The Annals of Probability
2002, Vol. 30, No. 3, 1492–1537

COMPOUND POISSON PROCESS APPROXIMATION

BY A. D. BARBOUR[1] AND MARIANNE MÅNSSON[2]

Universität Zürich and Chalmers University of Technology

Compound Poisson processes are often useful as approximate models, when describing the occurrence of rare events. In this paper, we develop a method for showing how close such approximations are. Our approach is to use Stein's method directly, rather than by way of declumping and a marked Poisson process; this has conceptual advantages, but entails technical difficulties. Several applications are given to illustrate the procedure.

1. Introduction. Rare events in dependent systems frequently appear in clusters. A typical example is that of extremes in meteorological or financial time series, but multidimensional scan statistics and more complicated, graph based phenomena also behave in this way. As exemplified in Aldous (1989), the occurrence of such events can be approximated by supposing that the positions of the clusters are the points of a Poisson process, and that the structures of the individual clusters are independent and identically distributed. In this paper, we model the rare events as the points of a dependent point process Ξ on some space Γ, and use compound Poisson processes as approximations, summarizing each cluster solely by the number of points that it contains.

If only the total number of points $\Xi(\Gamma)$ were of interest, the corresponding approximation would be by a compound Poisson distribution on $\mathbb{Z}_+$. However, a compound Poisson process approximation contains much more information. For instance, in examining the occurrence of certain motifs in a stretch of DNA, one may be interested in detecting regions in which the density of such motifs is unusually high, without having to specify the length of region in advance; a compound Poisson process approximation may then provide a tractable approximate null model with which to compare data. It is therefore useful to have some idea of how good such an approximation may actually be.

A very useful and widely applicable method of quantifying such approximations is by way of Stein's method for Poisson process approximation, as in Arratia, Goldstein and Gordon (1989). The original point process Ξ on Γ is replaced by a point process $\widetilde{\Xi}$ on $\Gamma \times \mathbb{N}$, in which a point at (α, i) denotes a cluster of size i at position α. The construction of such a mapping, "declumping," is not always easy, and is usually far from natural. However, the advantages of the procedure, when

Received October 2000; revised October 2001.

[1]Supported in part by Schweizer Nationalfonds Projekte 20-50686.97 and 20-61753.00.

[2]Supported in part by the Swedish Natural Science Research Council.

AMS 2000 subject classifications. Primary 60G55, 62E17; secondary 60F05.

Key words and phrases. Extreme values, point processes, compound Poisson process, Stein's method.

it can be carried through, are substantial. If $E\widetilde{\Xi}(\Gamma \times \mathbb{N})$ is not too large, it leads to good approximation with respect to total variation distance [Arratia, Goldstein and Gordon (1989), Theorem 2], and if $E\widetilde{\Xi}(\Gamma \times \mathbb{N})$ is large there is still good approximation [Barbour, Holst and Janson (1992), Theorem 10.F] with respect to the weaker, but still very useful, $d_2(\Gamma \times \mathbb{N})$ distance defined in (2.10) below; see Barbour, Holst and Janson [(1992), page 218] and Barbour, Novak and Xia (1999) for some general discussion of the d_2-distance.

In this paper, we use Stein's method to approximate the original point process Ξ directly, using a compound Poisson process on Γ, without invoking $\widetilde{\Xi}$. One reason for doing so is to avoid the necessity to declump. Another concerns the difference between $d_2(\Gamma \times \mathbb{N})$ and $d_2(\Gamma)$. A bound $d_2(\mathcal{X})(Q, R) \le \delta$ between probability measures Q and R over $\mathcal{X}$ implies that the difference $\int f\,dQ - \int f\,dR$ can be bounded for any $d_1(\mathcal{X})$-Lipschitz function f [see (2.7) below] in terms of δ and the Lipschitz constant of f. However, the $d_1(\Gamma \times \mathbb{N})$-Lipschitz functions are not necessarily the ones most relevant for approximating Ξ. For instance, the function f_A, defined on configurations ξ of points of Γ by $f_A(\xi) := \mathbb{1}_{\{\xi(\Gamma) \in A\}}$, is $d_1(\Gamma)$-Lipschitz with constant 1, so that $d_2(\Gamma)$-approximation to $\mathcal{L}(\Xi)$ implies the same accuracy of approximation in total variation to $\mathcal{L}(\Xi(\Gamma))$. In contrast, the corresponding function

$$\tilde{f}_A(\tilde{\xi}) := \mathbb{1}_{\{\sum_{j \ge 1} j\tilde{\xi}(\Gamma \times \{j\}) \in A\}}$$

of configurations in $\Gamma \times \mathbb{N}$ is not $d_1(\Gamma \times \mathbb{N})$-Lipschitz, so that $d_2(\Gamma \times \mathbb{N})$-approximation to $\mathcal{L}(\widetilde{\Xi})$ does not directly entail a corresponding approximation for $\mathcal{L}(\Xi(\Gamma))$—the analogous approximation would be to the total number of clusters $\widetilde{\Xi}(\Gamma \times \mathbb{N})$.

More generally, a d_2-distance measures the average d_1-distance between pairs of configurations in an optimal coupling. Now two configurations each containing c clusters, which are identical except for one cluster at α, which is of size $s \ge 2$ in one of them and of size 1 in the other, are at $d_1(\Gamma \times \mathbb{N})$-distance $1/c$ from one another, which is small when c is large, irrespective of the value of s; in contrast, the $d_1(\Gamma)$-distance takes its maximal value of 1. Then again, let the two configurations be as before, except that at α there is an s-cluster in the first, which is split into two clusters of sizes s_1 and s_2 at α and α', where $s_1 + s_2 = s$. In $d_1(\Gamma \times \mathbb{N})$-distance, the configurations are at the maximal distance of 1 from another whereas, in $d_1(\Gamma)$-distance, they are separated only by $s_2 d_0(\alpha, \alpha')/m$, where m is the total number of points in the configurations, being small if α and α' are close to one another or if $s_2 \ll m$. In either case, the distance $d_1(\Gamma)$ seems to represent a more useful and natural measure of discrepancy than $d_1(\Gamma \times \mathbb{N})$, implying in turn that $d_1(\Gamma)$ is better suited than $d_1(\Gamma \times \mathbb{N})$ for practical application.

Unfortunately, as has already been experienced in the approximation of random variables, there are severe technical difficulties involved in the direct approach to compound Poisson approximation using Stein's method. As a result, we are only able to prove good bounds in $d_2(\Gamma)$-distance under the additional condition (3.7), which is the point process analogue of the condition $j\theta_j \searrow 0$ of Barbour, Chen and Loh (1992), when approximating the distribution of a random variable by a compound Poisson distribution of the form $\mathcal{L}(jN_j)$, $N_j \sim \mathrm{Po}(\theta_j)$. There are many examples where the condition (3.7) holds, in particular when the compound Poisson process is close to a Poisson process, but, if it is not satisfied, it seems to be necessary to attempt a declumping; improvements along the lines of those made for random variable approximation in Barbour and Utev (1998, 1999) are not yet in sight.

The structure of the paper is as follows. The general setting is outlined in Section 2, properties of the solutions to the Stein equations are proved in Section 3, and the main approximation theorems are given in Section 4. One novel aspect of the argument, compared to that used in applying Stein's method to Poisson process approximation, is that the solutions g to the Stein equation, usually bounded functions from the configuration space $\mathcal{H} := \mathcal{H}(\Gamma)$ to $\mathbb{R}$, are now allowed to be functions: $\mathcal{H} \times \Gamma \to \mathbb{R}$. The paper concludes with some illustrative examples.

2. Generalities. Let Ξ denote a point process on Γ, whose mean measure $\boldsymbol{\mu}$ has density $\mu(\alpha)$, $\alpha \in \Gamma$, with respect to some measure $\boldsymbol{\nu}$ on Γ, and has $\boldsymbol{\mu}(\Gamma) < \infty$. We wish to approximate the distribution of Ξ by the distribution $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$ of a compound Poisson process on Γ; here, $\boldsymbol{\pi}_i$ denotes the mean measure for the positions of clumps of size i, $i \geq 1$. The $\boldsymbol{\pi}_i$ are defined in terms of a measurable family of decompositions $\mathcal{N}$ of the point process Ξ:

$$\mathcal{N} : \Gamma \times \Omega \to \mathcal{H}^3 : (\alpha, \omega) \mapsto \big(\Xi^{0,\alpha}(\omega), \Xi^{s,\alpha}(\omega), \Xi^{r,\alpha}(\omega)\big), \tag{2.1}$$

where, for each (α, ω),

$$\Xi(\omega) = \Xi^{0,\alpha}(\omega) + \Xi^{s,\alpha}(\omega) + \Xi^{r,\alpha}(\omega).$$

We always set $\Xi^{0,\alpha}(\omega) := \Xi(\omega)\{\alpha\}$; then $\Xi^{s,\alpha}$ is taken to represent that part of Ξ which is "significantly dependent" on $\Xi\{\alpha\}$, and $\Xi^{r,\alpha}$ is what remains. The latter two choices are essentially arbitrary, and there is no need to specify them further until applying the general results, when better choices lead to smaller error bounds. In particular, for the bounds that we derive to be useful, the mean measure of $\Xi^{r,\alpha}$ should be close to $\boldsymbol{\mu}$.

Let P^α and E^α refer to the Palm measures of Ξ at α, so that, as in Kallenberg [(1983), Section 10], for measurable $g : \mathcal{H} \times \Gamma \to [0, \infty)$ and B a Borel set in Γ,

$$E\left\{\int_B g(\Xi, \alpha)\Xi(d\alpha)\right\} = \int_B E^\alpha g(\Xi, \alpha)\boldsymbol{\mu}(d\alpha), \tag{2.2}$$

and hence

$$E[g(\Xi, \alpha)\Xi(d\alpha)] = E^\alpha[g(\Xi, \alpha)]\boldsymbol{\mu}(d\alpha), \qquad \boldsymbol{\mu} \text{ a.e.} \tag{2.3}$$

The intensity measures $\boldsymbol{\pi}_i$ for clumps of size i, $i \geq 1$, in the approximating compound Poisson process CP$(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, are then defined by

$$i\boldsymbol{\pi}_i(d\alpha) := P^{\alpha}\big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\big)\boldsymbol{\mu}(d\alpha), \tag{2.4}$$

and their densities $\mu_i(\alpha) = (d\boldsymbol{\pi}_i/d\boldsymbol{\nu})(\alpha)$ are given by

$$i\mu_i(\alpha) = P^{\alpha}\big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\big)\mu(\alpha). \tag{2.5}$$

Note that, from (2.4), $\sum_{i\geq 1} i\boldsymbol{\pi}_i(d\alpha) = \boldsymbol{\mu}(d\alpha)$, so that

$$\boldsymbol{\pi}(\Gamma) := \sum_{i\geq 1}\boldsymbol{\pi}_i(\Gamma) \leq \sum_{i\geq 1} i\boldsymbol{\pi}_i(\Gamma) = \boldsymbol{\mu}(\Gamma) < \infty. \tag{2.6}$$

2.1. *Distances.* Let $\mathcal{H} := \mathcal{H}(\mathcal{X})$ denote the space of finite point process configurations on a space $\mathcal{X}$, with metric d_0 bounded by 1. We shall almost always take $\mathcal{X} = \Gamma$, but for comparison with the Poisson process approach it is convenient also to allow other choices of $\mathcal{X}$. Let $\mathcal{K} := \mathcal{K}(\mathcal{X})$ denote the set of functions $k : \mathcal{X} \to \mathbb{R}$ such that

$$s_1(k) = \sup_{y_1 \neq y_2 \in \mathcal{X}} |k(y_1) - k(y_2)|/d_0(y_1, y_2) < \infty,$$

and define a distance $d_1 := d_1(\mathcal{X})$ between finite measures $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ over $\mathcal{X}$ by

$$d_1(\boldsymbol{\rho}, \boldsymbol{\sigma}) = \begin{cases} 1, & \text{if } \boldsymbol{\rho}(\mathcal{X}) \neq \boldsymbol{\sigma}(\mathcal{X}), \\ 0, & \text{if } \boldsymbol{\rho}(\mathcal{X}) = \boldsymbol{\sigma}(\mathcal{X}) = 0, \\ m^{-1} \sup_{k\in\mathcal{K}} \dfrac{1}{s_1(k)} \left| \displaystyle\int k\, d\boldsymbol{\rho} - \int k\, d\boldsymbol{\sigma} \right|, & \text{if } \boldsymbol{\rho}(\mathcal{X}) = \boldsymbol{\sigma}(\mathcal{X}) = m > 0. \end{cases} \tag{2.7}$$

An alternative interpretation of d_1, when considered as a distance between configurations $\xi_1, \xi_2 \in \mathcal{H}$, is

$$d_1(\xi_1, \xi_2) = \min_{\pi\in S_n} \left\{ n^{-1} \sum_{i=1}^{n} d_0(y_{1i}, y_{2\pi(i)}) \right\}, \tag{2.8}$$

where $(y_{11}, \ldots, y_{1n})$ and $(y_{21}, \ldots, y_{2n})$ are the points of ξ_1 and ξ_2, respectively, and S_n is the set of permutations of $\{1, \ldots, n\}$. Hence d_1 measures the average distance between the points of the two configurations under the closest matching. Then, letting $\mathcal{F} := \mathcal{F}(\mathcal{X})$ denote the set of functions $f : \mathcal{H} \to \mathbb{R}$ such that

$$s_2(f) = \sup_{\xi_1\neq\xi_2\in\mathcal{H}} |f(\xi_1) - f(\xi_2)|/d_1(\xi_1, \xi_2) < \infty, \tag{2.9}$$

we define a distance $d_2(\mathcal{X})$ between probability measures over $\mathcal{H}(\mathcal{X})$ by

$$d_2(Q, R) = \sup_{f\in\mathcal{F}} \frac{1}{s_2(f)} \left| \int f\, dQ - \int f\, dR \right|. \tag{2.10}$$

For finite measures $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ over $\mathcal{X}$, we also use the notation $\hat{d}_1(\boldsymbol{\rho}, \boldsymbol{\sigma})$, defined as follows:

$$
(2.11) \qquad \hat{d}_1(\boldsymbol{\rho}, \boldsymbol{\sigma}) := \begin{cases} \inf\limits_{\boldsymbol{\rho}' \le \boldsymbol{\rho};\, \boldsymbol{\rho}'(\mathcal{X}) = \boldsymbol{\sigma}(\mathcal{X})} d_1(\boldsymbol{\rho}', \boldsymbol{\sigma}), & \text{if } \boldsymbol{\rho}(\mathcal{X}) \ge \boldsymbol{\sigma}(\mathcal{X}), \\ \hat{d}_1(\boldsymbol{\sigma}, \boldsymbol{\rho}), & \text{if } \boldsymbol{\rho}(\mathcal{X}) < \boldsymbol{\sigma}(\mathcal{X}). \end{cases}
$$

3. Stein equations. To bound the d_2 distance between $\mathcal{L}(\Xi)$ and $\mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, where $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots$ are the measures defined in (2.4) and

$$
d_2\bigl(\mathcal{L}(\Xi), \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\bigr) = \sup_{f \in \mathcal{F}} \frac{1}{s_2(f)} |E[f(\Xi)] - \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)|,
$$

we need to find a Stein equation for the distribution $\mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$. By analogy with the cases of compound Poisson random variable [Barbour, Chen and Loh (1992)] and Poisson process approximation [Barbour and Brown (1992)], a candidate equation is

$$
(3.1) \qquad \sum_{l \ge 1} l \int g(\xi + l\delta_\alpha)\boldsymbol{\pi}_l(d\alpha) - g(\xi)|\xi| = f(\xi) - \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f),
$$

for $\xi \in \mathcal{H}$, where $|\xi|$ is used to denote $\xi(\Gamma)$. Note also that the identity

$$
(3.2) \qquad E\left\{\sum_{l \ge 1} l \int_B g(\Xi + l\delta_\alpha, \alpha)\boldsymbol{\pi}_l(d\alpha) - \int_B g(\Xi, \alpha)\Xi(d\alpha)\right\} = 0
$$

for all bounded measurable g, which is the Palm characterization of Ξ as a compound Poisson point process with distribution $\mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, dovetails neatly with (3.1), when the function g does not depend on its second argument. As we see in Section 3.1, there is a solution to the equation (3.1), but, as in general in the random variable case, the solution has useful uniform bounds only when $\boldsymbol{\pi}(\Gamma)$ is small. In the random variable case, under the condition that $i\theta_i \searrow 0$ as $i \to \infty$, where θ_i is the expected number of clumps of size i, much sharper uniform bounds can be found. The situation here is similar; under the additional condition that $i\mu_i(\alpha) \searrow 0$ for each $\alpha \in \Gamma$, we can derive better bounds. However, to do this, we need to consider an equation which looks slightly different from the Stein equation above, but is also related to the Palm characterization (3.2); see (3.9).

3.1. *The general case.* In this section, we prove two results about the solutions of the Stein equation (3.1) for bounded f, under quite general assumptions. The first gives uniform bounds on the function g_f and its differences at arguments ξ and $\xi + \eta$; it is the analogue of Theorems 1 and 2 of Barbour, Chen and Loh (1992) that relate to the random variable case.

LEMMA 3.1. *For any bounded $f : \mathcal{H} \to \mathbb{R}$ and any choice of measures $\boldsymbol{\pi}_l$, $l \geq 1$, there exists a solution $g = g_f : \mathcal{H} \to \mathbb{R}$ to the Stein equation* (3.1), *which satisfies*

$$\sup_{\xi \in \mathcal{H}} (|\xi| \vee 1)|g_f(\xi)| \leq 2\|f\| e^{\boldsymbol{\pi}(\Gamma)},$$

$$\sup_{\xi, \eta \in \mathcal{H}} (|\xi| \vee 1)|g_f(\xi + \eta) - g_f(\xi)| \leq 2\|f\| e^{\boldsymbol{\pi}(\Gamma)},$$

where $\boldsymbol{\pi}(\Gamma)$ is as in (2.6). *If $f \in \mathcal{F}$, the factor $2\|f\|$ can be replaced by $s_2(f)$.*

PROOF. Let $X, X_1, X_2, \ldots$ be independent point measures of the form $X = J\delta_Y$, such that (J, Y) takes values in $\mathbb{N} \times \Gamma$ and that $P((J, Y) \in \{j\} \times B) = \boldsymbol{\pi}_j(B)/\boldsymbol{\pi}(\Gamma)$, for $j \in \mathbb{N}$ and $B \subset \Gamma$. Note that $E|X| = EJ = \boldsymbol{\mu}(\Gamma)/\boldsymbol{\pi}(\Gamma) < \infty$. By letting $F(\xi) = \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f) - f(\xi)$, equation (3.1) can be rewritten for each fixed ξ with $|\xi| \geq 1$ as

$$g(\xi) = \frac{F(\xi)}{|\xi|} + \frac{\boldsymbol{\pi}(\Gamma)}{|\xi|} E[|X| g(\xi + X)], \tag{3.3}$$

where E is applied only to X. Equation (3.3) can then be solved in $|\xi| \geq 1$ by a recursive argument.

First, let $g_0(\xi) = F(\xi)/|\xi|$, and define $g_1, g_2, \ldots$ successively by

$$g_n(\xi) = \frac{F(\xi)}{|\xi|} + \frac{\boldsymbol{\pi}(\Gamma)}{|\xi|} E[|X_n| g_{n-1}(\xi + X_n)], \tag{3.4}$$

noting that the right-hand side only uses values of g_{n-1} at arguments η with $|\eta| \geq 1$. Then it follows that

$$g_n(\xi) = \frac{F(\xi)}{|\xi|} + \frac{1}{|\xi|} \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{F(\xi + S_r) \prod_{s=1}^{r} |X_s|}{|\xi + S_1| \cdots |\xi + S_r|}\right], \tag{3.5}$$

where $S_l := \sum_{j=1}^{l} X_j$. Hence, letting $\|f\| = \sup_\eta |f(\eta)|$, we have, for any $m \geq 1$,

$$\begin{aligned}
\sum_{n \geq m} |g_n(\xi) - g_{n-1}(\xi)| &\leq \sum_{n \geq m} \frac{1}{|\xi|} \boldsymbol{\pi}(\Gamma)^n E\left[\frac{|F(\xi + S_n)| \prod_{s=1}^{n} |X_s|}{|\xi + S_1| \cdots |\xi + S_n|}\right] \\
&\leq \sum_{n \geq m} \sup_\eta |F(\eta)| \boldsymbol{\pi}(\Gamma)^n E\left[\frac{\prod_{s=1}^{n} |X_s|}{|S_1| \cdots |S_n|}\right] \\
&\leq 2\|f\| \sum_{n \geq m} \frac{\boldsymbol{\pi}(\Gamma)^n}{n!} \leq 2\|f\| e^{\boldsymbol{\pi}(\Gamma)} < \infty,
\end{aligned}$$

where we have used the fact that

$$E\left[\frac{\prod_{s=1}^{n} |X_s|}{|S_1| \cdots |S_n|}\right] = \frac{1}{n!}$$

[see, e.g., Barbour, Holst and Janson (1992), page 179].

Hence g_n converges uniformly on $\mathcal{H} \setminus \{\varnothing\}$ as $n \to \infty$, where $\varnothing$ denotes the zero measure or empty configuration; we denote the limit by g_f, and set $g_f(\varnothing) = 0$. Hence, because $E|X| < \infty$, we have

$$\big|E\big(|X|g_n(\xi + X)\big) - E\big(|X|g_f(\xi + X)\big)\big| \le \|g_n - g_f\| E|X| \to 0$$

as $n \to \infty$; replacing X_n with X in (3.4) and letting $n \to \infty$, it thus follows that g_f satisfies (3.1).

Now, from (3.5), for any ξ with $|\xi| \ge 1$,

$$\begin{aligned}
|\xi||g_n(\xi)| &\le 2\|f\|\left(1 + \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{\prod_{s=1}^r |X_s|}{|\xi + S_1| \cdots |\xi + S_r|}\right]\right) \\
&\le 2\|f\|\left(1 + \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{\prod_{s=1}^r |X_s|}{|S_1| \cdots |S_r|}\right]\right) \\
&= 2\|f\|\left(1 + \sum_{r=1}^{n} \frac{\boldsymbol{\pi}(\Gamma)^r}{r!}\right) \\
&\le 2\|f\| e^{\boldsymbol{\pi}(\Gamma)},
\end{aligned}$$

for all n, so that g_f satisfies $|\xi||g_f(\xi)| \le 2\|f\|e^{\boldsymbol{\pi}(\Gamma)}$. This proves the first inequality.

Now, write

$$a(\xi, X_1, \ldots, X_r) = |\xi||\xi + S_1| \cdots |\xi + S_r|.$$

Then it follows by (3.5) that, for any $\eta \in \mathcal{H}$ with $|\eta| = k$ and for any $|\xi| \ge 1$,

$$\begin{aligned}
&|g_n(\xi + \eta) - g_n(\xi)| \\
&\quad \le \left|\frac{F(\xi + \eta)}{|\xi| + k} - \frac{F(\xi)}{|\xi|}\right| \\
&\qquad + \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\prod_{s=1}^{r} |X_s| \left|\frac{F(\xi + \eta + S_r)}{a(\xi + \eta, X_1, \ldots, X_r)} - \frac{F(\xi + S_r)}{a(\xi, X_1, \ldots, X_r)}\right|\right].
\end{aligned} \tag{3.6}$$

Since $F(\xi) = \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f) - f(\xi)$, we get

$$\begin{aligned}
\left|\frac{F(\xi + \eta)}{|\xi| + k} - \frac{F(\xi)}{|\xi|}\right| &\le \frac{|\xi||f(\xi) - f(\xi + \eta)| + k|f(\xi) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)|}{(|\xi| + k)|\xi|} \\
&\le \frac{2\|f\||\xi| + 2k\|f\|}{(|\xi| + k)|\xi|} = \frac{2\|f\|}{|\xi|},
\end{aligned}$$

and it follows similarly that the remaining part of the right-hand side of (3.6) is bounded by

$$2\|f\| \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{\prod_{s=1}^r |X_s|}{a(\xi, X_1, \ldots, X_r)}\right] \le \frac{2\|f\|}{|\xi|}\big(e^{\boldsymbol{\pi}(\Gamma)} - 1\big).$$

This, together with the previous bound if $\xi = \varnothing$, completes the proof of the second inequality.

Finally, the right-hand side $f(\xi) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)$ of the Stein equation (3.1) is not changed by subtracting $(\inf_{\xi\in\mathcal{H}} f(\xi) + \sup_{\xi\in\mathcal{H}} f(\xi))/2$ from f. Hence, for $f \in \mathcal{F}$, we may take $\|f\| \le s_2(f)/2$. $\square$

The second lemma bounds the differences of the values of the function g_f at arguments which have the same mass.

LEMMA 3.2. *If $f \in \mathcal{F}$ and $\xi, \eta \in \mathcal{H}$ are point configurations with $|\xi| = |\eta| = m$, then*

$$(m \vee 1)|g_f(\xi) - g_f(\eta)| \le s_2(f) d_1(\xi, \eta) e^{\boldsymbol{\pi}(\Gamma)}.$$

PROOF. For $m = 0$, the result is obvious. For $m \ge 1$, it follows from (3.5) that

$$\begin{aligned} m&|g_n(\xi) - g_n(\eta)| \\ &\le |f(\xi) - f(\eta)| + \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{|f(\xi + S_r) - f(\eta + S_r)| \prod_{s=1}^{r} |X_s|}{|\xi + S_1| \cdots |\xi + S_r|}\right]. \end{aligned}$$

Using the definitions of d_1 and $s_2(f)$ given in (2.7) and (2.9), respectively, we get

$$|f(\xi + S_r) - f(\eta + S_r)| \le s_2(f) d_1(\xi + S_r, \eta + S_r) \le s_2(f) d_1(\xi, \eta),$$

and hence, for $m \ge 1$,

$$\begin{aligned} m|g_n(\xi) - g_n(\eta)| &\le s_2(f) d_1(\xi, \eta)\left(1 + \sum_{r=1}^{n} \boldsymbol{\pi}(\Gamma)^r E\left[\frac{\prod_{s=1}^{r} |X_s|}{|S_1| \cdots |S_r|}\right]\right) \\ &\le s_2(f) d_1(\xi, \eta) e^{\boldsymbol{\pi}(\Gamma)}, \end{aligned}$$

for all n, from which the result follows. $\square$

3.2. *The case where $l\mu_l(\alpha) \searrow 0$ for each α.* In this section, we consider an alternative to the Stein equation (3.1), working under the assumption that

$$l\mu_l(\alpha) \searrow 0 \qquad \text{for each } \alpha \in \Gamma. \tag{3.7}$$

Set

$$\boldsymbol{\lambda}_l(d\alpha) = l\boldsymbol{\pi}_l(d\alpha) - (l+1)\boldsymbol{\pi}_{l+1}(d\alpha), \qquad l \ge 1,$$

and let Z be an immigration (in groups)—death process on Γ with immigration intensity measure $\boldsymbol{\lambda}_l$ for groups of size l and with unit per capita death rate. Then Z has equilibrium distribution $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, and its infinitesimal generator is given by

$$(\mathcal{A}h)(\xi) = \sum_{l\ge1} \int_\Gamma [h(\xi + l\delta_\alpha) - h(\xi)]\boldsymbol{\lambda}_l(d\alpha) - \int_\Gamma [h(\xi) - h(\xi - \delta_\alpha)]\xi(d\alpha),$$

where $\xi \in \mathcal{H}$. Here we solve the Stein equation

$$(\mathcal{A}h)(\xi) = f(\xi) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f), \tag{3.8}$$

for $f \in \mathcal{F}$.

If, for each $\alpha \in \Gamma$, we define a function g_α on $\mathcal{H} + \delta_\alpha$ by

$$g_\alpha(\xi + \delta_\alpha) = h(\xi + \delta_\alpha) - h(\xi), \qquad \xi \in \mathcal{H},$$

we have

$$\begin{aligned}
&\sum_{l\geq 1} \int [h(\xi + l\delta_\alpha) - h(\xi)]\boldsymbol{\lambda}_l(d\alpha) \\
&\qquad = \sum_{l\geq 1} \int \sum_{k=1}^{l} \bigl[h(\xi + k\delta_\alpha) - h\bigl(\xi + (k-1)\delta_\alpha\bigr)\bigr]\boldsymbol{\lambda}_l(d\alpha) \\
&\qquad = \sum_{k\geq 1} \int g_\alpha(\xi + k\delta_\alpha) \sum_{l\geq k} \boldsymbol{\lambda}_l(d\alpha) \\
&\qquad = \sum_{k\geq 1} \int g_\alpha(\xi + k\delta_\alpha) k\boldsymbol{\pi}_k(d\alpha),
\end{aligned}$$

and hence (3.8) can be written as

$$\sum_{l\geq 1} l \int g_\alpha(\xi + l\delta_\alpha)\boldsymbol{\pi}_l(d\alpha) - \int g_\alpha(\xi)\xi(d\alpha) = f(\xi) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f). \tag{3.9}$$

Note that this equation is the same as the Stein equation (3.1), except for the index α attached to the functions g_α.

Let P^ξ and E^ξ denote, respectively, the distribution and the expectation of the immigration–death process Z defined above, when $Z(0) = \xi$. We now show that

$$h_f(\xi) = -\int_0^\infty \bigl[E^\xi\bigl[f\bigl(Z(t)\bigr)\bigr] - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)\bigr]\,dt, \tag{3.10}$$

exists, is bounded and satisfies (3.8). The proofs are very similar to the proofs of Propositions 2.1 and 2.3 in Barbour and Brown (1992).

LEMMA 3.3. *For any bounded $f : \mathcal{H} \to \mathbb{R}$, the function $h_f : \mathcal{H} \to \mathbb{R}$ given in* (3.10) *is well defined.*

PROOF. Let Z_0 be an immigration–death process with immigration intensity $\boldsymbol{\lambda}_l$ for groups of size l, and unit per capita death rate, which is empty at time 0. Let D and $\tilde{D}$ be pure death processes with unit per capita death rates and with $D(0) = \xi$ and $\tilde{D}(0) \sim \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, independent of each other and of Z_0. Let $Z = Z_0 + D$, $\tilde{Z} = Z_0 + \tilde{D}$ and

$$\tau = \tau_\xi = \inf\{u \geq 0 : D(u) = \tilde{D}(u) = 0\},$$

so that $Z(t) = \tilde{Z}(t)$ for all $t \geq \tau$, and observe that, for each ξ,

$$E[\tau_\xi] = E\big[\big(|\xi| + |\tilde{D}(0)|\big)^{-1} + \big(|\xi| + |\tilde{D}(0)| - 1\big)^{-1} + \cdots + 1\big]$$
$$\leq E\big[1 + \log\big(|\xi| + |\tilde{D}(0)| + 1\big)\big] < \infty,$$

the finiteness of the expectation following because $|\tilde{D}(0)|$ has a compound Poisson distribution with finite mean.

Now define

$$h_{t,f}(\xi) = -\int_0^t \big[E^\xi f\big(Z(u)\big) - \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)\big]\,du. \tag{3.11}$$

Note that

$$\int_t^\infty \big|E^\xi\big[f\big(Z(t)\big)\big] - E\big[f\big(\tilde{Z}(t)\big)\big]\big|\,dt \leq 2\|f\| \int_t^\infty P[\tau_\xi > s]\,ds,$$

so that, since $E[\tau_\xi] < \infty$, $\lim_{t\to\infty} h_{t,f}(\xi)$ exists and is finite for each ξ. Hence (3.10) is well defined. $\square$

LEMMA 3.4. *For any bounded $f : \mathcal{H} \to \mathbb{R}$, the function h_f, defined in* (3.10), *satisfies the Stein equation* (3.8).

PROOF. The time for the first birth or death in the process Z under P^ξ is exponentially distributed with parameter $q = \xi(\Gamma) + \sum_{l\geq 1}\int_\Gamma \boldsymbol{\lambda}_l(d\alpha)$. Hence, from (3.11),

$$\begin{aligned} h_{t,f}(\xi) = {} & -[f(\xi) - \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)]e^{-qt}t \\ & + \int_0^t q e^{-qu}\Bigg\{ -u[f(\xi) - \mathrm{CP}\,(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)] \\ & \qquad + \sum_{l\geq 1}\int_\Gamma h_{t-u,f}(\xi + l\delta_\alpha)\frac{\boldsymbol{\lambda}_l(d\alpha)}{q} \\ & \qquad + \int_\Gamma h_{t-u,f}(\xi - \delta_\alpha)\frac{\xi(d\alpha)}{q}\Bigg\}\,du. \end{aligned} \tag{3.12}$$

We wish to let $t \to \infty$ on both sides of this equation.

Since, for each fixed ξ,

$$\int_0^\infty \int_\Gamma e^{-qu}\xi(d\alpha)\,du < \infty$$

and since, using the same notation and technique as in the proof of Lemma 3.3, it follows that

$$\big|\mathbb{1}_{\{u\leq t\}} h_{t-u,f}(\xi - \delta_\alpha)\big| \leq 2\|f\| E\left[\sum_{i=1}^{|\xi|+|\tilde{D}(0)|} i^{-1}\right] < \infty,$$

bounded convergence implies that

$$\int_0^t qe^{-qu}\int_\Gamma h_{t-u,f}(\xi-\delta_\alpha)\frac{\xi(d\alpha)}{q}\,du \;\to\; \frac{1}{q}\int_\Gamma h_f(\xi-\delta_\alpha)\xi(d\alpha) \tag{3.13}$$

as $t\to\infty$.

Furthermore, since $\sum_{i=1}^n i^{-1}\le 1+\log n$,

$$\begin{aligned}\left|\mathbb{1}_{\{u\le t\}}h_{t-u,f}(\xi+l\delta_\alpha)\right| &\le 2\|f\|E\left[\sum_{i=1}^{|\xi|+|\tilde D(0)|+l} i^{-1}\right]\\ &\le 2\|f\|E\bigl[1+\log\bigl(|\xi|+l+|\tilde D(0)|\bigr)\bigr]\end{aligned}$$

where the right-hand side does not depend on α or t. Next we need to show that

$$\begin{aligned}&\int_0^\infty\int_\Gamma\sum_{l\ge1}E\bigl[1+\log\bigl(|\xi|+l+|\tilde D(0)|\bigr)\bigr]\boldsymbol{\lambda}_l(d\alpha)e^{-qu}\,du\\ &\quad=q^{-1}\sum_{l\ge1}E\bigl[1+\log\bigl(|\xi|+l+|\tilde D(0)|\bigr)\bigr]\boldsymbol{\lambda}_l(\Gamma)\\ &\quad\le q^{-1}\left(\sum_{l\ge1}\boldsymbol{\lambda}_l(\Gamma)E\bigl\{1+\log\bigl(1+|\xi|+|\tilde D(0)|\bigr)\bigr\}+\sum_{l\ge1}\log(1+l)\boldsymbol{\lambda}_l(\Gamma)\right)\end{aligned}\tag{3.14}$$

is finite. Recall that $\boldsymbol{\lambda}_l(d\alpha)=l\boldsymbol{\pi}_l(d\alpha)-(l+1)\boldsymbol{\pi}_{l+1}(d\alpha)$, so that

$$0\le\sum_{l=1}^L\log(l+1)\boldsymbol{\lambda}_l(\Gamma)\le\sum_{l=1}^L\bigl(\log(l+1)-\log(l)\bigr)l\boldsymbol{\pi}_l(\Gamma)\le\sum_{l=1}^L\boldsymbol{\pi}_l(\Gamma),$$

giving

$$\lim_{L\to\infty}\sum_{l=1}^L\log(l+1)\boldsymbol{\lambda}_l(\Gamma)\le\boldsymbol{\pi}(\Gamma)<\infty.$$

Hence (3.14) is bounded, and by dominated convergence,

$$\int_0^t e^{-qu}\sum_{l\ge1}\int_\Gamma h_{t-u,f}(\xi+l\delta_\alpha)\boldsymbol{\lambda}_l(d\alpha)\,du \;\to\; \frac{1}{q}\sum_{l\ge1}\int_\Gamma h_f(\xi+l\delta_\alpha)\boldsymbol{\lambda}_l(d\alpha), \tag{3.15}$$

as $t\to\infty$. Furthermore, $e^{-qt}t\to0$ and $\int_0^t e^{-qu}qu\,du\to1/q$ as $t\to\infty$, and letting $t\to\infty$ in both sides of (3.12), it follows from (3.13) and (3.15) that

$$\begin{aligned}h_f(\xi)=\frac{1}{q}\Biggl\{&-[f(\xi)-\mathrm{CP}(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2,\dots)(f)]\\ &+\sum_{l\ge1}\int_\Gamma h_f(\xi+l\delta_\alpha)\boldsymbol{\lambda}_l(d\alpha)+\int_\Gamma h_f(\xi-\delta_\alpha)\xi(d\alpha)\Biggr\}.\end{aligned}\tag{3.16}$$

But now recall that $q = \xi(\Gamma) + \sum_{l\ge1}\int_\Gamma \boldsymbol{\lambda}_l(d\alpha)$, so that

$$h_f(\xi)q = \int_\Gamma h_f(\xi)\xi(d\alpha) + h_f(\xi)\sum_{l\ge1}\int_\Gamma \boldsymbol{\lambda}_l(d\alpha),$$

which combined with (3.16) proves the result. $\square$

The next lemma concerns the smoothness of the solution h_f, and of some functions derived from it.

LEMMA 3.5. *If*

$$g_{\alpha f}(\xi+\delta_\alpha) := h_f(\xi+\delta_\alpha) - h_f(\xi), \qquad \xi\in\mathcal{H}, \tag{3.17}$$

where h_f is defined in (3.10) *and $f\in\mathcal{F}$, then*

$$\text{(i)}\qquad \sup_{\alpha\in\Gamma,\xi\in\mathcal{H}} |g_{\alpha f}(\xi+\delta_\alpha)| \le s_2(f)\{1\wedge 1.65\lambda_1^{-1/2}\},$$

$$\begin{aligned}\text{(ii)}\qquad \sup_{\alpha,\beta\in\Gamma,\xi\in\mathcal{H}} &|g_{\alpha f}(\xi+\delta_\alpha+\delta_\beta) - g_{\alpha f}(\xi+\delta_\alpha)|\\ &\le s_2(f)\left\{1\wedge \frac{2}{\lambda_1}\big(1+2\log^+(\lambda_1/2)\big)\right\},\end{aligned}$$

where $\lambda_1 := \boldsymbol{\lambda}_1(\Gamma)$.

PROOF. The proof is modeled on that in Barbour, Holst and Janson [(1992), Lemmas 10.2.3 and 10.2.5]. We begin with (i). Let Z_0 and D be defined as in the proof of Lemma 3.3, and let E be an independent standard exponential random variable. Then $Z_1(t) = Z_0(t) + D(t) \sim P^\xi$ and $Z_2(t) = Z_0(t) + D(t) + \delta_\alpha 1\{E > t\} \sim P^{\xi+\delta_\alpha}$. Then, by the definition of $g_{\alpha f}$ and h given in (3.17) and (3.10), respectively,

$$\begin{aligned} g_{\alpha f}(\xi+\delta_\alpha) &= h_f(\xi+\delta_\alpha) - h_f(\xi)\\ &= \int_0^\infty \{E^\xi[f(Z(t))] - E^{\xi+\delta_\alpha}[f(Z(t))]\}\,dt\\ &= \int_0^\infty e^{-t}\sum_{\eta\le\xi} P\big(D(t)=\eta\big)\\ &\qquad\qquad \times E\big[f\big(Z_0(t)+\eta\big) - f\big(Z_0(t)+\eta+\delta_\alpha\big)\big]\,dt, \end{aligned}\tag{3.18}$$

where $\eta\in\mathcal{H}$ is a possible outcome of what remains of the point configuration $D(0)=\xi$ at time t.

By the definition of $s_2(f)$, and since $d_1(\xi_1,\xi_2) = 1$ if $|\xi_1|\ne|\xi_2|$, it follows that

$$\begin{aligned} &\big|f\big(Z_0(t)+\eta\big) - f\big(Z_0(t)+\eta+\delta_\alpha\big)\big|\\ &\qquad \le s_2(f)d_1\big(Z_0(t)+\eta, Z_0(t)+\eta+\delta_\alpha\big) = s_2(f), \end{aligned}\tag{3.19}$$

for all $\alpha \in \Gamma$, $\xi \in \mathcal{H}$, and thus one of the estimate in (i) is immediate from (3.18):

$$\sup_{\alpha \in \Gamma, \xi \in \mathcal{H}} |g_{\alpha f}(\xi + \delta_\alpha)| \le s_2(f). \tag{3.20}$$

Let $Y_j(t)$ denote the number of individuals which have immigrated in groups of size j and are still alive at time t. Then $|Z_0(t)| = \sum_{j \ge 1} Y_j(t)$ and $Y_1(t) \sim \text{Po}(\lambda_{1,t})$ where

$$\lambda_{1,t} = \int_0^t \boldsymbol{\lambda}_1(\Gamma) e^{-u}\, du = (1 - e^{-t})\lambda_1. \tag{3.21}$$

Fix t and let $Y_1 = Y_1(t)$, $Y^* = \sum_{j \ge 2} Y_j(t)$, and $p(i) = P(Y_1 = i)$. Furthermore, write

$$\begin{aligned} f_0 &= f(Z_0 + \eta); & f_\alpha &= f(Z_0 + \eta + \delta_\alpha), \\ f_\beta &= f(Z_0 + \eta + \delta_\beta); & f_{\alpha\beta} &= f(Z_0 + \eta + \delta_\alpha + \delta_\beta). \end{aligned}$$

Then

$$\begin{aligned} E[f_0 - f_\alpha] &= \sum_{k \ge 0} E[f_0 - f_\alpha \,|\, Y_1 = k] p(k) \\ &= E[f_0 \,|\, Y_1 = 0] p(0) \\ &\quad + \sum_{k \ge 0} \{E[f_0 \,|\, Y_1 = k+1] p(k+1) - E[f_\alpha \,|\, Y_1 = k] p(k)\}. \end{aligned} \tag{3.22}$$

As in the proof of Lemma 3.1, we may subtract $(\inf_{\xi \in \mathcal{H}} f(\xi) + \sup_{\xi \in \mathcal{H}} f(\xi))/2$ from f, and take $\sup_{\xi \in \mathcal{H}} |f(\xi)| \le s_2(f)/2$ if $f \in \mathcal{F}$. Hence the first part of the right-hand side of (3.22) is bounded as

$$\big|E[f_0 \,|\, Y_1 = 0]\big| p(0) \le p(0) \frac{s_2(f)}{2}. \tag{3.23}$$

Since the positions of the Y_1 points follow a Poisson process with intensity measure $(1 - e^{-t})\boldsymbol{\lambda}_1$, it follows that

$$E[f_0 \,|\, Y_1 = k+1] = \int_\Gamma E[f_\beta \,|\, Y_1 = k] \frac{\boldsymbol{\lambda}_1(d\beta)}{\boldsymbol{\lambda}_1(\Gamma)}.$$

Furthermore, by the definition of $s_2(f)$ and d_1, we have

$$|f(\xi + \delta_\beta) - f(\xi + \delta_\alpha)| \le s_2(f) d_1(\xi + \delta_\beta, \xi + \delta_\alpha) \le \frac{s_2(f)}{|\xi| + 1},$$

for any $\xi \in \mathcal{H}$. This will be used to bound the terms of the sum in (3.22):

$$\begin{aligned}
&\big|E[f_0 \,|\, Y_1 = k+1]p(k+1) - E[f_\alpha \,|\, Y_1 = k]p(k)\big| \\
&\quad = \big|E[f_0 \,|\, Y_1 = k+1]\{p(k+1) - p(k)\} \\
&\qquad - \big\{E[f_\alpha \,|\, Y_1 = k] - E[f_0 \,|\, Y_1 = k+1]\big\}p(k)\big| \\
&\quad \le |p(k+1) - p(k)|\frac{s_2(f)}{2} \qquad (3.24)\\
&\qquad + p(k)\int_\Gamma E\big[|f_\beta - f_\alpha|\big|Y_1 = k\big]\frac{\boldsymbol{\lambda}_1(d\beta)}{\boldsymbol{\lambda}_1(\Gamma)} \\
&\quad \le s_2(f)\left\{\frac{|p(k+1) - p(k)|}{2} + p(k)E[(k + Y^* + |\eta| + 1)^{-1}]\right\}.
\end{aligned}$$

Inserting (3.23) and (3.24) in (3.22) yields

$$\begin{aligned}
&|E[f_0 - f_\alpha]| \\
&\quad \le \frac{s_2(f)}{2}\Bigg(p(0) + \sum_{k\ge 0}|p(k+1) - p(k)| \\
&\qquad\qquad + 2\sum_{k\ge 0}E[(k + Y^* + 1 + |\eta|)^{-1}]p(k)\Bigg) \\
&\quad \le \frac{s_2(f)}{2}\left(p(0) + 2E[(Y_1 + 1)^{-1}] + \sum_{k\ge 0}|p(k+1) - p(k)|\right).
\end{aligned}$$

Since $p(i) = P(Y_1 = i)$ and $\mathcal{L}(Y_1) = \operatorname{Po}(\lambda_{1,t})$, where $\lambda_{1,t}$ is defined in (3.21), it follows that

$$\begin{aligned}
|E[f_0 - f_\alpha]| &\le s_2(f)\left(\max_{k\ge 0} p(k) + \frac{1 - e^{-\lambda_{1,t}}}{\lambda_{1,t}}\right) \\
&\le s_2(f)\left((2e\lambda_{1,t})^{-1/2} + \frac{1 - e^{-\lambda_{1,t}}}{\lambda_{1,t}}\right),
\end{aligned}$$

where the last inequality follows from Proposition A.2.7 in Barbour, Holst and Janson (1992).

We now have two possible bounds on $|E[f_0 - f_\alpha]|$:

$$|E[f_0 - f_\alpha]| \le \begin{cases} s_2(f), \\ s_2(f)\left((2e\lambda_{1,t})^{-1/2} + \dfrac{1 - e^{-\lambda_{1,t}}}{\lambda_{1,t}}\right), \end{cases}$$

neither of which depends on α or η. Choose t_1 such that $e^{-t_1} = 1 - \lambda_1^{-1}$. Then $\lambda_{1,t} = \lambda_1(1-e^{-t}) > \lambda_1(1-e^{-t_1}) = 1$ for $t > t_1$, and we get

$$\begin{aligned}
&|g_{\alpha f}(\xi+\delta_\alpha)| \\
&\quad \le \int_0^\infty e^{-t} \sum_{\eta\le\xi} P\big(D(t)=\eta\big)\big|E\big[f\big(Z_0(t)+\eta\big) - f\big(Z_0(t)+\eta+\delta_\alpha\big)\big]\big|\,dt \qquad (3.25)\\
&\quad \le s_2(f)\int_0^{t_1} e^{-t}\,dt + s_2(f)\int_{t_1}^\infty e^{-t}\big((2e\lambda_{1,t})^{-1/2} + (\lambda_{1,t})^{-1}\big)\,dt.
\end{aligned}$$

As in Barbour, Holst and Janson [(1992), page 222], computation of the integrals in (3.25) yields

$$|g_{\alpha f}(\xi+\delta_\alpha)| \le s_2(f)\frac{1.65}{\sqrt{\lambda_1}},$$

for all $\alpha \in \Gamma, \xi \in \mathcal{H}$, which together with (3.20) proves (i).

For the proof of part (ii), use a coupling much as in Part (i) to provide an expression for $g_{\alpha f}(\xi+\delta_\alpha+\delta_\beta) - g_{\alpha f}(\xi+\delta_\alpha)$, but now with two independent exponential random variables E_1 and E_2 and four Z-processes $Z_0 + D$, $Z_0 + D + \delta_\alpha 1\{E_1 > t\}$, $Z_0 + D + \delta_\beta 1\{E_2 > t\}$ and $Z_0 + D + \delta_\alpha 1\{E_1 > t\} + \delta_\beta 1\{E_2 > t\}$. This, with the earlier notation, yields the formula

$$\begin{aligned}
&g_{\alpha f}(\xi+\delta_\alpha+\delta_\beta) - g_{\alpha f}(\xi+\delta_\alpha)\\
&\qquad = h_f(\xi+\delta_\alpha+\delta_\beta) - h_f(\xi+\delta_\beta) - h_f(\xi+\delta_\alpha) + h_f(\xi) \qquad (3.26)\\
&\qquad = -\int_0^\infty e^{-2t}\sum_{\eta\le\xi} P\big(D(t)=\eta\big)E[f_{\alpha\beta} - f_\alpha - f_\beta + f_0]\,dt.
\end{aligned}$$

Since $E[|f_{\alpha\beta} - f_\alpha - f_\beta + f_0|] \le 2s_2(f)$ by (3.19), we get one of the bounds in (ii) directly:

$$|g_{\alpha f}(\xi+\delta_\alpha+\delta_\beta) - g_{\alpha f}(\xi+\delta_\alpha)| \le s_2(f). \qquad (3.27)$$

To get a second bound, we rewrite the integrand in the form

$$\begin{aligned}
&E[f_{\alpha\beta} - f_\alpha - f_\beta + f_0]\\
&\quad = \sum_{k\ge 0} E[f_{\alpha\beta} - f_\alpha - f_\beta + f_0 \mid Y_1 = k]p(k)\\
&\quad = -p(0)E[f_\alpha + f_\beta - f_0 \mid Y_1 = 0] + p(1)E[f_0 \mid Y_1 = 1]\\
&\qquad + \sum_{k\ge 0}\big\{p(k)E[f_{\alpha\beta} \mid Y_1 = k] - p(k+1)E[f_\beta + f_\alpha \mid Y_1 = k+1]\\
&\qquad\qquad + p(k+2)E[f_0 \mid Y_1 = k+2]\big\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k\ge 0}\Big\{\{p(k+2)-2p(k+1)+p(k)\}E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta \,|\, Y_1=k+1\big] \\
&\qquad\qquad + p(k)\big\{E[f_{\alpha\beta}\,|\,Y_1=k]-E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=k+1\big]\big\} \\
&\qquad\quad + p(k+2)\big\{E[f_0\,|\,Y_1=k+2]-E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=k+1\big]\big\}\Big\} \\
&\quad + \big(p(1)-2p(0)\big)E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=0\big] \\
&\quad + p(1)\big\{E[f_0\,|\,Y_1=1]-E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=0\big]\big\} \\
&\quad + p(0)E[f_0\,|\,Y_1=0] \\
&= \sum_{k\ge 0}\Big\{\big(p(k+1)-2p(k)+p(k-1)\big)E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=k\big] \\
&\qquad + p(k)\big\{E[f_{\alpha\beta}\,|\,Y_1=k]-E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=k+1\big]\big\} \\
&\qquad + p(k+1)\big\{E[f_0\,|\,Y_1=k+1]-E\big[\tfrac12 f_\alpha+\tfrac12 f_\beta\,|\,Y_1=k\big]\big\}\Big\} \\
&\qquad + p(0)E[f_0\,|\,Y_1=0],
\end{aligned}
$$

where $p(k):=0$ for $k<0$. Using the same technique as in (i), it thus follows that

$$
\begin{aligned}
&|E[f_{\alpha\beta}-f_\alpha-f_\beta+f_0]| \\
&\quad \le \frac{s_2(f)}{2}\left\{p(0)+\sum_{k\ge 0}|p(k+1)-2p(k)+p(k-1)|\right\} \\
&\qquad + s_2(f)E\sum_{k\ge 0}\left\{\frac{p(k)}{(k+2+|\eta|+Y^*)}+\frac{p(k+1)}{(k+1+|\eta|+Y^*)}\right\} \\
&\quad \le s_2(f)\left\{\tfrac12\sum_{k\ge 0}|p(k)-2p(k-1)+p(k-2)|+3E[(Y_1+1)^{-1}]\right\}.
\end{aligned}
$$

Recall that $p(i)=P(Y_1=i)$ and that $\mathcal{L}(Y_1)=\mathrm{Po}\,(\lambda_{1,t})$. Hence

$$
\begin{aligned}
&\sum_{k\ge 0}|p(k)-2p(k-1)+p(k-2)| \\
&\qquad = \sum_{k\ge 0}p(k)\left|\left(1-\frac{k}{\lambda_{1,t}}\right)^2-\frac{k}{\lambda_{1,t}^2}\right| \le \lambda_{1,t}^{-2}\{\mathrm{Var}\,Y_1+EY_1\}=2\lambda_{1,t}^{-1},
\end{aligned}
$$

and we get

$$
|E[f_{\alpha\beta}-f_\beta-f_\alpha+f_0]| \le s_2(f)\left(\left\{\frac{1}{\lambda_{1,t}}\right\}+\frac{3(1-e^{-\lambda_{1,t}})}{\lambda_{1,t}}\right). \tag{3.28}
$$

Combining (3.26) with (3.27) and (3.28) leads to

$$\begin{aligned}|g_{\alpha f}(\xi+\delta_\beta+\delta_\alpha)-g_{\alpha f}(\xi+\delta_\alpha)| &\le s_2(f)\left\{\int_0^{t_1} 2e^{-2t}\,dt+\int_{t_1}^{\infty}\frac{4e^{-2t}}{\lambda_{1,t}}\,dt\right\}\\ &\le \frac{2}{\lambda_1}\left(1+2\log^+\left(\frac{\lambda_1}{2}\right)\right)s_2(f),\end{aligned}$$

for all $\lambda_1 \ge 2$, and the proof is complete. □

REMARK. When compared with Lemma 3.1, the conclusion of Lemma 3.5 is important because of the factors $\lambda_1^{-1/2}$ and λ_1^{-1} appearing in the bounds. If both λ_1 and $\boldsymbol{\pi}(\Gamma)$ are large, as is often the case when Γ represents a long time interval, the bounds in Lemma 3.1 are made large by the factor $e^{\boldsymbol{\pi}(\Gamma)}$, whereas those of Lemma 3.5 are small for large λ_1.

For Poisson process approximation, the uniform bound in Lemma 3.5 (ii) can be replaced by a nonuniform bound of order $O(\lambda^{-1}+(|\xi|+1)^{-1})$, by means of Theorem 5.1 in Brown and Xia (2000). It is possible that something in the same spirit could also be done for compound Poisson process approximation.

For configurations ξ and η with the same number of points, there are slightly different results.

LEMMA 3.6. *If $\xi,\eta\in\mathcal{H}$ are point configurations with $|\xi|=|\eta|=m$, then*:

(i)

$$\begin{aligned}&|g_{\alpha f}(\xi+\delta_\alpha)-g_{\alpha f}(\eta+\delta_\alpha)|\\ &\qquad\le 2s_2(f)d_1(\xi,\eta)\min\left\{1, 2m\lambda_1^{-1}(1-e^{-\lambda_1})\left(1+\log\left[1+\frac{\lambda_1}{m+1}\right]\right)\right\};\end{aligned}$$

(ii)

$$\begin{aligned}&|g_{\alpha f}(\xi+s\delta_\alpha)-g_{\beta f}(\xi+s\delta_\beta)|\\ &\qquad\le s_2(f)d_0(\alpha,\beta)\min\left\{1, (2s-1)\lambda_1^{-1}(1-e^{-\lambda_1})\left(1+\log\left[1+\frac{\lambda_1}{m+1}\right]\right)\right\}.\end{aligned}$$

PROOF. For part (i), let Z_0 be an immigration–death process with the usual parameters, and with $Z_0(0)=\varnothing$. Furthermore, let D_1 and D_2 be pure death processes with $D_1(0)=\xi$ and $D_2(0)=\eta$, coupled so that pairs (ξ_i,η_i) of individuals whose indices are matched in the coupling definition of $d_1(\xi,\eta)$, given in (2.8), have identical lifetimes T_i, and let E be an independent standard exponential random variable. Then, from (3.17) and (3.10), and defining $Z_1 = D_1 + Z_0$, $Z_2 = D_2 + Z_0$, we can argue in the usual way to obtain

$$
\begin{aligned}
&|g_{\alpha f}(\xi + \delta_\alpha) - g_{\alpha f}(\eta + \delta_\alpha)| \\
&\quad = |h_f(\xi + \delta_\alpha) - h_f(\xi) - h_f(\eta + \delta_\alpha) + h_f(\eta)| \\
&\quad = \biggl| \int_0^\infty E^{\xi + \delta_\alpha}\bigl[f\bigl(Z(t)\bigr)\bigr] - E^{\xi}\bigl[f\bigl(Z(t)\bigr)\bigr] \\
&\qquad\qquad - E^{\eta + \delta_\alpha}\bigl[f\bigl(Z(t)\bigr)\bigr] + E^{\eta}\bigl[f\bigl(Z(t)\bigr)\bigr]\, dt \biggr| \\
&\quad = \biggl| \int_0^\infty E\bigl[f\bigl(Z_1(t) + \delta_\alpha I\{E > t\}\bigr) - f\bigl(Z_1(t)\bigr) \\
&\qquad\qquad - f\bigl(Z_2(t) + \delta_\alpha I\{E > t\}\bigr) + f\bigl(Z_2(t)\bigr)\bigr]\, dt \biggr| \\
&\quad = \biggl| \int_0^\infty e^{-t} E\bigl[f\bigl(Z_1(t) + \delta_\alpha\bigr) - f\bigl(Z_1(t)\bigr) \\
&\qquad\qquad - f\bigl(Z_2(t) + \delta_\alpha\bigr) + f\bigl(Z_2(t)\bigr)\bigr]\, dt \biggr| \\
&\quad \le \int_0^\infty e^{-t} \bigl\{E\bigl|f\bigl(Z_1(t) + \delta_\alpha\bigr) - f\bigl(Z_2(t) + \delta_\alpha\bigr)\bigr| \\
&\qquad\qquad + E\bigl|f\bigl(Z_1(t)\bigr) - f\bigl(Z_2(t)\bigr)\bigr|\bigr\}\, dt \\
&\quad \le s_2(f) \int_0^\infty 2e^{-t} E\bigl[d_1\bigl(Z_1(t), Z_2(t)\bigr) 1\{|D_1(t)| \ge 1\}\bigr]\, dt.
\end{aligned}
\tag{3.29}
$$

By the coupling of D_1 and D_2,

$$
\begin{aligned}
&E\bigl\{d_1\bigl(Z_1(t), Z_2(t)\bigr) I\{|D_1(t)| \ge 1\}\bigr\} \\
&\quad = E\biggl\{ \frac{I\{|D_1(t)| \ge 1\} \sum_{i=1}^m d_0(\xi_i, \eta_i) 1\{T_i > t\}}{|Z_0(t)| + \sum_{i=1}^m 1\{T_i > t\}} \biggr\} \\
&\quad = \sum_{r=1}^m \sum_{j \ge 0} P\bigl(|D_1(t)| = r\bigr) P\bigl(|Z_0(t)| = j\bigr) d_1(\xi, \eta) \frac{r}{r + j} \\
&\quad \le d_1(\xi, \eta) \min\Bigl(1, 2mE\bigl\{\bigl(1 + |D_1(t) + Z_0(t)|\bigr)^{-1}\bigr\}\Bigr).
\end{aligned}
$$

The final expectation is bounded by first replacing $Z_0(t)$ by the smaller $Y_1(t) \sim \mathrm{Po}\,(\lambda_{1,t})$ as in (3.21), and then applying Barbour, Holst and Janson [(1992), proof of Lemma 10.2.1]. Substituting this into (3.29) gives part (i).

The argument for Part (ii) is of very similar structure, with pure death processes D_1 and D_2 starting with configurations $D_1(0) = (s-1)\delta_\alpha$ and $D_2(0) = (s-1)\delta_\beta$, and with an additional independent D_0 starting with $D_0(0) = \xi$. This leads to the estimate

$$\begin{aligned}
&|g_{\alpha f}(\xi + s\delta_\alpha) - g_{\beta f}(\xi + s\delta_\beta)| \\
&\quad = \left| \int_0^\infty e^{-t} E\bigl[f\bigl(Z_1(t) + \delta_\alpha\bigr) - f\bigl(Z_1(t)\bigr) - f\bigl(Z_2(t) + \delta_\beta\bigr) + f\bigl(Z_2(t)\bigr)\bigr]\, dt \right| \\
&\quad \le s_2(f) \int_0^\infty e^{-t} d_0(\alpha, \beta)(2s - 1) E\bigl\{\bigl(1 + |Z_0(t) + D_0(t)|\bigr)^{-1}\bigr\},
\end{aligned}$$

where $Z_1 = Z_0 + D_0 + D_1$ and $Z_2 = Z_0 + D_0 + D_2$. □

REMARK. Combining Lemmas 3.5 and 3.6, it follows that, if $\xi, \eta \in \mathcal{H}$, with $|\xi| \le |\eta|$, and if the condition (3.7) holds, then

$$\begin{aligned}
&|g_{\alpha f}(\xi + \delta_\alpha) - g_{\alpha f}(\eta + \delta_\alpha)| \\
&\quad \le s_2(f)\biggl(|\xi(\Gamma) - \eta(\Gamma)| \min\biggl\{1, 2\lambda_1^{-1}\biggl(1 + 2\log^+ \frac{\lambda_1}{2}\biggr)\biggr\} \\
&\qquad + 2\hat{d}_1(\xi, \eta) \min\biggl\{1, 2\xi(\Gamma)\lambda_1^{-1}(1 - e^{-\lambda_1}) \\
&\qquad\qquad \times \biggl(1 + \log\biggl[1 + \frac{\lambda_1}{1 + \xi(\Gamma)}\biggr]\biggr)\biggr\}\biggr),
\end{aligned} \tag{3.30}$$

where $\hat{d}_1(\xi, \eta)$ is as in (2.11). In general, one still has the estimate

$$|g_f(\xi) - g_f(\eta)| \le s_2(f) e^{\boldsymbol{\pi}(\Gamma)} \frac{\mathbb{1}_{\{\xi(\Gamma) \ne \eta(\Gamma)\}} + \hat{d}_1(\xi, \eta)}{\min(\xi(\Gamma), \eta(\Gamma)) \vee 1}, \tag{3.31}$$

from Lemmas 3.1 and 3.2.

4. Process approximation. In order to derive a bound on the d_2 distance between the point process Ξ and a compound Poisson process Π with $\mathcal{L}(\Pi) = \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, we use either the Stein equation (3.1), which can be solved whatever the behaviour of $\mu_i(\alpha)$, $i \ge 1$, or Equation (3.9) if the condition (3.7) is satisfied. By letting $g_{\alpha f} = g_f$ for all $\alpha \in \Gamma$ in the general case, both of the equations (3.1) and (3.9) can be written as

$$\sum_{l \ge 1} l \int g_{\alpha f}(\xi + l\delta_\alpha)\boldsymbol{\pi}_l(d\alpha) - \int g_{\alpha f}(\xi)\xi(d\alpha) = f(\xi) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f),$$

and hence, by the definition of the d_2 metric given in (2.10),

$$\begin{aligned}
&d_2\bigl(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\bigr) \\
&\quad = \sup_{f \in \mathcal{F}} \frac{1}{s_2(f)} |E[f(\Xi)] - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f)| \\
&\quad = \sup_{f \in \mathcal{F}} \frac{1}{s_2(f)} \left| E\left[\sum_{l \ge 1} l \int g_{\alpha f}(\Xi + l\delta_\alpha)\boldsymbol{\pi}_l(d\alpha) - \int g_{\alpha f}(\Xi)\Xi(d\alpha)\right]\right|.
\end{aligned} \tag{4.1}$$

The following lemma shows how this latter quantity can be bounded.

LEMMA 4.1. *Let Ξ be a finite point process on Γ with mean measure $\boldsymbol{\mu}$, decomposed by $\mathcal{N}$ as in* (2.1), *and let the $\boldsymbol{\pi}_i$ be defined as in* (2.4). *For each $\alpha \in \Gamma$ and $i \geq 1$, suppose that $\Theta^r_{\alpha i}$ and $\Phi_{\alpha i}$ are point processes defined on a common probability space with distributions*

$$\mathcal{L}(\Theta^r_{\alpha i}) = P^\alpha\big(\Xi^{r,\alpha} \in \cdot \mid \Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\big), \qquad \mathcal{L}(\Phi_{\alpha i}) = \mathcal{L}(\Xi).$$

Then, for any family of bounded measurable functions $\mathcal{g}: \Gamma \times \mathcal{H} \to \mathbb{R} : (\alpha, \xi) \mapsto g_\alpha(\xi)$, we have

$$\left|\sum_{l\geq 1} l \int E[g_\alpha(\Xi + l\delta_\alpha)]\boldsymbol{\pi}_l(d\alpha) - E\left\{\int g_\alpha(\Xi)\Xi(d\alpha)\right\}\right| \leq b_{\mathcal{g}}(\Xi, \mathcal{N}) + c_{\mathcal{g}}(\Xi, \mathcal{N}), \tag{4.2}$$

where

$$b_{\mathcal{g}}(\Xi, \mathcal{N}) := \left|\int_\Gamma E^\alpha\big\{g_\alpha(\Xi) - g_\alpha\big(\Xi^{r,\alpha} + \big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\big)\delta_\alpha\big)\big\}\boldsymbol{\mu}(d\alpha)\right|,$$

$$c_{\mathcal{g}}(\Xi, \mathcal{N}) := \sum_{i\geq 1} i \int_\Gamma \big|E\big\{g_\alpha(\Theta^r_{\alpha i} + i\delta_\alpha) - g_\alpha(\Phi_{\alpha i} + i\delta_\alpha)\big\}\big|\boldsymbol{\pi}_i(d\alpha).$$

PROOF. Letting

$$B = \int_\Gamma E^\alpha g_\alpha(\Xi)\boldsymbol{\mu}(d\alpha) - \int_\Gamma E^\alpha g_\alpha\big(\Xi^{r,\alpha} + \big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\big)\delta_\alpha\big)\boldsymbol{\mu}(d\alpha), \tag{4.3}$$

$$\begin{aligned} C = {} & \int_\Gamma E^\alpha g_\alpha\big(\Xi^{r,\alpha} + \big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\big)\delta_\alpha\big)\boldsymbol{\mu}(d\alpha) \\ & - \sum_{i\geq 1} i \int_\Gamma E[g_\alpha(\Xi + i\delta_\alpha)]\boldsymbol{\pi}_i(d\alpha), \end{aligned} \tag{4.4}$$

we immediately get

$$\left|E\left\{\int_\Gamma g_\alpha(\Xi)\Xi(d\alpha)\right\} - \sum_{i\geq 1} i \int_\Gamma E[g_\alpha(\Xi + i\delta_\alpha)]\boldsymbol{\pi}_i(d\alpha)\right| \leq |B| + |C|, \tag{4.5}$$

by (2.2). That $|B| \leq b_{\mathcal{g}}(\Xi, \mathcal{N})$ is clear. For $|C|$, write

$$\begin{aligned} & \int_\Gamma E^\alpha g_\alpha\big(\Xi^{r,\alpha} + \big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\big)\delta_\alpha\big)\boldsymbol{\mu}(d\alpha) \\ & \quad = \sum_{i\geq 1} \int_\Gamma E^\alpha\big\{g_\alpha(\Xi^{r,\alpha} + i\delta_\alpha) \mid \Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\big\} \\ & \qquad \times P^\alpha\big(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\big)\boldsymbol{\mu}(d\alpha) \\ & \quad = \sum_{i\geq 1} \int_\Gamma E g_\alpha(\Theta^r_{\alpha i} + i\delta_\alpha) i\boldsymbol{\pi}_i(d\alpha), \end{aligned} \tag{4.6}$$

by (2.4), from which $|C| \leq c_{\mathcal{g}}(\Xi, \mathcal{N})$ follows immediately. $\square$

Lemma 4.1 and (4.1) can now be combined with the estimates (3.31) and (3.30) to give bounds in d_2-distance between $\mathcal{L}(\Xi)$ and $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$. Here, we give two slightly simplified and weakened versions of the result.

THEOREM 4.2. *Under the assumptions of Lemma* 4.1, *it follows that*

$$\begin{aligned} &d_2\big(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\big) \\ &\quad \le e^{\boldsymbol{\pi}(\Gamma)}\biggl\{ \int_\Gamma E^\alpha d_1\big(\Xi^{s,\alpha}, \Xi^{s,\alpha}(\Gamma)\delta_\alpha\big)\boldsymbol{\mu}(d\alpha) \\ &\qquad\qquad + \sum_{i\ge 1} i \int_\Gamma \big(P\big(\Theta^r_{\alpha i}(\Gamma) \ne \Phi_{\alpha i}(\Gamma)\big) + E\hat{d}_1(\Theta^r_{\alpha i}, \Phi_{\alpha i})\big)\biggr\}\boldsymbol{\pi}_i(d\alpha). \end{aligned}$$

If, *in addition*, *Condition* (3.7) *holds*, *then*

$$d_2\big(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\big) \le 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1)\{T_1 + T_2\},$$

where

$$\begin{aligned} T_1 &:= 2\int_\Gamma E^\alpha \biggl(\int_\Gamma \Xi^{s,\alpha}(d\beta) d_0(\alpha, \beta)\biggr)\boldsymbol{\mu}(d\alpha), \\ T_2 &:= \sum_{i\ge 1} i \int_\Gamma E\big(|\Theta^r_{\alpha i}(\Gamma) - \Phi_{\alpha i}(\Gamma)| \\ &\qquad + 2\min\big(\Theta^r_{\alpha i}(\Gamma), \Phi_{\alpha i}(\Gamma)\big)\hat{d}_1(\Theta^r_{\alpha i}, \Phi_{\alpha i})\big)\boldsymbol{\pi}_i(d\alpha). \end{aligned}$$

PROOF. First, we simplify the bound in (3.30) to the form

$$\begin{aligned} &|g_{\alpha f}(\xi + \delta_\alpha) - g_{\alpha f}(\eta + \delta_\alpha)| \\ &\quad \le s_2(f)\big(|\xi(\Gamma) - \eta(\Gamma)| 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1) \\ &\qquad\qquad + 2\hat{d}_1(\xi, \eta) 2\xi(\Gamma)\lambda_1^{-1}(1 + \log^+ \lambda_1)\big) \\ &\quad \le s_2(f) 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1)\big(|\xi(\Gamma) - \eta(\Gamma)| + 2\hat{d}_1(\xi, \eta)\xi(\Gamma)\big), \end{aligned} \tag{4.7}$$

valid for $\xi, \eta \in \mathcal{H}$ with $|\xi| \le |\eta|$ if Condition (3.7) holds. To bound $b_{\mathcal{G}}(\Xi, \mathcal{N})$ in Lemma 4.1, note that $\Xi(\Gamma) = \Xi^{r,\alpha}(\Gamma) + \Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}$, so that by (3.31), (4.7) and the definition of $\hat{d}_1$, given in (2.11),

$$\frac{b_{\mathcal{G}}(\Xi, \mathcal{N})}{s_2(f)} \le \begin{cases} 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1) T_1, & \text{if (3.7) holds,} \\ e^{\boldsymbol{\pi}(\Gamma)} \displaystyle\int_\Gamma E^\alpha d_1\big(\Xi^{s,\alpha}, \Xi^{s,\alpha}(\Gamma)\delta_\alpha\big)\boldsymbol{\mu}(d\alpha), & \text{otherwise.} \end{cases}$$

Bounding $c_{\mathcal{G}}(\Xi, \mathcal{N})$ by applying (3.31) and (4.7) again, and combining these bounds with (4.1), completes the proof. □

In both versions, there are two components to the bounds, corresponding to the elements $b_{\mathcal{G}}(\Xi, \mathcal{N})$ and $c_{\mathcal{G}}(\Xi, \mathcal{N})$ of Lemma 4.1. The first measures the effect of shifting the points of $\Xi^{s,\alpha}$ onto α; the second measures the closeness of $\Theta^r_{\alpha i}$ to $\Phi_{\alpha i}$. Thus, for approximation to be good, the decomposition $\mathcal{N}$ of Ξ should be chosen in such a way that the points of $\Xi^{s,\alpha}$ are close to α and that the P^α-distribution of $\Xi^{r,\alpha}$ is not very different from $\mathcal{L}(\Xi)$.

REMARK 1. Estimation of the final term can at times be helped by the observation that, for $\xi = \xi_1 + \xi_2$ and $\eta = \eta_1 + \eta_2$ in $\mathcal{H}$, then

$$\begin{aligned}(4.8)\quad &|\xi(\Gamma) - \eta(\Gamma)| + 2\min\{\xi(\Gamma), \eta(\Gamma)\}\hat{d}_1(\xi, \eta)\\ &\le \xi_2(\Gamma) + \eta_2(\Gamma) + |\xi_1(\Gamma) - \eta_1(\Gamma)| + 2\min\{\xi_1(\Gamma), \eta_1(\Gamma)\}\hat{d}_1(\xi_1, \eta_1);\end{aligned}$$

see Example 5.3. It may also be useful, as in Examples 5.2–5.4, to condition on the extra information contained in some random element Y, constructing $\Theta^{r,y}_{\alpha i}$ such that

$$\mathcal{L}(\Theta^{r,y}_{\alpha i}) = P^\alpha\big(\Xi^{r,\alpha} \in \cdot \mid \Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i,\ Y = y\big)$$

to match some $\Phi^y_{\alpha i}$ with $\mathcal{L}(\Phi^y_{\alpha i}) = \mathcal{L}(\Xi)$. Note also that

$$(4.9)\quad \big(|\Theta^r_{\alpha i}(\Gamma) - \Phi_{\alpha i}(\Gamma)| + \min\big(\Theta^r_{\alpha i}(\Gamma), \Phi_{\alpha i}(\Gamma)\big)\hat{d}_1(\Theta^r_{\alpha i}, \Phi_{\alpha i})\big) \le \|\Theta^r_{\alpha i} - \Phi_{\alpha i}\|,$$

where $\|\cdot\|$ applied to a measure denotes the variation norm; see Example 5.4.

REMARK 2. The d_2-distance depends on the choice of the underlying metric d_0 on Γ, and it is therefore not surprising that d_0 appears in both T_1 and T_2, in the latter through $\hat{d}_1$. In practice, an appropriate choice of d_0 has to be made. The standardization which seems most natural, where possible, is one which, loosely speaking, gives the process Ξ or the approximating compound Poisson process unit d_0-intensity; this has something of the flavor of the traditional idea of a limiting process. If Γ is a proper subset of some $\mathbb{R}^k$, and if $\boldsymbol{\nu}$ is uniform over Γ, one could for instance take

$$(4.10)\qquad d_0(\alpha, \beta) := \min\{1, (\bar{\mu})^{1/k}|\alpha - \beta|\},$$

where $\bar{\mu} = \boldsymbol{\mu}(\Gamma)/\boldsymbol{\nu}(\Gamma)$. However, for $k \ge 2$, it may be preferable to choose the scaling differently in different coordinate directions, if these have a particular, practical meaning.

Theorem 4.2 is very flexible, but rather abstract. We now consider some special settings, in which the various terms are more easily understood.

4.1. *Fixed neighborhoods of dependence.* For each $\alpha \in \Gamma$, let $\{\alpha\}$, N_α^s, N_α^b and N_α^w be a partition of Γ, such that sets of the form $\{(\alpha, \beta);\ \beta \in N_\alpha^s\}$ are product measurable in $\Gamma \times \Gamma$, and let

$$\Xi^{s,\alpha} := \sum_{\beta \in N_\alpha^s} \Xi\{\beta\}\delta_\beta, \qquad \Xi^{b,\alpha} := \sum_{\beta \in N_\alpha^b} \Xi\{\beta\}\delta_\beta, \qquad \Xi^{w,\alpha} := \sum_{\beta \in N_\alpha^w} \Xi\{\beta\}\delta_\beta,$$

so that $\Xi^{s,\alpha}$, $\Xi^{b,\alpha}$ and $\Xi^{w,\alpha}$ are the point processes resulting from restricting the process Ξ to N_α^s, N_α^b, and N_α^w, respectively. A typical choice is to define the sets $N_\alpha^s = B(\alpha, r(\alpha)) \setminus \{\alpha\}$ and $N_\alpha^b = B(\alpha, r'(\alpha)) \setminus B(\alpha, r(\alpha))$, where $B(\alpha, r)$ denotes the r-ball in Γ with centre α, r and r' are continuous functions, and $r(\alpha) < r'(\alpha)$ for all α. Then

$$\Xi = \Xi\{\alpha\} + \Xi^{s,\alpha} + \Xi^{b,\alpha} + \Xi^{w,\alpha}$$

defines a family $\mathcal{N}$ of decompositions as in (2.1), with $\Xi^{r,\alpha} = \Xi^{b,\alpha} + \Xi^{w,\alpha}$. In many cases, there is a natural way to make the partitioning in such a way that $\Xi^{s,\alpha}$ is strongly dependent on $\Xi\{\alpha\}\delta_\alpha$, $\Xi^{w,\alpha}$ is weakly dependent on $\Xi\{\alpha\}\delta_\alpha + \Xi^{s,\alpha}$, and the set $N_\alpha^b = \Gamma \setminus \{\{\alpha\} \cup N_\alpha^s \cup N_\alpha^w\}$ acts as a "boundary" between $\Xi^{s,\alpha}$ and $\Xi^{w,\alpha}$, much as in traditional "blocking" arguments. The intensity measures defined in (2.4) are now expressed as

$$\begin{aligned} i\boldsymbol{\pi}_i(d\alpha) &= P^\alpha\big(\Xi(N_\alpha^s \cup \{\alpha\}) = i\big)\boldsymbol{\mu}(d\alpha) \\ &= E\big\{I[\Xi(\{\alpha\} \cup N_\alpha^s) = i]\Xi(d\alpha)\big\}, \qquad \boldsymbol{\mu} \text{ a.e.,} \end{aligned}$$

by (2.3).

THEOREM 4.3. *Let $\mathcal{N}$ and $\boldsymbol{\pi}_i$, $i \geq 1$, be defined as above, and assume that, for each $\alpha \in \Gamma$ and $i \geq 1$, point processes $\Theta_{\alpha i}^w$ and $\Phi_{\alpha i}^w$ are defined on a common probability space in such a way that*

$$\mathcal{L}(\Theta_{\alpha i}^w) = P^\alpha\big(\Xi^{w,\alpha} \in \cdot \mid \Xi(N_\alpha^s \cup \{\alpha\}) = i\big); \qquad \mathcal{L}(\Phi_{\alpha i}^w) = \mathcal{L}(\Xi^{w,\alpha}).$$

Let $d_0(\mathcal{N}) := \sup_{\alpha \in \Gamma; \beta \in N_\alpha^s} d_0(\alpha, \beta)$. Then

$$\begin{aligned} &d_2\big(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\big) \\ &\quad \leq e^{\boldsymbol{\pi}(\Gamma)}\bigg\{d_0(\mathcal{N}) E\,\Xi(\Gamma) + E\bigg(\int_\Gamma I[\Xi(N_\alpha^b) \geq 1]\Xi(d\alpha)\bigg) \\ &\qquad + \int_\Gamma P\big(\Xi(N_\alpha^b \cup N_\alpha^s \cup \{\alpha\}) \geq 1\big)\boldsymbol{\mu}(d\alpha) \\ &\qquad + \sum_{i \geq 1} i \int_\Gamma \big(P\big(\Theta_{\alpha i}^w(\Gamma) \neq \Phi_{\alpha i}^w(\Gamma)\big) + E\hat{d}_1(\Theta_{\alpha i}^w, \Phi_{\alpha i}^w)\big)\boldsymbol{\pi}_i(d\alpha)\bigg\}. \end{aligned}$$

If, in addition, Condition (3.7) *holds, then*

$$
\begin{aligned}
&d_2\bigl(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)\bigr)\\
&\quad \le 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1)\\
&\qquad \times \biggl\{2d_0(\mathcal{N})E\biggl(\int_\Gamma \Xi(N_\alpha^s)\Xi(d\alpha)\biggr) + E\biggl(\int_\Gamma \Xi(N_\alpha^b)\Xi(d\alpha)\biggr)\\
&\qquad\qquad + \int_\Gamma E\,\Xi(N_\alpha^b \cup N_\alpha^s \cup \{\alpha\})\,\boldsymbol{\mu}(d\alpha) + \sum_{i\ge 1} i\int_\Gamma E\bigl(|\Theta_{\alpha i}^w(\Gamma) - \Phi_{\alpha i}^w(\Gamma)|\\
&\qquad\qquad\qquad + 2\min\bigl(\Theta_{\alpha i}^w(\Gamma), \Phi_{\alpha i}^w(\Gamma)\bigr)\hat{d}_1(\Theta_{\alpha i}^w, \Phi_{\alpha i}^w)\bigr)\,\boldsymbol{\pi}_i(d\alpha)\biggr\}.
\end{aligned}
$$

REMARK. Note that, in both bounds, the final and most complicated term is zero if $\Xi^{w,\alpha}$ and $\Xi^{s,\alpha} + \Xi\{\alpha\}$ are independent.

PROOF OF THEOREM 4.3. The proof is much as for Theorem 4.2, and involves estimating the right-hand side of (4.1). By (2.2), it follows that

$$
\left|E\biggl(\int_\Gamma g_\alpha(\Xi)\Xi(d\alpha)\biggr) - \sum_{i\ge 1} i\int_\Gamma E g_\alpha(\Xi + i\delta_\alpha)\,\boldsymbol{\pi}_i(d\alpha)\right| \le |B| + \sum_{l=1}^{3} |C^{(l)}|,
$$

where B is as introduced in (4.3), and C, defined in (4.4), can be split into the sum of

$$
\begin{aligned}
C^{(1)} := E\biggl\{\int_\Gamma \bigl\{&g_\alpha\bigl(\Xi^{w,\alpha} + \Xi^{b,\alpha} + \bigl(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\bigr)\delta_\alpha\bigr)\\
&- g_\alpha\bigl(\Xi^{w,\alpha} + \bigl(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\bigr)\delta_\alpha\bigr)\bigr\}\Xi(d\alpha)\biggr\},
\end{aligned}
\tag{4.11}
$$

$$
\begin{aligned}
C^{(2)} :=\ & E\biggl\{\int_\Gamma g_\alpha\bigl(\Xi^{w,\alpha} + \bigl(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\}\bigr)\delta_\alpha\bigr)\Xi(d\alpha)\biggr\}\\
&- \sum_{i\ge 1} i\int_\Gamma E g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)\,\boldsymbol{\pi}_i(d\alpha)
\end{aligned}
\tag{4.12}
$$

and

$$
C^{(3)} := \sum_{i\ge 1} i\int_\Gamma \bigl\{E g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha) - E g_\alpha(\Xi + i\delta_\alpha)\bigr\}\,\boldsymbol{\pi}_i(d\alpha).
\tag{4.13}
$$

The contribution from $|B|$ can be estimated as for Theorem 4.2; note that, by (2.2),

$$
\begin{aligned}
\int_\Gamma E^\alpha d_1\bigl(\Xi^{s,\alpha}, \Xi^{s,\alpha}(\Gamma)\delta_a\bigr)\,\boldsymbol{\mu}(d\alpha) &= E\biggl(\int_\Gamma d_1\bigl(\Xi^{s,\alpha}, \Xi^{s,\alpha}(\Gamma)\delta_a\bigr)\,\Xi(d\alpha)\biggr)\\
&\le d_0(\mathcal{N})E\,\Xi(\Gamma),
\end{aligned}
$$

and similarly

$$\int_\Gamma E^\alpha\left\{\int_\Gamma \Xi^{s,\alpha}(d\beta)d_0(\alpha,\beta)\right\}\boldsymbol{\mu}(d\alpha) \le d_0(\mathcal{N})E\left(\int_\Gamma \Xi(N_\alpha^s)\Xi(d\alpha)\right).$$

Furthermore, with $g_\alpha = g_f$ or $g_{\alpha f}$ as appropriate, for any $f \in \mathcal{F}$, the quantities $|C^{(1)}|$ and $|C^{(3)}|$ are easily bounded using Lemmas 3.1 and 3.5, and the remaining element $C^{(2)}$ is bounded in the same way that C was bounded for Theorem 4.2, by means of (3.31) and (3.30). □

REMARK. The fixed neighborhood structure above can be easily adapted to the restriction Ξ^A of the process Ξ to a subset $A \subset \Gamma$. For $\alpha \in A$, take $N_\alpha^l(A) := N_\alpha^l \cap A$ for $l = s, b, w$, and compute the corresponding bounds. There are some differences when $N_\alpha^l \not\subset A$, and integrals, including that implicit in the definition of $\lambda_1(A) := \boldsymbol{\pi}_1(A) - 2\boldsymbol{\pi}_2(A)$, are to be taken over A and not Γ. However, if the process Ξ is reasonably homogeneous, and if A is not such that edge effects play an important part, then the bounds thus obtained will not be greatly different from those for Ξ over the whole of Γ. In particular, bounds of similar accuracy then apply to the total variation approximation of $\mathcal{L}(\Xi(A))$ by the corresponding compound Poisson distribution on $\mathbb{Z}_+$, for any "reasonable" set A.

4.2. *Countable* Γ. If Γ is countable, Ξ can be written in the form $\Xi = \sum_{\alpha\in\Gamma} X_\alpha\delta_\alpha$, where the X_α are nonnegative integer valued random variables. Taking $\boldsymbol{\nu}$ to be counting measure, $\mu(\alpha)$ simply becomes EX_α, and the Palm measure P^α is given by the mixture of conditional distributions

$$P^\alpha(\cdot) = \sum_{l\ge 1} \frac{lP(X_\alpha = l)P(\Xi \in \cdot \mid X_\alpha = l)}{EX_\alpha}.$$

This makes possible some further simplification of the bounds in Theorems 4.2–4.3. We give a variation of the latter theorem, in a form reminiscent of the "local" version of (compound) Poisson approximation for random variables by Stein's method.

For each α, let N_α^s, N_α^b and N_α^w constitute an arbitrary partition of $\Gamma \setminus \{\alpha\}$. Define

$$\begin{aligned} U_\alpha &:= \sum_{\beta\in N_\alpha^s} X_\beta = \Xi(N_\alpha^s), \\ Z_\alpha &:= \sum_{\beta\in N_\alpha^b} X_\beta = \Xi(N_\alpha^b), \qquad (4.14) \\ W_\alpha &:= \sum_{\beta\in N_\alpha^w} X_\beta = \Xi(N_\alpha^w), \end{aligned}$$

so that $\Xi(\Gamma) = X_\alpha + U_\alpha + Z_\alpha + W_\alpha$, for each α. Then the intensities μ_i, $i \geq 1$, of clumps of size i for the approximating compound Poisson process are given by

$$i\mu_i(\alpha) = P^\alpha(X_\alpha + U_\alpha = i)EX_\alpha = \sum_{l=1}^{i} lP(X_\alpha = l, U_\alpha = i - l),$$

and $\boldsymbol{\pi}_i\{\alpha\} = \mu_i(\alpha)$ for each $i \geq 1$ and $\alpha \in \Gamma$. We now prove the following theorem.

THEOREM 4.4. *If Γ is countable, and $\Xi = \sum_{\alpha\in\Gamma} X_\alpha \delta_\alpha$ is decomposed as above, then*

$$\begin{aligned}
&d_2\bigl(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\bigr) \\
&\quad \leq e^{\boldsymbol{\pi}(\Gamma)}\Biggl\{d_0(\mathcal{N})E\,\Xi(\Gamma) + \sum_{\alpha\in\Gamma} E(X_\alpha I[Z_\alpha \geq 1]) \\
&\qquad + \sum_{\alpha\in\Gamma} P(X_\alpha + U_\alpha + Z_\alpha \geq 1)EX_\alpha \\
&\qquad + \sum_{\alpha\in\Gamma}\sum_{i\geq 1} E\bigl|E\{X_\alpha I[X_\alpha + U_\alpha = i] \mid \Xi^{w,\alpha}\} \\
&\qquad\qquad - E\{X_\alpha I[X_\alpha + U_\alpha = i]\}\bigr|\Biggr\}.
\end{aligned}$$

If, in addition, Condition (3.7) *holds, then*

$$\begin{aligned}
&d_2\bigl(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\bigr) \\
&\quad \leq 2\lambda_1^{-1}(1 + 2\log^+\lambda_1)\Biggl\{2d_0(\mathcal{N})\sum_{\alpha\in\Gamma} E(X_\alpha U_\alpha) + \sum_{\alpha\in\Gamma} E(X_\alpha Z_\alpha) \\
&\qquad\qquad + \sum_{\alpha\in\Gamma} E\{X_\alpha + U_\alpha + Z_\alpha\}EX_\alpha\Biggr\} \\
&\qquad + 1.65\lambda_1^{-1/2}\sum_{\alpha\in\Gamma}\sum_{i\geq 1} E\bigl|E\{X_\alpha I[X_\alpha + U_\alpha = i] \mid \Xi^{w,\alpha}\} \\
&\qquad\qquad - E\{X_\alpha I[X_\alpha + U_\alpha = i]\}\bigr|.
\end{aligned}$$

PROOF. All but the last elements in the bounds are direct translations of the corresponding terms in Theorem 4.3. The final term comes from an alternative bound for $C^{(2)}$ of (4.12):

$$
\begin{aligned}
|C^{(2)}| &= \Bigg| \sum_{\alpha\in\Gamma} \sum_{i\ge 1} \big(E\{ g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha) X_\alpha I[X_\alpha + U_\alpha = i]\} \\
&\qquad\qquad - E\{g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)\} E\{X_\alpha I[X_\alpha + U_\alpha = i]\}\big)\Bigg| \\
&\le \sum_{\alpha\in\Gamma} \sum_{i\ge 1} \big| E\big[g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)\{X_\alpha I[X_\alpha + U_\alpha = i] \\
&\qquad\qquad - E(X_\alpha I[X_\alpha + U_\alpha = i])\}\big]\big| \qquad (4.15)\\
&\le \sum_{\alpha\in\Gamma} \|g_\alpha\| \sum_{i\ge 1} E\big| E\{X_\alpha I[X_\alpha + U_\alpha = i] \mid \Xi^{w,\alpha}\} \\
&\qquad\qquad - E\{X_\alpha I[X_\alpha + U_\alpha = i]\}\big|.
\end{aligned}
$$

Taking $g_\alpha = g_f$ or $g_{\alpha f}$ as appropriate, and using Lemmas 3.1 and 3.5(i) respectively, gives the required bound. □

There are other variants of the final terms, analogous to those proved in Barbour and Chryssaphinou [(2000), (2.8)–(2.11)] for random variable approximation.

4.3. *Janossy densities.* In the case of fixed neighborhoods but uncountable Γ, the final element in the above bounds still has an analogue, if the Janossy densities $j_n : \Gamma^n \to [0,\infty)$ with respect to $\boldsymbol{\nu}^n$ exist for the process Ξ; this in particular requires the process to be simple. To simplify the notation, let $\overline{\alpha}_i = \alpha_1, \dots, \alpha_i$, $d\overline{\alpha}_i = d\alpha_1 \cdots d\alpha_i$, and similarly for $\overline{\beta}_i$, $d\overline{\beta}_i$, $\overline{\gamma}_i$, and $d\overline{\gamma}_i$. Then the density of the mean measure $\boldsymbol{\pi}$ of Ξ is given by

$$
\begin{aligned}
\mu(\alpha) &= \sum_{n\ge 0} \int_{\Gamma^n} \frac{1}{n!} j_{n+1}(\alpha, \overline{\alpha}_n)\boldsymbol{\nu}^n(d\overline{\alpha}_n) \\
&= \sum_{n\ge 0} \int_{(N_\alpha^w \cup N_\alpha^b)^n} \frac{1}{n!} \sum_{m\ge 0} \int_{(N_\alpha^s)^m} \frac{1}{m!} j_{n+m+1}(\overline{\alpha}_m, \alpha, \overline{\beta}_n)\boldsymbol{\nu}^{n+m}(d\overline{\alpha}_m, d\overline{\beta}_n).
\end{aligned}
$$

For each $\alpha \in \Gamma$ and $i \ge 1$, define the clump densities by

$$
\begin{aligned}
&i\mu_i(\alpha) \\
&\quad = \sum_{n\ge 0} \int_{(N_\alpha^w \cup N_\alpha^b)^n} \frac{1}{n!} \int_{(N_\alpha^s)^{i-1}} \frac{1}{(i-1)!} \qquad (4.16)\\
&\qquad\qquad \times j_{n+i}(\overline{\alpha}_{i-1}, \alpha, \overline{\beta}_n)\boldsymbol{\nu}^{n+i-1}(d\overline{\alpha}_{i-1}, d\overline{\beta}_n).
\end{aligned}
$$

This definition is equivalent to (2.5) and, as before, $\sum_{i\ge 1} i\mu_i(\alpha) = \mu(\alpha)$.

We give a version of Theorem 4.3, in which the contribution from $C^{(2)}$ of (4.12) is expressed in these terms. To do this, we need to introduce the conditional density of a clump of size l near α, given the configuration of Ξ on N_α^w; this is to be interpreted in the sense that we can find a function $\psi : \Gamma \times \mathcal{H} \to \mathbb{R}_+$ with the property that

$$E\left\{\int_\Gamma g\big(\Xi^{w,\alpha} + (U_\alpha + 1)\delta_\alpha, \alpha\big)\Xi(d\alpha)\right\} \tag{4.17}$$
$$= \sum_{i\ge 1}\int_\Gamma iE\big\{g(\Xi^{w,\alpha} + i\delta_\alpha, \alpha)\psi_i(\alpha|\Xi^{w,\alpha})\big\}\,\boldsymbol{\nu}(d\alpha),$$

for all bounded measurable functions $g : \mathscr{H} \times \Gamma \to \mathbb{R}$. For each $l \ge 1, m \ge 0, \alpha \in \Gamma$ and $\overline{\beta}_m \in (N_\alpha^w)^m$, we define

$$l\psi_l(\alpha\,|\,\overline{\beta}_m) \tag{4.18}$$
$$= \frac{\sum_{r\ge 0}\frac{1}{r!}\int_{(N_\alpha^b)^r}\frac{1}{(l-1)!}\int_{(N_\alpha^s)^{l-1}} j_{l+m+r}(\alpha, \overline{\alpha}_{l-1}, \overline{\beta}_m, \overline{\gamma}_r)\boldsymbol{\nu}^{r+l-1}(d\overline{\alpha}_{l-1}, d\overline{\gamma}_r)}{\sum_{n\ge 0}\frac{1}{n!}\int_{(N_\alpha^b)^n}\sum_{t\ge 0}\frac{1}{t!}\int_{(N_\alpha^s\cup\{\alpha\})^t} j_{m+n+t}(\overline{\alpha}_t, \overline{\beta}_m, \overline{\gamma}_n)\boldsymbol{\nu}^{t+n}(d\overline{\alpha}_t, d\overline{\gamma}_n)},$$

and show that this ψ satisfies (4.17).

If there are in total $n = k + m + r + 1$, $k, m, r \ge 0$, points in the process, then there are $\binom{k+m+r+1}{k,m,r,1}$ ways to divide these points into subsets consisting respectively of k, m, r, and 1 points. Hence, for any measurable nonnegative function $h : \mathscr{H} \to \mathbb{R}$,

$$\begin{aligned}
&E[h(\Xi)] \\
&\quad= \sum_{n\ge 0}\frac{1}{n!}\int_{\Gamma^n} h\left(\sum_{i=1}^n \delta_{\alpha_i}\right) j_n(\overline{\alpha}_n)\boldsymbol{\nu}^n(d\overline{\alpha}_n) \\
&\quad= j_0 h(\varnothing) + \sum_{k\ge 0}\sum_{m\ge 0}\sum_{r\ge 0}\binom{k+m+r+1}{k, m, r, 1} \\
&\qquad\qquad \times \int_\Gamma \int_{(N_\alpha^s)^k}\int_{(N_\alpha^w)^m}\int_{(N_\alpha^b)^r}\frac{1}{(k+m+r+1)!} \\
&\qquad\qquad\qquad \times h\left(\sum_{j=1}^r \delta_{\gamma_j} + \sum_{j=1}^m \delta_{\beta_j} + \sum_{j=1}^k \delta_{\alpha_j} + \delta_\alpha\right) \\
&\qquad\qquad\qquad \times j_{m+r+k+1}(\overline{\beta}_m, \overline{\gamma}_r, \overline{\alpha}_k, \alpha) \\
&\qquad\qquad\qquad \times \boldsymbol{\nu}^{r+m+k+1}(d\overline{\gamma}_r, d\overline{\beta}_m, d\overline{\alpha}_k, d\alpha) \\
&\quad= j_0 h(\varnothing) + \int_\Gamma \sum_{k\ge 0}\frac{1}{k!}\int_{(N_\alpha^s)^k}\sum_{m\ge 0}\frac{1}{m!}\int_{(N_\alpha^w)^m}\sum_{r\ge 0}\frac{1}{r!} \\
&\qquad\qquad \times \int_{(N_\alpha^b)^r} h\left(\sum_{j=1}^r \delta_{\gamma_j} + \sum_{j=1}^m \delta_{\beta_j} + \sum_{j=1}^k \delta_{\alpha_j} + \delta_\alpha\right) \\
&\qquad\qquad\qquad \times j_{m+r+k+1}(\overline{\beta}_m, \overline{\gamma}_r, \overline{\alpha}_k, \alpha) \\
&\qquad\qquad\qquad \times \boldsymbol{\nu}^{r+m+k+1}(d\overline{\gamma}_r, d\overline{\beta}_m, d\overline{\alpha}_k, d\alpha).
\end{aligned}$$

In particular, for the function $\int_\Gamma g(\Xi^{w,\alpha} + (U_\alpha + 1)\delta_\alpha, \alpha)\Xi(d\alpha)$, which equals zero if $\Xi = \varnothing$, we get

$$E\biggl[\int_\Gamma g\bigl(\Xi^{w,\alpha} + (U_\alpha+1)\delta_\alpha,\alpha\bigr)\Xi(d\alpha)\biggr]$$

$$= \int_\Gamma \sum_{m\ge 0}\frac{1}{m!}\int_{(N_\alpha^w)^m}\sum_{r\ge 0}\frac{1}{r!}\int_{(N_\alpha^b)^r}\sum_{i\ge 0}\frac{1}{i!}\int_{(N_\alpha^s)^i} g\Biggl(\sum_{j=1}^m \delta_{\beta_j} + (i+1)\delta_\alpha,\alpha\Biggr) \tag{4.19}$$

$$\times j_{m+r+i+1}(\overline{\beta}_m,\overline{\gamma}_r,\overline{\alpha}_i,\alpha)\boldsymbol{\nu}^{r+m+i+1}(d\overline{\alpha}_i, d\overline{\gamma}_r, d\overline{\beta}_m, d\alpha).$$

Now $l = i+1$ in (4.18) yields

$$E\biggl\{\int_\Gamma g\bigl(\Xi^{w,\alpha} + (U_\alpha+1)\delta_\alpha,\alpha\bigr)\Xi(d\alpha)\biggr\}$$

$$= \int_\Gamma \sum_{m\ge 0}\frac{1}{m!}\int_{(N_\alpha^w)^m}\sum_{i\ge 0} g\Biggl(\sum_{j=1}^m \delta_{\beta_j} + (i+1)\delta_\alpha,\alpha\Biggr)(i+1)\psi_{i+1}(\alpha|\overline{\beta}_m)$$

$$\times \sum_{n\ge 0}\frac{1}{n!}\int_{(N_\alpha^b)^n}\sum_{t\ge 0}\frac{1}{t!}\int_{(N_\alpha^s\cup\{\alpha\})^t} j_{m+n+t}(\overline{\alpha}_t,\overline{\beta}_m,\overline{\gamma}_n)$$

$$\times \boldsymbol{\nu}^{t+n}(d\overline{\alpha}_t, d\overline{\gamma}_n)\boldsymbol{\nu}^{m+1}(d\overline{\beta}_m, d\alpha)$$

$$= \int_\Gamma \sum_{i\ge 0}(i+1)E\bigl[g\bigl(\Xi^{w,\alpha} + (i+1)\delta_\alpha,\alpha\bigr)\psi_{i+1}(\alpha|\Xi^{w,\alpha})\bigr]\boldsymbol{\nu}(d\alpha)$$

$$= \sum_{i\ge 1} i\int_\Gamma E[g(\Xi^{w,\alpha} + i\delta_\alpha,\alpha)\psi_i(\alpha|\Xi^{w,\alpha})]\boldsymbol{\nu}(d\alpha),$$

and so (4.17) is indeed satisfied.

This enables us to bound $|C^{(2)}|$ as follows:

$$|C^{(2)}| = \biggl|E\biggl[\int_\Gamma g_\alpha\bigl(\Xi^{w,\alpha} + (U_\alpha+1)\delta_\alpha\bigr)\Xi(d\alpha)\biggr]$$

$$- \sum_{i\ge 1} i\int_\Gamma E[g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)]\boldsymbol{\pi}_i(d\alpha)\biggr|$$

$$\le \sum_{i\ge 1} i\int_\Gamma \bigl|E[g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)\psi_i(\alpha|\Xi^{w,\alpha})]$$

$$- E[g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)]\mu_i(\alpha)\bigr|\boldsymbol{\nu}(d\alpha)$$

$$= \sum_{i\ge 1} i\int_\Gamma \bigl|E[g_\alpha(\Xi^{w,\alpha} + i\delta_\alpha)\{\psi_i(\alpha|\Xi^{w,\alpha}) - \mu_i(\alpha)\}]\bigr|\boldsymbol{\nu}(d\alpha)$$

$$\le \sum_{i\ge 1} i\int_\Gamma \|g_\alpha\| E[|\psi_i(\alpha|\Xi^{w,\alpha}) - \mu_i(\alpha)|]\boldsymbol{\nu}(d\alpha). \tag{4.20}$$

Using Lemmas 3.1 and 3.5 to bound the contribution to (4.1) which results, we obtain the following theorem.

THEOREM 4.5. *Let $\mathcal{N}$ be defined as for Theorem* 4.3. *Suppose that the Janossy densities for Ξ exist, and are denoted as above. Let $\boldsymbol{\pi}_i(d\alpha) = \mu_i(\alpha)\boldsymbol{\nu}(d\alpha)$, $i \ge 1$, where μ_i is as in* (4.16), *and let $\psi_l(\alpha \mid \overline{\beta}_m)$ be as in* (4.18). *Then*

$$
\begin{aligned}
d_2(&\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)) \\
&\le e^{\boldsymbol{\pi}(\Gamma)} \biggl\{ d_0(\mathcal{N}) E\,\Xi(\Gamma) + E\biggl(\int_\Gamma I[\Xi(N_\alpha^b) \ge 1]\,\Xi(d\alpha)\biggr) \\
&\qquad + \int_\Gamma P\bigl(\Xi(N_\alpha^b \cup N_\alpha^s \cup \{\alpha\}) \ge 1\bigr)\boldsymbol{\mu}(d\alpha) \\
&\qquad + \sum_{i\ge 1} i \int_\Gamma \bigl(E|\psi_i(\alpha \mid \Xi^{w,\alpha}) - \mu_i(\alpha)|\bigr)\,\boldsymbol{\nu}(d\alpha)\biggr\}.
\end{aligned}
$$

If, in addition, Condition (3.7) *holds, then*

$$
\begin{aligned}
d_2(&\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)) \\
&\le 2\lambda_1^{-1}(1 + 2\log^+ \lambda_1) \biggl\{ 2d_0(\mathcal{N}) E\biggl(\int_\Gamma \Xi(N_\alpha^s)\,\Xi(d\alpha)\biggr) \\
&\qquad + E\biggl(\int_\Gamma \Xi(N_\alpha^b)\,\Xi(d\alpha)\biggr) \\
&\qquad + \int_\Gamma E\,\Xi(N_\alpha^b \cup N_\alpha^s \cup \{\alpha\})\,\boldsymbol{\mu}(d\alpha)\biggr\} \\
&\quad + 1.65\lambda_1^{-1/2} \sum_{i\ge 1} i \int_\Gamma \bigl(E|\psi_i(\alpha \mid \Xi^{w,\alpha}) - \mu_i(\alpha)|\bigr)\,\boldsymbol{\nu}(d\alpha).
\end{aligned}
$$

4.4. *Comparing compound Poisson processes.* The theorems given above presuppose that the point process Ξ is to be compared to a compound Poisson process $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$ which is derived from the properties of Ξ and the decomposition family $\mathcal{N}$ in a particular way. If one wants to compare instead with a perhaps nicer compound Poisson process, it is useful to be able also to bound the distance between the distributions of two compound Poisson processes. Now, if $\Xi \sim \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, the Palm characterization (3.2) implies that

$$
E\biggl\{\sum_{l\ge 1} l \int_\Gamma g(\Xi + l\delta_\alpha, \alpha)\boldsymbol{\pi}_l(d\alpha) - \int_\Gamma g(\Xi, \alpha)\,\Xi(d\alpha)\biggr\} = 0
$$

for all bounded g, including the functions g_f and $g_{\alpha f}$ found by solving the Stein equations (3.1) and (3.8) appropriate to $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots)$ for bounded functions f, so that

$$\begin{aligned}
&\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots)(f) - \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)(f) \\
&\qquad = E\left\{\sum_{l\ge 1} l \int_\Gamma g_\alpha(\Xi + l\delta_\alpha)\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\right\},
\end{aligned} \tag{4.21}$$

for $g_\alpha = g_f$ or $g_{\alpha f}$. Thus we can bound the d_2-difference between $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots)$ and $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)$, using (2.10), if we can control the right-hand side of (4.21) for all $f \in \mathcal{F}$. This leads to the following theorem.

THEOREM 4.6. *In general, we have the bound*

$$\begin{aligned}
&d_2\big(\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\big) \\
&\qquad \le e^{\boldsymbol{\rho}(\Gamma)} \sum_{l\ge 1} \big\{ l|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)| + \min\{\boldsymbol{\pi}_l(\Gamma), \boldsymbol{\rho}_l(\Gamma)\}\hat{d}_1(\boldsymbol{\pi}_l, \boldsymbol{\rho}_l)\big\}.
\end{aligned} \tag{4.22}$$

If $l(d\boldsymbol{\rho}_l/d\boldsymbol{\nu})(\alpha)$ *is decreasing in* l *for each* α *and if* $\sum_{l\ge 1} l^2\{\boldsymbol{\pi}_l(\Gamma) + \boldsymbol{\rho}_l(\Gamma)\} < \infty$, *then*

$$\begin{aligned}
&d_2\big(\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots)\big) \\
&\quad \le 1.65\sigma_1^{-1/2}|\hat{\boldsymbol{\pi}}(\Gamma) - \hat{\boldsymbol{\rho}}(\Gamma)| \\
&\qquad + 2\sigma_1^{-1}\big(1 + 2\log^+(\sigma_1/2)\big)\sum_{l\ge 2} l(l-1)|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)| \\
&\qquad + (1 - e^{-\sigma_1})\big\{\sigma_1^{-1} + \boldsymbol{\pi}(\Gamma)^{-1}(1 - e^{-\boldsymbol{\pi}(\Gamma)})\big\} \\
&\qquad\quad \times \Bigg(\min\{\hat{\boldsymbol{\pi}}(\Gamma), \hat{\boldsymbol{\rho}}(\Gamma)\}\hat{d}_1(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\rho}}) \\
&\qquad\qquad + 2\sum_{l\ge 2} l^2 \min\{\boldsymbol{\pi}_l(\Gamma), \boldsymbol{\rho}_l(\Gamma)\}\hat{d}_1(\boldsymbol{\pi}_l, \boldsymbol{\rho}_l)\Bigg),
\end{aligned} \tag{4.23}$$

where $\hat{\boldsymbol{\pi}} := \sum_{l\ge 1} l\boldsymbol{\pi}_l$, $\hat{\boldsymbol{\rho}} := \sum_{l\ge 1} l\boldsymbol{\rho}_l$ *and* $\sigma_1 := \boldsymbol{\rho}_1(\Gamma) - 2\boldsymbol{\rho}_2(\Gamma)$.

PROOF. For (4.22), if g_f solves the Stein equation (3.1) with $\boldsymbol{\rho}$ for $\boldsymbol{\pi}$, then it follows from Lemma 3.1 that $\|g_f\| \le s_2(f)e^{\boldsymbol{\rho}(\Gamma)}$ for all $f \in \mathcal{F}$; from Lemma 3.2, it also follows that that the function $\alpha \mapsto g_f(\xi + l\delta_\alpha)$ is d_0-Lipschitz with constant at most $s_2(f)l^{-1}e^{\boldsymbol{\rho}(\Gamma)}$ for all $f \in \mathcal{F}$, $l \ge 1$ and $\xi \in \mathcal{H}$. Hence

$$\begin{aligned}
&\frac{1}{s_2(f)}\left|\sum_{l\ge 1} l \int_\Gamma g_f(\xi + l\delta_\alpha)\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\right| \\
&\qquad \le e^{\boldsymbol{\rho}(\Gamma)} \sum_{l\ge 1} \big\{ l|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)| + \min\{\boldsymbol{\pi}_l(\Gamma), \boldsymbol{\rho}_l(\Gamma)\}\hat{d}_1(\boldsymbol{\pi}_l, \boldsymbol{\rho}_l)\big\},
\end{aligned}$$

and (4.22) follows from (4.21).

For (4.23), let the functions $g_{\alpha f}$ be those solving the Stein equation (3.9) for $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots)$, and write

$$\begin{aligned}&\sum_{l\ge 1} l \int_\Gamma g_{\alpha f}(\xi + l\delta_\alpha)\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\\ &\quad = \sum_{l\ge 2} l \int_\Gamma \{g_{\alpha f}(\xi + l\delta_\alpha) - g_{\alpha f}(\xi+\delta_\alpha)\}\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\\ &\qquad + \int_\Gamma g_{\alpha f}(\xi+\delta_\alpha)\big(\hat{\boldsymbol{\pi}}(d\alpha) - \hat{\boldsymbol{\rho}}(d\alpha)\big).\end{aligned} \tag{4.24}$$

Taking the second term first, observe that, for each $\xi \in \mathscr{H}$ and $f \in \mathscr{F}$, the function $\alpha \mapsto g_{\alpha f}(\xi + \delta_\alpha)$ is uniformly bounded by $s_2(f)(1 \wedge 1.65\sigma_1^{-1/2})$ by Lemma 3.5(i), and is d_0-Lipschitz with constant at most

$$s_2(f)\left\{1 \wedge \sigma_1^{-1}(1 - e^{-\sigma_1})\left(1 + \log\left(1 + \frac{\sigma_1}{1+|\xi|}\right)\right)\right\}$$

by Lemma 3.6(ii), so that

$$\begin{aligned}&\frac{1}{s_2(f)}\left|\int_\Gamma g_{\alpha f}(\xi+\delta_\alpha)\big(\hat{\boldsymbol{\pi}}(d\alpha) - \hat{\boldsymbol{\rho}}(d\alpha)\big)\right|\\ &\quad \le 1.65\sigma_1^{-1/2}|\hat{\boldsymbol{\pi}}(\Gamma) - \hat{\boldsymbol{\rho}}(\Gamma)|\\ &\qquad + (1 - e^{-\sigma_1})\big(\sigma_1^{-1} + (1+|\xi|)^{-1}\big)\min\{\hat{\boldsymbol{\pi}}(\Gamma), \hat{\boldsymbol{\rho}}(\Gamma)\}\hat{d}_1(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\rho}}).\end{aligned} \tag{4.25}$$

For the first term, treat each $l \ge 2$ separately. For each $\xi \in \mathscr{H}$ and $f \in \mathscr{F}$, the function

$$\alpha \mapsto g_{\alpha f}(\xi + l\delta_\alpha) - g_{\alpha f}(\xi + \delta_\alpha)$$

is uniformly bounded by

$$s_2(f)(l-1)\{1 \wedge 2\sigma_1^{-1}\big(1 + 2\log^+(\sigma_1/2)\big)\}$$

by (3.30), and is d_0-Lipschitz with constant at most

$$2s_2(f)\left\{1 \wedge l\sigma_1^{-1}(1 - e^{-\sigma_1})\left(1 + \log\left(1 + \frac{\sigma_1}{1+|\xi|}\right)\right)\right\}$$

by Lemma 3.6(ii), so that

$$\begin{aligned}&\frac{1}{s_2(f)}\left|\sum_{l\ge 2} l \int_\Gamma \big(g_{\alpha f}(\xi + l\delta_\alpha) - g_{\alpha f}(\xi+\delta_\alpha)\big)\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\right|\\ &\quad \le 2\sigma_1^{-1}\big(1 + 2\log^+(\sigma_1/2)\big)\sum_{l\ge 2} l(l-1)|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)|\\ &\qquad + 2(1 - e^{-\sigma_1})\big(\sigma_1^{-1} + (1+|\xi|)^{-1}\big)\sum_{l\ge 2} l^2 \min\{\boldsymbol{\pi}_l(\Gamma), \boldsymbol{\rho}_l(\Gamma)\}\hat{d}_1(\boldsymbol{\pi}_l, \boldsymbol{\rho}_l).\end{aligned} \tag{4.26}$$

Now take $\Xi \sim \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)$; then, from (4.24)–(4.26), and using the fact that

$$E\{(1+|\Xi|)^{-1}\} \le \frac{1}{\boldsymbol{\pi}(\Gamma)}(1-e^{-\boldsymbol{\pi}(\Gamma)}),$$

we have

$$\begin{aligned}
&\frac{1}{s_2(f)}\left| E\sum_{l\ge 1} l\int_\Gamma g_\alpha(\Xi + l\delta_\alpha)\big(\boldsymbol{\pi}_l(d\alpha) - \boldsymbol{\rho}_l(d\alpha)\big)\right| \\
&\quad \le 1.65\sigma_1^{-1/2}|\hat{\boldsymbol{\pi}}(\Gamma) - \hat{\boldsymbol{\rho}}(\Gamma)| \\
&\qquad + 2\sigma_1^{-1}\big(1+2\log^+(\sigma_1/2)\big)\sum_{l\ge 2} l(l-1)|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)| \\
&\qquad + (1-e^{-\sigma_1})\{\sigma_1^{-1} + \boldsymbol{\pi}(\Gamma)^{-1}(1-e^{-\boldsymbol{\pi}(\Gamma)})\} \\
&\qquad\quad \times \left(\min\{\hat{\boldsymbol{\pi}}(\Gamma), \hat{\boldsymbol{\rho}}(\Gamma)\}\hat{d}_1(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\rho}}) + 2\sum_{l\ge 2} l^2 \min\{\boldsymbol{\pi}_l(\Gamma), \boldsymbol{\rho}_l(\Gamma)\}\hat{d}_1(\boldsymbol{\pi}_l, \boldsymbol{\rho}_l)\right),
\end{aligned}$$

from which the theorem follows. □

REMARK. In the case of random variables, when $\mathrm{card}(\Gamma) = 1$, the terms involving $\hat{d}$ in Theorem 4.6 are zero, and the bound reduces to that of Barbour and Chryssaphinou [(2000), CPA 1B with $\varepsilon_0 = \varepsilon_1 = 0$, (2.15) and (2.19)], apart from differences in the constants. For Poisson process approximation, $\boldsymbol{\pi}_l(\Gamma) = \boldsymbol{\rho}_l(\Gamma) = 0$ for all $l \ge 2$, so that the l-sums are empty. If also $\boldsymbol{\pi}(\Gamma) = \boldsymbol{\rho}(\Gamma)$, this gives the same bound as is implied by Barbour, Holst and Janson [(1992), Theorem 10.F].

COROLLARY 4.7. *In the general bounds of Theorems* 4.2–4.5, $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)$ *can be replaced by* $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots)$ *if* $\boldsymbol{\pi}(\Gamma)$ *is replaced by* $\boldsymbol{\rho}(\Gamma)$ *and if the bound in* (4.22) *is added. If also* $l(d\boldsymbol{\rho}_l/d\boldsymbol{\nu})(\alpha)$ *is decreasing in* l *for each* α *and if* $\sum_{l\ge 1} l^2\{\boldsymbol{\pi}_l(\Gamma) + \boldsymbol{\rho}_l(\Gamma)\} < \infty$, *then* $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)$ *can be replaced by* $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots)$ *in the second bounds of Theorems* 4.2–4.5, *if* λ_1 *is replaced by* σ_1 *and if the bound in* (4.23) *is added.*

5. Examples.

5.1. *Declumping.* Suppose that $\Xi = \sum_{\alpha\in\Gamma} I_\alpha \delta_\alpha$ is a simple point process on a finite or countable set Γ, in which the points tend to occur in clusters. In this section, we discuss the approximation of a "declumped" version of Ξ, as in Arratia, Goldstein and Gordon (1989). If a declumping can be simply defined, approximation can be made by way of a Poisson process approximation on a larger carrier space, leading to a different measure of closeness of distributions. Here, we show that approximation is then also possible with respect to our usual d_2-distance, and show that the bounds are similar in form.

To best illustrate the comparison, we suppose that Ξ can be expressed in the form $\Xi_* = \sum_{\alpha\in\Gamma}\sum_{i\ge 1} i I_{\alpha i}\delta_\alpha$, where $I_{\alpha i} = 1$ is interpreted as the event that a cluster of size i occurs at α, $I_{\alpha i} = 0$ that there is none; hence, for each α, at most one of the $I_{\alpha i}$ can take the value 1. To reach such a form, some approximation has usually already been made, moving all points of the α-clump onto the one representative α, and the error involved corresponds to the first element in the bounds in Theorems 4.2–4.4. We then approximate Ξ_* by a compound Poisson process $\mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)$ with $\boldsymbol{\pi}_i\{\alpha\} = E[I_{\alpha i}]$ for $i \ge 1$ and $\alpha \in \Gamma$.

First, as in Arratia, Goldstein and Gordon (1989), we construct another point process on $\Gamma \times \mathbb{N}$ by letting $\widetilde{\Xi} = \sum_{\alpha\in\Gamma}\sum_{i\ge 1} I_{\alpha i}\delta_{\alpha i}$. For each $(\alpha, i) \in \Gamma \times \mathbb{N}$, let $B(\alpha, i) \subset \Gamma \times \mathbb{N}$ be a set containing (α, i). Set

$$b_1 = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}} \sum_{(\beta,j)\in B(\alpha,i)} E[I_{\alpha i}]E[I_{\beta j}],$$

$$b_2 = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}} \sum_{(\beta,j)\in B(\alpha,i),(\beta,j)\ne(\alpha,i)} E[I_{\alpha i} I_{\beta j}],$$

$$b_3 = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}} E\big|E\big[I_{\alpha i} \,|\, \sigma\{I_{\beta j}; (\beta, j) \notin B(\alpha, i)\}\big] - E[I_{\alpha i}]\big|.$$

Let $\mathrm{Po}(\boldsymbol{\lambda}_{\Gamma\times\mathbb{N}})$ denote a Poisson process on $\Gamma \times \mathbb{N}$ with intensity $\lambda(\alpha, i) = E[I_{\alpha i}]$. Using Stein's method for Poisson process approximation, with error bounds as given in Theorem 10.A in Barbour, Holst and Janson (1992), we get

$$d_{TV}\big(\mathcal{L}(\widetilde{\Xi}), \mathrm{Po}(\boldsymbol{\lambda}_{\Gamma\times\mathbb{N}})\big) \le b_1 + b_2 + b_3. \tag{5.1}$$

Then, since Ξ_* is a function of $\widetilde{\Xi}$, it follows immediately that

$$d_{TV}\big(\mathcal{L}(\Xi_*), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots)\big) \le b_1 + b_2 + b_3 \tag{5.2}$$

also.

This bound is simple and effective, as long as $\boldsymbol{\pi}(\Gamma)$ is not too large. Otherwise, one can take the metric $\tilde{d}_0$ on $\Gamma \times \mathbb{N}$ defined by

$$\tilde{d}_0\big((\alpha, i), (\beta, j)\big) := \begin{cases} 1, & \text{if } i \ne j, \\ d_0(\alpha, \beta), & \text{if } i = j, \end{cases}$$

and apply Theorem 10.F in Barbour, Holst and Janson (1992) for Poisson process approximation in the d_2-metric; this gives

$$\begin{aligned} &\tilde{d}_2\big(\mathcal{L}(\widetilde{\Xi}), \mathrm{Po}(\boldsymbol{\lambda}_{\Gamma\times\mathbb{N}})\big) \\ &\quad \le (b_1 + b_2)\big\{1 \wedge 2\lambda^{-1}\big(1 + 2\log^+(\lambda/2)\big)\big\} + b_3\big\{1 \wedge 1.65\lambda^{-1/2}\big\}, \end{aligned} \tag{5.3}$$

where $\lambda := \boldsymbol{\pi}(\Gamma)$ and $\tilde{d}_2 := d_2(\Gamma \times \mathbb{N})$.

As discussed in the Introduction, this is not the same as approximation in the metric $d_2(\Gamma)$, which is preferable for practical application. To obtain approximation in $d_2(\Gamma)$, we write $\Xi_* = \sum_{\alpha\in\Gamma} X_\alpha \delta_\alpha$ as for Theorem 4.4, with $X_\alpha := \sum_{i\ge 1} i I_{\alpha i} = \Xi_*\{\alpha\}$, and we define a decomposition family $\mathcal{N}$ by setting $\Xi_*^{s,\alpha} := \varnothing$, and hence $\Xi_*^{r,\alpha} := \Xi_* - X_\alpha\delta_\alpha$. Let $\{I_{\beta j}^{\alpha i},\ (\beta, j) \in \Gamma\times\mathbb{N}\}$ be distributed as $\{I_{\beta j},\ (\beta, j) \in \Gamma\times\mathbb{N}\}$ conditional on $I_{\alpha i} = 1$, and defined on the same probability space, enlarged if necessary, as Ξ. Set

$$\Theta_{\alpha i}^r := \sum_{\beta\ne\alpha}\sum_{j\ge 1} j I_{\beta j}^{\alpha i}\delta_\beta; \qquad \Theta_{\alpha i}^{r,2} := \sum_{(\beta,j)\in B(\alpha,i),\beta\ne\alpha} j I_{\beta j}^{\alpha i}\delta_\beta;$$
$$\Theta_{\alpha i}^{r,1} := \sum_{(\beta,j)\notin B(\alpha,i)} j I_{\beta j}^{\alpha i}\delta_\beta;$$
$$\Phi_{\alpha i} := \sum_{\beta}\sum_{j\ge 1} j I_{\beta j}\delta_\beta; \qquad \Phi_{\alpha i}^2 := \sum_{(\beta,j)\in B(\alpha,i)} j I_{\beta j}\delta_\beta;$$
$$\Phi_{\alpha i}^1 := \sum_{(\beta,j)\notin B(\alpha,i)} j I_{\beta j}\delta_\beta;$$

then we have $\mathcal{L}(\Theta_{\alpha i}^r) = \mathcal{L}(\Xi_*^{r,\alpha} \mid \Xi_*^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i)$ and $\mathcal{L}(\Phi_{\alpha i}) = \mathcal{L}(\Xi_*)$, and also $\Theta_{\alpha i}^r = \Theta_{\alpha i}^{r,1} + \Theta_{\alpha i}^{r,2}$ and $\Phi_{\alpha i} = \Phi_{\alpha i}^1 + \Phi_{\alpha i}^2$. Now note that $c_{\mathcal{g}}(\Xi, \mathcal{N})$ of Lemma 4.1 is just

$$\begin{aligned}\sum_{\alpha\in\Gamma}\sum_{i\ge 1} i E(I_{\alpha i})\, \big| E\big\{g_\alpha(\Theta_{\alpha i}^r + i\delta_\alpha) - g_\alpha(\Theta_{\alpha i}^{r,1} + i\delta_\alpha)\big\} \\ + E\big\{g_\alpha(\Theta_{\alpha i}^{r,1} + i\delta_\alpha) - g_\alpha(\Phi_{\alpha i}^1 + i\delta_\alpha)\big\} \\ + E\big\{g_\alpha(\Phi_{\alpha i}^1 + i\delta_\alpha) - g_\alpha(\Phi_{\alpha i} + i\delta_\alpha)\big\}\big|.\end{aligned} \tag{5.4}$$

Replace b_1, b_2 and b_3 by the modified quantities

$$b_1^* = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}}\ \sum_{(\beta,j)\in B(\alpha,i)} ij E[I_{\alpha i}]E[I_{\beta j}];$$

$$b_2^* = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}}\ \sum_{(\beta,j)\in B(\alpha,i),\beta\ne\alpha} ij E[I_{\alpha i} I_{\beta j}];$$

$$b_3^* = \sum_{(\alpha,i)\in\Gamma\times\mathbb{N}} i E\big|E\big[I_{\alpha i} \mid \sigma\{I_{\beta j}; (\beta, j)\notin B(\alpha, i)\}\big] - E[I_{\alpha i}]\big|.$$

Then, if condition (3.7) holds, the first and last terms in (5.4) are bounded using Lemma 3.5(ii) by

$$s_2(f)(b_2^* + b_1^*)\{1 \wedge 2\lambda_1^{-1}(1 + 2\log^+(\lambda_1/2))\};$$

the second, by the argument for $C^{(2)}$ in Theorem 4.4 and by Lemma 3.5(i), is bounded by $s_2(f)b_3^*\{1 \wedge 1.65\lambda_1^{-1/2}\}$; as usual, $\lambda_1 := \boldsymbol{\pi}_1(\Gamma) - 2\boldsymbol{\pi}_2(\Gamma)$. Hence the bound on $d_2(\Gamma)(\mathcal{L}(\Xi_*), \mathrm{CP}(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots))$ is the same as in (5.3), but with b_l^* in place of b_l, $1 \le l \le 3$, and with λ_1 in place of λ. Thus, with these differences, approximation in $d_2(\Gamma)$ for the process Ξ_* can also be simply established. However, if condition (3.7) does not hold, the bound is very much worse, because of the factor $e^{\boldsymbol{\pi}(\Gamma)}$ appearing in the general bounds, and the usual declumping approach is clearly better.

Note that, in this example, there is no mention of the underlying metric d_0. This is primarily because all points in a cluster have already been moved together in the definition of the "declumped" process Ξ_*. However, there are also the terms corresponding to T_2, which in principle involve d_0; here, the bounds are calculated as if the strongest possible metric, the discrete metric, were being used.

5.2. *Runs.* Let $Y_1, \dots, Y_n$ be independent $\mathrm{Be}(p)$-distributed random variables. Define $\Xi := \sum_{\alpha=1}^n I_\alpha \delta_\alpha$, where $I_\alpha := \prod_{l=0}^{k-1} Y_{\alpha+l}$, and where we suppose that the sequence is "tied together" as a circle so that $Y_{n+\alpha} = Y_\alpha$ and $Y_{1-\alpha} = Y_{n-\alpha+1}$ for $\alpha \geq 1$. A point of Ξ at α indicates that a run of k 1's starts at α; this run may be overlapped by others at either end. In this example, it is easy to declump, by taking the first index in a cluster of overlapping runs as its representative. However, it is just as easy to apply Theorem 4.2 directly.

If $I_\alpha = 1$, define a decomposition family $\mathcal{N}$ by setting

$$\Xi^{s,\alpha} := \sum_{\beta<\alpha} \delta_\beta \prod_{\gamma=\beta}^{\alpha} I_\gamma + \sum_{\beta>\alpha} \delta_\beta \prod_{\gamma=\alpha}^{\beta} I_\gamma \tag{5.5}$$

if $\prod_{\alpha=1}^n Y_\alpha = 0$, with the circle convention employed in all sums and products; and otherwise set $\Xi^{s,\alpha} + \delta_\alpha := \sum_{\beta=1}^n \delta_\beta$. In either case, set $\Xi^{r,\alpha} := \Xi - \Xi^{s,\alpha}$. Thus, if the last 0 before α occurs at $\alpha - l$ and the first 0 after α at $\alpha + k - 1 + m$, with $l, m \geq 1$ and $m + l - 1 = i \leq n - k$, then $\Xi^{s,\alpha} + \delta_\alpha = \sum_{\beta=\alpha-l+1}^{\alpha+m-1} \delta_\beta$ is the cluster containing α, and $\Xi^{r,\alpha}$ is determined by the values of Y_β, $\beta \notin [\alpha - l, \alpha + k - 1 + m]$ together with the 0's at $\alpha - l$ and at $\alpha + k - 1 + m$. It is now easy to check that $\boldsymbol{\mu}\{\alpha\} = p^k$ for all α, and that

$$P^\alpha\bigl(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\bigr) = \begin{cases} i(1-p)^2 p^{i-1}, & \text{if } 1 \leq i \leq n-k-1, \\ i(1-p)p^{i-1}, & \text{if } i = n-k, \\ p^{n-k}, & \text{if } i = n, \end{cases} \tag{5.6}$$

so that, in particular, $\boldsymbol{\pi}_i\{\alpha\} = (1-p)^2 p^{k+i-1}$ for $1 \leq i \leq n-k-1$. This suggests approximation by $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots)$ with $\boldsymbol{\rho}_i\{\alpha\} := (1-p)^2 p^{k+i-1}$ for all $i \geq 1$, satisfying $l(d\boldsymbol{\rho}_l/d\boldsymbol{\nu})(\alpha)$ decreasing in l for each α if $p < 1/2$, where $\boldsymbol{\nu}$ denotes counting measure. Using the notation in Theorem 4.6, we have $\sigma_1 = np^k(1-p)^2 (1-2p)$, $\hat{\boldsymbol{\rho}}\{\alpha\} = p^k = E(\Xi\{\alpha\}) = \hat{\boldsymbol{\pi}}\{\alpha\}$ for each α, $\boldsymbol{\pi}_i = \boldsymbol{\rho}_i$ for $1 \leq i \leq n-k-1$ and $\hat{d}_1(\boldsymbol{\pi}_i, \boldsymbol{\rho}_i) = 0$ for all i. Hence the contribution (4.23), required in addition to the bound of Theorem 4.2 by Corollary 4.7, reduces to

$$2\sigma_1^{-1}\bigl(1 + 2\log^+(\sigma_1/2)\bigr) \sum_{l\geq 2} l(l-1)|\boldsymbol{\pi}_l(\Gamma) - \boldsymbol{\rho}_l(\Gamma)|, \tag{5.7}$$

in which the sum is of order $O(n^2 p^{n-1})$.

To tackle the terms T_1 and T_2 in the second bound in Theorem 4.2, we couple $\Theta^r_{\alpha i}$ and $\Phi_{\alpha i}$ by also conditioning on the value of l as defined above, taking $\Theta^{r,l}_{\alpha i} := \Xi^{r,\alpha}$ and letting $\Phi^l_{\alpha i}$ be derived from indicators $(Y'_\beta,\, 1 \le \beta \le n)$, where we set $Y'_\beta = Y_\beta$ for $\beta \notin [\alpha - l, \alpha + k - 1 + m]$ and choose $Y'_\beta \sim \mathrm{Be}(p)$ for $\beta \in [\alpha - l, \alpha + k - 1 + m]$ independently of one another and of the Y_β's. With this construction, it is clear that $\mathcal{L}(\Phi^l_{\alpha i}) = \mathcal{L}(\Xi)$ and that $\Phi^l_{\alpha i}$ is close to $\Theta^{r,l}_{\alpha i}$; in fact, $\Theta^{r,l}_{\alpha i}(A) = \Phi^l_{\alpha i}(A)$ for all $A \subset \{1, 2, \dots, n\} \setminus [\alpha - l - k + 1, \alpha + k - 1 + m]$ and $E|\Theta^{r,l}_{\alpha i}(\Gamma) - \Phi^l_{\alpha i}(\Gamma)| \le (2k + i)p^k$. Furthermore, by the definition of $\Xi^{r,\alpha}$ and because $Y_{\alpha - l} = Y_{\alpha + k - 1 + m} = 0$, it follows that $\Theta^{r,l}_{\alpha i}([\alpha - l - k + 1, \alpha + k - 1 + m]) = \varnothing$, so that $\Theta^{r,l}_{\alpha i} \le \Phi^l_{\alpha i}$ a.s. and $\hat{d}_1(\Theta^{r,l}_{\alpha i}, \Phi^l_{\alpha i}) = 0$. Hence, for the second bound in Theorem 4.2, we have

$$\begin{aligned} T_2 &\le n(1-p)^2 \sum_{i \ge 1} i(2k+i) p^{2k+i-1} + n^2 p^n \\ &= n(2k+1)p^{2k} + 2n(1-p)^{-1} p^{2k+1} + n^2 p^n. \end{aligned}$$

For T_1, we must first specify d_0. By Remark 2 following Theorem 4.2, a natural choice is to take $d_0(\alpha, \beta) := (p^k|\alpha - \beta| \wedge 1)$ (with $|\cdot|$ measuring the shortest distance on the circle), so that Ξ has unit d_0-intensity. Then direct calculation shows that

$$E^\alpha\left(\int_\Gamma \Xi^{s,\alpha}(d\beta) d_0(\alpha, \beta)\right) \le 2p^{k+1}(1-p)^{-2} + np^{n-k} \mathbb{1}_{\{\alpha = 1\}},$$

and hence $T_1 \le 4np^{2k+1}(1-p)^{-2} + np^n$. Combining the various terms, we find that

$$\begin{aligned} d_2\big(\mathcal{L}(\Xi), \mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots)\big) &\le 2p^k \left\{ \frac{1 + 2\log^+ \sigma_1}{(1-p)^2(1-2p)} \right\} \\ &\quad \times \left\{ 2k + 1 + \frac{2p(3-p)}{(1-p)^2} + O(np^{n-2k-1}) \right\} \\ &= O\left(kp^k \frac{1 + \log^+(np^k)}{1 - 2p} \right), \end{aligned}$$

uniformly in $p < 1/2$.

5.3. *Sequence matching.* Assume that $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ are sequences of independent random variables taking values in a finite alphabet $\mathcal{A}$. Let the X_i and the Y_j be sampled from the distributions τ and ν, respectively, and let

$$0 < p := \sum_{a \in \mathcal{A}} \tau_a \nu_a < 1, \tag{5.8}$$

be the probability of a "match" between X_i and Y_j, i, $j \geq 1$. Define the index set $\Gamma = \{\alpha = (\alpha_1, \alpha_2) : 1 \leq \alpha_1 \leq m, 1 \leq \alpha_2 \leq n\}$, and set

$$I_\alpha := I\big[X_{\alpha_1} = Y_{\alpha_2}, X_{\alpha_1+1} = Y_{\alpha_2+1}, \dots, X_{\alpha_1+k-1} = Y_{\alpha_2+k-1}\big],$$

so that $I_\alpha = 1$ if a matching subsequence of length k between the two sequences starts at α; we use the torus convention, much as in the previous example, to avoid edge effects. Then define the point process

$$\Xi := \sum_{\alpha \in \Gamma} I_\alpha \delta_\alpha$$

of starting points of matchings of length k.

We start by defining a decomposition family $\mathcal{N}$. As with much of the argument below, this can be done in a fashion similar to that used in the runs example. Let $\beta < \alpha$ be interpreted as $\beta_1 < \alpha_1$ and $\beta_2 < \alpha_2$. If $I_\alpha = 1$, set

$$\Xi^{s,\alpha} := \sum_{\beta < \alpha, |\alpha_1 - \beta_1| = |\alpha_2 - \beta_2|} \delta_\beta \prod_{\gamma=\beta}^{\alpha} I_\gamma + \sum_{\beta > \alpha, |\alpha_1 - \beta_1| = |\alpha_2 - \beta_2|} \delta_\beta \prod_{\gamma=\alpha}^{\beta} I_\gamma$$

if $\prod_{j=-J}^{\min\{m,n\}-J} I[X_{\alpha_1+j} = Y_{\alpha_2+j}] = 0$, for all $J = 0, \dots, \min\{m,n\} - k$, and otherwise set $\Xi^{s,\alpha} + \delta_\alpha := \sum_{j=-J}^{\min\{m,n\}-J} \delta_{\alpha+j}$, for any J such that the product above is non-zero. In either case, set $\Xi^{r,\alpha} := \Xi - \Xi^{s,\alpha}$.

As in the runs example, $P^\alpha(\Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i)$ is given by (5.6), but now with p as defined in (5.8) and with n replaced by $\min\{m,n\}$. Furthermore, $\boldsymbol{\mu}\{\alpha\} = p^k$ and $\boldsymbol{\pi}_i(\alpha) = (1-p)^2 p^{k+i-1}$, $1 \leq i \leq \min\{m,n\} - k - 1$. The discussion following (5.6) concerning approximation with a compound Poisson process with the more convenient $\boldsymbol{\rho}$ measures, defined by $\boldsymbol{\rho}_i(\alpha) = (1-p)^2 p^{k+i-1}$, $i \geq 1$, rather than the $\boldsymbol{\pi}_i$s is valid here as well. The extra contribution we need to add in the d_2-bound, according to Corollary 4.7, is the same as in (5.7):

$$2\sigma_1^{-1}\big(1 + 2\log^+(\sigma_1/2)\big) \sum_{i \geq 2} i(i-1) |\boldsymbol{\pi}_i(\Gamma) - \boldsymbol{\rho}_i(\Gamma)|, \tag{5.9}$$

where $\sigma_1 = mnp^k(1-p)^2(1-2p)$, and the sum is of order $O(mnp^{\min\{m,n\}-1})$.

We couple $\Theta_{\alpha i}^r$ and $\Phi_{\alpha i}$ by also conditioning on the value of l, defined as follows. If $I_\alpha = 1$, we let l be such that $\alpha - l + 1$ is the starting position of the cluster containing α, which if the cluster size is i means that $1 \leq l \leq i$, $X_{\alpha_1 - l} \neq Y_{\alpha_2 - l}$, $X_{\alpha_1 - l + j} = Y_{\alpha_2 - l + j}$, $j = 1, \dots, k + i - 1$, and $X_{\alpha_1 - l + k + i} \neq Y_{\alpha_2 - l + k + i}$, and that $\Xi^{s,\alpha} + \delta_\alpha = \sum_{\beta = \alpha - l + 1}^{\alpha - l + i} \delta_\beta$. We then take $\Theta_{\alpha i}^{r,l} := \Xi^{r,\alpha}$ and let $\Phi_{\alpha i}^l$ be derived from $\{(X'_{\beta_1}, Y'_{\beta_2}), 1 \leq \beta_1 \leq m, 1 \leq \beta_2 \leq n\}$, where we set $X'_{\beta_1} = X_{\beta_1}$ and $Y'_{\beta_2} = Y_{\beta_2}$ for $\beta_j \notin [\alpha_j - l, \alpha_j + k + i - l]$, and choose X'_{β_1} and Y'_{β_2} from the distributions τ and ν, respectively, for $\beta_j \in [\alpha_j - l, \alpha_j + k + i - l]$, independently of one another and of the X_{β_1}'s and Y_{β_2}'s. Let I'_β be defined as I_β, but determined by the X'_i's and Y'_i's rather then by the X_i's and Y_i's.

Let $A_{\alpha_j,i,l} := [\alpha_j - l - k + 1, \alpha_j + k + i - l]$, $j = 1, 2$, and $A := \{\{1, 2, \dots, m\} \setminus A_{\alpha_1,i,l}\} \times \{\{1, 2, \dots, n\} \setminus A_{\alpha_2,i,l}\}$. Note that for $i > \min\{m, n\} - 2k$ either or both of $A_{\alpha_j,i,l}$ "overlaps itself," so that $A_{\alpha_j,i,l} = \Gamma$ and $A = \varnothing$. To bound T_2, defined in Theorem 4.2, we first note that $\Theta^{r,l}_{\alpha i}$ and $\Phi^l_{\alpha i}$ are identical on the set A. Then, by (4.8),

$$
\begin{aligned}
&|\Theta^{r,l}_{\alpha i}(\Gamma) - \Phi^l_{\alpha i}(\Gamma)| + 2\min\bigl(\Theta^{r,l}_{\alpha i}(\Gamma), \Phi^l_{\alpha i}(\Gamma)\bigr)\hat{d}_1(\Theta^{r,l}_{\alpha i}, \Phi^l_{\alpha i})\\
&\qquad \le \Theta^{r,l}_{\alpha i}(\Gamma \setminus A) + \Phi^l_{\alpha i}(\Gamma \setminus A).
\end{aligned}
$$

Now we divide $\Gamma \setminus A$ into four subsets:

$$
\begin{aligned}
\Gamma^0_{\alpha,i,l} &:= \bigl\{\beta : \beta_1 \in A_{\alpha_1,i,l}, \beta_2 \in A_{\alpha_2,i,l}, |\alpha_1 - \beta_1| = |\alpha_2 - \beta_2|\bigr\},\\
\Gamma^1_{\alpha,i,l} &:= \bigl\{\beta : \beta_1 \in A_{\alpha_1,i,l}, \beta_2 \in A_{\alpha_2,i,l}, |\alpha_1 - \beta_1| \ne |\alpha_2 - \beta_2|\bigr\},\\
\Gamma^2_{\alpha,i,l} &:= \{\beta : \beta_1 \in A_{\alpha_1,i,l}, \beta_2 \notin A_{\alpha_2,i,l}\}
\end{aligned}
$$

and

$$
\Gamma^3_{\alpha,i,l} := \{\beta : \beta_1 \notin A_{\alpha_1,i,l}, \beta_2 \in A_{\alpha_2,i,l}\}.
$$

Furthermore, let

$$
q_1 := \sum_{a \in \mathcal{A}} \tau_a^2 \nu_a; \qquad q_2 := \sum_{a \in \mathcal{A}} \tau_a \nu_a^2; \qquad \gamma_+ = \max_{a \in \mathcal{A}} \gamma_a, \text{ where } \gamma_a := \tau_a \nu_a / p,
$$

and note that $p\gamma_+ \ge q_i \ge p^2$, with equalities if and only if $\tau = \nu$ is the uniform distribution. Then it follows that

$$
E[I_\beta | I_\alpha = 1] p^k \le
\begin{cases}
\max\Bigl\{\sum_{a \in \mathcal{A}} \tau_a (\tau_a \nu_a)^k, \sum_{a \in \mathcal{A}} \nu_a (\tau_a \nu_a)^k\Bigr\} \le \gamma_+^k p^k, & \\
& \text{if } \beta \in \Gamma^1_{\alpha,i,l},\\
q_1^k, & \text{if } \beta \in \Gamma^2_{\alpha,i,l},\\
q_2^k, & \text{if } \beta \in \Gamma^3_{\alpha,i,l},
\end{cases}
$$

[arguments for this can be found in for instance Månsson (2000)] and

$$
\begin{aligned}
&E\bigl[\Theta^{r,l}_{\alpha i}(\Gamma \setminus A) + \Phi^l_{\alpha i}(\Gamma \setminus A)\bigr]\\
&\qquad \le |\Gamma^1_{\alpha,i,l}|\gamma_+^k + |\Gamma^2_{\alpha,i,l}| q_1^k / p^k + |\Gamma^3_{\alpha,i,l}| q_2^k / p^k\\
&\qquad\quad + \bigl[|\Gamma^0_{\alpha,i,l}| + |\Gamma^1_{\alpha,i,l}| + |\Gamma^2_{\alpha,i,l}| + |\Gamma^3_{\alpha,i,l}|\bigr] p^k.
\end{aligned}
$$

Letting $(2k + i)^- = (2k + i) \wedge \min\{m, n\}$,

$$
\begin{aligned}
|\Gamma^0_{\alpha,i,l}| &= (2k + i)^-, & |\Gamma^1_{\alpha,i,l}| &= (2k + i)^-((2k + i)^- - 1),\\
|\Gamma^2_{\alpha,i,l}| &= (2k + i)^-(n - (2k + i)^-), & |\Gamma^3_{\alpha,i,l}| &= (2k + i)^-(m - (2k + i)^-),
\end{aligned}
$$

so that

$$T_2 = O\big(mnp^k\big(k^2\gamma_+^k + nkq_1^k/p^k + mkq_2^k/p^k + (m+n)kp^k\big)\big). \tag{5.10}$$

Letting

$$d_0(\alpha, \beta) := \begin{cases} p^k|\alpha_1 - \beta_1| \wedge 1, & \text{if } |\alpha_1 - \beta_1| = |\alpha_2 - \beta_2|, \\ 1, & \text{otherwise,} \end{cases}$$

we get, as in the runs example,

$$\begin{aligned} &E^\alpha\bigg(\int_\Gamma \Xi^{s,\alpha}(d\beta) d_0(\alpha, \beta)\bigg) \\ &\qquad \le 2p^{k+1}(1-p)^{-2} + \min\{m,n\} p^{\min\{m,n\}-k} \mathbb{1}_{\{\alpha=(1,1)\}}(|m-n|+1), \end{aligned}$$

and hence $T_1 \le 2mnp^{2k+1}(1-p)^{-2} + \min\{m,n\}(|m-n|+1)p^{\min\{m,n\}}$. By this inequality, Corollary 4.7, (5.9) and (5.10) we get

$$\begin{aligned} &d_2\big(\mathscr{L}(\Xi), \mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots)\big) \\ &\qquad = O\bigg(\big(k^2\gamma_+^k + nkq_1^k/p^k + mkq_2^k/p^k + (m+n)kp^k\big)\frac{1 + \log^+(mnp^k)}{1-2p}\bigg), \end{aligned} \tag{5.11}$$

uniformly in $p < 1/2$.

An alternative to the above approach would be to use fixed neighborhoods of dependence and Theorem 4.3. If, for each $\alpha \in \Gamma$, we define

$$\begin{aligned} N_\alpha^s &= \big\{(\beta_1, \beta_2) \in \Gamma \backslash \{\alpha\} : -k < \alpha_1 - \beta_1 = \alpha_2 - \beta_2 < k\big\}, \\ N_\alpha^b &= \big\{(\beta_1, \beta_2) \in \Gamma \backslash \big\{\{\alpha\} \cup N_\alpha^s\big\} : |\alpha_1 - \beta_1| < 2k-1 \text{ or } |\alpha_2 - \beta_2| < 2k-1\big\}, \\ N_\alpha^w &= \big\{(\beta_1, \beta_2) \in \Gamma : |\alpha_1 - \beta_1| \ge 2k-1 \text{ and } |\alpha_2 - \beta_2| \ge 2k-1\big\}, \end{aligned}$$

we achieve a bound of the same order as (5.11), but with somewhat larger constants. For this bound to approach zero, thereby verifying a good approximation asymptotically, there are restrictions on the relative growth rate of m and n, and on how different the two distributions τ and ν are allowed to be. These restrictions are unnecessary stringent, as observed in Neuhauser (1996). Barbour and Chryssaphinou (2000) carry out compound Poisson approximation for the total number of, possibly overlapping, matching subsequences of length k, using an approach with fixed neighborhoods. They also use ideas from Neuhauser (1996) in order to get less restrictive conditions for the approximation bounds to tend to zero. Using fixed neighborhoods, a similar refinement could be achieved here as well.

5.4. *Rare sets in Markov chains.* Let $Y := (Y_t, t \in \mathbb{Z})$ be an irreducible, positive recurrent Markov chain with stationary distribution ψ. Fix $s_0 \in \mathbb{Z}$, preferably with $\psi_0 := \psi\{s_0\}$ relatively large, since it is to be used to define a sequence of regeneration points; and a subset $S_1 \subset \mathbb{Z} \setminus \{s_0\}$ of actual interest, which is "rare" in the sense that $\psi_1 := \psi(S_1)$ is small. Take $\Gamma = \{1, 2, \dots, n\}$, and define

$$\Xi := \sum_{\alpha \in \Gamma} I[Y_\alpha \in S_1]\delta_\alpha,$$

the point process of visits to S_1. Erhardsson (1999), in the more general context of Harris recurrent Markov chains, combines regeneration arguments, coupling and Stein's method in elegant fashion to show that $\mathcal{L}(\Xi(\Gamma))$ can be approximated in total variation by a compound Poisson distribution on $\mathbb{Z}_+$. Here, we consider d_2-approximation of the distribution $\mathcal{L}(\Xi)$ of the whole process, when Y is stationary, by way of Theorem 4.2 and couplings. As in Remark 2 following Theorem 4.2, a natural choice for the metric d_0 is given by

$$d_0(\alpha, \beta) := \min\{1, \psi_1|\alpha - \beta|\}. \tag{5.12}$$

To use Theorem 4.2, we first need to specify a decomposition family $\mathcal{N}$. Define

$$\begin{aligned} \tau_{s_0}^{\alpha,+} &:= \tau_{s_0}^{\alpha,+}(Y) := \min\{j \geq 0 : Y_{\alpha+j} = s_0\}, \\ \tau_{s_0}^{\alpha,-} &:= \tau_{s_0}^{\alpha,-}(Y) := \min\{j \geq 0 : Y_{\alpha-j} = s_0\}, \end{aligned} \tag{5.13}$$

so that $\alpha - \tau_{s_0}^{\alpha,-}$ is the index of the last visit of Y to s_0 before α and $\alpha + \tau_{s_0}^{\alpha,+}$ that of the first visit after α, and both are equal to α if $Y_\alpha = s_0$. Set

$$\Xi^{s,\alpha} := \sum_{\substack{\beta \in \Gamma \\ \beta \neq \alpha}} I[Y_\beta \in S_1] I\big[\alpha - \tau_{s_0}^{\alpha,-} < \beta < \alpha + \tau_{s_0}^{\alpha,+}\big]\delta_\beta, \tag{5.14}$$

and set $\Xi^{r,\alpha} := \Xi - \Xi^{s,\alpha}$. With this decomposition, from (2.4) it follows that

$$i\boldsymbol{\pi}_i\{\alpha\} := \psi_1 P\left(\sum_{j=-\alpha+1}^{n-\alpha} I[Y_j \in S_1] I\big[-\tau_{s_0}^{0,-} < j < \tau_{s_0}^{0,+}\big] = i \big| Y_0 \in S_1\right). \tag{5.15}$$

Note that, because of edge effects, the $\boldsymbol{\pi}_i\{\alpha\}$ are not equal for all α, but that $i\boldsymbol{\pi}_i\{\alpha\}$ is mostly close to

$$i\rho_i := \psi_1 P\left(\sum_{j \in \mathbb{Z}} I[Y_j \in S_1] I\big[-\tau_{s_0}^{0,-} < j < \tau_{s_0}^{0,+}\big] = i \big| Y_0 \in S_1\right), \tag{5.16}$$

suggesting approximating the stationary point process Ξ by $\mathrm{CP}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \dots)$ with $\boldsymbol{\rho}_i := \rho_i \boldsymbol{\nu}$ for $\boldsymbol{\nu}$ counting measure on Γ. To shorten the remaining discussion, we shall restrict ourselves to the case where $\sigma_1 := n(\rho_1 - 2\rho_2) > 0$ and $i\rho_i$ decreases with i. This is so, for instance, if S_1 is a singleton, when $i\rho_i = \psi_1 i p^{i-1}(1-p)^2$ for $p = P_{S_1}(\tau(S_1) < \tau_{s_0}^{0,+})$, provided that $p < 1/2$; here, $\tau(S_1)$ denotes the first strictly positive index at which S_1 is visited, and P_{S_1} denotes probability conditional on $Y_0 \in S_1$.

We consider two couplings of Y-processes which can be used to realize $\Theta^r_{\alpha i}$ and $\Phi_{\alpha i}$ of Theorem 4.2, in each of which we also condition on the values l and m taken by $\tau^{\alpha,-}_{s_0}$ and $\tau^{\alpha,+}_{s_0}$ respectively. In the first of them, very much as in Erhardsson (1999), we build the coupled processes $Y^{(1)}$ and $Y^{(2)}$ out of three independent realizations of Y-processes: a stationary realization Y^{ψ}, a realization $Y^{(0)}$ conditional on $Y_0 = s_0$ and a realization $Y^{(S_1)}$ conditional on $Y_\alpha \in S_1$, $\tau^{\alpha,-}_{s_0} = l$, $\tau^{\alpha,+}_{s_0} = m$ and $\sum_{\beta\in\Gamma} I[Y_\beta \in S_1] I[\alpha - l < \beta < \alpha + m] = i$. We set

$$(5.17) \qquad Y^{(1)}_t := \begin{cases} Y^{(S_1)}_t, & \text{if } t \in \Gamma \cap [\alpha - l, \alpha + m], \\ Y^{(0)}_t, & \text{if } t \in [1, \alpha - l), \\ Y^{(0)}_t, & \text{if } t \in (\alpha + m, n], \end{cases}$$

and on $\tau^{\alpha,-}_{s_0}(Y^{\psi}) = l'$, $\tau^{\alpha,+}_{s_0}(Y^{\psi}) = m'$ define

$$(5.18) \qquad Y^{(2)}_t := \begin{cases} Y^{\psi}_t, & \text{if } t \in \Gamma \cap [\alpha - l', \alpha + m'], \\ Y^{(0)}_t, & \text{if } t \in [1, \alpha - l'), \\ Y^{(0)}_t, & \text{if } t \in (\alpha + m', n], \end{cases}$$

for each $l', m' \geq 0$. Finally, set

$$\Theta^{r,l,m}_{\alpha i} := \sum_{\substack{\beta\in\Gamma \\ \beta\notin(\alpha-l,\alpha+m)}} I[Y^{(1)}_\beta \in S_1]\delta_\beta; \qquad \Phi_{\alpha i} := \sum_{\beta\in\Gamma} I[Y^{(2)}_\beta \in S_1]\delta_\beta.$$

With this coupling, the processes $\Theta^{r,l,m}_{\alpha i}$ and $\Phi_{\alpha i}$ are the restrictions to Γ of point processes which consist of left- and right-hand pieces which are identical except for a shift, together with middle segments which are different in both length and measure: the stretch from $\alpha - l$ to $\alpha + m$ is empty in $\Theta^{r,l,m}_{\alpha i}$, whereas the stretch from $\alpha - \tau^{\alpha,-}_{s_0}(Y^{\psi})$ to $\alpha + \tau^{\alpha,+}_{s_0}(Y^{\psi})$ in $\Phi_{\alpha i}$ need not be. Hence the discrepancy $|\Theta^{r,l,m}_{\alpha i}(\Gamma) - \Phi_{\alpha i}(\Gamma)|$ consists of a contribution from the differing middle segments, together with contributions from points of the left- and right-hand pieces which, because of the shifts, may belong to Γ for one but not both of $\Theta^{r,l,m}_{\alpha i}$ and $\Phi_{\alpha i}$. The calculations made in Erhardsson [(1999), Theorem 4.3] suffice to bound this contribution to T_2 of Theorem 4.2:

$$(5.19) \qquad \sum_{i\geq 1} i \sum_{\alpha\in\Gamma} E|\Theta^r_{\alpha i}(\Gamma) - \Phi_{\alpha i}(\Gamma)| \boldsymbol{\pi}_i\{\alpha\} \leq 2n\psi_1^2\{c_0 + c_1\},$$

where

$$c_0 := E_\psi(\tau^{0,+}_{s_0})/\psi_0 \quad \text{and} \quad c_1 := E_{S_1}\{\tau^{0,+}_{s_0} + \tau^{0,-}_{s_0}\}.$$

For the remaining elements of Theorem 4.2, it can be shown that

$$T_1 \le 2\psi_1^2 \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma} E\{|\beta-\alpha| I[Y_\beta\in S_1] I[\alpha-\tau_{s_0}^{\alpha,-}<\beta<\alpha+\tau_{s_0}^{\alpha,+}] \,|\, Y_\alpha\in S_1\} \tag{5.20}$$
$$\le 2n\psi_1^2 c_2,$$

where $c_2 := E_{S_1}\{(\tau_{s_0}^{0,+})^2+(\tau_{s_0}^{0,-})^2\}$; and since matched points in $\Theta_{\alpha i}^{r,l,m}$ and $\Phi_{\alpha i}$ are at most a d_0-distance $\psi_1\max\{|l-\tau_{s_0}^{\alpha,-}(Y^\psi)|, |m-\tau_{s_0}^{\alpha,+}(Y^\psi)|\}$ apart, it follows that the remaining contribution to T_2 satisfies

$$2\sum_{i\ge1} i\sum_{\alpha\in\Gamma}\boldsymbol{\pi}_i\{\alpha\}E\{\min(\Theta_{\alpha i}^r(\Gamma),\Phi_{\alpha i}(\Gamma))\hat d_1(\Theta_{\alpha i}^r,\Phi_{\alpha i})\} \tag{5.21}$$
$$\le 2n^2\psi_1^3\{2c_0+c_1\}.$$

Finally, using (5.15) and (5.16), we compute

$$\sum_{\alpha\in\Gamma}\sum_{i\ge1}|i\boldsymbol{\pi}_i\{\alpha\}-i\rho_i|\le\psi_1c_2, \tag{5.22}$$

so that, by (4.21) and Lemma 3.5(i), the extra contribution involved in using $\mathrm{CP}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\dots)$ in place of $\mathrm{CP}(\boldsymbol{\pi}_1,\boldsymbol{\pi}_2,\dots)$ is no more than $1.65\sigma_1^{-1/2}\psi_1c_2$.

Putting the estimates (5.19)–(5.21) into Theorem 4.2, it follows that if $i\rho_i$ decreases with i and $\rho_1>2\rho_2$ then

$$d_2(\mathscr{L}(\Xi),\mathrm{CP}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\dots))$$
$$\le 2\sigma_1^{-1}(1+\log^+\sigma_1)n\psi_1^2\{2(c_0+c_1)+2c_2+2n\psi_1(2c_0+c_1)\} \tag{5.23}$$
$$+\ 1.65\sigma_1^{-1/2}c_2\psi_1,$$

with $\sigma_1=n(\rho_1-2\rho_2)\asymp n\psi_1$. The estimate is small provided that the quantity $\psi_1\log(n\psi_1)\{1+n\psi_1\}$ is small, and that $c_2=E_{S_1}\{(\tau_{s_0}^{0,+})^2+(\tau_{s_0}^{0,-})^2\}<\infty$: in contrast, the Erhardsson (1999) bound on the approximation to $\mathscr{L}(\Xi(\Gamma))$ only involves the first moments of the hitting times being finite.

The element of order $n\psi_1^2\log(n\psi_1)$ arises from the general shift of points involved in matching $\Theta_{\alpha i}^r$ to $\Phi_{\alpha i}$, and is the major contribution whenever $E(\Xi(\Gamma))=n\psi_1$ is large. A more appropriate coupling, in which points are generally matched exactly, without shifts, can eliminate it. The construction is as follows: $Y^{(2')}$ is taken simply to be Y^ψ, and, with l,m and i as before, define

$$Y_t^{(1')} := \begin{cases} Y_t^{(S_1)}, & \text{if } t\in\Gamma\cap[\alpha-l,\alpha+m],\\ Y_t^{(3)}, & \text{if } t\in[1,\alpha-l),\\ Y_t^{(4)}, & \text{if } t\in(\alpha+m,n],\end{cases} \tag{5.24}$$

where now $(Y_t^{(3)},\,t\le\alpha-l)$ is a reversed Y-process starting with $Y_{\alpha-l}^{(3)}=s_0$, coupled to $(Y_t^\psi,\,t\le\alpha-l)$ so as eventually to coincide, and $(Y_t^{(4)},\,t\ge\alpha+m)$ is a Y-process starting with $Y^{(4)}_{\alpha+m} = s_0$, coupled to $(Y^\psi_t,\ t \ge \alpha + m)$ so as eventually to coincide. Let the coupling time of the reversed Y-chains be denoted by T^-, that of the forward chains by T^+. Then set

$$\Theta^{r,l,m}_{\alpha i} := \sum_{\substack{\beta\in\Gamma\\ \beta\notin(\alpha-l,\alpha+m)}} I[Y^{(1')}_\beta \in S_1]\delta_\beta; \qquad \Phi_{\alpha i} := \sum_{\beta\in\Gamma} I[Y^{(2')}_\beta \in S_1]\delta_\beta.$$

This again generates the right distributions, but now $\Theta^{r,l,m}_{\alpha i}(A) = \Phi_{\alpha i}(A)$ for all $A \subset \Gamma \setminus [\alpha - T^-, \alpha + T^+]$, so that

$$\begin{aligned}\|\Phi_{\alpha i} - \Theta^{r,l,m}_{\alpha i}\| &\\ \le \sum_{\beta\in\mathbb{Z}} \bigl\{ I[Y^{(1')}_\beta \in S_1]&(I[\alpha - T^- < \beta < \alpha - l] + I[\alpha + m < \beta < \alpha + T^+]) \\ &+ I[Y^{(2')}_\beta \in S_1] I[\alpha - T^- < \beta < \alpha + T^+]\bigr\}.\end{aligned}$$

Integrating out the conditioning on the values of $\tau^{\alpha,-}_{s_0}$ and $\tau^{\alpha,+}_{s_0}$, it follows that

$$E\|\Theta^r_{\alpha i} - \Phi_{\alpha i}\| \le \psi_1\bigl(c_3 + E^\alpha\{\tau^{\alpha,-}_{s_0} + \tau^{\alpha,+}_{s_0} \mid \Xi^{s,\alpha}(\Gamma) + \Xi\{\alpha\} = i\}\bigr),$$

where $c_3 := \psi_1^{-1}\{e^{(1)}_+ + e^{(1)}_- + e^{(2)}_+ + e^{(2)}_-\}$ and where $e^{(1)}_+$ and $e^{(2)}_+$ are respectively the expected numbers of visits to S_1 by coupled P_{s_0}- and P_ψ-chains before coupling, and $e^{(1)}_-$ and $e^{(2)}_-$ are the corresponding quantities in the reversed chains; all four are constant in i and α. Hence

$$\sum_{i\ge1} i\boldsymbol{\pi}_i\{\alpha\} E\|\Theta^r_{\alpha i} - \Phi_{\alpha i}\| \le (c_1 + c_3)\psi_1^2.$$

Thus the element $2(c_0 + c_1) + 2n\psi_1(2c_0 + c_1)$, appearing in (5.23) as a bound for T_2, can be replaced by $2(c_1 + c_3)$.

It remains to be shown that couplings can be found such that c_3 is not automatically large when ψ_1 is small. One example can be constructed as follows. Suppose that P_ψ- and P_{s_0}-chains are run independently, and that the coupling occurs at the first time T^+ that they are simultaneously in s_0. Taking the P_{s_0}-chain first, consider the set $\mathcal{T}$ of times at which the P_ψ-chain hits s_0 as given; then T^+ corresponds to a stopping time for the process $C_1, C_2, \ldots$ of s_0-regeneration cycles of the P_{s_0}-chain, and hence, by Wald's identity, it follows that $e^{(1)}_+ = \psi_1 E(T^+)$. Arguing similarly for the other three pieces, and taking into account the parts of the P_ψ-chains before their first visit to s_0, we obtain

$$c_3 \le 4E(T^+) + c_1.$$

This is enough, provided that $E(T^+) < \infty$, which, by Chapter II.4 in Lindvall (1992), is the case if $E_\psi[(\tau^{0,+}_{s_0})^2] < \infty$ and $E_{s_0}[\tau^{0,+}_{s_0}] < \infty$, and if Y is aperiodic. Better couplings can be expected to yield sharper bounds.

If Y has period r, an argument similar to that above can still be used. Defining

$$\Xi^{(s)} := \sum_{\beta\in\Gamma} I[Y_{\beta-s} \in S_1]\delta_\beta$$

to be the s-shift of Ξ on Γ, it follows that

$$d_1(\Xi^{(s)}, \Xi) \le \sum_{\beta=-s+1}^{0} I[Y_{\beta-s} \in S_1] + \sum_{\beta=n-s+1}^{n} I[Y_\beta \in S_1] + s\psi_1,$$

so that $d_2(\mathcal{L}(\Xi^{(s)}), \mathcal{L}(\Xi)) \le 3s\psi_1$. Thus, for a contribution of at most $3(r-1)\psi_1$ to the bound, $\mathcal{L}(\Xi^*)$ can be approximated in place of $\mathcal{L}(\Xi)$, where Ξ^* is constructed from a Y-process under P_{ψ^*}, where ψ^* is ψ conditioned to that of the r periodic sets which contains s_0. For this process, the argument runs much as above, and since the chains to be coupled are now synchronized with respect to the period, the coupling is successful.

Acknowledgments. The authors would like to thank Torgny Lindvall for a valuable discussion on coupling times in the rare sets example, and the referee for a number of helpful comments. The second author also acknowledges the warm hospitality of the Abteilung für Angewandte Mathematik, Universität Zürich, during a visit supported by the Swiss Nationalfonds.

REFERENCES

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

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 BROWN, T. C. (1992). Stein's method and point process approximation. *Stochastic Processes Appl.* **43** 9–31.

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 CHRYSSAPHINOU, O. (2001). Compound Poisson approximation: A user's guide. *Ann. Appl. Probab.* **11** 964–1002.

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

BARBOUR, A. D., NOVAK, S. Y. and XIA, A. (1999). Compound Poisson approximation for the distribution of extremes. TU Eindhoven, EURANDOM Research Report 99-040.

BARBOUR, A. D. and UTEV, S. (1998). Solving the Stein Equation in compound Poisson approximation. *Adv. Appl. Probab.* **30** 449–475.

BARBOUR, A. D. and UTEV, S. (1999). Compound Poisson approximation in total variation. *Stochastic Process. Appl.* **82** 89–125.

BROWN, T. and XIA, A. (2001). Stein's method and birth–death processes. *Ann. Probab.* **29** 1373–1403.

ERHARDSSON, T. (1999). Compound Poisson approximation for Markov chains using Stein's method. *Ann. Probab.* **27** 565–596.

KALLENBERG, O. (1983). *Random Measures,* 3rd ed. Akademie, Berlin.

LINDVALL, T. (1992). *Lectures on the Coupling Method.* Wiley, New York.

MÅNSSON, M. (2000). On compound Poisson approximation for sequence matching. *Combin. Probab. Comput.* **9** 529–548.

NEUHAUSER, C. (1996). A phase transition for the distribution of matching blocks. *Combin. Probab. Comput.* **5** 139–159.

ABTEILUNG FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT ZÜRICH
WINTERTHURERSTRASSE 190
CH-8057 ZÜRICH
SWITZERLAND
E-MAIL: adb@amath.unizh.ch

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