrium distribution is that of the Poisson process on $\mathscr{Y}$ with intensity Λ. Let $h_{\lambda,B}$ be the solution of

$$(\mathscr{A}h)(\xi) = I(\xi \in B) - \mathrm{Po}(\Lambda)(B),$$

which can be found in Proposition 10.1.1 in BHJ (1992). Then

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda)) = \sup_{B \subset \mathscr{Y}} \left| E\left[(\mathscr{A}h_{\lambda,B})(\Xi)\right] \right|.$$

In order to bound $|E[(\mathscr{A}h_{\lambda,B})(\Xi)]|$ we write

$$\begin{aligned} E\left[(\mathscr{A}h)(\Xi)\right] &= E\Bigg\{ \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} \left[h(\Xi + \delta_x) - h(\Xi)\right] \Lambda_\alpha(dx) \\ &\qquad + \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} \left[h(\Xi - \delta_x) - h(\Xi)\right] \left(I_\alpha \delta_{Y_\alpha}\right)\{dx\} \Bigg\} \\ &= \sum_{\alpha \in \Gamma} \left\{ \int_{\mathscr{Y}} \left[h(\Xi - \delta_x) - h(\Xi)\right] \left(I_\alpha \delta_{Y_\alpha}\right)\{dx\} \right\} \\ &\qquad - E\left[I_\alpha \left(h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_{Y_\alpha}\right) \right) \right] \\ &\qquad + \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} E\left[h(\Xi + \delta_x) - h(\Xi)\right] \Lambda_\alpha(dx) \\ &\qquad + \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} E\left[h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_x\right)\right] \Lambda_\alpha(dx) \\ &\qquad + \sum_{\alpha \in \Gamma} E\left[I_\alpha \left(h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_{Y_\alpha}\right) \right) \right] \\ &\qquad - \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} E\left[h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_x\right)\right] \Lambda_\alpha(dx). \end{aligned} \tag{2.5}$$

Arguing much the same as in the proof of Theorem 10.A in BHJ (1992), the absolute value of the four lines following the second equality in (2.5) can be bounded by

$$\sum_{\alpha \in \Gamma} E[\, I_\alpha Z_\alpha\,] + \sum_{\alpha \in \Gamma} E[\, I_\alpha + Z_\alpha\,] E[\, I_\alpha\,],$$

respectively. Furthermore, the absolute value of the two last lines of (2.5) is equal to

$$\begin{aligned} \Bigg| \sum_{\alpha \in \Gamma} \int_{\mathscr{Y}} \Big\{ & E\left[h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_x\right) \mid I_\alpha = 1, Y_\alpha = x\right] \\ & - E\left[h\left(\Xi_\alpha^w\right) - h\left(\Xi_\alpha^w + \delta_x\right)\right] \Big\} \Lambda_\alpha(dx) \Bigg|. \end{aligned} \tag{2.6}$$

By Lemma 10.1.3 in BHJ (1992), $|h(\Xi_\alpha^w) - h(\Xi_\alpha^w + \delta_x)| \le 1$, and hence, using the definition of total variation distance, (2.6) is bounded by

$$\sum_{\alpha\in\Gamma}\int_{\mathscr{Y}} d_{\mathrm{TV}}\big(\mathscr{L}(\Xi_\alpha^w \mid I_\alpha = 1, Y_\alpha = x), \mathscr{L}(\Xi_\alpha^w)\big)\Lambda_\alpha(dx),$$

and the theorem is proved. □

If Γ is a collection of k-subsets of $\{1, 2, \dots, n\}$, then the family $\{X_\alpha\colon \alpha \in \Gamma\}$ of random variables is said to be *dissociated*, as defined by McGinley and Sibson (1975), if $(X_\alpha\colon \alpha \in A)$ and $(X_\alpha\colon \alpha \in B)$ are independent whenever $(\bigcup_{\alpha\in A}\alpha) \cap (\bigcup_{\alpha\in B}\alpha) = \varnothing$. If $k = 1$, "dissociated" is equivalent to "independent." In the following theorem, families of indicator variables belonging to a somewhat wider group are considered; instead of just one value of k, we let Γ be a collection of arbitrary subsets of $\{1, 2, \dots, n\}$.

THEOREM 2.3. (i) *Let Γ be a collection of subsets of $\{1, 2, \dots, n\}$ and the family $\{I_\alpha\colon \alpha \in \Gamma\}$ be such that $(I_\alpha\colon \alpha \in A)$ and $(I_\alpha\colon \alpha \in B)$ are independent whenever $(\bigcup_{\alpha\in A}\alpha) \cap (\bigcup_{\alpha\in B}\alpha) = \varnothing$. If $\Gamma_\alpha^s = \{\beta \in \Gamma\colon \beta \ne \alpha,\ \beta \cap \alpha \ne \varnothing\}$, then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \le \sum_{\alpha\in\Gamma}\big(\pi_\alpha^2 + \pi_\alpha E[Z_\alpha] + E[I_\alpha Z_\alpha]\big)\lambda^{-1}(1 - e^{-\lambda}).$$

(ii) *Let Γ and Γ_α^s be defined as above, and assume that the family $\{I_\alpha \delta_{Y_\alpha}\colon \alpha \in \Gamma\}$ is such that $(I_\alpha \delta_{Y_\alpha}\colon \alpha \in A)$ and $(I_\alpha \delta_{Y_\alpha}\colon \alpha \in B)$ are independent whenever $(\bigcup_{\alpha\in A}\alpha) \cap (\bigcup_{\alpha\in B}\alpha) = \varnothing$. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda)) \le \sum_{\alpha\in T}\pi_\alpha^2 + \pi_\alpha E[Z_\alpha] + E[I_\alpha Z_\alpha].$$

PROOF. The result is an immediate consequence of Theorem 2.1 and 2.2, since by the definition of Γ_α^s it follows that I_β and $I_\beta \delta_{Y_\beta}$, $\beta \in \Gamma_\alpha^w$, are independent of I_α and $I_\alpha \delta_{Y_\alpha}$, respectively. □

REMARK 2.4. In the case of dissociated indicators, i.e. if all subsets of Γ are of the same size, Theorem 2.3(i) coincides with Theorem 2.N in BHJ (1992).

REMARK 2.5. Note that the bounds in Theorem 2.3(i) and (ii) are equal except for the factor $(1 - \exp\{-\lambda\})/\lambda$.

3. The probability of covering of all particles. Assume $p_1, \dots, p_k$ are independently and uniformly distributed particles in the rectangle $\mathbf{A} \subset \mathbb{R}^2$ and let $C \in \mathscr{K}$. In this subsection we will derive the probability

$$P(\exists\, x \in \mathbf{A}\colon p_1, \dots, p_k \in C + x) \tag{3.1}$$

by means of results in integral geometry and give some historical background.

First, the case of two particles will be considered, so as to give some understanding of the formula for the probability (3.1), given in Theorem 3.2. In this case, the probability can be found by fixing one of the particles in an arbitrary position, and considering all possible positions for the other one, such that both particles are covered by some translate of C. By the independence and the uniform distribution of p_1 and p_2, and the torus convention, it does not matter where the first particle lies, and the quotient of the "possible" area for the second particle and $\mu(\mathbf{A})$ equals $P(\exists\ x \in \mathbf{A}\colon p_1, p_2 \in C + x)$.

EXAMPLE 3.1. If $C \in \mathscr{K}$ is a circle of radius r, the possible area for the second particle is a circle of radius $2r$. Then

$$P(\exists\ x \in \mathbf{A}\colon p_1, p_2 \in C + x) = 4\pi r^2/\mu(\mathbf{A}) = 4\mu(C)/\mu(\mathbf{A}).$$

EXAMPLE 3.2. Let $C \in \mathscr{K}$ be a triangle. Figure 1 shows C and the possible area for the second particle, given the position of p_1. As seen by the figure, this area is six times as large as that of the original triangle. Hence

$$P(\exists\ x \in \mathbf{A}\colon p_1, p_2 \in C + x) = 6\mu(C)/\mu(\mathbf{A}).$$

By these examples (which will turn out to be extreme; see Corollary 3.3), we learn that for the probability of covering the particles, it is not only the area of the sets which is of importance.

To handle the case of a general $C \in \mathscr{K}$, the following simple lemma is of considerable use, by giving an equivalent way of viewing the problem. Recall the definition of the reflection of C at the origin; $\check{C} = \{-x\colon x \in C\}$.

LEMMA 3.1. *Let $x_1, \ldots, x_k$ be arbitrary fixed particles in $\mathbb{R}^d$ and $C \subset \mathbb{R}^d$. Then*

$$\exists\ x \in \mathbb{R}^d \quad \textit{such that } x_1, \ldots, x_k \in C + x$$

if and only if

$$\bigcap_{i=1}^{k} \left(\check{C} + x_i\right) \neq \varnothing.$$

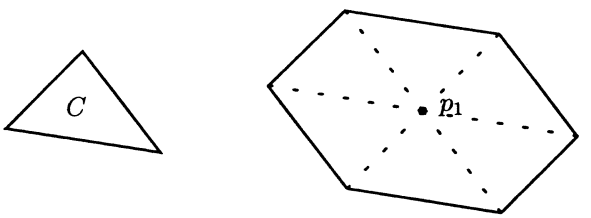

FIG. 1. *The possible area for the second particle.*

PROOF. First note that

$$x \in \check{C} + x_i \Leftrightarrow x - x_i \in \check{C} \Leftrightarrow x_i - x \in C \Leftrightarrow x_i \in C + x,$$

for every $i = 1, \dots, k$. Hence

$$\bigcap_{i=1}^{k} \left(\check{C} + x_i\right) \neq \varnothing \Leftrightarrow \exists\, x \in \mathbb{R}^d \colon x \in \bigcap_{i=1}^{k} \left(\check{C} + x_i\right)$$
$$\Leftrightarrow \exists\, x \in \mathbb{R}^d \colon x_1, \dots, x_k \in C + x. \qquad \square$$

By this lemma it follows that

$$P(\exists\, x \in \mathbf{A} \colon p_1, \dots, p_k \in C + x) = P\left(\bigcap_{i=1}^{k} (C + p_i) \neq \varnothing \right),$$

when the torus convention is used. It is now easy to derive an expression for the probability in case of two particles and a general convex set $C \in \mathscr{K}$. Since we may let $p_1 = O$, we are looking for the probability that p_2 belongs to the set $\{x \colon C \cap (C + x) \neq \varnothing\}$ which is precisely $C + \check{C}$ by (2.1). Thus

$$\begin{aligned} P(\exists\, x \in \mathbf{A} \colon p_1, p_2 \in C + x) &= P\big((C + p_1) \cap (C + p_2) \neq \varnothing\big) \\ &= \mu(C + \check{C})/\mu(\mathbf{A}) \\ &= 2\big(\mu(C) + \nu(C, \check{C})\big)/\mu(\mathbf{A}), \end{aligned}$$

where the last equality follows from (2.2). What determines the probability is hence the area and the mixed area of the set, where the latter is dependent on the shape of the set. This carries over to the case of more than two particles, as can be seen in Theorem 3.2.

A generalization of the problem to find an expression for $P((C + p_1) \cap \cdots \cap (C + p_k) \neq \varnothing)$, is to let sets $C_1, \dots, C_k$ of possibly unequal size and shape be translated by $p_1, \dots, p_k$, respectively. (Note that in the case of unequal sets there is obviously no equivalence corresponding to that in Lemma 3.1.) In fact it needs no more effort to handle this generalized case; thus we search the probability

$$P\big((C_1 + p_1) \cap (C_2 + p_2) \cap \cdots \cap (C_k + p_k) \neq \varnothing\big),$$

where $C_1, \dots, C_k \in \mathscr{K}$. Furthermore, for a while we extend the discussion to concern an arbitrary dimension, $\mathbb{R}^d$.

A consequence of using the torus convention is that we may assume that the first particle lies at the origin. Because of the assumptions that the origin lies in the centre of $\mathbf{A}$ and is an interior point of $C_1, \dots, C_k$ and of the restrictions on the sizes of $C_1, \dots, C_k$, all vectors $(x_2, \dots, x_k)$ such that

$$C_1 \cap (C_2 + x_2) \cap \cdots \cap (C_k + x_k) \neq \varnothing$$

satisfy $x_i \in \mathbf{A}$, $i = 2, \dots, k$, without using the torus convention. Hence, once the assumption that the first particle lies at the origin is made, we may treat $\mathbf{A}$ as a "normal" d-dimensional rectangle and not as a torus.

Since $p_2, \dots, p_k$ are independently and uniformly distributed in $\mathbf{A}$, the vector $(p_2, \dots, p_k)$ is uniformly distributed in the product space $\mathbf{A}^{k-1}$. Let μ_d^{k-1} denote the $(k-1)$-fold product measure of μ_d, the d-dimensional Lebesgue measure. Now $\mu_d^{k-1}(\mathbf{A}^{k-1}) = \mu_d(\mathbf{A})^{k-1}$ and

$$\begin{aligned} &P(C_1 \cap (C_2 + p_2) \cap \cdots \cap (C_k + p_k) \neq \varnothing) \\ &\quad = \mu_d^{k-1}\{(x_2, \dots, x_k) : x_i \in \mathbb{R}^d, \\ &\qquad\qquad C_1 \cap (C_2 + x_2) \cap \cdots \cap (C_k + x_k) \neq \varnothing\} / \mu_d(\mathbf{A})^{k-1}. \end{aligned}$$

To see the connection with integral geometry, we write

$$\begin{aligned} &\mu_d^{k-1}\{(x_2, \dots, x_k) : x_i \in \mathbb{R}^d, C_1 \cap (C_2 + x_2) \cap \cdots \cap (C_k + x_k) \neq \varnothing\} \\ &\quad = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} V_0(C_1 \cap (C_2 + x_2) \cap \cdots \cap (C_k + x_k))\, dx_2 \cdots dx_k, \end{aligned} \tag{3.2}$$

where

$$V_0(C) = \begin{cases} 1, & \text{if } C \neq \varnothing, \\ 0, & \text{if } C = \varnothing. \end{cases}$$

The functional $V_0(C)$ is one of the so-called intrinsic volumes of C, $V_i(C)$, $i = 0, \dots, d$, defined for compact, convex subsets of $\mathbb{R}^d$. The intrinsic volumes can be defined by the classical Steiner formula as follows. Let $\mathbf{B}^d$ be the d-dimensional unit ball, κ_d its volume and $\lambda \in \mathbb{R}^+$. Then

$$\mu_d(C + \lambda \mathbf{B}^d) = \sum_{i=0}^{d} \kappa_{d-i} \lambda^{d-i} V_i(C).$$

The most interesting, and therefore most studied, cases are $i = 0, d-1, d$. As already mentioned, V_0 is the indicator of nonempty sets, while $2V_{d-1}$ is surface area, and V_d is volume.

As early as 1937, explicit expressions for (3.2) were given for $k = 2, 3$ in two and three dimensions. Blaschke (1937) discusses both dimensions while Berwald and Varga (1937) handle three dimensions. In a probabilistic context, the planar case, including an iterated version (i.e., for an arbitrary number of sets), was rediscovered by Miles (1974), and the two- and three-dimensional cases by Månsson (1996).

In an arbitrary dimension and for general V_i, the integral in (3.2) is handled in Weil (1990), where it is a special case of an even more general situation. Hence the formula for $P(\bigcap_{i=1}^{k}(C_i + p_i) \neq \varnothing)$ follows directly in an arbitrary dimension. However, in higher dimensions, the formulas involve complicated functionals for which explicit descriptions are known only in special cases. Since the setting in this paper is two-dimensional, we present here only the probability in this case.

THEOREM 3.2. *Suppose $C_i \in \mathscr{K}$, $i = 1, \dots, k$, and that p_i, $i = 1, \dots, k$, are independently and uniformly distributed particles in $\mathbf{A}$, and $k = 2, 3, \dots$.*

Then, using the torus convention,

$$P\left(\bigcap_{i=1}^{k}(C_i + p_i) \neq \varnothing\right) = \left(\sum_{i=1}^{k}\prod_{\substack{j=1\\ j\neq i}}^{k}\mu(C_j) + \sum_{\substack{i,j=1\\ i\neq j}}^{k}\nu\left(C_i, \check{C}_j\right)\prod_{\substack{l=1\\ l\neq i,j}}^{k}\mu(C_l)\right) \times \frac{1}{\mu(\mathbf{A})^{k-1}}.$$

In particular, if $C_i = C$, $i = 1, \ldots, k$, *then*

$$P(\exists\, x \in \mathbf{A}\colon p_1, \ldots, p_k \in C + x) = P\left(\bigcap_{i=1}^{k}(C + p_i) \neq \varnothing\right) = \left(k + k(k-1)\frac{\nu(\check{C}, C)}{\mu(C)}\right)\frac{\mu(C)^{k-1}}{\mu(\mathbf{A})^{k-1}}.$$

It is the latter part of this theorem which will be useful in the rest of this paper. Corollary 3.3 follows directly from Theorem 3.2 and (2.3).

COROLLARY 3.3. *Under the assumptions of Theorem* 3.2,

$$k^2\frac{\mu(C)^{k-1}}{\mu(\mathbf{A})^{k-1}} \leq P(\exists\, x \in \mathbf{A}\colon p_1, \ldots, p_k \in C + x) \leq (2k^2 - k)\frac{\mu(C)^{k-1}}{\mu(\mathbf{A})^{k-1}},$$

where there is equality on the left if and only if C *is centrally symmetric and on the right if and only if* C *is a triangle.*

In this paper the discussions will be carried on in terms of covering particles. But throughout the paper we shall bear in mind that when reading, for instance, "k uniformly distributed particles are covered by C," we equally well can read, "k uniformly translated copies of $\check{C}$ have a nonempty intersection."

4. Poisson approximation. In this section we will consider Poisson approximation of the number of k-subsets of particles which are covered by translates of C, and Poisson process approximation of some point processes which arise in this connection. Results concerning one value of k in the variable and process cases are presented in Section 4.1.1 and 4.1.2, respectively, while Section 4.1.3 handles joint distributions. The proofs are deferred to Section 4.2. Without loss of generality, we will henceforth let $\mu(\mathbf{A}) = 1$.

4.1. *Main results.*

4.1.1. *The univariate case.* As noted in the introduction, the number of k-subsets which are covered by some translate of C can be written as

$$W = \sum_{i=1}^{\binom{n}{k}} I_i, \tag{4.1}$$

where

$$I_i = \begin{cases} 1, & \text{if there exists } x \in \mathbf{A} \text{ such that the } i\text{th } k\text{-subset} \\ & \quad \text{is covered by } C + x, \\ 0, & \text{otherwise.} \end{cases}$$

From Theorem 3.2 we know that

$$\begin{aligned} E[I_i] &= P(\exists\, x \in \mathbf{A}\colon p_1, \dots, p_k \in C + x) \\ &= \left(k + k(k-1)\frac{\nu(\check{C}, C)}{\mu(C)} \right) \mu(C)^{k-1}, \qquad i = 1, \dots, \binom{n}{k}, \end{aligned} \tag{4.2}$$

and hence that

$$\lambda = E[W] = \binom{n}{k} \left(k + k(k-1)\frac{\nu(\check{C}, C)}{\mu(C)} \right) \mu(C)^{k-1}. \tag{4.3}$$

W is a sum of indicators, where those pairs of indicators which concern k-subsets with common particles are dependent, while those with no particles in common are independent. In a situation such as this, the local version of the Stein–Chen method is a suitable mean to get a bound on the total variation distance between the distribution of W and a Poisson variable with parameter λ, and it leads to the following theorem.

THEOREM 4.1. *Let W and λ be defined by* (4.1) *and* (4.3), *respectively. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \le \left\{ \frac{\lambda k^2}{n} + \sum_{l=1}^{k-1} \binom{k}{l} \binom{n-k}{k-l} a_{k-l+1}\, \mu(C)^{k-l} \right\} (1 - e^{-\lambda}),$$

where $a_i = 2i^2 - i$.

REMARK 4.2. In case of $k = 2$ the bound can be somewhat improved by means of Theorem 2.O in BHJ (1992), which concerns families of strongly dissociated indicator variables. The idea is to reduce the number of indices in Γ_α^s.

4.1.2. *The process case.* The purpose of this section is to introduce three point processes determined by position and sizes of the k-subsets which are covered by some $C + x$ and consider approximation of these processes by Poisson processes. The point processes are defined as follows.

Let the leftmost particles (the lowest of these in case of ambiguity) in the k-subsets which actually are covered by some translate of C constitute the points of the point process $\Xi_{\mathbf{A}}$ on $\mathbf{A}$. The *size* of a k-subset we define to be the smallest s such that for some $x \in \mathbf{A}$, $sC + x$ covers the k-subset. If sizes are attached to the points of $\Xi_{\mathbf{A}}$, we get a point process on the space $\mathbf{A} \times [0, 1]$, which we denote by Ξ. These sizes are identically, but not independ-

ently, distributed with distribution function

$$
\begin{aligned}
F(y) &= P\big(\exists\, x \in \mathbf{A}\colon p_1, \dots, p_k \in yC + x \,\big|\, \exists\, x \in \mathbf{A}\colon p_1, \dots, p_k \in C + x\big) \\
&= \frac{\big(k + k(k-1)\nu\big(y\check{C}, yC\big)/\mu(yC)\big)\mu(yC)^{k-1}}{\big(k + k(k-1)\nu(\check{C}, C)/\mu(C)\big)\mu(C)^{k-1}} \qquad (4.4) \\
&= y^{2(k-1)},
\end{aligned}
$$

$0 \le y \le 1$, by (4.2) and since $\nu(y\check{C}, yC) = y^2\nu(\check{C}, C)$ and $\mu(yC) = y^2\mu(C)$. If we drop the positions and just consider the sizes, the result is a point process on $[0, 1]$, which we denote by $\Xi_{[0,1]}$.

All these three processes can be written as

$$
\sum_{i=1}^{\binom{n}{k}} I_i \delta_{Y_i},
$$

where the state space for $\{Y_i\}$, denoted by $\mathscr{Y}$, is $\mathbf{A}$, $\mathbf{A} \times [0, 1]$ or $[0, 1]$. We will first consider Ξ, for which $\mathscr{Y} = \mathbf{A} \times [0, 1]$, and we need to derive the measure on $\mathscr{Y}$ introduced in (2.4),

$$
\Lambda_i(A) = P(I_i = 1, Y_i \in A).
$$

The position of a k-subset which is covered by some translate of C, that is, its leftmost particle, is uniformly distributed on $\mathbf{A}$, since the k particles themselves are uniformly and independently distributed on $\mathbf{A}$, and the torus convention is used. The size of a k-subset which is covered takes its value in $[0, 1]$ and has density function

$$
2(k-1)y^{2k-3}
$$

by (4.4). Furthermore, the size is independent of the position of the k-subset. With $\pi = E[I_i]$ and λ defined as in (4.2) and (4.3), respectively, the measure Λ_i is thus given by

$$
\begin{aligned}
d\Lambda_i(x, y) &= \pi \frac{1}{\mu(\mathbf{A})} 2(k-1)y^{2k-3}\, dx\, dy \qquad (4.5) \\
&= 2\pi(k-1)y^{2k-3}\, dx\, dy, \qquad i = 1, \dots, \binom{n}{k},
\end{aligned}
$$

and we get

$$
\begin{aligned}
d\Lambda(x, y) &= \sum_{i=1}^{\binom{n}{k}} d\Lambda_i(x, y) = \binom{n}{k} 2\pi(k-1)y^{2k-3}\, dx\, dy \qquad (4.6) \\
&= 2\lambda(k-1)y^{2k-3}\, dx\, dy.
\end{aligned}
$$

The bound given below on the total variation distance between $\mathscr{L}(\Xi)$ and a Poisson process with intensity Λ equals the bound on the distance between $\mathscr{L}(W)$ and a Poisson variable with parameter λ given in Theorem 4.1 if $\lambda \le 1$. If $\lambda > 1$, it is unfortunately not as good in the process case.

THEOREM 4.3. *Let* $\Xi = \sum_{i=1}^{\binom{n}{k}} I_i \delta_{Y_i}$ *be the point process on* $\mathbf{A} \times [0,1]$ *defined above and let* Λ *be given by* (4.6). *Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda)) \le \frac{\lambda^2 k^2}{n} + \lambda \sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} a_{k-l+1}\, \mu(C)^{k-l},$$

where $a_i = 2i^2 - i$.

REMARK 4.4. Since $\Xi_{\mathbf{A}}$ and $\Xi_{[0,1]}$ and the corresponding Poisson processes are obtained as measurable mappings from Ξ and $\mathrm{Po}(\Lambda)$, respectively, it follows that

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi_{\mathscr{Y}}), \mathrm{Po}(\Lambda_{\mathscr{Y}})) \le d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda)),$$

where $\mathscr{Y} = \mathbf{A}$ or $[0,1]$, and $\Lambda_{\mathscr{Y}}$ is the measure corresponding to (4.6). Hence the bound in Theorem 4.3 holds also when these processes are concerned. □

4.1.3. *The multivariate case.* For $C^{(k_j)} \in \mathscr{K}$, where $j = 1, 2, \dots, m$, $k_j \in \{2, 3, \dots\}$ and $k_i \ne k_j$ if $i \ne j$, let

$$W^{(k_j)} = \sum_{i=1}^{\binom{n}{k_j}} I_{k_j, i} \tag{4.7}$$

and

$$\Xi^{(k_j)} = \sum_{i=1}^{\binom{n}{k_j}} I_{k_j, i}\, \delta_{Y_{k_j, i}}, \tag{4.8}$$

where $I_{k_j,i} = 1$ if the ith k_j-subset is covered by some translate of $C^{(k_j)}$ and 0 otherwise. From (4.3) we know that

$$\lambda^{(k_j)} = E[W^{(k_j)}] = \binom{n}{k_j} \pi_{k_j}, \tag{4.9}$$

where

$$\pi_{k_j} = E\big[I_{k_j,1}\big] = \left(k_j + k_j(k_j-1)\frac{\nu(\check{C}^{(k_j)}, C^{(k_j)})}{\mu(C^{(k_j)})}\right)\mu(C^{(k_j)})^{k_j-1}, \tag{4.10}$$

and furthermore, by (4.6),

$$d\Lambda^{(k_j)}(x,y) = \sum_{i=1}^{\binom{n}{k_j}} d\Lambda_i^{(k_j)}(x,y) = 2\lambda^{(k_j)}(k_j-1)y^{2k_j-3}\,dx\,dy, \tag{4.11}$$

is the measure on $\mathbf{A} \times [0,1]$, which is connected to $\Xi^{(k_j)}$.

We will start to approximate $\sum_{j=1}^m W^{(k_j)}$ by a Poisson variable with expectation

$$E\left[\sum_{j=1}^m W^{(k_j)}\right] = \sum_{j=1}^m \lambda^{(k_j)}.$$

THEOREM 4.5. *Let* $W^{(k_j)}$, $\lambda^{(k_j)}$ *and* π_{k_j} *be defined by* (4.7), (4.9) *and* (4.10), *respectively. Then*

$$
\begin{aligned}
&d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{j=1}^{m} W^{(k_j)}\right), \mathrm{Po}\left(\sum_{j=1}^{m} \lambda^{(k_j)}\right)\right) \\
&\quad \le \left(\sum_{j=1}^{m} \binom{n}{k_j}\left(\pi_{k_j}^2 + \pi_{k_j}\beta_{1j} + \beta_{2j}\right)\right) \frac{1 - \exp\left\{-\sum_{j=1}^{m}\lambda^{(k_j)}\right\}}{\sum_{j=1}^{m}\lambda^{(k_j)}},
\end{aligned} \tag{4.12}
$$

where

$$
\beta_{1j} = \sum_{l=1}^{k_j-1} \binom{k_j}{l}\binom{n-k_j}{k_j-l}\pi_{k_j} + \sum_{\substack{s=1\\ s\neq j}}^{m} \sum_{l=1}^{\min\{k_j,k_s\}} \binom{k_j}{l}\binom{n-k_j}{k_s-l}\pi_{k_s}, \tag{4.13}
$$

$$
\begin{aligned}
\beta_{2j} = {} & \sum_{l=1}^{k_j-1} \binom{k_j}{l}\binom{n-k_j}{k_j-l}\pi_{k_j}\mu\left(C^{(k_j)}\right)^{k_j-l} a_{k_j-l+1} \\
& + \sum_{\substack{s=1\\ s\neq j}}^{m} \sum_{l=1}^{\min\{k_j,k_s\}} \binom{k_j}{l}\binom{n-k_j}{k_s-l} \min\Big\{\pi_{k_j}\mu\left(C^{(k_s)}\right)^{k_s-l} a_{k_s-l+1}, \\
& \qquad\qquad \pi_{k_s}\mu\left(C^{(k_j)}\right)^{k_j-l} a_{k_j-l+1}\Big\},
\end{aligned} \tag{4.14}
$$

and $a_i = 2i^2 - i$.

We will now consider approximation of the distribution of the vectors $(W^{(k_1)}, \ldots, W^{(k_m)})$ and $(\Xi^{(k_1)}, \ldots, \Xi^{(k_m)})$ by the vectors of corresponding independent Poisson variables and Poisson processes, respectively.

THEOREM 4.6. *Let* $W^{(k_j)}$, $\Xi^{(k_j)}$, $\lambda^{(k_j)}$ *and* $\Lambda^{(k_j)}$ *be defined by* (4.7), (4.8), (4.9) *and* (4.11), *respectively. Then*

$$
\begin{aligned}
d_{\mathrm{TV}}\left(\mathscr{L}\left(\{W^{(k_j)}\}_{j=1}^{m}\right), \prod_{j=1}^{m}\mathrm{Po}(\lambda^{(k_j)})\right) &\le d_{\mathrm{TV}}\left(\mathscr{L}\left(\{\Xi^{(k_j)}\}_{j=1}^{m}\right), \prod_{j=1}^{m}\mathrm{Po}(\Lambda^{(k_j)})\right) \\
&\le \sum_{j=1}^{m}\binom{n}{k_j}\left(\pi_{k_j}^2 + \pi_{k_j}\beta_{1j} + \beta_{2j}\right),
\end{aligned}
$$

where π_{k_j}, β_{1j} *and* β_{2j} *are given by* (4.10), (4.13) *and* (4.14), *respectively.*

REMARK 4.7. Note that for $m = 1$ the bound in Theorem 4.6 coincides (after some rewriting) with the bound in Theorem 4.3, which concerns approximation of only one process. However, it is not as good as the bound in the univariate approximation in Theorem 4.1 (it lacks the factor $(1 - e^{-\lambda})/\lambda$).

4.2. *Proofs.* Proofs of the results in the previous subsection are given below. We start with Theorem 4.5, since that theorem is used in the rest of the proofs.

PROOF OF THEOREM 4.5. First we must introduce some new, local notation. Let

$$\Gamma_{j,i}^{j} = \left\{ l\colon l = 1,\dots,\binom{n}{k_j},\ l \neq i,\ (i\text{th } k_j\text{-subset} \cap l\text{th } k_j\text{-subset}) \neq \varnothing \right\},$$
$$j = 1,\dots,m,$$

$$\Gamma_{s,i}^{j} = \left\{ l\colon l = 1,\dots,\binom{n}{k_s},\ (i\text{th } k_j\text{-subset} \cap l\text{th } k_s\text{-subset}) \neq \varnothing \right\},$$
$$j = 1,\dots,m,\ s = 1,\dots,m,\ s \neq j,$$

$$Z_{k_j,i} = \sum_{s=1}^{m} \sum_{l \in \Gamma_{s,i}^{j}} I_{k_s,l}, \qquad j = 1,\dots,m,$$

so that $Z_{k_j,i}$ equals the sum of the number of k_s-subsets which are covered by some $C^{(k_s)} + x$, $s = 1,\dots,m$, and have at least one particle in common with the ith k_j-subset.

Note that if we let Γ consist of the indices of all k_j-subsets of $\{1,2,\dots,n\}$, $j = 1,\dots,m$, and $\Gamma_\alpha^s = \{\beta \in \Gamma\colon \beta \neq \alpha,\ \beta \cap \alpha \neq \varnothing\}$, then we can write

$$\sum_{j=1}^{m} W^{(k_j)} = \sum_{j=1}^{m} \sum_{i=1}^{\binom{n}{k_j}} I_{k_j,i} = \sum_{\alpha \in \Gamma} I_\alpha,$$

where $I_\alpha = 1$ if all p_i, $i \in \alpha$, are covered by some $C^{(|\alpha|)} + x$, $x \in \mathbf{A}$. Furthermore, if α is the index of the ith k_j-subset, then

$$Z_{k_j,i} = \sum_{s=1}^{m} \sum_{l \in \Gamma_{s,i}^{j}} I_{k_s,l} = \sum_{\beta \in \Gamma_\alpha^s} I_\beta.$$

Hence it is clear that the assumptions of Theorem 2.3(i) are satisfied. By noting that $Z_{k_j,i}$ as well as $I_{k_j,i}$ are equally distributed for all i, it follows that

$$d_{\mathrm{TV}}\left(\mathscr{L}\left(\sum_{j=1}^{m} W^{(k_j)} \right), \mathrm{Po}\left(\sum_{j=1}^{m} \lambda^{(k_j)} \right) \right)$$
$$\leq \left(\sum_{j=1}^{m} \binom{n}{k_j} \left(\pi_{k_j}^2 + \pi_{k_j} E\left[Z_{k_j,1} \right] + E\left[I_{k_j,1} Z_{k_j,1} \right] \right) \right) \frac{1 - \exp\left\{ -\sum_{j=1}^{m} \lambda^{(k_j)} \right\}}{\sum_{j=1}^{m} \lambda^{(k_j)}}.$$

We need now to derive bounds for $E[Z_{k_j,1}]$ and $E[I_{k_j,1} Z_{k_j,1}]$. Since $E[I_{k_s,i}] = \pi_{k_s}$ for all i by (4.10),

$$E\left[Z_{k_j,1} \right] = \sum_{s=1}^{m} \sum_{i \in \Gamma_{s,1}^{j}} E\left[I_{k_s,i} \right] = \sum_{s=1}^{m} \sum_{i \in \Gamma_{s,1}^{j}} \pi_{k_s}. \tag{4.15}$$

There are $\binom{k_j}{l}\binom{n-k_j}{k_j-l}$ different k_j-subsets with exactly l particles in common with the first k_j-subset, $l = 1, \dots, k_j - 1$, and there are $\binom{k_j}{l}\binom{n-k_j}{k_s-l}$ different k_s-subsets with exactly l particles in common with the first k_j-subset, $l = 1, \dots, \min\{k_j, k_s\}$. Hence (4.15) equals β_{1j} in (4.13).

Now, for each $l = 1, \dots, k_j - 1$, choose any of the indicators pertaining to a k_j-subset which has l particles in common with the first k_j-subset, and denote it by $I_{k_j}^l$. Denote by $I_{k_s}^l$ a corresponding variable, but concerning a k_s-subset which has l particles in common with the first k_j-subset, where $l = 1, \dots, \min\{k_j, k_s\}$. Then $E[Z_{k_j,1}I_{k_j,1}]$ can be rewritten as

$$
\begin{aligned}
E\left[Z_{k_j,1}I_{k_j,1}\right] = {} & \sum_{l=1}^{k_j-1}\binom{k_j}{l}\binom{n-k_j}{k_j-l}E\left[I_{k_j,1}I_{k_j}^l\right] \\
& + \sum_{\substack{s=1\\ s\neq j}}^{m}\sum_{l=1}^{\min\{k_j,k_s\}}\binom{k_j}{l}\binom{n-k_j}{k_s-l}E\left[I_{k_j,1}I_{k_s}^l\right].
\end{aligned} \tag{4.16}
$$

If $I_{k_j,1} = 1$ is given, we know that l of the particles which determine $I_{k_j}^l$ are close enough to be covered by some translate of $C^{(k_j)}$. The closer these l particles are, the larger is the probability that these l and the remaining $k_j - l$ particles which determine $I_{k_j}^l$ are all covered by some translate of $C^{(k_j)}$, that is, that $I_{k_j}^l = 1$. To get an upper bound on $P(I_{k_j}^l = 1 \mid I_{k_j,1} = 1)$, we can therefore think of the l common particles as having the same position, say $y \in \mathbf{A}$. By the discussion in Section 3, we then get

$$
\begin{aligned}
P\left(I_{k_j}^l = 1 \mid I_{k_j,1} = 1\right) &\le P\left(\exists\, x \in \mathbf{A}\colon p_1, \dots, p_{k_j-l},\, y \in C^{(k_j)} + x\right) \\
&= P\left(\exists\, x \in \mathbf{A}\colon p_1, \dots, p_{k_j-l+1} \in C^{(k_j)} + x\right) \\
&\le \mu\left(C^{(k_j)}\right)^{k_j-1} a_{k_j-l+1},
\end{aligned}
$$

where $a_i = 2i^2 - i$, by Corollary 3.3. Hence

$$
\begin{aligned}
E\left[I_{k_j,1}I_{k_j}^l\right] &= \pi_{k_j}P\left(I_{k_j}^l = 1 \mid I_{k_j,1} = 1\right) \\
&\le \pi_{k_j}\mu\left(C^{(k_j)}\right)^{k_j-l} a_{k_j-l+1}.
\end{aligned} \tag{4.17}
$$

As above,

$$
E\left[I_{k_j,1}I_{k_s}^l\right] \le \pi_{k_j}\mu\left(C^{(k_s)}\right)^{k_s-l} a_{k_s-l+1}, \tag{4.18}
$$

and by symmetry, another bound for $E[I_{k_j,1}I_{k_s}^l]$ is

$$
E\left[I_{k_j,1}I_{k_s}^l\right] \le \pi_{k_s}\mu\left(C^{(k_j)}\right)^{k_j-l} a_{k_j-l+1}. \tag{4.19}
$$

Insert (4.17)–(4.19) in (4.16) to obtain $E[Z_{k_j,1}I_{k_j,1}] \le \beta_{2j}$, where β_{2j} is given in (4.14).

PROOF OF THEOREM 4.1. Letting $m = 1$ and $k_1 = k$ in Theorem 4.5 and noting that $\pi_k = \lambda/\binom{n}{k}$ gives

$$
\begin{aligned}
&d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \\
&\quad = \left\{ \frac{\lambda}{\binom{n}{k}} + \frac{\lambda}{\binom{n}{k}} \sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} \right. \\
&\qquad \left. + \sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} a_{k-l+1}\, \mu(C)^{k-l} \right\} (1 - e^{-\lambda}).
\end{aligned} \tag{4.20}
$$

Let X be a hypergeometric random variable such that

$$
P(X = l) = \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}}, \quad l = 0, 1, \ldots, k \qquad \text{and} \qquad E[X] = \frac{k^2}{n}.
$$

Then

$$
\frac{1}{\binom{n}{k}} \sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} = P(X \ge 1) - \frac{1}{\binom{n}{k}} \le \frac{k^2}{n} - \frac{1}{\binom{n}{k}}
$$

by Markov's inequality, and the first two terms in the major brackets of (4.20) can be bounded by $\lambda k^2/n$, by which the theorem is proved. □

REMARK 4.8. Theorem 4.1 could be derived more directly by using (3.2) p. 35 in BHJ (1992).

PROOF OF THEOREM 4.3. Position pertaining to k-subsets with common particles are dependent, otherwise they are independent. The same holds for the sizes. Furthermore, this dependence structure is the same as that of $\left\{ I_i : i = 1, \ldots, \binom{n}{k} \right\}$. Hence the family $\left\{ I_i \delta_{Y_i} : i = 1, \ldots, \binom{n}{k} \right\}$ satisfies the as - sumptions of Theorem 2.3(ii). As noted in Remark 2.5, in such a case the bound of $d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda))$ given in Theorem 2.3(ii) coincides with the bound of $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda))$ given in Theorem 2.3(i), except for the factor $(1 - \exp(\lambda))/\lambda$. We used Theorem 4.5 in the special case where $j = 1$ to bound $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda))$. However, in that case Theorem 4.5 and Theorem 2.3(i) coincide, and the result follows without further calculations. □

PROOF OF THEOREM 4.6. To prove this theorem, we use another point process $\tilde{\Xi}$ which is the sum of the processes $\Xi^{(k_j)}$, $j = 1, \ldots, m$, $k_j \in \{2, 3, \ldots\}$, but with the additional information of which process $\Xi^{(k_j)}$ each point originates from. Hence, if $\Gamma^{(k_j)}$ contains the indices of the k_j-subsets, and $\Gamma = \bigcup_{j=1}^{m} \Gamma^{(k_j)}$, let

$$\tilde{\Xi} = \sum_{\alpha \in \Gamma} I_\alpha \delta_{\tilde{Y}_\alpha},$$

where the $\tilde{Y}_\alpha$'s take their values in

$$\tilde{\mathscr{Y}} = \mathbf{A} \times [0, 1] \times \{k_1, k_2, \ldots, k_m\}.$$

The first step is to derive $\tilde{\Lambda}_\alpha(A) = P(I_\alpha = 1,\ \tilde{Y}_\alpha \in A)$, where $A \subset \tilde{\mathscr{Y}}$. Note that if $\alpha \in \Gamma^{(k_j)}$, then

$$d\tilde{\Lambda}_\alpha(x, y, k_j) = 2\pi_{k_j}(k_j - 1) y^{2k_j - 3}\, dx\, dy,$$

by (4.5), and $d\tilde{\Lambda}_\alpha(x, y, z) = 0$ for $z \neq k_j$. Hence, for $(x, y, z) \in \tilde{\mathscr{Y}}$, let

$$d\tilde{\Lambda}_\alpha(x, y, z) = 2\pi_z(z - 1) y^{2z-3} 1\{\alpha \in \Gamma^{(z)}\}\, dx\, dy,$$

so that

$$\begin{aligned} d\tilde{\Lambda}(x, y, z) &= \sum_{\alpha \in \Gamma} d\tilde{\Lambda}_\alpha(x, y, z) \\ &= \sum_{\alpha \in \Gamma^{(z)}} d\tilde{\Lambda}_\alpha(x, y, z) = 2\lambda^{(z)}(z - 1) y^{2z-3}\, dx\, dy. \end{aligned}$$

Since

$$W^{(k_j)} = \int_A \int_{[0,1]} \Xi^{(k_j)}\{dx, dy\}$$

is a functional of $\Xi^{(k_j)}$, and

$$\Xi^{(k_j)}\{dx, dy\} = \tilde{\Xi}\{dx, dy, k_j\}$$

in turn is a measurable mapping of $\tilde{\Xi}$, it follows that

$$\begin{aligned} d_{\mathrm{TV}}\left(\mathscr{L}\left(\{W^{(k_j)}\}_{j=1}^{m}\right),\ \prod_{j=1}^{m} \mathrm{Po}(\lambda^{(k_j)})\right) &\le d_{\mathrm{TV}}\left(\mathscr{L}\left(\{\Xi^{(k_j)}\}_{j=1}^{m}\right),\ \prod_{j=1}^{m} \mathrm{Po}(\Lambda^{(k_j)})\right) \\ &\le d_{\mathrm{TV}}\left(\mathscr{L}(\tilde{\Xi}), \mathrm{Po}(\tilde{\Lambda})\right). \end{aligned}$$

For a fixed k we could use the bound on $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda))$ to get a bound on $d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathrm{Po}(\Lambda))$, because of the dependence structure of $\left\{I_i \delta_{Y_i},\ i = 1, \ldots, \binom{n}{k}\right\}$. Similarly we get a bound on $d_{\mathrm{TV}}(\mathscr{L}(\tilde{\Xi}), \mathrm{Po}\ (\tilde{\Lambda}))$ by means of the bound on $d_{\mathrm{TV}}(\mathscr{L}(\sum_{j=1}^{m} W^{(k_j)}), \mathrm{Po}(\sum_{j=1}^{m} \lambda^{(k_j)}))$ given in Theorem 4.5. □

5. Asymptotic results. For a fixed C and k, the errors made by Poisson approximating of course get large when n does. We will now consider asymptotic aspects of the previous results as k is kept fixed, but the areas of the sets decrease as n grows. Since n is no longer fixed, the quantities which change with n are here indexed by n.

Particular attention will be paid to the case where $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$, is a sequence of sets such that $\mu(C_n) = O(n^{-t})$, $t \in \mathbb{R}$. If

$$\mu(C_n) = cn^{-t},$$

where $c > 0$ and $t \in \mathbb{R}$ are constants, then the expectation of the number of k-subsets which are covered by some translate of C_n is

$$\begin{aligned} \lambda_n = E[W_n] &= \binom{n}{k}\left(k + k(k-1)\frac{\nu(C_n, \check{C}_n)}{\mu(C_n)}\right)c^{k-1}n^{-t(k-1)} \\ &= O(n^{k-t(k-1)}). \end{aligned} \tag{5.1}$$

The limit of this expectation depends on the value of t,

$$\lambda_n \to \begin{cases} 0, & \text{if } t > k/(k-1), \\ \infty, & \text{if } t < k/(k-1) \end{cases} \quad \text{as } n \to \infty, \tag{5.2}$$

and if $t = k/(k-1)$,

$$\frac{k^2c^{k-1}}{k!} \le \liminf_{n\to\infty} \lambda_n \le \limsup_{n\to\infty} \lambda_n \le \frac{(2k^2-k)c^{k-1}}{k!},$$

where the bounds follow by Corollary 3.3. If the sets C_n are such that $\nu(C_n, \check{C}_n)/\mu(C_n)$ is constant for all n, for instance if they are of equal shape, and $t = k/(k-1)$, then

$$\lambda_n \to \alpha_{C,k}c^{k-1}/k!,$$

where $\alpha_{C,k} = k + k(k-1)\nu(C_n, \check{C}_n)/\mu(C_n) \in [k^2, 2k^2 - k]$.

In Section 5.1 and 5.2 we will see that:

1. If $\lambda_n \to 0$, $(t > k/(k-1))$, then the total variation distance tends to zero at the same rate both in the univariate and in the process case.
2. If λ_n stays bounded away from 0 and ∞, $(t = k/(k-1))$, then the total variation distance tends to zero at the same rate in both cases.
3. If $\lambda_n \to \infty$, $(t < k/(k-1))$, but not too fast, the distance also tends to zero in both cases. However, here the univariate case is doing better: the rate of convergence is faster, and the range of values of t for which the distance really tends to zero is larger.

In Section 5.3 is shown that to achieve convergence in total variation in the multivariate case (where there are more than one sequence of sets), the conditions on each sequence are somewhat stronger than in the process case. Finally, in Section 5.4 an example, which further illustrates the role of $t = k/(k-1)$ as threshold values, is given.

5.1. *The univariate case.* Let $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$, be a sequence of sets. To get a bound on the variation distance $d_{\mathrm{TV}}(\mathscr{L}(W_n), \mathrm{Po}(\lambda_n))$, which is given in Theorem 4.1, valid for all shapes of the set C_n we use that the expectation

of W_n, given in (4.3), can be bounded by

$$\lambda_n \le \frac{n^k(2k^2-k)\mu(C_n)^{k-1}}{k!},$$

by Corollary 3.3. This bound, together with $(1-e^{-\lambda_n}) \le \min(1,\lambda_n)$, inserted in the bound in Theorem 4.1 yields

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(W_n),\mathrm{Po}(\lambda_n)) \\ &\quad\le \left\{ \frac{a_k k(n\mu(C_n))^{k-1}}{(k-1)!} + \sum_{l=1}^{k-1} \binom{k}{l} \frac{a_{k-l+1}(n\mu(C_n))^{k-l}}{(k-l)!} \right\} \\ &\qquad \times \min\left\{1, \frac{n^k a_k\, \mu(C_n)^{k-1}}{k!}\right\}, \end{aligned} \tag{5.3}$$

where $a_i = 2i^2 - i$, valid for all $C_n \in \mathscr{K}$.

THEOREM 5.1. *Let W_n and λ_n be defined as in* (4.1) *and* (4.3), *respectively*.

(i) *For any sequence of sets*, $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$,

$$d_{\mathrm{TV}}(\mathscr{L}(W_n),\mathrm{Po}(\lambda_n)) = O\left(\min\left\{1, n^k\mu(C_n)^{k-1}\right\} \sum_{l=1}^{k-1} (n\mu(C_n))^{k-l}\right). \tag{5.4}$$

(ii) *For a sequence of sets* $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$, *with* $\mu(C_n) = O(n^{-t})$, *where* $t > 1$ *is constant*, *the bound tends to zero and is of the order*

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(W_n),\mathrm{Po}(\lambda_n)) \\ &\quad= \begin{cases} O(n^{1-t}), & \text{if } 1 < t < k/(k-1),\ (\lambda_n \to \infty), \\ O(n^{-1/(k-1)}), & \text{if } t = k/(k-1),\ (\lambda_n \text{ stays away from } 0,\infty), \\ O(n^{k(1-t)+1}), & \text{if } t > k/(k-1),\ (\lambda_n \to 0). \end{cases} \end{aligned}$$

PROOF. (i) Follows directly from (5.3).

(ii) Since $t > 1$, it follows that $n\mu(C_n) < 1$ asymptotically. Hence it is the $(k-1)$st term (i.e., $l = k-1$) in the sum in (5.4) which is dominating, and it is $O(n^{1-t})$. Furthermore

$$\min\left\{1, n^k\mu(C_n)^{k-1}\right\} = \begin{cases} O(1), & \text{if } 1 < t \le k/(k-1), \\ O(n^{k(1-t)+t}), & \text{if } t > k/(k-1). \end{cases}$$

Combining these facts yields the result. □

REMARK 5.2. Note that it is the $(k-1)$st term of the sum which determines the rates in the second part of the theorem. This term concerns the dependence between two indicators connected to k-subsets with $k-1$ common particles.

COROLLARY 5.3. *(i) Let* $\{C_n\}_{n=1}^{\infty}$ *be a sequence of sets such that* $C_n \in \mathscr{K}$ *and* $\lambda_n \to \lambda$ *as* $n \to \infty$, *where* $0 < \lambda < \infty$. *Then*

$$\mathscr{L}(W_n) \to_{\mathscr{D}} \mathrm{Po}(\lambda) \quad \text{as } n \to \infty.$$

(ii) *For instance, if* $\mu(C_n) = cn^{-k/(k-1)}$, $c > 0$ *and* C_n *is of equal shape for all* n, *then*

$$\lambda = a_{C,k} c^{k-1}/k!,$$

where $a_{C,k} = k + k(k-1)\nu(C_n, \check{C}_n)/\mu(C_n) \in [k^2, 2k^2 - k]$ *does not depend on* n, *and*

$$\mathscr{L}(W_n) \to_{\mathscr{D}} \mathrm{Po}(\lambda) \quad \text{as } n \to \infty,$$

at the rate $O(n^{-1/(k-1)})$.

PROOF. (i) Since

$$\lambda_n = \binom{n}{k}\left(k + k(k-1)\nu\left(C_n, \check{C}_n\right)/\mu(C_n)\right)\mu(C_n)^{k-1} \to \lambda \quad \text{as } n \to \infty,$$

and $1 \le \nu(C_n, \check{C}_n)/\mu(C_n) \le 2$, for all n, by (2.3), there exists $0 \le c < \infty$ such that

$$\mu(C_n) \le cn^{-k/(k-1)}.$$

By Theorem 5.1(ii), the total variation distance between $\mathscr{L}(W_n)$ and $\mathrm{Po}(\lambda_n)$ tends to zero at the rate $O(n^{-1/(k-1)})$ in this case. This combined with the fact that

$$d_{\mathrm{TV}}(\mathrm{Po}(\lambda_n), \mathrm{Po}(\lambda)) \le |\lambda_n - \lambda| \to 0$$

if $\lambda_n \to \lambda$, implies convergence in total variation distance, and hence in distribution, by the triangle inequality.

(ii) In this case,

$$\lambda_n = \binom{n}{k} a_{C,k}\, \mu(C_n)^{k-1} = \lambda\binom{n}{k} n^{-k} k!.$$

Hence

$$|\lambda_n - \lambda| = \lambda\left|\frac{n(n-1)\cdots(n-k+1) - n^k}{n^k}\right| = O(n^{-1}),$$

and the total variation distance is determined by $d_{\mathrm{TV}}(\mathscr{L}(W_n), \mathrm{Po}(\lambda_n))$. □

5.2. *The process case.* Let $\{\Xi_n\}_{n=1}^{\infty}$ and $\{\Lambda_n\}_{n=1}^{\infty}$ be the sequences of processes and measures, respectively, introduced in Section 4.1.2, which correspond to the sequence of sets $\{C_n\}_{n=1}^{\infty}$. The bound on $d_{\mathrm{TV}}(\mathscr{L}(\Xi_n), \mathrm{Po}(\Lambda_n))$ in Theorem 4.3 equals the bound on $d_{\mathrm{TV}}(\mathscr{L}(W_n), \mathrm{Po}(\lambda_n))$ in Theorem 4.1, except that it lacks the factor $\min(1, 1/\lambda_n)$, and is hence not as good when $\lambda_n > 1$. To obtain convergence of $d_{\mathrm{TV}}(\mathscr{L}(\Xi_n), \mathrm{Po}(\Lambda_n))$ to zero, we can therefore not allow the areas of the sets to decrease as slowly as in the previous case of sequences of variables. The order terms in Theorem 5.1 valid for $\lambda_n \le 1$, that

is, when $\min\{1, n^k\mu(C_n)^{k-1}\} = n^k\mu(C_n)^{k-1}$ in (i), and when $t \geq k/(k-1)$ in (ii), will be valid for all λ_n here, and the counterpart of Theorem 5.1 then reads as follows.

THEOREM 5.4. *Let Ξ_n and Λ_n be defined as in Theorem* 4.3. *Then, for any sequence of sets $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$,*

$$d_{\mathrm{TV}}\big(\mathscr{L}(\Xi_n), \mathrm{Po}(\Lambda_n)\big) = O\left(\sum_{l=1}^{k-1} n^{2k-l}\mu(C_n)^{2k-l-1}\right). \tag{5.5}$$

For a sequence of sets $\{C_n\}_{n=1}^{\infty}$, $C_n \in \mathscr{K}$, with $\mu(C_n) = O(n^{-t})$, where $t > (k+1)/k$ is constant, this bound tends to zero and is of the order

$$d_{\mathrm{TV}}\big(\mathscr{L}(\Xi_n), \mathrm{Po}(\Lambda_n)\big) = O\big(n^{k(1-t)+1}\big).$$

Note that the condition for convergence in the univariate case, $\mu(C_n) = O(n^{-t})$, $t > 1$ does not depend on k. Here the condition is $t > (k+1)/k$; the smaller k, the faster must the areas decrease.

5.3. *The multivariate case.* By studying the bound on the total variation distance in the multivariate case, given in Theorem 4.6, we find that it is $\binom{n}{k_j}\beta_{2j}$ which has to be examined to achieve conditions for convergence. Note that it is obvious that we need at least that

$$n\,\mu\big(C_n^{(k_j)}\big) \to 0, \tag{5.6}$$

which was the condition for convergence in the univariate case [Theorem 5.1(ii)]. Hence it is enough to consider the terms corresponding to $l = k_j - 1$ and $l = \min\{k_j, k_s\}$ in β_{2j}, given in (4.14), since they are dominating if (5.6) is fulfilled. Collect the n-dependent parts of these terms, to get the following conditions on $\mu(C_n^{(k_j)})$ for convergence of $d_{\mathrm{TV}}(\mathscr{L}\{\Xi_n^{(k_j)}\}_{j=1}^m), \Pi_{j=1}^m \mathrm{Po}(\Lambda_n^{(k_j)}))$ to zero:

$$n^{k_j+1}\mu\big(C_n^{(k_j)}\big)^{k_j} \to 0 \tag{5.7}$$

and

$$\begin{aligned} n^{k_s+k_j-\min\{k_j,k_s\}}\min\Big\{ &\mu\big(C_n^{(k_s)}\big)^{k_s-1}\mu\big(C_n^{(k_j)}\big)^{k_j-\min\{k_j,k_s\}}, \\ &\mu\big(C_n^{(k_j)}\big)^{k_j-1}\mu\big(C_n^{(k_s)}\big)^{k_s-\min\{k_j,k_s\}}\Big\} \to 0 \end{aligned} \tag{5.8}$$

as $n \to \infty$, for all $j, s = 1, \ldots, m$, $j \neq s$.

In the previous case, concerning one point process, it was the $(k_j - 1)$st term in the sum in (5.5) which decided at which rate the areas of the sets in the sequence $\{C_n^{(k_j)}\}$ must decrease for the distance to tend to zero. The important part of this term originates from the expectation of the product of two indicators connected to k_j-subsets with $k_j - 1$ common particles [i.e., $E[I_{k_j,1}I_{k_j}^{k_j-1}]$ in (4.16), using the notation in the proof of Theorem 4.5]. This

expectation is of importance also in this multivariate case and turns up as condition (5.7). The other condition, (5.8), comes from the expectation of products of indicators pertaining to k_j-subsets and k_s-subsets with $\min\{k_j, k_s\}$ particles in common ($E[I_{k_s,1} I_{k_j}^{\min\{k_j,k_s\}}]$).

If $k_j < k_s$, condition (5.8) reads

$$\min\Big\{n^{k_s}\mu\big(C_n^{(k_s)}\big)^{k_s-1}, n^{k_s}\mu\big(C_n^{(k_j)}\big)^{k_j-1}\mu\big(C_n^{(k_s)}\big)^{k_s-k_j}\Big\} \to 0 \quad \text{as } n \to \infty,$$

and we can state the following theorem.

THEOREM 5.5. *If* $\{C_n^{(k_j)}\}_{n=1}^{\infty}$, $j = 1, \dots, m$, $k_j \in \{2, 3, \dots\}$, *are sequences of sets such that*

$$n^{k_j+1}\mu\big(C_n^{(k_j)}\big)^{k_j} \to 0 \quad \text{as } n \to \infty, \tag{5.9}$$

for all $j = 1, \dots, m$, *and*

$$\min\Big\{n^{k_s}\mu\big(C_n^{(k_s)}\big)^{k_s-1}, n^{k_s}\mu\big(C_n^{(k_j)}\big)^{k_j-1}\mu\big(C_n^{(k_s)}\big)^{k_s-k_j}\Big\} \to 0 \quad \text{as } n \to \infty, \tag{5.10}$$

for all $j, s = 1, \dots, m$, *such that* $k_j < k_s$, *then*

$$d_{\mathrm{TV}}\left(\mathscr{L}\Big(\big\{W_n^{(k_j)}\big\}_{j=1}^m\Big), \prod_{j=1}^m \mathrm{Po}\big(\lambda_n^{(k_j)}\big)\right) \le d_{\mathrm{TV}}\left(\mathscr{L}\Big(\big\{\Xi_n^{(k_j)}\big\}_{j=1}^m\Big), \prod_{j=1}^m \mathrm{Po}\big(\Lambda_n^{(k_j)}\big)\right) \to 0$$

as $n \to \infty$.

As expected, for sequences $\{C_n^{(k_j)}\}$ with $\mu(C_n^{(k_j)}) = O(n^{-t})$, it is now not enough that $t > (k_j + 1)/k_j$, $j = 1, \dots, m$, which was the critical rate for convergence in total variation norm in case of one process; (5.9) is fulfilled, but (5.10) is not. An example which satisfies both conditions is

$$\mu\big(C_n^{(k_j)}\big) = O\big(n^{-(k_j+1)/k_j - 1/(k_j^3-1)}\big),$$

$j = 1, \dots, m$, while

$$\mu\big(C_n^{(k_j)}\big) = O\big(n^{-(k_j+1)/k_j - 1/k_j^3}\big),$$

$j = 1, \dots, m$, do not decrease fast enough.

If the expected numbers of covered k_j-subsets, $W_n^{(k_j)}$, $j = 1, \dots, m$, converge to finite values, then joint convergence in distribution follows immediately.

COROLLARY 5.6. *If* $E[W_n^{(k_j)}] \to \lambda^{(k_j)}$, $0 \le \lambda^{(k_j)} < \infty$, $j = 1, \dots, m$, *and* $\mathscr{L}(Z^{(k_j)}) = \mathrm{Po}(\lambda^{(k_j)})$ *are independent, then*

$$\big(W_n^{(k_1)}, \dots, W_n^{(k_m)}\big) \to_{\mathscr{D}} \big(Z^{(k_1)}, \dots, Z^{(k_m)}\big) \quad \text{as } n \to \infty.$$

Furthermore, let $\Pi^{(k_j)}$, $j = 1, \dots, m$, *be independent Poisson processes on* $\mathbf{A} \times [0, 1]$ *with intensity measure* $d\Lambda^{(k_j)}(x, y) = 2\lambda^{(k_j)}(k_j - 1)y^{2k_j-3}\,dx\,dy$. *Then*

$$\big(\Xi_n^{(k_1)}, \dots, \Xi_n^{(k_m)}\big) \to_{\mathscr{D}} \big(\Pi^{(k_1)}, \dots, \Pi^{(k_m)}\big).$$

For the proof, note that $\mu(C_n^{(k_j)}) = O(n^{-k_j/(k_j-1)})$ and hence the conditions of Theorem 5.5 are fulfilled. The result is then proved in the same manner as Corollary 5.3(i).

5.4. The maximal number of covered particles. This section is concluded by an application of Theorem 5.1. Let M_{C_n} be the maximal number of n independent and uniformly distributed particles on $\mathbf{A}$ which are covered by $C_n + x$ for some $x \in \mathbf{A}$, that is, M_{C_n} is the so-called scan statistic, which was mentioned in the introduction. We will study the distribution of M_{C_n} for different sequences of sets and how its asymptotic distribution depends on the chosen sequence. The sequences $\{C_n\}_{n=1}^{\infty}$ are such that $\mu(C_n) = cn^{-t}$, where $c > 0$ and $t \in \mathbb{R}$. Recall from the beginning of this section how the expected number of covered k-subsets asymptotically depends on t. The theorem below further displays the threshold behaviour at the values $t = k/(k-1)$, $k \in \{2, 3, \ldots, \}$: it implies for instance that:

1. If $t > 2$, then $P(M_{C_n} = 1) \to 1$.
2. If $t = 2$, then $P(M_{C_n} = 1 \text{ or } 2) \to 1$.
3. If $3/2 < t < 2$, then $P(M_{C_n} = 2) \to 1$.
4. If $t = 3/2$, then $P(M_{C_n} = 2 \text{ or } 3) \to 1$,

and so on.

THEOREM 5.7. *Suppose that* $\{C_n\}_1^{\infty}$, $C_n \in \mathscr{K}$, *is a sequence of sets with* $\mu(C_n) = cn^{-t}$, $c > 0$.

(i) *If* $t = k/(k-1)$, $k \in \{2, 3, \ldots\}$, *and* $\lambda_n \to \lambda$ *as* $n \to \infty$, *then*

$$P\big(M_{C_n} = i\big) \to \begin{cases} e^{-\lambda}, & i = k-1, \\ 1 - e^{-\lambda}, & i = k, \\ 0, & i \neq k-1, k \end{cases} \qquad \text{as } n \to \infty.$$

(ii) *If* $1 < t < 2$, $t \neq k/(k-1)$ *for all* $k \geq 2$, *then*

$$P\big(M_{C_n} = k\big) \to 1 \quad \text{as } n \to \infty,$$

where k *satisfies* $(k+1)/k < t < k/(k-1)$.

(iii) *If* $t > 2$, *then* $P(M_{C_n} = 1) \to 1$ *as* $n \to \infty$.

(iv) *If* $t \leq 1$ *then* $P(M_{C_n} < k) \to 0$ *as* $n \to \infty$, *for all* k.

REMARK 5.8. If all C_n are of the same shape and $t = k/(k-1)$, then $\lambda_n \to \lambda$, since for some $1 \leq a \leq 2$, $\nu(\check{C}_n, C_n)/\mu(C_n) = a$ for all n.

PROOF. (i) First we let $t = k/(k-1)$. It then follows from Corollary 5.3 that

$$P\big(M_{C_n} < k\big) = P(W_n = 0) \to \exp\{-\lambda\} \tag{5.11}$$

as $n \to \infty$.

Let $W_{n,k-1}$ and $W_{n,k+1}$ denote the number of covered $(k-1)$- and $(k+1)$-subsets, respectively, and let $\lambda_{n,k-1}$ and $\lambda_{n,k+1}$ be their expectations. If

$k = 2$, then $\lambda_{n,k-1} \to \infty$ since every single particle is of course covered by some $C_n + x$. Furthermore, since $t < (k-1)/(k-2)$, $k > 2$, and $t > (k+1)/k$ we get by (5.2) that

$$\lambda_{n,k-1} \to \infty \quad \text{and} \quad \lambda_{n,k+1} \to 0 \quad \text{as } n \to \infty. \tag{5.12}$$

Hence the probability that M_{C_n} is strictly less than $k-1$ approaches 0 since $\mu(C_n) = cn^{-k/(k-1)}$, so by Theorem 5.1(ii),

$$\begin{aligned} P\big(M_{C_n} < k-1\big) &= P(W_{N,k-1} = 0) \\ &= \exp\{-\lambda_{n,k-1}\} + O(n^{-1/(k-1)}) \to 0 \quad \text{as } n \to \infty, \end{aligned}$$

which together with (5.11) yields

$$P\big(M_{C_n} = k-1\big) \to \exp\{-\lambda\}.$$

Next we consider the probability that M_{C_n} is greater than k:

$$\begin{aligned} P\big(M_{C_n} \geq k+1\big) &= P(W_{n,k=1} \geq 1) \\ &\leq E[W_{n,k+1}] \to 0 \quad \text{as } n \to \infty, \end{aligned}$$

by Markov's inequality and (5.12). The only possibility that remains for M_{C_n} in the limit is to be equal to k,

$$P\big(M_{C_n} = k\big) = 1 - P\big(M_{C_n} < k\big) - P\big(M_{C_n} > k\big) \to 1 - \exp\{-\lambda\}$$

as $n \to \infty$. Hence the theorem is proved for $t = k/(k-1)$.

(ii), (iii) The cases where $1 < t < 2$, $t \neq k/(k-1)$, and $t > 2$ follow by arguments similar to those above.

(iv) In the case where $t \leq 1$ we cannot use the Poisson approximation directly, since the sets are decreasing too slowly for the total variation distance to tend to zero. Instead we use sequences of slightly smaller sets. For fixed k, let $\{C'_n\}$ consist of sets such that $C'_n \subset C_n$ and $\mu(C'_n) = cn^{-(k+1)/k}$. Then $W'_{n,k}$ = the number of k-subsets covered by some translate of C'_n is approximately Poisson distributed with an expectation which tends to ∞, and $d_{\mathrm{TV}}(\mathscr{L}(W'_{n,k}), \mathrm{Po}(E[W'_{n,k}])) \to 0$ by Theorem 5.1(ii), so that

$$P\big(M_{C_n} < k\big) \leq P\big(M_{C'_n} < k\big) = P(W'_{n,k} = 0) \to 0 \quad \text{as } n \to \infty. \ \square$$

6. Possible extensions, improvements and applications.

6.1. *Extensions.* A natural extension of this paper is to generalize the results to higher dimensions. All the approximations make use of the probability of covering a number of independently and uniformly distributed particles with some translate of a convex set. In three dimensions this probability can be found in Månsson (1996) and, as shown in Section 3, the probability in an arbitrary dimension can be obtained directly from results in Weil (1990). It should be straightforward, but tedious, to extend the approximation results to an arbitrary dimension.

The starting point can be changed in various directions. An obvious variation is to let the particles on **A** constitute a Poisson process rather than being a fixed number. To derive results, corresponding to those of the present paper, in that case should be easy, and it would be surprising if there would be any difference asymptotically. Variations, more difficult to handle, are for

instance to let the particles follow some distribution other than the uniform or to let the sets be non-convex.

6.2. *Compound Poisson approximation.* The indicators pertaining to k-subsets with common particles are not independent, but have a positive dependence, and the k-subsets which are covered tend to occur in clumps. The more common particles and the larger sets, the stronger is the dependence. Then it seems natural to approximate the process of the positions of the k-aggregates by some process in which clumps are more likely to occur than in the usual Poisson process and to approximate the number of k-aggregates by some distribution other than Poisson. Natural candidates are the compound Poisson process and the compound Poisson distribution. It seems possible that such approximation can handle larger sets and improve the error bounds, especially in the case of large k.

6.3. *Statistical applications.* As mentioned in the introduction, a common test statistic when testing whether a point pattern originates from a Poisson process is the number of pairs of points closer than some distance r. This number equals W in our terminology, in the case where $k = 2$ and C is a disc with diameter r. An idea is to study if the number of k-subsets, $k = 2, 3, \ldots$, which are covered by a set C, which is not necessarily a disc, would be reasonable as a statistic when testing the randomness of a sample. The crucial element in this matter is the rate of convergence.

Acknowledgments. For all the support and guidance during work on my thesis, on which this paper is based, I am deeply grateful to Olle Nerman. I also thank Andrew Barbour for advice and fruitful discussions and a referee for helpful comments. Finally, I would like to acknowledge the warm hospitality of the Statistics Section at Monash University and of the Abteilung für Angewandte Mathematik, Universität Zürich, where parts of this work were undertaken.

REFERENCES

ALDOUS, D. (1989). *Probability Approximation Via the Poisson Clumping Heuristic*. Springer, New York.

ALM, S. E. (1983). On the distribution of the scan statistic of a Poisson process. In *Probability and Mathematical Statistics. Essays in Honour of Carl-Gustav Esseen* 1–10. (A. Gut and L. Holst, eds.), Dept. Mathematics, Uppsala Univ.

ALM, S. E. (1997). On the distribution of scan statistics of a two-dimensional Poisson process. *Adv. in Appl. Probab.* **29** 1–18.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Probab.* **17** 9–25.

BARBOUR, A. D. and EAGLESON, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials. *Adv. in Appl. Probab.* **15** 585–600.

BARBOUR, A. D. and EAGLESON, G. K. (1984). Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B* **46** 397–402.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

BERWALD, W. and VARGA, O. (1937). Integralgeometrie 24, über die Schiebungen im Raum. *Math. Z.* **42** 710–736.

BLASCHKE, W. (1937). Integralgeometrie 21, über Schiebungen. *Math. Z.* **42** 399–410.
BONNESEN, T. and FENCHEL, W. (1948). *Theorie der Konvexen Körper*. Chelsea, New York.
CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.
EGGLETON, P. and KERMACK, W. O. (1944). A problem in the random distribution of particles, *Proc. Roy. Soc. Edinburgh Sec. A* **62** 103–115.
GLAZ, J. (1989). Approximation and bounds for the distribution of the scan statistic. *J. Amer. Statist. Assoc.* **84** 560–566.
JANSON, S. (1984). Bounds on the distributions of extremal values of a scanning process. *Stochastic Proces. Appl.* **18** 313–328.
KRYSCIO, R. J. and SAUNDERS, R. (1983). On interpoint distances for planar Poisson cluster processes, *J. Appl. Probab.* **20** 513–528.
LOADER, C. R. (1991). Large-deviation approximations to the distribution of scan statistics. *Adv. Appl. Probab.* **23** 751–771.
MACK, C. (1948). An exact formula for $Q_k(n)$, the probable number of k-aggregates in a random distribution of n points. *Philos. Mag.* **39** 778–790.
MACK, C. (1949). The expected number of aggregates in a random distribution of n points. *Proc. Cambridge Philos. Soc.* **46** 285–292.
MÅNSSON, M. (1996). On clustering of random points in the plane and in space. Thesis, Dept. Mathematics, Chalmers Univ. Technology.
MILES, R. E. (1974). The fundamental formula of Blaschke in integral geometry and geometrical probability and its iteration, for domains with fixed orientations. *Austral. J. Statist.* **16** 111–118.
MCGINLEY, W. G. and SIBSON, R. (1975). Dissociated random variables. *Math. Proc. Cambridge Philos. Soc.* **77** 185–188.
NAUS, J. I. (1982). Approximations for distributions of scan statistics. *J. Amer. Statist. Assoc.* **77** 177–183.
SILBERSTEIN, L. (1945). The probable number of aggregates in random distributions of points. *Philos. Mag.* **36** 319–336.
SILVERMAN, B. and BROWN, T. (1978). Short distances, flat triangles and Poisson limits. *J. Appl. Probab.* **15** 815–825.
SILVERMAN, B. and BROWN, T. (1979). Rates of Poisson convergence for U-statistics. *J. Appl. Probab.* **16** 428–432.
WEIL, W. (1990). Iterations of translative formulae and non-isotropic Poisson processes of particles. *Math. Z.* **205** 531–549.

DEPARTMENT OF MATHEMATICS
CHALMERS UNIVERSITY OF TECHNOLOGY
S-412 96 GÖTEBORG
SWEDEN
E-MAIL: marianne@math.chalmers.se