Electronic Journal of Probability

Vol. 10 (2005), Paper no. 5, pages 165-201.

Journal URL
http://www.math.washington.edu/~ejpecp/

Distance Estimates for Poisson Process Approximations of Dependent Thinnings

Dominic Schuhmacher*
Institut für Mathematik, Universität Zürich
Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
schumi@amath.unizh.ch

Abstract: It is well known, that under certain conditions, gradual thinning of a point process on $\mathbb{R}_+^d$, accompanied by a contraction of space to compensate for the thinning, leads in the weak limit to a Cox process. In this article, we apply discretization and a result based on Stein's method to give estimates of the Barbour-Brown distance d_2 between the distribution of a thinned point process and an approximating Poisson process, and evaluate the estimates in concrete examples. We work in terms of two, somewhat different, thinning models. The main model is based on the usual thinning notion of deleting points independently according to probabilities supplied by a random field. In Section 4, however, we use an alternative thinning model, which can be more straightforward to apply if the thinning is determined by point interactions.

Keywords and phrases: Point processes, probability metrics, Barbour-Brown distance, Wasserstein distance.

AMS Subject Classifications (2000): Primary 60G55, 60E99; secondary 60D05.

Submitted to EJP on June 14, 2004. Final version accepted on December 29, 2004.

*Partially supported by Schweizerischer Nationalfonds, Project Nos. 20-61753.00 and 20-67909.02.

1 Introduction

Thinning is one of the fundamental operations to construct a new point process from a given one. Consider a point process ξ on $\mathbb{R}_+^d$, and a $[0,1]$-valued measurable random field π on $\mathbb{R}_+^d$. A π-thinning ξ_π of ξ can then be obtained in the following way: for any realizations $\xi(\omega)$ (point measure on $\mathbb{R}_+^d$) and $\pi(\omega,\cdot)$ (function $\mathbb{R}_+^d \to [0,1]$), look at each point s_i of $\xi(\omega)$ in turn, and retain it with probability $\pi(\omega,s_i)$, or delete it with probability $1-\pi(\omega,s_i)$, independently of any retention/deletion decisions of other points. Note that multiple points $s_{i_1} = \ldots = s_{i_k}$ at the same location are retained independently of one another with equal probabilities. Regard the points left over by this procedure as a realization of the thinned point process ξ_π (for a more formal definition see Section 2). We will usually refer to ξ as "the original process", and to π as "the retention field".

Consider now a sequence $(\pi_n)_{n\in\mathbb{N}}$ of retention fields, such that $\sup_{s\in\mathbb{R}_+^d} \pi_n(s) \xrightarrow{\mathcal{D}} 0$, corresponding to the idea of ξ being gradually thinned away. To compensate for this effect, we contract the Euclidean space by $\kappa_n : \mathbb{R}_+^d \to \mathbb{R}_+^d$, $x \mapsto (1/n)x$. Then, Theorem 1.A is a standard result (see e.g. Daley and Vere-Jones (1988), Theorem 9.3.III). Convergence in distribution of random measures, and in particular, of point processes, is defined via the convergence of expectations of bounded continuous functions, where continuity is in terms of the vague topology on the corresponding space of (all respectively only $\mathbb{Z}_+ \cup \{\infty\}$-valued) boundedly finite measures (for details see Kallenberg (1986), Section 4.1).

Theorem 1.A. We obtain convergence in distribution of the thinned and contracted process $\xi_\pi\kappa_n^{-1}$ to a point process η if and only if, with $\Lambda_n(A) := \int_{\kappa_n^{-1}(A)} \pi_n(s)\,\xi(ds)$ for every measurable set $A \subset \mathbb{R}_+^d$, we have

$$\Lambda_n \xrightarrow{\mathcal{D}} \Lambda \quad \text{for } n \to \infty \tag{1.1}$$

for some random measure Λ on $\mathbb{R}_+^d$. In this case $\eta \sim \mathrm{Cox}(\Lambda)$, i.e. η is a Cox process with directing measure Λ.

The main goal of the present article is to examine the rate of convergence in the above theorem. For that purpose, we use the Barbour-Brown distance d_2 between the distributions of point processes, which is a very natural choice (see Section 2 for the definition and some elementary properties). Due to the method of proof that we use, we will always assume that a mixing condition holds for the random measures Λ_n (see Assumption 2 in Section 3). Since Condition (1.1) can be interpreted as the statement of a weak ergodic theorem, it is natural, in view of the usual chain "*mixing* implies *ergodic* implies *constant limit in the ergodic theorem*", that we get a deterministic limiting measure $\Lambda = \lambda$, and hence even a Poisson process as the limiting process η of the contracted thinning. These heuristics can easily be made rigorous if ξ is stationary, and $\pi_n = (1/n^d)\tilde{\pi}$ for a stationary random field $\tilde{\pi}$.

The method of our proof is a combination of discretization and an application of the Barbour-Brown theorem for distance estimates for discrete Poisson process approximations where the approximated process has a local dependence property. The same method has been successfully used in Schuhmacher (2005), and in fact, by applying Theorem 2.A. from that paper to a suitable randomization of ξ, we obtain directly upper bounds for the distance we seek here. While these bounds are quite good in special cases (e.g. if our retention field is deterministic, and even more so if it is constant), a direct application of the above method yields considerably better and more intuitive results in the general case.

The thinning of point processes was first studied in Rényi (1957): a renewal process on $\mathbb{R}_+$ was subjected to an independent thinning with constant retention probability p, and a Poisson limit theorem was obtained for $p \to 0$ upon a change of time scale by a factor $1/p$. There have been many generalizations within the theory of point processes on the real line since then, with some of the most comprehensive found in Jagers and Lindvall (1974), Serfozo (1977), and Böker and Serfozo (1983).

Also in Kallenberg (1975) (alternatively, see Kallenberg (1986), Section 8.3), independent thinnings with constant retention probability p were considered, but this time the processes to be thinned were arbitrary point processes on general locally compact, second countable, Hausdorff spaces. Necessary and sufficent conditions were derived for the convergence of thinnings of increasingly "dense" point processes to a Cox limit. This result was generalized in Brown (1979) to position dependent, random retention probabilities, which yielded, up to some negligible details in the setting, exactly the statement of Theorem 9.3.III in Daley and Vere-Jones (1988), from which Theorem 1.A is a direct consequence.

A result regarding distance estimates in thinning theorems may also be found in Daley and Vere-Jones (1988). In Proposition 9.3.IV, the authors give a quite abstract upper bound for the total variation distance between the distributions of the point counts $\xi_{\pi_n}\kappa_n^{-1}(A)$ and $\eta(A)$ for any bounded Borel set A. By a rather similar argument, it is possible to obtain a corresponding upper bound for the d_2-distance between the distributions of the restricted point processes $\xi_{\pi_n}\kappa_n^{-1}|_K$ and $\eta|_K$, where $K \subset \mathbb{R}_+^d$ is an arbitrary compact set. Write d_W for the Wasserstein distance on $\mathbb{R}_+^d$ with respect to the Euclidean distance truncated at one (which is denoted by d_0 in this paper), that is

$$d_W(P, Q) = \sup_f \Bigl| \int f\, dP - \int f\, dQ \Bigr|$$

for probability measures P and Q on $\mathbb{R}_+^d$, where the supremum ranges over all functions $f : \mathbb{R}_+^d \to \mathbb{R}$ with $|f(s_1) - f(s_2)| \leq |s_1 - s_2| \wedge 1$ (see e.g. Barbour, Holst and Janson (1992), Appendix A.1 for the general definition and results). The proof of the following proposition can be found in Appendix A.2.

Proposition 1.B. In the notation of Theorem 1.A we have for compact $K \subset \mathbb{R}_+^d$

$$\begin{aligned} &d_2\bigl(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_K), \mathrm{Cox}(\Lambda|_K)\bigr) \\ &\qquad \leq \mathbb{E}\Bigl(8.45 \cdot \sup_{s \in \kappa_n^{-1}(K)} \pi_n(s) \;+\; \min\bigl(1, \Lambda_n(K), \Lambda(K)\bigr)\, d_W\bigl(\tfrac{\Lambda_n|_K(\cdot)}{\Lambda_n(K)}, \tfrac{\Lambda|_K(\cdot)}{\Lambda(K)}\bigr) \\ &\qquad\qquad + 1 \wedge \Bigl(\min\bigl(1, \tfrac{1.65}{\sqrt{\Lambda_n(K)}}, \tfrac{1.65}{\sqrt{\Lambda(K)}}\bigr)\, \bigl|\Lambda_n(K) - \Lambda(K)\bigr|\Bigr)\Bigr), \end{aligned} \tag{1.2}$$

where the second summand is defined to be zero if either $\Lambda_n(K)$ or $\Lambda(K)$ is zero.

While such an upper bound is very nice from a theoretical point of view, because it is of a comparatively simple form, and because it can be shown that it goes to zero under the general Condition (1.1) if K satisfies $\mathbb{P}[\Lambda(\partial K) > 0] = 0$, it is usually not so easy to calculate. In what follows, we are interested in more explicit upper bounds for the thinning approximation, i.e. upper bounds that can be calculated directly in terms of certain characteristic quantities of the point process ξ and the retention field π. Since, in the (still quite general) case where ξ and π satisfy mixing conditions, and the approximating process is Poisson, we have available well-developed tools from the field of Stein's method for point process approximation, we prefer to make use of these, rather than trying to reduce the right hand side of Inequality (1.2) to more fundamental characteristics of ξ and π.

We start out in Section 2, by providing the necessary definitions and notation along with some technical background. Section 3 contains the main results, namely upper bounds for the d_2-distance between $\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_K)$ and a Poisson process law under different conditions, as well as an application of these results. Finally, in Section 4, a slightly different notion of a thinning, called Q-thinning, is introduced. A corresponding upper bound is given, and the benefit of the new definition is demonstrated in another application.

The difference between the two thinning models in Section 1–3 and in Section 4, is conceptual rather than a difference in terms of the modeled objects (see Remarks 4.B and 4.E for details on how the resulting thinnings differ). Under π-thinning, points are more or less likely to be deleted according to the conditions they encounter in a random environment (which itself may respond to the point configuration); under Q-thinning, points are more or less likely to be deleted according directly to the interactions among themselves. This difference is nicely illustrated by the two applications in Subsection 3.3 and Subsection 4.2. In both situations, we start with a point process ξ on $\mathbb{R}_+^d$ having "reasonable" first and second factorial moment measures, and obeying an appropriate mixing condition. We have a "basic" thinning effect given by the constant retention probability $q_0^{(n)}$, and an additional "characteristic" thinning effect for each of the two examples. The basic and the characteristic effect are combined, and $\mathbb{R}_+^d$ is contracted by a factor $1/n$ to obtain a point process that is compared to a Poisson process.

In the first example (Subsection 3.3), we consider a random environment given by a union Ξ of balls whose centers form a stationary Poisson process on $\mathbb{R}^d$, and whose radii are i.i.d. and bounded with L_d-norm r_n. The characteristic thinning effect is then provided by deleting all the points that are covered by Ξ (we illustrate this situation with the image of stars that are covered by clouds). In the second example (Subsection 4.2), we give point interactions by assigning i.i.d. real-valued marks to the points. The characteristic thinning effect is then provided by deleting all the points whose distance to some other point having a larger mark is at most r_n (we illustrate this situation with the scenario of competition within a plant population).

Roughly speaking, Poisson approximation is good in both examples if the retention probabilities $q_0^{(n)}$ become appropriately small, or the ranges r_n become appropriately large.

It should be noted, that the illustrations of the examples as "visibility of stars" and "plant competition", respectively, may well provide inspiration for modeling similar situations, but are not meant to be serious modeling attempts in themselves.

We end this section by giving an indication of the type of results that are obtained in Section 3. The following proposition covers the important special case when $\pi_n = p_n$ is non-random, which follows directly from either Theorem 3.C or Theorem 3.F. The situation corresponds to an independent thinning of the points of ξ with deterministic retention probabilities, each of which may depend on the location of the point it is attached to.

Proposition 1.C (Non-random retention field). Let ξ be a point process on $\mathbb{R}_+^d$ which has an expectation measure μ_1 that has a bounded density with respect to Lebesgue measure, and a second factorial moment measure μ_2 that is bounded on the set of all unit cubes in $\mathbb{R}_+^{2d}$. Let $(p_n)_{n\in\mathbb{N}}$ be a sequence of functions $\mathbb{R}_+^d \to [0,1]$, and let $\bar{p}_n := \sup_{s\in\mathbb{R}_+^d} p_n(s)$. Suppose that "long range" covariances in ξ are controlled by a decreasing function $\breve{\beta} : \mathbb{R}_+ \to \mathbb{R}_+$, such that for every open cube $A_{\mathrm{int}} = (a, a+h\mathbf{1})$ of side length $h \in (0,1]$ and with minimal corner in $a \in \mathbb{R}_+^d$, and every surrounding set $A_{\mathrm{ext}}^{(t)} := \mathbb{R}_+^d \setminus [a - t\mathbf{1}, a + (t+h)\mathbf{1}]$ with $t \in \mathbb{R}_+$, we have

$$\sup_{D,Z} \bigl|\mathrm{cov}(Z, 1_D)\bigr| \leq \breve{\beta}(t),$$

where the supremum ranges over all $D \in \sigma(\xi|_{A_{\mathrm{ext}}^{(t)}})$ and $Z \in \mathcal{L}_2(\sigma(\xi|_{A_{\mathrm{int}}}))$ with $0 \leq Z \leq \xi(A_{\mathrm{int}})/|A_{\mathrm{int}}|$. Set furthermore $J := [0,1]^d$. Then we obtain for arbitrarily chosen $m := m(n) \in \mathbb{N}$

$$d_2\bigl(\mathcal{L}(\xi_{p_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\bigr) = O\bigl(n^d\bar{p}_n\bigl(m^d\bar{p}_n \vee \breve{\beta}(m)\bigr)\bigr) \quad \text{for } n \to \infty$$

with $\nu_n(\cdot) := \int_{\cdot} p_n(s)\ \mu_1(ds)$. In the most natural case, where $\bar{p}_n = O(1/n^d)$, this implies

$$d_2\bigl(\mathcal{L}(\xi_{p_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\bigr) = O\bigl(n^{-sd/(s+1)}\bigr)$$

if $\breve{\beta}(t) = O(t^{-sd})$ as $t \to \infty$ for some fixed $s > 0$.

Note that we maintain a great deal of flexibility, and avoid evaluating awkward distances between general probability measures, by approximating the thinning with a Poisson process whose intensity measure depends on n. If we must have a fixed approximating Poisson process, we get by Inequality (A.3) from Appendix A.2 the additional term of

$$\begin{aligned} d_2\bigl(\mathrm{Po}(\nu_n\kappa_n^{-1}|_J), \mathrm{Po}(\lambda|_J)\bigr) &\leq \min\bigl(1, \nu_n\kappa_n^{-1}(J), \lambda(J)\bigr) d_W\Bigl(\tfrac{\nu_n\kappa_n^{-1}|_J(\cdot)}{\nu_n\kappa_n^{-1}(J)}, \tfrac{\lambda|_J(\cdot)}{\lambda(J)}\Bigr) \\ &\quad + \min\Bigl(1, \tfrac{1.65}{\sqrt{\nu_n\kappa_n^{-1}(J)}}, \tfrac{1.65}{\sqrt{\lambda(J)}}\Bigr)\bigl|\nu_n\kappa_n^{-1}(J) - \lambda(J)\bigr|, \end{aligned} \tag{1.3}$$

which is, in most cases, still much more convenient to estimate than the two corresponding terms in (1.2), inasmuch as the quantities appearing in (1.3) are no longer random.

2 Preliminaries

Assume that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete in order to avoid measurability problems. Denote by $\mathcal{B}_+^d$ the Borel σ-field on $\mathbb{R}_+^d$, and by $\mathcal{B}_A$ for any set $A \subset \mathbb{R}_+^d$ the trace σ-field $\mathcal{B}_+^d|_A$. Furthermore, write $\mathcal{M}$ for the space of boundedly finite measures on $\mathbb{R}_+^d$, that is measures μ with $\mu(B) < \infty$ for any bounded $B \in \mathcal{B}_+^d$, and equip it with the usual topology (i.e. the vague one, see Kallenberg (1986), Section 15.7) and the smallest σ-field $\mathfrak{M}$ that renders the masses of bounded Borel sets measurable (see Kallenberg (1986), Section 1.1). Do the same for the subspace $\mathcal{N} \subset \mathcal{M}$ of boundedly finite point measures, and denote the corresponding σ-field by $\mathfrak{N}$. A random measure is a random element of $\mathcal{M}$, and a point process a random element of $\mathcal{N}$. By $\mathrm{Po}(\lambda)$ we denote the distribution of the Poisson process with intensity measure λ if $\lambda \in \mathcal{M}$, and the Poisson distribution with parameter λ if λ is a positive real number.

In view of the Barbour-Brown distance d_2 that we use, we always restrict our point processes to the unit cube $J = [0,1]^d$. Write furthermore $J_n := \kappa_n^{-1}(J) = [0,n]^d$, $\widetilde{J}_n := [0,n)^d$, and $C_{\boldsymbol{k}} := \prod_{i=1}^d [k_i - 1, k_i)$ for all $\boldsymbol{k} = (k_1, \ldots, k_d) \in \{1, 2, \ldots, n\}^d$ for the unit hypercubes that make up $\widetilde{J}_n$. We make use of a simplified multi-index notation by denoting properties of all the individual components as though they were properties of the whole multi-index, for example writing $\widetilde{J}_n = \bigcup_{\boldsymbol{k}=1}^n C_{\boldsymbol{k}}$ instead of $\widetilde{J}_n = \bigcup_{\boldsymbol{k}:k_1,\ldots k_d=1}^n C_{\boldsymbol{k}}$. Furthermore, we use the distance between multi-indices that is defined by $|\boldsymbol{l}-\boldsymbol{k}| := \max_{1\leq i\leq d}|l_i - k_i|$. As a last convention, all equations and inequalities between random variables in this article hold almost surely, if it is not explicitly stated that they hold pointwise.

In what follows, let ξ always be a point process on $\mathbb{R}_+^d$, and $\pi := (\pi(\cdot, s); s \in \mathbb{R}_+^d)$ a $[0,1]$-valued random field on $\mathbb{R}_+^d$ that satisfies the following measurability condition: for any bounded rectangle R that is open in $\mathbb{R}_+^d$, the mapping $\pi_R^* : \Omega \times R \to [0,1]$, $(\omega, s) \mapsto \pi(\omega, s)$ is $\sigma(\pi|_R) \otimes \mathcal{B}_R$-measurable (we say in this case that π *is locally evaluable*). Assuming this technical condition simplifies part of the notation and the arguments in Section 3 considerably. However, if one accepts a more involved presentation, in particular larger and more complicated σ-fields in Assumption 2, and a stronger independence property for Theorem 3.F, all that is needed to prove the corresponding theorems is the measurability of π as a mapping $\Omega \times \mathbb{R}_+^d \to [0,1]$. The latter is necessary to ensure that for any random point S in $\mathbb{R}_+^d$, $\pi(S)$ is a random variable. Note that local evaluability is satisfied for many desirable random fields, as for example, for random fields which are pathwise continuous or for indicators of (fully) separable random closed sets (see Appendix A.3 for a more detailed discussion of locally evaluable random fields with proofs).

The π-thinning of ξ can now be defined as follows.

Definition (Thinning). First, assume that $\xi = \sigma = \sum_{i=1}^v \delta_{s_i}$ and $\pi = p$ are non-random, where $v \in \overline{\mathbb{Z}}_+ := \{0, 1, 2, \ldots\} \cup \{\infty\}$, $s_i \in \mathbb{R}_+^d$, and p is a function $\mathbb{R}_+^d \to [0,1]$. Then, a π-thinning of ξ is defined as $\xi_\pi = \sum_{i=1}^v X_i \delta_{s_i}$, where the X_i are independent indicator random variables with expectations $p(s_i)$, respectively. Under these circumstances, ξ_π has a distribution $P_{(\sigma,p)}$ that does not depend on the chosen enumeration of σ. We obtain the general π-thinning from this by randomization, that is by the condition $\mathbb{P}[\xi_\pi \in \cdot \,|\, \xi, \pi] = P_{(\xi,\pi)}$ (under our conditions on π it is straightforward to see that $P_{(\xi,\pi)}$ is a $\sigma(\xi,\pi)$-measurable family in the sense that $P_{(\xi,\pi)}(D)$ is $\sigma(\xi,\pi)$-measurable for every $D \in \mathfrak{N}$). Note that the distribution of ξ_π is uniquely determined by this procedure.

This definition of thinning can be found in Kallenberg (1986) for non-random π, and was generalized in Serfozo (1984) to random π. The definition of Brown (1979), and the less formal definitions of Stoyan, Kendall and Mecke (1987) and Daley and Vere-Jones (1988) also yield the same distribution for ξ_π.

The following remark simplifies the presentation of the proofs.

Remark 2.A (Numberings of points / definition of retention decisions). Given a countable partition $(B_j)_{j\in\mathbb{N}}$ of $\mathbb{R}_+^d$ into bounded measurable sets, there is always a representation of ξ as $\sum_{i=1}^{\xi(\mathbb{R}_+^d)} \delta_{S_i}$ with $\mathbb{R}_+^d$-valued, $\sigma(\xi)$-measurable random elements S_i, such that $S_1, \dots, S_{\xi(B_1)} \in B_1$, $S_{\xi(B_1)+1}, \dots, S_{\xi(B_1)+\xi(B_2)} \in B_2$, and so on (see e.g. the proof of Lemma 2.3 in Kallenberg (1986)). We will make tacit use of this fact in connection with the thinning definition on various occasions. For example, for a point process ξ and a bounded Borel set A, we may write $\xi_\pi(A) = \sum_{i=1}^{\xi(A)} X_i$, and hereby imply that we define $\sigma(\xi)$-measurable "point random elements" S_i for ξ in such a way that the first $\xi(A)$ points $S_1, \dots, S_{\xi(A)}$ always lie in A, and all other points in A^c, and define "retention decisions" X_i which, conditional on ξ and π_n, are independent with expectations $\pi_n(S_i)$, respectively.

We measure distances between distributions of point processes on a compact subset $K \subset \mathbb{R}^d$ by means of the Barbour-Brown distance d_2, a variant of a Wasserstein distance, which has proved to be a useful metric between distributions of point processes in many examples. It can be defined in the following way: Let d_0 be the minimum of the usual Euclidean distance on $\mathbb{R}^d$ and 1. Denote by $\mathcal{N}(K)$ the set of all finite point measures on K, set $\mathcal{F}_1 := \{k : K \to \mathbb{R};\ |k(s_1) - k(s_2)| \leq d_0(s_1, s_2)\}$, and define the d_1-*distance* (w.r.t. d_0) between $\varrho_1, \varrho_2 \in \mathcal{N}(K)$ as

$$d_1(\varrho_1, \varrho_2) := \begin{cases} 1 & \text{if } |\varrho_1| \neq |\varrho_2|, \\ \frac{1}{|\varrho_1|} \sup_{k\in\mathcal{F}_1} \left| \int k \, d\varrho_1 - \int k \, d\varrho_2 \right| & \text{if } |\varrho_1| = |\varrho_2| > 0, \\ 0 & \text{if } |\varrho_1| = |\varrho_2| = 0, \end{cases} \tag{2.1}$$

where $|\varrho_i| := \varrho_i(K) < \infty$. It can be seen that $(\mathcal{N}(K), d_1)$ is a complete, separable metric space and that d_1 is bounded by 1. Furthermore, for $\varrho_1 = \sum_{i=1}^v \delta_{s_{1,i}}$, $\varrho_2 = \sum_{i=1}^v \delta_{s_{2,i}}$ with $v \geq 1$, the Kantorovich-Rubinstein theorem (see Dudley (1989), Section 11.8) yields that

$$d_1(\varrho_1, \varrho_2) = \min_{\tau\in\Sigma_v} \left[\frac{1}{v} \sum_{i=1}^v d_0(s_{1,i}, s_{2,\tau(i)}) \right], \tag{2.2}$$

where Σ_v is the set of permutations of $\{1, 2, \dots, v\}$. Now let $\mathcal{F}_2 := \{f : \mathcal{N}(K) \to$

$\mathbb{R}$; $|f(\varrho_1) - f(\varrho_2)| \leq d_1(\varrho_1, \varrho_2)\}$, and define the d_2-*distance* (w.r.t. d_0) between probability measures P and Q on $\mathcal{N}(K)$ (distributions of point processes on K) as

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

By the Kantorovich-Rubinstein theorem, one obtains that

$$d_2(P, Q) = \min_{\substack{\xi_1 \sim P \\ \xi_2 \sim Q}} \mathbb{E} d_1(\xi_1, \xi_2). \tag{2.3}$$

Furthermore, because of the bound on the d_1-distance, the d_2-distance can also be interpreted as a variant of a bounded Wasserstein distance. Hence Theorem 11.3.3 in Dudley (1989) yields that d_2 metrizes the weak convergence of point process distributions. In other words, for point processes $\xi, \xi_1, \xi_2, \ldots$ on K, we have

$$\xi_n \xrightarrow{\mathcal{D}} \xi \quad \text{iff} \quad d_2\big(\mathcal{L}(\xi_n), \mathcal{L}(\xi)\big) \longrightarrow 0, \tag{2.4}$$

where the convergence in distribution for point processes is defined in terms of the vague topology (see also the explanation before Theorem 1.A), which, since K is compact, is the same as the weak topology. The fact that is crucial here is that, for d_0 as defined, the topology generated by the metric d_1 on $\mathcal{N}(K)$ is equal to the weak topology.

For further information on the d_2-distance see Barbour, Holst and Janson (1992), Section 10.2. For applications of d_2 upper bounds see also Schuhmacher (2005), Section 3.

3 The main results

For this whole section, let ξ be a point process on $\mathbb{R}_+^d$, and, for each $n \in \mathbb{N}$, let $\pi_n := (\pi_n(s)\,;\, s \in \mathbb{R}_+^d) := (\pi_n(\cdot, s)\,;\, s \in \mathbb{R}_+^d)$ be a $[0,1]$-valued, locally evaluable random field on $\mathbb{R}_+^d$.

3.1 Results

Recall that the expectation measure μ_1 of ξ is given by $\mu_1(B) := \mathbb{E}\big(\xi(B)\big)$ for any $B \in \mathcal{B}_+^d$, and that the second factorial moment measure of ξ is the measure on $(\mathcal{B}_+^d)^2$ defined by $\mu_2(B_1^2) := \mathbb{E}\big(\xi(B_1)(\xi(B_1) - 1)\big)$ and $\mu_2(B_1 \times B_2) := \mathbb{E}\big(\xi(B_1)\xi(B_2)\big)$ for any $B_1, B_2 \in \mathcal{B}_+^d$ that are disjoint. The following two assumptions, in one or the other form, are used several times in this article.

Assumption 1 (Control of moment measures).

a) ξ has an expectation measure $\mu_1 \ll Leb^d$ with bounded density $h_1 : \mathbb{R}_+^d \to \mathbb{R}_+$, and $\bar{h}_1 := \|h_1\|_\infty$;

b) ξ has a second factorial moment measure μ_2 which is bounded on the set of unit cubes in $\mathbb{R}_+^{2d}$, i.e. there is $\bar{h}_2 \geq 0$ such that $\mu_2(C) \leq \bar{h}_2$ for any unit cube $C \subset \mathbb{R}_+^{2d}$.

Assumption 2 (Mixing property). For each $n \in \mathbb{N}$, let $\breve{\beta}_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that for every cube $A_{\mathrm{int}} = (a, a+h\mathbf{1})$ of side length $h \in (0,1]$ and with minimal corner in $a \in \mathbb{R}_+^d$, and every surrounding set $A_{\mathrm{ext}}^{(t)} := \mathbb{R}_+^d \setminus [a - t\mathbf{1}, a + (t+h)\mathbf{1}]$ with $t \in \mathbb{R}_+$ we have that

$$\sup_{\substack{D \in \mathcal{F}_{\mathrm{ext}}^{(n,t)} \\ Z \in \mathcal{L}_2(\mathcal{F}_{\mathrm{int}}^{(n)}),\, 0 \le Z \le \Pi}} \bigl|\mathrm{cov}\bigl(Z, 1_D\bigr)\bigr| \le \breve{\beta}_n(t),$$

where

$$\mathcal{F}_{\mathrm{int}}^{(n)} := \sigma\bigl(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}}\bigr), \quad \mathcal{F}_{\mathrm{ext}}^{(n,t)} := \sigma\bigl(\xi|_{A_{\mathrm{ext}}^{(t)}}, \pi_n|_{A_{\mathrm{ext}}^{(t)}}\bigr),$$

and

$$\Pi := \Bigl(\xi(A_{\mathrm{int}}) \sup_{s \in A_{\mathrm{int}}} \pi_n(s)\Bigr) \Big/ \mathbb{E}\Bigl(\xi(A_{\mathrm{int}}) \sup_{s \in A_{\mathrm{int}}} \pi_n(s)\Bigr)$$

if $\mathbb{E}\bigl(\xi(A_{\mathrm{int}}) \sup_{s \in A_{\mathrm{int}}} \pi_n(s)\bigr) > 0$, and $\Pi :\equiv 0$ otherwise.

Remark 3.A. Since π_n is a measurable random field and $\mathcal{F}$ is complete, it can be shown by a standard argument involving analytic sets (using the Lusin-Choquet-Meyer theorem from Kallenberg (2002), Appendix A1) that $\sup_{s \in A} \pi_n(s)$ is a random variable for any $A \in \mathcal{B}_+^d$.

We now state the main theorem, first in its most general form, and then in weaker but less involved versions. In all the results, we use the notation $O\bigl(f_1(n), \ldots, f_j(n)\bigr)$ as short hand for $O\bigl(\max\{f_1(n), \ldots, f_j(n)\}\bigr)$. Quantitative versions of the upper bounds can be found in the proofs in Subsection 3.2.

Theorem 3.B. Suppose that the point process ξ and the sequence $(\pi_n)_{n \in \mathbb{N}}$ of retention fields satisfy Assumptions 1 and 2 above. Set

$$w_1 := \sup_{1 \le \boldsymbol{k} \le n} \mathbb{E}\Bigl(\xi(C_{\boldsymbol{k}}) \sup_{s \in C_{\boldsymbol{k}}} \pi_n(s)\Bigr),$$

$$w_{[2]} := \sup_{1 \le \boldsymbol{k}, \boldsymbol{l} \le n} \mathbb{E}\Bigl(\xi(C_{\boldsymbol{k}})\bigl(\xi(C_{\boldsymbol{l}}) - \delta_{\boldsymbol{k}\boldsymbol{l}}\bigr) \sup_{s \in C_{\boldsymbol{k}}} \pi_n(s) \sup_{s \in C_{\boldsymbol{l}}} \pi_n(s)\Bigr),$$

$$\underline{w}_1 := \inf_{1 \le \boldsymbol{k} \le n} \mathbb{E}\Bigl(1_{\{\xi(C_{\boldsymbol{k}}) \ge 1\}} \inf_{s \in C_{\boldsymbol{k}}} \pi_n(s)\Bigr),$$

and let $\nu_n(\cdot) := \mathbb{E}\bigl(\int_{\cdot} \pi_n(s)\, \xi(ds)\bigr)$, which is the expectation measure of ξ_{π_n}.
Then we obtain for arbitrary $m := m(n) \in \mathbb{N}$

$$\begin{aligned} &d_2\bigl(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\bigr) \\ &\quad = O\Bigl(n^d w_{[2]}\,,\ m^d\bigl(n^d \wedge \tfrac{\log^{\uparrow}(n^d w_1)}{\underline{w}_1}\bigr)(w_1^2 \vee w_{[2]})\,,\ \sqrt{n^d}\bigl(\sqrt{n^d} \wedge \tfrac{1}{\sqrt{\underline{w}_1}}\bigr) w_1 \breve{\beta}_n(m)\Bigr) \end{aligned}$$

for $n \to \infty$

with $\log^{\uparrow}(x) := 1 + \bigl(\log(x) \vee 0\bigr)$ for $x \ge 0$.

The next few results represent different attempts to simplify the above theorem by separating ξ from π_n in the various terms involved. For the first result only a slight modification in the proof of Theorem 3.B is necessary.

Theorem 3.C (L_∞-version: upper bound and convergence). Suppose that the prerequisites of Theorem 3.B hold, but now replace Π in the mixing condition 2 by $\widetilde{\Pi} := \xi(A_{\mathrm{int}})/|A_{\mathrm{int}}|$, and write $\breve{\beta}_n^{(\infty)}$, instead of $\breve{\beta}_n$, for the function bounding the covariances. Furthermore set

$$p_n^{(\infty)} := \big\| \sup_{s \in J_n} \pi_n(s) \big\|_{L_\infty}.$$

Then we obtain for arbitrary $m := m(n) \in \mathbb{N}$

$$\begin{aligned} & d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\big) \\ & \quad = O\Big(n^d(p_n^{(\infty)})^2\,,\ m^d\big(n^d \wedge \tfrac{\log^{\uparrow}(n^d p_n^{(\infty)})}{\underline{w}_1}\big)(p_n^{(\infty)})^2\,,\ \sqrt{n^d}\big(\sqrt{n^d} \wedge \tfrac{1}{\sqrt{\underline{w}_1}}\big) p_n^{(\infty)} \breve{\beta}_n^{(\infty)}(m)\Big) \\ & \hspace{30em} \text{for } n \to \infty. \end{aligned}$$

Convergence Condition: The right hand side goes to 0 if, for example, $p_n^{(\infty)} = O\big(1/n^d\big)$ and there is a sequence $(m(n))_n$ with $m(n) = o(n)$ such that $\breve{\beta}_n^{(\infty)}(m(n)) = o(1)$ for $n \to \infty$.

Remark 3.D (Convergence towards a fixed Poisson process). If in fact ξ and π_n are such that

$$\xi_{\pi_n}\kappa_n^{-1}|_J \xrightarrow{\mathcal{D}} \mathrm{Po}(\lambda) \quad (n \to \infty)$$

for some finite measure λ on J, then we obtain by Theorem 1.A that, on J,

$$\int_{\kappa_n^{-1}(\cdot)} \pi_n(s)\, \xi(ds) \xrightarrow{\mathcal{D}} \lambda(\cdot) \quad (n \to \infty),$$

which implies under the convergence condition in Theorem 3.C that also

$$\nu_n\kappa_n^{-1}(A) \longrightarrow \lambda(A) \quad (n \to \infty)$$

for any Borel set $A \subset J$ with $\lambda(\partial A) = 0$. Thus, by Inequality (1.3) and Theorem 3.C, we get an upper bound for $d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\lambda)\big)$ that goes to zero under the convergence condition. Of course, the conditions are stronger than the ones for Proposition 1.B, but in return the upper bound is much more explicit and easier to apply.

The next result is a direct consequence of Theorem 3.B. This time, we try to gain a convergence rate that goes to zero under the weaker assumption on the sequence (π_n) that was used for Theorem 1.A, namely that $\sup_{s \in \mathbb{R}_+^d} \pi_n(s) \xrightarrow{\mathcal{D}} 0$. It should be noted, that the simple but rather crude estimates we are using here, are based on the assumption that the $\xi(C_{\boldsymbol{k}})$ have a generous number of moments. It is obviously by no means the best result attainable based on Theorem 3.B.

Corollary 3.E (L_1-version: upper bound and convergence). Suppose that the prerequisites of Theorem 3.B hold, and set

$$p_n^{(1)} := \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s)\Big), \quad p_n^{(2)} := \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s)^2\Big).$$

Then we obtain for arbitrary $m := m(n) \in \mathbb{N}$, and $Q_1 := Q_1(n)$, $Q_2 := Q_2(n) \in \mathbb{Z}_+$:

$$\begin{aligned}
&d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\big)\\
&\quad = O\Big(n^d\big(Q_2 p_n^{(2)} + \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2 1_{\{\xi(C_{\boldsymbol{k}})^2>Q_2\}}\big)\big),\\
&\qquad\quad m^d\big(n^d \wedge \tfrac{\log^\uparrow(n^d w_1)}{\underline{w}_1}\big)\big(Q_2 p_n^{(2)} + \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2 1_{\{\xi(C_{\boldsymbol{k}})^2>Q_2\}}\big)\big),\\
&\qquad\quad m^d\big(n^d \wedge \tfrac{\log^\uparrow(n^d w_1)}{\underline{w}_1}\big)\big(Q_1 p_n^{(1)} + \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}}) 1_{\{\xi(C_{\boldsymbol{k}})>Q_1\}}\big)\big)^2,\\
&\qquad\quad \sqrt{n^d}\big(\sqrt{n^d} \wedge \tfrac{1}{\sqrt{\underline{w}_1}}\big)\big(Q_1 p_n^{(1)} + \sup_{1\le \boldsymbol{k}\le n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}}) 1_{\{\xi(C_{\boldsymbol{k}})>Q_1\}}\big)\big)\breve{\beta}_n(m)\Big)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } n\to\infty.
\end{aligned}$$

Convergence Condition: Provided that $\xi(C_{\boldsymbol{k}})$ has sufficiently many moments that are bounded uniformly in $\boldsymbol{k}$, we can choose $m := m(n)$, $Q_1 := Q_1(n)$, and $Q_2 := Q_2(n)$ in such a way that the right hand side goes to 0. For example, under the assumptions that $p_n^{(1)} = O\big(1/n^d\big)$, $p_n^{(2)} = O\big(1/n^{(1+x)d}\big)$ $(n\to\infty)$ for some $x \in (0,1]$, and $\breve{\beta}_n(t) = \breve{\beta}(t) = O(1/t^{sd})$ $(t\to\infty)$ for some $s > 0$, we get convergence to zero of the upper bound if

$$\sup_{\boldsymbol{k}\in\mathbb{N}^d} \mathbb{E}\big(\xi(C_{\boldsymbol{k}})^r\big) < \infty \text{ for some } r > 2 + \frac{2s+1}{sx}.$$

We now examine how certain independence properties can be exploited. The main benefit of this is a more convenient mixing condition, that takes only the point process into account, allowing the retention field to be dealt with separately.

Assumption 2′ (Mixing property). Let $\breve{\beta}^{(\mathrm{ind})} : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that for every open cube $A_{\mathrm{int}} = (a, a+h\mathbf{1})$ of side length $h \in (0,1]$, and with minimal corner in $a \in \mathbb{R}_+^d$, and every surrounding set $A_{\mathrm{ext}}^{(t)} := \mathbb{R}_+^d \setminus [a - t\mathbf{1}, a + (t+h)\mathbf{1}]$ with $t \in \mathbb{R}_+$ we have that

$$\sup_{\substack{D\in\mathcal{F}_{\mathrm{ext}}^{(t)}\\ Z\in\mathcal{L}_2(\mathcal{F}_{\mathrm{int}}),\, 0\le Z\le\widetilde{\Pi}}} \big|\mathrm{cov}\big(Z, 1_D\big)\big| \le \breve{\beta}^{(\mathrm{ind})}(t),$$

where $\mathcal{F}_{\mathrm{int}} = \sigma(\xi|_{A_{\mathrm{int}}})$, $\mathcal{F}_{\mathrm{ext}}^{(t)} = \sigma(\xi|_{A_{\mathrm{ext}}^{(t)}})$, and $\widetilde{\Pi} := \xi(A_{\mathrm{int}})/|A_{\mathrm{int}}|$.

Theorem 3.F (Independent thinning: upper bound and convergence). Suppose that the prerequisites of Theorem 3.B hold with Assumption 2′, instead of Assumption 2. Let ξ and π_n be independent for any $n \in \mathbb{N}$. Note that we now have

$$\nu_n(\cdot) = \mathbb{E}\Big(\int_\cdot \pi_n(s)\,\xi(ds)\Big) = \int_\cdot \mathbb{E}\pi_n(s)\,\mu_1(ds).$$

Choose $m := m(n) \in \mathbb{N}$ in such a way that $\pi_n|_{\widetilde{A}_{\mathrm{int}}}$ and $\pi_n|_{\widetilde{A}^{(m)}_{\mathrm{ext}}}$ are independent for every unit cube $\widetilde{A}_{\mathrm{int}} = (a, a+\mathbf{1})$ and every surrounding set $\widetilde{A}^{(m)}_{\mathrm{ext}} := \mathbb{R}^d_+ \setminus [a-m\mathbf{1}, a+(m+1)\mathbf{1}]$, $a \in \mathbb{R}^d_+$. Then

$$\begin{aligned} &d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\big) \\ &\quad = O\Big(n^d p_n^{(2)}\,,\; m^d\big(n^d \wedge \tfrac{\log^{\uparrow}(n^d p_n^{(1)})}{\underline{w}_1}\big)p_n^{(2)}\,,\; \sqrt{n^d}\big(\sqrt{n^d} \wedge \tfrac{1}{\sqrt{\underline{w}_1}}\big)p_n^{(1)}\breve{\beta}^{(\mathrm{ind})}(m)\Big) \quad \text{for } n \to \infty. \end{aligned}$$

Convergence Condition: The right hand side goes to 0 if, for example, $p_n^{(1)} = O\big(1/n^d\big)$, $p_n^{(2)} = o\big(1/n^d\big)$ for $n \to \infty$, $\breve{\beta}^{(\mathrm{ind})}(t) = o(1)$ for $t \to \infty$, and there is a sequence $(m(n))_n$ with $m(n) = o\big(1/\big(n(p_n^{(2)})^{1/d}\big)\big)$ such that $\pi_n|_{\widetilde{A}_{\mathrm{int}}}$ and $\pi_n|_{\widetilde{A}^{(m(n))}_{\mathrm{ext}}}$ are independent for all sets $\widetilde{A}_{\mathrm{int}}$ and $\widetilde{A}^{(m(n))}_{\mathrm{ext}}$ of the above form.

3.2 Proofs

A complete proof is presented only for Theorem 3.B. For the other statements the corresponding modifications are given.

Proof of Theorem 3.B. Let $\eta_n \sim \mathrm{Po}(\nu_n)$. Our processes $\xi_{\pi_n}\kappa_n^{-1}|_J$ and $\eta_n\kappa_n^{-1}|_J$ are discretized in the following way. By Assumption 1 we may suppose that $\xi(\omega)(J_n \setminus \widetilde{J}_n) = \eta(\omega)(J_n \setminus \widetilde{J}_n) = 0$ for every $\omega \in \Omega$, without changing the distributions of the point processes. Then, subdivide $\widetilde{J}_n$ in the domain of the contraction κ_n into the hypercubes $C_{\boldsymbol{k}}$, which were introduced in Section 2. Choose an arbitrary $\tilde{n} \in \mathbb{N}$, and further subdivide every $C_{\boldsymbol{k}}$ into cubes $C_{\boldsymbol{kr}} := \prod_{i=1}^d [k_i - 1 + (r_i-1)/\tilde{n}, k_i - 1 + r_i/\tilde{n})$ of side length $1/\tilde{n}$ for $\boldsymbol{r} = (r_1, \ldots, r_d) \in \{1, 2, \ldots, \tilde{n}\}^d$. The concrete "inner" and "outer" sets used for Assumption 2 are given by

$$A_{\mathrm{int}}(\boldsymbol{k}, \boldsymbol{r}) := \mathring{C}_{\boldsymbol{kr}}\,,\; A^{(t)}_{\mathrm{ext}}(\boldsymbol{k}) := \Big(\bigcup\nolimits_{\substack{\boldsymbol{j}=1\\ |\boldsymbol{j}-\boldsymbol{k}|>t}}^n C_{\boldsymbol{j}}\Big)^{\circ}$$

for $t \in \mathbb{Z}_+$, $\boldsymbol{k} \in \{1, 2, \ldots, n\}^d$, and $\boldsymbol{r} \in \{1, 2, \ldots, \tilde{n}\}^d$, where $\mathring{C}$ denotes the interior of $C \subset \mathbb{R}^d$. We denote by $\alpha_{\boldsymbol{kr}}$ the center of $C_{\boldsymbol{kr}}$, and by $\theta_{\boldsymbol{kr}} := \kappa_n(\alpha_{\boldsymbol{kr}})$ the center of the contracted hypercube $\kappa_n(C_{\boldsymbol{kr}})$. Set furthermore

$$\begin{gathered} I_{\boldsymbol{kr}} := 1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) \geq 1\}}, \quad q_{\boldsymbol{kr}} := \mathbb{E}I_{\boldsymbol{kr}}, \text{ and let} \\ Y_{\boldsymbol{kr}} \sim \mathrm{Po}(q_{\boldsymbol{kr}}), \boldsymbol{k} \in \{1, 2, \ldots, n\}^d, \boldsymbol{r} \in \{1, 2, \ldots, \tilde{n}\}^d, \text{ independent.} \end{gathered}$$

Construct the discretized processes (note that for the Poisson process our discretization is only one "in distribution") as

$$\Xi_n := \sum_{\boldsymbol{k}=1}^{n} \sum_{\boldsymbol{r}=1}^{\tilde{n}} I_{\boldsymbol{kr}} \delta_{\alpha_{\boldsymbol{kr}}}, \quad \mathrm{H}_n := \sum_{\boldsymbol{k}=1}^{n} \sum_{\boldsymbol{r}=1}^{\tilde{n}} Y_{\boldsymbol{kr}} \delta_{\alpha_{\boldsymbol{kr}}}.$$

Note that the $\boldsymbol{r}$-sum over $q_{\boldsymbol{kr}}$ can be estimated from above as

$$\begin{aligned}\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}} \leq \mathbb{E}\xi_{\pi_n}(C_{\boldsymbol{k}}) &= \mathbb{E}\Big(\mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}}) \bigm| \xi, \pi_n\big)\Big) \\ &= \mathbb{E}\bigg(\mathbb{E}\bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} X_i \biggm| \xi, \pi_n\bigg)\bigg) \\ &= \mathbb{E}\bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} \pi_n(S_i)\bigg) \leq w_1, \end{aligned} \tag{3.1}$$

where we used the "tacit numbering of points" and the "tacit definition of the retention decisions X_i" as announced in Remark 2.A. In the analogous way, the same sum can be estimated from below as

$$\begin{aligned}\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}} &\geq \mathbb{E}\bigg(\mathbb{P}\bigg[\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} X_i \geq 1 \biggm| \xi, \pi_n\bigg]\bigg) \\ &\geq \mathbb{E}\Big(1_{\{\xi(C_{\boldsymbol{k}})\geq 1\}} \sup_{1\leq i\leq \xi(C_{\boldsymbol{k}})} \pi_n(S_i)\Big) \geq \underline{w}_1.\end{aligned}$$

The initial distance is split up as follows:

$$\begin{aligned}d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathcal{L}(\eta_n\kappa_n^{-1}|_J)\big) &\leq d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathcal{L}(\Xi_n\kappa_n^{-1})\big) \\ &\quad + d_2\big(\mathcal{L}(\Xi_n\kappa_n^{-1}), \mathcal{L}(\mathrm{H}_n\kappa_n^{-1})\big) \\ &\quad + d_2\big(\mathcal{L}(\mathrm{H}_n\kappa_n^{-1}), \mathcal{L}(\eta_n\kappa_n^{-1}|_J)\big).\end{aligned} \tag{3.2}$$

We first attend to the discretization errors, which are represented by the first and third terms on the right hand side. We have by Equations (2.3) and (2.2)

$$\begin{aligned}&d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathcal{L}(\Xi_n\kappa_n^{-1})\big) \\ &\quad\leq \mathbb{E}\big(d_1(\xi_{\pi_n}\kappa_n^{-1}|_J, \Xi_n\kappa_n^{-1})1_{\{\xi_{\pi_n}(J_n)=\Xi_n(J_n)\}}\big) + \mathbb{P}\big[\xi_{\pi_n}(J_n) \neq \Xi_n(J_n)\big] \\ &\quad= \mathbb{E}\big(d_1(\xi_{\pi_n}\kappa_n^{-1}|_J, \Xi_n\kappa_n^{-1})1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})=\Xi_n(C_{\boldsymbol{kr}})\,\forall \boldsymbol{k},\boldsymbol{r}\}}\big) + \mathbb{P}\big[\textstyle\bigcup_{\boldsymbol{k},\boldsymbol{r}}\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) \geq 2\}\big] \\ &\quad\leq \sqrt{d}/(2n\tilde{n}) + \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{P}[\xi_{\pi_n}(C_{\boldsymbol{kr}}) \geq 2].\end{aligned}$$

The first summand in the last line is obtained because, for estimating the d_1-distance on $\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) = \Xi_n(C_{\boldsymbol{kr}})\ \forall \boldsymbol{k}, \boldsymbol{r}\}$, we can pair every point of $\xi_{\pi_n}\kappa_n^{-1}|_J$ with the center of that cube $\kappa_n(C_{\boldsymbol{kr}})$ in which it lies; this center is at most $\sqrt{d}/(2n\tilde{n})$ (half a body diagonal of $\kappa_n(C_{\boldsymbol{kr}})$) apart, and a point of $\Xi_n\kappa_n^{-1}$. A similar argument is used to obtain $\sqrt{d}/(2n\tilde{n})$ in the fourth line of the next formula, with the difference that now there might be more than one point pairing in each $\kappa_n(C_{\boldsymbol{kr}})$. Thus

$$
\begin{aligned}
&d_2\big(\mathcal{L}(\eta_n\kappa_n^{-1}|_J),\mathcal{L}(\mathrm{H}_n\kappa_n^{-1})\big)\\
&\qquad\qquad \le d_2\bigg(\mathcal{L}\big(\eta_n\kappa_n^{-1}|_J\big),\mathcal{L}\Big(\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\theta_{\boldsymbol{kr}}}\Big)\bigg)\\
&\qquad\qquad\qquad + d_2\bigg(\mathcal{L}\Big(\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\theta_{\boldsymbol{kr}}}\Big),\mathcal{L}\Big(\sum_{\boldsymbol{k},\boldsymbol{r}}Y_{\boldsymbol{kr}}\delta_{\theta_{\boldsymbol{kr}}}\Big)\bigg)\\
&\qquad\qquad \le \sqrt{d}/(2n\tilde{n}) + d_{TV}\bigg(\mathcal{L}\Big(\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\theta_{\boldsymbol{kr}}}\Big),\mathcal{L}\Big(\sum_{\boldsymbol{k},\boldsymbol{r}}Y_{\boldsymbol{kr}}\delta_{\theta_{\boldsymbol{kr}}}\Big)\bigg)\\
&\qquad\qquad \le \sqrt{d}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}} d_{TV}\big(\mathcal{L}(\eta_n(C_{\boldsymbol{kr}})),\mathcal{L}(Y_{\boldsymbol{kr}})\big)\\
&\qquad\qquad \le \sqrt{d}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}}\big|\mathbb{E}\xi_{\pi_n}(C_{\boldsymbol{kr}}) - \mathbb{P}[\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 1]\big|\\
&\qquad\qquad = \sqrt{d}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}}\mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{kr}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 2\}}\big) - \sum_{\boldsymbol{k},\boldsymbol{r}}\mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 2\big],
\end{aligned}
$$

where d_{TV} denotes the total variation distance between probability distributions (see e.g. Barbour, Holst and Janson (1992), Appendix A.1 for definition and results). Since the sum over $\boldsymbol{r}$ in the second term can be estimated as

$$
\begin{aligned}
\sum_{\boldsymbol{r}=1}^{\tilde{n}}\mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{kr}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 2\}}\big) &\le \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{k}})\ge 2\}}\big)\\
&\le \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})(\xi_{\pi_n}(C_{\boldsymbol{k}})-1)\big)\\
&= \mathbb{E}\bigg(\mathbb{E}\Big(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{\substack{j=1\\ j\ne i}}^{\xi(C_{\boldsymbol{k}})} X_iX_j \;\Big|\; \xi,\pi_n\Big)\bigg)\\
&= \mathbb{E}\Big(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{\substack{j=1\\ j\ne i}}^{\xi(C_{\boldsymbol{k}})}\pi_n(S_i)\pi_n(S_j)\Big)\\
&\le \mathbb{E}\Big(\xi(C_{\boldsymbol{k}})\big(\xi(C_{\boldsymbol{k}})-1\big)\Big(\sup_{1\le i\le \xi(C_{\boldsymbol{k}})}\pi_n(S_i,\cdot)\Big)^2\Big)\\
&\le w_{[2]},
\end{aligned}
\tag{3.3}
$$

we obtain, as an overall bound for the discretization error terms,

$$
\sqrt{d}/(n\tilde{n}) + n^d w_{[2]}. \tag{3.4}
$$

Next we consider the left over term in (3.2), which is estimated by application of the Barbour-Brown theorem A.A (see Appendix A.1). We choose $m := m(n) \in \mathbb{N}$

arbitrarily, and set in the notation of the appendix

$$\begin{aligned}
\Theta &= \{1,2,\ldots,n\}^d \times \{1,2,\ldots,\tilde{n}\}^d \\
\Theta_{\boldsymbol{kr}} &= \Theta \setminus \{(\boldsymbol{k},\boldsymbol{r})\} \\
\Theta^s_{\boldsymbol{kr}} &= \{(\boldsymbol{l},\boldsymbol{s}) \in \Theta_{\boldsymbol{kr}}\,;\ |\boldsymbol{l}-\boldsymbol{k}| \le m\} \\
\Theta^w_{\boldsymbol{kr}} &= \{(\boldsymbol{l},\boldsymbol{s}) \in \Theta_{\boldsymbol{kr}}\,;\ |\boldsymbol{l}-\boldsymbol{k}| > m\}.
\end{aligned}$$

Since

$$\mathbb{E}\,\Xi_n(\{\alpha_{\boldsymbol{kr}}\}) = q_{\boldsymbol{kr}} = \mathbb{E}\,\mathrm{H}_n(\{\alpha_{\boldsymbol{kr}}\}),$$

we can apply Theorem A.A and obtain (once more in the notation of the appendix):

$$d_2\big(\mathcal{L}(\Xi_n\kappa_n^{-1}), \mathcal{L}(\mathrm{H}_n\kappa_n^{-1})\big) \le M_1(\lambda)\sum_{\boldsymbol{k},\boldsymbol{r}}\big(q_{\boldsymbol{kr}}\mathbb{E}\widetilde{Z}_{\boldsymbol{kr}} + \mathbb{E}(I_{\boldsymbol{kr}}Z_{\boldsymbol{kr}})\big) + M_2(\lambda)\sum_{\boldsymbol{k},\boldsymbol{r}} e_{\boldsymbol{kr}}.$$

Further estimation of the various terms yields

$$M_1(\lambda) \le 1 \wedge \big[\tfrac{2}{n^d\underline{w}_1}\big(1 + 2\log^+\big(\tfrac{n^d w_1}{2}\big)\big)\big]; \tag{3.5}$$

$$M_2(\lambda) \le 1 \wedge \big[1.65/(n^d\underline{w}_1)^{1/2}\big]; \tag{3.6}$$

$$\sum_{\boldsymbol{k},\boldsymbol{r}} q_{\boldsymbol{kr}}\mathbb{E}\widetilde{Z}_{\boldsymbol{kr}} = \sum_{\boldsymbol{k}=1}^{n}\ \sum_{\substack{\boldsymbol{l}=1\\|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n}\bigg(\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}}\bigg)\bigg(\sum_{\boldsymbol{s}=1}^{\tilde{n}} q_{\boldsymbol{ls}}\bigg) \le (2m+1)^d n^d w_1^2; \tag{3.7}$$

and, using part of Inequality (3.3) for the second line,

$$\begin{aligned}
\sum_{\boldsymbol{k},\boldsymbol{r}}\mathbb{E}(I_{\boldsymbol{kr}}Z_{\boldsymbol{kr}}) &\le \sum_{\boldsymbol{k}=1}^{n}\Bigg(\sum_{\substack{\boldsymbol{l}=1\\1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n}\mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})\,\xi_{\pi_n}(C_{\boldsymbol{l}})\big) \;+\; \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})(\xi_{\pi_n}(C_{\boldsymbol{k}})-1)\big)\Bigg) \\
&\le \sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n}\mathbb{E}\Bigg(\mathbb{E}\bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\ \sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})} X_iX_j\ \bigg|\ \xi,\pi_n\bigg)\Bigg) \;+\; n^d w_{[2]} \\
&= \sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n}\mathbb{E}\bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\ \sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})}\pi_n(S_i)\pi_n(S_j)\bigg) \;+\; n^d w_{[2]} \\
&\le (2m+1)^d n^d w_{[2]},
\end{aligned} \tag{3.8}$$

where our numbering in Inequality (3.8) is such that $S_1,\ldots,S_{\xi(C_{\boldsymbol{k}})}$ lie in $C_{\boldsymbol{k}}$, and $S_{\xi(C_{\boldsymbol{k}})+1},\ldots,S_{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})}$ in $C_{\boldsymbol{l}}$.

A little more work is needed to estimate $e_{\boldsymbol{kr}}$. First, note that by Assumption 1, the probability that any points of ξ lie on the grid $G := \bigcup_{\boldsymbol{kr}}\partial C_{\boldsymbol{kr}}$ is zero. Since we are only interested in distributional properties of ξ_π, we may therefore assume w.l.o.g. that $\xi(\omega)(G) = 0$ for all $\omega \in \Omega$. For any set U, denote its power set by $\mathcal{P}(U)$, and write

$$\mathcal{F}^w_{\boldsymbol{k}} := \mathcal{P}\big(\{0,1\}^{\Theta^w_{\boldsymbol{kr}}}\big)\,,\ \text{and } W_{\boldsymbol{k}} := (I_{\boldsymbol{ls}})_{(\boldsymbol{l},\boldsymbol{s})\in\Theta^w_{\boldsymbol{kr}}}$$

(note that $\Theta^w_{\boldsymbol{kr}}$ does not depend on $\boldsymbol{r}$). If $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}})\sup_{s\in C_{\boldsymbol{kr}}}\pi_n(s)\big)$ is zero, it is evident that $e_{\boldsymbol{kr}} = 0$, so assume that the expectation is non-zero. We use the "formula of total covariance", that is, the relation

$$\operatorname{cov}(X,Y) = \mathbb{E}(\operatorname{cov}(X,Y \,|\, Z)) + \operatorname{cov}(\mathbb{E}(X \,|\, Z), \mathbb{E}(Y \,|\, Z))$$

for random variables $X, Y \in L_2$ and an arbitrary random variable Z, along with the conditional independence of the retention decisions, to obtain

$$\begin{aligned}
e_{\boldsymbol{kr}}/2 &= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\big(1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1\}}\,,\ 1_{\{W_{\boldsymbol{k}}\in B\}}\big)\Big| \\
&= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\big(\mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1 \;\big|\; \xi,\pi_n\big]\,,\ \mathbb{P}\big[W_{\boldsymbol{k}}\in B \;\big|\; \xi,\pi_n\big]\big)\Big|. \qquad (3.9)
\end{aligned}$$

Since no realization of ξ has any points in G, the arguments of the covariance are conditional probabilities of the form $\mathbb{P}[\xi_{\pi_n} \in D \,|\, \xi, \pi_n]$ with $D \in \mathfrak{N}(A) := \sigma\big(\big\{\{\varrho \in \mathcal{N}; \varrho(B) = l\}\,;\ B \in \mathcal{B}_A, l \in \mathbb{Z}_+\big\}\big) \subset \mathfrak{N}$, where $A = A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})$ and $A = A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})$, respectively. We then may condition as well only on ξ and π_n restricted to the corresponding set A (the proof of this is rather technical and has therefore been placed in Appendix A.4). Hence

$$\begin{aligned}
\frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde n} e_{\boldsymbol{kr}} &= \sum_{\boldsymbol{r}=1}^{\tilde n}\max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\big(\mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1 \;\big|\; \xi|_{A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})}, \pi_n|_{A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})}\big], \\
&\qquad\qquad\qquad\qquad \mathbb{P}\big[W_{\boldsymbol{k}}\in B \;\big|\; \xi|_{A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})}, \pi_n|_{A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})}\big]\big)\Big| \\
&\leq w_1\,\breve\beta_n(m), \qquad (3.10)
\end{aligned}$$

where for the inequality, the factor $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}})\sup_{s\in C_{\boldsymbol{kr}}}\pi_n(s)\big)$ was extracted from the first argument of the covariance. Note that in the second argument we do not have an indicator as required, but a general $[0,1]$-valued random variable. The upper bound from Assumption 2 still holds, as can be seen from the proof of Equation ($1'$) in Doukhan (1994), Section 1.1.

We assemble the different parts from Result (3.4), Inequalities (3.5) to (3.8), and Inequality (3.10), and let $\tilde n$ go to infinity to obtain the overall estimate

$$\begin{aligned}
&d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathcal{L}(\eta_n\kappa_n^{-1}|_J)\big) \\
&\qquad\leq\ n^d w_{[2]} + \Big(1 \wedge \Big[\tfrac{2}{n^d\underline{w}_1}\big(1 + 2\log^+\big(\tfrac{n^d w_1}{2}\big)\big)\Big]\Big)(2m+1)^d n^d (w_1^2 + w_{[2]}) \\
&\qquad\quad + \Big(1 \wedge \tfrac{1.65}{\sqrt{n^d\underline{w}_1}}\Big) 2n^d w_1 \breve\beta_n(m),
\end{aligned}$$

which is of the required order for $n \to \infty$. $\square$

Proof of Theorem 3.C. Use, in the upper bound of Theorem 3.B, the estimates $w_1 \leq \bar{h}_1 p_n^{(\infty)}$ and $w_{[2]} \leq \bar{h}_2 (p_n^{(\infty)})^2$, and modify the estimation of the $e_{\boldsymbol{kr}}$ by extracting $p_n^{(\infty)} |\mathrm{C}_{\boldsymbol{kr}}|$ instead of $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}}) \sup_{s \in C_{\boldsymbol{kr}}} \pi_n(s)\big)$ in Inequality (3.10), such that we get

$$\frac{1}{2} \sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \leq p_n^{(\infty)} \breve{\beta}_n^{(\infty)}(m).$$

□

Proof of Corollary 3.E. In the upper bound of Theorem 3.B, use the estimates

$$w_1 \leq Q_1 p_n^{(1)} + \sup_{1 \leq \boldsymbol{k} \leq n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}}) 1_{\{\xi(C_{\boldsymbol{k}}) > Q_1\}}\big), \;\; w_{[2]} \leq Q_2 p_n^{(2)} + \sup_{1 \leq \boldsymbol{k} \leq n} \mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2 1_{\{\xi(C_{\boldsymbol{k}})^2 > Q_2\}}\big),$$

from which we obtain the required order. For the convergence condition, set $m(n) := n^y$, $Q_1^{(\varepsilon)}(n) := n^{(sy-\varepsilon)d}$ and $Q_2^{(\varepsilon)}(n) := n^{(2s(x+1)y-\varepsilon)d}$, where $y := x/\big(2s(x+1)+1\big)$ and $\varepsilon > 0$ is arbitrary. Use that $\mathbb{E}(Y 1_{\{Y > Q\}}) = o\big((1/Q)^{r-1}\big)$ for any non-negative random variable Y with finite r-th moment, in order to show that for any $r > 2 + (2s+1)/(sx)$, there is still a choice of $\varepsilon > 0$, such that the upper bound in Corollary 3.E goes to zero. □

Proof of Theorem 3.F. Use in the upper bound of Theorem 3.B the estimates $w_1 \leq \bar{h}_1 p_n^{(1)}$ and $w_{[2]} \leq \bar{h}_2 p_n^{(2)}$, and modify the estimation of the $e_{\boldsymbol{kr}}$ in the following way: for better readability, suppress $\boldsymbol{k}$ and $\boldsymbol{r}$ in the expressions $A_{\mathrm{int}}(\boldsymbol{k}, \boldsymbol{r})$, $A_{\mathrm{ext}}^{(m)}(\boldsymbol{k})$, and write $f(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}})$ and $g(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}}) = g_B(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}})$ for the conditional probabilities in (3.10), such that

$$e_{\boldsymbol{kr}}/2 = \max_{B \in \mathcal{F}_{\boldsymbol{k}}^w} \big|\mathrm{cov}\big(f(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}}) \,,\; g_B(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}})\big)\big|.$$

Then, because of the independence of $(\xi, \pi_n|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}})$,

$$\begin{aligned}
&\big|\mathrm{cov}\big(f(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}}) \,,\; g(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}})\big)\big| \\
&\quad \leq \mathbb{E}\Big|\mathrm{cov}\big(f(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}}) \,,\; g(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}}) \;\big|\; \pi_n|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)\Big| \\
&\quad \leq |A_{\mathrm{int}}| \, \mathbb{E}\Big(\sup_{s \in C_{\boldsymbol{k}}} \pi_n(s) \; \Big|\mathrm{cov}\Big(\tfrac{f(\xi|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{int}}})}{|A_{\mathrm{int}}| \sup_{s \in A_{\mathrm{int}}} \pi_n(s)} \,,\; g(\xi|_{A_{\mathrm{ext}}^{(m)}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}}) \;\Big|\; \pi_n|_{A_{\mathrm{int}}}, \pi_n|_{A_{\mathrm{ext}}^{(m)}}\Big)\Big|\Big) \\
&\quad \leq |A_{\mathrm{int}}| \, p_n^{(1)} \breve{\beta}^{(\mathrm{ind})}(m),
\end{aligned}$$

hence

$$\frac{1}{2} \sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \leq p_n^{(1)} \breve{\beta}^{(\mathrm{ind})}(m),$$

which yields the desired result. □

3.3 Example of thinning according to a random environment (clouds in the starry sky)

Suppose that the stars within a certain distance $\bar{r}$ to earth, and lying in a large window $J_n \subset \mathbb{R}_+^2$ of the night sky, are part of a point process ξ on $\mathbb{R}_+^2$ that fulfills Assumptions 1 and $2'$. Whether you can actually see a particular star in this window depends, among other things, upon the distance and the brightness of the star, and upon whether any other object (here, a cloud) covers the star. Suppose, for the sake of simplicity, that the distributions of distance and brightness of stars do not depend on their position in the window, so that there is a basic probability $q_0 = q_0^{(n)}$ that you can see any fixed star in that part of the window that is not covered by clouds. Suppose furthermore, that the clouds in the sky as seen from earth in the upper right area of some reference point, form a separable RACS (random closed set) $\Xi \subset \mathbb{R}_+^2$ that is independent of ξ. By Remark A.E from the appendix, Theorem 3.F may then be applied for the retention field given by $\pi_n(\omega, s) = q_0(1-1_{\Xi(\omega)}(s))$. In the following, we admit a general dimension $d \in \mathbb{N}$ for the sky.

In order to make things more concrete, let us consider a toy example. Suppose the cloud RACS $\Xi := \Xi_n$ is a very simple Boolean model consisting only of discs of positive i.i.d. radii whose centers form a homogeneous Poisson process on $\mathbb{R}^d$: Denote by $\oplus$ the (Minkowski-)addition of subsets of $\mathbb{R}^d$, i.e. $A \oplus B = \{a + b;\, a \in A, b \in B\}$, and set

$$\Xi := \bigcup_{i=1}^{\infty} \bigl(Y_i \oplus \overline{\mathbb{B}}(0, R_i^{(n)})\bigr) \cap \mathbb{R}_+^d, \tag{3.11}$$

where $(Y_i)_{i\in\mathbb{N}}$ are the points of a $\mathrm{Po}(\lambda Leb^d)$-process on $\mathbb{R}^d$, $\lambda > 0$, and $\overline{\mathbb{B}}(0, R_i^{(n)})$ is the closed Euclidean ball with center at 0 and radius $R_i^{(n)}$. Here $R_i^{(n)} \in L_\infty$ are i.i.d. $(0,\infty)$-valued random variables that are independent also of the Poisson process with $\|R_i^{(n)}\|_{L_d} =: r_n$, $\bigl(\mathbb{E}\bigl((R_i^{(n)} - \sqrt{d})^d 1_{\{R_i^{(n)} > \sqrt{d}\}}\bigr)\bigr)^{1/d} =: \check{r}_n$, and $\|R_i^{(n)}\|_{L_\infty} =: r_n^{(\infty)}$. Note that Ξ is a separable RACS. Its capacity functional is given by

$$T_\Xi(K) := \mathbb{P}[\Xi \cap K \neq \emptyset] = 1 - \exp\bigl(-\lambda \mathbb{E}\bigl(Leb^d(\overline{\mathbb{B}}(0, R_1^{(n)}) \oplus K)\bigr)\bigr) \tag{3.12}$$

for any compact subset $K \subset \mathbb{R}^d$; see e.g. Stoyan, Kendall and Mecke (1987), Section 3.1. Let $\widetilde{\Xi}$ be the RACS that is obtained from Ξ by decreasing the radii $R_i^{(n)}$ by $\sqrt{d}$ of those balls which have $R_i^{(n)} > \sqrt{d}$, and deleting those centers whose balls have $R_i^{(n)} \le \sqrt{d}$. Then, applying (3.12) on $C_{\boldsymbol{k}}$ (by continuity of the measure from below),

$$\begin{aligned}
\mathbb{P}\bigl[\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s) = q_0\bigr] &= \mathbb{P}\bigl[C_{\boldsymbol{k}} \not\subset \Xi\bigr] \le 1 - \mathbb{P}\bigl[\widetilde{\Xi} \cap C_{\boldsymbol{k}} \neq \emptyset\bigr] \\
&\le \exp\bigl(-\lambda \mathbb{E}\bigl(Leb^d(\overline{\mathbb{B}}(0, R_1^{(n)} - \sqrt{d}) \oplus C_{\boldsymbol{k}}) 1_{\{R_i^{(n)} > \sqrt{d}\}}\bigr)\bigr) \\
&\le \exp(-\lambda \alpha_d \check{r}_n^d),
\end{aligned}$$

where $\alpha_d := \pi^{d/2} / \Gamma(d/2+1)$ is the volume of the d-dimensional unit ball. Therefore

$$\mathbb{E}\bigl(\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s)\bigr) = q_0 \mathbb{P}\bigl[\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s) = q_0\bigr] \le q_0 \exp(-\lambda \alpha_d \check{r}_n^d),$$

and

$$\mathbb{E}\big(\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s)^2\big) = q_0^2\mathbb{P}\big[\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s) = q_0\big] \le q_0^2 \exp(-\lambda\alpha_d \check{r}_n^d).$$

Furthermore, we obtain for the approximating expectation measure in Theorem 3.F

$$\begin{aligned}\nu_n\kappa_n^{-1} &= \Big(\int_{\cdot} \mathbb{E}\pi_n(s)\ \mu_1(ds)\Big)\kappa_n^{-1} = q_0 \exp(-\lambda\alpha_d r_n^d)\mu_1\kappa_n^{-1}\\ &= q_0 \exp(-\lambda\alpha_d r_n^d) n^d \int_{\cdot} \big(h_1 \circ \kappa_n^{-1}\big)(s)\ Leb^d(ds),\end{aligned}$$

where the second equality is obtained by Equation (3.12) with $K = \{s\}$, yielding in total the following result.

Proposition 3.G. Let ξ be a point process on $\mathbb{R}_+^d$ that satisfies Assumptions 1 and 2′. Furthermore, let $\pi_n(\omega, s) = q_0(1 - 1_{\Xi_n(\omega)}(s))$ for all $s \in \mathbb{R}_+^d$, $\omega \in \Omega$, where $q_0 = q_0^{(n)} \in [0,1]$ and Ξ_n is the Boolean model given by Equation (3.11), i.e. the $\mathbb{R}_+^d$-part of a union of balls in $\mathbb{R}^d$ whose centers form a $\mathrm{Po}(\lambda Leb^d)$-process and whose radii $R_i^{(n)}$ are positive-valued L_∞ random variables with $r_n^{(\infty)} = \big\|R_i^{(n)}\big\|_{L_\infty}$, $r_n = \big\|R_i^{(n)}\big\|_{L_d}$, and $\check{r}_n = \big(\mathbb{E}\big((R_i^{(n)} - \sqrt{d})^d 1_{\{R_i^{(n)} > \sqrt{d}\}}\big)\big)^{1/d}$.
We then obtain, for $m = m(n) \in \mathbb{N}$ with $m(n) \ge 2r_n^{(\infty)}$, that

$$\begin{aligned}&d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\nu_n\kappa_n^{-1}|_J)\big)\\ &\quad = O\Big(m^d n^d q_0^2 \exp(-\lambda\alpha_d \check{r}_n^d)\,,\ n^d q_0 \exp(-\lambda\alpha_d \check{r}_n^d)\breve{\beta}^{(\mathrm{ind})}(m)\Big) \quad \text{for } n \to \infty,\end{aligned}$$

where $\nu_n = q_0 \exp(-\lambda\alpha_d r_n^d)\mu_1$, and α_d is the volume of the d-dimensional unit ball.

From an asymptotical point of view, clearly the interesting cases are those in which $\nu_n\kappa_n^{-1}|_J$ does not fade away to the zero measure as n tends to infinity, giving us an artificial benefit for our distance estimate. In order to prevent this behavior, we must avoid choosing r_n of a higher than logarithmic order. To get a quick idea, consider the special case in which there is a $n_0 \in \mathbb{N}$ such that $R_1^{(n)} > \sqrt{d}$ for $n \ge n_0$, and $\mathbb{E}(R_1^{(n)})^{d-1} = o\big(\mathbb{E}(R_1^{(n)})^d\big)$. This situation allows us to choose an arbitrary $\zeta > 0$, and we still find $n_1 \in \mathbb{N}$ such that $\check{r}_n^d \ge (1-\zeta) r_n^d$ for all $n \ge n_1$. Let us furthermore arrange for a constant non-zero $\nu_n\kappa_n^{-1}$: let $\mu_1 := \mu_0 Leb^d$ with $\mu_0 > 0$, choose $q_0 \ge 1/n^d$ and set

$$r_n := \Big(\frac{1}{\lambda\alpha_d} \log(q_0 n^d)\Big)^{1/d},$$

such that $\nu_n\kappa_n^{-1} = \mu_0 Leb^d$ for every $n \in \mathbb{N}$. The result is as follows.

Corollary 3.H. Under the conditions of Proposition 3.G, as well as the additional conditions above, we have for any $\zeta > 0$

$$d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\mu_0 Leb^d|_J)\big) = O\Big(m^d q_0 (n^d q_0)^\zeta\,,\ (n^d q_0)^\zeta \breve{\beta}^{(\mathrm{ind})}(m)\Big) \quad \text{for } n \to \infty.$$

The upper bound in the above proposition goes to zero under an appropriate choice of $m \geq 2r_n^{(\infty)}$ and $\zeta > 0$ if $q_0 = O(n^{-\varepsilon_1 d})$ as $n \to \infty$ and $\breve{\beta}^{(\mathrm{ind})}(t) = O(t^{-\varepsilon_2 d})$ as $t \to \infty$ for some $\varepsilon_1, \varepsilon_2 > 0$. Note that it is always possible to choose m appropriately, provided that $r_n^{(\infty)} = O(n^{\varepsilon_3})$ for some $\varepsilon_3 \leq \varepsilon_1$. In the case of convergence of the above bound to zero we obtain furthermore by Result (2.4) that

$$\xi_{\pi_n} \kappa_n^{-1}|_J \xrightarrow{\mathcal{D}} \mathrm{Po}(\mu_0 Leb^d|_J) \quad \text{for } n \to \infty.$$

4 An alternative thinning definition

In this section, we consider a different thinning concept, where in place of a random retention field, we have deterministic retention kernels by which we directly model dependences between retention decisions. This will lead us, by means of the same method of proof as before, to a theorem that is similar to Theorem 3.B, less appealing from a theoretical and a typographical point of view, but sometimes more intuitively applied, because it permits us to look at the situation from a different angle: rather than thinking of a thinning as the result of point deletions according to a (potentially inscrutable) random environment, we now understand it as the result of point deletions according to (potentially more transparent) point interactions.

4.1 Definition, requirements and results

We first present the new thinning definition. Let ξ be a point process on $\mathbb{R}_+^d$. To simplify the presentation we assume, for this whole section, that all realizations of ξ have infinitely many points in $\mathbb{R}_+^d$.

Definition ("*Q*-Thinning"). For any $u \in \mathbb{N}$ let $D_u := \{(s_1, s_2, \ldots, s_u; \sigma) \in (\mathbb{R}_+^d)^u \times \mathcal{N};\ \sum_{i=1}^u \delta_{s_i} \leq \sigma\}$, equipped with the trace σ-field of $(\mathcal{B}_+^d)^u \otimes \mathfrak{N}$, and denote by $D_u(\sigma) := \{\boldsymbol{s} \in (\mathbb{R}_+^d)^u; (\boldsymbol{s}; \sigma) \in D_u\}$ the section of D_u at $\sigma \in \mathcal{N}$. Let Q_u be a probability kernel from D_u to $\{0,1\}^u$. We call $(Q_u)_{u \in \mathbb{N}}$ an *admissible sequence of retention kernels* if

(a) the Q_u are "simultaneously symmetrical" in the sense that for any permutation τ on $\{1, \ldots, u\}$ and its corresponding linear transformations $T_1 : \big(\mathbb{R}_+^d\big)^u \to \big(\mathbb{R}_+^d\big)^u$, $(s_1, \ldots, s_u) \mapsto (s_{\tau(1)}, \ldots, s_{\tau(u)})$ and $T_2 : \{0,1\}^u \to \{0,1\}^u$, $(e_1, \ldots, e_u) \mapsto (e_{\tau(1)}, \ldots, e_{\tau(u)})$ we have

$$Q_u\big((T_1\boldsymbol{s}; \sigma), T_2(A)\big) = Q_u\big((\boldsymbol{s}; \sigma), A\big)$$

for every $(\boldsymbol{s}, \sigma) \in D_u$ and every $A \subset \{0,1\}^u$;

(b) the following compatibility condition holds between the Q_u: let $u \geq 2$ be arbitrary; then we have

$$Q_u\big((s_1, \ldots, s_{u-1}, s_u; \sigma), A \times \{0,1\}\big) = Q_{u-1}\big((s_1, \ldots, s_{u-1}; \sigma), A\big)$$

for every $(s_1, \ldots, s_{u-1}, s_u; \sigma) \in D_u$ and every $A \subset \{0,1\}^{u-1}$.

Assume now that $Q = (Q_u)_{u\in\mathbb{N}}$ is an admissible sequence of retention kernels. In analogy with Section 2, the Q-thinning of ξ can now be defined as follows. First, assume that $\xi = \sigma = \sum_{i=1}^{\infty} \delta_{s_i}$ is non-random, and define a Q-thinning of ξ in this case as $\xi_Q := \sum_{i=1}^{\infty} X_i \delta_{s_i}$, where the X_i are indicator random variables whose joint distribution is given by the fidi-distributions

$$\mathcal{L}(X_1, \dots, X_u) = Q_u\big((s_1, \dots, s_u; \sigma), \cdot\,\big)$$

for every $u \in \mathbb{N}$. It is easy to see that, due to Properties (a) and (b) from above, ξ_Q has a distribution P_σ that is well-defined and does not depend on the enumeration of σ. We obtain the general Q-thinning from this by randomization, as in Section 2.

Remark 4.A. Let $f_1, f_2, \dots$ be $\mathfrak{N}$-$\mathcal{B}_+^d$-measurable functions such that $\sigma = \sum_{i=1}^{\sigma(\mathbb{R}_+^d)} \delta_{f_i(\sigma)}$ for every $\sigma \in \mathcal{N}$. Such functions exist by Remark 2.A. With any such sequence of functions it is enough to define $Q_u\big((f_1(\sigma), \dots, f_u(\sigma); \sigma), A\big)$ for every $\sigma \in \mathcal{N}$, every $A \subset \{0,1\}^u$, and every $u \in \mathbb{N}$, in such a way that Properties (a) and (b) from the thinning definition are satisfied, and the above term is a $\mathfrak{N}$-measurable mapping in σ and a probability measure in A. There is then a unique admissible sequence $(\widetilde{Q}_u)_{u\in\mathbb{N}}$ of retention kernels such that $\widetilde{Q}_u$ extends Q_u to the whole of D_u for every $u \in \mathbb{N}$. A short proof of this statement is the topic of Appendix A.5.

Remark 4.B. The new Q-thinning concept generalizes the thinning concept from Section 2. That is to say, for any combination of a point process ξ and a locally evaluable random field π, the thinning ξ_π can be modeled as a Q-thinning ξ_Q. As in Remark 4.A, let $f_1, f_2, \dots$ be $\mathfrak{N}$-$\mathcal{B}_+^d$-measurable functions with $\sigma = \sum_i \delta_{f_i(\sigma)}$ for every $\sigma \in \mathcal{N}$. Define then for any $u \in \mathbb{N}$, and for $e_1, \dots, e_u \in \{0,1\}$,

$$Q_u\big((f_1(\sigma), \dots, f_u(\sigma); \sigma), \{(e_1, \dots, e_u)\}\big) := \mathbb{E}\Big(\prod_{i=1}^{u} \big(\pi(f_i(\xi))\big)^{e_i} \big(1 - \pi(f_i(\xi))\big)^{1-e_i} \,\Big|\, \xi = \sigma\Big) \tag{4.1}$$

for almost every σ, which, upon adaptation on a $\mathbb{P}\xi^{-1}$-null set, yields by Remark 4.A, a well-defined sequence of retention kernels. It follows then from (4.1) that for every $D \in \mathfrak{N}$

$$\mathbb{P}[\xi_Q \in D \,|\, \xi] = \mathbb{E}\big(\mathbb{P}[\xi_\pi \in D \,|\, \xi, \pi] \,\big|\, \xi\big) = \mathbb{P}[\xi_\pi \in D \,|\, \xi] \quad \text{a.s.},$$

and hence that ξ_Q has the same distribution as ξ_π.

One can prove a theorem corresponding to Theorem 3.B, which now relies on separate control of a mixing coefficient with respect to ξ alone and of the conditional covariances between functions of retention decisions. This seems intuitively more appealing, but is also quite a bit more inconvenient to formulate than the more abstract way via the σ-fields $\mathcal{F}_{\mathrm{int}}^{(n)}$ and $\mathcal{F}_{\mathrm{ext}}^{(n,t)}$ used for Theorem 3.B. To keep things reasonably neat, we only state a special case here, which is basically the analogue of Theorem 3.C, the L_∞-version.

We first present the additional assumptions we need. Let $Q^{(n)} := (Q_u^{(n)})_{u\in\mathbb{N}}$ for each $n \in \mathbb{N}$ be an admissible sequence of retention kernels, and set

$$\bar{p}_n := \operatorname{ess\,sup}\Bigl(\sup_{s\in D_1(\xi)\cap J_n} Q_1^{(n)}\bigl((s;\xi),\{1\}\bigr)\Bigr) \in [0,1],$$

$$\underline{p}_n := \operatorname{ess\,inf}\Bigl(\inf_{s\in D_1(\xi)\cap J_n} Q_1^{(n)}\bigl((s;\xi),\{1\}\bigr)\Bigr) \in [0,1],$$

where we define the supremum over the empty set as 0, and the infimum as 1. Note that the role of $\pi_n(\omega,s)$ is now always taken by $Q_1^{(n)}\bigl((s;\xi(\omega)),\{1\}\bigr)$, and that $D_1(\sigma)$ is just the set of all points of σ. The additional assumptions are as follows.

Assumption 2″ (Mixing property of $\boldsymbol{\xi}$). Let $\breve{\beta}^{(\infty)} : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that for any cubes $A_{\text{int}} = (a, a+h\mathbf{1})$, $A_{\text{int}}^{(\tilde{t})} = (a-\tilde{t}\mathbf{1}, a+(\tilde{t}+h)\mathbf{1})$ and surrounding sets $A_{\text{ext}}^{(\tilde{t},t)} := \mathbb{R}_+^d \setminus [a-(\tilde{t}+t)\mathbf{1}, a+(\tilde{t}+t+h)\mathbf{1}]$ with $h\in(0,1]$, $a\in\mathbb{R}_+^d$, and $\tilde{t}, t\in\mathbb{R}_+$ we have that

$$\sup_{\substack{D\in\mathcal{F}_{\text{ext}}^{(\tilde{t},t)}\\ Z\in\mathcal{L}_2(\mathcal{F}_{\text{int}}^{(\tilde{t})}),\,0\le Z\le\widetilde{\Pi}}} \bigl|\operatorname{cov}\bigl(Z,1_D\bigr)\bigr| \le \breve{\beta}^{(\infty)}(t),$$

where $\mathcal{F}_{\text{int}}^{(\tilde{t})} := \sigma\bigl(\xi|_{A_{\text{int}}^{(\tilde{t})}}\bigr)$, $\mathcal{F}_{\text{ext}}^{(\tilde{t},t)} := \sigma\bigl(\xi|_{A_{\text{ext}}^{(\tilde{t},t)}}\bigr)$, and $\widetilde{\Pi} := \xi(A_{\text{int}})/|A_{\text{int}}|$.

Assumption 3 (Local functional dependence of $\boldsymbol{Q_u^{(n)}(\cdot,A)}$ on $\boldsymbol{\sigma}$). For each $n\in\mathbb{N}$ there is a $\tilde{t}_n\in\mathbb{R}_+$, such that $Q_u^{(n)}(\cdot,A)$ is only a function of $\bigl(s_1,\ldots,s_u;\sigma|_{\bigcup_{i=1}^u \mathbb{B}(s_i,\tilde{t}_n)}\bigr)$ for $A\subset\{0,1\}^u$. That is, more exactly, $Q_u^{(n)}\bigl((s_1,\ldots,s_u;\sigma'),A\bigr) = Q_u^{(n)}\bigl((s_1,\ldots,s_u;\sigma''),A\bigr)$ for $(s_1,\ldots,s_u;\sigma'),(s_1,\ldots,s_u;\sigma'')\in D_u$ and $A\subset\{0,1\}^u$ whenever $\sigma'|_{\bigcup_{i=1}^u \mathbb{B}(s_i,\tilde{t}_n)} = \sigma''|_{\bigcup_{i=1}^u \mathbb{B}(s_i,\tilde{t}_n)}$.

Assumption 4 (Control of the short range positive covariances between the retention decisions given $\boldsymbol{\xi}$). For each $n\in\mathbb{N}$, let $\breve{\delta}_n : \mathbb{R}_+\to\mathbb{R}_+$ be an increasing function such that for every $t\in\mathbb{R}_+$ and for $(s_1,s_2;\sigma)\in D_2$ with $|s_1-s_2|<\sqrt{d}(t+1)$, we have

$$Q_2^{(n)}\bigl((s_1,s_2;\sigma),\{(1,1)\}\bigr) \le \bar{p}_n\breve{\delta}_n(t).$$

Assumption 5 (Control of the long range dependence between the retention decisions given $\boldsymbol{\xi}$). For each $n\in\mathbb{N}$, let $\breve{\gamma}_n : \mathbb{R}_+\to\mathbb{R}_+$ be a decreasing function such that the following property holds: for every cube $A_{\text{int}} = (a,a+h\mathbf{1})$ with $a\in\mathbb{R}_+^d, h\in(0,1]$, for $t\in\mathbb{R}_+$, for every $(s_1,\ldots,s_l,s_{l+1},\ldots,s_u;\sigma)\in D_u$ with $l\ge1$, $u>l$, $\sum_{i=1}^l \delta_{s_i} = \sigma|_{A_{\text{int}}}$ and $|s_i-s_j|>t$ for $i\in\{1,\ldots,l\}$, $j\in\{l+1,\ldots,u\}$, and for any set $B\subset\{0,1\}^{\{l+1,\ldots,u\}}$, we have

$$\begin{aligned}\Bigl|&Q_u^{(n)}\bigl((s_1,\ldots,s_l,s_{l+1},\ldots,s_u;\sigma),\{(0,\ldots,0)\}\times B\bigr)\\ &- Q_l^{(n)}\bigl((s_1,\ldots,s_l;\sigma),\{(0,\ldots,0)\}\bigr)\,Q_{u-l}^{(n)}\bigl((s_{l+1},\ldots,s_u;\sigma),B\bigr)\Bigr| \;\le\; l\,\bar{p}_n\breve{\gamma}_n(t).\end{aligned}$$

Remark 4.C. Assumptions 4 and 5 amount to the statement that, for any $t \in \mathbb{R}_+$ and for any set A_{int} of the above form, given a representation $\sum_{i=1}^{\infty} \delta_{f_i(\xi)}$ of ξ with measurable functions $f_1, f_2, \ldots$ which first enumerate all the points in A_{int} and have images $s_i := f_i(\sigma)$, and given an associated sequence (X_i) of retention decisions with respect to $Q^{(n)}$, we have with $i \neq j$

$$\mathbb{E}(X_i X_j \,|\, \xi = \sigma) \leq \bar{p}_n \breve{\delta}_n(t)$$

whenever $|s_i - s_j| < \sqrt{d}(t+1)$, and with $B \subset \{0,1\}^{\{l+1,\ldots,u\}}$

$$\left|\operatorname{cov}\left(1_{\{\sum_{i=1}^{l} X_i \geq 1\}}, 1_{\{(X_{l+1},\ldots,X_u)\in B\}} \,|\, \xi = \sigma\right)\right| \leq l\, \bar{p}_n \breve{\gamma}_n(t)$$

whenever $|s_i - s_j| > t$ for all $i \in \{1,\ldots,l\}$ and $j \in \{l+1,\ldots,u\}$, where $l = \sigma(A_{\text{int}}) \geq 0$.

We are now in the position to formulate the theorem. Again, the corresponding quantitative upper bound can be found in the proof.

Theorem 4.D (L_∞-version: upper bound and convergence). Let ξ be a point process on $\mathbb{R}_+^d$ that satisfies Assumptions 1 and 2″ above, and for each $n \in \mathbb{N}$, let $Q^{(n)} := (Q_u^{(n)})$ be an admissible sequence of retention kernels that satisfies Assumptions 3, 4, and 5.
Set $\underline{\varphi} := \inf_{\boldsymbol{k}\in\mathbb{N}^d} \mathbb{P}[\xi(C_{\boldsymbol{k}}) \geq 1] \in [0,1]$, and let $\tilde{\nu}_n(\cdot) := \mathbb{E}\big(\int_{\cdot} Q_1^{(n)}\big((s;\xi),\{1\}\big)\ \xi(ds)\big)$, which is the expectation measure of $\xi_{Q^{(n)}}$.
Then we obtain for any $m := m(n) \in \mathbb{N}$ with $m \geq 2\tilde{t}_n$

$$\begin{aligned} &d_2\big(\mathcal{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \operatorname{Po}(\tilde{\nu}_n\kappa_n^{-1}|_J)\big) \\ &\qquad = O\Big(n^d \bar{p}_n \breve{\delta}_n(0)\,,\ m^d\big(n^d \wedge \tfrac{\log^{\uparrow}(n^d\bar{p}_n)}{\underline{\varphi}\,\underline{p}_n}\big)\bar{p}_n(\bar{p}_n \vee \breve{\delta}_n(m))\,, \\ &\qquad\qquad \sqrt{n^d}\big(\sqrt{n^d} \wedge \tfrac{1}{\sqrt{\underline{\varphi}\,\underline{p}_n}}\big)\bar{p}_n\big(\breve{\beta}^{(\infty)}(m - 2\tilde{t}_n) \vee \breve{\gamma}_n(m)\big)\Big) \qquad \text{for } n \to \infty. \end{aligned}$$

The right hand side goes to 0 if, for example, $\bar{p}_n = O(1/n^d)$, $\breve{\delta}_n(0) = o(1)$, and there is a sequence $(m(n))_n$ with $m(n) \geq 2\tilde{t}_n$, $m(n) = o\big(1/\bar{p}_n^{1/d}\big)$, such that $\breve{\delta}_n(m(n)) = o\big(1/m(n)^d\big)$, $\breve{\gamma}_n(m(n)) = o(1)$, and $\breve{\beta}^{(\infty)}(m(n) - 2\tilde{t}_n) = o(1)$ for $n \to \infty$.

Remark 4.E. The main ideas of Theorem 4.D and Theorem 3.C are closely related. For example we have formulated in two different ways — once in Assumption 2 and once in Assumptions 2″, 3, and 5 — what is essentially the decreasing dependence between $(\xi|_{A_{\text{int}}}, X_1^{(n)}, \ldots, X_{\xi(A_{\text{int}})}^{(n)})$ and $(\xi|_{A_{\text{ext}}}, X_{\xi(A_{\text{int}})+1}^{(n)}, \ldots, X_{\xi(A_{\text{int}})+\xi(A_{\text{ext}})}^{(n)})$ with increasing distance between an inner set A_{int} and an outer set A_{ext} and increasing n (where $X_1^{(n)}, \ldots, X_{\xi(A_{\text{int}})}^{(n)}$ and $X_{\xi(A_{\text{int}})+1}^{(n)}, \ldots, X_{\xi(A_{\text{int}})+\xi(A_{\text{ext}})}^{(n)}$ denote the retention decisions for the points in A_{int} and in A_{ext}, respectively).

Nevertheless, there are also substantial differences between the two theorems. We mention briefly two of the more important ones, leaving the confirmation of the formal details for the reader. First, in Theorem 3.C the thinning is location-based (i.e. the location in the state space), whereas in Theorem 4.D the thinning is point-based (i.e. the point of the original process ξ), that is to say that two or more points of ξ which occupy the same location in the state space must be thinned with the same dependence of other retention decisions and with equal probabilities if we want to apply Theorem 3.B, but they can be thinned much more generally if we want to apply Theorem 4.D.

Secondly, in Theorem 4.D the conditional distribution $\mathcal{L}(X_1, \ldots, X_u \,|\, \xi = \sigma)$ is only allowed to be a function of $\sigma|_{\bigcup_{i=1}^{u} \mathbb{B}(s_i, \tilde{t}_n)}$, whereas in Theorem 3.C it may be a function of all of σ. As an example, consider the π-thinning of a homogenous Poisson process on the real half line with points $S_1 < S_2 < \ldots$, using a retention field of the form

$$\pi_n(\omega, t) := h_1(n) \exp\Big(-h_2(n) \Big/ \min_{\substack{i \in \mathbb{N} \\ S_i(\omega) > t}} \big(S_i(\omega) - t\big)\Big)$$

with functions $h_1 : \mathbb{N} \to [0,1]$ and $h_2 : \mathbb{N} \to \mathbb{R}_+$, where the idea is that $h_1(n) \to 0$ and/or $h_2(n) \to \infty$ as $n \to \infty$. Since π_n is only via ξ a function of ω, this situation is easily translated into a Q-thinning model. Obviously, Condition 3 is not satisfied by this model, because the retention probabilities may depend on arbitrarily distant regions. On the other hand, this long range dependence is very weak, and it can in fact be shown that the conditions for Theorem 3.C are met with a mixing coefficient $\breve{\beta}^{(\infty)}(t)$ that does not depend on n and that goes to zero exponentially fast as $t \to \infty$.

Proof of Theorem 4.D. Let $\eta_n \sim \mathrm{Po}(\tilde{\nu}_n)$, and choose an arbitrary $\tilde{n} \in \mathbb{N}$. We use the notation and the conventions from Subsection 3.2, replacing only ξ_{π_n} by $\xi_{Q^{(n)}}$, and define the concrete "inner" and "outer" sets needed by

$$A_{\mathrm{int}}(\boldsymbol{k}, \boldsymbol{r}) := \mathring{C}_{\boldsymbol{kr}}\,,\ A_{\mathrm{int}}^{(\tilde{t})}(\boldsymbol{k}) := \Big(\bigcup\nolimits^{n}_{\substack{\boldsymbol{j}=1 \\ |\boldsymbol{j}-\boldsymbol{k}| \leq \tilde{t}}} C_{\boldsymbol{j}}\Big)^{\circ},\ A_{\mathrm{ext}}^{(\tilde{t},t)}(\boldsymbol{k}) := \Big(\bigcup\nolimits^{n}_{\substack{\boldsymbol{j}=1 \\ |\boldsymbol{j}-\boldsymbol{k}| > \tilde{t}+t}} C_{\boldsymbol{j}}\Big)^{\circ}$$

for $\tilde{t}, t \in \mathbb{Z}_+$, $\boldsymbol{k} \in \{1, 2, \ldots, n\}^d$, and $\boldsymbol{r} \in \{1, 2, \ldots, \tilde{n}\}^d$. Most of the estimates from the proof of Theorem 3.B are still valid for the new thinning $\xi_{Q^{(n)}}$ if we condition on ξ alone instead of ξ and π_n together and replace in the bounds w_1 by $\bar{h}_1 \bar{p}_n$ and $\underline{w}_1$ by $\underline{\varphi}\,\underline{p}_n$. Thus, e.g.

$$\underline{\varphi}\,\underline{p}_n \leq \sum_{r=1}^{\tilde{n}} q_{\boldsymbol{kr}} \leq \bar{h}_1 \bar{p}_n,$$

$$\begin{aligned} d_2\big(\mathcal{L}(\xi_{Q^{(n)}} \kappa_n^{-1}|_J), \mathcal{L}(\Xi_n \kappa_n^{-1})\big) &+ d_2\big(\mathcal{L}(\mathrm{H}_n \kappa_n^{-1}), \mathcal{L}(\eta_n \kappa_n^{-1}|_J)\big) \\ &\leq \sqrt{d}/(n\tilde{n}) + \sum_{\boldsymbol{k}=1}^{n} \sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{E}\big(\xi_{Q^{(n)}}(C_{\boldsymbol{kr}}) 1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}}) \geq 2\}}\big), \end{aligned}$$

and also $M_1(\lambda)$, $M_2(\lambda)$ and $\sum_{\boldsymbol{k},\boldsymbol{r}} q_{\boldsymbol{kr}} \mathbb{E}\widetilde{Z}_{\boldsymbol{kr}}$ from the Barbour-Brown theorem follow exactly this pattern.

The remaining terms can be estimated in a very similar fashion as in the proof of Theorem 3.B, namely

$$\begin{aligned}\sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{E}\big(\xi_{Q^{(n)}}(C_{\boldsymbol{kr}}) 1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 2\}}\big) &\leq \mathbb{E}\big(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})-1)\big) \\ &\leq \mathbb{E}\biggl(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} \sum_{\substack{j=1\\ j\neq i}}^{\xi(C_{\boldsymbol{k}})} \mathbb{E}\big(X_i X_j \,|\, \xi\big)\biggr) \\ &\leq \bar{h}_2 \bar{p}_n \breve{\delta}_n(0),\end{aligned}$$

and in the Barbour-Brown theorem

$$\begin{aligned}\sum_{\boldsymbol{k},\boldsymbol{r}} \mathbb{E}(I_{\boldsymbol{kr}} Z_{\boldsymbol{kr}}) \leq \sum_{\boldsymbol{k}=1}^{n} \biggl(&\sum_{\substack{\boldsymbol{l}=1\\ 1\leq|\boldsymbol{l}-\boldsymbol{k}|\leq m}}^{n} \mathbb{E}\big(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})\, \xi_{Q^{(n)}}(C_{\boldsymbol{l}})\big) \;+\; \mathbb{E}\big(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})-1)\big)\biggr) \\ \leq \quad &\sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\ 1\leq|\boldsymbol{l}-\boldsymbol{k}|\leq m}}^{n} \mathbb{E}\biggl(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} \sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})} \mathbb{E}\big(X_i X_j \;|\; \xi\big)\biggr) \;+\; \bar{h}_2 n^d \bar{p}_n \breve{\delta}_n(0) \\ \leq\; &\bar{h}_2 (2m+1)^d n^d \bar{p}_n \breve{\delta}_n(m),\end{aligned}$$

using the same partition of the index set Θ as in Subsection 3.2.

Again we have to argue a bit more carefully for the estimation of the $e_{\boldsymbol{kr}}$. If we again set $\mathcal{F}_{\boldsymbol{k}}^{w} := \mathcal{P}\big(\{0,1\}^{\Theta_{\boldsymbol{kr}}^{w}}\big)$ and $W_{\boldsymbol{k}} := (I_{\boldsymbol{ls}})_{(\boldsymbol{l},\boldsymbol{s})\in\Theta_{\boldsymbol{kr}}^{w}}$, we obtain, in a similar fashion as in the proof of Theorem 3.B,

$$\begin{aligned} e_{\boldsymbol{kr}}/2 &= \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}} \Big|\mathrm{cov}\big(1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1\}}\,,\; 1_{\{W_{\boldsymbol{k}}\in B\}}\big)\Big| \\ &\leq \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}} \Big|\mathrm{cov}\big(\mathbb{P}\big[\xi_{Q^{(n)}}(C_{\boldsymbol{kr}}) \geq 1 \;\big|\; \xi\big]\,,\; \mathbb{P}\big[W_{\boldsymbol{k}} \in B \;\big|\; \xi\big]\big)\Big| \\ &\qquad + \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}} \Big|\mathbb{E}\Big(\mathrm{cov}\big(1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1\}}\,,\; 1_{\{W_{\boldsymbol{k}}\in B\}} \;\big|\; \xi\big)\Big)\Big| \\ &= \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}} \Big|\mathrm{cov}\big(\mathbb{P}\big[\xi_{Q^{(n)}}(C_{\boldsymbol{kr}}) \geq 1 \;\big|\; \xi|_{A_{\mathrm{int}}^{(\tilde{t}_n)}(\boldsymbol{k})}\big]\,,\; \mathbb{P}\big[W_{\boldsymbol{k}} \in B \;\big|\; \xi|_{A_{\mathrm{ext}}^{(\tilde{t}_n, m-2\tilde{t}_n)}(\boldsymbol{k})}\big]\big)\Big| \\ &\qquad + \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}} \mathbb{E}\Big|\mathrm{cov}\big(1_{\{\sum_{i=1}^{\xi(C_{\boldsymbol{kr}})} X_i\geq 1\}}\,,\; 1_{\{W_{\boldsymbol{k}}\in B\}} \;\big|\; \xi\big)\Big|, \end{aligned} \tag{4.2}$$

where this time Lemma A.F(ii) from Appendix A.4 was used. Compared with the earlier proofs, we now have a second term that is in general not zero, because this time the retention decisions need not be independent under the conditioning we have. In the first summand, we extract $\bar{p}_n|C_{\boldsymbol{kr}}|$ from the first argument of the covariance, and use Assumption $2''$, in the second summand we use Assumption 5, and obtain altogether

$$\frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \leq \bar{p}_n\big(\breve{\beta}^{(\infty)}(m-2\tilde{t}_n) + \bar{h}_1 \breve{\gamma}_n(m)\big). \tag{4.3}$$

Assembling once more the different parts, and letting $\tilde{n}$ go to infinity, gives the overall estimate

$$
\begin{aligned}
&d_2\big(\mathcal{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \mathcal{L}(\eta_n\kappa_n^{-1}|_J)\big) \\
&\qquad \leq \bar{h}_2 n^d \bar{p}_n \breve{\delta}_n(0) \\
&\qquad\quad + \Big(1 \wedge \Big[\tfrac{2}{\underline{\varphi} n^d \underline{p}_n}\big(1 + 2\log^+\big(\tfrac{\bar{h}_1}{2} n^d \bar{p}_n\big)\big)\Big]\Big)(2m+1)^d n^d \big(\bar{h}_1^2 \bar{p}_n + \bar{h}_2 \breve{\delta}_n(m)\big)\bar{p}_n \\
&\qquad\quad + \Big(1 \wedge \tfrac{1.65}{\sqrt{\underline{\varphi} n^d \underline{p}_n}}\Big) 2n^d \bar{p}_n \big(\breve{\beta}^{(\infty)}(m - 2\tilde{t}_n) + \bar{h}_1 \breve{\gamma}_n(m)\big),
\end{aligned}
$$

which is of the required order for $n \to \infty$. $\square$

4.2 Example of thinning according to point interactions (competition in a plant population)

Suppose that the individuals of a certain kind of plant that grow in a large piece $J_n \subset \mathbb{R}_+^2$ of soil are part of a point process $\xi = \sum_{i=1}^{\infty} \delta_{S_i}$ on $\mathbb{R}_+^2$ which has $\sigma(\xi)$-measurable points S_i with realizations s_i, and fulfills the Assumptions 1 and $2''$ above. As before we will do our analysis for a general dimension $d \in \mathbb{N}$. Assume that the plants have certain "fitness parameters" Ψ_i, one per plant, which are i.i.d. $(0,1]$-valued random variables, independent also of everything else and following a continuous distribution function.

Whether a given plant survives until some time t_0, depends firstly on the overall environmental conditions, which we require to be the same for all plants (say, each plant has a basic survival probability $q_0 = q_0^{(n)}$), and secondly on the influence of other plants in its immediate surroundings. Suppose that the competition is such, that an individual plant survives it, independently of whether it survives the environmental effect, if there are no plants with a higher degree of fitness within a radius of $r_n > 0$. We model this situation by assuming that we have for each S_i a retention decision

$$X_i := X_i^{(n)} := Y_i^{(n)} Z_i^{(n)},$$

where the $Y_i^{(n)}$ are i.i.d. $\mathrm{Be}(q_0)$-random variables that are independent of everything else and determine survival due to the environmental effect, and the $Z_i^{(n)}$ are defined by

$$Z_i^{(n)} := \prod_{\substack{j=1 \\ |S_j - S_i| \leq r_n}}^{\infty} 1_{\{\Psi_i \geq \Psi_j\}}$$

and determine survival due to the competition effect. This second thinning effect is the same one used for the construction of the Matern hard core process (see Stoyan, Kendall and Mecke (1987), Section 5.4). We obtain from symmetry considerations

$$\mathbb{E}(X_i \,|\, \xi = \sigma) = q_0 / \sigma\big(\overline{\mathbb{B}}(s_i, r_n)\big), \tag{4.4}$$

and for $i \neq j$,

$$\mathbb{E}(X_i X_j \,|\, \xi = \sigma) = \begin{cases} 0 & \text{if } |s_i - s_j| \leq r_n, \\ \Big(1 + \frac{\sigma(\overline{\mathbb{B}}(s_i,r_n) \cap \overline{\mathbb{B}}(s_j,r_n))}{\sigma(\overline{\mathbb{B}}(s_i,r_n) \cup \overline{\mathbb{B}}(s_j,r_n)}\Big) q_0^2 \frac{1}{\sigma(\overline{\mathbb{B}}(s_i,r_n))} \frac{1}{\sigma(\overline{\mathbb{B}}(s_j,r_n))} & \text{if } |s_i - s_j| > r_n. \end{cases} \tag{4.5}$$

For the second result a bit of computation is necessary, which basically consists of counting the number of orderings of the fitness values of the individuals in $\overline{\mathbb{B}}(s_i, r_n) \cup \overline{\mathbb{B}}(s_j, r_n)$ that leave the values of s_i and s_j as highest in their respective "competition ball". Note that for $|s_i - s_j| > 2r_n$ the retention decisions X_i and X_j are independent given ξ, because the "competition balls" $\overline{\mathbb{B}}(s_i, r_n)$ and $\overline{\mathbb{B}}(s_j, r_n)$ are disjoint. By Remark 4.A we can define probability kernels by

$$Q_u^{(n)}\big((s_1, \ldots, s_u; \sigma), \{(e_1, \ldots, e_u)\}\big) := \mathbb{P}[X_1 = e_1, \ldots X_u = e_u \,|\, \xi = \sigma]$$

for almost every σ, in such a way that $(Q_u^{(n)})_{u \in \mathbb{N}}$ is an admissible family of retention kernels for the sequence $(X_i)_{i \in \mathbb{N}}$.

In view of Theorem 4.D, Condition 3 is satisfied with $\tilde{t}_n = r_n$, and for $m(n) \geq 2r_n$ we can always set $\breve{\gamma}_n(m(n))$ to zero in Condition 5. Furthermore, we have by Equation (4.4)

$$\bar{p}_n = q_0 \Big\| \sup_{s \in D_1(\xi) \cap J_n} 1/\xi(\overline{\mathbb{B}}(s, r_n)) \Big\|_{L_\infty},$$

and obtain, by (4.5) for $\breve{\delta}_n$ in Assumption 4, the possible choice of

$$\breve{\delta}_n :\equiv 2q_0 \Big\| \sup_{s \in D_1(\xi) \cap J_n} 1/\xi(\overline{\mathbb{B}}(s, r_n)) \Big\|_{L_\infty}.$$

Obviously, for most point processes ξ, the norms on the right hand side of the above equations are one, and we then do not get very interesting upper bounds for the d_2-distance between the law of the thinned process and a Poisson process law, because the bounds take into account only the environmental thinning effect and not the competition effect.

There are two ways in which this can be rescued. First, we could use the L_1-version of Theorem 4.D (i.e. the Q-thinning analogue of Corollary 3.E), which in this case is much more promising, because we have expectations instead of the L_∞-norms above. However, as we have not formulated this more involved version, we do not pursue this idea here any further. Secondly, we can look at a specific ξ-process where we can be sure of having a certain number of competitors in every competition ball: let ξ be the point process on all of $\mathbb{R}^d$ (in order to avoid edge effects) that has in every unit square $[\boldsymbol{k}, \boldsymbol{k}+\boldsymbol{1})$, $\boldsymbol{k} \in \mathbb{Z}^d$, exactly one point, which is uniformly distributed over the square and independent of the locations of all the other points. We might think of a gardener who sows seeds, exactly one per square, by just throwing each carelessly over its square; the fact that the distribution over each square is the uniform is by no means crucial to the essence of the following explanations.

We then have, for all $s \in J_n$,

$$\xi\big(\overline{\mathbb{B}}(s,r_n)\big) \geq \big((r_n-1)/\sqrt{2}\big)^d \geq (r_n/2)^d$$

for $r_n \geq 4$, and

$$\xi\big(\overline{\mathbb{B}}(s,r_n)\big) \leq \big(2(r_n+1)\big)^d \leq (3r_n)^d$$

for $r_n \geq 2$. Therefore, in the notation of Theorem 4.D,

$$\bar{p}_n \leq 2^d\big(q_0/r_n^d\big)\,, \quad \underline{p}_n \geq (1/3)^d\big(q_0/r_n^d\big)\,, \text{ and } \quad \breve{\delta}_n :\equiv 2\bar{p}_n \leq 2^{d+1}q_0/r_n^d.$$

Furthermore, we obviously have

$$\mathbb{E}\xi = Leb^d,$$

and we can calculate for the second factorial moment measure

$$\mu_2(A \times B) \leq Leb^{2d}(A \times B) \quad \text{for all } A, B \in \mathcal{B}^d,$$

such that $\mu_2(C) \leq 1$ for every unit cube $C \subset \mathbb{R}^{2d}$. Finally, the mixing property $2''$ is also met with $\breve{\beta}^{(\infty)}(t) = 0$ for $t \geq 1$. With $m(n) := \lceil 2r_n \rceil + 1$, this yields the following result.

Proposition 4.F. Let $U_{\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^d$, be independent random variables that are uniformly distributed on $[0,1)^d$, and $\xi := \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \delta_{\boldsymbol{k}+U_{\boldsymbol{k}}}$. Furthermore let $(S_i)_{i\in\mathbb{N}}$ be a $\sigma(\xi)$-measurable enumeration of the points of ξ, and $(\Psi_i)_{i\in\mathbb{N}}$ an i.i.d. sequence of continuous, $(0,1]$-valued random variables independent of ξ. Define retention decisions $X_i := X_i^{(n)} := Y_i^{(n)} Z_i^{(n)}$, where $Y_i^{(n)}$, $i \in \mathbb{N}$, are i.i.d. Bernoulli random variables with parameter $q_0 = q_0^{(n)} \in [0,1]$ that are independent of everything else, and $Z_i^{(n)}$, $i \in \mathbb{N}$, are given by

$$Z_i^{(n)} := \prod_{\substack{j=1 \\ |S_j - S_i| \leq r_n}}^{\infty} 1_{\{\Psi_i \geq \Psi_j\}}$$

for some $r_n > 0$.
We then have that the point process $\sum_{i=1}^{\infty} X_i \delta_{S_i}$ can be constructed as a Q-thinning $\xi_{Q^{(n)}}$ of ξ for an appropriate sequence $Q = Q^{(n)}$, and obtain

$$d_2\big(\mathcal{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \mathrm{Po}(\tilde{\nu}_n\kappa_n^{-1}|_J)\big) \;=\; O\Big(n^d\big(q_0/r_n^d\big)^2\,,\; q_0\Big(n^d\big(q_0/r_n^d\big) \wedge \log^{\uparrow}\big(n^d\big(q_0/r_n^d\big)\big)\Big)\Big) \quad \text{for } n \to \infty$$

with

$$\tilde{\nu}_n(\cdot) = \mathbb{E}\Big(\int_{\cdot} Q_1^{(n)}\big((s,\xi),\{1\}\big)\,\xi(ds)\Big) = q_0\mathbb{E}\Big(\int_{\cdot} \frac{\xi(ds)}{\xi(\overline{\mathbb{B}}(s,r_n))}\Big).$$

Note that

$$\tilde{\nu}_n\kappa_n^{-1}(J) = q_0\mathbb{E}\Big(\int_{J_n} \frac{\xi(ds)}{\xi(\overline{\mathbb{B}}(s,r_n))}\Big) \asymp n^d\big(q_0/r_n^d\big) \quad \text{for } n \to \infty,$$

and therefore, in view of obtaining a stable Poisson process, $r_n \asymp nq_0^{1/d}$ is a natural choice, with q_0 not going faster to zero than $1/n^d$. Then

$$d_2\big(\mathcal{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \mathrm{Po}(\tilde{\nu}_n\kappa_n^{-1}|_J)\big) = O\big(q_0\big).$$

Suppose that we would like to compare the distribution of $\xi_{Q^{(n)}}\kappa_n^{-1}|_J$ to the fixed $\mathrm{Po}(Leb^d|_J)$-distribution (any stationary intensity measure $\mu_0 Leb^d$ can be used as long as r_n is adapted accordingly). This requires a more precise calculation of $\tilde{\nu}\kappa_n^{-1}$. For any set $A \in \mathcal{B}_J$, we obtain

$$\tilde{\nu}_n\kappa_n^{-1}(A) = q_0\mathbb{E}\Big(\int_{nA} \frac{\xi(ds)}{\xi(\overline{\mathbb{B}}(s,r_n))}\Big) = q_0\mathbb{E}\Big(\int_{nA} \frac{\xi(ds)}{\alpha_d r_n^d + \varepsilon_{d,n}(s,\xi)}\Big),$$

where $\alpha_d = \pi^{d/2}/\Gamma(d/2+1)$ is the volume of the d-dimensional unit ball and $\varepsilon_{d,n}(s,\xi)$ is the error we make by just setting $\xi(\overline{\mathbb{B}}(s,r_n)) = Leb^d(\overline{\mathbb{B}}(s,r_n))$. Note that

$$\big|\varepsilon_{d,n}(s,\xi)\big| \leq 3^d\omega_d r_n^{d-1} \quad \text{a.s.}$$

with $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$, the surface of the d-dimensional unit ball. Hence

$$\begin{aligned}\frac{|A|}{\alpha_d}n^d\frac{q_0}{r_n^d}\Big(\frac{\alpha_d r_n^d}{\alpha_d r_n^d + 3^d\omega_d r_n^{d-1}}\Big) &\leq \tilde{\nu}_n\kappa_n^{-1}(A) \\ &\leq \frac{|A|}{\alpha_d}n^d\frac{q_0}{r_n^d}\Big(\frac{\alpha_d r_n^d}{(\alpha_d r_n^d - 3^d\omega_d r_n^{d-1}) \vee 0}\Big).\end{aligned}$$

Thus, if $q_0 = \omega(1/n^d)$ (i.e. $n^d q_0 \to \infty$) for $n \to \infty$, and if we set $r_n := n\big((1/\alpha_d)q_0\big)^{1/d}$ for all $n \in \mathbb{N}$, we obtain

$$\tilde{\nu}_n\kappa_n^{-1}(A) \longrightarrow Leb^d(A) \quad \text{for } n \to \infty$$

for all $A \in \mathcal{B}_J$. Hence

$$\mathrm{Po}(\tilde{\nu}_n\kappa_n^{-1}|_J) \xrightarrow{\mathcal{D}} \mathrm{Po}(Leb^d|_J),$$

and more exactly, by an application of Inequality (A.3) from Appendix A.2 (using $d_W \leq d_{TV}$, which follows from Formula (1.2) in Appendix A.1 of Barbour, Holst and Janson (1992)),

$$\begin{aligned}d_2\big(\mathrm{Po}(\tilde{\nu}_n\kappa_n^{-1}|_J), \mathrm{Po}(Leb^d|_J)\big) &\leq \big|\tilde{\nu}_n\kappa_n^{-1}(J) - 1\big| + \sup_{A\in\mathcal{B}_J}\Big|\frac{\tilde{\nu}_n\kappa_n^{-1}(A)}{\tilde{\nu}_n\kappa_n^{-1}(J)} - Leb^d(A)\Big| \\ &\leq \frac{3^d\omega_d r_n^{d-1}}{(\alpha_d r_n^d - 3^d\omega_d r_n^{d-1}) \vee 0} + \frac{2\cdot 3^d\omega_d r_n^{d-1}}{(\alpha_d r_n^d - 3^d\omega_d r_n^{d-1}) \vee 0} \\ &= O\big(1/r_n\big) \qquad \text{for } n \to \infty.\end{aligned}$$

Thus we have proved the following result.

Corollary 4.G. Under the conditions of Proposition 4.F, with $q_0 = \omega(1/n^d)$ for $n \to \infty$ and $r_n := n\big((1/\alpha_d)q_0\big)^{1/d}$ for every $n \in \mathbb{N}$, we have that

$$d_2\big(\mathcal{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \mathrm{Po}(Leb^d|_J)\big) = O\big(1/\big(nq_0^{1/d}\big)\ ,\ q_0\big) \quad \text{for } n \to \infty.$$

Appendix

We give here an overview of some auxiliary results needed in this article. The Barbour-Brown theorem, which we make vital use of in our main proofs, is stated in Section A.1, along with a slight improvement for a special case, needed for Proposition 1.B. Sections A.2 to A.5 deal with additional proofs and technical details that were left out in the main text for better readability.

A.1 Theorems obtained by Stein's method

Let Θ be any finite non-empty index set, and $(I_k)_{k\in\Theta}$ a sequence of indicator random variables with a local dependence property, that is for every $k \in \Theta$ the set $\Theta_k := \Theta \setminus \{k\}$ can be partitioned as $\Theta_k = \Theta_k^s \,\dot\cup\, \Theta_k^w$ into a set Θ_k^s of indices l, for which I_l depends "strongly" on I_k, and a set Θ_k^w of indices l, for which I_l depends "weakly" on I_k. Herein, the terms "strongly" and "weakly" are not meant as a restriction to the partition of Θ_k, but serve only illustrative purposes. The same holds true for the term "local dependence", which does not have to possess any representation in the spatial structure of Θ (though it always does in our applications). We now write $Z_k := \sum_{l\in\Theta_k^s} I_l$, $\widetilde{Z}_k := Z_k + I_k$, $q_k := \mathbb{E}I_k$ for every $k \in \Theta$, and set $W := \sum_{k\in\Theta} I_k$ and $\lambda := \mathbb{E}W = \sum_{k\in\Theta} q_k$. To exclude a trivial case, assume that $\lambda > 0$. Furthermore, we choose arbitrary points $(\alpha_k)_{k\in\Theta}$ in any desired complete, separable metric space $(\mathcal{X}, d_0)$ with $d_0 \leq 1$ and set $\Xi := \sum_{k\in\Theta} I_k\delta_{\alpha_k}$, $\boldsymbol{\lambda} := \sum_{k\in\Theta} q_k\delta_{\alpha_k}$.

By applying a natural generalization of the Stein-Chen method, the following result is obtained (cf. Barbour, Holst and Janson (1992), Theorem 10.F).

Theorem A.A (Barbour and Brown (1992)). With the above definitions, we have

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})\big) \leq M_1(\lambda)\sum_{k\in\Theta}\big(q_k\mathbb{E}\widetilde{Z}_k + \mathbb{E}(I_kZ_k)\big) + M_2(\lambda)\sum_{k\in\Theta} e_k,$$

where

$$M_1(\lambda) = 1 \wedge \Big[\frac{2}{\lambda}\Big(1 + 2\log^+\Big(\frac{\lambda}{2}\Big)\Big)\Big], \quad M_2(\lambda) = 1 \wedge \Big[1.65\frac{1}{\sqrt{\lambda}}\Big],$$

and

$$e_k = \mathbb{E}\big|\mathbb{E}\big(I_k \bigm| (I_l\,;\ l \in \Theta_k^w)\big) - q_k\big| = 2\max_{B\in\sigma(I_l\,;\ l\in\Theta_k^w)}\big|\mathrm{cov}(I_k, 1_B)\big|.$$

Remark A.B. Note that the upper bound in Theorem A.A depends neither on the points α_k, $k \in \Theta$, nor on the specific choice of the metric d_0, as long as it is bounded by 1.

In Brown, Weinberg and Xia (2000), a method was proposed to dispose of the factor $\log^+(\lambda/2)$ at some cost (compare also Brown and Xia (2001), Section 5, where the estimates needed were considerably simplified). Since this requires a more specialized treatment of the terms obtained in the upper bound, and since the logarithm above is negligible for almost all practical purposes, we do not pursue this method here much further. We only give an adaptation of Proposition 4.1 from Brown, Weinberg and Xia (2000), which is needed for the proof of Proposition 1.B, where a logarithmic term would be slightly annoying.

Proposition A.C. Suppose that $(I_k)_{k\in\Theta}$ is an independent sequence of indicators, and set $q_k := \mathbb{E}I_k$, $\Xi := \sum_k I_k\delta_{\alpha_k}$, $\boldsymbol{\lambda} := \sum_k q_k\delta_{\alpha_k}$, and $\lambda := \sum_k q_k > 0$ as above. Then

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})\big) \leq \frac{8.45}{\lambda}\sum_{k\in\Theta} q_k^2 \leq 8.45 \max_{k\in\Theta} q_k.$$

Proof. Suppose that $\lambda > 10$. By applying Theorem 3.1 from Brown, Weinberg and Xia (2000) in the same way as the authors did to obtain the first inequality of Proposition 4.1, we get

$$d_2(\mathcal{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})) \leq \sum_{k\in\Theta} q_k^2\Big(\frac{5.65}{\lambda} + \frac{1.65}{\lambda^{1.5}} + \frac{0.5}{\lambda^2} + \frac{2}{\lambda - q_k}\Big).$$

Note that this estimate does not need any requirements on the underlying space (i.e. the positions of the points α_k and the metric d_0), or on the individual q_k, and the only requirement on λ is that $\lambda > q_k$ for every k. Using $q_k \leq 1 < \lambda/10$ for the last term, we thus obtain

$$d_2(\mathcal{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})) \leq \frac{8.45}{\lambda}\sum_{k\in\Theta} q_k^2.$$

For $\lambda \leq 10$ we simply use Theorem A.A with $\Theta_k^s = \emptyset$ and $\Theta_k^w = \Theta_k$ to obtain

$$d_2(\mathcal{L}(\Xi), \mathrm{Po}(\boldsymbol{\lambda})) \leq M_1(\lambda)\sum_{k\in\Theta} q_k^2 \leq \frac{8.45}{\lambda}\sum_{k\in\Theta} q_k^2.$$

□

A.2 Proof of Proposition 1.B

We use the notation and some results from Section 2. Let $n \in \mathbb{N}$ be fixed, and let $\sum_{i=1}^{\xi(\mathbb{R}_+^d)} \delta_{S_i}$ be a representation of ξ with $\sigma(\xi)$-measurable random elements $S_1, S_2, \ldots$ numbered in such a way that all the points in $\kappa_n^{-1}(K)$ come first. Let $(X_i)_{i\in\mathbb{N}}$, $(Y_i)_{i\in\mathbb{N}}$ be sequences of random variables such that, given ξ and π_n, the X_i are independent indicators with expectations $\pi_n(S_i)$, and the Y_i are independent and $\mathrm{Po}(\pi_n(S_i))$-distributed (i.e. the X_i are retention decisions as usual, and the Y_i are "Poissonized" variants of retention decisions). We have

$$\begin{aligned} &d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_K), \mathrm{Cox}(\Lambda|_K)\big) \\ &\qquad \leq d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_K), \mathrm{Cox}(\Lambda_n|_K)\big) + d_2\big(\mathrm{Cox}(\Lambda_n|_K), \mathrm{Cox}(\Lambda|_K)\big). \end{aligned}$$

Choosing an arbitrary $\mathrm{Cox}(\Lambda_n)$-distributed point process η_n, the first term is estimated as

$$\begin{aligned}
&d_2\big(\mathcal{L}(\xi_{\pi_n}\kappa_n^{-1}|_K), \mathrm{Cox}(\Lambda_n|_K)\big) \\
&\qquad \leq \mathbb{E}\Big(\sup_{f\in\mathcal{F}_2}\big|\mathbb{E}\big(f(\xi_{\pi_n}\kappa_n^{-1}|_K) - f(\eta_n|_K) \bigm| \xi, \pi_n\big)\big|\Big) \\
&\qquad = \mathbb{E}\Big(d_2\Big(\mathcal{L}\big(\textstyle\sum_{i=1}^{\xi\kappa_n^{-1}(K)} X_i\delta_{\kappa_n(S_i)} \bigm| \xi, \pi_n\big), \mathcal{L}\big(\textstyle\sum_{i=1}^{\xi\kappa_n^{-1}(K)} Y_i\delta_{\kappa_n(S_i)} \bigm| \xi, \pi_n\big)\Big)\Big) \\
&\qquad \leq 8.45\,\mathbb{E}\Big(\sup_{s\in\kappa_n^{-1}(K)} \pi_n(s)\Big),
\end{aligned} \tag{A.1}$$

where for the last line we used the fact that, given ξ and π_n, each of the sequences (X_i) and (Y_i) is independent, so that we can apply Proposition A.C.

For the second term, we just use an upper bound for the d_2-distance between two general Cox processes. Let Λ and M be arbitrary finite random measures on K, and write $\Phi(\lambda,\mu) := d_2\big(\mathrm{Po}(\lambda), \mathrm{Po}(\mu)\big)$ for finite measures λ, μ on K. Then, by conditioning on Λ and M,

$$d_2\big(\mathrm{Cox}(\Lambda), \mathrm{Cox}(\mathrm{M})\big) \leq \mathbb{E}\big(\Phi(\Lambda, \mathrm{M})\big). \tag{A.2}$$

Inequality (2.8) in Brown and Xia (1995) tells us that

$$\Phi(\lambda,\mu) \;\leq\; (1 - e^{-\min(\lambda(K),\mu(K))})d_W\big(\tfrac{\lambda}{\lambda(K)}, \tfrac{\mu}{\mu(K)}\big) + \min\big(1, \tfrac{1.65}{\sqrt{\lambda(K)}}, \tfrac{1.65}{\sqrt{\mu(K)}}\big)|\lambda(K) - \mu(K)| \tag{A.3}$$

(note that the first d_1 in Inequality (2.8) should be d_2; also, there is no formal difference between the two measures in Inequality (2.8), so they can be exchanged in the upper bound). Combining of the estimates (A.1), (A.2), and (A.3) yields the desired result.

A.3 Locally evaluable random fields

In Section 2, the notion of a locally evaluable random field was introduced. Essentially, any measurable random field which owes its measurability to some local feature (such as continuity of paths) has this property. In what follows we show local evaluability for two important classes of random fields. We first give the corresponding definitions.

Definition. Let $f : \mathbb{R}_+^d \to [0,1]$ be a function, and $\pi := (\pi(\cdot, s); s \in \mathbb{R}_+^d)$ a $[0,1]$-valued random fields on $\mathbb{R}_+^d$.

(a) Let $A \subset \mathbb{R}_+^d$ be a closed convex cone, that is a closed convex set for which $v \in A$ implies $\alpha v \in A$ for every $\alpha \in \mathbb{R}_+$. We call f *continuous from A at a point $s \in \mathbb{R}_+^d$* if $\lim_{n\to\infty} f(s_n) = f(s)$ for any sequence (s_n) in $s + A$ with $s_n \to s$ $(n \to \infty)$. We call f *continuous from A* if it is continuous from A at every point.

(b) We say f is *lower semicontinuous at a point $s \in \mathbb{R}_+^d$* if $\liminf_{n\to\infty} f(s_n) \geq f(s)$ for every sequence (s_n) in $\mathbb{R}_+^d$ with $s_n \to s$ $(n \to \infty)$. We say f is *upper semicontinuous at s* if $-f$ is lower semicontinuous at s. Furthermore f is called *lower [resp. upper] semicontinuous* if it is lower [resp. upper] semicontinuous at every point.

(c) Let $\mathcal{C}$ be a class of subsets of $\mathbb{R}$. We call π *separable for* $\mathcal{C}$ if there exists a countable set $\Sigma \subset \mathbb{R}_+^d$, and a fixed set $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$, such that for any set $C \in \mathcal{C}$ and for any rectangle R that is open in $\mathbb{R}_+^d$, we have

$$\big\{\omega \in \Omega;\ \pi(\omega, s) \in C, s \in R\big\} \,\Delta\, \big\{\omega \in \Omega;\ \pi(\omega, s) \in C, s \in R \cap \Sigma\big\} \subset N.$$

In this case we call Σ a *separant* for π. We call π *fully separable for* $\mathcal{C}$ and accordingly Σ a *full separant* for π if the above property holds with $N = \emptyset$. Accordingly, every separable random field π can be made fully separable by adjustment on a set of probability zero. Note that such an adjustment does not change the distribution of the thinning ξ_π.

Proposition A.D. The random field $\pi := (\pi(\cdot, s); s \in \mathbb{R}_+^d)$ is locally evaluable under each of the following conditions:

(a) There is a closed convex cone $A \subset \mathbb{R}_+^d$ of positive volume, such that the paths $\pi(\omega, \cdot)$ are all continuous from A;

(b) the paths $\pi(\omega, \cdot)$ are all lower semicontinuous, and π is fully separable with respect to the class $\{(y, \infty);\ y \in \mathbb{R}\}$;

(c) the paths $\pi(\omega, \cdot)$ are all upper semicontinuous, and π is fully separable with respect to the class $\{(-\infty, y);\ y \in \mathbb{R}\}$.

Remark A.E. A special case of a random field with all upper semicontinuous paths is the indicator of a random closed set (RACS) F. It is straightforward to see that such an indicator is separable with respect to the class $\{(-\infty, y);\ y \in \mathbb{R}\}$, if and only if F is a separable RACS, meaning that there is a countable, dense subset Σ of $\mathbb{R}_+^d$ such that $F = \overline{F \cap \Sigma}$ a.s. Hence Proposition A.D guarantees us, if $\pi(\omega, s) := 1_{F(\omega)}(s)$ for a separable RACS F, that by adjusting π on a null set, we obtain a locally evaluable random field. For more details on the relationship between indicator random fields and RACS see Matheron (1975), Section 2-4.

Proof of Proposition A.D. (a) Take d linearly independent vectors $v_1, v_2, \ldots, v_d \in A$ of unit length, and let $E := \{\sum_{i=1}^d \alpha_i v_i;\ \alpha_i \in (0, 1]\}$ be the half-open parallelepiped spanned by these vectors. Define the lattice $L^{(N)} := \{2^{-N} \sum_{i=1}^d k_i v_i\,;\ k_i \in \mathbb{Z}\}$, and denote by $E_v^{(N)}$ the shifted copy of $2^{-N}E$ with the maximal corner in $v \in L^{(N)}$ (i.e. of all the corners $2^{-N} \sum_{i=1}^d k_i v_i$ of $E_v^{(N)}$ the corner v maximizes the coefficients $k_1, \ldots, k_d$). Note that $\{E_v^{(N)};\ v \in L^{(N)}\}$ is, for any N, a partition of $\mathbb{R}^d$. We may now, for any open set $R \subset \mathbb{R}_+^d$, approximate π_R^* by the random fields π_N given as

$$\pi_N(\omega, s) := \sum_{v \in L^{(N)} \cap R} \pi(\omega, v) 1_{E_v^{(N)}}(s)$$

for every $\omega \in \Omega$ and every $s \in R$. It is easy to see that π_N is $(\sigma(\pi|_R) \otimes \mathcal{B}_R)$-measurable. Furthermore, π_N converges pointwise to π_R^* for $N \to \infty$, because for any (ω, s), the sequence $(\pi_N(\omega, s))_N$ has (with the possible exception of the first finitely many elements) the same values as $(\pi(\omega, s_N))_N$, where s_N are defined via the condition $s \in E_{s_N}^{(N)}$ for every $N \in \mathbb{N}$, such that $s_N \to s$ from the cone A. Hence, π_R^* is $(\sigma(\pi|_R) \otimes \mathcal{B}_R)$-measurable, which completes the proof of part (a).

(b) We show that $(\pi_R^*)^{-1}((y, \infty)) \in \sigma(\pi|_R) \otimes \mathcal{B}_R$ for any rectangle R that is open in $\mathbb{R}_+^d$ and for any $y \in [0, 1)$. Since π is pathwise lower continuous, we get for all $\omega \in \Omega$ that $(\pi(\omega, \cdot))^{-1}((y, \infty)) \cap R$ is open. Let Σ be a full separant for π, and set $\mathcal{G} := \{(a, b) \subset R;\, a, b \in \mathbb{Q}^d, a < b\}$. We then can write

$$(\pi_R^*)^{-1}((y, \infty)) = \bigcup_{G \in \mathcal{G}} \bigcap_{s \in G \cap \Sigma} (\pi(\cdot, s))^{-1}((y, \infty)) \times G$$

for any $y \in [0, 1)$, which lies in $\sigma(\pi|_R) \otimes \mathcal{B}_R$, because the union and the intersection are countable.

(c) Apply (b) to $1 - \pi$.

$\square$

A.4 Thinnings are determined locally

We show here that "local conditioning" is enough to determine the "local distribution" of the thinnings, a result which is required for the estimation of the $e_{\boldsymbol{kr}}$-terms in both Section 3 and Section 4. For any set $A \subset \mathbb{R}_+^d$ write

$$\mathfrak{N}(A) := \sigma\big(\big\{\{\varrho \in \mathcal{N}; \varrho(B) = l\}\,;\ B \in \mathcal{B}_A, l \in \mathbb{Z}_+\big\}\big).$$

We then obtain

Lemma A.F. Suppose that ξ is a point process on $\mathbb{R}_+^d$.

(i) **For the $\boldsymbol{\pi}$-thinning:**
Let π be a locally evaluable $[0, 1]$-valued random field on $\mathbb{R}_+^d$. Then we have for any bounded open set $A \subset \mathbb{R}_+^d$, and for any $D \in \mathfrak{N}(A)$, that

$$\mathbb{P}\big[\xi_\pi \in D \bigm| \xi, \pi\big] = \mathbb{P}\big[\xi_\pi \in D \bigm| \xi|_A, \pi|_A\big] \quad \text{a.s.}$$

(ii) **For the $\boldsymbol{Q}$-thinning:**
Let $(Q_u)_{u \in \mathbb{N}}$ be an admissible sequence of retention kernels which satisfy Assumption 3 from Section 4. Then we have for any bounded measurable set $A \subset \mathbb{R}_+^d$, and for any $D \in \mathfrak{N}(A)$, that

$$\mathbb{P}\big[\xi_Q \in D \bigm| \xi\big] = \mathbb{P}\big[\xi_Q \in D \bigm| \xi|_{H^{(\tilde{t}_n)}(A)}\big] \quad \text{a.s.,}$$

where $H^{(\tilde{t}_n)}(A) := \{s \in \mathbb{R}_+^d;\, \exists s' \in A \text{ s.t. } |s' - s| \leq \tilde{t}_n\}$ denotes the $\tilde{t}_n$-halo set for A.

Proof. We only prove (i), since the proof of (ii) is very similar, requiring only a few and very obvious changes in notation. Let $A \subset \mathbb{R}_+^d$ be bounded and open, and let $D \in \mathfrak{N}(A)$. We show that $P_{(\xi,\pi)}(D)$, the version of $\mathbb{P}\big[\xi_\pi \in D \bigm| \xi, \pi\big]$ used for the π-thinning definition in Section 2, is $\sigma(\xi|_A, \pi|_A)$-measurable, whence the statement follows.

Let $k \in \mathbb{N}$, $B_1, \ldots B_k \in \mathcal{B}_A$ be pairwise disjoint, and $l_1, \ldots, l_k \in \mathbb{Z}_+$. Choose a representation $\sum_{i=1}^{\xi(A)} \delta_{S'_i}$ of $\xi|_A$ with $\sigma(\xi|_A)$-measurable random elements $S'_1, S'_2, \ldots$, numbered "B_i-wise" (first all points in B_1, then all points in B_2, a.s.o., until B_k). Any such representation can easily be extended to a representation $\sum_{i=1}^{\xi(\mathbb{R}_+^d)} \delta_{S_i}$ of ξ with $\sigma(\xi)$-measurable random elements $S_1, S_2, \ldots$, numbered "B_i-wise". We then have for $D = \{\sigma \in \mathcal{N}; \sigma(B_1) = l_1, \ldots, \sigma(B_k) = l_k\}$, setting $V_j := \sum_{i=1}^{j} \xi(B_i)$ for $j \in \{0, 1, \ldots, k\}$ and $t_0 := 0$,

$$
\begin{aligned}
P_{(\xi,\pi)}(D) \;=\; \sum_{\substack{t_1,\ldots,t_k=0\\ t_1\leq\ldots\leq t_k}}^{\infty} & \mathrm{I}\big[V_j = t_j\,;\ 1 \leq j \leq k\big] \cdot \\
& \sum_{e_1,\ldots,e_{t_k} \in \{0,1\}} \mathrm{I}\big[\textstyle\sum_{i=t_{j-1}+1}^{t_j} e_i = l_j\,;\ 1 \leq j \leq k\big] \displaystyle\prod_{i=1}^{t_k} \pi(S_i)^{e_i} \big(1 - \pi(S_i)\big)^{1-e_i},
\end{aligned}
$$

which by $\pi(S_i) \equiv \pi|_A(S'_i)$, and the local evaluability of π, is obviously $\sigma(\xi|_A, \pi|_A)$-measurable (note that the measurability property required for local evaluability extends easily to arbitrary bounded open sets, instead of only bounded open rectangles). Since any set of the form $D = \{\sigma \in \mathcal{N}; \sigma(\widetilde{B}_1) = l_1, \ldots, \sigma(\widetilde{B}_k) = l_k\}$ with arbitrary $\widetilde{B}_1, \ldots \widetilde{B}_k \in \mathcal{B}_A$ can be written as a finite disjoint union of sets of the form $\{\sigma \in \mathcal{N}; \sigma(B_1) = l_1, \ldots, \sigma(B_k) = l_k\}$ with pairwise disjoint $B_1, \ldots B_k \in \mathcal{B}_A$, we obtain immediately the $\sigma(\xi|_A, \pi|_A)$-measurability of $P_{(\xi,\pi)}(D)$ for sets D of this new form. The measurability for general D is now obtained by a standard extension argument. $\square$

A.5 Simplification for the retention kernels in the Q-thinning (Proof of Remark 4.A)

We use the notation from Section 4. Let $(f_i)_{i\in\mathbb{N}}$ be a sequence of $\mathfrak{N}$-$\mathcal{B}_+^d$-measurable functions, and suppose that $Q_u\big((f_1(\sigma), \ldots, f_u(\sigma); \sigma), A\big)$ is defined for every $\sigma \in \mathcal{N}$, every $A \subset \{0,1\}^u$, and every $u \in \mathbb{N}$, subject to the same conditions as in Remark 4.A. For $u \in \mathbb{N}$ and $(s_1, s_2, \ldots, s_u; \sigma) \in D_u$, choose pairwise different $i_1, i_2, \ldots, i_u \in \mathbb{N}$ with $f_{i_j}(\sigma) = s_j$, $1 \leq j \leq u$, and let $\bar{u} := \max\{i_j;\, 1 \leq j \leq n\}$. By Properties (a) and (b) it is clear that if an extension $\widetilde{Q}_u$ of Q_u with the required features exists, it must have the form

$$
\widetilde{Q}_u\big((s_1, \ldots, s_u; \sigma), A\big) = Q_{\bar{u}}\big((f_1(\sigma), \ldots, f_{\bar{u}}(\sigma); \sigma), T_2(A \times \{0,1\}^{\bar{u}-u})\big),
$$

where $A \subset \{0,1\}^u$ and $T_2 : \{0,1\}^{\bar{u}} \to \{0,1\}^{\bar{u}}$ is the transformation that belongs (in the sense of Property (a) from the Q-thinning definition) to an arbitrary permutation τ on $\{1, 2, \ldots, \bar{u}\}$ with $\tau(i_j) := j$ for $1 \leq j \leq u$. It is easy to see, that the $\widetilde{Q}_u$ are well-defined, extend the Q_u, and satisfy Properties (a) and (b) from the Q-thinning definition. In order to obtain the measurability in the first argument, note that for any $A \subset \{0,1\}^u$ and any measurable $B \subset [0,1]$,

$$\big(\widetilde{Q}_u(\cdot, A)\big)^{-1}(B) = \bigcup_{\substack{i_1,\ldots,i_u \in \mathbb{N} \\ i_j \neq i_k \text{ for } j \neq k}} \big\{\big(f_{i_1}(\sigma), \ldots, f_{i_u}(\sigma); \sigma\big)\,;\ \sigma \in \varphi^{-1}_{i_1,\ldots,i_u}(B)\big\},$$

where $\varphi_{i_1,\ldots,i_u}(\sigma) := \widetilde{Q}_u\big((f_{i_1}(\sigma), \ldots, f_{i_u}(\sigma); \sigma), A\big)$ for every $\sigma \in \mathcal{N}$. By the definition of the $\widetilde{Q}_u$, and the measurability requirement on the Q_u, it follows that $\varphi^{-1}_{i_1,\ldots,i_u}(B) \in \mathfrak{N}$. Hence the above set is a countable union of graphs of $\mathfrak{N}$-$(\mathcal{B}_+^d)^u$-measurable functions each of which is intersected by a $(\mathcal{B}_+^d)^u \otimes \mathfrak{N}$-measurable set. Thus $\big(\widetilde{Q}_u(\cdot, A)\big)^{-1}(B) \in (\mathcal{B}_+^d)^u \otimes \mathfrak{N}$, which completes the proof.

Acknowledgement.

The author wishes to thank Ilya Molchanov for helpful discussions, which have substantially improved the presentation of the paper, and Andrew Barbour for contributing valuable comments and suggestions. Further thanks go to a referee and an associate editor (both anonymous) for many useful remarks.

References

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.

Böker, F., Serfozo, R. (1983). *Ordered thinnings of point processes and random measures.* Stochastic Process. Appl. **15** 113–132.

Brown, T.C. (1979). *Position dependent and stochastic thinning of point processes.* Stochastic Process. Appl. **9** 189–193.

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.

Brown, T.C., Xia, A. (2001). *Stein's method and birth-death processes.* Ann. Probab. **29** 1373–1403.

Daley, D.J., Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

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.

Jagers, P., Lindvall, T. (1974). *Thinning and rare events in point processes.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **28** 89–98.

Kallenberg, O. (1975). *Limits of compound and thinned point processes.* J. Appl. Probability **12** 269–278.

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.

Matheron, G. (1975). *Random sets and integral geometry.* Wiley, New York.

Rényi, A. (1957). *A characterization of Poisson processes (in Hungarian with summaries in English and Russian).* Magyar Tud. Akad. Mat. Kutató Int. Közl. **1** 519–527.
Translated in *Selected papers of Alfréd Rényi*, Vol. 1, ed. Pál Turán, Akadémiai Kiadó, Budapest, 1976.

Schuhmacher, D. (2005) *Upper bounds for spatial point process approximations.* To appear in Ann. Appl. Probab. **15**, No. 1B.
Preprint available at http://www.math.unizh.ch/~schumi/papers/sppa.pdf

Serfozo, R.F. (1977). *Compositions, inverses and thinnings of random measures.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **37** 253–265.

Serfozo, R.F. (1984). *Rarefactions of compound point processes.* J. Appl. Probability **21** 710–719.

Stoyan, D., Kendall, W.S., Mecke, J. (1987). *Stochastic geometry and its applications.* Wiley, Chichester.