ESTIMATION OF ENTROPY FOR POISSON MARKED POINT PROCESSES

P. ALONSO-RUIZ,* and

E. SPODAREV,** *Ulm University*

Abstract

In this paper a kernel estimator of the differential entropy of the mark distribution of a homogeneous Poisson marked point process is proposed. The marks have an absolutely continuous distribution on a compact Riemannian manifold without boundary. We investigate L^2 and the almost surely consistency of this estimator as well as its asymptotic normality.

Keywords: Marked point process; kernel density estimator; entropy; central limit theorem; Boolean model

2010 Mathematics Subject Classification: Primary 62M30
Secondary 60G60; 60F05

1. Introduction

The concept of entropy was introduced by Shannon in the context of information theory [24] and its origin lies in the classical Boltzmann entropy of thermodynamics. In Shannon's original paper, entropy was defined both for discrete and continuous distributions in $\mathbb{R}^d$. In the last case it is called *differential entropy* and this notion can be naturally generalized as follows. Let P be a probability distribution of a random element X on an abstract measurable phase space (M, μ) with probability density f with respect to μ. The entropy of X is given by

$$\mathcal{E}_f = -\mathbb{E}_P(\log f(X)) = - \int_M f(x) \log f(x) \, \mu(\mathrm{d}x),$$

where the expectation $\mathbb{E}_P$ is taken with respect to the probability measure P.

In this paper we consider a homogeneous Poisson marked point process (MPP) with marks from a compact Riemannian manifold of dimension $p \geq 1$ without boundary that are assumed to be independent of the process, and investigate the differential entropy of the mark distribution $\mathcal{E}_f$. Our motivation for the study of this quantity is its applicability to detect inhomogeneities in materials modeled by MPPs such as fibre-reinforced plastics, where the direction of each fibre corresponds to a mark of the MPP. During the production process of such materials, the direction of the fibres may deviate from the predefined one and thus give rise to undesirable clusters or deformations. If the deviation is strong, a significant change on the (local) entropy of the directional distribution can be expected. Considering marks with values in a Riemannian manifold makes this method applicable not only to directions but to any other

characteristic of interest, for instance fibre length or fibre curvature. Asymptotic properties of such an estimator are important, in particular, for hypothesis testing.

In the present work we propose a nonparametric plug-in estimator of the differential entropy $\mathcal{E}_f$ based on [1]. It requires estimating the density of the distribution of interest in a nonparametric way, which we perform by means of *kernel density estimation*. This technique was introduced for stationary sequences of real random variables by Rosenblatt [21] and Parzen [16], and extended to stationary real random fields in [9]. In the case of finite samples of independent and identically distributed (i.i.d.) random vectors on the sphere, nonparametric kernel estimation methods were studied in [2], [10] and extended to Riemannian manifolds in [14], [18]. Alternative nonparametric estimators for the directional distribution in line and fibre processes were presented in [15].

The main result of our paper, Theorem 5.3, gives a central limit theorem (CLT) for an estimator of the differential entropy of the mark distribution density f of a homogeneous Poisson MPP as the observation window grows to $\mathbb{R}_+^d$ in a regular manner. This result is an application of a more general result (cf. Corollary 5.1) of this type for sequences of m_n-dependent random fields proved in Section 5. A CLT for nearest-neighbor estimates of entropy of a sample with values in a manifold is given in [19], see also the references therein.

The paper is organized as follows. Notation and basics of the theory of MPPs are given in Section 2. In Section 3 we construct a nonparametric kernel density estimator of f and give conditions for its L^2 and almost sure consistency. In Section 4 we introduce the nonparametric estimator $\widehat{\mathcal{E}}_f(B_n)$ of the entropy $\mathcal{E}_f$ in an observation window $B_n \subset \mathbb{R}^d$ and prove its L^2-consistency when the window size grows appropriately. Finally, in Section 5 we present a CLT for random sums of m_n-dependent random fields (cf. Corollary 5.1) where independence between the random number of summands and the summands themselves is not assumed. A special case of this result is applied to obtain a CLT of the entropy estimator.

2. Preliminaries

In this section we briefly review basic notions from the theory of MPPs. For an introduction and summary on these and other models of stochastic geometry we refer the reader to, e.g. [25], and [26].

2.1. Poisson MPPs

In the following, $\Pi := \{Y_i\}_{i \geq 1}$ will denote a *homogeneous Poisson point process* on $\mathbb{R}^d$ of intensity $\lambda > 0$ and (M, g) a compact smooth Riemannian manifold of dimension p without boundary and with Riemannian metric g. We further assume that (M, g) is complete, i.e. (M, d_g) is a complete metric space, where d_g denotes the geodesic distance induced by the Riemannian metric g. The associated Riemannian measure will be denoted by υ_g. A detailed construction of this measure can be found in, e.g. [22, p. 61]. Note that since M is compact, $\upsilon_g(M)$ is finite.

To each point $Y_i \in \Pi$ we attach a mark $\xi_i \in M$ and assume that marks are i.i.d. random variables independent of the location of the points in Π. The *Poisson MPP* $\Psi := \{(Y_i, \xi_i), Y_i \in \Pi\}$ we will work with is a random variable with values in $\mathcal{N} := \{\varphi$ locally finite counting measure on $\mathbb{R}^d \times M\}$. An important property of this process is *stationarity*, meaning that $T_y \Psi \overset{\mathrm{D}}{=} \Psi$ for all $y \in \mathbb{R}^d$, where the translation operator T_y is defined as $T_y \varphi(B \times L) := \varphi((B + y) \times L)$ for any Borel set $B \times L \subset \mathbb{R}^d \times M$ and $\varphi \in \mathcal{N}$. We denote equality in distribution by '$\overset{\mathrm{D}}{=}$'. We will assume that the distribution of a typical mark ξ_0 has a density $f : M \to \mathbb{R}$ with respect to the Riemannian volume measure υ_g.

Example 2.1. Poisson fibre process (cf. [26, Section 8]). A *fibre* $F \colon [0, 1] \to \mathbb{R}^2$ is a sufficiently smooth simple curve of finite length and a *fibre process* Φ is a random closed subset of $\mathbb{R}^2$ that can be represented as the union of at most countable many fibres F. To each fibre, we can attach a mark $\xi_F \in [0, \ell]$ that represents its (random) length. If the fibre process is Poisson distributed then $\Psi = \{(F, \xi_F), \ F \in \Phi\}$ and $M = [0, \ell]$.

Example 2.2. A Boolean model. Assume that $d \geq 3$ and consider, for each $1 \leq k \leq d - 1$, the Grassmannian $G(k, d)$, i.e. the set of all nonoriented k-dimensional flats in $\mathbb{R}^d$ that contain the origin (see, e.g. [22, p. 186]). This is a compact manifold of dimension $k(d - k)$. Furthermore, denote by $B(o, r)$ the ball of radius r centered at the origin $o \in \mathbb{R}^d$. The homogeneous Poisson point process $\Pi \subset \mathbb{R}^d$ leads to the Boolean model

$$\Phi := \bigcup_{Y_i \in \Pi} ((B(o, R_i) \cap Z_i) + Y_i),$$

where R_i and Z_i are independent copies of the random radius $R \colon \Omega \to [0, r]$ and the random Grassmannian $Z \colon \Omega \to G(k, d)$, respectively. The particular case $k = d - 1$ is used in applications to model lamellae structures, whereas the $k = 1$ case corresponds to a Poisson fibre process with straight fibres. In both cases, $G(k, d)$ is isomorphic to the half-sphere S_+^{d-1}. Based on this model, one can directly work with the MPP $\Psi = \{(Y_i, B(o, R_i) \cap Z_i)\}_{i \geq 1}$, with $M = [0, r] \times G(k, d)$ and $p = k(d - k)$. Here, one may be interested in the entropy of some specific characteristics of the grains, for instance their radius R and direction Z.

2.2. Space of marks

Since our mark space is a manifold, we recall in this section some useful concepts from Riemannian geometry. For further details we refer the reader to [5] and [22].

Let $\mathcal{T}_\eta M$ denote the tangent space of M at $\eta \in M$ and let $\exp_\eta \colon \mathcal{T}_\eta M \to M$ denote the exponential map. For any $r > 0$, $B_M(\eta, r) := \{v \in M \mid d_g(v, \eta) < r\}$ defines a neighborhood of η, that we call a *normal neighborhood of η* if there exists an open ball $V \subset \mathcal{T}_\eta M$ such that $\exp_\eta \colon V \to B_M(\eta, r)$ is a diffeomorphism. The *injectivity radius of M* is defined as $\operatorname{inj}_g M := \inf_{\eta \in M} \sup\{r \geq 0 \mid B_M(\eta, r) \text{ is a normal neighborhood of } \eta\}$.

Let U be a normal neighborhood of $\eta \in M$ and let (U, ψ) be the induced exponential chart of (M, g). For any $v \in U$, the *volume density function* introduced by Besse in [4, p. 154] is given by

$$\theta_\eta(v) := \left| \det\left(g_v\left(\frac{\partial}{\partial \psi_i}(v), \frac{\partial}{\partial \psi_j}(v) \right) \right)_{i, j = 1}^p \right|^{1/2},$$

where $g_v(\partial(v)/\partial \psi_i, \partial(v)/\partial \psi_j)$ denotes the metric g in normal coordinates at the point $\exp_\eta^{-1} v$ (see, e.g. [22, p. 24]). Note that this function is only defined for points $v \in U$ such that $d_g(\eta, v) < \operatorname{inj}_g M$. Since M is smooth, θ_η is continuous on M.

3. Kernel density estimator of the mark distribution

In this section we introduce a kernel density estimator for the density of the mark distribution on an observation window $B_n' \subset \mathbb{R}^d$. More precisely, we consider a sequence $\{B_n'\}_{n \in \mathbb{N}}$ of bounded Borel sets of $\mathbb{R}^d$ *growing in the Van Hove sense (VH-growing sequence)*. This means that

$$\lim_{n \to \infty} |B_n'| = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{|\partial B_n' \oplus B(o, r)|}{|B_n'|} = 0,$$

where $B(o, r)$ denotes the ball of radius $r > 0$ centered at the origin o. Given a set $B \subset \mathbb{R}^d$, $|B|$ will denote its d-dimensional Lebesgue measure, where d is the 'correct' dimension of B, i.e. the one for which B is a d-set. In this particular case, $|B'_n|$ is the d-dimensional volume of B'_n.

3.1. The estimator

Let $\Psi = \{(Y_i, \xi_i)\}_{i \geq 1}$ be an homogeneous Poisson MPP of intensity $\lambda > 0$. We define the kernel density estimator

$$\hat{f}_n(\eta) := \frac{1}{\lambda |B'_n|} \sum_{i \geq 1} \frac{1_{\{Y_i \in B'_n\}}}{b_n^p \theta_\eta(\xi_i)} K\left(\frac{d_g(\eta, \xi_i)}{b_n}\right),$$

where $1_{\{.\}}$ denotes the indicator function.

This is an extension of the estimator given in [18]. The sequence of bandwidths $\{b_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$ satisfies

(B1) $b_n < r_0$ for all $n \in \mathbb{N}$, with $0 < r_0 < \mathrm{inj}_g M$ and $\inf_{\eta \in B_M(z, r_0)} \theta_z(\eta) > 0$ for any $z \in M$,

(B2) $b_n \downarrow 0$,

(B3) $\lim_{n \to \infty} b_n^p |B'_n| = \infty$.

The kernel $K \colon \mathbb{R}_+ \to \mathbb{R}$ is a bounded nonnegative function satisfying

(K1) $\mathrm{supp}\, K = [0, 1]$,

(K2) $\int_{B(o,1)} K(\|x\|)\, \mathrm{d}x = 1$,

(K3) $0 < \int_{B(o,1)} K(\|x\|)\|x\|^2\, \mathrm{d}x =: K_2 < \infty$,

(K4) $\sup_{r \geq 0} K(r) =: K_0 < \infty$,

(K5) $\int_{B(o,1)} K(\|x\|)x\, \mathrm{d}x = o$.

We further assume that

(F1) $f \in L^2(M)$, i.e. $\|f\|_2^2 := \int_M |f(\eta)|^2\, \mathrm{d}v_g(\eta) < \infty$,

(F2) f is twice continuously differentiable.

Property (F2), in particular, means that f has bounded Hessian on any normal neighborhood $U \subset M$, i.e. there exists $C_2 > 0$ such that $\|D^2 f\| \leq C_2$.

Assumptions on the kernel are standard when dealing with nonparametric density estimation [18], [27]. For ease of notation, we will usually write

$$F_n(\eta, \xi) := \frac{1}{b_n^p \theta_\eta(\xi)} K\left(\frac{d_g(\eta, \xi)}{b_n}\right), \qquad \eta, \xi \in M.$$

In case the observation window B'_n needs to be explicitly indicated in the notation, we will write $\hat{f}_{B'_n}$ instead of $\hat{f}_n$.

3.2. Consistency

In this section we prove L^2 and almost sure consistency of $\hat{f}_n$. In what follows, ω_p will denote the volume of the unit ball in $\mathbb{R}^p$ and we will write $x \cdot y$ for the Euclidean scalar product of any two vectors $x, y \in \mathbb{R}^p$.

Note that in the classical (Euclidean) setting one could shorten proofs by applying Fourier methods [27]. However, in the general case of manifolds, this approach does not seem to be possible.

Theorem 3.1. *Under the assumptions* $(B1)$–$(B3)$, $(K1)$–$(K5)$, $(F1)$ *and* $(F2)$, *we have*

$$\mathbb{E}[\|\hat{f}_n - f\|_2^2] \leq \frac{C_\theta \omega_p K_0^2}{\lambda |B_n'| b_n^p} + b_n^4 C_2^2 K_2^2 \upsilon_g(M),$$

where $C_\theta := \sup_{z \in M} \sup_{\eta \in B_M(z, r_0)} \theta_z(\eta)^{-1}$.

Corollary 3.1. *Under the above assumptions, it follows directly from Theorem 3.1 that $\hat{f}_n$ is an L^2-consistent estimator of f, i.e.* $\mathbb{E}[\|\hat{f}_n - f\|_2^2] \to 0$ *as* $n \to \infty$.

Corollary 3.2. *Under the assumptions of Theorem 3.1, it holds that*

$$\mathbb{E}[|\hat{f}_n(\xi_0) - f(\xi_0)|^2] \to 0 \quad as\ n \to \infty.$$

In order to prove these results, we establish some auxiliary lemmas.

Lemma 3.1. *For each* $\eta \in M$ *and* $n \in \mathbb{N}$,

$$\int_{B_M(\eta, b_n)} \frac{1}{b_n^p \theta_\eta(z)} K\left(\frac{d_g(\eta, z)}{b_n}\right) \mathrm{d}\upsilon_g(z) = 1. \tag{3.1}$$

Proof. Consider the exponential chart (U, ψ) of (M, g) introduced in Section 2.2 and set $z := \exp_\eta(x)$, $B(0, b_n) := \exp_\eta B_M(\eta, b_n)$. Note that by definition (see [22, p. 65] for details) the Jacobian of the transformation $\|g(x)\|^{1/2}$ coincides with $\theta_\eta(\exp_\eta(x))$. The integral in (3.1) can thus be expressed as

$$\int_{B(0, b_n)} \frac{1}{b_n^p \theta_\eta(\exp_\eta(x))} K\left(\frac{\|x\|}{b_n}\right) \|g(x)\|^{1/2} \mathrm{d}x = \int_{B(0,1)} K(\|y\|) \mathrm{d}y = 1. \qquad \square$$

The calculations in the proof of this lemma lead to the useful equality

$$\int_{B_M(\eta, b_n)} \frac{1}{\theta_\eta(z)} \mathrm{d}\upsilon_g(z) = \int_{B(0, b_n)} \frac{\|g(x)\|^{1/2}}{\theta_\eta(\exp_\eta(x))} \mathrm{d}x = \int_{B(0, b_n)} \mathrm{d}x = b_n^p \omega_p. \tag{3.2}$$

We next give an asymptotic bound for the bias of $\hat{f}_n$.

Lemma 3.2. *For any* $\eta \in \operatorname{supp} f$ *and* $n \in \mathbb{N}$,

$$\operatorname{bias} \hat{f}_n(\eta) := |\mathbb{E}[\hat{f}_n(\eta)] - f(\eta)| \leq b_n^2 C_2 K_2.$$

Proof. Let $\eta \in \operatorname{supp} f$. By the Campbell theorem,

$$\mathbb{E}[\hat{f}_n(\eta)] = \int_M F_n(\eta, z) f(z) \mathrm{d}\upsilon_g(z) = \mathbb{E}[F_n(\eta, \xi_0)]. \tag{3.3}$$

Due to Lemma 3.1, and (K2), we have

$$|\mathbb{E}[F_n(\eta, \xi_0)] - f(\eta)| = \left| \int_{B_M(\eta, b_n)} \frac{1}{b_n^p \theta_\eta(z)} K\left(\frac{d_g(\eta, z)}{b_n}\right)(f(z) - f(\eta))\, \mathrm{d}\upsilon_g(z) \right|.$$

Consider now a normal neighborhood $\eta \in U \subset M$ and a point $x = (x^1, \ldots, x^p) \in \mathcal{T}_\eta M$ in normal coordinates, i.e. $z = \exp_\eta(x)$. Furthermore, define $\tilde{f} := f \circ \exp_\eta$. The Taylor expansion of $f(z)$ around η in normal coordinates is

$$f(z) = \tilde{f}(x) = \tilde{f}(0) + \nabla \tilde{f}(0) \cdot x + R_2(0, x),$$

where $R_2(0, x) = O(x^\top D^2 \tilde{f}(0)x)$ is the second-order remainder. From assumption (F2), we have $|R_2(0, x)| \leq C_2 \|x\|^2$ for all $x \in B(0, b_n)$, hence passing to the exponential chart as in the proof of Lemma 3.1 yields

$$\left| \int_{B_M(\eta, b_n)} \frac{1}{b_n^p \theta_\eta(z)} K\left(\frac{d_g(\eta, z)}{b_n}\right)(f(z) - f(\eta))\, \mathrm{d}\upsilon_g(z) \right|$$

$$= \left| \int_{B(0, b_n)} \frac{1}{b_n^p} \frac{1}{\theta_\eta(\exp_\eta(x))} K\left(\frac{\|x\|}{b_n}\right)(\tilde{f}(x) - \tilde{f}(0))\|g(x)\|^{1/2}\, \mathrm{d}x \right| \tag{3.4}$$

$$= \left| \int_{B(0, b_n)} \frac{1}{b_n^p} \frac{1}{\theta_\eta(\exp_\eta(x))} K\left(\frac{\|x\|}{b_n}\right) R_2(0, x)\|g(x)\|^{1/2}\, \mathrm{d}x \right| \tag{3.5}$$

$$\leq C_2 \int_{B(0, b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right)\|x\|^2\, \mathrm{d}x$$

$$= C_2 b_n^2 K_2.$$

Equality (3.5) follows from (K5) due to

$$\int_{B(0, b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) \nabla \tilde{f}(0) x\, \mathrm{d}x = \sum_{i=1}^{d} \int_{B(0, b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) \nabla \tilde{f}(0)_i x^i\, \mathrm{d}x$$

$$= \sum_{i=1}^{d} \nabla \tilde{f}(0)_i \int_{B(0, b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) x^i\, \mathrm{d}x$$

$$= \nabla \tilde{f}(0) \underbrace{\int_{B(0, b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) x\, \mathrm{d}x}_{=0}$$

$$= 0. \qquad \square$$

Lemma 3.3. *For any $n \in \mathbb{N}$,*

$$\int_M \mathbb{E}[F_n^2(\eta, \xi_0)]\, \mathrm{d}\upsilon_g(\eta) \leq \frac{C_\theta\, \omega_p K_0^2}{b_n^p},$$

with C_θ as in Theorem 3.1.

Proof. Applying Fubini's theorem, we write

$$\int_M \mathbb{E}[F_n^2(\eta, \xi_0)]\, \mathrm{d}\upsilon_g(\eta) = \int_M I(z) f(z)\, \mathrm{d}\upsilon_g(z), \tag{3.6}$$

where

$$I(z) = \int_{B_M(z,b_n)} \frac{1}{b_n^{2p}\theta_z^2(\eta)} K^2\left(\frac{d_g(\eta,z)}{b_n}\right) \mathrm{d}\upsilon_g(\eta).$$

Define $C_\theta(z) := \sup_{\eta \in B_M(z,r_0)} \theta_z(\eta)^{-1}$, which is finite because of (B1). By assumption (K4) and (3.2),

$$I(z) \le \frac{C_\theta(z)K_0^2}{b_n^p} \int_{B_M(z,b_n)} \frac{1}{b_n^p\theta_z(\eta)} \mathrm{d}\upsilon_g(\eta) = \frac{C_\theta(z)\omega_p K_0^2}{b_n^p}.$$

Substituting this estimate into (3.6) completes the proof. □

Proof of Theorem 3.1. By Fubini's theorem,

$$\mathbb{E}[\|\hat{f}_n - f\|_2^2] = \int_M \mathbb{E}[|\hat{f}_n(\eta) - f(\eta)|^2] \mathrm{d}\upsilon_g(\eta) =: \int_M J(\eta) \mathrm{d}\upsilon_g(\eta).$$

Note that $J(\eta) = \mathrm{var}(\hat{f}_n(\eta)) + (\mathrm{bias}\,\hat{f}_n(\eta))^2$. In view of (3.3) and the Campbell theorem, we obtain

$$
\begin{aligned}
\mathrm{var}(\hat{f}_n(\eta)) &= \mathbb{E}[\hat{f}_n^2(\eta)] - (\mathbb{E}[\hat{f}_n(\eta)])^2 \\
&= \frac{1}{\lambda^2|B_n'|^2} \mathbb{E}\left[\sum_{i \ge 1} 1_{\{Y_i \in B_n'\}} F_n^2(\eta, \xi_i)\right] \\
&\quad + \frac{1}{\lambda^2|B_n'|^2} \mathbb{E}\left[\sum_{i,j \ge 1}^{\ne} 1_{\{Y_i,Y_j \in B_n'\}} F_n(\eta,\xi_i) F_n(\eta,\xi_j)\right] - \mathbb{E}[F_n(\eta,\xi_0)]^2 \\
&= \frac{1}{\lambda|B_n'|} \mathbb{E}[F_n^2(\eta,\xi_0)] + \frac{\alpha^{(2)}(B_n' \times B_n')}{\lambda^2|B_n'|^2} \mathbb{E}[F_n(\eta,\xi_0)]^2 - \mathbb{E}[F_n(\eta,\xi_0)]^2 \\
&= \frac{1}{\lambda|B_n'|} \mathbb{E}[F_n^2(\eta,\xi_0)].
\end{aligned}
\tag{3.7}
$$

Here, $\alpha^{(2)}(\cdot)$ denotes the second-order factorial moment measure of the Poisson point process $\Pi := \{Y_i\}_{i \ge 1}$. We refer the reader to [26, Chapter 1] for further definitions and formulas related to this measure in the Poisson case. Lemmas 3.2 and 3.3 yield the existence of constants $C_\theta, C_2 > 0$ such that

$$\mathbb{E}[\|\hat{f}_n - f\|_2^2] \le \frac{C_\theta \omega_p K_0^2}{\lambda|B_n'|b_n^p} + b_n^4 C_2^2 K_2^2 \upsilon_g(M). \qquad \qquad \square$$

Analogous arguments show the L^2-convergence of $\hat{f}_n(\xi_0)$ to $f(\xi_0)$.

Proof of Corollary 3.2. Passing to normal coordinates as in (3.4) and (3.5) and setting $\tilde{f} := f \circ \exp_\eta$ lead to

$$\mathbb{E}[F_n(\eta,\xi_0)] = \int_{B(0,b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) \tilde{f}(x) \mathrm{d}x = (1 + o(1))f(\eta). \tag{3.8}$$

From the proof of Lemma 3.3, we thus obtain

$$\mathbb{E}[F_n^2(\eta,\xi_0)] \le \frac{K_0 C_\theta(\eta)}{b_n^p} \mathbb{E}[F_n(\eta,\xi_0)] \le \frac{2K_0 C_\theta(\eta)}{b_n^p} f(\eta) \quad \text{for any } \eta \in \mathrm{supp}\, f. \tag{3.9}$$

In view of (3.7) and Lemma 3.2, this yields

$$\mathbb{E}[|\hat{f}_n(\xi_0) - f(\xi_0)|^2] \leq \frac{2C_\theta K_0 \|f\|_2^2}{\lambda b_n^p |B_n'|} + b_n^4 C_2^2 K_2^2,$$

which tends to 0 as $n \to \infty$. $\qquad\square$

Remark 3.1. The problem of finding an optimal sequence of bandwidths $\{b_n\}_{n \in \mathbb{N}}$ can be understood as a special case of regularization [23] and the bound of the estimation error given in Theorem 3.1 can be used in order to find it. For any fixed $n \in \mathbb{N}$, the optimal bandwidth will be $\arg \min_{b_n} \mathbb{E}[\|\hat{f}_n - f\|_2^2]$. Applying Theorem 3.1, we can approximate the order of magnitude of this optimal b_n by minimizing the upper bound of the mean-square error

$$e(b_n) := \frac{C_\theta \omega_p K_0^2}{\lambda |B_n'| b_n^p} + b_n^4 C_2^2 K_2^2 \upsilon_g(M).$$

A simple calculation leads to the unique minimum point

$$b_{n,\text{opt}} = \left(\frac{p C_\theta \omega_p K_0^2}{4 C_2^2 K_2^2 \upsilon_g(M) \lambda |B_n'|} \right)^{1/(p+4)}.$$

Note that $b_{n,\text{opt}} \downarrow 0$ and $b_{n,\text{opt}}^p |B_n'| \to \infty$ as $n \to \infty$.

We finish this section by proving that if the observation window B_n' is large enough, then the previous bounds provide the almost surely (a.s.) consistency of $\hat{f}_n$.

Theorem 3.2. *Under the assumptions of Theorem 3.1, choosing $b_n = o(n^{-(1+\delta)/4})$ and B_n' such that $b_n^p |B_n'| > n^{1+\delta}$ for some $\delta > 0$,*

$$|\hat{f}_n(\eta) - f(\eta)| \to 0 \quad \text{as } n \to \infty \text{ a.s.}$$

for any $\eta \in M$ such that $f(\eta) < \infty$.

Proof. For each $\varepsilon > 0$, Chebyshev's inequality and the bounds used in the proof of Corollary 3.2 yield

$$\mathbb{P}(|\hat{f}_n(\eta) - f(\eta)| > \varepsilon) \leq \frac{\mathbb{E}[|\hat{f}_n(\eta) - f(\eta)|^2]}{\varepsilon^2} \leq \frac{2C_\theta K_0 f(\eta)}{\varepsilon^2 \lambda b_n^p |B_n'|} + \frac{b_n^4 C_2^2 K_2^2}{\varepsilon^2}.$$

Due to the choice of b_n, we have $b_n^p |B_n'| > n^{1+\delta}$ and $b_n^4 < n^{-(1+\delta)}$, hence,

$$\sum_{n=1}^{\infty} \mathbb{P}(|\hat{f}_n(\eta) - f(\eta)| > \varepsilon) \leq c_1 f(\eta) \sum_{n=1}^{\infty} \frac{1}{n^{1+\delta}} < \infty \quad \text{for some } c_1 < \infty.$$

The almost sure convergence follows from Borel–Cantelli's lemma. $\qquad\square$

4. Entropy estimator

As already mentioned in the introduction, we measure the diversity of the distribution of interest by analyzing its Kolmogorov entropy defined as

$$\mathcal{E}_f := - \int_M f(\eta) \log f(\eta) \, d\upsilon_g(\eta),$$

where f is the density of the distribution. This section is devoted to the construction of a consistent estimator for $\mathcal{E}_f$.

4.1. Definition of the estimator and consistency

For each $n \in \mathbb{N}$ we define

$$\widehat{\mathcal{E}}_f(B_n) := -\frac{1}{\lambda|B_n|} \sum_{i \geq 1} 1_{\{Y_i \in B_n\}} \log \hat{f}_{B'_n + Y_i}(\xi_i),$$

where $B'_n + y$ denotes the translation of B'_n by $y \in \mathbb{R}^d$ and $B'_n \subseteq B_n$. The window B'_n is introduced for the purpose of notation and it will become relevant when proving the CLT in Section 5. Throughout this section we have no restrictions on it and we can assume that $B_n = B'_n$.

From now on, we substitute the previous assumption (F1) by f being continuous. Note that since M is compact, the new (F1), in particular, implies the former. With the additional assumptions for a typical mark ξ_0,

(L1) $\mathbb{E}[\log^2 f(\xi_0)] =: L_1 < \infty$ and

(L2) $\mathbb{E}[(\|\nabla f(\xi_0)\| / f(\xi_0))^2] =: L_2 < \infty,$

we can prove L^2-consistency of the estimator.

Theorem 4.1. *For each $n \in \mathbb{N}$, let $\{B_n\}_{n \in \mathbb{N}}$ and $\{B'_n\}_{n \in \mathbb{N}}$ be sequences of VH-growing Borel sets satisfying $(B1)$–$(B3)$. Furthermore, assume that conditions $(K1)$–$(K5)$, $(F1)$, $(F2)$, $(L1)$, and $(L2)$ hold. Then,*

$$\mathbb{E}[|\widehat{\mathcal{E}}_f(B_n) - \mathcal{E}_f|^2] \leq 3\left(\frac{8K_0 C_\theta \upsilon_g(M)}{\lambda^2 |B_n||B'_n| b_n^p} + \frac{4}{\lambda^2 |B'_n|} + 32b_n^2 L_2 + \frac{L_1}{\lambda|B_n|}\right)$$

for sufficiently large $n \in \mathbb{N}$.

Corollary 4.1. *Under the above assumptions, it follows directly from Theorem 4.1 that $\widehat{\mathcal{E}}_f(B_n)$ is an L^2-consistent estimator of $\mathcal{E}_f$, i.e. $\mathbb{E}[|\widehat{\mathcal{E}}_f(B_n) - \mathcal{E}_f|^2] \to 0$ as $n \to \infty$.*

4.2. Proof of Theorem 4.1

We start by proving the following lemma assuming that all conditions of Theorem 4.1 are satisfied.

Lemma 4.1. *For sufficiently large $n \in \mathbb{N}$, it holds that*

$$\int_{\text{supp } f} \frac{(\mathbb{E}[\hat{f}_{B'_n}(\eta)] - f(\eta))^2}{f(\eta)} \, d\upsilon_g(\eta) \leq 4b_n^2 L_2.$$

Proof. Recall from (3.3) that $\mathbb{E}[\hat{f}_{B'_n}(\eta)] = \mathbb{E}[F_n(\eta, \xi_0)]$. Using normal coordinates analogously to (3.4) and (3.5) with $\tilde{f} := f \circ \exp_\eta$, we obtain

$$|\mathbb{E}[F_n(\eta, \xi_0)] - f(\eta)| = \left|\int_{B(0,b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) x \cdot \int_0^1 \nabla\tilde{f}(tx) \, dt \, dx\right|$$

$$\leq b_n \int_{B(0,1)} K(\|y\|)\|y\| \int_0^1 \|\nabla\tilde{f}(tb_n y)\| \, dt \, dy.$$

Since $b_n \downarrow 0$, we have $\|\nabla \tilde{f}(tb_n y)\| = \|\nabla \tilde{f}(0)\|(1 + o(1))$ for sufficiently large $n \in \mathbb{N}$ and in view of (K2), the last expression can be bounded by

$$2b_n \|\nabla \tilde{f}(0)\| \int_{B(0,1)} K(\|y\|) \|y\| \, \mathrm{d}y \le 2b_n \|\nabla \tilde{f}(0)\|.$$

Hence, $|\mathbb{E}[\hat{f}_{B'_n}(\eta)] - f(\eta)| \le 2b_n \|\nabla f(\eta)\|$ for sufficiently large $n \in \mathbb{N}$ and (L2) yields

$$\int_{\operatorname{supp} f} \frac{(\mathbb{E}[\hat{f}_{B'_n}(\eta)] - f(\eta))^2}{f(\eta)} \, \mathrm{d}\upsilon_g(\eta) \le 4b_n^2 \int_{\operatorname{supp} f} \frac{\|\nabla f(\eta)\|^2}{f(\eta)} \, \mathrm{d}\upsilon_g(\eta) = 4b_n^2 L_2. \qquad \square$$

We now proceed to prove Theorem 4.1. Based on [1], we introduce the quantities

$$L_n := -\frac{1}{\lambda |B_n|} \sum_{i \ge 1} 1_{\{Y_i \in B_n\}} \log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)], \qquad M_n := -\frac{1}{\lambda |B_n|} \sum_{i \ge 1} 1_{\{Y_i \in B_n\}} \log f(\xi_i).$$

Applying $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$, $a, b, c \in \mathbb{R}$, leads to

$$\mathbb{E}[|\widehat{\mathscr{E}}_f(B_n) - \mathscr{E}_f|^2] \le 3 \Big(\underbrace{\mathbb{E}[|\widehat{\mathscr{E}}_f(B_n) - L_n|^2]}_{=:I_{1,n}} + \underbrace{\mathbb{E}[|L_n - M_n|^2]}_{=:I_{2,n}} + \underbrace{\mathbb{E}[|M_n - \mathscr{E}_f|^2]}_{=:I_{3,n}} \Big),$$

hence, our aim is to compute an upper bound for $I_{i,n}$ and each $i = 1, 2, 3$. First,

$$\begin{aligned}
I_{1,n} = \;& \frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\left[\sum_{i \ge 1} 1_{\{Y_i \in B_n\}} (\log \hat{f}_{B'_n + Y_i}(\xi_i) - \log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)])^2 \right] \\
& + \frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\left[\sum_{i,j \ge 1}^{\ne} 1_{\{Y_i, Y_j \in B_n\}} (\log \hat{f}_{B'_n + Y_i}(\xi_i) - \log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)]) \right. \\
& \hspace{5cm} \left. \times (\log \hat{f}_{B'_n + Y_j}(\xi_j) - \log \mathbb{E}[\hat{f}_{B'_n + Y_j}(\xi_j)]) \right] \\
=: \;& J_1 + J_2.
\end{aligned}$$

On the one hand, note that, by definition,

$$h(Y_i, \xi_i, T_{Y_i} \Psi - \delta_{(o,\xi_i)}) := 1_{\{Y_i \in B_n\}} (\log \hat{f}_{B'_n + Y_i}(\xi_i) - \log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)])^2$$

depends on (Y_i, ξ_i) and $T_{Y_i} \Psi - \delta_{(o,\xi_i)}$. Since Ψ is an independently marked Poisson MPP, the Campbell–Mecke-type formula in [25, p. 129] yields

$$\frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\left[\sum_{i \ge 1} h(Y_i, \xi_i, T_{Y_i} \Psi - \delta_{(o,\xi_i)}) \right] = \frac{1}{\lambda |B_n|^2} \int_{\mathbb{R}^d} \int_M \mathbb{E}_{P_\eta^{o!}}[h(y, \eta, \Psi)] f(\eta) \, \mathrm{d}\upsilon_g(\eta) \, \mathrm{d}y,$$

where $\mathbb{E}_{P_{(o,\eta)}^!}$ denotes expectation with respect to the reduced Palm distribution of Ψ. Again because Ψ is an independently marked Poisson MPP, $P_{(o,\eta)}^!$ coincides with the distribution of Ψ and we obtain

$$J_1 = \frac{1}{\lambda |B_n|^2} \int_{B_n} \int_M \mathbb{E}[(\log \hat{f}_{B'_n + y}(\eta) - \log \mathbb{E}[\hat{f}_{B'_n + y}(\eta)])^2] f(\eta) \, \mathrm{d}\upsilon_g(\eta) \, \mathrm{d}y.$$

Note that $\log x$ is a differentiable function, hence the mean value theorem yields

$$|\log x - \log z| = \frac{|x - z|}{|(1 - \gamma)x + \gamma z|} \le \frac{|x - z|}{\min\{x, z\}}, \qquad x, z > 0 \tag{4.1}$$

for some $\gamma \in (0, 1)$. Since Ψ is stationary and, by assumption (F1), f is continuous, $\hat{f}_{B'_n+y}(\eta)$ converges to $f(\eta)$ a.s. for any $y \in \mathbb{R}^d$ and $\eta \in M$ by Theorem 3.2. Furthermore, in view of (3.8), $\mathbb{E}[F_n(\eta, \xi_0)] = (1 + o(1))f(\eta)$, hence, for $n \in \mathbb{N}$ large enough,

$$\min\{\hat{f}_{B'_n+y}(\eta), \mathbb{E}[\hat{f}_{B'_n+y}(\eta)]\} \ge \tfrac{1}{2}f(\eta). \tag{4.2}$$

Applying (4.1) with $x = \hat{f}_{B'_n+y}(\eta)$ and $z = \mathbb{E}[\hat{f}_{B'_n+y}(\eta)] = \mathbb{E}[F_n(\eta, \xi_0)]$, we obtain

$$J_1 \le \frac{4}{\lambda|B_n|^2} \int_{B_n} \int_M \frac{\mathbb{E}[(\hat{f}_{B'_n+y}(\eta) - \mathbb{E}[\hat{f}_{B'_n+y}(\eta)])^2]}{f(\eta)^2} f(\eta) \, d\upsilon_g(\eta) \, dy.$$

Due to (3.7) and (3.9),

$$J_1 \le \frac{4}{\lambda|B_n|} \int_M \frac{\mathbb{E}[F_n^2(\eta, \xi_0)]}{f(\eta)^2 \lambda|B'_n|} f(\eta) \, d\upsilon_g(\eta) \, dy \le \frac{8K_0 C_\theta \upsilon_g(M)}{\lambda^2|B_n||B'_n|b_n^p}. \tag{4.3}$$

Analogously, each summand in J_2 can be expressed as a function h depending on (Y_i, ξ_i), (Y_j, ξ_j), and $T_{Y_i}\Psi - \delta_{(o,\xi_i)} - \delta_{(Y_j,\xi_j)}$. Hence, the Campbell–Mecke-type formula in [25, p. 129] in the independently marked Poisson case yields

$$J_2 = \mathbb{E}\left[\sum_{i,j \ge 1}^{\ne} h(Y_i, \xi_i, Y_j, \xi_j, T_{Y_i}\Psi - \delta_{(o,\xi_i)} - \delta_{(Y_j,\xi_j)}) \right]$$

$$= \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{M^2} \mathbb{E}_{P_{\eta_1,\eta_2}^{o,y_2!}}[h(y_1, \eta_1, y_2, \eta_2, \Psi)]f(\eta_1)f(\eta_2) \, d\upsilon_g(\eta_2) \, d\upsilon_g(\eta_1) \, dy_1 \, dy_2$$

$$= \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{M^2} \mathbb{E}[h(y_1, \eta_1, y_2, \eta_2, \Psi)]f(\eta_1)f(\eta_2) \, d\upsilon_g(\eta_2) \, d\upsilon_g(\eta_1) \, dy_1 \, dy_2,$$

where the last inequality follows from the independent marking of the Poisson MPP. Applying again Theorem 3.2, (4.1), and (3.8), we obtain, for $n \in \mathbb{N}$ large enough,

$$J_2 \le \frac{4}{\lambda|B_n|^2} \int_{(B_n \times M)^2} \frac{\text{cov}(\hat{f}_{B'_n+y_1}(\eta_1), \hat{f}_{B'_n+y_2}(\eta_2))}{f(\eta_1)f(\eta_2)} f(\eta_1)f(\eta_2) \, d\upsilon_g(\eta_2) \, d\upsilon_g(\eta_1) \, dy_1 \, dy_2.$$

In view of (3.3) and the Campbell theorem,

$$\text{cov}(\hat{f}_{B'_n+y_1}(\eta_1), \hat{f}_{B'_n+y_2}(\eta_2))$$

$$= \mathbb{E}[\hat{f}_{B'_n+y_1}(\eta_1)\hat{f}_{B'_n+y_2}(\eta_2)] - \mathbb{E}[\hat{f}_{B'_n+y_1}(\eta_1)]\mathbb{E}[\hat{f}_{B'_n+y_2}(\eta_2)]$$

$$= \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\left[\sum_{i \ge 1} 1_{\{Y_i \in (B'_n+y_1) \cap (B'_n+y_2)\}} F_n(\eta_1, \xi_i)F_n(\eta_2, \xi_i) \right]$$

$$+ \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\left[\sum_{i,j \ge 1}^{\ne} 1_{\{Y_i \in B'_n+y_1\}} 1_{\{Y_j \in B'_n+y_2\}} F_n(\eta_1, \xi_i)F_n(\eta_2, \xi_j) \right]$$

$$- \mathbb{E}[F_n(\eta_1, \xi_0)]\mathbb{E}[F_n(\eta_2, \xi_0)]$$

$$= \frac{|(B'_n + y_1) \cap (B'_n + y_2)|}{\lambda |B'_n|^2} \mathbb{E}[F_n(\eta_1, \xi_0) F_n(\eta_2, \xi_0)]$$

$$\leq \frac{1}{\lambda |B'_n|} \mathbb{E}[F_n(\eta_1, \xi_0) F_n(\eta_2, \xi_0)].$$

Fubini's theorem and Lemma 3.1 yield

$$J_2 \leq \frac{4}{\lambda^2 |B'_n|} \int_{M^2} \mathbb{E}[F_n(\eta_1, \xi_0) F_n(\eta_2, \xi_0)] \, \mathrm{d}\upsilon_g(\eta_1) \, \mathrm{d}\upsilon_g(\eta_2) = \frac{4}{\lambda^2 |B'_n|},$$

which together with (4.3) leads to

$$I_{1,n} \leq \frac{8 K_0 C_\theta \upsilon_g(M)}{\lambda^2 |B_n| |B'_n| b_n^p} + \frac{4}{\lambda^2 |B'_n|}.$$

Secondly, due to the stationarity of Ψ and the Campbell theorem, we have, for large $n \in \mathbb{N}$,

$$I_{2,n} = \frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\left[\sum_{i \geq 1} 1_{\{Y_i \in B_n\}} (\log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)] - \log f(\xi_i))^2 \right]$$

$$+ \frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\left[\sum_{i,j \geq 1}^{\neq} 1_{\{Y_i, Y_j \in B_n\}} (\log \mathbb{E}[\hat{f}_{B'_n + Y_i}(\xi_i)] - \log f(\xi_i))(\log \mathbb{E}[\hat{f}_{B'_n + Y_j}(\xi_j)] \right.$$

$$\left. - \log f(\xi_j)) \right]$$

$$= \frac{1}{\lambda |B_n|} \mathbb{E}[(\log \mathbb{E}[\hat{f}_{B'_n}(\xi_0)] - \log f(\xi_0))^2] + (\mathbb{E}[\log \mathbb{E}[\hat{f}_{B'_n}(\xi_0)] - \log f(\xi_0)])^2$$

$$\leq 2 \mathbb{E}[(\log \mathbb{E}[\hat{f}_{B'_n}(\xi_0)] - \log f(\xi_0))^2].$$

On the other hand, by (4.2) and Lemma 4.1, we obtain

$$\mathbb{E}[(\log \mathbb{E}[\hat{f}_{B'_n}(\xi_0)] - \log f(\xi_0))^2] \leq 4 \int_{\mathrm{supp}\, f} \frac{(\mathbb{E}[\hat{f}_{B'_n}(\eta)] - f(\eta))^2}{f(\eta)} \, \mathrm{d}\upsilon(\eta) \leq 16 b_n^2 L_2,$$

so that $I_{2,n} \leq 32 b_n^2 L_2$.

Finally, note that $\mathcal{E}_f = -\mathbb{E}[\log f(\xi_0)]$. Applying once again the Campbell theorem, we obtain

$$I_{3,n} = \frac{1}{\lambda^2 |B_n|^2} \left(\mathbb{E}\left[\sum_{i \geq 1} 1_{\{Y_i \in B_n\}} \log^2 f(\xi_i) \right] + \mathbb{E}\left[\sum_{i,j \geq 1}^{\neq} 1_{\{Y_i, Y_j \in B_n\}} \log f(\xi_i) \log f(\xi_j) \right] \right)$$

$$+ \frac{2}{\lambda |B_n|} \mathbb{E}\left[\sum_{i \geq 1} 1_{\{Y_i \in B_n\}} \log f(\xi_i) \right] \mathcal{E}_f + \mathcal{E}_f^2$$

$$= \frac{1}{\lambda |B_n|} \mathbb{E}[\log^2 f(\xi_0)] + (\mathbb{E}[\log f(\xi_0)])^2 + 2\mathbb{E}[\log f(\xi_0)] \mathcal{E}_f + \mathcal{E}_f^2$$

$$= \frac{1}{\lambda |B_n|} \mathbb{E}[\log^2 f(\xi_0)]$$

$$= \frac{L_1}{\lambda |B_n|}.$$

This completes the proof of Theorem 4.1. □

Remark 4.1. The proof of Theorem 4.1 gives an explicit bound of the error that can be used to find an optimal sequence of bandwidths. In this case analogous calculations to Remark 3.1 lead to

$$b_{n,\mathrm{opt}} = \left(\frac{pK_0 C_\theta v_g(M)}{4L_2 \lambda^2 |B_n||B_n'|} \right)^{1/(p+2)}.$$

5. A CLT for entropy

If the window B_n' satisfies $B_n' \subset B_n$, and m_n is the diameter of B_n', the estimator $\widehat{\mathcal{E}}_f(B_n)$ can be seen as a normalized random sum of elements of a stationary m_n-*dependent random field*. In this section we present a CLT for a modified version of the original estimator.

We start by fixing some notation. In general, we use uppercase for coordinates and lowercase for enumerating elements. For $\mathbb{K} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{R}\}$, any $j \in \mathbb{K}^d$ will therefore be written as $j = (j^1, \ldots, j^d)$, while $j_1, j_2, \ldots$ will denote a sequence in $\mathbb{K}^d$. Moreover, we write $\boldsymbol{t} = (t, \ldots, t) \in \mathbb{K}^d$ for any $t \in \mathbb{K}$. We set $C_y := \times_{k=1}^d [0, y^k)$ for any $y \in \mathbb{R}_+^d$ and $V_j := C_j \cap \mathbb{N}^d$ for $j \in \mathbb{N}^d$. In particular, $C_{\boldsymbol{t}} = [0, t)^d$ for $t \in \mathbb{R}_+$.

A random field $\{X_j, \ j \in \mathbb{K}^d\}$ is said to be *m-dependent* for some $m > 0$ if for any finite sets $I, J \subset \mathbb{K}^d$ the random vectors $(X_i)_{i \in I}$ and $(X_j)_{j \in J}$ are independent whenever $\|i - j\|_\infty > m$ for all $i \in I$ and $j \in J$.

In stochastic geometry, m-dependent random fields often appear in connection with models based on independently marked MPPs. A CLT for sums of m-dependent random fields was first investigated by Rosén [20] and improved by Heinrich [11]. These results have been extended in the last years to weaker dependence structures (see [8] and [25], and references therein).

5.1. Theoretical results

Our CLT is based on the following result by Chen and Shao [8] for deterministic sums of m-dependent random fields.

Theorem 5.1. ([8, Theorem 2.6].) *Let $\{X_i\}_{i \in I}$, $I \subseteq \mathbb{N}^d$, be a centered m-dependent random field such that $\mathbb{E}[|X_i|^q] < \infty$ for some $2 < q \leq 3$ and any $i \in I$. Then,*

$$\sup_{x \in \mathbb{R}} |F(x) - \Phi(x)| \leq 75(10m + 1)^{(q-1)d} \left(\mathrm{var} \sum_{i \in I} X_i \right)^{-q/2} \sum_{i \in I} \mathbb{E}[|X_i|^q],$$

where F is the distribution function of $(\mathrm{var} \sum_{i \in I} X_i)^{-1/2} \sum_{i \in I} X_i$.

We give an extension of this theorem to random sums of stationary m_n-dependent random fields indexed in $\mathbb{R}_+^d$. For simplicity, we assume that our observation windows are cubic, i.e. $B_n := C_{\mathbf{p_n}}$ with $p_n \to \infty$ as $n \to \infty$.

Corollary 5.1. *Let $\{X_{n,y}, \ y \in B_n\}_{n \in \mathbb{N}}$ be a sequence of stationary centered m_n-dependent random fields and let Π be a stationary Poisson point process on $\mathbb{R}_+^d$. Assume that*

(A) $\sup_{n \in \mathbb{N}} \mathbb{E}[|\sum_{y \in \Pi \cap C_{\mathbf{1}}} X_{n,y}|^q] < \infty$ *for some $2 < q \leq 3$.*

Then,

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq 75(10m_n + 11)^{(q-1)d} |B_n| \sigma_n^{-q} \mathbb{E}\left[\left| \sum_{y \in \Pi \cap C_{\mathbf{1}}} X_{n,y} \right|^q \right],$$

where $\sigma_n^2 = \mathrm{var} \sum_{y \in \Pi \cap B_n} X_{n,y}$ and F_n is the distribution function of $\sum_{y \in \Pi \cap B_n} X_{n,y}/\sigma_n$.

Proof. For each $j \in \mathbb{N}^d$ and $n \in \mathbb{N}$, define $Z_{n,j} := \sum_{y \in \Pi \cap (C_1 + j)} X_{n,y}$. Obviously, $\{Z_{n,j}\}_{j \in V_{\mathbf{p_n}}}$ is a stationary centered $(m_n + 1)$-dependent random field with $\sup_{n \in \mathbb{N}} \mathbb{E}[|Z_{n,j}|^q] < \infty$ for any $j \in V_{\mathbf{p_n}}$ and $2 < q \leq 3$. Hence, an application of Theorem 5.1 with $I = V_{\mathbf{p_n}}$ yields the stated bound. $\qquad\square$

Remark 5.1. Note that Corollary 5.1 does not require independence between the random fields $\{X_{n,y}\}_{y \in B_n}$ and the point process Π. If independence is provided, the Campbell theorem together with the generalized Cauchy–Schwartz inequality and the stationarity of Π lead to

$$
\mathbb{E}\left[\left| \sum_{y \in \Pi \cap C_1} X_{n,y} \right|^3 \right] \leq \lambda \int_{C_1} \mathbb{E}[|X_{n,y}|^3]\,\mathrm{d}y + \lambda \int_{C_1^2} \mathbb{E}[X_{n,y_1}^2 |X_{n,y_2}|]\,\alpha^{(2)}(\mathrm{d}y_1,\,\mathrm{d}y_2)
$$

$$
+ \lambda \int_{C_1^3} \mathbb{E}[|X_{n,y_1} X_{n,y_2} X_{n,y_3}|]\,\alpha^{(3)}(\mathrm{d}y_1,\,\mathrm{d}y_2,\,\mathrm{d}y_3)
$$

$$
\leq \lambda \mathbb{E}[|X_{n,\mathbf{0}}|^3](1 + \alpha^{(2)}(C_1^2) + \alpha^{(3)}(C_1^3))
$$

$$
= \lambda \mathbb{E}[|X_{n,\mathbf{0}}|^3](1 + \lambda^2 + \lambda^3),
$$

where $\lambda > 0$ is the intensity of Π and $\alpha^{(k)}$, $k = 2, 3$, denotes the kth-order factorial moment measure of Π (see [26, Chapter 1] for explicit formulas in the Poisson case). Thus, we may substitute condition (A) by

(A′) $\sup_{n \in \mathbb{N}} \mathbb{E}[|X_{n,\mathbf{0}}|^3] < \infty$

and obtain Corollary 5.1 in the $q = 3$ case.

Before applying Corollary 5.1 and Remark 5.1 to our entropy estimator, we want to investigate under which conditions the limiting variance exists. The following theorem is an extension of [6, Theorem 1.8, p. 175] to random sums of wide-sense stationary random fields indexed in $\mathbb{R}^d$.

Theorem 5.2. *Let $\{X_{n,y}, \ y \in \mathbb{R}^d\}_{n \in \mathbb{N}}$ be a sequence of wide-sense stationary measurable centered random fields and let Π be a homogeneous Poisson point process of intensity $\lambda > 0$ independent of $\{X_{n,y}, \ y \in \mathbb{R}^d\}$. Assume that*

$$
\lim_{p \to \infty} \limsup_{n \to \infty} \int_{\mathbb{R}^d \setminus (-p,p)^d} |\mathrm{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y = 0, \tag{5.1}
$$

and

$$
\sup_{n \in \mathbb{N}} \int_{\mathbb{R}^d} |\mathrm{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y < \infty. \tag{5.2}
$$

If the limit

$$
\sigma^2 := \lim_{n \to \infty} \left(\lambda \mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2 \int_{\mathbb{R}^d} \mathrm{cov}(X_{n,\mathbf{0}}, X_{n,y})\,\mathrm{d}y \right)
$$

exists and is positive, then

$$
\frac{1}{|U_n|} \mathrm{var}\left(\sum_{y \in \Pi \cap U_n} X_{n,y} \right) \to \sigma^2 \quad \text{as } n \to \infty \tag{5.3}
$$

for any VH-growing sequence $\{U_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}^d$.

Proof. Since Π is a Poisson point process independent of $\{X_{n,y}\}_{y\in\mathbb{R}^d}$, it follows from the Campbell theorem and the wide-sense stationarity that

$$\text{var}\left(\sum_{y\in\Pi\cap U_n} X_{n,y}\right) = \lambda|U_n|\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2|U_n|\int_{\mathbb{R}^d}\text{cov}(X_{n,\mathbf{0}}, X_{n,y})\,\mathrm{d}y$$

$$-\lambda^2\int_{U_n}\int_{U_n^c}\text{cov}(X_{n,y_1}, X_{n,y_2})\,\mathrm{d}y_1\,\mathrm{d}y_2.$$

Following the proof of [6, Theorem 1.8], let $p > 0$ be arbitrary and set $G_n := U_n \cap (\partial U_n)_p$, $W_n := U_n \setminus G_n$, where $(\partial U_n)_p := \partial U_n \oplus B(\mathbf{0}, p)$ denotes the p-neighborhood of $\partial U_n \subset \mathbb{R}^d$. From the previous calculation, we have

$$\lambda|U_n|\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2|U_n|\int_{\mathbb{R}^d}\text{cov}(X_{n,\mathbf{0}}, X_{n,y})\,\mathrm{d}y - \text{var}\left(\sum_{y\in\Pi\cap U_n} X_{n,y}\right)$$

$$= \lambda^2\int_{G_n}\int_{U_n^c}\text{cov}(X_{n,y_1}, X_{n,y_2})\,\mathrm{d}y_1\,\mathrm{d}y_2 + \lambda^2\int_{W_n}\int_{U_n^c}\text{cov}(X_{n,y_1}, X_{n,y_2})\,\mathrm{d}y_1\,\mathrm{d}y_2$$

$$=: R_{n,1} + R_{n,2}.$$

On the one hand, $|G_n| \leq |(\partial U_n)_p|$ and since $\{U_n\}_{n\in\mathbb{N}}$ is VH-growing, assumption (5.2) yields

$$\frac{|R_{n,1}|}{|U_n|} \leq \frac{|(\partial U_n)_p|}{|U_n|}\lambda^2\int_{\mathbb{R}^d}|\text{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y \to 0 \quad \text{as } n \to \infty.$$

On the other hand, $\text{dist}(W_n, U_n^c) \geq p$ and $|W_n| \leq |U_n|$, hence,

$$\frac{|R_{n,2}|}{|U_n|} \leq \frac{|W_n|}{|U_n|}\lambda^2\int_{\mathbb{R}^d\setminus(-p,p)^d}|\text{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y$$

$$\leq \lambda^2\int_{\mathbb{R}^d\setminus(-p,p)^d}|\text{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y$$

and in view of assumption (5.1) the convergence in (5.3) is established. $\square$

The same holds under weaker assumptions if the random fields $\{X_{n,y}, \ y \in \mathbb{R}^d\}_{n\in\mathbb{N}}$ are m_n-dependent.

Corollary 5.2. *Let $\{X_{n,y}, \ y \in \mathbb{R}^d\}_{n\in\mathbb{N}}$ be a sequence of wide-sense stationary measurable centered m_n-dependent random fields and let Π be a homogeneous Poisson point process of intensity $\lambda > 0$ independent of $\{X_{n,y}, \ y \in \mathbb{R}^d\}_{n\in\mathbb{N}}$. Assume that*

$$\sup_{n\in\mathbb{N}}\int_{\mathbb{R}^d}|\text{cov}(X_{n,\mathbf{0}}, X_{n,y})|\,\mathrm{d}y < \infty. \tag{5.4}$$

If the limit

$$\sigma^2 := \lim_{n\to\infty}\left(\lambda\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2\int_{\mathbb{R}^d}\text{cov}(X_{n,\mathbf{0}}, X_{n,y})\,\mathrm{d}y\right)$$

exists and is positive, then

$$\lim_{n\to\infty}\frac{1}{|U_n|}\text{var}\left(\sum_{y\in\Pi\cap U_n} X_{n,y}\right) \to \sigma^2 \quad \text{as } n \to \infty$$

for any sequence of subsets $\{U_n\}_{n\in\mathbb{N}}$ satisfying $|(\partial U_n)_{m_n}|/|U_n| \to 0$ as $n \to \infty$.

Remark 5.2. The result holds, for instance, by taking cubic windows $U_n = C_{\mathbf{u_n}}$ with $m_n/u_n \to 0$ as $n \to \infty$.

Proof. Set $p = m_n$ in the proof of Theorem 5.2. Due to m_n-dependence, condition (5.1) is trivially fulfilled and, therefore, $\lim \sup_{n\to\infty} |R_{n,2}|/|U_n| = 0$. On the other hand,

$$\frac{|R_{n,1}|}{|U_n|} \leq \frac{|(\partial U_n)_{m_n}|}{|U_n|} \int_{\mathbb{R}^d} |\mathrm{cov}(X_{n,\mathbf{0}}, X_{n,y})| \, dy \to 0 \quad \text{as } n \to \infty$$

in view of assumption (5.4) and the choice of U_n. □

5.2. Application to entropy

The results of last paragraph evince that the independence between the Poisson point process Π and the sequence $\{X_{n,y}, \ y \in \mathbb{R}^d_+\}_{n\in\mathbb{N}}$ is crucial in order to perform calculations. Therefore, we need to consider the modified estimator

$$\widehat{\mathcal{E}}^*_f(B_n) := -\frac{1}{\lambda|B_n|} \sum_{i\geq 1} 1_{\{Y^*_i \in B_n\}} \log \hat{f}_{B'_n+y}(\xi^*_i),$$

where $\Psi^* := \{(Y^*_i, \xi^*_i)\}_{i\geq 1}$ is an independent copy of the original Poisson MPP Ψ. The study of the original estimator is the subject of further research and it involves MPPs whose marks depend of their location (we refer the reader to [12], [13], and [17] for some investigations in this direction). Moreover, we also need to assume that

(F3) $\inf_{\eta\in\mathrm{supp}\, f} f(\eta) := c_0 > 0.$

This assumption, although being very restrictive, is usual in the context of entropy estimation (see, e.g. [3]). We could substitute it by a set of slightly milder yet cumbersome assumptions and opted for the former for ease of proofs. The aim of this section is to apply Corollary 5.1 in order to obtain a CLT for $\widehat{\mathcal{E}}^*_f(B_n)$.

Theorem 5.3. *Let $\{B_n\}_{n\in\mathbb{N}}$ and $\{B'_n\}_{n\in\mathbb{N}}$ be sequences of observation windows in $\mathbb{R}^d_+$ with $B_n = C_{\mathbf{p_n}}, B'_n = C_{\mathbf{m_n}}$ for some $p_n, m_n > 0$. Under the conditions of Theorem 4.1, there exists $a > 0$ such that, for any $n \in \mathbb{N}$,*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{600 a\lambda(1+\lambda^2+\lambda^3)(10|B'_n|^{1/d}+11)^{2d}}{|B_n|^{1/2}}, \qquad (5.5)$$

where F_n is the distribution function of

$$\sqrt{|B_n|} \, \frac{\widehat{\mathcal{E}}^*_f(B_n) - \hat{\mu}_{B_n}}{\sigma_n}$$

with

$$\hat{\mu}_{B_n} := -\frac{\Pi^*(B_n)}{\lambda|B_n|} \mathbb{E}[\log \hat{f}_{B'_n}(\xi_{\mathbf{0}})]$$

and

$$\sigma_n^2 := \lambda \, \mathrm{var}(\log \hat{f}_{B'_n}(\xi_{\mathbf{0}})) + \lambda^2 \int_{B'_n} \mathrm{cov}(\log \hat{f}_{B'_n}(\xi_{\mathbf{0}}), \log \hat{f}_{B'_n}(\xi'_y)) \, dy,$$

where $\{\xi'_y\}_{y\in\mathbb{R}^d_+}$ are independent copies of $\xi_{\mathbf{0}}$.

Choosing a suitable size relation between B_n and B'_n leads to the desired CLT.

Corollary 5.3. *If the side-lengths of the observation windows satisfy $p_n = m_n^{4+\delta}$ for some $\delta > 0$ and any $n \in \mathbb{N}$, then*

$$\sqrt{|B_n|} \, \frac{\widehat{\mathcal{E}}_f^*(B_n) - \hat{\mu}_{B_n}}{\sigma_n} \xrightarrow{\text{D}} \mathcal{N}(0, 1) \quad \text{as } n \to \infty,$$

where '$\xrightarrow{\text{D}}$' denotes convergence in distribution with the uniform rate of convergence of order $m_n^{-\delta d/2}$ given in (5.5).

5.3. Proof of Theorem 5.3

First of all, note that

$$\sqrt{|B_n|} \, \frac{\widehat{\mathcal{E}}_f^*(B_n) - \hat{\mu}_{B_n}}{\sigma_n} =: \sum_{y \in \Pi^* \cap B_n} X_{n,y},$$

where

$$X_{n,y} = \frac{1}{\sqrt{|B_n|}\,\sigma_n} (-\log \hat{f}_{B_n'+y}(\xi_y^*) + \mathbb{E}[\log \hat{f}_{B_n'}(\xi_0)])$$

is a stationary centered m_n-dependent random field with variance 1. Our strategy will thus consist in verifying condition (A′) and computing the bound given by Corollary 5.1. In order to do so we prove next some helpful lemmas.

For the ease of notation, we use $\hat{f}_{B_n'+y}$ instead of $\hat{f}_{B_n'+y}(\xi_y')$ and only refer explicitly to the argument when confusion may occur. Moreover, we assume that the conditions of Theorem 5.3 hold in the subsequent lemmas without mentioning them explicitly.

We begin by proving the uniform boundedness of the third moment.

Lemma 5.1. *There exists a constant $c_1 > 0$ such that for any $y \in \mathbb{R}_+^d$ and $n \in \mathbb{N}$*

$$\mathbb{E}[|\log \hat{f}_{B_n'+y}|^3] \leq c_1.$$

Proof. Due to stationarity, it suffices to show that the assertion holds for $\mathbb{E}[|\log \hat{f}_{B_n'}|^3]$. On the one hand, by adding and subtracting $\log \mathbb{E}[\hat{f}_{B_n'}]$, we have

$$\mathbb{E}[|\log \hat{f}_{B_n'}|^3] \leq \mathbb{E}[|\log \hat{f}_{B_n'} - \log \mathbb{E}[\hat{f}_{B_n'}]|^3] + 3\,|\log \mathbb{E}[\hat{f}_{B_n'}]|\mathbb{E}[|\log \hat{f}_{B_n'} - \log \mathbb{E}[\hat{f}_{B_n'}]|^2]$$
$$+ 3(\log \mathbb{E}[\hat{f}_{B_n'}])^2 \mathbb{E}[|\log \hat{f}_{B_n'} - \log \mathbb{E}[\hat{f}_{B_n'}]|] + |\log \mathbb{E}[\hat{f}_{B_n'}]|^3.$$

By Corollary 3.2, $\log \mathbb{E}[\hat{f}_{B_n'}] \to \log \mathbb{E}[f(\xi_0)]$ as $n \to \infty$. In view of (F3) and since f is continuous, any power of this quantity is also bounded. Thus, it suffices to show that $\mathbb{E}[|\log \hat{f}_{B_n'} - \log \mathbb{E}[\hat{f}_{B_n'}]|^3] < \infty$. For $n \in \mathbb{N}$ large, (4.1), (4.2), and assumption (F3) yield

$$\mathbb{E}[|\log \hat{f}_{B_n'} - \log \mathbb{E}[\hat{f}_{B_n'}]|^3] \leq \frac{8\mathbb{E}[|\hat{f}_{B_n'} - \mathbb{E}[\hat{f}_{B_n'}]|^3]}{c_0^3} \quad \text{for } n \in \mathbb{N} \text{ large,}$$

hence, it suffices to prove that $\mathbb{E}[|\hat{f}_{B_n'}|^3]$ is finite. Due to the Campbell theorem,

$$\mathbb{E}[|\hat{f}_{B_n'}(\xi_0')|^3] = \frac{1}{\lambda^2|B_n'|^2}\mathbb{E}[F_n^3(\xi_0', \xi_1)] + \frac{1}{\lambda|B_n'|}\mathbb{E}[F_n^2(\xi_0', \xi_1)F_n(\xi_0', \xi_2)]$$
$$+ \mathbb{E}[F_n(\xi_0', \xi_1)F_n(\xi_0', \xi_2)F_n(\xi_0', \xi_3)], \tag{5.6}$$

where ξ_1, ξ_2, ξ_3 are independent copies of $\xi'_{\mathbf{0}}$. Moreover, following the proof of Lemma 3.3 we find constants C_θ, $K_0 > 0$ such that, for $n \in \mathbb{N}$ large enough,

$$\mathbb{E}[F_n^3(\xi'_{\mathbf{0}}, \xi_1)] \leq \frac{C_\theta^2 K_0^2}{b_n^{2p}} (1 + o(1)) \mathbb{E}[f(\xi'_{\mathbf{0}})],$$

$$\mathbb{E}[F_n^2(\xi'_{\mathbf{0}}, \xi_1) F_n(\xi'_{\mathbf{0}}, \xi_2)] \leq \frac{C_\theta K_0}{b_n^p} (1 + o(1)) \mathbb{E}[f^2(\xi'_{\mathbf{0}})],$$

as well as

$$\mathbb{E}[F_n(\xi'_{\mathbf{0}}, \xi_1) F_n(\xi'_{\mathbf{0}}, \xi_2) F_n(\xi'_{\mathbf{0}}, \xi_3)] \leq (1 + o(1)) \int_M f(\eta)^4 \, \mathrm{d}\upsilon_g(\eta) = (1 + o(1)) \mathbb{E}[f^3(\xi'_{\mathbf{0}})].$$

Substituting this into (5.6), we obtain

$$\mathbb{E}[|\hat{f}_{B'_n}|^3] \leq \frac{2C_\theta^2 K_0^2}{\lambda^2 b_n^{2p} |B'_n|^2} \mathbb{E}[f(\xi'_{\mathbf{0}})] + \frac{2C_\theta K_0}{b_n^p \lambda |B'_n|} \mathbb{E}[f^2(\xi'_{\mathbf{0}})] + 2\mathbb{E}[f^3(\xi'_{\mathbf{0}})]$$

for $n \in \mathbb{N}$ sufficiently large. This quantity is bounded because all expressions depending on n tend to 0 as $n \to \infty$. $\qquad\square$

In the consequent lemmas we show that σ_n^2 is uniformly bounded.

Lemma 5.2. *There exists $c_2 > 0$ such that, for any $x_1, x_2 \in B_n$ and $n \in \mathbb{N}$,*

$$\mathrm{cov}(\log \hat{f}_{B'_n + x_1}, \log \hat{f}_{B'_n + x_2}) \leq c_2 \, \mathrm{cov}(\hat{f}_{B'_n + x_1}, \hat{f}_{B'_n + x_2}).$$

Proof. Adding and subtracting $\log \mathbb{E}[\hat{f}_{B'_n + y_1}]$ respectively $\log \mathbb{E}[\hat{f}_{B'_n + y_2}]$, Theorem 3.2, (4.2), and assumption (F3) lead to

$$\begin{aligned}
\mathrm{cov}(\log \hat{f}_{B'_n + x_1}, \log \hat{f}_{B'_n + x_2}) &= \mathbb{E}[(\log \hat{f}_{B'_n + x_1} - \log \mathbb{E}[\hat{f}_{B'_n}])(\log \hat{f}_{B'_n + x_2} - \log \mathbb{E}[\hat{f}_{B'_n}])] \\
&\quad - (\mathbb{E}[\log \hat{f}_{B'_n}] - \log \mathbb{E}[\hat{f}_{B'_n}])^2 \\
&\leq \frac{4}{c_0^2} \, \mathrm{cov}(\hat{f}_{B'_n + x_1}, \hat{f}_{B'_n + x_2})
\end{aligned}$$

for $n \in \mathbb{N}$ sufficiently large. The result now follows for any $n \in \mathbb{N}$ with a constant $c_2 > 0$ (maybe different from $4/c_0^2$). $\qquad\square$

Lemma 5.3. *There exists $c_3 > 0$ such that, for any $n \in \mathbb{N}$ and $x_1, x_2 \in B_n$,*

$$\mathrm{cov}(\hat{f}_{B'_n + x_1}, \hat{f}_{B'_n + x_2}) \leq \frac{c_3 |(B'_n + x_1) \cap (B'_n + x_2)|}{\lambda |B'_n|^2}.$$

Proof. Applying the Campbell theorem,

$$\mathrm{cov}(\hat{f}_{B'_n + x_1}, \hat{f}_{B'_n + x_2}) = \mathbb{E}[\hat{f}_{B'_n + x_1} \hat{f}_{B'_n + x_2}] - (\mathbb{E}[\hat{f}_{B'_n}])^2$$

$$
\begin{aligned}
&= \frac{1}{\lambda^2 |B_n'|^2} \mathbb{E}\left[\sum_{y \in \Pi \cap (B_n' + x_1) \cap (B_n' + x_2)} F_n(\xi_y, \xi_{x_1}') F_n(\xi_y, \xi_{x_2}') \right] \\
&\quad + \frac{1}{\lambda^2 |B_n'|^2} \mathbb{E}\left[\sum_{y_1 \in \Pi \cap (B_n' + x_1),\, y_2 \in \Pi \cap (B_n' + x_2)}^{\neq} F_n(\xi_{y_1}, \xi_{x_1}') F_n(\xi_{y_2}, \xi_{x_2}') \right] \\
&\quad - (\mathbb{E}[F_n(\xi_0, \xi_{x_1}')])^2 \\
&= \frac{|(B_n' + x_1) \cap (B_n' + x_2)|}{\lambda |B_n'|^2} \mathbb{E}[F_n(\xi_0, \xi_{x_1}') F_n(\xi_0, \xi_{x_2}')] \\
&\quad + \frac{|(B_n' + x_1) \cap (B_n' + x_2)|}{\lambda |B_n'|^2} (\mathbb{E}[F_n(\xi_0, \xi_{x_1}')])^2.
\end{aligned}
$$

Furthermore, it follows from (3.8) that, for $n \in \mathbb{N}$ large enough,

$$
\begin{aligned}
\mathbb{E}[F_n(\xi_0, \xi_{x_1}') F_n(\xi_0, \xi_{x_2}')] &= \int_{M^3} F_n(\mu, z) F_n(z, \eta) f(\mu) f(z) f(\eta) \, \mathrm{d}\upsilon_g(\mu, z, \eta) \\
&= (1 + o(1)) \int_M f(z)^3 \, \mathrm{d}\upsilon_g(z) \\
&= (1 + o(1)) \mathbb{E}[f^2(\xi_0)]
\end{aligned}
$$

as well as

$$
\mathbb{E}[F_n(\xi_0, \xi_{x_1}')] = \int_{M^2} F_n(\mu, z) f(\mu) f(z) \, \mathrm{d}\upsilon_g(\mu, z) = (1 + o(1)) \mathbb{E}[f(\xi_0)].
$$

Thus, the assertion holds with $c_3 = 2\mathbb{E}[f^2(\xi_0)] + 4(\mathbb{E}[f(\xi_0)])^2 > 0$ for $n \in \mathbb{N}$ large and for any $n \in \mathbb{N}$ with maybe a different constant $c_3 > 0$. $\qquad\square$

Finally, by Corollary 3.2 and analogous arguments involved in (4.1)–(4.3), it follows that $\log \hat{f}_{B_n'}(\xi_0)$ converges to $\log f(\xi_0)$ in L^2. Therefore, $\mathbb{E}[\log^2 \hat{f}_{B_n'}] \to \mathbb{E}[\log^2 f(\xi_0)]$ as $n \to \infty$ and since $\mathbb{E}[\log^2 f(\xi_0)] < L_1$ by assumption (L1), $\mathbb{E}[\log^2 \hat{f}_{B_n'}]$ can be bounded by some constant $\tilde{L}_1 > 0$ uniformly on $n \in \mathbb{N}$. On the other hand, Lemmas 5.2 and 5.3, and the m_n-dependence yield

$$
\begin{aligned}
\int_{\mathbb{R}^d} |\mathrm{cov}(\log \hat{f}_{B_n'}, \log \hat{f}_{B_n'+y})| \, \mathrm{d}y &= \int_{B_n'} |\mathrm{cov}(\log \hat{f}_{B_n'}, \log \hat{f}_{B_n'+y})| \, \mathrm{d}y \\
&\leq \frac{c_1 c_2}{\lambda |B_n'|^2} \int_{B_n'} |B_n' \cap (B_n' + y)| \, \mathrm{d}y \\
&= \frac{c_1 c_2}{\lambda^2 2^d} < \infty.
\end{aligned}
$$

In the next lemmas we ensure that σ_n^2 can be uniformly bounded from below. Recall that we are assuming that the density f is continuous.

Lemma 5.4. *The estimator $\hat{f}_{B_n'+y}(\xi_y')$ is uniformly bounded with respect to $y \in \mathbb{R}_+^d$ and $n \in \mathbb{N}$ almost surely.*

Proof. By stationarity it suffices to prove the assertion for $\hat{f}_{B_n'}(\xi_0)$. Note that ξ_0 is a generic mark that is independent of the MPP Ψ. From Theorem 3.2 and since M is compact and f continuous, we have $\hat{f}_n(\eta) \to f(\eta) \leq \|f\|_\infty$ as $n \to \infty$ a.s., and, hence, $\hat{f}_n(\eta) \leq \|f\|_\infty + \varepsilon$ a.s. for any $\varepsilon > 0$ and $n \in \mathbb{N}$. The same holds for $\hat{f}_{B_n'}(\xi_0)$. $\qquad\square$

Lemma 5.5. *There exists $c_4 > 0$ such that*

$$\liminf_{n \to \infty} \int_{B'_n} \mathrm{cov}(\log \hat{f}_{B'_n}(\xi_0), \log \hat{f}_{B'_n + y}(\xi'_y)) \, \mathrm{d}y \geq c_4.$$

Proof. Since Π is a Poisson point process, we know from [7] that it is positively associated. On the other hand, the random variables $\{\xi'_y\}_{y \in \mathbb{R}^d}$ are positively associated as well because they are i.i.d. (see [6, Theorem 1.8]). Therefore, by [6, Corollary 1.9], the random field $\{\hat{f}_{B'_n + y}(\xi'_y)\}_{y \in \mathbb{R}^d_+}$ is positively associated. Using the characterization of positively associated random fields given in [6, Remark 1.4], this means that, for any nondecreasing functions $h, g : \mathbb{R} \to \mathbb{R}$ such that the expectations forming the covariance $\mathrm{cov}(h(\hat{f}_{B'_n + y_1}), g(\hat{f}_{B'_n + y_2}))$ exist, $\mathrm{cov}(h(\hat{f}_{B'_n + y_1}), g(\hat{f}_{B'_n + y_2})) \geq 0$. In view of Lemma 5.1, we thus have $\mathrm{cov}(\log \hat{f}_{B'_n + y_1}, \log \hat{f}_{B'_n + y_2}) \geq 0$ and since $\log$ is an increasing function, the random field $\{\log \hat{f}_{B'_n + y}\}_{y \in \mathbb{R}^d_+}$ is also positively associated.

From Lemma 5.4 we know that $\hat{f}_{B'_n} \leq \|f\|_\infty + \varepsilon$ a.s. for large $n \in \mathbb{N}$, and following the proof of [6, Theorem 5.3] with the exponential function, we obtain

$$\mathrm{cov}(\log \hat{f}_{B'_n}, \log \hat{f}_{B'_n + y}) \geq \frac{1}{2(\|f\|_\infty + \varepsilon)^2} \mathrm{cov}(\hat{f}_{B'_n}, \hat{f}_{B'_n + y}).$$

Together with the calculations in the proof of Lemma 5.3, this yields

$$\begin{aligned}
\int_{\mathbb{R}^d} \mathrm{cov}(\log \hat{f}_{B'_n}, \log \hat{f}_{B'_n + y}) \, \mathrm{d}y &\geq \frac{\mathbb{E}[f^2(\xi_0)] + (\mathbb{E}[f(\xi_0)])^2}{4(\|f\|_\infty + \varepsilon)^2 \lambda |B'_n|^2} \int_{B'_n} |B'_n \cap (B'_n + y)| \, \mathrm{d}y \\
&= \frac{\mathbb{E}[f^2(\xi_0)] + (\mathbb{E}[f(\xi_0)])^2}{4(\|f\|_\infty + \varepsilon)^2 \lambda m_n^{2d}} \left(\int_0^{m_n} (m_n - y) \, \mathrm{d}y \right)^d \\
&= \frac{\mathbb{E}[f^2(\xi_0)] + (\mathbb{E}[f(\xi_0)])^2}{(\|f\|_\infty + \varepsilon)^2 \lambda 2^{d+2}} \\
&=: c_4 > 0
\end{aligned}$$

and the result follows with maybe a different constant c_4. $\square$

Proof of Theorem 5.3. Recall that

$$X_{n, y} = \frac{1}{\sqrt{|B_n|} \sigma_n} (-\log \hat{f}_{B'_n + y}(\xi^*_y) + \mathbb{E}[\log \hat{f}_{B'_n}(\xi_0)]).$$

On the one hand, applying the Cauchy–Schwartz inequality, Lemma 5.1 and Lemma 5.5, we obtain, for $n \in \mathbb{N}$ large,

$$\mathbb{E}[|X_{n, 0}|^3] \leq \frac{8\mathbb{E}[|\log \hat{f}_{B'_n}(\xi^*_0)|^3]}{|B_n|^{3/2} \sigma_n^3} \leq \frac{8c_1}{|B_n|^{3/2} \sigma_n^3} \leq \frac{8a}{|B_n|^{3/2}}$$

with $a \geq c_1 (\lambda c_4)^{-3/2}$. Corollary 5.1 and the bound in Remark 5.1 finally yield

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{600 a \lambda (1 + \lambda^2 + \lambda^3)(10|B'_n|^{1/d} + 11)^{2d}}{|B_n|^{1/2}}$$

as desired.

This completes the proof of Theorem 5.3. $\square$

Acknowledgement

References

[1] AHMAD, I. A. AND LIN, P. E. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Inf. Theory* **IT-22,** 372–375.

[2] BAI, Z. D., RADHAKRISHNA RAO, C. AND ZHAO, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27,** 24–39.

[3] BEIRLANT, J., DUDEWICZ, E. J., GYÖRFI, L. AND VAN DER MEULEN, E. C. (1997). Nonparametric entropy estimation: an overview. *Int. J. Math. Statist. Sci.* **6,** 17–39.

[4] BESSE, A. L. (1978). *Manifolds All of Whose Geodesics are Closed* (Results Math. Rel. Areas **93**). Springer, Berlin.

[5] BOOTHBY, W. M. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Pure App. Math. **120**), 2nd edn. Academic Press, Orlando, FL.

[6] BULINSKI, A. AND SHASHKIN, A. (2007). *Limit Theorems for Associated Random Fields and Related Systems* (Adv. Ser. Statist. Sci. App. Prob. **10**). World Scientific Publishing, Hackensack, NJ.

[7] BURTON, R. AND WAYMIRE, E. (1985). Scaling limits for associated random measures. *Ann. Prob.* **13,** 1267–1278.

[8] CHEN, L. H. Y. AND SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Prob.* **32,** 1985–2028.

[9] EL MACHKOURI, M. (2014). Kernel density estimation for stationary random fields. *ALEA Lat. Amer. J. Prob. Math. Statist.* **11,** 259–279.

[10] HALL, P., WATSON, G. S. AND CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74,** 751–762.

[11] HEINRICH, L. (1988). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary Poisson cluster processes. *Math. Nachr.* **136,** 131–148.

[12] HEINRICH, L., KLEIN, S. AND MOSER, M. (2014). Empirical Mark covariance and product density function of stationary marked point processes—a survey on asymptotic results. *Methodology. Comput. Appl. Prob.* **16,** 283–293.

[13] HEINRICH, L., LÜCK, S. AND SCHMIDT, V. (2014). Asymptotic goodness-of-fit tests for the Palm mark distribution of stationary point processes with correlated marks. *Bernoulli* **20,** 1673–1697.

[14] HENRY, G. AND RODRIGUEZ, D. (2009). Kernel density estimation on Riemannian manifolds: asymptotic results. *J. Math. Imaging Vision* **34,** 235–239.

[15] KIDERLEN, M. (2001). Non-parametric estimation of the directional distribution of stationary line and fibre processes. *Adv. Appl. Prob.* **33,** 6–24.

[16] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33,** 1065–1076.

[17] PAWLAS, Z. (2009). Empirical distributions in marked point processes. *Stoch. Process. App.* **119,** 4194–4209.

[18] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Prob. Lett.* **73,** 297–304.

[19] PENROSE, M. D. AND YUKICH, J. E. (2013). Limit theory for point processes in manifolds. *Ann. Appl. Prob.* **23,** 2161–2211.

[20] ROSÉN, B. (1969). A note on asymptotic normality of sums of higher-dimensionally indexed random variables. *Ark. Mat.* **8,** 33–43.

[21] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. USA* **42,** 43–47.

[22] SAKAI, T. (1996). *Riemannian Geometry* (Transl. Math. Monogr. **149**). American Mathematical Society, Providence, RI.

[23] SCHUSTER, T. (2007). *The Method of Approximate Inverse: Theory and Applications* (Lecture Notes Math. **1906**). Springer, Berlin.

[24] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27,** 379–423, 623–656.

[25] SPODAREV, E. (ed.) (2013). *Stochastic Geometry, Spatial Statistics and Random Fields* (Lecture Notes Math. **2068**). Springer, Heidelberg.

[26] STOYAN, D., KENDALL, W. S. AND MECKE, J. (1987). *Stochastic Geometry and Its Applications*. John Wiley, Chichester.

[27] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.