A universal strong law of large numbers for conditional expectations via nearest neighbors

Harrow Walk

Preprint 2006/005
A universal strong law of large numbers for conditional expectations via nearest neighbors

Harro Walk

Preprint 2006/005
A universal strong law of large numbers for conditional expectations via nearest neighbors

Harro Walk
Stuttgart, Germany

Abstract
For $k_n$-nearest neighbor estimates of a regression $Y$ on $X$ ($d$-dimensional random vector $X$, integrable real random variable $Y$) based on observed independent copies of $(X,Y)$, strong universal pointwise consistency is shown, i.e., strong consistency $P_X$-almost everywhere for general distribution of $(X,Y)$. With tie-breaking by indices, this means validity of a universal strong law of large numbers for conditional expectations $E(Y|X = x)$.

AMS 1991 subject classification: 62G07; 62H12; 60F15

Keywords: Conditional expectation; Nearest neighbor regression estimation; Strong universal pointwise consistency; Strong law of large numbers

1 Introduction

The estimation sequence $(m_n)$ is called strongly universally pointwise consistent, if

$$\text{almost surely } m_n(x) \rightarrow m(x) \mod \mu \quad (1)$$

for all distributions of $(X,Y)$ with $E|Y| < \infty$. If, in the case that $\mu$ is concentrated in a single point $x^* \in \mathbb{R}^d$, (1) immediately yields Kolmogorov’s strong law of large numbers (SLLN) $(Y_1 + \ldots + Y_n)/n \rightarrow EY_1 = m(x^*)$ almost surely, then the strong universal consistency result can be considered as a universal strong law of large numbers for conditional expectations. In literature one finds several results on strong universal pointwise consistency which concern modifications of averaging estimates: kernel estimates with truncated $Y_i$’s in Kozek et al. [9], modified recursive partitioning estimate in Algoet [1], (modified) truncated kernel estimate, modified recursive kernel estimate and modified truncated nearest neighbor estimate in Algoet and Györfi [2], (semi-)recursive partitioning and (semi-)recursive kernel estimate in Walk [14]. Strong pointwise consistency of the classical Nadaraya-Watson kernel estimate was established under boundedness or moment conditions (stronger than $E|Y| < \infty$) by Devroye [3], Greblicki et al. [7], Zhao and Fang [16], Stute [13] and Kozek et al. [9] or under regularity conditions in Mukerjee [10] and Kozek et al. [9]. Györfi et
al. [8] mention strong universal pointwise consistency of the classical kernel estimate and
the classical nearest neighbor estimate as open problems. This paper gives an affirmative
answer for the latter estimate, i.e., it states (1) for \( k_n \)-nearest neighbor estimates for
suitable \( (k_n) \) under the only condition \( E|Y| < \infty \) (Theorem 1). This result comprehends
a universal strong law of large numbers for conditional expectations. Tools in the proof
are Etemadi’s [6] device to prove classical strong laws of large numbers, a variant of the
generalized Lebesgue density theorem concerning \( E m_n(x) \rightarrow m(x) \mod \mu \) (Lemma 1),
the Efron-Stein inequality for variances in Steele’s [11] version (Lemma 2), a sharpened
covering lemma for nearest neighbors (Lemma 3) with corollaries (Lemmas 4 - 8).

2 Result

For the definition \( m_n(x) \) of the \( k_n \)-nearest neighbor estimate, \( x \in \mathbb{R}^d \) fixed, the data
\((X_1, Y_1), \ldots, (X_n, Y_n)\) are reordered according to increasing values of \( \|X_i - x\| \) (euclidean
norm) where the reordered data sequence is denoted by

\[
(X_{1,n}(x), Y_{1,n}(x)), \ldots, (X_{n,n}(x), Y_{n,n}(x))
\]

with \( X_{k,n}(x) \) as the so-called \( k \)th nearest neighbor (\( k \)-NN) of \( x \) in \( \{X_1, \ldots, X_n\} \). The \( k_n \)-NN
regression function estimate is defined by

\[
m_n(x) := \frac{1}{k_n} \sum_{i=1}^{k_n} Y_{i,n}(x) = \frac{1}{k_n} \sum_{i=1}^{n} Y_i I[X_i \text{ is among the } k_n \text{ NNS of } x \text{ in } \{X_1, \ldots, X_n\}]
\]

with \( k_n \in \{1, \ldots, n-1\}, n \geq 2 \), where \( I \) denotes an indicator function.

We use two rules for breaking a so-called tie \( \|x_{i_1} - x\| = \ldots = \|x_{i_j} - x\| \). As to the
first rule (called purely random tie-breaking), let \((X, V)\) be a random vector, where \( V \)
is independent of \((X,Y)\) and uniformly distributed on \([0, 1]\). We also artificially enlarge
the random data set by introducing real random variables \( V_1, V_2, \ldots \) such that the \((d + 2)\)-dimensional random vectors \((X, V, Y), (X_1, V_1, Y_1), (X_2, V_2, Y_2), \ldots \) are independent and
identically distributed. Especially the \( V_i \)'s have uniform distribution on \([0, 1]\), and each
\((X_i, V_i)\) is distributed as \((X, V)\). Ties, now in context with \( \|(x_i, V_i) - (x, V)\| \) instead of
\( \|x_i - x\| \), appear only with probability zero. In contrast to the global rule described in
Györfi et al [8], pp. 86, 87, we use this enlargement only in the above context, i.e., only in
the realized tie situation. The second rule for breaking the tie consists in declaring \( x_{i_\nu} \)
to be “closer” than \( x_{i_\nu'} \) if \( i_\nu < i_\nu' \) (tie-breaking by indices). The formulations in this paper
concern both rules, except one of the rules is expressly mentioned.

The following theorem states strong universal consistency of the \( k_n \)-nearest neighbor
estimates.
Theorem 1. Assume $E|Y| < \infty$, and let $k_n \in \{1, \ldots, n - 1\}$, $n \geq 2$, with $k_n \uparrow, k_n/n^\beta \to c \in (0, \infty)$ with $0 < \beta < 1$. Let the $k_n$-nearest neighbor estimation be defined by (2) with purely random tie-breaking or tie-breaking by indices. Then (1) holds.

Remark 1. If in Theorem 1 especially $\mu$ is concentrated on $\{x^*\}$ for some $x^* \in \mathbb{R}^d$, then almost surely $m_n(x^*) \to m(x^*) = EY$. In the case of tie-breaking by indices, this means

almost surely \( \frac{Y_1 + \ldots + Y_{k_n}}{k_n} \to EY \),

thus, because \( \{k_n; n \geq 2\} \subset \{n_0, n_0 + 1, \ldots\} \) for some \( n_0 \),

almost surely \( \frac{Y_1 + \ldots + Y_n}{n} \to EY \).

Therefore Theorem 1 is a universal strong law of large numbers for conditional expectations.

3 Proofs

First we give some tools (Lemmas 1 - 8) and then prove Theorem 1. Let the assumptions of Theorem 1 be fulfilled.

Lemma 1.

\[ E m_n(x) \to m(x) \mod \mu. \]

Proof. We shall use $k_n/n \to 0$. We notice $E(Y_i|X_i) = m(X_i)$ and thus

\[ E m_n(x) = \frac{1}{k_n} \sum_{i=1}^{n} E m(X_i) I_{\{X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}\}}. \]

Because the random vectors $(X_1, Y_i)$ are independent and identically distributed, under both rules of tie breaking the left-hand side has the same value. Therefore we may restrict to the case of purely random tie-breaking and obtain

\[ E m_n(x) = \frac{n}{k_n} E m(X_1) I_{\{X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}\}}. \]

We shall use the argument in the proof of the generalized pointwise Lebesgue density theorem (see, e.g., Wheeden and Zygmund ([15], chapter 10), and Győrfi et al. ([8], section 24.2)) and of a further generalization due to Greblicki et al. ([7] (see also Győrfi et al. ([8], Lemma 24.8)).

In the first step, for an arbitrary $\mu$-integrable $f : \mathbb{R}^d \to \mathbb{R}$ we show existence of a constant $c$ depending on $d$ such that

\[ \mu \left\{ x \in \mathbb{R}^d : \sup_n \frac{n}{k_n} E |f(X_1)| I_{\{X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}\}} > \alpha \right\} \leq \frac{c}{\alpha} \int |f(y)| \mu(dy) \]

for any $\alpha > 0$. Set

\[ A^{(n)}_t := \left\{ y \in \mathbb{R}^d : P \{y \text{ is among the } k_n \text{ NNs of } x \text{ in } \{y, X_2, \ldots, X_n\} \} > t \}, \ t \in (0, 1), \]

\[ \mu \left\{ x \in \mathbb{R}^d : \sup_n \frac{n}{k_n} E |f(X_1)| I_{\{X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}\}} > \alpha \right\} \leq \frac{c}{\alpha} \int |f(y)| \mu(dy) \]

for any $\alpha > 0$. Set

\[ A^{(n)}_t := \left\{ y \in \mathbb{R}^d : P \{y \text{ is among the } k_n \text{ NNs of } x \text{ in } \{y, X_2, \ldots, X_n\} \} > t \}, \ t \in (0, 1), \]

\[ \mu \left\{ x \in \mathbb{R}^d : \sup_n \frac{n}{k_n} E |f(X_1)| I_{\{X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}\}} > \alpha \right\} \leq \frac{c}{\alpha} \int |f(y)| \mu(dy) \]
which is $\emptyset$ or a ball in $\mathbb{R}^d$ centered at $x$. Then
\[
\frac{n}{k_n}E[f(X_1)|I[X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}] = \frac{n}{k_n} \int |f(y)| \int_0^1 I_{A_t^{(n)}}(y) dt \mu(dy)
\]
\[
= \frac{n}{k_n} \int_0^1 \left[ \int_{A_t^{(n)}} |f(y)| \mu(dy) \right] dt
\]
\[
= \int_0^1 \left[ \int_{A_t^{(n)}} |f(y)| \mu(dy) \right] dt / \int_0^1 \mu(A_t^{(n)}) dt
\]
\[
\leq \sup_{\text{rational } t>0} \int_{A_t^{(n)}} |f(y)| \mu(dy) / \mu(A_t^{(n)}),
\]
\[
\leq \sup_{h \in H} \int_{S_{x,h}} |f(y)| \mu(dy) / \mu(S_{x,h})
\]
for a suitable countable set $H \subset (0, \infty)$. This together with the well-known fact that

\[
\mu\left( \{x \in \mathbb{R}^d : \sup_{h \in H} \int_{S_{x,h}} |f(y)| \mu(dy) / \mu(S_{x,h}) > \alpha \} \right) \leq \frac{c}{\alpha} \int |f(y)| \mu(dy)
\]

for any $\alpha > 0$ with some constant $c$ depending on $d$ (see, e.g., Wheeden and Zygmund ([15], Lemma 10.47), and Györfi et al. ([8], Lemma 24.4)), yields the desired auxiliary result.

In the second step, for an arbitrary fixed $\epsilon > 0$ we choose a continuous function $g : \mathbb{R}^d \to \mathbb{R}$ with compact support such that $\int |m(y) - g(y)| \mu(dy) < \frac{c^2}{2(\epsilon+1)}$ with constant $c$ from the first step. For each $x \in \text{support}(\mu)$, because of $k_n/n \to 0$, one has

almost surely $||X_{(k_n,n)}(x) - x|| \to 0$,

which is a consequence of the strong law of large numbers (see Györfi et al. ([8], Lemma 6.1)), thus

almost surely $\frac{1}{k_n} \sum_{i=1}^n |g(X_i) - g(x)| I[X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}] \to 0$

and

\[
d_n(x) := \frac{n}{k_n}E[|g(X_1) - g(x)| I[X_1 \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}] \to 0.
\]
Lemma 3.

Proof. One has, for $x \in \text{support}(\mu)$,

\[
|E m_n(x) - m(x)| \leq \frac{n}{k_n} E|m(X_1) - m(x)| I[X_1 \text{ is among the } k_n \text{ NN's of } x \text{ in } \{X_1, \ldots, X_n\}] \\
\leq \frac{n}{k_n} E|m(X_1) - g(X_1)| I[X_1 \text{ is among the } k_n \text{ NN's of } x \text{ in } \{X_1, \ldots, X_n\}] + |m(x) - g(x)| + d_n(x) \\
= p_n(x) + |m(x) - g(x)| + d_n(x) \\
= p_n(x) + |m(x) - g(x)| + o(1).
\]

Define the set

\[
T_\varepsilon := \{x \in \mathbb{R}^d : \sup_n p_n(x) + |m(x) - g(x)| > \varepsilon\}.
\]

By the first step and the Markov inequality

\[
\mu(T_\varepsilon) \leq \mu\left(\{x \in \mathbb{R}^d : \sup_n p_n(x) > \varepsilon/2\}\right) + \mu\left(\{x \in \mathbb{R}^d : |m(x) - g(x)| > \varepsilon/2\}\right) \\
\leq \frac{2c + 2}{\varepsilon} \int |m(x) - g(x)| \mu(dx) \leq \varepsilon.
\]

Now $\varepsilon \to 0$ yields the assertion. \hfill \Box

The Efron-Stein [5] inequality on variances in Steele’s [11] version (see Györfi et al. [8] for further references) will be formulated for the independent identically distributed random vectors $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n), \tilde{Z}_1 = (\hat{X}_1, \hat{Y}_1), \ldots, \tilde{Z}_n = (\hat{X}_n, \hat{Y}_n)$.

**Lemma 2.** Let $f : \mathbb{R}^{(d+1)n} \to \mathbb{R}$ be measurable with square integrability of $f(Z_1, \ldots, Z_n)$. Then

\[
\text{Var}(f(Z_1, \ldots, Z_n)) \leq \frac{1}{2} \sum_{j=1}^{n} E|f(Z_1, \ldots, z_j, \ldots, Z_n) - f(Z_1, \ldots, \tilde{z}_j, \ldots, Z_n)|^2.
\]

Let $\gamma_d$ be the minimal number of closed cones $C_1, \ldots, C_{\gamma_d}$ of angle $\pi/4$ which are centered at 0 with different central directions such that their union covers $\mathbb{R}^d$. The following lemma sharpens Corollary 6.1 in Györfi et al. [8], which deals with $P[x \text{ is among the } k \text{ NNs of } X \text{ in } \{x, X_2, \ldots, X_n\}]$.

**Lemma 3.** Let $x \in \mathbb{R}^d$, $1 \leq k < n$. With purely random tie-breaking,

\[
P[x \text{ is } k\text{-NN of } X \text{ in } \{x, X_2, \ldots, X_n\}] \leq \frac{\gamma_d}{n}.
\]

**Proof.** We use ideas from the proof of Corollary 6.1 in Györfi et al. [8]. Let $k \geq 2$. The treatment of the case $k = 1$ is analogous, but simpler. For $i \in \{1, \ldots, n\}$ let $X_{J_1(i)}, \ldots, X_{J_{k-1}(i)}$ with random indices $J_1(i) < \ldots < J_{k-1}(i)$ in $\{1, \ldots, i-1, i+1, \ldots, n\}$ be the $k-1$ NNs of $X_i$ in $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$. Then,

\[
P[x \text{ is } k\text{-NN of } X \text{ in } \{x, X_2, \ldots, X_n\}] = \frac{\gamma_d}{n}.
\]
implies

\[ \text{by symmetry} \]

\[ \leq \frac{1}{n} \sum_{i=1}^{n} P \{ x \text{ is 1st NN of } X_i \text{ in } \{ x, X_1, \ldots, X_{i+1}, X_{i+1}, \ldots, X_n \} \setminus \{ X_{j_1(i)}, \ldots, X_{j_{k-1}(i)} \} \} \]

\[ = \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} I \{ x \text{ is 1st NN of } X_i \text{ in } \{ x, X_1, \ldots, X_{i+1}, X_{i+1}, \ldots, X_n \} \setminus \{ X_{j_1(i)}, \ldots, X_{j_{k-1}(i)} \} \}. \]

Because for \( u, u' \in x + C_j \ (j \in \{1, \ldots, \gamma_d\}) \) with \( u \neq x \) the inequality \( ||u - x|| \leq ||u' - x|| \) implies \( ||u - u'|| < ||u' - x|| \) and thus \( ||u - u'|| \geq ||u' - x|| \) implies \( ||u - x|| > ||u' - x|| \), we can notice: if \( x \) is the 1st NN of some \( X_i \) in \( x + C_j \) \((i = 1, \ldots, n)\) in the set \( A_{i,j} \cup \{x\} \) with \( A_{i,j} \) consisting of those \( X_k \) \((l \in \{1, \ldots, i - 1, i + 1, \ldots, n\}\) \(\{j_1(i), \ldots, j_{k-1}(i)\})\) falling into \( x + (C_j \setminus \{0\}) \), then \( X_k \) \((in x + C_j)\) is the unique 1st NN of \( x \) in \( A_{i,j} \). Thus the number of such \( X_i \)'s is at most \( \gamma_d \) and the expected sum above is bounded by \( \gamma_d \). This yields the assertion. \( \square \)

**Lemma 4.** Let \( 1 \leq k < n \) and \( f : \mathbb{R}^d \rightarrow \mathbb{R}_+ \) be measurable. Then

\[ \sum_{j=1}^{n} \mathbb{E} f(X_j) I \{ X_j \text{ is } k\text{-NN of } X \in \{X_1, \ldots, X_n\} \} \leq \gamma_d \mathbb{E} f(X). \]

**Proof.** Because the random vectors \( (X_i, Y_i) \) are independent and identically distributed, under both rules of tie-breaking the left-hand side has the same value. Therefore we may restrict to the case of purely random tie-breaking and obtain

\[ \sum_{j=1}^{n} \mathbb{E} f(X_j) I \{ X_j \text{ is } k\text{-NN of } X \in \{X_1, \ldots, X_n\} \} \]

\[ = \sum_{j=1}^{n} \int f(x) P \{ x \text{ is } k\text{-NN of } X \in \{X_1, \ldots, X_{j-1}, x, X_{j+1}, \ldots, X_n\} \} \mu(dx) \]

\[ \leq \gamma_d \int f(x) \mu(dx) \]

(by Lemma 3)

\[ = \gamma_d \mathbb{E} f(X). \]

\( \square \)

**Lemma 5.** Let \( 1 \leq k < n \) and \( Y_j \) be square integrable.

a) \( \sum_{j=1}^{n} \mathbb{E} Y_j^2 I \{ X_j \text{ is } k\text{-NN of } X \in \{X_1, \ldots, X_n\} \} \leq \gamma_d \mathbb{E} Y^2. \)
b) \( EY_{k,n}(X)^2 \leq \gamma_d EY^2. \)

**Proof.**

a) The inequality is obtained by Lemma 4 with \( f(X_j) = E(Y_j^2|X_j). \)

b) We obtain

\[
EY_{k,n}(X)^2 = \sum_{j=1}^{n} EY_{k,n}(X)^2 I_{X_j \text{ is } k-\text{NN of } X \text{ in } \{X_1, \ldots, X_n\}}
\]

\[
= \sum_{j=1}^{n} EY_j^2 I_{X_j \text{ is } k-\text{NN of } X \text{ in } \{X_1, \ldots, X_n\}}
\]

\[
\leq \gamma_d EY^2
\]

by part a) \( \square \)

**Lemma 6.** Let \( q > 0, \ 1 \leq k < n. \)

a) \( P[Y_n > q, \ X_n \text{ is } k-\text{NN of } X \text{ in } \{X_1, \ldots, X_n\}] \leq \frac{\gamma_d}{n} P[Y > q]. \)

b) \( P[Y_n > q, \ X_n \text{ is among the } k \ \text{NNs of } X \text{ in } \{X_1, \ldots, X_n\}] \leq \frac{k}{n} \gamma_d P[Y > q]. \)

**Proof.**

a) The left-hand side below concerns tie-breaking by indices as well as purely random tie-breaking, with differing probability values. We obtain

\[
P[Y_n > q, \ X_n \text{ is } k-\text{NN of } X \text{ in } \{X_1, \ldots, X_n\}]
\]

\[
\leq \int P[Y_n > q|X_n = x] P[x \text{ is } k-\text{NN of } X \text{ in } \{X_1, \ldots, X_{n-1}, x\}]
\]

under purely random tie-breaking \( \mu(dx) \)

\[
\leq \frac{\gamma_d}{n} \int P[Y_n > q|X_n = x] \mu(dx)
\]

(by Lemma 3)

\[
= \frac{\gamma_d}{n} P[Y > q].
\]

b) Immediately by part a). \( \square \)
Lemma 7. Let $Y_j \geq 0$ be square integrable. Let $1 \leq k < M \leq N$. Then

$$
\int \text{Var} \left( \sum_{j=1}^{M} Y_j I_{X_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_N\}} \right) \mu(dx)
\leq 2\gamma d k EY^2.
$$

Proof. Let $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_N, \tilde{Y}_N)$ be $(d+1)$-dimensional random vectors such that $(X_1, Y_1), \ldots, (X_N, Y_N), (\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_n, \tilde{Y}_N)$ are independent and identically distributed. With

$$
F_{N,j}(x) := [X_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_N\}],
$$

$$
G_{N,j}(x) := \left[\tilde{X}_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{j-1}, \tilde{X}_j, X_{j+1}, \ldots, X_N\}\right],
$$

by Lemma 2 we obtain

$$
\text{Var} \left( \sum_{j=1}^{M} Y_j I_{F_{N,j}(x)} \right)
\leq \frac{1}{2} \sum_{j=1}^{M} E \left( Y_j - \tilde{Y}_j \right)^2 I_{F_{N,j}(x) \cap G_{N,j}(x)} + \frac{1}{2} \sum_{j=1}^{M} E \left( Y_j^2 + Y_{k+1,N}(x)^2 \right) I_{F_{N,j}(x) \cap G_{N,j}(x)}
\quad + \frac{1}{2} \sum_{j=1}^{M} \sum_{j=M+1}^{N} E Y_{k+1,N}(x)^2 I_{F_{N,j}(x) \cap G_{N,j}(x)}
\quad + \frac{1}{2} \sum_{j=M+1}^{N} E Y_{k,N}(x)^2 I_{F_{N,j}(x) \cap G_{N,j}(x)},
$$

where on $F_{N,j}(x) \cap G_{N,j}(x)$ $X_{k,N}(x)$ is only the $(k+1)$-NN of $x$ in $\{X_1, \ldots, X_{j-1}, \tilde{X}_j, X_{j+1}, \ldots, X_N\}$. Thus, by symmetry,

$$
\text{Var} \left( \sum_{j=1}^{M} Y_j I_{F_{N,j}(x)} \right)
\leq \sum_{j=1}^{M} E Y_j^2 I_{F_{N,j}(x)} + \sum_{j=1}^{M} E Y_{k+1,N}(x)^2 I_{F_{N,j}(x)} + \sum_{j=M+1}^{N} E Y_{k+1,N}(x)^2 I_{F_{N,j}(x)}
\leq \sum_{j=1}^{M} E Y_j^2 I_{F_{M,j}(x)} + k E Y_{k+1,N}(x)^2.
$$

Now

$$
\int \text{Var} \left( \sum_{j=1}^{M} Y_j I_{F_{N,j}(x)} \right) \mu(dx)
\leq 2\gamma d k EY^2.
$$
\[ \leq \sum_{l=1}^{k} \sum_{j=1}^{M} EY_j^2 I_{X_j \text{ is } \ell\text{-NN of } X \text{ in } \{X_1, \ldots, X_N\}} + kEY_{k+1,N}(X)^2 \]
\[ \leq \gamma_d \sum_{l=1}^{k} EY^2 + \gamma_d kEY^2 \]
(by Lemma 5a,b)
\[ \leq 2\gamma_d kEY^2. \]

**Lemma 8.** Let \( Y_j \geq 0 \) be square integrable. Let \( 1 \leq k < M < N \leq (1+\rho)M \), with \( \rho > 0 \). Then
\[
\int \text{Var} \left( \sum_{j=M+1}^{N} Y_j I_{X_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right) \mu(dx)
\]
\[ \leq 4\rho(1+\rho)\gamma_d kEY^2. \]

**Proof.** Let \((\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_N, \tilde{Y}_N)\) be \((d+1)\)-dimensional random vectors such that \((X_1, Y_1), \ldots, (X_N, Y_N), (\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_N, \tilde{Y}_N)\) are independent and identically distributed. By Lemma 2 we have
\[
\text{Var} \left( \sum_{j=M+1}^{N} Y_j I_{X_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right)
\]
\[ \leq \frac{1}{2} \sum_{l=1}^{M} E \left| \sum_{j=M+1}^{N} Y_j \left( I_{X_j \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right) - I_{X_l \text{ is among the } k+1 \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l-1}, X_l, X_{l+1}, \ldots, X_M, X_j\}} \right|^2
\]
\[ + \frac{1}{2} \sum_{l=M+1}^{N} E \left| Y_l I_{X_l \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_l\}} \right| \]
\[ - \tilde{Y}_l I_{\tilde{X}_l \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, \tilde{X}_l\}} \right|^2
\]
\[ =: V_1(x) + V_2(x). \]

Then, similarly to the proof of Lemma 7, by symmetry
\[
V_1(x) \leq 2 \sum_{l=1}^{M} E \left( \sum_{j=M+1}^{N} Y_j I_{X_l \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right)^2
\]
\[ \leq 2E \left( \sum_{j=M+1}^{N} Y_j I_{X_j \text{ is } (k+1)\text{-NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right)^2 \]
\[ \leq 2E \left( \sum_{j=M+1}^{N} Y_j I_{X_j \text{ is } (k+1)\text{-NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right)^2 \]

13
Let \[\sum_{l=1}^{M} I_{X_l \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M\}}\]

\[= 2kE \left( \sum_{j=M+1}^{N} Y_j I_{X_j \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}} \right)^2\]

\[= 2k \sum_{j=M+1}^{N} EY_j^2 I_{X_j \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}}\]

\[+ 2k \sum_{i,j \in \{M+1, \ldots, N\}} EY_i Y_j I_{X_i \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_i\}}\]

\[I_{X_j \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}}\]

\[= 2kW_1(x) + 2kW_2(x).\]

Now

\[\int W_1(x) \mu(dx)\]

\[= \sum_{j=M+1}^{N} EY_j^2 I_{X_j \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_j\}}\]

\[= \sum_{j=M+1}^{N} \int E(Y_j^2 | X_j = x) P[x \text{ is } (k+1)-\text{NN of } X \text{ in } \{X_1, \ldots, X_M, x\}] \mu(dx)\]

(with label \(j\) for \(x\) in case of tie-breaking by indices)

\[\leq (N - M) \int E(Y^2 | X = x) P[x \text{ is } (k+1)-\text{NN of } X \text{ in } \{X_1, \ldots, X_M, x\}]\]

under purely random tie-breaking\(\mu(dx)\)

\[\leq \frac{N - M}{M + 1} \gamma_\text{d} \int E(Y^2 | X = x) \mu(dx)\]

(by Lemma 3)

\[\leq \rho \gamma_\text{d} EY^2.\]

Further

\[\int W_2(x) \mu(dx)\]

\[= 2E \sum_{i,j \in \{M+1, \ldots, N\}} Y_i Y_j I_{X_i \text{ is } (k+1)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_i, X_j\}}\]

\[I_{X_j \text{ is } (k+2)-\text{NN of } x \text{ in } \{X_1, \ldots, X_M, X_i, X_j\}}\]

(by symmetry)
\[ \leq 2 \int \int \sum_{i,j \in \{M+1, \ldots, N\} \setminus \{i \neq j\}} E(Y_i|X_i = x)E(Y_j|X_j = \tilde{x})P[x \text{ is (}k+1\text{-NN of } X \\
\text{ in } \{X_1, \ldots, X_M, x, \tilde{x}\}, \tilde{x} \text{ is (}k+2\text{-NN of } X \text{ in } \{X_1, \ldots, X_M, x, \tilde{x}\} \\
\text{ under purely random tie-breaking}] \mu(dx)\mu(d\tilde{x}) \\
\text{(as before)} \]

\[ = 2(N - M)(N - M - 1)E(Y_{M+1}Y_{M+2} \\
I_{[X_{M+1} \text{ is (}k+1\text{-NN of } X \text{ in } \{X_1, \ldots, X_{M+2}\}, X_{M+2} \text{ is (}k+2\text{-NN of } X \text{ in } \{X_1, \ldots, X_{M+3}\}] \\
\leq \frac{(N - M)^2}{M^2} E \left[ (Y_{k+1,M+2}(X)^2 + Y_{k+2,M+2}(X)^2) \right. \\
\left. \sum_{j=1}^{M+2} I_{[X_j \text{ is (}k+1\text{-NN of } X \text{ in } \{X_1, \ldots, X_{M+2}\}]} \sum_{j=1}^{M+2} I_{[X_j \text{ is (}k+2\text{-NN of } X \text{ in } \{X_1, \ldots, X_{M+2}\}]} \right] \\
= \frac{(N - M)^2}{M^2} \left[ EY_{k+1,M+2}(X)^2 + EY_{k+2,M+2}(X)^2 \right] \\
\text{(each time under purely random tie-breaking)} \]

\[ \leq 2\rho^2\gamma_dEY^2 \]

by Lemma 5b. Finally

\[ V_2(x) \leq 2 \sum_{l=M+1}^{N} EY_l^2 I_{[X_l \text{ is among the } k \text{ NNs of } x \text{ in } \{X_1, \ldots, X_M, X_l\}]} \]

thus

\[ \int V_2(x)\mu(dx) \]

\[ \leq 2 \sum_{l=M+1}^{N} \int E(Y_l^2|X_l = x)P[x \text{ is among the } k \text{ NNs of } X \text{ in } \{X_1, \ldots, X_M, x\}] \mu(dx) \\
\text{(with label } l \text{ for } x \text{ in case of tie-breaking by indices)} \]

\[ \leq 2 \int E(Y^2|X = x) \sum_{l=M+1}^{N} P[x \text{ is among the } k \text{ NNs of } X \text{ in } \{X_1, \ldots, X_M, x\}] \\
\text{under purely random tie-breaking}] \mu(dx) \\
\text{(as before)} \]

\[ \leq 2\frac{N - M}{M + 1} \gamma_dk \int E(Y^2|X = x)\mu(dx) \\
\text{(by Lemma 3)} \]

\[ \leq 2\rho\gamma_dkEY^2. \]
Thus the assertion is obtained.

\[ \square \]

Proof of Theorem 1. We use Etemadi’s [6] device to prove strong laws of large numbers. Without loss of generality assume \( Y_i \ge 0 \). For \( c > 0 \) set \( Y_{i}^{[c]} := Y_i I_{[Y_i \le c]} \). Further set

\[
m_n^{(n)}(x) := \frac{\sum_{i=1}^{n} Y_{i}^{[k_n]} I_{[X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}]} }{k_n}, \quad x \in \mathbb{R}^d.
\]

In the first step we show that almost surely for \( \mu \)-almost all \( x \in \mathbb{R}^d \) the event

\[ B_{i}(x) := \text{[for some } n \ge i : Y_i > k_n, \ X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } \{X_1, \ldots, X_n\}] \]

occurs for only finitely many \( i \in \mathbb{N} \). Thus

\[
\text{almost surely } m_n^{(n)}(x) - m_n(x) \to 0 \mod \mu. \quad (3)
\]

Let \( r_l := \min\{j \in \mathbb{N}; k_j = l\}, \ l \in \mathbb{N} \). For \( i \in \mathbb{N} \) we notice

\[
\int P(B_i(x)) \mu(dx) = \sum_{l=k_i+1}^{\infty} P[Y_l > l, X_i \text{ is among the } l \text{ NNs of } X \text{ in } \{X_1, \ldots, X_{r_l}\}] \]

(with pairwise disjoint events)

\[
\leq \gamma_d \frac{k_i}{i} P[Y > k_i] + \sum_{l=k_i+1}^{\infty} P[Y_l > l, X_i \text{ is } l\text{-NN of } X \text{ in } \{X_1, \ldots, X_{r_l}\}]
\]

(by Lemma 6b)

\[
= A_i + D_i.
\]

Further

\[
\sum_{i=1}^{\infty} A_i = \gamma_d \sum_{i=1}^{\infty} \frac{k_i}{i} P[Y > k_i]
\]
\[
\leq c_1 \sum_{i=1}^{\infty} t^{i-1} P \left[ Y > \frac{1}{c_1} t^i \right]
\]
\[
\leq \frac{c_1^2}{\beta} \int_0^{\infty} P [Y > t] dt
\]
\[
= \frac{c_1^2}{\beta} EY < \infty
\]

with some constant \( c_1 \in (0, \infty) \), and

\[
\sum_{i=1}^{\infty} D_i = \sum_{l=1}^{\infty} \sum_{i=1}^{\gamma_{l-1}} P [Y_i > l, X_i \text{ is } l-\text{NN of } X \text{ in } \{X_1, \ldots, X_{\gamma_l}\}]
\]
\[
\leq \gamma d \sum_{l=1}^{\infty} P [Y > l]
\]
(by Lemma 4 with \( f(X_i) = E(I[Y_i > l]|X_i) \))
\[
\leq \gamma d EY < \infty.
\]

Thus

\[
\sum_{i=1}^{\infty} \int P(B_i(x))\mu(dx) < \infty.
\]

Now the Borel-Cantelli lemma yields the desired result.

In the second step we show

almost surely \( m_N(x) \to m(x) \mod \mu \). \hspace{1cm} (4)

Set \( l_n := \lfloor a^n \rfloor \) for fixed \( a > 1 \). For \( N, n \) so large that \( k_{l_n+1} \leq l_n < N \leq l_{n+1} \), we have

\[
m^*_n(x) := \frac{1}{k_{l_n+1}} \sum_{i=1}^{l_n} Y_i^{[k_{l_n}]} I_{[X_i \text{ is among the } k_{l_n} \text{ NNSs of } x \text{ in } \{X_1, \ldots, X_{l_n+1}\}]} \leq m^{(N)}_N(x)
\]
\[
\leq \frac{1}{k_{l_n}} \sum_{i=1}^{l_{n+1}} Y_i^{[k_{l_{n+1}}]} I_{[X_i \text{ is among the } k_{l_{n+1}} \text{ NNSs of } x \text{ in } \{X_1, \ldots, X_{l_n}\}]} + \frac{1}{k_{l_{n+1}}} \sum_{i=l_{n+1}+1}^{l_{n+1}} Y_i^{[k_{l_{n+1}}]} I_{[X_i \text{ is among the } k_{l_{n+1}} \text{ NNSs of } x \text{ in } \{X_{l_n+1}, \ldots, X_{l_{n+1}}\}]} =: m'_n(x) + m''_n(x). \hspace{1cm} (5)
\]

First we show

almost surely \( m^*_n(x) - E m^*_n(x) \to 0 \mod \mu \). \hspace{1cm} (6)
almost surely \( m'_n(x) - E m'_n(x) \to 0 \mod \mu, \) \hspace{1cm} (7)
almost surely \( m''_n(x) - E m''_n(x) \to 0 \mod \mu. \) \hspace{1cm} (8)

It suffices to show
\[
\int \sum Var(m'_n(x)) \mu(dx) < \infty, \tag{9}
\]
\[
\int \sum Var(m'_n(x)) \mu(dx) < \infty, \tag{10}
\]
\[
\int \sum Var(m''_n(x)) \mu(dx) < \infty. \tag{11}
\]

By Lemma 7 we obtain
\[
\sum \int Var(m'_n(x)) \mu(dx) \leq 2\gamma d \sum \frac{1}{k^2_{n+1}} k_n E (Y[k_n])^2 \leq c_2 \sum \frac{1}{a^{(n+1)\beta}} \int_0^\infty t^{2\beta} P_Y(dt) 
\]
\[
\leq c_2 \int_0^\infty \frac{1}{a^{\beta}} \int_0^\infty t^{2\beta} P_Y(dt) \int_0^\infty a^{-s\beta} ds = \frac{c_2 c'}{\beta \ln a} \int t P_Y(dt) = \frac{c_2 c'}{\beta \ln a} E Y < \infty
\]

with suitable constants \( c', c_2 \in (0, \infty) \), thus (9). Analogously, by Lemmas 7 and 8, we obtain (10) and (11), respectively. Now for \( \delta > 0 \) choose \( k'_n \in \{1, \ldots, n - 1\} \) such that \( k'_n = \lceil (1 + \delta)a^\beta k_n \rceil \) for large \( n \). By Lemma 1
\[
\lim \sup E m'_n(x) \leq \lim \sum_{i=1}^{k_n} \frac{1}{k_n k'_n} \sum_{i=1}^{l_n} Y_i I[\text{X is among the } k'_n \text{ NNs of } x] = (1 + \delta)a^\beta m(x) \mod \mu.
\]

Further
\[
\lim \sup E m_n'(x) \\
\leq \lim \sup \frac{1}{k_n} \sum_{i=l_{n+1}}^{l_{n+1}+1} E Y_i [X_i \text{ is among the } k_{l_{n+1}} \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l_{n+1}}\}] \\
\leq \lim \sup \frac{l_{n+1} - l_n}{k_n} E Y_i [X \text{ is among the } k_{l_{n+1}} \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l_{n+1}}\}] \\
\quad \text{(the latter expectation under purely random tie-breaking)} \\
= \lim \frac{l_{n+1} - l_n}{k_n} \frac{k_{l_{n+1}}}{l_n + 1} E \sum_{i=1}^{l_{n+1}} Y_i [X_i \text{ is among the } k_{l_{n+1}} \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l_{n+1}}\}] \\
\quad \text{(the expectation under purely random tie-breaking)} \\
= a^\beta (a - 1) m(x) \mod \mu
\]
by Lemma 1. We notice that for arbitrary \( C > 0 \) one has \( k_n > C \) for \( n \) sufficiently large, further
\[
\frac{1}{l_{n+1}} E \sum_{i=1}^{l_{n+1}} Y_i^{[C]} [X_i \text{ is among the } k_{l_{n}} \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l_{n+1}}\}] \\
\leq \frac{1}{l_n} E \sum_{i=1}^{l_n} Y_i^{[C]} [X_i \text{ is among the } k_{l_{n}} \text{ NNs of } x \text{ in } \{X_1, \ldots, X_{l_{n+1}}\}]
\]
(with equality in the case of purely random tie-breaking). Once more by Lemma 1 together with (5),(6),(7),(8), we then obtain
\[
\text{almost surely } \frac{1}{a} E Y^{[C]} | X = x \leq \lim \inf E m_n'(x) \\
= \lim \inf m_n^*(x) \leq \lim \inf m_N^{(N)}(x) \\
\leq \lim \sup m_N^{(N)}(x) \leq \lim \sup m_n'(x) + \lim \sup m_n''(x) \\
= \lim \sup E m_n'(x) + \lim \sup E m_n''(x) \\
\leq [(1 + \delta) + (a - 1)] a^\beta m(x) \mod \mu.
\]

Letting \( \delta \downarrow 0, a \downarrow 1 \) and \( C \uparrow \infty \) we obtain (4).

Now (3) and (4) yield the assertion. \( \square \)

References


Harro Walk
Pfaffenwaldring 57
70569 Stuttgart
Germany

E-Mail: Harro.Walk@mathematik.uni-stuttgart.de
WWW: http://www.isa.uni-stuttgart.de/LstStoch/Walk
Erschienene Preprints ab Nummer 2004/001

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


2004/010  Dillen, F., Kühnel, W.:  Total curvature of complete submanifolds of Euclidean space.


2004/014  Kühnel, W.:  Tight embeddings of simply connected 4-manifolds.


2005/001 Mielke, A.; Schmid, F.: Vortex pinning in super-conductivity as a rate-independent model


2005/003 Höfert, C.; Kimmerle, W.: On Torsion Units of Integral Group Rings of Groups of Small Order


2005/005 Borisov, D.; Ekholm, T; Kovarik, H.: Spectrum of the Magnetic Schrödinger Operator in a Waveguide with Combined Boundary Conditions

2005/006 Zelik, S.: Spatially nondecaying solutions of 2D Navier-Stokes equation in a strip

2005/007 Meister, A.: Deconvolving compactly supported densities

2005/008 Förster, C., Weidl, T.: Trapped modes for an elastic strip with perturbation of the material properties

2006/001 Dippon, J., Schiemert, D.: Stochastic differential equations driven by Gaussian processes with dependent increments

2006/002 Lesky, P.A.: Orthogonale Polynomlösungen von Differenzengleichungen vierter Ordnung

2006/003 Dippon, J., Schiemert, D.: Option Pricing in a Black-Scholes Market with Memory

2006/004 Banchoff, T., Kühnel, W.: Tight polyhedral models of isoparametric families, and PL-taut submanifolds

2006/005 Walk, H.: A universal strong law of large numbers for conditional expectations via nearest neighbors