Compound Poisson approximation and the clustering of random points

A. D. Barbour[*] and Marianne Månsson[†]

October 26, 1998

Abstract. Let n random points be uniformly and independently distributed in the unit square, and count the number W of subsets of k of the points which are covered by some translate of a small square C. If $n|C|$ is small, the number of such clusters is approximately Poisson distributed, but the quality of the approximation is poor. In this paper, we show that the distribution of W can be much more closely approximated by an appropriate compound Poisson distribution $\mathrm{CP}\,(\lambda_1, \lambda_2, \ldots)$. The argument is based on Stein's method, and is far from routine, largely because the approximating distribution does not satisfy the simplifying condition that $i\lambda_i$ be decreasing.

AMS 1991 Classification: 62E17, 60F05, 60G55.
Key words and phrases: Compound Poisson approximation, Stein's method.

1 Introduction

Assume that n points, denoted by $\xi_1, \ldots, \xi_n$, are uniformly and independently distributed in the unit square $\mathbf{A}$ in $\mathbb{R}^2$, and let $C \subset \mathbf{A}$ be a small square of side c. A subset consisting of k points, where $1 < k < n$, will be called a k–subset. There are $\binom{n}{k}$ different k–subsets of points, some of which are covered by translates $C^a := C + a$ of C. Number the k–subsets in some way, independently of the positions of the points, and let W denote the number of k–subsets which actually are covered by some C^a. Then W can be written as

$$W = \sum_{i=1}^{\binom{n}{k}} I_i, \tag{1.1}$$

where

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

To avoid problems with the boundaries of $\mathbf{A}$, the torus convention is used throughout the paper. Furthermore, for convenience, we let C have its lower left corner at the

origin; thus C^a has its lower left corner at a. We shall study the distribution of W when k is fixed, n is large and the density of points is not too great, in the sense that $P(C \ni \text{some } k \text{ subset of } \xi_1, \ldots, \xi_n)$ is small. For asymptotics, we therefore always assume that $n|C_n| \to 0$, where $|\cdot|$ denotes area.

In Månsson (1997), Poisson approximation of the quantity W, in some contexts called a multiple scan statistic, is studied by means of the Stein-Chen method. Some of the results in that paper are presented in Section 3.1. As can be seen in Theorem 3.2, the total variation distance tends to zero whenever $\lim_{n\to\infty} n|C_n| = 0$. However, the rate of convergence is typically of order $n|C_n|$, which is slower than might be hoped, if $k \geq 3$. To get an idea of what an ideal convergence rate might be, let $J_i = I[\xi_i$ is the leftmost member of a k–cluster], $1 \leq i \leq n$; then $p = E[J_i] \asymp (n|C_n|)^{k-1}$. If the J_i were *independent,* then approximation to $U = \sum_{i=1}^{n} J_i$ by Po $(E[U])$ would have accuracy of order $(n|C_n|)^{k-1}$ if $E[U] \geq 1$. Thus we shall be satisfied if we can achieve this order of accuracy for our sum W of dependent random indicators; clearly, for $k \geq 3$, a rate of order $n|C_n|$ is markedly inferior.

That it is not only the bounds for the Poisson approximation which are crude, but in fact the approximations that are bad, can be seen in Månsson (1999), where simulations are carried out. The inaccuracy stems from the strong dependence between k–subsets with many common points: if two k–subsets have $k-1$ points in common, then given that one of these k–subsets is covered, the probability is high that also the other one is covered. To be a little more precise: as we will see in (3.1), $P(I_i = 1) = k^2|C|^{k-1}$, while if the ith and jth k–subsets have $k-1$ points in common, then $P(I_j = 1 \mid I_i = 1) \geq |C|$. This latter probability is critical in determining the order $n|C_n|$ of the bound on the total variation distance given in Theorem 3.2, when $E[W] \geq 1$. Hence the troublesome part in the Poisson approximation arises because the covered k–subsets tend to occur in clumps. A natural step is thus to approximate W by some distribution in which clumps can be taken into account, and an obvious candidate is a compound Poisson distribution. The aim of this paper is to investigate such approximations.

Some early references concerning the distribution of W are Eggleton and Kormack (1944), Silberstein (1945) and Mack (1948,1949). The main focus of these articles is the expectation of W for different shapes of C, but Mack (1948) also argues for a Poisson limit of W as the total number of points tends to infinity. A more recent reference of interest is Aldous (1989), who handles the case where C is a disc or a square. In the special case where $k = 2$ and C is circular, the number of k–subsets that are covered equals the number of pairs of points with interpoint distance less than the diameter of C. Convergence of this number to a Poisson limit is discussed in Silverman and Brown (1978, 1979).

In Silverman and Brown (1978), the number of close pairs is given as an example of a classical U-statistic. Poisson approximation for U-statistics and sums of dissociated variables 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 to the sums.

In terms of the scan statistic, the maximal number of points covered by C, the problem is investigated by Loader (1991) when C is rectangular. In Alm (1997), C can be any convex set, and the suggested approximations are verified by means of simulations.

Another problem related to ours, but one-dimensional, is that of m-spacings, defined as follows. Take n points from a uniform distribution on the unit interval, and let

$X_{(1)}, \ldots, X_{(n)}$ denote the ordered sample. Then the m-spacings are defined by $S_{im} = X_{(m+i)} - X_{(i)}$, $i = 1, \ldots, n-m$, and, using the torus convention, $S_{im} = 1 - X_{(i)} + X_{(m+i-n)}$, $i = n-m+1, \ldots, n$. Barbour et al. (1992) showed that the total variation distance between the number of m-spacings smaller than a, say, $W = \sum_{i=1}^{n} 1\{S_{im} < a\}$ and $\mathcal{L}(\mathrm{Po}(E[W]))$ tends to zero at the rate na whenever $na \to 0$ and $E[W] \geq 1$. In Roos (1993) it is shown that a suitable chosen compound Poisson approximation for W yields rates of order $O(\pi(1 + \log^+(2n\pi)))$, where $\pi = O((na)^m/m!)$. Note that $S_{im} < a$ means that $m+1$ points lie in an interval of length a; hence, except for the logarithmic term, the order Roos achieved corresponds to the order we aim at in our problem.

2 Compound Poisson approximation by Stein's method

Let $W = \sum_{\alpha \in \Gamma} X_\alpha$, where Γ is a finite family of indices, and the X_α take values in $\mathbb{N}$; in our case, the X_α's are indicator variables. Here, we consider the compound Poisson approximation of W by means of Stein's method. The relevant compound Poisson distributions are defined by

$$\mathrm{CP}\,(\Lambda) = \mathcal{L}\left(\sum_{i \geq 1} i N_i\right), \qquad (2.1)$$

where $(N_i, \; i \geq 1)$ are independent with $\mathcal{L}(N_i) = \mathrm{Po}\,(\lambda_i)$, and $\Lambda = \sum_{i \geq 1} \delta_i \lambda_i$, with δ_i the unit mass at the point i. Note that this definition is equivalent to

$$\mathrm{CP}\,(\Lambda) = \mathcal{L}\left(\sum_{i=1}^{N} V_i\right), \qquad (2.2)$$

where $\mathcal{L}(N) = \mathrm{Po}\,(\sum_{i \geq 1} \lambda_i)$ and V_i, $i = 1, 2, \ldots$, are independent of each other and of N, with $P(V_i = j) = \lambda_j / \sum_{i \geq 1} \lambda_i$.

Barbour, Chen and Loh (1992) showed that there exists a bounded solution $g : \mathbb{N} \to \mathbb{R}$ of

$$\sum_{i \geq 1} i \lambda_i g(j+i) - j g(j) = f(j),$$

for all $j \geq 0$, if and only if $E[f(Z)] = 0$ where $\mathcal{L}(Z) = \mathrm{CP}\,(\Lambda)$. Letting g_A be the solution when $f(j) = 1\{j \in A\} - \mathrm{CP}\,(\Lambda)\{A\}$, it follows that

$$
\begin{aligned}
d_{TV}(\mathcal{L}(W), \mathrm{CP}(\Lambda)) &= \sup_{A \subset \mathbf{Z}^+} |P(W \in A) - \mathrm{CP}\,(\Lambda)\{A\}| \\
&= \sup_{A \subset \mathbf{Z}^+} \left| E\left[\sum_{i \geq 1} i \lambda_i g_A(W+i) - W g_A(W)\right]\right|. \qquad (2.3)
\end{aligned}
$$

Roos (1994a, 1994b) showed that it is often possible to find ε_0 and ε_1 such that

$$\left| E\left[\sum_{i \geq 1} i \lambda_i g(W+i) - W g(W)\right]\right| \leq \varepsilon_0 |g| + \varepsilon_1 |\Delta g|, \qquad (2.4)$$

for all bounded $g : \mathbb{N} \to \mathbb{R}$, where

$$|g| = \sup_{j \geq 1} |g(j)| \quad \text{and} \quad |\Delta g| = \sup_{j \geq 1} |g(j+1) - g(j)|,$$

and the quantities λ_i are appropriately defined; and if (2.4) is satisfied, then

$$d_{TV}(\mathcal{L}(W), \text{CP}(\Lambda)) \quad \leq \quad \varepsilon_0 H_0(\Lambda) + \varepsilon_1 H_1(\Lambda), \tag{2.5}$$

from (2.3), where

$$H_0(\Lambda) = \sup_{A \subset \mathbf{Z}^+} |g_A| \quad \text{and} \quad H_1(\Lambda) = \sup_{A \subset \mathbf{Z}^+} |\Delta g_A|.$$

Note that letting $\lambda_1 = \lambda$ and $\lambda_i = 0$, $i \geq 2$, the above reduces to the Poisson case:

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \quad \leq \quad \varepsilon_0 H_0(\lambda) + \varepsilon_1 H_1(\lambda),$$

in which case there exist good bounds on $H_0(\lambda)$ and $H_1(\lambda)$, given respectively by $\min(1, \lambda^{-1/2})$ and $\lambda^{-1}(1 - e^{-\lambda})$. Bounds on $H_0(\Lambda)$ and $H_1(\Lambda)$ as sharp as these cannot be found in the compound Poisson case for general Λ. In Barbour, Chen and Loh (1992) the following bounds are given:

$$H_0(\Lambda), H_1(\Lambda) \leq \min\left(1, \frac{1}{\lambda_1}\right) \exp\left\{\sum_{i=1}^{\infty} \lambda_i\right\}, \tag{2.6}$$

valid for any $\Lambda = \sum_{i=1}^{\infty} \lambda_i \delta_i$, and the considerably better bounds

$$H_0(\Lambda) \leq \min\left\{1, \frac{1}{\sqrt{\lambda_1 - 2\lambda_2}}\left[2 - \frac{1}{\sqrt{\lambda_1 - 2\lambda_2}}\right]\right\} \tag{2.7}$$

and

$$H_1(\Lambda) \leq \min\left\{1, \frac{1}{\lambda_1 - 2\lambda_2}\left[\frac{1}{4(\lambda_1 - 2\lambda_2)} + \log^+ 2(\lambda_1 - 2\lambda_2)\right]\right\}, \tag{2.8}$$

valid only if $i\lambda_i \searrow 0$. The difficulties in deriving bounds in the compound Poisson approximation case is thoroughly discussed in Barbour (1995).

If $i\lambda_i$ does not decrease, as is the case in the current problem, (2.7) and (2.8) unfortunately cannot be used. Because of the exponential term, (2.6) is unsatisfactory when $\sum_{i \geq 1} \lambda_i$ is large, and in particular it is useless in the regime where it tends to ∞. There is however an alternative way to handle the case when $\sum_{i \geq 1} \lambda_i$ is large. First we introduce the notation $\Omega_n = \sum_{i \geq 1} \lambda_{in}$ and $\mu_{in} = \lambda_{in}/\Omega_n$, with the index n explicit for now; then the following theorem can easily be deduced from Theorems 1.10 and TV and Equations (1.24)–(1.28) of Barbour and Utev (1998).

Theorem 2.1 *If a sequence $(\Lambda_n, \ n \geq 1)$ satisfies the conditions*

(i) $\lim_{n \to \infty} \mu_{in} = \mu_i$ *for each $i \geq 1$;*

(ii) $\sup_{n \geq 1} \sum_{i \geq 1} \mu_{in} r_0^i < \infty$ *for some $r_0 > 1$;*

(iii) $\inf_{n \geq 1} \Omega_n > 2; \quad \inf_{n \geq 1} \mu_{1n} > 0,$

and if, for each n, (2.4) is satisfied with $\varepsilon_0 = \varepsilon_{0n} = 0$ for some non–negative integer valued random variable $W = W_n$, then there exist positive constants $K < \infty$ and $c_2 < 1$ such that, for any x satisfying $c_2 \leq x < 1$ and any n such that $E[W_n] \geq (x - c_2)^{-1}$,

$$d_{TV}(\mathcal{L}(W_n), \text{CP}(\Lambda_n)) \quad \leq \quad K(1-x)^{-1}\left(\Omega_n^{-1}\varepsilon_{1n} + P(W_n \leq \tfrac{1}{2}(1+x)E[W_n])\right). \tag{2.9}$$

4

3 Results

3.1 Poisson approximation

In this subsection we discuss the approximation of W, defined in (1.1), by a Poisson variable with parameter $E[W]$. Hence we need the expectation of W,

$$E[W] \;=\; \binom{n}{k} P(I_1 = 1),$$

where $P(I_1 = 1) = P(\exists a \in \mathbf{A} : \xi_1, \ldots, \xi_k \in C^a)$. Since C is a square, $P(I_1 = 1)$ is easy to derive, because the x–coordinates of the points ξ_i are independent of the y–coordinates, and therefore each dimension can be considered separately. Hence it is enough to find the probability that k points uniformly distributed on a circle (by the torus convention) with circumference 1 are covered by some interval of length c. If $c < \frac{1}{2}$, this is easily seen to be kc^{k-1}, since any one of the points can be the leftmost, and c^{k-1} is the probability that the rest of the points lie in an interval of length c to its right. Thus we get

$$P(I_1 = 1) = (kc^{k-1})^2 = k^2 |C|^{k-1}, \tag{3.1}$$

when using the torus convention; the result would be similar for rectangular sets. Hence

$$E[W] \;=\; \binom{n}{k} k^2 |C|^{k-1}, \tag{3.2}$$

a result true also for rectangles, and easily generalized to $\mathbb{R}^d$ as $P(I_1 = 1) = k^d |C|^{k-1}$.

The probability $P(I_1 = 1)$ for any convex set C can be found in Månsson (1997). It depends on the shape of C, to be more precise on the so-called mixed area of C and its reflection at the origin. However, it is of minor importance here, and we just state bounds for the probability:

$$k^2 |C|^{k-1} \leq P(I_1 = 1) \leq (2k^2 - k)|C|^{k-1},$$

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.

The following two theorems, which are proved in Månsson (1997) by means of the Stein-Chen method, hold for any convex set C which is not too large, by which we mean that $\{a : C \cap C^a \neq \emptyset\} \subset \mathbf{A}$.

Theorem 3.1 *Let W be defined by (1.1). Then*

$$d_{TV}(\mathcal{L}(W), \mathrm{Po}(E[W])) \leq$$
$$\left\{ \frac{E[W]k^2}{n} + \sum_{l=1}^{k-1} \binom{k}{l} \binom{n-k}{k-l} a_{k-l+1} |C|^{k-l} \right\} (1 - e^{-E[W]}),$$

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

The next result concerns the rate of convergence when $\{C_n\}_{n=1}^{\infty}$ is a sequence of sets with areas decreasing with n.

Theorem 3.2 (*i*) *For any sequence of sets* $\{C_n\}_{n \geq 1}$ *such that* $n|C_n| \to 0$,

$$d_{TV}(\mathcal{L}(W_n), \mathrm{Po}(E[W_n])) = O\left(\min\{1, n^k|C_n|^{k-1}\} n|C_n|\right).$$

(*ii*) *If* $|C_n| = O(n^{-t})$, *where* $t > 1$ *is constant, the bound tends to zero and is of order*

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

Extensions of this theorem to associated point processes were also proved in Månsson (1997). As can be seen, the rates obtained above in the case $E[W_n] \geq 1$ are slow for $k \geq 3$, when compared to the ideal rate $(n|C_n|)^{k-1}$.

3.2 Compound Poisson approximation

We now turn to compound Poisson approximation. Recall the definitions of the compound Poisson distribution given in (2.1) and (2.2). In our case, it is natural to think of N as the total number of clumps (where clumps are yet to be defined) of covered k–subsets, V_i as the number of covered k–subsets in the ith clump, and N_j as the number of clumps of size j. In the Poisson case, the choice $E[W]$ of parameter in the approximating distribution was obvious, and easy to derive. Here the situation is different, since we first need to define the clumps, and this definition can be rather arbitrary. The one that we choose is not intuitively clear, but it is convenient for the purpose of bounding the total variation distance. A drawback is that it seems to be difficult to find exact formulae for the parameters; however, determining them numerically is a very much smaller problem than the original distributional approximation.

Let Γ be the family of subsets $\alpha = \{\alpha_1, \ldots, \alpha_k\} \subset \{1, \ldots, n\}$. Then $W = \sum_{\alpha \in \Gamma} I_\alpha$ where $I_\alpha = 1$ if $\xi_{\alpha_1}, \ldots, \xi_{\alpha_k}$ are covered by some C^a, and 0 otherwise. For each $\alpha \in \Gamma$, pick one of the α_i arbitrarily, and denote it by α^*. Defining $\mathbf{R}_i^{(m)}$ to be the square of side mc centred at ξ_i, let

$$Z_\alpha = \sum_{\beta \in \Gamma} I_\beta 1\{\xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)} \text{ for all } \beta_i \in \beta\}.$$

Then Z_α is the number of k–subsets which are covered and have all their points in $\mathbf{R}_{\alpha^*}^{(4)}$, which means that they are in one sense close to the k–subset α. These k–subsets constitute the α–clump.

Our first attempt is to approximate W by a compound Poisson distribution with parameters

$$\lambda_i = \sum_{\alpha \in \Gamma} \frac{1}{i} E[I_\alpha I\{Z_\alpha = i\}] = \frac{1}{i} EW P(Z_1 = i | I_1 = 1), \qquad (3.3)$$

by means of (2.5). It turns out that this approach works well for small $E[W]$. When the expectation is large, we need to use Theorem 2.1 instead, but its Condition (ii) is not satisfied with the λ_i defined in (3.3). In order to define the alternative parameters that we use, another random variable has to be introduced: let S_α be the number of points falling in $\mathbf{R}_{\alpha^*}^{(4)}$, apart from those belonging to the k–subset α. Then we define

$$\lambda_i' = i^{-1} EW P(Z_1 = i | I_1 = 1, S_1 \leq s_0), \qquad (3.4)$$

6

for a suitably chosen s_0 which is independent of n, and we set

$$
\begin{aligned}
\varepsilon_1 \;=\; & \binom{n}{k}\left\{ 16\binom{n-k}{k}k^4|C|^{2k-1} \right.\\
& +k^4|C|^{2(k-1)}\left(\sum_{l=1}^{k-1}\binom{k}{l}\binom{n-k}{k-l}+25\binom{n-k}{k}|C|+1\right)\\
& \left. + 32\,k^4|C|^k\left(\frac{|C|}{1-16|C|}\right)^{k-1}(n-k)\binom{n-k}{k-1}\right\}.
\end{aligned}
\tag{3.5}
$$

Note that $E[W]=\binom{n}{k}E[I_1]$ and that

$$
E[W]\geq\Omega=\sum_{i\geq 1}\lambda_i\geq\binom{n}{k}E[I_1]P[S_1=0]\geq\binom{n}{k}E[I_1](1-16|C|)^n\sim E[W]
\tag{3.6}
$$

if $n|C_n|\to 0$, so that $E[W]$ and Ω are then asymptotically equivalent. We are now in a position to state the main results of the paper. We begin with the theorems which are relevant when $E[W]$ is not too large.

Theorem 3.3 *With the above notation, let $\Lambda=\sum_{i\geq 1}\lambda_i\delta_i$, where λ_i is defined as in (3.3). Then*

$$
d_{TV}(\mathcal{L}(W),\mathrm{CP}(\Lambda))\;\leq\;\left(1\wedge\frac{1}{\lambda_1}\right)\exp\left\{\sum_{i=1}^{\infty}\lambda_i\right\}\varepsilon_1.
\tag{3.7}
$$

If $\{C_n\}$ is a sequence of sets such that $E[W_n]$ stays bounded away from ∞ as $n\to\infty$, then

$$
d_{TV}(\mathcal{L}(W_n),\mathrm{CP}(\Lambda_n))\;=\;O(n^{2k-1}|C_n|^{2k-2}).
\tag{3.8}
$$

Corollary 3.4
(i) If $|C_n|=O(n^{-k/(k-1)})$ then $E[W_n]$ stays bounded away from 0 and ∞, and

$$
d_{TV}(\mathcal{L}(W_n),\mathrm{CP}(\Lambda_n))\;=\;O(n^{-1})\to 0.
$$

(ii) If $|C_n|=O(n^{-t})$, $t>k/(k-1)$, then $E[W_n]\to 0$ and

$$
d_{TV}(\mathcal{L}(W_n),\mathrm{CP}(\Lambda_n))\;=\;O(n^{2k(1-t)+2t-1})\to 0.
$$

The next pair of results are useful when $E[W]$ is large.

Theorem 3.5 *Assume that $n|C_n|\to 0$ and $E[W_n]=\binom{n}{k}k^2|C_n|^{k-1}\to\infty$ as $n\to\infty$.*
(i) If $\Lambda'_n=\sum_{i\geq 1}\lambda'_{in}\delta_i$, where λ'_{in} is defined as in (3.4), then there exists a constant $c'_2\in(0,1)$ such that for $x'=(1+c'_2)/2$, $y'=(1+x')/2$ and any $z\in(y'^{1/k},1)$

$$
d_{TV}(\mathcal{L}(W_n),\mathrm{CP}(\Lambda'_n))\;=\;O((n|C_n|)^{k-1}+\exp\{-E[W_n]D_2(y',z)\}),
$$

where D_2 is defined in (4.31) below.

7

(ii) If $\Lambda_n = \sum_{i \geq 1} \lambda_{in} \delta_i$, where λ_{in} is defined as in (3.3), then there exist constants $c_2, c_2' \in (0, 1)$ such that for y' and z defined as in (i), and $x = (1 + c_2)/2$

$$d_{TV}(\mathcal{L}(W_n), \mathrm{CP}(\Lambda_n)) = O\left((n|C_n|)^{k-1} + \exp\left\{-E[W_n]\left(D_2(y', z) + \frac{(1-x)^2 k!}{2(2k)^k}\right)\right\}\right).$$

Corollary 3.6 *If $|C_n| = O(n^{-t})$, $1 < t < k/(k-1)$, then $E[W_n] \to \infty$ and*

$$\begin{aligned}
d_{TV}(\mathcal{L}(W_n), \mathrm{CP}(\Lambda_n)) &= d_{TV}(\mathcal{L}(W_n), \mathrm{CP}(\Lambda_n')) \\
&= O((n|C_n|)^{k-1}) = O(n^{(1-t)(k-1)}) \to 0.
\end{aligned}$$

In all cases, we obtain the 'ideal' rate of $(n|C_n|)^{k-1} \min\{1, E[W_n]\}$, unless, in Theorem 3.5, the term which decays exponentially with $E[W_n]$ is actually the larger, which is only the case when $E[W_n]$ grows extremely slowly.

4 Proofs

The proofs of Theorem 3.3 and Theorem 3.5 are based on the following two lemmas. The first, Lemma 4.1, is the main ingredient in establishing (2.4) in the two theorems; the second is used to bound $P(W < \frac{1}{2}(1 + x)E[W])$ in the application of Theorem 2.1.

Lemma 4.1 *For any bounded function $g : \mathbb{N} \to \mathbb{R}$,*

$$\left| E[W g(W) - \sum_{i \geq 1} i \lambda_i g(W + i)] \right| \leq |\Delta g| \varepsilon_1,$$

where ε_1 is given in (3.5).

Proof. To use the results in Roos (1994a,1994b) to derive bounds on ε_0 and ε_1 in (2.4), it is required that for each $\alpha \in \Gamma$ the indices can be divided into four disjoint subsets, $\{\alpha\}, \Gamma_\alpha^{vs}, \Gamma_\alpha^b, \Gamma_\alpha^{vw}$, in such a way that I_β, $\beta \in \Gamma_\alpha^{vs}$, is in some sense strongly related to I_α, and that, I_β, $\beta \in \Gamma_\alpha^{vw}$, is weakly related to I_γ, $\gamma \in \{\alpha\} \cup \Gamma_\alpha^{vs}$. In our case there is no obvious way to make such a division. There is a natural way to divide the indices into three groups; $\{\alpha\}$, $\{\beta \in \Gamma \setminus \{\alpha\} : \beta \cap \alpha \neq \emptyset\}$ and $\{\beta \in \Gamma \setminus \{\alpha\} : \beta \cap \alpha = \emptyset\}$, which is used in Månsson (1997) to bound the error when approximating $\mathcal{L}(W)$ with a Poisson distribution; see Theorem 3.1. A finer division can however be achieved by taking the positions of the points into account, but this cannot be done in advance. Therefore the results of Roos cannot be used directly, but much of the reasoning below is in the same spirit as in her articles.

Recall from Section 3.2 that α^* is an arbitrarily chosen number in the k–subset α, and that $\mathbf{R}_{\alpha^*}^{(4)}$ is the square centred at ξ_{α^*} with side length $4c$, where c is the side length of C. The reason for this choice of $\mathbf{R}_{\alpha^*}^{(4)}$ will become clear later. We use the following simple equality

$$\begin{aligned}
I_\beta &= I_\beta 1\{\forall \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}\} + I_\beta 1\{\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\} \\
&\quad + I_\beta 1\{\forall \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\},
\end{aligned}$$

8

and let

$$U_\alpha = \sum_{\beta \in \Gamma \setminus \{\alpha\}} I_\beta 1\{\forall \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}\}, \tag{4.1}$$

$$X_\alpha = \sum_{\beta \in \Gamma \setminus \{\alpha\}} I_\beta 1\{\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\}$$
$$+ \sum_{\beta \in \Gamma \setminus \{\alpha\}, \alpha \cap \beta \neq \emptyset} I_\beta 1\{\forall \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\}, \tag{4.2}$$

$$Y_\alpha = \sum_{\beta \in \Gamma \setminus \{\alpha\} : \alpha \cap \beta = \emptyset} I_\beta 1\{\forall \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\}.$$

and

$$Z_\alpha = I_\alpha + U_\alpha.$$

Then Z_α is the number of covered k–subsets such that all points lie in $\mathbf{R}_{\alpha^*}^{(4)}$, our α-clump. Y_α is the number of covered k–subsets 'far' from the α-clump; all their points lie in $\mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}$ and no point is common with the k–subset α. X_α handles the 'boundary' of $\mathbf{R}_{\alpha^*}^{(4)}$, and k–subsets outside $\mathbf{R}_{\alpha^*}^{(4)}$ with points in common with α. With these definitions, W can be split as

$$W = I_\alpha + U_\alpha + X_\alpha + Y_\alpha$$

for each $\alpha \in \Gamma$.

In order to bound $|E[Wg(W) - \sum_{i=1}^{\binom{n}{k}} i\lambda_i g(W+i)]|$, we make the following decomposition:

$$\left| E[Wg(W) - \sum_{i=1}^{\binom{n}{k}} i\lambda_i g(W+i)] \right|$$

$$\leq \left| \sum_{\alpha \in \Gamma} \sum_{i=1}^{\binom{n}{k}} E[I_\alpha 1\{Z_\alpha = i\} (g(Y_\alpha + X_\alpha + i) - g(Y_\alpha + i))] \right|$$

$$+ \left| \sum_{\alpha \in \Gamma} \sum_{i=1}^{\binom{n}{k}} E[I_\alpha 1\{Z_\alpha = i\}] E[g(Y_\alpha + i) - g(W + i)] \right|$$

$$+ \left| \sum_{\alpha \in \Gamma} \sum_{i=1}^{\binom{n}{k}} E[I_\alpha 1\{Z_\alpha = i\} g(Y_\alpha + i)] - E[I_\alpha 1\{Z_\alpha = i\}] E[g(Y_\alpha + i)] \right|. \tag{4.3}$$

The first two expressions on the right-hand side are bounded by

$$|\Delta g| \left\{ \sum_{\alpha \in \Gamma} E[I_\alpha X_\alpha] + \sum_{\alpha \in \Gamma} E[I_\alpha] E[X_\alpha + U_\alpha + I_\alpha] \right\}. \tag{4.4}$$

To bound the third, we need the random variable S_α, introduced in Section 3.2, which is the number of points ξ_i, $i \notin \alpha$, which lie in $\mathbf{R}_{\alpha^*}^{(4)}$, and which is independent of I_α.

9

Note also that the distribution of Y_α given S_α is independent of $I_\alpha 1\{Z_\alpha = i\}$. Thus we find that

$$\left| E[I_\alpha 1\{Z_\alpha = i\}g(Y_\alpha + i)] - E[I_\alpha 1\{Z_\alpha = i\}]E[g(Y_\alpha + i)] \right|$$

$$= \left| E[E[I_\alpha 1\{Z_\alpha = i\}g(Y_\alpha + i) \mid I_\alpha, Z_\alpha, S_\alpha]] - E[I_\alpha 1\{Z_\alpha = i\}]E[g(Y_\alpha + i)] \right|$$

$$\leq \sum_{s=0}^{n-k} P(I_\alpha = 1, Z_\alpha = i, S_\alpha = s)$$

$$\times \left| E[g(Y_\alpha + i) \mid I_\alpha = 1, Z_\alpha = i, S_\alpha = s] - E[g(Y_\alpha + i)] \right|. \tag{4.5}$$

We bound $|E[g(Y_\alpha + i) \mid I_\alpha = 1, Z_\alpha = i, S_\alpha = s] - E[g(Y_\alpha + i)]|$ by means of a coupling. For each $\alpha \in \Gamma$ and $s = 0, \ldots, n-k$, let $Y''_{\alpha s}$ and Y'_α be random variables defined on the same probability space such that

$$\mathcal{L}(Y''_{\alpha s}) = \mathcal{L}(Y_\alpha \mid S_\alpha = s) \quad \text{and} \quad \mathcal{L}(Y'_\alpha) = \mathcal{L}(Y_\alpha),$$

constructed as follows. First note that $Y''_{\alpha s}$ is determined by $n - k - s$ independent uniformly distributed points in $\mathbf{A} \setminus \mathbf{R}^{(4)}_{\alpha*}$, while Y'_α is determined by $n - k - N_\alpha$ such points, where $\mathcal{L}(N_\alpha) = \text{Bi}(n - k, 16|C|)$. Set

$$M_{\alpha s} = \max\{n - k - s, n - k - N_\alpha\}; \quad m_{\alpha s} = \min\{n - k - s, n - k - N_\alpha\},$$

and let $(\eta_i, \ i \geq 1)$ be uniformly and independently distributed on $\mathbf{A} \setminus \mathbf{R}^{(4)}_{\alpha*}$, independently of N_α. Let the k–subsets consisting of points $\eta_1, \ldots, \eta_{m_{\alpha s}}$ be numbered from 1 to $\binom{m_{\alpha s}}{k}$, and the k–subsets also involving the $M_{\alpha s} - m_{\alpha s}$ remaining points from $\binom{m_{\alpha s}}{k} + 1$ to $\binom{M_{\alpha s}}{k}$. Then the variables below have the required distributions:

$$Y'_\alpha = \sum_{i=1}^{\binom{n-k-N_\alpha}{k}} J_{\alpha i} \quad \text{and} \quad Y''_{\alpha s} = \sum_{i=1}^{\binom{n-k-s}{k}} J_{\alpha i}, \tag{4.6}$$

where $J_{\alpha i}$ is 1 if the ith of the k–subsets of the points $\eta_1, \ldots, \eta_{M_{\alpha s}}$ is covered by some C^a, and 0 otherwise. Then, for $s = 0, \ldots, n-k$, it follows that

$$\left| E[g(Y_\alpha + i) \mid I_\alpha = 1, Z_\alpha = i, S_\alpha = s] - E[g(Y_\alpha + i)] \right| \leq |\Delta g| E|Y''_{\alpha s} - Y'_\alpha|, \tag{4.7}$$

since the conditioning on $I_\alpha = 1$ and $Z_\alpha = i$ is superfluous when $S_\alpha = s$ is given. Hence we obtain

$$\left| \sum_{i=1}^{\binom{n}{k}} E[I_\alpha 1\{Z_\alpha = i\}g(Y_\alpha + i)] - E[I_\alpha 1\{Z_\alpha = i\}]E[g(Y_\alpha + i)] \right|$$

$$\leq |\Delta g| \sum_{i=1}^{\binom{n}{k}} \sum_{s=0}^{n-k} P(I_\alpha = 1, Z_\alpha = i, S_\alpha = s) E|Y''_{\alpha s} - Y'_\alpha|$$

10

$$
= |\Delta g| \sum_{s=0}^{n-k} E|Y_{\alpha s}'' - Y_\alpha'| \sum_{i=1}^{\binom{n}{k}} P(I_\alpha = 1, Z_\alpha = i, S_\alpha = s)
$$

$$
= |\Delta g| E[I_\alpha] \sum_{s=0}^{n-k} E|Y_{\alpha s}'' - Y_\alpha'| P(S_\alpha = s), \tag{4.8}
$$

where the last equality follows from the independence of I_α and S_α. (4.4) and (4.8) inserted in (4.3) now give the following bound valid for any bounded $g : \mathbb{N} \to \mathbb{R}$:

$$
\left| E[W g(W)] - \sum_{i=1}^{\binom{n}{k}} i\lambda_i g(W + i) \right|
$$

$$
\leq |\Delta g| \left(\sum_{\alpha \in \Gamma} E[I_\alpha X_\alpha] + \sum_{\alpha \in \Gamma} E[I_\alpha] E[X_\alpha + U_\alpha + I_\alpha] \right.
$$

$$
\left. + \sum_{\alpha \in \Gamma} E[I_\alpha] \sum_{s=0}^{n-k} P(S_\alpha = s) E|Y_{\alpha s}'' - Y_\alpha'| \right). \tag{4.9}
$$

The next step is to bound the probabilities and expectations involved in (4.9).

The summands in X_α, defined in (4.2), concern k–subsets with at least one point in $\mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}$. When $I_\alpha = 1$, all the points ξ_{α_i}, $\alpha_i \in \alpha$, lie clustered close to the midpoint ξ_{α^*} of $\mathbf{R}_{\alpha^*}^{(4)}$. Because of the choice of size of $\mathbf{R}_{\alpha^*}^{(4)}$, this means that if $I_\alpha = 1$, $\beta \cap \alpha \neq \emptyset$ and $\xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}$ for at least one $\beta_j \in \beta$, then $I_\beta = 0$. Hence only those β for which $\alpha \cap \beta = \emptyset$ contribute to $E[I_\alpha X_\alpha]$. If $\alpha \cap \beta = \emptyset$, then I_β is independent of I_α, and, since there are $\binom{n-k}{k}$ such β, we get

$$
E[I_\alpha X_\alpha] = E[X_\alpha \mid I_\alpha = 1] P(I_\alpha = 1)
$$

$$
= E[\sum_{\beta \in \Gamma \setminus \{\alpha\}: \alpha \cap \beta = \emptyset} I_\beta \mathbf{1}\{\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\} \mid I_\alpha = 1] P(I_\alpha = 1)
$$

$$
= \binom{n-k}{k} E[I_\beta \mathbf{1}\{\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\}] P(I_\alpha = 1)
$$

$$
= \binom{n-k}{k} P(\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)} \mid I_\beta = 1) P(I_\alpha = 1)^2. \tag{4.10}
$$

If $I_\beta = 1$, then C^X, for X with coordinates the (torus) minima of the coordinates of ξ_{β_i}, $\beta_i \in \beta$, in each direction, covers all ξ_{β_i}. Furthermore, since all ξ_{β_i} are uniformly distributed on $\mathbf{A}$, X is also uniformly distributed on $\mathbf{A}$. The conditional probability above is therefore less than the probability that C^X intersects any of the edges of $\mathbf{R}_{\alpha^*}^{(4)}$. Thus

$$
P(\exists \beta_i, \beta_j \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}, \xi_{\beta_j} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)} \mid I_\beta = 1) \leq P(C^X \cap (\text{edges of } \mathbf{R}_{\alpha^*}^{(4)}) \neq \emptyset)
$$

$$
= 16|C|, \tag{4.11}
$$

which together with (3.1) and (4.10) yields

$$
E[I_\alpha X_\alpha] \leq 16 \binom{n-k}{k} k^4 |C|^{2k-1}. \tag{4.12}
$$

11

By the definitions of U_α and X_α, given in (4.1) and (4.2), respectively,

$$E[X_\alpha + U_\alpha] = E\left[\sum_{\beta \in \Gamma \setminus \{\alpha\}, \beta \cap \alpha \neq \emptyset} I_\beta + \sum_{\beta \in \Gamma \setminus \{\alpha\}, \beta \cap \alpha = \emptyset} I_\beta 1\{\exists \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}\} \right]. \quad (4.13)$$

There are $\binom{k}{l}\binom{n-k}{k-l}$ different k–subsets with exactly l points in common with the k–subset α, $l = 1, \ldots, k-1$, and there are $\binom{n-k}{k}$ different k–subsets with no point in common with α. By changing the index of summation in (4.13), we get the following equality:

$$E[X_\alpha + U_\alpha] = \sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} E[I_\beta] + \binom{n-k}{k} E[I_\beta 1\{\exists \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)}\}].$$

By arguments similar those that led to (4.11),

$$P(\exists \beta_i \in \beta : \xi_{\beta_i} \in \mathbf{R}_{\alpha^*}^{(4)} \mid I_\beta = 1) \leq 25|C|,$$

and hence we get

$$E[I_\alpha]E[X_\alpha + U_\alpha + I_\alpha] \leq$$
$$k^4 |C|^{2(k-1)} \left(\sum_{l=1}^{k-1} \binom{k}{l}\binom{n-k}{k-l} + 25 \binom{n-k}{k} |C| + 1 \right). \quad (4.14)$$

It remains to bound

$$E[I_\alpha] \sum_{s=0}^{n-k} P(S_\alpha = s) E|Y_{\alpha s}'' - Y_\alpha'|$$

$$= E[I_\alpha] \sum_{s=0}^{n-k} P(S_\alpha = s) \left(E[Y_{\alpha s}'' - Y_\alpha'] + 2E[(Y_\alpha' - Y_{\alpha s}'')1\{Y_{\alpha s}'' < Y_\alpha'\}] \right)$$

$$\leq E[I_\alpha] \sum_{s=0}^{n-k} P(S_\alpha = s) \left(E[Y_{\alpha s}'' - Y_\alpha'] + 2E[(Y_\alpha' - Y_{\alpha s}'')1\{N_\alpha < s\}] \right). \quad (4.15)$$

First, it follows directly from the definitions of $Y_{\alpha s}''$ and Y_α', given in (4.6), that

$$\sum_{s=0}^{n-k} P(S_\alpha = s) E[Y_\alpha' - Y_{\alpha s}''] = 0, \quad (4.16)$$

since $\mathcal{L}(S_\alpha) = \mathcal{L}(N_\alpha)$, and that

$$E[(Y_\alpha' - Y_{\alpha s}'')1\{N_\alpha < s\}]$$

$$= \sum_{r=0}^{s-1} E[Y_\alpha' - Y_{\alpha s}'' \mid N_\alpha = r] P(N_\alpha = r)$$

$$= \sum_{r=0}^{s-1} \left[\binom{n-k-r}{k} - \binom{n-k-s}{k} \right] P(J_{\alpha 1} = 1) P(N_\alpha = r). \quad (4.17)$$

12

Then, from the definition of $J_{\alpha 1}$, given below (4.6), it follows that

$$P(J_{\alpha 1} = 1) = P(I_\beta = 1 \mid \xi_{\beta_1}, \ldots, \xi_{\beta_k} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)})$$

$$= \frac{P(\{I_\beta = 1\} \cap \{\xi_{\beta_2}, \ldots, \xi_{\beta_k} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}\} \mid \xi_{\beta_1} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}) \, P(\xi_{\beta_1} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)})}{P(\xi_{\beta_1}, \ldots, \xi_{\beta_k} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)})}$$

$$\leq \frac{P(I_\beta = 1 \mid \xi_{\beta_1} \in \mathbf{A} \setminus \mathbf{R}_{\alpha^*}^{(4)}) \, (1 - 16|C|)}{(1 - 16|C|)^k}$$

$$= \frac{P(I_\beta = 1)}{(1 - 16|C|)^{k-1}}$$

$$= k^2 \left(\frac{|C|}{1 - 16|C|} \right)^{k-1}, \tag{4.18}$$

where the second last equality follows by the torus convention and the uniform distribution of the points, and the last equality by (3.1).

Furthermore,

$$\binom{n-k-r}{k} = \sum_{i=1}^{s-r} \binom{n-k-r-i}{k-1} + \binom{n-k-r-(s-r)}{k}$$

$$\leq \binom{n-k-r-1}{k-1}(s-r) + \binom{n-k-s}{k}$$

and hence

$$\binom{n-k-r}{k} - \binom{n-k-s}{k} \leq \binom{n-k-r-1}{k-1}(s-r). \tag{4.19}$$

Combining (4.15)–(4.19) yields

$$E[I_\alpha] \sum_{s=0}^{n-k} P(S_\alpha = s) E|Y_\alpha' - Y_{\alpha s}''|$$

$$\leq E[I_\alpha] 2k^2 \left(\frac{|C|}{1 - 16|C|} \right)^{k-1} \sum_{s=0}^{n-k} \sum_{r=0}^{l-1} \binom{n-k-r-1}{k-1}(s-r)P(N_\alpha = r)P(S_\alpha = s)$$

$$\leq E[I_\alpha] 2k^2 \left(\frac{|C|}{1 - 16|C|} \right)^{k-1} E[S_\alpha] \binom{n-k}{k-1}$$

$$= 32 \, k^4 |C|^k \left(\frac{|C|}{1 - 16|C|} \right)^{k-1} (n-k) \binom{n-k}{k-1}, \tag{4.20}$$

where the last equality follows from (3.1) and since $\mathcal{L}(S_\alpha) = \text{Bin}(n - k, |\mathbf{R}_{\alpha^*}^{(4)}|)$ and $|\mathbf{R}_{\alpha^*}^{(4)}| = 16|C|$. Since (4.9), (4.12), (4.14) and (4.20) hold for all $\alpha \in \Gamma$, Lemma 4.1 is now proved. ∎

We now turn our attention to estimating $P(W < \frac{1}{2}(1 + x)E[W])$, for suitable values of $x \in (0, 1)$. Our argument is based on:

Janson's inequality. (Janson (1990)) *Consider a collection $(J_i, i \in Q)$ of independent indicator variables and a (finite) family $(Q(\alpha), \alpha \in \Gamma_\alpha)$ of subsets of the index set Q,*

and define $I_\alpha = \Pi_{i \in Q(\alpha)} J_i$ and $W = \sum_{\alpha \in \Gamma_\alpha} I_\alpha$. Partition Γ_α into $\Gamma_\alpha^+ \cup \Gamma_\alpha^i$, where $\Gamma_\alpha^+ = \{\beta \neq \alpha : Q(\alpha) \cap Q(\beta) \neq \emptyset\}$. Then, if $0 \leq \varepsilon \leq 1$,

$$P(W \leq (1 - \varepsilon)E[W]) \leq \exp\left\{-\frac{1}{2(1+\delta)}\varepsilon^2 E[W]\right\}, \qquad (4.21)$$

where

$$\delta = \frac{1}{E[W]} \sum_\alpha \sum_{\beta \in \Gamma_\alpha^+} E[I_\alpha I_\beta].$$

Using the inequality (4.21), we establish the following lemma.

Lemma 4.2 *For any $0 < y < 1$ and $y^{1/k} < z < 1$, we have*

$$P(W_n \leq yE[W_n]) \leq (2\pi n)^{-1/2}(1 - z)^{-1}e^{-nD_1(z)} + e^{-E[W_n]D_2(y,z)}, \qquad (4.22)$$

where $D_1(z)$ and $D_2(y, z)$ are as given in (4.24) and (4.31) below.

Proof. To apply Janson's inequality to establish (4.22), we proceed in a roundabout way. First, we construct another random variable W_M from the sequence $(\xi_i, i \geq 1)$, in the same way as W_n is constructed, but now based on the M points $\xi_1, \ldots, \xi_M$, where $M \sim \text{Po}(nz)$ is independent of the ξ_i; note that $W_M \leq W_n$ whenever $M \leq n$. Thus we have

$$P(W \leq yE[W]) \leq P(W \leq W_M) + P(W_M \leq yE[W]).$$

Now, for any fixed $t \in \mathbb{N}$, divide $\mathbf{A}$ into t^2 small squares $K_i^{(t)}$, $1 \leq i \leq t^2$, of equal size, with

$$Y_i^{(t)} = \sum_{j=1}^{M} I[\xi_j \in K_i^{(t)}] \sim \text{Po}(nzt^{-2})$$

the number of points in square i, $i = 1, \ldots, t^2$. Then, defining $B_t = \cap_{i=1}^{t^2}\{Y_i^{(t)} \leq 1\}$, and letting P_{B_t} denote probability conditional on the event B_t, we have

$$\begin{aligned}
&P(W \leq yE[W]) \\
&\leq P(W \leq W_M) + P_{B_t}(W_M \leq yE[W])P(B_t) + P(W_M \leq yE[W] \mid B_t^c)P(B_t^c) \\
&\leq P(M \geq n) + P_{B_t}(W_M \leq yE[W]) + P(B_t^c). \qquad (4.23)
\end{aligned}$$

The first and third probabilities on the right-hand side of the above inequality are easily bounded; defining

$$D_1(z) = (1 - z)^2/\{2(1 + z)\}, \qquad (4.24)$$

using Barbour, Holst and Janson (1992, p259), and letting $m = nz$, we have

$$\begin{aligned}
P(M \geq n) &\leq \left(\frac{n}{n - m}\right) \frac{\exp\{-m\}m^n}{n!} \\
&\leq \left(\frac{n}{n - m}\right) \exp\{-(n - m)^2/(2(m + n))\}/\sqrt{2\pi n} \\
&= (2\pi n)^{-1/2}(1 - z)^{-1}e^{-nD_1(z)}, \qquad (4.25)
\end{aligned}$$

and then

$$
\begin{aligned}
P(B_t^c) &\leq t^2 P(Y_1^{(t)} \geq 2) = t^2(1 - e^{-m/t^2}(1 + m/t^2)) \\
&\leq t^2(1 - (1 - m/t^2)(1 + m/t^2)) \\
&= m^2/t^2 = n^2 z^2/t^2,
\end{aligned}
\tag{4.26}
$$

so that $\lim_{t\to\infty} P(B_t^c) = 0$.

We now use Janson's inequality to bound the remaining part of the right hand side of (4.23). Conditional on B_t, $\{Y_1^{(t)}, \ldots, Y_{t^2}^{(t)}\}$ are independent Bernoulli Be $(m/(t^2+m))$ random variables, and can be used for the J_i in Janson's inequality. To exploit this idea, let Γ_α consist of all k–subsets $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ of $\{1, \ldots, t^2\}$, such that $\cup_{j=1}^k K_{\alpha_j}^{(t)}$ is covered by some C^a. Then, on the event B_t,

$$
W_M \geq \sum_{\alpha \in \Gamma_\alpha} \tilde{I}_\alpha := W_{M,t},
$$

where

$$
\tilde{I}_\alpha = \prod_{\alpha_i \in \alpha} Y_{\alpha_i}^{(t)},
$$

and

$$
P_{B_t}(\tilde{I}_\alpha = 1) = \left(\frac{m/t^2}{1 + m/t^2} \right)^k.
$$

Hence, if we set

$$
\theta_t = 1 - yE[W]/E_{B_t}[W_{M,t}],
$$

then, by Janson's inequality (4.21), if $0 \leq \theta_t \leq 1$, it follows that

$$
\begin{aligned}
P_{B_t}(W_M \leq yE[W]) &\leq P_{B_t}(W_{M,t} \leq (1 - \theta_t)E_{B_t}[W_{M,t}]) \\
&\leq \exp\left\{ -\frac{\theta_t^2 E_{B_t}[W_{M,t}]}{2(1 + \delta_t)} \right\},
\end{aligned}
\tag{4.27}
$$

where

$$
\delta_t = \frac{1}{E_{B_t}[W_{M,t}]} \sum_\alpha \sum_{\beta \in \Gamma_\alpha^+} E_{B_t}[\tilde{I}_\alpha \tilde{I}_\beta] = \sum_{\beta \in \Gamma_1^+} P_{B_t}(\tilde{I}_\beta = 1 \mid \tilde{I}_1 = 1),
$$

and Γ_α^+ consists of the indices of those k–subsets of squares which have at least one square in common with the k–subset α.

Now observe that $\lim_{t\to\infty} P(B_t) = 1$, that $\lim_{t\to\infty} W_{M,t} I[B_t] = W_M$ a.s. and that $W_{M,t} I[B_t] \leq \binom{M}{k}$, so that

$$
\begin{aligned}
\lim_{t\to\infty} E_{B_t}[W_{M,t}] &= E[W_M] = E\left[\binom{M}{k} \right] P(I_1 = 1) \\
&= \frac{m^k}{k!} k^2 |C|^{k-1} \geq E[W] z^k,
\end{aligned}
$$

by (3.1) and (3.2), since $m = nz$, and hence

$$
\lim_{t\to\infty} \theta_t^2 E_{B_t}[W_{M,t}] \geq \left(1 - yz^{-k} \right)^2 E[W] z^k.
\tag{4.28}
$$

To bound δ_t, let r_t^2 be the maximal number of small squares that fit in C. Then $r_t^2/t^2 \leq |C|$, and there exist less than $\binom{k}{l}\binom{4r_t^2-k}{k-l}$ k–subsets of small squares which have l squares in common with the first k–subset. Hence, since if α and $\beta \in \Gamma_\alpha^+$ have l common squares, then

$$P_{B_t}(\tilde{I}_\beta = 1 \mid \tilde{I}_\alpha = 1) = \left(\frac{m/t^2}{1+m/t^2}\right)^{k-l},$$

we find that

$$
\begin{aligned}
\delta_t &\leq \sum_{l=1}^{k-1}\binom{k}{l}\binom{4r_t^2-k}{k-l}\left(\frac{m/t^2}{1+m/t^2}\right)^{k-l} \\
&\leq \sum_{l=1}^{k-1}\binom{k}{l}\frac{(4r_t^2)^{k-l}}{(k-l)!}(m/t^2)^{k-l} \\
&\leq \sum_{l=1}^{k-1}\binom{k}{l}\frac{4^{k-l}}{(k-l)!}(m|C|)^{k-l} \\
&\leq \sum_{l=1}^{k-1}\binom{k}{l}\frac{4^{k-l}}{(k-l)!},
\end{aligned}
\tag{4.29}
$$

where the right-hand side is independent of t. Inserting (4.29) and (4.28) into (4.27) yields

$$\lim_{t\to\infty} P_{B_t}(W_M \leq yE[W]) \leq e^{-E[W]D_2(y,z)}, \tag{4.30}$$

where

$$D_2(y,z) = \frac{z^k(1-yz^{-k})^2}{2+2\sum_{l=1}^{k-1}\binom{k}{l}4^{k-l}/(k-l)!}, \tag{4.31}$$

provided that $z^k > y$. Now insert (4.25), (4.26) and (4.30) into (4.23), to get

$$P(W \leq yE[W]) \leq (2\pi n)^{-1/2}(1-z)^{-1}e^{-nD_1(z)} + e^{-E[W]D_2(y,z)},$$

for any z such that $y^{1/k} < z < 1$. ∎

Proof of Theorem 3.3. By (2.5) and Lemma 4.1,

$$d_{TV}(\mathcal{L}(W), \mathrm{CP}(\Lambda)) \leq \varepsilon_1 H_1(\Lambda),$$

where ε_1 is given in (3.5), and all that remains is thus to bound $H_1(\Lambda)$. There are two possibilities: (2.6) and (2.8). In our case, $i\lambda_i$ does not in general decrease, and so (2.8) does not apply. This leaves us with (2.6), from which (3.7) is obtained directly. If $E[W_n]$ stays bounded away from ∞, then the bound in (3.7) is of order $O(\varepsilon_1)$; the critical part of ε_1 is

$$\binom{n}{k}k^4|C|^{2(k-1)}k\binom{n-k}{k-1} = O(n^{2k-1}|C|^{2(k-1)}),$$

from which Theorem 3.3 and Corollary 3.4 follow.

Proof of Theorem 3.5 (*i*). In trying to apply Theorem 2.1, we immediately encounter problems; the probability of 'large' clumps is not small enough for Condition (ii) of Theorem 2.1 to be satisfied. To circumvent this problem, we truncate the clump size distribution as follows. Recall that S_α is the number of points in $\mathbf{R}_{\alpha *}^{(4)}$, not counting those in the k–subset α. As in (3.4), and suppressing the index n where possible, define

$$\lambda_i' = i^{-1} \sum_{\alpha \in \Gamma} E[I_\alpha 1\{Z_\alpha = i\} \mid S_\alpha \leq s_0] = i^{-1} E[W] P(Z_1 = i \mid I_1 = 1, S_1 \leq s_0), \quad (4.32)$$

where we now fix

$$s_0 > 2 \max\{k, \sup_{n \geq 1} n|C_n|\},$$

and set $\Omega' = \sum_{i \geq 1} \lambda_i'$ and $\mu_i' = \lambda_i'/\Omega'$. Conditioning on $S_1 \leq s_0$ implies that there are at most $s_0 + k$ points in $\mathbf{R}_{1*}^{(4)}$, from which there can be at most $\binom{s_0+k}{k}$ covered k–subsets; thus $Z_1 \leq \binom{s_0+k}{k}$, and $\lambda_i' = \mu_i' = 0$ for $i > \binom{s_0+k}{k}$, so that Condition (*ii*) of Theorem 2.1 is satisfied. With this definition, we still have $\sum i\lambda_i' = E[W]$, because I_α and S_α are independent, while the expected number of clumps Ω' has increased, and the expected number of covered k–subsets in a clump, $E[W]/\Omega'$, has decreased. Furthermore

$$\Omega' = \sum_{i=1}^{\binom{s_0+k}{k}} \lambda_i' = E[W] \sum_{i=1}^{\binom{s_0+k}{k}} \frac{1}{i} P(Z_1 = i \mid I_1 = 1, S_1 \leq s_0) \left\{ \begin{array}{ll} \leq & E[W], \\ \geq & E[W]/\binom{s_0+k}{k}, \end{array} \right. \quad (4.33)$$

so that Ω' is of the same order as $E[W]$. Conditions (*i*) and (*iii*) of Theorem 2.1 are satisfied because $\sum_{i \geq 1} \mu_{in}' = 1$ for all n, and because

$$\begin{aligned} \mu_{1n}' &= (E[W]/\Omega') P(Z_1 = 1 \mid I_1 = 1, S_1 \leq s_0) \\ &\geq P(S_1 = 0 \mid I_1 = 1, S_1 \leq s_0) \geq P(S_1 = 0), \end{aligned}$$

with $\mathcal{L}(S_1) = \mathrm{Bi}\,(n - k, 16|C|)$, so that $\lim_{n \to \infty} \mu_{1n}' = 1$ if $n|C_n| \to 0$.

It remains to bound the right hand side of (2.9) in Theorem 2.1 for our particular sequence W_n. The first step is to show that (2.4) is satisfied with $\varepsilon_0 = 0$ and with some suitable ε_1, and now with λ_i' instead of λ_i. By the triangle inequality,

$$\left| E\left[W g(W) - \sum_{i \geq 1} i\lambda_i' g(W + i) \right] \right| \leq \quad (4.34)$$

$$\left| E\left[W g(W) - \sum_{i \geq 1} i\lambda_i g(W + i) \right] \right| + \left| E\left[\sum_{i \geq 1} i\lambda_i g(W + i) - \sum_{i \geq 1} i\lambda_i' g(W + i) \right] \right|,$$

in which the first term is bounded by $|\Delta g|\varepsilon_1$ by Lemma 4.1. For the second, from the definition $d_W(P, Q) = \sup_{g:\,|\Delta g| \leq 1} \left| \int g\, dP - \int g\, dQ \right|$ of the Wasserstein distance between probability measures on $\mathbf{Z}^+$, we have

$$\left| E\left[\sum_{i \geq 1} i\lambda_i g(W + i) - \sum_{i \geq 1} i\lambda_i' g(W + i) \right] \right| \leq |\Delta g| E[W]\, d_W(P, P'), \quad (4.35)$$

where $P\{i\} = i\lambda_i/E[W]$ and $P'\{i\} = i\lambda_i'/E[W]$, $i \geq 1$: and, from the definitions of λ_i and λ_i',

$$d_W(P, P') = d_W(\mathcal{L}(Z_1 \mid I_1 = 1), \mathcal{L}(Z_1 \mid I_1 = 1,\ S_1 \leq s_0)).$$

Observe that, in view of the natural coupling, $\mathcal{L}(Z_1 \,|\, I_1 = 1, \ S_1 = s)$ is stochastically increasing in s, and hence that

$$
\begin{aligned}
d_W(\mathcal{L}(Z_1 \,|\, I_1 = 1), \mathcal{L}(Z_1 \,|\, I_1 = 1, \ S_1 \le s_0)) &\le\ E(Z_1 I[S_1 > s_0] \,|\, I_1 = 1) \\
&\le\ E\left\{ \binom{S_1 + k}{k} I[S_1 > s_0] \right\},
\end{aligned}
$$

where $S_1 \sim \mathrm{Bi}\,(n - k, 16|C|)$. But now $\mathrm{Bi}\,(m, p)\{j\} \le \mathrm{Po}\,(mp)\{j\}$ for all $j \ge 2mp + 1$, and $\{(j + k)(j + k - 1) \ldots (j + 1)\}/\{j(j - 1) \ldots (j - k + 1)\} \le 2^k$ whenever $j \ge 2k - 1$, giving

$$
d_W(\mathcal{L}(Z_1 \,|\, I_1 = 1), \mathcal{L}(Z_1 \,|\, I_1 = 1, \ S_1 \le s_0)) \le (2^k/k!)(16n|C|)^k \mathrm{Po}\,(16n|C|)(\tfrac{1}{2}s_0, \infty),
\tag{4.36}
$$

since $s_0 > 2\max\{k, n|C_n|\}$.

Combining (4.34)–(4.36), it thus follows that

$$
\left| E\left[Wg(W) - \sum_{i \ge 1} i\lambda_i' g(W + i) \right] \right| \le \varepsilon_1' |\Delta g|,
$$

where

$$
\varepsilon_1' = \varepsilon_1 + (2^k/k!)(16n|C|)^k \mathrm{Po}\,(16n|C|)(\tfrac{1}{2}s_0, \infty) E[W],
$$

and ε_1 is given in (3.5). Thus, and using (4.33), (2.4) is satisfied with $\varepsilon_0' = 0$, and $\varepsilon_1' = O((n|C_n|)^{k-1}\Omega_n')$ as $n \to \infty$.

The remaining element, a bound on $P(W_n \le \tfrac{1}{2}(1 + x)E[W_n])$ for suitable values of $x \in (0, 1)$, is given in Lemma 4.2, the proof of Theorem 3.5 being completed by choosing $x = \tfrac{1}{2}(1 + c_2)$, $y = \tfrac{1}{2}(1 + x)$ and z any value between $y^{1/k}$ and 1.

Proof of Theorem 3.5 (*ii*). By the triangle inequality

$$
d_{TV}(\mathcal{L}(W_n), \mathrm{CP}(\Lambda_n)) \ \le\ d_{TV}(\mathcal{L}(W_n), \mathrm{CP}(\Lambda_n')) + d_{TV}(\mathrm{CP}(\Lambda_n), \mathrm{CP}(\Lambda_n')). \tag{4.37}
$$

In order to bound $d_{TV}(\mathrm{CP}(\Lambda_n), \mathrm{CP}(\Lambda_n'))$, let $\widetilde{W}_n = \sum_{i \ge 1} iN_i$, where N_i, $i \ge 1$, are independent and $\mathcal{L}(N_i) = \mathrm{Po}(\lambda_{in})$, so that $\mathcal{L}(\widetilde{W}_n) = \mathrm{CP}\,(\Lambda_n)$. We will use Theorem 2.1, and hence we need a bound on

$$
\left| E\left[\sum_{i \ge 1} i\lambda_{in}' g(\widetilde{W}_n + i) - \widetilde{W}_n g(\widetilde{W}_n) \right] \right| \le
$$

$$
\left| E\left[\sum_{i \ge 1} i\lambda_{in}' g(\widetilde{W}_n + i) - i\lambda_{in} g(\widetilde{W}_n + i) \right] \right| + \left| E\left[\sum_{i \ge 1} i\lambda_{in} g(\widetilde{W}_n + i) - \widetilde{W}_n g(\widetilde{W}_n) \right] \right|.
$$

The first term on the right-hand side of the above inequality is bounded by

$$
|\Delta g|(2^k/k!)(16n|C_n|)^k \mathrm{Po}\,(16n|C_n|)(\tfrac{1}{2}s_0, \infty),
$$

as in the proof of (i), while the second term is zero for those g of interest. In the proof of (i) we also confirmed that assumptions (i)–(iii) of Theorem 2.1 are satisfied for $(\Lambda'_n, n \geq 1)$, and hence

$$
\begin{aligned}
d_{TV}(\mathrm{CP}(\Lambda_n), \mathrm{CP}(\Lambda'_n)) &= d_{TV}(\mathcal{L}(\widetilde{W}_n), \mathrm{CP}(\Lambda'_n)) \\
&\leq O\left((n|C_n|)^k + P(\widetilde{W}_n \leq \tfrac{1}{2}(1+x)E[\widetilde{W}_n])\right). \quad (4.38)
\end{aligned}
$$

By Markov's inequality and by the definition of $\widetilde{W}_n$, it follows that

$$
\begin{aligned}
P(\widetilde{W}_n \leq yE[\widetilde{W}_n]) &= P\left(\exp\{-t\widetilde{W}_n\} \geq \exp\{-tyE[\widetilde{W}_n]\}\right) \\
&\leq E[\exp\{-t\widetilde{W}_n\}]\exp\{tyE[\widetilde{W}_n]\} \\
&\leq \exp\{\sum_{i \geq 1} \lambda_{in}(e^{-ti} - 1 + ity)\} \\
&\leq \exp\{\sum_{i \geq 1} \lambda_{in}(it(y - 1) + (ti)^2/2)\}, \quad (4.39)
\end{aligned}
$$

for $t > 0$. Furthermore, by the definition of λ_{in}, given in (3.3),

$$
\begin{aligned}
\sum_{i \geq 1} \lambda_{in} i^2 &= E[W_n]E[Z_1 | I_1 = 1] \\
&\leq E[W_n]E\left[\binom{S_1 + k}{k}\right] \\
&\leq E[W_n]\left(\frac{(2k)^k}{k!} + E\left[\binom{S_1 + k}{k}\mathbf{1}\{S_1 > k\}\right]\right).
\end{aligned}
$$

Since

$$
\frac{\binom{j+k}{k}}{\binom{j}{k}} = \frac{(2k)!}{(k!)^2}\prod_{i=0}^{j-k-1}\frac{(j - i + k)(j - i - k)}{(j - i)^2} \leq \frac{(2k)!}{(k!)^2} \leq \frac{(2k)^k}{k!},
$$

if $j > k$, and $E[\binom{S_1}{k}] \leq (16(n - k)|C_n|)^k \leq 1$ as $n \to \infty$, it follows that

$$
\sum_{i \geq 1} \lambda_{in} i^2 \leq E[W_n]\frac{2(2k)^k}{k!},
$$

which inserted in (4.39) yields

$$
P(\widetilde{W}_n \leq yE[\widetilde{W}_n]) \leq \exp\{E[W_n](t(y - 1) + t^2(2k)^k/k!)\}. \quad (4.40)
$$

Letting $y = (1 + x)/2$, where $x = (1 + c_2)/2$, and $t = (1 - y)k!/(2k)^k$, the right-hand side of (4.40) inserted in (4.38), yields

$$
d_{TV}(\mathrm{CP}(\Lambda_n), \mathrm{CP}(\Lambda'_n)) \leq O\left((n|C_n|)^k + \exp\left\{-E[W_n]\frac{(1 - x)^2 k!}{2(2k)^k}\right\}\right),
$$

which together with the bound achieved in (i) and (4.37) completes the proof.

References

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

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

BARBOUR, A.D. (1997). Stein's method. *Encyclopaedia of Statistical Sciences*, Update, Vol. 1, (Eds. S. Kotz, C. B. Read and D. L. Banks), Wiley, New York, 513–521.

BARBOUR, A.D., CHEN, L.H.Y. AND LOH, W.-L. (1992). Compound Poisson approximation for non–negative random variables via Stein's method. *Ann. Probab.* **20**, 1843–1866.

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 University Press.

BARBOUR, A.D. AND UTEV, S. (1998). Compound Poisson approximation in total variation. Preprint.

EGGLETON, P. AND KERMACK, W. O. (1944). A problem in the random distribution of particles. *Proc. Roy. Soc. Edinburgh Sect. A* **62**, 103–115.

JANSON, S. (1990). Poisson approximation for large deviations. *Random Structures Algorithms* **1**, 221–230.

LOADER, C. R. (1991). Large-deviation approximations to the distribution of scan statistics. *Adv. in 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. *Phil. Mag.* **39**, 778–790.

MACK, C. (1949). The expected number of aggregates in a random distribution of n points. *Proc. Cambridge Phil. Soc.* **46**, 285–292.

MÅNSSON, M. (1997). Poisson approximation in connection with clustering of random points. *Ann. Appl. Probab.*, (to appear).

MÅNSSON, M. (1999). On Poisson approximation for continuous multiple scan statistics in two dimensions. *Scan Statistics and Applications*, (Eds. J. Glaz and N. Balakrishnan), Birkhauser, Boston, (to appear).

ROOS, M. (1993). Compound Poisson approximations for the numbers of extreme spacings. *Adv. in Appl. Probab.* **25**, 847–874.

ROOS, M. (1994a). Stein's method for compound Poisson approximation: the local approach. *Ann. Appl. Probab.*, **4**, 1177–1187.

ROOS, M. (1994b). Stein–Chen method for compound Poisson approximation: the coupling approach. *Probability Theory and Mathematical Statistics*, Proceedings of the Sixth Vilnius Conference, 645–660.

SILBERSTEIN, L. (1945). The probable number of in random distributions of points. *Phil. Mag.* **36**, 319–336.

SILVERMAN, B.W. AND BROWN, T.C. (1978). Short distances, flat triangles and Poisson limits. *J. Appl. Probab.* **15**, 815–825.

SILVERMAN, B.W. AND BROWN, T.C. (1979). Rates of Poisson convergence for U-statistics. *J. Appl. Probab.* **16**, 428–432.