Ann Inst Stat Math (2016) 68:1–24
DOI 10.1007/s10463-014-0485-6

Intrinsically weighted means and non-ergodic marked point processes

Alexander Malinowski · Martin Schlather · Zhengjun Zhang

Received: 21 August 2013 / Revised: 17 June 2014 / Published online: 21 September 2014

Abstract Mean marks form a versatile toolbox in the analysis of marked point processes (MPPs). For ergodic processes, their definition is straightforward and practical application is well established. In the stationary non-ergodic case, though, different definitions of mark averages are possible and might be practically relevant. In this paper, the classical definition of mean marks is compared to a set of new characteristics for non-ergodic MPPs, which stand out due to the weighting of ergodicity classes. Another weighting can be introduced on the single-point level via weights given by the marks themselves. These intrinsically given weights and the weighting of ergodicity classes are closely related to each other meaning that for suitable choices of weights, a mean mark characteristic can be expressed in either way. Estimators for the different definitions of mean marks are discussed and their consistency and asymptotic normality are shown under certain conditions.

Keywords Ergodic decomposition · Hierarchical modeling · Mark-location interaction · Moment measure · Non-ergodicity · Weighted mark mean

1 Introduction

Marked point processes (MPPs) provide an adequate framework for modeling irregularly scattered events in space or time in that they incorporate the joint distri-

A. Malinowski (✉) · M. Schlather
Institute for Mathematics, University of Mannheim, A5 6, 68131 Mannheim, Germany
e-mail: malinows@math.uni-goettingen.de

M. Schlather
e-mail: schlather@math.uni-mannheim.de

Z. Zhang
Department of Statistics, University of Wisconsin at Madison, 1300 University Avenue, Madison, WI 53706, USA
e-mail: zjz@stat.wisc.edu

bution of the observed values and the corresponding locations (e.g., Karr 1991; Møller,Waagepetersen 2003; Schlather et al. 2004; Myllymäki and Penttinen 2009; Daley and Vere-Jones 2008; Diggle et al. 2010). Due to the variety of possible forms of dependence between marks and locations, already the notion of the mark mean, which is usually considered as being the simplest summary statistic, involves delicate questions. Additionally, the points of a MPP may carry information, e.g., in terms of a second mark component, which naturally lead to a weighting of points when considering averages. These *intrinsically given* weights may further complicate the situation.

For an introductory example, let us consider the trading process in financial markets. Transactions are irregularly spaced in time and can thus be regarded as a point process. They are typically characterized by the two quantities *price* and *volume*, which form the marks in the MPP context. A benchmark quantity is the so-called volume-weighted average price (VWAP, e.g., Madhavan 2002; Bialkowski et al. 2008), which is defined as $p_{\mathrm{VWAP}} = \sum_{i=1}^{k}(p_i v_i)/\sum v_i$ for prices p_i and traded volumes v_i, $i = 1, \ldots, k$. Here, the weights $v_i/\sum v_i$ and the prices clearly depend on each other. Since the dynamics of the trading process usually differ between different trading periods (e.g., volatile vs. calm periods, fast trading vs. slow trading periods), non-ergodicity appears naturally. Each day of trading, for instance, might represent a different ergodicity class.

This example can be embedded in a broader family of mark means of MPPs: Let, for simplicity,

$$\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$$

be a stationary MPP on $\mathbb{R}^d$, where $t_i \in \mathbb{R}^d$ is the point location, $y_i \in \mathbb{R}$ is the first mark and $z_i \in [0, \infty)$ is a second mark of the ith point of Φ. Let $\Phi_{\mathrm{g}} = \{t : (t, y, z) \in \Phi\}$ denote the ground process of point locations of Φ and let the marks at location $t \in \Phi_{\mathrm{g}}$ be denoted by $y(t)$ and $z(t)$. Assume that the z-component is normalized such that its mean value is 1. Then, using an intuitive notation, a weighted mean for a measurable transformation of the first mark component, $f(y(t))$, is given by

$$\mu_f^{(1)} = \mathbb{E}[z(t) f(y(t)) \,|\, t \in \Phi_{\mathrm{g}}]. \tag{1}$$

The conditioning on "$t \in \Phi_{\mathrm{g}}$" is understood in the sense of the Palm mark distribution and due to the stationarity assumption, the quantity in (1) does not depend on t. Since the weights $z(t)$ are provided by the MPP itself and may depend on both the marks $y(t)$ and the point locations $t \in \Phi_{\mathrm{g}}$, we refer to $\mu_f^{(1)}$ as *intrinsically weighted mark mean* of Φ. The formal definition of $\mu_f^{(1)}$ and related quantities will be given in Sect. 2.

When a system of randomly distributed objects is described by an MPP, different choices of intrinsic weights $z(t)$ make sense, leading to different weighted mark means that are relevant for different statistical questions:

– Average height of trees: Consider n forest areas of about equal size, say 1 hectare each, which, together, form the forest of interest. For simplicity, it is assumed that the trees in all of the forest areas are so dense such that the whole ground surface is covered by tree canopy. The n forest areas might represent up to n different ergodicity classes with different properties concerning size, shape, species and

distribution of trees. The unweighted average of the height of all trees in all forest areas is relevant for forest inventory applications and amounts to $z(t) = 1$ in (1), where $y(t)$ is the height of a tree at t and $f = \text{id}$ is the identity function. By way of contrast, the average height of the canopy of a typical *forest area* should be understood as a quantity that does not depend on the densities of the trees within the ergodicity classes. Then, to define a suitable quantity for the average height of a typical forest area, the weight $z(t)$ will be inverse-proportional to the density of trees in the neighborhood of t. Intuitively speaking, individual trees are aggregated to forest areas in such a way that each forest area (and not each tree) receives the same total weight in (1).
– Sampling of regionalized variables: Measurements of continuous-space or continuous-time processes at random locations usually aim at estimating the underlying process, e.g., the spatial or temporal mean over the whole domain of the process. Since measurement locations are not necessarily independent of the underlying process, knowledge of the pattern of point locations might already provide information about the values of the process. Such a situation is commonly referred to as *biased* or *preferential sampling* and different weighting approaches exist to correct for this form of biases (e.g., Isaaks and Srivastava 1989). Then $z(t)$ usually depends on the pattern of point locations around t. The issue of dependence between measurement locations and the underlying process is particularly relevant when the sampling area is composed of realizations that belong to different ergodicity classes.

While ergodicity of an MPP allows for a straightforward interpretation of the mark distribution as the distribution of the mark of a typical point, non-ergodicity implies that different realizations can have different stochastic behavior. Then, the dependence between marks and point locations can be rather subtle: For a MPP that results from sampling a regionalized variable, the pattern of point locations can be independent of the underlying process for each of its ergodic subclasses, although there might be a strong dependence between the pattern of measurement locations and the process if multiple realizations are considered. For a simple example, consider a Gaussian random field $\{Y(t), t \in \mathbb{R}\}$ with a random mean combined with a stationary Poisson point process of measurement locations whose intensity is a function of $\{Y(t), t \in \mathbb{R}\}$, e.g., a log-Gaussian Cox process (Møller et al. 1998).

To treat non-ergodic MPPs adequately, this paper proposes intrinsically weighted mark means as a special case of (1), in which the weights $z(t)$ are constant within each ergodicity class but allow for compensating for differences between the different ergodicity classes.

The remainder of the article is organized as follows: In Sect. 2 moment-based characteristics for MPPs are recalled and generalized. Their behavior and interpretation for non-ergodic processes are studied, and, following the idea of the above examples, alternative definitions of moment-based summary statistics are proposed in Sect. 3. Different estimators for the above characteristics and their asymptotic properties are discussed in Sect. 4; in the conclusion, also a comment on the relation to inference in geostatistical applications is given. The Appendix contains basic results from ergodic theory and some of the proofs of Sect. 4.

2 MPP moment-measures and measurement of interaction effects

Throughout the paper, $\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$ is a stationary and simple marked point process on $\mathbb{R}^d$ with marks $(y(t_i), z(t_i)) = (y_i, z_i) \in \mathbb{R} \times [0, \infty)$, and $\Phi_{\mathrm{g}} = \{t : (t, y, z) \in \Phi\}$ is its ground process of point locations. For the general theory of point processes, the reader is referred to Stoyan et al. (1995) and Daley and Vere-Jones (2003, 2008), for example.

The classical formal definition of $\mu_f^{(1)}$ in (1) is

$$\mu_f^{(1)} = \frac{\mathbb{E}\sum_{(t,y,z)\in\Phi} zf(y)\mathbf{1}_B(t)}{\mathbb{E}\sum_{(t,y,z)\in\Phi} z\mathbf{1}_B(t)} \tag{2}$$

for any Borel set $B \subset \mathbb{R}^d$ with $\nu(B) > 0$, where ν is the Lebesgue measure (Daley and Vere-Jones 2008, Chap. 13). The numerator and denominator are denoted by $\alpha_f^{(1)}(B)$ and $\alpha^{(1)}(B)$, respectively. Due to the stationarity of Φ, the definition in (2) does not depend on the choice of B. The most relevant example of f in practical application is $f(y) = y^m$ for $m = 1, 2, \ldots$ Then, if $z(t) \equiv 1$, $\mu_f^{(1)}$ simply represents the m-th moment of the (Palm) mark distribution. Similarly, dependency structures within MPPs (cf. Stoyan and Stoyan 1994, Chap. 14) can be investigated by mark means, conditioned on the existence of another point at a certain distance, e.g., $\mathbb{E}[f(y(t_1)) \,|\, t_1, t_2 \in \Phi_{\mathrm{g}}, t_1 \neq t_2]$ (cf. Schlather 2001).

The superscripts $^{(1)}$ and $^{(2)}$ are used to indicate whether first- or second-order measures are meant.

Definition 1 For any non-negative function f on $\mathbb{R}\times\mathbb{R}$, a σ-finite measure on $\mathbb{R}^d \times \mathbb{R}^d$ is defined by

$$\alpha_f^{(2)}(C) = \mathbb{E} \sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi}^{\neq} z_1 f(y_1, y_2)\mathbf{1}_C((t_1, t_2)), \;\; C \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d), \tag{3}$$

which we call *weighted second moment measure*. Here, "$\neq$" indicates that the sum runs over all pairs of points with $(t_1, y_1, z_1) \neq (t_2, y_2, z_2)$.

A related but simpler measure than (3) can be defined by introducing the following sets in $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$: For $B \in \mathcal{B}(\mathbb{R}^d)$, $I \in \mathcal{B}(\mathbb{R})$ and $t = (t_1, \ldots, t_d) \in [0, \infty)^d$ let

$$\begin{aligned} C(B, I) &= \begin{cases} \big\{(t_1, t_2) : t_1 \in B,\ t_2 \in t_1 + I\big\}, & d = 1, \\ \big\{(t_1, t_2) : t_1 \in B,\ t_2 \in t_1 + \{x \in \mathbb{R}^d : \|x\| \in I\}\big\}, & d > 1, \end{cases} \\ C(t, I) &= C([\mathbf{0}, t], I), \\ C(I) &= C([\mathbf{0}, \mathbf{1}], I). \end{aligned}$$

Here, $[\mathbf{0}, t] = [0, t_1] \times \cdots \times [0, t_d]$ and $\mathbf{1}$ denotes the vector $(1, \ldots, 1) \in \mathbb{R}^d$. Then

$$\alpha_f^{(2)}(C(I)), \quad I \in \mathcal{B}(\mathbb{R}),$$

defines a σ-finite measure on $\mathbb{R}$, which represents a typical pair of points whose distance is contained in I. This interpretation becomes clearer for I being a small interval in $\mathbb{R}^+$. Note that the distinction between $d = 1$ and $d > 1$ in the definition of the set $C(B, I)$ allows to capture a possibly non-symmetric behavior of $\alpha_f^{(2)}$ in the one-dimensional case. In particular,

$$\alpha_f^{(2)}(C(I)) = \begin{cases} \mathbb{E} \sum^{\neq}_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi,\ t_1\in[0,\,1]} z_1 f(y_1, y_2)\mathbf{1}_{t_2-t_1\in I}, & d = 1, \\ \mathbb{E} \sum^{\neq}_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi,\ t_1\in[\mathbf{0},\,\mathbf{1}]} z_1 f(y_1, y_2)\mathbf{1}_{\|t_2-t_1\|\in I}, & d > 1. \end{cases}$$

For notational convenience, it is assumed that the density of $\alpha_f^{(2)}(C(\cdot))$ w.r.t. the one-dimensional Lebesgue measure exists, which is then denoted by ρ_f. In the case $f \equiv 1$, the index $_f$ is dropped.

Definition 2 For a measurable non-negative function f on $\mathbb{R} \times \mathbb{R}$, the *(weighted) second-order mark mean* is defined as

$$\mu_f^{(2)}(I) = \frac{\alpha_f^{(2)}(C(I))}{\alpha^{(2)}(C(I))}, \quad I \in \mathcal{B}(\mathbb{R}), \tag{4}$$

if $\alpha^{(2)}(C(I)) > 0$.

Obviously, $\alpha_f^{(2)}(C(\cdot))$ is dominated by $\alpha^{(2)}(C(\cdot))$, and, according to Radon–Nikodym, the density of $\alpha_f^{(2)}(C(\cdot))$ exists and can be expressed as

$$\mu_f^{(2)}(r) = \frac{\rho_f(r)}{\rho(r)}, \tag{5}$$

for $r \neq 0$. With a slight abuse of notation, the terms in (4) and (5) are both referred to as $\mu_f^{(2)}$. For $r \neq 0$ and f only depending on its first argument, $\mu_f^{(2)}(r)$ can be interpreted as the (weighted) expectation of the mark at location t subject to the conditioning that Φ has a point at location t and at some point on the sphere with radius r around t, i.e.,

$$\mu_f^{(2)}(r) = \mathbb{E}\Big[z(t) f(y(t)) \Big| t \in \Phi_g,\ \#\big(\{s : \|s - t\| = r\} \cap \Phi_g\big) = 1\Big].$$

Remark 1 (a) The extension to moment measures of higher order is straightforward and allows to condition on arbitrary point constellations. In practice, however, mostly first- and second-order statistics are considered.
(b) The non-negativity condition on f can be weakened by considering the restriction of $\mu_f^{(2)}(\cdot)$ to some bounded set $J \in \mathcal{B}(\mathbb{R})$. Then it suffices that $\alpha_h^{(2)}(C(J)) < \infty$ for $h = f_+ = \max\{f, 0\}$ *or* for $h = f_- = -\min\{f, 0\}$.

Well-known examples of second-order mark characteristics for stationary and isotropic MPPs are Cressie's mark variogram and covariance function (Cressie 1993), Stoyan's k_{mm}-function (Stoyan 1984), the mark correlation function (Isham 1985), the J-function (Lieshout 2006) and the E- and V-function (Schlather 2001), which can all be expressed in terms of (5) with a constant z-component. Here, only the functions $f(y_1, y_2) = y_1 y_2$, $f(y_1, y_2) = y_1$ and $f(y_1, y_2) = y_1^2$ are used. For the purpose of model diagnostics, Adelfio and Schoenberg (2009) introduce weighted second-order statistics in which each point is weighted by the inverse of the conditional intensity of points.

3 Moment measures for non-ergodic MPPs

Ergodicity makes spatial averages over suitably increasing observation windows of a single realization converge to the corresponding expectation over the state space:

$$\nu(W)^{-1} \int_W X(T_x \Phi)\, \nu(dx) \xrightarrow{\text{a.s.}} \mathbb{E}(X(\Phi)), \quad \text{for } \nu(W) \to \infty \text{ suitably},$$

for any integrable function X on the space of all locally finite counting measures. Here, T_x denotes the shift operator, i.e., if $\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$, then $T_x \Phi = \{(t_i + x, y_i, z_i) : i \in \mathbb{N}\}$. In essence, ergodicity enables consistent estimation of MPP moment measures by observing a single realization on a suitably increasing domain. In this section, though, the opposite situation is considered, namely where Φ is a non-ergodic process.

A non-ergodic MPP can be seen as a hierarchical or doubly stochastic model, which, in a first step, draws an ergodic class out of which the final realization is drawn in a second step. In some cases, the set of ergodic classes is finite but, in general, can be infinite.

Proposition 1 *Let Φ be a non-ergodic MPP with probability law P, let $\mathbb{M}_0$ and $\mathcal{M}_0$ denote the space of all locally finite counting measures on $\mathbb{R}^d \times \mathbb{R} \times [0, \infty)$ and the usual σ-algebra, respectively. (See Appendix A for more details.) Then, for $B \in \mathcal{B}(\mathbb{R}^d)$ and $I \in \mathcal{B}(\mathbb{R})$, with $\alpha^{(1)}(B) > 0$ and $\alpha^{(2)}(C(I)) > 0$,*

$$\mu_f^{(1)} = \frac{\mathbb{E}_Q\left[\mu^{(1)}_{f,\Phi|Q} \cdot \alpha^{(1)}_{\Phi|Q}(B)\right]}{\alpha^{(1)}(B)} = \int \mu^{(1)}_{f,\Phi|Q=q} \frac{\alpha^{(1)}_{\Phi|Q=q}(B)}{\alpha^{(1)}(B)} \lambda(dq), \tag{6}$$

$$\mu_f^{(2)}(I) = \frac{\mathbb{E}_Q\left[\mu^{(2)}_{f,\Phi|Q}(I)\alpha^{(2)}_{\Phi|Q}(C(I)) \,\middle|\, \alpha^{(2)}_{\Phi|Q}(C(I)) > 0\right]}{\alpha^{(2)}(C(I))} \lambda\left(\left\{\alpha^{(2)}_{\Phi|Q}(C(I)) > 0\right\}\right). \tag{7}$$

Here, Q is a random variable with values in the set $\mathcal{P}_{\text{erg}}$ of all ergodic MPP probability laws, distributed according to some probability measure λ such that $P(M) = \int_{\mathcal{P}_{\text{erg}}} q(M)\lambda(dq)$, $M \in \mathcal{M}_0$.

Further, if λ-almost all Lebesgue densities $\rho_{\Phi|Q}(r)$ of $\alpha^{(2)}_{\Phi|Q}(C(\cdot))$ exist, for $r \neq 0$ with $\rho(r) > 0$, it is

$$\mu_f^{(2)}(r) = \frac{\mathbb{E}_Q\left[\mu^{(2)}_{f,\Phi|Q}(r) \cdot \rho_{\Phi|Q}(r) \,\middle|\, \rho_{\Phi|Q}(r) > 0\right]}{\rho(r)} \times \lambda\left(\left\{\rho_{\Phi|Q}(r) > 0\right\}\right). \tag{8}$$

Proof The ergodic decomposition theorem (cf. Theorem 3 in the Appendix) guarantees the existence of a unique decomposition $P(\cdot) = \int_{\mathcal{P}_{\text{erg}}} q(\cdot)\lambda(dq)$ and a corresponding mixing random variable $Q \sim \lambda$. Conditioning Φ on Q, the moment measures $\alpha_f^{(i)}$ can be decomposed.

In the first-order case (cf. (2)), since $[\Phi|Q]$ is stationary by definition, $\alpha^{(1)}_{\Phi|Q}(B)$ is positive λ-a.s. if and only if $\alpha^{(1)}(B)$ is so. Thus, $\mu^{(1)}_{f,\Phi|Q}$ is well-defined and

$$\mu_f^{(1)} = \frac{\mathbb{E}_Q\left[\alpha^{(1)}_{f,\Phi|Q}(B)\right]}{\alpha^{(1)}(B)} = \frac{\mathbb{E}_Q\left[\mu^{(1)}_{f,\Phi|Q} \times \alpha^{(1)}_{\Phi|Q}(B)\right]}{\alpha^{(1)}(B)}.$$

In the second-order case (cf. Definition 2), $\alpha^{(2)}_{\Phi|Q}(C(I))$ can take the value 0 with positive probability even though $\alpha^{(2)}(C(I))$ is positive. For those realizations of Q for which $\alpha^{(2)}_{\Phi|Q}(C(I)) > 0$ holds, $\alpha^{(2)}_{f,\Phi|Q}(C(I))$ can be replaced by $\mu^{(2)}_{f,\Phi|Q}(I) \cdot \alpha^{(2)}_{\Phi|Q}(C(I))$, which yields (7). Equation (8) is obtained similarly. □

The decomposition of non-ergodic MPPs given in Proposition 1 is illustrated by the following examples. For notational convenience, the intrinsic weights, i.e., the z-components of the marks, are assumed to be constant and can thus be omitted.

In the most elementary form of non-ergodicity only two "ergodic regimes" exist. Then, intuitively speaking, prior to each realization, a coin is tossed and depending on the result, the realization is generated according to a certain set of parameters.

Example 1 Let $p_1, p_2 > 0$ be chosen such that $p_1 + p_2 = 1$. With probability p_1 let Φ be a stationary Poisson point process on $\mathbb{R}$ with intensity λ_1 and independent marks with mean m_1. With probability p_2, let the process be constructed in the same way but with parameters λ_2 and m_2 instead of λ_1 and m_1, with $m_1 \neq m_2$. Let the two ergodicity classes be denoted by q_1 and q_2. Due to the Poisson property and independence of the marks in this example, there is no interaction between the points of the process. Therefore, only the first-order mark mean $\mu_f^{(1)}$ needs to be considered. Then, for $i = 1, 2$,

$$\alpha^{(1)}_{\Phi|Q=q_i}(B) = \lambda_i \times \nu(B),$$
$$\mu^{(1)}_{f,\Phi|Q=q_i} = f(m_i).$$

The latter equality is due to independence between point locations and marks. Thus,

$$\mu_f^{(1)} = \frac{p_1\lambda_1\nu(B)f(m_1) + p_2\lambda_2\nu(B)f(m_2)}{\alpha^{(1)}(B)}.$$

Since also the denominator can be decomposed into $\alpha^{(1)}(B) = \mathbb{E}_Q\alpha^{(1)}_{\Phi|Q}(B)$, it is

$$\mu_f^{(1)} = \frac{p_1\lambda_1 f(m_1) + p_2\lambda_2 f(m_2)}{p_1\lambda_1 + p_2\lambda_2}. \tag{9}$$

Obviously, the mark mean in its classical definition corresponds to a weighted average over the different ergodicity classes, where the weights not only depend on the probabilities associated with the ergodicity classes, p_1 and p_2, but also on the intensity of points within each of the classes, λ_1 and λ_2.

Although the decomposition in (9) is instructive from a theoretical point of view, it is less relevant for estimation problems since in most practical applications, not only p_1, p_2, λ_1 and λ_2 are unknown, but also the ergodicity class of a given realization would have to be estimated. Let us still mention that the components on the right-hand side of (9) could be approached by hierarchical Bayesian modeling. Our example is a special case of the common latent process model in Diggle et al. (2010). To be explicit, let p_1, λ_1, λ_2, m_1 and m_2 be random variables with density functions $g_{p,1}$, $g_{l,1}$, $g_{l,2}$, $g_{m,1}$ and $g_{m,2}$, respectively, such that $g_{l,1}$ and $g_{l,2}$ as well as $g_{m,1}$ and $g_{m,2}$ have distinct support. For $i = 1, 2$, let $h_{\mu,i}$ be a family of density functions for the marks of the ergodic class i, parametrized by their mean μ. Provided that the hyperparameters are known and that the j-th realization belongs to the ergodic class i, the log likelihood for the j-th realization with n_j points $\left\{(t_1^{(j)}, y_1^{(j)}), \ldots, (t_{n_j}^{(j)}, y_{n_j}^{(j)})\right\}$ is

$$u_j(i) = n_j \log(\lambda_i) - \lambda_i \nu(B) + \sum_{k=1}^{n_j} \log h_{m_i,i}\left(y_k^{(j)}\right).$$

Let $\xi^{(j)} \in \{1, 2\}$ be the random variable indicating the ergodic class of sample j, $j = 1, \ldots, n$. Then the overall log likelihood equals

$$\sum_{j=1}^{n} \left[u_j(\xi_j) + (2 - \xi^{(j)}) \log(p_1) + (\xi^{(j)} - 1) \log(p_2)\right]$$
$$+ \log(g_{p,1}(p_1)) + \log(g_{l,1}(\lambda_1)) + \log(g_{l,2}(\lambda_2))$$
$$+ \log(g_{m,1}(m_1)) + \log(g_{m,2}(m_2))$$

and has to be maximized with respect to all the parameters, including the $\xi^{(j)}$.

The following example is a straightforward generalization of Example 1. It has an (uncountably) infinite number of ergodicity classes:

Example 2 Let (Λ, M) be a $(\mathbb{R}^+ \times \mathbb{R})$-valued random vector and, conditionally on that, let Φ be a stationary Poisson point process on $\mathbb{R}$ with intensity Λ and with random marks with mean M that are independent of the point locations. Then,

$$\mu_f^{(1)} = \frac{\mathbb{E}[\Lambda f(M)]}{\mathbb{E}\Lambda}.$$

Here, in principle, a hierarchical Bayesian approach might be still feasible, but additional strong assumptions will be necessary. Instead, non-parametric estimators for $\mu_f^{(1)}$ and $\mu_f^{(2)}$ are discussed in Sect. 4.

Example 3 A common model for point processes with clusters is the log-Gaussian Cox process (Møller et al. 1998). Amongst others, Diggle et al. (2010) and Myllymäki and Penttinen (2009) use log-Gaussian Cox processes, combined with an intensity-dependent marking, as parametric models for preferential sampling applications. The log-Gaussian Cox process is based on a Gaussian random field $Y = \{Y(t) : t \in \mathbb{R}^d\}$: Conditionally on Y, the points are given by a Poisson point process with intensity measure $\Xi(B) = \int_B \exp(Y(t))dt$, $B \in \mathcal{B}(\mathbb{R}^d)$. It is known that a log-Gaussian Cox process with an underlying *stationary* Gaussian field Y is ergodic if and only if Y is so. A sufficient condition for Y being ergodic is that the covariance function decays to zero (Adler and Taylor 2007).

Now, let Λ be a real-valued random variable with a positive variance and, conditionally on Λ, let $\{Y(t) : t \in \mathbb{R}\}$ be a stationary Gaussian random field with mean Λ. Consider a generalized log-Gaussian Cox process Φ on $\mathbb{R}$ with intensity measure $\Xi(B) = \int_B \exp(Y(t))dt$ and with marks $y(t) = Y(t)$ for $t \in \Phi_{\mathrm{g}}$. Then Φ is a non-ergodic MPP. For each ergodicity class (which is in this example determined by the realization of Λ), the second-order mark mean $\mu^{(2)}_{f|\Lambda}(\cdot)$ is a non-constant function and, in general, so is the overall second-order mark mean $\mu^{(2)}_f(\cdot)$. In particular, by applying some basics of MPP Palm theory, one obtains

$$\mu^{(2)}_f(r) = \frac{\rho_f(r)}{\rho(r)} = \frac{\mathbb{E}\big[f(Y(0), Y(t))\exp(Y(0) + Y(r))\big]}{\mathbb{E}\exp(Y(0) + Y(r))}.$$

Since the random mean Λ is the mixing variable and since $\mu^{(2)}_{f,\Phi|\Lambda}(r) \cdot \rho_{\Phi|\Lambda}(r) = \rho_{f,\Phi|\Lambda}(r)$, the decomposition of $\mu^{(2)}_f(r)$ according to Proposition 1 is given by

$$\mu^{(2)}_f(r) = \frac{\mathbb{E}_\Lambda\Big[\mu^{(2)}_{f,\Phi|\Lambda}(r) \times \mathbb{E}_{Y|\Lambda}\exp(Y(0) + Y(r))\Big]}{\mathbb{E}\exp(Y(0) + Y(r))}.$$

For the first-order mark mean, analogously, it is

$$\mu^{(1)}_f = \frac{\mathbb{E}\big[f(Y(0))\exp(Y(0))\big]}{\mathbb{E}\exp(Y(0))} = \frac{\mathbb{E}_\Lambda\big[\mu^{(1)}_{f,\Phi|\Lambda} \times \mathbb{E}_{Y|\Lambda}\exp(Y(0))\big]}{\mathbb{E}\exp(Y(0))}.$$

Note that Example 2 with the choice $\Lambda = \exp(M)$ and M normally distributed is obtained as a special case of Example 3 when the correlation function of the Gaussian field Y is chosen to be 1 everywhere, i.e., the field is realization-wise constant.

Proposition 1 and the subsequent examples show that, in case of non-ergodicity, $\mu^{(i)}_f$ is an average of its counterparts of the ergodic subclasses, in which each class q is implicitly weighted by the respective intensity $\alpha^{(i)}_{\Phi|Q=q}$ (in addition to the weighting according to the probability measure of Q). If all ergodic subprocesses $[\Phi|Q = q]$ have the same intensity measure, the additional weights cancel out and $\mu^{(i)}_f = \mathbb{E}_Q\mu^{(i)}_{f,\Phi|Q}$. In the general case, however, these weights do not cancel out and a single ergodicity class with low probability may exhibit an immense value of $\alpha^{(i)}_{\Phi|Q=q}$ and thus drive the value of $\mu_f^{(i)}$. This rises the demand for a new characteristic $\tilde{\mu}_f^{(i)}$ that summarizes the properties of all ergodicity classes, irrespectively of how the processes of point locations differ between the ergodicity classes. These requirements are met by a definition that excludes the implicit weighting proportional to the ith-order intensities:

Definition 3 Let λ and Q be the ergodic decomposition mixture measure and mixture variable, respectively, of Φ, and let $\mathbb{E}_Q\left|\mu_{f,\Phi|Q}^{(i)}\right| < \infty$. Then we call

$$\tilde{\mu}_f^{(i)} = \mathbb{E}_Q \mu_{f,\Phi|Q}^{(i)} = \int_{\mathcal{P}_{\mathrm{erg}}} \mu_{f,\Phi|Q=q}^{(i)} \, \lambda(dq) \tag{10}$$

the *(equally-weighted) average ith-order mark mean* of Φ.

Relating to the introductory forest example, the classical definition of the mark mean in (2) corresponds to the average height of *all* trees, irrespective of different tree densities in the different forest areas, while the new definition in (10) refers to the average height of a typical forest (mean of a typical realization).

Example 4 (Continuation of Examples 1, 2 and 3) For the MPP defined in Example 1,

$$\tilde{\mu}_f^{(1)} = p_1 f(m_1) + p_2 f(m_2).$$

Both $\tilde{\mu}_f^{(1)}$ and $\mu_f^{(1)}$ are convex combinations of $f(m_1)$ and $f(m_2)$. Whilst $\tilde{\mu}_f^{(1)}$ only depends on p, the classical average $\mu_f^{(1)}$ additionally depends on the values of λ_1 and λ_2.

For the MPP defined in Example 2,

$$\tilde{\mu}_f^{(1)} = \mathbb{E} f(M).$$

Under the assumptions of Example 3 it is

$$\tilde{\mu}_f^{(1)} = \mathbb{E}_\Lambda \left[\frac{\mathbb{E}_{Y|\Lambda}\big[f(Y(0))\exp(Y(0))\big]}{\mathbb{E}_{Y|\Lambda}\exp(Y(0))} \right],$$
$$\tilde{\mu}_f^{(2)}(r) = \mathbb{E}_\Lambda \left[\frac{\mathbb{E}_{Y|\Lambda}\big[f(Y(0), Y(t))\exp(Y(0)+Y(r))\big]}{\mathbb{E}_{Y|\Lambda}\exp(Y(0)+Y(r))} \right].$$

Clearly, an ergodic decomposition as in Definition 3 can be applied to any expectation-based functional of an MPP including the Palm mark distribution itself.

The following remark is obvious:

Remark 2 The quantity $\tilde{\mu}_f^{(i)}$ coincides with $\mu_f^{(i)}$ if $\alpha_{\Phi|Q}^{(i)}$ is λ-a.s. constant, i.e., if all ergodicity classes exhibit the same intensity measure. This is the case when Φ is ergodic.

4 Estimators and properties

First note that, due to $\lim_{I\to\mathbb{R}} \mu_f^{(2)}(I) = \mu_{\tilde{f}}^{(1)}$ for $f(y_1, y_2) = \tilde{f}(y_1)$, it suffices to consider the second-order statistics based on $\mu_f^{(2)}$. For readability reasons, the superscript $^{(2)}$ is dropped in all the estimators of $\mu_f^{(2)}$.

4.1 The ergodic case

For ergodic processes Φ, the pointwise ergodic theorem for MPPs (Proposition 3 in the Appendix) yields that

$$\mathbb{E}\left[\sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi}^{\neq} z_1 f(y_1, y_2)\mathbf{1}_{(t_1,t_2)\in C(I)}\right] = \lim_{n\to\infty}\left[n^{-d}\sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\varphi}^{\neq} z_1 f(y_1, y_2)\mathbf{1}_{(t_1,t_2)\in C(n\mathbf{1},I)}\right]$$

for almost all realizations φ of Φ, which builds the basis for the estimators being discussed in this section.

Applying the standard estimator for MPP moment measures (e.g., Baddeley 1999) to a realization of Φ observed on the set $[\mathbf{0}, \mathbf{T}]$, $\mathbf{T} \in (0, \infty)^d$, gives

$$\hat{\mu}_f(I, \Phi, \mathbf{T}) = \frac{\hat{\alpha}_f(I, \Phi, \mathbf{T})}{\hat{\alpha}_1(I, \Phi, \mathbf{T})}, \tag{11}$$

where $\hat{\alpha}_f(I, \Phi, \mathbf{T}) = \sum^{\neq}_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi} z_1 f(y_1, y_2)\mathbf{1}_{(t_1,t_2)\in C(\mathbf{T},I)}$.

Lemma 1 *If Φ is ergodic, $\hat{\mu}_f(I, \Phi, \mathbf{T})$ is consistent for $\mu_f^{(2)}(I)$ as $\mathbf{T} \to \infty$ componentwise. If Φ is non-ergodic, $\hat{\mu}_f(I, \Phi, \mathbf{T})$ is consistent if and only if $\mu^{(2)}_{f,\Phi|Q=q}(I)$ is constant w.r.t. q.*

Proof By Proposition 3, the vector consisting of the numerator and the denominator of (11), each normalized by the volume of $[0, \mathbf{T}]$, converges a.s. to the vector $\left(\alpha_f^{(2)}(C(I)), \alpha^{(2)}(C(I))\right)$ if Φ is ergodic. The first assertion thus follows from the continuous mapping theorem. In the non-ergodic case, clearly only $\mu^{(2)}_{f,\Phi|Q=q}(I)$ can be estimated consistently for q being the respective ergodicity class. Though, if $\mu^{(2)}_{f,\Phi|Q=q}(I)$ is constant w.r.t. q it is $\mu_f^{(2)}(I) = \mu^{(2)}_{f,\Phi|Q=q}(I)$ for any $q \in \mathcal{P}_{\text{erg}}$. $\square$

To establish asymptotic normality of $\hat{\mu}_f(I, \Phi, \mathbf{T})$, for technical reasons, only MPPs on $\mathbb{R}$ are considered. While the following CLT-like result remains valid also under weaker mixing conditions, our assumption of a certain form of memorylessness allows for a fairly simple proof. The following property is to ensure that the limiting Gaussian distribution has a non-zero variance.

Property 1 Let Φ be a stationary MPP on $\mathbb{R}$ and let $L > 0$ and $m \in \mathbb{N}$. For $j \in \mathbb{Z}$, define X_j as $X_j = \hat{\alpha}_f([(j-1)L, jL]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(j-1)L, jL])$, where $\hat{\alpha}_f([a, b])$ is short notation for $\hat{\alpha}_f(I, \Phi, b) - \hat{\alpha}_f(I, \Phi, a)$. Then $(X_j)_{j\in\mathbb{Z}}$ is a stationary time series with autocovariance function $C(j) = \operatorname{Cov}(X_0, X_j)$, $j \in \mathbb{Z}$. We call the process Φ *non-degenerate for L and m* if $\sum_{j=-m}^{m} C(j) > 0$.

Theorem 1 *Let Φ be an ergodic MPP on $\mathbb{R}$ with the property that there exists a constant $L_0 > 0$ such that $0 < \operatorname{Var} \hat{\alpha}_f(I, \Phi, L_0) < \infty$ and the restrictions of Φ,*

$$\begin{aligned} &\Phi \cap \left\{(-\infty, 0) \times \mathbb{R} \times [0, \infty)\right\} \\ \textit{and} \quad &\Phi \cap \left\{(L_0, \infty) \times \mathbb{R} \times [0, \infty)\right\}, \end{aligned}$$

are stochastically independent of each other. Then, for $T \to \infty$,

$$\sqrt{\hat{\alpha}_1(I, \Phi, T)} \left\{ \frac{\hat{\alpha}_f(I, \Phi, T)}{\hat{\alpha}_1(I, \Phi, T)} - \mu_f^{(2)}(I) \right\} \xrightarrow{d} \mathcal{N}(0, s), \tag{12}$$

where

$$s = \lim_{T\to\infty} \frac{\operatorname{Var}\left[\hat{\alpha}_f(I, \Phi, T) - \mu_f^{(2)}(I)\hat{\alpha}_1(I, \Phi, T)\right]}{T\alpha^{(2)}(C(I))} \in [0, \infty).$$

If Φ satisfies Property 1 for L_0 and $m_0 = \min\left\{m \in \mathbb{N} : m \geq \frac{L_0+\max\{|x|:x\in I\}}{L_0}\right\}$, then $0 < s < \infty$.

Note that, to the authors knowledge, only pathological examples violate Property 1. The independence assumption in Theorem 1 is satisfied, for instance, for a Poisson Point process with marks $y(t)$ only depending on the point pattern within a radius of L_0 around t.

Proof In the following, I is a fixed interval, and a single realization of Φ is considered. With regard to (11), the short notation $\hat{\alpha}_f(T) = \hat{\alpha}_f(I, \Phi, T)$, for $T > 0$, and $\hat{\alpha}_f([a, b]) = \hat{\alpha}_f(b) - \hat{\alpha}_f(a)$, for $b > a > 0$, is used. Then, for $T = nL_0$, $n \in \mathbb{N}$, decompose $[0, T]$ into sub-intervals of length L_0, which gives

$$\begin{aligned} &\sqrt{\hat{\alpha}_1(nL_0)} \left\{ \frac{\hat{\alpha}_f(nL_0)}{\hat{\alpha}_1(nL_0)} - \mu_f^{(2)}(I) \right\} \\ &\quad = \frac{\sum_{i=1}^{n} \left\{\hat{\alpha}_f([(i-1)L_0, iL_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(i-1)L_0, iL_0])\right\}}{\sqrt{nL_0\alpha^{(2)}(C(I))}} \sqrt{\frac{nL_0\alpha^{(2)}(C(I))}{\hat{\alpha}_1(nL_0)}}. \end{aligned} \tag{13}$$

The latter factor of the RHS of (13) converges to 1 in probability. The summands in (13) have zero mean. The sequence of summands is m_0-dependent with $m_0 = \min\left\{m \in \mathbb{N} : m \geq \frac{L_0+\max\{|x|:x\in I\}}{L_0}\right\}$, and the classical CLT for m-dependent variables yields that the first factor of (13) converges to a centered Gaussian distribution with variance

$$s = \lim_{n\to\infty} \frac{1}{nL_0\alpha^{(2)}(C(I))} \operatorname{Var}\left(\hat{\alpha}_f(nL_0) - \mu_f^{(2)}(I)\hat{\alpha}_1(nL_0)\right).$$

It remains to show that $0 < s < \infty$. It is

$$\begin{aligned}
s &= \lim_{n\to\infty} \frac{1}{nL_0\alpha^{(2)}(C(I))} \\
&\quad \times \operatorname{Var}\left[\sum_{i=1}^{n} \left\{\hat{\alpha}_f([(i-1)L_0, iL_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(i-1)L_0, iL_0])\right\}\right] \\
&= \lim_{n\to\infty} \frac{1}{nL_0\alpha^{(2)}(C(I))} \\
&\quad \times \left\{\sum_{i=m_0+1}^{n-m_0} \sum_{j=i-m_0}^{i+m_0} \operatorname{Cov}\Big(\hat{\alpha}_f([(i-1)L_0, iL_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(i-1)L_0, iL_0]),\right. \\
&\qquad\qquad \left.\hat{\alpha}_f([(j-1)L_0, jL_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(j-1)L_0, jL_0])\Big)\right\} \\
&= \frac{1}{L_0\alpha^{(2)}(C(I))} \\
&\quad \times \left\{\sum_{j=-m_0}^{m_0} \operatorname{Cov}\Big(\hat{\alpha}_f([0, L_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([0, L_0]),\right. \\
&\qquad\qquad \left.\hat{\alpha}_f([(j-1)L_0, jL_0]) - \mu_f^{(2)}(I)\hat{\alpha}_1([(j-1)L_0, jL_0])\Big)\right\}.
\end{aligned}$$

The latter sum is finite and, if Property 1 is fulfilled, also non-zero. □

4.2 The non-ergodic case

If $\Phi \sim P$ is non-ergodic, drawing iid realizations $\Phi_1, \ldots, \Phi_n$ of Φ corresponds to drawing ergodicity classes according to the mixture measure λ. Consistent estimation in this case clearly requires the number n of realizations tend to infinity. Then, by the law of large numbers,

$$\hat{\mu}_f^n(I) := \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_f(I, \Phi_i, \mathbf{T}) \longrightarrow \mathbb{E}\hat{\mu}_f(I, \Phi, \mathbf{T}) = \mathbb{E}\frac{\hat{\alpha}_f(I, \Phi, \mathbf{T})}{\hat{\alpha}_1(I, \Phi, \mathbf{T})}. \tag{14}$$

However, as $\mu_f^{(i)}$ is defined as a ratio of moments (e.g., $\mu_f^{(2)}(I) = \frac{\mathbb{E}\hat{\alpha}_f(I,\Phi,\mathbf{T})}{\mathbb{E}\hat{\alpha}_1(I,\Phi,\mathbf{T})}$), the RHS of (14) is not equal to $\mu_f^{(2)}(I)$. We need additionally that $\mathbf{T} \to \infty$. Further, while in

(14) all realizations receive the weight $\frac{1}{n}$, here, more general weights are included. Let $\mathbf{w} = (w_1, \ldots, w_n)$ denote a vector of weight functions $w_i : \mathbb{M}_0 \times [0, \infty)^d \to [0, \infty)$, which might also depend on I. It is assumed that for λ-almost all ergodic MPP laws q there exist constants $w_i^*(q) \geq 0$ with $\sum_{i=1}^n w_i^*(q) > 0$ to which the weights converge stochastically within the respective ergodicity class, i.e.,

$$P_{\Phi|Q=q}\left(|w_i(\Phi, \mathbf{T}) - w_i^*(q)| > \varepsilon\right) \longrightarrow 0 \quad (\mathbf{T} \to \infty) \tag{15}$$

for all $\varepsilon > 0$. Then we consider estimators of the form

$$\begin{aligned} \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}) &= \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T}) \\ &= \left(\sum w_i(\Phi_i, \mathbf{T})\right)^{-1} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T}) \hat{\mu}_f(I, \Phi_i, \mathbf{T}). \end{aligned} \tag{16}$$

In order to estimate $\mu_f^{(2)}(I)$ consistently, according to the decomposition in (7), the weights have essentially to be chosen as

$$w_i(\Phi_i, \mathbf{T}) = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)/v_{\mathbf{T}} = \sum_{t_1, t_2 \in \Phi_{i,g}}^{\neq} \mathbf{1}_{(t_1, t_2) \in C(\mathbf{T}, I)}/v_{\mathbf{T}}, \tag{17}$$

where $v_{\mathbf{T}}$ is the volume of the cube $[\mathbf{0}, \mathbf{T}]$. By Proposition 3, the fraction $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)/v_{\mathbf{T}}$ converges to $\alpha_{\Phi|Q=Q_i}^{(2)}(C(I))$ a.s. as $\mathbf{T} \to \infty$, where Q_i is the realized ergodicity class of Φ_i. With $\mathbf{w}$ being the vector of weights from (17), we define

$$\hat{\mu}_f^{\alpha}(I, (\Phi_1, \ldots, \Phi_n), \mathbf{T}) = \hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T}), \tag{18}$$

which, in a sense, represents the family of *all* pairs of points with a distance contained in I from all realizations. This choice of weights satisfies the above stochastic convergence condition (15) and is sufficient, but not necessary, for consistency. The following theorem gives a weaker set of conditions that is still sufficient for consistency.

Theorem 2 *Let Φ_i, $i \in \mathbb{N}$, be iid copies of a possibly non-ergodic MPP Φ and let Q_i denote the respective ergodicity classes. For weight functions $\tilde{w}_i : \mathbb{M}_0 \times [0, \infty)^d \to [0, \infty)$, $i \in \mathbb{N}$, and iid random factors $W_i > 0$ with $\mathbb{E}W_1 < \infty$, let $w_i(\Phi_i, \mathbf{T}) = W_i \cdot \tilde{w}_i(\Phi_i, \mathbf{T})$ and $\mathbf{w} = (w_1(\Phi_1, \mathbf{T}), \ldots, w_n(\Phi_n, \mathbf{T}))$. Let $\mathbf{T} = \mathbf{T}(n)$ and $\lim_{n\to\infty} \mathbf{T}(n) = \infty$ at an arbitrary rate. Then, $\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w})$ is consistent for*

$\mu_f^{(2)}(I)$ *if the following conditions hold:*

$$\operatorname{Var} \tilde{w}_i(\Phi_i, \mathbf{T}) \leq c_1 \quad \textit{for some } c_1 > 0, \tag{19}$$

$$\tfrac{1}{n}\mathbb{E}\sum_{i=1}^n \tilde{w}_i(\Phi_i, \mathbf{T}) \geq c_2 > 0 \quad \forall n \geq n_0 \textit{ for some } n_0 \in \mathbb{N}, \tag{20}$$

$$\mathbb{E}[W_i \cdot Y] = \mathbb{E}W_i \cdot \mathbb{E}Y \quad \textit{for } Y = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i) \tag{21}$$

$$\textit{and } Y = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)\, \mu_{f,\Phi|Q=Q_i}^{(2)}(I) \tag{22}$$

$$n\mathbb{P}\left\{\left|\frac{\tilde{w}_i(\Phi_i, \mathbf{T})\sum_{j=1}^n \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_j)}{\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)\sum_{j=1}^n \tilde{w}_j(\Phi_j, \mathbf{T})} - 1\right| > \varepsilon\right\} \overset{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0. \tag{23}$$

The proof is given in Appendix B. Note that if $\tilde{w}_i = \tilde{w}$ for all $i \in \mathbb{N}$ for some weight function $\tilde{w}$ with $\mathbb{E}|\tilde{w}(\Phi, \mathbf{T})| < \infty$, the $\tilde{w}_i(\Phi_i, \mathbf{T})$ are iid and conditions (19) and (20) become obsolete.

As regards estimation of $\tilde{\mu}_f^{(2)}(I)$, the equally weighted estimator $\hat{\mu}_f^n(I)$, defined in (14), is consistent by the law of large numbers (cf. (14) and Definition 3). When I is thought of as a small interval around $r > 0$, the value of $\hat{\mu}_f^n(I)$ reflects an average pair of points with distance r within a randomly chosen ergodicity class. This contrasts the idea of $\hat{\mu}_f^\alpha(I)$, in which each ergodicity class is additionally weighted by the intensity of points. Analogously to Theorem 2, for consistent estimation of $\tilde{\mu}_f^{(2)}(I)$, also other choices of weights are feasible, apart from the choice $w_i(\Phi_i, \mathbf{T}) = 1$:

Corollary 1 *Under the assumptions of Theorem 2 with $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ being replaced by the constant 1, $\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w})$ is consistent for $\tilde{\mu}_f^{(2)}(I)$.*

Remark 3 If Φ is ergodic, $\hat{\mu}_f^{n,\mathrm{wght}}(I, \mathbf{w})$ is consistent for $\mu_f^{(2)}(I)$ (as $\mathbf{T} \to \infty$) for any choice of weights $\mathbf{w}$ that satisfies (15). Note that in this case, consistency is independent of n, which can be fixed to any finite value.

The weak convergence result of Theorem 1 directly extends to the non-ergodic case:

Corollary 2 *Let the assumptions of Theorem 1 be satisfied for λ-almost all ergodicity classes. Let $w_i(\Phi_i, T)$, $i = 1, 2, \ldots$, be iid weights satisfying condition (15). Let $T = T(n) \to \infty$ fast enough such that*

$$\sqrt{n} \times \mathbb{E}\left[w_i(\Phi_i, T)\sqrt{\hat{\alpha}_1(I, \Phi_i, T)}\left(\hat{\mu}_f(I, \Phi_i, T) - \mu_{f,\Phi|Q=Q_i}^{(2)}(I)\right)\right] \to 0 \tag{24}$$

as $n \to \infty$. Then,

$$\frac{\sum_{i=1}^n w_i(\Phi_i, T)\sqrt{\hat{\alpha}_1(I, \Phi_i, T)}\left(\hat{\mu}_f(I, \Phi_i, T) - \mu_{f,\Phi|Q=Q_i}^{(2)}(I)\right)}{\sqrt{\sum_{i=1}^n w_i(\Phi_i, T)}} \xrightarrow{d} \mathcal{N}(0, s),$$

where

$$s = \frac{\mathbb{E}\left[w^*(Q)^2\right]}{\mathbb{E}\left[w^*(Q)\right]} \times \mathbb{E}\left\{\lim_{T\to\infty} \frac{\mathrm{Var}\left[\hat{\alpha}_f(I,\Phi,T) - \mu^{(2)}_{f,\Phi|Q}(I)\hat{\alpha}_1(I,\Phi,T)\Big| Q\right]}{T\alpha^{(2)}_{\Phi|Q}(C(I))}\right\}.$$

Let us remark that, if the marks are independent of the process of point locations and also of the weights $w(\Phi, T)$, then the convergence in (24) is satisfied even for constant T.

4.3 Minimal variance approach

Here, another estimator for $\tilde{\mu}^{(2)}_f(I)$ is defined via a special choice of weights in the generic estimator $\hat{\mu}^{n,\mathrm{wght}}_f(I, \mathbf{w})$, given in (16). Let $\mathcal{A}^*_n$ denote the σ-algebra generated by the unmarked ground processes $\Phi_{1,\mathrm{g}}, \ldots, \Phi_{n,\mathrm{g}}$, i.e., $\mathcal{A}^*_n = \sigma(\{\{\omega : \Phi_{i,\mathrm{g}}(\omega)(B) = k\} : k \in \mathbb{N}, B \in \mathcal{B}, i = 1, \ldots, n\})$. With

$$w_i(\Phi_i, \mathbf{T}) = \mathrm{Var}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T})\middle| \mathcal{A}^*_n\right]^{-1}$$

and $\mathbf{w}$ being the vector composed of these weights, define

$$\hat{\mu}^{n,\mathrm{Var}}_f(I) = \hat{\mu}^{n,\mathrm{wght}}_f(I, \mathbf{w}). \tag{25}$$

Proposition 2 *Under the assumptions that* $\mathbb{E}[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \mid \mathcal{A}^*_n]$ *is a.s. constant, the estimator* $\hat{\mu}^{n,\mathrm{Var}}_f(I)$ *uniquely minimizes the variance amongst all estimators of the form* $\hat{\mu}^{n,\mathrm{wght}}_f(I, \mathbf{w})$ *with* $\mathcal{A}^*_n$*-measurable weights. If the conditional variance* $\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, \mathbf{T})\middle| \mathcal{A}^*_n\right]$ *is independent of the random ergodicity class Q, the estimator* $\hat{\mu}^{n,\mathrm{Var}}_f(I)$ *is consistent for* $\tilde{\mu}^{(2)}_f(I)$.

Proof For general $\mathcal{A}^*_n$-measurable weights $w_i(\Phi_i, \mathbf{T})$, $i = 1, \ldots, n$, it is

$$\begin{aligned}
&\mathrm{Var}\left[\hat{\mu}^{n,\mathrm{wght}}_f(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), \mathbf{T})\right] \\
&= \mathbb{E}\left[\frac{1}{\left(\sum w_i(\Phi_i, \mathbf{T})\right)^2} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})^2 \,\mathrm{Var}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T})\middle| \mathcal{A}^*_n\right]\right] \\
&\quad + \mathrm{Var}\left[\frac{1}{\sum w_i(\Phi_i, \mathbf{T})} \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})\mathbb{E}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T})\middle| \mathcal{A}^*_n\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n w^{\mathrm{rel}}_i(\Phi_i, \mathbf{T})^2 \,\mathrm{Var}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T})\middle| \mathcal{A}^*_n\right]\right] \quad + \quad 0
\end{aligned} \tag{26}$$

with $w^{\mathrm{rel}}_i(\Phi_i, \mathbf{T}) = w_i(\Phi_i, \mathbf{T}) / \sum_{i=1}^n w_i(\Phi_i, \mathbf{T})$. The variance term in (26) vanishes because of the assumption on $\mathbb{E}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T})\middle| \mathcal{A}^*_n\right]$. Since any weighted average

$\sum v_i^2 x_i$ with $x_i > 0$ and $\sum v_i = 1$ is uniquely minimized by $v_i = x_i^{-1}/\sum x_i^{-1}$ (Lagrange multiplier), the unconditional variance (26) is minimized by choosing

$$w_i(\Phi_i, \mathbf{T}) = \mathrm{Var}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \middle| \mathcal{A}_n^*\right]^{-1}.$$

The $w_i(\Phi_i, \mathbf{T})$ are $\mathcal{A}_n^*$-measurable by definition of the conditional variance. If $\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, \mathbf{T}) \middle| \mathcal{A}_n^*\right]$ is independent of the random ergodicity class Q, the weights satisfy (19)–(23) with $W_i \equiv 1$ and $w_i(\Phi_i, \mathbf{T}) = \mathrm{Var}\left[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \middle| \mathcal{A}_n^*\right]^{-1}$ and $\hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ being replaced by 1. $\square$

Note that an analog variance minimizing procedure could be included into the estimator $\hat{\mu}_f^{\alpha}$ of $\mu_f^{(2)}(I)$ as well. The result can also be generalized to arbitrary sub-σ-algebras $\mathcal{B}_n^*$ for which $\mathbb{E}[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \mid \mathcal{B}_n^*]$ is a.s. constant and $\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, \mathbf{T}) \middle| \mathcal{B}_n^*\right]$ is independent of the random ergodicity class Q. However, the above choice of $\mathcal{A}_n^*$ might be most relevant, particularly for models in which the marks are suitably independent of the point locations.

If there exist interaction effects in the MPP that are of higher than second order, the assumption on $\mathbb{E}[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) \mid \mathcal{A}_n^*]$ might not be satisfied anymore, and the estimator $\hat{\mu}_f^{n,\mathrm{Var}}(I)$ should be applied with care. Clusters of point locations which tend to increase the conditional variance of $\hat{\mu}_f$ given the ground process, can additionally influence the mean of other marks in excess of the bivariate interaction measured by $\mu_f^{(2)}(I)$. Then, a bias will be introduced using the above random weights. More generally, the more is known about the relation between $\hat{\mu}_f(I, \Phi, \mathbf{T})$ and the ground process Φ_{g}, the more can be gained from using different (random) weights while preserving consistency of the estimator. Without any assumption, only independent weights are feasible and then $w_i(\Phi_i, \mathbf{T}) = 1$ is naturally the best choice, i.e., the use of $\hat{\mu}_f^n(I)$.

In the following example, the marks are assumed to be independent of the locations:

Example 5 Let $\tilde{\Phi}$ be a one-dimensional, stationary unmarked point process and $\{Y(t) : t \in \mathbb{R}\}$ a stationary process, independent of $\tilde{\Phi}$, such that $\{f(Y(t)) : t \in \mathbb{R}\}$ has finite second moments. We consider the MPP $\Phi = \{(t, Y(t), 1) : t \in \tilde{\Phi}\}$. Then

$$\mathrm{Var}\left[\hat{\mu}_f(I, \Phi, T) \middle| \mathcal{A}_n^*\right] = \frac{\sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0,T]} \sum_{s_1 \in \Phi_{\mathrm{g}} \cap [0,T]} \mathrm{Cov}\big[f(Y(t_1)), f(Y(s_1))\big] n(t_1, \Phi_{\mathrm{g}}, I) n(s_1, \Phi_{\mathrm{g}}, I)}{\left[\sum_{t_1 \in \Phi_{\mathrm{g}} \cap [0,T]} n(t_1, \Phi_{\mathrm{g}}, I)\right]^2},$$

where $n(t_1, \Phi_{\mathrm{g}}, I) = \sum_{t_2 \in \Phi_{\mathrm{g}} \setminus \{t_1\}} \mathbf{1}_{t_2 - t_1 \in I}$.

The proof is given in Appendix B.

5 Conclusion

The MPP summary statistics considered in this paper are (weighted) mark means. Formally, we distinguish between two layers of weighting, namely the intrinsic weights

z_i in $\Phi = \{(t_i, y_i, z_i) : i \in \mathbb{N}\}$ by which single points are weighted, and the weights $w_j(\Phi_j, \mathbf{T})$ by which different realizations are weighted. In practical applications, though, these two might become indistinct: revisiting the forest example from the beginning, when a particular dataset of tree locations and marks is considered, this can be either treated as one single realization of a random forest or it can be treated as a collection of realizations, which may represent different forest areas with different characteristics. Independently of ergodicity properties of the corresponding stochastic process, the resulting estimators $\hat{\mu}_f$ and $\hat{\mu}_f^{n,\mathrm{wght}}$ coincide, if the z-component in the former interpretation is set to $z_i = w_j(\Phi_j, \mathbf{T})$ whenever the ith tree belongs to the jth realization of the latter interpretation.

The practical choice of weights depends on the statistical question and, once non-ergodicity is included into the model, on whether the classical moment measures $\mu_f^{(i)}$ or the two-stage expectations $\tilde{\mu}_f^{(i)}$ are to be estimated. Using the latter quantity particularly allows to exclude effects on the mark average caused by systematic differences in the pattern of point locations between the ergodicity classes.

An important family of MPP models, often used as null models, is obtained by geostatistical marking, i.e., by measuring a regionalized variable at random and stochastically independent locations (Illian et al. 2008). What might then be of interest is the average of the latent process over the whole index space and the weights z_i allow to compensate for clustering of point locations, for instance.

Whenever the points represent physical objects that influence each other, the assumption of a continuous-space background process becomes problematic. Trees in a forest, for example, compete for resources and if another tree had been added at some point, the measured characteristics of the surrounding trees would have likely changed. With increasing distance, however, these interaction effects between single objects of an MPP may become negligible and the latent process assumption can be sensible on a larger scale. These considerations motivate combining classical mark mean estimators for MPPs on the small scale (i.e., z being locally constant) with a weighting procedure that accounts for the irregular distribution of points on the larger scale. In some sense, this corresponds again to the non-ergodic approach with the two-step expectation, applied to the data without z-component.

Appendix A: Ergodic theory

Ergodicity is a mixing property that can be defined in the very general context of dynamical systems. A MPP on $\mathbb{R}^d$ together with the group of $\mathbb{R}^d$-indexed shift operators is a special case of a dynamical system.

We denote by $\mathbb{M}_0$ the set of all locally finite counting measures on $\mathbb{R}^d \times \mathbb{R}$, and by $\mathcal{M}_0$ the smallest σ-algebra on $\mathbb{M}_0$ that makes all mappings $\mathbb{M}_0 \to \mathbb{N}_0 \cup \infty$, $\varphi \mapsto \varphi(S)$, measurable. Formally, a MPP Φ is a measurable mapping from some probability space $(\Omega, \mathcal{A}, P)$ into $(\mathbb{M}_0, \mathcal{M}_0)$ and we can identify $(\Omega, \mathcal{A})$ with $(\mathbb{M}_0, \mathcal{M}_0)$ in the usual way. Let $\mathbf{T} = \{T_x : x \in \mathbb{R}^d\}$ with

$$(T_x\varphi)(B \times L) = \varphi((B + x), L), \qquad B \in \mathcal{B}^d, L \in \mathbb{R}. \tag{27}$$

Recall that Φ is said to be stationary if the induced probability measure P^{Φ} is $\mathbf{T}$-invariant. Further, a stationary MPP Φ is called ergodic if $P^{\Phi}(A)$ is either zero or one for all $\mathbf{T}$-invariant sets $A \in \mathcal{M}_0$. Let $\mathcal{A}_0 \subset \mathcal{M}_0$ be the sub-σ-algebra of all $\mathbf{T}$-invariant sets in $\mathcal{M}_0$, i.e., $A = T^{-1}A$ for all $A \in \mathcal{A}_0$ and $T \in \mathbf{T}$.

For the basic results in ergodic theory, the reader is referred to (Daley and Vere-Jones 2008, Chap. 12). The following Proposition provides an ergodic theorem for the point process context. The proof is based on a simple sandwich argument, which can also be used for other consistency statements. We include the proof here, because to our knowledge, it is not available in this form in pertinent literature. A similar assertion can be found in Daley and Vere-Jones (2008, Thm. 12.2.IV).

Proposition 3 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathbf{T} = \{T_x : x \in \mathbb{R}^d\}$ a group of measure-preserving transformations acting on $(\Omega, \mathcal{A}, P)$ such that the mapping $(T_x, \omega) \mapsto T_x\omega$ is jointly measurable, i.e., $(\mathcal{B}(\mathbf{T})\otimes\mathcal{A}, \mathcal{A})$-measurable. (Multiplication in $\mathbf{T}$ is given by $T_x T_y = T_{x+y}$.) Let $\{W_n\}_{n\in\mathbb{N}}$ be a convex averaging sequence in $\mathbb{R}^d$ and $\mathcal{A}_0$ the σ-algebra of $\mathbf{T}$-invariant events. Let Φ be stationary and ergodic and let $f : \mathbb{R}^d \times \mathbb{R} \times \mathbb{M}_0 \to \mathbb{R}$ be a non-negative function that satisfies $f(t-x, y, T_x\varphi) = f(t, y, \varphi)$ for all $t, x \in \mathbb{R}^d$, $y \in \mathbb{R}$, and that is integrable w.r.t. to the marked Campbell measure $C(B\times L\times M) = \mathbb{E}\big[\Phi((B\cap[0,1]^d)\times L)\mathbf{1}_M(\Phi)\big]$, $B \in \mathcal{B}^d$, $L \in \mathcal{L}$, $M \in \mathcal{M}_0$. We define random variables $X, X_n : \mathbb{M}_0 \to \mathbb{R}$ by*

$$
\begin{aligned}
X(\varphi) &= \sum_{(t,y)\in\varphi,\ t\in[0,1]^d} f(t, y, \varphi)\\
X_n(\varphi) &= \frac{1}{n^d} \sum_{(t,y)\in\varphi,\ t\in[0,n]^d} f(t, y, \varphi).
\end{aligned}
$$

Then X_n converges to $\mathbb{E}X$ almost surely if $n \to \infty$.

Proof An extension of the classical Campbell theorem (e.g., Daley and Vere-Jones 2008, Lem. 13.1.II) guarantees that $\mathbb{E}|X| < \infty$ if f is integrable w.r.t. the Campbell measure. The $W_n = [0, n]^d$ obviously form an averaging sequence and

$$
\begin{aligned}
X_n(\varphi) &= \frac{1}{\nu(W_n)} \sum_{(t,y)\in\varphi,\ t\in W_n} f(t, y, \varphi) \int_{\mathbb{R}^d} \mathbf{1}_{[t,\,t+1]}(x)\, \nu(dx)\\
&= \frac{1}{\nu(W_n)} \int_{\mathbb{R}^d} \sum_{(t,y)\in\varphi,\ t\in W_n\cap[x-1,\,x]} f(t, y, \varphi)\, \nu(dx), \qquad (28)
\end{aligned}
$$

where $x \pm 1$ for $x \in \mathbb{R}^d$ is defined component-wise. Note that the integrand on the RHS equals 0 whenever $W_n \cap [x-1,\ x] = \varnothing$, which means that x is not contained in $W_n \oplus [0, 1]^d$, which is, on its part, a subset of W_{n+1}. Thus, we can shrink the region of integration to W_{n+1} without changing the integral. If we then drop the condition '$t \in W_n$' under the summation sign, we enlarge the whole expression since f is non-negative, i.e.,

$$\begin{aligned} X_n(\varphi) &\leq \frac{1}{\nu(W_n)} \int_{W_{n+1}} \sum_{(t,y)\in\varphi,\ t\in[x-1,\,x]} f(t,\,y,\,\varphi)\,\nu(dx) \\ &= \frac{1}{\nu(W_n)} \int_{W_{n+1}} \sum_{(t,y)\in T_{x-1}\varphi,\ t\in[0,\,1]^d} f(t,\,y,\,T_{x-1}\varphi)\,\nu(dx) \\ &= \frac{\nu(W_{n+1})}{\nu(W_n)} \frac{1}{\nu(W_{n+1})} \int_{W_{n+1}-1} X(T_x\varphi)\,\nu(dx), \end{aligned} \tag{29}$$

where the second equation uses that $f(t-x,\,y,\,T_x\varphi) = f(t,\,y,\,\varphi)$ and the last equation uses that ν is shift-invariant. Since the ratio $\nu(W_{n+1})/\nu(W_n)$ converges to 1, the classical individual ergodic theorem (cf. Daley and Vere-Jones 2008, Prop. 12.2.II) yields that the RHS of (29) converges to $\mathbb{E}(X \mid \mathcal{A}_0)$ for almost all $\varphi \in \mathbb{M}_0$. Since Φ was assumed to be ergodic, this conditional expectation equals $\mathbb{E}X$. Similarly, if we restrict integration in (28) to the set W_{n-1}, we reduce the value of the integral. Since $W_{n-1} \oplus [-1,0]^d \subset W_n$, we can again drop the condition '$t \in W_n$' under the summation sign and by the same argument as before, we have

$$X_n(\varphi) \geq \frac{1}{\nu(W_n)} \int_{W_{n-1}} \sum_{(t,y)\in\varphi,\ t\in[x,\,x+1]} f(t,\,y,\,\varphi)\,\nu(dx) \overset{n\to\infty}{\longrightarrow} \mathbb{E}X$$

for almost all $\varphi \in \mathbb{M}_0$. Thus, we have a sandwich relation for $X_n(\varphi)$ and can conclude that $X_n \to \mathbb{E}X$ a.s. □

In case that Φ is not ergodic, the following results provide a representation of Φ as a mixture of a set of ergodic MPPs. To this end, let $\mathcal{P}$ ($\mathcal{P}_{\mathrm{erg}}$ resp.) denote the set of all probability measures on $(\mathbb{M}_0, \mathcal{M}_0)$ induced by stationary (and ergodic) MPPs and let Π_{erg} be the smallest σ-algebra making all mappings $\mathcal{P}_{\mathrm{erg}} \to [0,1]$, $P \mapsto P(A)$, measurable. We say that $\mathbf{T}$ fulfills the condition (LocCompGrp) if $\mathbf{T}$ is a locally compact, second-countable Hausdorff group of jointly measurable, surjective transformations. From Farrell (1962) we can extract a very general result:

Theorem 3 *Let $(\Omega, \mathcal{A})$ be a measurable space with Ω a complete separable metric space and $\mathcal{A}$ its Borel-σ-algebra. Let $\mathbf{T}$ be a set of measurable transformations of Ω satisfying the condition (LocCompGrp) and let $P \in \mathcal{P}$. Here, $\mathcal{P}$ ($\mathcal{P}_{\mathrm{erg}}$ resp.) is the set of all $\mathbf{T}$-invariant (and ergodic) probability measures on $(\Omega, \mathcal{A})$. Then there is a unique probability measure λ_P on $(\mathcal{P}_{\mathrm{erg}}, \Pi_{\mathrm{erg}})$ and a $\mathcal{P}_{\mathrm{erg}}$-valued random variable Q_P s.t.*

$$P(A) = \int_{\mathcal{P}_{\mathrm{erg}}} Q(A)\,\lambda_P(dQ) = \int_\Omega Q_P(\omega)(A)\,P(d\omega) \qquad \forall A \in \mathcal{A},$$

i.e., λ_P is the distribution of Q_P.

In the context of MPPs on $\mathbb{R}^d$, the group $\mathbf{T}$ of shifts, as defined in (27), obviously fulfills the condition (LocCompGrp), and since $\mathbb{M}_0$ is a complete separable metric space and $\mathcal{M}_0$ its Borel-σ-algebra (e.g., Kallenberg 1986), Theorem 3 can directly be applied, which yields a decomposition of the non-ergodic MPP $\Phi \sim P$:

$$P(M) = \int_{\mathcal{P}_{\text{erg}}} Q(M)\,\lambda(dQ) \qquad \forall M \in \mathcal{M}_0.$$

Note that each Q induces a new ergodic MPP $\Phi_Q : \Omega \to \mathbb{M}_0$ which is given implicitly by $P(\Phi_Q \in M) = Q(M)$, $M \in \mathcal{M}_0$. By the second representation in Theorem 3, we can also consider Q as a random variable on $(\mathbb{M}_0, \mathcal{M}_0, P)$ with distribution $\lambda = \lambda_P$. Thus, Φ and Q^{Φ} have a joint distribution and the conditional distribution of Φ given Q is well-defined:

$$P(\cdot \mid Q = q) = q(\cdot).$$

Appendix B: Proofs of Sect. 4

Proof (of Theorem 2) We consider

$$\begin{aligned}
&\left|\frac{\sum_{i=1}^n w_i(\Phi_i, \mathbf{T})\hat{\mu}_f(I, \Phi_i, \mathbf{T})}{\sum_{i=1}^n w_i(\Phi_i, \mathbf{T})} - \mu_f^{(2)}(I)\right| \\
&\quad \le \left|\frac{\sum_{i=1}^n w_i(\Phi_i, \mathbf{T})\big[\hat{\mu}_f(I, \Phi_i, \mathbf{T}) - \mu^{(2)}_{f,\Phi|Q=Q_i}(I)\big]}{\sum_{i=1}^n w_i(\Phi_i, \mathbf{T})}\right| && (30)\\
&\qquad + \left|\frac{\sum_{i=1}^n W_i\tilde{w}_i(\Phi_i, \mathbf{T})\mu^{(2)}_{f,\Phi|Q=Q_i}(I)}{\sum_{i=1}^n W_i\tilde{w}_i(\Phi_i, \mathbf{T})} - \mu_f^{(2)}(I)\right|. && (31)
\end{aligned}$$

By Lemma 1, $\hat{\mu}_f(I, \Phi_i, \mathbf{T})$ is consistent (for $\mathbf{T} \to \infty$) within the respective ergodicity class. Thus, (30) converges to 0 in probability if $\mathbf{T} \to \infty$. With the short notation $\alpha_i = \hat{\alpha}^{(2)}(C(\mathbf{T}, I), \Phi_i)$ and $\tilde{w}_i = \tilde{w}_i(\Phi_i, \mathbf{T})$, since α_i, W_i and $\tilde{w}_i$ are non-negative, we have

$$\begin{aligned}
(31) = &\left|\sum_{i=1}^n \frac{W_i\alpha_i\big[\mu^{(2)}_{f,\Phi|Q=Q_i}(I) - \mu_f^{(2)}(I)\big]}{\sum_{j=1}^n W_j\alpha_j} \times \frac{\tilde{w}_i\sum_{j=1}^n \alpha_j}{\alpha_i\sum_{j=1}^n \tilde{w}_j}\right| \\
&\times \frac{\sum_{j=1}^n W_j\alpha_j}{\sum_{j=1}^n \alpha_j} \times \frac{\sum_{j=1}^n \tilde{w}_j}{\sum_{j=1}^n W_j\tilde{w}_j}.
\end{aligned}$$

Since by assumption (20), $(n^{-1}\mathbb{E}\sum_{i=1}^n \tilde{w}_i)_{n\in\mathbb{N}}$ is eventually bounded away from 0 and the variance of the $\tilde{w}_i$ is uniformly bounded by (19), the law of large numbers yields that $\sum_{j=1}^n \tilde{w}_j/\mathbb{E}\sum_{j=1}^n \tilde{w}_j$ and $\sum_{j=1}^n \tilde{W}_j w_j/\mathbb{E}\sum_{j=1}^n W_j\tilde{w}_j$ converge to 1 in probability. Since the W_i are assumed to be iid, each W_i is either independent of $\tilde{w}_i$ or the functions $\tilde{w}_i(\cdot,\cdot)$ are identical for those indices for which W_i is not independent of $\tilde{w}_i$. Then $\frac{\mathbb{E}\sum_{j=1}^n \tilde{w}_j}{\mathbb{E}\sum_{j=1}^n W_j\tilde{w}_j} \to d_1$ for some $d_1 \in \mathbb{R}$ and we get

$$\frac{\sum_{j=1}^n \tilde{w}_j}{\sum_{j=1}^n W_j \tilde{w}_j} = \frac{\sum_{j=1}^n \tilde{w}_j / \mathbb{E}\sum_{j=1}^n \tilde{w}_j}{\sum_{j=1}^n W_j \tilde{w}_j / \mathbb{E}\sum_{j=1}^n W_j \tilde{w}_j} \times \frac{\mathbb{E}\sum_{j=1}^n \tilde{w}_j}{\mathbb{E}\sum_{j=1}^n W_j \tilde{w}_j} \xrightarrow{p} d_1$$

as $n \to \infty$. Similarly, due to (21) and (22), we have

$$\frac{\sum_{j=1}^n W_j \alpha_j}{\sum_{j=1}^n \alpha_j} \xrightarrow{p} \mathbb{E}W_1,$$

$$\frac{\sum_{i=1}^n W_i \alpha_i \mu^{(2)}_{f,\Phi|Q=Q_i}(I)}{\sum_{i=1}^n W_i \alpha_i} \xrightarrow{p} \frac{\mathbb{E}\left[\alpha^{(2)}_{\Phi|Q=Q_i}(C(I)) \times \mu^{(2)}_{f,\Phi|Q=Q_i}(I)\right]}{\mathbb{E}\left[\alpha^{(2)}_{\Phi|Q=Q_i}(C(I))\right]} = \mu_f^{(2)}(I). \tag{32}$$

It remains to show that $\sum_{i=1}^n A_i B_i \xrightarrow{p} 0$, where

$$A_i = \frac{W_i \alpha_i \left[\mu^{(2)}_{f,\Phi|Q=Q_i}(I) - \mu_f^{(2)}(I)\right]}{\sum_{j=1}^n W_j \alpha_j}, \quad B_i = \frac{\tilde{w}_i \sum_{j=1}^n \alpha_j}{\alpha_i \sum_{j=1}^n \tilde{w}_j}.$$

Equation (32) now reads as $\sum_{i=1}^n A_i \xrightarrow{p} 0$ and with $A_i^+ = \max(A_i, 0)$ and $A_i^- = -\min(A_i, 0)$, we have $\sum_{i=1}^n A_i^+ \xrightarrow{p} d_2$ and $\sum_{i=1}^n A_i^- \xrightarrow{p} d_2$ for some $d_2 \geq 0$. Since the B_i are non-negative, the following inequalities hold:

$$\min_{i=1}^n B_i \sum_{i=1}^n A_i^+ - \max_{i=1}^n B_i \sum_{i=1}^n A_i^- \leq \sum_{i=1}^n A_i B_i \leq \max_{i=1}^n B_i \sum_{i=1}^n A_i^+ - \min_{i=1}^n B_i \sum_{i=1}^n A_i^-. \tag{33}$$

Now

$$\mathbb{P}(\min B_i - 1 \geq -\varepsilon) = (1 - \mathbb{P}(B_1 - 1 < -\varepsilon))^n \geq (1 - \mathbb{P}(|B_1 - 1| > \varepsilon))^n \tag{34}$$

$$\mathbb{P}(\max B_i - 1 \leq \varepsilon) = (1 - \mathbb{P}(B_1 - 1 > \varepsilon))^n \geq (1 - \mathbb{P}(|B_1 - 1| > \varepsilon))^n. \tag{35}$$

By assumption (23), the RHS of (34) and (35) converge to 1, which means that $\min B_i \xrightarrow{p} 1$ and $\max B_i \xrightarrow{p} 1$. Hence, both the upper bound and the lower bound of $\sum_{i=1}^n A_i B_i$ in (33) converge to 0 in probability which finishes the proof. □

Proof (of Example 5) We have

$$\begin{aligned}
&\mathbb{E}\left[\hat{\alpha}_f(I, \Phi, T)/\hat{\alpha}_1(I, \Phi, T) \,\middle|\, \mathcal{A}_n^*\right] \\
&\quad = \hat{\alpha}_1(I, \Phi, T)^{-1} \times \mathbb{E}\left[\textstyle\sum_{(t_1,y_1,z_1),(t_2,y_2,z_2)\in\Phi,\ t_1\in[0,T]} z_1 f(y_1) \times \mathbf{1}_{t_2-t_1\in I} \,\middle|\, \mathcal{A}_n^*\right] \\
&\quad = \hat{\alpha}_1(I, \Phi, T)^{-1} \times \textstyle\sum_{t_1\in\Phi_g\cap[0,T]} \#\{t_2 \in \Phi_g : t_2 - t_1 \in I\} \times \mathbb{E}\left[f(Y(t_1))|\mathcal{A}_n^*\right] \\
&\quad = \mathbb{E}f(Y(0)).
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}\big[\hat{\alpha}_f(I,\Phi,T)^2 \mid \mathcal{A}_n^*\big] \\
&= \mathbb{E}\Big[\sum_{t_1,s_1\in\Phi_{\mathrm{g}}\cap[0,T]} f(Y(t_1)f(Y(s_1)) \\
&\quad \times \#\{t_2\in\Phi_{\mathrm{g}}: t_2-t_1\in I\}\times\#\{s_2\in\Phi_{\mathrm{g}}: s_2-s_1\in I\}\Big|\mathcal{A}_n^*\Big] \\
&= \sum_{t_1,s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I)n(s_1,\Phi_{\mathrm{g}},I) \\
&\quad \times \mathbb{E}\big[f(Y(t_1)f(Y(s_1))\mid\mathcal{A}_n^*\big] \\
&= \sum_{t_1,s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I)n(s_1,\Phi_{\mathrm{g}},I) \\
&\quad \times \Big[\mathbb{E}\big[f(Y(0))|\mathcal{A}_n^*\big]^2 + \mathrm{Cov}\big[f(Y(t_1)),f(Y(s_1))\mid\mathcal{A}_n^*\big]\Big] \\
&= \sum_{t_1,s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I)n(s_1,\Phi_{\mathrm{g}},I)\times\mathrm{Cov}\big[f(Y(t_1)),f(Y(s_1))\big] \\
&\quad + (\mathbb{E}f(Y(0)))^2\hat{\alpha}_1(I,\Phi,T)^2.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\mathrm{Var}\big[\hat{\alpha}_f(I,\Phi,T)/\hat{\alpha}_1(I,\Phi,T)\mid\mathcal{A}_n^*\big] \\
&\quad = \mathbb{E}\big[(\hat{\alpha}_f(I,\Phi,T)/\hat{\alpha}_1(I,\Phi,T)^2\mid\mathcal{A}_n^*\big] - \big(\mathbb{E}[\hat{\alpha}_f(I,\Phi,T)/\hat{\alpha}_1(I,\Phi,T)\mid\mathcal{A}_n^*]\big)^2 \\
&\quad = \hat{\alpha}_1(I,\Phi,T)^{-2}\times\mathbb{E}\big[\hat{\alpha}_f(I,\Phi,T)^2\mid\mathcal{A}_n^*\big] - (\mathbb{E}f(Y(0)))^2 \\
&\quad = \hat{\alpha}_1(I,\Phi,T)^{-2}\sum_{t_1,s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I)n(s_1,\Phi_{\mathrm{g}},I) \\
&\qquad \times\mathrm{Cov}\big[f(Y(t_1)),f(Y(s_1))\big].
\end{aligned}
$$

□

Acknowledgments We are grateful to two anonymous referees for many helpful comments. We are indebted to Katrin Meyer for useful discussions. A. Malinowski has been financially supported by the German Science Foundation (DFG), Research Training Group 1644 ‘Scaling problems in Statistics’.

References

Adelfio, G., Schoenberg, F. (2009). Point process diagnostics based on weighted second-order statistics and their asymptotic properties. *Annals of the Institute of Statistical Mathematics*, *61*(4), 929–948.

Adler, R. J., Taylor, J. E. (2007). *Random fields and geometry*. Springer Monographs in Mathematics. New York: Springer.

Baddeley, A. J. (1999). Spatial sampling and censoring. In O. E. Barndorff-Nielsen, W. S. Kendall, M. N. M. van Lieshout (Eds.), *Stochastic geometry, likelihood and computation* (pp. 37–78). Boca Raton: Chapman & Hall/CRC.

Bialkowski, J., Darolles, S., Le Fol, G. (2008). Improving VWAP strategies: A dynamic volume approach. *Journal of Banking & Finance*, *32*(9), 1709–1722.

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: Wiley.

Daley, D. J., Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I: elementary theory and methods* (2nd ed.). Probability and its Applications. New York: Springer.

Daley, D. J., Vere-Jones, D. (2008). *An introduction to the theory of point processes. Vol. II: general theory and structure* (2nd ed.). Probability and its Applications. New York: Springer.

Diggle, P. J., Menezes, R., Su, T. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C*, *59*(2), 191–232.

Farrell, R. H. (1962). Representation of invariant measures. *Illinois Journal of Mathematics*, *6*(3), 447–467.

Illian, J., Penttinen, P. A., Stoyan, H., Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Statistics in Practice. Chichester: Wiley.

Isaaks, E. H., Srivastava, R. M. (1989). *Applied geostatistics*. New York: Oxford University Press.

Isham, V. (1985). Marked point processes and their correlations. In F. Droesbeke (Ed.), *Spatial processes and spatial time series analysis* (pp. 63–75). Brussels: Publications des Facultés Universitaires Saint-Louis.

Kallenberg, O. (1986). *Random measures* (4th ed.). New York: Academic Press.

Karr, A. F. (1991). *Point processes and their statistical inference*. New York: Marcel Dekker Inc.

Madhavan, A. (2002). VWAP strategies. In B. Bruce (Ed.) *Investment guides, transaction performance: the changing face of trading* (pp. 32–38). New York: Institutional Investor Inc.

Møller, J., Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Boca Raton: Chapman & Hall/CRC.

Møller, J., Syversveen, A. R., Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, *25*(3), 451–482.

Myllymäki, M., Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica*, *63*(4), 450–473.

Schlather, M. (2001). On the second-order characteristics of marked point processes. *Bernoulli*, *7*(1), 99–117.

Schlather, M., Ribeiro, P. J, Jr, Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society, Series B*, *66*(1), 79–93.

Stoyan, D. (1984). On correlations of marked point processes. *Mathematische Nachrichten*, *116*, 197–207.

Stoyan, D., Stoyan, H. (1994). *Fractals, random shapes and point fields*. Chichester: Wiley.

Stoyan, D., Kendall, W. S., Mecke, J. (1995). *Stochastic geometry and its applications* (2nd ed.). Chichester: Wiley.

van Lieshout, M. N. M. (2006). A J-function for marked point patterns. *Annals of the Institute of Statistical Mathematics*, *58*(2), 235–259.