



**HAL**  
open science

## Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy

Salim Bouzebda, Issam Elhattab

► **To cite this version:**

Salim Bouzebda, Issam Elhattab. Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electronic Journal of Statistics*, 2011. hal-00296641v2

**HAL Id: hal-00296641**

**<https://hal.science/hal-00296641v2>**

Submitted on 27 Aug 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Kernel-Type Estimators of Shannon's Entropy

Salim BOUZEBDA<sup>1\*</sup> and Issam ELHATTAB<sup>1†</sup>

Laboratoire de Statistique Théorique et Appliquée (L.S.T.A.)

Université de Paris VI<sup>1</sup>

## Abstract

We establish uniform-in-bandwidth consistency for kernel-type estimators of the differential entropy. Precisely, in this work, we consider two kernel-type estimators of Shannon's entropy. As a consequence, an asymptotic 100% confidence interval of entropy is provided.

*AMS Subject Classification:* 62F12 ; 62F03 ; 62G30 ; 60F17 ; 62E20.

*Keywords:* Entropy; Kernel estimation; Uniform in bandwidth; Consistency.

## 1 Introduction and estimation

Let  $(\mathbf{X}_n)_{n \geq 1}$  be a sequence of i.i.d.  $\mathbb{R}^d$ -valued random vectors,  $d \geq 1$ , with distribution function  $\mathbb{F}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x})$  for  $\mathbf{x} \in \mathbb{R}^d$ . We set here  $\mathbf{X} = (X_1, \dots, X_d) \leq \mathbf{x} = (x_1, \dots, x_d)$  whenever  $X_i \leq x_i$ , for all  $i = 1, \dots, d$ . Assume that  $\mathbb{F}(\cdot)$  has a density function  $f(\cdot)$  with respect to Lebesgue measure on  $\mathbb{R}^d$ . The differential entropy of  $f(\cdot)$  is defined by the quantity

$$H(f) = - \int_{\mathbb{R}^d} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} \quad (1.1)$$

$$= - \int_{\mathbb{R}^d} \log(f(\mathbf{x})) d\mathbb{F}(\mathbf{x}), \quad (1.2)$$

whenever this integral is meaningful, and where  $d\mathbf{x}$  denotes Lebesgue measure in  $\mathbb{R}^d$ . We will use the convention that  $0 \log(0) = 0$  since  $u \log(u) \rightarrow 0$  as  $u \rightarrow 0$ . The differential entropy concept was introduced by Shannon (1948). Since then and because of numerous potential applications, the subject has received a considerable interest handling various problems as estimating the quantity  $H(f)$ . We refer to Beirlant *et al.* (1997), and the references therein, for details. The main purpose of the present paper is to establish uniform in bandwidth consistency of the so-called kernel estimator

---

\*e-mail: salim.bouzebda@upmc.fr

†e-mail: issam.elhattab@upmc.fr

of the entropy functional  $H(f)$ .

As a first step of our study, we gather hereafter hypotheses needed to establish our results.

**(F.1)** The functional  $H(f)$  is well-defined by (1.1), in the sense that

$$|H(f)| < \infty. \quad (1.3)$$

A kernel  $K(\cdot)$  will be any measurable function fulfilling the following conditions.

**(K.1)**  $K(\cdot)$  is of bounded variation on  $\mathbb{R}^d$ ;

**(K.2)**  $K(\cdot)$  is right continuous on  $\mathbb{R}^d$ , i.e., for any  $\mathbf{t} = (t_1, \dots, t_d)$ , we have

$$K(t_1, \dots, t_d) = \lim_{\varepsilon_1 \downarrow 0, \dots, \varepsilon_d \downarrow 0} K(t_1 + \varepsilon_1, \dots, t_d + \varepsilon_d);$$

**(K.3)**  $\|K\|_\infty := \sup_{\mathbf{t} \in \mathbb{R}^d} |K(\mathbf{t})| =: \kappa < \infty$ ;

**(K.4)**  $\int_{\mathbb{R}^d} K(\mathbf{t}) d\mathbf{t} = 1$ .

The well known Akaike-Parzen-Rosenblatt (refer to Akaike (1954), Parzen (1962) and Rosenblatt (1956)) kernel estimator of  $f(\cdot)$  is defined, for any  $\mathbf{x} \in \mathbb{R}^d$ , by

$$f_{n,h_n}(\mathbf{x}) = (nh_n^d)^{-1} \sum_{i=1}^n K((\mathbf{x} - \mathbf{X}_i)/h_n), \quad (1.4)$$

where  $0 < h_n \leq 1$  is the smoothing parameter.

In a second step, given  $f_{n,h_n}(\cdot)$ , we estimate  $H(f)$  using the representation (1.1), by setting

$$H_{n,h_n}^{(1)}(f) = - \int_{A_n} f_{n,h_n}(\mathbf{x}) \log(f_{n,h_n}(\mathbf{x})) d\mathbf{x}, \quad (1.5)$$

where  $A_n := \{\mathbf{x} : f_{n,h_n}(\mathbf{x}) \geq \gamma_n\}$ , and  $\gamma_n \downarrow 0$  is a sequence of positive constant. The *plug-in* estimator  $H_{n,h_n}^{(1)}(f)$  was introduced by Dmitriev and Tarasenko (1973) for  $d = 1$  and  $A_n = [-b_n, b_n]$ , where  $b_n$  is a specified sequence of constants. The integral estimator  $H_{n,h_n}^{(1)}(f)$  can however be easily calculated if, for example,  $f_n(\cdot)$  is a histogram. We will consider also the *resubstitution* estimate proposed Ahmad and Lin (1976). In this case, the estimator of  $H(f)$  based on the representation (1.2) is given by

$$H_{n,h_n}^{(2)}(f) = -\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \log(f_{n,h_n}(\mathbf{X}_i)), \quad (1.6)$$

where  $\Omega_{n,i} := \{f_{n,h_n}(\mathbf{X}_i) \geq \gamma_n\}$ .

The limiting behavior of  $f_{n,h_n}(\cdot)$ , for appropriate choices of the bandwidth  $h_n$ , has been extensively investigated in the literature (refer to Devroye and Györfi (1985) and Prakasa Rao (1983)).

In particular, under our assumptions, the condition that  $h_n \rightarrow 0$  together with  $nh_n \rightarrow \infty$  is necessary and sufficient for the convergence in probability of  $f_{n,h_n}(\mathbf{x})$  towards the limit  $f(\mathbf{x})$ , independently of  $\mathbf{x} \in \mathbb{R}^d$  and the density  $f(\cdot)$ . Various uniform consistency results involving the estimator  $f_{n,h_n}(\cdot)$  have been recently established. We refer to Deheuvels and Mason (2004), Einmahl and Mason (2005), Giné and Mason (2008). In this paper we will use their methods to establish convergence results for the estimates  $H_{n,h_n}^{(1)}(f)$  and  $H_{n,h_n}^{(2)}(f)$  of  $H(f)$  in the similar spirit of Bouzebda and Elhattab (2009, 2010). To prove the strong consistency of  $H_{n,h_n}^{(1)}$ , we shall consider another, but more appropriate and more computationally convenient, centering factor than the expectation  $\mathbb{E}H_{n,h_n}^{(1)}$ , which is delicate to handle. This is given by

$$\widehat{\mathbb{E}}H_{n,h_n}^{(1)}(f) = - \int_{A_n} \mathbb{E}f_{n,h_n}(\mathbf{x}) \log(\mathbb{E}f_{n,h_n}(\mathbf{x})) d\mathbf{x}.$$

The remainder of this paper is organized as follows. In Section 2, we state our main results concerning the limiting behavior of  $H_{n,h_n}^{(1)}(f)$  and  $H_{n,h_n}^{(2)}(f)$ . Some concluding remarks and possible future developments are mentioned in Section 3. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to Section 4.

## 2 Main results

The main result, concerning  $H_{n,h}^{(1)}$ , to be proved here may now be stated precisely as follows.

**Theorem 2.1** *Let  $K(\cdot)$  satisfy (K.1-2-3-4), and let  $f(\cdot)$  be a bounded density fulfill (F.1). Let  $c > 0$  and  $\{h_n\}_{n \geq 1}$  be a sequence of positive constants such that,  $cn^{-1}\gamma_n^{-4}(\log n) \leq h_n < 1$ . Then there exists a positive constant  $\Upsilon$ , such that*

$$\limsup_{n \rightarrow \infty} \sup_{h_n \leq h \leq 1} \frac{\sqrt{nh\gamma_n^4} |H_{n,h}^{(1)}(f) - \widehat{\mathbb{E}}H_{n,h}^{(1)}(f)|}{\sqrt{(\log(1/h) \vee \log \log n)}} \leq \Upsilon \text{ a.s.}$$

The proof of Theorem 2.1 is postponed until §4.

Let  $(h'_n)_{n \geq 1}$  and  $(h''_n)_{n \geq 1}$  be two sequences of constants such that  $0 < h'_n < h''_n < 1$ , together with  $h''_n \rightarrow 0$  and for any  $\beta > 0$ ,  $nh'_n\gamma_n^4 / \log n \rightarrow \infty$ , as  $n \rightarrow \infty$ . A direct application of Theorem 2.1 shows that, with probability 1,

$$\sup_{h'_n \leq h \leq h''_n} |H_{n,h}^{(1)}(f) - \widehat{\mathbb{E}}H_{n,h}^{(1)}(f)| = O\left(\sqrt{\frac{(\log(1/h'_n) \vee \log \log n)}{nh'_n\gamma_n^4}}\right).$$

This, in turn, implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |H_{n,h}^{(1)}(f) - \widehat{\mathbb{E}}H_{n,h}^{(1)}(f)| = 0 \text{ a.s.} \quad (2.1)$$

The following result handles the uniform deviation of the estimate  $H_{n,h_n}^{(1)}(f)$  with respect to  $H(f)$ .

**Corollary 2.2** Let  $K(\cdot)$  satisfy (K.1-2-3-4), and let  $f(\cdot)$  be a uniformly Lipschitz continuous and bounded density on  $\mathbb{R}^d$ , fulfilling (F.1). Then for any  $\beta > 0$ , and for each pair of sequences  $0 < h'_n < h''_n \leq 1$  with  $h''_n \rightarrow 0$ ,  $nh'_n \gamma_n^4 / \log n \rightarrow \infty$  and  $|\log(h''_n)| / \log \log n \rightarrow \infty$  as  $n \rightarrow \infty$ , we have

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |H_{n,h}^{(1)}(f) - H(f)| = 0 \quad a.s. \quad (2.2)$$

The proof of Corollary 2.2 is postponed until §4.

We note that the main problem in using entropy estimates such as (1.5) is to choose properly the smoothing parameter  $h_n$ . The uniform in bandwidth consistency result given in (2.2) shows that any choice of  $h$  between  $h'_n$  and  $h''_n$  ensures the consistency of  $H_{n,h}^{(1)}(f)$ . Now, we shall establish another result in a similar direction for a class of compactly supported densities. We need the following additional conditions.

**(F.2)**  $f(\cdot)$  has a compact support say  $\mathbb{I}$ , and there exists a constant  $0 < \mathfrak{M} < \infty$  such as

$$\sup_{\mathbf{x} \in \mathbb{I}} \left| \frac{\partial^s f(\mathbf{x})}{\partial x_1^{j_1} \dots \partial x_d^{j_d}} \right| \leq \mathfrak{M}, \quad j_1 + \dots + j_d = s.$$

**(K.5)**  $K(\cdot)$  is of order  $s$ , i.e., for some constant  $\mathfrak{S} \neq 0$ ,

$$\begin{aligned} \int_{\mathbb{R}^d} t_1^{j_1} \dots t_d^{j_d} K(\mathbf{t}) d\mathbf{t} &= 0, \quad j_1, \dots, j_d \geq 0, \quad j_1 + \dots + j_d = 1, \dots, s-1, \\ \int_{\mathbb{R}^d} |t_1^{j_1} \dots t_d^{j_d}| K(\mathbf{t}) d\mathbf{t} &= \mathfrak{S}, \quad j_1, \dots, j_d \geq 0, \quad j_1 + \dots + j_d = s. \end{aligned}$$

Under the condition (F.2), the differential entropy of  $f(\cdot)$  may be written as follows

$$H(f) = - \int_{\mathbb{I}} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x}.$$

**Theorem 2.3** Let  $K(\cdot)$  satisfy (K.1-2-3-4-5), and let  $f(\cdot)$  fulfill (F.1-2). Then for any  $\beta > 0$ , and for each pair of sequences  $0 < h'_n < h''_n \leq 1$  with  $h''_n \rightarrow 0$  and  $nh'_n \gamma_n^4 / \log n \rightarrow \infty$  as  $n \rightarrow \infty$ , we have

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh \gamma_n^4} |H_{n,h}^{(1)}(f) - H(f)|}{\sqrt{\log(1/h)} \vee \log \log n} \leq \zeta(\mathbb{I}) \quad a.s.,$$

where

$$\zeta(\mathbb{I}) = \sup_{\mathbf{x} \in \mathbb{I}} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u} \right\}^{1/2}.$$

The proof of Theorem 2.3 is postponed until §4.

To state the our result concerning  $H_{n,h_n}^{(2)}(f)$  we need the following additional conditions.

$$\mathbf{(F.3)} \quad \mathbb{E} \left[ \log^2 (f(\mathbf{X})) \right] < \infty.$$

**Remark 2.4** *Condition (F.3) is extremely weak and is satisfied by all commonly encountered distributions including many important heavy tailed distributions for which the moments do not exists (see. e.g. Song (2000)) for more details and references on the subject.*

To prove the strong consistency of  $H_{n,h_n}^{(2)}$  we consider the following centering factor

$$\widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \log (\mathbb{E}(f_{n,h_n}(\mathbf{x}) \mid \mathbf{X}_i = \mathbf{x})).$$

The main results concerning  $H_{n,h}^{(2)}(f)$  is given in the following Theorems.

**Theorem 2.5** *Let  $K(\cdot)$  satisfy (K.1-2-3-4), and let  $f(\cdot)$  be a bounded density fulfill (F.1). Let  $c > 0$  and  $\{h_n\}_{n \geq 1}$  be a sequence of positive constants such that,  $cn^{-1}\gamma_n^{-2}(\log n) \leq h_n < 1$ . Then there exists a positive constant  $\Upsilon'$ , such that*

$$\limsup_{n \rightarrow \infty} \sup_{h_n \leq h \leq 1} \frac{\sqrt{nh\gamma_n^2} |H_{n,h}^{(2)}(f) - \widehat{\mathbb{E}}H_{n,h}^{(2)}(f)|}{\sqrt{(\log(1/h) \vee \log \log n)}} \leq \Upsilon' \text{ a.s.}$$

The proof of Theorem 2.5 is postponed until §4.

**Theorem 2.6** *Assuming that the kernel function  $K(\cdot)$  is compactly supported and satisfies the conditions (K.1-2-3-4-5). Let  $f(\cdot)$  be a bounded density function fulfilling the conditions (F.1-2-3). Let  $\{h'_n\}_{n \geq 1}$  and  $\{h''_n\}_{n \geq 1}$  such that  $h'_n = An^{-\delta}$  and  $h''_n = Bn^{-\delta}$  with arbitrary choices of  $0 < A < B < \infty$  and  $(1/(d+4)) \leq \delta < 1$ . Then for each  $\beta > 0$  and  $\gamma > 0$ , we have with probability one,*

$$\limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh\gamma_n^2} |H_{n,h}^{(2)}(f) - H(f)|}{\sqrt{2 \log(1/h)}} \leq \sigma_{\mathbb{I}}, \quad (2.3)$$

where

$$\sigma_{\mathbb{I}} = \frac{1}{\gamma} \left\{ \sup_{\mathbf{x} \in \mathbb{I}} f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u} \right\}^{1/2},$$

where  $\mathbb{I}$  is given in (F.2).

The proof of Theorem 2.6 is postponed until §4.

**Remark 2.7** *Theorem 2.3 leads to the construction of asymptotic 100% certainty interval for the true entropy  $H(f)$ , using the techniques developed in Deheuvels and Mason (2004). We give in what follows, the idea how to construct this interval. Throughout, we let  $h \in [h'_n, h''_n]$ , where  $h'_n$  and  $h''_n$  are as in Theorem 2.3. We infer from Theorem 2.3 that, for suitably chosen data-dependent functions  $L_n = L_n(X_1, \dots, X_n) > 0$ , for each  $0 < \varepsilon < 1$  and for any  $\beta > 0$ , we have, as  $n \rightarrow \infty$ ,*

$$\mathbb{P} \left( \frac{1}{L_n} |H_{n,h}^{(1)}(f) - H(f)| \geq 1 + \varepsilon \right) \rightarrow 0. \quad (2.4)$$

*Assuming the validity of the statement (2.4), we obtain asymptotic certainty interval for  $H(f)$  in the following sense. For each  $0 < \varepsilon < 1$ , we have, as  $n \rightarrow \infty$ ,*

$$\mathbb{P} \left( H(f) \in \left[ H_{n,h}^{(1)}(f) - (1 + \varepsilon)L_n, H_{n,h}^{(1)}(f) + (1 + \varepsilon)L_n \right] \right) \rightarrow 1. \quad (2.5)$$

*Whenever (2.5) holds for each  $0 < \varepsilon < 1$ , we will say that the interval*

$$\left[ H_{n,h}^{(1)}(f) - L_n, H_{n,h}^{(1)}(f) + L_n \right],$$

*provides asymptotic 100% certainty interval for  $H(f)$ .*

*To construct  $L_n$  we proceed as follows. Assume that there exists a sequence  $\{\mathbb{I}_n\}_{n \geq 1}$  of strictly nondecreasing compact subsets of  $\mathbb{I}$ , such that*

$$\bigcup_{n \geq 1} \mathbb{I}_n = \mathbb{I}$$

*(for the estimation of the support  $\mathbb{I}$  we may refer to Devroye and Wise (1980) and the references therein). Furthermore, suppose that there exists a sequence (possibly random)  $\{\zeta_n(\mathbb{I}_n)\}$ ,  $n = 1, 2, \dots$ , converging to  $\zeta(\mathbb{I})$  in the sense that*

$$\mathbb{P} \left( \left| \frac{\zeta_n(\mathbb{I}_n)}{\zeta(\mathbb{I})} - 1 \right| \geq \varepsilon \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ for each } \varepsilon > 0. \quad (2.6)$$

*Observe that the statement (2.6) is satisfied when the choice*

$$\zeta_n(\mathbb{I}_n) = \sup_{\mathbf{x} \in \mathbb{I}_n} \sqrt{f_{n,h}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{u}) d\mathbf{u}}$$

*is considered. Consequently, we may define the quantity  $L_n$  displayed in the statement (2.4) by*

$$L_n = \sqrt{\frac{\gamma_n^4 (\log(1/h) \vee \log \log n)}{nh}} \times \zeta_n(\mathbb{I}_n).$$

**Remark 2.8** *A practical choice of  $\gamma_n$  is  $\beta(\log n)^{-\alpha}$  where  $\beta > 0$  and  $\alpha \geq 0$ . In the case of the density which is bounded away from 0,  $\alpha$  is equal to 0.*

**Remark 2.9** *Giné and Mason (2008) establish uniform in bandwidth consistency for the one-live-out entropy estimator, which is defined by*

$$\widehat{H}_{n,h_n} = -\frac{1}{n} \sum_{i=1}^n \log \{f_{n,h_n,-i}(X_i)\},$$

where

$$f_{n,h_n,-i}(X_i) = 1/((n-1)h_n) \sum_{1 \leq j \neq i \leq n} K((X_i - X_j)/h_n).$$

Their results hold subject to the condition that the density  $f(\cdot)$  is bounded away from 0 on its support.

### 3 Concluding remarks and future works

We have addressed the problem of nonparametric estimation of Shannon's entropy. The results presented in this work are general, since the required conditions are fulfilled by a large class of densities. Furthermore, if we assume that the density  $f(\cdot)$  is bounded away from 0 on its support, then the rate of the strong convergence is of order  $\{\{\log(1/h_n)\}/\{nh_n\}\}^{1/2}$  which is the same rate of the strong convergence for the density kernel-type estimators. The evaluation of the integral in (1.5) requires numerical integration and is not easy if  $f_{n,h_n}(\cdot)$  is a kernel density estimator but it does not involve any stochastic aspects. The integral estimator can however be easily calculated if we approximate  $f_{n,h_n}(\cdot)$  by piecewise-constant functions on a fine enough partition, for example,  $f_{n,h_n}(\cdot)$  is a histogram. We mention that in some particular case ( $K(\cdot)$  is a double exponential kernel), the approximations are easily calculated since the distribution function corresponding to the kernel  $K(\cdot)$  is available, confer Eggermont and LaRiccia (1999). An interesting aspect of the  $H_{n,h_n}^{(2)}(f)$  is that its rate of convergence is faster than that of  $H_{n,h_n}^{(1)}(f)$  and that is very easy to compute. It will be interesting to enrich our results presented here by an additional uniformity in term of  $\gamma_n$  in the supremum appearing in all our theorems, which requires non trivial mathematics, this would go well beyond the scope of the present paper. Another direction of research is to obtain results similar to that in Giné and Mason (2008) for entropy estimator under general conditions.



## 4 Proofs

**Proof of Theorem 2.1.** We first decompose  $H_{n,h_n}^{(1)}(f) - \widehat{\mathbb{E}}H_{n,h_n}^{(1)}(f)$  into the sum of two components, by writing

$$\begin{aligned}
& H_{n,h_n}^{(1)}(f) - \widehat{\mathbb{E}}H_{n,h_n}^{(1)}(f) \\
&= - \int_{A_n} f_{n,h_n}(\mathbf{x}) \log(f_{n,h_n}(\mathbf{x})) d\mathbf{x} \\
&\quad + \int_{A_n} \mathbb{E}f_{n,h_n}(\mathbf{x}) \log(\mathbb{E}f_{n,h_n}(\mathbf{x})) d\mathbf{x} \\
&= - \int_{A_n} \{\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E}f_{n,h_n}(\mathbf{x})\} \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\
&\quad - \int_{A_n} \{f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})\} \log f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\
&:= \Delta_{1,n,h_n} + \Delta_{2,n,h_n}. \tag{4.1}
\end{aligned}$$

We observe that for all  $z > 0$ ,  $|\log z| \leq \left|\frac{1}{z} - 1\right| + |z - 1|$ . Therefore, we get

$$\begin{aligned}
& |\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E}f_{n,h_n}(\mathbf{x})| = \left| \log \frac{f_{n,h_n}(\mathbf{x})}{\mathbb{E}f_{n,h_n}(\mathbf{x})} \right| \\
&\leq \left| \frac{\mathbb{E}f_{n,h_n}(\mathbf{x})}{f_{n,h_n}(\mathbf{x})} - 1 \right| + \left| \frac{f_{n,h_n}(\mathbf{x})}{\mathbb{E}f_{n,h_n}(\mathbf{x})} - 1 \right| \\
&= \frac{|\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})|}{f_{n,h_n}(\mathbf{x})} + \frac{|f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})|}{\mathbb{E}f_{n,h_n}(\mathbf{x})}.
\end{aligned}$$

Recalling that  $A_n := \{\mathbf{x} : f_{n,h_n}(\mathbf{x}) \geq \gamma_n\}$ , we readily obtain from these relations that, for any  $\mathbf{x} \in A_n$ ,

$$|\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E}f_{n,h_n}(\mathbf{x})| \leq \frac{2}{\gamma_n} |f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})|.$$

We can therefore write, for any  $n \geq 1$ , the inequalities

$$\begin{aligned}
|\Delta_{1,n,h_n}| &= \left| \int_{A_n} \{\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E}f_{n,h_n}(\mathbf{x})\} \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x} \right| \\
&\leq \int_{A_n} |\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E}f_{n,h_n}(\mathbf{x})| \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\
&\leq \frac{2}{\gamma_n} \int_{A_n} |f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})| \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\
&\leq \frac{2}{\gamma_n} \sup_{\mathbf{x} \in A_n} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})| \int_{A_n} \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\
&\leq \frac{2}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})| \int_{\mathbb{R}^d} \mathbb{E}f_{n,h_n}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

In view of (K.4), we have,

$$\int_{\mathbb{R}^d} \mathbb{E}f_{n,h}(\mathbf{x}) d\mathbf{x} = 1.$$

Thus, for any  $n \geq 1$ , we have

$$|\Delta_{1,n,h_n}| \leq \frac{2}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})|. \quad (4.2)$$

We next evaluate the second term  $\Delta_{2,n,h_n}$  in the right side of (4.1). Since  $|\log z| \leq \frac{1}{z} + z$ , for all  $z > 0$ , we see that

$$\begin{aligned} |\Delta_{2,n,h_n}| &= \left| \int_{A_n} \{f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})\} \log f_{n,h_n}(\mathbf{x}) d\mathbf{x} \right| \\ &\leq \int_{A_n} |f_{n,h_n}(\mathbf{x}) - \mathbb{E}f_{n,h_n}(\mathbf{x})| \left[ \frac{1}{f_{n,h_n}(\mathbf{x})} + f_{n,h_n}(\mathbf{x}) \right] d\mathbf{x}. \end{aligned}$$

Similarly as above, we get, for any  $\mathbf{x} \in A_n$ ,

$$\begin{aligned} \frac{1}{f_{n,h_n}(\mathbf{x})} + f_{n,h_n}(\mathbf{x}) &= \left( \frac{1}{f_{n,h_n}(\mathbf{x})f_{n,h_n}(\mathbf{x})} + 1 \right) f_{n,h_n}(\mathbf{x}) \\ &\leq \left( \frac{1}{\gamma_n^2} + 1 \right) f_{n,h_n}(\mathbf{x}). \end{aligned}$$

We can therefore write, for any  $n \geq 1$ ,

$$\begin{aligned} |\Delta_{2,n,h_n}| &\leq \left( \frac{1}{\gamma_n^2} + 1 \right) \int_{A_n} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})| f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\ &\leq \left( \frac{1}{\gamma_n^2} + 1 \right) \sup_{\mathbf{x} \in A_n} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})| \int_{A_n} f_{n,h_n}(\mathbf{x}) d\mathbf{x} \\ &\leq \left( \frac{1}{\gamma_n^2} + 1 \right) \sup_{\mathbf{x} \in A_n} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})| \int_{\mathbb{R}^d} f_{n,h_n}(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

In view of (K.4), we have,

$$\int_{\mathbb{R}^d} f_{n,h}(\mathbf{x}) d\mathbf{x} = 1.$$

Thus, for any  $n \geq 1$ , we have

$$|\Delta_{2,n,h_n}| \leq \left( \frac{1}{\gamma_n^2} + 1 \right) \sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f_{n,h_n}(\mathbf{x})|. \quad (4.3)$$

We now impose some slightly more general assumptions on the kernel  $K(\cdot)$  than that of Theorem 2.1. Consider the class of functions

$$\mathcal{K} = \left\{ K((\mathbf{x} - \cdot)/h^{1/d}) : h > 0, \mathbf{x} \in \mathbb{R}^d \right\}.$$

For  $\varepsilon > 0$ , set  $N(\varepsilon, \mathcal{K}) = \sup_Q N(\varepsilon, \mathcal{K}, d_Q)$ , where the supremum is taken over all probability measures  $Q$  on  $(\mathbb{R}^d, \mathcal{B})$ . Here,  $d_Q$  denotes the  $L_2(Q)$ -metric and  $N(\varepsilon, \mathcal{K}, d_Q)$  is the minimal number of balls  $\{g : d_Q(g, g') < \varepsilon\}$  of  $d_Q$ -radius  $\varepsilon$  needed to cover  $\mathcal{K}$ . We assume that  $\mathcal{K}$  satisfies the following uniform entropy condition.

(K.6) for some  $C > 0$  and  $\nu > 0$ ,

$$N(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-\nu}, 0 < \varepsilon < 1. \quad (4.4)$$

Finally, to avoid using outer probability measures in all of statements, we impose the following measurability assumption.

(K.7)  $\mathcal{K}$  is a pointwise measurable class, that is, there exists a countable subclass  $\mathcal{K}_0$  of  $\mathcal{K}$  such that we can find for any function  $g \in \mathcal{K}$  a sequence of functions  $\{g_m : m \geq 1\}$  in  $\mathcal{K}_0$  for which

$$g_m(\mathbf{z}) \longrightarrow g(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d.$$

Remark that condition (K.6) is satisfied whenever (K.1) holds, i.e.,  $K(\cdot)$  is of bounded variation on  $\mathbb{R}^d$  (in the sense of Hardy and Kauser, see, e.g. Clarkson and Adams (1933), Vituškin (1955) and Hobson (1958)). Condition (K.7) is satisfied whenever (K.2) holds, i.e.,  $K(\cdot)$  is right continuous (refer to Deheuvels and Mason (2004) and Einmahl and Mason (2005) and the references therein).

By Theorem 1 of Einmahl and Mason (2005), whenever  $K(\cdot)$  is measurable and satisfies (K.3-4-6-7), and when  $f(\cdot)$  is bounded, we have for each  $c > 0$ , and for a suitable function  $\Sigma(c)$ , with probability 1,

$$\limsup_{n \rightarrow \infty} \sup_{cn^{-1} \log n \leq h \leq 1} \frac{\sqrt{nh} \|f_{n,h} - \mathbb{E}f_{n,h}\|_\infty}{\sqrt{\log(1/h) \vee \log \log n}} = \Sigma(c) < \infty, \quad (4.5)$$

which implies, in view of (4.2) and (4.3), that, with probability 1,

$$\limsup_{n \rightarrow \infty} \sup_{h_n \leq h < 1} \frac{\sqrt{nh\gamma_n^4} |\Delta_{1,n,h}|}{\sqrt{(\log(1/h) \vee \log \log n)}} = 0, \quad (4.6)$$

and

$$\limsup_{n \rightarrow \infty} \sup_{h_n \leq h < 1} \frac{\sqrt{nh\gamma_n^4} |\Delta_{2,n,h}|}{\sqrt{(\log(1/h) \vee \log \log n)}} \leq \Upsilon(c). \quad (4.7)$$

Recalling (4.1), the proof of Theorem 2.1 is completed by combining (4.6) with (4.7).  $\blacksquare$

**Proof of Corollary 2.2.** Recall  $A_n = \{\mathbf{x} : f_{n,h_n}(\mathbf{x}) \geq \gamma_n\}$  and let  $A_{n,\beta}^c$  the complement of  $A_n$  in  $\mathbb{R}^d$  (i.e.,  $A_{n,\beta}^c = \{\mathbf{x} : f_{n,h_n}(\mathbf{x}) < \gamma_n\}$ ). Observe that

$$|f(\mathbf{x})| \geq |f_{n,h_n}(\mathbf{x})| - |f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})| \geq \gamma_n + O(h_n''^{1/d}).$$

Recall that  $|\log(h_n'')|/\log \log n \rightarrow \infty$  as  $n \rightarrow \infty$ , thus, for  $n$  enough large, the second term of the last inequality is dominated by the first one, then, we obtain

$$|f(\mathbf{x})| \geq \gamma_n.$$

We repeat the arguments above with the formal change of  $H_{n,h_n}^{(1)}(f)$  by  $H(f)$ . We show that, for any  $n \geq 1$ ,

$$\begin{aligned}
& |\widehat{\mathbb{E}}H_{n,h_n}^{(1)}(f) - H(f)| \\
& \leq \left| \int_{A_{n,\beta}^c} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} \right| \\
& \quad + \frac{1}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})| \\
& \quad + \left( \frac{1}{\gamma_n^2} + 1 \right) \sup_{\mathbf{x} \in \mathbb{R}^d} |\mathbb{E}f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})|. \tag{4.8}
\end{aligned}$$

We have

$$\begin{aligned}
\int_{A_{n,\beta}^c} f(\mathbf{x}) d\mathbf{x} & \leq \int_{\frac{1}{2}f(\mathbf{x}) \leq \gamma_n} f(\mathbf{x}) d\mathbf{x} + \int_{f_{n,h}(\mathbf{x}) \leq \gamma_n \leq \frac{1}{2}f(\mathbf{x})} f(\mathbf{x}) d\mathbf{x} \\
& \leq \int_{\frac{1}{2}f(\mathbf{x}) \leq \gamma_n} f(\mathbf{x}) d\mathbf{x} + 2 \int_{\mathbb{R}^d} |f_{n,h}(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x}.
\end{aligned}$$

Observe that we have

$$\mathbb{1}_{\{\frac{1}{2}f(\mathbf{x}) \leq \gamma_n\}} f(\mathbf{x}) \leq f(\mathbf{x})$$

and  $\mathbb{1}_{\{\frac{1}{2}f(\mathbf{x}) \leq \gamma_n\}} f(\mathbf{x}) \rightarrow 0$  as  $n \rightarrow \infty$ , thus an application of Lebesgue dominated convergence theorem gives

$$\lim_{n \rightarrow \infty} \int_{\frac{1}{2}f(\mathbf{x}) \leq \gamma_n} f(\mathbf{x}) d\mathbf{x} = 0. \tag{4.9}$$

Recall that the conditions  $h_n \rightarrow 0$  together with  $nh_n \rightarrow \infty$  as  $n \rightarrow \infty$ , ensure that (see e.g., Devroye and Györfi (1985))

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} = 0 \quad a.s.$$

Thus, for all  $h \in [h'_n, h''_n]$  such that  $nh'_n \rightarrow \infty$  and  $h''_n \rightarrow 0$ , as  $n \rightarrow \infty$ , we have

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \int_{\mathbb{R}^d} |f_{n,h}(\mathbf{x}) - f(\mathbf{x})| d\mathbf{x} = 0 \quad a.s. \tag{4.10}$$

By (4.10) and (4.9) we have

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \int_{A_{n,\beta}^c} f(\mathbf{x}) d\mathbf{x} = 0 \quad a.s. \tag{4.11}$$

Since  $H(f)$  is finite, the measure

$$\nu(A) = \int_A |\log(f(\mathbf{x}))| d\mathbb{F}(\mathbf{x}),$$

is absolutely continuous with respect to the measure  $\mu(A) = \int_A d\mathbb{F}(\mathbf{x})$ , which implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \int_{A_{n,\beta}^c} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} = 0 \quad a.s. \quad (4.12)$$

In the other hand, we know (see, e.g, Einmahl and Mason (2005)), that when the density  $f(\cdot)$  is uniformly Lipschitz and continuous, we have for each  $h'_n < h < h''_n$ , as  $n \rightarrow \infty$ ,

$$\|\mathbb{E}f_{n,h}(\mathbf{x}) - f(\mathbf{x})\|_\infty = O(h_n''^{1/d}). \quad (4.13)$$

Thus, we have

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \gamma_n^{-2} \|\mathbb{E}f_{n,h}(\mathbf{x}) - f(\mathbf{x})\|_\infty = 0.$$

This when combined with (4.8), entails that, as  $n \rightarrow \infty$ ,

$$\sup_{h'_n \leq h \leq h''_n} \|\widehat{\mathbb{E}}H_{n,h}^{(1)}(f) - H(f)\| \rightarrow 0. \quad (4.14)$$

By (4.11) and (4.14) in connection with (2.1) imply (2.2). ■

**Proof of Theorem 2.3.** Under conditions (F.2), (K.5) and using Taylor expansion of order  $s$  we get, for  $\mathbf{x} \in \mathbb{I}$ ,

$$|\mathbb{E}f_{n,h}(\mathbf{x}) - f(\mathbf{x})| = \frac{h^{s/d}}{s!} \left| \int \sum_{k_1 + \dots + k_d = s} t_1^{k_1} \dots t_d^{k_d} \frac{\partial^s f(\mathbf{x} - h\theta\mathbf{t})}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} K(\mathbf{t}) d\mathbf{t} \right|,$$

where  $\theta = (\theta_1, \dots, \theta_d)$  and  $0 < \theta_i < 1$ ,  $i = 1, \dots, d$ . Thus a straightforward application of Lebesgue dominated convergence theorem gives, for  $n$  large enough,

$$\sup_{\mathbf{x} \in \mathbb{I}} |\mathbb{E}f_{n,h}(\mathbf{x}) - f(\mathbf{x})| = O(h_n''^{s/d}). \quad (4.15)$$

Let  $\mathbb{J}$  be a nonempty compact subset of the interior of  $\mathbb{I}$  (say  $\mathring{\mathbb{I}}$ ). First, note that we have

$$\limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \sup_{\mathbf{x} \in \mathbb{J}} \frac{\sqrt{nh} |f_{n,h}(\mathbf{x}) - f(\mathbf{x})|}{\sqrt{\log(1/h)} \vee \log \log n} = \sup_{\mathbf{x} \in \mathbb{J}} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}. \quad (4.16)$$

Set, for all  $n \geq 1$ ,

$$\pi_n(\mathbb{J}) = \left| \int_{\mathbb{J}} f_{n,h_n}(\mathbf{x}) \log(f_{n,h_n}(\mathbf{x})) d\mathbf{x} - \int_{\mathbb{J}} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x} \right|. \quad (4.17)$$

Using condition (F.2) ( $f(\cdot)$  is compactly supported),  $f(\cdot)$  is bounded away from zero on its support, thus, we have for  $n$  enough large,  $f(\mathbf{x}) > 2_n$ , for all  $\mathbf{x}$  in the support of  $f(\cdot)$ . By the same previous arguments we have, for  $n$  enough large,

$$\begin{aligned} \pi_n(\mathbb{J}) &\leq \frac{1}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{J}} |f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})| \\ &\quad + \left( \frac{1}{\gamma_n^2} + 1 \right) \sup_{\mathbf{x} \in \mathbb{J}} |f_{n,h_n}(\mathbf{x}) - f(\mathbf{x})|. \end{aligned}$$

One finds, by combining the last equation with (4.16),

$$\limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh} \gamma_n^4 \pi_n(\mathbb{J})}{\sqrt{(\log(1/h) \vee \log \log n)}} \leq \sup_{\mathbf{x} \in \mathbb{J}} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}. \quad (4.18)$$

Let  $\{\mathbb{J}_\ell\}$ ,  $\ell = 1, 2, \dots$ , be a sequence of nondecreasing nonempty compact subsets of  $\mathbb{I}$  such that

$$\bigcup_{\ell \geq 1} \mathbb{J}_\ell = \mathbb{I}.$$

Now, from (4.18), it is straightforward to observe that

$$\begin{aligned} \lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh} \gamma_n^4 \pi_n(\mathbb{J}_\ell)}{\sqrt{(\log(1/h) \vee \log \log n)}} &\leq \lim_{\ell \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{J}_\ell} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2} \\ &\leq \sup_{\mathbf{x} \in \mathbb{I}} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t} \right\}^{1/2}. \end{aligned}$$

The proof of Corollary 2.3 is completed. ■

**Proof of Theorem 2.5.** Let  $\varphi_{n,h_n}(\mathbf{x}) = \mathbb{E}(f_{n,h_n}(\mathbf{x}))$ . Recall that

$$\begin{aligned} H_{n,h_n}^{(2)}(f) - \widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f) &= -\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \log(f_{n,h_n}(\mathbf{X}_i)) + \mathbb{1}_{\Omega_{n,i}} \log(\varphi_{n,h_n}(\mathbf{X}_i)) \\ &=: \Xi_{n,h_n}. \end{aligned}$$

Using a Taylor-Lagrange expansion of the  $\log(\cdot)$  function, we have, for some random sequence  $\theta_n \in (0, 1)$ ,

$$\Xi_{n,h_n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \left[ \frac{f_{n,h_n}(\mathbf{X}_i) - \varphi_{n,h_n}(\mathbf{X}_i)}{(1 - \theta_n)f_{n,h_n}(\mathbf{X}_i) + \theta_n \varphi_{n,h_n}(\mathbf{X}_i)} \right].$$

Recalling that  $\Omega_{n,i} = \{f_{n,h_n}(\mathbf{X}_i) \geq \gamma_n\}$ , we readily obtain, with probability 1,

$$\begin{aligned} |\Xi_{n,h_n}| &\leq \frac{1}{n\gamma_n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} |f_{n,h_n}(\mathbf{X}_i) - \varphi_{n,h_n}(\mathbf{X}_i)| \\ &\leq \frac{1}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{I}} |f_{n,h_n}(\mathbf{x}) - \varphi_{n,h_n}(\mathbf{x})| \\ &= \frac{1}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{I}} |f_{n,h_n}(\mathbf{x}) - \mathbb{E}(f_{n,h_n}(\mathbf{x}))|. \end{aligned}$$

Combining the last inequality with (4.5), we readily obtain the desired result. ■

**Proof of Theorem 2.6.** We have

$$H_{n,h_n}^{(2)}(f) - H(f) = \{H_{n,h_n}^{(2)}(f) - \widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f)\} + \{\widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f) - H(f)\}.$$

Since the first term in the right hand of the last equality is controlled in the preceding proof, it remains only to evaluate the second one. To simplify our exposition, we will decompose  $\widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f) -$

$H(f)$  into the sum of three components,

$$\begin{aligned}
\widehat{\mathbb{E}}H_{n,h_n}^{(2)}(f) - H(f) &= -\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \log(\varphi_{n,h_n}(\mathbf{X}_i)) + \mathbb{E}(\log(f(\mathbf{X}_i))) \\
&= -\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} (\log(\varphi_{n,h_n}(\mathbf{X}_i)) - \log(f(\mathbf{X}_i))) \\
&\quad -\frac{1}{n} \sum_{i=1}^n (\mathbb{1}_{\Omega_{n,i}} \log(f(\mathbf{X}_i)) - \log(f(\mathbf{X}_i))) \\
&\quad -\frac{1}{n} \sum_{i=1}^n (\log(f(\mathbf{X}_i)) - \mathbb{E}(\log(f(\mathbf{X}_i)))) \\
&=: -\nabla_{1,n,h_n} - \nabla_{2,n,h_n} - \nabla_{3,n,h_n}.
\end{aligned} \tag{4.19}$$

In view of (4.19), we have

$$\nabla_{1,n,h_n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} (\log(\varphi_{n,h_n}(\mathbf{X}_i)) - \log(f(\mathbf{X}_i))).$$

Using a Taylor-Lagrange expansion of the  $\log(\cdot)$  function, we have, for some random sequence  $\theta_n \in (0, 1)$ ,

$$\nabla_{1,n,h_n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} \left[ \frac{\varphi_{n,h_n}(\mathbf{X}_i) - f(\mathbf{X}_i)}{(1 - \theta_n)\varphi_{n,h_n}(\mathbf{X}_i) + \theta_n f(\mathbf{X}_i)} \right].$$

By (F.2), there exists an  $\eta_{\mathbb{I}}$  such that  $f(\mathbf{x}) > \eta_{\mathbb{I}}$  for all  $\mathbf{x} \in \mathbb{I}$ . It follows that for  $n$  enough large that,  $f(\mathbf{x}) > \gamma_n$  for all  $\mathbf{x} \in \mathbb{I}$ . Recalling that  $\Omega_{n,i} = \{f_{n,h_n}(\mathbf{X}_i) \geq \gamma_n\}$ , we readily obtain, with probability 1,

$$\begin{aligned}
|\nabla_{1,n,h_n}| &\leq \frac{1}{n\gamma_n} \sum_{i=1}^n \mathbb{1}_{\Omega_{n,i}} |\varphi_{n,h_n}(\mathbf{X}_i) - f(\mathbf{X}_i)| \\
&\leq \frac{1}{\gamma_n} \sup_{\mathbf{x} \in \mathbb{I}} |\varphi_{n,h_n}(\mathbf{x}) - f(\mathbf{x})|.
\end{aligned}$$

We mention that the bandwidth  $h$  is to be chosen in such a way that the bias of  $f_{n,h}(\mathbf{x})$  may be neglected, in the sense that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \left\{ \frac{nh}{2 \log(1/h)} \right\}^{1/2} \sup_{\mathbf{x} \in \mathbb{I}} |\varphi_{n,h}(\mathbf{x}) - f(\mathbf{x})| = 0, \tag{4.20}$$

which is implied by (4.15). Thus,

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh\gamma_n^2} |\nabla_{1,n,h}|}{\sqrt{2 \log(1/h)}} = 0. \tag{4.21}$$

We next evaluate the second term  $\nabla_{2,n,h_n}$  in the right side of (4.19). We have from (4.15) and (4.5)

$$\sup_{h'_n \leq h \leq h''_n} \sup_{\mathbf{x} \in \mathbb{I}} |f_{n,h}(\mathbf{x}) - f(\mathbf{x})| = O\left(\sqrt{\frac{(\log(1/h'_n))}{nh'_n}}\right).$$

Thus, for  $n$  sufficiently large, almost surely,  $f_{n,h}(\mathbf{x}) \geq (1/2)f(\mathbf{x})$  for all  $\mathbf{x} \in \mathbb{I}$  and all  $h \in [h'_n, h''_n]$ . Note that under condition (F.2), the density  $f(\cdot)$  is compactly supported, it is possible to find a positive constant  $\eta_{\mathbb{I}}$  such as  $f(\mathbf{x}) > \eta_{\mathbb{I}}$ . This implies that  $f_{n,h}(\mathbf{x}) \geq \eta_{\mathbb{I}}/2$ , and thus, for all  $n$  enough large, we have, almost surely,

$$\mathbb{1}_{\Omega_{n,i}} = 1, \quad (4.22)$$

which implies that, for all  $n$  enough large, almost surely,

$$\nabla_{2,n,h_n} = 0. \quad (4.23)$$

We finally evaluate the second term  $\nabla_{3,n,h_n}$  in the right side of (4.19). We have,

$$\nabla_{3,n,h_n} = -\frac{1}{n} \sum_{i=1}^n \xi_i,$$

where, for  $i = 1, \dots, n$ ,

$$\xi_i := \log\{f(\mathbf{X}_i)\} - \mathbb{E}\left(\log\{f(\mathbf{X}_i)\}\right),$$

are a centered i.i.d. random variables with finite variance  $\text{Var}(\log(f(\mathbf{X}_i)))$  (condition (F.3)). Observe that

$$\frac{\gamma_n \sqrt{nh_n} \sum_{i=1}^n \xi_i}{n \sqrt{2 \log(1/h_n)}} = \frac{\gamma_n \sqrt{h_n \log \log n} \sum_{i=1}^n \xi_i}{\sqrt{\log(1/h_n)} \sqrt{2n \log \log n}}$$

which, by the law of the iterated logarithm, tends to 0 as  $n$  tends to infinity. Namely,

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh \gamma_n^2} |\nabla_{3,n,h}|}{\sqrt{2 \log(1/h)}} = 0. \quad (4.24)$$

Using (4.24) and (4.23) in connection with Fact 1 completes the proof of Theorem 2.6. ■

## References

- Ahmad, I. A. and Lin, P. E. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Information Theory*, **IT-22**(3), 372–375.
- Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math., Tokyo*, **6**, 127–132.
- Beirlant, J., Dudewicz, E. J., Györfi, L., and van der Meulen, E. C. (1997). Nonparametric entropy estimation: an overview. *Int. J. Math. Stat. Sci.*, **6**(1), 17–39.



- Bouzebda, S. and Elhattab, I. (2009). A strong consistency of a nonparametric estimate of entropy under random censorship. *C. R. Math. Acad. Sci. Paris*, **347**(13-14), 821–826.
- Bouzebda, S. and Elhattab, I. (2010). Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy. *C. R. Math. Acad. Sci. Paris*, **348**(5-6), 317–321.
- Clarkson, J. A. and Adams, C. R. (1933). On definitions of bounded variation for functions of two variables. *Trans. Amer. Math. Soc.*, **35**(4), 824–854.
- Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, **7**(3), 225–277.
- Devroye, L. and Györfi, L. (1985). *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The  $L_1$  view.
- Devroye, L. and Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.*, **38**(3), 480–488.
- Dmitriev, J. G. and Tarasenko, F. P. (1973). The estimation of functionals of a probability density and its derivatives. *Teor. Veroyatnost. i Primenen.*, **18**, 662–668.
- Eggermont, P. P. B. and LaRiccia, V. N. (1999). Best asymptotic normality of the kernel density entropy estimator for smooth densities. *IEEE Trans. Inform. Theory*, **45**(4), 1321–1326.
- Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**(3), 1380–1403.
- Giné, E. and Mason, D. M. (2008). Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.*, **35**(4), 739–761.
- Hobson, E. W. (1958). *The theory of functions of a real variable and the theory of Fourier's series*. Vol. I. Dover Publications Inc., New York, N.Y.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
- Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation*. Probability and Mathematical Statistics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–837.

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

Song, K.-S. (2000). Limit theorems for nonparametric sample entropy estimators. *Statist. Probab. Lett.*, **49**(1), 9–18.

Vituškin, A. G. (1955). *O mnogomernyh variaciyah*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow.