Annals of the Institute of Statistical Mathematics (2024) 76:263–287
https://doi.org/10.1007/s10463-023-00883-5

Multivariate frequency polygon for stationary random fields

Michel Carbon[1] **· Thierry Duchesne**[1]

Received: 9 September 2022 / Revised: 3 August 2023 / Accepted: 29 August 2023 /
Published online: 8 November 2023

Abstract
The purpose of this paper is to investigate the multivariate frequency polygon as a density estimator for stationary random fields indexed by multidimensional lattice points space. Optimal cell widths that asymptotically minimize integrated mean square error (IMSE) are derived. Under weak conditions, the IMSE of frequency polygons achieves the same rate of convergence to zero as that of kernel estimators. The frequency polygon can also attain the optimal uniform rate of convergence and the almost sure convergence under general conditions. Finally, a result of L^1 convergence is given. Frequency polygons thus appear to be very good density estimators with respect to the criteria of IMSE, of uniform convergence, of almost sure convergence and of L^1 convergence. We apply our results to simulated data and real data.

Keywords Bandwidth · Density estimation · Frequency polygons · Mixing field · Random field

1 Introduction

Our goal in this paper is to study the multivariate frequency polygon as a density estimator for multivariate random variables which show spatial interaction. We sense a practical need for nonparametric spatial estimation for situations in which parametric families cannot be adopted with confidence. There are several ways for defining a linear interpolant of a multivariate histogram with hyper-rectangular cells, where $\mathbf{x} \in \mathbb{R}^d$. To define a frequency polygon in the multivariate case, we choose the so-called *linear blend frequency polygon* (LBFP), initially introduced by Terell (1983) and Hjort (1986). We'll define precisely the LBFP in Section 2. In general

✉ Thierry Duchesne
thierry.duchesne@mat.ulaval.ca

1 Department of Mathematics and Statistics, Université Laval, 1056, avenue de la Médecine, Québec, QC G1V 0A6, Canada

dimension d, a single portion of a linear blend extends over a hyper-rectangle with 2^d vertices, defined by the centers of the 2^d adjacent histogram cells. Any cut of the surface parallel to a coordinate axis gives a linear fit.

The LBFP is a good estimator of the density and its construction involves a computational effort about equivalent to that of the multivariate histogram, it is easier to explain to non-mathematician users than kernel estimators, and it surprisingly has the same rates of convergence as the kernel estimator. This paper is a generalization of results of Scott (1985) in the i.i.d. case, of Carbon et al. (1997a) for random processes, and of Carbon (2006) for random fields. The proofs herewith thus use techniques very specific to random fields. We also add a rate of almost sure (a.s.) convergence and a result of L^1 convergence.

Compared to kernel estimation, we cannot benefit from assumptions about the regularity of kernels, which are intensively used to obtain convergence results in that context. For example, the LBFP has no derivative at some points, which makes it more difficult to prove convergence. However, it will be shown that the LBFP still achieves the same rates of convergence as the kernel estimator.

Denote the integer lattice points in the N-dimensional Euclidean space by $Z^N, N \geq 1$. Consider a strictly stationary random field $\{X_{\mathbf{n}}\}$ indexed by $\mathbf{n}$ in Z^N and defined on some probability space $(\Omega, \mathcal{F}, P)$. A point $\mathbf{n}$ in Z^N will be referred to as a site. For a site $\mathbf{i} = (i_1, \ldots, i_N)$, we denote $||\mathbf{i}|| = (i_1^2 + \cdots + i_N^2)^{1/2}$. We will write n instead of $\mathbf{n}$ when $N = 1$. For two finite sets of sites S and S', the Borel fields $\mathcal{B}(S) = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S)$ and $\mathcal{B}(S') = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S')$ are the σ-fields generated by the random variables $X_{\mathbf{n}}$ with $\mathbf{n}$ ranging over S and S', respectively. Denote the Euclidean distance between S and S' by dist(S, S'). We will assume that $X_{\mathbf{n}}$ satisfies the following mixing condition: there exists a function $\varphi(t) \downarrow 0$ as $t \to \infty$, such that whenever $S, S' \subset Z^N$,

$$\begin{aligned} \alpha\{\mathcal{B}(S), \mathcal{B}(S')\} &= \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}(S')\} \\ &\leq h\{\text{Card}(S), \text{Card}(S')\}\varphi\{\text{ dist } (S, S')\}, \end{aligned} \tag{1}$$

where Card(S) denotes the cardinality of S. Here h is a symmetric positive function nondecreasing in each variable. Throughout the paper, assume that h satisfies either

$$h(n, m) \leq \min\{m, n\} \tag{2}$$

or

$$h(n, m) \leq C(n + m + 1)^{\tilde{k}} \tag{3}$$

for some $\tilde{k} \geq 1$ and some $C > 0$. If $h \equiv 1$, then $X_{\mathbf{n}}$ is called strongly mixing. Conditions (2) and (3) are the same as the mixing conditions used by, respectively, Neaderhouser (1980) and Takahata (1983) and are weaker than the uniform mixing condition used by Nahapetian (1980). Conditions (2) and (3) are satisfied by many spatial models. Examples can be found in Neaderhouser (1980), Rosenblatt (1985) and Guyon (1987). You can refer to the basing book on mixing properties of Doukhan (1994).

For relevant works on random fields, see e.g. (Neaderhouser 1980; Bolthausen 1982; Guyon and Richardson 1984; Guyon 1987; Nahapetian 1987; Tran 1990; Tran and Yakowitz 1993; Carbon et al. 1996; Carbon et al. 1997b; Biau 2003; Carbon 2006, 2014; Hallin et al. 2001, 2004a, b; Harel, Lenain and Ngatchou-Wandji 2016; Robinson 1983, 2011).

Denote by $\mathcal{I}_{\mathbf{n}}$ a rectangular region defined by

$$\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} : \mathbf{i} \in Z^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}.$$

Assume that we observe $\{X_{\mathbf{n}}\}$ on $\mathcal{I}_{\mathbf{n}}$. Suppose $X_{\mathbf{n}}$ takes values in $\mathbb{R}^d$ and has an uniformly continuous density f. We write $\mathbf{n} \to \infty$ if

$$\min\{n_k\} \to \infty \quad \text{and} \quad |n_j/n_k| < C \tag{4}$$

for some $0 < C < \infty$, $1 \leq j, k \leq N$. All limits are taken as $\mathbf{n} \to \infty$ unless indicated otherwise. Define $\hat{\mathbf{n}} = n_1 \dots n_N$.

Under weak dependence conditions, the integrated mean squared error (IMSE) of LBFP $\hat{f}_{\mathbf{n}}$ is shown to achieve the rate of convergence to zero of order $\hat{\mathbf{n}}^{-4/(4+d)}$. In the case $N = d = 1$, histograms can only achieve the slower rate of convergence of the IMSE of order $n^{-2/3}$. It is also established that frequency polygons attain a uniform rate of convergence under appropriate smoothness conditions. It is the optimal rate of convergence for nonparametric estimators of a density function in the i.i.d. case for $N = d = 1$ (see Stone 1983). We herein obtain similar results for random fields. The LBFP thus appears to be a very good density estimator with respect to both criteria of IMSE and uniform convergence. We also establish an optimal rate of a.s. convergence and finally a result of L^1 convergence.

Our paper is organized as follows. Section 2 provides assumptions and main results. The optimal choice of the cell width which asymptotically minimizes the IMSE is given in Theorem 1, which generalizes results of Scott (1985), Carbon et al. (1997a, b) and Carbon (2006). Weak conditions for the uniform convergence of $\hat{f}_{\mathbf{n}}$ on $\mathbb{R}^d$ are also provided. Sharp rates of uniform convergence and a rate of a.s. convergence of $\hat{f}_{\mathbf{n}}$ are established. A result of L^1 convergence is finally obtained. In Section 3, we provide numerical applications of the method; we first illustrate the convergence of the estimator through simulations then show an application of the LBFP to hydrological data. Section 4 is devoted to some preliminary lemmas. The proofs of the main theorems are postponed to Section 5. Online supplementary materials include a web appendix with additional technical details and R code to compute the LBFP in dimension $d = 2$.

We use x to denote a fixed point of $\mathbb{R}^d$. The integer part of a number a is denoted by $[a]$. The letter C will be used to denote constants whose values are unimportant. The letter D denotes an arbitrary compact set in $\mathbb{R}^d$.

2 Assumptions and main results

We define a grid of separate cells I_k in $\mathbb{R}^d$ with centres x_k and volume $b_1 \times \cdots \times b_d$. Then, for any $x \in \mathbb{R}^d$, there exists a unique k such that $x \in I_k$. A standard multivariate histogram estimate can be defined for $x \in I_0(k)$ by

$$f_0(x) = \frac{1}{\hat{\mathbf{n}} b_1 \dots b_d} \nu_k, \tag{5}$$

where $I_0(k) = \prod_{s=1}^{d} \left[x_{k,s} - \frac{1}{2}b_s\,;\, x_{k,s} + \frac{1}{2}b_s\right)$ and ν_k is the number of observations $X_{\mathbf{i}}$ falling in this cell. For f_0 to be consistent, it is necessary that $b_1 = b_{1,\mathbf{n}} \to 0, \dots, b_d = b_{d,\mathbf{n}} \to 0$ and that $\hat{\mathbf{n}} b_1 \dots b_d \to \infty$ as the number of observations tends to infinity; these requirements are standard and we will assume them to be true in the following.

To generalize the standard notion of an univariate frequency polygon, we want to linearly combine nearby the 2^d values of f_0, defined in (5). Here, we first fix a particular k and consider the new cell

$$I(k) = \prod_{s=1}^{d} \left[x_{k,s}\,,\, x_{k,s} + b_s\right) = \prod_{s=1}^{d} \left[a_s - \frac{1}{2}b_s\,,\, a_s + \frac{1}{2}b_s\right),$$

where $a = x_k + \frac{1}{2}b$ is the middle of the 2^d histogram cells. The generalized frequency polygon cell $I(k)$ lies within 2^d histogram cells $I_0(k;j_1, \dots, j_d)$ with $j_1, \dots, j_d \in \{0, 1\}^d$.

Following (Scott 2015), the *linear blend frequency polygon* (LBFP) is defined by

$$\hat{f}_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}}\, b_1 \dots b_d} \sum_{j_1,\dots,j_d \in \{0,1\}^d} c_{j_1,\dots,j_d}(x)\, \nu_{k_1+j_1,\dots,k_d+j_d}, \quad x \in I(k), \tag{6}$$

where $\nu_{k_1+j_1,\dots,k_d+j_d}$ is the number of observations falling in the cell

$$\begin{aligned} I(k,j) &= \prod_{s=1}^{d} \left[x_{k,s} + \left(j_s - \frac{1}{2}\right)b_s\,,\, x_{k,s} + \left(j_s + \frac{1}{2}\right)b_s\right) \\ &= \prod_{s=1}^{d} \left[a_s + (j_s - 1)b_s\,,\, a_s + j_s b_s\right), \end{aligned}$$

and

$$c_{j_1,\dots,j_d}(x) = \prod_{s=1}^{d} u_s^{j_s}(1 - u_s)^{1-j_s}$$

with

$$u_s = u_s(x) = \frac{(x_s - x_{k,s})}{b_s}, \quad x_s \in \left[x_{k,s}\,,\, x_{k,s} + b_s\right) = \left[a_s - \frac{1}{2}b_s\,,\, a_s + \frac{1}{2}b_s\right), s = 1, \dots, d.$$

We can remark that u_s goes linearly from 0 to 1 as x_s moves from the left to the right side of the s-th side of the cell $I(k)$. We also notice that all the $c_{j_1,\dots,j_d}(x)$ are non negative and that their sum is equal to 1. Furthermore, we can see that $\hat{f}_{\mathbf{n}} = f_0$ at each of the 2^d corners of $I(k)$, that is to say $c_{j_1,\dots,j_d}(x) = 1$ when $(u_1, \dots, u_d) = (j_1, \dots, j_d)$, $\hat{f}_{\mathbf{n}}$ is the plain average of the 2^d nearby corner values at the centre point a. So, in some sense, the coefficient $c_{j_1,\dots,j_d}(x)$ measures the closeness of x to corner $j_1,\dots,j_d$. In general, approximating a function in a rectangle by linear interpolation, with weights $c_{j_1,\dots,j_d}$, is known in numerical analyst circles as linear blend interpolation. So the proposed LBFP is accordingly the *linear blend* of the histogram. For background material on LBFP, see Scott (2015).

Noting $j = (j_1,\dots,j_d)$, we define

$$Y_{\mathbf{i},k,j} = \begin{cases} 1 & \text{if } X_{\mathbf{i}} \in I(k,j) \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad p_k(j) = \int_{I(k,j)} f(x)\,dx. \tag{7}$$

Then $\nu_{k_1+j_1,\dots,k_d+j_d} = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k,j}$ and

$$E\big(\nu_{k_1+j_1,\dots,k_d+j_d}\big) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} P\big(X_{\mathbf{i}} \in I(k,j)\big) = \hat{\mathbf{n}} \int_{I(k,j)} f(x)\,dx = \hat{\mathbf{n}}\, p_k(j).$$

We suppose that $(X_{\mathbf{i}}, X_{\mathbf{j}})$ admits a joint density. Denote the conditional density of $X_{\mathbf{j}}$ given $X_{\mathbf{i}}$ by $f_{\mathbf{j}|\mathbf{i}}$ for simplicity.

Assumption 1 For all $\mathbf{i}$, $\mathbf{j}$ and some constant M_1,

$$\sup_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d} f_{\mathbf{j}|\mathbf{i}}(y|x) \le M_1.$$

Assumption 2 f is three times differentiable, f''' is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, f has continuous and square integrable partial derivatives $f_s, f_{s,t}, f_{s,t,m}$ with $s,t,m \in \{1,\dots,d\}$ and $f^{1/2} \in L^1$.

Assumption 3 There exists a constant $C > 0$ such that

$$\big|f(x) - f(x')\big| \le C\big\|x - x'\big\| \quad \text{for any} \quad x, x' \in \mathbb{R}^d.$$

Even though the proof uses techniques specific to random fields, the following theorem is a generalization of results found in Scott (1985) and Carbon et al. (1997a), and it hopefully simplifies to the same results when $N = d = 1$.

Remember that the definition of the IMSE is

$$IMSE = \int_{\mathbb{R}^d} E\big[\hat{f}_{\mathbf{n}}(x) - f(x)\big]^2 dx.$$

Theorem 1 *If Assumptions* 1 *and* 2 *hold and* $\varphi(k) = O(k^{-\rho})$ *for some* $\rho > 2N + (3/2)$, *the value of the cell width that asymptotically minimizes the IMSE of the LBFP is*

$$b_i = b_{\mathbf{n},i} = O\left(\hat{\mathbf{n}}^{-1/(4+d)}\right) \qquad i = 1, \dots, d$$

with corresponding

$$IMSE = O\left(\hat{\mathbf{n}}^{-4/(4+d)}\right).$$

It follows from Theorem 1 that the integrated mean square of $\hat{f}_{\mathbf{n}} - f$ can achieve the rate of convergence to 0 of order $\hat{\mathbf{n}}^{-4/(4+d)}$. This is the same optimal rate of convergence to zero of the IMSE as that of non-negative kernel estimators in the i.i.d. case for $N = d = 1$.

To study the uniform convergence of the LBFP estimator, we let

$$\theta_1 = \frac{\rho + 3N}{\rho - 3N}, \quad \theta_2 = \frac{\rho + 4N}{\rho - 2N(\tilde{k} + 1)}, \quad \theta_3 = \frac{2N - \rho}{\rho - 2N(\tilde{k} + 1)}.$$

We also let $\tilde{b}_{\mathbf{n}} = \max_{i=1,\dots,d} b_{i,\mathbf{n}}$, $\tilde{\Psi}_{\mathbf{n}} = \{\log \hat{\mathbf{n}}(\hat{\mathbf{n}} b_1 \dots b_d)^{-1}\}^{1/2}$ and $\Psi_{\mathbf{n}} = \max\left(\tilde{b}_{\mathbf{n}}; \tilde{\Psi}_{\mathbf{n}}\right)$.

Theorem 2 *Suppose that $\varphi(k) = O(k^{-\rho})$ for some $\rho > 0$ and that Assumption 3 holds.*

(*i*) *If* (2) *is satisfied*, $\rho > 3N$ *and*

$$\hat{\mathbf{n}}(b_1 \dots b_d)^{\theta_1} (\log \hat{\mathbf{n}})^{-1} \to \infty, \tag{8}$$

(*ii*) *or if* (3) *is satisfied*, $\rho > 2N(\tilde{k} + 1)$ *and*

$$\hat{\mathbf{n}}(b_1 \dots b_d)^{\theta_2} (\log \hat{\mathbf{n}})^{\theta_3} \to \infty, \tag{9}$$

then

$$\sup_{x \in D} |\hat{f}_{\mathbf{n}}(x) - f(x)| = O(\Psi_{\mathbf{n}}) \quad \text{in probability.} \tag{10}$$

Remark 1 Under the assumptions of Theorem 2, (i) implies that $\theta_1 > 1$ and (8) implies that

$$\hat{\mathbf{n}}(b_1 \dots b_d) \to \infty, \tag{11}$$

which is a condition for $f_{\mathbf{n}}(x)$ to converge to $f(x)$ in the case $N = d = 1$. Similarly, it can be shown that (ii) implies (11).

Remark 2 Such strongly mixing stationary random fields do exist. For instance consider the two-dimensional autoregressive model

$$X_{t_1,t_2} = \phi_1 X_{t_1-1,t_2} + \phi_2\left(X_{t_1,t_2-1} - \phi_1 X_{t_1-1,t_2-1}\right) + \varepsilon_{t_1,t_2}$$

where $(t_1, t_2) \in \mathbb{Z}^2$, $|\phi_1| < 1$, $|\phi_2| < 1$ and where ε_{t_1,t_2} is a white noise with $Var\left(\varepsilon_{t_1,t_2}\right) = \sigma_\varepsilon^2$. Then Kulkarni (1992) proves that

$$Corr\left(X_{t_1,t_2}, X_{t'_1,t'_2}\right) = \phi_1^{|t_1-t'_1|}\phi_2^{|t_2-t'_2|}.$$

Example 1 (i) Take $b_{i,\mathbf{n}} = \left(\hat{\mathbf{n}}^{-1}\log\hat{\mathbf{n}}\right)^{1/3}$ for $i = 1, \dots, d$. Then $\Psi_{\mathbf{n}} = \left(\hat{\mathbf{n}}^{-1}\log\hat{\mathbf{n}}\right)^{(3-d)/6}$, which is the optimal rate $\left(n^{-1}\log n\right)^{1/3}$ for the *i.i.d.* case when $N = d = 1$.

(ii) Take the optimal cell width $b_{i,\mathbf{n}} = \hat{\mathbf{n}}^{-1/(4+d)}$ for $i = 1, \dots, d$. Then (8) is satisfied if $\rho > (3/2)Nd + 3N$ and (9) is satisfied if $\rho > (N/2)\{3d + \tilde{k}(d+4) + 4\}$.

In the following theorem, we give a rate of the a.s. convergence of the LBFP estimator. Let ε be an arbitrary small positive number and denote $g(\mathbf{n}) = \prod_{i=1}^{N}(\log n_i)(\log\log n_i)^{1+\varepsilon}$. Clearly, $\sum \frac{1}{\hat{\mathbf{n}}g(\mathbf{n})} < \infty$, where the summation is over all $\mathbf{n}$ in $\mathbb{Z}^N$. Define

$$\theta_1^* = \frac{\rho + 3N}{\rho - 5N}, \quad \theta_2^* = \frac{3N - \rho}{\rho - 5N}$$
$$\theta_3^* = \frac{\rho + 3N}{\rho - (2\tilde{k}+3)N}, \quad \theta_4^* = \frac{N - \rho}{\rho - (2\tilde{k}+3)N}.$$

Theorem 3 *Suppose $\varphi(k) = O(k^{-\rho})$ for some $\rho > 0$ and Assumption 3 hold.*

(i) If (2) is satisfied, $\rho > 5N$ and

$$\hat{\mathbf{n}}(b_1 \dots b_d)^{\theta_1^*}(\log\hat{\mathbf{n}})^{\theta_2^*}(g(\mathbf{n}))^{-2N/(\rho-5N)} \to \infty, \tag{12}$$

(ii) or if (3) is satisfied, $\rho > (2\tilde{k}+3)N$ and

$$\hat{\mathbf{n}}(b_1 \dots b_d)^{\theta_3^*}(\log\hat{\mathbf{n}})^{\theta_4^*}(g(\mathbf{n}))^{-2N/\{\rho-(2\tilde{k}+3)N\}} \to \infty. \tag{13}$$

then

$$\sup_{x\in D}|\hat{f}_{\mathbf{n}}(x) - f(x)| = O(\Psi_{\mathbf{n}}) \quad \text{a.s.} \tag{14}$$

The following result states that the L^1 convergence is a consequence of the previous theorem.

Theorem 4 *Under the assumptions of Theorem 3 we have*

$$\int_{x\in\mathbb{R}^d}|\hat{f}_{\mathbf{n}}(x) - f(x)|\,dx \to 0 \quad \text{a.s.}$$

Remark 3 The hypothesis on the mixing condition and the condition on the increasing size of the sample could be relaxed if, for example, it were replaced by a condition on the form of the underlying generating process. As it is done in the article of Carbon (2014) with a linear stationary field, by adding conditions on the coefficients

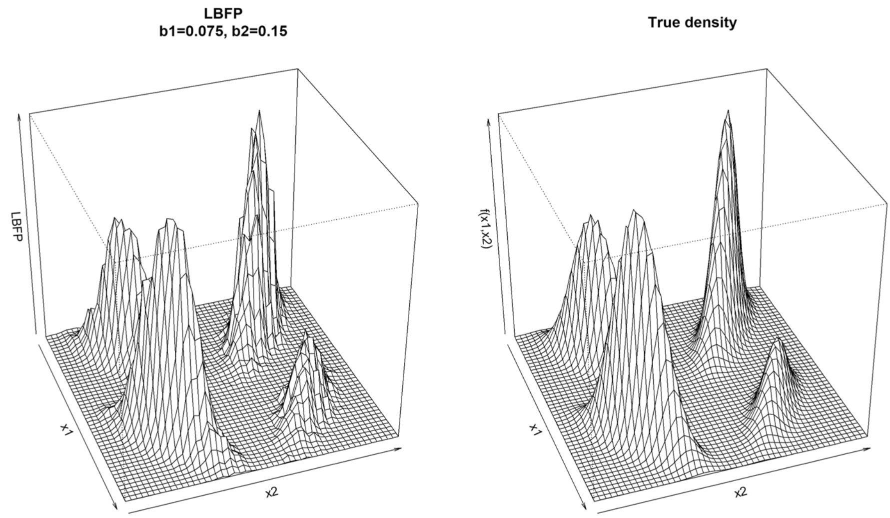

Fig. 1 LBFP estimation (left) and true density (right) of an autocorrelated mixture of bivariate normals

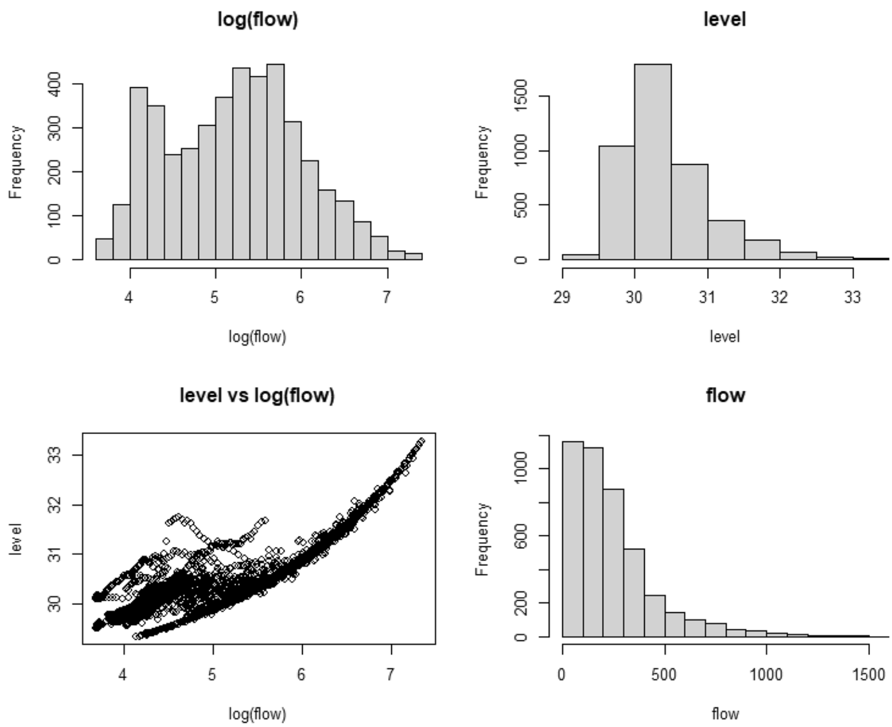

Fig. 2 Descriptive graphs for the level and flow variables. Top left: histogram of log(flow). Top right: histogram of level. Bottom left: scatter plot of level vs log(flow). Bottom right: histogram of flow

of this linear field. We could also draw inspiration from the article by Robinson (2011) for other general conditions.

3 Applications

In order to illustrate the previous results, two examples are presented here, one with a simulation and another with real data. We chose $d = 2$ to be able to make graphs. The first example with simulations makes it possible to compare a known true density with its LBFP estimate. The second example is an application of the LBFP to real hydrological data. General R code to implement the LBFP with $d = 2$ is provided in the online supplement.

3.1 Simulations from an autocorrelated mixture of bivariate normal laws

Let $\{Y_t; t = 0, 1, 2, \dots\}$ be an irreducible aperiodic Markov chain on the state space $\{1, \dots, 4\}$ with stationary distribution $\tilde{\pi} = (\pi_1, \dots, \pi_4)$. Suppose $X_t = (X_{t1}, X_{t2}) \mid Y_t = y \sim \mathcal{N}_2(\mu_y, V_y)$. When drawing Y_0 from $\tilde{\pi}$, the (stationary) distribution of X_t is easily shown to be a mixture of the four two-dimensional normal distributions with weights given by the elements of $\tilde{\pi}$. The Markov chain then dictates which normal law one should draw the next pair X_t from.

We first draw Y_0 from $\tilde{\pi}$, then we draw 100 000 realizations of Y_t and X_t from the model above. The parameters are chosen so that there is strong autocorrelation between consecutive pairs. The transition matrix of the chain is

$$\mathcal{P} = \begin{pmatrix} 0.7 & 0.2 & 0.05 & 0.05 \\ 0.15 & 0.7 & 0.1 & 0.05 \\ 0 & 0.15 & 0.8 & 0.05 \\ 0.15 & 0.05 & 0.2 & 0.6 \end{pmatrix},$$

leading to $\tilde{\pi} = (2/9, 1/3, 1/3, 1/9)$. The means of the normals are $\mu_1 = \binom{-1.5}{-1.5}$, $\mu_2 = \binom{-1.5}{1.5}$, $\mu_3 = \binom{1.5}{-1.5}$ and $\mu_4 = \binom{1.5}{1.5}$. The variance matrices are $V_1 = V_2 = \begin{pmatrix} 0.15 & 0.0415 \\ 0.0415 & 0.25 \end{pmatrix}$ and $V_3 = V_4 = \begin{pmatrix} 0.15 & -0.0415 \\ -0.0415 & 0.25 \end{pmatrix}$. Once the 100 000 pairs (X_{t1}, X_{t2}) are generated, we compute the sample lag 1 serial autocorrelations of 0.66 for X_{t1} and 0.45 for X_{t2}. The sample correlation between the simulated X_{t1} and X_{t2} is −0.31.

The LBFP estimator was then calculated for these simulated data on a 50×50 grid on the $(-3, 3)^2$ square with the bin widths $b_1 = 0.075$ and $b_2 = 0.15$. These values of b_1 and b_2 minimize the ISE, which is estimated at 3.73×10^{-5} when taking the mean of the squared errors over the grid. The size of the grid was chosen to have a good graphical view of the true density and the LBFP. Examination of the graphs depicted in Fig. 1 clearly shows that the LBFP estimator approximates the true simulated density very well when the sample size is large.

3.2 Real data

We have 4389 daily readings of river Ashuapmushuan's flow (in m^3/s) and level (in m), between 22 March 1992 and 30 September 2007 (with a few missing days), a dataset extracted from the Environment and Climate Change Canada Historical Hydrometric Data web site (https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html) on 26 July 2023.

For the flow rate variable, we have a quasi-exponential marginal law, but if we take its logarithm, it seems to be bimodal, see panels 1 and 4 in Fig. 2.

As for the level variable, it is unimodal but left asymmetrical, see panel 2 in Fig. 2.

The correlation between the two variables flow and level is 0.69, as seen on panel 3 of Fig. 2.

We chose $b_1 = 0.19$ and $b_2 = 0.15$ as bandwidths; as in the previous example, we took $b_i = s_i \times n^{-1/6}$ where s_i is the estimated standard deviation in the direction i.

Finally, we plotted the LBFP estimator on a 50×50 grid on the rectangle $(3.7, 6.5) \times (29.4, 32.25)$ for readability (see Fig. 3), which gives an idea of the shape of the underlying density of the log flow and the level.

4 Preliminary lemmas

Lemma 1 *Suppose 1 holds. Let* $\mathcal{L}_r(\mathcal{F})$ *denote the class of* $\mathcal{F}$*-measurable r.v.'s* U *satisfying* $\|U\|_r = (E|U|^r)^{1/r} < \infty$. *Let* $U \in \mathcal{L}_r(\mathcal{B}(S))$ *and* $V \in \mathcal{L}_r(\mathcal{B}(S'))$. *If* r, s *and* τ *are positive numbers and* $r^{-1} + s^{-1} + \tau^{-1} = 1$, *then*

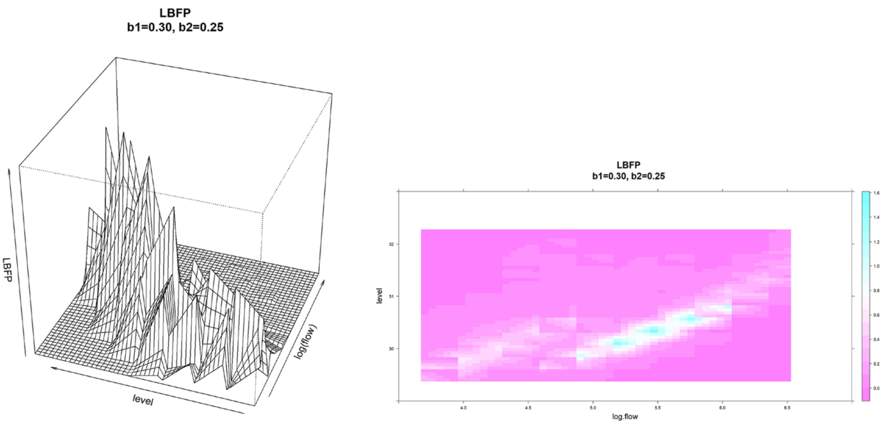

Fig. 3 Perspective (left) and contour plot (right) of the LBFP estimator of the joint density of log(flow) and level on a 50×50 grid

$$|EUV - EUEV| \leq C\|U\|_r\|V\|_s\left[h\{\text{Card}(S), \text{Card}(S')\}\varphi\{\text{ dist } (S, S')\}\right]^{1/\tau}. \quad (15)$$

One or both of r and s can be taken to be ∞ for bounded random variables, and then the right-hand side of (15) can be replaced by $Ch\{\text{Card}(S), \text{Card}(S')\}\varphi\{\text{ dist } (S, S')\}$. For the proof of different versions of the Davydov inequality in Lemma 1, see Davydov (1970), Deo (1973), Hall and Heyde (1980) or Tran (1990).

Using (7), denote $\eta_{\mathbf{i},k,j} = Y_{\mathbf{i},k,j} - EY_{\mathbf{i},k,j}$.

Corollary 1 *For each* $k = \{k_1, \dots, k_d\}$, $j = (j_1, \dots, j_d) \in \{0, 1\}^d$ *and* $j' = (j'_1, \dots, j'_d) \in \{0, 1\}^d$, *there exists some* $\xi_{k,j} \in I(k, j)$ *such that*

$$\begin{aligned} &(i) \ \left|\text{cov}\left(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}\right)\right| \leq C\left(f(\xi_{k,j})b_1 \dots b_d\right)^{1/2}\left(\varphi\left(\|\mathbf{i}-\mathbf{i}'\|\right)\right)^{1/2}, \\ &(ii) \ \left|\text{cov}\left(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j'}\right)\right| \leq C\left(\min\{f(\xi_{k,j}), f(\xi_{k,j'})\}b_1 \dots b_d\right)^{1/2}\left(\varphi\left(\|\mathbf{i}-\mathbf{i}'\|\right)\right)^{1/2}, \ \text{for } j \neq j'. \end{aligned} \quad (16)$$

Proof (i) Taking $r = 2$, $s = \infty$ and $\tau = 2$, Lemma 1 leads to the following result: if $E|U|^2 < +\infty$ and $P|V| > 1) = 0$, then

$$|EUV - EU\,EV| < C\|U\|_2\left(\varphi\left(\|\mathbf{i} - \mathbf{i}'\|\right)\right)^{1/2}.$$

Taking $U = Y_{\mathbf{i},k,j}$ and $V = Y_{\mathbf{i}',k,j}$, and using the generalized mean-value theorem we have

$$\|U\|_2 = \left(EY^2_{\mathbf{i},k,j}\right)^{1/2} = \left[P\{X_{\mathbf{i}} \in I(k,j)\}\right]^{1/2} = \left\{f(\xi_{k,j})b_1 \dots b_d\right\}^{1/2}, \text{ where } \xi_{k,j} \in I(k,j).$$

The proof of (i) thus follows.

(ii) The proof can be handled in the same way.

Note that $\xi_{k,j}$ and $\xi_{k,j'}$ are independent of $\mathbf{i}$ and $\mathbf{i}'$. □

Lemma 2 *If Assumption* 1 *is satisfied, then there exists* $\zeta_{k,j} \in I(k, j)$ *such that*

$$\int\int_{I(k,j)\times I(k,j)} \left|f_{\mathbf{i},\mathbf{j}}(x, y) - f(x)f(y)\right| \, dx\, dy \leq Mf(\zeta_{k,j})\left(b_1 \dots b_d\right)^2. \quad (17)$$

Proof Since f is uniformly continuous and integrable,

$$\sup_{x\in\mathbb{R}^d} f(x) \equiv \|f\| < \infty.$$

By Assumption 1,

$$\begin{aligned}
&\int\int_{I(k,j)\times I(k,j)} \left|f_{\mathbf{i},\mathbf{j}}(x,y)-f(x)f(y)\right| \, dx\,dy\\
&=\int\int_{I(k,j)\times I(k,j)} f(x)\left|f_{\mathbf{j}|\mathbf{i}}(y|x)-f(y)\right| \, dx\,dy\\
&\leq Mb_1\ldots b_d\int_{I(k,j)} f(x)dx,
\end{aligned}$$

where M can be taken to be $\max\{M_1, \|f\|\}$. The lemma follows by the generalized mean-value theorem. □

Now, we study the integrated mean squared error and the optimal cell width. As usual, the IMSE is defined as the sum of two terms: the integrated pointwise squared bias, and the integrated pointwise variance.

For convenience, define

$$r_{1\mathbf{n}} = \frac{1}{\left(\hat{\mathbf{n}}b_1\ldots b_d\right)^2}\sum_{\mathbf{i}\neq\mathbf{i}'}\operatorname{cov}\left(Y_{\mathbf{i},k,j},\, Y_{\mathbf{i}',k,j}\right),$$

and, for $j \neq j'$,

$$r_{2\mathbf{n}} = \frac{1}{\left(\hat{\mathbf{n}}b_1\ldots b_d\right)^2}\sum_{\mathbf{i}\neq\mathbf{i}'}\operatorname{cov}\left(Y_{\mathbf{i},k,j},\, Y_{\mathbf{i}',k,j'}\right).$$

Lemma 3 *The variance of the LBFP estimator $\hat{f}_{\mathbf{n}}(x)$ defined in (6) is given by*

$$\begin{aligned}
\operatorname{var}\hat{f}_{\mathbf{n}}(x) \;=\; &\frac{1}{\hat{\mathbf{n}}\left(b_1\ldots b_d\right)^2}\sum_{j\in\{0,1\}^d} c_j^2(x)p_k(j) - \frac{1}{\hat{\mathbf{n}}\left(b_1\ldots b_d\right)^2}\left(\sum_{j\in\{0,1\}^d} c_j(x)p_k(j)\right)^2\\
&+\sum_{j\in\{0,1\}^d} c_j^2(x)\, r_{1\mathbf{n}} + \sum_{j\neq j'} c_j(x)c_{j'}(x)\, r_{2\mathbf{n}}.
\end{aligned}$$

Proof Clearly,

$$\begin{aligned}
\operatorname{var}\hat{f}_{\mathbf{n}}(x) \;=\; &\frac{1}{\left(\hat{\mathbf{n}}b_1\ldots b_d\right)^2}\sum_{j\in\{0,1\}^d}\left\{c_j(x)\right\}^2\operatorname{var}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k,j}\right)\\
&+\frac{1}{\left(\hat{\mathbf{n}}b_1\ldots b_d\right)^2}\sum_{j\neq j'} c_j(x)c_{j'}(x)\operatorname{cov}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k,j},\, \sum_{\mathbf{i}'\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}',k,j'}\right).
\end{aligned}\tag{18}$$

First of all, we have

$$\operatorname{var}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k,j}\right) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}\operatorname{var}\left(Y_{\mathbf{i},k,j}\right) + \sum_{\mathbf{i}\neq\mathbf{i}'}\operatorname{cov}\left(Y_{\mathbf{i},k,j},\, Y_{\mathbf{i}',k,j}\right),$$

with

$$\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \operatorname{var}\left(Y_{\mathbf{i},k,j}\right) = \hat{\mathbf{n}} p_k(j)\left\{1 - p_k(j)\right\}.$$

Therefore the first term of the right member of (18) is

$$\frac{1}{\hat{\mathbf{n}}\left(b_1 \dots b_d\right)^2} \sum_{j\in\{0,1\}^d} \left\{c_j(x)\right\}^2 p_k(j)\left\{1 - p_k(j)\right\} + \sum_{j\in\{0,1\}^d} \left\{c_j(x)\right\}^2 r_{1\mathbf{n}}.$$

Now, for $j \neq j'$, we have

$$\operatorname{cov}\left(\sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k,j}, \sum_{\mathbf{i}'\in\mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}',k,j'}\right) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \operatorname{cov}\left(Y_{\mathbf{i},k,j}, Y_{\mathbf{i},k,j'}\right) + \sum_{\mathbf{i}\neq\mathbf{i}'} \operatorname{cov}\left(Y_{\mathbf{i},k,j}, Y_{\mathbf{i}',k,j'}\right).$$

For $j \neq j'$, we have

$$\operatorname{cov}\left(Y_{\mathbf{i},k,j}, Y_{\mathbf{i},k,j'}\right) = E\left(Y_{\mathbf{i},k,j} Y_{\mathbf{i},k,j'}\right) - E\left(Y_{\mathbf{i},k,j}\right) E\left(Y_{\mathbf{i},k,j'}\right) = -p_k(j) p_k(j').$$

Then the second term of the right member of (18) is

$$-\frac{1}{\left(\hat{\mathbf{n}} b_1 \dots b_d\right)^2} \sum_{j\neq j'} c_j(x) c_{j'}(x) \frac{1}{\hat{\mathbf{n}}\left(b_1 \dots b_d\right)^2} p_k(j) p_k(j') + \sum_{j\neq j'} c_j(x) c_{j'}(x)\, r_{2\mathbf{n}}.$$

Finally, we obtain

$$\begin{aligned} \operatorname{var} \hat{f}_{\mathbf{n}}(x) &= \frac{1}{\hat{\mathbf{n}}\left(b_1 \dots b_d\right)^2} \sum_{j\in\{0,1\}^d} c_j^2(x) p_k(j) - \frac{1}{\hat{\mathbf{n}}\left(b_1 \dots b_d\right)^2} \left(\sum_{j\in\{0,1\}^d} c_j(x) p_k(j)\right)^2 \\ &\quad + \sum_{j\in\{0,1\}^d} c_j^2(x)\, r_{1\mathbf{n}} + \sum_{j\neq j'} c_j(x) c_{j'}(x)\, r_{2\mathbf{n}}. \end{aligned}$$

□

Let

$$w_{k,j} = \max\left\{(f(\xi_{k,j}))^{1/2}, Mf(\zeta_{k,j})\right\}$$

where $\xi_{k,j}$ and $\zeta_{k,j}$ are defined in Corollary 1 and Lemma 2.

Lemma 4 *Assume that* $\varphi(k) = O(k^{-\rho})$ *for some* $\rho > 2N + (3/2)$. *Let* $0 < \varepsilon \leq (2N-1)(8N-1)^{-1}$. *Then*

$$|r_{1\mathbf{n}}| \leq C\hat{\mathbf{n}}^{-1}(b_1 \dots b_d)^{-1+\varepsilon} w_{k,j}.$$

Proof By Corollary 1 and Lemma 2

$$|\operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j})| \leq \min\{C(f(\xi_{k,j}))^{1/2}(b_1 \dots b_d)^{1/2}(\varphi(\|\mathbf{i}-\mathbf{i}'\|))^{1/2}, Mf(\zeta_{k,j})b^2\}.$$

Let $K_{\mathbf{n}} = (b_1 \ldots b_d)^{(\varepsilon-1)/N}$. Define

$$\begin{aligned} S_1 &= \{\mathbf{i}, \mathbf{i}' \in \mathcal{I}_{\mathbf{n}} \quad | \quad 0 < \|\mathbf{i} - \mathbf{i}'\| \le K_{\mathbf{n}}\} \\ S_2 &= \{\mathbf{i}, \mathbf{i}' \in \mathcal{I}_{\mathbf{n}} \quad | \quad \|\mathbf{i} - \mathbf{i}'\| > K_{\mathbf{n}}\}, \end{aligned}$$

split $\sum_{\mathbf{i} \neq \mathbf{i}'} \left| \operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}) \right|$ into two separate summations A_1 and A_2 over sites S_1 and S_2. Then

$$\sum_{\mathbf{i} \neq \mathbf{i}'} \left| \operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}) \right| = A_1 + A_2.$$

Now, we have the following upper bounds.

$$A_1 = \sum_{\mathbf{i},\mathbf{i}' \in S_1} \left| \operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}) \right| \le Mf(\zeta_{k,j}) \sum_{\mathbf{i},\mathbf{i}' \in S_1} (b_1 \ldots b_d)^2 \le Cf(\zeta_{k,j})(b_1 \ldots b_d)^2 \hat{\mathbf{n}} K_{\mathbf{n}}^N.$$

Thus

$$A_1 \le Cf(\zeta_{k,j})\, \hat{\mathbf{n}}\, (b_1 \ldots b_d)^{1+\varepsilon}.$$

Let now $\nu = \frac{N}{2} \cdot \frac{1 + 2\varepsilon}{1 - \varepsilon}$. Clearly

$$\begin{aligned} A_2 &= \sum_{\mathbf{i},\mathbf{i}' \in S_2} \left| \operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}) \right| \\ &\le C\, \{f(\xi_{k,j})\}^{1/2} (b_1 \ldots b_d)^{1/2} \sum_{\mathbf{i},\mathbf{i}' \in S_2} \{\varphi(\|\mathbf{i} - \mathbf{j}\|)\}^{1/2} \\ &\le C\, \{f(\xi_{k,j})\}^{1/2} (b_1 \ldots b_d)^{1/2}\, \hat{\mathbf{n}} \sum_{\|\mathbf{i}\| > K_{\mathbf{n}}} \{\varphi(\|\mathbf{i}\|)\}^{1/2} \\ &\le C\, \{f(\xi_{k,j})\}^{1/2} (b_1 \ldots b_d)^{1/2}\, \hat{\mathbf{n}}\, K_{\mathbf{n}}^{-\nu} \sum_{\|\mathbf{i}\| > K_{\mathbf{n}}} \|\mathbf{i}\|^{\nu} \{\varphi(\|\mathbf{i}\|)\}^{1/2}. \end{aligned}$$

Since $0 < \varepsilon \le (2N - 1)(8N - 1)^{-1}$, we have $\nu \le N - \frac{1}{4}$. Thus

$$\sum_{\|\mathbf{i}\| > K_{\mathbf{n}}} \|\mathbf{i}\|^{\nu} \{\varphi(\|\mathbf{i}\|)\}^{1/2} \le \left(\sum_{i=1}^{\infty} i^{-(2\rho - 4N + 1)/4} \right)^N < +\infty$$

since $\rho > 2N + 3/2$. Finally

$$A_2 \le C\, \{f(\xi_{k,j})\}^{1/2} (b_1 \ldots b_d)^{1/2}\, \hat{\mathbf{n}}\, K_{\mathbf{n}}^{-\nu} \le C\, \{f(\xi_{k,j})\}^{1/2}\, \hat{\mathbf{n}} (b_1 \ldots b_d)^{1+\varepsilon}$$

and

$$\sum_{\mathbf{i}\neq\mathbf{i}'} \left| \operatorname{cov}(\eta_{\mathbf{i},k,j}, \eta_{\mathbf{i}',k,j}) \right| \leq C\, w_{k,j}\, \hat{\mathbf{n}}\, (b_1 \dots b_d)^{1+\varepsilon}.$$

Thus $r_{1\mathbf{n}}$ is bounded by $C\, w_{k,j}\, \hat{\mathbf{n}}^{-1}\, b^{-1+\varepsilon}$. Using Corollary 1 (ii) and a slight variation of Lemma 2 it can be shown that $r_{2\mathbf{n}}$ is also bounded by the same quantity. □

We'll need the following lemma which is a multidimensional generalization of results reached by Friedman and Diaconis (1981).

Lemma 5 *Suppose that $\phi : \mathbb{R}^d \to \mathbb{R}$ is continuous and its first order partial derivatives $\phi_1', \dots, \phi_d'$ are continuous and integrable. Then*

$$\sum_k \phi(\xi_k) b_1 \dots b_d = \int_{\mathbb{R}^d} \phi(x)\, dx + O\left(\sum_{s=1}^{d} b_s \int_{\mathbb{R}^d} \left| \phi_s'(x) \right| dx \right)$$

where the sum is of over all cells, the union of which is $\mathbb{R}^d$, each cell has volume $b_1 \dots b_d$ and ξ_k is an arbitrary point in cell number k.

After some algebraic efforts and using Lemma 5, we obtain the following lemma.

Lemma 6 *Suppose Assumptions* 1 *and* 2 *are satisfied, and suppose that $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$. Let ε be defined as in Lemma 4. Then*

$$\left| \int_{\mathbb{R}^d} \operatorname{var} \hat{f}_{\mathbf{n}}(x)\, dx - \frac{1}{\hat{\mathbf{n}}(b_1 \dots b_d)} \left(\frac{2}{3}\right)^d \right| \leq \frac{1}{\hat{\mathbf{n}}} \int_{\mathbb{R}^d} f^2(x)\, dx$$

$$+ O\left(\hat{\mathbf{n}}^{-1} (b_1 \dots b_d)^{-1} \sum_{s=1}^{d} b_s \|f_s'\|_1\, dx \right) + O\left(\hat{\mathbf{n}}^{-1} (b_1 \dots b_d)^{-1} \sum_{s=1}^{d} b_s^2 \right)$$

$$+ O\left(\hat{\mathbf{n}}^{-1} (b_1 \dots b_d)^{-1} \sum_{s=1}^{d} b_s \left\| (f^2)' \right\|_1 \right) + O\left(\hat{\mathbf{n}}^{-1} (b_1 \dots b_d)^{-1+\varepsilon} \left[\left\| f^{1/2} \right\|_1 \right] \right)$$

$$+ O\left\{ \hat{\mathbf{n}}^{-1} (b_1 \dots b_d)^{-1+\varepsilon} \left[\left\| f^{1/2} \right\|_1 + O\left(\sum_{s=1}^{d} b_s \left\| (f^{1/2})' \right\|_1 \right) + \|f\|_1 + O\left(\sum_{s=1}^{d} b_s \|f'\|_1 \right) \right] \right\}.$$

Let $bias(x) = E\big(\hat{f}_{\mathbf{n}}(x)\big) - f(x)$. In the following lemma we obtain a result for the integrated squared bias that is similar to the result in Hjort (1986).

Lemma 7 *If Assumptions* 1 *and* 2 *hold, then*

$$\int_{\mathbb{R}^d} \{bias(x)\}^2\, dx = \sum_{s=1}^{d} \frac{49}{2880} b_s^4 \int_{\mathbb{R}^d} \left\{ f_{s,s}''(x) \right\}^2 dx$$
$$+ \sum_{1 \leq s < t \leq d}^{d} \frac{1}{32} b_s^2 b_t^2 \int_{\mathbb{R}^d} f_{s,s}''(x) f_{t,t}''(x)\, dx + O\left(\sum_{s=1}^{d} b_s^5 \right).$$

To obtain the uniform convergence of the LBFP estimator, we need to define some random variables linked to the LBFP estimator defined in (6). Let D be an arbitrary compact set in $\mathbb{R}^d$. For $x \in D$, let

$$S_{\mathbf{n}}(x) = \hat{f}_{\mathbf{n}}(x) - E\hat{f}_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}}(x),$$

with $\Delta_{\mathbf{i}}(x) = (\hat{\mathbf{n}} b_1 \dots b_d)^{-1} \sum_{j\in\{0,1\}^d} c_j(x)\eta_{\mathbf{i},k,j}$. Also define

$$I_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} E(\Delta_{\mathbf{i}}(x))^2 \quad \text{and} \quad R_{\mathbf{n}}(x) = \sum_{\mathbf{j}\in\mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} |\mathrm{Cov}\{\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x)\}|.$$

$$_{i_k \neq j_k} \text{ for some } k$$

Lemma 8 *If $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$, then*

$$\lim_{\mathbf{n}\to+\infty} \hat{\mathbf{n}} b_1 \dots b_d \left\{ I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x) \right\} < C,$$

where C is a constant independent of x.

Proof Lemma 8 follows by a careful analysis of the proof of Lemmas 3-4. □

The following lemma of Rio will be needed in the sequel. Its proof is found in Rio (1995) (see Theorem 4).

Lemma 9 *Suppose $\mathcal{A}$ is a σ-field of $(\Omega, \mathcal{F}, P)$ and X is a real-valued random variable taking almost surely its values in $[a, b]$. Suppose furthermore that there exists a random variable U with uniform distribution over $[0, 1]$, independent of $\mathcal{A} \vee \sigma(X)$. Then there exists some random variable X^* independent of $\mathcal{A}$ and with the same distribution as X such that*

$$E|X - X^*| \leq 2(b - a)\alpha\{\mathcal{A}, \sigma(X)\}.$$

Moreover, X^* is a $\mathcal{A} \vee \sigma(X) \vee \sigma(U)$-measurable random variable.

The approximation of strongly mixing r.v.'s by independent ones used later is presented below.

Lemma 10 *Let $S_1,\dots,S_r$ be sets containing m sites each with $\mathrm{dist}(S_i, S_j) \geq \delta$ for all $i \neq j$ where $1 \leq i \leq r$ and $1 \leq j \leq r$. Suppose that $Y_1,\dots,Y_r$ is a sequence of real-valued r.v.'s measurable with respect to $\mathcal{B}(S_1),\dots,\mathcal{B}(S_r)$, respectively, and that Y_i takes values in $[a, b]$. Then there exists a sequence of independent r.v.'s $Y_1^*,\dots,Y_r^*$ independent of $Y_1,\dots,Y_r$ such that Y_i^* has the same distribution as Y_i and satisfies*

$$\sum_{i=1}^{r} E|Y_i - Y_i^*| \leq 2r(b-a)h\{(r-1)m, m\}\varphi(\delta). \tag{19}$$

For a proof, see Carbon et al. (1997b).

Recall that

$$S_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}} \Delta_{\mathbf{i}}(x) = \sum_{\substack{i_k=1 \\ k=1,\ldots,N}}^{n_k} \Delta_{\mathbf{i}}(x).$$

Without loss of generality assume that $n_i = 2pq_i$ for $1 \leq i \leq N$. The random variables $\Delta_{\mathbf{i}}(x)$ can be grouped into $2^N q_1 \times \cdots \times q_N$ cubic blocks of side p.

Denote

$$U(1, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1 \\ k=1,\ldots,N}}^{(2j_k+1)p} \Delta_{\mathbf{i}}(x),$$

$$U(2, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1 \\ k=1,\ldots,N-1}}^{(2j_k+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{i}}(x),$$

$$U(3, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1 \\ k=1,\ldots,N-2}}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{i}}(x),$$

$$U(4, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1 \\ k=1,\ldots,N-2}}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{i}}(x),$$

and so on. Note that

$$U(2^{N-1}, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=(2j_k+1)p+1 \\ k=1,\ldots,N-1}}^{2(j_k+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{i}}(x).$$

Finally

$$U(2^N, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=(2j_k+1)p+1 \\ k=1,\ldots,N}}^{2(j_k+1)p} \Delta_{\mathbf{i}}(x).$$

For each integer $1 \leq i \leq 2^N$, define

$$T(\mathbf{n}, i, x) = \sum_{\substack{j_{k=0} \\ k=1,\ldots,N}}^{q_k-1} U(i, \mathbf{n}, \mathbf{j}, x).$$

Clearly

$$S_{\mathbf{n}}(x) = \sum_{i=1}^{2^N} T(\mathbf{n}, i, x). \tag{20}$$

The blocking idea here is reminiscent of the blocking scheme in Tran (1990) and Politis and Romano (1993).

Without loss of generality let us focus on the terms with $i = 1$. Now, $T(\mathbf{n}, 1, x)$ is the sum of

$$r = q_1 \times \cdots \times q_N$$

of the $U(1, \mathbf{n}, \mathbf{j}, x)$'s. Note that $U(1, \mathbf{n}, \mathbf{j}, x)$ is measurable with respect to the σ-field generated by $X_{\mathbf{i}}$, with $\mathbf{i}$ belonging to the set of sites

$$\{\mathbf{i} : 2j_k p + 1 \le i_k \le (2j_k + 1)p, \quad k = 1, ..., N\}.$$

For each j_k, these sets of sites are separated by a distance of at least p. Enumerate the r.v.'s $U(1, \mathbf{n}, \mathbf{j}, x)$ and the corresponding σ-fields with respect to which they are measurable in an arbitrary manner and refer to them respectively as $Y_1, ..., Y_r$ and $S_1, ..., S_r$. Approximate $Y_1, ..., Y_r$ by the r.v.'s $Y_1^*, ..., Y_r^*$ as was done in Lemma 10. Clearly,

$$|Y_i| < Cp^N(\hat{\mathbf{n}} b_1 \ldots b_d)^{-1}. \tag{21}$$

Denote

$$\tilde{\Psi}_{\mathbf{n}} = \left\{\log \hat{\mathbf{n}}(\hat{\mathbf{n}} b_1 \ldots b_d)^{-1}\right\}^{1/2}, \quad \varepsilon_{\mathbf{n}} = \eta \tilde{\Psi}_{\mathbf{n}},$$

where η is a positive constant to be chosen later. Define

$$\alpha_{\mathbf{n}} = (b_1 \ldots b_d)^{-1}\, h(\hat{\mathbf{n}}, p^N)\varphi(p)\tilde{\Psi}_{\mathbf{n}}^{-1}. \tag{22}$$

Lemma 11 *Given $x \in D$ and an arbitrarily large positive constant M, there exist positive constants C and η such that*

$$P\left[|T(\mathbf{n}, 1, x)| > \eta \tilde{\Psi}_{\mathbf{n}}\right] \le C\left(\hat{\mathbf{n}}^{-M} + \alpha_{\mathbf{n}}\right).$$

Proof Since $T(\mathbf{n}, 1, x)$ is equal to $\sum_{i=1}^{r} Y_i$, we have

$$P[|T(\mathbf{n},1,x)| > \varepsilon_{\mathbf{n}}] \leq P\left[\left|\sum_{i=1}^{r} Y_i^*\right| > \varepsilon_{\mathbf{n}}/2\right] + P\Big[\sum_{i=1}^{r} |Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}/2\Big]. \quad (23)$$

We now obtain bounds for the two terms on the right hand side of (23). By Markov's inequality and using (19–21) and recalling that the sets of sites with respect to which the Y_i's are measurable are separated by a distance of at least p,

$$P\left[\sum_{i=1}^{r} |Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}\right] \leq Crp^N(\hat{\mathbf{n}}b_1 \dots b_d)^{-1} h(\hat{\mathbf{n}}, p^N)\varphi(p)\varepsilon_{\mathbf{n}}^{-1} \sim \alpha_{\mathbf{n}}. \quad (24)$$

Setting

$$\lambda_{\mathbf{n}} = (\hat{\mathbf{n}}b_1 \dots b_d \log \hat{\mathbf{n}})^{1/2},$$

and

$$p = \left[\left(\frac{\hat{\mathbf{n}}b_1 \dots b_d}{4\lambda_{\mathbf{n}}}\right)^{1/N}\right] \sim \left(\frac{\hat{\mathbf{n}}b_1 \dots b_d}{\log \hat{\mathbf{n}}}\right)^{\frac{1}{2N}}, \quad (25)$$

we have

$$\lambda_{\mathbf{n}}\varepsilon_{\mathbf{n}} = \eta \log \hat{\mathbf{n}},$$

and by Lemma 8

$$\lambda_{\mathbf{n}}^2 \sum_{i=0}^{r} E(Y_i^*)^2 \leq C\hat{\mathbf{n}}b_1 \dots b_d \big\{I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)\big\} \log \hat{\mathbf{n}} < C \log \hat{\mathbf{n}}.$$

Using (21), we have $|\lambda_{\mathbf{n}} Y_i^*| < 1/2$ for large $\hat{\mathbf{n}}$. Applying the standard Bernstein's inequality,

$$P\left[\left|\sum_{i=0}^{r} Y_i^*\right| > \varepsilon_{\mathbf{n}}\right] \leq 2\exp\Big(-\lambda_{\mathbf{n}}\varepsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2 \sum_{i=0}^{r} E(Y_i^*)^2\Big) \quad (26)$$

$$\leq 2\exp\big\{(-\eta + C)\log \hat{\mathbf{n}}\big\} \leq \hat{\mathbf{n}}^{-M},$$

for sufficiently large $\hat{\mathbf{n}}$.

The conclusion then follows from (23), (24) and (26). □

Lemma 12 *Suppose Assumption 3 holds. Let D be a compact set of $\mathbb{R}^d$ and let $(b_1, \dots, b_d) = (b_{1,\mathbf{n}}, \dots, b_{d,\mathbf{n}})$ be a sequence of cell widths such that*

$$\tilde{\Psi}_{\mathbf{n}} := \big\{\log \hat{\mathbf{n}}(\hat{\mathbf{n}}b_1 \dots b_d)^{-1}\big\}^{1/2} \to 0 \qquad \text{as} \qquad \mathbf{n} \to \infty. \quad (27)$$

If there exists a constant M such that

$$(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\left(\hat{\mathbf{n}}^{-M} + \alpha_{\mathbf{n}}\right) = o(1) \quad \text{(respectively}$$
$$\sum_{\mathbf{n}\in Z^N}(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\left(\hat{\mathbf{n}}^{-M} + \alpha_{\mathbf{n}}\right) < \infty)$$

then we have

$$\sup_{x\in D}|\hat{f}_{\mathbf{n}}(x) - E\hat{f}_{\mathbf{n}}(x)| = O\left(\tilde{\Psi}_{\mathbf{n}}\right) \quad \text{in probability} \quad \text{(respectively a.s.).}$$

Proof The compact set D can be covered with n_D cubic cells I_k of volume $v_D = w_D^d$ and center x_k. Choose the cells such that $v_D < b_1 \dots b_d$ and $n_D \leq C/v_D$. We have

$$\sup_{x\in D}|\hat{f}_{\mathbf{n}}(x) - E\hat{f}_{\mathbf{n}}(x)| \leq c_1 + c_2 + c_3$$

where

$$c_1 = \max_{1\leq k\leq n_D}\sup_{x\in I_k}|\hat{f}_{\mathbf{n}}(x) - \hat{f}_{\mathbf{n}}(x_k)|,$$
$$c_2 = \max_{1\leq k\leq n_D}|\hat{f}_{\mathbf{n}}(x_k) - E\hat{f}_{\mathbf{n}}(x_k)|,$$
$$c_3 = \max_{1\leq k\leq n_D}\sup_{x\in I_k}|E\hat{f}_{\mathbf{n}}(x_k) - E\hat{f}_{\mathbf{n}}(x)|.$$

By (6), after some calculations, we have, for $x, x_k \in I_k$,

$$|\hat{f}_{\mathbf{n}}(x) - \hat{f}_{\mathbf{n}}(x_k)| \leq \frac{C}{\hat{\mathbf{n}}(b_1 \dots b_d)^2}\left(\sum_{j_1,\dots,j_d,j'_1,\dots,j'_d\in\{0,1\}^{2d}}\left|v_{k_1+j_1,\dots,k_d+j_d} - v_{k_1+j'_1,\dots,k_d+j'_d}\right|\right)\|x - x_k\|, \tag{28}$$

and

$$|E\hat{f}_{\mathbf{n}}(x) - E\hat{f}_{\mathbf{n}}(x_k)| \leq \frac{C}{(b_1 \dots b_d)^2}\left(\sum_{j_1,\dots,j_d,j'_1,\dots,j'_d\in\{0,1\}^{2d}}\left|p_k(j_1,\dots,j_d) - p_k(j'_1,\dots,j'_d)\right|\right)\|x - x_k\|. \tag{29}$$

The generalized mean-value theorem shows that there exist $\zeta_{k;j_1,\dots,j_d}$ and $\zeta_{k;j'_1,\dots,j'_d}$ such that $p_k(j_1,\dots,j_d)/(b_1 \dots b_d) = f(\zeta_{k;j_1,\dots,j_d})$ and $p_k(j'_1,\dots,j'_d)/(b_1 \dots b_d) = f(\zeta_{k;j'_1,\dots,j'_d})$. It follows that, under Assumption 2*,

$$|E\hat{f}_{\mathbf{n}}(x) - E\hat{f}_{\mathbf{n}}(x_k)| \leq \frac{C}{b_1 \dots b_d}\left(\sum_{j_1,\dots,j_d,j'_1,\dots,j'_d\in\{0,1\}^{2d}}\left|f(\zeta_{k;j_1,\dots,j_d}) - f(\zeta_{k;j'_1,\dots,j'_d})\right|\right)\|x - x_k\|$$
$$\leq \frac{C}{b_1 \dots b_d}\|x - x_k\|.$$

In view of (28)

$$|\hat{f}_n(x) - \hat{f}_n(x_k)| \leq 2(b_1 \ldots b_d)^{-2}\|x_k - x\| \quad \text{a.s.},$$

and $c_1 = O\{(b_1 \ldots b_d)b^{-2}w_D\}$ a.s. Choosing $w_D = (b_1 \ldots b_d)^2\tilde{\Psi}_{\mathbf{n}} < b_1 \ldots b_d$ implies $v_D < b_1 \ldots b_d$ and we then obtain $c_1 = O(\tilde{\Psi}_{\mathbf{n}})$ a.s. and $c_3 = O(\tilde{\Psi}_{\mathbf{n}})$.

In view of (20),

$$c_2 = \max_{1\leq k\leq n_D} |\hat{f}_{\mathbf{n}}(x_k) - E\hat{f}_{\mathbf{n}}(x_k)| = \max_{1\leq k\leq n_D} |S_{\mathbf{n}}(x_k)| = O(\tilde{\Psi}_{\mathbf{n}}) \quad \text{in probability (or a.s.)}$$

if

$$\max_{1\leq k\leq n_D} |T(\mathbf{n}, i, x_k)| = O(\tilde{\Psi}_{\mathbf{n}}) \quad \text{in probability (or a.s.)}$$

for each $1 \leq i \leq 2^N$. Lemma 11 entails

$$P\left\{\max_{1\leq k\leq n_D} |T(\mathbf{n}, i, x_k)| > \eta\tilde{\Psi}_{\mathbf{n}}\right\} \leq Cn_D\left(\hat{\mathbf{n}}^{-M} + \alpha_{\mathbf{n}}\right)$$

for $i = 1$, and it is easy to obtain the same result for all $i = 2, \ldots, 2^N$. Using the Borel-Cantelli lemma, the conclusion follows. □

Lemma 13 *Assume that Assumption* 3 *holds. Then*

$$\left|E\hat{f}_{\mathbf{n}}(x) - f(x)\right| = O(\Psi_{\mathbf{n}}).$$

Proof For $x \in I(k)$, as $\sum_{j_1,\ldots,j_d\in\{0,1\}^d} c_{j_1,\ldots,j_d}(x) = 1$, we have

$$E\hat{f}_{\mathbf{n}}(x) - f(x) = \sum_{j_1,\ldots,j_d\in\{0,1\}^d} c_{j_1,\ldots,j_d}(x)\left(\frac{p_k(j)}{b_1 \ldots b_d} - f(x)\right).$$

Under Assumption 3, we have

$$\begin{aligned}\left|\frac{p_k(j)}{b_1 \ldots b_d} - f(x)\right| &\leq \frac{1}{b_1 \ldots b_d}\int_{I(k,j)} |f(y) - f(x)|\, dy \\ &\leq \frac{C}{b_1 \ldots b_d}\int_{I(k,j)} \|y - x\|\, dy \leq C\tilde{b}_{\mathbf{n}}.\end{aligned}$$

Then recalling the definition of $\Psi_{\mathbf{n}}$ concludes the proof. □

5 Proofs

Proof of Theorem 1 The IMSE is found by combining the results of Lemmas 6 and 7. The leading terms of the IMSE are found to be

$$\frac{1}{\hat{\mathbf{n}}(b_1 \dots b_d)}\left(\frac{2}{3}\right)^d + \sum_{s=1}^{d} \frac{49}{2880} b_s^4 \int_{\mathbb{R}^d} \left(f''_{s,s}(x)\right)^2 dx$$
$$+ \sum_{1 \le s < t \le d} \frac{1}{32} b_s^2 b_t^2 \int_{\mathbb{R}^d} f''_{s,s}(x) f''_{t,t}(x)\, dx.$$

It is thus sufficient to minimize this function with respect to $(b_1, \dots, b_d)$. □

Proof of Theorem 2 In view of Lemma 12, it suffices to show that $(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\hat{\mathbf{n}}^{-M}$ and $(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}}$ converge to zero. First note that (8) implies (27). Now (27) implies $\hat{\mathbf{n}} > C(b_1 \dots b_d)^{-1}$. Therefore

$$(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\hat{\mathbf{n}}^{-M} = \hat{\mathbf{n}}^{-M+1/2}(b_1 \dots b_d)^{-3/2} \log \hat{\mathbf{n}}^{-1/2} < C(b_1 \dots b_d)^{M-2} \log \hat{\mathbf{n}}^{-1/2} \to 0$$

when $M > 2$. From (22) we have

$$(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}} \le (b_1 \dots b_d)^{-3}\, h(\hat{\mathbf{n}}, p^N)p^{-\rho}\tilde{\Psi}_{\mathbf{n}}^{-2}.$$

(i) By (25), $p \sim \tilde{\Psi}_{\mathbf{n}}^{-1/N}$. Thus, if (2) is satisfied

$$(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}} \le (b_1 \dots b_d)^{-3}\left(\frac{\hat{\mathbf{n}} b_1 \dots b_d}{\log \hat{\mathbf{n}}}\right)^{\frac{3}{2}-\frac{\rho}{2N}} = \left\{\frac{\hat{\mathbf{n}}(b_1 \dots b_d)^{1-3\frac{2N}{3N-\rho}}}{\log \hat{\mathbf{n}}}\right\}^{\frac{3N-\rho}{2N}}. \tag{30}$$

When $\rho > 3N$ the right-hand side of (30) converges to zero if and only if (8) holds. When $\rho \le 3N$, the term between round brackets at the center of (30) always tends to infinity and the right-hand side of (30) never converges to zero.

(ii) If (3) is satisfied

$$\begin{aligned}(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}} &\sim C(b_1 \dots b_d)^{-3}\, \hat{\mathbf{n}}^{\tilde{k}}\left(\frac{\hat{\mathbf{n}} b_1 \dots b_d}{\log \hat{\mathbf{n}}}\right)^{1-\frac{\rho}{2N}} \\ &= C\left\{\frac{\hat{\mathbf{n}}(b_1 \dots b_d)^{\frac{\rho+4N}{\rho-2N-2N\tilde{k}}}}{\log \hat{\mathbf{n}}^{\frac{\rho-2N}{\rho-2N-2N\tilde{k}}}}\right\}^{\tilde{k}+1-\frac{\rho}{2N}}.\end{aligned} \tag{31}$$

When $\rho > 2N(\tilde{k}+1)$ and (9) holds, the right-hand side of (31) converges to zero. □

Proof of Theorem 3 (i) Condition (12) implies

$$(b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}}\hat{\mathbf{n}} g(\mathbf{n}) \to 0,$$

which entails

$$\sum_{\mathbf{n} \in \mathbb{Z}^N} (b_1 \dots b_d)^{-2}\tilde{\Psi}_{\mathbf{n}}^{-1}\alpha_{\mathbf{n}} < \infty.$$

The theorem follows easily by Lemmas 12 and 13.

(ii) The proof of (ii) is similar to that of (i) and is omitted. □

Proof of Theorem 4 We use the arguments of Lemma 5.6 in Francq and Tran (2002). More precisely, for any $\varepsilon > 0$ there exists a compact set D (which depends on ε) such that $\int_D f(x)dx \geq 1 - \varepsilon$. Such a compact exists because, by the Lebesgue theorem,

$$\lim_{k\to\infty} \int_{([-k,k]^c)^d} f(x)dx = 0.$$

Denote by $\lambda(D)$ the Lebesgue measure of D. Theorem 3 entails that almost surely

$$\lambda(D) \sup_{x\in D} |\hat{f}_{\mathbf{n}}(x) - f(x)| \leq \varepsilon$$

for sufficiently large $\mathbf{n}$. Thus

$$\int_D \hat{f}_{\mathbf{n}}(x)dx \geq \int_D f(x)dx - \lambda(D) \sup_{x\in D} |\hat{f}_{\mathbf{n}}(x) - f(x)| \geq 1 - 2\varepsilon$$

and

$$\begin{aligned} \int_{x\in\mathbb{R}^d} |\hat{f}_{\mathbf{n}}(x) - f(x)|dx &= \int_D |\hat{f}_{\mathbf{n}}(x) - f(x)|dx + \int_{D^c} |\hat{f}_{\mathbf{n}}(x) - f(x)|dx \\ &\leq \lambda(D) \sup_{x\in D} |\hat{f}_{\mathbf{n}}(x) - f(x)| + \int_{D^c} \hat{f}_{\mathbf{n}}(x)dx + \int_{D^c} f(x)dx \leq 4\varepsilon. \end{aligned}$$

Since ε is arbitrarily small, the conclusion follows. □

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-023-00883-5.

Acknowledgements The authors wish to thank the Editor and two anonymous reviewers whose comments have lead to an improved version of this manuscript.

Data availability statement The dataset analyzed in Section 3.2 is publicly available. It was extracted on 26 July 2023 from the Environment and Climate Change Canada Historical Hydrometric Data web site, https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Biau, G. (2003). Spatial kernel density estimation. *Mathematical Methods of Statistics, 12*, 371–390.
Bolthausen, E. (1982). On the central limit theorem for stationary random fields. *Annals of Probability*, *10*, 1047–1050.
Carbon, M. (2006). Polygone des fréquences pour des champs aléatoires. *Note aux Comptes Rendus de l'Académie des Sciences de Paris, Série I*, *342*, 693–696.
Carbon, M. (2014). Histograms for stationary linear random files. *Statistical Inference for Stochastic Processes 17*, 245–266.
Carbon, M., Hallin, M., Tran, L.T. (1996). Kernel density estimation for random fields: The L_1 theory. *Journal of Non Parametric Statistics*, *6*, 157–170.
Carbon, M., Garel, B., Tran, L.T. (1997a). Frequency polygons for weakly dependent processes. *Statistics and Probability Letters*, *33*, 1–13.
Carbon, M., Tran, L.T., Wu, B. (1997b). Kernel density estimation for random fields (Density estimation for random fields). *Statistics and Probability Letters*, *36*, 115–125.
Davydov, Y.A. (1970). The invariant principle for stationary processes. *Theory of Probability and Its Applications*, *14*, 487–498.
Deo, C.M. (1973). A note on empirical processes of strong mixing sequences. *Annals of Probability*, *1*, 870–875.
Doukhan, P. (1994). *Mixing: properties and examples*. Springer.
Francq, C., Tran, L.T. (2002) Nonparametric estimation of density, regression and dependence coefficients. *Nonparametric Statistics*, *14*, 729–747.
Friedman, D., Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrschenlichtkeitstheory und Verwandte Gebiete*, *67*, 453–476.
Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas Markovien. *Proceedings of the 6th Franco-Belgian Meeting of Statisticians*.
Guyon, X., Richardson, S. (1984). Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants. *Zeitschrift für Wahrschenlichtkeitstheory und Verwandte Gebiete*, *66*, 297–314.
Hall, P., Heyde, C.C. (1980). *Martingale limit theory and its application*. Academic Press.
Hallin, M., Lu, Z., Tran, L.T. (2001). Density estimation for spatial processes. *Bernouilli*, *7*, 657–688.
Hallin, M., Lu, Z., Tran, L.T. (2004a). Local linear spatial regression. *Annals of Statistics*, *32*, 2469–2500.
Hallin, M., Lu, Z., Tran, L.T. (2004b). Kernel density estimation for spatail processes : The L_1 theory. *Journal of Multivariate Analysis*, *88*, 61–75.
Harel, M., Lenain, J.-F., Ngatchou-Wandji, J. (2016). Asymptotic behaviour of binned kernel density estimators for locally non-stationary random fields. *Journal of Nonparametric Statistics*, *28*(2), 296–321.
Hjort, N.L. (1986). On frequency polygons and averaged shifted histograms in higher dimensions. Technical Report, No. 22, Stanford University.
Nahapetian, B.S. (1980). The central limit theorem for random fields with mixing conditions. *Advances in Probability*, *6*, 531–548.
Nahapetian, B.S. (1987). An approach to proving limit theorems for dependent random variables. *Theory of Probability and Its Applications*, *32*, 535–539.
Neaderhouser, C.C. (1980) Convergence of block spins defined on random fields. *Journal of Statistical Physics*, *22*, 673–684.
Politis, D.N., Romano, J.P. (1993) Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis*, *47*, 301–328.
Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Annals of Probability*, *23*, 1188–1203.
Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, *4*, 185–207.
Robinson, P.M. (2011). Asymptotic theory for non parametric regression with spatial data. *Journal of Econometrics*, *165*, 5–19.
Rosenblatt, M. (1985). *Stationary sequences and random fields*. Birkhauser.
Scott, D.W. (1985) Frequency polygons, theory and applications. *Journal of the American Statistical Association*, *80*, 348–354.
Scott, D.W. (2015). *Multivariate density estimation: Theory, practice and visualization*. Wiley.

Stone, C.J. (1983). Optimal uniform rate of convergence for non parametric estimators of a density function and its derivative. In Revzi, M.H., Rustagi, J.S., Siegmund D. (Eds.) *Recent advances in statistics: Papers in Honor of H. Chernoff*. (pp. 393–406). Academic Press.
Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. *Zeitschrift für Wahrschenlichtkeitstheory und Verwandte Gebiete*, *62*, 477–480.
Terrell, G.R. (1983). The multilinear frequency spline. Technical Report, Department of Math Sciences, Rice University, Houston.
Tran, L.T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, *34*, 37–53.
Tran, L.T., Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis*, *44*, 23–46.