

**Universität  
Stuttgart**

**Fachbereich  
Mathematik**

---

**Strong universal consistent estimate of the minimum  
mean squared error**

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

---

**Preprint 2013/002**



**Universität  
Stuttgart**

**Fachbereich  
Mathematik**

---

Strong universal consistent estimate of the minimum  
mean squared error

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

---

**Preprint 2013/002**

Fachbereich Mathematik  
Fakultät Mathematik und Physik  
Universität Stuttgart  
Pfaffenwaldring 57  
D-70 569 Stuttgart

**E-Mail:** [preprints@mathematik.uni-stuttgart.de](mailto:preprints@mathematik.uni-stuttgart.de)  
**WWW:** <http://www.mathematik.uni-stuttgart.de/preprints>

ISSN **1613-8309**

© Alle Rechte vorbehalten. Nachdruck nur mit Genehmigung des Autors.  
L<sup>A</sup>T<sub>E</sub>X-Style: Winfried Geis, Thomas Merkle

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).

## 1 Introduction

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

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

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], [10], Liitiäinen et al. [11], Müller and Stadtmüller [12], Neumann [14], Stadtmüller and Tsybakov [15], 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 guaranteed rate of convergence. They introduced a modified nearest neighbor 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 neighbor of  $\mathbf{X}_i$  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

$$|m(\mathbf{x}) - m(\mathbf{z})| \leq C \|\mathbf{x} - \mathbf{z}\|, \quad (1)$$

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

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

Liitiäinen et al. ([9], [10]) introduced another estimate of the minimum mean squared error  $L^*$  by the first and second nearest neighbor 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 neighbors  $\mathbf{X}_{n,i,1}$  and  $\mathbf{X}_{n,i,2}$  of  $\mathbf{X}_i$  among  $\mathbf{X}_1, \dots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \dots, \mathbf{X}_n$ , resp. (In the sequel, assume that for calculating the first and second nearest neighbors, ties occur with probability zero. When  $\mathbf{X}$  has a density, the case of ties among nearest neighbor distances occurs with probability zero.) 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 (2).

In this paper 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.

## 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 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 1.** *Assume that ties occur with probability zero. If  $|Y|$  is bounded then*

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

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

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

*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  $Y$ ,

$$L_n \rightarrow L^* \tag{3}$$

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

$$\frac{1}{n} \sum_{k=1}^n L_k \rightarrow L^* \tag{4}$$

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  $Y$ ,

$$\frac{1}{n} \sum_{i=1}^n Y_i Y_{n,i,2} \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \tag{5}$$

a.s. and

$$\frac{1}{n} \sum_{i=1}^n Y_{n,i,1} Y_{n,i,2} \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \tag{6}$$

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

$$\frac{1}{n} \sum_{k=1}^n \frac{1}{k} \sum_{i=1}^k Y_i Y_{k,i,2} \rightarrow L^* \tag{7}$$

a.s. and

$$\frac{1}{n} \sum_{k=1}^n \frac{1}{k} \sum_{i=1}^k Y_{k,i,1} Y_{k,i,2} \rightarrow L^* \tag{8}$$

a.s. Now the statements of the theorem follow from (3), (5), (6), and from (4), (7) and (8), respectively.  $\square$

Next we consider a recursive estimate

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

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

$$L'_n \rightarrow L^* \quad \text{a.s.}$$

*Proof.* We have to show that

$$\frac{1}{n} \sum_{i=1}^n Y_i Y_{i,1} \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \quad \text{a.s.} \quad (9)$$

For  $a > 0$ , introduce the truncation function

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

Similarly to the proof of Theorem 2.5 in Ferrario and Walk [6] one can check that in order to show (9), it suffices to prove that

$$\frac{1}{n} \sum_{i=1}^n T_{\sqrt{i}}(Y_i) T_{\sqrt{i}}(Y_{i,1}) \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \quad (10)$$

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

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

where

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

and

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

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

$$I_n \rightarrow 0 \quad (11)$$

can be derived from the Chow [1] theorem if

$$\sum_{n=1}^{\infty} \frac{\mathbf{Var}\{T_{\sqrt{n}}(Y_n) T_{\sqrt{n}}(Y_{n,n,1})\}}{n^2} < \infty. \quad (12)$$

We have that

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

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

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

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

Lemma 1 implies that

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

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 (12) is verified, which implies (11). Concerning the term  $J_n$ , the derivations below are based on the facts, 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 neighbor) of  $\mathbf{x}$  among  $\{\mathbf{X}_1, \dots, \mathbf{X}_{i-1}\}$  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

$$J_n \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \quad \text{a.s.} \tag{13}$$

Before proving (13) we use two lemmas. Let  $\mu$  denote the distribution of  $\mathbf{X}$ .

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

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



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

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

then

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

*Proof of (13).* It suffices to show

$$J_n^* := \frac{1}{n} \sum_{i=1}^n \int_{R^d} m(\mathbf{x}) T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x})) \mu(d\mathbf{x}) \rightarrow \mathbf{E}\{m(\mathbf{X})^2\} \quad a.s. \quad (14)$$

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

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

If we can show

$$\limsup \frac{1}{n} \sum_{i=1}^n \int_{R^d} T_i(Y_{i-1,1}(\mathbf{x}))^2 \mu(d\mathbf{x}) \leq c \mathbf{E}\{Y^2\} \quad a.s., \quad (15)$$

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

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

But  $\alpha \rightarrow 0$  yields that left-hand side equals 0 a.s. This, together with (14), implies (13). Therefore, to complete the proof it remains to show (15) and (14). In the first part we show (15). 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 (15) it is enough to show

$$\frac{1}{n} \sum_{i=1}^n \int_{R^d} (T_i(Y_{i-1,1}(\mathbf{x}))^2 - r_i(\mathbf{X}_{i-1,1}(\mathbf{x}))) \mu(d\mathbf{x}) \rightarrow 0 \quad a.s., \quad (16)$$

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

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

The latter follows from

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

a.s. (where the first equality holds by Lemma 2), which further yields that the sequence

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

$n = 1, 2, \dots$ , is a.s. bounded from below. In order to get (16) and therefore (15), by Lemma 3, it suffices to show

$$\sum_{n=1}^{\infty} \mathbf{E} \left\{ \frac{\left[ \sum_{i=1}^n \int (T_i(Y_{i-1,1}(\mathbf{x})^2) - r_i(\mathbf{X}_{i-1}(\mathbf{x}))) \mu(d\mathbf{x}) \right]^2}{n^3} \right\} < \infty. \quad (17)$$

We now show (17). Set for it  $A_{i,j} := \{\mathbf{x}; \mathbf{X}_{i-1,1}(\mathbf{x}) = \mathbf{X}_j\}$ . We note

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

where the  $(n-1)$  summands in brackets are orthogonal, because  $\mathbf{E}\{T_i(Y_j^2) - r_i(\mathbf{X}_j) \mid \mathbf{X}_1, \dots, \mathbf{X}_{n-1}, Y_j'\} = 0$  for all  $i$  and all  $j' \neq j$  ( $j, j' \in \{1, \dots, n-1\}$ ). Thus (17) is equivalent to

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left( \sum_{i=j+1}^n \mu(A_{i,j}) (T_i(Y_j^2) - r_i(\mathbf{X}_j)) \right)^2 \right\} < \infty. \quad (18)$$

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

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

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

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left( \sum_{i=j+1}^n \mu(B_{i,j,l}) (T_i(Y_j^2) - r_i(\mathbf{X}_j)) \right)^2 \right\} < \infty. \quad (19)$$

We have that

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

$$= \int \left( \sum_{i=j+1}^n \sqrt{\mathbf{E}\{\mu(B_{i,j,l})^2 | \mathbf{X}_j = \mathbf{x}\}} \sqrt{\mathbf{E}\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \right)^2 \mu(d\mathbf{x})$$

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

$$\begin{aligned} & \mathbf{E} \left\{ \left( \sum_{i=j+1}^n \mu(B_{i,j,l}) (T_i(Y_j^2) - r_i(\mathbf{X}_j)) \right)^2 \right\} \\ & \leq \int \left( \sum_{i=j+1}^n \sqrt{\int_0^1 \mathbf{P}\{\mu(B_{i,j,l}) > \sqrt{p} | \mathbf{X}_j = \mathbf{x}\} dp} \sqrt{\mathbf{E}\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \right)^2 \mu(d\mathbf{x}) \\ & = \int \left( \sum_{i=j+1}^n \sqrt{\int_0^1 (1 - \sqrt{p})^{i-2} dp} \sqrt{\mathbf{E}\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \right)^2 \mu(d\mathbf{x}) \\ & = \int \left( \sum_{i=j+1}^n \sqrt{\frac{2}{(i-1)i}} \sqrt{\mathbf{E}\{(T_i(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \right)^2 \mu(d\mathbf{x}) \\ & \leq 8 \int \left( \sum_{i=j+1}^n \frac{1}{i} \sqrt{\mathbf{E}\{(T_n(Y^2))^2 | \mathbf{X} = \mathbf{x}\}} \right)^2 \mu(d\mathbf{x}) \\ & \leq 8 \left( \ln \frac{n}{j} \right)^2 \int \mathbf{E}\{(T_n(Y^2))^2 | \mathbf{X} = \mathbf{x}\} \mu(d\mathbf{x}) \\ & = 8 \left( \ln \frac{n}{j} \right)^2 \mathbf{E}\{(T_n(Y^2))^2\}. \end{aligned}$$

Thus the left-hand side of (19) is bounded by

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

(because of  $\mathbf{E}\{Y^2\} < \infty$ ), having used the fact that  $\frac{1}{n} \sum_{j=1}^{n-1} \left( \ln \frac{n}{j} \right)^2 \rightarrow \int_0^1 (\ln \frac{1}{t})^2 dt = \int_0^1 (\ln t)^2 dt < \infty$ . Thus (19), and therefore (15), is proved. In the second part, it remains to show (14). In order to get it, according to the proof of Lemma 23.3 in Györfi et al. [7], it suffices to show

$$\limsup \frac{1}{n} \sum_{i=1}^n \int |m(\mathbf{x}) T_{\sqrt{i}}(Y_{i-1,1}(\mathbf{x}))| \mu(d\mathbf{x}) \leq c^* \mathbf{E}\{Y^2\} \quad \text{a.s.} \quad (20)$$

for some constant  $c^*$  and to show (14) for bounded  $Y$ . We prove first (20). Notice that

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

From  $\int m(\mathbf{x})^2 \mu(d\mathbf{x}) \leq \mathbf{E}\{Y^2\}$  and from (15) we obtain (20), 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 2 it suffices to show

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

where  $m(\mathbf{X}_{0,1}(\mathbf{x})) := 0$ . By boundedness, because of Lemma 3 it is enough to show

$$\sum_{n=1}^{\infty} \mathbf{E} \left\{ \frac{[\sum_{i=1}^n \int_{R^d} m(\mathbf{x}) (Y_{i-1,1}(\mathbf{x}) - m(\mathbf{X}_{i-1,1}(\mathbf{x}))) \mu(d\mathbf{x})]^2}{n^3} \right\} < \infty. \quad (21)$$

Noticing

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

we obtain, with suitable constants  $c'$  and  $c''$ , that the left-hand side of (21) equals

$$\begin{aligned} & \sum_{n=1}^{\infty} \frac{1}{n^3} \mathbf{E} \left\{ \left[ \sum_{i=1}^n \sum_{j=1}^{i-1} \int_{A_{i,j}} m(\mathbf{x}) \mu(d\mathbf{x}) (Y_j - m(\mathbf{X}_j)) \right]^2 \right\} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^3} \mathbf{E} \left\{ \left[ \sum_{j=1}^{n-1} \left( \sum_{i=j+1}^n \int_{A_{i,j}} m(\mathbf{x}) \mu(d\mathbf{x}) \right) (Y_j - m(\mathbf{X}_j)) \right]^2 \right\} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left( \sum_{i=j+1}^n \int_{A_{i,j}} m(\mathbf{x}) \mu(d\mathbf{x}) (Y_j - m(\mathbf{X}_j)) \right)^2 \right\} \\ & \quad \text{(by orthogonality of the } (n-1) \text{ summands in the brackets)} \\ &\leq c' \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \mathbf{E} \left\{ \left( \sum_{i=j+1}^n \mu(A_{i,j}) \right)^2 \right\} \\ &\leq c'' \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n-1} \left( \ln \frac{n}{j} \right)^2 < \infty, \end{aligned}$$

the latter is as in the proof of (15). Thus (14) is proved for bounded  $Y$ . Therefore altogether (13) and thus the assertion have been verified.  $\square$

### 3 Rate of convergence

Next we bound the rate of convergence:

**Theorem 3.** *Assume that  $Y$  and  $\mathbf{X}$  are bounded ( $|Y| < L$ ,  $\|\mathbf{X}\| < K$ ) and  $m$  is Lipschitz continuous and ties occur with probability zero. In addition, suppose that*

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

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

with  $\gamma > 0$ .

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

$$\mathbf{E}\{|\tilde{L}_n - L^*|\} \leq c_1 n^{-1/2} + c_2 n^{-2/d}. \quad (22)$$

*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

$$\mathbf{Var}(R_n) \leq \frac{2(1+2\gamma_d)^2 L^4}{n}, \quad (23)$$

from which we get that

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

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 (23) using the Efron-Stein inequality [5]. Replacement of  $(\mathbf{X}_j, Y_j)$  by  $(\mathbf{X}'_j, Y'_j)$  for fixed  $j \in \{1, \dots, n\}$  (where  $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n), (\mathbf{X}'_1, Y'_1), \dots, (\mathbf{X}'_n, Y'_n)$  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  $R_n - R_{n,1}$ :

$$\begin{aligned} R_n - R_{n,1} &= \frac{1}{n} \left( Y_1 Y_{n,1,1} + \sum_{i \neq 1} Y_i Y_{n,i,1} \right) - \frac{1}{n} \left( Y'_1 Y'_{n,1,1} + \sum_{i \neq 1} Y_i Y'_{n,i,1} \right) \\ &= \frac{1}{n} (Y_1 Y_{n,1,1} - Y'_1 Y'_{n,1,1}) + \frac{1}{n} \sum_{i \neq 1} Y_i (Y_{n,i,1} - Y'_{n,i,1}). \end{aligned}$$

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 neighbor of  $\mathbf{X}_i$  from the set  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\} \setminus \{\mathbf{X}_i\}$ . Similarly, let  $n'[i]$  be the index of the first nearest neighbor of  $\mathbf{X}_i$  from the set  $\{\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n\} \setminus \{\mathbf{X}_i\}$ . 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

$$\left| \sum_{i \neq 1} Y_i (Y_{n,i,1} - Y'_{n,i,1}) \right| \leq L \sum_{i \neq 1} |Y_{n,i,1} - Y'_{n,i,1}|$$

$$\leq 2L^2 \left( \sum_{i \neq 1} \mathbf{1}_{n[i]=1} + \sum_{i \neq 1} \mathbf{1}_{n'[i]=1} \right) \leq 2L^2 (\gamma_d + \gamma_d) = 4L^2 \gamma_d$$

a.s., where in the last step we applied Lemma 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 (23) is complete. For the bias term  $\mathbf{E}\{\tilde{L}_n\} - L^*$ , notice that

$$\mathbf{E}\{\tilde{L}_n\} - L^* = \mathbf{E}\{m(\mathbf{X}_1)m(\mathbf{X}_{n,1,1})\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}. \quad (24)$$

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 (1) implies that

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

where  $C$  is the Lipschitz constant in (1). For  $d \geq 3$ , Lemma 6.4 in Györfi et al. [7] and for  $d \geq 2$ , Theorem 3.2 in Liitiäinen et al. [10] say that

$$\mathbf{E}\{\|\mathbf{X}_{n,1,1} - \mathbf{X}_1\|^2\} \leq c_3 n^{-2/d}. \quad (25)$$

Therefore

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

and so we have to prove that

$$|\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\}| \leq c_4 n^{-1/2} + c_5 n^{-2/d}. \quad (26)$$

In order to show (26), 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

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

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

$$\mathbf{E}\{m(\mathbf{X}_{n,1,1})^2\} - \mathbf{E}\{m(\mathbf{X}_1)^2\} = \frac{1}{2} \mathbf{E}\{(m(\mathbf{X}_1)^2 - m(\mathbf{X}_2)^2)(f_n(\mathbf{X}_2, \mathbf{X}_1) - f_n(\mathbf{X}_1, \mathbf{X}_2))\}. \quad (27)$$

$m$  satisfies the Lipschitz condition (1). Therefore

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

and so (27) implies that

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

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

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

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

$$\begin{aligned} \left| \mu(S_{\mathbf{X}_1, \|\mathbf{X}_1 - \mathbf{X}_2\|}) - \mu(S_{\mathbf{X}_2, \|\mathbf{X}_1 - \mathbf{X}_2\|}) \right| & \leq c_d \|\mathbf{X}_1 - \mathbf{X}_2\|^d \max_{\|\mathbf{x} - \mathbf{z}\| \leq 2\|\mathbf{X}_1 - \mathbf{X}_2\|} |f(\mathbf{x}) - f(\mathbf{z})| \\ & \leq c_9 \|\mathbf{X}_1 - \mathbf{X}_2\|^{d+1}. \end{aligned}$$

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

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

□

## References

1. Chow, Y. S.: Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics*, **36**, 552–558 (1965).
2. Devroye, L., Györfi, L., Krzyżak, A., Lugosi, G.: On the strong universal consistency of nearest neighbor regression function estimation. *Annals of Statistics*, **22**, 1371–1385 (1994).
3. Devroye, L., Schäfer, D., Györfi, L., Walk, H.: The estimation problem of minimum mean squared error. *Statistics and Decisions*, **21**, 15–28 (2003)
4. Dudoit, S., van der Laan, M.J.: Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, **2**, 131–154 (2005)
5. Efron, B., Stein, C.: The jackknife estimate of variance. *Annals of Statistics*, **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, *Journal of Nonparametric Statistics*, **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**, pp. 91-177 (1910).
9. Liitiäinen, E., Corona, F., Lendasse, A.: On nonparametric residual variance estimation. *Neural Processing Letters*, **28**, 155–167 (2008)
10. Liitiäinen, E., Corona, F., Lendasse, A.: Residual variance estimation using a nearest neighbor statistic. *Journal of Multivariate Analysis*, **101**, 811–823 (2010)
11. Liitiäinen, E., Verleysen, M., Corona, F., Lendasse, A.: Residual variance estimation in machine learning. *Neurocomputing*, **72**, 3692–3703 (2009)
12. Müller, H.-G., Stadtmüller, U.: Estimation of heteroscedasticity in regression analysis, *Annals of Statistics*, **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)



Luc Devroye  
School of Computer Science McGill University  
3480 University Street Montreal  
Canada H3A OE9  
**E-Mail:** lucdevroye@gmail.com

Paola G. Ferrario  
Department of Mathematics  
University of Stuttgart  
Pfaffenwaldring 57  
70565 Stuttgart  
Germany  
**E-Mail:** paola.ferrario@mathematik.uni-stuttgart.de

László 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

Harro Walk  
Department of Mathematics  
University of Stuttgart  
Pfaffenwaldring 57  
70565 Stuttgart  
Germany  
**E-Mail:** walk@mathematik.uni-stuttgart.de



## Erschienenene Preprints ab Nummer 2007/2007-001

Komplette Liste: <http://www.mathematik.uni-stuttgart.de/preprints>

- 2012-002 *Devroye, L.; Ferrario, P.G.; Györfi, L.; Walk, H.:* Strong universal consistent estimate of the minimum mean squared error
- 2013-001 *Kohls, K.; Rsch, A.; Siebert, K.G.:* A Posteriori Error Analysis of Optimal Control Problems with Control Constraints
- 2012-018 *Kimmerle, W.; Konovalov, A.:* On the Prime Graph of the Unit Group of Integral Group Rings of Finite Groups II
- 2012-017 *Stroppel, B.; Stroppel, M.:* Desargues, Doily, Dualities, and Exceptional Isomorphisms
- 2012-016 *Moroianu, A.; Pilca, M.; Semmelmann, U.:* Homogeneous almost quaternion-Hermitian manifolds
- 2012-015 *Steinke, G.F.; Stroppel, M.J.:* Simple groups acting two-transitively on the set of generators of a finite elation Laguerre plane
- 2012-014 *Steinke, G.F.; Stroppel, M.J.:* Finite elation Laguerre planes admitting a two-transitive group on their set of generators
- 2012-013 *Diaz Ramos, J.C.; Dominguez Vázquez, M.; Kollross, A.:* Polar actions on complex hyperbolic spaces
- 2012-012 *Moroianu, A.; Semmelmann, U.:* Weakly complex homogeneous spaces
- 2012-011 *Moroianu, A.; Semmelmann, U.:* Invariant four-forms and symmetric pairs
- 2012-010 *Hamilton, M.J.D.:* The closure of the symplectic cone of elliptic surfaces
- 2012-009 *Hamilton, M.J.D.:* Iterated fibre sums of algebraic Lefschetz fibrations
- 2012-008 *Hamilton, M.J.D.:* The minimal genus problem for elliptic surfaces
- 2012-007 *Ferrario, P.:* Partitioning estimation of local variance based on nearest neighbors under censoring
- 2012-006 *Stroppel, M.:* Buttons, Holes and Loops of String: Lacing the Doily
- 2012-005 *Hantsch, F.:* Existence of Minimizers in Restricted Hartree-Fock Theory
- 2012-004 *Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.:* Unitals admitting all translations
- 2012-003 *Hamilton, M.J.D.:* Representing homology classes by symplectic surfaces
- 2012-002 *Hamilton, M.J.D.:* On certain exotic 4-manifolds of Akhmedov and Park
- 2012-001 *Jentsch, T.:* Parallel submanifolds of the real 2-Grassmannian
- 2011-028 *Spreer, J.:* Combinatorial 3-manifolds with cyclic automorphism group
- 2011-027 *Griesemer, M.; Hantsch, F.; Wellig, D.:* On the Magnetic Pekar Functional and the Existence of Bipolarons
- 2011-026 *Müller, S.:* Bootstrapping for Bandwidth Selection in Functional Data Regression
- 2011-025 *Felber, T.; Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent static forecasting of stationary and ergodic time series via local averaging and least squares estimates
- 2011-024 *Jones, D.; Kohler, M.; Walk, H.:* Weakly universally consistent forecasting of stationary and ergodic time series
- 2011-023 *Györfi, L.; Walk, H.:* Strongly consistent nonparametric tests of conditional independence
- 2011-022 *Ferrario, P.G.; Walk, H.:* Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors
- 2011-021 *Eberts, M.; Steinwart, I.:* Optimal regression rates for SVMs using Gaussian kernels
- 2011-020 *Frank, R.L.; Geisinger, L.:* Refined Semiclassical Asymptotics for Fractional Powers of the Laplace Operator
- 2011-019 *Frank, R.L.; Geisinger, L.:* Two-term spectral asymptotics for the Dirichlet Laplacian on a bounded domain
- 2011-018 *Hänel, A.; Schulz, C.; Wirth, J.:* Embedded eigenvalues for the elastic strip with cracks
- 2011-017 *Wirth, J.:* Thermo-elasticity for anisotropic media in higher dimensions
- 2011-016 *Höllig, K.; Hörner, J.:* Programming Multigrid Methods with B-Splines

- 2011-015 *Ferrario, P.*: Nonparametric Local Averaging Estimation of the Local Variance Function
- 2011-014 *Müller, S.; Dippon, J.*: k-NN Kernel Estimate for Nonparametric Functional Regression in Time Series Analysis
- 2011-013 *Knarr, N.; Stroppel, M.*: Unitals over composition algebras
- 2011-012 *Knarr, N.; Stroppel, M.*: Baer involutions and polarities in Moufang planes of characteristic two
- 2011-011 *Knarr, N.; Stroppel, M.*: Polarities and planar collineations of Moufang planes
- 2011-010 *Jentsch, T.; Moroianu, A.; Semmelmann, U.*: Extrinsic hyperspheres in manifolds with special holonomy
- 2011-009 *Wirth, J.*: Asymptotic Behaviour of Solutions to Hyperbolic Partial Differential Equations
- 2011-008 *Stroppel, M.*: Orthogonal polar spaces and unitals
- 2011-007 *Nagl, M.*: Charakterisierung der Symmetrischen Gruppen durch ihre komplexe Gruppenalgebra
- 2011-006 *Solanes, G.; Teufel, E.*: Horo-tightness and total (absolute) curvatures in hyperbolic spaces
- 2011-005 *Ginoux, N.; Semmelmann, U.*: Imaginary Kählerian Killing spinors I
- 2011-004 *Scherer, C.W.; Köse, I.E.*: Control Synthesis using Dynamic  $D$ -Scales: Part II — Gain-Scheduled Control
- 2011-003 *Scherer, C.W.; Köse, I.E.*: Control Synthesis using Dynamic  $D$ -Scales: Part I — Robust Control
- 2011-002 *Alexandrov, B.; Semmelmann, U.*: Deformations of nearly parallel  $G_2$ -structures
- 2011-001 *Geisinger, L.; Weidl, T.*: Sharp spectral estimates in domains of infinite volume
- 2010-018 *Kimmerle, W.; Konovalov, A.*: On integral-like units of modular group rings
- 2010-017 *Gauduchon, P.; Moroianu, A.; Semmelmann, U.*: Almost complex structures on quaternion-Kähler manifolds and inner symmetric spaces
- 2010-016 *Moroianu, A.; Semmelmann, U.*: Clifford structures on Riemannian manifolds
- 2010-015 *Grafarend, E.W.; Kühnel, W.*: A minimal atlas for the rotation group  $SO(3)$
- 2010-014 *Weidl, T.*: Semiclassical Spectral Bounds and Beyond
- 2010-013 *Stroppel, M.*: Early explicit examples of non-desarguesian plane geometries
- 2010-012 *Effenberger, F.*: Stacked polytopes and tight triangulations of manifolds
- 2010-011 *Györfi, L.; Walk, H.*: Empirical portfolio selection strategies with proportional transaction costs
- 2010-010 *Kohler, M.; Krzyżak, A.; Walk, H.*: Estimation of the essential supremum of a regression function
- 2010-009 *Geisinger, L.; Laptev, A.; Weidl, T.*: Geometrical Versions of improved Berezin-Li-Yau Inequalities
- 2010-008 *Poppitz, S.; Stroppel, M.*: Polarities of Schellhammer Planes
- 2010-007 *Grundhöfer, T.; Krinn, B.; Stroppel, M.*: Non-existence of isomorphisms between certain unitals
- 2010-006 *Höllig, K.; Hörner, J.; Hoffacker, A.*: Finite Element Analysis with B-Splines: Weighted and Isogeometric Methods
- 2010-005 *Kaltenbacher, B.; Walk, H.*: On convergence of local averaging regression function estimates for the regularization of inverse problems
- 2010-004 *Kühnel, W.; Solanes, G.*: Tight surfaces with boundary
- 2010-003 *Kohler, M.; Walk, H.*: On optimal exercising of American options in discrete time for stationary and ergodic data
- 2010-002 *Gulde, M.; Stroppel, M.*: Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras
- 2010-001 *Leitner, F.*: Examples of almost Einstein structures on products and in cohomogeneity one
- 2009-008 *Griesemer, M.; Zenk, H.*: On the atomic photoeffect in non-relativistic QED

- 2009-007 *Griesemer, M.; Moeller, J.S.:* Bounds on the minimal energy of translation invariant n-polaron systems
- 2009-006 *Demirel, S.; Harrell II, E.M.:* On semiclassical and universal inequalities for eigenvalues of quantum graphs
- 2009-005 *Bächle, A, Kimmerle, W.:* Torsion subgroups in integral group rings of finite groups
- 2009-004 *Geisinger, L.; Weidl, T.:* Universal bounds for traces of the Dirichlet Laplace operator
- 2009-003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation
- 2009-002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces
- 2009-001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of  $E^3$  and of the 3-torus
- 2008-006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps
- 2008-005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems
- 2008-004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities
- 2008-003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope
- 2008-002 *Hertweck, M.; Höfert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups  $PSL(2, q)$
- 2008-001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term
- 2007-006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term
- 2007-005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field
- 2007-004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees
- 2007-003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides
- 2007-002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry
- 2007-001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions