Chapter 14
Strong Universal Consistent Estimate of the Minimum Mean Squared Error

Luc Devroye, Paola G. Ferrario, László Györfi, and Harro Walk

Abstract Consider the regression problem with a response variable Y and a feature vector $\mathbf{X}$. For the regression function $m(\mathbf{x}) = \mathbf{E}\{Y \mid \mathbf{X} = \mathbf{x}\}$, we introduce new and simple estimators of the minimum mean squared error $L^* = \mathbf{E}\{(Y - m(\mathbf{X}))^2\}$, and prove their strong consistencies. We bound the rate of convergence, too.

14.1 Introduction

Let the label Y be a real-valued random variable and let the feature vector $\mathbf{X} = (X_1, \ldots, X_d)$ be a d-dimensional random vector. The regression function m is defined by

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

L. Devroye (✉)
School of Computer Science, McGill University, 3480 University Street, Montreal, H3A OE9, Canada
e-mail: lucdevroye@gmail.com

P.G. Ferrario · H. Walk
Department of Mathematics, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany
e-mail: paola.ferrario@mathematik.uni-stuttgart.de; walk@mathematik.uni-stuttgart.de

L. Györfi
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Stoczek u. 2, 1521 Budapest, Hungary
e-mail: gyorfi@cs.bme.hu

B. Schölkopf et al. (eds.), *Empirical Inference*, DOI 10.1007/978-3-642-41136-6_14,

The minimum mean squared error, called also variance of the residual $Y - m(\mathbf{X})$, is denoted by

$$L^* := \mathbf{E}\{(Y - m(\mathbf{X}))^2\} = \min_f \mathbf{E}\{(Y - f(\mathbf{X}))^2\}.$$

The regression function m and the minimum mean squared error L^* cannot be calculated when the distribution of $(\mathbf{X}, Y)$ is unknown. Assume, however, that we observe data $D_n = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)\}$ consisting of independent and identically distributed copies of $(\mathbf{X}, Y)$. D_n can be used to produce an estimate of L^*.

For nonparametric estimates of the minimum mean squared error $L^* = \mathbf{E}\{(Y - m(\mathbf{X}))^2\}$ see, e.g., Dudoit and van der Laan [4], Liitiäinen et al. [9–11], Müller and Stadtmüller [12], Neumann [14], Stadtmüller and Tsybakov [15], and Müller, Schick and Wefelmeyer [13] and the literature cited there.

Devroye et al. [3] proved that without any tail and smoothness condition, L^* cannot be estimated with a guaranteed rate of convergence. They introduced a modified nearest neighbour cross-validation estimate

$$\hat{L}_n = \frac{1}{2n} \sum_{i=1}^{n} (Y_i - Y_{j(i)})^2, \quad n \geq 2,$$

where $Y_{j(i)}$ is the label of the modified first nearest neighbour of $\mathbf{X}_i$ from among $\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n$. If Y and $\mathbf{X}$ are bounded, and m is Lipschitz continuous

$$|m(\mathbf{x}) - m(\mathbf{z})| \leq C \|\mathbf{x} - \mathbf{z}\|, \tag{14.1}$$

then for $d \geq 3$, they proved that

$$\mathbf{E}\{|\hat{L}_n - L^*|\} \leq c_1 n^{-1/2} + c_2 n^{-2/d}. \tag{14.2}$$

Liitiäinen et al. [9, 11] introduced another estimate of the minimum mean squared error L^* by the first and second nearest neighbour cross-validation

$$L_n = \frac{1}{n} \sum_{i=1}^{n} (Y_i - Y_{n,i,1})(Y_i - Y_{n,i,2}), \quad n \geq 3,$$

where $Y_{n,i,1}$ and $Y_{n,i,2}$ are the labels of the first and second nearest neighbours $\mathbf{X}_{n,i,1}$ and $\mathbf{X}_{n,i,2}$ of $\mathbf{X}_i$ from among $\mathbf{X}_1, \dots, \mathbf{X}_{i-1}$, and $\mathbf{X}_{i+1}, \dots, \mathbf{X}_n$, resp. (In the sequel, assume that for calculating the first and second nearest neighbours, ties occur with probability 0. When $\mathbf{X}$ has a density, the case of ties among nearest neighbour distances occurs with probability 0.) If Y and $\mathbf{X}$ are bounded and m is Lipschitz continuous, then for $d \geq 2$, they proved the rate of convergence of order in the inequality (14.2).

In this chapter we introduce a non-recursive and a recursive estimator of the minimum mean squared error L^*, and prove their distribution-free strong consistencies. Under some mild conditions on the regression function m and on the distribution of $(\mathbf{X}, Y)$, we bound the rate of convergence of the non-recursive estimate.

14.2 Strong Universal Consistency

One can derive a new and simple estimator of L^*, considering the definition

$$L^* = \mathbf{E}\{(Y - m(\mathbf{X}))^2\} = \mathbf{E}\{Y^2\} - \mathbf{E}\{m(\mathbf{X})^2\}.$$

Obviously, $\mathbf{E}\{Y^2\}$ can be estimated by $\frac{1}{n}\sum_{i=1}^n Y_i^2$, while we estimate the term $\mathbf{E}\{m(\mathbf{X})^2\}$ by $\frac{1}{n}\sum_{i=1}^n Y_i Y_{n,i,1}$. Thus we estimate L^* by

$$\tilde{L}_n := \frac{1}{n}\sum_{i=1}^n Y_i^2 - \frac{1}{n}\sum_{i=1}^n Y_i Y_{n,i,1}.$$

Theorem 14.1. *Assume that ties occur with probability 0. If $|Y|$ is bounded then*

$$\tilde{L}_n \to L^* \quad a.s.$$

If $\mathbf{E}\{Y^2\} < \infty$ then

$$\bar{L}_n := \frac{1}{n}\sum_{k=1}^n \tilde{L}_k \to L^* \quad a.s.$$

Proof. This theorem says that, for bounded $|Y|$, the estimate $\tilde{L}_n$ is strongly consistent, while the estimate $\bar{L}_n$ is strongly universally consistent. The theorem is an easy consequence of Ferrario and Walk [6] (Theorems 2.1 and 2.5), who proved that, for bounded Y,

$$L_n \to L^* \tag{14.3}$$

a.s., and moreover, under the only condition $\mathbf{E}\{Y^2\} < \infty$,

$$\frac{1}{n}\sum_{k=1}^n L_k \to L^* \tag{14.4}$$

a.s. We simply use the decomposition

$$L_n = \tilde{L}_n - \frac{1}{n}\sum_{i=1}^{n} Y_i Y_{n,i,2} + \frac{1}{n}\sum_{i=1}^{n} Y_{n,i,1} Y_{n,i,2}.$$

Then, as in the proof of Theorem 2.1 in Ferrario and Walk [6], on the basis of (21)–(25) in [9], one can show that, for bounded Y,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i Y_{n,i,2} \to \mathbf{E}\{m(\mathbf{X})^2\} \tag{14.5}$$

a.s. and

$$\frac{1}{n}\sum_{i=1}^{n} Y_{n,i,1} Y_{n,i,2} \to \mathbf{E}\{m(\mathbf{X})^2\} \tag{14.6}$$

a.s. Similarly, as in the proof of Theorem 2.5 in [6], for $\mathbf{E}\{Y^2\} < \infty$, one can show that

$$\frac{1}{n}\sum_{k=1}^{n} \frac{1}{k}\sum_{i=1}^{k} Y_i Y_{k,i,2} \to L^* \tag{14.7}$$

a.s. and

$$\frac{1}{n}\sum_{k=1}^{n} \frac{1}{k}\sum_{i=1}^{k} Y_{k,i,1} Y_{k,i,2} \to L^* \tag{14.8}$$

a.s. Now the statements of the theorem follow from (14.3), (14.5), and (14.6), and from (14.4), (14.7) and (14.8), respectively. □

Next we consider a recursive estimate

$$L'_n := \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\sum_{i=1}^{n} Y_i Y_{i,i,1}, \quad n \geq 2,$$

where $Y_{1,1,1} := 0$. It is really recursive since

$$L'_n = \left(1 - \frac{1}{n}\right) L'_{n-1} + \frac{1}{n}(Y_n^2 - Y_n Y_{n,n,1}).$$

Theorem 14.2. *Assume that ties occur with probability 0. If* $\mathbf{E}\{Y^2\} < \infty$ *then*

$$L'_n \to L^* \quad a.s.$$

Proof. We have to show that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i Y_{i,i,1} \to \mathbf{E}\{m(\mathbf{X})^2\} \quad \text{a.s.} \tag{14.9}$$

For $a > 0$, introduce the truncation function

$$T_a(z) = \begin{cases} a & \text{if } z > a; \\ z & \text{if } |z| < a; \\ -a & \text{if } z < -a. \end{cases}$$

As in to the proof of Theorem 2.5 in Ferrario and Walk [6], one can check that in order to show (14.9), it suffices to prove that

$$\frac{1}{n}\sum_{i=1}^{n} T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,i,1}) \to \mathbf{E}\{m(\mathbf{X})^2\}$$

a.s. Let $\mathcal{F}_{i-1}$ be the σ-algebra generated by $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_{i-1}, Y_{i-1})$. Introduce the decomposition

$$\frac{1}{n}\sum_{i=1}^{n} T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,i,1}) = I_n + J_n,$$

where

$$I_n = \frac{1}{n}\sum_{i=1}^{n} (T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,i,1}) - \mathbf{E}\{T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,i,1}) \mid \mathcal{F}_{i-1}\})$$

and

$$J_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\{T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,i,1}) \mid \mathcal{F}_{i-1}\}.$$

I_n is an average of martingale differences such that the a.s. convergence

$$I_n \to 0 \tag{14.10}$$

can be derived from the Chow [1] theorem if

$$\sum_{n=1}^{\infty} \frac{\mathbf{Var}\{T_{\sqrt{n}}(Y_n) T_{\sqrt{n}}(Y_{n,n,1})\}}{n^2} < \infty. \tag{14.11}$$

We have that

$$\mathbf{Var}\{T_{\sqrt{n}}(Y_n)T_{\sqrt{n}}(Y_{n,n,1})\} \leq \mathbf{E}\{(T_{\sqrt{n}}(Y_n))^2(T_{\sqrt{n}}(Y_{n,n,1}))^2\}$$
$$\leq \frac{\mathbf{E}\{(T_{\sqrt{n}}(Y_n))^4\} + \mathbf{E}\{(T_{\sqrt{n}}(Y_{n,n,1}))^4\}}{2}.$$

Because of $\mathbf{E}\{Y_1^2\} < \infty$,

$$\sum_{n=1}^{\infty} \frac{\mathbf{E}\{(T_{\sqrt{n}}(Y_n))^4\}}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbf{E}\{T_{n^2}(Y_n^4)\}}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbf{E}\{T_{n^2}(Y_1^4)\}}{n^2} < \infty.$$

Recall now the following useful lemma.

Lemma 14.1. *(Györfi et al. [7], Corollary 6.1) Under the assumption that ties occur with probability 0,*

$$\sum_{i=1}^{n} \mathbf{I}_{\{\mathbf{X} \text{ is the first NN of } \mathbf{X}_i \text{ in } \{\mathbf{X}_1,\ldots,\mathbf{X}_{i-1},\mathbf{X},\mathbf{X}_{i+1},\ldots,\mathbf{X}_n\}\}} \leq \gamma_d$$

a.s., where $\mathbf{I}$ *denotes the indicator and* $\gamma_d < \infty$ *depends only on* d.

Lemma 14.1 implies that

$$\begin{aligned}
&\mathbf{E}\{(T_{\sqrt{n}}(Y_{n,n,1})^4\} \\
&= \mathbf{E}\left\{\sum_{j=1}^{n-1}(T_{\sqrt{n}}(Y_j))^4 \mathbf{I}_{\{\mathbf{X}_j \text{ is the first NN of } \mathbf{X}_n \text{ in } \{\mathbf{X}_1,\ldots,\mathbf{X}_{n-1}\}\}}\right\} \\
&= (n-1)\mathbf{E}\left\{(T_{\sqrt{n}}(Y_1))^4 \mathbf{I}_{\{\mathbf{X}_1 \text{ is the first NN of } \mathbf{X}_n \text{ in } \{\mathbf{X}_1,\ldots,\mathbf{X}_{n-1}\}\}}\right\} \\
&= (n-1)\mathbf{E}\left\{(T_{\sqrt{n}}(Y_n))^4 \mathbf{I}_{\{\mathbf{X}_n \text{ is the first NN of } \mathbf{X}_1 \text{ in } \{\mathbf{X}_2,\ldots,\mathbf{X}_n\}\}}\right\} \\
&= \mathbf{E}\left\{(T_{\sqrt{n}}(Y_n))^4 \sum_{j=1}^{n-1} \mathbf{I}_{\{\mathbf{X}_n \text{ is the first NN of } \mathbf{X}_j \text{ in } \{\mathbf{X}_1,\ldots,\mathbf{X}_{j-1},\mathbf{X}_{j+1},\ldots,\mathbf{X}_n\}\}}\right\} \\
&\leq \mathbf{E}\left\{(T_{\sqrt{n}}(Y_n))^4 \gamma_d\right\}.
\end{aligned}$$

Therefore

$$\sum_{n=1}^{\infty} \frac{\mathbf{E}\{(T_{\sqrt{n}}(Y_{n,n,1})^4\}}{n^2} \leq \sum_{n=1}^{\infty} \frac{\gamma_d \mathbf{E}\{T_{n^2}(Y_n^4)\}}{n^2} = \gamma_d \sum_{n=1}^{\infty} \frac{\mathbf{E}\{T_{n^2}(Y_1^4)\}}{n^2} < \infty,$$

and so (14.11) is verified, which implies (14.10). Concerning the term J_n, the derivations below are based on the fact that the ordinary 1-NN regression estimate is not universally consistent; however, it is strongly Cesàro convergent in the weak topology, and for noiseless observations ($Y_i = m(\mathbf{X}_i)$) it is strongly convergent in L_2. Introduce the notation

$$m_i(\mathbf{x}) := \mathbf{E}\{T_{\sqrt{i}}(Y) \mid \mathbf{X} = \mathbf{x}\}.$$

Let $\mathbf{X}_{i-1,1}(\mathbf{x})$ denote the 1-NN (first nearest neighbour) of $\mathbf{x}$ from among $\{\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}\}$ and $Y_{i-1,1}(\mathbf{x})$ denote the corresponding label ($x \in \mathbf{R}^d$, $i \geq 2$); then

$$Y_{i-1,1}(\mathbf{X}_i) = Y_{i,i,1} \quad \text{and} \quad \mathbf{X}_{i-1,1}(\mathbf{X}_i) = \mathbf{X}_{i,i,1}.$$

The representation

$$J_n = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbf{R}^d} m_i(\mathbf{x}) T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x})) \mu(d\mathbf{x})$$

holds, where $Y_{0,1}(\mathbf{x}) := 0$. It remains to show that

$$J_n \to \mathbf{E}\{m(\mathbf{X})^2\} \quad \text{a.s.} \tag{14.12}$$

Before proving (14.12) we use two lemmas. Let μ denote the distribution of $\mathbf{X}$.

Lemma 14.2. *If* $\mathbf{E}\{Y^2\} < \infty$ *then*

$$\int \mid m(\mathbf{X}_{n-1,1}(\mathbf{x})) - m(\mathbf{x}) \mid^2 \mu(d\mathbf{x}) \to 0 \quad \textit{a.s.}$$

Proof. The proof is in the spirit of the proof of Theorem 4.1 and Problems 4.5 and 6.3 in Györfi et al. [7]. □

The following lemma is a reformulation of a classic deterministic Tauberian theorem of Landau [8] in summability theory. For a proof and further references, see Lemma 1 in Walk [16].

Lemma 14.3. *If the sequence* a_n, $n = 1, 2, \ldots$ *of real numbers is bounded from below and satisfies*

$$\sum_{n=1}^{\infty} \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{n^3} < \infty,$$

then

$$\frac{1}{n} \sum_{i=1}^{n} a_i \to 0.$$

Proof of (14.12). It suffices to show

$$J_n^* := \frac{1}{n}\sum_{i=1}^{n}\int_{R^d} m(\mathbf{x})T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x}))\mu(d\mathbf{x}) \to \mathbf{E}\{m(\mathbf{X})^2\} \quad a.s. \tag{14.13}$$

In fact, we notice that for each $\alpha > 0$

$$\begin{aligned}&\int_{R^d} \mid m_i(\mathbf{x}) - m(\mathbf{x}) \mid\mid T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x})) \mid \mu(d\mathbf{x}) \\ &\leq \frac{1}{2}\frac{1}{\alpha}\int_{R^d} \mid m_i(\mathbf{x}) - m(\mathbf{x}) \mid^2 \mu(d\mathbf{x}) + \frac{1}{2}\alpha\int_{R^d} T_i(Y_{i-1,1}(\mathbf{x})^2)\mu(d\mathbf{x}).\end{aligned}$$

If we can show

$$\limsup \frac{1}{n}\sum_{i=1}^{n}\int T_i(Y_{i-1,1}(\mathbf{x})^2)\mu(d\mathbf{x}) \leq c\mathbf{E}\{Y^2\} \quad a.s., \tag{14.14}$$

for some constant c, then this together with $\int_{R^d} \mid m_i(\mathbf{x}) - m(\mathbf{x}) \mid^2 \mu(d\mathbf{x}) \to 0$ implies

$$\limsup \frac{1}{n}\sum_{i=1}^{n}\int_{R^d} \mid m_i(\mathbf{x}) - m(\mathbf{x}) \mid\mid T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x})) \mid \mu(d\mathbf{x}) \leq \frac{1}{2}\alpha c\mathbf{E}\{Y^2\} \quad a.s.$$

But $\alpha \to 0$ yields that left-hand side equals 0 a.s. This, together with (14.13), implies (14.12). Therefore, to complete the proof it remains to show (14.14) and (14.13). In the first part we show (14.14). Set $r(\mathbf{x}) := \mathbf{E}\{Y^2|\mathbf{X} = \mathbf{x}\}$, $r_i(\mathbf{x}) := \mathbf{E}\{T_i(Y^2)|\mathbf{X} = \mathbf{x}\}$. In order to get (14.14) it is enough to show

$$\frac{1}{n}\sum_{i=1}^{n}\int \left(T_i(Y_{i-1,1}(\mathbf{x})^2) - r_i(\mathbf{X}_{i-1,1}(\mathbf{x}))\right)\mu(d\mathbf{x}) \to 0 \quad a.s., \tag{14.15}$$

where $r_1(\mathbf{X}_{0,1}(\mathbf{x})) := 0$, and

$$\limsup \frac{1}{n}\sum_{i=1}^{n}\int r_i(\mathbf{X}_{i-1,1}(\mathbf{x}))\mu(d\mathbf{x}) \leq c\mathbf{E}\{Y^2\} \quad a.s.$$

The latter follows from

$$\limsup \int r_n(\mathbf{X}_{n-1,1}(\mathbf{x}))\mu(d\mathbf{x}) \leq \lim \int r(\mathbf{X}_{n-1,1}(\mathbf{x}))\mu(d\mathbf{x}) = \mathbf{E}\{r(\mathbf{X})\} = \mathbf{E}\{Y^2\}$$

a.s. (where the first equality holds by Lemma 14.2), which further yields that the sequence

$$\int (T_n(Y_{n-1,1}(\mathbf{x})^2) - r_n(\mathbf{X}_{n-1,1}(\mathbf{x})))\mu(d\mathbf{x}),$$

$n = 1, 2, \ldots,$ is a.s. bounded from below. In order to get (14.15) and therefore (14.14) by Lemma 14.3, it suffices to show

$$\sum_{n=1}^{\infty} \frac{\mathbf{E}\left\{\left[\sum_{i=1}^{n} \int \left(T_i(Y_{i-1,1}(\mathbf{x})^2) - r_i(\mathbf{X}_{i-1}(\mathbf{x}))\right) \mu(d\mathbf{x})\right]^2\right\}}{n^3} < \infty. \tag{14.16}$$

We now show (14.16). Set for it $A_{i,j} := \left\{\mathbf{x};\ \mathbf{X}_{i-1,1}(\mathbf{x}) = \mathbf{X}_j\right\}$. We note

$$\begin{aligned}
&\sum_{i=1}^{n} \int_{R^d} \left(T_i(Y_{i-1,1}(\mathbf{x})^2) - r_i(\mathbf{X}_{i-1,1}(\mathbf{x}))\right) \mu(d\mathbf{x}) \\
&= \sum_{i=1}^{n} \int_{R^d} \sum_{j=1}^{i-1} \mathbf{I}_{\{\mathbf{X}_{i-1,1}(\mathbf{x})=\mathbf{X}_j\}} \mu(d\mathbf{x}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j)\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mu(A_{i,j}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j)\right) \\
&= \sum_{j=1}^{n-1} \left(\sum_{i=j+1}^{n} \mu(A_{i,j}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j)\right) \right),
\end{aligned}$$

where the $(n-1)$ summands in brackets are orthogonal, because $\mathbf{E}\{T_i(Y_j^2) - r_i(\mathbf{X}_j) \mid \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}, Y_{j'}^{\prime}\} = 0$ for all i and all $j' \neq j$ $(j, j' \in \{1, \ldots, n-1\})$. Thus (14.16) is equivalent to

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E}\left\{\left(\sum_{i=j+1}^{n} \mu(A_{i,j}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j)\right)\right)^2\right\} < \infty.$$

Let the cones $C_1, \ldots, C_{\gamma_d}$ have top $\mathbf{0}$ and angle $\frac{\pi}{3}$, which cover $\mathbf{R}^d$, and let $B_{i,j,l}$ be the subset of $C_{j,l} := \mathbf{X}_j + C_l$ $(j = 1, \ldots, i-1;\ l = 1, \ldots, \gamma_d)$ consisting of all $\mathbf{x}$ that are closer to $\mathbf{X}_j$ than the 1-NN of $\mathbf{X}_j$ in $\{\mathbf{X}_1, \ldots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \ldots, \mathbf{X}_{i-1}\} \cap C_{j,l}$. For $j \leq i-1$, a covering result of Devroye et al. [2], and also of pp. 489 and 490 in Györfi et al. [7], holds as follows:

$$\mu(A_{i,j}) \leq \sum_{l=1}^{\gamma_d} \mu(B_{i,j,l}).$$

It suffices to show, for each $l \in \{1, \ldots, \gamma_d\}$,

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left(\sum_{i=j+1}^{n} \mu(B_{i,j,l}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j) \right) \right)^2 \right\} < \infty. \tag{14.17}$$

We have that

$$\begin{aligned}
&\mathbf{E} \left\{ \left(\sum_{i=j+1}^{n} \mu(B_{i,j,l}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j) \right) \right)^2 \right\} \\
&= \mathbf{E} \left\{ \sum_{i,i'=j+1}^{n} \mathbf{E} \left\{ \mu(B_{i,j,l}) \mu(B_{i',j,l}) (T_i(Y_j^2) - r_i(\mathbf{X}_j))(T_{i'}(Y_j^2) - r_{i'}(\mathbf{X}_j)) \mid \mathbf{X}_j \right\} \right\} \\
&= \mathbf{E} \left\{ \sum_{i,i'=j+1}^{n} \mathbf{E} \left\{ \mu(B_{i,j,l}) \mu(B_{i',j,l}) \mid \mathbf{X}_j \right\} \mathbf{E} \left\{ (T_i(Y_j^2) - r_i(\mathbf{X}_j))(T_{i'}(Y_j^2) - r_{i'}(\mathbf{X}_j)) \mid \mathbf{X}_j \right\} \right\} \\
&\leq \int \sum_{i,i'=j+1}^{n} \sqrt{\mathbf{E}\{\mu(B_{i,j,l})^2 | \mathbf{X}_j = \mathbf{x}\}} \sqrt{\mathbf{E}\{\mu(B_{i',j,l})^2 | \mathbf{X}_j = \mathbf{x}\}} \\
&\quad \cdot \sqrt{\mathbf{E}\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \sqrt{\mathbf{E}\{(T_{i'}(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \mu(d\mathbf{x}) \\
&= \int \left(\sum_{i=j+1}^{n} \sqrt{\mathbf{E}\{\mu(B_{i,j,l})^2 | \mathbf{X}_j = \mathbf{x}\}} \sqrt{\mathbf{E}\left\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\}} \right)^2 \mu(d\mathbf{x}).
\end{aligned}$$

According to [2] and pp. 489 and 490 in [7], one has that $\mathbf{P}\{\mu(B_{i,j,l}) > \sqrt{p}\}$ equals the probability that a *Binom*$(i-2, \sqrt{p})$-distributed random variable takes the value 0, i.e., $(1 - \sqrt{p})^{i-2}$ $(0 < p < 1)$. Thus,

$$\begin{aligned}
&\mathbf{E} \left\{ \left(\sum_{i=j+1}^{n} \mu(B_{i,j,l}) \left(T_i(Y_j^2) - r_i(\mathbf{X}_j) \right) \right)^2 \right\} \\
&\leq \int \left(\sum_{i=j+1}^{n} \sqrt{\int_0^1 \mathbf{P}\{\mu(B_{i,j,l}) > \sqrt{p} \mid \mathbf{X}_j = \mathbf{x}\} dp} \sqrt{\mathbf{E}\left\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\}} \right)^2 \mu(d\mathbf{x}) \\
&= \int \left(\sum_{i=j+1}^{n} \sqrt{\int_0^1 (1 - \sqrt{p})^{i-2} dp} \sqrt{\mathbf{E}\left\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\}} \right)^2 \mu(d\mathbf{x}) \\
&= \int \left(\sum_{i=j+1}^{n} \sqrt{\frac{2}{(i-1)i}} \sqrt{\mathbf{E}\left\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\}} \right)^2 \mu(d\mathbf{x})
\end{aligned}$$

$$\le 8 \int \left(\sum_{i=j+1}^{n} \frac{1}{i} \sqrt{\mathbf{E}\left\{(T_n(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\}} \right)^2 \mu(d\mathbf{x})$$

$$\le 8 \left(\ln \frac{n}{j} \right)^2 \int \mathbf{E}\left\{(T_n(Y^2))^2 | \mathbf{X} = \mathbf{x}\right\} \mu(d\mathbf{x})$$

$$= 8 \left(\ln \frac{n}{j} \right)^2 \mathbf{E}\left\{(T_n(Y^2))^2\right\}.$$

Thus the left-hand side of (14.17) is bounded by

$$8 \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \left(\ln \frac{n}{j} \right)^2 \mathbf{E}\{(T_n(Y^2))^2\} \le 8\, const \sum_{n=1}^{\infty} \frac{\mathbf{E}\{(T_n\left(Y^2\right))^2\}}{n^2} < \infty$$

(because of $\mathbf{E}\{Y^2\} < \infty$), our having used the fact that $\frac{1}{n} \sum_{j=1}^{n-1} \left(\ln \frac{n}{j} \right)^2 \to \int_0^1 \left(\ln \frac{1}{t} \right)^2 dt = \int_0^1 (\ln t)^2 dt < \infty$. Thus (14.17), and therefore (14.14) is proved. In the second part, it remains to show (14.13). In order to get it, according to the proof of Lemma 23.3 in Györfi et al. [7] it suffices to show

$$\limsup \frac{1}{n} \sum_{i=1}^{n} \int \mid m(\mathbf{x}) T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x})) \mid \mu(d\mathbf{x}) \le c^* \mathbf{E}\{Y^2\} \quad \text{a.s.} \quad (14.18)$$

for some constant c^* and to show (14.13) for bounded Y. We prove first (14.18). Notice that

$$\int \mid m(\mathbf{x}) T_{\sqrt{i}}\left(Y_{i-1,1}(\mathbf{x})\right) \mid \mu(d\mathbf{x}) \le \frac{1}{2} \int m(\mathbf{x})^2 \mu(d\mathbf{x}) + \frac{1}{2} \int T_i \left(Y_{i-1,1}(\mathbf{x})\right)^2 \mu(d\mathbf{x}).$$

From $\int m(\mathbf{x})^2 \mu(d\mathbf{x}) \le \mathbf{E}\{Y^2\}$ and from (14.14) we obtain (14.18), with $c^* = \frac{1}{2} + \frac{1}{2}c$. By boundedness of Y, from some index on we have that $T_{\sqrt{i}}(Y) = Y$. Therefore, and because of Lemma 14.2, it suffices to show

$$\frac{1}{n} \sum_{i=1}^{n} \int_{R^d} m(\mathbf{x}) \left(Y_{i-1,1}(\mathbf{x}) - m(\mathbf{X}_{i-1,1}(\mathbf{x}))\right) \mu(d\mathbf{x}) \to 0 \quad a.s.,$$

where $m(\mathbf{X}_{0,1}(\mathbf{x})) := 0$. By boundedness, because of Lemma 14.3 it is enough to show

$$\sum_{n=1}^{\infty} \frac{\mathbf{E}\{[\sum_{i=1}^{n} \int_{R^d} m(\mathbf{x}) \left(Y_{i-1,1}(\mathbf{x}) - m(\mathbf{X}_{i-1,1}(\mathbf{x}))\right) \mu(d\mathbf{x})]^2\}}{n^3} < \infty. \quad (14.19)$$

Noticing

$$\int_{R^d} m(\mathbf{x})\,(Y_{i-1,1}(\mathbf{x}) - m(\mathbf{X}_{i-1,1}(\mathbf{x})))\,\mu(d\mathbf{x})$$
$$= \int_{R^d} m(\mathbf{x}) \sum_{j=1}^{i-1} \mathbf{I}_{\{\mathbf{X}_{i-1,1}(\mathbf{x})=\mathbf{X}_j\}} \mu(d\mathbf{x}) \left(Y_j - m(\mathbf{X}_j)\right)$$
$$= \sum_{j=1}^{i-1} \int_{A_{i,j}} m(\mathbf{x})\mu(d\mathbf{x}) \left(Y_j - m(\mathbf{X}_j)\right),$$

we obtain, with suitable constants c' and c'', that the left-hand side of (14.19) equals

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \mathbf{E} \left\{ \left[\sum_{i=1}^{n} \sum_{j=1}^{i-1} \int_{A_{i,j}} m(\mathbf{x})\mu(d\mathbf{x}) \left(Y_j - m(\mathbf{X}_j)\right) \right]^2 \right\}$$
$$= \sum_{n=1}^{\infty} \frac{1}{n^3} \mathbf{E} \left\{ \left[\sum_{j=1}^{n-1} \left(\sum_{i=j+1}^{n} \int_{A_{i,j}} m(\mathbf{x})\mu(d\mathbf{x}) \right) \left(Y_j - m(\mathbf{X}_j)\right) \right]^2 \right\}$$
$$= \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left(\sum_{i=j+1}^{n} \int_{A_{i,j}} m(\mathbf{x})\mu(d\mathbf{x}) \left(Y_j - m(\mathbf{X}_j)\right) \right)^2 \right\}$$

(by orthogonality of the $(n-1)$ summands in the brackets)

$$\le c' \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left(\sum_{i=j+1}^{n} \mu(A_{i,j}) \right)^2 \right\}$$
$$\le c'' \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \left(\ln \frac{n}{j} \right)^2 < \infty;$$

the latter is as in the proof of (14.14). Thus (14.13) is proved for bounded Y. Therefore (14.12) and thus the assertion have been verified. □

14.3 Rate of Convergence

Next we bound the rate of convergence:

Theorem 14.3. *Assume that Y and $\mathbf{X}$ are bounded ($|Y| < L$, $\|\mathbf{X}\| < K$) and m is Lipschitz continuous and ties occur with probability 0. In addition, suppose that*

(i) *μ has a Lipschitz continuous density f,*
(ii) *For any $\mathbf{x}$ from the support of μ and $0 < r < 2K$,*

$$\mu(S_{\mathbf{x},r}) \geq \gamma r^d,$$

with $\gamma > 0$.

Then for $d \geq 2$, we have that

$$\mathbf{E}\{|\tilde{L}_n - L^*|\} \leq c_1 n^{-1/2} + c_2 n^{-2/d}.$$

Proof. Apply the decomposition

$$\mathbf{E}\{|\tilde{L}_n - L^*|\} \leq \mathbf{E}\{|\tilde{L}_n - \mathbf{E}\{\tilde{L}_n\}|\} + |\mathbf{E}\{\tilde{L}_n\} - L^*| \leq \sqrt{\mathbf{Var}(\tilde{L}_n)} + |\mathbf{E}\{\tilde{L}_n\} - L^*|.$$

For the variance term $\mathbf{Var}(\tilde{L}_n)$, introduce the notation

$$R_n := -\frac{1}{n}\sum_{i=1}^{n} Y_i Y_{n,i,1}.$$

For bounded Y $(|Y| \leq L)$, we show that

$$\mathbf{Var}(R_n) \leq \frac{2(1+2\gamma_d)^2 L^4}{n}, \tag{14.20}$$

from which we get that

$$\begin{aligned}\mathbf{Var}(\tilde{L}_n) &= \mathbf{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 + R_n\right)\\ &\leq 2\mathbf{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right) + 2\mathbf{Var}\left(R_n\right) \leq \frac{2L^4}{n} + \frac{4(1+2\gamma_d)^2 L^4}{n},\end{aligned}$$

and thus,

$$\sqrt{\mathbf{Var}(\tilde{L}_n)} \leq \frac{c_1}{\sqrt{n}}.$$

In the same way as in Liitiäinen et al. [9], we show (14.20) using the Efron–Stein inequality [5]. Replacement of $(\mathbf{X}_j, Y_j)$ by $(\mathbf{X}'_j, Y'_j)$ for fixed $j \in \{1, \ldots, n\}$ (where $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n), (\mathbf{X}'_1, Y'_1), \ldots, (\mathbf{X}'_n, Y'_n)$ are independent and identically distributed) leads to the estimator

$$R_{n,j} = \frac{1}{n}\left(Y_j' Y_{n,j,1}' + \sum_{i \neq j} Y_i Y_{n,i,1}'\right).$$

According to the Efron–Stein inequality we have that

$$\mathbf{Var}(R_n) \leq \frac{1}{2}\sum_{i=1}^{n} \mathbf{E}\{(R_n - R_{n,i})^2\} = \frac{n}{2}\mathbf{E}\{(R_n - R_{n,1})^2\}.$$

Evaluate the difference $R_n - R_{n,1}$:

$$\begin{aligned} R_n - R_{n,1} &= \frac{1}{n}\left(Y_1 Y_{n,1,1} + \sum_{i \neq 1} Y_i Y_{n,i,1}\right) - \frac{1}{n}\left(Y_1' Y_{n,1,1}' + \sum_{i \neq 1} Y_i Y_{n,i,1}'\right) \\ &= \frac{1}{n}\left(Y_1 Y_{n,1,1} - Y_1' Y_{n,1,1}'\right) + \frac{1}{n}\sum_{i \neq 1} Y_i (Y_{n,i,1} - Y_{n,i,1}'). \end{aligned}$$

One can check that $|Y_1 Y_{n,1,1} - Y_1' Y_{n,1,1}'| \leq 2L^2$. Introduce the following notations. Let $n[i]$ be the index of the first nearest neighbour of $\mathbf{X}_i$ from the set $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\} \backslash \{\mathbf{X}_i\}$. Similarly, let $n'[i]$ be the index of the first nearest neighbour of $\mathbf{X}_i$ from the set $\{\mathbf{X}_1', \mathbf{X}_2, \ldots, \mathbf{X}_n\} \setminus \{\mathbf{X}_i\}$. For fixed $i \neq 1$, notice

$$\{Y_{n,i,1} - Y_{n,i,1}' \neq 0\} \subset \{n[i] = 1\} \cup \{n'[i] = 1\}.$$

Thus

$$\begin{aligned} |\sum_{i \neq 1} Y_i (Y_{n,i,1} - Y_{n,i,1}')| &\leq L \sum_{i \neq 1} |Y_{n,i,1} - Y_{n,i,1}'| \\ &\leq 2L^2 \left(\sum_{i \neq 1} \mathbf{1}_{n[i]=1} + \sum_{i \neq 1} \mathbf{1}_{n'[i]=1}\right) \leq 2L^2(\gamma_d + \gamma_d) = 4L^2\gamma_d \end{aligned}$$

a.s., where in the last step we applied Lemma 14.1. Summarizing these bounds we get that

$$\mathbf{Var}(R_n) \leq \frac{n}{2}\left(\frac{1}{n}2L^2 + \frac{1}{n}4L^2\gamma_d\right)^2 = \frac{2(1+2\gamma_d)^2 L^4}{n}$$

a.s., and the proof of (14.20) is complete. For the bias term $\mathbf{E}\{\tilde{L}_n\} - L^*$, notice that

$$\mathbf{E}\{\tilde{L}_n\} - L^* = \mathbf{E}\{m(\mathbf{X}_1)m(\mathbf{X}_{n,1,1})\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}.$$

Because of

$$m(\mathbf{X}_1)m(\mathbf{X}_{n,1,1}) - m(\mathbf{X}_1)^2 = \frac{1}{2}(m(\mathbf{X}_{n,1,1})^2 - m(\mathbf{X}_1)^2) - \frac{1}{2}(m(\mathbf{X}_{n,1,1}) - m(\mathbf{X}_1))^2,$$

the Lipschitz condition (14.1) implies that

$$\begin{aligned}
&|\mathbf{E}\{m(\mathbf{X}_1)m(\mathbf{X}_{n,1,1})\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \\
&\le \frac{1}{2}|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| + \frac{1}{2}\mathbf{E}\{(m(\mathbf{X}_{n,1,1}) - m(\mathbf{X}_1))^2\} \\
&\le \frac{1}{2}|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| + \frac{C}{2}\mathbf{E}\{\|\mathbf{X}_{n,1,1} - \mathbf{X}_1\|^2\},
\end{aligned}$$

where C is the Lipschitz constant in (14.1). For $d \ge 3$, Lemma 6.4 in Györfi et al. [7], and for $d \ge 2$, Theorem 3.2 in Liitiäinen et al. [11] say that

$$\mathbf{E}\{\|\mathbf{X}_{n,1,1} - \mathbf{X}_1\|^2\} \le c_3 n^{-2/d}.$$

Therefore

$$|\mathbf{E}\{m(\mathbf{X}_1)m(\mathbf{X}_{n,1,1})\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \le \frac{1}{2}|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| + c_4 n^{-2/d},$$

and so we have to prove that

$$|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \le c_4 n^{-1/2} + c_5 n^{-2/d}. \tag{14.21}$$

In order to show (14.21), let's calculate the density f_n of $(\mathbf{X}_1, \mathbf{X}_{n,1,1})$ with respect to $\mu \times \mu$. We have that

$$\begin{aligned}
&\mathbf{P}\{\mathbf{X}_1 \in A, \mathbf{X}_{n,1,1} \in B\} \\
&= \sum_{i=2}^{n} \mathbf{P}\{\mathbf{X}_1 \in A, \mathbf{X}_i \in B, \mathbf{X}_{n,1,1} = \mathbf{X}_i\} \\
&= (n-1)\mathbf{P}\{\mathbf{X}_1 \in A, \mathbf{X}_2 \in B, \mathbf{X}_{n,1,1} = \mathbf{X}_2\} \\
&= (n-1)\mathbf{E}\{\mathbf{P}\{\mathbf{X}_1 \in A, \mathbf{X}_2 \in B, \mathbf{X}_{n,1,1} = \mathbf{X}_2 \mid \mathbf{X}_1, \mathbf{X}_2\}\} \\
&= (n-1)\mathbf{E}\left\{\mathbf{P}\{\cap_{i=3}^{n}\{\|\mathbf{X}_1 - \mathbf{X}_i\| \ge \|\mathbf{X}_1 - \mathbf{X}_2\|\} \mid \mathbf{X}_1, \mathbf{X}_2\}\, \mathbf{I}_{\{\mathbf{X}_1 \in A, \mathbf{X}_2 \in B\}}\right\} \\
&= (n-1)\mathbf{E}\left\{\left(1 - \mu(S_{\mathbf{X}_1, \|\mathbf{X}_1 - \mathbf{X}_2\|})\right)^{n-2} \mathbf{I}_{\{\mathbf{X}_1 \in A, \mathbf{X}_2 \in B\}}\right\}.
\end{aligned}$$

Therefore

$$f_n(\mathbf{x}_1, \mathbf{x}_2) = (n-1)\left(1 - \mu(S_{\mathbf{x}_1, \|\mathbf{x}_1 - \mathbf{x}_2\|})\right)^{n-2}.$$

It implies that

$$\mathbf{E}\{m(\mathbf{X}_1)^2\} = \mathbf{E}\{m(\mathbf{X}_1)^2 f_n(\mathbf{X}_1, \mathbf{X}_2)\}$$

and

$$\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} = \mathbf{E}\{m(\mathbf{X}_2)^2 f_n(\mathbf{X}_1, \mathbf{X}_2)\} = \mathbf{E}\{m(\mathbf{X}_1)^2 f_n(\mathbf{X}_2, \mathbf{X}_1)\}.$$

Thus,

$$\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\} = \mathbf{E}\{m(\mathbf{X}_1)^2(f_n(\mathbf{X}_2, \mathbf{X}_1) - f_n(\mathbf{X}_1, \mathbf{X}_2))\},$$

and interchanging $\mathbf{X}_1$ and $\mathbf{X}_2$ we get that

$$\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\} = -\mathbf{E}\{m(\mathbf{X}_2)^2(f_n(\mathbf{X}_2, \mathbf{X}_1) - f_n(\mathbf{X}_1, \mathbf{X}_2))\},$$

and so

$$\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\} = \frac{1}{2}\mathbf{E}\{(m(\mathbf{X}_1)^2 - m(\mathbf{X}_2)^2)(f_n(\mathbf{X}_2, \mathbf{X}_1) - f_n(\mathbf{X}_1, \mathbf{X}_2))\}. \tag{14.22}$$

m satisfies the Lipschitz condition (14.1). Therefore

$$|m(\mathbf{x})^2 - m(\mathbf{z})^2| \le |m(\mathbf{x}) - m(\mathbf{z})|(|m(\mathbf{x})| + |m(\mathbf{z})|) \le 2LC\|\mathbf{x} - \mathbf{z}\|,$$

and so (14.22) implies that

$$|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \le LC\mathbf{E}\{\|\mathbf{X}_1 - \mathbf{X}_2\| \cdot |f_n(\mathbf{X}_2, \mathbf{X}_1) - f_n(\mathbf{X}_1, \mathbf{X}_2)|\}.$$

For any $0 < a < b < 1$, we have the inequality

$$0 < (1-a)^n - (1-b)^n < n(b-a)(1-a)^{n-1}.$$

Therefore

$$\begin{aligned} &|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \\ &\le LCn^2\mathbf{E}\Big\{\|\mathbf{X}_1 - \mathbf{X}_2\| \cdot \left|\mu(S_{\mathbf{X}_1,\|\mathbf{X}_1-\mathbf{X}_2\|}) - \mu(S_{\mathbf{X}_2,\|\mathbf{X}_1-\mathbf{X}_2\|})\right| \\ &\quad \cdot \left(e^{-(n-2)\mu(S_{\mathbf{X}_1,\|\mathbf{X}_1-\mathbf{X}_2\|})} + e^{-(n-2)\mu(S_{\mathbf{X}_2,\|\mathbf{X}_1-\mathbf{X}_2\|})}\right)\Big\}. \end{aligned}$$

If $c_d := Vol(S_{\mathbf{0},1})$ then condition (i) implies that

$$\left|\mu(S_{\mathbf{X}_1,\|\mathbf{X}_1-\mathbf{X}_2\|}) - \mu(S_{\mathbf{X}_2,\|\mathbf{X}_1-\mathbf{X}_2\|})\right| \le c_d\|\mathbf{X}_1 - \mathbf{X}_2\|^d \max_{\|\mathbf{x}-\mathbf{z}\| \le 2\|\mathbf{X}_1-\mathbf{X}_2\|} |f(\mathbf{x}) - f(\mathbf{z})|$$

$$\leq c_9 \|\mathbf{X}_1 - \mathbf{X}_2\|^{d+1}.$$

Because of condition (ii), both $e^{-(n-2)\mu(S_{\mathbf{X}_1,\|\mathbf{X}_1-\mathbf{X}_2\|})}$ and $e^{-(n-2)\mu(S_{\mathbf{X}_2,\|\mathbf{X}_1-\mathbf{X}_2\|})}$ are upper bounded by $e^{-(n-1)\gamma\|\mathbf{X}_1-\mathbf{X}_2\|^d}$. Therefore

$$|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \leq c_{10} n^2 \mathbf{E}\left\{\|\mathbf{X}_1 - \mathbf{X}_2\|^{d+2} e^{-(n-1)\gamma\|\mathbf{X}_1-\mathbf{X}_2\|^d}\right\}.$$

Note that the random variable $R := \|\mathbf{X}_1 - \mathbf{X}_2\|$ has a density on $[0, 2K]$ bounded above by $c_{11} r^{d-1}$. Therefore

$$\begin{aligned} |\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| &\leq c_{12} n^2 \int_0^{2K} r^{d+2} e^{-n\gamma r^d} r^{d-1} dr \\ &\leq c_{12} n^{-2/d} \int_0^{\infty} r^{1+2/d} e^{-\gamma r} dr. \end{aligned}$$

□

Acknowledgements This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

References

1. Chow, Y.S.: Local convergence of martingales and the law of large numbers. Ann. Math. Stat. **36**, 552–558 (1965)
2. Devroye, L., Györfi, L., Krzyżak, A., Lugosi, G.: On the strong universal consistency of nearest neighbor regression function estimation. Ann. Stat. **22**, 1371–1385 (1994)
3. Devroye, L., Schäfer, D., Györfi, L., Walk, H.: The estimation problem of minimum mean squared error. Stat. Decis. **21**, 15–28 (2003)
4. Dudoit, S., van der Laan, M.: Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. Stat. Methodol. **2**, 131–154 (2005)
5. Efron, B., Stein, C.: The jackknife estimate of variance. Ann. Stat. **9**, 586–596 (1981)
6. Ferrario, P.G., Walk, H.: Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors. J. Nonparametric Stat. **24**, 1019–1039 (2012)
7. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: A Distribution-Free Theory of Nonparametric Regression. Springer, New York (2002)
8. Landau, E.: Über die Bedeutung einiger neuer Grenzwertsätze der Herren Hardy und Axer. Prace mat.-fiz. **21**, 91–177 (1910)
9. Liitiäinen, E., Corona, F., Lendasse, A.: On nonparametric residual variance estimation. Neural Process. Lett. **28**, 155–167 (2008)
10. Liitiäinen, E., Verleysen, M., Corona, F., Lendasse, A.: Residual variance estimation in machine learning. Neurocomputing **72**, 3692–3703 (2009)
11. Liitiäinen, E., Corona, F., Lendasse, A.: Residual variance estimation using a nearest neighbor statistic. J. Multivar. Anal. **101**, 811–823 (2010)
12. Müller, H.G., Stadtmüller, U.: Estimation of heteroscedasticity in regression analysis. Ann. Stat. **15**, 610–625 (1987)

13. Müller, U., Schick, A., Wefelmeyer, W.: Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. Statistics **37**, 179–188 (2003)
14. Neumann, M.H.: Fully data-driven nonparametric variance estimators. Statistics **25**, 189–212 (1994)
15. Stadtmüller, U., Tsybakov, A.: Nonparametric recursive variance estimation. Statistics **27**, 55–63 (1995)
16. Walk, H.: Strong laws of large numbers by elementary Tauberian arguments. Monatsh. Math. **144**, 329–346 (2005)