Journal of Multivariate Analysis 123 (2014) 201–213

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On the asymptotic normality of kernel density estimators for causal linear random fields

Yizao Wang [a,*], Michael Woodroofe [b]

[a] Department of Mathematical Sciences, The University of Cincinnati, 2815 Commons Way, ML#0025, Cincinnati, OH 45221–0025, USA

[b] Department of Statistics, The University of Michigan, 439 West Hall, 1085 S. University, Ann Arbor, MI 48109–1107, USA

ARTICLE INFO

Article history:
Received 23 June 2012
Available online 27 September 2013

AMS 2010 subject classifications:
primary 60F05
62G07
secondary 60G10

Keywords:
Central limit theorem
Causal linear random field
Kernel density estimation
m-dependence
Moment inequality

ABSTRACT

We establish sufficient conditions for the asymptotic normality of kernel density estimators applied to causal linear random fields, by m-dependent approximation. Our conditions on the coefficients of linear random fields are weaker than the known results, although our assumption on the bandwidth is not minimal. We also establish a convergence rate of Berry–Esseen's type.

1. Introduction

Let $\{X_i\}_{i\in\mathbb{Z}^d}$, $d \in \mathbb{N}$ be a stationary zero-mean random field, such that the marginal probability density function $f(\cdot)$ exists. We are interested in the Parzen–Rosenblatt kernel density estimator of $p(x)$ in the form of

$$f_n(x) = \frac{1}{n^d b_n} \sum_{i\in[\![1,n]\!]^d} K\left(\frac{x - X_i}{b_n}\right), \quad x \in \mathbb{R}. \tag{1}$$

Throughout this paper, we assume that the kernel $K : \mathbb{R} \to \mathbb{R}$ is a bounded Lipschitz-continuous density function, and the bandwidth b_n satisfies

$$b_n \to 0 \quad \text{and} \quad n^d b_n \to \infty \quad \text{as } n \to \infty. \tag{2}$$

We also write, for $a, b \in \mathbb{Z}$, $[\![a, b]\!] \equiv \{a, a + 1, \ldots, b\}$.

This problem was first considered by Rosenblatt [18] and Parzen [15], in the case that X_i's are independent and identically distributed (i.i.d.) random variables: in particular, one can show the *consistency*

$$\lim_{n\to\infty} f_n(x) = p(x),$$

---

* Corresponding author.
E-mail addresses: yizao.wang@uc.edu (Y. Wang), michaelw@umich.edu (M. Woodroofe).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.09.008

and the *asymptotic normality*

$$(n^d b_n)^{1/2}(f_n(x) - \mathbb{E}f_n(x)) \Rightarrow \mathcal{N}(0, \sigma_x^2) \quad \text{as } n \to \infty, \tag{3}$$

where $\sigma_x^2 = p(x) \int K^2(s)\mathrm{d}s$. See for example Silverman [19] for more references on density estimation problems with i.i.d. data.

The case that X_i's are dependent, however, has presented more challenges, and we focus on establishing the asymptotic normality (3) in this paper. The dependent one-dimensional case has been considered by Robinson [17], Castellana and Leadbetter [2], Bosq et al. [1], Wu and Mielniczuk [23] and Dedecker and Merlevède [6], among others. In particular,

Wu and Mielniczuk [23] investigated thoroughly the case when $\{X_i\}_{i\in\mathbb{Z}}$ is a *causal linear process*. Recall that a linear process $\{X_n\}_{n\in\mathbb{N}}$ has the form

$$X_i = \sum_{k=-\infty}^{\infty} a_k \epsilon_{i-k}, \quad i \in \mathbb{N},$$

where $\sum_k a_k^2 < \infty$ and the *innovations* $\{\epsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. random variables. A causal linear process has in addition $a_k = 0$ for all $k < 0$. Wu and Mielniczuk [23] showed that when ϵ_0 has zero mean and finite variance, the asymptotic normality holds if

$$\sum_{k=0}^{\infty} |a_k| < \infty \quad \text{and} \quad a_k = 0 \quad \text{for all } k < 0, \tag{4}$$

with bandwidth satisfying (2) ($d = 1$).

The asymptotic normality of kernel density estimators for random fields has been considered by Tran [20], Hallin et al. [11], Cheng et al. [4] and El Machkouri [8,9], among others. The extension of results in one dimension to high dimensions, however, is not trivial. As summarized in Hallin et al. [11], 'the points of $\mathbb{Z}^d$ do not have a natural ordering. As a result, most techniques available for one-dimensional processes do not extend to random fields.' Indeed, the *martingale approximation* approach taken by Wu and Mielniczuk seems difficult to be extended to high dimension. One would naturally think of a scheme of multiparameter martingale approximation. There are a few different definitions of multiparameter martingales in the literature. The notion of *orthomartingale* [14] is the most general and it seems to suit our model here. However, to the best of our knowledge, establishing a general central limit theorem for triangular arrays of orthomartingales remains an open problem. At the same time, other notions of multiparameter martingales, with general central limit theorems established, have more restrictive assumptions on the underlying filtration structure. See for example the discussion in [21].

In particular, a notorious difficulty for the kernel density estimation of random fields is that one often needs more assumptions on the bandwidth b_n than the *minimal* one (2). This condition is minimal in the sense that it is the natural condition for the asymptotic normality (3) to hold when X_i's are i.i.d. To the best of our knowledge, only the recent results by El Machkouri [8,9] assume no other but minimal condition (2) on b_n for dependent random fields.

In this paper, we focus on the kernel density estimation for *causal* linear random fields $\{X_i\}_{i\in\mathbb{Z}^d}$ ($d \in \mathbb{N}$) in form of

$$X_i = \sum_{k\in\mathbb{Z}^d, k\succeq\mathbf{0}} a_k \epsilon_{i-k}, \quad i \in \mathbb{Z}^d, \tag{5}$$

where $\sum_{i\succeq\mathbf{0}} a_i^2 < \infty$ and $\{\epsilon_i\}_{i\in\mathbb{Z}^d}$ are i.i.d. zero-mean random variables with finite second moments. Throughout this paper, we let '$i \succeq k$' denote '$i_\tau \geq k_\tau$ for all $\tau = 1, \ldots, d$' for $i, k \in \mathbb{Z}^d$ and write $\mathbf{0} = (0, \ldots, 0)$, $\mathbf{1} = (1, \ldots, 1) \in \mathbb{Z}^d$. The causality assumption is technically crucial in establishing weak conditions for the asymptotic normality. This has been known in the one-dimensional case: somehow surprisingly, with the method of Wu and Mielniczuk [23], it remains open to weaken the condition (4) to the non-causal version $\sum_{k\in\mathbb{Z}} |a_k| < \infty$.

We provide new conditions on the coefficient $\{a_i\}_{i\in\mathbb{Z}^d}$ such that the asymptotic normality (3) holds (see Theorem 1 below) and compare with results obtained by Hallin et al. [11] and El Machkouri [9]. In both cases, our conditions are weaker on the *coefficients* $\{a_i\}_{i\in\mathbb{Z}^d}$. On the other hand, our condition on the *bandwidth* improves the one in [11], but it is still stronger than the minimal one (2) assumed in [9]. Note also that for the index sets, [9] considers arbitrary shapes other than $[\![1, n]\!]^d$. We do not compare our result with Cheng et al. [4], as there is a mistake in their proof. We briefly explain the problem and how this might be fixed in Remark 6 below. At the end, we also establish a convergence rate of Berry–Esseen's type.

Our proof is based on the m-approximation approach. As a key step of our approach, we establish a central limit theorem for triangular arrays of stationary m-dependent random fields with unbounded m (Theorem 2). In the case that the partial sums are taken over rectangles, this result improves a central limit theorem established by Heinrich [12] (it however treated partial sums over general index sets). We also applied certain moment inequalities for stationary random fields established in [21].

The paper is organized as follows. Our assumptions and main results are presented in Section 2. Examples and comparison with other results are provided in Section 3. Section 4 is devoted to the central limit theorem for triangular arrays of m-dependent random fields. Section 5 establishes asymptotic normality by m-approximation. Section 6 investigates the convergence rate. Auxiliary proofs are given in Section 7.