Let \((X,Y), (X_1,Y_1), \ldots, (X_n,Y_n)\) be i.i.d. \(\mathbb{R}^r \times \mathbb{R}\)-valued random vectors with \(E|Y| < \infty\), and let \(Q_n(x)\) be a kernel estimate of the regression function \(Q(x) = E(Y|X = x)\). In this paper, we establish an exponential bound of the mean deviation between \(Q_n(x)\) and \(Q(x)\) given the training sample \(Z^n = (X_1,Y_1,\ldots,X_n,Y_n)\), under the conditions as weak as possible.
EXponential BoundS of Mean Error for thE Kernal Estimates of Regression Functions

L. C. Zhao
Center for Multivariate Analysis
University of Pittsburgh

Technical Report No. 85-50

December 1985

Center for Multivariate Analysis
University of Pittsburgh
515 Thackeray Hall
Pittsburgh, PA 15260

*Research sponsored by the Air Force Office of Scientific Research (AFSC)
under Contract F49620-85-C-0008. The United States Government is authorized
to reproduce and distribute reprints for governmental purposes notwithstanding
any copyright notation herein.
Let \((X,Y), (X_1,Y_1), \ldots, (X_n,Y_n)\) be i.i.d. \(R^r \times R\)-valued random vectors with \(E|Y| < \infty\), and let \(Q_n(x)\) be a kernel estimate of the regression function \(Q(x) = E(Y|X = x)\). In this paper, we establish an exponential bound of the mean deviation between \(Q_n(x)\) and \(Q(x)\) given the training sample \(Z^n = (X_1,Y_1, \ldots, X_n,Y_n)\), under the conditions as weak as possible.
1. INTRODUCTION

Let \((X,Y), (X_1,Y_1),\ldots,(X_n,Y_n)\) be i.i.d. \(\mathbb{R}^r \times \mathbb{R}\)-valued random vectors with \(E|Y| < \infty\). To estimate \(Q(x) = E(Y|X=x)\), the regression function of \(Y\) with respect to \(X\), Nadaraya (1964) and Watson (1964) proposed a class of kernel estimates of the form

\[
Q_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i - x}{h} Y_i / \sum_{j=1}^{n} \frac{1}{h} K\left(\frac{X_i - x}{h}\right),
\]

where \(K\) is a probability density function on \(\mathbb{R}^r\), and \(h = h_n\) is a sequence of positive numbers. Write

\[
W_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i - x}{h} / \sum_{j=1}^{n} \frac{1}{h} K\left(\frac{X_i - x}{h}\right),
\]

we define \(W_n(x) = 1/n, i=1,2,\ldots,n\), when \(0/0\) appears. Many scholars studied convergence problems of these estimates from various points of view. For the universal consistency, one can refer to, for example, Devroye and Wagner (1980I), Spiegelman and Sacks (1980). For the pointwise moment-consistency, see Devroye (1981). For the pointwise a.s. consistency, see Devroye (1981), Greblicki, Krzyzak and Pawlak (1984), Zhao and Fang (1985). In this paper, we study another convergence of these estimates.

Let \(Z^n = (X_1,Y_1),\ldots,(X_n,Y_n)\) be a training sample, \(g_n = g_n(x,Z^n)\) be an estimate of \(Q(x)\). In some problems, we are interested in the following mean deviation of \(g_n\) given the training sample \(Z^n\):

\[
D(g_n) = E|g_n(x,Z^n) - Q(x)||Z^n|
\]

\[
= \int_{\mathbb{R}^r} |g_n(x,Z^n) - Q(x)| F(dx),
\]

where \(F\) denotes the distribution of \(X\). Devroye and Wagner (1980II) proved
that
\[ \lim_{n \to \infty} D(Q_n) = 0 \text{ a.s.} \]

for the kernel estimates \( Q_n(x) \) of \( Q(x) \), if the following conditions are satisfied:

(i) \( Y \) is bounded,
(ii) \( F \) has a density \( f \),
(iii) \( K \) is bounded and
\[ \int_{\mathbb{R}^r} \psi(x) \, dx < \infty, \]
where
\[ \psi(x) = \sup_{||u|| > ||x||} K(u), \, x \in \mathbb{R}^r \]
and \( ||.|| \) is the \( L_2 \) norm or \( L_\infty \) norm on \( \mathbb{R}^r \),
(iv) \( h_n \to 0 \) and \( \sum_{n} \exp(-\alpha nh_n^r) < \infty \) for any \( \alpha > 0 \).

In this paper, we establish an exponential bound for the above mentioned deviation of \( Q_n \). Take \( ||.|| \) as \( L_2 \) or \( L_\infty \) norm, and denote by \( I(A) \) or \( I_A \) the indicator of set \( A \). We establish the following

**Theorem.** Let \( Q_n(x) \) be a kernel estimate defined by (1). Suppose that the following conditions are satisfied:

(i) \( Y \) is bounded.
(ii) \( F \), the distribution of \( X \), has a density of \( f \).
(iii) There exist positive constant \( \alpha \) and \( \rho_0 \) such that
\[ K(x) \geq \alpha I(||x|| \leq \rho_0), \, x \in \mathbb{R}^r. \]
(iv) \( h \to 0 \) and \( nh^r \to \infty \) as \( n \to \infty \).
Then for any given $\varepsilon > 0$, we have

$$P(D(Q_n) \geq \varepsilon) \leq e^{-cn}.$$  \hspace{1cm} (6)

where $C > 0$ is a constant independent of $n$.

Obviously, we need only to give the proof for $L_\infty$ norm. We shall introduce some lemmas in section 2, and give a proof of the theorem in section 3.
2. SOME LEMMAS

For simplicity we use the following convention: \( \varepsilon, \varepsilon_1, \varepsilon_2, \ldots, \alpha, \beta_1, \beta_2, \delta, \) etc., are all constants independent of \( n. \) \( \#(A) \) denotes the cardinal of set \( A. \) \( S_{x, \rho} = \{ u \in \mathbb{R}^d : ||u-x|| \leq \rho \}. \) \( F^* \) and \( \lambda^* \) denote the outer measure generated by \( F \) and the Lebesgue measure \( \lambda \) (on \( \mathbb{R}^d \)), respectively. \( F_n \) denotes the empirical measure of \( X_n = (X_1, \ldots, X_n) \). We now give four lemmas which are needed in the sequel.

**Lemma 1 (Besicovitch Covering Lemma).** Let \( E \) be a bounded subset of \( \mathbb{R}^d \), and let \( K \) be a family of cubes covering \( E \) which contain a cube \( D_x \) with center \( x \) for each \( x \in E \). Then there exist points \( \{x_k\} \) in \( E \) such that

(i) \( E \subset \bigcup_{x_k} D_{x_k}, \)

(ii) there exists a constant \( \sigma \) depending only on \( d \) such that

\[
\inf_{x_k} \sum_{x_k} I(D_{x_k}) \leq \sigma.
\]

For the proof, refer to Wheeden and Zygmund (1977), pp. 185-187.

**Lemma 2.** Let \( T > 0 \) be a given constant. Suppose that \( F \) has a density \( f \). Then for any given \( \varepsilon > 0 \), we can choose \( \beta_1 > 0 \) small enough and \( \beta_2 > 0 \) large enough such that the set

\[
E^* = \{ x \in S_{0,T} : \beta_1 \leq f(x) \leq \beta_2 \}
\]

satisfies \( P^*(S_{0,T} - E^*) < \varepsilon. \)

Note that for any Borel-measurable set \( E \subset E^* \), we have

\[
\beta_1 \leq f(x) \leq \beta_2, \quad \text{for almost all } x \in E \text{(with respect to } \lambda). \]

**Proof.** Set
\[ E_1 = \{ x \in S_{0,T} : \text{Sup}_{0<\rho<1} \frac{\lambda(S_{x,\rho})}{F(S_{x,\rho})} > 1/\beta_1 \}, \]
\[ E_2 = \{ x \in S_{0,T} : \text{Sup}_{0<\rho<1} \frac{\lambda(S_{x,\rho})}{F(S_{x,\rho})} > \beta_2 \}. \]

For any \( x \in E_1 \) there exists a cube \( S_{x,\rho} \) with \( \rho \in (0,1) \) such that \( \lambda(S_{x,\rho}) > F(S_{x,\rho})/\beta_1 \). By Lemma 1 there exist \( x_k \in E_1 \) and \( S_{x_k,\rho_k} \) such that \( \lambda(S_k) > F(S_k)/\beta_1 \), \( E_1 \subseteq \bigcup_k S_k \) and \( \int I(S_k) \leq \sigma \). Thus
\[
F^*(E_1) \leq F(\bigcup_k S_k) \leq \sum_k F(S_k) < \beta_1 \sum_k \lambda(S_k)
\]
\[
= \beta_1 \int_{\bigcup_k S_k} \int I(S_k) d\lambda \leq \beta_1 \sigma \lambda(\bigcup_k S_k) \leq \beta_1 \sigma \lambda(S_{0,2T}).
\]

Taking \( \beta_1 > 0 \) small enough, we have \( F^*(E_1) < \varepsilon/2 \). In the same way, we have \( \lambda^*(E_2) \leq F(R^d)/\beta_2 = \sigma/\beta_2 \). Taking \( \beta_2 \) large enough, we can make \( \lambda^*(E_2) \) small enough and, by the absolute continuity of \( F \) with respect to \( \lambda \), \( F^*(E_2) < \varepsilon/2 \).

The lemma is proved.

Fix \( \delta \in (0,1/2\sigma) \) and assume that \( h = h_\delta \in (0,1) \). Set
\[ G_n^* = \{ x \in S_{0,T} : F_n(S_{x,h}) < \delta F(S_{x,h}) \}. \tag{8} \]

Lemma 3. Suppose that \( F \) has a density \( f \), \( h = h_\delta \in (0,1) \) and \( \lim_{n \to \infty} \).

for any given \( \varepsilon > 0 \) we have
\[ P\{ F^*(G^*_n) > \varepsilon \} < e^{-C_1n}. \]

Proof. By Lemma 1, there exist \( x_k \in G_n^* \) and \( S_k = S_{x_k,h} \) such that
\[ G_n^* \subseteq \bigcup_k S_k \triangle G, \int I(S_k) \leq \sigma. \]
Partition \( R^r \) into sets with the form
\[
\Pi_j [(i_j-1)e_h, i_j e_h), \text{ where } i_1, \ldots, i_r = 0, \pm 1, \pm 2, \ldots \text{ and } e \text{ is a fixed constant to be chosen later. Call the partition } \phi, \text{ and write}
\]
\[ \phi' = \{ B \in S_{0,2T} \}, \]
\[ G = \bigcup_{B \in G, B \subseteq \phi'} S_{0,1} = \bigcap_{j=1}^r [-1+e, 1+e). \]
\[ C_{x_k} \triangleq S_k - \bigcup_{B \in \Phi, B \subseteq S_k} B \leq x_k + h(S_0, 1 - S_0, 1) \]
\[ \triangleq C_{x_k}^* \]

Then
\[ \lambda(C_{x_k}^*) = h^{r} \lambda(S_0, 1 - S_0, 1) = (2h)^{r}[1-(1-e)^{r}] \]
\[ \leq \text{rel}(S_k). \]

Since
\[ G - \tilde{G} \subseteq \bigcup_{k} (S_k - \bigcup_{B \in \Phi, B \subseteq S_k} B) \subseteq \bigcup_{k} C_{x_k}^* \]

we see that
\[ \lambda(G - \tilde{G}) \leq \sum_{k} \lambda(C_{x_k}^*) \leq \text{rel} \sum_{k} \lambda(S_k) \]
\[ \leq \text{rel} \int_{S_k} I(S_k) d\lambda \leq \text{rel} \lambda(U_k, S_k) \]
\[ \leq \text{rel} \lambda(S_{0, 2^n}). \]

Hence we can choose \( \epsilon \) small enough to render \( \lambda(G - \tilde{G}) \) small enough and \( F(G - \tilde{G}) < \epsilon/4 \). By (8) and the fact that \( \delta \alpha < 1/2 \), we get

\[ F_n(G) \leq F_n(\tilde{G}) \leq \sum_{k} F_n(S_k) < \delta \sum_{k} F(S_k) \]
\[ = \delta \int_{U_k} I(S_k) dF(\leq \delta \alpha F(U_k) = \delta \alpha F(G) \]
\[ < \frac{1}{4} F(G). \]

Therefore
\[ F(\tilde{G}) - F_n(\tilde{G}) > F(G) - \epsilon/4 - \frac{1}{4} F(G) = \frac{1}{4} F(G) - \epsilon/4, \]

and
\[ F_n^*(G_n^*) \geq \epsilon \quad \text{implies} \quad F(\tilde{G}) - F_n(\tilde{G}) > \epsilon/4. \]

For any \( H \in \Phi \), we write \( U_H = \bigcup_{B \subseteq H} B \). Then

\[ \{ F_n^*(G_n^*) \geq \epsilon \} \subseteq \bigcup_{H \in \Phi \setminus H} \{ F(U_H) - F_n(U_H) > \epsilon/4 \}. \]

Assume that \( \epsilon \in (0, 1) \). By Hoeffding's inequality,
$$\sup_A P(F(A) - F_n(A) > \varepsilon/4) \leq \sup_A \exp\{-n(\varepsilon/4)^2/[2F(A)+\varepsilon/4]\}$$

$$\leq \exp\{-n\varepsilon^2/[16(2+\varepsilon/4)]\} \leq \exp(-n\varepsilon^2/36).$$
Noticing that $\#\{H: H \subseteq \Phi\} \leq 2^0 h^r$ and $h^{-r} = o(n)$, we get

$$C_0 h^r \sup_A P(F(A) - F_n(A) > \varepsilon/4) \leq \exp(-C_1 n),$$
and the lemma is proved.

Lemma 4. Suppose that $\int_{\mathbb{R}^r} |g(x)| P_F(dx) < \infty$ for some $p > 0$, then

$$\lim_{n \to 0} \int_{S_{x, h}} |g(u) - g(x)| P_F(du)/F(S_{x, h}) = 0$$

for almost all $x$ (with respect to $F$).

Refer to Wheeden and Zygmund (1977), p. 191, example 20.
3. PROOF OF THE THEOREM

Assume that $|Y| \leq M$. Without loss of generality, we can assume that $\rho_0 = 1$ in (5)(iii). It is enough to prove that for each fixed $T > 0$,
\[
P\left\{ \int_{S_{0,T}} |Q_n(x) - Q(x)| F(dx) > \varepsilon \right\} < e^{-cn}. \tag{9}
\]

By Lemma 2, there exist $\beta_1 = \beta_1(\varepsilon)$, $\beta_2 = \beta_2(\varepsilon)$ and a compact set $E \subset E^*$ such that $F(S_{0,T} - E) < \varepsilon/8M$, where $E^*$ is defined by (7). Hence
\[
\int_{S_{0,T} - E} |Q_n(x) - Q(x)| F(dx) \leq 2MF(S_{0,T} - E) < \varepsilon/4.
\]

Fix $\delta \in (0,1/2)$. By Lemma 3, there exists a compact set $H_n$ such that
\[
H_n \subset \left\{ x \in S_{0,T} : F_n(S_{x,h}) \geq \delta F(S_{x,h}) \right\}, \tag{10}
\]
and
\[
P\left\{ F(S_{0,T} - H_n) \geq \varepsilon/(8M) \right\} < e^{-cn}. \tag{11}
\]

Hence
\[
P\left\{ \int_{S_{0,T} - H_n} |Q_n(x) - Q(x)| F(dx) > \varepsilon/4 \right\} < e^{-cn}.
\]

Therefore, we need only to prove that
\[
P\left\{ \int_{H_n \cap E} |Q_n(x) - Q(x)| F(dx) > \varepsilon/2 \right\} < e^{-cn}. \tag{12}
\]

For $x \in H_n \cap E$, by (5)(iii), (7) and (10), we have
\[
\sum_{j=1}^{n} K(\frac{X_j - x}{h}) \geq n\delta F_n(S_{x,h}) \geq n\delta F(S_{x,h})
\]
\[
\geq n\delta \beta_1 2^{r} h^{r},
\]
and $f(x) \leq \beta_2$. Write $C_3 = \beta_2/(a\delta \beta_1 2^{r})$, we see that
\[
\int_{H_n \cap E} |Q_n(x) - Q(x)| f(x) dx 
\]
\[
\leq C_3 (nh^{r})^{-1} \int_{H_n \cap E} \sum_{i=1}^{n} K(\frac{X_i - x}{h})(Y_i - Q(x)) dx.
\]
There exist finite positive constants $m, a_1, \ldots, a_m$ and disjoint regular cubes $A_1, \ldots, A_m$ such that $K^*(x) = \sum_{i=1}^{m} a_i I_{A_i}(x)$ satisfies
\[
\int_{\mathbb{R}^r} |K(x) - K^*(x)| \, dx < \varepsilon/(8C_3 M).
\]

Here a regular cube means a $r$-fold product of one-dimensional compact intervals.

Thus
\[
P((nh^r)^{-1} \int_{H \cap E} \left| \sum_{i=1}^{n} \frac{X_i - x}{h} \left( K\left(\frac{X_i - x}{h}\right) - K\left(\frac{X_i - x}{h}\right)\right) (Y_i - Q(x)) \right| \, dx
\leq 2C_3 M (nh^r)^{-1} \int \left| \sum_{i=1}^{n} \frac{X_i - x}{h} \left( K\left(\frac{X_i - x}{h}\right) - K\left(\frac{X_i - x}{h}\right)\right) \right| \, dx
\leq 2C_3 M \int |K(x) - K^*(x)| \, dx < \varepsilon/4.
\]

Take $\varepsilon_1 = \varepsilon/(4C_3)$. To prove (12), it is enough to prove that
\[
P((nh^r)^{-1} \int_{H \cap E} \left| \sum_{i=1}^{n} \frac{X_i - x}{h} \left( K\left(\frac{X_i - x}{h}\right) + K\left(\frac{X_i - x}{h}\right)\right) (Y_i - Q(x)) \right| \, dx
< \varepsilon_1.
\]

It is sufficient for any $\varepsilon_2 > 0$ and any regular cube $A$ to prove that
\[
P((nh^r)^{-1} \int_{H \cap E} \left| \sum_{i=1}^{n} I_{x+hA}(X_i) (Y_i - Q(X_i)) \right| \, dx \geq \varepsilon_2) \leq C_3 n
< \varepsilon_2.
\]

We proceed to prove (13). To this end, we construct the partition $\phi$ of $\mathbb{R}^r$ mentioned in the proof of Lemma 3. Assume that $A = \prod_{i=1}^{r} [x_i, x_i + a_i]$ and $\min a_i \geq 2 \varepsilon$. Set $A = \prod_{i=1}^{r} [x_i + e, x_i + a_i - e]$, $A_x = \bigcup_{B \in \phi} B \subset X + hA$.  

\[ C_x = x + hA - A_x + h(A - A) = C^*. \]

It is easy to see that we can make \( \lambda(A - A) \) arbitrarily small by choosing \( \epsilon \) small enough. We have

\[
(nh^r)^{-1} \int_{H^r \cap E} \left| \sum_{i=1}^{n} I_x + hA \right| (Y_i - Q(x_i)) \, dx
\]

\[
\leq (nh^r)^{-1} \int_{H^r \cap E} \left| \sum_{i=1}^{n} I_x \right| (Y_i - Q(x_i)) \, dx
\]

\[
+ 2Mh^{-r} \int_{H^r} f_n(x) \, dx
\]

\[
\leq (nh^r)^{-1} \int_{H^r \cap E} \sum_{i=1}^{n} I_x \left| (Y_i - Q(x_i)) \right| \, dx
\]

\[
+ 2M \lambda(A - A).
\]

Here we use the fact that \( \int \nu(x + hD) \, dx = h^r \lambda(D) \) for any \( r \)-dimensional probability measure \( \nu \) and any Borel set \( D \subset \mathbb{R}^r \). We can choose \( \epsilon \) such that \( 2M \lambda(A - A) < \epsilon_2 / 2 \). Note that for \( B \in \Phi, \lambda \{ x \in B \subseteq x + hA \} \leq C_4 h^r \), and \( U_{x \in H^r \cap E} \{ x + hA \} \subseteq S_{0,2T} \) for small \( h \).

Hence, for large \( n \), we have

\[
(15) \leq 4^{n-1} \int_{B \in \Phi} \left| \sum_{i=1}^{n} I_B(x_i)(Y_i - Q(x_i)) \right| + \epsilon_2 / 2.
\]

Set \( \epsilon_3 = \epsilon_2 / (4C_4) \). To prove (13), we need only to prove that

\[
P\left( \sum_{B \in \Phi} \left| \sum_{i=1}^{n} I_B(x_i)(Y_i - Q(x_i)) \right| \geq \frac{n \epsilon_3}{C_5 n} \right) < e^{-C_5 n}, \tag{16}
\]

\[
P\left( \sum_{B \in \Phi} \left| \sum_{i=1}^{n} I_B(x_i)Q(x_i) \right| - n \int_{B} Q(x) \, dF \geq n \epsilon_3 \right) < e^{-C_5 n}. \tag{17}
\]

Let \( N \) be a Poisson random variable with mean value \( n \), which is independent of \( (X_1,Y_1),(X_2,Y_2), \ldots \). In the sequel, we use \( \sum^\cdot \) for \( \sum_{B \in \Phi} \). Notice
That B's are disjoint, we see that for $t_B \in (-\infty, \infty)$,

$$E\left\{ \exp\left( \sum_{i=1}^{N} t_B \sum_{i}^{N} I_B(X_i)Y_i \right) \right\}$$

$$= \sum_{t=0}^{\infty} \frac{n^t}{t!} E\left\{ \exp\left( \sum_{i}^{N} t_B I_B(X_i)Y_i \right) \right\}^t$$

$$= \exp\{n\sum_{i}^{N} E[I_B(X_i)(e^{t_B Y_i - 1})]\},$$

So that, $\{ \sum_{i}^{N} I_B(X_i)Y_i - n \int_B Q(x) dF \}$ is a group of mutually independent variables. Set

$$Z(B,N) = \sum_{i}^{N} I_B(X_i)Y_i - n \int_B Q(x) dF.$$  

for $t > 0$, notice that $e^{-t} > e^{-t} + t$, we have

$$P\left\{ \sum_{i}^{N} Z(B,N) \geq \frac{1}{2} n \epsilon_3 \right\} \leq \exp(-\frac{1}{2} n \epsilon_3)E\left\{ \exp\left( t \sum_{i}^{N} Z(B,N) \right) \right\}$$

$$= \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} E\left\{ \exp\left( t \sum_{B \in \Phi} Z(B,N) \right) \right\}$$

$$\leq \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \left[ E\left\{ \exp(tZ(B,N) + E\{\exp(-tZ(B,N))\} \right\} \right.$$  

$$= \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \left[ \exp\{nE[I_B(X_1)(e^{t_B Y_1 - 1})] \right.$$  

$$+ \exp\{nE[I_B(X_1)(e^{-t_B Y_1 + t_B Y_1 - 1})] \right.$$  

$$\leq \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \left[ 2\exp\{nE[I_B(X_1)(e^{t_B Y_1 - 1})] \right.$$  

$$\leq \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \exp\left\{ n \sum_{i}^{N} E[I_B(X_i)(e^{t_B Y_i - 1})] \right\}$$

$$\leq \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \exp\left\{ n \sum_{i}^{N} E[I_B(X_i)(e^{t_B Y_i - 1})] \right\}$$

$$\leq \exp(-\frac{1}{2} n \epsilon_3) \prod_{B \in \Phi} \exp\left\{ n \sum_{i}^{N} E[I_B(X_i)(e^{t_B Y_i - 1})] \right\}$$

Take $t \in (0,1/M)$, we see that $e^{t_B Y_i - 1} \leq e^{t_B Y_i}$. Hence, we can take $t > 0$ such that

$$P\left\{ \sum_{i}^{N} I_B(X_i)Y_i - n \int_B Q(x) dF \geq \frac{1}{2} n \epsilon_3 \right\} \leq \exp(-C_0 n). \quad (18)$$
Write $\Delta = (X_1, X_2, \ldots)$. By Jensen's inequality, for $t > 0$ we have

$$P\left\{ \left\| \sum_{i=1}^{N} I_B(X_i)Q(X_i) - n \int_B Q(x) dP \right\| \geq \frac{1}{2} \|n\|_3 \right\}$$

$$= P\left\{ \left\| E(Z(B,N)|\Delta) \right\| \geq \frac{1}{2} \|n\|_3 \right\}$$

$$\leq P\{E\left( \sum |Z(B,N)| |\Delta) \geq \frac{1}{2} \|n\|_3 \right\}$$

$$\leq \exp(-\frac{\|n\|_3}{2})E\left[ \exp(tE\left( \sum |Z(B,N)| |\Delta) \right) \right]$$

$$= \exp(-\frac{\|n\|_3}{2})E\left[ \exp(t \sum |Z(B,N)| \right]$$

By (18) and (19), we can take $t > 0$ such that

$$P\left\{ \left\| \sum_{i=1}^{N} I_B(X_i)Q(X_i) - n \int_B Q(x) dP \right\| \geq \frac{1}{2} \|n\|_3 \right\} < e^{-C_6 n}. \quad (20)$$

Note that

$$\left| \sum_{i=1}^{N} I_B(X_i)Y_i - \sum_{i=1}^{N} I_B(X_i)Y_i \right| \leq M|N-n|,$$

$$\left| \sum_{i=1}^{N} I_B(X_i)Q(X_i) - \sum_{i=1}^{N} I_B(X_i)Q(X_i) \right| \leq M|N-n|,$$

we have

$$P\left\{ \left\| \sum_{i=1}^{N} I_B(X_i)Y_i - \sum_{i=1}^{N} I_B(X_i)Y_i \right\| \geq \frac{1}{2} \|n\|_3 \right\} \quad (21)$$

$$\leq P\{|N-n| \geq n\|_3/(2M)\} < e^{-C_7 n},$$

$$P\left\{ \left\| \sum_{i=1}^{N} I_B(X_i)Q(X_i) - \sum_{i=1}^{N} I_B(X_i)Q(X_i) \right\| \geq \frac{1}{2} \|n\|_3 \right\} \quad (22)$$

$$< e^{-C_7 n}.$$
From (19) - (22), (16) and (17) follows, and (13) is proved. It remains to prove (14). To this end, we need only to prove

\[ P\left( \bigcup_{i=1}^{n} Z(X_i) \geq \frac{2}{3} \varepsilon \right) < e^{-C_3 n}, \tag{23} \]

where

\[ Z(u) = h^{-1} \int_{\mathbb{R}} I_{x+hA}(u) \left( Q(u) - Q(x) \right) dx. \tag{24} \]

Hence, we have

\[ \left| \sum_{i=1}^{N} Z(X_i) - \sum_{i=1}^{n} Z(X_i) \right| \leq C_8 |N-n|, \]

and

\[ P\left( \sum_{i=1}^{n} Z(X_i) - \sum_{i=1}^{N} Z(X_i) \geq \frac{1}{4} \varepsilon \right) \leq P\left( |N-n| \geq \frac{1}{4} \varepsilon / (2C_8) \right) < e^{-C_9 n} \tag{25} \]

For \( t > 0 \), we have

\[ P\left( \sum_{i=1}^{N} Z(X_i) > \frac{1}{4} \varepsilon \right) \leq \exp\left( -\frac{1}{4} \varepsilon \right) E\left\{ \exp(t \sum_{i=1}^{N} Z(X_i)) \right\} \]

\[ = \exp\left( -\frac{1}{4} \varepsilon t \right) + \int_{0}^{\infty} \left( e^{t Z(u)} - 1 \right) F(du). \tag{26} \]

Take \( t \in (0, 1/C_8) \). By \( 0 \leq t Z(u) \leq 1 \) we get

\[ \int_{0}^{\infty} \left( e^{t Z(u)} - 1 \right) F(du) \leq 2nt \int Z(u) F(du) \tag{27} \]

Take \( \rho > 0 \) so large that \( A \geq S_{0, \rho} \). Then, by Lemma 4, we have

\[ Z(u) = h^{-1} \int_{u-hA} \left| m(x) - m(u) \right| dx \]

\[ \leq \lambda(S_{0, \rho}) \int_{S_{u, h\rho}} \left| m(x) - m(u) \right| dx / \lambda(S_{u, h\rho}) \]
\[ \rightarrow 0 \text{ as } h \rightarrow 0, \text{ for almost all } x(\lambda). \]

In view of (24), from the dominated convergence theorem, we see that
\[ \lim_{h \to 0} \int Z(u) F(du) = 0. \tag{28} \]

By (26) - (28), we can take \( t > 0 \) sufficiently small such that
\[
P\left( \bigcup_{i=1}^{N} \{ Z(X_i) > \frac{1}{2} n \varepsilon_2 \} \right) \leq \exp(-\frac{1}{2} n t \varepsilon_2 + o(nt))
\]
\[< \exp(-C_10^n). \]

From (25) and (29), we obtain (23), and (14) follows. Up to now, the theorem is proved.
REFERENCES


