

**Universität
Stuttgart**

**Fachbereich
Mathematik**

On the asymptotic normality of an estimate of a
regression functional

László Györfi, Harro Walk

Preprint 2015/011

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

E-Mail: 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^AT_EX-Style: Winfried Geis, Thomas Merkle

On the asymptotic normality of an estimate of a regression functional

László Györfi *

GYORFI@CS.BME.HU

*Department of Computer Science and Information Theory
Budapest University of Technology and Economics
Magyar Tudósok körútja 2., H-1117 Budapest, Hungary*

Harro Walk

WALK@MATHEMATIK.UNI-STUTTGAERT.DE

*Department of Mathematics
University of Stuttgart
Pfaffenwaldring 57, D-70569 Stuttgart, Germany*

Abstract

An estimate of the second moment of the regression function is introduced. Its asymptotic normality is proved such that the asymptotic variance depends neither on the dimension of the observation vector, nor on the smoothness properties of the regression function.

Keywords: nonparametric estimation, regression functional, central limit theorem, partitioning estimate

1. Introduction

Let Y be a real valued random variable and let $X = (X^{(1)}, \dots, X^{(d)})$ be a d -dimensional random observational vector. In regression analysis one wishes to estimate Y given X , i.e., one wants to find a function g defined on the range of X so that $g(X)$ is “close” to Y . Assume that the main aim of the analysis is to minimize the mean squared error :

$$\min_g \mathbb{E}\{(g(X) - Y)^2\}.$$

As is well-known, this minimum is achieved by the regression function $m(x)$, which is defined by

$$m(x) = \mathbb{E}\{Y \mid X = x\}. \quad (1)$$

For each measurable function g one has

$$\begin{aligned} \mathbb{E}\{(g(X) - Y)^2\} &= \mathbb{E}\{(m(X) - Y)^2\} + \mathbb{E}\{(m(X) - g(X))^2\} \\ &= \mathbb{E}\{(m(X) - Y)^2\} + \int |m(x) - g(x)|^2 \mu(dx), \end{aligned}$$

*. This research has been partially supported by the European Union and Hungary and co-financed by the European Social Fund through the project TMOP-4.2.2.C-11/1/KONV-2012-0004 - National Research Center for Development and Market Introduction of Advanced Information and Communication Technologies.

where μ stands for the distribution of the observation X .

It is of great importance to be able to estimate the minimum mean squared error

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

accurately, even before a regression estimate is applied: in a standard nonparametric regression design process, one considers a finite number of real-valued features $X^{(i)}$, $i \in I$, and evaluates whether these suffice to explain Y . In case they suffice for the given explanatory task, an estimation method can be applied on the basis of the features already under consideration, if not, more or different features must be considered. The quality of a subvector $\{X^{(i)}, i \in I\}$ of X is measured by the minimum mean squared error

$$L^*(I) := \mathbb{E}\left(Y - \mathbb{E}\{Y|X^{(i)}, i \in I\}\right)^2$$

that can be achieved using the features as explanatory variables. $L^*(I)$ depends upon the unknown distribution of $(Y, X^{(i)} : i \in I)$. The first phase of any regression estimation process therefore heavily relies on estimates of L^* (even *before* a regression estimate is picked).

Concerning dimension reduction the related testing problem is on the hypothesis

$$L^* = L^*(I).$$

This testing problem can be managed such that we estimate both L^* and $L^*(I)$, and accept the hypothesis if the two estimates are close to each other. (Cf. De Brabanter et al. (2014).)

Devroye et al. (2003), Evans and Jones (2008), Liitiäinen et al. (2008), Liitiäinen et al. (2009), Liitiäinen et al. (2010), and Ferrario and Walk (2012) introduced nearest neighbor based estimates of L^* , proved strong universal consistency and calculated the (fast) rate of convergence.

Because of

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

estimating L^* is equivalent to estimating the second moment S^* of the regression function:

$$S^* = \mathbb{E}\{m(X)^2\} = \int m(x)^2 \mu(dx).$$

In this paper we introduce a partitioning based estimator of S^* , and show its asymptotic normality. It turns out that the asymptotic variance depends neither on the dimension of the observation vector, nor on the smoothness properties of the regression function.

2. A splitting estimate

We suppose that the regression estimation problem is based on a sequence

$$(X_1, Y_1), (X_2, Y_2), \dots$$

of i.i.d. random vectors distributed as (X, Y) . Let

$$\mathcal{P}_n = \{A_{n,j}, j = 1, 2, \dots\}$$

be a cubic partition of \mathbb{R}^d of size $h_n > 0$.

The partitioning estimator of the regression function m is defined as

$$m_n(x) = \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} \text{ if } x \in A_{n,j}, \quad (2)$$

(interpreting $0/0 = 0$) with

$$\nu_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \in A\}} Y_i$$

and

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \in A\}}.$$

(Here \mathbb{I} denotes the indicator function.)

If for cubic partition

$$nh_n^d \rightarrow \infty \quad \text{and} \quad h_n \rightarrow 0 \quad (3)$$

as $n \rightarrow \infty$, then the partitioning regression estimate (2) is weakly universally consistent, which means that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \int (m_n(x) - m(x))^2 \mu(dx) \right\} = 0 \quad (4)$$

for any distribution of (X, Y) with $\mathbb{E}\{Y^2\} < \infty$, and for bounded Y it holds

$$\lim_{n \rightarrow \infty} \int (m_n(x) - m(x))^2 \mu(dx) = 0 \quad (5)$$

a.s. (Cf. Theorems 4.2 and 23.1 in Györfi et al. (2002).)

Assume splitting data

$$Z_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$$

and

$$D'_n = \{(X'_1, Y'_1), \dots, (X'_n, Y'_n)\}$$

such that $(X_1, Y_1), \dots, (X_n, Y_n), (X'_1, Y'_1), \dots, (X'_n, Y'_n)$ are i.i.d.

The splitting data estimate of S^* is defined as

$$S_n = \frac{1}{n} \sum_{i=1}^n Y'_i m_n(X'_i).$$

Put

$$\nu'_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X'_i \in A\}} Y'_i,$$

then S_n has the equivalent form

$$S_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\infty} \mathbb{I}_{\{X'_i \in A_{n,j}\}} Y'_i \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} = \sum_{j=1}^{\infty} \nu'_n(A_{n,j}) \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})}. \quad (6)$$

Theorem 1 *Assume (3) and that μ is non-atomic and has bounded support. Suppose that there is a finite constant C such that*

$$\mathbb{E}\{|Y|^3 \mid X\} < C. \quad (7)$$

Then

$$\sqrt{n}(S_n - \mathbb{E}\{S_n\})/\sigma \xrightarrow{\mathcal{D}} N(0, 1),$$

where

$$\sigma^2 = 2 \int M_2(x)m(x)^2\mu(dx) - \left(\int m(x)^2\mu(dx) \right)^2 - \int m(x)^4\mu(dx),$$

with

$$M_2(X) = \mathbb{E}\{Y^2 \mid X\}.$$

We prove Theorem 1 in the next section.

3. Proof of Theorem 1

Introduce the notations

$$U_n = \sqrt{n}(S_n - \mathbb{E}\{S_n \mid Z_n\})$$

and

$$V_n = \sqrt{n}(\mathbb{E}\{S_n \mid Z_n\} - \mathbb{E}\{S_n\}),$$

then

$$\sqrt{n}(S_n - \mathbb{E}\{S_n\}) = U_n + V_n.$$

We prove Theorem 1 by showing that for any $u, v \in \mathbb{R}$

$$\mathbb{P}\{U_n \leq u, V_n \leq v\} \rightarrow \Phi\left(\frac{u}{\sigma_1}\right) \Phi\left(\frac{v}{\sigma_2}\right) \quad (8)$$

where Φ denotes the standard normal distribution function, and

$$\sigma_1^2 = \int M_2(x)m(x)^2\mu(dx) - \left(\int m(x)^2\mu(dx) \right)^2$$

and

$$\sigma_2^2 = \int M_2(x)m(x)^2\mu(dx) - \int m(x)^4\mu(dx).$$

Notice that V_n is measurable with respect to Z_n , therefore

$$\begin{aligned} & \left| \mathbb{P}\{U_n \leq u, V_n \leq v\} - \Phi\left(\frac{u}{\sigma_1}\right) \Phi\left(\frac{v}{\sigma_2}\right) \right| \\ &= \left| \mathbb{E}\{\mathbb{I}_{\{V_n \leq v\}} \mathbb{P}\{U_n \leq u \mid Z_n\}\} - \Phi\left(\frac{u}{\sigma_1}\right) \Phi\left(\frac{v}{\sigma_2}\right) \right| \\ &\leq \left| \mathbb{E}\left\{ \mathbb{I}_{\{V_n \leq v\}} \left(\mathbb{P}\{U_n \leq u \mid Z_n\} - \Phi\left(\frac{u}{\sigma_1}\right) \right) \right\} \right| \\ &+ \left| \left(\mathbb{P}\{V_n \leq v\} - \Phi\left(\frac{v}{\sigma_2}\right) \right) \Phi\left(\frac{u}{\sigma_1}\right) \right| \\ &\leq \mathbb{E}\left\{ \left| \mathbb{P}\{U_n \leq u \mid Z_n\} - \Phi\left(\frac{u}{\sigma_1}\right) \right| \right\} + \left| \mathbb{P}\{V_n \leq v\} - \Phi\left(\frac{v}{\sigma_2}\right) \right|. \end{aligned}$$

Thus, (8) is satisfied if

$$\mathbb{P}\{U_n \leq u \mid Z_n\} \rightarrow \Phi\left(\frac{u}{\sigma_1}\right) \quad (9)$$

in probability and

$$\mathbb{P}\{V_n \leq v\} \rightarrow \Phi\left(\frac{v}{\sigma_2}\right). \quad (10)$$

Proof of (9).

Let's start with the representation

$$\begin{aligned} U_n &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (Y_i' m_n(X_i') - \mathbb{E}\{Y_i' m_n(X_i') \mid Z_n\}) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i' m_n(X_i') - \mathbb{E}\{Y_i' m_n(X_i') \mid Z_n\}). \end{aligned}$$

Because of (7) and the Jensen inequality, for any $1 \leq s \leq 3$, we get

$$M_s(X) := \mathbb{E}\{|Y|^s \mid X\} = (\mathbb{E}\{|Y|^s \mid X\}^{1/s})^s \leq (\mathbb{E}\{|Y|^3 \mid X\}^{1/3})^s \leq C^{s/3}, \quad (11)$$

especially

$$|m(X)| \leq C^{1/3}$$

and

$$\mathbb{E}\{|Y|^3\} \leq C.$$

Next we apply a Berry-Esseen type central limit theorem (see Theorem 14 in Petrov (1975)). It implies that

$$\left| \mathbb{P}\{U_n \leq u \mid Z_n\} - \Phi\left(\frac{u}{\sqrt{\text{Var}(Y_1' m_n(X_1') \mid Z_n)}}\right) \right| \leq \frac{c}{\sqrt{n}} \frac{\mathbb{E}\{|Y_1' m_n(X_1')|^3 \mid Z_n\}}{\sqrt{\text{Var}(Y_1' m_n(X_1') \mid Z_n)}^3}$$

with the universal constant $c > 0$. Because of

$$\mathbb{E}\{Y_1' m_n(X_1') \mid Z_n\} = \int m(x) m_n(x) \mu(dx),$$

we get that

$$\begin{aligned} \text{Var}(Y_1' m_n(X_1') \mid Z_n) &= \mathbb{E}\{Y_1'^2 m_n(X_1')^2 \mid Z_n\} - \mathbb{E}\{Y_1' m_n(X_1') \mid Z_n\}^2 \\ &= \int M_2(x) m_n(x)^2 \mu(dx) - \left(\int m(x) m_n(x) \mu(dx) \right)^2. \end{aligned}$$

Now (4) together with (11) implies that

$$\text{Var}(Y_1' m_n(X_1') \mid Z_n) \rightarrow \sigma_1^2$$

in probability. Further

$$\mathbb{E}\{|Y_1' m_n(X_1')|^3 \mid Z_n\} \leq C \int |m_n(x)|^3 \mu(dx).$$

Put

$$A_n(x) = A_{n,j} \text{ if } x \in A_{n,j}.$$

Again, applying the Jensen inequality we get

$$|m_n(x)|^3 \leq \left| \frac{\sum_{i=1}^n \mathbb{I}_{\{X_i \in A_n(x)\}} |Y_i|^{3/2}}{\sum_{i=1}^n \mathbb{I}_{\{X_i \in A_n(x)\}}} \right|^2,$$

the right hand side of which is the square of the regression estimate, where Y is replaced by $|Y|^{3/2}$. Thus, (4) together with $\mathbb{E}\{|Y|^3\} < \infty$ implies that

$$\int \left| \frac{\sum_{i=1}^n \mathbb{I}_{\{X_i \in A_n(x)\}} |Y_i|^{3/2}}{\sum_{i=1}^n \mathbb{I}_{\{X_i \in A_n(x)\}}} \right|^2 \mu(dx) \rightarrow \mathbb{E}\{\mathbb{E}\{|Y|^{3/2} \mid X\}^2\} < C$$

in probability. These limit relations imply (9).

Proof of (10).

Assuming that the support S of μ is bounded, let l_n be such that $S \subset \cup_{j=1}^{l_n} A_{n,j}$. Also we re-index the partition so that

$$\mu(A_{n,j}) \geq \mu(A_{n,j+1}),$$

with $\mu(A_{n,j}) > 0$ for $j \leq l_n$, and $\mu(A_{n,j}) = 0$ otherwise. Then,

$$S_n = \sum_{j=1}^{l_n} \nu'_n(A_{n,j}) \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})}, \tag{12}$$

and

$$l_n \leq \frac{c}{h_n^d}.$$

The condition $nh_n^d \rightarrow \infty$ implies that

$$l_n/n \rightarrow 0.$$

Because of (12) we have that

$$V_n = \sqrt{n} \sum_{j=1}^{l_n} \nu(A_{n,j}) \left(\frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} \right\} \right).$$

Observe that we have to show the asymptotic normality for a finite sum of dependent random variables. In order to prove (10), we follow the lines of the proof in Beirlant and Györfi (1998) such that use a Poissonization argument. We introduce the notation N_n for a Poisson(n) random variable independent of $(X_1, Y_1), (X_2, Y_2), \dots$. Moreover we put

$$n\tilde{\nu}_n(A) = \sum_{i=1}^{N_n} I_{\{X_i \in A\}} Y_i$$

and

$$n\tilde{\mu}_n(A) = \sum_{i=1}^{N_n} I_{\{X_i \in A\}}.$$

The key result in this step is the following property:

Proposition 2 (*Beirlant and Mason (1995), Beirlant et al. (1994).*) Put

$$\tilde{M}_n = \sqrt{n} \sum_{j=1}^{l_n} \nu(A_{n,j}) \left(\frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \right), \quad (13)$$

and

$$M_n = \sqrt{n} \sum_{j=1}^{l_n} \nu(A_{n,j}) \left(\frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \right). \quad (14)$$

Assume that

$$\Phi_n(t, v) = \mathbb{E} \left(\exp \left(it\tilde{M}_n + iv \frac{N_n - n}{\sqrt{n}} \right) \right) \rightarrow e^{-(t^2\rho^2 + v^2)/2}$$

for a constant $\rho > 0$. Then

$$M_n/\rho \xrightarrow{D} N(0, 1).$$

Put

$$T_n = t\tilde{M}_n + v \frac{N_n - n}{\sqrt{n}},$$

for which a central limit result is to hold:

$$T_n \xrightarrow{D} N(0, t^2\rho^2 + v^2) \quad (15)$$

as $n \rightarrow \infty$. Remark that

$$\text{Var}(T_n) = t^2 \text{Var}(\tilde{M}_n) + 2tv \mathbb{E} \left\{ \tilde{M}_n \frac{N_n - n}{\sqrt{n}} \right\} + v^2.$$

For a cell $A = A_{n,j}$ from the partition with $\mu(A) > 0$, let $Y(A)$ be a random variable such that

$$\mathbb{P}\{Y(A) \in B\} = \mathbb{P}\{Y \in B | X \in A\},$$

where B is an arbitrary Borel set.

Introduce the notations

$$q_{n,k} = \mathbb{P}\{n\mu_n(A) = k\} = \binom{n}{k} \mu(A)^k (1 - \mu(A))^{n-k}$$

and

$$\tilde{q}_{n,k} = \mathbb{P}\{n\tilde{\mu}_n(A) = k\} = \frac{(n\mu(A))^k}{k!} e^{-n\mu(A)}.$$

Concerning the expectation, with $(Y_1(A), Y_2(A), \dots)$ an i.i.d. sequence of random variables distributed as $Y(A)$ we find that

$$\begin{aligned} \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} \right\} &= \sum_{k=0}^{\infty} \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} \mid n\tilde{\mu}_n(A) = k \right\} \mathbb{P}\{n\tilde{\mu}_n(A) = k\} \\ &= \sum_{k=1}^{\infty} \mathbb{E} \left\{ \frac{\sum_{i=1}^k Y_i(A)}{k} \right\} \tilde{q}_{n,k} \\ &= \mathbb{E}\{Y_1(A)\} (1 - \tilde{q}_{n,0}) \\ &= \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}), \end{aligned} \quad (16)$$

further, by (21)

$$\mathbb{E} \left\{ \frac{\nu_n(A)}{\mu_n(A)} \right\} = n \mathbb{E} \left\{ \frac{Y_n(A)}{1 + (n-1)\mu_{n-1}(A)} \right\} = \frac{\nu(A)}{\mu(A)} (1 - (1 - \mu(A))^n), \quad (17)$$

Moreover,

$$\begin{aligned} \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A)^2}{\tilde{\mu}_n(A)^2} \right\} &= \sum_{k=0}^{\infty} \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A)^2}{\tilde{\mu}_n(A)^2} \mid n\tilde{\mu}_n(A) = k \right\} \mathbb{P}\{n\tilde{\mu}_n(A) = k\} \\ &= \sum_{k=1}^{\infty} \mathbb{E} \left\{ \frac{\left(\sum_{i=1}^k Y_i(A) \right)^2}{k^2} \right\} \tilde{q}_{n,k} \\ &= \sum_{k=1}^{\infty} \frac{k \mathbb{E} \{Y_1(A)^2\} + k(k-1) \mathbb{E} \{Y_1(A)\}^2}{k^2} \tilde{q}_{n,k} \\ &= \text{Var}(Y_1(A)) \sum_{k=1}^{\infty} \frac{1}{k} \tilde{q}_{n,k} + \mathbb{E} \{Y_1(A)\}^2 (1 - \tilde{q}_{n,0}), \end{aligned}$$

and

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{1}{k} \tilde{q}_{n,k} &= \sum_{k=1}^{\infty} \frac{1}{k} \frac{(n\mu(A))^k}{k!} e^{-n\mu(A)} \\ &= \sum_{k=1}^{\infty} \frac{1}{k+1} \frac{(n\mu(A))^k}{k!} e^{-n\mu(A)} + \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \frac{(n\mu(A))^k}{k!} e^{-n\mu(A)} \\ &\leq \frac{1}{n\mu(A)} (1 - \tilde{q}_{n,0}) + \frac{3}{n^2\mu(A)^2} (1 - \tilde{q}_{n,0}). \end{aligned}$$

The independence of the Poisson masses over different cells leads to

$$\begin{aligned} \text{Var}(\tilde{M}_n) &= n \sum_{j=1}^{l_n} \nu(A_{n,j})^2 \text{Var} \left(\frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right) \\ &\leq n \sum_{j=1}^{l_n} \nu(A_{n,j})^2 \left(\text{Var}(Y_1(A_{n,j})) \left(\frac{1}{n\mu(A_{n,j})} (1 - e^{-n\mu(A_{n,j})}) \right) \right. \\ &\quad \left. + \frac{3}{n^2\mu(A_{n,j})^2} (1 - e^{-n\mu(A_{n,j})}) \right) \\ &\quad + \mathbb{E} \{Y_1(A_{n,j})\}^2 (1 - e^{-n\mu(A_{n,j})}) - \mathbb{E} \{Y_1(A_{n,j})\}^2 (1 - e^{-n\mu(A_{n,j})})^2 \\ &\leq \sum_{j=1}^{l_n} \frac{\nu(A_{n,j})^2}{\mu(A_{n,j})^2} \text{Var}(Y_1(A_{n,j})) \mu(A_{n,j}) \\ &\quad + \sum_{j=1}^{l_n} \frac{3 \text{Var}(Y_1(A_{n,j})) \nu(A_{n,j})^2}{n\mu(A_{n,j})^2} \\ &\quad + n \sum_{j=1}^{l_n} \nu(A_{n,j})^2 \mathbb{E} \{Y_1(A_{n,j})\}^2 e^{-n\mu(A_{n,j})} \end{aligned}$$

such that the bounding error in these inequalities is of order $O(l_n/n)$. (4) together with the boundedness of M_2 and m implies that

$$\begin{aligned} & \sum_{j=1}^{l_n} \frac{\nu(A_{n,j})^2}{\mu(A_{n,j})^2} \text{Var}(Y_1(A_{n,j})) \mu(A_{n,j}) \\ &= \int \frac{\int_{A_n(x)} M_2(z) \mu(dz)}{\mu(A_n(x))} \left(\frac{\int_{A_n(x)} m(z) \mu(dz)}{\mu(A_n(x))} \right)^2 \mu(dx) - \int \left(\frac{\int_{A_n(x)} m(z) \mu(dz)}{\mu(A_n(x))} \right)^4 \mu(dx) \\ &= \sigma_2^2 + o(1). \end{aligned}$$

Moreover,

$$\sum_{j=1}^{l_n} \frac{3 \text{Var}(Y_1(A_{n,j})) \nu(A_{n,j})^2}{n \mu(A_{n,j})^2} \leq \frac{3C^{4/3} l_n}{n} \rightarrow 0.$$

Then

$$\begin{aligned} & n \sum_{j=1}^{l_n} \nu(A_{n,j})^2 \mathbb{E} \{Y_1(A_{n,j})\}^2 e^{-n\mu(A_{n,j})} \\ &= \sum_{j=1}^{l_n} \frac{\nu(A_{n,j})^2}{\mu(A_{n,j})^2} \mathbb{E} \{Y_1(A_{n,j})\}^2 n \mu(A_{n,j}) e^{-n\mu(A_{n,j})} \mu(A_{n,j}) \\ &\leq C^{4/3} \sum_{j=1}^{l_n} n \mu(A_{n,j})^2 e^{-n\mu(A_{n,j})} \\ &\leq C^{4/3} (\max_{z>0} z^2 e^{-z}) l_n / n \rightarrow 0. \end{aligned}$$

So we proved that

$$\text{Var}(\tilde{M}_n) \rightarrow \sigma_2^2.$$

To complete the asymptotics for $\text{Var}(T_n)$, it remains to show that

$$\mathbb{E} \left\{ \tilde{M}_n \frac{N_n - n}{\sqrt{n}} \right\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Because of

$$N_n = n \sum_{j=1}^{l_n} \tilde{\mu}_n(A_{n,j})$$

and

$$n = n \sum_{j=1}^{l_n} \mu(A_{n,j}),$$

we have that

$$\begin{aligned}
 & \mathbb{E} \left\{ \tilde{M}_n \frac{N_n - n}{\sqrt{n}} \right\} \\
 &= n \sum_{j=1}^{l_n} \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \nu(A_{n,j}) (\tilde{\mu}_n(A_{n,j}) - \mu(A_{n,j})) \right\} \\
 &= n \sum_{j=1}^{l_n} \nu(A_{n,j}) \left(\mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \mu(A_{n,j}) \right) \\
 &= n \sum_{j=1}^{l_n} \nu(A_{n,j}) \left(\nu(A_{n,j}) - \frac{\nu(A_{n,j})}{\mu(A_{n,j})} (1 - e^{-n\mu(A_{n,j})}) \mu(A_{n,j}) \right) \\
 &= n \sum_{j=1}^{l_n} \nu(A_{n,j})^2 e^{-n\mu(A_{n,j})} \\
 &\leq C^{2/3} (\max_{z>0} z^2 e^{-z}) l_n / n \rightarrow 0.
 \end{aligned}$$

To finish the proof of (15) by Lyapunov's central limit theorem, it suffices to prove that

$$n^{3/2} \sum_{j=1}^{l_n} \mathbb{E} \left\{ \left| t \left(\frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \right) \nu(A_{n,j}) + v(\tilde{\mu}_n(A_{n,j}) - \mu(A_{n,j})) \right|^3 \right\} \rightarrow 0$$

or, by invoking the c_r inequality, that

$$n^{3/2} \sum_{j=1}^{l_n} \mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \right|^3 \right\} \nu(A_{n,j})^3 \rightarrow 0 \quad (18)$$

and

$$n^{3/2} \sum_{j=1}^{l_n} \mathbb{E} \left\{ |\tilde{\mu}_n(A_{n,j}) - \mu(A_{n,j})|^3 \right\} \rightarrow 0. \quad (19)$$

In view of (18), because of (11) it suffices to prove

$$D_n := n^{3/2} \sum_{j=1}^{l_n} \mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} \right|^3 \right\} \mu(A_{n,j})^3 \rightarrow 0 \quad (20)$$

For a cell A , (16) implies that

$$\begin{aligned}
 \mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} - \mathbb{E} \left\{ \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} \right\} \right|^3 \right\} &\leq 4\mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} - \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \mathbb{I}_{\{\tilde{\mu}_n(A) > 0\}} \right|^3 \right\} \\
 &\quad + 4\mathbb{E} \left\{ \left| \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \mathbb{I}_{\{\tilde{\mu}_n(A) > 0\}} - \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \right|^3 \right\}.
 \end{aligned}$$

On the one hand, (16), (11) and (22) imply that, for a constant K ,

$$\begin{aligned}
 & \mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} - \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \mathbb{I}_{\{\tilde{\mu}_n(A) > 0\}} \right|^3 \right\} \\
 &= \sum_{k=0}^{\infty} \mathbb{E} \left\{ \left| \frac{\tilde{\nu}_n(A)}{\tilde{\mu}_n(A)} - \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \mathbb{I}_{\{\tilde{\mu}_n(A) > 0\}} \right|^3 \mid n\tilde{\mu}_n(A) = k \right\} \mathbb{P}\{n\tilde{\mu}_n(A) = k\} \\
 &= \sum_{k=1}^{\infty} \mathbb{E} \left\{ \left| \frac{\sum_{i=1}^k (Y_i(A) - \mathbb{E}\{Y_i(A)\})}{k^3} \right|^3 \right\} \tilde{q}_{n,k} \\
 &\leq K \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} \tilde{q}_{n,k} \\
 &\leq c_1 \frac{1}{n^{3/2} \mu(A)^{3/2}},
 \end{aligned}$$

where we applied the Marcinkiewicz and Zygmund (1937) inequality for absolute central moments of sums of i.i.d. random variables. On the other hand

$$\mathbb{E} \left\{ \left| \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \mathbb{I}_{\{\tilde{\mu}_n(A) > 0\}} - \frac{\nu(A)}{\mu(A)} (1 - \tilde{q}_{n,0}) \right|^3 \right\} \leq C \tilde{q}_{n,0}.$$

Therefore

$$\begin{aligned}
 D_n &\leq n^{3/2} c_2 \sum_{j=1}^{l_n} \left(\frac{1}{n^{3/2} \mu(A_{n,j})^{3/2}} + e^{-n\mu(A_{n,j})} \right) \mu(A_{n,j})^3 \\
 &\leq c_2 \left(\sum_{j=1}^{l_n} \mu(A_{n,j})^{3/2} + \sum_{j=1}^{l_n} n^{3/2} e^{-n\mu(A_{n,j})} \mu(A_{n,j})^3 \right) \\
 &\leq c_2 \sum_{j=1}^{l_n} \mu(A_{n,j})^{3/2} \left(1 + \max_{z>0} z^{3/2} e^{-z} \right) \\
 &= c_3 \int \mu(A_n(x))^{1/2} \mu(dx) \\
 &\rightarrow 0,
 \end{aligned}$$

where we used the assumption that μ is non-atomic. Thus, (18) is proved.

The proof of (19) is easier. Notice that (19) means

$$F_n := n^{-3/2} \sum_{j=1}^{l_n} \mathbb{E} \left\{ \left| \sum_{i=1}^{N_n} \mathbb{I}_{\{X_i \in A_{n,j}\}} - n\mu(A_{n,j}) \right|^3 \right\} \rightarrow 0.$$

One has

$$\begin{aligned}
 & \mathbb{E} \left\{ \left| \sum_{i=1}^{N_n} \mathbb{I}_{\{X_i \in A_{n,j}\}} - n\mu(A_{n,j}) \right|^3 \right\} \\
 & \leq 4\mathbb{E} \left\{ \left| \sum_{i=1}^{N_n} (\mathbb{I}_{\{X_i \in A_{n,j}\}} - \mu(A_{n,j})) \right|^3 \right\} + 4\mathbb{E} \left\{ |(N_n - n)\mu(A_{n,j})|^3 \right\} \\
 & \leq c_4 \left(\sum_{k=1}^{\infty} k^{3/2} \mu(A_{n,j})^{3/2} e^{-n} \frac{n^k}{k!} + \mathbb{E} \left\{ |N_n - n|^3 \right\} \mu(A_{n,j})^3 \right) \\
 & \leq c_5 \left(n^{3/2} \mu(A_{n,j})^{3/2} + n^{3/2} \mu(A_{n,j})^3 \right).
 \end{aligned}$$

Therefore

$$F_n \leq 2c_5 \sum_{j=1}^{l_n} \mu(A_{n,j})^{3/2} \rightarrow 0,$$

and so (19) is proved, too.

The remaining step in the proof of (10) is to show that

$$\Delta_n := V_n - M_n = n^{1/2} \sum_{j=1}^{l_n} \left(\mathbb{E} \left\{ \frac{\tilde{\nu}_n(A_{n,j})}{\tilde{\mu}_n(A_{n,j})} \right\} - \mathbb{E} \left\{ \frac{\nu_n(A_{n,j})}{\mu_n(A_{n,j})} \right\} \right) \nu(A_{n,j}) \rightarrow 0$$

By (16) and (17) have that

$$\begin{aligned}
 |\Delta_n| &= \left| n^{1/2} \sum_{j=1}^{l_n} \frac{\nu(A_{n,j})}{\mu(A_{n,j})} (e^{-n\mu(A_{n,j})} - (1 - \mu(A_{n,j}))^n) \nu(A_{n,j}) \right| \\
 &= n^{1/2} \sum_{j=1}^{l_n} \frac{\nu(A_{n,j})^2}{\mu(A_{n,j})^2} (e^{-n\mu(A_{n,j})} - (1 - \mu(A_{n,j}))^n) \mu(A_{n,j}) \\
 &\leq C^{2/3} n^{1/2} \sum_{j=1}^{l_n} (e^{-n\mu(A_{n,j})} - (1 - \mu(A_{n,j}))^n) \mu(A_{n,j}).
 \end{aligned}$$

For $0 \leq z \leq 1$, using the elementary inequalities

$$1 - z \leq e^{-z} \leq 1 - z + z^2$$

we have that

$$e^{-nz} - (1 - z)^n = (e^{-z} - (1 - z)) \sum_{k=0}^{n-1} e^{-kz} (1 - z)^{n-1-k} \leq nz^2 e^{-(n-1)z},$$

and thus we get that

$$\begin{aligned}
 |\Delta_n| &\leq C^{2/3} n^{1/2} \sum_{j=1}^{l_n} (e^{-n\mu(A_{n,j})} - (1 - \mu(A_{n,j}))^n) \mu(A_{n,j}) \\
 &\leq C^{2/3} n^{1/2} \sum_{j=1}^{l_n} n\mu(A_{n,j})^3 e^{-(n-1)\mu(A_{n,j})} \\
 &\leq \frac{C^{2/3}}{n^{1/2}} \sum_{j=1}^{l_n} \mu(A_{n,j}) \left([n\mu(A_{n,j})]^2 e^{-n\mu(A_{n,j})} \right) e \\
 &\leq \frac{C^{2/3}}{n^{1/2}} \sum_{j=1}^{l_n} \mu(A_{n,j}) \max_{z \geq 0} (z^2 e^{-z}) e \\
 &\rightarrow 0.
 \end{aligned}$$

This ends the proof of (10) and so the proof of Theorem 1 is complete. \square

Next we give two lemmas, which are used above.

Lemma 3 *If $B(n, p)$ is a binomial random variable with parameters (n, p) , then*

$$\mathbb{E} \left\{ \frac{1}{1 + B(n, p)} \right\} = \frac{1 - (1 - p)^{n+1}}{(n + 1)p}. \quad (21)$$

Lemma 4 *If $Po(\lambda)$ is a Poisson random variable with parameter λ , then*

$$\mathbb{E} \left\{ \frac{1}{Po(\lambda)^3} \mathbb{1}_{\{Po(\lambda) > 0\}} \right\} \leq \frac{24}{\lambda^3}. \quad (22)$$

References

- J. Beirlant and L. Györfi. On the asymptotic L_2 -error in partitioning regression estimation. *Journal of Statistical Planning and Inference*, 71:93–107, 1998.
- J. Beirlant and D. Mason. On the asymptotic normality of l_p -norms of empirical functionals. *Mathematical Methods of Statistics*, 4:1–19, 1995.
- J. Beirlant, L. Györfi, and G. Lugosi. On the asymptotic normality of the l_1 - and l_2 - errors in histogram density estimation. *Canadian J. Statistics*, 22:309–318, 1994.
- K. De Brabanter, P. G. Ferrario, and L. Györfi. Detecting ineffective features for nonparametric regression. In A. Argyriou, J. A. K. Suykens, M. Signoretto, editor, *Regularization, Optimization, Kernels, and Support Vector Machines*, pages 177–194. Chapman & Hall/CRC Machine Learning and Pattern Recognition Series, 2014.
- L. Devroye, D. Schäfer, L. Györfi, and H. Walk. The estimation problem of minimum mean squared error. *Statistics and Decisions*, 21:15–28, 2003.

- D. Evans and A. J. Jones. Non-parametric estimation of residual moments and covariance. *Proceedings of the Royal Society, A* 464:2831–2846, 2008.
- P. G. Ferrario and H. Walk. Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors. *Journal of Nonparametric Statistics*, 24:1019–1039, 2012.
- L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer–Verlag, New York, 2002.
- E. Liitiäinen, F. Corona, and A. Lendasse. On nonparametric residual variance estimation. *Neural Processing Letters*, 28:155–167, 2008.
- E. Liitiäinen, M. Verleysen, F. Corona, and A. Lendasse. Residual variance estimation in machine learning. *Neurocomputing*, 72:3692–3703, 2009.
- E. Liitiäinen, F. Corona, and A. Lendasse. Residual variance estimation using a nearest neighbor statistic. *Journal of Multivariate Analysis*, 101:811–823, 2010.
- J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29:60–90, 1937.
- V. V. Petrov. *Sums of Independent Random Variables*. Springer-Verlag, Berlin, 1975.

László Györfi

Department of Computer Science and Information Theory Budapest University of Technology and
Economics Magyar Tudósok körútja 2., H-1117 Budapest, Hungary

E-Mail: gyorfi@cs.bme.hu

Harro Walk

Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany

E-Mail: walk@mathematik.uni-stuttgart.de

WWW: <http://www.isa.uni-stuttgart.de/LstStoch/Walk/>

Erschienenene Preprints ab Nummer 2012-001

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

- 2015-011 *Györfi, L.; Walk, H.:* On the asymptotic normality of an estimate of a regression functional
- 2015-010 *Gorodski, C, Kollross, A.:* Some remarks on polar actions
- 2015-009 *Apprich, C.; Höllig, K.; Hörner, J.; Reif, U.:* Collocation with WEB-Splines
- 2015-008 *Kabil, B.; Rodrigues, M.:* Spectral Validation of the Whitham Equations for Periodic Waves of Lattice Dynamical Systems
- 2015-007 *Kollross, A.:* Hyperpolar actions on reducible symmetric spaces
- 2015-006 *Schmid, J.; Griesemer, M.:* Well-posedness of Non-autonomous Linear Evolution Equations in Uniformly Convex Spaces
- 2015-005 *Hinrichs, A.; Markhasin, L.; Oettershagen, J.; Ullrich, T.:* Optimal quasi-Monte Carlo rules on higher order digital nets for the numerical integration of multivariate periodic functions
- 2015-004 *Kutter, M.; Rohde, C.; Sändig, A.-M.:* Well-Posedness of a Two Scale Model for Liquid Phase Epitaxy with Elasticity
- 2015-003 *Rossi, E.; Schleper, V.:* Convergence of a numerical scheme for a mixed hyperbolic-parabolic system in two space dimensions
- 2015-002 *Döring, M.; Györfi, L.; Walk, H.:* Exact rate of convergence of kernel-based classification rule
- 2015-001 *Kohler, M.; Müller, F.; Walk, H.:* Estimation of a regression function corresponding to latent variables
- 2014-021 *Neusser, J.; Rohde, C.; Schleper, V.:* Relaxed Navier-Stokes-Korteweg Equations for Compressible Two-Phase Flow with Phase Transition
- 2014-020 *Kabil, B.; Rohde, C.:* Persistence of undercompressive phase boundaries for isothermal Euler equations including configurational forces and surface tension
- 2014-019 *Bilyk, D.; Markhasin, L.:* BMO and exponential Orlicz space estimates of the discrepancy function in arbitrary dimension
- 2014-018 *Schmid, J.:* Well-posedness of non-autonomous linear evolution equations for generators whose commutators are scalar
- 2014-017 *Margolis, L.:* A Sylow theorem for the integral group ring of $PSL(2, q)$
- 2014-016 *Rybak, I.; Magiera, J.; Helmig, R.; Rohde, C.:* Multirate time integration for coupled saturated/unsaturated porous medium and free flow systems
- 2014-015 *Gaspoz, F.D.; Heine, C.-J.; Siebert, K.G.:* Optimal Grading of the Newest Vertex Bisection and H^1 -Stability of the L_2 -Projection
- 2014-014 *Kohler, M.; Krzyżak, A.; Walk, H.:* Nonparametric recursive quantile estimation
- 2014-013 *Kohler, M.; Krzyżak, A.; Tent, R.; Walk, H.:* Nonparametric quantile estimation using importance sampling
- 2014-012 *Györfi, L.; Ottucsák, G.; Walk, H.:* The growth optimal investment strategy is secure, too.
- 2014-011 *Györfi, L.; Walk, H.:* Strongly consistent detection for nonparametric hypotheses
- 2014-010 *Köster, I.:* Finite Groups with Sylow numbers $\{q^x, a, b\}$
- 2014-009 *Kahnert, D.:* Hausdorff Dimension of Rings
- 2014-008 *Steinwart, I.:* Measuring the Capacity of Sets of Functions in the Analysis of ERM

- 2014-007 *Steinwart, I.:* Convergence Types and Rates in Generic Karhunen-Loève Expansions with Applications to Sample Path Properties
- 2014-006 *Steinwart, I.; Pasin, C.; Williamson, R.; Zhang, S.:* Elicitation and Identification of Properties
- 2014-005 *Schmid, J.; Griesemer, M.:* Integration of Non-Autonomous Linear Evolution Equations
- 2014-004 *Markhasin, L.:* L_2 - and $S_{p,q}^r B$ -discrepancy of (order 2) digital nets
- 2014-003 *Markhasin, L.:* Discrepancy and integration in function spaces with dominating mixed smoothness
- 2014-002 *Eberts, M.; Steinwart, I.:* Optimal Learning Rates for Localized SVMs
- 2014-001 *Giesselmann, J.:* A relative entropy approach to convergence of a low order approximation to a nonlinear elasticity model with viscosity and capillarity
- 2013-016 *Steinwart, I.:* Fully Adaptive Density-Based Clustering
- 2013-015 *Steinwart, I.:* Some Remarks on the Statistical Analysis of SVMs and Related Methods
- 2013-014 *Rohde, C.; Zeiler, C.:* A Relaxation Riemann Solver for Compressible Two-Phase Flow with Phase Transition and Surface Tension
- 2013-013 *Moroianu, A.; Semmelmann, U.:* Generalized Killing spinors on Einstein manifolds
- 2013-012 *Moroianu, A.; Semmelmann, U.:* Generalized Killing Spinors on Spheres
- 2013-011 *Kohls, K.; Rösch, A.; Siebert, K.G.:* Convergence of Adaptive Finite Elements for Control Constrained Optimal Control Problems
- 2013-010 *Corli, A.; Rohde, C.; Schleper, V.:* Parabolic Approximations of Diffusive-Dispersive Equations
- 2013-009 *Nava-Yazdani, E.; Polthier, K.:* De Castelju's Algorithm on Manifolds
- 2013-008 *Bächle, A.; Margolis, L.:* Rational conjugacy of torsion units in integral group rings of non-solvable groups
- 2013-007 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups over composition algebras
- 2013-006 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups, semifields, and translation planes
- 2013-005 *Eck, C.; Kutter, M.; Sändig, A.-M.; Rohde, C.:* A Two Scale Model for Liquid Phase Epitaxy with Elasticity: An Iterative Procedure
- 2013-004 *Griesemer, M.; Wellig, D.:* The Strong-Coupling Polaron in Electromagnetic Fields
- 2013-003 *Kabil, B.; Rohde, C.:* The Influence of Surface Tension and Configurational Forces on the Stability of Liquid-Vapor Interfaces
- 2013-002 *Devroye, L.; Ferrario, P.G.; Györfi, L.; Walk, H.:* Strong universal consistent estimate of the minimum mean squared error
- 2013-001 *Kohls, K.; Rösch, A.; Siebert, K.G.:* A Posteriori Error Analysis of Optimal Control Problems with Control Constraints
- 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