Consistency of the frequency polygon estimators of density mode for strongly mixing processes

Ahmad Younso

Consistency of the frequency polygon estimators of density mode for strongly mixing processes

Ahmad Younso

Department of mathematical statistics, Faculty of sciences, Damascus and Tishreen universities/Syria

Abstract

We consider a simple estimator of the density mode using the frequency polygon estimate. We investigate strong consistency of the estimator for strong mixing sequence of real variables under mild assumptions. We study the almost sure rate of convergence and we show that the estimator can achieve optimal almost sure rates of convergence for appropriate choices of the bin widths. The asymptotic normality of the simple estimator is given and a simulation study is performed. Our asymptotic results are obtained without any differentiability condition assumed on the density around the mode.

Keywords: Frequency polygon; Mode estimate; Mixing condition; Consistency; Rate of convergence.

1. Introduction

The problem of estimating the modes of a density function has generated a considerable amount of interest in many areas. For example, in unsupervised problems where modes are used as measure of typicality of a set of data. In particular, in modern applications, mode estimation is often used in clustering, with the modes representing cluster centers. There is an extensive literature on mode estimation in the independent case, see the key references: (Parzen, 1962), (Konakov, 1973), (Samanta, 1973), (Devroye, 1979), (Romano, 1988) and the references therein. The common approaches consist of estimating the density mode by maximizing an estimate of the unknown density (usually a kernel estimate) on $\mathbb{R}^d$ or $\mathbb{R}$. (Abraham et al., 2004, 2003) deal with a simple estimate of the mode by maximizing the kernel density estimate on data. More recently, (Dasgupta and Kpotufe, 2014) investigate the k-NN mode estimation. Most of the existing works are concerned with the consistency and asymptotic normality of the estimators and rates achievable by various approaches. Despite the easy computation, there is only a very few literature dedicated to

Email address: ahyounso@yahoo.fr (Ahmad Younso)

the frequency polygon as density estimator. This estimator is constructed by connecting with straight lines the mid-bin values of a histogram, for a comprehensive overview, see (Scott, 1985) in the independent case and (Carbon et al., 1997) and (X.Yang, 2015) in the dependent case. (Scott, 1985) shows that the frequency polygon estimate has rates of convergence similar to those of non-negative kernel estimators with respect to the criterion of integrated mean squared error. (Carbon et al., 1997) and (X.Yang, 2015) extend the results of (Scott, 1985) to the weakly dependent case and investigate the uniform strong consistency. It is important to note that all the asymptotic results on the frequency polygon estimator of density are obtained on the real line. In this paper, we consider the problem of estimating the mode of an unknown unimodal density by maximizing the frequency polygon estimate of the density. Let $\{X_i : i \geq 1\}$ be a sequence of identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathbb{R}$. Suppose X_1 has a density f which is unknown. Denote by θ the unknown mode of f. Consider a partition $... < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < ...$ of the real line into equal intervals $I_{n,j} \equiv [(j-1)b_n, jb_n[$ of length b_n, where b_n, the bin width, is a strictly positive number decreasing to 0 as $n \to \infty$, and $j = 0, \pm 1, \pm 2, ...$. Let $J_{n,j} = [(j-1/2)b_n, (j+1/2)b_n[$ and consider the two adjacent histogram bins $I_{n,j}$ and $I_{n,j+1}$. Consider a set of observations $S_n = \{X_1, ..., X_n\}$ and denote the number of observations falling in the intervals $I_{n,j}$ and $I_{n,j+1}$ respectively by $\nu_{n,j}$ and $\nu_{n,j+1}$. Therefore, $\nu_{n,k} = \sum_{i=1}^{n} Y_{i,k}$ with

$$Y_{i,k} = \begin{cases} 1, & \text{if } X_i \in I_{n,k} \\ 0, & \text{otherwise,} \end{cases} \tag{1.1}$$

for $k = j$ or $j+1$. Then, the values of the histogram in these previous bins are given by

$$f_{n,k} = \nu_{n,k}/(nb_n),$$

for $k = j$ or $j+1$. Thus, the frequency polygon estimator of the density function f, for each $x \in J_{n,j}$, is defined by

$$f_n(x) = \left(\frac{1}{2} + j - \frac{x}{b_n}\right) f_{n,j} + \left(\frac{1}{2} - j + \frac{x}{b_n}\right) f_{n,j+1}. \tag{1.2}$$

We first let the estimate θ_n of the mode θ be defined as

$$\theta_n \in \arg\max_{\mathbb{R}} f_n. \tag{1.3}$$

The estimate θ_n is classified as indirect estimate since we first estimate f by f_n, then θ_n is taken to be any point of $\mathbb{R}$ for which (1.3) is satisfied . When f_n is the kernel estimator, θ_n is considered by many authors in the independent case, see for example (Parzen, 1962), (Konakov, 1973), (Samanta, 1973), (Devroye, 1979), (Romano, 1988) and (Shi et al., 2009). More recently, (Hwang and Shin, 2016) consider it in the ψ-weakly dependent case. As noticed by (Devroye, 1979), the estimate defined by (1.3) is of small practical value because a time-consuming search is necessary. Also, classical search methods perform satisfactorily only when f_n is sufficiently "regular" (continuous, unimodal, etc). An estimate of the mode which eliminates these problems is originally defined by (Devroye, 1979) and then considered by (Abraham et al., 2004, 2003). This later estimator estimates the mode by maximizing the density estimate on data. When f_n is the frequency polygon estimator, we let the estimate $\widehat{\theta}_n$ of the mode θ be defined as

$$\widehat{\theta}_n \in \arg\max_{S_n} f_n. \tag{1.4}$$

i.e.,

$$\widehat{\theta}_n \in \{x \in S_n : f_n(x) = \max_{i=1,\dots,n} f_n(X_i)\}.$$

The estimate (1.4) is classified as direct (simple or sample) estimate since there is a simple recipe to obtain the estimate $\widehat{\theta}_n$ from the data (see (Scott, 1992) for another definition of the sample mode). Since the sample points are naturally concentrated in high-density areas, the set S_n can be regarded as the most natural (random) grid for approximating the mode. Clearly, the sharper the density around the mode, the more the data will concentrate around it, and the better $\widehat{\theta}_n$ will perform. Our aim is to show some consistency and asymptotic results concerning the estimates (1.3) and (1.4) under a classical mixing condition.

2. Notations and assumptions

We first introduce some notations. A sequence $\{X_i : i \geq 1\}$ is said to be α-mixing (or strongly mixing) if

$$\alpha(n) = \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+n}^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \downarrow 0, \text{ as } n \to \infty, \tag{2.1}$$

where $\mathcal{F}_1^k$ and $\mathcal{F}_{k+n}^\infty$ are the sub σ-algebras generated by $\{X_i, i = 1, ..., k\}$ and $\{X_i, i = k+n, ...\}$, respectively. The α-mixing coefficient is one of the most general mixing coefficients (see (Bradley, 2005) for further details). It is often used to obtain asymptotic results for some estimators in nonparametric functional estimation. We suppose that

$$\alpha(n) = O(n^{-\rho}), \quad \text{for } \rho > 0. \tag{2.2}$$

This means that $\alpha(n)$ tends to 0 at polynomial rate. Observe that if $\alpha(n) = 0$ for all $n \in \mathbb{N}^*$, the two sub σ-algebras $\mathcal{F}_1^k$ and $\mathcal{F}_{k+n}^\infty$ are independent for each $k \in \mathbb{N}^*$, and this implies that $\{X_i : i \geq 1\}$ is a sequence of independent random variables. We consider for some $\epsilon > 0$, the level set

$$A(\epsilon) = \{x \in \mathbb{R} : f(x) > f(\theta) - \epsilon\}$$

which will play crucial rule throughout. Clearly, $\theta \in A(\epsilon)$ for any $\epsilon > 0$. We denote by *diam* $A(\epsilon)$ the diameter of $A(\epsilon)$ and

$$V(\delta) = \{x \in \mathbb{R} : |x - \theta| < \delta\}.$$

We shall assume that there exists $\delta_0 > 0$ such that f is continuous on $V \equiv V(\delta_0)$. Without loss of generality, we assume throughout the paper that δ_0 is small enough to ensure $\inf_V f > 0$. Finally, for any set B, its complement is denoted by B^c. Before we state the main results, we need the following basic assumptions.

H1. As $n \to \infty$, $b_n \to 0$ and $nb_n/\log n \to \infty$.

H2. For any $i \neq j$, the joint density $f_{i,j}(x, y)$ of (X_i, X_j) exists and satisfies

$$\sup_{(x,y)\in\mathbb{R}^2} |f_{i,j}(x, y) - f(x)f(y)| < M, \text{ for some } M > 0.$$

H3. for any $\delta > 0$, $\sup_{x\in V(\delta)^c} f(x) < f(\theta)$.

H4. As $\epsilon \to 0$, *diam* $A(\epsilon) \to 0$.

H5. There exists a constant $C > 0$ such that $|f(x) - f(x')| \leq C|x - x'|$ for all $x, x' \in \mathbb{R}$.

H6. $\mathbb{E}|X_1|^{2/T} < \infty$ for some $T > 0$.

H7. There exists $L > 0$ and $\beta > 0$ such that *diam* $A(\epsilon) \leq L\epsilon^\beta$ for $\epsilon \leq \epsilon_0$ where $\epsilon_0 > 0$ is small enough.

Note that hypotheses H1-H2 are mild regularity assumptions that are used by (X.Yang, 2015) to establish the strong uniform consistency of the frequency polygon estimator of the density. H3 is in line with the assumption that f is unimodal. H4 is introduced to avoid high density areas arbitrarily far from θ. It is used by (Abraham et al., 2003, 2004) to obtain strong consistency and asymptotic normality of the simple kernel estimator of the mode in the independent case. Assumptions H5-H6 are used by (X.Yang, 2015) to establish the optimal rate of convergence for the density estimator. Assumptions H6 is used by (Abraham et al., 2003) to investigate the rate of convergence for the simple kernel estimator of the mode in the independent case. If H7 holds, we say that the density f admits a peak index β. This peak index measures the sharpness of the density around the mode. Roughly, the sharper the density around θ, the larger the peak index is. For example, $\beta = 1/2$ corresponds to the family of normal densities and $\beta = 1$ corresponds to the family of Laplace densities (see (Abraham et al., 2003) for further details).

3. Preliminaries

For the proofs of main results, we need to state the following lemmas.

Lemma 3.1. *If H4 is satisfied, then, for any $\delta > 0$, $\sup_{V(\delta)^c} f < f(\theta)$.*

Lemma 3.2. *For any $\epsilon > 0$ and $\delta > 0$, we have $\mathbb{P}(X \in A(\epsilon) \cap V(\delta)) > 0$.*

For the proofs of above two lemmas, we refer the reader to (Abraham et al., 2003).

Lemma 3.3. *If H4 and (2.1) are verified then, a.s., for any* $\delta > 0$,

$$\max_{S_n \cap V(\delta)} f \longrightarrow f(\theta) \text{ as } n \to \infty. \tag{3.1}$$

Note that the result of Lemma 3.3 is shown without any condition on $\alpha(n)$. Extension of Lemma 3.3 to the general multivariate case is immediate. The general version of Lemma 3.3 extends the result of (Abraham et al., 2003) to the α-mixing case.

4. Main results

In this section, we will state the main results. The following theorem shows the almost sure convergence of the indirect estimator (1.3).

Theorem 4.1. *Suppose that H1-H3 and (2.2) are satisfied for some* $\rho > 2$*. If there exists* $\xi > 0$ *such that as* $n \to \infty$,

$$nb_n^{-2}(nb_n/\log n)^{-\rho/2+1/2}\log n(\log\log n)^{1+\xi} \longrightarrow 0, \tag{4.1}$$

then as $n \to \infty$,

$$\theta_n \longrightarrow \theta \qquad a.s.$$

The following theorem shows the almost sure convergence of the direct estimator.

Theorem 4.2. *Suppose that H1-H4, (2.2) and (4.1) are satisfied for some* $\rho > 2$*. Then, as* $n \to \infty$,

$$\widehat{\theta}_n \longrightarrow \theta \qquad a.s.$$

Note that condition (4.1) is the same as the one used by (X.Yang, 2015) to obtain the uniform consistency of $f_n(x)$ on $\mathbb{R}$. It is important to note that the results of Theorems 4.1 and 4.2 are obtained without assuming any differentiability condition on the density f around the mode θ. Similar result for the simple estimator of the mode issued by kernel method is obtained by (Abraham et al., 2003) under the independence assumption.
Now, we consider the almost sure rate of convergence for the estimate (1.3) in the following theorem.

Theorem 4.3. *Suppose that H1-H7 and (2.2) are satisfied for some* $\rho > 2$*. If there exists* $\xi > 0$ *such that as* $n \to \infty$,

$$n^{T+1}b_n^{-2}(nb_n/\log n)^{-\rho/2+1/2}\log n(\log\log n)^{1+\xi} \longrightarrow 0, \tag{4.2}$$

where $T > 0$ *is the constant defined in H6, then,*

$$\left|\hat{\theta}_n - \theta\right| = O\left(\psi_n^{\beta}\right) \qquad a.s.,$$

where $\psi_n = \max\{b_n, (\log n/(nb_n))^{1/2}\}$ *and* β *is the peak index.*

Remark. Condition (4.2) is the same as the one used by (X.Yang, 2015) to obtain the uniform consistency rate of $f_n(x)$ on $\mathbb{R}$. Theorem 4.3 indicates that the sharper the density around θ the faster the rate is. For example, if $\beta = 1$, the rate of convergence achieved by the simple estimator of the mode is the same as that of the density estimator while it is slower when $\beta < 1$. For $\beta = 1$, the rate of Theorem 4.3 is optimal $\psi_n = O\left(1/n^{1/5}\right)$ for a setting of $b_n \sim 1/n^{1/5}$ which is the optimal bin width derived in (Carbon et al., 1997). It is obtained for $\rho > 21/6$. It is the same optimal rate obtained by (Dasgupta and Kpotufe, 2014) for the K-NN mode estimator and faster than the rate $O\left((\log n)^{2/5}/n^{1/5}\right)$ established by (Abraham et al., 2003) for the simple kernel estimator. However, if we set $b_n \sim \left(n^{-1}\log n\right)^{1/3}$, then $\psi_n = O\left(n^{-1}\log n\right)^{1/3}$ which is the optimal rate of convergence for the density estimator in the *i.i.d.* case (see (Tran, 1994)). This rate can be achieved if $\rho > 7$.
The proof of the following proposition is immediate by the fact that the expectation $\mathbb{E}$ is a positive linear operator.

Proposition 4.1. *We have*

$$\left|\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right| = O\left(b_n\right) \qquad a.s.$$

The proof of the following corollary is immediate from Theorem 4.3.

Corollary 4.1. *Suppose that H1-H7, (4.2) and (2.2) are satisfied for some $\rho > 2$. Then,*

$$\mathbb{E}\left\{\left|\hat{\theta}_n - \theta\right|\right\} = O\left(\psi_n^{\beta}\right).$$

Theorem 4.4. *Suppose that H2-H7, (4.1) and (2.2) are satisfied for some $\rho > 2$. If as $n \to \infty$,*

$$\varphi(n) := n^{1+\tau/2} b_n^{-\tau/2} (nb_n)^{-\tau(2\tau\rho-1)} \longrightarrow 0, \tag{4.3}$$

for some $0 < \tau < 1/2$, then if $\hat{\theta}_n \in J_{n,j}$, a.s. and $\theta \in J_{n,j_0}$,

$$\sqrt{\frac{n}{b_n}}\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)(f_{n,j+1} - f_{n,j})/\sigma_n(\theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

with $\sigma_n^2(\theta) = \left(\frac{1}{2} + \left(2j_0 - \frac{\theta}{b_n}\right)^2\right) f(\theta)$ and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

Note that assumption (4.3) is weaker than (4.3) of (Tran, 1994).

4.1. Simulation study

First, to illustrate the asymptotic result of Lemma 3.3, we simulate a sample of size 700 drawn from a correlated Gaussian process (X_i) with $X_i \sim \mathcal{N}(0,1)$ and

$$\operatorname{cov}(X_i, X_j) = O\left(|i-j|^{-5}\right)$$

for any $i \neq j$. We set $\delta = 0.5$. For each $n = 1, ..., 700$, we replicate the simulation for 100 times. In each replication, we calculate $\max f(x)$ on $S_n \cap V(\delta)$.

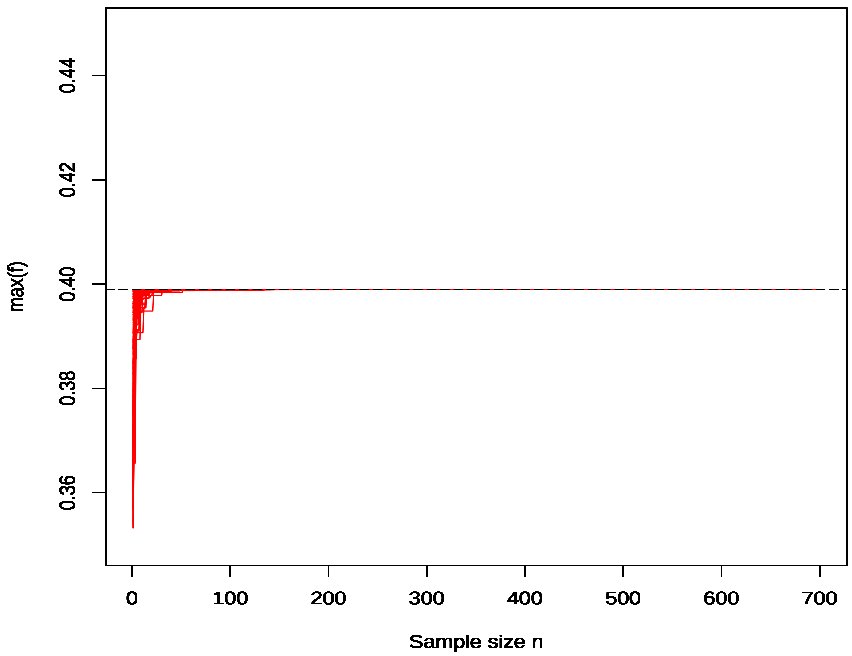
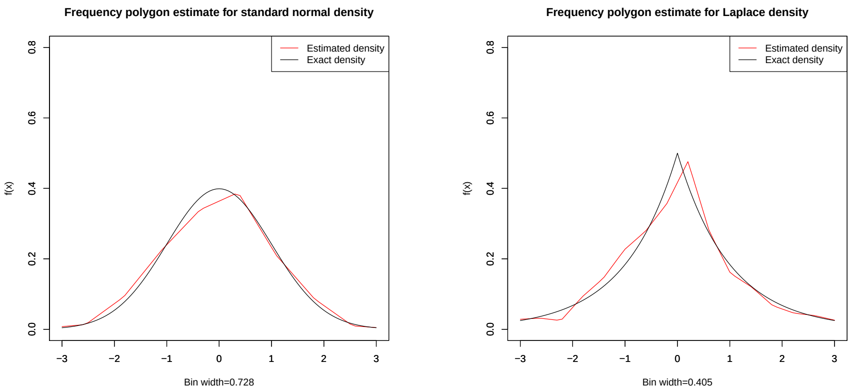


Figure 1: The red lines show how $\max\{f(X_i), X_i \in S_n \cap V(\delta),\ i = 1, ..., n\}$ varies around $f(0) \approx 0.4$ in function of $n = 1, ..., 700$.

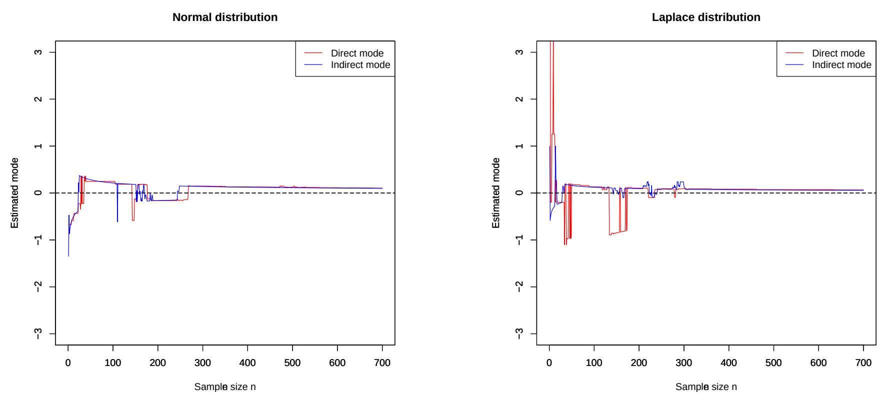
Figure 1 shows that $\max\{f(X_i), i = 1, ..., n\}$ tends quickly to $f(0)$ as n increases form 1 to 700. To illustrate the consistency results of Theorem 4.1-4.3, we first reconsider the above simulated Gaussian model and estimate the mode θ by both the direct and the indirect modes using the optimal bin width obtained by ((Carbon et al., 1997), Theorem 3.1). Then, we generate a samples of sizes $n = 700$ from the standard Laplace distribution also based on the above correlated model. We use this sample to estimate the density of this distribution. The following figure includes frequency polygon estimates of both the standard normal and Laplace densities.



(a) (b)

Figure 2: Frequency polygon estimators of the standard normal (a) and Laplace (b) densities based on $n = 700$ draws.

The following figure shows how the two estimators θ_n and $\hat{\theta}_n$ vary around the exact mode θ.



(a) (b)

Figure 3: Typical trajectories of θ_n and $\hat{\theta}_n$ for estimating the mode ($\theta = 0$) of both the standard normal (a) and Laplace (b) densities with $n = 1, ..., 700$.

In the one hand, in both (a) and (b) of Figure 3, we note that the similarity between the trajectory of θ_n and that of $\hat{\theta}_n$ increases as the sample size increases. In the other hand, if we compare (a) with (b) in Figure 3, we note that when the sample size increases, the trajectories in (b) becomes slightly closer to the exact mode than that of (a) since the peak index of Laplace density ($\beta = 1$) is more than that of normal density ($\beta = 0.5$) which is in line with Theorem 4.3.

5. Proofs

Proof of Lemma 3.3. Let $S_n^* = \{X_1^*, ..., X_n^*\}$ be a *ghost* set of independent and identically distributed random variables with density f. Recall that one can use Bradly's coupling lemma (see (Bradley, 1983)) to approximate a sequence of (stationary or identically distributed) strongly mixing variables $\{X_i : i \geq 1\}$ such that (2.1) by a sequence of *i.i.d.* variables $\{X_i^* : i \geq 1\}$. Thus, if H3 is verified, by ((Abraham et al., 2003), Theorem 1), we have *a.s.*, for any $\delta > 0$,

$$\max_{S_n^* \cap V(\delta)} f \longrightarrow f(\theta) \text{ as } n \to \infty. \tag{5.1}$$

By contradiction, we suppose that (3.1) is not satisfied, *i.e.*, there are $\epsilon_0 > 0$ and $\delta_0 > 0$ such that for each integer $n_0 \in \mathbb{N}^*$ there is an integer $n \geq n_0$ such that

$$\mathbb{P}\left(f(\theta) - \max_{S_n \cap V(\delta_0)} f \geq \epsilon_0\right) > 0.$$

Hence,

$$\mathbb{P}\left(f(\theta) - \max_{S_n \cap V(\delta_0)} f \geq \epsilon_0\right) = \mathbb{P}\left(\max_{S_n \cap V(\delta_0)} f \leq f(\theta) - \epsilon_0\right)$$
$$= \mathbb{P}\left(\bigcap_{i=1}^{n} ((f(X_i) \leq f(\theta) - \epsilon_0, X_i \in V(\delta_0)) \cup (X_i \in V(\delta_0)^c)\right) > 0.$$

Then, for each $i = 1, ..., n$,

$$\mathbb{P}\left((f(X_i) \leq f(\theta) - \epsilon_0, X_i \in V(\delta_0)) \cup (X_i \in V(\delta_0)^c)\right) > 0.$$

Consequently, since X_i^* has the same density f as X_i, for each $i = 1, ..., n$,

$$\mathbb{P}\left((f(X_i^*) \leq f(\theta) - \epsilon_0, X_i^* \in V(\delta_0)) \cup (X_i^* \in V(\delta_0)^c)\right) > 0.$$

Since $X_1^*, ..., X_n^*$ are *i.i.d.*, we can write

$$\mathbb{P}\left(\bigcap_{i=1}^{n} ((f(X_i^*) \leq f(\theta) - \epsilon_0, X_i^* \in V(\delta_0)) \cup (X_i^* \in V(\delta_0)^c)\right) > 0.$$

Thus, there are $\epsilon_0 > 0$ and $\delta_0 > 0$ such that for each integer $n_0 \in \mathbb{N}^*$ there is an integer $n \geq n_0$ such that

$$\mathbb{P}\left(f(\theta) - \max_{S_n^* \cap V(\delta_0)} (f) \geq \epsilon_0\right) > 0. \tag{5.2}$$

Finally, (5.2) leads to contradiction with (5.1) and the proof is completed. □

Proof of Theorem 4.1. The frequency polygon estimator (1.2) of f can be written as follows, for each $x \in J_{n,j} = [(j-1/2)b_n, (j+1/2)b_n[$:

$$f_n(x) = a_{n,1} f_{n,j} + a_{n,2} f_{n,j+1},$$

where $a_{n,1} = 1/2 + j - x/b_n$, $a_{n,2} = 1/2 - j + x/b_n$, $f_{n,j} = (nb_n)^{-1} \sum_{i=1}^{n} Y_{i,j}$, $f_{n,j+1} = (nb_n)^{-1} \sum_{i=1}^{n} Y_{i,j+1}$ with $Y_{i,j}$ and $Y_{i,j+1}$ are defined in (1.1). Since $x \in J_{n,j}$, $0 < a_{n,1}, a_{n,2} < 1$ and $a_{n,1} + a_{n,2} = 1$. Now, we have, for any $x \in J_{n,j}$ and $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\left(f_n(x)\right) &= a_{n,1} b_n^{-1} \mathbb{P}(X_1 \in I_{n,j}) + a_{n,2} b_n^{-1} \mathbb{P}(X_1 \in I_{n,j+1}) \\
&= a_{n,1} b_n^{-1} \int_{(j-1)b_n}^{jb_n} f(t)dt + a_{n,2} b_n^{-1} \int_{jb_n}^{(j+1)b_n} f(t)dt \\
&= a_{n,1} b_n^{-1} \Big(\int_{(j-1)b_n}^{(j-1/2)b_n} f(t)dt + \int_{(j-1/2)b_n}^{jb_n} f(t)dt \Big) \\
&+ a_{n,2} b_n^{-1} \Big(\int_{jb_n}^{(j+1/2)b_n} f(t)dt + \int_{(j+1/2)b_n}^{(j+1)b_n} f(t)dt \Big). \qquad (5.3)
\end{aligned}$$

Let $\delta > 0$, since $x \in J_{n,j}$, so in the one hand, if $|x - \theta| > \delta$ with $\theta < x < (j+1/2)b_n$, then $|t - \theta| > \delta$ for all $t \in [(j+1/2)b_n, (j+1)b_n[$ and this implies according to Lemma 3.1:

$$\sup_{t \in [(j+1/2)b_n, (j+1)b_n[} f(t) < f(\theta). \qquad (5.4)$$

In the other hand, if $|x - \theta| > \delta$ with $(j-1/2)b_n \leq x < \theta$, then $|t - \theta| > \delta$ for all $t \in [(j-1)b_n, (j-1/2)b_n[$ which implies also according to Lemma 3.1:

$$\sup_{t \in [(j-1)b_n, (j-1/2)b_n[} f(t) < f(\theta). \qquad (5.5)$$

Consequently, since $a_{n,1} + a_{n,2} = 1$, by (5.3)-(5.5), we get

$$\sup_{x \in V(\delta)^c} \mathbb{E}\left(f_n(x)\right) = \sup_{j} \sup_{x \in J_{n,j} \cap V(\delta)^c} \mathbb{E}\left(f_n(x)\right) < f(\theta).$$

So, we can write

$$\limsup_{n} \sup_{x \in V(\delta)^c} \mathbb{E}\left(f_n(x)\right) < f(\theta).$$

By ((X.Yang, 2015), Lemma 3.1 and (4.12)),

$$\sup_{x \in \mathbb{R}} |f_n(x) - \mathbb{E}\left(f_n(x)\right)| \longrightarrow 0 \ \ a.s.$$

Consequently, almost surely, for all $\delta > 0$,

$$\limsup_{n} \sup_{x \in V(\delta)^c} f_n(x) < f(\theta). \qquad (5.6)$$

We can write, almost surely, for all $\delta > 0$,

$$\limsup_n \sup_{V(\delta)^c} f_n < f(\theta).$$

By ((X.Yang, 2015),Theorem 2.1), we have for some $\delta_0 > 0$,

$$\sup_{x \in V(\delta_0)} |f_n(x) - f(x)| \longrightarrow 0 \;\; a.s.$$

Then, we have $a.s.$, for all $\delta \leq \delta_0$,

$$\limsup_n \sup_{V(\delta)^c} f_n < \lim_n \sup_{V(\delta)} f_n.$$

Finally, since δ is as small as desired, the last inequality shows that

$$\theta_n \longrightarrow \theta \;\; a.s. \qquad \square$$

Proof of Theorem 4.2. By (5.6), we have almost surely, for all $\delta > 0$,

$$\limsup_n \max_{S_n \cap V(\delta)^c} f_n < f(\theta).$$

By ((X.Yang, 2015), Theorem 2.1), we have for some $\delta_0 > 0$,

$$\sup_{x \in V(\delta_0)} |f_n(x) - f(x)| \longrightarrow 0 \;\; a.s.$$

Applying Lemma 3.3, we have $a.s.$, for all $\delta \leq \delta_0$,

$$\limsup_n \max_{S_n \cap V(\delta)^c} f_n < \lim_n \max_{S_n \cap V(\delta)} f_n.$$

Finally, since δ is as small as desired, the last inequality shows that

$$\hat{\theta}_n \longrightarrow \theta \;\; a.s. \qquad \square$$

Proof of Theorem 4.3. Let $\psi_n = \max\left\{b_n, (\log n/(nb_n))^{1/2}\right\}$ and $M' > 0$ such that for n large enough:

$$\|f_n - \mathbb{E}f_n\|_\infty \leq M' \left(\log n/(nb_n)\right)^{1/2} \tag{5.7}$$

and

$$\|\mathbb{E}f_n - f\|_V \leq M' b_n, \tag{5.8}$$

with $\|.\|_\infty$ and $\|.\|_V$ denote the supremum norms on $\mathbb{R}$ and V, respectively. The existence of M' is ensured by ((X.Yang, 2015), Theorem 2.2 and Lemma 3.1). Set, for $n \geq 1$, $\gamma_n = 2M'\psi_n$ and $k_n = 2^\beta L \left(1.5Cb_n + 4\gamma_n\right)^\beta$ where L and C are the constants defined in H5 and H7, respectively and $\eta > 0$ a constant to be chosen later. So, $\gamma_n \to 0$ and $k_n \to 0$ as $n \to \infty$. We prove the theorem if we show that for n large enough

$$\mathbb{P}\left(|\theta_n - \theta| \geq k_n\right) = 0. \tag{5.9}$$

For n large enough, to ensure that $V(k_n) \subset V$, we have

$$
\begin{aligned}
&\mathbb{P}\left(|\theta_n - \theta| \geq k_n\right) \\
&\quad \leq \mathbb{P}\left(\max_{S_n \cap V(k_n)} (f_n) \leq \max_{S_n \cap V(k_n)^c} (f_n) \right) \\
&\quad \leq \mathbb{P}\left(-\|f_n - \mathbb{E}f_n\|_\infty + \max_{S_n \cap V(k_n)} (f_n) \leq \max_{S_n \cap V(k_n)^c} (f_n) + \|f_n - \mathbb{E}f_n\|_\infty \right) \\
&\quad = \mathbb{P}\left(\max_{S_n \cap V(k_n)} (f_n) \leq \max_{S_n \cap V(k_n)^c} (f_n) + 2\|f_n - \mathbb{E}f_n\|_\infty \right) \\
&\quad \leq \mathbb{P}\left(\max_{S_n \cap V(k_n)} (f) \leq \sup_{V(k_n)^c} (f_n) + 2\|f_n - \mathbb{E}f_n\|_\infty + \|\mathbb{E}f_n - f\|_V \right) \\
&\quad \leq \mathbb{P}\left(\max_{S_n \cap V(k_n)} (f) \leq \sup_{V(k_n)^c} (f_n) + 3\gamma_n \right) + \mathbb{P}\left(\|f_n - \mathbb{E}f_n\|_\infty \geq \gamma_n\right) \\
&\qquad\qquad + \mathbb{P}\left(\|\mathbb{E}f_n - f\|_V \geq \gamma_n\right).
\end{aligned}
$$

Clearly, according to (5.7) and (5.8), we have for n large enough

$$\mathbb{P}\left(\|f_n - \mathbb{E}f_n\|_\infty \geq \gamma_n\right) = 0$$

and

$$\mathbb{P}\left(\|\mathbb{E}f_n - f\|_V \geq \gamma_n\right) = 0.$$

Consequently, it remains to prove that for n large enough

$$\mathbb{P}\left(\max_{S_n \cap V(k_n)} (f) \leq \sup_{V(k_n)^c} \mathbb{E}(f_n) + 3\gamma_n \right) = 0. \tag{5.10}$$

For each $x \in J_{n,j} = [(j-1/2)b_n, (j+1/2)b_n[$, we can write

$$
\begin{aligned}
\mathbb{E}f_n(x) &= a_{n,1} b_n^{-1} \int_{(j-1)b_n}^{jb_n} f(t)dt + a_{n,2} b_n^{-1} \int_{jb_n}^{(j+1)b_n} f(t)dt \\
&\leq a_{n,1} b_n^{-1} \int_{(j-1)b_n}^{jb_n} |f(t) - f(x)|dt + a_{n,1} b_n^{-1} \int_{(j-1)b_n}^{jb_n} f(x)dt \\
&\quad + a_{n,2} b_n^{-1} \int_{jb_n}^{(j+1)b_n} |f(t) - f(x)|dt + a_{n,2} b_n^{-1} \int_{jb_n}^{(j+1)b_n} f(x)dt,
\end{aligned}
$$

By H5, we have $|f(t) - f(x)| \leq 1.5Cb_n$ since $|x - t| \leq 1.5b_n$. Now, since $a_{n,1} + a_{n,2} = 1$, we get

$$\sup_{V(k_n)^c} \mathbb{E}(f_n) = \sup_j \sup_{V(k_n)^c \cap J_{n,j}} \mathbb{E}(f_n) \leq \sup_{V(k_n)^c} (f) + 1.5Cb_n. \tag{5.11}$$

Choosing n large enough, we may assume that $k_n \leq \epsilon_0$. Fix $t \in V(k_n)^c$. If t meets the condition $f(\theta) - f(t) \leq \epsilon_0/2$, then by H7, we have

$$k_n \leq |t - \theta| \leq diam\ A(2(f(\theta) - f(t))) \leq 2^{\beta} L(f(\theta) - f(t))^{\beta}.$$

Consequently,

$$f(\theta) - f(t) \geq 2^{-1} \left(k_n/L\right)^{1/\beta} = 1.5Cb_n + 4\gamma_n.$$

It implies that

$$f(t) \leq f(\theta) - 1.5Cb_n - 4\gamma_n. \tag{5.12}$$

Now, choosing n large enough, we may assume that $1.5Cb_n + 4\gamma_n \leq \epsilon_0/2$. If $t \in V(k_n)^c$ with $f(\theta) - f(t) \geq \epsilon_0/2$, we have

$$f(t) \leq f(\theta) - 1.5Cb_n - 4\gamma_n. \tag{5.13}$$

By (5.12)-(5.13), we can write

$$\sup_{V(k_n)^c}(f) \leq f(\theta) - 1.5Cb_n - 4\gamma_n. \tag{5.14}$$

Therefore, by (5.11) and (5.14),

$$\begin{aligned}
&\mathbb{P}\left(\max_{S_n \cap V(k_n)}(f) \leq \sup_{V(k_n)^c} \mathbb{E}(f_n) + 3\gamma_n\right) \\
&\leq \mathbb{P}\left(\max_{S_n \cap V(k_n)}(f) \leq \sup_{V(k_n)^c}(f) + 1.5Cb_n + 3\gamma_n\right) \\
&\leq \mathbb{P}\left(f(\theta) - \max_{S_n \cap V(k_n)}(f) \geq \gamma_n\right).
\end{aligned}$$

Then, it suffices to prove that for n large enough,

$$\mathbb{P}\left(f(\theta) - \max_{S_n \cap V(k_n)}(f) \geq \gamma_n\right) = 0. \tag{5.15}$$

Let $S_n^* = \{X_1^*, ..., X_n^*\}$ be a *ghost* set of independent and identically distributed random variables with density f. By (5.12)-(5.13), we get

$$A(1.5Cb_n + 4\gamma_n) \subset V(k_n),$$

then obviously,

$$A(\gamma_n) \subset V(k_n).$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left(f(\theta) - \max_{S_n^* \cap V(\gamma_n)} f \geq \gamma_n\right) \\
&= \mathbb{P}\left(\bigcap_{i=1}^{n} ((f(X_i^*) \leq f(\theta) - \gamma_n, X_i^* \in V(\gamma_n)) \cup (X_i^* \in V(\gamma_n)^c)\right) \\
&\leq \prod_{i=1}^{n} \mathbb{P}\left(((f(X_i^*) \leq f(\theta) - \gamma_n, X_i^* \in V(\gamma_n)) \cup (X_i^* \in V(k_n)^c)\right) \\
&= \left[1 - \mathbb{P}\left\{X \in A(\gamma_n) \cap V(k_n)\right\}\right]^n = \left[1 - \mathbb{P}\left\{X \in A(\gamma_n)\right\}\right]^n. \qquad (5.16)
\end{aligned}$$

Without loss of generality, suppose that ϵ_0 is small enough so that

$$\inf_{A(\epsilon_0)}(f) > 0 \text{ and } \exists l > 0: \ \lambda(A(\epsilon)) \geq l\epsilon^\beta, \ \forall \epsilon \leq \epsilon_0,$$

with λ denotes the Lebesgue measure. Thus, by (5.16),

$$\mathbb{P}\left(f(\theta) - \max_{S_n^* \cap V(\gamma_n)} f \geq \gamma_n\right) \leq \left(1 - l\gamma_n^\beta \inf_{A(\epsilon_0)}(f)\right)^n.$$

By Borel-Cantelli Lemma, we have for n large enough

$$\mathbb{P}\left(f(\theta) - \max_{S_n^* \cap V(k_n)}(f) \geq \gamma_n\right) = 0. \qquad (5.17)$$

As was done in the proof of Lemma 3.3, if we assume that (5.15) is not true, then we will arrive to a contradiction with (5.17) and the proof is completed.□

Proof of Theorem 4.4. Let $\hat{\theta}_n \in J_{n,j}$ *a.s.*, then $\mathbb{E}(\hat{\theta}_n) \in J_{n,j}$. Consequently, by (1.2),

$$\frac{1}{b_n}\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)(f_{n,j+1} - f_{n,j}) = f_n(\hat{\theta}_n) - f_n(\mathbb{E}(\hat{\theta}_n)). \qquad (5.18)$$

Then, we can write

$$\begin{aligned}
&\sqrt{\frac{n}{b_n}}\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)(f_{n,j+1} - f_{n,j}) \\
&= \sqrt{nb_n}\left(f_n(\hat{\theta}_n) - \mathbb{E}\left(f_n(\mathbb{E}(\hat{\theta}_n))\right)\right) + \sqrt{nb_n}\left(\mathbb{E}\left(f_n(\mathbb{E}(\hat{\theta}_n))\right) - f_n(\mathbb{E}(\hat{\theta}_n))\right). \qquad (5.19)
\end{aligned}$$

We will first show that

$$\sqrt{nb_n}\left(f_n(\hat{\theta}_n) - \mathbb{E}\left(f_n(\mathbb{E}(\hat{\theta}_n))\right)\right) \xrightarrow{P} 0 \qquad (5.20)$$

with $\xrightarrow{P}$ denotes the convergence in probability. We have for any $\epsilon > 0$ and n large enough

$$\mathbb{P}\left(\left|f_n(\hat{\theta}_n) - \mathbb{E}(f_n(\mathbb{E}(\hat{\theta}_n))\right| > \epsilon/\sqrt{nb_n}\right) \leq \sup_{x \in J_{n,j}} \mathbb{P}\left(|f_n(x) - \mathbb{E}(f_n(x))| > \epsilon/\sqrt{nb_n}\right).$$

Therefore, it suffices to prove that as $n \to \infty$,

$$\mathbb{P}\left(|f_n(x) - \mathbb{E}(f_n(x))| > \epsilon/\sqrt{nb_n}\right) \longrightarrow 0 \text{ for } x \in J_{n,j}. \tag{5.21}$$

For this aim, we write

$$f_n(x) = \frac{1}{nb_n}\sum_{i=1}^{n} Z_i(x)$$

where

$$Z_i(x) = \left(\frac{1}{2} + j - \frac{x}{b_n}\right) Y_{i,j} + \left(\frac{1}{2} - j + \frac{x}{b_n}\right) Y_{i,j+1},$$

with $Y_{i,j}$ and $Y_{i,j+1}$ are given in (1.1). Then,

$$\sqrt{nb_n}(f_n(x) - \mathbb{E}(f_n(x))) = \sum_{i=1}^{n} \Delta_i(x),$$

with $\Delta_i(x) = \frac{1}{\sqrt{nb_n}}(Z_i(x) - \mathbb{E}(Z_i(x)))$. We now refer to $\Delta_i(x)$ simply as Δ_i for simplicity. Without loss of generality, let $n = 2pq$ for $p = p_n, q = q_n \in [1, n/2]$ such that $p_n \to \infty$ as $n \to \infty$ and let us define blocks as follow

$$\begin{aligned}
W_1 &= \sum_{i=1}^{p} \Delta_i, & V_1 &= \sum_{i=p+1}^{2p} \Delta_i \\
W_2 &= \sum_{i=2p+1}^{3p} \Delta_i, & V_2 &= \sum_{i=3p+1}^{4p} \Delta_i \\
&\vdots & &\vdots \\
W_q &= \sum_{i=2(q-1)p+1}^{(2q-1)p} \Delta_i, & V_q &= \sum_{i=(2q-1)p+1}^{2pq} \Delta_i.
\end{aligned}$$

Observe that W_i and V_i are measurable with respect to the σ-algebras: $\sigma(\Delta_i, i \in I_k)$ and $\sigma(\Delta_i, i \in \tilde{I}_k)$, respectively, where $I_k = \{i :\ 2(k-1)p+1 \le i \le (2k-1)p\}$ and $\tilde{I}_k = \{i :\ (2k-1)p + 1 \le i \le 2kp\}$ for all $k = 1, ..., q$. Furthermore, we have $|i - i^{'}| > p$ for any $i \in I_k$ and $i^{'} \in I_{k^{'}}$ if $k \neq k^{'}$. In the same way, one can show that $|i - i^{'}| > p$ for any $i \in \tilde{I}_k$ and $i^{'} \in \tilde{I}_{k^{'}}$ if $k \neq k^{'}$. Consequently,

$$\sqrt{nb_n}(f_n(x) - \mathbb{E}(f_n(x))) = \sum_{k=1}^{q} W_k + \sum_{k=1}^{q} V_k.$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left(|f_n(x) - \mathbb{E}(f_n(x))| > \epsilon\right) \\
&\quad \le \mathbb{P}\left(\left|\sum_{k=1}^{q} W_k\right| > \epsilon/2\right) + \mathbb{P}\left(\left|\sum_{k=1}^{q} V_k\right| > \epsilon/2\right).
\end{aligned} \tag{5.22}$$

We will find an upper bound for each term in the right hand-side of the above inequality. Without loss of generality, we will only find an upper bound for $\mathbb{P}\left(|\sum_{k=1}^{q} W_k| > \epsilon/2\right)$ since the other term can be similarly treated. According to Bradly's lemma (see (Bradley, 1983)), we can find mutually independent random variables $W_1^*, ..., W_q^*$ such that for all $k = 1, ..., q$, W_k^* has the same probability distribution as W_k and

$$\mathbb{P}\left(|W_k - W_k^*| > \zeta\right) \leq 18 \left(\|W_k\|_\nu/\zeta\right)^{\nu/(2\nu+1)} [\alpha(p)]^{2\nu/(2\nu+1)}, \tag{5.23}$$

for any positive numbers ζ and ν such that $0 < \zeta \leq \|W_k\|_\nu < \infty$. Clearly,

$$\begin{aligned} &\mathbb{P}\left(\left|\sum_{k=1}^{q} W_k\right| > \epsilon/2\right) \\ &\quad \leq \mathbb{P}\left(\left|\sum_{k=1}^{q}(W_k - W_k^*)\right| > \epsilon/4\right) + \mathbb{P}\left(\left|\sum_{k=1}^{q} W_k^*\right| > \epsilon/4\right). \end{aligned} \tag{5.24}$$

Let $\tau = \nu/(2\nu + 1)$ and choose $p = (nb_n)^\tau$. Clearly,

$$\max_{1\leq k\leq q} \|W_k\|_\nu \leq 2p(nb_n)^{-1/2}. \tag{5.25}$$

In the one hand, if $\epsilon/4q \leq \|W_k\|_\nu$, then by (5.23),

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{k=1}^{q}(W_k - W_k^*)\right| > \epsilon/4\right) &\leq q \max_{1\leq k\leq q} \mathbb{P}\left(|W_k - W_k^*| > \epsilon/4q\right) \\ &\leq 18q(4q\epsilon^{-1})^\tau \max_{1\leq k\leq q} \|W_k\|_\nu^\tau [\alpha(p)]^{2\tau} \\ &\leq 18(8\epsilon^{-1})^\tau q^{1+\tau} p^\tau (nb_n)^{-\tau/2} p^{-2\tau\rho} \\ &\leq Cn^{1+\tau/2} b_n^{-\tau/2} (nb_n)^{-\tau(2\tau\rho+1)} \\ &= C\varphi(n), \end{aligned} \tag{5.26}$$

for some generic constant $C > 0$. In the other hand, if $\epsilon/4q > \|W_k\|_\nu$, then by (5.23),

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{k=1}^{q}(W_k - W_k^*)\right| > \epsilon/4\right) &\leq q \max_{1\leq k\leq q} \mathbb{P}\left(|W_k - W_k^*| > \epsilon/4q\right) \\ &\leq 18q[\alpha(p)]^{2\tau} \leq C\varphi(n). \end{aligned} \tag{5.27}$$

A bound for $\mathbb{P}\left(|\sum_{k=1}^{q} W_k^*| > \epsilon/4\right)$ will now be obtained. By Bernstein inequality, we have

$$\mathbb{P}\left(\left|\sum_{k=1}^{q} W_k^*\right| > \epsilon/4\right) \leq \exp\left(-\frac{\epsilon^2}{64\sum_{k=1}^{q} \mathrm{var}(W_k^*) + 8\|W_k^*\|_\infty \epsilon}\right).$$

Since by ((Carbon et al., 1997), Lemma 5.2),

$$\sum_{k=1}^{q} \mathrm{var}(W_k^*) = \sum_{k=1}^{q} \mathrm{var}(W_k) = O\left(\frac{1}{(nb_n)^2}\right)$$

then, by (5.25),

$$\mathbb{P}\left(\left|\sum_{k=1}^{q} W_k^*\right| > \epsilon/4\right) \leq \exp\left(-C(nb_n)^{1/2-\tau}\right). \tag{5.28}$$

Combining (4.3), (5.22), (5.24) and (5.26)-(5.28), we get (5.21) which implies (5.20). Using the same proof as (5.20), we can show

$$\sqrt{nb_n}\left(\mathbb{E}f_n(\mathbb{E}(\hat{\theta}_n)) - f_n(\mathbb{E}(\hat{\theta}_n))\right) \xrightarrow{P} 0$$

and

$$\sqrt{nb_n}\left(\mathbb{E}\left(f_n(\theta)\right) - f_n(\theta)\right) \xrightarrow{P} 0.$$

Then,

$$\sqrt{nb_n}\left(\mathbb{E}\left(f_n(\mathbb{E}(\hat{\theta}_n))\right) - f_n(\mathbb{E}(\hat{\theta}_n))\right) - \sqrt{nb_n}\left(\mathbb{E}\left(f_n(\theta)\right) - f_n(\theta)\right) \xrightarrow{P} 0 \tag{5.29}$$

It is well known that if $\theta \in J_{n,j_0}$ (see (Carbon et al., 2010)),

$$\sqrt{nb_n}\left(\mathbb{E}\left(f_n(\theta)\right) - f_n(\theta)\right)/\sigma_n(\theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{5.30}$$

with $\sigma_n^2(\theta) = \left(\frac{1}{2} + \left(2j_0 - \frac{\theta}{b_n}\right)^2\right) f(\theta)$. By (5.19) together with (5.29)-(5.30), the theorem is proved. □

Acknowledgments

The author would like to thank the anonymous referees whose valuable comments led to an improved version of the paper.

References

Abraham, C., Biau, G., Cadre, B., 2003. Simple estimation of the mode of a multivariate density. The Canadian Journal of Statistics 31, 23–34.

Abraham, C., Biau, G., Cadre, B., 2004. On the asymptotic properties of a simple estimate of the mode. ESAIM: PS 8, 1–11.

Bradley, R.C., 1983. Approximation theorems for strongly mixing random variables. Michigan Math. J. 30, 69–81.

Bradley, R.C., 2005. Basic properties of strong mixing conditions. A survey and some open questions. Probability Surveys 2, 107–144.

Carbon, M., Francq, C., Tran, L.T., 2010. Asymptotic normality of frequency polygons for random fields. Statist. Plann. Inference 2, 502–514.

Carbon, M., Garel, B., Tran, L.T., 1997. Frequency polygons for weakly dependent processes. Statistics & Probability Letters 33, 1–13.

Dasgupta, S., Kpotufe, S., 2014. Optimal rates for k-nn density and mode estimation. In Advances in Neural Information Processing Systems , 2555–2563.

Devroye, L., 1979. Recursive estimation of the mode of a multivariate density. The Canadian Journal of Statistics 7, 159–167.

Hwang, E., Shin, D.W., 2016. Kernel estimators of mode under ψ-weak dependence,. Annals of the Institute of Statistical Mathematics 68, 301–327.

Konakov, V.D., 1973. On asymptotic normality of the sample mode of multivariate distributions. Theory of Probability and its Applications 18, 836–842.

Parzen, E., 1962. On estimation of a probability density function and mode. The Annals of Mathematica1 Statistics 33, 1065–1076.

Romano, J.P., 1988. On weak convergence and optimality of kernel density estimates of the mode. The Annals of Statistics 16, 629–647.

Samanta, M., 1973. Nonparamebic estimation of the mode of a multivariate density. South African Statistical Journal 7, 109–117.

Scott, D.W., 1985. Frequency polygons: theory and application. Journal of the American Statistical Association 80, 348–354.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley.

Shi, X., Wu, Y., Miao, B., 2009. A note of the convergence rate of the kernel density estimator of the mode. Statistics & Probability Letters 79, 1866–1871.

Tran, L., 1994. Density estimation for time series by histograms. Statist. Plann. Inference 40, 61–79.

X.Yang, 2015. Frequency polygon estimation of density function for dependent samples. Journal of the Korean Statistical Society 44, 530–537.