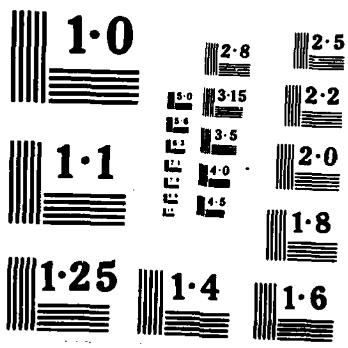
AD-A170 059 ALMOST SURE L(1)-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMA..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS X R CHEN ET AL. MAR 86 1/1

UNCLASSIFIED TR-86-07 AFOSR-TR-86-0394 F49620-85-C-0008 F/G 12/1 NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6

(2)

ALMOST SURE L_1-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMATES

X. R. Chen

and

L. C. Zhao[1]

University of Science and Technology of China
Beijing, China

Center for Multivariate Analysis

University of Pittsburgh

AD-A170059

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR- 86-0394	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ALMOST SURE L_1-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMATES		5. TYPE OF REPORT & PERIOD COVERED Technical - March 1986 6. PERFORMING ORG. REPORT NUMBER 86-07
7. AUTHOR(s) X.R. Chen and L.C. Zhao		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis 515 Thackeray Hall University of Pittsburgh, Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force Base, DC 20332		12. REPORT DATE March 1986 13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Data-based, density estimator.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The main result of this paper is summarized in Theorem 1, which states that when certain conditions of a general nature are satisfied, the data-based histogram density estimator is strongly consistent in the sense that the mean absolute deviation of the estimator and the density function converges to zero almost surely for any density function, as the sample size increases to infinity.

ALMOST SURE L_1-NORM CONVERGENCE FOR DATA-BASED

HISTOGRAM DENSITY ESTIMATES

X. R. Chen

and

L. C. Zhao[1]

University of Science and Technology of China
Beijing, China

March 1986

Technical Report No. 86-07

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

[1] Part of the work of this author was sponsored by the Air Force Office if Scientific Research (AFOSR), under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

ALMOST SURE L_1-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMATES

X. R. Chen

and

L. C. Zhao[1]

University of Science and Technology of China
Beijing, China

ABSTRACT

The main result of this paper is summarized in Theorem 1, which states that when certain conditions of a general nature are satisfied, the data-based histogram density estimator is strongly consistent in the sense that the mean absolute deviation of the estimator and the density function converges to zero almost surely for any density function, as the sample size increases to infinity.

AMS 1980 Subject Classification. Primary 62F10: Secondary 62H99

Key Words: Data-based, density estimator.

[1] Part of the work of this author was sponsored by the Air Force Office of Scientific Research (AFOSR), under Contract F49620-85-C-0008. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

1. INTRODUCTION AND MAIN RESULTS

Suppose that X is a one-dimensional random variables with distribution F and density f. Let $X_1,\ldots,X_n$ be i.i.d. observations of X. It is desired to estimate f(x) by these samples. For this purpose we introduce a partition of $R^1 = (-\infty,\infty)$:

$$R^1 = \bigcup_{i=1}^{\infty} I(i;X_1,\ldots,X_n), \tag{1}$$

where $I(i;X_1,\ldots,X_n)$, $i=1,2,\ldots$ are intervals with lengths greater than zero, and $I(i;X_1,\ldots,X_n) \cap I(j;X_1,\ldots,X_n) = \emptyset$ for $i \neq j$. Write

$K_n = K_n(X_1,\ldots,X_n) = \{I(i;X_1,\ldots,X_n): i = 1,2,\ldots\}$,

$I_n(x)$ = The interval $I(i;X_1,\ldots,X_n)$ containing x,

$p_n(x) = \#(\{i: 1 \le i \le n,\ X_i \in I_n(x)\})$, where $\#(A)$ denotes the number of elements belonging to A,

and define an estimate of f(x) as follows:

$$f_n(x) = f_n(x;X_1,\ldots,X_n) = p_n(x)/(n|I_n(x)|). \tag{2}$$

Here and in the following we write $|A|$ for the Lebesgue measure of the set $A \subset R^1$. $f_n(x)$ is the so-called data-based histogram estimate based on the partition K_n. "Data-based" means that K_n depends on the sample $X_1,\ldots,X_n$, while in the ordinary histogram estimate, the partition is predetermined before the samples were drawn.

Write $\Delta_n = \Delta_n(X_1,\ldots,X_n)$ for the L_1-norm of f_n:

$$\Delta_n = \int_{-\infty}^{\infty} |f_n(x) - f(x)|\,dx = \int |f_n - f|\,dx. \tag{3}$$

A number of papers appeared dealing with the weak (i.e. in probability) convergence of Δ_n to zero. Among these we mention the recent paper [2] by J. Chen and H. Rubin, in which they prove that $\Delta_n \xrightarrow{P} 0$ under quite general conditions imposed on K_n. In the present article we deal with the problem of a.s. convergence. Specifically speaking, we prove the following theorem.

Theorem 1. Suppose that K_n satisfies the following two conditions:

1. $\lim_{n\to\infty} |I_n(x)| = 0$, a.s. for $x \in R^1$, a.e.L. (4)

2. Denote by C_{nt} the number of intervals in K_n having at least one common point with $[-t,t]$, we have for any fixed $t > 0$,

$$C_{nt} = o(n/\log n), \text{ a.s.} \tag{5}$$

Then, for the estimator f_n defined by (2) it is true that

$$\lim_{n\to\infty} \Delta_n = 0, \text{ a.s.} \tag{6}$$

where Δ_n is the L_1-norm of f_n defined by (3).

Essentially speaking, Chen and Rubin proved that $\Delta_n \xrightarrow{P} 0$ under our condition 1 and $C_{nt} = o_p(n)$, and another condition with a more complicated nature. In order to prove (6) we pay a price that the condition $C_{nt} = o_p(n)$ is strengthend to (5). It is easy to show by an example that (5) cannot be replaced by $C_{nt} = o_p(n/\log n)$. Judging from the known results in density estimation, it seems doubtful that the condition (5) can be substaintially improved.

We also remark that Y. S. Chow and others [3] gave a result concerning the truth of (6), where K_n has a form $K_n = \{(i/\lambda_n,(i+1)/\lambda_n, i = 0,\pm 1,\pm 2,\ldots\}$, $\lambda_n = \lambda_n(X_1,\ldots,X_n)$ is determined by the sample in a way described in [3], but the condition imposed on f is rather stringent.

Chen and Rubin also considered in [2] the case that X is multidimensional, again for weak convergence. For the problem of strong convergence, Wang and Chen [8] obtained some results which are of more complicated form. Recently, we improved these results and obtain a simple one which includes the above one-dimensional result as a special case.

In this paper we give a proof in detail for one-dimensional case only. In Section 2 we introduce a lemma which is needed in the sequel. The proof of the main result is given in Section 2. In Section 3 we discuss a generalization to the multidimensional case.

2. A LEMMA

The proof of Theorem 1 depends on a lemma which is proved in this section. The lemma has independent interest, and may be useful in some related problems.

Lemma 1. Suppose that $X_1, X_2, \ldots$ is a sequence of independent one-dimensional random variables, and the distribution function F of X_1 is continuous everywhere on R^1. Denote by F_n the empirical distribution function of $\{X_1, \ldots, X_n\}$. Then there exist absolute constants $C_i > 0$, $i = 0,1,\ldots,4$, such that for any $\varepsilon > 0$ we have

$$P\{\sup_{A\in F}|F_n(A) - F(A)| \geq \varepsilon\}$$

$$\leq C_1\left(\frac{\sqrt{b}}{\varepsilon\sqrt{n}} + \frac{1}{b}\right)\exp(-C_2 n\varepsilon^2/b) + C_3\exp(-C_4 n\varepsilon), \tag{7}$$

where F is a set consisting of some intervals $A \subset R^1$ with

$$\sup_{A\in F} F(A) \leq b \leq 1$$

and $n/\log n$ is greater than C_o/ε.

Proof. By a result of Komlós, Major and Tusnády [6], we can find a suitable probability space in which we can define an i.i.d. sequence (which for convenience will be denoted again by $X_1, \ldots, X_n$) whose common distribution is F, and a Brownian bridge $B_n(t)$, such that

$$P\{\sup_x|n(F_n(x) - F(x)) - \sqrt{n}B_n(F(x))| > C\log n + y\}$$

$$< \tilde{C}\exp(-\lambda y), \tag{8}$$

where C, $\tilde{C}$ and λ are positive absolute constants. Put $B_n(t) = W_n(t) - tW_n(1)$, where W_n is a Brownian motion process, and write

$$n(F_n(x) - F(x)) = \sqrt{n}B_n(F(x)) + e_n(x).$$

Then for $A = [a_1, a_2) \in \mathcal{F}$ we have

$$F_n(A) - F(A) = n^{-1/2}[W_n(F(a_2)) - W_n(F(a_1))]$$
$$- n^{-1/2} F(A)W_n(1) + n^{-1}[e_n(a_2) - e_n(a_1)]. \tag{9}$$

From (8), when $n\varepsilon/6 > 2C \log n$ we have

$$P\{\sup_{A\in\mathcal{F}} \frac{1}{n}|e_n(a_2) - e_n(a_1)| \geq \varepsilon/3\}$$
$$\leq P\{\sup_x |e_n(x)| \geq n\varepsilon/6\} \leq \tilde{C} \exp(-\lambda n\varepsilon/12). \tag{10}$$

Further,

$$P(\sup_{A\in\mathcal{F}} n^{-1/2}F(A)|W_n(1)| \geq \varepsilon/3)$$
$$\leq P(|W_n(1)| \geq \varepsilon\sqrt{n}/(3b))$$
$$\leq \frac{6b}{\sqrt{2\pi}\,\varepsilon\,\sqrt{n}} \exp(-\frac{n\varepsilon^2}{18b^2}) \leq \frac{6\sqrt{b}}{\sqrt{2\pi}\,\varepsilon\,\sqrt{n}} \exp(-\frac{n\varepsilon^2}{18b}). \tag{11}$$

By Lemma 1,2.1 in [4] and $\sup_{A\in\mathcal{F}} F(A) \leq b$,

$$P\{\sup_{A\in\mathcal{F}} n^{-1/2}|W_n(F(a_2)) - W_n(F(a_1))| \geq \varepsilon/3\}$$
$$\leq P\{\sup_{x_2-x_1\leq b,\ 0\leq x_1<x_2\leq 1} |W_n(x_2) - W_n(x_1)| \geq \varepsilon\sqrt{n}/3\}$$
$$\leq P\{\sup_{0\leq s\leq 2-b}\ \sup_{0<t\leq b} |W_n(s+t) - W_n(s)| \geq (\varepsilon\sqrt{n}/3\sqrt{b})\sqrt{b}\}$$
$$\leq \frac{C}{b} \exp(-n\varepsilon^2/18b), \tag{12}$$

where C is a constant. The lemma now follows from (9)-(12).

Remark. The lemma is an essential improvement of a similar result given by Devroye and Wagner in [5] for the special case that F consists of one-dimensional intervals. In their result, b is given by

$$b = \sup_{A \in F'} F(A), \text{ and } F' = \{A = [a, a+h) : a \in R^1,\ h = 2 \sup_{A \in F} |A|\}.$$

3. PROOF OF THE THEOREM 1

First we note that it is enough to show that

$$\lim_{n\to\infty} \int_{-t}^{t} |f - f_n| dx = 0, \text{ a.s. for each } t > 0. \tag{13}$$

For if (13) has been proved, denote by $E_t \subset R^\infty$ the set on which (13) is not true. Then $P(E_t) = 0$. Put $E = \bigcup_{t=1}^{\infty} E_t$. By an easy argument it is seen that $\lim_{n\to\infty} \int |f - f_n| dx \to 0$ on $R^\infty - E$.

Next define

$$Q_n(x) = \int_{I_n(x)} f(u)du/|I_n(x)|.$$

In order to prove (13) it is enough to prove that

$$\lim_{n\to\infty} \int_{-t}^{t} |f - Q_n| dx = 0, \text{ a.s. for each } t > 0, \tag{14}$$

$$\lim_{n\to\infty} \int_{-t}^{t} |f_n - Q_n| dx = 0, \text{ a.s. for each } t > 0. \tag{15}$$

By assumption 1 of Theorem 1 it is easily seen that there exists a set $A \subset R^\infty$ such that $P(A) = 0$ and for $(X_1, X_2, \ldots) \bar{\in} A$ we have $\lim_{n\to\infty} |I_n(x)| = 0$ for $x \in R^1$, a.e.L, and in turn it follows that $\lim_{n\to\infty} Q_n(x) = f(x)$ for x, a.e.L. Since $Q_n(x)$ is a density function when $(X_1, X_2, \ldots)$ is fixed, by a well-known theorem due to Schéffe, it follows that $\lim_{n\to\infty} \int |f - Q_n| dx = 0$ for each fixed $(X_1, X_2, \ldots) \bar{\in} A$. Thus (14) is proved.

Now we proceed to prove (15). Denote by $\Delta_{n1}, \ldots, \Delta_{nm_n}$ those intervals belonging to K_n and having common points with $[-t, t]$. Put

$$F = \bigcup_{(x_1,\ldots,x_n) \in R^n} \{I(i;x_1,\ldots,x_n): i = 1,2,\ldots\}$$

and denote by $F(b)(0 \le b \le 1)$ the set of intervals belonging to F and satisfying $F(I) \le b$. Then $\bigcup_{i=1}^{m_n} \Delta_{ni} \supset [-t,t]$, and by Assumption 2 of Theorem 1 we have

$$m_n = o(n/\log n), \text{ a.s.} \tag{16}$$

Denote by #(A) the number of elements belonging to A, and

$$q_{ni} = \#(\{j: 1 \le j \le n, X_j \in \Delta_{ni}\}),$$

$$Z_{ni} = \int_{\Delta_{ni}} f dx. \tag{17}$$

Then we have

$$\int_{-t}^{t} |f_n - Q_n| dx \le \sum_{i=1}^{m_n} \int_{\Delta_{ni}} |f_n - Q_n| dx$$

$$= \frac{1}{n} \sum_{i=1}^{m_n} |q_{ni} - nZ_{ni}|. \tag{18}$$

Given $\epsilon > 0$, choose $M > \max\{64, (C_0/\epsilon)^{3/2}, (\frac{4}{C_4\epsilon})^{3/2}, (\frac{4}{C_2\epsilon^2})^3\}$, where C_0, C_2, C_4 are the constants mentioned in Lemma 1. Divide $\{1,2,\ldots,m_n\}$ into a number of nonintersecting sets $J_0, J_1, \ldots$ in the following way:

$$J_0 = \{i: 1 \le i \le m_n, Z_{ni} \le M \log n/n\},$$

$$J_r = \{i: 1 \le i \le m_n, \frac{M + r - 1}{n} \log n < Z_{ni} \le \frac{M + r}{n} \log n\}, \ r = 1,2,\ldots .$$

Define $a_i = \#(J_i)$, $i = 0,1,2,\ldots$. Since $\int f dx = 1$, we have $\sum_{i=1}^{\infty} a_i (M + i - 1) \log n/n \le 1$. From this and $M > 1$ we have

$$\sum_{i=1}^{\infty} a_i (M + 1)\log n/n \le 2. \tag{19}$$

By $(M + r - 1)\log n/n < 1$, we can restrict ourselves to the cases where $r < n/\log n$. Thus

$$\frac{1}{n}\sum_{i=1}^{m_n} |q_{ni} - nZ_{ni}| = \sum_{r=o}^{n/\log n} \sum_{i \in J_r} \left|\frac{q_{ni}}{n} - Z_{ni}\right|$$

$$\le \frac{1}{n}\sum_{r=o}^{n/\log n} a_i \sup_{I \in F_r} |F_n(I) - F(I)|, \tag{20}$$

where

$$F_r = F((M + r)\log n/n), \; r \ge 0.$$

Write

$$B_o = \{\sup_{I \in F_o} |F_n(I) - F(I)| \ge \varepsilon M \log n/n \,,$$

$$B_r = \{\sup_{I \in F_r} |F_n(I) - F(I)| \ge \varepsilon M^{-1/3}(M + r) \log n/n\}, \; r > 1,$$

and

$B = \bigcup_{r=o}^{n/\log n} B_r$. Using Lemma 1, we get

$$P(B_o) \le C_1\left(\frac{1}{\varepsilon\sqrt{M \log n}} + \frac{n}{M \log n}\right)\exp(-C_2\varepsilon^2 M \log n)$$

$$+ C_3 \exp(- C_4 \varepsilon M \log n), \tag{21}$$

and

$$P(B_r) \le C_1\left(\frac{1}{\varepsilon} M^{-1/6}(\log n)^{-1/2} + n/M \log n\right)\exp(-C_2\varepsilon^2 M^{1/3} \log n)$$

$$+ C_3 \exp(- C_4 M^{2/3}\varepsilon \log n), \tag{22}$$

which implies

$$P(B) \le C/n^2,$$

where C does not depend on n.

But when the event B does not happen, we have

$$\frac{1}{n}\sum_{i=1}^{m_n}|q_{ni} - nZ_{ni}| \le \{a_o\frac{M \log n}{n} + \sum_{r\geq 1} a_r\frac{(M+r)\log n}{n} M^{-1/3}\}\varepsilon.$$

By (19) and $M > 64$ it follows that

$$P(\frac{1}{n}\sum_{i=1}^{m_n}|q_{ni} - nZ_{ni}| \ge a_o\varepsilon M \log n/n + \varepsilon/2) \le P(B) \le C/n^2.$$

Hence by Borel-Cantelli's Lemma we have

$$P(\frac{1}{n}\sum_{i=1}^{m_n}|Q_{ni} - nZ_{ni}| \ge a_o\varepsilon M \log n/n + \varepsilon/2,\ i.o.) = 0.$$

Therefore, with probability one, we can assert that

$$\frac{1}{n}\sum_{i=1}^{m_n}|Q_{ni} - nZ_{ni}| \le a_o\varepsilon M \log n/n + \varepsilon/2 \tag{23}$$

for n sufficiently large. But by (16)

$$a_o \le m_n = o(n/\log n),\ \text{a.s.}$$

From this and (23), it follows that with probability one, we have

$$\frac{1}{n}\sum_{i=1}^{m_n}|Q_{ni} - nZ_{ni}| \le \varepsilon \tag{24}$$

for n sufficiently large. Since $\varepsilon > 0$ is arbitrarily given, (15) follows from (18) and (24), and Theorem 1 is proved.

4. MULTIDIMENSIONAL CASE

We now consider multidimensional extension of the result in Section 3. Let us assume that X, X_1, ..., X_n are i.i.d. d-dimensional random vectors, and replace density, partition, interval in R^1 and so on by the analogues in R^d. In particular, by an interval in R^d we mean a set in R^d having the form $\prod_{i=1}^{d} A_i$, where A_i's are all one-dimensional intervals. Now C_{nt} in (5) is defined as the number of intervals in K_n having at least one common point with the interval $V_t = \{(x_1,\ldots,x_d) : |x_i| \leq t,\ i = 1,\ldots,d\}$. Also, condition (4) is replaced by the following:

$$\text{(I)} \quad \lim_{n\to\infty} D(I_n(x)) = 0, \text{ a.s. for } x \in R^d, \text{ a.e.L}, \tag{25}$$

where $D(I)$ denotes the diameter of set $I \subset R^d$.

For the case where $d > 1$, Chen and Rubin [2] proved that $\Delta_n \xrightarrow{P} 0$ under (25), $C_{nt} = o_p(\sqrt{n})$ for any $t > 0$, and another condition with a more complicated nature. Wang and Chen [8] studied the problem of strong convergence of Δ_n. They proved that $\lim_{n\to\infty} \Delta_n = 0$, a.s. if (25) holds and one among the following sets of conditions is satisfied:

II'. $C_{nt} = o(\sqrt{n/\log n})$, a.s. for any $t > 0$,

II''. f is bounded on any bounded subset of R^d,

$C_{nt} = o(n/\log n)$, a.s. for any $t > 0$,

$\limsup_{n\to\infty} \sigma_n(t) < \infty$, a.s. for any $t > 0$,

where

$$\sigma_n(t) = \sup\{D(I) : I \in K_n \text{ and } I \cap V_t \neq \emptyset\}.$$

II'''. For $a > 0$ large enough, the set $\{x : f(x) < a\}$ differs only by a null Lebesgue measurable set from an open set,

$C_{nt} = o(n/\log n)$, a.s. for any $t > 0$,

$\lim_{n\to\infty} \sigma_n(t) = 0$, a.s. for any $t > 0$.

Recently, we find an inequality by which we obtain the following.

Theorem 2. Suppose that K_n satisfies the condition (I) and

(II) $C_{nt} = o(n/\log n)$, a.s. for any $t > 0$.

Then (6) is true for the d-dimensional case.

Since the proof is parallel to that of Theorem 1, we only introduce related inequality.

To this end, let $x_1,\ldots,x_r$ be r points in R^d, and A be a class of Borel sets in R^d. Denote by $\Delta^A(x_1,\ldots,x_r)$ the number of distinct sets in $\{\{x_1,\ldots,x_r\} \cap A, A \in A\}$. Define

$$m^A(r) = \max_{x_1,\ldots,x_r \in R^d} \Delta^A(x_1,\ldots,x_r).$$

Vapnik and Chervonenkis' [7] showed that either $m^A(r) = 2^r$ for any positive integer r or $m^A(r) \leq r^s + 1$, where s is the smallest k such that $m^A(k) \neq 2^k$. A class of sets A for which the latter case holds will be called a V-C class with index s.

Suppose that μ is a probability measure on R^d. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random vectors with common distribution μ, and μ_n be the empirical distribution of $X_1,\ldots,X_n$. Denote a "distance" between μ_n and μ by

$$D_n(A,\mu) = \sup_{A \in A} |\mu_n(A) - \mu(A)|.$$

Here we assume that $D_n(A,\mu)$, $\sup_{A \in A} |\mu_n(A) - \mu_{2n}(A)|$ and $\sup_{A \in A} \mu_n(A)$ are all random variables. We have the following.

Lemma 2. Let A be a V-C class with index s such that

$$\sup_{A \in \mathcal{A}} \mu(A) \leq \delta \leq 1/8. \tag{26}$$

Then for any $\varepsilon > 0$ we have

$$\begin{aligned} P\{D_n(\mathcal{A},\mu) > \varepsilon\} &\leq 5(2n)^s \exp(-n\varepsilon^2/(91\delta + 4\varepsilon)) \\ &+ 7(2n)^s \exp(-\delta n/68) \\ &+ 2^{2+s} n^{1+2s} \exp(-\delta n/8), \end{aligned} \tag{27}$$

provided $n \geq \max(12\sigma/\varepsilon^2, 68(1 + s)(\log 2)/\delta)$.

Proof. See [9].

In the present case, we should take $\mathcal{A}$ as some interval class in R^d which is a V-C class obviously. Also, there is no problem with measurability mentioned above. As an alternative lemma, we can also use the corollary 2.9 in [1].

Acknowledgement. The authors are grateful to the referee for his helpful suggestions.

REFERENCES

[1] Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. Ann. Prob. 12, 1041-1067.

[2] Chen, J. and Rubin, H. (1984). On the consistency of the data-based histogram density estimator. Technical Report 84-11, Purdue University.

[3] Chow, Y. S., Geman, S. and Wu, L. D. (1983). Consistent cross-validated density estimates. Ann. Statist. 11, 25-38.

[4] Csörgő, M. and Révész, P. (1981). Strong Approximations in Probability and Statistics. Akadémiai Kiadó, Budapest.

[5] Devroye, L. P. and Wagner, T. J. (1980). The strong uniform consistency of kernel density estimates. In Multivariate Analysis V (edited by P. Krishnaiah), pp.59-77: North Holland.

[6] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF, I. Z. Wahrscheinlichkeistheorie verw.Gebiete. 32, 111-131.

[7] Vapnik, V. N. and Chervonenkis, A. Ya. (1971). On uniform convergence of the frequencies of events to their probabilities. Theor. Probab. Appl. 16, 264-280.

[8] Wang, Shou-ren and Chen, Xi-ru (1985). L_1-norm strong consistency for data-based histogram estimates of multidimensional density. (Personal communication.)

[9] Zhao, Lin-Cheng (1985). An inequality concerning the deviation between theoretical and empirical distributions. Technical Report 85-30. Center for Multivariate Analysis, Univ. of Pittsburgh.

END
DTIC
8-86