Distance estimates for dependent superpositions of point processes

Dominic Schuhmacher [*,1]

Institute of Mathematics, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Abstract

In this article, superpositions of possibly dependent point processes on a general space $\mathcal{X}$ are considered. Using Stein's method for Poisson process approximation, an estimate is given for the Wasserstein distance d_2 between the distribution of such a superposition and an appropriate Poisson process distribution. This estimate is compared to a modern version of Grigelionis' theorem, and to results of Banys (1980), Arratia, Goldstein, and Gordon (1989), and Barbour, Holst, and Janson (1992). Furthermore, an application to a spatial birth-death model is presented.

Key words: Point processes, Poisson process approximation, Stein's method, superposition, Wasserstein distance, Barbour-Brown distance

2000 MSC: primary 60G55; secondary 62E20.

1 Introduction

Superposition is the historical term for *sum* when the summands are point processes. It is a standard result that the superposition of independent and uniformly sparse processes converges in distribution to a Poisson process as the number of processes and the sparseness increase; a fact which forms for example the theoretical backing for many Poisson models of random occurences in time.

* Tel.: +41-44-6355852; fax: +41-44-6355706.
Email address: `dominic.schuhmacher@math.unizh.ch` (Dominic Schuhmacher).
URL: `www.math.unizh.ch/~schumi/` (Dominic Schuhmacher).

[1] Work supported in part by Schweizerischer Nationalfonds, Project Nos. 20-61753.00 and 20-67909.02, and by the Institute for Mathematical Sciences of the National University of Singapore.

Convergence results of this type were first examined in the context of mass service in telecommunications, with Palm (1943) and Khinchin (1955) being the first sources of formal proofs for Poisson limit theorems, albeit under quite strong assumptions. A general Poisson limit theorem for independent superpositions was then obtained in Grigelionis (1963) for the state space $\mathbb{R}_+$, versions for more universal state spaces in Goldman (1967) and Jagers (1972). We formulate Kallenberg's version of Grigelionis' theorem below. A discussion of results with general infinitely divisible point processes in the limit can be found in Matthes, Kerstan, and Mecke (1978). Note in particular Theorem 3.4.2, which contains Grigelionis' theorem as a special case. All the sources mentioned so far restrict themselves to superpositions of independent point processes. Corresponding results for dependent (mixing) point processes with Poisson and compound Poisson processes in the limit can be found in Banys (1980). A first weak distance estimate for the finite dimensional distributions of an independent superposition on the real line was obtained in Theorem 2 of Grigelionis (1963). A much stronger result in total variation distance for superpositions of processes with dependent numbers but independent positions of points is an immediate consequence of Theorem 10.H in Barbour, Holst, and Janson (1992). Many rather specialized contributions have to go unmentioned here. For a listing of authors of the more general results left out above the reader is referred to the historical remarks for Chapter 6 and 7 in Kallenberg (1986) and the introduction of Serfozo (1984) (note the article itself).

In the present study, we focus on Poisson process approximation of dependent superpositions, which is the same setting as in Banys (1980), Section 2. We give an explicit upper bound for a Wasserstein-type of distance (denoted by d_2 and called the Barbour-Brown distance) between the law of such a superposition and a Poisson process law. d_2 has proved to be a useful metric between point process distributions in many instances. See the references at the end of this section, given after the definition and some elementary results.

Our proof is based on Stein's method for Poisson process approximation as presented in Barbour (1988). A short account of this method can be found in the appendix. Stein's method in general is known especially for making no fundamental distinction between independent random elements and different degrees of weakly dependent random elements, which makes it possible to analyze the situation of moderately dependent point processes (where the dependence is controlled by a mixing condition) without too much difficulty.

After the removal of all but at most one point from each of the superimposed point processes, Stein's method is applied directly in our proof. Another strategy would be to align the point processes in a bigger state space, assembling a marked process of the form $\sum_{i \in \mathbb{N}} \delta_i \otimes \xi_i$ instead of the superposition $\sum_{i \in \mathbb{N}} \xi_i$, so that the dependence between the point processes ξ_i is expressed as spatial dependence. There are various theorems then that can be used for such a situation, like Theorem 3.6 in Barbour and Brown (1992), Theorem 2.3 in Chen and Xia (2004), or, in special cases, Theorem 2.A in Schuhmacher (2005a). It can easily be seen that the bounds obtained by these theorems for the approximation of the marked process are then also valid for the corresponding superposition approximation. The results derived in such a manner have similar flavor as the estimate presented in this paper, but are still quite different in certain aspects, especially in the way the dependence between the point processes enters the upper bound.

A detailed comparison with other related results is given in Remark 2.4. Our estimate typically performs well compared to these results. It is much more generally applicable than the distance estimates previously obtained (usually in the stronger total variation distance), and it yields bounds that imply convergence under conditions very similar to those of previous limit theorems.

In Section 3, we give an application of our upper bound in the context of a spatial birth-death model. We examine the development of an animal population modeled by assigning a birth-death process to each of the individuals. These processes may depend on each other according to the spatial arrangement of the animals. We show that the events occurring in the population over a short period of time are approximately composed of two independent Poisson processes, one for the births and one for the deaths, and give explicit bounds for the approximation.

In the remainder of this introduction we give the necessary definitions and notation along with some basic results. For the whole article let $\mathcal{X}$ be a locally compact, second countable Hausdorff space (lcscH space). Denote by $\mathcal{B}$ the Borel σ-algebra on $\mathcal{X}$, and by $\mathcal{B}_\lambda$ for any measure λ on $\mathcal{X}$ the algebra $\{B \in \mathcal{B}; \lambda(\partial B) = 0\}$. Write furthermore $\mathcal{M}$ and $\mathcal{N}$ for the space of boundedly finite measures on $\mathcal{X}$ (i.e. measures which take finite values at every relatively compact set) and the subspace of boundedly finite point measures on $\mathcal{X}$, respectively, and denote the usual σ-algebras on these spaces by $\mathfrak{M}$ resp. $\mathfrak{N}$ (see Kallenberg (1986), Section 1.1 for the corresponding definitions). A point process is then defined as a random element of $(\mathcal{N}, \mathfrak{N})$. Convergence in distribution of point processes is defined with respect to the vague topology on $\mathcal{N}$ (see Kallenberg (1986), Section 15.7) and written in the form $\xi_n \xrightarrow{\mathcal{D}} \xi$ for $n \to \infty$, where $\xi, \xi_1, \xi_2, \ldots$ are point processes. The next definition is used to formulate Grigelionis' theorem.

Definition 1.1 *For every $n, i \in \mathbb{N}$ let ξ_{ni} be a point process on $\mathcal{X}$. The collection $(\xi_{ni})_{n,i}$ is called a* null array *if*

(a) $(\xi_{ni})_{i\in\mathbb{N}}$ *is an independent sequence of point processes for every $n \in \mathbb{N}$;*
(b) $\sup_{i\in\mathbb{N}} \mathbb{P}[\xi_{ni}(B) \geq 1] \longrightarrow 0$ *as $n \to \infty$ for every bounded $B \in \mathcal{B}$.*

We now state Grigelionis' theorem in the version of Kallenberg (2002), Theorem 16.18, in order to have some basic possibility of comparison for our result in Section 2.

Theorem 1.2 (Grigelionis) *Let $(\xi_{ni})_{n,i\in\mathbb{N}}$ be a null array of point processes on $\mathcal{X}$. Furthermore let λ be a boundedly finite measure on $\mathcal{X}$ and denote by η the Poisson process on $\mathcal{X}$ with parameter measure λ. Then $\sum_{i=1}^{\infty}\xi_{ni} \xrightarrow{\mathcal{D}} \eta$ for $n \to \infty$ iff the following conditions hold:*

(i) $\sum_i \mathbb{P}[\xi_{ni}(B) \geq 1] \longrightarrow \lambda(B)$ $(n \to \infty)$ *for every bounded* $B \in \mathcal{B}_\lambda$*;*
(ii) $\sum_i \mathbb{P}[\xi_{ni}(B) \geq 2] \longrightarrow 0$ $(n \to \infty)$ *for every bounded* $B \in \mathcal{B}$*.*

Finally, we give a brief description of the distance d_2 we are going to use, along with some basic results. We call this distance the Barbour-Brown distance, according to its introduction in Barbour and Brown (1992). It can be constructed essentially as two Wasserstein metrics, one on top of the other. Suppose $\mathcal{X}$ is now compact, and d_0 is a metric on $\mathcal{X}$ that generates the topology on $\mathcal{X}$ and is bounded by one. It is always possible to find such a metric since any lcscH is Polish and trimming of the metric has no influence on the generated topology. We denote for any point measure $\varrho \in \mathcal{N}$ by $|\varrho| := \varrho(\mathcal{X}) < \infty$ its total number of points.

Definition 1.3 *(a) The* d_1-distance *(w.r.t. d_0) between point measures $\varrho_1, \varrho_2 \in \mathcal{N}$ is defined by*

$$d_1(\varrho_1,\varrho_2) := \begin{cases} 1 & \text{if } |\varrho_1| \neq |\varrho_2| \\ \frac{1}{|\varrho_1|}\sup_{g\in\mathcal{F}_1}\left|\int g\, d\varrho_1 - \int g\, d\varrho_2\right| & \text{if } |\varrho_1| = |\varrho_2| \geq 1 \\ 0 & \text{if } |\varrho_1| = |\varrho_2| = 0 \end{cases},$$

where $\mathcal{F}_1 := \{g : \mathcal{X} \to \mathbb{R};\ |g(x_1) - g(x_2)| \leq d_0(x_1, x_2)\}$.

(b) The Barbour-Brown distance d_2 *(w.r.t. d_0) between probability measures P and Q on $\mathcal{N}$ is defined by*

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

where $\mathcal{F}_2 := \{f : \mathcal{N} \to \mathbb{R};\ |f(\varrho_1) - f(\varrho_2)| \leq d_1(\varrho_1, \varrho_2)\}$.

The following proposition summarizes some results in connection with the Barbour-Brown distance.

Proposition 1.4 *(i) For point measures ϱ_1, ϱ_2 on $\mathcal{X}$ with $\varrho_1 = \sum_{i=1}^{v}\delta_{s_{1,i}}$, $\varrho_2 = \sum_{i=1}^{v}\delta_{s_{2,i}}$, $v \geq 1$, we have that*

$$d_1(\varrho_1,\varrho_2) = \min_{\pi\in\Sigma_v} \tfrac{1}{v}\textstyle\sum_{i=1}^{v} d_0(s_{1,i}, s_{2,\pi(i)}),$$

where Σ_v denotes the set of permutations on $\{1, 2, \ldots, v\}$.

(ii) For probability measures P, Q on $\mathcal{N}$ we have that

$$d_2(P,Q) = \min_{\substack{\xi_1\sim P\\ \xi_2\sim Q}} \mathbb{E}\, d_1(\xi_1,\xi_2).$$

(iii) The Barbour-Brown distance metrizes the weak convergence of point process distributions, that is for point processes $\xi, \xi_1, \xi_2, \ldots$ on $\mathcal{X}$ we have that $\xi_n \xrightarrow{\mathcal{D}} \xi$ iff $d_2\Big(\mathcal{L}(\xi_n), \mathcal{L}(\xi)\Big) \longrightarrow 0$.

For a more detailed account of the distance including proofs of the above results, the reader is referred to Barbour, Holst, and Janson (1992), Section 10.2, and to Schuhmacher (2005a), Sections 1 and 3. Further details and recent applications of d_2 include the results of Brown, Weinberg, and Xia (2000), Barbour, Novak, and Xia (2002), Barbour and Månsson (2002), and Chen and Xia (2004).

2 The distance estimates

We state in this section the main theorem, which gives an upper bound for the distance between the distribution of a superposition and a corresponding Poisson process distribution. Note that this is a static result, so there is no need to have n in our notation, nor is there anything else going to infinity. To demonstrate the usefulness of our result we compare the upper bound to the convergence conditions in Grigelionis' theorem and in Theorem 4 in Banys (1980), as well as to the related results of Theorem 2 in Arratia, Goldstein, and Gordon (1989), and Theorem 10.H in Barbour, Holst, and Janson (1992).

Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of point processes on the compact metric space $(\mathcal{X}, d_0)$ which satisfies $0 < \sum_{i=1}^{\infty} \mathbb{P}[|\xi_i| \geq 1] < \infty$. For each $i \in \mathbb{N}$, partition $\mathbb{N}$ as $\Big\{\{i\}, \Gamma_i^{\mathrm{s}}, \Gamma_i^{\mathrm{w}}\Big\}$, where the idea is that ξ_j depends "strongly" on ξ_i for $j \in \Gamma_i^{\mathrm{s}}$, and ξ_j depends "weakly" on ξ_i for $j \in \Gamma_i^{\mathrm{w}}$. There is, however, no formal requirement for these partitions: if $\mathbb{N} \setminus \{i\}$ is split up "unnaturally" for a large part of the i, the bound below is still true, but can be very bad.

Choose for each point process ξ_i a representation as $\xi_i = \sum_{k=1}^{|\xi_i|} \delta_{S_i^{(k)}}$, where $|\xi_i| := \xi_i(\mathcal{X})$ and $S_i^{(k)}$ are $\sigma(\xi_i)$-measurable random variables for $k \in \mathbb{N}$. That such representations exist is a direct consequence of Lemma 2.3 in Kallenberg (1986). Set furthermore $S_i := S_i^{(1)}$. For any finite measure λ on $\mathcal{X}$ denote by $\mathrm{Po}(\lambda)$ the distribution of the Poisson process with parameter measure λ.

We have the following main result.

Theorem 2.1 *Let* $p_i := \mathbb{P}[|\xi_i| \geq 1]$, $p_i' := \mathbb{P}[|\xi_i| \geq 2]$ *and define the measure* μ *on* $\mathcal{X}$ *by* $\mu(B) := \sum_{i=1}^{\infty} \mathbb{P}[|\xi_i| \geq 1, S_i \in B]$ *for every* $B \in \mathcal{B}$. *Then*

$$\begin{aligned} & d_2\Big(\mathcal{L}\Big(\sum_{i=1}^{\infty} \xi_i\Big), \operatorname{Po}(\mu)\Big) \\ & \qquad \leq \sum_{i=1}^{\infty} p_i' + M_2 \sum_{i=1}^{\infty} p_i^2 + M_2 \sum_{i=1}^{\infty} \sum_{j \in \Gamma_i^{\mathrm{s}}} \Big(p_i p_j + \mathbb{P}\Big[|\xi_i| \geq 1, |\xi_j| \geq 1\Big]\Big) \\ & \qquad\quad + (M_1 + M_2) \sum_{i=1}^{\infty} \mathbb{E}\Big|\mathbb{P}\Big[|\xi_i| \geq 1 \;\Big|\; (\xi_j)_{j \in \Gamma_i^{\mathrm{w}}}\Big] - p_i\Big| \\ & \qquad\quad + M_2 \sum_{i=1}^{\infty} p_i \, \mathbb{E} \, d_W\Big(\mathcal{L}\Big(S_i \;\Big|\; |\xi_i| \geq 1\Big), \mathcal{L}\Big(S_i \;\Big|\; |\xi_i| \geq 1, (\xi_j)_{j \in \Gamma_i^{\mathrm{w}}}\Big)\Big), \end{aligned}$$

where M_1, M_2 *are given by Formulae* (A.5) *and* (A.6) *in the appendix, and* d_W *denotes the Wasserstein distance on* $\mathcal{X}$ *with respect to* d_0 *(see e.g. Barbour, Holst, and Janson (1992), Appendix A.1 for the definition and some elementary results).*

Remark 2.2 (Poisson process with slightly different parameter measure) We obtain a similar result for $d_2\big(\mathcal{L}(\sum_{i=1}^{\infty} \xi_i), \operatorname{Po}(\tilde{\mu})\big)$ with $\tilde{\mu}(B) = \sum_{i=1}^{\infty} \mathbb{P}[|\xi_i| = 1, \xi_i(B) = 1]$ for every $B \in \mathcal{B}$. Just replace in the above theorem "$|\xi_i| \geq 1$" with "$|\xi_i| = 1$" every time it occurs. The advantage of this alternative result is that no explicit representations of the ξ_i are needed for its formulation.

Corollary 2.3 *Let* $(\xi_{ni})_{n,i}$ *be a null array of point processes, and* λ *a measure on* $\mathcal{X}$ *with* $|\lambda| := \lambda(\mathcal{X}) < \infty$. *In the notation of Theorem 2.1, with* p_i, p_i' *and* μ *depending now on* n, *we have*

$$\begin{aligned} & d_2\Big(\mathcal{L}\Big(\sum_{i=1}^{\infty} \xi_{ni}\Big), \operatorname{Po}(\lambda)\Big) \\ & \leq \sum_{i=1}^{\infty} p_i' + M_2 \sum_{i=1}^{\infty} p_i^2 + \min\Big(1, \tfrac{1.65}{\sqrt{|\mu|}}, \tfrac{1.65}{\sqrt{|\lambda|}}\Big)\Big||\mu| - |\lambda|\Big| + \Big(1 - e^{\min(|\mu|, |\lambda|)}\Big) d_W\Big(\tfrac{\mu}{|\mu|}, \tfrac{\lambda}{|\lambda|}\Big), \end{aligned}$$

which under the Conditions (i) and (ii) of Grigelionis' theorem goes to zero as $n \to \infty$.

Remark 2.4 (Comparisons with other results)

(a) The sufficiency of Conditions (i) and (ii) in Grigelionis' theorem 1.2 is by Proposition 1.4(iii) an immediate consequence of Corollary 2.3.

(b) Theorem 4 of Banys (1980), which like Grigelionis' theorem is a mere convergence result, is not implied directly by Theorem 2.1, but the two theorems have very similar flavor. Banys uses in indirect form also the concepts of an index set Γ_i^{s} of strong dependence and an index set Γ_i^{w} of weak dependence, his assumptions of them being weaker in so far as for every index i only $\{1, 2, \ldots, i-1\}$ has to be partitioned, but stronger in so far as there is less freedom in the choice of these partitions. Apart from this difference however, the summands in our upper bound correspond directly to the terms that have to go to zero in Banys' theorem in order to ensure convergence of the superposition. They are even exactly the same, except for the last two summands, which capture the weak long range dependence: in Banys' theorem this dependence is controlled by the smallness of terms of the form (in our notation)

$$\sum_{i=1}^{\infty} \mathbb{E}\Big|\mathbb{P}[\xi_i(B) \geq 1 \,|\, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}}] - \mathbb{P}[\xi_i(B) \geq 1]\Big|$$

for every $B \in \mathcal{B}$.

(c) The setting of Theorem 10.H from Barbour, Holst, and Janson (1992) is a special case of the setting of Theorem 2.1 from this paper except for the stronger total variation distance that was used there. To cope with this distance, strong assumptions about the independence of the point positions were made in Theorem 10.H, which we do not need for our Theorem 2.1 (note that the motivation for Theorem 10.H was a very different one). The basic ideas of the proofs are in both theorems the same. Under the more restrictive setting of Theorem 10.H, the upper bounds obtained for the two distances are also the same, up to some rather slight differences in the factors M_1 and M_2.

(d) In Arratia, Goldstein, and Gordon (1989), Theorem 2, an upper bound is given for the total variation distance between a dependent Bernoulli process $(X_{\alpha_i})_{i\in\mathbb{N}}$ and a Poisson process $(Y_{\alpha_i})_{i\in\mathbb{N}}$ on an arbitrary set $\{\alpha_i; i \in \mathbb{N}\}$. There are two ways in which this situation can be related to the setting of Theorem 2.1. The more obvious way is by contrasting the sequence $(X_{\alpha_i})_{i\in\mathbb{N}}$ with the sequence $(\xi_i)_{i\in\mathbb{N}}$. Thus, where Arratia, Goldstein, and Gordon use dependent indicator random variables, we use point processes with exactly the same local dependence structure. Where they examine the common distribution of the indicators (and also their sum) in the total variation distance, we examine the sum of our point processes in the Barbour-Brown distance. The other way to relate the two situations, is by setting $\xi_i := X_{\alpha_i}\delta_{\alpha_i}$. Thus, we obtain the Bernoulli process $(X_{\alpha_i})_{i\in\mathbb{N}}$ as a very special case of our superposition $\sum_i \xi_i$ (there is at most one point per point process, and its position is deterministic). Since we use the weaker Barbour-Brown distance the results are not comparable. However, a comparison with a corresponding d_2 upper bound for the Bernoulli situation, such as the one in Theorem 10.F of Barbour, Holst, and Janson (1992) (ignoring the last summand in their upper bound, which just stems from an additional comparison between two Poisson processes), yields that our upper bound is qualitatively exactly the same, and differs in absolute terms only by having the factor $(M_1 + M_2)$ instead of only M_1 in front of the fourth summand. So, up to this changed factor, the bound in Theorem 10.F can be obtained as a special case of our Theorem 2.1. Our result is strictly more general, among other things in that the neighborhoods of strong dependence Γ_i^{s} are not bound to fixed regions of the state space.

Proof of Theorem 2.1

Our strategy is to reduce the point processes ξ_i to their first points, and then apply Stein's method to the superposition of these reduced point processes. Let η be a Poisson point process with parameter measure μ, and split up the initial distance as

$$d_2\Big(\mathcal{L}(\textstyle\sum_i \xi_i), \mathcal{L}(\eta)\Big) \;\leq\; d_2\Big(\mathcal{L}(\textstyle\sum_i \xi_i), \mathcal{L}(\textstyle\sum_i I_i \delta_{S_i})\Big) + d_2\Big(\mathcal{L}(\textstyle\sum_i I_i \delta_{S_i}), \mathcal{L}(\eta)\Big),$$

where $I_i := 1_{\{|\xi_i| \geq 1\}}$ for every $i \in \mathbb{N}$.

The reduction term is very easily estimated, using the "natural coupling" of the two processes. By Proposition 1.4(ii) we have

$$\begin{aligned} & d_2\Big(\mathcal{L}(\textstyle\sum_i \xi_i), \mathcal{L}(\textstyle\sum_i I_i \delta_{S_i})\Big) \\ &\qquad \leq\; \mathbb{E}\, d_1(\textstyle\sum_i \xi_i, \sum_i I_i \delta_{S_i}) \\ &\qquad =\; \mathbb{P}\Big[|\textstyle\sum_i \xi_i| \neq |\sum_i I_i \delta_{S_i}|\Big] + \mathbb{E}\Big(d_1(\textstyle\sum_i \xi_i, \sum_i I_i \delta_{S_i})\, 1_{\{|\sum_i \xi_i| \,=\, |\sum_i I_i \delta_{S_i}|\}}\Big) \\ &\qquad =\; \mathbb{P}\Big[\textstyle\bigcup_i \{|\xi_i| \geq 2\}\Big] \\ &\qquad \leq\; \sum_i p_i', \end{aligned} \tag{2.1}$$

where the expectation in the third line is zero, because $I_i \delta_{S_i} \leq \xi_i$ for every $i \in \mathbb{N}$.

For the distance between the distributions of the reduced superposition and the Poisson process we apply Stein's method for Poisson process approximation (Barbour (1988)). A short sketch of this method can be found in the appendix. Set $\Xi := \sum_{i=1}^{\infty} I_i \delta_{S_i}$, and $\Xi_i := \sum_{j \in \mathbb{N} \setminus \{i\}} I_j \delta_{S_j}$ and $\Xi_i^{\mathrm{w}} := \sum_{j \in \Gamma_i^{\mathrm{w}}} I_j \delta_{S_j}$ for every $i \in \mathbb{N}$. Choose random elements $\widetilde{S}_1, \widetilde{S}_2, \ldots$ in $\mathcal{X}$ that are independent among each other and of anything else, such that $\widetilde{S}_i \sim \mathcal{L}(S_i \,|\, I_i = 1)$. Fix $f \in \mathcal{F}_2$ and let $h = h_f$ be the solution to the Stein equation (A.4) given by (A.3). Then we have that

$$\begin{aligned} & \Big|\mathbb{E} f(\Xi) - \mathbb{E} f(\eta)\Big| \\ &\qquad = \Big|\mathbb{E} \int_{\mathcal{X}} \Big[h(\Xi - \delta_s) - h(\Xi)\Big]\; \Xi(ds) + \mathbb{E} \int_{\mathcal{X}} \Big[h(\Xi + \delta_s) - h(\Xi)\Big]\; \mu(ds)\Big| \\ &\qquad = \Big|\mathbb{E}\Big(\sum_{i=1}^{\infty} I_i \Big[h(\Xi - \delta_{S_i}) - h(\Xi)\Big]\Big) + \mathbb{E}\Big(\sum_{i=1}^{\infty} p_i\, \mathbb{E}\Big(h(\Xi + \delta_{\widetilde{S}_i}) - h(\Xi) \;\Big|\; \Xi\Big)\Big)\Big| \\ &\qquad \leq \sum_{i=1}^{\infty} \Big|\mathbb{E}\Big(I_i \Big[h(\Xi - \delta_{S_i}) - h(\Xi)\Big]\Big) + \mathbb{E}\Big(p_i \Big[h(\Xi + \delta_{\widetilde{S}_i}) - h(\Xi)\Big]\Big)\Big|, \end{aligned}$$

where we used that $\int g(x)\, \mu(dx) = \sum_{i=1}^{\infty} \int g(x)\, \mu_i(dx)$ for $\mu_1, \mu_2, \ldots$ and $\mu = \sum_{i=1}^{\infty} \mu_i$ finite measures on $\mathcal{X}$ and $g \in \mathcal{L}_1(\mu)$ in the third line, and Fubini's Theorem in the last line (both based on $\sum_i \mathbb{P}[|\xi_i| \geq 1] < \infty$ and $\Delta_1 h \leq 1$). The i-th summand can then be split up further as

$$\Big|\mathbb{E}\Big(I_i\Big[h(\Xi-\delta_{S_i})-h(\Xi)\Big]\Big)-\mathbb{E}\Big(p_i\Big[h(\Xi)-h(\Xi+\delta_{\widetilde{S}_i})\Big]\Big)\Big|$$

$$\begin{aligned}
&\le \Big|\mathbb{E}\Big(I_i\Big[h(\Xi-\delta_{S_i})-h(\Xi)\Big]\Big)-\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{S_i})\Big]\Big)\Big| \\
&\quad+\Big|\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{S_i})\Big]\Big)-\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{\widetilde{S}_i})\Big]\Big)\Big| \\
&\quad+\Big|\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{\widetilde{S}_i})\Big]\Big)-\mathbb{E}\Big(p_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{\widetilde{S}_i})\Big]\Big)\Big| \\
&\quad+\Big|\mathbb{E}\Big(p_i\Big[h(\Xi_i^{\mathrm{w}})-h(\Xi_i^{\mathrm{w}}+\delta_{\widetilde{S}_i})\Big]\Big)-\mathbb{E}\Big(p_i\Big[h(\Xi)-h(\Xi+\delta_{\widetilde{S}_i})\Big]\Big)\Big|. \qquad (2.2)
\end{aligned}$$

The first summand in Inequality (2.2)

Assume without loss of generality that Γ_i^{s} is an infinite set, for if it is not, we can add infinitely many 0-processes to the superposition and put all their indices into Γ_i^{s}. Enumerate the elements in Γ_i^{s} by $r(1), r(2), \ldots$ and write $\Xi_i^{\mathrm{w},l} := \Xi_i^{\mathrm{w}} + \sum_{j=1}^{l} I_{r(j)}\delta_{S_{r(j)}}$ for $l \ge 0$. Since by Borel-Cantelli $\sum_{j=1}^{\infty} I_{r(j)}$ is almost surely finite, the first summand can be expanded into a telescopic sum, and is hence equal to

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_i\sum_{l=1}^{\infty}\Big(\Big[h(\Xi_i^{\mathrm{w},l})-h(\Xi_i^{\mathrm{w},l}+\delta_{S_i})\Big]-\Big[h(\Xi_i^{\mathrm{w},l-1})-h(\Xi_i^{\mathrm{w},l-1}+\delta_{S_i})\Big]\Big)\Big)\Big| \\
&\qquad\le \sum_{l=1}^{\infty}\Big|\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w},l})-h(\Xi_i^{\mathrm{w},l}+\delta_{S_i})-h(\Xi_i^{\mathrm{w},l-1})+h(\Xi_i^{\mathrm{w},l-1}+\delta_{S_i})\Big]\Big)\Big|.
\end{aligned}$$

The moduli can be further bounded by

$$\mathbb{E}\Big(I_iI_{r(l)}\Big|h(\Xi_i^{\mathrm{w},l-1}+\delta_{S_{r(l)}}+\delta_{S_i})-h(\Xi_i^{\mathrm{w},l-1}+\delta_{S_{r(l)}})-h(\Xi_i^{\mathrm{w},l-1}+\delta_{S_i})+h(\Xi_i^{\mathrm{w},l-1})\Big|\Big),$$

such that by Inequality (A.6) the total bound for the first summand in Inequality (2.2) is

$$M_2\sum_{l=1}^{\infty}\mathbb{E}(I_iI_{r(l)}) = M_2\sum_{j\in\Gamma_i^{\mathrm{s}}}\mathbb{E}(I_iI_j).$$

The second summand in Inequality (2.2)

We first show that for any $\varrho \in \mathcal{N}$ the function $g_\varrho : \mathcal{X} \to \mathbb{R}$ given by

$$g_\varrho(s) := h(\varrho + \delta_s) \quad \text{for all } s \in \mathcal{X}$$

is d_0-Lipschitz continuous with constant $C := 1 \wedge \frac{1}{|\mu|}\Big(\log^+(|\mu|)+1\Big)$. This is done in a similar way as the bounds (A.5) and (A.6) are obtained. Write the spatial immigration-death processes Z and Z' with (deterministic) initial configurations $\varrho+\delta_s$ and $\varrho+\delta_{s'}$ as $Z_1+\delta_s 1_{\{E>t\}}$ and $Z_1+\delta_{s'} 1_{\{E>t\}}$, respectively, where E is a standard exponentially distributed random variable that is independent of everything else, and Z_1 is the immigration-death process with immigration measure μ and unit per-capita death rate that starts with configuration ϱ. We furthermore write Z_0 for the same immigration-death process that starts with zero points. Note that $Z_0(t) \sim \mathrm{Po}((1-e^{-t})\mu)$, and write $\mu_t := (1-e^{-t})|\mu|$. Then, using the explicit form of h given by Equation (A.3), we have that

$$\begin{aligned}\Big|g_\varrho(s) - g_\varrho(s')\Big| &= \Big|\int_0^\infty \Big[\mathbb{E}f(Z(t)) - \mathbb{E}f(Z'(t))\Big]\,dt\Big| \\ &= \Big|\int_0^\infty \Big[\mathbb{E}f(Z_1(t)+\delta_s) - \mathbb{E}f(Z_1(t)+\delta_{s'})\Big]\,\mathbb{P}[E>t]\,dt\Big| \\ &\le \int_0^\infty \mathbb{E}\,d_1\Big(Z_1(t)+\delta_s, Z_1(t)+\delta_{s'}\Big)\,e^{-t}\,dt \\ &= d_0(s,s')\int_0^\infty \mathbb{E}(|Z_1(t)|+1)^{-1}e^{-t}\,dt \end{aligned} \tag{2.3}$$

by Proposition 1.4(i), where furthermore

$$\mathbb{E}\Big(\frac{1}{|Z_1(t)|+1}\Big) \le \mathbb{E}\Big(\frac{1}{|Z_0(t)|+1}\Big) = \frac{1-e^{-\mu_t}}{\mu_t}.$$

Hence it follows that the integral at the end of Inequality (2.3) is bounded by C, which yields the required Lipschitz continuity.

The second term in Inequality (2.2) is now estimated for $p_i > 0$ as

$$\begin{aligned}&\Big|\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}}+\delta_{S_i})\Big]\Big) - \mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}}+\delta_{\widetilde{S}_i})\Big]\Big)\Big| \\ &\qquad = \Big|\mathbb{E}\Big(I_i\Big[g_{\Xi_i^{\mathrm{w}}}(\widetilde{S}_i) - g_{\Xi_i^{\mathrm{w}}}(S_i)\Big]\Big)\Big| \\ &\qquad = \Big|\mathbb{E}\Big(I_i\,\mathbb{E}\Big(g_{\Xi_i^{\mathrm{w}}}(\widetilde{S}_i) - g_{\Xi_i^{\mathrm{w}}}(S_i)\;\Big|\; I_i=1, \Xi_i^{\mathrm{w}}\Big)\Big)\Big| \\ &\qquad \le C\,\mathbb{E}\Big(I_i\,d_W\Big(\mathcal{L}(S_i\,|\,I_i=1), \mathcal{L}(S_i\,|\,I_i=1,\Xi_i^{\mathrm{w}})\Big)\Big) \\ &\qquad \le C\,p_i\,\mathbb{E}\,d_W\Big(\mathcal{L}(S_i\,|\,I_i=1), \mathcal{L}(S_i\,|\,I_i=1,\Xi_i^{\mathrm{w}})\Big) + C\,\mathbb{E}\Big|\mathbb{E}(I_i\,|\,\Xi_i^{\mathrm{w}}) - p_i\Big|,\end{aligned}$$

where for the third line we used that $I_i\,F(I_i, X) = I_i\,F(1, X)$ for any random variable X and any function F. Note that the d_W-term is a measurable function in Ξ_i^{w}, because the supremum in its definition can be substituted by the supremum over a countable set of functions. The overall bound above is trivially true for $p_i = 0$ as well.

The third summand in Inequality (2.2)

Since $\widetilde{S}_i$ is independent of $(I_i, \Xi_i^{\mathrm{w}})$ and hence $\widetilde{S}_i \perp\!\!\!\perp_{\Xi_i^{\mathrm{w}}} I_i$ (i.e. $\widetilde{S}_i$ is independent of I_i given Ξ_i^{w}), we obtain for the third summand

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_i\Big[h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}} + \delta_{\widetilde{S}_i})\Big]\Big) - \mathbb{E}\Big(p_i\Big[h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}} + \delta_{\widetilde{S}_i})\Big]\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(\mathbb{E}\Big((I_i - p_i)\Big[h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}} + \delta_{\widetilde{S}_i})\Big] \;\Big|\; \Xi_i^{\mathrm{w}}\Big)\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(\mathbb{E}(I_i - p_i \,|\, \Xi_i^{\mathrm{w}})\, \mathbb{E}\Big(h(\Xi_i^{\mathrm{w}}) - h(\Xi_i^{\mathrm{w}} + \delta_{\widetilde{S}_i}) \;\Big|\; \Xi_i^{\mathrm{w}}\Big)\Big)\Big| \\
&\qquad \leq M_1\, \mathbb{E}\Big|\mathbb{E}(I_i \,|\, \Xi_i^{\mathrm{w}}) - p_i\Big|
\end{aligned}$$

by Inequality (A.5).

The fourth summand in Inequality (2.2)

We proceed in the same way as for the first summand and use the corresponding notation. Thus the fourth summand can be expanded into a telescopic sum and hence estimated by

$$\begin{aligned}
&p_i \sum_{l=1}^{\infty}\Big|\mathbb{E}\Big(\Big[h(\Xi_i^{\mathrm{w},l}) - h(\Xi_i^{\mathrm{w},l} + \delta_{\widetilde{S}_i})\Big] - \Big[h(\Xi_i^{\mathrm{w},l-1}) - h(\Xi_i^{\mathrm{w},l-1} + \delta_{\widetilde{S}_i})\Big]\Big)\Big| \\
&\qquad + p_i\Big|\mathbb{E}\Big(\Big[h(\Xi) - h(\Xi + \delta_{\widetilde{S}_i})\Big] - \Big[h(\Xi_i) - h(\Xi_i + \delta_{\widetilde{S}_i})\Big]\Big)\Big|.
\end{aligned}$$

We bound the moduli in the analogous way as for the first summand and have thus

$$M_2\Big(\sum_{j\in\Gamma_i^{\mathrm{s}}} p_i p_j + p_i^2\Big)$$

as the total bound for the fourth summand.

Assembling of the four estimates for the summands in Inequality (2.2) yields

$$\begin{aligned}
d_2\Big(\mathcal{L}(\textstyle\sum_i I_i \delta_{S_i}), \mathcal{L}(\eta)\Big) &= \sup_{f\in\mathcal{F}_2}\Big|\mathbb{E}f(\Xi) - \mathbb{E}f(\eta)\Big| \\
&\leq M_2 \sum_{i=1}^{\infty} p_i^2 + M_2 \sum_{i=1}^{\infty}\sum_{j\in\Gamma_i^{\mathrm{s}}}\Big(p_i p_j + \mathbb{E}(I_i I_j)\Big) \\
&\quad + (M_1 + C)\sum_{i=1}^{\infty}\mathbb{E}\Big|\mathbb{E}(I_i \,|\, \Xi_i^{\mathrm{w}}) - p_i\Big| \\
&\quad + C\sum_{i=1}^{\infty} p_i\, \mathbb{E}\, d_W\Big(\mathcal{L}(S_i \,|\, I_i = 1), \mathcal{L}(S_i \,|\, I_i = 1, \Xi_i^{\mathrm{w}})\Big).
\end{aligned}$$

Together with Inequality (2.1), and noting that $C \leq M_2$ and that Ξ_i^{w} is a measurable function of $(\xi_j)_{j \in \Gamma_i^{\mathrm{w}}}$, we obtain the required result. □

Proof of Corollary 2.3

Choose $\Gamma_i^{\mathrm{s}} := \emptyset$ and $\Gamma_i^{\mathrm{w}} := \Gamma_i = \mathbb{N} \setminus \{i\}$ in Theorem 2.1, so that the last three terms in the upper bound of the theorem disappear. We are left with the first two terms and the additional term of $d_2(\mathrm{Po}(\mu), \mathrm{Po}(\lambda))$, which can be estimated by Inequality (A.3) in Schuhmacher (2005b) (this goes back to Brown and Xia (1995), Inequality (2.8), but contains a few clarifications). Thus we obtain the required bound.

For the convergence statement note that Condition (i) in Theorem 1.2 implies for $|\mu|, |\lambda| > 0$ that $\frac{\mu}{|\mu|} = \frac{\mu_n}{|\mu_n|}$ converges weakly to $\frac{\lambda}{|\lambda|}$ for $n \to \infty$. Hence the fourth summand in the upper bound goes to zero (see e.g. Dudley (1989), Theorem 11.3.3, taking into account that $d_0 \leq 1$). The first and third summands go to zero directly by Conditions (ii) and (i), respectively. Finally, the second summand can be estimated as $M_2 |\mu| \sup_{i \geq 1} p_i$, which goes to zero, because $(\xi_{ni})_{n,i}$ is a null array, and $|\mu|$ goes to $|\lambda|$. □

3 Application: short term behavior of a spatial birth-death model

Consider a population of n animals that are more or less evenly spread over a certain area. The i-th animal and its offspring at time t are described by a birth-death process (BDP) ζ_i with starting state 1 (see e.g. Feller (1968), Section XVII.5 for the definition). The corresponding processes may be rather strongly dependent if two animals live close together, but the dependence between processes is expected to decay with the distance between animals.

In what follows, we provide a rather general modeling framework for this situation and derive statements about the short-term behavior of populations of the above form. We use biological terminology (such as "animals" or "predators") for illustrative purposes, but it should be noted that the model we present is still somewhat too abstract for a serious modeling attempt of any concrete biological situation. On the other hand, the model is flexible enough to be adapted to many other contexts where a birth-death paradigm is reasonable. Examples include the failure and repair of components in complex systems, attacks in computer networks, absorption and emission of photons or other particles, or arrivals and departures in a large system of queues.

Our concrete model is as follows. Let there be an infinite number of animals in $\mathbb{R}_+^d$ represented by a point measure $\varrho = \sum_{i=1}^{\infty} \delta_{z_i}$. The population described above will consist of the first n of these points, so it is preferable to number them in a reasonable way, e.g. according to their distance from the origin. We assume that this infinite group of animals is "evenly spread", meaning that there is a constant $\kappa > 0$ such that

$$\varrho(\mathbb{B}(z_i, r)) - 1 \leq \kappa r^d \tag{3.1}$$

for all $r \geq 0$ and $i \in \mathbb{N}$, where $\mathbb{B}(z, r)$ denotes the closed Euclidean ball with center in z and radius r. Depending on the situation, other metrics on the set $\{z_i; i \in \mathbb{N}\}$ and more general functions in r bounding the left hand side of (3.1) might be more appropriate.

Let ζ_i, $i \in \mathbb{N}$, be identically distributed BDPs with birth rates $(\alpha_k)_{k\in\mathbb{Z}_+}$ and death rates $(\beta_k)_{k\in\mathbb{Z}_+}$ which all start with one individual at time 0. We think of ζ_i as the process that belongs to the original animal at z_i. The dependence between these processes is controlled by functions $\phi : [0, 1] \to [0, 1]$ and $\chi_\alpha, \chi_\beta : [0, 1] \times \mathbb{R}_+ \to [0, 1]$ which are chosen in such a way that

$$\mathbb{P}[\zeta_i \text{ and } \zeta_j \text{ each have a jump in } [0, t]] \leq t\phi(t) \tag{3.2}$$

for $i \neq j$, and

$$\sup_{F\in\mathcal{F}_i(t)} \mathbb{E}\Big|\mathbb{P}[F \,|\, \mathcal{G}_i(t, r)] - \mathbb{P}[F]\Big| \leq \chi_\alpha(t, r) \tag{3.3}$$

$$\text{and} \quad \mathbb{E}\Big(\operatorname*{ess\,sup}_{F\in\mathcal{F}_i(t)}\Big|\mathbb{P}[F \,|\, \mathcal{G}_i(t, r)] - \mathbb{P}[F]\Big|\Big) \leq \chi_\beta(t, r) \tag{3.4}$$

for every i, where $\mathcal{F}_i(t) := \sigma(\zeta_i|_{[0,t]})$ and $\mathcal{G}_i(t, r) := \sigma(\zeta_j|_{[0,t]}\,;\ j \in \mathbb{N}, |z_j - z_i| > r)$. The idea behind this dependence structure is as follows: ϕ controls the short-term positive correlation of events (births or deaths) happening at points close together. In the biological setting, deaths of animals living close together, for example, might be rather strongly positively correlated, because they might be caused (among other reasons) by predators roaming the neighborhood or by fights among the animals. On the other hand, the χ-functions control the short-term dependence over long distances by providing bounds for the α- and β-mixing coefficients between the evolution of a single process and the evolution of all the processes far enough away. See Doukhan (1994), Section 1.1 for an introduction to mixing coefficients. Note that the term bounded by χ_α in Inequality (3.3) is in fact twice the α-mixing condition used by Doukhan and elsewhere. In the animal framework, the long-range dependence might be caused by the abundance or scarcity of prey or by the environmental conditions (such as climate or vegetation).

Consider now the population of the first n animals. Theorem 2.1 yields a result concerning the aggregated population process, observed over a short period of time $h > 0$. Denote by $\widetilde{\zeta}_i$ the event point process for the BDP ζ_i, which we define as the point process on $\mathbb{R}_+ \times \{0, 1\}$ that has a point in (t, e) (in other words: a point in t with mark e) if ζ_i has an event of type e at time t, where $e = 1$ codes for a birth and $e = 0$ codes for a death. We choose the distance d_0 on $\mathcal{X} := [0, 1] \times \{0, 1\}$ that is defined by $d_0((t_1, e_1), (t_2, e_2)) := \max(|t_2 - t_1|, |e_2 - e_1|)$ for all $(t_1, e_1), (t_2, e_2) \in \mathcal{X}$.

Proposition 3.1 *Suppose that the conditions above hold, that is, let $(\zeta_i)_{i\in\mathbb{N}}$ be identically distributed birth-death processes, attached to the points z_i of a point measure ϱ that satisfies Inequality (3.1) for some $\kappa > 0$, and with birth and death rates $(\alpha_k)_{k\in\mathbb{Z}_+}$ and $(\beta_k)_{k\in\mathbb{Z}_+}$, respectively, and let the dependence between the ζ_i be controlled by Inequalities (3.2), (3.3), and (3.4). Set $\sigma_k := \alpha_k + \beta_k$ for all $k \in \mathbb{Z}_+$, and define furthermore, for $n \in \mathbb{N}$ and $h > 0$, the Poisson intensity measure $\lambda_{n,h}$ on $\mathcal{X}$ by $\lambda_{n,h} := nh(\beta_1(\mathrm{Leb}\otimes\delta_0) + \alpha_1(\mathrm{Leb}\otimes\delta_1))$ and the time dilation function $\theta_h : \mathbb{R}_+ \times \{0,1\} \to \mathbb{R}_+ \times \{0,1\}$ by $\theta_h(t,e) := (t/h, e)$. Write $\xi_i^{(h)} := \widetilde{\zeta}_i\theta_h^{-1}|_{\mathcal{X}}$ for the dilated point process of the events of ζ_i up to time h. Then there is a constant $K := K(\kappa, \sigma_0, \sigma_1, \sigma_2) > 0$, such that*

$$d_2\Bigl(\mathcal{L}\Bigl(\sum_{i=1}^n \xi_i^{(h)}\Bigr), \mathrm{Po}(\lambda_{n,h})\Bigr) \leq K \inf_{r\geq 0}\Bigl(nh^2 + \log^{\uparrow}(nh)r^d\bigl(h \vee \phi(h)\bigr) + \sqrt{nh}\frac{\chi_\alpha(h,r)}{h} + \log^{\uparrow}(nh)\frac{\chi_\beta(h,r)}{h}\Bigr)$$

for any $n \in \mathbb{N}$ and any $h \in (0, 1/\sigma_1)$, where $\log^{\uparrow}(x) := 1 + \log^+(x)$ for $x > 0$.

An upper bound with explicit constants for general $h > 0$, which furthermore improves considerably on the above bound for small nh, can be found at the end of the proof, in Inequality (3.11). To make the result more transparent, we consider the special case where $h = h_n = 1/n$ and some of the other conditions are simplified as well.

Corollary 3.2 *Suppose that the conditions of Proposition 3.1 hold, that additionally $h = h_n = 1/n$ for all $n \in \mathbb{N}$, and that there is a function $\chi : \mathbb{R}_+ \to [0,1]$ with $\chi_\beta(t,r) \leq t\chi(r)$ for $t \in [0,1]$, $r \in \mathbb{R}_+$. Note that $\lambda_{n,1/n} =: \lambda$ does not depend on n now. Then there is a constant $K := K(\kappa, \sigma_0, \sigma_1, \sigma_2) > 0$, such that*

$$d_2\Bigl(\mathcal{L}\Bigl(\sum_{i=1}^n \xi_i^{(1/n)}\Bigr), \mathrm{Po}(\lambda)\Bigr) \leq K \inf_{r\geq 0}\Bigl(\frac{r^d+1}{n} + r^d\phi(1/n) + \chi(r)\Bigr)$$

for any $n \in \mathbb{N}$. For $\phi(t) = O(t^a)$ $(t \to 0)$ and $\chi(r) = O(r^{-bd})$ $(r \to \infty)$ with constants $a, b > 0$, we have

$$d_2\Bigl(\mathcal{L}\Bigl(\sum_{i=1}^n \xi_i^{(1/n)}\Bigr), \mathrm{Po}(\lambda)\Bigr) = O\Bigl(n^{-(a\wedge 1)b/(1+b)}\Bigr) \quad \text{for } n \to \infty,$$

and by Proposition 1.4(iii) that

$$\sum_{i=1}^n \xi_i^{(1/n)} \xrightarrow{\mathcal{D}} \mathrm{Po}(\lambda) \quad \text{for } n \to \infty.$$

□

PROOF. [Proposition 3.1] A few adaptations are necessary for the application of Theorem 2.1. Fix $n \in \mathbb{N}$ and $h > 0$. We exclude a trivial case by assuming that

$\sigma_1 > 0$. Write furthermore T_i for the time of the first event of ζ_i, T_i' for the time between the first and the second event of ζ_i, and E_i and E_i' for the types (0 or 1) of these events, respectively. Note that T_i' might be infinite if σ_0 or σ_2 is zero. Set $S_i := \theta_h(T_i \wedge h, E_i)$ and $S_i' := \theta_h((T_i + T_i') \wedge h, E_i')$, and write $\xi_i := \xi_i^{(h)}$ and $\lambda := \lambda_{n,h}$. In order to obtain the right Poisson intensity measure for the theorem, we split up the initial distance as

$$d_2\Big(\mathcal{L}\Big(\sum_{i=1}^n \xi_i\Big), \mathrm{Po}(\lambda)\Big) \leq d_2\Big(\mathcal{L}\Big(\sum_{i=1}^n \xi_i\Big), \mathrm{Po}(\mu)\Big) + d_2\Big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\Big) \tag{3.5}$$

with $\mu(B) = \sum_{i=1}^n \mathbb{P}[|\xi_i| \geq 1, S_i \in B]$ for any Borel set $B \subset \mathcal{X}$.

The second summand is estimated by the Brown-Xia inequality that was used for the proof of Corollary 2.3, as

$$d_2\Big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\Big) \leq \min\Big(1, \tfrac{1.65}{\sqrt{|\mu|}}, \tfrac{1.65}{\sqrt{|\lambda|}}\Big)\Big||\mu| - |\lambda|\Big| + \Big(1 - e^{-\min(|\mu|,|\lambda|)}\Big) d_W\Big(\tfrac{\mu}{|\mu|}, \tfrac{\lambda}{|\lambda|}\Big). \tag{3.6}$$

We have

$$|\mu| = n\Big(1 - e^{-\sigma_1 h}\Big), \quad |\lambda| = \sigma_1 n h\,,$$

and

$$\begin{aligned} \frac{\mu}{|\mu|}\Big([0,t] \times C\Big) &= \frac{1 - e^{-\sigma_1 h t}}{1 - e^{-\sigma_1 h}}\Big(\tfrac{\beta_1}{\sigma_1}\delta_0 + \tfrac{\alpha_1}{\sigma_1}\delta_1\Big)(C), \\ \frac{\lambda}{|\lambda|}\Big([0,t] \times C\Big) &= t\Big(\tfrac{\beta_1}{\sigma_1}\delta_0 + \tfrac{\alpha_1}{\sigma_1}\delta_1\Big)(C), \end{aligned}$$

for any $t \in [0,1]$ and $C \subset \{0,1\}$. The Wasserstein term in (3.6) can easily be estimated by noting that, since $\frac{\mu}{|\mu|}$ and $\frac{\lambda}{|\lambda|}$ are product measures that put both the same mass on $[0,1] \times \{0\}$, as well as on $[0,1] \times \{1\}$,

$$d_W\Big(\tfrac{\mu}{|\mu|}, \tfrac{\lambda}{|\lambda|}\Big) = d_W\Big(\tfrac{\mu}{|\mu|}(\cdot \times \{0,1\}), \tfrac{\lambda}{|\lambda|}(\cdot \times \{0,1\})\Big),$$

where the underlying distances are d_0 on the left hand side, and the Euclidean distance on the right hand side. Using then the fact that, for real-valued random variables X and Y, the Wasserstein distance between their distributions can be represented as

$$d_W(\mathcal{L}(X), \mathcal{L}(Y)) = \int_{-\infty}^{\infty} \Big|\mathbb{P}[X \leq x] - \mathbb{P}[Y \leq x]\Big|\, dx$$

(see e.g. Problem 1 in Section 11.8 of Dudley (1989)), yields

$$d_W\Big(\tfrac{\mu}{|\mu|}, \tfrac{\lambda}{|\lambda|}\Big) = \int_0^1 \left|\frac{1 - e^{-\sigma_1 h t}}{1 - e^{-\sigma_1 h}} - t\right| dt \leq \frac{(\sigma_1 h)^2}{4(1 - e^{-\sigma_1 h})}.$$

Thus, we obtain in Inequality (3.6)

$$d_2\Big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\Big) \le \min\Big(1, \tfrac{1.65}{\sqrt{\sigma_1 nh}}\Big)\frac{\sigma_1^2}{2}nh^2 + \frac{1-e^{-\sigma_1 nh}}{1-e^{-\sigma_1 h}}\frac{\sigma_1^2 h^2}{4} \le \frac{3\sigma_1^2}{4}nh^2. \tag{3.7}$$

The first summand in Inequality (3.5) is suited for the application of Theorem 2.1. For the terms in the upper bound of that theorem, we obtain

$$\begin{aligned}
p_i &= 1 - e^{-\sigma_1 h} \le \sigma_1 h, \text{ and} \\
p_i' &= \mathbb{P}[T_i + T_i' \le h] \\
&= \int_0^h \Big(\mathbb{P}[T_i' \le h - t \,|\, E_i = 1]\,\mathbb{P}[E_i = 1] \\
&\qquad + \mathbb{P}[T_i' \le h - t \,|\, E_i = 0]\,\mathbb{P}[E_i = 0]\Big)\sigma_1 e^{-\sigma_1 t}\, dt \\
&= \int_0^h \Big((1 - e^{-\sigma_2(h-t)})\alpha_1 + (1 - e^{-\sigma_0(h-t)})\beta_1\Big)e^{-\sigma_1 t}\, dt \\
&\le \frac{1}{2}(\alpha_1\sigma_2 + \beta_1\sigma_0)h^2.
\end{aligned} \tag{3.8}$$

Choosing an arbitrary $r \ge 0$, and setting $\Gamma_i^{\mathrm{s}} := \{j \in \mathbb{N} \setminus \{i\}; |z_j - z_i| \le r\}$, and $\Gamma_i^{\mathrm{w}} := \{j \in \mathbb{N}; |z_j - z_i| > r\}$ for the neighborhoods of strongly and weakly dependent processes, respectively, yields furthermore

$$\begin{aligned}
\mathbb{P}[|\xi_i| \ge 1, |\xi_j| \ge 1] &\le h\phi(h) \text{ and} \\
\mathbb{E}\Big|\mathbb{P}[|\xi_i| \ge 1 \,|\, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}}] - p_i\Big| &\le \chi_\alpha(h, r),
\end{aligned} \tag{3.9}$$

because ξ_j is a measurable function of $\zeta_j|_{[0,h]}$ for every $j \in \mathbb{N}$, and

$$\begin{aligned}
&p_i\,\mathbb{E}\,d_W\Big(\mathcal{L}(S_i \,|\, |\xi_i| \ge 1), \mathcal{L}(S_i \,|\, |\xi_i| \ge 1, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}})\Big) \\
&\quad \le p_i\,\mathbb{E}\,d_{TV}\Big(\mathcal{L}(S_i \,|\, |\xi_i| \ge 1), \mathcal{L}(S_i \,|\, |\xi_i| \ge 1, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}})\Big) \\
&\quad \le \mathbb{E} \sup_{B\in\mathcal{B}_{[0,1]}} \Big|\mathbb{P}[S_i \in B, |\xi_i| \ge 1 \,|\, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}}] - \mathbb{P}[S_i \in B, |\xi_i| \ge 1]\Big| \\
&\qquad + \mathbb{E} \sup_{B\in\mathcal{B}_{[0,1]}} \Big|\mathbb{P}[S_i \in B \,|\, |\xi_i| \ge 1, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}}]\Big(\mathbb{P}[|\xi_i| \ge 1] - \mathbb{P}[|\xi_i| \ge 1 \,|\, (\xi_j)_{j\in\Gamma_i^{\mathrm{w}}}]\Big)\Big| \\
&\quad \le \chi_\beta(h, r) + \chi_\alpha(h, r)
\end{aligned} \tag{3.10}$$

for the same reason. The expectations above are well-defined, because the suprema can all be replaced by suprema over countable sets, e.g. in lines 2 to 4 by the suprema over all finite unions of intervals with endpoints in $\mathbb{Q} \cap [0, 1]$ (which can be shown by using an elementary approximation property for finite measures). This fact also justifies the inequality between the supremum and the essential supremum used for the last line.

Combining the estimates from Inequalities (3.7) to (3.10), we obtain

$$\begin{aligned} d_2\Big(\mathcal{L}\Big(\sum_{i=1}^{n}\xi_i^{(h)}\Big), \mathrm{Po}(\lambda_{n,h})\Big) \leq &\Big(\tfrac{3\sigma_1^2}{4} + \tfrac{1}{2}(\alpha_1\sigma_2 + \beta_1\sigma_0) + M_2\sigma_1^2\Big)nh^2 \\ &+ M_2\kappa\sigma_1^2 r^d nh^2 + M_2\kappa r^d nh\phi(h) \\ &+ (M_1 + 2M_2)n\chi_\alpha(h,r) + M_2 n\chi_\beta(h,r), \end{aligned} \tag{3.11}$$

where

$$\begin{aligned} &M_1 = 1 \wedge \Big(\tfrac{1.65}{\sqrt{n(1-e^{-\sigma_1 h})}}\Big) \leq 1 \wedge \frac{5}{2\sqrt{\sigma_1 nh}} \quad \text{and} \\ &M_2 = 1 \wedge \Big[\tfrac{2}{n(1-e^{-\sigma_1 h})}\Big(1 + 2\log^+(n(1-e^{-\sigma_1 h})/2)\Big)\Big] \leq 1 \wedge \Big[\tfrac{4}{\sigma_1 nh}\Big(1 + 2\log^+\Big(\tfrac{\sigma_1 nh}{2}\Big)\Big)\Big] \end{aligned}$$

for $h \leq 1/\sigma_1$. Since $r \geq 0$ was arbitrary, this yields the required upper bound. □

Remark 3.3 (A sketch for the model with randomly positioned animals) It might be desirable to model also the positions of the animals as random. Then an upper bound can be calculated in a similar fashion as above, but with a few important differences: Typically, one wants to drop Condition (3.1) in this situation and work with $\varrho(\mathbb{B}(z_i, r)) - 1$ directly, where ϱ is now a point process and z_i its i-th point (in a suitable enumeration). Accordingly, the index sets Γ_i^{s} and Γ_i^{w} defined after Inequality (3.8) are now random. It is no problem to adapt Theorem 2.1 so that it comprises random index sets: Γ_i^{w} appears only via the random variable $\Xi_i^{\mathrm{w}} = \sum_{j\in\Gamma_i^{\mathrm{w}}} I_j\delta_{S_j}$ in the proof of Theorem 2.1, and it is easily seen that the few properties of Ξ_i^{w} we used remain unchanged for random Γ_i^{w}. The set Γ_i^{s} on the other hand, appears only as a summation set in the estimation of the first and the fourth summand in Inequality (2.2). There, the only difference is that the summation and the expectation cannot be exchanged. In total, we get the same bound in Theorem 2.1 for the case of random index sets as for the case of deterministic index sets, except for the third summand, which is, in the random case,

$$M_2\sum_{i=1}^{\infty}\mathbb{E}\Big(\sum_{j\in\Gamma_i^{\mathrm{s}}}\Big(p_i 1_{\{|\xi_j|\geq 1\}} + 1_{\{|\xi_i|\geq 1, |\xi_j|\geq 1\}}\Big)\Big). \tag{3.12}$$

Thus a very similar upper bound for the d_2-distance in Proposition 3.1 can be obtained if ϱ is random, but we have to replace Condition (3.2) by suitable conditions that control the term (3.12).

A Appendix: Sketch of Stein's method for Poisson process approximation

In 1972, Stein published his ingenious method for the normal approximation of dependent random variables. In Chen (1975) a variant of this method was developed for the Poisson case, which was generalized by Barbour (1988) to the Poisson process case. We only describe the most important ideas of this last version for the case of the Barbour-Brown distance. For a detailed presentation in a more general context see Barbour, Holst, and Janson (1992), Chapter 10.

Our goal is to bound

$$d_2\Big(\mathcal{L}(\Xi), \mathcal{L}(\eta)\Big) = \sup_{f \in \mathcal{F}_2} \Big|\mathbb{E}f(\Xi) - \mathbb{E}f(\eta)\Big|,$$

where Ξ is an arbitrary point process on $\mathcal{X}$, and η is a $\mathrm{Po}(\mu)$-process. The essential idea consists in writing

$$\Big|\mathbb{E}f(\Xi) - \mathbb{E}f(\eta)\Big| = \Big|\mathbb{E}\mathcal{A}h(\Xi)\Big| \tag{A.1}$$

for $f \in \mathcal{F}_2$ and then bounding the right hand side instead of the left hand side uniformly in f. In Equation (A.1), $\mathcal{A}$ is the generator of the spatial immigration-death process $(Z(t))_{t\geq 0}$ on $\mathcal{X}$ (that is, the state space is $\mathcal{N}$) with immigration measure μ and unit per capita death rate, a pure jump Markov process which has our approximating distribution $\mathrm{Po}(\mu)$ as equilibrium distribution. $\mathcal{A}$ is given by

$$\mathcal{A}\tilde{h}(\varrho) = \int_{\mathcal{X}} \Big[\tilde{h}(\varrho - \delta_s) - \tilde{h}(\varrho)\Big]\, \varrho(ds) + \int_{\mathcal{X}} \Big[\tilde{h}(\varrho + \delta_s) - \tilde{h}(\varrho)\Big]\, \mu(ds)\,, \quad \varrho \in \mathcal{N}, \tag{A.2}$$

for suitable functions $\tilde{h} : \mathcal{N} \to \mathbb{R}$. Furthermore $h = h_f$ in Equation (A.1), defined by

$$h(\varrho) = h_f(\varrho) = -\int_0^\infty \Big[\mathbb{E}\Big(f(Z(t)) \;\Big|\; Z(0) = \varrho\Big) - \mathbb{E}f(\eta)\Big]\, dt\,, \quad \varrho \in \mathcal{N}, \tag{A.3}$$

is the solution from Proposition 10.1.1 in Barbour, Holst, and Janson (1992) to the *Stein equation*

$$f(\varrho) - \mathbb{E}f(\eta) = \mathcal{A}h(\varrho) \quad \text{for } \varrho \in \mathcal{N}. \tag{A.4}$$

In Lemmas 10.2.3 and 10.2.5 of the same book bounds for the first and second differences of h are given as

$$\Delta_1 h := \sup_{\varrho \in \mathcal{N},\, s \in \mathcal{X}} \Big|h(\varrho + \delta_s) - h(\varrho)\Big| \leq 1 \wedge \tfrac{1.65}{\sqrt{|\mu|}} =: M_1, \tag{A.5}$$

and

$$\begin{aligned}\Delta_2 h :&= \sup_{\varrho \in \mathcal{N},\, s_1, s_2 \in \mathcal{X}} \Big|h(\varrho + \delta_{s_1} + \delta_{s_2}) - h(\varrho + \delta_{s_1}) - h(\varrho + \delta_{s_2}) + h(\varrho)\Big| \\ &\leq 1 \wedge \Big[\tfrac{2}{|\mu|}\Big(1 + 2\log^+(|\mu|/2)\Big)\Big] =: M_2. \end{aligned} \tag{A.6}$$

These quantities are usually needed to obtain the bounds for the term $|\mathbb{E}\mathcal{A}h(\Xi)|$ in Equation (A.1). Brown, Weinberg, and Xia (2000) present a way of bounding quantities similar to $\Delta_2 h$ in such a way that the $\log^+$-term above can sometimes be disposed of. Because the logarithmic term will be negligible for the purposes of this paper, we avoid doing this more involved considerations.

With the above ingredients it often turns out (like in the main proof of this article) that the right hand side of Equation (A.1) is surprisingly easy to bound.

Acknowledgements

I would like to thank Andrew Barbour for contributing valuable comments and suggestions to this article. Part of the work was carried out while I was visiting the National University of Singapore, and I am very grateful to Louis Chen and the Institute for Mathematical Sciences for their hospitality and support.

References

Arratia, R., Goldstein, L., Gordon, L., 1989. Two moments suffice for Poisson approximations: the Chen-Stein method. Ann. Probab. 17, 9–25.

Banys, R., 1980. On superpositions of random measures and point processes, in: Mathematical statistics and probability theory (Proc. Sixth Internat. Conf., Wisła, 1978), Lecture Notes in Statist. 2, Springer, New York, 26–37.

Barbour, A.D., 1988. Stein's method and Poisson process convergence, J. Appl. Probab. Special Vol. 25A, 175–184.

Barbour, A.D., Brown T.C., 1992. Stein's method and point process approximation, Stochastic Process. Appl. 43, 9–31.

Barbour, A.D., Holst, L., Janson, S., 1992. Poisson approximation. Oxford University Press, Oxford.

Barbour, A.D., Månsson, M., 2002. Compound Poisson process approximation, Ann. Probab. 30, 1492–1537.

Barbour, A.D., Novak, S.Y., Xia, A., 2002. Compound Poisson approximation for the distribution of extremes, Adv. in Appl. Probab. 34, 223–240.

Brown, T.C., Weinberg, G.V., Xia, A., 2000. Removing logarithms from Poisson process error bounds, Stochastic Process. Appl. 87, 149–165.

Brown, T.C., Xia, A., 1995. On metrics in point process approximation, Stochastics Stochastics Rep. 52, 247–263.

Chen, L.H.Y., 1975. Poisson approximation for dependent trials, Ann. Probability 3, 534–545.

Chen, L.H.Y., Xia, A., 2004. Stein's method, Palm theory and Poisson process approximation, Ann. Probab. 32, 2545–2569.

Doukhan, P., 1994. Mixing. Properties and examples. Lecture Notes in Statistics, 85. Springer, New York.

Dudley, R.M., 1989. Real analysis and probability. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Feller, W. 1968. An introduction to probability theory and its applications, vol. I. Third edition. Wiley, New York.

Goldman, J.R., 1967. Stochastic point processes: limit theorems, Ann. Math. Statist. 38, 771–779.

Grigelionis, B., 1963. On the convergence of sums of random step processes to a Poisson process, Theory Probab. Appl. 8, 177–182.

Jagers, P., 1972. On the weak convergence of superpositions of point processes, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 22, 1–7.

Kallenberg, O., 1986. Random measures, 4th ed. Akademie-Verlag, Berlin; Academic Press, London.

Kallenberg, O., 2002. Foundations of modern probability, 2nd ed. Springer, New York.

Khinchin, A.Y., 1955. Mathematical methods in the theory of queueing (in Russian). Trudy Mat. Inst. Steklov v. 49, Izdat. Akad. Nauk SSSR, Moscow.
Translated: Griffin, London, 1960.

Matthes, K., Kerstan, J., Mecke, J., 1978. Infinitely divisible point processes. Wiley, Chichester.

Palm, C., 1943. Intensitätsschwankungen im Fernsprechverkehr, Ericsson Technics no. 44.

Schuhmacher, D., 2005a. Upper bounds for spatial point process approximations, Ann. Appl. Probab. 15, 615–51.

Schuhmacher, D., 2005b. Distance estimates for Poisson process approximations of dependent thinnings, Electron. J. Probab. 10, 165–201.

Serfozo, R., 1984. Thinning of cluster processes: convergence of sums of thinned point processes, Math. Oper. Res. 9, 522–33.

Stein, C., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, in: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2. Univ. California Press, Berkeley, CA, 583–602.