



# Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes

Salim Bouzebda<sup>1</sup> · Sultana Didi<sup>2</sup>

Received: 9 January 2020 / Accepted: 27 July 2020 / Published online: 17 August 2020  
© Universidad Complutense de Madrid 2020

## Abstract

In the present paper, we consider the nonparametric regression model with random design based on  $(\mathbf{X}_t, \mathbf{Y}_t)_{t \geq 0}$  a  $\mathbb{R}^d \times \mathbb{R}^q$ -valued strictly stationary and ergodic continuous time process, where the regression function is given by  $m(\mathbf{x}, \psi) = \mathbb{E}(\psi(\mathbf{Y}) | \mathbf{X} = \mathbf{x})$ , for a measurable function  $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ . We focus on the estimation of the location  $\Theta$  (mode) of a unique maximum of  $m(\cdot, \psi)$  by the location  $\hat{\Theta}_T$  of a maximum of the Nadaraya–Watson kernel estimator  $\hat{m}_T(\cdot, \psi)$  for the curve  $m(\cdot, \psi)$ . Within this context, we obtain the consistency with rate and the asymptotic normality results for  $\hat{\Theta}_T$  under mild local smoothness assumptions on  $m(\cdot, \psi)$  and the design density  $f(\cdot)$  of  $\mathbf{X}$ . Beyond ergodicity, any other assumption is imposed on the data. This paper extends the scope of some previous results established under the mixing condition. The usefulness of our results will be illustrated in the construction of confidence regions.

**Keywords** Conditional density · Conditional mode · Kernel estimate · Nadaraya–Watson estimators · Ergodic processes · Martingale difference arrays · Prediction · Continuous time processes · Confidence regions · Rate of convergence · Strong consistency

**Mathematics Subject Classification** 62G05 · 62G07 · 62G08 · 62E20 · 60F05 · 60E05 · 60A10

---

✉ Salim Bouzebda  
salim.bouzebda@utc.fr

Sultana Didi  
s.biha@qu.edu.sa

<sup>1</sup> Alliance Sorbonne Université, Université de Technologie de Compiègne, L.M.A.C., Compiègne, France

<sup>2</sup> College of Sciences, Qassim University, PO Box 6688, 51452 Buraydah, Saudi Arabia

## 1 Introduction

Nonparametric estimation has been the subject of intense investigation for many years and this has led to the development of a large variety of methods. Because of numerous applications and their important role in mathematical statistics, the problem of estimating the density and regression function has been the subject of considerable interest during the last decades. One of the most commonly used classes of estimators is that formed by the so-called kernel-type estimators. For more theoretical aspects along with statistical applications the interested reader is referred to Tapia and Thompson [71], Wertz [74], Devroye and Györfi [23], Devroye [22], Nadaraya [57], Härdle [38], Wand and Jones [73], Eggermont and LaRiccia [30], Devroye and Lugosi [24] and the references therein. Recently, a number of statistical problems has found an unexpected solution being investigated by a “modal point of view”. This investigation includes classical processes such as clustering. This has led to a renewed interest in the estimation and the inference for the mode. The estimation of the conditional mode of an outcome variable given the regressors, is called modal regression. Modal regression is an alternative approach to the usual regression methods for exploring the relationship between a response variable  $\mathbf{Y}$  and a predictor variable  $\mathbf{X}$ . Unlike conventional regression, which is based on the conditional mean of  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ , modal regression estimates conditional modes of  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ . Modal regression is a more reasonable modelling approach than the usual regression at least in two scenarios. Firstly when the conditional density function is skewed or has a heavy tail. When the conditional density function has skewness, the conditional mean may not provide a good representation for summarising the relations between the response and the covariate. The other scenario is when the conditional density function has multiple local modes. This occurs when the relation of  $\mathbf{X}$  and  $\mathbf{Y}$  contains multiple patterns. The conditional mean may not capture any of these patterns, so it can be a very bad summary; see Chen et al. [17] for an example. This situation has already been pointed out in Tarter and Lock [72]. Modal regression has a wide variety of applications including the analysis of traffic and forest fire data [31,75], econometrics [45,50,51], and machine learning [33,68]. For example, Kemp and Santos Silva [45] argue that the mode is the most intuitive measure of central tendency for positively skewed data found in many econometric applications such as wages, prices, and expenditures [45, p. 93]. For more recent reviews and further details on the subject the reader is referred to Chen [16] and Chacón [14].

We will start by providing some notation and definitions that are needed for the forthcoming sections. Let  $(\mathbf{X}_t, \mathbf{Y}_t)_{t \geq 0}$  be a  $\mathbb{R}^d \times \mathbb{R}^q$ -valued strictly stationary and ergodic<sup>1</sup> continuous time process defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $g(\cdot, \cdot)$  be the density function of the random vector  $(\mathbf{X}_t, \mathbf{Y}_t)$ ,  $f(\cdot)$  be the density of  $\mathbf{X}_t$  and  $\rho(\cdot)$  the density of  $\mathbf{Y}$ . For a given measurable function  $\psi(\cdot)$  and  $\mathbf{x} \in \mathbb{R}^d$  the regression function, whenever it exists, is defined to be

$$m(\mathbf{x}, \psi) = \mathbb{E}(\psi(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}).$$

<sup>1</sup> The definition is provided in the “Appendix”.

In this situation, we have the random design regression model and  $\mathbf{X}$  is called the design variable and  $\mathbf{Y}$  the response variable. The random design model is very important in clinical studies, where the design variable usually represents the age of a particular individual receiving treatment, and  $\mathbf{Y}$  is the quantity whose dependence on the age of the patient is investigated. A typical example (from forensic medicine) is given by Härdle and Marron [39], where  $\mathbf{Y}$  stands for the liver weight of female persons (depending on their age). Inequalities  $\mathbf{x} \leq \mathbf{y}$  holds for all the components, i.e.,  $x_j \leq y_j$  for all  $j = 1, \dots, d$ . The introduction of the function  $\psi(\cdot)$  allows us to include some important special cases:

- $\psi(\mathbf{Y}) = \mathbb{1}\{\mathbf{Y} \leq \mathbf{y}\}$  gives the conditional distribution of  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ .
- $\psi(\mathbf{Y}) = \mathbf{Y}^k$  gives the conditional moments of  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ .

In the present paper, we focus on estimating the location  $\Theta$  and the size  $m(\Theta, \psi)$  of a unique maximum (mode, peak) of the (unknown) function  $m(\cdot, \psi)$ . Our method is indirect in the sense that the estimators of  $\Theta$  and  $m(\Theta, \psi)$  are based on a kernel estimator  $\widehat{m}_T(\mathbf{x}, \psi)$  of the regression curve  $m(\mathbf{x}, \psi)$ . We will use the Nadaraya–Watson estimator which is defined by

$$\widehat{m}_T(\mathbf{x}, \psi) := \begin{cases} \frac{\frac{1}{Th_T^d} \int_0^T \psi(\mathbf{Y}_t) K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt}{\frac{1}{Th_T^d} \int_0^T K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt}, & \text{if } \frac{1}{Th_T^d} \int_0^T K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt \neq 0, \\ \frac{1}{T} \int_0^T \psi(\mathbf{Y}_t) dt, & \text{if } \frac{1}{Th_T^d} \int_0^T K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt = 0, \end{cases}$$

where  $K(\cdot)$  is a kernel,  $h_T$  is a positive sequence of real numbers such that

$$(i) \lim_{T \rightarrow \infty} h_T = 0, \quad (ii) \lim_{T \rightarrow \infty} Th_T^d = +\infty, \quad \text{or} \quad (iii) \lim_{T \rightarrow \infty} \frac{Th_T^d}{\log T} = +\infty. \quad (1.1)$$

The condition (i) is used to obtain the asymptotic unbiasedness of the kernel (density or regression) type estimators. We need more restrictive assumption on  $h_T$  for the consistency, this is given by the condition (ii), one can refer to Parzen [61]. In general, the strong consistency fails to hold when either (i) or (iii) is not satisfied. Now the location  $\Theta$  (mode) and the size  $m(\Theta, \psi)$  are estimated by the respective functionals  $\widehat{\Theta}_T$  and  $\widehat{m}_T(\widehat{\Theta}_T, \psi)$  pertaining to  $\widehat{m}_T(\cdot, \psi)$ , i.e.,  $\widehat{\Theta}_T$  is chosen through the equation

$$\widehat{m}_T(\widehat{\Theta}_T, \psi) = \sup_{\mathbf{x} \in \mathcal{C}} \widehat{m}_T(\mathbf{x}, \psi), \quad (1.2)$$

where the maximum is running over some compact set  $\mathcal{C} \subset \mathbb{R}^d$ . Note that  $\widehat{\Theta}_T$  exists if  $K(\cdot)$  is continuous; however, it may not be unique. In fact, it is known that kernel estimators tend to produce some additional and superfluous modality. In this context, one can consider

$$\widehat{\Theta}_T = \inf \left\{ \mathbf{t} \in \mathcal{C} \text{ such that } \widehat{m}_T(\mathbf{t}, \psi) = \sup_{\mathbf{x} \in \mathcal{C}} \widehat{m}_T(\mathbf{x}, \psi) \right\},$$

where the infimum is taken with respect to the lexicographic order on  $\mathbb{R}^d$ . However, this has no bearing on asymptotic theory; our results are valid for any choice of  $\hat{\Theta}_T$  satisfying (1.2). To ensure both uniqueness and measurability of  $\hat{\Theta}_T$ , one could use the so-called mode functional on  $C(\mathcal{C})$  apparently introduced by Eddy [28] which considers the infimum of the maximised locations and whose measurability is also proved in a paper of Eddy [28]. Alternatively, Grund and Hall [35] have suggested to break ties at random if necessary. Anyway, the validity of our proofs will not be affected by potential non-measurability of  $\hat{\Theta}_T$ , since we can always replace probabilities by outer probabilities when necessary with no further changes in the proofs, this issue is discussed in Ziegler [79], and also Ziegler [77,78], Herrmann and Ziegler [41]. As it is mentioned in Ziegler [79], estimating the mode and size of a maximum of a nonparametric curve by the corresponding functionals of a kernel estimator of the curve is not new; it stems from the closely related problem of estimating the mode of a density. In continuation of Parzen [61] pioneering work on density estimation and estimation of the mode, Eddy [28,29] and Romano [66] tackled optimality questions of the kernel density estimators of the mode. Romano [66] seems also to be the first to consider data-dependent bandwidths in this framework. In another paper, Romano [65] examined the limiting behaviour of bootstrap estimators of the location of the mode, an idea used later by Grund and Hall [35] in the context of bandwidth selection by minimising the bootstrapped  $L_p$ -error for the mode estimator. It is worth noticing that the conditional mode function estimate of the predictor is used for the first time by Collomb et al. [18]. The kernel type estimators were studied extensively in different setting of dependencies, we cite among many others Samanta and Thavaneswaran [67], Ould-Saïd [59], Quintela-Del-Río and Vieu [64], Berlinet et al. [5], Ferraty et al. [34], Ezzahrioui and Ould-Saïd [32], Benrabah et al. [3] and the references therein. Quintela-Del-Río and Vieu [64] motivated the use of the conditional mode by pointing out that the prediction of  $\mathbf{Y}$ -values given the  $\mathbf{X}$ -values is achieved through the regression function estimation. Finally, when the process is considered to be i.i.d., the almost sure convergence along with the mean convergence of the conditional density were obtained by Youndjé [76]. Ota et al. [60] proposed a new estimator of the conditional mode that is able to avoid the curse of dimensionality and at the same time is computationally scalable, thereby complementing the above existing methods. Within the framework described above, our aim is to establish consistency and asymptotic normality results (which in turn can be exploited for the construction of confidence intervals) for the estimators  $\hat{\Theta}_T$  and  $\hat{m}_T(\hat{\Theta}_T, \psi)$  for the location and the size of the peak under some mild local smoothness conditions on the regression function  $m(\cdot, \psi)$  and the design density  $f(\cdot)$  (mostly imposed locally in a neighbourhood of  $\Theta$ ). Those results will be valid for a wide class of kernels not necessarily having compact support. This includes, in particular, the Gaussian kernel which is widely used in practice. Mixing is some kind of asymptotic independence assumption which is commonly used for seek of simplicity but which can be unrealistic in situations where there is strong dependence between the data. Extending non-parametric functional ideas to general dependence structure is a rather underdeveloped field. Note that the ergodic framework avoids the widely used strong mixing condition and its variants to measure the dependency and the involved probabilistic calculations that it implies (see, for instance, Masry [55]). It is worth noticing that the ergodicity is implied by all mixing

conditions, being weaker than all of them. Further motivations to consider ergodic data are discussed in Laib and Louani [48,49], Didi and Louani [27], Bouzebda et al. [12], Bouzebda et al. [8], Bouzebda and Didi [9–11] and Krebs [46], in some of these references the definitions of the ergodic property of continuous time processes are given. In the present work, we do not assume anything beyond ergodicity of the underlying process. It is worth noticing that strong mixing implies ergodicity; see e.g., Remark 2.6 on page 50 in combination with Proposition 2.8 on page 51 in Bradley [13]. Hence the present work extends the scope of applications compared to the existing works. On the other hand, we mention that there exist interesting processes which are ergodic but not mixing according to Andrews [1] and Bradley [13]. An example of an ergodic and non-mixing process was considered in Sect. 5.3 of Leucht and Neumann [52]. Indeed, assume that the process  $\{(T_i, \lambda_i) : i \in \mathbb{Z}\}$  is strictly stationary with  $T_i \mid \mathcal{T}_{i-1} \sim \text{Poisson}(\lambda_i)$ , let  $\mathcal{T}_i$  be the  $\sigma$ -field generated by  $(T_i, \lambda_i, T_{i-1}, \lambda_{i-1}, \dots)$ . We assume that  $\lambda_i = \kappa(\lambda_{i-1}, T_{i-1})$ , where  $\kappa : [0, \infty) \times \mathbb{N} \rightarrow (0, \infty)$ . However, this process is not mixing in general; see Remark 3 of Neumann [58] for a counterexample. We refer to Leucht and Neumann [52] for further details and motivations for the use of the ergodicity assumption. One of their arguments, is that for certain classes of processes, it can be much easier to prove ergodicity rather than mixing assumption. It is known that any sequence  $\{\varepsilon_t : t \in \mathbb{Z}\}$  of i.i.d. random variables is ergodic. Hence, it is immediately clear that  $\{\mathbf{Y}_t : t \in \mathbb{Z}\}$  with

$$\mathbf{Y}_t = \vartheta((\dots, \varepsilon_{t-1}, \varepsilon_t), (\varepsilon_{t+1}, \varepsilon_{t+2}, \dots))$$

is also ergodic. Didi [25] has constructed an example of a non-mixing ergodic continuous time process. It is well known that the fractional Brownian motion  $\{W_t^H : t \geq 0\}$  with parameter  $H \in (0, 1)$  has strictly stationary increments. Otherwise, the fractional Gaussian noise, defined for every  $s > 0$  by

$$\{G_t^H : t \geq 0\} := \{W_{t+s}^H - W_t^H : t \geq 0\},$$

is a strictly stationary centered long memory process when  $H \in (\frac{1}{2}, 1)$  (see for instance, Beran [4, p.55] and Lu [53, p.17]), hence the condition of strong mixing is not satisfied. Let  $\{G_t : t \geq 0\}$  be a strictly stationary centered Gaussian process with correlation function

$$R(t) = \mathbb{E}[G_0 G_t].$$

Relying on the work of Maslowski and Pospíšil [54], Lemma 4.2, it follows that the process  $\{G_t : t \geq 0\}$  is ergodic whenever

$$\lim_{t \rightarrow \infty} R(t) = 0,$$

which is the case for the process  $\{G_t^H : t \geq 0\}$ . The ergodicity hypothesis seems to be the most naturally adapted and provides a better framework to study data series, for example, generated by noisy chaos.

To the best of our knowledge, the results presented here, respond to a problem that has not been studied systematically until recently, and it gives the main motivation to this paper. Indeed, we establish the exact rate of strong uniform consistency of the estimators  $\widehat{\Theta}_T$  and we characterise the limiting law. To prove our results, we base our methodology upon the martingale approximation which allows to provide a unified nonparametric time series analysis framework enabling one to study systematically dependent data. This methodology is a quite different approach, in the i.i.d. context, compared to the existing ones.

The layout of the article is as follows. The assumptions and asymptotic properties of the estimators are given in Sect. 2, which includes the optimal convergence rates and the asymptotic normality of the estimators  $\widehat{\Theta}_T$ . Some concluding remarks and possible future developments are mentioned in Sect. 3. To avoid interrupting the flow of the presentation, all mathematical proofs are presented in Sect. 4.

### 2 Main results

Let us introduce some notation and definitions. Let  $\alpha = (\alpha_1, \dots, \alpha_d)$  be a multi-index of the nonnegative integers  $\alpha_i$ , set  $|\alpha| = \sum_{i=1}^d \alpha_i$ , and let

$$D^\alpha = \frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1} \dots (\partial x_d)^{\alpha_d}}$$

denote the partial differential operator of order  $\alpha$ . For  $\alpha = 0$  set  $D^\alpha = id$ , for identity. For continuous real-valued functions  $\zeta_1(\cdot)$  and  $\zeta_2(\cdot)$  that are  $s$ -times continuously differentiable on  $\mathbb{R}^d$ ,

$$D^\alpha(\zeta_1 \zeta_2) = \sum_{\{\beta : \beta \leq \alpha\}} \frac{\alpha!}{(\alpha - \beta)! \beta!} \binom{\alpha}{\beta} (D^\beta \zeta_2)(D^{\alpha - \beta} \zeta_1).$$

We will use the notation

$$D^i \zeta_1 = \zeta_1^{(i)} \text{ for } i = 1, \dots, s.$$

Let us define the partial derivatives of order one of the regression estimator by

$$\begin{aligned} \widehat{m}_T^{(1)}(\mathbf{x}, \psi) &= \left( \frac{M_T(\mathbf{x}, \psi)}{f_T(\mathbf{x})} \right)^{(1)} \\ &= \frac{M_T^{(1)}(\mathbf{x}, \psi) f_T(\mathbf{x}) - f_T^{(1)}(\mathbf{x}) M_T(\mathbf{x}, \psi)}{f_T^2(\mathbf{x})}. \end{aligned}$$

The derivatives of order  $\alpha = 1, 2$  of the estimators  $f_T(\mathbf{x})$  and  $M_T(\mathbf{x}, \psi)$  are defined as follows

$$f_T^{(\alpha)}(\mathbf{x}) = \frac{1}{Th_T^{d+\alpha}} \int_0^T K^{(\alpha)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt,$$

and

$$M_T^{(\alpha)}(\mathbf{x}, \psi) = \frac{1}{Th_T^{d+\alpha}} \int_0^T \psi(\mathbf{Y}_t) K^{(\alpha)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt.$$

We denote by  $m^{(1)}(\cdot, \psi)$  the gradient of the function  $m(\cdot, \psi) : \mathbb{R}^d \rightarrow \mathbb{R}$ , that is,  $m^{(1)}(\cdot, \psi)$  is the  $d \times 1$ -vector of the partial derivatives of  $m(\cdot, \psi)$

$$m^{(1)}(\cdot, \psi) = \left( \frac{\partial}{\partial x_1} m(\cdot, \psi), \dots, \frac{\partial}{\partial x_d} m(\cdot, \psi) \right)^\top.$$

Using the definition of the conditional mode function, i.e. the mode of  $m(\cdot, \psi)$ , we have

$$m^{(1)}(\Theta, \psi) = \left( \frac{\partial}{\partial x_1} m(\Theta, \psi), \dots, \frac{\partial}{\partial x_d} m(\Theta, \psi) \right)^\top = 0. \tag{2.1}$$

Similarly, it follows from the statement (2.1) that

$$\widehat{m}_T^{(1)}(\widehat{\Theta}_T, \psi) = \left( \frac{\partial}{\partial x_1} \widehat{m}_T(\widehat{\Theta}_T, \psi), \dots, \frac{\partial}{\partial x_d} \widehat{m}_T(\widehat{\Theta}_T, \psi) \right)^\top = 0.$$

We denote by  $m^{(2)}(\cdot, \psi)$  the Hessian of the function  $m(\cdot, \psi)$ , that is,  $m^{(2)}(\cdot, \psi)$  is the  $d \times d$ -matrix of the second partial derivatives of  $m(\cdot, \psi)$ . Furthermore, assumption (A.7) implies that

$$m^{(2)}(\Theta, \psi) < 0, \quad \text{and} \quad \widehat{m}_T^{(2)}(\widehat{\Theta}_T, \psi) < 0.$$

By the definition of  $\widehat{\Theta}_T$ , we have  $\widehat{m}_T^{(1)}(\widehat{\Theta}_T, \psi) = 0$  so that

$$\widehat{m}_T^{(1)}(\widehat{\Theta}_T, \psi) - \widehat{m}_T^{(1)}(\Theta, \psi) = -\widehat{m}_T^{(1)}(\Theta, \psi). \tag{2.2}$$

For each  $i \in \{1, \dots, d\}$ , Taylor's expansion applied to the real-valued application  $\frac{\partial}{\partial x_i} \widehat{m}_T^{(1)}(\cdot, \psi)$  implies the existence of  $\Theta_T^*(i) = (\Theta_{T,1}^*(i), \dots, \Theta_{T,d}^*(i))^\top$

$$\begin{cases} \frac{\partial}{\partial x_i} \widehat{m}_T(\widehat{\Theta}_T, \psi) - \frac{\partial}{\partial x_i} \widehat{m}_T(\Theta, \psi) = \sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} \widehat{m}_T(\Theta_T^*(i), \psi) (\widehat{\Theta}_{T,j} - \Theta_j), \\ \left| \Theta_{T,1}^*(i) - \Theta_j \right| \leq |\widehat{\Theta}_{T,j} - \Theta_j|, \quad j \in \{1, \dots, d\}. \end{cases} \tag{2.3}$$

Define the  $d \times d$  matrix  $H_T = (H_{T,i,j})_{1 \leq i, j \leq d}$  by setting

$$H_{T,i,j} = \frac{\partial^2}{\partial x_i \partial x_j} \widehat{m}_T(\Theta_T^*(i), \psi).$$

Equation (2.2) can then be rewritten as

$$H_T(\widehat{\Theta}_T - \Theta) = -\widehat{m}_T^{(1)}(\Theta, \psi). \tag{2.4}$$

The last relation will play an important role in our proofs, in particular for the asymptotic normality. To formulate our assumptions, some additional notation is required, for some constant  $\delta > 0$  small enough, let  $n \in \mathbb{N}$  be such that  $T = \delta n$ , and  $T_j = j\delta$ , for  $j = 1, \dots, n$ . Let  $\mathcal{F}_t$  be the  $\sigma$ -field defined by

$$\mathcal{F}_t := \sigma\{(\mathbf{X}_s, \mathbf{Y}_s) : 0 \leq s < t\}.$$

Set  $\mathcal{F}_j$  to be the  $\sigma$ -field defined by

$$\mathcal{F}_j := \sigma\{(\mathbf{X}_s, \mathbf{Y}_s) : 0 \leq s \leq T_j\}.$$

Let  $\mathcal{S}_{t,\delta}$  be the  $\sigma$ -field defined by

$$\mathcal{S}_{t,\delta} := \sigma\{(\mathbf{X}_s, \mathbf{Y}_s), (\mathbf{X}_r) : 0 \leq s < t; t \leq r \leq t + \delta\}.$$

Let  $\mathcal{G}_t := \sigma\{(\mathbf{X}_s, \mathbf{Y}_s) : 0 \leq s \leq t\}$ , and for  $\delta > 0$  small enough, let  $g^{\mathcal{G}_{t-\delta}}(\cdot)$  and  $\rho^{\mathcal{G}_{t-\delta}}(\cdot)$  be the conditional densities of  $(\mathbf{X}, \mathbf{Y})$  and  $\mathbf{Y}$  respectively, given the  $\sigma$ -field  $\mathcal{G}_{t-\delta}$ . Finally, if  $\zeta(\cdot)$  is a real-valued random function which satisfies  $\zeta(u)/u \rightarrow 0$  a.s. as  $u \rightarrow 0$ , we write  $\zeta(u) = o_{\text{a.s.}}(u)$ . In the same way, we say that  $\zeta(u)$  is  $O_{\text{a.s.}}(u)$  if  $\zeta(u)/u$  is a.s. bounded as  $u \rightarrow 0$ .

### 2.1 Assumptions

In our analysis, the following assumptions are needed.

(A.1) The kernel  $K(\cdot)$ , is a probability density function compactly supported,

(i) Kernel  $K$  is assumed to be Lipschitz with ratio  $C_K < \infty$  and order  $\gamma$ , i.e.,

$$|K(\mathbf{x}) - K(\mathbf{x}')| \leq C_K \|\mathbf{x} - \mathbf{x}'\|^\gamma, \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{2d};$$

(ii)  $\int_{\mathbb{R}^d} \|\mathbf{x}\| K(\mathbf{x}) d\mathbf{x} < \infty$ ;

(A.2) There exists  $\Gamma < \infty$ , such that, for all  $\mathbf{x} \in \mathcal{C}$ ,

$$\sup_{\mathbf{x} \in \mathcal{C}} |m(\mathbf{x}, \psi)| < \Gamma;$$

(A.3) (i) Recall that  $\mathcal{C}$  is a compact set of  $\mathbb{R}^d$ . Assume, for all  $\mathbf{x} \in \mathcal{C}$ , that there exists  $\lambda > 0$  and finite constant  $0 < \eta$  such that

$$\lambda \leq f(\mathbf{x}) \leq \eta;$$

(ii) the density  $f(\cdot)$  is an element of  $\mathcal{C}^2(\mathbb{R}^d)$ ;

(A.4) For every  $t \in \mathbb{R}_+$ , for every  $\mathbf{x} \in \mathbb{R}^d$ ,

(i) The conditional density  $f_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}}(\cdot)$  of  $\mathbf{X}_t$  given the  $\sigma$ -field  $\mathcal{F}_{t-\delta}$  exists a.s. and is an element of  $\mathcal{C}^2(\mathbb{R}^d)$ ;



(ii) For any  $\delta > 0$  small enough

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt = f(\mathbf{x}), \quad \text{a.s.};$$

(A.5) For every  $t \in \mathbb{R}_+$ , for every  $\mathbf{x} \in \mathbb{R}^d$ ,

$$\lim_{T \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T f_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt - f(\mathbf{x}) \right| = 0, \quad \text{a.s.},$$

for any  $\delta > 0$  small enough;

(A.6) For any  $t$  and  $r$  such that  $t \in [0, T]$  and  $t \leq r \leq t + \delta$  we have

(i)

$$\mathbb{E}(|\psi(\mathbf{Y}_r)| | \mathcal{S}_{t,\delta}) = \mathbb{E}(|\psi(\mathbf{Y}_r)| | \mathbf{X}_r) = m(\mathbf{X}_r);$$

(ii) there exist constants  $C_\psi > 0$  and  $\beta > 0$  such that, for any couple  $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{2d}$ ,

$$|m(\mathbf{x}, \psi) - m(\mathbf{x}', \psi)| \leq C_\psi \|\mathbf{x} - \mathbf{x}'\|^\beta;$$

(iii) For any  $k \geq 2$  and any  $\delta > 0$ ,

$$\mathbb{E} \left( \left| \psi^k(\mathbf{Y}_r) \right| | \mathcal{S}_{t,\delta} \right) = \mathbb{E} \left( \left| \psi^k(\mathbf{Y}_r) \right| | \mathbf{X}_r \right),$$

and the function  $\Phi_k(\mathbf{x}, \psi) = \mathbb{E} \left( \left| \psi^k(\mathbf{Y}) \right| | \mathbf{X} = \mathbf{x} \right)$  is continuous in the neighbourhood of  $\mathbf{x}$ ;

(A.7) For any fixed  $\mathbf{x} \in \mathbb{R}^d$ ,

- (i)  $m(\mathbf{x}, \psi)$  is twice differentiable on  $\mathbb{R}^d$ , the matrix  $m^{(2)}(\mathbf{x}, \psi)$  is continuous in a neighbourhood of  $\Theta$ , and  $m^{(2)}(\Theta, \psi)$  is nonsingular;
- (ii)  $m^{(2)}(\Theta, \psi)$  is bounded on  $\mathbb{R}^d$ .

### Comments on hypotheses

Conditions (A.1) are very common in the nonparametric function estimation literature. Notice that the condition (A.1) is classical in the nonparametric estimation procedures. In particular, by imposing the condition (A.1)(i), the kernel function exploits the smoothness of the density function or the regression function. If we loose the condition that the kernel function  $K(\cdot)$  must be a density, the convergence rate could be faster. Indeed, the convergence rate can be made arbitrarily close to the parametric  $n^{-1}$  as the order increases. In fact, Chac3n et al. [15] showed that the parametric rate  $n^{-1}$  can be attained by the use of super-kernels, and that super-kernel density estimators automatically adapt to the unknown degree of smoothness of the density. The main

drawback of higher-order kernels in this situation is the negative impact of the kernel may make the estimated density not a density itself. The interested reader is referred to, e.g., Jones et al. [44], Jones and Signorini [43] and Jones [42]. They set some kind of regularity upon the kernels used in our estimates. Notice that the transform of the stationary ergodic process  $(\mathbf{X}_t, \mathbf{Y}_t)_{t \geq 0}$  into the process  $(\psi^2(\mathbf{Y}_t))_{t \geq 0}$  is a measurable function. Therefore, making use of Proposition 4.3 of Krengel [47] and then the ergodic Theorem, we obtain

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \psi^2(\mathbf{Y}_t) dt = \mathbb{E} \left[ \psi^2(\mathbf{Y}_0) \right] \quad \text{a.s.}, \tag{2.5}$$

Condition (A.3)(ii), is a technical condition that simplifies our proofs, precisely, we assume that the density function  $f(\cdot)$  is bounded away from zero and infinity on the compact set  $\mathcal{C}$  in a similar way as in Ziegler [79], Stute [70], Harel and Puri [40], Debbarh [20]. For any set  $B \subset \mathbb{R}^d$  and  $\epsilon > 0$ , denote by  $B^\epsilon$  the set of all  $\mathbf{x} \in \mathbb{R}^d$  such that there exists  $\mathbf{y} \in B$  with  $\|\mathbf{x} - \mathbf{y}\| < \epsilon$ . One can use that  $f(\cdot)$  is continuous and strictly positive on  $\mathcal{C}^\epsilon$ , but this will add much extra complexity to the proofs. Condition (A.4) involves the ergodic nature of the data as given, for instance, in Györfi et al. [36]. Assuming that  $\rho^{\mathcal{G}_{t-\delta}}(\cdot)$  and  $g^{\mathcal{G}_{t-\delta}}(\cdot)$  belong to the space  $\mathcal{C}^0$ , at least, of continuous functions, which is a separable Banach space. Moreover, approximating the integral  $\int_0^T \rho^{\mathcal{G}_{t-\delta}}(\mathbf{y}) dt$  and  $\int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, \mathbf{y}) dt$  by their Riemann's sums, it follows that

$$\begin{aligned} T^{-1} \int_0^T \rho^{\mathcal{G}_{t-\delta}}(\mathbf{y}) dt &\simeq n^{-1} \sum_{i=1}^n \rho^{\mathcal{G}_{t_i-\delta}}(\mathbf{y}) \\ &= n^{-1} \sum_{j=1}^n \rho^{\mathcal{G}_{(j-1)\delta}}(\mathbf{y}), \end{aligned}$$

and

$$\begin{aligned} T^{-1} \int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, \mathbf{y}) dt &\simeq n^{-1} \sum_{i=1}^n g^{\mathcal{G}_{t_i-\delta}}(\mathbf{x}, \mathbf{y}) \\ &= n^{-1} \sum_{j=1}^n g^{\mathcal{G}_{(j-1)\delta}}(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Since the processes  $(\mathbf{X}_{T_j}, \mathbf{Y}_{T_j})_{j \geq 1}$  and  $(\mathbf{Y}_{T_j})_{j \geq 1}$  are stationary and ergodic (see Proposition 4.3 of Krengel [47]) following Delecroix [21] (see, Lemma 4 and Corollary 1 along with their proofs), one may prove that the sequences  $(\rho^{\mathcal{G}_{(j-1)\delta}}(\mathbf{y}))_{j \geq 1}$  and  $(g^{\mathcal{G}_{(j-1)\delta}}(\mathbf{x}, \mathbf{y}))_{j \geq 1}$  of conditional densities are stationary and ergodic. Moreover, making use of Beck [2]'s theorem (see, for instance, Györfi et al. [36], Theorem 2.1.1), it

follows that

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \sup_{\mathbf{y} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T \rho^{\mathcal{G}_{t-\delta}}(\mathbf{y}) dt - \mathbb{E}(\rho^{\mathcal{G}_{-\delta}}(\mathbf{y})) \right| \\ &= \limsup_{T \rightarrow \infty} \sup_{\mathbf{y} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T \rho^{\mathcal{G}_{t-\delta}}(\mathbf{y}) dt - \rho(\mathbf{y}) \right| = 0, a.s. \end{aligned}$$

and

$$\begin{aligned} & \limsup_{T \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, \mathbf{y}) dt - \mathbb{E}(g^{\mathcal{G}_{-\delta}}(\mathbf{x}, \mathbf{y})) \right| \\ &= \limsup_{T \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x}, \mathbf{y}) dt - g(\mathbf{x}, \mathbf{y}) \right| = 0, a.s. \end{aligned}$$

It is then clear that both the conditions (A.4) and (A.5) are satisfied. Condition (A.6)(i) is usual in the literature dealing with the study of ergodic processes. The hypothesis (A.6)(ii) is a regularity condition upon the regression function. For the condition (A.6)(iii), we can refer to the following examples.

**Example 2.1** Consider the regression model  $Y_t = m(X_t) + \epsilon_t$ , where the random variables  $\epsilon_t$ 's stand as martingale differences with respect to the  $\sigma$ -field  $\mathcal{S}_{r,\delta}$ ,  $r \leq t \leq r + \delta$ , generated by  $\{(X_s, \epsilon_s), (X_t) : 0 \leq s < r, r \leq t \leq r + \delta\}$ . Clearly, we have

$$\mathbb{E}[Y_t | \mathcal{S}_{r,\delta}] = m(X_t),$$

almost surely.

**Example 2.2** Consider the regression model  $Y_t = m(X_t) + \sigma(X_t)\epsilon_t$ , where the random variables  $\epsilon_t$  are centered and independent of the process  $(X_t)_{t \geq 0}$ . Taking  $\mathcal{S}_{r,\delta}$  as the  $\sigma$ -field generated by  $\{(X_s) : 0 \leq s \leq r\}$ , it follows, for  $t \leq r$ , that

$$\mathbb{E}[Y_t | \mathcal{S}_{r,\delta}] = \mathbb{E}[m(X_t) + \sigma(X_t)\epsilon_t | \mathcal{S}_{r,\delta}] = m(X_t) + \sigma(X_t)\mathbb{E}[\epsilon_t] = m(X_t),$$

almost surely.

**Remark 2.3** For notational convenience, we have chosen the same bandwidth sequence for all margins. This assumption can be dropped easily. If one wants to make use of the vector bandwidths (see, in particular, Chapter 12 of Devroye and Lugosi [24]). With obvious changes in the notation, our results and their proofs remain true when  $h_T$  is replaced by a vector bandwidth  $\mathbf{h}_T = (h_T^{(1)}, \dots, h_T^{(d)})$ , where  $\min h_T^{(i)} > 0$ . In this situation we set  $h_T = \prod_{i=1}^d h_T^{(i)}$ , and for any vector  $\mathbf{v} = (v_1, \dots, v_d)$  we replace  $\mathbf{v}/h$  by  $(v_1/h_T^{(1)}, \dots, v_d/h_T^{(d)})$ . For a better understanding we will use real-valued bandwidths throughout the text.

### 2.2 Theoretical properties

Below, we write  $Z \stackrel{D}{=} \mathcal{N}(\mu, \sigma^2)$  whenever the random variable  $Z$  follows a normal law with expectation  $\mu$  and variance matrix  $\sigma^2$ ,  $\xrightarrow{D}$  denotes the convergence in distribution and  $\xrightarrow{\mathbb{P}}$  the convergence in probability.

#### 2.2.1 Consistency

The following theorem gives the almost sure consistency result.

**Theorem 2.4** *Under the hypotheses (A.1)–(A.4) and (A.6), for any  $n$  large enough, we have*

$$\|\widehat{\Theta}_T - \Theta\| = O(h_T^\beta) + O\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right), \text{ a.s.}$$

The proof of Theorem 2.4 is postponed to the Sect. 4.

#### 2.2.2 Asymptotic normality

To establish the asymptotic normality of  $\widehat{\Theta}_T$ , observing the statement (2.4), we have to prove that the numerator suitably normalised is asymptotically normally distributed and that the denominator converges in probability to  $m_T^{(2)}(\widehat{\Theta}_T, \psi)$ . Let  $G$  be the  $d \times d$  matrix defined by, for  $i, j = 1, \dots, d$ ,

$$G_{i,j} = \int_{\mathbb{R}^d} \frac{\partial}{\partial u_i} K(\mathbf{u}) \frac{\partial}{\partial u_j} K(\mathbf{u}) d\mathbf{u}.$$

Let us introduce the matrix  $V(\Theta, \psi)$ , the  $d \times d$  matrix defined by, for  $i, j = 1, \dots, d$

$$V_{i,j}(\Theta, \psi) = \frac{\mathbb{E}(|\psi^2(\mathbf{Y})| | \mathbf{X} = \Theta)}{f(\Theta)} G_{i,j}$$

The main result to be proved here may now be stated precisely as follows.

**Theorem 2.5** 1. *Under the assumptions (A.1), (A.3)(i)–(ii), (A.4) and (A.6), for any  $n$  large enough, we have*

$$\sqrt{Th_T^{d+1}} \widehat{m}_T^{(1)}(\Theta, \psi) \xrightarrow{D} N(0, V(\Theta, \psi)).$$

2. *If the assumptions (A.1)–(A.5), (A.6)(i) and (A.7) are fulfilled, we have, as  $T \rightarrow \infty$ ,  $\widehat{m}_T^{(2)}(\cdot, \psi)$  converges uniformly to  $m^{(2)}(\cdot, \psi)$  on the compact set  $\mathfrak{C}$ . Then, for any  $n$  large enough, we have*

$$\sqrt{Th_T^{d+1}} (\widehat{\Theta}_T - \Theta) \xrightarrow{D} N(0, [m^{(2)}(\Theta, \psi)]^{-1} V(\Theta, \psi) [m^{(2)}(\Theta, \psi)]^{-1}). \tag{2.6}$$

The proof of Theorem 2.5 is postponed to the Sect. 4.

**Confidence set**

The asymptotic variance in the central limit theorem depends on the unknown functions, which should be estimated in practice. Let us introduce the matrix  $\widehat{V}(\widehat{\Theta}_T, \psi)$  an estimate of  $V(\Theta, \psi)$ , that is a  $d \times d$  matrix defined for  $i, j = 1, \dots, n$ , by

$$\widehat{V}_{i,j}(\Theta, \psi) = \frac{\widehat{m}_T(\widehat{\Theta}_T, \psi^2)}{f_T(\widehat{\Theta}_T)} G_{i,j}.$$

The asymptotic variance is estimated by

$$[\widehat{m}_T^{(2)}(\widehat{\Theta}_T, \psi)]^{-1} \widehat{V}(\widehat{\Theta}_T, \psi) [\widehat{m}_T^{(2)}(\widehat{\Theta}_T, \psi)]^{-1}.$$

Furthermore, from (2.6), the approximate confidence region of  $\Theta$  can be obtained as

$$\Theta \in \left[ \widehat{\Theta}_T \pm c_\alpha \frac{[\widehat{m}_T^{(2)}(\widehat{\Theta}_T, \psi)]^{-1} \widehat{V}(\widehat{\Theta}_T, \psi) [\widehat{m}_T^{(2)}(\widehat{\Theta}_T, \psi)]^{-1}^{1/2}}{\sqrt{Th_T^{d+1}}} \right],$$

where  $c_\alpha$ , denotes the  $(1 - \alpha)$ -quantile of the multivariate normal distribution. Note that  $c_\alpha$  is not unique since  $\Theta$  is assumed to be a vector. Sinotina and Vogel [69] used a different approach to construct confidence sets derived as suitable neighbourhoods for maximum points of a regression estimator. The approach relies on the concentration-of-measure inequalities for the regression estimators.

**Remark 2.6** It can be observed that our proofs constitute a generalisation of those used in the kernel density mode. Hence, one can obtain easily the corresponding results for the mode density estimators as a particular case of our setting. More precisely, one can consider the kernel estimator of the conditional density of  $Y$  given  $X = x$ , defined by

$$\widehat{g}_T(y | x) := \frac{\frac{1}{Th_T^d} \int_0^T \mathbf{K} \left( \frac{y - Y_t}{\check{h}_T} \right) K \left( \frac{x - X_t}{h_T} \right) dt}{\frac{1}{Th_T^d} \int_0^T K \left( \frac{x - X_t}{h_T} \right) dt}, \text{ for } \frac{1}{Th_T^d} \int_0^T K \left( \frac{x - X_t}{h_T} \right) dt \neq 0,$$

where  $\mathbf{K}(\cdot)$  is a kernel,  $\check{h}_T$  is a positive sequence of real numbers tending to 0 at a specific rate. We refer to Bouzebda et al. [8] for more details about the framework of functional ergodic discrete time processes.

**Remark 2.7** Chen et al. [17] considered that the conditional (or local) mode set at  $x$  is defined as

$$M(x) = \left\{ y : \frac{\partial}{\partial y} p(y | x) = 0, \frac{\partial^2}{\partial y^2} p(y | x) < 0 \right\}, \tag{2.7}$$

where  $p(y | x) = p(x, y)/f(x)$  is the conditional density of  $Y$  given  $X = x$ . As a simplification, the set  $M(x)$  can be expressed in terms of the joint density as:

$$M(x) = \left\{ y : \frac{\partial}{\partial y} p(x, y) = 0, \frac{\partial^2}{\partial y^2} p(x, y) < 0 \right\}. \quad (2.8)$$

At each  $x$ , the local mode set  $M(x)$  may consist of several points, and so  $M(x)$  is in general a multivalued function. Under appropriate conditions, as we will show, these modes change smoothly as  $x$  changes. Thus, local modes behave like a collection of surfaces called *modal manifolds* in Chen et al. [17]. In our setting, we have considered the extension of the work of Ziegler [79] to the multivariate ergodic setting. The approaches are different and the extension of Chen et al. [17] to the ergodic setting is of interest. The proof of such a statement, however, should require a different methodology than that used in the present paper, and we leave this problem open for future research.

**Remark 2.8** In continuous time, data are often collected by using a sampling scheme. Several discretisation schemes have been proposed throughout the literature including deterministic and randomised sampling. The interested reader is referred to Masry [56], Prakasa Rao [62,63], Bosq [7] and Blanke and Pumo [6]. To simplify the idea, we consider the density estimator of  $f(\cdot)$  based on  $\{\mathbf{X}_t : t \in [0, T]\}$  and let  $\{\mathbf{X}(t_k) : k = 1, \dots, n\}$  be its sampled discrete sequence. The sampled estimator of the density  $f(\cdot)$  is then

$$f_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_{t_i}}{h_n}\right).$$

As in Masry [56], we only recall two cases of designs: irregular sampling and random sampling.

**Deterministic sampling.** Consider the case where the instants  $(t_k)_{1 \leq k \leq n}$  are deterministic irregularly spaced with

$$\inf_{1 \leq k \leq n} |t_{j+1} - t_j| = \frac{1}{\tau},$$

for some  $\tau > 0$ . For  $1 \leq k \leq n$ , consider  $\mathcal{G}_k := \sigma(X(t_k))$  the  $\sigma$ -field generated by  $\{\mathbf{X}_s : 0 \leq s \leq t_k\}$ . Obviously,  $(\mathcal{G}_k)_{1 \leq k \leq n}$  is an increasing family of  $\sigma$ -fields.

**Random sampling.** Assume that the instants  $(t_k)_{1 \leq k \leq n}$  form a sequence of uniform random variables in the interval  $[0, T]$  independent of the process  $\{\mathbf{X}_t : t \in [0, T]\}$ . Define

$$0 \leq \tau_1 < \dots < \tau_n \leq T$$

as the associated order statistics. Notice that  $(\tau_k)_{1 \leq k \leq n}$  are the process observation points. Obviously, the spacings

between these points are all positive. As a consequence, taking  $\mathcal{G}_k := \sigma(X(t_k))$  the  $\sigma$ -field generated by  $\{\mathbf{X}_s : 0 \leq s \leq \tau_k\}$ , it follows that  $(\mathcal{G}_k)_{1 \leq k \leq n}$  is a sequence of increasing  $\sigma$ -fields.

We would like here to mention that the penalisation procedure for the choice of the mesh  $\delta$  of the observations gives an optimal rate of convergence as demonstrated in Comte and Merlevède [19], we leave this problem open for future research in the framework of ergodic processes.

### 3 Concluding remarks

In the present paper, we are mainly concerned with the nonparametric regression model, where the regression function  $m(\cdot, \psi)$  is given by  $m(\mathbf{x}, \psi) = \mathbb{E}(\psi(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x})$ . For a measurable function  $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ , estimation of the location  $\Theta$  (mode) of a unique maximum of  $m(\cdot, \psi)$  by the location  $\hat{\Theta}_T$  of a maximum of the Nadaraya–Watson kernel estimator  $\hat{m}_T(\cdot, \psi)$  for the curve  $m(\cdot, \psi)$  is considered. Within this context, we obtain consistency and asymptotic normality results for  $\hat{\Theta}_T$  under mild local smoothness assumptions on  $m(\cdot, \psi)$  and the design density of  $\mathbf{X}$ . It is worth noticing that the ergodic framework covers and completes various situations compared to the mixing case and is more convenient to use in practice, in this sense our work extends the already existing research in the literature. We have illustrated how to use our results to construct the confidence set for the mode  $\Theta$ . In a future research one could consider the same estimation problem for stationary and ergodic discrete time processes in the case of censored data. It will be of interest to relax the stationarity to the local stationarity and establish similar results to those presented in this work, which requires a different mathematical methodology than the one used in this document. We leave this problem open for further investigation.

### 4 Proofs

This section is devoted to the proofs of our results. The previously defined notation continues to be used in what follows.

From the definition of  $\hat{\Theta}_T$  in (1.2) and  $\Theta$ , we have

$$\begin{aligned}
 |m(\hat{\Theta}_T, \psi) - m(\Theta, \psi)| &\leq |\hat{m}_T(\hat{\Theta}_T, \psi) - m(\hat{\Theta}_T, \psi)| \\
 &\quad + |\hat{m}_T(\hat{\Theta}_T, \psi) - m(\Theta, \psi)| \\
 &\leq \sup_{\mathbf{x} \in \mathcal{C}} |\hat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| \\
 &\quad + r + \left| \sup_{\mathbf{x} \in \mathcal{C}} \hat{m}_T(\mathbf{x}, \psi) - \sup_{\mathbf{x} \in \mathcal{C}} m(\mathbf{x}, \psi) \right| \\
 &\leq 2 \sup_{\mathbf{x} \in \mathcal{C}} |\hat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)|. \tag{4.1}
 \end{aligned}$$

Consider the following decomposition

$$Q_T(\mathbf{x}) := (\Psi_T(\mathbf{x}, \psi) - \bar{\Psi}_T(\mathbf{x}, \psi)) - m(\mathbf{x}, \psi)(f_T(\mathbf{x}) - \bar{f}_T(\mathbf{x})), \quad (4.2)$$

$$R_T(\mathbf{x}, \psi) := -B_T(\mathbf{x}, \psi)(f_T(\mathbf{x}) - \bar{f}_T(\mathbf{x})), \quad (4.3)$$

$$B_T(\mathbf{x}, \psi) := \frac{\bar{\Psi}_T(\mathbf{x}, \psi)}{\bar{f}_T(\mathbf{x})} - m(\mathbf{x}, \psi), \quad (4.4)$$

$$\hat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi) = B_T(\mathbf{x}, \psi) + \frac{Q_T(\mathbf{x}, \psi) + R_T(\mathbf{x}, \psi)}{f(\mathbf{x})} \frac{f(\mathbf{x})}{\bar{f}_T(\mathbf{x})}, \quad (4.5)$$

where

$$\begin{aligned} \bar{f}_T(\mathbf{x}) &= \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt, \\ \Psi_T(\mathbf{x}, \psi) &= \frac{1}{Th_T^d} \int_0^T \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt, \end{aligned}$$

and

$$\bar{\Psi}_T(\mathbf{x}, \psi) = \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt.$$

The following simple lemmas will play an instrumental role in the sequel.

**Lemma 4.1** *Let  $(Z_n)_{n \geq 1}$  be a sequence of real martingale differences with respect to the sequence of  $\sigma$ -fields  $(\mathcal{F}_n = \sigma(Z_1, \dots, Z_n))_{n \geq 1}$ , where  $\mathcal{F}_n$  is the  $\sigma$ -field generated by the random variables  $Z_1, \dots, Z_n$ . Set*

$$S_n = \sum_{i=1}^n Z_i.$$

For any  $p \geq 2$  and any  $n \geq 1$ , assume that there exist some nonnegative constants  $C$  and  $d_n$  such that

$$\mathbb{E} [Z_n^p \mid \mathcal{F}_{n-1}] \leq C^{p-1} p! d_n^2, \quad \text{almost sure.}$$

Then, for any  $\epsilon > 0$ , we have

$$\mathbb{P} (|S_n| > \epsilon) \leq 2 \exp \left\{ -\frac{\epsilon^2}{2(D_n + C\epsilon)} \right\}.$$

where

$$D_n = \sum_{i=1}^n d_i^2.$$



**Lemma 4.2** *Let  $\Lambda \times \Lambda'$  be an index set and for each  $(\eta, \eta') \in \Lambda \times \Lambda'$ , let  $\{Z_i(\eta, \eta'), i \geq 1\}$ , be a sequence of a martingale difference such that  $|Z_i(\eta, \eta')| \leq B$  a.s. then, for all  $\epsilon > 0$  and all sufficiently large  $n$ , we have*

$$P \left\{ \left| \sum_{i=1}^n Z_i(\eta, \eta') \right| > \epsilon \right\} \leq 2 \exp \left\{ -\frac{\epsilon^2}{2nB^2} \right\}.$$

The following proposition describes the almost sure consistency of  $\widehat{m}_T(\mathbf{x}, \psi)$  with rate.

**Proposition 4.3** *Under assumptions (A.1)–(A.4) and (A.6), we have*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| = O(h_T^\beta) + O \left( \left( \frac{\log T}{Th_T^d} \right)^{1/2} \right), \quad a.s. \quad (4.6)$$

**Proof of Proposition 4.3.**

Making use of conditions (A.2) and (A.3), we infer readily that

$$\begin{aligned} & \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| \\ &= \sup_{\mathbf{x} \in \mathcal{C}} \left| B_T(\mathbf{x}, \psi) + \frac{Q_T(\mathbf{x}, \psi) + R_T(\mathbf{x}, \psi)}{f(\mathbf{x})} \frac{f(\mathbf{x})}{f_T(\mathbf{x})} \right| \\ &\leq \sup_{\mathbf{x} \in \mathcal{C}} |B_T(\mathbf{x}, \psi)| + \frac{1}{\lambda} \sup_{\mathbf{x} \in \mathcal{C}} \left| \frac{Q_T(\mathbf{x}, \psi) + R_T(\mathbf{x}, \psi)}{\frac{f_T(\mathbf{x})}{f(\mathbf{x})}} \right|. \end{aligned} \quad (4.7)$$

**Lemma 4.4** *Didi and Louani [26] Let  $(\mathbf{X}_i)_{i \geq 0}$  be a strictly stationary and ergodic process, under (A.1) and (A.4), we have then*

$$\sup_{\mathbf{x} \in \mathcal{C}} \left| \frac{f_T(\mathbf{x})}{f(\mathbf{x})} - 1 \right| = o_{a.s.}(1), \quad \text{as } T \rightarrow \infty. \quad (4.8)$$

**Proof of Lemma 4.4**

Notice that we have the following decomposition

$$\begin{aligned} \frac{f_T(\mathbf{x}) - f(\mathbf{x})}{f(\mathbf{x})} &= \frac{f_T(\mathbf{x}) - \bar{f}_T(\mathbf{x}) + \bar{f}_T(\mathbf{x}) - f(\mathbf{x})}{f(\mathbf{x})} \\ &= \frac{1}{f(\mathbf{x})} \{F_{1,T}(\mathbf{x}) + F_{2,T}(\mathbf{x})\}, \end{aligned} \quad (4.9)$$

where

$$\bar{f}_T(\mathbf{x}) = \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt.$$

Two terms to be investigated, we first take a closer look to the the second term  $F_{2,T}(\mathbf{x})$ . We have

$$\begin{aligned} F_{2,T}(\mathbf{x}) &= \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt - f(\mathbf{x}) \\ &= \frac{1}{Th_T^d} \int_0^T \int_{\mathbb{R}^d} K \left( \frac{\mathbf{x} - \mathbf{u}}{h_T} \right) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) d\mathbf{u} dt - f(\mathbf{x}) \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} K(\mathbf{r}) f^{\mathcal{F}_{t-\delta}}(\mathbf{x} - h_T \mathbf{r}) d\mathbf{r} dt - f(\mathbf{x}). \end{aligned}$$

Taylor expansion of  $f^{\mathcal{F}_{t-\delta}}(\mathbf{x} - h_T \mathbf{r})$  in neighbourhood of  $\mathbf{x}$  with assumption (A.4)(i), yields

$$f^{\mathcal{F}_{t-\delta}}(\mathbf{x} - h_T \mathbf{r}) = f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) + h_T \nabla f^{\mathcal{F}_{t-\delta}}(\mathbf{x}^*),$$

where  $\mathbf{x}^*$  is between  $\mathbf{x}$  and  $\mathbf{x} - a_T \mathbf{r}$ . It follows from assumption (A.4)(i) that

$$\left| f^{\mathcal{F}_{t-\delta}}(\mathbf{x} - h_T \mathbf{r}) - f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) \right| \leq Ch_T \|\mathbf{r}\|.$$

Making use of assumptions (A.1)(iii) and (A.4)(ii), it follows that

$$\begin{aligned} F_{2,T}(\mathbf{x}) &= Ch_T \int_{\mathbb{R}^d} \|\mathbf{r}\| K(\mathbf{r}) d\mathbf{r} + \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt - f(\mathbf{x}) \\ &= o(1), \quad \text{a.s.} \end{aligned} \tag{4.10}$$

Now, we will focus on the first term of decomposition (4.9),  $G_{1,T}(\mathbf{x})$ , it is clear that

$$\begin{aligned} F_{1,T}(\mathbf{x}) &= \frac{1}{Th_T^d} \int_0^T \left( K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] \right) dt \\ &= \frac{1}{Th_T^d} \sum_{k=1}^n Z_{T,k}(\mathbf{x}), \end{aligned}$$

where

$$T = n\delta, T_k = k\delta$$

and

$$Z_{T,k}(\mathbf{x}) = \int_{T_{k-1}}^{T_k} \left( K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{G}_{t-\delta} \right] \right) dt.$$

We observe that the sequence  $\{Z_{T,k}(\mathbf{x})\}, k = 1, \dots, n$ , is a sequence of martingale differences with respect to the  $\sigma$ -field

$$\mathcal{F}_{k-1} = \sigma(X_s : 0 \leq s < T_{k-1}).$$

Under assumption (A.1), the kernel  $K(\cdot)$  is a compactly supported probability function, then we obtain

$$\begin{aligned} |Z_{T,k}(\mathbf{x})| &\leq \int_{T_{k-1}}^{T_k} \left| K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right| dt \\ &\leq 2\delta \sup_{\mathbf{x} \in \mathcal{C}} |K(\mathbf{x})| \\ &= 2\delta \tilde{K}, \end{aligned}$$

where

$$\tilde{K} = \sup_{\mathbf{x} \in \mathcal{C}} |K_1(\mathbf{x})|.$$

Now, for any  $\epsilon > 0$ , making use of Lemma 4.2 we obtain

$$\begin{aligned} \mathbb{P} \left\{ \left| \sum_{k=1}^n Z_{T,k}(\mathbf{x}) \right| > \epsilon(T h_T^d) \right\} &\leq 2 \exp \left\{ -\frac{\epsilon^2 (T h_T^d)^2}{8n\delta^2 \tilde{K}^2} \right\} \\ &= 2 \exp \left\{ -\frac{\epsilon^2 T h_T^{2d}}{8\delta \tilde{K}^2} \right\}. \end{aligned}$$

The right-hand side of the last inequality is the general term of a convergent series, hence, for sufficiently large  $T$  we conclude by Borel-Cantelli lemma that

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \left| \sum_{k=1}^n Z_{T,k}(\mathbf{x}) \right| > \epsilon(T h_T^d) \right\} < \infty,$$

which means that

$$F_{1,T}(\mathbf{x}) = 0, \quad \text{a.s.} \tag{4.11}$$

The proof is achieved by combining the statements (4.10) and (4.11). □

The following lemma gives the rate of convergence of  $f_T(\mathbf{x})$  over a compact set  $\mathcal{C}$ .

**Lemma 4.5** Didi and Louani [26] *Under assumption (A.1), we have*

$$\sup_{\mathbf{x} \in \mathcal{C}} |f_T(\mathbf{x}) - \tilde{f}_T(\mathbf{x})| = O \left( \left( \frac{\log T}{Th_T^d} \right)^{1/2} \right), \quad a.s. \tag{4.12}$$

**Proof of Lemma 4.5**

We refer to the Theorem 1 of Didi and Louani [26]. As in the proof of (4.11) in Lemma 4.4 we obtain the result by using Lemma 4.1 instead of Lemma 4.2.  $\square$

In order to complete the proof of Proposition 4.3, we will will show Lemma 4.6 and Lemma 4.7 given hereafter.

**Lemma 4.6** *If hypothesis (A.1)(i), (A.3), (A.4)(i), (A.6)(ii)-(iii) are fulfilled, we have*

$$\sup_{\mathbf{x} \in \mathcal{C}} |\Psi_T(\mathbf{x}, \psi) - \bar{\Psi}_T(\mathbf{x}, \psi)| = O \left( \left( \frac{\log T}{Th_T^d} \right)^{1/2} \right), \quad a.s. \tag{4.13}$$

**Proof of Lemma 4.6**

For  $k = 1, \dots, l$ , let  $\mathbf{x}_k \in \mathcal{C}$ . Consider a covering of the compact set  $\mathcal{C}$  by a finite number  $l$  of spheres  $\mathcal{S}_k$  centered upon by  $\mathbf{x}_k$ , with radius

$$r = h_T^{d+q+1} / T,$$

we have that

$$\mathcal{C} \subset \bigcup_{k=1}^l \mathcal{S}_k.$$

Then we have

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{C}} |\Psi_T(\mathbf{x}, \psi) - \bar{\Psi}_T(\mathbf{x}, \psi)| &\leq \max_{1 \leq k \leq l} \sup_{\mathbf{x} \in \mathcal{S}_k} |\Psi_T(\mathbf{x}, \psi) - \Psi_T(\mathbf{x}_k, \psi)| \\ &\quad + \max_{1 \leq k \leq l} |\Psi_T(\mathbf{x}_k, \psi) - \bar{\Psi}_T(\mathbf{x}_k, \psi)| \\ &\quad + \max_{1 \leq k \leq l} \sup_{\mathbf{x} \in \mathcal{S}_k} |\bar{\Psi}_T(\mathbf{x}_k, \psi) - \bar{\Psi}_T(\mathbf{x}, \psi)| \\ &= \Psi_{1,T}(\mathbf{x}, \mathbf{x}_k) + \Psi_{2,T}(\mathbf{x}_k) + \Psi_{3,T}(\mathbf{x}, \mathbf{x}_k). \end{aligned}$$

Making use of the Cauchy-Schwarz inequality together with assumption (A.1)(i), (A.3), (A.6)(iv) and Lemma 4.4, we readily obtain

$$|\Psi_T(\mathbf{x}, \psi) - \Psi_T(\mathbf{x}_k, \psi)|$$

$$\begin{aligned}
 &\leq \frac{1}{Th_T^d f_T(\mathbf{x})} \int_0^T |\psi(\mathbf{Y}_t)| \left| K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) - K\left(\frac{\mathbf{x}_k - \mathbf{X}_t}{h_T}\right) \right| dt \\
 &\leq \frac{1}{Th_T^d f_T(\mathbf{x})} \left( \int_0^T \psi^2(\mathbf{Y}_t) dt \right)^{1/2} \left( \int_0^T \left( K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) - K\left(\frac{\mathbf{x}_k - \mathbf{X}_t}{h_T}\right) \right)^2 dt \right)^{1/2} \\
 &\leq \frac{1}{\sqrt{T} h_T^d f_T(\mathbf{x})} \left( \frac{1}{T} \int_0^T \psi^2(\mathbf{Y}_t) dt \right)^{1/2} \frac{\sqrt{T} C_K}{h_T} \|\mathbf{x} - \mathbf{x}_k\| \\
 &\leq \frac{C_K}{h_T^{d+1} f(\mathbf{x})} \|\mathbf{x} - \mathbf{x}_k\| O_{a.s.}(\mathbb{E}[\psi^2(\mathbf{Y}_0)]) \\
 &\leq \frac{C_K}{h_T^{d+1} \lambda} \|\mathbf{x} - \mathbf{x}_k\| O_{a.s.}(\mathbb{E}[\psi^2(\mathbf{Y}_0)]) \\
 &\leq \frac{C_K}{h_T^{d+1} \lambda} \frac{h_T}{T} O_{a.s.}(\mathbb{E}[\psi^2(\mathbf{Y}_0)]) \\
 &= \frac{C_K}{Th_T^d \lambda} O_{a.s.}(\mathbb{E}[\psi^2(\mathbf{Y}_0)]). \tag{4.14}
 \end{aligned}$$

Considering the right hand side of statement (4.14) together with the fact that  $\mathbb{E}[\psi^2(\mathbf{Y}_0)] < \infty$ , we obtain for

$$\epsilon_T = \epsilon_0 \left( \log T / Th_T^d \right)^{1/2},$$

that

$$\epsilon_T^{-1} \Psi_{1,T}(\mathbf{x}, \mathbf{x}_k) = O_{a.s.} \left( \left( \frac{1}{Th_T^d \log T} \right)^{1/2} \right). \tag{4.15}$$

Making use of similar arguments as those used for  $\Psi_{1,T}(\mathbf{x}, \mathbf{x}_k)$ , we infer that

$$\begin{aligned}
 &|\tilde{\Psi}_T(\mathbf{x}) - \tilde{\Psi}_T(\mathbf{x}_k)| \\
 &\leq \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ |\psi(\mathbf{Y}_t)| \left| K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) - K\left(\frac{\mathbf{x}_k - \mathbf{X}_t}{h_T}\right) \right| \middle| \mathcal{F}_{t-\delta} \right] dt \\
 &\leq \frac{C_K}{Th_T^d} \int_0^T \mathbb{E} \left[ |\psi(\mathbf{Y}_t)| \left\| \frac{\mathbf{x} - \mathbf{x}_k}{h_T} \right\| \middle| \mathcal{F}_{t-\delta} \right] dt \\
 &\leq \left\| \frac{\mathbf{x} - \mathbf{x}_k}{h_T} \right\| \frac{C_K}{Th_T^{d+1}} \int_0^T \mathbb{E} [|\psi(\mathbf{Y}_t)| \middle| \mathcal{F}_{t-\delta}] dt \\
 &\leq \frac{h_T}{T} \frac{C_K}{h_T^{d+1}} \left( \frac{1}{T} \int_0^T \mathbb{E} [|\psi(\mathbf{Y}_t)| \middle| \mathcal{F}_{t-\delta}] dt \right) \\
 &\leq \frac{C_K}{Th_T^d} O(\mathbb{E}[|\psi(\mathbf{Y}_0)|]).
 \end{aligned}$$

Using the fact in (2.5), we get

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{E} [\psi(\mathbf{Y}_t) | \mathcal{F}_{t-\delta}] dt = \mathbb{E} [\psi(\mathbf{Y}_0)].$$

This implies that we have

$$\epsilon_T^{-1} \Psi_{3,T}(\mathbf{x}, \mathbf{x}_k) = O \left( \left( \frac{1}{Th_T^d \log T} \right)^{1/2} \right), \quad \text{a.s.} \tag{4.16}$$

Now we deal with  $\Psi_{2,T}(\mathbf{x}, y, s_k)$ . Observe that

$$\begin{aligned} \Psi_{2,T}(\mathbf{x}_k) &= \max_{1 \leq k \leq l} \left| \frac{1}{Th_T^d} \int_0^T \psi(\mathbf{Y}_t) \left( K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] \right) dt \right| \\ &= \frac{1}{Th_T^d} \max_{1 \leq k \leq l_T} \left| \sum_{j=1}^n R_{T,j}(\mathbf{x}_k) \right|, \end{aligned}$$

where

$$R_{T,j}(\mathbf{x}_k) = \int_{T_{j-i}}^{T_j} \left( K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] \right) dt,$$

where

$$T = n\delta \quad \text{and} \quad T_j = j\delta.$$

We observe that the sequence  $\{R_{T,j}(\mathbf{x}_k)\}_{0 \leq j \leq n}$  is a sequence of martingale differences adapted to the filtration

$$\mathcal{F}_{j-1} = \sigma((X_s, \mathbf{Y}_s) : 0 \leq s < T_{j-1}).$$

For  $p \geq 2$ , making use of Jensen and Minkowski’s inequalities, we get

$$\begin{aligned} & \left| \mathbb{E} \left[ R_{T,j}^p(\mathbf{x}_k) \mid \mathcal{F}_{j-2} \right] \right| \\ &= \left| \mathbb{E} \left[ \left( \int_{T_{j-i}}^{T_j} \psi(\mathbf{Y}_t) \left( K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] \right) dt \right)^p \mid \mathcal{F}_{j-2} \right] \right| \\ &\leq \int_{T_{j-i}}^{T_j} \mathbb{E} \left[ \left| K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] \right|^p \mid \mathcal{F}_{j-2} \right] dt \\ &\leq \int_{T_{j-i}}^{T_j} \left( \mathbb{E} \left[ \psi^p(\mathbf{Y}_t) K^p \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) \mid \mathcal{F}_{j-2} \right] \right)^{1/p} \end{aligned}$$

$$\begin{aligned}
 & -\mathbb{E} \left[ \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{t-\delta} \right]^p \Big| \mathcal{F}_{j-2} \right]^{1/p} \Big)^p dt \\
 & \leq \int_{T_{j-i}}^{T_j} \left( 2\mathbb{E} \left[ \psi^p(\mathbf{Y}_t) K^p \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right]^{1/p} \right)^p dt \\
 & = 2^p \int_{T_{j-i}}^{T_j} \mathbb{E} \left[ \psi^p(\mathbf{Y}_t) K^p \left( \frac{\mathbf{x}_k - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] dt, \tag{4.17}
 \end{aligned}$$

Furthermore, by assumption (A.6)(iii), we get

$$\begin{aligned}
 & \mathbb{E} \left[ \left| \psi^p(\mathbf{Y}_t) K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right| \Big| \mathcal{F}_{j-2} \right] \\
 & = \mathbb{E} \left[ \left| \mathbb{E} \left[ \psi^p(\mathbf{Y}_t) K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{S}_{t,\delta} \right] \right| \Big| \mathcal{F}_{j-2} \right] \\
 & = \mathbb{E} \left[ \left| \mathbb{E} \left[ \psi^p(\mathbf{Y}_t) \Big| \mathcal{S}_{t,\delta} \right] K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right| \Big| \mathcal{F}_{j-2} \right] \\
 & = \mathbb{E} \left[ |h_p(\mathbf{X}_t)| K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] \\
 & \leq \mathbb{E} \left[ |h_p(\mathbf{X}_t) - h_p(\mathbf{x})| K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] \\
 & \quad + \mathbb{E} \left[ |h_p(\mathbf{x})| K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] \\
 & \leq \mathbb{E} \left[ K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] \left( \sup_{\|\mathbf{x}-\mathbf{u}\| \leq \lambda h_T} |h_p(\mathbf{X}_t) - h_p(\mathbf{x})| + |h_p(\mathbf{x})| \right) \\
 & \leq \eta(\mathbf{x}) \mathbb{E} \left[ K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right],
 \end{aligned}$$

where  $\eta(\mathbf{x})$  is a constant. We infer from condition (A.4)(i) that

$$\begin{aligned}
 \mathbb{E} \left[ K^p \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \Big| \mathcal{F}_{j-2} \right] & = \int_{\mathbb{R}^{d+1}} K^p \left( \frac{\mathbf{x} - \mathbf{v}}{h_T} \right) f_T^{\mathcal{F}_{j-2}}(\mathbf{v}) d\mathbf{v} \\
 & = h_T^d \int_{\mathbb{R}^{d+1}} K_1^p(\mathbf{w}) f_T^{\mathcal{F}_{j-2}}(\mathbf{x} - a_T \mathbf{w}) d\mathbf{w} \\
 & \leq h_T^d \|K\|^p. \tag{4.18}
 \end{aligned}$$

Notice that the Eq. (4.17) can be rewritten using Eq. (4.18) as follows

$$\begin{aligned}
 \mathbb{E} \left[ R_{T,j}^p(\mathbf{x}_k) \Big| \mathcal{F}_{j-2} \right] & \leq 2^p C(\mathbf{x}) \delta h_T^d \|K_1\|^p \\
 & \leq p! C^{p-2} d_j^2, \tag{4.19}
 \end{aligned}$$

where  $C = 2 \|K\|$  and

$$d_j^2 = 2\delta C(\mathbf{x})h_T^d \|K\|^2.$$

Let

$$D_n = \sum_{j=1}^n d_j^2 = \sum_{j=1}^n 2\delta h_T^d \|K\| = O(Th_T^d).$$

An application of Lemma 4.1 and keeping in mind that  $\epsilon_T = \epsilon_0 (\log T / Th_T^d)^{1/2}$ , we get, for any  $\epsilon_0 > 0$ ,

$$\begin{aligned} & \mathbb{P} \left\{ \max_{1 \leq k \leq l} \left| \sum_{j=1}^n R_{T,j}(\mathbf{x}_k) \right| > \epsilon_T (Th_T^d) \right\} \\ & \leq \sum_{k=1}^l \mathbb{P} \left\{ \left| \sum_{j=1}^n R_{T,j}(\mathbf{x}_k) \right| > \epsilon_T (Th_T^d) \right\} \\ & \leq 2l \exp \left\{ -\frac{\epsilon_T^2 (Th_T^d)^2}{2(D_n + C(Th_T^d)\epsilon_T)} \right\} \\ & \leq 2l \exp \left\{ -\frac{\epsilon_0^2 (Th_T^d) (1/Th_T^d)}{O(Th_T^d) + 2C(Th_T^d)\epsilon_0 \left(\frac{\log T}{Th_T^d}\right)^{1/2}} \right\} \\ & \leq 2l \exp \left\{ -\frac{\epsilon_0^2 (Th_T^d)^2 \log T / Th_T^d}{O(Th_T^d) \left(1 + \epsilon_0 \left(\frac{\log T}{Th_T^d}\right)^{1/2}\right)} \right\} \\ & := 2l \exp \left\{ \log T^{-2C\epsilon_0^2 C_1} \right\} \\ & = 2l T^{-\epsilon_0^2 C_1}, \end{aligned}$$

where  $C_1$  is a positive constant. The right-hand side of the previous inequality, is the general term of convergent series, hence for  $T$  large enough, we obtain the following statement via the Borel-Cantelli lemma

$$\sum_{n=1}^{\infty} \mathbb{P} \left\{ \max_{1 \leq k \leq l} \left| \sum_{j=1}^n R_{T,j}(\mathbf{x}_k) \right| > \epsilon_T (Th_T^d)^{-1} \right\} < \infty.$$



This, in turn, implies that

$$\frac{1}{Th_T^d} \max_{1 \leq k \leq l} \left| \sum_{j=1}^n R_{T,j}(\mathbf{x}_k) \right| = O \left( \left( \frac{\log T}{Th_T^d} \right)^{1/2} \right), \quad \text{a.s.} \tag{4.20}$$

By combining (4.15), (4.16) and (4.20) with Lemma 4.4, we obtain that,

$$\sup_{\mathbf{x} \in \mathcal{C}} |\Psi_T(\mathbf{x}, \psi) - \bar{\Psi}_T(\mathbf{x}, \psi)| = O \left( \left( \frac{\log T}{Th_T^d} \right)^{1/2} \right), \quad \text{a.s.} \tag{4.21}$$

Therefore the proof is complete. □

We next evaluate the term  $B_T(\mathbf{x}, \psi)$  defined in (4.5).

**Lemma 4.7** *Under assumptions (A.1) and (A.6)(i)–(ii), we have*

$$\sup_{\mathbf{x} \in \mathcal{C}} |B_T(\mathbf{x}, \psi)| = O \left( h_T^\beta \right). \tag{4.22}$$

**Proof of Lemma 4.7**

First, we will use the notation

$$K_{h_T}(\cdot) = \frac{1}{h_T^d} K \left( \frac{\cdot}{h_T} \right).$$

We let

$$\sup_{\mathbf{x} \in \mathcal{C}} |B_T(\mathbf{x}, \psi)| = \sup_{\mathbf{x} \in \mathcal{C}} \left| \frac{B_T^*(\mathbf{x}, \psi)}{\bar{f}_T(\mathbf{x})} \right|.$$

Observe that assumption (A.6)(i) implies that

$$\begin{aligned} B_T^*(\mathbf{x}, \psi) &= \bar{\Psi}_T(\mathbf{x}, \psi) - \bar{f}_T(\mathbf{x})m(\mathbf{x}, \psi) \\ &= \frac{1}{T} \int_0^T \mathbb{E} [(\psi(\mathbf{Y}_t) - m(\mathbf{x}, \psi)) K_{h_T}(\mathbf{x} - \mathbf{X}_t) \mid \mathcal{F}_{t-\delta}] dt \\ &= \frac{1}{T} \int_0^T \mathbb{E} [K_{h_T}(\mathbf{x} - \mathbf{X}_t) \mathbb{E} [(\psi(\mathbf{Y}_t) - m(\mathbf{x}, \psi)) \mid \mathcal{S}_{T-\delta, \delta}] \mid \mathcal{F}_{t-\delta}] dt \\ &= \frac{1}{T} \int_0^T \mathbb{E} [K_{h_T}(\mathbf{x} - \mathbf{X}_t) (\mathbb{E} [\psi(\mathbf{Y}_t) \mid \mathbf{X}_t] - m(\mathbf{x}, \psi)) \mid \mathcal{F}_{t-\delta}] dt, \\ &= \frac{1}{T} \int_0^T \mathbb{E} [K_{h_T}(\mathbf{x} - \mathbf{X}_t) (m(\mathbf{X}_t, \psi) - m(\mathbf{x}, \psi)) \mid \mathcal{F}_{t-\delta}] dt. \end{aligned}$$

Under assumption (A.6)(ii), we have

$$\begin{aligned}
 |B_T^*(\mathbf{x}, \psi)| &\leq \sup_{\|\mathbf{u}-\mathbf{x}\| \leq h_T \lambda} |m(\mathbf{X}_t, \psi) - m(\mathbf{x}, \psi)| \frac{1}{T} \int_0^T \mathbb{E} [K_{h_T}(\mathbf{x} - \mathbf{X}_t) \mid \mathcal{F}_{t-\delta}] dt \\
 &\leq C_\psi \lambda^\beta h_T^\beta \frac{1}{T h_T^d} \int_0^T \mathbb{E} \left[ K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt \\
 &= C_\psi \lambda^\beta h_T^\beta \bar{f}_T(\mathbf{x}).
 \end{aligned}$$

We obtain that

$$\sup_{\mathbf{x} \in \mathcal{C}} |B_T(\mathbf{x}, \psi)| = O \left( h_T^\beta \right), \quad \text{a.s.} \tag{4.23}$$

The proof of the lemma is therefore completed. □

Recalling (4.21), the proof of Theorem 4.3 is completed by combining Lemmas 4.4, 4.5 and 4.7. □

In the following lemma, we give the almost sure convergence of  $\widehat{\Theta}_T$ .

**Lemma 4.8** *Under the hypotheses of Theorem (2.4), we have, as  $T \rightarrow \infty$ ,*

$$\|\widehat{\Theta}_T - \Theta\| \xrightarrow{a.s.} 0.$$

**Proof of Lemma 4.8**

The uniqueness hypothesis of the conditional mode of the regression gives

$$\forall \epsilon > 0, \exists \eta(\epsilon) > 0; \forall \xi : \|\Theta - \xi\| \geq \epsilon \Rightarrow |\widehat{m}_T(\Theta, \psi) - m(\xi, \psi)| \geq \eta(\epsilon). \tag{4.24}$$

Combining conditions (4.24) and (4.1), we obtain, for any fixed  $\mathbf{x} \in \mathcal{C}$  all  $\epsilon > 0$ , that there exists a  $\xi > 0$  such that

$$\mathbb{P} \{ \|\Theta_T - \Theta\| \geq \epsilon \} \leq \mathbb{P} \left\{ \sup_{\mathbf{x} \in \mathcal{C}} |m_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| \geq \xi \right\}. \tag{4.25}$$

Which gives the desired result provided that the right-hand side of Eq. (4.25) converges almost surely to zero. The proof is therefore completed by using Proposition 4.3. □

The following lemma gives the uniform convergence of  $\widehat{m}_T^{(2)}(\mathbf{x}, \psi)$  over the compact set  $\mathcal{C}$ . To simplify our reasoning, from now on, all our will be given in the univariate setting. The extension to the multivariate setting follows easily.

**Lemma 4.9** *If assumptions (A.1)(ii), (A.3), (A.4)(i), (A.5), (A.6)(i) and (A.7)(i) are fulfilled, we have, as  $T \rightarrow \infty$ ,*

$$\sup_{\mathbf{x} \in \mathcal{C}} \left\| \widehat{m}_T^{(2)}(\mathbf{x}, \psi) - m^{(2)}(\mathbf{x}, \psi) \right\| \longrightarrow 0, \quad \text{almost surely.} \tag{4.26}$$

**Proof of Lemma 4.9**

We first observe that we have

$$\begin{aligned} \widehat{m}_T^{(2)}(\mathbf{x}, \psi) &= \left( \frac{\widehat{M}_T(\mathbf{x}, \psi)}{f(\mathbf{x})} \right)^{(2)} \\ &= \frac{\left( \widehat{M}_T^{(1)}(\mathbf{x}, \psi) f(\mathbf{x}) - f^{(1)}(\mathbf{x}) \widehat{M}_T(\mathbf{x}, \psi) \right)^{(1)}}{f^2(\mathbf{x})} \\ &\quad - \frac{2f^{(1)}(\mathbf{x}) \left( \widehat{M}_T^{(1)}(\mathbf{x}, \psi) f(\mathbf{x}) - f^{(1)}(\mathbf{x}) \widehat{M}_T(\mathbf{x}, \psi) \right)}{f^3(\mathbf{x})} \\ &= \frac{\widehat{M}_T^{(2)}(\mathbf{x}, \psi)}{f(\mathbf{x})} - \frac{2f^{(1)}(\mathbf{x}) \widehat{M}_T^{(1)}(\mathbf{x}, \psi)}{f^2(\mathbf{x})} \\ &\quad + \frac{\widehat{M}_T(\mathbf{x}, \psi) (2(f^{(1)}(\mathbf{x}))^2 f(\mathbf{x}) - f^{(2)}(\mathbf{x}))}{f^3(\mathbf{x})} \\ &= \frac{1}{f(\mathbf{x})} \frac{1}{Th_T^{d+2}} \int_0^T \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt \\ &\quad - \frac{2f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \frac{1}{Th_T^{d+1}} \int_0^T \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt \\ &\quad + \left( \frac{2(f^{(1)}(\mathbf{x}))^2}{f^3(\mathbf{x})} - \frac{f^{(2)}(\mathbf{x})}{f^2(\mathbf{x})} \right) \frac{1}{Th_T^d} \int_0^T \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt. \end{aligned}$$

Let us define

$$\begin{aligned} \widetilde{m}_T^{(2)}(\mathbf{x}, \psi) &= \frac{1}{f(\mathbf{x})} \frac{1}{Th_T^{d+2}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt \\ &\quad - \frac{2f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt \\ &\quad + \left( \frac{2(f^{(1)}(\mathbf{x}))^2}{f^3(\mathbf{x})} - \frac{f^{(2)}(\mathbf{x})}{f^2(\mathbf{x})} \right) \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt. \end{aligned}$$

Consider the following decomposition

$$\begin{aligned} \widehat{m}_T^{(2)}(\mathbf{x}, \psi) - m^{(2)}(\mathbf{x}, \psi) &= \widehat{m}_T^{(2)}(\mathbf{x}, \psi) - \widetilde{m}_T^{(2)}(\mathbf{x}, \psi) \\ &\quad + \widetilde{m}_T^{(2)}(\mathbf{x}, \psi) - m^{(2)}(\mathbf{x}, \psi) \\ &= A_{T,1}(\mathbf{x}, \psi) + A_{T,2}(\mathbf{x}, \psi). \end{aligned} \tag{4.27}$$

To achieve the asymptotic uniform convergence over the compact set  $\mathcal{C}$  of the term  $A_{T,1}(\mathbf{x}, \psi)$  in the decomposition (4.27), we have to prove that

$$\sup_{\mathbf{x} \in \mathcal{C}} \left\| \frac{1}{Th_T^{d+2}} \left( \int_0^T \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt - \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \right) \right\| = o_{a.s.}(1), \quad (4.28)$$

$$\sup_{\mathbf{x} \in \mathcal{C}} \left\| \frac{1}{Th_T^{d+1}} \left( \int_0^T \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt - \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \right) \right\| = o_{a.s.}(1), \quad (4.29)$$

$$\sup_{\mathbf{x} \in \mathcal{C}} \left\| \frac{1}{Th_T^d} \left( \int_0^T \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt - \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \right) \right\| = o_{a.s.}(1), \quad (4.30)$$

Using a simple integration par parts and Lemma 4.6, we obtain proof of (4.28)–(4.30), and combining Assumptions (A.1), (A.3), (A.4), (A.6)(i)–(ii) and statements (4.28)–(4.30) we obtain

$$\sup_{\mathbf{x} \in \mathcal{C}} \|A_{T,1}(\mathbf{x}, \psi)\| = o_{a.s.}(1). \quad (4.31)$$

Remark that

$$\begin{aligned} m^{(2)}(\mathbf{x}, \psi) &= \frac{M^{(2)}(\mathbf{x}, \psi)}{f(\mathbf{x})} - \frac{2f^{(2)}(\mathbf{x})M_T^{(1)}(\mathbf{x}, \psi)}{f^2(\mathbf{x})} \\ &\quad + \frac{(2(f^{(1)}(\mathbf{x}))^2 f(\mathbf{x}) - f^{(2)}(\mathbf{x})) M(\mathbf{x}, \psi)}{f^3(\mathbf{x})}. \end{aligned} \quad (4.32)$$

We now treat the second term  $A_{T,2}(\mathbf{x}, \psi)$  in (4.27). We have

$$\begin{aligned} &A_{T,2}(\mathbf{x}, \psi) \\ &= \frac{1}{f(\mathbf{x})} \frac{1}{Th_T^{d+2}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \\ &\quad - \frac{2f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \\ &\quad + \left( \frac{2(f^{(1)}(\mathbf{x}))^2 f(\mathbf{x}) - f^{(2)}(\mathbf{x})}{f^3(\mathbf{x})} \right) \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt \\ &\quad - \frac{M^{(2)}(\mathbf{x}, \psi)}{f(\mathbf{x})} + \frac{f^{(2)}(\mathbf{x})M_T^{(1)}(\mathbf{x}, \psi)}{f^2(\mathbf{x})} - \frac{(2(f^{(1)}(\mathbf{x}))^2 f(\mathbf{x}) - f^{(2)}(\mathbf{x})) M(\mathbf{x}, \psi)}{f^3(\mathbf{x})} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{f(\mathbf{x})} \left( \frac{1}{Th_T^{d+2}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M^{(2)}(\mathbf{x}, \psi) \right) \\
 &\quad - \frac{2f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \left( \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M_T^{(1)}(\mathbf{x}, \psi) \right) \\
 &\quad + \left( \frac{2(f^{(1)}(\mathbf{x}))^2 f(\mathbf{x}) - f^{(2)}(\mathbf{x})}{f^3(\mathbf{x})} \right) \left( \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M(\mathbf{x}, \psi) \right).
 \end{aligned}$$

To achieve the asymptotic uniform convergence over the compact set  $\mathfrak{C}$  of the term  $A_{T,2}(\mathbf{x}, \psi)$ , we have to show the following statements

$$\sup_{\mathbf{x} \in \mathfrak{C}} \left| \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M(\mathbf{x}, \psi) \right| = o_{a.s}(1), \tag{4.33}$$

$$\sup_{\mathbf{x} \in \mathfrak{C}} \left\| \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M_T^{(1)}(\mathbf{x}, \psi) \right\| = o_{a.s}(1), \tag{4.34}$$

$$\sup_{\mathbf{x} \in \mathfrak{C}} \left\| \frac{1}{Th_T^{d+2}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(2)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M^{(2)}(\mathbf{x}, \psi) \right\| = o_{a.s}(1). \tag{4.35}$$

Observe that statement (4.33) may be rewritten as follows

$$\begin{aligned}
 &\sup_{\mathbf{x} \in \mathfrak{C}} \left| \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - M(\mathbf{x}, \psi) \right| \\
 &= \sup_{\mathbf{x} \in \mathfrak{C}} \left| \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt - m(\mathbf{x}, \psi) f(\mathbf{x}) \right|.
 \end{aligned}$$

The desired result can be obtained in similar way to the statement (4.23). We have

$$\begin{aligned}
 M^{(1)}(\mathbf{x}, \psi) &= (m(\mathbf{x}, \psi) f(\mathbf{x}))^{(1)} \\
 &= m^{(1)}(\mathbf{x}, \psi) f(\mathbf{x}) + m(\mathbf{x}, \psi) f^{(1)}(\mathbf{x}).
 \end{aligned} \tag{4.36}$$

On the other hand, under assumption (A.6)(i), we have

$$\begin{aligned}
 &\frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt \\
 &= \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{S}_{t,\delta} \right] \middle| \mathcal{F}_{t-\delta} \right] dt \\
 &= \frac{1}{Th_T^{d+1}} \int_0^T \mathbb{E} \left[ m(\mathbf{X}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt
 \end{aligned}$$

$$= \frac{1}{Th_T^{d+1}} \int_0^T \int_{\mathbb{R}^d} m(\mathbf{u}) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) d\mathbf{u} dt.$$

By integrating by parts we infer that

$$U(\mathbf{u}) = m(\mathbf{u}) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) \rightarrow U^{(1)}(\mathbf{u}) = m^{(1)}(\mathbf{u}) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) + m(\mathbf{u}) \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{u}),$$

$$V^{(1)}(\mathbf{u}) = \frac{1}{h_T} K^{(1)}\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right) \rightarrow V(\mathbf{u}) = -K\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right).$$

By integrating by parts and the change of variable  $\mathbf{y} = \frac{\mathbf{x} - \mathbf{u}}{h_T}$  combined with Taylor expansions of order one, under assumptions (A.4)(i), (A.5) and (A.7)(i), we readily obtain

$$\begin{aligned} & \frac{1}{Th_T^d} \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) | \mathcal{F}_{t-\delta} \right] dt \\ &= \frac{1}{Th_T^d} \int_0^T \left( \int_{\mathbb{R}^d} \left[ m(\mathbf{u}) K\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) \right] d\mathbf{u} \right. \\ & \quad \left. + \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right) \left( m^{(1)}(\mathbf{u}) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) + m(\mathbf{u}) \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{u}) \right) d\mathbf{u} \right) dt \\ &= \frac{1}{Th_T^d} \int_0^T \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x} - \mathbf{u}}{h_T}\right) \left( m^{(1)}(\mathbf{u}) f^{\mathcal{F}_{t-\delta}}(\mathbf{u}) + m(\mathbf{u}) \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{u}) \right) d\mathbf{u} dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} K(\mathbf{y}) \left( m^{(1)}(\mathbf{x}) + O(h_T) \right) \left( f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) + O(h_T) \right) d\mathbf{y} dt \\ & \quad + \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} K(\mathbf{y}) \left( m(\mathbf{x}) + O(h_T) \right) \left( \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{x}) + O(h_T) \right) d\mathbf{y} dt \\ &= \left( m^{(1)}(\mathbf{x}) \left( \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt \right) \right. \\ & \quad \left. + m(\mathbf{x}) \left( \frac{1}{T} \int_0^T \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{x}) dt \right) \right) \int_{\mathbb{R}^d} K(\mathbf{y}) d\mathbf{y} + O(h_T) \\ &= m^{(1)}(\mathbf{x}) \left( \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt \right) + m(\mathbf{x}) \left( \frac{1}{T} \int_0^T \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{x}) dt \right) + O(h_T) \end{aligned} \tag{4.37}$$

where

$$g_i^{\mathcal{F}_{t-\delta}}(\mathbf{x}) = \left(f^{\mathcal{F}_{t-\delta}}\right)^{(1)}(\mathbf{x})$$

is a stationary and ergodic process. Therefore, one have (see Krengel [47], Theorem 4.4),

$$\lim_{T \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T g_i^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt - \mathbb{E} \left[ g_1^{\mathcal{F}_{t-\delta}}(\mathbf{x}) \right] \right| = 0, \tag{4.38}$$

where

$$\mathbb{E} \left[ g_1^{\mathcal{F}-\delta}(\mathbf{x}) \right] = f^{(1)}(\mathbf{x}).$$

By combining the statements (4.37) and (4.38) we conclude proof of (4.34). Moreover, statement (4.35) may be proved in the same way as in statement (4.34), keeping in mind that

$$\begin{aligned} M^{(2)}(\mathbf{x}, \psi) &= \left( m^{(1)}(\mathbf{x}, \psi) f(\mathbf{x}) + m(\mathbf{x}, \psi) f^{(1)}(\mathbf{x}) \right)^{(2)} \\ &= m^{(2)}(\mathbf{x}, \psi) f(\mathbf{x}) + 2m^{(1)}(\mathbf{x}, \psi) f^{(1)}(\mathbf{x}) \\ &\quad + m(\mathbf{x}, \psi) f^{(2)}(\mathbf{x}). \end{aligned}$$

By applying integration by parts twice, we obtain (4.34). Combining statement (4.33), (4.34) and (4.35), yields to

$$\sup_{\mathbf{x} \in \mathcal{C}} |A_{T,2}(\mathbf{x}, \psi)| = o_{a.s}(1). \tag{4.39}$$

Statements (4.31) and (4.39) complete the proof of Lemma 4.9. □

**Proof of Theorem 2.4**

Under assumption (A.3)(ii) and using Taylor expansion of  $m(\Theta_T, \psi)$  around  $\Theta$  we obtain

$$m(\Theta_T, \psi) = m(\Theta, \psi) + (\Theta_T - \Theta)m^{(2)}(\Theta_T^*, \psi)(\Theta_T - \Theta), \tag{4.40}$$

where  $\Theta_T^*$  is between  $\Theta_T$  and  $\Theta$ , it follows from equations (4.1) and (4.40) that

$$\|\Theta_T - \Theta\|^2 \left\| m^{(2)}(\Theta_T^*, \psi) \right\| = O \left( \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| \right).$$

Using Lemma 4.8 and condition (A.7)(ii), one obtains

$$\lim_{T \rightarrow \infty} \left\| m^{(2)}(\Theta_T^*, \psi) \right\| = \left\| m^{(2)}(\Theta, \psi) \right\| \neq 0.$$

Therefore,

$$\|\Theta_T - \Theta\|^2 = O \left( \sup_{\mathbf{x} \in \mathcal{C}} |\widehat{m}_T(\mathbf{x}, \psi) - m(\mathbf{x}, \psi)| \right), \tag{4.41}$$

which is enough, while considering Proposition 4.3, to complete the proof. □

**Proof of Theorem 2.5**

By using formula (2.4), we readily obtain

$$\sqrt{Th_T^{d+2}} \widehat{m}_T^{(1)}(\Theta, \psi) = \sqrt{Th_T^{d+2}} (\widehat{\Theta}_T - \Theta) \widehat{m}_T^{(2)}(\widehat{\Theta}_T^*, \psi),$$

where  $\widehat{\Theta}_T^*$  is a random variable taking his values between  $\Theta$  and  $\widehat{\Theta}_T$ . From the hypothesis made on  $\widehat{\Theta}_T$  it results that  $\widehat{\Theta}_T^*$  also converges a.s. towards  $\Theta$ . The continuity of function  $m^{(2)}(\cdot, \psi)$  leads

$$\lim_{T \rightarrow \infty} m^{(2)}(\widehat{\Theta}_T^*, \psi) = m^{(2)}(\Theta, \psi).$$

For  $T$  large enough, we have almost surely

$$\begin{aligned} \left| \widehat{m}_T^{(2)}(\widehat{\Theta}_T^*, \psi) - m^{(2)}(\Theta, \psi) \right| &\leq \sup_{x \in \mathcal{C}} \left| \widehat{m}_T^{(2)}(\mathbf{x}, \psi) - m^{(2)}(\mathbf{x}, \psi) \right| \\ &\quad + \left| m^{(2)}(\widehat{\Theta}_T^*, \psi) - m^{(2)}(\Theta, \psi) \right|. \end{aligned}$$

The uniform convergence in probability  $\widehat{m}_T^{(2)}(\cdot, \psi)$  to  $m^{(2)}(\cdot, \psi)$  over  $\mathcal{C}$  implies the convergence of the sequence  $\widehat{m}_T^{(2)}(\widehat{\Theta}_T^*, \psi)$  in probability to the non-null real  $m^{(2)}(\Theta, \psi)$ . The conclusion results from the asymptotic normality of  $m_T^{(1)}(\Theta, \psi)$ . Since

$$\lim_{T \rightarrow \infty} f_T(\mathbf{x}) = f(\mathbf{x})$$

almost surely and uniformly on the set  $\mathcal{C}$ , refer for details [26]. Notice that we have

$$\begin{aligned} \widehat{m}_T^{(1)}(\mathbf{x}, \psi) &= \left( \frac{M_T(\mathbf{x}, \psi)}{f(\mathbf{x})} \right)^{(1)} \\ &= \frac{M_T^{(1)}(\mathbf{x}, \psi) f(\mathbf{x}) - f^{(1)}(\mathbf{x}) M_T(\mathbf{x}, \psi)}{f^2(\mathbf{x})} \\ &= \frac{1}{f^2(\mathbf{x})} \left( \frac{f(\mathbf{x})}{Th_T^{d+1}} \int_0^T \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt \right. \\ &\quad \left. - \frac{f^{(1)}(\mathbf{x})}{Th_T^d} \int_0^T \psi(\mathbf{Y}_s) K \left( \frac{\mathbf{x} - \mathbf{X}_s}{h_T} \right) ds \right), \end{aligned} \tag{4.42}$$

where

$$\widehat{m}_T^{(1)}(\mathbf{x}, \psi) = \frac{1}{Th_T^{d+1} f^2(\mathbf{x})} \left[ f(\mathbf{x}) \int_0^T \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt \right]$$



$$-h_T f^{(1)}(\mathbf{x}) \int_0^T \psi(\mathbf{Y}_t) K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt \Big],$$

and

$$\begin{aligned} \tilde{m}_T^{(1)}(\mathbf{x}, \psi) &= \frac{1}{Th_T^{d+1} f^2(\mathbf{x})} \left[ f(\mathbf{x}) \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2} \right] dt \right. \\ &\quad \left. - h_T f^{(1)}(\mathbf{x}) \int_0^T \mathbb{E} \left[ \psi(\mathbf{Y}_t) K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{t-\delta} \right] dt \right]. \end{aligned}$$

We will make use of the following additional notation

$$\begin{aligned} W_i(\mathbf{x}, \psi) &= \frac{(f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x}))}{(Th_T^d)^{1/2} f^2(\mathbf{x})} \int_{T_{i-1}}^{T_i} \psi(\mathbf{Y}_t) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt, \\ \Delta_i(\mathbf{x}, \psi) &= (W_i(\mathbf{x}, \psi) - \mathbb{E}[W_i(\mathbf{x}, \psi) | \mathcal{F}_{i-2}]), \\ \sigma^2(\mathbf{x}, \psi) &= \frac{\Psi_2(\mathbf{x}, \psi)}{f(\mathbf{x})} \int_{\mathbb{R}^d} [K^{(1)}(\mathbf{u})]^2 d\mathbf{u}, \end{aligned}$$

where

$$\Psi_2(\mathbf{x}, \psi) = \mathbb{E}(|\psi^2(\mathbf{Y})| | \mathbf{X} = \mathbf{x}).$$

Observe that

$$\begin{aligned} &(Th_T^{d+2})^{1/2} \left( \widehat{m}_T^{(1)}(\mathbf{x}, \psi) - \tilde{m}_T^{(1)}(\mathbf{x}, \psi) \right) \\ &= \frac{(Th_T^{d+2})^{1/2} (f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x}))}{Th_T^{d+1} f^2(\mathbf{x})} \\ &\quad \times \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \left( \psi(\mathbf{Y}_t) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) - \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)}\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2} \right] \right) dt \\ &= \sum_{i=1}^n \Delta_i(\mathbf{x}, \psi). \tag{4.43} \end{aligned}$$

Lemma 4.10, stated below, will play an instrumental role in the proof of Theorem 2.5.

**Lemma 4.10** *Under assumptions (A.1), (A.3)(i)–(ii), (A.4) and (A.6), as  $n \rightarrow \infty$ , we have*

$$\sum_{i=1}^n \Delta_i(\mathbf{x}, \psi) = \sum_{i=1}^n (W_i(\mathbf{x}, \psi) - \mathbb{E}[W_i(\mathbf{x}, \psi) | \mathcal{F}_{i-2}]) \xrightarrow{D} N(0, \sigma^2(\mathbf{x}, \psi)).$$

**4.1 Proof of Lemma 4.10**

It is easily seen that  $(W_i(\mathbf{x}, \psi))_{1 \leq i \leq n}$  is a sequence of martingale differences with respect to the sequence of  $\sigma$ -fields  $(\mathcal{F}_{i-1})_{1 \leq i \leq n}$ . Therefore, we have to check the following two conditions.

(a)

$$\sum_{i=1}^n \mathbb{E} \left[ \Delta_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2} \right] \xrightarrow{\mathbb{P}} \sigma^2(\mathbf{x}, \psi);$$

(b)

$$n \mathbb{E} \left[ \Delta_i^2(\mathbf{x}, \psi) \mathbb{1}_{\{|\Delta_i(\mathbf{x}, \psi)| > \epsilon\}} \right] = o(1) \text{ holds, for any } \epsilon > 0.$$

These conditions are necessary to establish the asymptotic normality related to discrete time martingale difference sequences (see, for instance, Hall and Heyde [37]).

**Proof of (a)**

First, observe that

$$\left| \sum_{i=1}^n \mathbb{E} \left[ W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2} \right] - \sum_{i=1}^n \mathbb{E} \left[ \Delta_{T,i}^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2} \right] \right| = \left| \sum_{i=1}^n (\mathbb{E} [W_i(\mathbf{x}, \psi) | \mathcal{F}_{i-2}])^2 \right|.$$

Note that, for  $T_{i-1} \leq t \leq T_i$ , we have  $\mathcal{S}_{T-\delta, \delta} \subset \mathcal{F}_{i-2}$ . Therefore, making use of condition (A.6)(i), we obtain

$$\begin{aligned} & \left| \mathbb{E} [W_i(\mathbf{x}, \psi) | \mathcal{F}_{i-2}] \right| \\ &= \left| \frac{1}{(Th_T^d)^{1/2}} \frac{(f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x}))}{f^2(\mathbf{x})} \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{i-2} \right] dt \right| \\ &\leq \frac{1}{(Th_T^d)^{1/2}} \frac{(f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x}))}{f^2(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ |\psi(\mathbf{Y}_t)| K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_T}{h_T} \right) | \mathcal{F}_{i-2} \right] dt \\ &\leq \frac{\mathcal{M}_\psi}{(Th_T^d)^{1/2}} \frac{(f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x}))}{f^2(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_T}{h_T} \right) | \mathcal{F}_{i-2} \right] dt. \end{aligned}$$

Using Taylor’s formula combined with assumption (A.4), we obtain

$$\begin{aligned} & \left| \sum_{i=1}^n (\mathbb{E} [W_i(\mathbf{x}, \psi) | \mathcal{F}_{i-2}])^2 \right| \\ &\leq \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{\mathcal{M}_\psi^2}{Th_T^d} \sum_{i=1}^n \left( \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K^{(1)} \left( \frac{\mathbf{x} - \mathbf{y}}{h_T} \right) f_T^{\mathcal{F}_{i-2}}(\mathbf{y}) dy dt \right)^2 \end{aligned}$$

$$\begin{aligned}
 &= \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{\mathcal{M}_\psi^2 h_T^d}{T} \sum_{i=1}^n \left( \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K^{(1)}(\mathbf{z}) f_T^{\mathcal{F}_{i-2}}(\mathbf{x} - h_T \mathbf{z}) d\mathbf{z} dt \right)^2 \\
 &= \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{\mathcal{M}_\psi^2 h_T^d}{\delta} \left( \frac{1}{n} \sum_{i=1}^n \left( \int_{T_{i-1}}^{T_i} f_T^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right)^2 + O(h_T) \right) \\
 &:= \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{\mathcal{M}_\psi^2 h_T^d}{\delta} \frac{1}{n} \sum_{i=1}^n \left( g_i^{\mathcal{F}_{i-2}}(\mathbf{x}) \right)^2 + O(h_T^d) \\
 &= O(h_T^d),
 \end{aligned}$$

where

$$g_i^{\mathcal{F}_{i-2}}(\mathbf{x}) = \left( \int_{T_{i-1}}^{T_i} f_T^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right)^2,$$

is a stationary and ergodic process. So the sum  $\frac{1}{n} \sum_{i=1}^n g_i^{\mathcal{F}_{i-2}}(\mathbf{x})$  has a finite limit, (see Krengel [47, Theorem 4.4]), which is

$$\mathbb{E} \left[ g_1^{\mathcal{F}_{-\delta}}(\mathbf{x}) \right] = g_1(\mathbf{x}) = \left( \int_0^\delta f_T(\mathbf{x}) dt \right)^2 = \delta^2 f^2(\mathbf{x}). \tag{4.44}$$

Moreover, observe that by assumptions (A.3), we have

$$\frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} = \frac{1}{f(\mathbf{x})} + O(h_T) = \frac{1}{f(\mathbf{x})} + o(1).$$

Using Jensen inequality and Assumption (A.6)(iii), we obtain

$$\begin{aligned}
 &\sum_{i=1}^n \mathbb{E} [W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2}] \\
 &= \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{T h_T^d} \mathbb{E} \left[ \left( \int_{T_{i-1}}^{T_i} \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) dt \right)^2 \middle| \mathcal{F}_{i-2} \right] \\
 &\leq \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{T h_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \mathbb{E} \left[ \psi^2(\mathbf{Y}_t) \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 \middle| \mathcal{S}_{t,\delta} \right] \middle| \mathcal{F}_{i-2} \right] dt \\
 &= \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{T h_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \Phi_2(\mathbf{X}_t, \psi) \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 \middle| \mathcal{F}_{i-2} \right] dt.
 \end{aligned}$$

According to Assumption (A.6)(iii), the function  $\Phi_2(\cdot, \psi)$  is continuous in the neighbourhood of  $\mathbf{x}$ , we have

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2}] \\ & \leq \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ |\Phi_2(\mathbf{X}_t, \psi) - \Phi_2(\mathbf{x}, \psi)| \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt \\ & \quad + \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \Phi_2(\mathbf{x}, \psi) \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt \\ & \leq \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{1}{Th_T^d} \sup_{\|\mathbf{x} - \mathbf{v}\| \leq h_T} |\Phi_2(\mathbf{v}, \psi) - \Phi_2(\mathbf{x}, \psi)| \\ & \quad \times \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt \\ & \quad + \left( \frac{f(\mathbf{x}) - h_T f^{(1)}(\mathbf{x})}{f^2(\mathbf{x})} \right)^2 \frac{\Phi_2(\mathbf{x}, \psi)}{Th_T^d} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt \\ & = \frac{1}{Th_T^d f^2(\mathbf{x})} (\Phi_2(\mathbf{x}, \psi) + o(h_T)) \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt. \end{aligned}$$

By a first order Taylor expansion of the function  $f_{T, T_{i-2}}$ , for  $\mathbf{x}^*$  in  $[\mathbf{x} - h_T \mathbf{v}, \mathbf{x}]$ , we obtain

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2}] \\ & \leq \frac{\Phi_2(\mathbf{x}, \psi)}{Th_T^d f^2(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left( K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right)^2 | \mathcal{F}_{i-2} \right] dt \\ & = \frac{\Phi_2(\mathbf{x}, \psi)}{Th_T^d f^2(\mathbf{x})} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2 \left( \frac{\mathbf{x} - \mathbf{u}}{h_T} \right) f_T^{\mathcal{F}_{T_{i-2}}}(\mathbf{u}) d\mathbf{u} dt \\ & = \frac{\Phi_2(\mathbf{x}, \psi)}{T f^2(\mathbf{x})} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2(\mathbf{v}) f_T^{\mathcal{F}_{T_{i-2}}}(\mathbf{x} - h_T \mathbf{v}) d\mathbf{v} dt \\ & = \frac{\Phi_2(\mathbf{x}, \psi)}{\delta f^2(\mathbf{x})} \left( \frac{1}{n} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} f_T^{\mathcal{F}_{T_{i-2}}}(\mathbf{x}) dt + O(h_T) \right) \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2(\mathbf{v}) d\mathbf{v}. \end{aligned}$$

It is clear, whenever  $\delta$  is small enough, that the quantities

$$\left( \int_{T_{i-1}}^{T_i} f_T^{\mathcal{F}_{T_{i-2}}}(\mathbf{x}) dt \right)_{i \in \mathbb{N}}$$

may be approximated by

$$\left( \delta f_{\mathcal{F}_{T_{i-1}}^{\mathcal{F}_{T_i-2}}}(\mathbf{x}) \right)_{i \in \mathbb{N}}.$$

Consequently, using the ergodic and stationarity properties of the process  $(\mathbf{X}_t)_{T \geq 0}$ , it follows that

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \left( \int_{T_{j-1}}^{T_j} f_T^{\mathcal{F}_{T_i-2}}(\mathbf{x}) dt \right) &= \mathbb{E} \left( \int_{T_0}^{T_1} f_T(\mathbf{x}) dt \right) + o(1) \\ &= \int_0^\delta \mathbb{E}(f_T(\mathbf{x})) dt + o(1) \\ &= \delta f(\mathbf{x}) + o(1). \end{aligned}$$

It follows that

$$\sum_{i=1}^n \mathbb{E} \left[ W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2} \right] = \frac{\Phi_2(\mathbf{x}, \psi)}{f(\mathbf{x})} \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2(\mathbf{y}) dy + O(h_T).$$

This implies that

$$\sum_{i=1}^n \mathbb{E} \left[ W_i^2(\mathbf{x}, \psi) | \mathcal{F}_{i-2} \right] = \frac{\Phi_2(\mathbf{x}, \psi)}{f(\mathbf{x})} \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2(\mathbf{y}) dy, \text{ as } T \rightarrow \infty. \tag{4.45}$$

**Proof of (b)**

Using the inequalities of Holder, Markov, Jensen and Minkowski’s inequalities, together with Assumption (A.6)(iii), we obtain for all  $\epsilon > 0$  and all  $p$  and  $q$ , such that

$$\frac{1}{p} + \frac{1}{q} = 1,$$

that

$$\begin{aligned} &\mathbb{E}[\Delta_{T,i}^2(\mathbf{x}) \mathbb{1}_{\{|\Delta_{T,i}(\mathbf{x})| > \epsilon\}}] \\ &\leq (\mathbb{E}[\Delta_{T,i}^{2q}(\mathbf{x})])^{1/q} (P\{|\Delta_{T,i}(\mathbf{x})| > \epsilon\})^{1/p} \\ &\leq \epsilon^{-2q/p} \mathbb{E}[|\Delta_{T,i}(\mathbf{x})|^{2q}] \\ &= \frac{\epsilon^{-2q/p}}{(Th_T^q)^q f^{2q}(\mathbf{x})} \mathbb{E} \left[ \int_{T_{i-1}}^{T_i} \left| \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[ \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) | \mathcal{F}_{i-2} \right] \right|^{2q} dt \right] \\ &\leq \frac{2^{2q} \epsilon^{-2q/p}}{(Th_T^q)^q f^{2q}(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left| \psi(\mathbf{Y}_t) K^{(1)} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right|^{2q} \right] dt \\ &= \frac{2^{2q} \epsilon^{-2q/p}}{(Th_T^q)^q f^{2q}(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \mathbb{E}[|\psi(\mathbf{Y}_t)|^{2q} | \mathcal{S}_{t,\delta}] \left( K^{(1)} \right)^{2q} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right] dt \end{aligned}$$

$$\begin{aligned}
 &= \frac{2^{2q} \epsilon^{-2q/p}}{(Th_T^d)^q f^{2q}(\mathbf{x})} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ h_{2q}(\mathbf{X}_t, \psi) \left( K^{(1)} \right)^{2q} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right] dt \\
 &\leq \frac{2^{2q} \epsilon^{-2q/p}}{(Th_T^d)^q f^{2q}(\mathbf{x})} \left( \sup_{\|\mathbf{x}-\mathbf{v}\| \leq h_T} |h_{2q}(\mathbf{v}, \psi) - h_{2q}(\mathbf{x}, \psi)| + h_{2q}(\mathbf{x}, \psi) \right) \\
 &\quad \times \int_{T_{i-1}}^{T_i} \mathbb{E} \left[ \left( K^{(1)} \right)^{2q} \left( \frac{\mathbf{x} - \mathbf{X}_t}{h_T} \right) \right] dt \\
 &= \frac{2^{2q} \epsilon^{-2q/p}}{(Th_T^d)^q f^{2q}(\mathbf{x})} (h_{2q}(\mathbf{x}, \psi) + o(1)) \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}} \left( K^{(1)} \right)^{2q} \left( \frac{\mathbf{x} - \mathbf{u}}{h_T} \right) f(\mathbf{u}) d\mathbf{u} dt \\
 &= \frac{2^{2q} \delta \epsilon^{-2q/p}}{(Th_T^d)^q f^{2q}(\mathbf{x})} (h_{2q}(\mathbf{x}, \psi) + o(1)) \int_{\mathbb{R}} \left( K^{(1)} \right)^{2q}(\mathbf{u}) f(\mathbf{x} - h_T \mathbf{u}) d\mathbf{u}.
 \end{aligned}$$

By a first order Taylor’s expansion, we have

$$\begin{aligned}
 &\mathbb{E}[\Delta_{T,i}^2(\mathbf{x}) \mathbb{1}_{\{|\Delta_{T,i}(\mathbf{x})| > \epsilon\}}] \\
 &= \frac{2^{2q} \epsilon^{-2q/p} \left\| \left( K^{(1)} \right)^{2q}(\mathbf{v}) \right\|_{\infty}}{(Th_T^d)^{q(q-1)} f^{2q-1}(\mathbf{x})} (h_{2q}(\mathbf{x}, \psi) + o(1)) \\
 &= o(1).
 \end{aligned} \tag{4.46}$$

Combining statements (4.45) and (4.46), we obtain

$$(Th_T^{d+2})^{1/2} \left( \widehat{m}_T^{(1)}(\mathbf{x}, \psi) - \widetilde{m}_T^{(1)}(\mathbf{x}, \psi) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( 0, \frac{\Phi_2(\mathbf{x}, \psi)}{f(\mathbf{x})} \int_{\mathbb{R}^d} \left( K^{(1)} \right)^2(\mathbf{y}) d\mathbf{y} \right). \tag{4.47}$$

Lemmas 4.9 and 4.10 combined with Theorem 2.4 complete the proof of Theorem 2.5. □

**Acknowledgements** The authors would like to thank the two referees for their constructive suggestions that have led to improve the presentation of this paper. The authors would like to express their appreciation for everybody fighting against the COVID-19 pandemic.

**Author contributions** SD, SB: conceptualization, methodology, investigation, writing—original draft, writing—review and Editing. Both authors contributed equally to this work.

### 5 Appendix

For the sake of clarity, introduce some details defining the ergodic property of continuous time processes. Let  $\mathbf{Z} = \{\mathbf{Z}_t\}_{t \geq 0}$  be a continuous time process taking values in some measurable space  $(E, \mathbf{Z})$  on which is defined a probability measure  $\mu$ . For  $\delta \geq 0$ , let  $\Upsilon^\delta$  be a  $\delta$ -shift transformation, i.e.,  $(\Upsilon^\delta(x))_s = x_{s+\delta}$ . A measurable set  $\mathcal{A}$  is  $\delta$ -invariant if it does not change under  $\delta$ -shift transformation, i.e.,  $\Upsilon^\delta(\mathcal{A}) = \mathcal{A}$ .

**Definition 5.1** ( $\delta$ -ergodicity) A continuous time process  $\mathbf{Z} = \{\mathbf{Z}_t\}_{t \geq 0}$  is  $\delta$ -ergodic if every measurable  $\delta$ -invariant set related to the process  $\mathbf{Z}$  has probability either 1 or 0, in other words, for any  $\delta$ -invariant set  $\mathcal{A}$ ,  $\mu(\mathcal{A}) = (\mu(\mathcal{A}))^2$ .

This definition means that if we take the process  $X$  and slice it into time blocks of length  $\delta$  then the new discrete time process  $(\mathbf{Z}_0^\delta, \mathbf{Z}_\delta^{2\delta}, \mathbf{Z}_{2\delta}^{3\delta}, \mathbf{Z}_{3\delta}^{4\delta}, \dots)$  is ergodic. For discrete time processes, we refer, for instance, to Krengel [47] for the definition and details on the ergodic property.

**Definition 5.2** (*Ergodicity*) A continuous time process  $\mathbf{Z} = \{\mathbf{Z}_t\}_{t \geq 0}$  is ergodic if it is  $\delta$ -ergodic for every  $\delta > 0$ .

It is well known from the ergodic theorem that, for a measurable function  $g$  and a stationary ergodic process  $\mathbf{Z} = \{\mathbf{Z}_t\}_{t \geq 0}$ , we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(\mathbf{Z}_t) dt = \mathbb{E}(g(\mathbf{Z}_0)). \quad (5.1)$$

We refer to the book of Krengel [47] for an account of details and results on the ergodic theory.

## References

1. Andrews, D.W.K.: Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21**(4), 930–934 (1984)
2. Beck, A.: On the strong law of large numbers. In: *Ergodic Theory* (Proc. Internat. Sympos., Tulane Univ., New Orleans, La., 1961), pp. 21–53. Academic Press, New York (1963)
3. Benrabah, O., Ould Saïd, E., Tatachak, A.: A kernel mode estimate under random left truncation and time series model: asymptotic normality. *Stat. Pap.* **56**(3), 887–910 (2015)
4. Beran, J.: *Statistics for Long-Memory Processes*. Volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York (1994)
5. Berline, A., Gannoun, A., Matzner-Løber, E.: Normalité asymptotique d'estimateurs convergents du mode conditionnel. *Can. J. Stat.* **26**(2), 365–380 (1998)
6. Blanke, D., Pumo, B.: Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.* **24**(1), 1–23 (2003)
7. Bosq, D.: *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Volume 110 of *Lecture Notes in Statistics*. Springer, New York (1998)
8. Bouzebda, S., Chaouch, M., Laïb, N.: Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Math. Methods Stat.* **25**(3), 168–195 (2016)
9. Bouzebda, S., Didi, S.: Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: asymptotic results. *Commun. Stat. Theory Methods* **46**(3), 1367–1406 (2017a)
10. Bouzebda, S., Didi, S.: Additive regression model for stationary and ergodic continuous time processes. *Commun. Stat. Theory Methods* **46**(5), 2454–2493 (2017b)
11. Bouzebda, S., Didi, S.: Some results about kernel estimators for function derivatives based on stationary and ergodic continuous time processes with applications. *Commun. Stat. Theory Methods* (2020). (**to appear**)
12. Bouzebda, S., Didi, S., El Hajj, L.: Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: asymptotic results. *Math. Methods Stat.* **24**(3), 163–199 (2015)
13. Bradley, R.C.: *Introduction to Strong Mixing Conditions*, vol. 1. Kendrick Press, Heber City (2007)
14. Chacón, J.E.: The modal age of statistics (2018). [arXiv:1807.02789](https://arxiv.org/abs/1807.02789)
15. Chacón, J.E., Montanero, J., Nogales, A.G.: A note on kernel density estimation at a parametric rate. *J. Nonparametric Stat.* **19**(1), 13–21 (2007)
16. Chen, Y.-C.: Modal regression using kernel density estimation: a review. *Wiley Interdiscip. Rev. Comput. Stat.* **10**(4), e1431 (2018)
17. Chen, Y.-C., Genovese, C.R., Tibshirani, R.J., Wasserman, L.: Nonparametric modal regression. *Ann. Stat.* **44**(2), 489–514 (2016)
18. Collomb, G., Härdle, W., Hassani, S.: A note on prediction via estimation of the conditional mode function. *J. Stat. Plan. Inference* **15**(2), 227–236 (1987)

19. Comte, F., Merlevède, F.: Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Stat.* **6**, 211–238 (2002). (**New directions in time series analysis (Luminy, 2001)**)
20. Debarh, M.: Normalité asymptotique de l'estimateur par ondelettes des composantes d'un modèle additif de régression. *C. R. Math. Acad. Sci. Paris* **343**(9), 601–606 (2006)
21. Delecroix, M.: Sur l'estimation et la prévision non-paramétrique des processus ergodiques. Université des sciences de Lille, Flandre-Artois, Doctorat d'État (1987)
22. Devroye, L.: A Course in Density Estimation. Volume 14 of Progress in Probability and Statistics. Birkhäuser Boston Inc., Boston (1987)
23. Devroye, L., Györfi, L.: Nonparametric Density Estimation. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. Wiley, New York (1985). (**The  $L_1$  view**)
24. Devroye, L., Lugosi, G.: Combinatorial Methods in Density Estimation. Springer Series in Statistics. Springer, New York (2001)
25. Didi, S.: Quelques propriétés asymptotiques en estimation non paramétrique de fonctionnelles de processus stationnaires en temps continu. Ph.D. thesis. Thèse de doctorat de Statistique, Paris 6 2014 (2014)
26. Didi, S., Louani, D.: Consistency results for the kernel density estimate on continuous time stationary and dependent data. *Stat. Probab. Lett.* **83**(4), 1262–1270 (2013)
27. Didi, S., Louani, D.: Asymptotic results for the regression function estimate on continuous time stationary and ergodic data. *Stat. Risk Model.* **31**(2), 129–150 (2014)
28. Eddy, W.F.: Optimum kernel estimators of the mode. *Ann. Stat.* **8**(4), 870–882 (1980)
29. Eddy, W.F.: The asymptotic distributions of kernel estimators of the mode. *Z. Wahrsch. Verw. Gebiete* **59**(3), 279–290 (1982)
30. Eggermont, P.P.B., LaRiccia, V.N.: Maximum Penalized Likelihood Estimation. Density Estimation. Springer Series in Statistics, vol. I. Springer, New York (2001)
31. Einbeck, J., Tutz, G.: Modelling beyond regression functions: an application of multimodal regression to speed-flow data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **55**(4), 461–475 (2006)
32. Ezzahrioui, M., Ould-Saïd, E.: Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *J. Nonparametric Stat.* **20**(1), 3–18 (2008)
33. Feng, Y., Fan, J., Suykens, J.A.K.: A statistical learning approach to modal regression (2017). [arXiv:1702.05960](https://arxiv.org/abs/1702.05960)
34. Ferraty, F., Laksaci, A., Vieu, P.: Functional time series prediction via conditional mode estimation. *C. R. Math. Acad. Sci. Paris* **340**(5), 389–392 (2005)
35. Grund, B., Hall, P.: On the minimisation of  $L^p$  error in mode estimation. *Ann. Stat.* **23**(6), 2264–2284 (1995)
36. Györfi, L., Härdle, W., Sarda, P., Vieu, P.: Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics, vol. 60. Springer, Berlin (1989)
37. Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (1980)
38. Härdle, W.: Applied Nonparametric Regression. Volume 19 of Econometric Society Monographs. Cambridge University Press, Cambridge (1990)
39. Härdle, W., Marron, J.S.: Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Stat.* **13**(4), 1465–1481 (1985)
40. Harel, M., Puri, M.L.: Conditional  $U$ -statistics for dependent random variables. *J. Multivariate Anal.* **57**(1), 84–100 (1996)
41. Herrmann, E., Ziegler, K.: Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Stat. Probab. Lett.* **68**(4), 359–368 (2004)
42. Jones, M.C.: On higher order kernels. *J. Nonparametric Stat.* **5**(2), 215–221 (1995)
43. Jones, M.C., Signorini, D.F.: A comparison of higher-order bias kernel density estimators. *J. Am. Stat. Assoc.* **92**(439), 1063–1073 (1997)
44. Jones, M.C., Linton, O., Nielsen, J.P.: A simple bias reduction method for density estimation. *Biometrika* **82**(2), 327–338 (1995)
45. Kemp, G.C.R., Santos Silva, J.M.C.: Regression towards the mode. *J. Econom.* **170**(1), 92–101 (2012)
46. Krebs, J.T.N.: The bootstrap in kernel regression for stationary ergodic data when both response and predictor are functions. *J. Multivar. Anal.* **173**, 620–639 (2019)
47. Krengel, U.: Ergodic Theorems. Volume 6 of de Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin (1985). (**With a supplement by Antoine Brunel**)



48. Laib, N., Louani, D.: Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivar. Anal.* **101**(10), 2266–2281 (2010)
49. Laib, N., Louani, D.: Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Stat. Plan. Inference* **141**(1), 359–372 (2011)
50. Lee, M.-J.: Mode regression. *J. Econom.* **42**(3), 337–349 (1989)
51. Lee, M.-J.: Quadratic mode regression. *J. Econom.* **57**(1–3), 1–19 (1993)
52. Leucht, A., Neumann, M.H.: Degenerate  $U$ - and  $V$ -statistics under ergodicity: asymptotics, bootstrap and applications in statistics. *Ann. Inst. Stat. Math.* **65**, 349–386 (2013)
53. Lu, Z.: Analyse des processus longue mémoire stationnaires et non-stationnaires : estimations, applications et prévisions. Ph.D. thesis. Thèse de doctorat de Mathématiques, Mathématiques financières et statistiques appliquées, Cachan, Ecole normale supérieure 2009 (2009)
54. Maslowski, B., Pospíšil, J.: Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.* **57**(3), 401–429 (2008)
55. Masry, E.: Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stoch. Process. Appl.* **115**(1), 155–177 (2005)
56. Masry, E.: Probability density estimation from sampled data. *IEEE Trans. Inf. Theory* **29**(5), 696–709 (1983)
57. Nadaraya, È.A.: Nonparametric Estimation of Probability Densities and Regression Curves. Volume 20 of Mathematics and its Applications (Soviet Series). Kluwer Academic Publishers Group, Dordrecht (1989). (Translated from the Russian by Samuel Kotz)
58. Neumann, M.H.: Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17**(4), 1268–1284 (2011)
59. Ould-Saïd, E.: A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Stat.* **24**(2), 231–239 (1997)
60. Ota, H., Kato, K., Hara, S.: Quantile regression approach to conditional mode estimation. *Electron. J. Stat.* **13**(2), 3120–3160 (2019)
61. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065–1076 (1962)
62. Prakasa Rao, B.L.S.: Statistical Inference for Diffusion Type Processes, Volume 8 of Kendall's Library of Statistics. Edward Arnold, London; Oxford University Press, New York (1999)
63. Prakasa Rao, B.L.S.: Nonparametric density estimation for stochastic processes from sampled data. *Publ. Inst. Statist. Univ. Paris.* **35**(3), 51–83 (1991) (1990)
64. Quintela-Del-Río, A., Vieu, P.: A nonparametric conditional mode estimate. *J. Nonparametric Stat.* **8**(3), 253–266 (1997)
65. Romano, J.P.: Bootstrapping the mode. *Ann. Inst. Stat. Math.* **40**(3), 565–586 (1988a)
66. Romano, J.P.: On weak convergence and optimality of kernel density estimates of the mode. *Ann. Stat.* **16**(2), 629–647 (1988b)
67. Samanta, M., Thavaneswaran, A.: Nonparametric estimation of the conditional mode. *Commun. Stat. Theory Methods* **19**(12), 4515–4524 (1990)
68. Sasaki, H., Ono, Y., Sugiyama, M.: Modal regression via direct log-density derivative estimation. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds.) *Neural Information Processing*, pp. 108–116. Springer, Cham (2016)
69. Sinotina, T., Vogel, S.: Universal confidence sets for the mode of a regression function. *IMA J. Manag. Math.* **23**(4), 309–323 (2012)
70. Stute, W.: Conditional  $U$ -statistics. *Ann. Probab.* **19**(2), 812–825 (1991)
71. Tapia, R.A., Thompson, J.R.: Nonparametric Probability Density Estimation. Volume 1 of Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore (1978)
72. Tarter, M.E., Lock, M.D.: Model-Free Curve Estimation. Volume 56 of Monographs on Statistics and Applied Probability. Chapman & Hall, New York (1993)
73. Wand, M.P., Jones, M.C.: Kernel Smoothing. Volume 60 of Monographs on Statistics and Applied Probability. Chapman and Hall Ltd., London (1995)
74. Wertz, W.: Statistical Density Estimation: A Survey, Volume 13 of *Angewandte Statistik und Ökonometrie [Applied Statistics and Econometrics]*. Vandenhoeck & Ruprecht, Göttingen. With German and French Summaries (1978)
75. Yao, W., Li, L.: A new regression model: modal linear regression. *Scand. J. Stat.* **41**(3), 656–671 (2014)
76. Youndjé, E.: Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. Ph.D. thesis. Thèse de doctorat de Sciences et techniques communes Rouen 1993 (1993)

77. Ziegler, K.: On bootstrapping the mode in the nonparametric regression model with random design. *Metrika* **53**(2), 141–170 (2001)
78. Ziegler, K.: On nonparametric kernel estimation of the mode of the regression function in the random design model. *J. Nonparametric Stat.* **14**(6), 749–774 (2002)
79. Ziegler, K.: On the asymptotic normality of kernel regression estimators of the mode in the nonparametric random design model. *J. Stat. Plan. Inference* **115**(1), 123–144 (2003)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.