

A NEW METRIC BETWEEN DISTRIBUTIONS OF POINT PROCESSES

DOMINIC SCHUHMACHER,* *The University of Western Australia*

AIHUA XIA,** *The University of Melbourne*

Abstract

Most metrics between finite point measures currently used in the literature have the flaw that they do not treat differing total masses in an adequate manner for applications. This paper introduces a new metric \bar{d}_1 that combines positional differences of points under a closest match with the relative difference in total mass in a way that fixes this flaw. A comprehensive collection of theoretical results about \bar{d}_1 and its induced Wasserstein metric \bar{d}_2 for point process distributions are given, including examples of useful \bar{d}_1 -Lipschitz continuous functions, \bar{d}_2 upper bounds for the Poisson process approximation, and \bar{d}_2 upper and lower bounds between distributions of point processes of independent and identically distributed points. Furthermore, we present a statistical test for multiple point pattern data that demonstrates the potential of \bar{d}_1 in applications.

Keywords: Wasserstein metric; point process; Poisson point process; Stein's method; distributional approximation; statistical analysis of point pattern data

2000 Mathematics Subject Classification: Primary 60G55

Secondary 60F05; 62M30

1. Introduction

The two metrics most widely used on the space \mathfrak{N} of finite point measures on a compact metric space (X, d_0) are the Prohorov metric ϱ and the metric d_1 that was introduced in [3]. We use δ_x to stand for the Dirac measure at x . For $\xi = \sum_{i=1}^m \delta_{x_i}$, $\eta = \sum_{i=1}^n \delta_{y_i} \in \mathfrak{N}$, and $d_0 \leq 1$, the metric d_1 is given by

$$d_1(\xi, \eta) = \min_{\pi \in \Pi_n} \frac{1}{n} \sum_{i=1}^n d_0(x_i, y_{\pi(i)}) \quad (1.1)$$

if $m = n \geq 1$ and $d_1(\xi, \eta) = 1$ if $m \neq n$, where Π_n denotes the set of permutations of $\{1, 2, \dots, n\}$. The gap between $d_1 = d_1^{(1)}$ and $\varrho \wedge 1 = d_1^{(\infty)}$ can be bridged by metrics $d_1^{(p)}$, where the average in (1.1) is replaced by a general p th order average (see [23]).

All of these metrics are good choices from a theoretical point of view, because they metrize the natural vague topology on \mathfrak{N} . Furthermore, d_1 has especially been highly successful as an underlying metric for defining a Wasserstein metric d_2 between point process distributions: letting $\mathcal{F}_2 = \{f: \mathfrak{N} \rightarrow [0, 1]; |f(\xi) - f(\eta)| \leq d_1(\xi, \eta) \text{ for all } \xi, \eta \in \mathfrak{N}\}$, we set

$$d_2(P, Q) = \sup_{f \in \mathcal{F}_2} \left| \int f \, dP - \int f \, dQ \right|$$

Received 21 August 2007; revision received 26 March 2008.

* Current address: Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, CH-3012 Bern, Switzerland. Email address: dominic.schuhmacher@stat.unibe.ch

** Postal address: Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. Email address: xia@ms.unimelb.edu.au

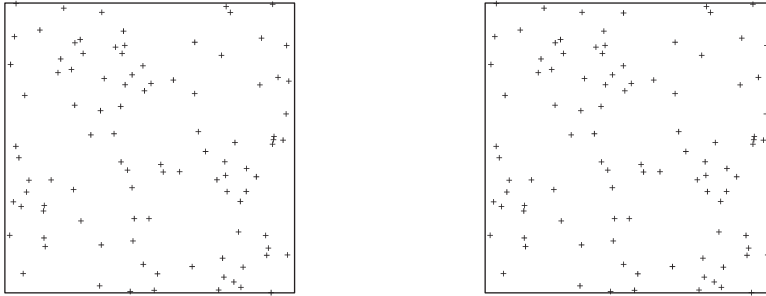


FIGURE 1: The left is a realization of 99 independent and uniformly distributed points and the right is the same as the left except that an additional point is added. Intuitively, we would say both point patterns are very similar. However, the d_1 -distance between the two is maximal, whereas the \bar{d}_1 -distance is only 0.01 (out of a possible range of $[0, 1]$).

for any two probability measures P and Q on \mathfrak{N} . Numerous useful upper bounds in this metric have been obtained, including the results of [3], [6], [9], [11], [12], [22], and [23], which for the most part assume that one of the probability measures involved is a Poisson (or compound Poisson) process distribution. Such estimates can be used to compare the distributions of point pattern statistics $S(\Xi)$, where $S \in \mathcal{F}_2$, for different underlying point process models, since the Wasserstein distance $d_W(\mathcal{L}(S(\Xi)), \mathcal{L}(S(\Xi')))$ (see [8, pp. 254–255]) is easily seen to be bounded by $d_2(\mathcal{L}(\Xi), \mathcal{L}(\Xi'))$. For a concrete example where this was exploited, see [21, Section 3.2].

However, there are certain limitations with respect to the practical applications of the metric d_1 (as well as of the other metrics between point measures that were mentioned), which are mainly due to the fact that $d_1(\xi, \eta)$ is always set to the maximal distance 1 if the total numbers of points of the point patterns ξ and η disagree. Such crude treatment results in a metric that does not usually reflect very well our intuitive idea of two-point patterns being ‘far apart’ from one another if the cardinalities of the point patterns are different, as can be seen from the extreme case illustrated in Figure 1. This flaw is, in our opinion, the main reason why such metrics have not been taken up in more application-oriented fields, such as spatial statistics.

In the present paper we introduce a new metric \bar{d}_1 , which refines the metric d_1 in the sense that $\bar{d}_1(\xi, \eta) = d_1(\xi, \eta)$ if the cardinalities of the two-point patterns ξ and η agree, but $\bar{d}_1(\xi, \eta)$ can take general values in $(0, 1]$ if the cardinalities disagree. In particular, \bar{d}_1 assigns a large distance if the difference in the numbers of points is large compared to the total number of points in the point pattern with more points, and it takes the quality of point matchings into account even if the total numbers are not the same.

While \bar{d}_1 is a slightly weaker metric than d_1 , it still metrizes the same topology as d_1 , and its induced Wasserstein metric \bar{d}_2 still metrizes convergence in distribution of point processes and provides an upper bound for the Wasserstein distance $d_W(\mathcal{L}(S(\Xi)), \mathcal{L}(S(\Xi')))$ for many of the useful point-pattern statistics S that d_2 does. As far as Poisson process approximation is concerned, we are able to obtain better bounds in the \bar{d}_2 -metric than in the stronger d_2 -metric for a wide range of situations. Furthermore, we present a simulation study that assesses the powers of certain tests based on \bar{d}_1 and demonstrates its usefulness in spatial statistics.

2. Definition and elementary properties

Let (X, d_0) be a compact metric space with $d_0 \leq 1$, on which we always consider the Borel σ -algebra \mathcal{B} . Denote the space of all finite point measures on X by \mathfrak{N} and equip it as usual with the vague topology and the σ -algebra \mathcal{N} generated by this topology, which is the smallest σ -algebra that renders the point counts on measurable sets measurable (see [15, Section 1.1, Lemma 4.1, and Section 15.7]). Recall that a point process is just a random element of \mathfrak{N} .

Definition. Let \bar{d}_1 be the symmetric map $\mathfrak{N}^2 \rightarrow \mathbb{R}_+$ that is given by

$$\bar{d}_1(\xi, \eta) = \frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^m d_0(x_i, y_{\pi(i)}) + (n - m) \right)$$

for $\xi = \sum_{i=1}^m \delta_{x_i}, \eta = \sum_{j=1}^n \delta_{y_j} \in \mathfrak{N}$ with $n \geq \max(m, 1)$, and $\bar{d}_1(0, 0) = 0$.

In essence, we arrange for ξ and η to have the same number of points by introducing extra points located at distance 1 from X , and then take the average distance between the points under a closest match (which is the d_1 -distance).

Proposition 2.1. *The map \bar{d}_1 is a metric that is bounded by 1.*

The proof of this proposition, as well as further proofs that are of a more technical nature and would otherwise disrupt the flow of the main text can be found in Appendix A. It is convenient to introduce the ‘relative difference metric’ d_R on \mathbb{Z}_+ , which is given by $d_R(m, n) = |m - n| / \max(m, n)$ for $\max(m, n) > 0$. The triangle inequality for d_R follows immediately from the triangle inequality for \bar{d}_1 , because we have $d_R(m, n) = \bar{d}_1(m\delta_x, n\delta_x)$.

Proposition 2.2. *The following statements about \bar{d}_1 hold.*

- (i) $d_R(|\xi|, |\eta|) \leq \bar{d}_1(\xi, \eta) \leq d_1(\xi, \eta)$ for all $\xi, \eta \in \mathfrak{N}$.
- (ii) \bar{d}_1 metrizes the vague (weak) topology on \mathfrak{N} .
- (iii) The metric space $(\mathfrak{N}, \bar{d}_1)$ is locally compact, complete, and separable.

Next we define the metric \bar{d}_2 on the space $\mathfrak{P}(\mathfrak{N})$ of probability distributions on $(\mathfrak{N}, \mathcal{N})$ just as the Wasserstein metric with respect to \bar{d}_1 .

Definition. Let $\bar{\mathcal{F}}_2 = \{f : \mathfrak{N} \rightarrow [0, 1]; |f(\xi) - f(\eta)| \leq \bar{d}_1(\xi, \eta) \text{ for all } \xi, \eta \in \mathfrak{N}\}$. Then set

$$\bar{d}_2(P, Q) = \sup_{f \in \bar{\mathcal{F}}_2} \left| \int_{\mathfrak{N}} f \, dP - \int_{\mathfrak{N}} f \, dQ \right|$$

for $P, Q \in \mathfrak{P}(\mathfrak{N})$.

Since this is exactly the Wasserstein construction (the fact that we restrict the functions in $\bar{\mathcal{F}}_2$ to be $[0, 1]$ -valued has no influence on the supremum, because the underlying \bar{d}_1 -metric is bounded by 1), it is clear that \bar{d}_2 is a metric that is obviously bounded by 1, and we can easily derive basic properties. For two probability distributions μ and ν on \mathbb{Z}_+ , write $d_{RW}(\mu, \nu) = \min_{M \sim \mu, N \sim \nu} E[d_R(M, N)]$, which is the Wasserstein distance with respect to d_R (compare Proposition 2.3(i), below). The minimum exists because (\mathbb{Z}_+, d_R) is complete (compare the proof of Proposition 2.3(i) in Appendix A).

Proposition 2.3. *The metric \bar{d}_2 satisfies*

- (i) $\bar{d}_2(P, Q) = \min_{\Xi \sim P, H \sim Q} E[\bar{d}_1(\Xi, H)]$ for all $P, Q \in \mathfrak{P}(\mathfrak{N})$;
- (ii) $d_{RW}(\mathcal{L}(|\Xi|), \mathcal{L}(|H|)) \leq \bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H)) \leq d_2(\mathcal{L}(\Xi), \mathcal{L}(H))$ for any point processes Ξ and H ;
- (iii) \bar{d}_2 metrizes the weak topology on $\mathfrak{P}(\mathfrak{N})$, so that $\Xi_n \xrightarrow{D} \Xi$ if and only if $\bar{d}_2(\mathcal{L}(\Xi_n), \mathcal{L}(\Xi)) \rightarrow 0$ as $n \rightarrow \infty$, where ‘ \xrightarrow{D} ’ denotes convergence in distribution.

3. Lipschitz continuous functions

By the definition of \bar{d}_2 , upper bounds for a distance $\bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H))$ also bound the difference $|E[f(\Xi)] - E[f(H)]|$ for any $f \in \mathcal{F}_2$. It is thus of considerable interest for the application of estimates such as those obtained in Section 4 to have a certain supply of ‘meaningful’ \bar{d}_1 -Lipschitz continuous statistics of point patterns (where we do not worry too much about the Lipschitz constant, as it will only appear as an additional factor in the upper bound).

For the case of the d_1 -metric, a selection of such statistics was given in Section 10.2 of [8] and in Subsection 3.3.1 of [20]. Since \bar{d}_1 is in general strictly smaller than d_1 , we cannot reasonably expect all of these functions to lie in \mathcal{F}_2 . However, we are able to recover many of the most important examples, which is illustrated by the two propositions below. This is mainly due to the fact that these functions take all the points in the pattern into account without fundamentally distinguishing how many there are, which is a situation where a d_1 -Lipschitz condition typically provides too much room in the upper bound.

Our first proposition concerns certain U -statistics with Lipschitz continuous kernels (the former are usually considered for a fixed number of points, but the extension is obvious). See [16] for detailed results about such statistics.

Proposition 3.1. *Suppose that $\mathcal{Y} \supset \mathcal{X}$, and extend the metric d_0 to \mathcal{Y} in such a way that it is still bounded by 1. Fix $l \in \mathbb{N} = \{1, 2, \dots\}$, and write $\mathfrak{N}_{l+} = \{\xi \in \mathfrak{N}; |\xi| \geq l\}$. Let $K : \mathcal{Y}^l \rightarrow [0, 1]$ be a symmetric function that satisfies*

(i)

$$|K(u_1, \dots, u_l) - K(v_1, \dots, v_l)| \leq \frac{1}{l} \sum_{i=1}^l d_0(u_i, v_i)$$

for $u_1, \dots, u_l, v_1, \dots, v_l \in \mathcal{Y}$;

(ii) for every $N \in \mathbb{N}$, there exist $\bar{u}_1, \dots, \bar{u}_N \in \mathcal{Y}$ such that, for any $k \in \{1, \dots, l\}$ and any selection $1 \leq i_1 < \dots < i_k \leq N$ of k indices,

$$K(\bar{u}_{i_1}, \bar{u}_{i_2}, \dots, \bar{u}_{i_k}, u_{k+1}, u_{k+2}, \dots, u_l) \geq K(u_1, u_2, \dots, u_k, u_{k+1}, u_{k+2}, \dots, u_l)$$

for all $u_1, u_2, \dots, u_l \in \mathcal{X}$;

(iii) for every $k \in \{1, \dots, l\}$, we have

$$K(u_1, u_1, \dots, u_1, u_{k+1}, u_{k+2}, \dots, u_l) \leq K(u_1, u_2, \dots, u_k, u_{k+1}, u_{k+2}, \dots, u_l)$$

for all $u_1, u_2, \dots, u_l \in \mathcal{X}$.

Define $f: \mathfrak{N}_{l+} \rightarrow [0, 1]$ by

$$f(\xi) = \binom{m}{l}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_l \leq m} K(x_{i_1}, x_{i_2}, \dots, x_{i_l})$$

for $\xi = \sum_{i=1}^m \delta_{x_i} \in \mathfrak{N}$ with $m \geq l$. Then there exists an extension F of f to the whole of \mathfrak{N} such that $F \in \mathcal{F}_2$.

One possible choice for the function K in the above result is half the interpoint distance, i.e. $K(u_1, u_2) = \frac{1}{2}d_0(u_1, u_2)$ for all $u_1, u_2 \in \mathcal{X}$. If $\mathcal{X} \subset \mathbb{R}^D = \mathcal{Y}$ for some $D \in \mathbb{N}$ and $d_0(x, y) = |x - y| \wedge 1$ for all $x, y \in \mathbb{R}^D$, we can consider more generally the diameter of the minimal bounding ball, defining

$$K(u_1, \dots, u_l) = \frac{1}{l} \min\{\text{diam}_0(B); B \subset \mathbb{R}^D \text{ closed Euclidean ball with } u_1, \dots, u_l \in B\}$$

for $l \geq 2$ and $u_1, \dots, u_l \in \mathbb{R}^D$, where $\text{diam}_0(B) = \sup\{d_0(x, y); x, y \in B\}$. It can be shown that this again yields a function K that satisfies (i)–(iii) of Proposition 3.1.

The second proposition looks at the average nearest-neighbor distance in a finite point pattern on \mathbb{R}^D . This statistic gives important information about the amount of clustering in the pattern.

Proposition 3.2. Let $\mathcal{X} \subset \mathbb{R}^D$, and let $d_0(x, y) = |x - y| \wedge 1$ for all $x, y \in \mathbb{R}^D$. Define the function $f: \mathfrak{N}_{2+} \rightarrow [0, 1]$ by

$$f(\xi) = \frac{1}{m} \sum_{i=1}^m \min_{\substack{j \in \{1, \dots, m\} \\ j \neq i}} d_0(x_i, x_j)$$

for $\xi = \sum_{i=1}^m \delta_{x_i} \in \mathfrak{N}$ with $m \geq 2$. Then there exists an extension F of f to the whole of \mathfrak{N} that is \bar{d}_1 -Lipschitz continuous with constant $\tau_D + 1$, where τ_D denotes the kissing number in D dimensions (i.e. the maximal number of unit balls that can touch a unit ball in \mathbb{R}^D without producing any overlaps of the interiors; see [13, Section 1.2] for details).

Proof of Proposition 3.1. Fix a point $x_0 \in \mathcal{X}$, and define

$$F(\xi') = f(\xi' + (l - |\xi'|)^+ \delta_{x_0})$$

for every $\xi' \in \mathfrak{N}$. It suffices to show that $|f(\xi) - f(\eta)| \leq \bar{d}_1(\xi, \eta)$ for $\xi, \eta \in \mathfrak{N}$ with $|\xi|, |\eta| \geq l$, because this implies that

$$\begin{aligned} |F(\xi') - F(\eta')| &= |f(\xi' + (l - |\xi'|)^+ \delta_{x_0}) - f(\eta' + (l - |\eta'|)^+ \delta_{x_0})| \\ &\leq \bar{d}_1(\xi' + (l - |\xi'|)^+ \delta_{x_0}, \eta' + (l - |\eta'|)^+ \delta_{x_0}) \\ &\leq \bar{d}_1(\xi', \eta') \end{aligned}$$

for every $\xi', \eta' \in \mathfrak{N}$. Then let $\xi = \sum_{i=1}^m \delta_{x_i}$ and $\eta = \sum_{i=1}^n \delta_{y_i}$, where $m, n \geq l$ and, without loss of generality, $m \leq n$ (because of the symmetry of the inequality that we would like to show). We add $n - m$ points x_{m+1}, \dots, x_n to ξ in one of the following two ways depending on whether $f(\xi) \geq f(\eta)$ or $f(\xi) < f(\eta)$, and call the result $\tilde{\xi} = \sum_{i=1}^n \delta_{x_i}$.

If $f(\xi) \geq f(\eta)$, let $x_{m+r} = \bar{u}_r$, $1 \leq r \leq n - m$, for points $\bar{u}_1, \dots, \bar{u}_{n-m}$ chosen as in assumption (ii) with $N = n - m$. It follows that

$$\begin{aligned} f(\bar{\xi}) &= \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} K(x_{i_1}, \dots, x_{i_l}) \\ &= \left(\sum_{j=0}^l \binom{m}{j} \binom{n-m}{l-j} \right)^{-1} \sum_{j=0}^l \sum_{\substack{1 \leq i_1 < \dots < i_j \leq m \\ m+1 \leq i_{j+1} < \dots < i_l \leq n}} K(x_{i_1}, \dots, x_{i_l}) \\ &\geq \binom{m}{l}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq m} K(x_{i_1}, \dots, x_{i_l}) \\ &= f(\xi). \end{aligned}$$

The inequality is a consequence of the fact that $(\sum_{j=0}^l a_j) / (\sum_{j=0}^l b_j) \geq a_l / b_l$ if $a_j / b_j \geq a_l / b_l$ for every j ; and the latter condition holds because, for $\max(0, l - n + m) \leq j \leq l - 1$ (since $a_j = b_j = 0$ if $j < l - n + m$, these pairs can be ignored altogether),

$$\begin{aligned} &\binom{m}{j}^{-1} \binom{n-m}{l-j}^{-1} \sum_{\substack{1 \leq i_1 < \dots < i_j \leq m \\ m+1 \leq i_{j+1} < \dots < i_l \leq n}} K(x_{i_1}, \dots, x_{i_l}) \\ &\geq \binom{m}{j}^{-1} \binom{n-m}{l-j}^{-1} \sum_{\substack{1 \leq i_1 < \dots < i_j \leq m \\ m+1 \leq i_{j+1} < \dots < i_l \leq n}} \binom{m-j}{l-j}^{-1} \\ &\quad \times \sum_{\substack{1 \leq r_{j+1} < \dots < r_l \leq m \\ \{r_{j+1}, \dots, r_l\} \cap \{i_1, \dots, i_j\} = \emptyset}} K(x_{i_1}, \dots, x_{i_j}, x_{r_{j+1}}, \dots, x_{r_l}) \\ &= \binom{m}{j}^{-1} \binom{m-j}{l-j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq m} \sum_{\substack{1 \leq r_{j+1} < \dots < r_l \leq m \\ \{r_{j+1}, \dots, r_l\} \cap \{i_1, \dots, i_j\} = \emptyset}} K(x_{i_1}, \dots, x_{i_j}, x_{r_{j+1}}, \dots, x_{r_l}) \\ &= \binom{l}{j} \binom{m}{j}^{-1} \binom{m-j}{l-j}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq m} K(x_{i_1}, \dots, x_{i_l}) \\ &= \binom{m}{l}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq m} K(x_{i_1}, \dots, x_{i_l}), \end{aligned}$$

where the inequality follows by assumption (ii) and the symmetry of K .

If, on the other hand, $f(\xi) < f(\eta)$, let $x_{m+r} = x_1$, $1 \leq r \leq n - m$. It follows, in exactly the same way as for the first case, only this time with ‘ \geq ’ replaced by ‘ \leq ’ and using assumption (iii) instead of assumption (ii), that $f(\bar{\xi}) \leq f(\xi)$.

In total, we thus obtain

$$|f(\xi) - f(\eta)| \leq |f(\bar{\xi}) - f(\eta)| \leq d_1(\bar{\xi}, \eta) = \bar{d}_1(\bar{\xi}, \eta) \leq \bar{d}_1(\xi, \eta),$$

where the second inequality follows from the d_1 -Lipschitz continuity of the functions considered in Proposition 2.A of [23].

Proof of Proposition 3.2. Fix arbitrary $\alpha_0, \alpha_1 \in [0, 1]$, and define $F(\xi) = \alpha_i$ if $|\xi| = i \in \{0, 1\}$ and $F(\xi) = f(\xi)$ otherwise. Let $\xi = \sum_{i=1}^m \delta_{x_i}$ and $\eta = \sum_{i=1}^n \delta_{y_i}$, where, without loss of generality, we assume that $m \leq n$. Since $|F(\xi) - F(\eta)| \leq 1 \leq (\tau_D + 1)\frac{1}{2} \leq (\tau_D + 1)\bar{d}_1(\xi, \eta)$ if $m \in \{0, 1\}$ and $n > m$, the Lipschitz inequality remains to be shown for $n \geq m \geq 2$ only.

As before, we bring the cardinalities to the same level. Let $\bar{\xi} = \sum_{i=1}^n \delta_{x_i}$, where the points x_{m+1}, \dots, x_n are chosen in the following way. If $f(\xi) \geq f(\eta)$, let x_{m+1}, \dots, x_n be arbitrary pairwise distinct points in \mathbb{R}^D that are at d_0 -distance 1 from each other and from \mathcal{X} . Hence, $f(\bar{\xi}) \geq f(\xi)$ because, for each of the added points, the distance to its nearest neighbor is 1, which is maximal. If, on the other hand, $f(\xi) < f(\eta)$, let $x_{m+1} = \dots = x_n = x_1$, whence it is immediately clear that $f(\bar{\xi}) \leq f(\xi)$ because, for each of the added points, the distance to its nearest neighbor is 0.

In total, we obtain

$$|f(\xi) - f(\eta)| \leq |f(\bar{\xi}) - f(\eta)| \leq (\tau_D + 1)d_1(\bar{\xi}, \eta) = (\tau_D + 1)\bar{d}_1(\bar{\xi}, \eta) \leq (\tau_D + 1)\bar{d}_1(\xi, \eta),$$

where the second inequality follows from the d_1 -Lipschitz continuity of the average nearest-neighbor distance considered in Proposition 2.C of [23].

4. Distance estimates in \bar{d}_2

In this section we present upper bounds for some essential \bar{d}_2 -distances, which all clearly improve on the bounds that are available for the corresponding d_2 -distances. However, the improvement in general results of this type is not always as much as we would hope it to be, and it seems that considerably better bounds can be obtained by a more specialized treatment (see, for example, Subsection 4.2).

4.1. Poisson process approximation of a general point process

Using the fact that

$$\mathcal{A}h(\xi) = \int_{\mathcal{X}} (h(\xi + \delta_\alpha) - h(\xi))\hat{\lambda}(d\alpha) + \int_{\mathcal{X}} (h(\xi - \delta_\alpha) - h(\xi))\xi(d\alpha), \quad \xi \in \mathfrak{N}, \quad (4.1)$$

is the generator of the spatial immigration-death process whose steady state distribution is the Poisson process with expectation measure $\hat{\lambda}$, Barbour and Brown [3] established the Stein identity for the Poisson process approximation as

$$\mathcal{A}h(\xi) = f(\xi) - \text{Po}(\hat{\lambda})(f) \quad (4.2)$$

for suitable test functions f on \mathfrak{N} . The solution of (4.2) is given by

$$h_f(\xi) = - \int_0^\infty (\mathbb{E}[f(\mathbb{Z}_\xi(t))] - \text{Po}(\hat{\lambda})(f)) dt, \quad (4.3)$$

where \mathbb{Z}_ξ is an immigration-death process with generator \mathcal{A} and initial point pattern $\mathbb{Z}_\xi(0) = \xi$.

Using (4.2) and different characteristics of point processes, we can establish various versions of Poisson process approximation error bounds (see [3], [4], [7], and [12]). To keep our text concise, we present here a slightly simplified version of the main result in [12] only; it is an obvious exercise to apply our estimates (4.4) and (4.5), below, to get parallel results in the other papers mentioned above.

We assume that, for each $\alpha \in \mathcal{X}$, there is a Borel set $A_\alpha \subset \mathcal{X}$ such that $\alpha \in A_\alpha$ and the mapping

$$\mathcal{X} \times \mathfrak{N} \rightarrow \mathcal{X} \times \mathfrak{N}: (\alpha, \xi) \mapsto (\alpha, \xi|_{A_\alpha^c})$$

is product measurable, where $\xi|_{A_\alpha^c}$ stands for the point pattern of ξ restricted to A_α^c (see [15, Section 1.1]). Such a requirement can be ensured by $A = \{(x, y); y \in A_x, x \in \mathcal{X}\}$ measurable in \mathcal{X}^2 (see [12]). We define, for any function h on \mathfrak{N} ,

$$\begin{aligned} \Delta h(\xi) &= \sup_{\alpha \in \mathcal{X}} |h(\xi + \delta_\alpha) - h(\xi)|, \\ \Delta^2 h(\xi) &= \sup_{\eta - \xi \in \mathfrak{N}, \alpha, \beta \in \mathcal{X}} |h(\eta + \delta_\alpha + \delta_\beta) - h(\eta + \delta_\alpha) - h(\eta + \delta_\beta) + h(\eta)|, \quad \xi \in \mathfrak{N}. \end{aligned}$$

Theorem 4.1. (Chen and Xia [12].) *For each bounded measurable function $f : \mathfrak{N} \rightarrow \mathbb{R}_+$, let h_f be solution (4.3) of (4.2). If Ξ is a point process on \mathcal{X} with expectation measure $\hat{\lambda}$ then*

$$\begin{aligned} &|E[f(\Xi)] - \text{Po}(\hat{\lambda})(f)| \\ &\leq E \int_{\mathcal{X}} \Delta^2 h_f(\Xi|_{A_\alpha^c})(\Xi(A_\alpha) - 1)\Xi(d\alpha) \\ &\quad + E \int_{\mathcal{X}} |(h_f(\Xi|_{A_\alpha^c}) - h_f(\Xi|_{A_\alpha^c} + \delta_\alpha)) - (h_f(\Xi_\alpha|_{A_\alpha^c}) - h_f(\Xi_\alpha|_{A_\alpha^c} + \delta_\alpha))|\hat{\lambda}(d\alpha) \\ &\quad + E \int_{\mathcal{X}} \Delta^2 h_f(\Xi|_{A_\alpha^c})\Xi(A_\alpha)\hat{\lambda}(d\alpha), \end{aligned}$$

where Ξ_α is the Palm process of Ξ at location $\alpha \in \mathcal{X}$ (see [15, Chapter 10]).

Error bounds for the Poisson process approximation as in Theorem 4.1 (see [3], [4], [7], and [12] for full details) pivot on the estimates of Δh_f and $\Delta^2 h_f$. The following proposition summarizes these estimates for \bar{d}_2 .

Proposition 4.1. *Let*

$$\begin{aligned} \Delta h(\xi; \alpha) &= h(\xi + \delta_\alpha) - h(\xi), \\ \Delta^2 h(\xi; \alpha, \beta) &= h(\xi + \delta_\alpha + \delta_\beta) - h(\xi + \delta_\alpha) - h(\xi + \delta_\beta) + h(\xi), \quad \xi \in \mathfrak{N}, \alpha, \beta \in \mathcal{X}. \end{aligned}$$

Then, for each \bar{d}_1 -Lipschitz function f , we have

$$\begin{aligned} |\Delta h_f(\xi; \alpha)| &\leq \min \left\{ 1, \frac{0.95 + \ln^+ \lambda}{\lambda}, \frac{1 - e^{-|\xi| \wedge \lambda}}{|\xi| \wedge \lambda} \right\}, \tag{4.4} \\ |\Delta^2 h_f(\xi; \alpha, \beta)| &\leq \min \left\{ 0.75, \frac{1}{|\xi| \wedge \lambda}, \frac{1.09}{|\xi| + 1} + \frac{1}{\lambda}, \frac{2 \ln \lambda}{\lambda} \mathbf{1}_{\{\lambda \geq 1.76\}} + 0.75 \mathbf{1}_{\{\lambda < 1.76\}} \right\}, \tag{4.5} \end{aligned}$$

where $(1 - e^0)/0 = 1$ and $\lambda = \hat{\lambda}(\mathcal{X})$.

Proof. For convenience, we write $|\xi| = n$ and $|\mathbb{Z}_\xi(t)| = Z_\xi(t)$. Let τ_1 and τ_2 be independent exponential random variables with mean 1 which are also independent of \mathbb{Z}_ξ ; then we can write

$$\begin{aligned} \mathbb{Z}_{\xi + \delta_\alpha}(t) &= \mathbb{Z}_\xi(t) + \delta_\alpha \mathbf{1}_{\{\tau_1 > t\}}, & \mathbb{Z}_{\xi + \delta_\beta}(t) &= \mathbb{Z}_\xi(t) + \delta_\beta \mathbf{1}_{\{\tau_2 > t\}}, \\ \text{and } \mathbb{Z}_{\xi + \delta_\alpha + \delta_\beta}(t) &= \mathbb{Z}_\xi(t) + \delta_\alpha \mathbf{1}_{\{\tau_1 > t\}} + \delta_\beta \mathbf{1}_{\{\tau_2 > t\}}. \end{aligned}$$

Hence, it follows, from (4.3) and the \bar{d}_1 -Lipschitz property of f , that

$$|\Delta h_f(\xi; \alpha)| = \left| \int_0^\infty e^{-t} \mathbb{E}[f(\mathbb{Z}_\xi(t) + \delta_\alpha) - f(\mathbb{Z}_\xi(t))] dt \right| \tag{4.6}$$

$$\leq \int_0^\infty e^{-t} \mathbb{E} \left[\frac{1}{Z_\xi(t) + 1} \right] dt \tag{4.7}$$

$$\leq \int_0^\infty e^{-t} dt = 1.$$

Also,

$$|\Delta^2 h_f(\xi; \alpha, \beta)| = \left| \int_0^\infty e^{-2t} \mathbb{E} \left[f(\mathbb{Z}_\xi(t) + \delta_\alpha + \delta_\beta) - f(\mathbb{Z}_\xi(t) + \delta_\alpha) - f(\mathbb{Z}_\xi(t) + \delta_\beta) + f(\mathbb{Z}_\xi(t)) \right] dt \right| \tag{4.8}$$

$$\leq \int_0^\infty e^{-2t} \mathbb{E} \left[\frac{1}{Z_\xi(t) + 2} + \frac{1}{Z_\xi(t) + 1} \right] dt \tag{4.9}$$

$$\leq 1.5 \int_0^\infty e^{-2t} dt = 0.75. \tag{4.10}$$

However, since \mathbb{Z}_ξ has constant immigration rate $\hat{\lambda}$ and unit per capita death rate, it is possible to write

$$\mathbb{Z}_\xi(t) = \mathbb{Z}_\emptyset(t) + \mathbb{D}_\xi(t),$$

where \mathbb{D}_ξ is a pure-death process with unit per capita death rate independent of \mathbb{Z}_\emptyset . Direct verification shows that $\mathbb{Z}_\emptyset(t)$ follows the Poisson distribution with mean $\lambda_t = \lambda(1 - e^{-t})$, while $|\mathbb{D}_\xi(t)|$ follows $\text{Bi}(|\xi|, e^{-t})$. Hence,

$$\mathbb{E} \left[\frac{1}{Z_\xi(t) + 1} \right] \leq \mathbb{E} \left[\frac{1}{Z_\emptyset(t) + 1} \right] = \frac{1 - \exp(-\lambda_t)}{\lambda_t}, \tag{4.11}$$

$$\begin{aligned} \mathbb{E} \left[\frac{1}{Z_\xi(t) + 1} \right] &= \int_0^1 \mathbb{E}[x^{Z_\xi(t)}] dx \\ &= \int_0^1 (1 - e^{-t}(1-x))^n \exp(-\lambda_t(1-x)) dx \\ &\leq \int_0^1 \exp(-(ne^{-t} + \lambda_t)(1-x)) dx \\ &\leq \int_0^1 e^{-(n \wedge \lambda)(1-x)} dx \\ &= \frac{1 - e^{-n \wedge \lambda}}{n \wedge \lambda}, \end{aligned} \tag{4.12}$$

and, similarly,

$$\begin{aligned}
 E\left[\frac{1}{Z_{\xi}(t) + 2}\right] &= \int_0^1 E[x^{Z_{\xi}(t)+1}] dx \\
 &= \int_0^1 x(1 - e^{-t}(1-x))^n \exp(-\lambda_t(1-x)) dx \\
 &\leq \int_0^1 x e^{-(n \wedge \lambda)(1-x)} dx \\
 &= \frac{1}{n \wedge \lambda} - \frac{1}{(n \wedge \lambda)^2}(1 - e^{-n \wedge \lambda}).
 \end{aligned}
 \tag{4.13}$$

The claim

$$|\Delta h_f(\xi; \alpha)| \leq \frac{0.95 + \ln^+ \lambda}{\lambda} \tag{4.14}$$

is obvious for $\lambda < 0.95$ as the right-hand side is already greater than 1, so it remains to show (4.14) for $\lambda \geq 0.95$. Combining (4.7) and (4.11), with $s = 1 - e^{-t}$, we obtain

$$|\Delta h_f(\xi; \alpha)| \leq \int_0^\infty e^{-t} \frac{1 - \exp(-\lambda t)}{\lambda t} dt = \int_0^1 \frac{1 - e^{-\lambda s}}{\lambda s} ds \leq \frac{1}{\lambda} \left(\frac{e^{-\lambda}}{\lambda} + \ln \lambda + \gamma \right),$$

where γ is the Euler constant and the last inequality is due to items 5.1.39 and 5.1.19 of [1]. For $0.95 \leq \lambda \leq 1$, $e^{-\lambda}/\lambda + \ln \lambda + \gamma \leq e^{-1} + \gamma < 0.95$, since $e^{-\lambda}/\lambda + \ln \lambda + \gamma$ is increasing for $\lambda \geq 0.95$, and, for $\lambda > 1$, $e^{-\lambda}/\lambda + \gamma < e^{-1} + \gamma < 0.95$, because the function $e^{-\lambda}/\lambda + \gamma$ is decreasing, completing the proof of (4.14). The last claim in (4.4) is easily obtained from (4.7) and (4.12).

We then apply (4.12) and (4.13) in (4.9) to obtain

$$|\Delta^2 h_f(\xi; \alpha, \beta)| \leq \frac{0.5}{n \wedge \lambda} \left(2 - e^{-n \wedge \lambda} - \frac{1}{n \wedge \lambda} (1 - e^{-n \wedge \lambda}) \right) \tag{4.15}$$

$$\leq \frac{1}{n \wedge \lambda}. \tag{4.16}$$

Now, we show that

$$|\Delta^2 h_f(\xi; \alpha, \beta)| \leq \frac{1.09}{n+1} + \frac{1}{\lambda}. \tag{4.17}$$

As a matter of fact, by (4.10) and (4.16), (4.17) clearly holds for $n = 0$ and $n \geq \lambda$; hence, it remains to show (4.17) for $1 \leq n < \lambda$. Using (4.15), it suffices to prove that

$$\frac{0.5(n+1)}{n} \left(2 - e^{-n} - \frac{1}{n} (1 - e^{-n}) \right) \leq 1.09. \tag{4.18}$$

However, for $n \geq 12$,

$$\frac{0.5(n+1)}{n} \left(2 - e^{-n} - \frac{1}{n} (1 - e^{-n}) \right) < \frac{n+1}{n} \leq \frac{13}{12} < 1.09,$$

while, for $1 \leq n \leq 11$, we can verify (4.18) for each value of n .

Finally, we prove that

$$|\Delta^2 h_f(\xi; \alpha, \beta)| \leq \frac{2 \ln \lambda}{\lambda} \mathbf{1}_{\{\lambda \geq 1.76\}} + 0.75 \mathbf{1}_{\{\lambda < 1.76\}}. \tag{4.19}$$

Claim (4.19) is evident for $\lambda < 1.76$, so we assume that $\lambda \geq 1.76$. On the other hand, if Y follows $Po(\nu)$ then

$$E\left[\frac{1}{Y+2}\right] = E\left[\frac{1}{Y+1} - \frac{1}{(Y+1)(Y+2)}\right] = \frac{\nu - 1 + e^{-\nu}}{\nu^2}.$$

Therefore,

$$E\left[\frac{1}{Z_\xi(t)+1} + \frac{1}{Z_\xi(t)+2}\right] \leq E\left[\frac{1}{Z_\emptyset(t)+1} + \frac{1}{Z_\emptyset(t)+2}\right] = \frac{2\lambda_t - 1 + (1 - \lambda_t) \exp(-\lambda_t)}{\lambda_t^2},$$

which, together with (4.9), ensures that

$$\begin{aligned} |\Delta^2 h_f(\xi; \alpha, \beta)| &\leq \int_0^\infty e^{-2t} \frac{2\lambda_t - 1 + (1 - \lambda_t) \exp(-\lambda_t)}{\lambda_t^2} dt \\ &= \int_0^1 (1-s) \frac{2\lambda s - 1 + (1 - \lambda s) e^{-\lambda s}}{\lambda^2 s^2} ds \\ &= -\frac{3}{\lambda} + \frac{2(1 - e^{-\lambda})}{\lambda^2} + \left(\frac{2}{\lambda} + \frac{1}{\lambda^2}\right) \int_0^\lambda \frac{1 - e^{-t}}{t} dt \\ &\leq -\frac{3}{\lambda} + \frac{2(1 - e^{-\lambda})}{\lambda^2} + \left(\frac{2}{\lambda} + \frac{1}{\lambda^2}\right) \left(\frac{e^{-\lambda}}{\lambda} + \ln \lambda + \gamma\right) \\ &= -\frac{3}{\lambda} + \frac{2}{\lambda^2} + \frac{e^{-\lambda}}{\lambda^3} + \left(\frac{2}{\lambda} + \frac{1}{\lambda^2}\right) (\ln \lambda + \gamma) \\ &= a(\lambda), \end{aligned}$$

where the first equality is obtained by the change of variable $s = 1 - e^{-t}$ and the last inequality is obtained from items 5.1.39 and 5.1.19 of [1]. Now, $b(\lambda) = a(\lambda)\lambda - 2 \ln \lambda$ is decreasing in λ for $\lambda > 1$ and $b(1.76) < 0$, which implies that $a(\lambda) \leq 2 \ln \lambda / \lambda$ for $\lambda \geq 1.76$, completing the proof of (4.19).

The following counterexample, adapted from [10], shows that the logarithmic factors in (4.4) and (4.5) cannot be removed.

Example. Let $\mathcal{X} = \{0, 1\}$ with metric $d_0(x, y) = |x - y|$, let $\hat{\lambda}$ satisfy $\hat{\lambda}\{1\} = 1$ and $\hat{\lambda}\{0\} = \lambda - 1 > 0$, and define a \hat{d}_1 -Lipschitz function on \mathfrak{N} as

$$f(\xi) = \begin{cases} \frac{1}{|\xi| + 1} & \text{if } \xi\{1\} = 0, \\ 0 & \text{if } \xi\{1\} > 0. \end{cases}$$

Using the facts that $Z_\emptyset(t)\{0\}$ follows $Po((\lambda - 1)(1 - e^{-t}))$ and that $Z_\emptyset(t)\{1\}$ follows $Po(1 - e^{-t})$, we have, from (4.6) and (4.8), as $\lambda \rightarrow \infty$,

$$\begin{aligned} |\Delta h_f(\emptyset; 1)| &= \int_0^\infty e^{-t} E\left[\frac{1}{Z_\emptyset(t)\{0\} + 1}\right] P[Z_\emptyset(t)\{1\} = 0] dt \\ &= \int_0^\infty e^{-t} \frac{1 - \exp(-(\lambda - 1)(1 - e^{-t}))}{(\lambda - 1)(1 - e^{-t})} \exp(-(1 - e^{-t})) dt \end{aligned}$$

$$\begin{aligned}
 &= \int_0^1 \frac{1 - e^{-(\lambda-1)s}}{(\lambda-1)s} e^{-s} ds \quad (\text{where } s = 1 - e^{-t}) \\
 &\geq \frac{e^{-1}}{\lambda-1} \int_0^{\lambda-1} \frac{1 - e^{-u}}{u} du \\
 &\asymp \frac{\ln \lambda}{\lambda}
 \end{aligned}$$

and

$$\begin{aligned}
 |\Delta^2 h(\emptyset; 1, 1)| &= \int_0^\infty e^{-2t} E \left[\frac{1}{\mathbb{Z}_\emptyset(t)\{0\} + 1} \right] P[\mathbb{Z}_\emptyset(t)\{1\} = 0] dt \\
 &= \int_0^\infty e^{-2t} \frac{1 - \exp(-(\lambda-1)(1 - e^{-t}))}{(\lambda-1)(1 - e^{-t})} \exp(-(1 - e^{-t})) dt \\
 &= \int_0^1 (1-s) \frac{1 - e^{-(\lambda-1)s}}{(\lambda-1)s} e^{-s} ds \quad (\text{where } s = 1 - e^{-t}) \\
 &\geq \frac{e^{-1}}{\lambda-1} \int_0^{\lambda-1} \left(1 - \frac{u}{\lambda-1}\right) \frac{1 - e^{-u}}{u} du \\
 &\asymp \frac{\ln \lambda}{\lambda}.
 \end{aligned}$$

In both inequalities notation of the form $a(\lambda) \asymp b(\lambda)$ means that there are constants $k_1, k_2 > 0$ and $\lambda_0 > 0$ so that $k_1 b(\lambda) \leq a(\lambda) \leq k_2 b(\lambda)$ for every $\lambda \geq \lambda_0$.

Remark. Comparing the Stein factors for d_2 (see [26, p. 146]) with (4.4) and (4.5), we can easily see that there is a substantial improvement for the Stein factor in (4.4), while the Stein factor in (4.5) is essentially of the same order as the counterpart for d_2 . To capitalize on the improvement in (4.4), it is advisable to use the ‘first difference approach’ (see [25]) in bounding Poisson process approximation errors (see Remark 4.1, below, for more discussion).

As noted before, \bar{d}_1 is the same as d_1 when the point patterns have the same number of points, while it is smoother than d_1 when the point patterns do not have the same number of points. On the other hand, for any two-point processes Ξ and H on \mathcal{X} , we have

$$E[d_1(\Xi, H)] = E[d_1(\Xi, H) \mid |\Xi| = |H|] P[|\Xi| = |H|] + P[|\Xi| \neq |H|]. \tag{4.20}$$

When we consider $P[|\Xi| \neq |H|]$, which corresponds to the total variation distance between the distributions of the total number of points of the two-point processes (see [4]), there is no such logarithmic component in Stein’s factor, which means that the logarithmic component in d_1 was brought in only by the discrepancies in the locations of points when the point patterns have the same number of points. However, this problem is shared by \bar{d}_1 , that is, the Stein factors for \bar{d}_1 will inevitably have the logarithmic component as well.

It is also worthwhile to note that, since $E[\bar{d}_1(\Xi, H)]$ replaces the term $P[|\Xi| \neq |H|]$ in (4.20) with a smaller $E[d_R(|\Xi|, |H|)]$, we would expect a bigger improvement on bounding $\bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H))$ when $P[|\Xi| \neq |H|]$ is ‘dominant’ at the right-hand side of (4.20) under the best coupling. Such an improvement is obtained in the next two subsections.

4.2. Poisson process approximation of a Bernoulli process

Let $\mathcal{X} = [0, 1]$ with $d_0(x, y) = |x - y|$, and let X_1, \dots, X_n be independent and identically distributed (i.i.d.) Bernoulli random variables with $P[X_1 = 1] = p$. Then $\Xi = \sum_{i=1}^n X_i \delta_{i/n}$

defines a Bernoulli process on \mathcal{X} . If we let T_0, T_1, \dots, T_n be i.i.d. uniform random variables on \mathcal{X} which are also independent of $\{X_1, \dots, X_n\}$, then

$$Y = \sum_{i=1}^n X_i \delta_{T_i}$$

defines a binomial process on \mathcal{X} (see [19, p. 29]). By [27],

$$d_2(\mathcal{L}(\Xi), \mathcal{L}(Y)) \leq \left(\frac{1}{2n} + \frac{p}{2}\right) \wedge \frac{1}{\sqrt{3np}}. \tag{4.21}$$

To estimate $\bar{d}_2(\mathcal{L}(Y), \text{Po}(\hat{\lambda}))$ with $\hat{\lambda}(dx) = np dx$, we employ Stein’s method for the Poisson process approximation. As a matter of fact, it follows from (4.1) that

$$\begin{aligned} \mathbb{E}[\mathcal{A}h(Y)] &= \mathbb{E}\left[\int_{\mathcal{X}} (h(Y + \delta_\alpha) - h(Y))\hat{\lambda}(d\alpha) + \int_{\mathcal{X}} (h(Y - \delta_\alpha) - h(Y))Y(d\alpha)\right] \\ &= np \mathbb{E}[h(Y + \delta_{T_0}) - h(Y)] + \sum_{i=1}^n \mathbb{E}[h(Y^i) - h(Y^i + \delta_{T_i})]p \\ &= np \mathbb{E}[(h(Y + \delta_{T_0}) - h(Y)) - (h(Y^1 + \delta_{T_0}) - h(Y^1))], \end{aligned}$$

where $Y^i = Y - X_i \delta_{T_i}$. Define

$$g(i) = \mathbb{E}[h(Y + \delta_{T_0}) - h(Y) \mid |Y| = i] = \mathbb{E}\left[h\left(\sum_{j=0}^i \delta_{T_j}\right) - h\left(\sum_{j=1}^i \delta_{T_j}\right)\right].$$

Then

$$\begin{aligned} |\mathbb{E}[\mathcal{A}h(Y)]| &= np |\mathbb{E}[g(|Y|) - g(|Y^1|)]| \\ &= np^2 |\mathbb{E}[g(|Y^1| + 1) - g(|Y^1|)]| \\ &\leq 2np^2 \|g\| d_{\text{TV}}(\mathcal{L}(|Y^1|), \mathcal{L}(|Y^1| + 1)), \end{aligned} \tag{4.22}$$

where $\|\cdot\|$ denotes the supremum norm and, for any two nonnegative integer-valued random variables U_1 and U_2 ,

$$d_{\text{TV}}(\mathcal{L}(U_1), \mathcal{L}(U_2)) = \frac{1}{2} \sup_{\tilde{g}: \mathbb{Z}_+ \rightarrow [-1,1]} |\mathbb{E}[\tilde{g}(U_1)] - \mathbb{E}[\tilde{g}(U_2)]|.$$

On the other hand, by Lemma 1 of [5],

$$d_{\text{TV}}(\mathcal{L}(|Y^1|), \mathcal{L}(|Y^1| + 1)) \leq \max_{0 \leq i \leq n-1} \mathbb{P}[|Y^1| = i] \leq 1 \wedge \frac{1}{2\sqrt{(n-1)p(1-p)}},$$

and using (4.4), we have, for $f \in \tilde{\mathcal{F}}_2$,

$$|\mathbb{E}[\mathcal{A}h_f(Y)]| \leq \frac{(0.95 + \ln^+(np))p}{1/2 \sqrt{(n-1)p(1-p)}},$$

which implies from (4.2) that

$$\bar{d}_2(\mathcal{L}(Y), \text{Po}(\hat{\lambda})) = \sup_{f \in \tilde{\mathcal{F}}_2} |\mathbb{E}[\mathcal{A}h_f(Y)]| \leq \frac{(0.95 + \ln^+(np))p}{1/2 \sqrt{(n-1)p(1-p)}}. \tag{4.23}$$

Now, collecting (4.21) and (4.23) gives the following theorem.

Theorem 4.2. *With the above setup, we have*

$$\bar{d}_2(\mathcal{L}(\Xi), \text{Po}(\hat{\lambda})) \leq \left(\frac{1}{2n} + \frac{p}{2}\right) \wedge \frac{1}{\sqrt{3np}} + \frac{(0.95 + \ln^+(np))p}{1/2 \vee \sqrt{(n-1)p(1-p)}}.$$

Remark 4.1. The main idea for getting the sharp bound in Theorem 4.2 is to use the ‘first difference (of h) approach’ to obtain (4.22). If we use (4.5) in Theorem 4.1 then the bound obtained will be much the same as that for the d_2 -distance except with a smaller constant.

Remark. An immediate consequence of Theorem 4.2 is that, if n is large, it is almost impossible to distinguish between the distributions of the two processes. It is quite a contrast to the conclusion under d_2 , where it is essential to have a very small p as well as a large n to ensure a valid Poisson process approximation (see [24]). In practice, statisticians would use a Poisson process rather than a Bernoulli process when n is large, confirming our conclusion under \bar{d}_2 .

Remark. It is a tantalizing problem to remove the $\ln^+ \lambda$ term in the upper bound. We conjecture that, at the cost of more complexity, the actual bound should be of order $(1/n + p)/(1 \vee \sqrt{np})$.

4.3. Point processes of i.i.d. points

In this subsection we pursue a more direct approach for computing bounds of \bar{d}_2 -distances. While Stein’s method, which was used in the previous two subsections, excels by its versatility, the direct approach performs well in the special case where both point processes have i.i.d. points and yields an upper as well as a lower bound.

Let $\Xi = \sum_{i=1}^M \delta_{X_i}$ and $H = \sum_{i=1}^N \delta_{Y_i}$, where M and N are integer-valued random variables, $(X_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. \mathcal{X} -valued random elements that is independent of M , and $(Y_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. \mathcal{X} -valued random elements that is independent of N . Denote by d_W the Wasserstein metric between random elements of \mathcal{X} with respect to d_0 .

Proposition 4.2. *We have*

$$\begin{aligned} & \max(d_{RW}(\mathcal{L}(M), \mathcal{L}(N)), c_1 d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1))) \\ & \leq \bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H)) \\ & \leq d_{RW}(\mathcal{L}(M), \mathcal{L}(N)) + c_2 d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1)), \end{aligned}$$

where

$$c_1 = c_1(\mathcal{L}(M), \mathcal{L}(N)) = \max(\mathbb{P}[M > 0], \mathbb{P}[N > 0])$$

and

$$c_2 = c_2(\mathcal{L}(M), \mathcal{L}(N)) = \mathbb{E} \left[\frac{\tilde{M} \wedge \tilde{N}}{\tilde{M} \vee \tilde{N}} \mathbf{1}_{\{\tilde{M} \vee \tilde{N} > 0\}} \right] \leq \min(\mathbb{P}[M > 0], \mathbb{P}[N > 0])$$

for random variables \tilde{M} and \tilde{N} that are coupled in such a way that

$$\mathbb{E}[d_R(\tilde{M}, \tilde{N})] = d_{RW}(\mathcal{L}(M), \mathcal{L}(N)).$$

Remark 4.2. An interesting special case is given if Ξ and H are Poisson processes. For finite measures $\hat{\mu}$ and $\hat{\nu}$ on \mathcal{X} , we obtain, from Proposition 4.2,

$$\bar{d}_2(\text{Po}(\hat{\mu}), \text{Po}(\hat{\nu})) \leq \frac{|\mu - \nu|}{\mu \vee \nu} + (1 - e^{-(\mu \wedge \nu)}) d_W \left(\frac{\hat{\mu}}{\mu}, \frac{\hat{\nu}}{\nu} \right),$$

which is an improvement by a factor of order $1/\sqrt{\mu \vee \nu}$ for $\mu, \nu \rightarrow \infty$ in the first summand when compared to a corresponding d_2 -bound (see, for example, [9, Equation (2.8)]). Estimation of the d_{RW} -term was achieved by considering a Poisson process Z on \mathbb{R}_+ with intensity 1 and defining a coupling pair by $\tilde{M} = Z((0, \mu])$ and $\tilde{N} = Z((0, \nu])$.

Proof of Proposition 4.2. Upper bound. Let $\tilde{M} \stackrel{D}{=} M$ and $\tilde{N} \stackrel{D}{=} N$ (where ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution) be coupled according to a d_{RW} -coupling, so that

$$E \left[\frac{|\tilde{M} - \tilde{N}|}{\tilde{M} \vee \tilde{N}} \mathbf{1}_{\{\tilde{M} \vee \tilde{N} > 0\}} \right] = d_{RW}(\mathcal{L}(M), \mathcal{L}(N)),$$

and let $\tilde{X}_i \stackrel{D}{=} X_i$ and $\tilde{Y}_i \stackrel{D}{=} Y_i$ with $E[d_0(\tilde{X}_i, \tilde{Y}_i)] = d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1))$ for every $i \in \mathbb{N}$ in such a way that the pairs $(\tilde{M}, \tilde{N}), (\tilde{X}_1, \tilde{Y}_1), (\tilde{X}_2, \tilde{Y}_2), \dots$ are independent. We then obtain

$$\begin{aligned} \bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H)) &\leq E \left[\bar{d}_1 \left(\sum_{i=1}^{\tilde{M}} \delta_{\tilde{X}_i}, \sum_{j=1}^{\tilde{N}} \delta_{\tilde{Y}_j} \right) \right] \\ &\leq E \left[\frac{|\tilde{M} - \tilde{N}|}{\tilde{M} \vee \tilde{N}} \mathbf{1}_{\{\tilde{M} \vee \tilde{N} > 0\}} \right] + E \left[\frac{\mathbf{1}_{\{\tilde{M} \vee \tilde{N} > 0\}}}{\tilde{M} \vee \tilde{N}} \sum_{i=1}^{\tilde{M} \wedge \tilde{N}} d_0(\tilde{X}_i, \tilde{Y}_i) \right], \end{aligned}$$

which, by the independence between (\tilde{M}, \tilde{N}) and $\{(\tilde{X}_i, \tilde{Y}_i), i \geq 1\}$, and the assumptions on the distributions of those pairs, yields the upper bound claimed.

The bound for the factor c_2 follows from

$$E \left[\frac{\tilde{M} \wedge \tilde{N}}{\tilde{M} \vee \tilde{N}} \mathbf{1}_{\{\tilde{M} \vee \tilde{N} > 0\}} \right] \leq P[\tilde{M} > 0, \tilde{N} > 0]$$

and $P[\tilde{M} = \tilde{N} = 0] = \min(P[M = 0], P[N = 0])$, the proof of which is straightforward.

Lower bound. Let $\mathcal{F}_W^* = \{g: \mathcal{X} \rightarrow [0, 1]; |g(x) - g(y)| \leq d_0(x, y) \text{ for all } x, y \in \mathcal{X}\}$, and let $\tilde{g} \in \mathcal{F}_W^*$ be a mapping with $|E[\tilde{g}(X_1)] - E[\tilde{g}(Y_1)]| = d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1))$. Such a mapping exists by $d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1)) = \sup_{g \in \mathcal{F}_W^*} |E[g(X_1)] - E[g(Y_1)]|$, where the supremum is attained because \mathcal{F}_W^* is a compact subset of $C(\mathcal{X}, \mathbb{R})$ by the Arzelà–Ascoli theorem, and the mapping $[g \mapsto |E[g(X_1)] - E[g(Y_1)]|]$ is continuous (both statements are with respect to the topology of uniform convergence).

Define $\tilde{f}: \mathfrak{N} \rightarrow [0, 1]$ by

$$\tilde{f}(\xi) = \frac{1}{|\xi|} \int_{\mathcal{X}} \tilde{g}(x) \xi(dx)$$

for $\xi \in \mathfrak{N} \setminus \{0\}$ and $\tilde{f}(0) = E[\tilde{g}(X_1)]$. Next we check that $\tilde{f} \in \tilde{\mathcal{F}}_2$. It is immediately clear that $|\tilde{f}(\xi) - \tilde{f}(0)| \leq 1 = \bar{d}_1(\xi, 0)$ if $\xi \in \mathfrak{N} \setminus \{0\}$. Then let $\xi = \sum_{i=1}^m \delta_{x_i}$ and $\eta = \sum_{j=1}^n \delta_{y_j}$ both be in $\mathfrak{N} \setminus \{0\}$, where we assume, without loss of generality, that $m \leq n$ and $\tilde{f}(\xi) \geq \tilde{f}(\eta)$ (otherwise interchange ξ and η and/or replace \tilde{g} by $1 - \tilde{g} \in \mathcal{F}_W^*$) and that the points are numbered according to a \bar{d}_1 -pairing such that

$$\frac{1}{n} \left(\sum_{i=1}^m d_0(x_i, y_i) + (n - m) \right) = \bar{d}_1(\xi, \eta).$$

Let $k \in \arg \max_{1 \leq i \leq m} \tilde{g}(x_i)$ and $x_i = x_k$ for $m + 1 \leq i \leq n$, which implies that

$$\begin{aligned} |\tilde{f}(\xi) - \tilde{f}(\eta)| &= \frac{1}{m} \sum_{i=1}^m \tilde{g}(x_i) - \frac{1}{n} \sum_{i=1}^n \tilde{g}(y_i) \\ &\leq \frac{1}{n} \sum_{i=1}^n \tilde{g}(x_i) - \frac{1}{n} \sum_{i=1}^n \tilde{g}(y_i) \\ &\leq \frac{1}{n} \sum_{i=1}^n d_0(x_i, y_i) \\ &\leq \bar{d}_1(\xi, \eta), \end{aligned}$$

and, therefore, $\tilde{f} \in \tilde{\mathcal{F}}_2$.

Choose pairs $(\tilde{M}, \tilde{N}), (\tilde{X}_1, \tilde{Y}_1), (\tilde{X}_2, \tilde{Y}_2), \dots$ in the same way as for the proof of the upper bound (although the coupling of \tilde{X}_i and \tilde{Y}_i in each of the pairs is not important now). We obtain

$$\begin{aligned} \bar{d}_2(\mathcal{L}(\Xi), \mathcal{L}(H)) &\geq \left| \mathbb{E} \left[\tilde{f} \left(\sum_{i=1}^{\tilde{M}} \delta_{\tilde{X}_i} \right) \right] - \mathbb{E} \left[\tilde{f} \left(\sum_{j=1}^{\tilde{N}} \delta_{\tilde{Y}_j} \right) \right] \right| \\ &= \left| \mathbb{E} \left[\left(\frac{1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \tilde{g}(\tilde{X}_i) - \frac{1}{\tilde{N}} \sum_{j=1}^{\tilde{N}} \tilde{g}(\tilde{Y}_j) \right) \mathbf{1}_{\{\tilde{M} > 0, \tilde{N} > 0\}} \right] \right. \\ &\quad \left. + \mathbb{E} \left[\left(\frac{1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} \tilde{g}(\tilde{X}_i) - \mathbb{E} \tilde{g}(X_1) \right) \mathbf{1}_{\{\tilde{M} > 0, \tilde{N} = 0\}} \right] \right. \\ &\quad \left. + \mathbb{E} \left[\left(\mathbb{E}[\tilde{g}(X_1)] - \frac{1}{\tilde{N}} \sum_{j=1}^{\tilde{N}} \tilde{g}(\tilde{Y}_j) \right) \mathbf{1}_{\{\tilde{M} = 0, \tilde{N} > 0\}} \right] \right| \\ &= |(\mathbb{E}[\tilde{g}(X_1)] - \mathbb{E}[\tilde{g}(Y_1)]) \mathbb{P}[\tilde{M} > 0, \tilde{N} > 0]| \\ &\quad + (\mathbb{E}[\tilde{g}(X_1)] - \mathbb{E}[\tilde{g}(Y_1)]) \mathbb{P}[\tilde{M} = 0, \tilde{N} > 0]| \\ &= d_W(\mathcal{L}(X_1), \mathcal{L}(Y_1)) \mathbb{P}[N > 0]. \end{aligned}$$

Since the above argument is symmetric in Ξ and H , we obtain the lower bound when combining it with Proposition 2.3(ii).

5. A statistical application

In order to show the potential of \bar{d}_1 and \bar{d}_2 in statistical applications, we propose a test procedure based on these two metrics. For the following construction, only some of the elementary statements in Section 2 are required, but none of the results from Sections 3 and 4 are used. Suppose that our data consists of a few i.i.d. realizations of a point process Ξ , and we would like to test if $\Xi \sim P$ for a certain probability measure P on \mathfrak{N} . Such multiple point-pattern data may arise, among other examples, from recording degenerate cells in tissue samples or plants in a large population that is sampled only via a few comparatively small windows.

In what follows, we restrict our attention to a test for spatial homogeneity under the assumption that Ξ is a Poisson process on $W = [0, 1]^2$ with unknown expectation measure $\hat{\lambda}$. This limits the alternative hypothesis sufficiently to keep our simulation study within the scope of this paper. Suppose that ξ_1, \dots, ξ_N are realizations of i.i.d. copies Ξ_1, \dots, Ξ_N of Ξ and that the total mass $\lambda = \hat{\lambda}([0, 1]^2)$ of the expectation measure $\hat{\lambda}$ is known (otherwise we just take the canonical estimate $(1/N) \sum_{i=1}^N |\xi_i|$). Our null hypothesis is then $\Xi \sim \text{Po}(\lambda \text{Leb}^2)$, where Leb^2 denotes the Lebesgue measure on $[0, 1]^2$. Write

$$P_N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i} \in \mathfrak{P}(\mathfrak{N})$$

for the empirical distribution of our data. We perform a Monte Carlo test, where the test statistic would ideally be

$$T(\xi_1, \dots, \xi_N) = \bar{d}_2(P_N, \text{Po}(\lambda \text{Leb}^2)),$$

but since this is computationally intractable, we replace it by the randomized test statistic

$$T(\xi_1, \dots, \xi_N; \eta_1, \dots, \eta_N) = \bar{d}_2(P_N, Q_N),$$

where $Q_N = (1/N) \sum_{i=1}^N \delta_{\eta_i}$ for realizations η_i of $\text{Po}(\lambda \text{Leb}^2)$ -processes H_i that are independent amongst each other and of the Ξ_i . For the metric d_0 , we simply choose the Euclidean metric cut off at 1. The null hypothesis is rejected at significance level $\alpha = 0.05$ if $T(\xi_1, \dots, \xi_N; \eta_1, \dots, \eta_N)$ ranks among the five highest values when pooled with 99 simulations of $T(\tilde{\eta}_1, \dots, \tilde{\eta}_N; \eta_1, \dots, \eta_N)$, where $\tilde{\eta}_1, \dots, \tilde{\eta}_N, \eta_1, \dots, \eta_N$ are all independent $\text{Po}(\lambda \text{Leb}^2)$ -realizations.

We choose $N = 12$ and $\lambda = 30$ for the simulation study, which is both realistic for actual data and keeps computation times at a tolerable level. One single test of two series of 12 point patterns takes less than three seconds (given the simulated null hypothesis distribution) on an ordinary laptop computer using the library spatstat (see [2]) that supplies tools for the analysis of spatial point patterns within the statistical computing environment R (see [18]). Increasing either N to 50 or λ to 110 while keeping the other parameter fixed, still keeps the computation time well under one minute. Note that the optimal point assignments needed for computing \bar{d}_1 , and also \bar{d}_2 between empirical measures, can be found efficiently (in $O((m \vee n)^3)$ steps, where m and n are the cardinalities of the point patterns), by using the so-called Hungarian method from linear programming (see [17, Section 11.2]).

Table 1 summarizes the results of our simulations. The second of the five columns gives the Monte Carlo powers of our test against $\text{Po}(\lambda f_\kappa(x, y) \text{Leb}^2(d(x, y)))$ -alternatives, where

$$f_\kappa(x, y) = \frac{\kappa e^{-\kappa x}}{1 - e^{-\kappa}}$$

for $x, y \in [0, 1]$ and $\kappa = 1, \dots, 4$. See Figure 2 to obtain an impression of the corresponding distributions. By Monte Carlo power we mean the fraction of the number of rejected tests in 100 independent simulations of the alternative.

While it is natural to base d_0 on the Euclidean metric for our test, the choice of the cut off at 1 is very arbitrary and only yields reasonable results because it happens to be of the same order of magnitude as typical interpoint distances in our observation window W . Often we would like to be able to choose other cutoff values according to the size and geometry of the observation window and possibly the expected number of points. In order to do this, we generalize our

TABLE 1: Powers of the tests for two different cutoff values c against increasingly conspicuous alternatives. The last two columns give the corresponding results when the test is based on the metric d_1 instead of \bar{d}_1 and are listed for comparison only.

κ	\bar{d}_1		d_1	
	$c = 1$	$c = 0.3$	$c = 1$	$c = 0.3$
1	0.10	0.23	0.08	0.02
2	0.41	0.97	0.12	0.06
3	0.93	1.00	0.06	0.04
4	1.00	1.00	0.10	0.10

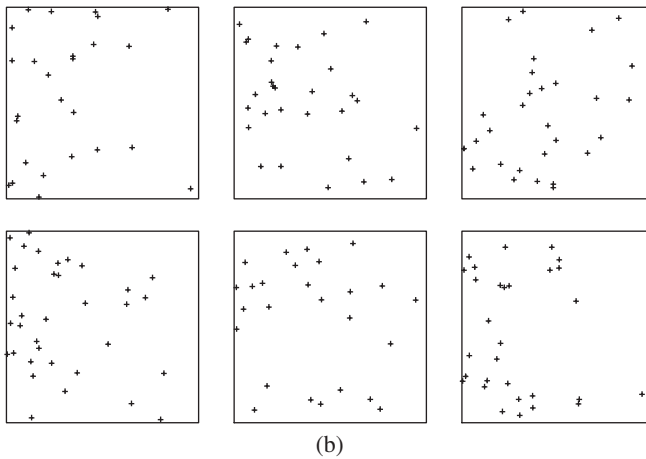
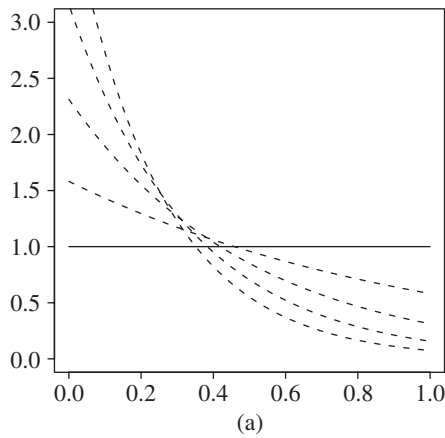


FIGURE 2: (a) Normalized intensity functions f_κ plotted against their first coordinate; (b) six independent realizations from $\text{Po}(f_2 \text{Leb}^2)$.

metric \bar{d}_1 somewhat by introducing an arbitrary distance bound $c > 0$. Furthermore, we include an order parameter $p \geq 1$, following the idea in [23] for d_1 , which may be useful in some applications. Assuming that $d_0 \leq c$, we thus set

$$\bar{d}_1^{(p,c)}(\xi, \eta) = \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^m d_0(x_i, y_{\pi(i)})^p + c^p(n - m) \right) \right)^{1/p}$$

for $\xi = \sum_{i=1}^m \delta_{x_i}$ and $\eta = \sum_{j=1}^n \delta_{y_j} \in \mathfrak{N}$ with $n \geq \max(m, 1)$, requiring again that $\bar{d}_1^{(p,c)}$ is symmetric and that $\bar{d}_1^{(p,c)}(0, 0) = 0$.

Remark 5.1. It is straightforward to adapt the statements and proofs of Propositions 2.1–2.3 for $\bar{d}_1^{(p,c)}$, using the more general results of [23] about $d_1^{(p)}$ instead of the corresponding results about d_1 where necessary. In this way, it can be seen in particular that $\bar{d}_1^{(p,c)}$ is a metric that is bounded by c , metrizes the vague topology, and thus induces a Wasserstein metric $\bar{d}_2^{(p,c)}$ on $\mathfrak{P}(\mathfrak{N})$ that metrizes convergence in distribution of point processes.

Regarding Sections 3 and 4, it is quite easy to accommodate the results for a general $c > 0$. In many cases this can be done by employing the fact that

$$\bar{d}_1^{(p,c)}(\xi, \eta) = c \bar{d}_1^{(p,1)}(\xi, \eta)$$

if $\bar{d}_1^{(p,c)}$ is based on $d_0 \leq c$ and $\bar{d}_1^{(p,1)}$ is based on $\tilde{d}_0 = (1/c)d_0$, although sometimes a somewhat different strategy is necessary. In particular, we find that the Lipschitz continuous functions for \bar{d}_1 in Section 3 are Lipschitz continuous with the same constants for $\bar{d}_1^{(1,c)}$ if we consider an arbitrary $d_0 \leq c$ for Proposition 3.1 and $d_0(x, y) = |x - y| \wedge c$ for Proposition 3.2. Furthermore, the Stein factors in Proposition 4.1 hold true for the corresponding $\bar{d}_2^{(1,c)}$ -distances after multiplication of these factors by c , which means that in applications the estimates of the errors for the Poisson process approximation in terms of $\bar{d}_2^{(1,c)}$ -distances can be derived in the same way with only minor modifications.

It can be seen that introducing a general p in Section 3 yields the corresponding p th-order averages as Lipschitz continuous functions, whereas in Section 4 considerably more work is necessary to adapt the distance estimates given (both situations are similar to the introduction of p in the metrics d_1 and d_2 that was considered in [23]).

For our hypothesis test, we stick to the case in which $p = 1$, but give in the third column of Table 1 the corresponding Monte Carlo powers if d_0 is the Euclidean metric cut off at $c = 0.3$ instead of 1, so that our test now puts less emphasis on cardinalities and more emphasis on positional differences in the compared point patterns than it did before. There is no strong reason for choosing exactly $c = 0.3$; the value reflects the somewhat vague idea that in an optimal pairing of about 30 points each, the pairing distances are ‘usually’ still below 0.3. As we can see from Table 1, the power improvement is very noticeable, and this second test especially detects the inhomogeneity quite well, even though it is not clearly visible in the simulated data.

For comparison, we have also added the results of the corresponding tests if \bar{d}_1 is replaced by d_1 . Since there is typically a wide range of values for the cardinalities of realizations of a Poisson process with 30 expected points, and since differing cardinalities are not appropriately addressed by d_1 , these tests perform very poorly (for $c = 0.3$, powers seem to increase substantially from $\kappa = 9$ on).

In summary, the above procedure is rather successful for testing spatial homogeneity from multiple point patterns. We have also obtained promising first results when testing for spatial dependence, but a more extensive further study will be necessary in order to establish the

possibilities and limitations of this test procedure and of tests or other statistical analyses based on the \bar{d}_1 -metric in general.

Appendix A. Proofs left out in the main text

Proof of Proposition 2.1. From the definition, it is clear that $0 \leq \bar{d}_1(\xi, \eta) \leq 1$, $\bar{d}_1(\xi, \eta) = 0$ if and only if $\xi = \eta$, and that $\bar{d}_1(\xi, \eta) = \bar{d}_1(\eta, \xi)$. To show the triangle inequality, let

$$\xi = \sum_{i=1}^l \delta_{x_i}, \quad \eta = \sum_{j=1}^m \delta_{y_j}, \quad \text{and} \quad \zeta = \sum_{k=1}^n \delta_{z_k} \in \mathfrak{N},$$

and add two points u_1 and u_2 to \mathcal{X} , extending d_0 by $d_0(u_1, u_2) = d_0(u_1, u) = d_0(u_2, u) = 1$ for every $u \in \mathcal{X}$.

Note that, for $l = m = n$, it is immediately clear that

$$\bar{d}_1(\xi, \eta) \leq \bar{d}_1(\xi, \zeta) + \bar{d}_1(\zeta, \eta), \tag{A.1}$$

by the triangle inequality for d_1 . We now prove (A.1) in general, assuming that at most one of the point patterns is empty (otherwise it is obvious). Since this inequality is symmetric in ξ and η , we assume, without loss of generality, that $l \leq m$ in what follows. We show two separate cases.

Case 1: $m \leq n$. Let $x_i = u_1$ for $l + 1 \leq i \leq n$ and let $y_j = u_2$ for $m + 1 \leq j \leq n$, and write $\bar{\xi} = \sum_{i=1}^n \delta_{x_i}$ and $\bar{\eta} = \sum_{j=1}^n \delta_{y_j}$. We then have

$$\bar{d}_1(\xi, \eta) \leq \bar{d}_1(\xi, \bar{\eta}) = \bar{d}_1(\bar{\xi}, \bar{\eta}) \leq \bar{d}_1(\bar{\xi}, \zeta) + \bar{d}_1(\zeta, \bar{\eta}) = \bar{d}_1(\xi, \zeta) + \bar{d}_1(\zeta, \eta),$$

using, for the first inequality, the fact that $a \leq m$ implies that $a/m \leq (a+n-m)/n$, and, for the second inequality, the fact that (A.1) holds if all three point patterns have the same cardinality.

Case 2: $m > n$. Let $x_i = z_k = u_1$ for $l + 1 \leq i \leq m$ and $n + 1 \leq k \leq m$, and write

$$\bar{\xi} = \sum_{i=1}^{l \vee n} \delta_{x_i}, \quad \bar{\xi} = \sum_{i=1}^m \delta_{x_i}, \quad \check{\zeta} = \sum_{k=1}^{l \vee n} \delta_{z_k}, \quad \text{and} \quad \bar{\zeta} = \sum_{k=1}^m \delta_{z_k}.$$

Note that $x_i = z_i$ for $(l \vee n) + 1 \leq i \leq m$, by which it can be seen that $\bar{d}_1(\bar{\xi}, \bar{\zeta}) \leq \bar{d}_1(\check{\xi}, \check{\zeta})$. We then have

$$\bar{d}_1(\xi, \eta) = \bar{d}_1(\bar{\xi}, \eta) \leq \bar{d}_1(\bar{\xi}, \bar{\zeta}) + \bar{d}_1(\bar{\zeta}, \eta) \leq \bar{d}_1(\check{\xi}, \check{\zeta}) + \bar{d}_1(\bar{\zeta}, \eta) = \bar{d}_1(\xi, \zeta) + \bar{d}_1(\zeta, \eta),$$

where we used, for the first inequality, the fact that (A.1) holds if all three-point patterns have the same cardinality.

Proof of Proposition 2.2. Statement (i) is straightforward from the definitions of d_R , \bar{d}_1 , and d_1 .

(ii) Proposition 4.2 of [26] states that $\xi_n \rightarrow \xi$ vaguely if and only if $d_1(\xi_n, \xi) \rightarrow 0$ as $n \rightarrow \infty$; so all we need to show is that the latter is equivalent to $\bar{d}_1(\xi_n, \xi) \rightarrow 0$.

If $\bar{d}_1(\xi_n, \xi) \rightarrow 0$, by (i) we have $d_R(|\xi_n|, |\xi|) \rightarrow 0$, from which it is easily seen that $|\xi_n| \rightarrow |\xi|$, i.e. there is an $n_0 \in \mathbb{N}$ such that $|\xi_n| = |\xi|$ and, hence, $\bar{d}_1(\xi_n, \xi) = d_1(\xi_n, \xi)$ for every $n \geq n_0$. Thus, $d_1(\xi_n, \xi) \rightarrow 0$. The converse direction follows immediately from $\bar{d}_1 \leq d_1$.

(iii) The local compactness and separability properties depend only on the generated topology. See, for example, Proposition 4.3 of [26] for the proof. Note that, by the compactness of \mathcal{X} , the sets $\mathfrak{N}_l = \{\xi \in \mathfrak{N}; |\xi| = l\}$ are compact for all $l \in \mathbb{Z}_+$.

It remains to show the completeness. Let $(\xi_n)_{n \in \mathbb{N}}$ be a \bar{d}_1 -Cauchy sequence in \mathfrak{N} . It is straightforward to see that this implies the existence of an $n_0 \in \mathbb{N}$ such that $|\xi_n| = |\xi_m|$ for every $n, m \geq n_0$, which means that there exists an $l \in \mathbb{Z}_+$ such that the tail of $(\xi_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in \mathfrak{N}_l . By the compactness of \mathfrak{N}_l , this tail converges.

Proof of Proposition 2.3. Statement (i) is a direct consequence of the Kantorovich–Rubinstein theorem, where the minimum is attained, because $(\mathfrak{N}, \bar{d}_1)$ is complete. See [14, Section 11.8] for details. Statement (ii) follows by taking expectations and minima in Proposition 2.2(i). The last statement follows from [14, Theorem 11.3.3] using Proposition 2.2(ii) and noting that \bar{d}_2 is an instance of Dudley’s β -metric.

Acknowledgements

This work was supported by the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems (AX), the ARC Discovery Grant number DP0345215 (DS), and the Schweizerischer Nationalfonds Fellowship number PBZH2-111668 (DS).

DS would like to thank Adrian Baddeley for stimulating discussions about spatial statistics.

References

- [1] ABRAMOWITZ, M. AND STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. Dover, New York.
- [2] BADDELEY, A. AND TURNER, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Software* **12**, 1–42.
- [3] BARBOUR, A. D. AND BROWN, T. C. (1992a). Stein’s method and point process approximation. *Stoch. Process. Appl.* **43**, 9–31.
- [4] BARBOUR, A. D. AND BROWN, T. C. (1992b). The Stein–Chen method, point processes and compensators. *Ann. Prob.* **20**, 1504–1527.
- [5] BARBOUR, A. D. AND JENSEN, J. L. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16**, 75–87.
- [6] BARBOUR, A. D. AND MÅNSSON, M. (2002). Compound Poisson process approximation. *Ann. Prob.* **30**, 1492–1537.
- [7] BARBOUR, A. D., BROWN, T. C. AND XIA, A. (1998). Point processes in time and Stein’s method. *Stoch. Stoch. Reports* **65**, 127–151.
- [8] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation* (Oxford Studies Prob. **2**). Oxford University Press.
- [9] BROWN, T. C. AND XIA, A. (1995a). On metrics in point process approximation. *Stoch. Stoch. Reports* **52**, 247–263.
- [10] BROWN, T. C. AND XIA, A. (1995b). On Stein–Chen factors for Poisson approximation. *Statist. Prob. Lett.* **23**, 327–332.
- [11] BROWN, T. C. AND XIA, A. (2001). Stein’s method and birth–death processes. *Ann. Prob.* **29**, 1373–1403.
- [12] CHEN, L. H. Y. AND XIA, A. (2004). Stein’s method, Palm theory and Poisson process approximation. *Ann. Prob.* **32**, 2545–2569.
- [13] CONWAY, J. H. AND SLOANE, N. J. A. (1999). *Sphere Packings, Lattices and Groups* (Fundamental Principles Math. Sci. **290**), 3rd edn. Springer, New York.
- [14] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- [15] KALLENBERG, O. (1986). *Random Measures*, 4th edn. Academic Press, London.
- [16] LEE, A. J. (1990). *U-Statistics* (Statistics: Textbooks Monogr. **110**). Marcel Dekker, New York.
- [17] PAPADIMITRIOU, C. H. AND STEIGLITZ, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola, NY.
- [18] R DEVELOPMENT CORE TEAM (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at <http://www.r-project.org>.
- [19] REISS, R.-D. (1993). *A Course on Point Processes*. Springer, New York.

- [20] SCHUHMACHER, D. (2005a). Estimation of distances between point process distributions. Doctoral Thesis, University of Zurich. Available at <http://www.dissertationen.unizh.ch/2006/schuhmacher/diss.pdf>.
- [21] SCHUHMACHER, D. (2005b). Upper bounds for spatial point process approximations. *Ann. Appl. Prob.* **15**, 615–651.
- [22] SCHUHMACHER, D. (2007a). Distance estimates for dependent thinnings of point processes with densities. Preprint. Available at <http://arxiv.org/abs/math/0701728>.
- [23] SCHUHMACHER, D. (2007b). Stein's method and Poisson process approximation for a class of Wasserstein metrics. Preprint. Available at <http://arxiv.org/abs/0706.1172>.
- [24] XIA, A. (1997a). On the rate of Poisson process approximation to a Bernoulli process. *J. Appl. Prob.* **34**, 898–907.
- [25] XIA, A. (1997b). On using the first difference in the Stein–Chen method. *Ann. Appl. Prob.* **7**, 899–916.
- [26] XIA, A. (2005). Stein's method and Poisson process approximation. In *An Introduction to Stein's Method* (Lecture Notes Ser. Inst. Math. Sci. Nat. Univ. Singapore **4**), Singapore University Press, pp. 115–181.
- [27] XIA, A. AND ZHANG, F. (2008). A polynomial birth–death point process approximation to the Bernoulli process. To appear in *Stoch. Process. Appl.*