Estimation of entropy for Poisson marked point processes

Patricia Alonso-Ruiz and Evgeny Spodarev

November 13, 2015

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. L^2 and almost surely consistency of this estimator as well as its asymptotic normality are investigated.

Keywords: marked point process, kernel density estimator, central limit theorem, fiber process, Boolean model.

1 Introduction

We consider a homogeneous Poisson marked point process (MPP) with marks from a compact Riemannian manifold without boundary that are assumed to be independent of the process. We are interested in detecting inhomogeneities in the distribution of the marks by studying its differential (Shannon, Kolmogorov) entropy.

The concept of entropy was introduced by Shannon in [25] for the needs of information theory and its origin lies in the classic 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*. This notion can be naturally generalized (see e.g. [8, Section 14.8]) as follows: Let P be a probability distribution of a random element X on an abstract measurable phase space M with probability density f with respect to a certain reference measure μ on M. The entropy of X is given by

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

where the expectation $\mathbb{E}_P$ is taken with respect to the probability measure P. An overview of nonparametric approaches to estimate the differential entropy for $M = \mathbb{R}^d$ can be found in [3]. Entropy estimation in discrete random settings has been treated in [16].

In the present work, we consider a nonparametric plug-in estimate of the differential entropy based on [1], which requires estimating the density of the distribution of interest also in a nonparametric way. To this purpose we use *kernel density estimation*, a technique introduced for stationary sequences of real random variables by Rosenblatt [21] and Parzen [17], extended to stationary real random fields in [9]. In the case of finite samples of i.i.d random vectors on the sphere, nonparametric kernel estimation methods have been studied in [10, 2], and extended to Riemannian manifolds in [19, 14]. Alternative nonparametric estimators for the directional distribution in line and fiber processes have been presented in [15].

The main result of our paper, Theorem 5.7, 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 the more general Theorem 5.1 of this type for sequences of m-dependent random fields proved in Section 5.

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 surely consistency. In Section 4 we introduce the nonparametric estimator $\widehat{\mathcal{E}}_f(B_n)$ of the entropy $\mathcal{E}_f$ of the density f in an observation window $B_n \subset \mathbb{R}^d$ and prove L^2-consistency of this estimator when the window size grows appropriately. Finally, we present in Section 5 a CLT for random sums of m_n-dependent random fields (cf. Corollary 5.2) 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 marked point processes. For an introduction and summary on these and other models of stochastic geometry we refer the reader to e.g. [27, 26, 23].

2.1 Poisson marked point processes

In the following, we consider $\Pi := \{Y_i\}_{i\geq 1}$ to be a *homogeneous Poisson point process* on $\mathbb{R}^d$ of intensity $\lambda > 0$. Moreover, we denote by (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 v_g. A detailed construction of this measure can be found e.g. in [22, p. 61]. Note that since M is compact, $v_g(M)$ is finite.

To each point $Y_i \in \Pi$ we attach a mark $\xi_i \in M$ and assume that the marks are i.i.d. random variables independent of the location of the points in Π. The collection $\Psi :=$

$\{(Y_i, \xi_i),\ Y_i \in \Pi\}$ is the *Poisson marked point process* we will work with. Ψ can be seen as 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 $T_y\Psi \stackrel{d}{=} \Psi$ 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 will assume that the probability law of a typical mark $\xi_{\mathbf{0}}$ has a density $f\colon M \to \mathbb{R}$ with respect to the Riemannian volume measure υ_g.

Example 2.1. Let F be a random 1-dimensional line segment of finite length $\ell > 0$ in $\mathbb{R}^d$ centered at the origin. We consider the *Boolean model*

$$\Phi := \bigcup_{Y \in \Pi} (Y \oplus F_Y),$$

where F_Y is an independent copy of F for each $Y \in \Pi$ and $\oplus$ denotes Minkowski addition. The random set Φ represents a system of independent fibers F_Y whereas $Y \in \Pi$ is assumed to be the middle point of F_Y. To each point $Y \in \Pi$ we attach a mark $\xi_Y \in S^{d-1}$ that represents the (random) unit direction vector of the fiber F_Y. Here, $M = S^{d-1}$, $p = d-1$, and υ_g is the surface area measure on S^{d-1}.

2.2 Space of marks

Since our mark space is a manifold, we need to recall some useful concepts from Riemannian geometry. For further details we refer to [5, 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 $r > 0$, $B_M(\eta, r) := \{\nu \in M \mid d_g(\nu, \eta) < r\}$ defines a neighborhood of η, and we call it 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 $\operatorname{inj}_g M := \inf_{\eta \in M} \sup\{r \geq 0 \mid B_M(\eta, r)$ is a normal nbhd. 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 $\nu \in U$, the *volume density function* introduced by Besse in [4, p.154] is given by

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

where $g_\nu(\frac{\partial}{\partial \psi_i}(\nu), \frac{\partial}{\partial \psi_j}(\nu))$ denotes the metric g in normal coordinates at the point $\exp_\eta^{-1} \nu$ (see e.g. [22, p.24]). Note that this function is only defined for points $\nu \in U$ such that $d_g(\eta, \nu) < \operatorname{inj}_g M$. 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 a growing 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 (v.H.-growing sequence)*. This means that

$$\lim_{n\to\infty} |B'_n| = \infty \qquad \text{and} \qquad \lim_{n\to\infty} \frac{|\partial B'_n \oplus B_r(o)|}{|B'_n|} = 0$$

where $B_r(o)$ 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 marked point process of intensity $\lambda > 0$. We define the kernel density estimator

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

This is an extension of the estimator given by Pelletier in [19]. The sequence of bandwidths $\{b_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ satisfies

(b1) $b_n < r_0 \; \forall\, n \in \mathbb{N}$, with $0 < r_0 < \operatorname{inj}_g M$ and $\inf_{\eta\in B_M(z,r_o)} \theta_z(\eta) > 0$ 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) $\int_{\mathbb{R}^p} K(\|x\|)dx = 1$,

(K2) $0 < \int_{\mathbb{R}^p} K(\|x\|) \, \|x\|^2 \, dx =: K_2 < \infty$,

(K3) $\operatorname{supp} K = [0, 1]$,

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

(K5) $\int_{\mathbb{R}^p} K(\|x\|) x \, dx = o$.

We further assume that

(f1) $f \in L^2(M)$, i.e. $\|f\|_2^2 := \int_M |f(\eta)|^2 \, d\upsilon_g(\eta) < \infty$,

(f2) f has bounded Hessian on any normal neighborhood $U \subset M$, i.e. $\exists\, C_2 > 0$ such that $\|D^2 f\| \leq C_2$.

Assumptions on the kernel are standard when dealing with nonparametric density estimation [19, 28]. For the 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 surely consistency of $\hat{f}_n$. In what follows, ω_p will denote the surface area 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 [28]. 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 that*

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

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

Corollary 3.2. *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.* $\mathbb{E}[\|\hat{f}_n - f\|_2^2] \xrightarrow{n \to \infty} 0$.

Corollary 3.3. *Under the assumptions of Theorem 3.3 it holds that*

$$\mathbb{E}[|\hat{f}_n(\xi_\mathbf{0}) - f(\xi_\mathbf{0})|^2] \xrightarrow{n \to \infty} 0.$$

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

Lemma 3.4. *For each* $\eta \in M$ *and* $n \in \mathbb{N}$,

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

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) thus becomes

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

$\square$

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

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

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

Lemma 3.5. *For any $\eta \in \operatorname{supp} f$ and $n \in \mathbb{N}$*

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

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

$$\mathbb{E}[\hat{f}_n(\eta)] = \int_M F_n(\eta, z) f(z)\, dv_g(z) = \mathbb{E}[F_n(\eta, \xi_\mathbf{o})]. \tag{3.3}$$

Due to Lemma 3.4 and (K3) we have that

$$|\mathbb{E}\left[F_n(\eta, \xi_\mathbf{o})\right] - f(\eta)| = \left| \int_{B_M(\eta,b_n)} \frac{1}{b_n^p \theta_\eta(z)} K\left(\frac{d_g(\eta, z)}{b_n}\right) \big(f(z) - f(\eta)\big)\, dv_g(z) \right|.$$

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. $z = \exp_\eta(x)$. Further, 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 $R_2(0, x) = O(x^T D^2 \tilde{f}(0) x)$ is the second order reminder. From assumption (f2) we have that $|R_2(0, x)| \le C_2 \|x\|^2$ for all $x \in B(0, b_n)$, hence passing to the exponential chart as in the proof of Lemma 3.4 yields

$$\begin{aligned}
&\left| \int_{B_M(\eta,b_n)} \frac{1}{b_n^p \theta_\eta(z)} K\left(\frac{d_g(\eta, z)}{b_n}\right) \big(f(z) - f(\eta)\big) dv_g(z) \right| \\
&= \left| \int_{B(0,b_n)} \frac{1}{b_n^p} \frac{1}{\theta_\eta(\exp_\eta(x))} K\left(\frac{\|x\|}{b_n}\right) \big(\tilde{f}(x) - \tilde{f}(0)\big) \|g(x)\|^{1/2}\, dx \right| && (3.4) \\
&= \left| \int_{B(0,b_n)} \frac{1}{b_n^p} \frac{1}{\theta_\eta(\exp_\eta(x))} K\left(\frac{\|x\|}{b_n}\right) R_2(0, x) \|g(x)\|^{1/2}\, dx \right| && (3.5) \\
&\le C_2 \int_{B(0,b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) \|x\|^2\, dx = C_2 b_n^2 K_2.
\end{aligned}$$

Equality (3.5) follows from (K5) because

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

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

$$= 0.$$

□

Lemma 3.6. *For any $n \in \mathbb{N}$,*

$$\int_M \mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})] \, d\upsilon_g(\eta) \leq \frac{C_\theta \, \omega_p K_0^2}{b_n^p},$$

with C_θ as in Theorem 3.1.

Proof. Applying Fubini's theorem we write

$$\int_M \mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})] \, d\upsilon_g(\eta) = \int_M I(z) f(z) \, d\upsilon_g(z), \tag{3.6}$$

where

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

Let us 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),

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

Plugging this estimate into (3.6) finishes the proof. □

We proceed to prove Theorem 3.1.

Proof of Theorem 3.1. By Fubini's theorem,

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

Note that $J(\eta) = \mathrm{Var}(\hat{f}_{B_n'}(\eta)) + \mathrm{Bias}(\hat{f}_{B_n'}, \eta)^2$. In view of (3.3) and the Campbell theorem we get

$$\mathrm{Var}(\hat{f}_n(\eta)) = \mathbb{E}[\hat{f}_n^2(\eta)] - (\mathbb{E}[\hat{f}_n(\eta)])^2$$

$$= \frac{1}{\lambda^2 |B_n'|^2} \mathbb{E}\Big[\sum_{i \geq 1} \mathbb{1}_{\{Y_i \in B_n'\}} F_n^2(\eta, \xi_i)\Big] + \frac{1}{\lambda^2 |B_n'|^2} \mathbb{E}\Big[\sum_{i,j \geq 1}^{\neq} \mathbb{1}_{\{Y_i, Y_j \in B_n'\}} F_n(\eta, \xi_i) F_n(\eta, \xi_j)\Big]$$

$$- \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})]^2 = \frac{1}{\lambda |B_n'|} \mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})] + \frac{\alpha^{(2)}(B_n' \times B_n')}{\lambda^2 |B_n'|^2} \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})]^2 - \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})]^2$$

$$= \frac{1}{\lambda |B_n'|} \mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})]. \tag{3.7}$$

Here, $\alpha^{(2)}(\cdot)$ denotes the 2nd-order factorial moment measure of the Poisson point process $\Pi := \{Y_i\}_{i \geq 1}$. We refer to [27, Chapter 1] for further definitions and formulas related to this measure in the Poisson case. Corollary 3.5 and Lemma 3.6 yield the existence of constants $C_\theta, C_2 > 0$ such that

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

□

Analogous arguments show the L^2-convergence of $\hat{f}_n(\xi_{\mathbf{0}})$ to $f(\xi_{\mathbf{0}})$.

Proof of Corollary 3.3. Passing to normal coordinates as in (3.4) and (3.5) and setting $\tilde{f} := f \circ \exp_\eta$, we obtain

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

From the proof of Lemma3.6 we thus obtain

$$\mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})] \leq \frac{K_0 C_\theta(\eta)}{b_n^p} \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})] \leq \frac{2 K_0 C_\theta(\eta)}{b_n^p} f(\eta) \tag{3.9}$$

for any $\eta \in \operatorname{supp} f$. In view of (3.7) and Lemma 3.5 this yields

$$\mathbb{E}[|\hat{f}_n(\xi_{\mathbf{0}}) - f(\xi_{\mathbf{0}})|^2] \leq \frac{2 C_\theta K_0\, \|f\|_2^2}{\lambda b_n^p |B_n'|} + b_n^4 C_2^2 K_2^2,$$

which tends to zero as $n \to \infty$. □

Remark 3.7. The problem of finding an optimal sequence of bandwidths $\{b_n\}_{n \in \mathbb{N}}$ can be understood as a special case of regularization [24] and we can use the bound of the estimation error given in Theorem 3.1 in order to find it. We start by considering the bandwidth $b > 0$ being independent of n. For any fixed $n \in \mathbb{N}$ the optimal bandwidth will be

$$\operatorname{argmin}_b \mathbb{E}[\|\hat{f}_n - f\|_2^2].$$

By Theorem 3.1, we can approximate the order of magnitude of this optimal b by minimizing the upper bound of the mean square error $e(b) := \frac{C_\theta \omega_p K_0^2}{\lambda |B_n'| b^p} + b^4 C_2^2 K_2^2 \upsilon_g(M)$. A simple calculation leads to the unique minimum point $b_{opt} = \left(\frac{p C_\theta \omega_p K_0^2}{4 C_2^2 K_2^2 \upsilon_g(M) \lambda |B_n'|} \right)^{\frac{1}{p+4}}$. Note that this value depends on n, so that we have $b_{opt} \downarrow 0$ and $b_{opt}^p\, |B_n'| \to \infty$ 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 consistency of $\hat{f}_n$.

Theorem 3.8. *Under the assumptions of Theorem 3.1 and if $|B'_n| > n^{\frac{(4+p)(1+\delta)}{4p}}$ for some $\delta > 0$, then*

$$|\hat{f}_n(\eta) - f(\eta)| \xrightarrow{n\to\infty} 0 \qquad a.s.$$

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

Proof. Notice that the assumption on $|B'_n|$ can be rewritten as $\left(\frac{n^{1+\delta}}{|B'_n|}\right)^{1/p} < n^{-\frac{1+\delta}{4}}$ and we can therefore choose b_n satisfying assumptions (b1) - (b3) and $\left(\frac{n^{1+\delta}}{|B'_n|}\right)^{1/p} < b_n < n^{-\frac{1+\delta}{4}}$. For each $\varepsilon > 0$, Chebyshev's inequality and the bounds used in the proof of Corollary 3.3 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

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

for some $c_1 < \infty$. The almost sure convergence follows from Borel-Cantelli's lemma. □

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

$$\widehat{\mathcal{E}}_f(B_n) := -\frac{1}{\lambda|B_n|} \sum_{i\geq 1} \mathbb{1}_{\{Y_i \in B_n\}} \log \hat{f}_{B'_n + Y_i}(\xi_i), \tag{4.1}$$

where $B'_n + y$ denotes the translation of B'_n by $y \in \mathbb{R}^d$. Moreover, we assume that $B'_n \subset B_n$, whereas their size relation will be determined later. From now on, we substitute the previous assumption (f1) by (f1) f is continuous.

Note that since M is compact, the new (f1) in particular implies the former. With the additional assumptions for a typical mark $\xi_{\mathbf{0}}$,

$$\text{(L1) } \mathbb{E}\left[\log^2 f(\xi_{\mathbf{0}})\right] =: L_1 < \infty \qquad \text{and} \qquad \text{(L2) } \mathbb{E}\left[\left(\frac{\|\nabla f(\xi_{\mathbf{0}})\|}{f(\xi_{\mathbf{0}})}\right)^2\right] =: L_2 < \infty,$$

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

Theorem 4.1. *Let $B'_n := (0, m_n)^d$, $B_n := (-p_n, p_n)^d$, $n \in \mathbb{N}$, for some $p_n > m_n > 0$, and let B'_n satisfy $(b1)-(b3)$. Further, assume that conditions $(K1)-(K5)$, $(f1)$, $(f2)$, $(L1)$ and $(L2)$ hold. Then,*

$$\mathbb{E}[|\widehat{\mathcal{E}}_f(B_n) - \mathcal{E}_f|^2] \le 3\left(\frac{8K_0 C_\theta \upsilon_g(M)}{\lambda^2 |B_n||B'_n| b_n^p} + \frac{4}{\lambda^2 |B'_n|} + 16 b_n^2 L_2 + \frac{L_1}{\lambda |B_n|}\right)$$

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

Corollary 4.2. *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. $\mathbb{E}[|\widehat{\mathcal{E}}_f(B_n) - \mathcal{E}_f|^2] \xrightarrow{n\to\infty} 0$.*

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

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

Proof. Recall from (3.3) that $\mathbb{E}[\hat{f}_{B'_n}(\eta)] = \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})]$. Using normal coordinates analogously to (3.4) and (3.5) with $\tilde{f} := f \circ \exp_\eta$ we obtain

$$\begin{aligned} |\mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})] - f(\eta)| &= \Big| \int_{B(0,b_n)} \frac{1}{b_n^p} K\left(\frac{\|x\|}{b_n}\right) x \cdot \int_0^1 \nabla \tilde{f}(tx)\, dt\, dx \Big| \\ &\le b_n \int_{B(0,1)} K\left(\|y\|\right) \|y\| \int_0^1 \left\|\nabla \tilde{f}(t b_n y)\right\| dt\, dy. \end{aligned}$$

Since $b_n \downarrow 0$, we have $\left\|\nabla \tilde{f}(t b_n y)\right\| = \left\|\nabla \tilde{f}(0)\right\|(1 + o(1))$ for sufficiently large $n \in \mathbb{N}$ and in view of (K1) last expression can be bounded by

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

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

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

□

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

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

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

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

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

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

On the one hand, notice that by definition,

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

depends on (Y_i, ξ_i) and $T_{Y_i}\Psi - \delta_{(o,\xi_i)}$. Thus, since Ψ is an independently marked Poisson MPP, the Campbell-Mecke type formula in [26, p.129] yields

$$\frac{1}{\lambda^2 |B_n|^2} \mathbb{E}\big[\sum_{i \geq 1} h(Y_i, \xi_i, T_{Y_i}\Psi - \delta_{(o,\xi_i)})\big] = \frac{1}{\lambda |B_n|^2} \int_{\mathbb{R}^d} \int_M \mathbb{E}_{P^{o!}_{\eta}}[h(y, \eta, \Psi)] f(\eta) dv_g(\eta) dy,$$

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

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

Now, recall that for any differentiable function g and any $x, z \in \mathbb{R}$ we have

$$|g(x) - g(z)| = |x - z||g'((1-\gamma)x + \gamma z)|, \qquad \gamma \in (0,1),$$

which in the particular case of $g(x) = \log x$ yields

$$|\log x - \log z| = \frac{|x-z|}{|(1-\gamma)x + \gamma z|} \leq \frac{|x-z|}{\min\{x, z\}}, \qquad x, z > 0. \tag{4.2}$$

Note that Ψ is stationary and by assumption (f1) f is continuous, thus by Theorem 3.8 $\hat{f}_{B'_n+y}(\eta)$ converges to $f(\eta)$ a.s. for any $y \in \mathbb{R}^d$ and $\eta \in M$. Furthermore, in view of (3.8), $\mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})] = (1 + o(1))f(\eta)$, hence for $n \in \mathbb{N}$ large enough

$$\min\{\hat{f}_{B'_n+y}(\eta), \mathbb{E}[\hat{f}_{B'_n+y}(\eta)]\} \geq \frac{1}{2} f(\eta). \tag{4.3}$$

Applying inequality (4.2) with $x = \hat{f}_{B'_n+y}(\eta)$ and $z = \mathbb{E}[\hat{f}_{B'_n+y}(\eta)] = \mathbb{E}[F_n(\eta, \xi_{\mathbf{0}})]$ we obtain

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

Due to (3.7) and (3.9),

$$J_1 \leq \frac{4}{\lambda |B_n|} \int_M \frac{\mathbb{E}[F_n^2(\eta, \xi_{\mathbf{0}})]}{f(\eta)^2 \lambda |B'_n|} f(\eta) d\upsilon_g(\eta) dy \leq \frac{8 K_0 C_\theta \upsilon_g(M)}{\lambda^2 |B_n||B'_n| b_n^p}. \tag{4.4}$$

Analogously, each summand in J_2 can be expressed as a function h depending of (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 [26, p.129] in the independently marked Poisson case yields

$$\begin{aligned} J_2 &= \mathbb{E}\Big[\sum_{i,j\geq 1}^{\neq} h(Y_i, \xi_i, Y_j, \xi_j, T_{Y_i}\Psi - \delta_{(o,\xi_i)} - \delta_{(Y_j,\xi_j)})\Big] \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{M^2} \mathbb{E}_{P^{o,y_2!}_{\eta_1,\eta_2}}[h(y_1, \eta_1, y_2, \eta_2, \Psi)] f(\eta_1) f(\eta_2)\, d\upsilon_g(\eta_2) d\upsilon_g(\eta_1) dy_1\, dy_2 \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{M^2} \mathbb{E}[h(y_1, \eta_1, y_2, \eta_2, \Psi)] f(\eta_1) f(\eta_2)\, d\upsilon_g(\eta_2) d\upsilon_g(\eta_1) dy_1\, dy_2, \end{aligned}$$

where last inequality follows from the independent marking of the Poisson MPP. Applying again Theorem 3.8, (4.2) and (3.8), we obtain for $n \in \mathbb{N}$ large enough

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

In view of (3.3) and the Campbell theorem,

$$\begin{aligned}
&\operatorname{Cov}\big(\hat f_{B'_n+y_1}(\eta_1), \hat f_{B'_n+y_2}(\eta_2)\big) = \mathbb{E}[\hat f_{B'_n+y_1}(\eta_1)\hat f_{B'_n+y_2}(\eta_2)] - \mathbb{E}[\hat f_{B'_n+y_1}(\eta_1)]\mathbb{E}[\hat f_{B'_n+y_2}(\eta_2)] \\
&= \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\big[\sum_{i\geq 1} \mathbb{1}_{\{Y_i\in(B'_n+y_1)\cap(B'_n+y_2)\}}F_n(\eta_1,\xi_i)F_n(\eta_2,\xi_i)\big] \\
&+ \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\big[\sum_{i,j\geq 1}^{\neq} \mathbb{1}_{\{Y_i\in B'_n+y_1\}}\mathbb{1}_{\{Y_j\in B'_n+y_2\}}F_n(\eta_1,\xi_i)F_n(\eta_2,\xi_j)\big] \\
&- \mathbb{E}[F_n(\eta_1,\xi_{\mathbf{0}})]\mathbb{E}[F_n(\eta_2,\xi_{\mathbf{0}})] \\
&= \frac{|(B'_n+y_1)\cap(B'_n+y_2)|}{\lambda|B'_n|^2}\mathbb{E}[F_n(\eta_1,\xi_{\mathbf{0}})F_n(\eta_2,\xi_{\mathbf{0}})] \leq \frac{1}{\lambda|B'_n|}\mathbb{E}[F_n(\eta_1,\xi_{\mathbf{0}})F_n(\eta_2,\xi_{\mathbf{0}})].
\end{aligned}$$

Fubini's theorem and Lemma 3.4 yield

$$J_2 \leq \frac{4}{\lambda^2|B'_n|}\int_{M^2}\mathbb{E}[F_n(\eta_1,\xi_{\mathbf{0}})F_n(\eta_2,\xi_{\mathbf{0}})]d\upsilon_g(\eta_1)d\upsilon_g(\eta_2) = \frac{4}{\lambda^2|B'_n|},$$

which together with (4.4) leads to

$$I_{1,n} \leq \frac{8K_0C_\theta\upsilon_g(M)}{\lambda^2|B_n||B'_n|b_n^p} + \frac{4}{\lambda^2|B'_n|}.$$

Secondly, due to the stationarity of Ψ and the Campbell theorem,

$$\begin{aligned}
I_{2,n} &= \frac{1}{\lambda^2|B_n|^2}\mathbb{E}\big[\sum_{i\geq 1}\mathbb{1}_{\{Y_i\in B_n\}}(\log\mathbb{E}[\hat f_{B'_n+Y_i}(\xi_i)] - \log f(\xi_i))^2\big] \\
&+ \frac{1}{\lambda^2|B_n|^2}\mathbb{E}\big[\sum_{i,j\geq 1}^{\neq}\mathbb{1}_{\{Y_i,Y_j\in B_n\}}(\log\mathbb{E}[\hat f_{B'_n+Y_i}(\xi_i)] - \log f(\xi_i))(\log\mathbb{E}[\hat f_{B'_n+Y_j}(\xi_j)] - \log f(\xi_j))\big] \\
&= \frac{1}{\lambda|B_n|}\mathbb{E}[(\log\mathbb{E}[\hat f_{B'_n}(\xi_{\mathbf{0}})] - \log f(\xi_{\mathbf{0}}))^2] + (\mathbb{E}[\log\mathbb{E}[\hat f_{B'_n}(\xi_{\mathbf{0}})] - \log f(\xi_{\mathbf{0}})])^2 \\
&\leq 2\mathbb{E}[(\log\mathbb{E}[\hat f_{B'_n}(\xi_{\mathbf{0}})] - \log f(\xi_{\mathbf{0}}))^2]
\end{aligned}$$

for large $n\in\mathbb{N}$. On the other hand, by (4.3) and Lemma 4.3 we get

$$\begin{aligned}
&\mathbb{E}[(\log\mathbb{E}[\hat f_{B'_n}(\xi_{\mathbf{0}})] - \log f(\xi_{\mathbf{0}}))^2] = \int_{\operatorname{supp} f}(\log\mathbb{E}[\hat f_{B'_n}(\eta)] - \log f(\eta))^2 f(\eta)d\upsilon(\eta) \\
&\leq 4\int_{\operatorname{supp} f}\frac{(\mathbb{E}[\hat f_{B'_n}(\eta)] - f(\eta))^2}{f(\eta)}d\upsilon(\eta) \leq 16b_n^2L_2,
\end{aligned}$$

so that $I_2 \leq 16b_n^2L_2$.

Finally, note that $\mathcal{E}_f = -\mathbb{E}[\log f(\xi_{\mathbf{0}})]$. Applying once more the Campbell theorem we obtain

$$
\begin{aligned}
I_{3,n} &= \frac{1}{\lambda^2|B_n|^2}\mathbb{E}\big[\sum_{i\geq 1}\mathbb{1}_{\{Y_i\in B_n\}}\log^2 f(\xi_i)\big] + \frac{1}{\lambda^2|B_n|^2}\mathbb{E}\big[\sum_{i,j\geq 1}^{\neq}\mathbb{1}_{\{Y_i,Y_j\in B_n\}}\log f(\xi_i)\log f(\xi_j)\big] \\
&\quad + \frac{2}{\lambda|B_n|}\mathbb{E}\big[\sum_{i\geq 1}\mathbb{1}_{\{Y_i\in B_n\}}\log f(\xi_i)\big]\mathcal{E}_f + \mathcal{E}_f^2 \\
&= \frac{1}{\lambda|B_n|}\mathbb{E}[\log^2 f(\xi_{\mathbf{0}})] + \big(\mathbb{E}[\log f(\xi_{\mathbf{0}})]\big)^2 + 2\mathbb{E}[\log f(\xi_{\mathbf{0}})]\mathcal{E}_f + \mathcal{E}_f^2 \\
&= \frac{1}{\lambda|B_n|}\mathbb{E}[\log^2 f(\xi_{\mathbf{0}})] = \frac{L_1}{\lambda|B_n|}.
\end{aligned}
$$

Remark 4.4. 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 3.7 lead to $b_{opt} = \left(\frac{pK_0C_\theta \upsilon_g(M)}{4L_2\lambda^2|B_n||B'_n|}\right)^{\frac{1}{p+2}}$.

5 Central limit theorem for entropy

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 $j = (j^1, \dots, j^d)$, while $j_1, j_2, \dots$ will denote a sequence in $\mathbb{K}^d$. Moreover, we write $\mathbf{t} = (t, \dots, t) \in \mathbb{K}^d$ 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, $C_{\mathbf{t}} = [0,t)^d$.

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*, where m_n is the diameter of $B'_n \subset B_n$ from the definition of $\widehat{\mathcal{E}}_f(B_n)$ in (4.1). We will present in this section a CLT for a modified version of the original estimator.

A random field $\{X_j : j \in \mathbb{K}^d\}$ is said to be $m-$*dependent* for some $m \in \mathbb{N}$ 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 point processes. 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. We refer to [26] for an overview on the subject.

5.1 Theoretical results

Our CLT is based on a slightly weaker version of the result obtained by Wang and Woodroofe [29] for deterministic sums of stationary m-dependent random fields.

Theorem 5.1. *Let $\{X_{n,j} : j \in \mathbb{N}^d\}_{n\in\mathbb{N}}$ be a sequence of stationary m_n-dependent zero mean random fields observed on a sequence of windows $\{V_{\mathbf{p_n}}\}_{n\in\mathbb{N}} \subset \mathbb{N}^d$, with $m_n, p_n \to \infty$ as $n \to \infty$. Suppose that there exists $C > 0$ such that*

$$\mathbb{E}\big[\big(\sum_{i\in V_j} X_{n,i}\big)^2\big] \le C\, j^1 \cdots j^d \qquad \forall\, n \in \mathbb{N},\ j \in \mathbb{N}^d, \tag{D1}$$

and there exists a sequence $\{q_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ such that

$$\frac{m_n}{q_n} \xrightarrow{n\to\infty} 0 \qquad \textit{and} \qquad \frac{q_n}{p_n} \xrightarrow{n\to\infty} 0, \tag{D2}$$

$$\liminf_{n\to\infty} \frac{\sigma_n^2}{q_n^d} =: L > 0, \qquad \textit{where} \qquad \sigma_n^2 := \mathbb{E}\Big[\Big(\sum_{j\in V_{\mathbf{q_n}}} X_{n,j}\Big)^2\Big], \tag{D3}$$

$$\lim_{n\to\infty} \frac{1}{q_n^d}\mathbb{E}\big[\big(\sum_{j\in V_{\mathbf{q_n}}} X_{n,j}\big)^2 \mathbb{1}\big(\big|\sum_{j\in V_{\mathbf{q_n}}} X_{n,j}\big| > p_n^{d/2}\varepsilon\big)\big] = 0 \qquad \forall\, \varepsilon > 0. \tag{D4}$$

Then,

$$\frac{\sum_{j\in V_{\mathbf{p_n}}} X_{n,j}}{p_n^{d/2}\sigma_n} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Proof. This follows by dividing by σ and replacing it by L in the original proof. □

We give an extension of this theorem to random sums of stationary $m-$dependent random fields indexed in $\mathbb{R}_+^d$. For simplicity, we will 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 5.2. *Let $\{X_{n,y} : y \in \mathbb{R}_+^d\}_{n\in\mathbb{N}}$ be a sequence of stationary m_n-dependent zero mean random fields observed on a sequence $\{C_{\mathbf{p_n}}\}_{n\in\mathbb{N}} \subset \mathbb{R}_+^d$ with $p_n, m_n \to \infty$ as $n \to \infty$, and let Π be a stationary Poisson point process on $\mathbb{R}_+^d$. Suppose that there exists $C > 0$ such that*

$$\mathbb{E}\big[\big(\sum_{y\in\Pi\cap C_j} X_{n,y}\big)^2\big] \le C|C_j| \qquad \forall\, n \in \mathbb{N},\ j \in \mathbb{N}^d, \tag{A1}$$

and there exists a sequence $\{q_n\}_{n\in\mathbb{N}} \subset \mathbb{N}$ such that

$$\frac{m_n}{q_n} \xrightarrow{n\to\infty} 0 \qquad \textit{and} \qquad \frac{q_n}{p_n} \xrightarrow{n\to\infty} 0, \tag{A2}$$

$$\liminf_{n\to\infty} \frac{\sigma_n^2}{q_n^d} =: L > 0, \qquad \textit{where} \qquad \sigma_n^2 := \mathbb{E}\big[\big(\sum_{y\in\Pi\cap C_{\mathbf{q_n}}} X_{n,y}\big)^2\big], \tag{A3}$$

$$\lim_{n\to\infty} \frac{1}{q_n^d}\mathbb{E}\big[\big(\sum_{y\in\Pi\cap C_{\mathbf{q_n}}} X_{n,y}\big)^2 \mathbb{1}\big(\big|\sum_{y\in\Pi\cap C_{\mathbf{q_n}}} X_{n,y}\big| > \sqrt{|B_n|\varepsilon}\big)\big] = 0 \qquad \forall\, \varepsilon > 0. \tag{A4}$$

Then,

$$\frac{\sum_{y\in\Pi\cap B_n} X_{n,y}}{\sqrt{|B_n|\,\sigma_n^2}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Proof. Define $Z_{n,j} := \sum_{y\in\Pi\cap(j\oplus C_{\mathbf{1}})} X_{n,y}$ for each $j \in \mathbb{N}^d$ and $n \in \mathbb{N}$. Obviously, $\{Z_{n,j}\}_{j\in\mathbb{N}^d}$ is (m_n+1)-dependent. Moreover, $\sum_{y\in\Pi\cap B_n} X_{n,y} = \sum_{j\in V_{\mathbf{p_n-1}}} Z_{n,j}$, and Theorem 5.1 holds for the sequence $\{Z_{n,j} :\ j \in \mathbb{N}^d\}_{n\in\mathbb{N}}$. □

Remark 5.3. Note that Corollary 5.2 does not require independence between the random fields $\{X_{n,y}\}_{y\in\mathbb{R}^d_+}$ and the point process Π. If independence is provided, applying the Campbell theorem and the stationarity of Π we get

$$\begin{aligned}\mathbb{E}\big[\big(\sum_{y\in\Pi\cap C_j} X_{n,y}\big)^2\big] &= \mathbb{E}\big[\sum_{y\in\Pi\cap C_j} X_{n,y}^2\big] + \mathbb{E}\big[\sum_{y_1,y_2\in\Pi\cap C_j}^{\neq} X_{n,y_1}X_{n,y_2}\big]\\ &= \lambda\int_{C_j}\mathbb{E}[X_{n,\mathbf{0}}^2]dy + \lambda\int_{C_j\times C_j}\mathbb{E}[X_{n,\mathbf{0}}X_{n,y_2-y_1}]\,\alpha^{(2)}(dy_1,dy_2)\\ &= \lambda|C_j|\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2\int_{\mathbb{R}^d}|C_j\cap(C_j-y)|\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})dy.\end{aligned}$$

Thus, condition (A1) can be substituted by the existence of $c > 0$ such that

$$\sup_{n\in\mathbb{N}}\Big(\mathbb{E}[X_{n,\mathbf{0}}^2] + \int_{\mathbb{R}^d_+}|\operatorname{Cov}(X_{n,\mathbf{0}}X_{n,y})|dy\Big) < c. \tag{A1'}$$

Condition (A3) can now be formulated as

$$\liminf_{n\to\infty}\Big(\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda\int_{\mathbb{R}^d_+}\frac{|C_{\mathbf{q_n}}\cap(C_{\mathbf{q_n}}-y)|}{|C_{\mathbf{q_n}}|}\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})\,dy\Big) > 0. \tag{A3'}$$

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

Theorem 5.4. *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*

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

and

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

If the limit

$$\sigma^2 := \lim_{n\to\infty}\Big(\lambda\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2\int_{\mathbb{R}^d}\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})\,dy\Big)$$

exists and is positive, then

$$\frac{1}{|U_n|}\operatorname{Var}\Big(\sum_{y\in\Pi\cap U_n}X_{n,y}\Big)\xrightarrow{n\to\infty}\sigma^2 \tag{5.3}$$

for any sequence of subsets $\{U_n\}_{n\in\mathbb{N}}$ growing in the van Hove sense.

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

$$\operatorname{Var}\Big(\sum_{y\in\Pi\cap U_n} X_{n,y}\Big) = \lambda|U_n|\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2|U_n|\int_{\mathbb{R}^d}\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})\,dy \\ - \lambda^2\int_{U_n}\int_{U_n^c}\operatorname{Cov}(X_{n,y_1},X_{n,y_2})\,dy_1dy_2.$$

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 $(\partial U_n)_p := \partial U_n \oplus B_p(o)$ denotes the $p-$neighborhood of $\partial U_n \subset \mathbb{R}^d$. From the previous calculation we have

$$\lambda|U_n|\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2|U_n|\int_{\mathbb{R}^d}\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})\,dy - \operatorname{Var}\Big(\sum_{y\in\Pi\cap U_n} X_{n,y}\Big)$$
$$= \lambda^2\int_{G_n}\int_{U_n^c}\operatorname{Cov}(X_{n,y_1},X_{n,y_2})\,dy_1dy_2 + \lambda^2\int_{W_n}\int_{U_n^c}\operatorname{Cov}(X_{n,y_1},X_{n,y_2})\,dy_1dy_2$$
$$=: R_{n,1}+R_{n,2}.$$

On the one hand, $|G_n|\le|(\partial U_n)_p|$ and assumption (5.2) yield

$$\frac{|R_{n,1}|}{|U_n|} \le \frac{|(\partial U_n|)_p}{|U_n|}\lambda^2\int_{\mathbb{R}^d}|\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})|\,dy \xrightarrow{n\to\infty} 0$$

since $\{U_n\}_{n\in\mathbb{N}}$ is v.H.-growing. On the other hand, $\operatorname{dist}(W_n,U_n^c)\ge p$ and $|W_n|\le|U_n|$, hence

$$\frac{|R_{n,2}|}{|U_n|} \le \frac{|W_n|}{|U_n|}\lambda^2\int_{\mathbb{R}^d\setminus(-p,p)^d}|\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})|\,dy \le \lambda^2\int_{\mathbb{R}^d\setminus(-p,p)^d}|\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})|\,dy$$

and in view of assumption (5.1) the convergence in (5.3) is established. □

The same result 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 5.5. *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*

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

If the limit

$$\sigma^2 := \lim_{n\to\infty}\Big(\lambda\mathbb{E}[X_{n,\mathbf{0}}^2] + \lambda^2\int_{\mathbb{R}^d}\operatorname{Cov}(X_{n,\mathbf{0}},X_{n,y})\,dy\Big)$$

exists and is positive, then

$$\lim_{n\to\infty}\frac{1}{|U_n|}\operatorname{Var}\Big(\sum_{y\in\Pi\cap U_n} X_{n,y}\Big) \xrightarrow{n\to\infty} \sigma^2$$

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

Remark 5.6. For instance, the result holds taking cubic windows $U_n = (-u_n, u_n)^d$ with $\frac{m_n}{u_n} \xrightarrow{n\to\infty} 0$.

Proof. Set $p = m_n$ in the proof of Theorem 5.4. Due to m_n-dependence, condition (5.1) is trivially fulfilled and therefore $\limsup_{n\to\infty} \frac{|R_{n,2}|}{|U_n|} = 0$. On the other hand,

$$\frac{|R_{n,1}|}{|U_n|} \leq \frac{|(\partial U_n)_{m_n}|}{|U_n|} \int_{\mathbb{R}^d} |\operatorname{Cov}(X_{n,\mathbf{0}}, X_{n,y})|\, dy \xrightarrow{n\to\infty} 0$$

due to 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 and the sequence $\{X_{n,y} : y \in \mathbb{R}^d\}_{n\in\mathbb{N}}$ is crucial to perform calculations. Therefore, we need to consider the modified estimator

$$\widehat{\mathcal{E}}_f^*(B_n) := -\frac{1}{\lambda|B_n|} \sum_{y\in\Pi^*\cap B_n} \log \hat{f}_{B_n'+y}(\xi_y^*)\, \mathbb{1}_{\{\Pi^*(B_n)>0\}},$$

where $\Psi^* := \{(Y, \xi_Y^*), Y \in \Pi^*\}$ is an independent copy of the original marked Poisson point process Ψ. The study of the original estimator is subject of further research and it involves marked point processes where the marks depend of their location (we refer to [18, 13, 12] for some investigations in this direction). Moreover, we also need to assume

(f3) $\inf_{\eta\in\operatorname{supp} f} f(\eta) := c_0 > 0$.

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 we opted for the former for ease of proofs. The aim of this section is to apply Corollary 5.2 in order to obtain a CLT for $\widehat{\mathcal{E}}_f^*(B_n)$.

Theorem 5.7. *Let $\{B_n\}_{n\in\mathbb{N}}$, $\{B_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_+^d$ be sequences of observation windows such that $B_n = C_{\mathbf{p_n}}$, $B_n' = C_{\mathbf{m_n}}$, $p_n = m_n^{4+\delta}$ for some $\delta > 0$, and $m_n \to \infty$ as $n \to \infty$. Under the conditions of Theorem 4.1,*

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

where

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

and

$$\sigma_n^2 := \lambda\mathbb{E}[\log \hat{f}_{B_n'}(\xi_{\mathbf{0}})^2] + \lambda^2 \int_{B_n'} \frac{|C_{\mathbf{q_n}} \cap (C_{\mathbf{q_n}} - y)|}{|C_{\mathbf{q_n}}|} \operatorname{Cov}(\log \hat{f}_{B_n'}(\xi_{\mathbf{0}}), \log \hat{f}_{B_n'+y}(\xi_y'))\, dy,$$

where $\{\xi_y'\}_{y\in\mathbb{R}_+^d}$ are independent copies of $\xi_{\mathbf{0}}$ and $q_n = m_n^{1+\delta'}$ with $4\delta' < \delta$.

Notice that $\widehat{\mathcal{E}}_f^*(B_n)$ can be expressed as a sum of m_n-dependent random fields and therefore

$$\begin{aligned}\sqrt{|B_n|}\big(\widehat{\mathcal{E}}_f^*(B_n)-\hat{\mu}_{B_n}\big)&=\frac{1}{\lambda\sqrt{|B_n|}}\sum_{y\in\Pi^*\cap B_n}\Big(-\log\hat{f}_{B_n'+y}(\xi_y^*)-\mathbb{E}\big[-\log\hat{f}_{B_n'}(\xi_{\mathbf{0}})\big]\Big)\\&=:\frac{1}{\lambda\sqrt{|B_n|}}\sum_{y\in\Pi^*\cap B_n}X_{n,y},\end{aligned}\tag{5.5}$$

where, by construction, $\{X_{n,y}\}_{y\in\mathbb{R}_+^d}$ is a stationary centered m_n-dependent random field. The CLT follows directly from Corollary 5.2 if the required conditions are satisfied. The next paragraphs are thus devoted to the proof of assumptions (A1'),(A2),(A3') and (A4) of Corollary 5.2 for $\{-\log\hat{f}_{B_n'+y}(\xi_y')\}_{y\in\mathbb{R}_+^d}$.

For the ease of reading, we use the notation $\hat{f}_{B_n'+y}$ 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 5.7 hold in the subsequent lemmata without mentioning them explicitly.

5.2.1 Condition (A1')

We start by establishing some helpful bounds.

Lemma 5.8. *There exists $c_1>0$ such that for any $x_1,x_2\in B_n$ and $n\in\mathbb{N}$,*

$$\operatorname{Cov}\big(\log\hat{f}_{B_n'+x_1},\log\hat{f}_{B_n'+x_2}\big)\leq c_1\operatorname{Cov}\big(\hat{f}_{B_n'+x_1},\hat{f}_{B_n'+x_2}\big).$$

Proof. Adding and subtracting $\log\mathbb{E}[\hat{f}_{B_n'+y_1}]$ resp. $\log\mathbb{E}[\hat{f}_{B_n'+y_2}]$, Theorem 3.8, (4.3) and assumption (f3) lead to

$$\begin{aligned}&\operatorname{Cov}\big(\log\hat{f}_{B_n'+x_1},\log\hat{f}_{B_n'+x_2}\big)\\&=\mathbb{E}[(\log\hat{f}_{B_n'+x_1}-\log\mathbb{E}[\hat{f}_{B_n'}])(\log\hat{f}_{B_n'+x_2}-\log\mathbb{E}[\hat{f}_{B_n'}])]-(\mathbb{E}[\log\hat{f}_{B_n'}]-\log\mathbb{E}[\hat{f}_{B_n'}])^2\\&\leq\frac{4}{c_0^2}\operatorname{Cov}(\hat{f}_{B_n'+x_1},\hat{f}_{B_n'+x_2})\end{aligned}$$

for $n\in\mathbb{N}$ sufficiently large. Thus the Lemma follows for any $n\in\mathbb{N}$ with a constant $c_1>0$ (maybe different from $4/c_0^2$). ☐

Lemma 5.9. *There exists $c_2>0$ such that for any $n\in\mathbb{N}$ and $x_1,x_2\in B_n$,*

$$\operatorname{Cov}(\hat{f}_{B_n'+x_1},\hat{f}_{B_n'+x_2})\leq\frac{c_2|(B_n'+x_1)\cap(B_n'+x_2)|}{\lambda|B_n'|^2}.$$

Proof. Applying the Campbell theorem,

$$\begin{aligned}\operatorname{Cov}(\hat f_{B'_n+x_1},\hat f_{B'_n+x_2}) &= \mathbb{E}[\hat f_{B'_n+x_1}\hat f_{B'_n+x_2}] - (\mathbb{E}[\hat f_{B'_n}])^2\\ &= \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\Big[\sum_{y\in\Pi\cap(B'_n+x_1)\cap(B'_n+x_2)} F_n(\xi_y,\xi'_{x_1})F_n(\xi_y,\xi'_{x_2})\Big]\\ &+ \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}\Big[\sum^{\neq}_{\substack{y_1\in\Pi\cap(B'_n+x_1)\\ y_2\in\Pi\cap(B'_n+x_2)}} F_n(\xi_{y_1},\xi'_{x_1})F_n(\xi_{y_2},\xi'_{x_2})\Big] - (\mathbb{E}[F_n(\xi_{\mathbf{0}},\xi'_{x_1})])^2\\ &= \frac{|(B'_n+x_1)\cap(B'_n+x_2)|}{\lambda|B'_n|^2}\mathbb{E}[F_n(\xi_{\mathbf{0}},\xi'_{x_1})F_n(\xi_{\mathbf{0}},\xi'_{x_2})]\\ &+ \frac{|(B'_n+x_1)\cap(B'_n+x_2)|}{\lambda|B'_n|^2}(\mathbb{E}[F_n(\xi_{\mathbf{0}},\xi'_{x_1})])^2.\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}[F_n(\xi_{\mathbf{0}},\xi'_{x_1})F_n(\xi_{\mathbf{0}},\xi'_{x_2})] &= \int_{M^3} F_n(\mu,z)F_n(z,\eta)f(\mu)f(z)f(\eta)\,dv_g(\mu,z,\eta)\\ &= (1+o(1))\int_M f(z)^3 dv_g(z) = (1+o(1))\mathbb{E}[f^2(\xi_{\mathbf{0}})]\end{aligned}$$

as well as

$$\mathbb{E}[F_n(\xi_{\mathbf{0}},\xi'_{x_1})] = \int_{M^2} F_n(\mu,z)f(\mu)f(z)\,dv_g(\mu,z) = (1+o(1))\mathbb{E}[f(\xi_{\mathbf{0}})].$$

Thus the Lemma holds with $c_2 = 2\mathbb{E}[f^2(\xi_{\mathbf{0}})] + (\mathbb{E}[f(\xi_{\mathbf{0}})])^2 > 0$ for $n\in\mathbb{N}$ large and for any $n\in\mathbb{N}$ with maybe a different constant $c_2>0$. □

In order to prove condition (A1'), notice first that Corollary 3.3 and analogous arguments involved in (4.2)-(4.4) imply that $\log\hat f_{B'_n}(\xi_{\mathbf{0}})$ converges to $\log f(\xi_{\mathbf{0}})$ in L^2. Therefore, $\mathbb{E}[\log^2\hat f_{B'_n}] \xrightarrow{n\to\infty} \mathbb{E}[\log^2 f(\xi_{\mathbf{0}})]$, and since $\mathbb{E}[\log^2 f(\xi_{\mathbf{0}})] < L_1$ 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, Lemmata 5.8 and 5.9 and the m_n-dependence yield

$$\begin{aligned}&\int_{\mathbb{R}^d_+} |\operatorname{Cov}(\log\hat f_{B'_n},\log\hat f_{B'_n+y})|dy = \int_{B'_n} |\operatorname{Cov}(\log\hat f_{B'_n},\log\hat f_{B'_n+y})|dy\\ &\le \frac{c_1c_2}{\lambda|B'_n|^2}\int_{B'_n}|B'_n\cap(B'_n+y)|\,dy = \frac{c_1c_2}{\lambda^2 2^d} < \infty.\end{aligned}$$

Condition (A2)

Since $q_n = m_n^{1+\delta'}$ with $4\delta' < \delta$, we have

$$\frac{m_n}{q_n}\xrightarrow{n\to\infty} 0 \qquad\text{and}\qquad \frac{q_n}{p_n} = \frac{m_n^{1+\delta'}}{m_n^{4+\delta}}\xrightarrow{n\to\infty} 0. \tag{5.6}$$

Condition (A3')

Recall that we are assuming that the density f is continuous.

Lemma 5.10. *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 $\hat{f}_{B'_n}(\xi_{\mathbf{0}})$. Note that $\xi_{\mathbf{0}}$ is a generic mark that is independent of the MPP Ψ. From Theorem 3.8 and since M is compact and f continuous, we have that $\hat{f}_n(\eta) \xrightarrow{n\to\infty} f(\eta) \leq \|f\|_\infty$ 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 $\hat{f}_{B'_n}(\xi_{\mathbf{0}})$. $\square$

Lemma 5.11. *It holds that*

$$\liminf_{n\to\infty} \int_{\mathbb{R}^d_+} \frac{|C_{\mathbf{q_n}} \cap (C_{\mathbf{q_n}} - y)|}{|C_{\mathbf{q_n}}|} \operatorname{Cov}(\log \hat{f}_{B'_n}(\xi_{\mathbf{0}}), \log \hat{f}_{B'_n+y}(\xi'_y))\, dy > 0.$$

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 non-decreasing functions $h, g\colon \mathbb{R} \to \mathbb{R}$ such that the expectations forming the covariance $\operatorname{Cov}(h(\hat{f}_{B'_n+y_1}), g(\hat{f}_{B'_n+y_2}))$ exist, $\operatorname{Cov}(h(\hat{f}_{B'_n+y_1}), g(\hat{f}_{B'_n+y_2})) \geq 0$. In view of Lemma 5.12 we thus have $\operatorname{Cov}(\log \hat{f}_{B'_n+y_1}, \log \hat{f}_{B'_n+y_2}) \geq 0$ 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 5.10 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

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

Together with the calculations in the proof of Lemma 5.9, this yields

$$\begin{aligned}
&\int_{\mathbb{R}^d_+} \frac{|C_{\mathbf{q_n}} \cap (C_{\mathbf{q_n}} - y)|}{|C_{\mathbf{q_n}}|} \operatorname{Cov}(\log \hat{f}_{B'_n}, \log \hat{f}_{B'_n+y})\, dy \\
&\geq \frac{c_3}{(\|f\|_\infty + \varepsilon)^2 \lambda |B'_n|^2} \int_{B'_n} \frac{|C_{\mathbf{q_n}} \cap (C_{\mathbf{q_n}} - y)|}{|C_{\mathbf{q_n}}|} |B'_n \cap (B'_n + y)| dy \\
&= \frac{c_3}{(\|f\|_\infty + \varepsilon)^2 \lambda m_n^{2d} q_n^d} \Big(\int_0^{m_n} (q_n - y)(m_n - y) dy \Big)^d \\
&= \frac{c_3}{(\|f\|_\infty + \varepsilon)^2 \lambda m_n^{2d} q_n^d} \Big(\frac{q_n m_n^2}{2} - \frac{m_n^2}{2} + \frac{m_n^3}{3} \Big)^d
\end{aligned}$$

for some constant $c_3 > 0$, and the whole expression tends to $\frac{c_3}{2^d \|f\|_\infty^2} > 0$ as $n \to \infty$. $\square$

5.2.2 Condition (A4)

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

Lemma 5.12. *There exists a constant $c_4 > 0$ such that for any $y \in \mathbb{R}^d$ and $n \in \mathbb{N}$*

$$\mathbb{E}\big[\big|\log \hat{f}_{B'_n+y}\big|^3\big] \le c_4.$$

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

$$\begin{aligned}\mathbb{E}\big[\big|\log \hat{f}_{B'_n}\big|^3\big] \le \mathbb{E}\big[\big|\log \hat{f}_{B'_n} - \log \mathbb{E}[\hat{f}_{B'_n}]\big|^3\big] + 3\,\big|\log \mathbb{E}[\hat{f}_{B'_n}]\big|\mathbb{E}\big[\big|\log \hat{f}_{B'_n} - \log \mathbb{E}[\hat{f}_{B'_n}]\big|^2\big]\\ + 3(\log \mathbb{E}[\hat{f}_{B'_n}])^2\mathbb{E}\big[\big|\log \hat{f}_{B'_n} - \log \mathbb{E}[\hat{f}_{B'_n}]\big|\big] + \big|\log \mathbb{E}[\hat{f}_{B'_n}]\big|^3.\end{aligned}$$

By Corollary 3.3, $\log \mathbb{E}[\hat{f}_{B'_n}] \xrightarrow{n\to\infty} \log \mathbb{E}[f(\xi_{\mathbf{0}})]$, and since f is continuous, any power of this quantity is also bounded. Hence 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, Theorem 3.8 and (4.3) yield

$$\mathbb{E}[|\log \hat{f}_{B'_n} - \log \mathbb{E}[\hat{f}_{B'_n}]|^3] \le \frac{8\mathbb{E}\big[\big|\hat{f}_{B'_n} - \mathbb{E}[\hat{f}_{B'_n}]\big|^3\big]}{c_0^3}$$

and it suffices to prove that $\mathbb{E}[|\hat{f}_{B'_n}|^3]$ is finite for large $n \in \mathbb{N}$. In view of the Campbell theorem,

$$\begin{aligned}\mathbb{E}[|\hat{f}_{B'_n}(\xi'_{\mathbf{0}})|^3] = \frac{1}{\lambda^2|B'_n|^2}\mathbb{E}[F_n^3(\xi'_{\mathbf{0}},\xi_1)] + \frac{1}{\lambda|B'_n|}\mathbb{E}[F_n^2(\xi'_{\mathbf{0}},\xi_1)F_n(\xi'_{\mathbf{0}},\xi_2)]\\ + \mathbb{E}[F_n(\xi'_{\mathbf{0}},\xi_1)F_n(\xi'_{\mathbf{0}},\xi_2)F_n(\xi'_{\mathbf{0}},\xi_3)],\end{aligned} \tag{5.7}$$

where ξ_1, ξ_2, ξ_3 are independent copies of $\xi'_{\mathbf{0}}$. Moreover, following the proof of Lemma 3.6 we find constants $C_\theta, K_0 > 0$ such that for $n \in \mathbb{N}$ large enough,

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

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

as well as

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

Plugging this into (5.7) we obtain

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

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

For the ease of reading, we set $X_{n,y} := \log \hat{f}_{B_n'+y} - \mathbb{E}[\log \hat{f}_{B_n'+y}]$, $y \in \mathbb{R}_+^d$. The aim of this paragraph is to show

$$\lim_{n\to\infty} \frac{1}{q_n^d} \mathbb{E}\Big[\Big(\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\Big)^2 \mathbb{1}_{\{|\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}|>\varepsilon p_n^{d/2}\}}\Big] = 0,$$

with q_n as in (5.6). Applying Hölder inequality, the expectation in the limit can be bounded by

$$\Big(\mathbb{E}\Big[\Big(\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\Big)^3\Big]\Big)^{2/3}\Big(\mathbb{P}\big(\big|\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\big| > \varepsilon p_n^{d/2}\big)\Big)^{1/3}.$$

On the one hand, repeated application of the Campbell theorem, the stationarity of Π^* and the generalized Hölder inequality lead to

$$\begin{aligned}\mathbb{E}\Big[\Big(\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\Big)^3\Big] &\le \lambda q_n^d \mathbb{E}[X_{n,\mathbf{0}}^3] + \int_{C_{\mathbf{q_n}}^2} (\mathbb{E}[X_{n,y_1}^3])^{1/3}(\mathbb{E}[X_{n,y_2}^3])^{2/3}\,\alpha^{(2)}(dy_1,dy_2)\\ &\quad + \int_{C_{\mathbf{q_n}}^3} (\mathbb{E}[X_{n,y_1}^3])^{1/3}(\mathbb{E}[X_{n,y_2}^3])^{1/3}(\mathbb{E}[X_{n,y_3}^3])^{1/3}\,\alpha^{(3)}(dy_1,dy_2,dy_3)\\ &= \mathbb{E}[X_{n,\mathbf{0}}^3](\lambda q_n^d + \lambda^2 q_n^{2d} + \lambda^3 q_n^{3d}) \le c_4(\lambda q_n^d + \lambda^2 q_n^{2d} + \lambda^3 q_n^{3d}),\end{aligned}$$

where last inequality follows from Lemma 5.12. On the other hand, Chebyshev's inequality and condition (A1') yield

$$\mathbb{P}\big(\big|\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\big| > \varepsilon p_n^{d/2}\big) \le \frac{1}{\varepsilon^2 p_n^d}\operatorname{Var}\Big(\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\Big) \le \frac{\lambda^2 q_n^d c}{\varepsilon^2 p_n^d}.$$

Hence we get

$$\frac{1}{q_n^d}\mathbb{E}\Big[\Big(\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}\Big)^2 \mathbb{1}_{\{|\sum_{y\in\Pi^*\cap C_{\mathbf{q_n}}} X_{n,y}|>\varepsilon p_n^{d/2}\}}\Big] \le \left(\frac{c_4\sqrt{c}}{\varepsilon}\right)^{\frac{2}{3}} \left(\frac{\lambda^2 + \lambda^3 q_n^d + \lambda^4 q_n^{2d}}{p_n^{d/2}}\right)^{\frac{2}{3}},$$

which tends to zero as $n \to \infty$ due to the choice of q_n.

References

[1] I.A. Ahmad and P.E. Lin, *A nonparametric estimation of the entropy for absolutely continuous distributions*, IEEE Trans. Information Theory **IT-22** (1976), no. 3, 372–375.

[2] Z. D. Bai, C. Radhakrishna Rao, and L. C. Zhao, *Kernel estimators of density function of directional data*, J. Multivariate Anal. **27** (1988), no. 1, 24–39.

[3] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. van der Meulen, *Nonparametric entropy estimation: an overview*, Int. J. Math. Stat. Sci. **6** (1997), no. 1, 17–39.

[4] A. L. Besse, *Manifolds all of whose geodesics are closed*, Results in Mathematics and Related Areas, vol. 93, Springer-Verlag, Berlin, 1978.

[5] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, second ed., Pure and Applied Mathematics, vol. 120, Academic Press Inc., Orlando, FL, 1986.

[6] A. Bulinski and A. Shashkin, *Limit theorems for associated random fields and related systems*, Advanced Series on Statistical Science & Applied Probability, 10, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[7] R. Burton and E. Waymire, *Scaling limits for associated random measures*, Ann. Probab. **13** (1985), no. 4, 1267–1278.

[8] D.J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. II*, second ed., Probability and its Applications (New York), Springer, New York, 2008, General theory and structure.

[9] M. El Machkouri, *Kernel density estimation for stationary random fields*, ALEA Lat. Am. J. Probab. Math. Stat. **11** (2014), no. 1, 259–279.

[10] P. Hall, G. S. Watson, and J. Cabrera, *Kernel density estimation with spherical data*, Biometrika **74** (1987), no. 4, 751–762.

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

[12] L. Heinrich, S. Klein, and M. Moser, *Empirical Mark covariance and product density function of stationary marked point processes—a survey on asymptotic results*, Methodol. Comput. Appl. Probab. **16** (2014), no. 2, 283–293.

[13] L. Heinrich, S. Lück, and V. Schmidt, *Asymptotic goodness-of-fit tests for the palm mark distribution of stationary point processes with correlated marks*, Bernoulli **20** (2014), no. 4, 1673–1697.

[14] G. Henry and D. Rodriguez, *Kernel density estimation on Riemannian manifolds: asymptotic results*, J. Math. Imaging Vision **34** (2009), no. 3, 235–239.

[15] M. Kiderlen, *Non-parametric estimation of the directional distribution of stationary line and fibre processes*, Adv. in Appl. Probab. **33** (2001), no. 1, 6–24.

[16] I. Kontoyiannis, P. H. Algoet, Yu. M. Suhov, and A. J. Wyner, *Nonparametric entropy estimation for stationary processes and random fields, with applications to English text*, IEEE Trans. Inform. Theory **44** (1998), no. 3, 1319–1327.

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

[18] Z. Pawlas, *Empirical distributions in marked point processes*, Stochastic Processes and their Applications **119** (2009), no. 12, 4194 – 4209.

[19] B. Pelletier, *Kernel density estimation on Riemannian manifolds*, Statist. Probab. Lett. **73** (2005), no. 3, 297–304.

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

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

[22] T. Sakai, *Riemannian geometry*, Translations of Mathematical Monographs, vol. 149, American Mathematical Society, Providence, RI, 1996.

[23] V. Schmidt (ed.), *Stochastic geometry, spatial statistics and random fields. models and algorithms*, Lecture Notes in Mathematics, vol. 2120, Springer, Heidelberg, 2015.

[24] T. Schuster, *The method of approximate inverse: theory and applications*, Lecture Notes in Mathematics, vol. 1906, Springer, Berlin, 2007.

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

[26] E. Spodarev (ed.), *Stochastic geometry, spatial statistics and random fields. Asymptotic methods*, Lecture Notes in Mathematics, vol. 2068, Springer, Heidelberg, 2013.

[27] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1987.

[28] A. B. Tsybakov, *Introduction to nonparametric estimation*, Springer Series in Statistics, Springer, New York, 2009.

[29] Y. Wang and M. Woodroofe, *On the asymptotic normality of kernel density estimators for causal linear random fields*, J. Multivariate Anal. **123** (2014), 201–213.