ALMOST SURE L(\Gamma)-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ES (U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS L C ZHAO ET AL AUG 87 TR-87-30 UNCLASSIFIED AFOSR-TR-87-1843 F49620-85-C-0000 F/G 12/3
ALMOST SURE $L_r$-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMATES

L. C. Zhao, P. R. Krishnaiah, X. R. Chen

Center for Multivariate Analysis
University of Pittsburgh
Pittsburgh PA 15260

August 1987

Technical Report Number 87-30

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 hereon.

DISTRIBUTION STATEMENT A

Suitable for public release; by author's discretion.
ALMOST SURE $L^r$-NORM CONVERGENCE FOR DATA-BASED HISTOGRAM DENSITY ESTIMATES

L. C. Zhao, P. R. Krishnaiah, X. R. Chen

Center for Multivariate Analysis
University of Pittsburgh
Pittsburgh PA 15260

ABSTRACT

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn from a $d$-dimensional distribution with density $f$. Partition the space $\mathbb{R}^d$ into a union of disjoint intervals

$$I_\ell = I(\ell, X_1, \ldots, X_n) = \left\{ x = (x(1), \ldots, x(d)) \mid -\infty < a_{\ell i} < x(i) < b_{\ell i} < \infty, \ i = 1, \ldots, d \right\}. $$

Define the data-based histogram estimate of $f(x)$ based on this partition as

$$f_n(x) = \text{The number of } X_1, \ldots, X_n \text{ falling into } I_\ell$$

$$\div n \text{ times the volume of } I_\ell, \ \text{for } x \in I_\ell, \ \ell = 1, 2, \ldots$$

For given constant $r > 1$ we obtain the sufficient condition for

$$\lim_{n \to \infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)|^r \, dx = 0.$$  

The results give substantial improvements upon the existing results.


Key words and phrases: Data-based, density estimator, empirical distribution, histogram.
Almost Sure Lr-Norm Convergence for Data-Based Histogram Density Estimates

L. C. Zhao, P. R. Krishnaiah, and X. R. Chen

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

Air Force Office of Scientific Research
Department of the Air Force
Bolling Air Force Base, DC 20332

Approved for public release; distribution unlimited

key words and phrases: Data-based, density estimator, empirical distribution, histogram.
Let \( X_1, \ldots, X_n \) be i.i.d. samples drawn from a \( d \)-dimensional distribution with density \( f \). Partition the space \( \mathbb{R}^d \) into a union of disjoint intervals \( \{ I_\ell = I(x_1, \ldots, x_n) \} \) with the form \( I_\ell = \{ x = (x(1), \ldots, x(d)) : -\infty < a_{\ell_1} < x(1) < b_{\ell_1} < \ldots < b_{\ell_d} < \infty, \ell = 1, \ldots, d \} \). Define the data-based histogram estimate of \( f(x) \) based on this partition as

\[
f_n(x) = \frac{L}{n} \text{ for } x \in I_\ell, \quad \ell = 1, \ldots,
\]

where \( L \) is the number of \( X_1, \ldots, X_n \) falling into \( I_\ell \) times the volume of \( I_\ell \), for \( x \in I_\ell, \ell = 1, 2, \ldots \).

For given constant \( r > 1 \) we obtain the sufficient condition for

\[
\lim_{n \to \infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)|^r \, dx = 0.
\]

The results give substantial improvements upon the existing results.
1. INTRODUCTION AND SUMMARY

Suppose that \( X_1, \ldots, X_n \) are i.i.d. samples of a d-dimensional random vector \( X \). Throughout this paper, we shall denote by \( F \) the distribution of \( X \), \( f \) the probability density function of \( X \), \( x^n = (X_1, \ldots, X_n) \), and \( F_n \) the empirical distribution of \( x^n \).

Let \( f_n = f_n(x) = f_n(x; x^n) \) be an estimate of \( f \) based on \( x^n \). For any constant \( r \geq 1 \), define
\[
m_{nr} = m_{nr}(x^n) = \int |f_n(x) - f(x)|^r dx.
\]
(1)

Here and in the sequel, \( \int \) means \( \int_{\mathbb{R}^d} \frac{1}{r} m_{nr} \), to be called the \( L_r \)-norm of \( f_n - f \), is a much-studied criterion in evaluating the performance of a density estimator. Quite a number of works have been done on the problem of convergence (to zero) of \( m_{nr} \) as the sample size \( n \) tends to infinity. We say that \( f_n \) is a \( L_r \)-norm consistent estimator of \( f \) if \( m_{nr} \to 0 \) as \( n \to \infty \) in some sense.

For the kernel estimator
\[
f_n(x) = (nh_n^d)^{-1} \sum_{i=1}^{n} K(h_n^{-1}(x - X_i)),
\]
where the kernel is assumed to be a probability density, Devroye [8] proved that the necessary and sufficient conditions for
\[
\lim_{n \to \infty} m_{n1} = 0, \text{ a.s.}
\]
are that \( h_n \to 0 \), \( nh_n^d \to \infty \). Bai and Chen [3] solved the general case of \( r \geq 1 \), proving that the necessary and sufficient conditions for
\[ \lim_{n \to \infty} m_n r = 0, \text{ a.s. for some } r > 1 \quad (3) \]

are that

\[ h_n + 0, \quad nh_n^d + \infty, \quad \int f^r(x)dx < \infty, \quad \int k^r(u)du < \infty. \]

In the case of \( k_n \)-nearest neighbor estimator proposed by Loftsgarden and Quesenberry \([10]\), Zhao Yue \([12]\) proved that a sufficient condition for \((3)\) in case of \( r > 1 \) is that

\[ k_n/n \to 0, \quad k_n/\log n \to \infty, \quad \int f^r(x)dx < \infty \quad (4) \]

while the first and last in \((4)\), and also that \( k_n \to \infty \), are necessary for the truth of \((3)\) \((r > 1)\).

Another important type of density-estimator is the histogram - ordinary histogram and data-based histogram. In ordinary histogram, the partitioning of range space of \( X, \mathbb{R}^d \), is done prior to the drawing of samples \( X^n \). For this case, Abou-Jaoude \([1], [2]\) (see also Devroye and Györfi \([9]\), pp.19-23) obtained the necessary and sufficient conditions (imposed on the partition) for the truth of \((2)\). Chen and Zhao \([7]\) solved the general case of \( r > 1 \), for the particular partition

\[ \mathbb{R}^d = \bigcup_{k_1, \ldots, k_d = -\infty}^\infty \{ x = (x^{(1)}, \ldots, x^{(d)}): a_i + k_i h_n \leq x^{(i)} < a_i + (k_i+1)h_n, \]

\[ 1 \leq i \leq d \}. \]

Data-based histogram differs from the ordinary one in that the partition of space \( \mathbb{R}^d \) defining the density estimate depends on the observations \( X^n \). Thus, after obtaining \( X^n \), we make a partition of \( \mathbb{R}^d \):

\[ \Phi_n \equiv \{ I(\ell, X^n): \ell = 1, 2, \ldots \} \]

\[ \bigcup_{\ell=1}^\infty I(\ell, X^n) = \mathbb{R}^d, \quad I(j, X^n) \cap I(k, X^n) = \emptyset \text{ when } j \neq k. \]
In this paper we consider only the case that \( I(\ell, X^n) \), \( \ell = 1, 2, \ldots \), are intervals in \( \mathbb{R}^d \) of the form
\[
[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d], \quad -\infty < a_i < b_i < \infty, \quad 1 \leq i \leq d.
\]
For each \( x \in \mathbb{R}^d \), denote by \( I_n(x) \) the unique interval in \( \phi_n \) containing \( x \), and by \( \lambda(I_n(x)) \) the Lebesgue measure of \( I_n(x) \). The data-based histogram estimate \( f_n \), based on the partition \( \phi_n \), is defined by
\[
f_n(x) = \frac{F_n(I_n(x))/\lambda(I_n(x))}{\lambda(I_n(x))}.
\]
For this estimate, the problem of \( L_r \)-norm consistency is much more complicated as compared with the ordinary histogram case. To begin with, for each positive integer \( n \) and positive constant \( t \), denote by \( C_{nt} \) the number of intervals in \( \phi_n = \{I(\ell, X^n)\} \) fulfilling the condition
\[
I(\ell, X^n) \cap \{x = (x^{(1)}, \ldots, x^{(d)}): |x^{(i)}| \leq t, \quad 1 \leq i \leq d\} \neq \emptyset
\]
and denote by \( D(A) \) the diameter for any set \( A \subseteq \mathbb{R}^d \). Chen and Rubin [4] proved that
\[
\lim_{n \to \infty} m_{nl} = 0, \quad \text{in probability}
\]
under three conditions, two of them are:
\[
\lim_{n \to \infty} D(I_n(x)) = 0, \quad \text{in probability, for } x \in \mathbb{R}^d, \text{ a.e.} \lambda \quad (7)
\]
\[
\begin{cases}
\lim_{n \to \infty} \frac{C_{nt}}{n} = 0, & d = 1; \\
\lim_{n \to \infty} \frac{C_{nt}}{\sqrt{n}} = 0, & d > 1.
\end{cases}
\]
while the third one is of a rather complicated nature. Chen and Zhao [6] studied the strong consistency for the case of general \( d \), proving
the truth of (2) under the conditions:

\[ \lim_{n \to \infty} D(I_n(x)) = 0, \text{ a.s., for } x \in \mathbb{R}^d, \text{ a.e.} \lambda \quad (7*) \]

\[ \lim_{n \to \infty} C_{nt} \log n / n = 0, \text{ a.s., for any } t > 0 \quad (8*) \]

while Chen and Wang [5] obtained analogous result for this problem.

By comparing (8) and (8*) we see that although in case \( d > 1 \) \( (8*) \) is an improvement of (8), but in case \( d = 1 \), in achieving strong consistency, we see that in case \( d > 1 \) we have not only made the improvement by establishing a.s. convergence instead of convergence in probability, but also succeeded in some sense in weakening the conditions required, since \( (8*) \) requires a lower rate of convergence to zero than (8) - Of course, strictly speaking, (8) and (8*) are mutually exclusive. In case \( d = 1 \), in achieving strong consistency, we pay a price by requiring that \( C_{nt} \) is of the order \( O(n/\log n) \) instead of \( O(n) \). Motivated by the works of Devroye [8], Bai and Chen [3] and Chen and Zhao [7], we expect that the order \( O(n) \) should be sufficient. In section 3, we shall prove that this is indeed the case:

**THEOREM 1.** Suppose that \( f_n \) is defined by (5), then (2) is true if (7*) and the following condition (9) are both true:

\[ \lim_{n \to \infty} C_{nt} / n = 0, \text{ a.s. for any } t > 0 \quad (9) \]

The general case of \( r > 1 \) is considered in section 4. To formulate our result, for any interval \( I = [a_1, b_1] \times \ldots \times [a_d, b_d] \) belonging to the partition \( \phi_n \), write \( a(I) \) for \( \min_{1 \leq i \leq d} (b_i - a_i) \). Also, for any \( t > 0 \), write

\[ Q_t = \{ x = (x(1), \ldots, x(d)) : |x(i)| \leq t, i = 1, \ldots, d \} \]
THEOREM 2. Suppose that \( f_n \) is defined by (5), then (3) is true if the following three conditions are satisfied:

\[
\int f^P(x)dx < \infty \quad (10)
\]

\[
\sup\{D(I): I \in \Phi_n, I \cap Q_t \neq \emptyset\} \rightarrow 0 \text{ a.s. for any } t > 0, \quad (11)
\]

\[
n(\inf_{I \in \Phi_n} a(I))^d \rightarrow \infty, \text{ a.s.} \quad (12)
\]

The basic tool in our argument is an inequality establishing the exponential bound for the deviation between theoretical and empirical distributions over a class of partitions of \( \mathbb{R}^d \). The inequality is of independent interest and is the subject of section 2.
2. AN INEQUALITY

Suppose that \( \mu \) is a probability measure on \( \mathcal{B}^d \) - the \( \sigma \)-field of all Borel sets in \( \mathbb{R}^d \). Let \( X_1, \ldots, X_n, \ldots \) be i.i.d. random vectors with a common probability distribution \( \mu \), and \( \nu_n \) be the empirical distribution of \( X_1, \ldots, X_n \).

We call \( \phi = \{A_1, \ldots, A_k\} \) a partition of \( \mathbb{R}^d \), if \( A_1, \ldots, A_k \) are disjoint intervals of the form

\[
[a_1, b_1) \times \cdots \times [a_d, b_d) : -\infty < a_i < b_i < \infty, \quad i = 1, \ldots, d
\]

and \( \bigcup_{i=1}^k A_i = \mathbb{R}^d \). For fixed positive integer \( k \), denote by \( F = F_k \) the collection of all such partitions, and define

\[
D_n = D_n(X^n) = \sup \{ \sum_{A \in \phi} |\nu_n(A) - \mu(A)| : \phi \in F \}
\]

It can easily be seen that there exists a countable subset \( \{\phi_i, i = 1, 2, \ldots\} \) of \( F \), such that \( D_n = \sup \{ \sum_{A \in \phi_i} |\nu_n(A) - \mu(A)| : i = 1, 2, \ldots\} \), and \( \{\phi_i\} \) is independent of \( X^n \). This shows that \( D_n \) is a random variable. We are now going to establish the following exponential bound for \( D_n \):

**THEOREM A.** Given \( \epsilon \in (0,1) \), we have

\[
P(D_n > \epsilon) < 6 \exp(-n\epsilon^2 - 9)
\]

when \( n \geq \max(k, 100 \log 6/\epsilon^2) \), and \( (\frac{k}{n}) \log(\frac{3en}{\epsilon k}) < \epsilon^2 2^{-9(d+1)-1} \).

In proving the theorem, we borrow some idea from a work of Vapnik and Chervonenkis [11]. We also note that Theorem A extends a work of Devroye [8], which we formulate below as a lemma:

**LEMMA 1** (Devroye [8]) Suppose that \( \mathbb{R}^d = \bigcup_{i=1}^k A_i, A_i \in \mathcal{B}^d \),
\[ i = 1, \ldots, k, \text{ and } A_i \cap A_j = \emptyset \text{ when } i \neq j. \] Then for given \( \varepsilon > 0 \) we have
\[
P\left( \sum_{i=1}^{k} |\mu_n(A_i) - \mu(A_i)| \geq \varepsilon \right) \leq 3 \exp\left(-n\varepsilon^2/25\right), \text{ when } k/n \leq \varepsilon^2/20
\]

The following simple fact is also needed in the proof:

**LEMMA 2.** Let \( q_i, \lambda_i, i = 1, \ldots, k \), be positive numbers. Write \( a = \prod_{i=1}^{k} \lambda_i, \quad b = \prod_{i=1}^{k} \lambda_i q_i. \) We have
\[
\prod_{i=1}^{k} \lambda_i q_i > \left(\frac{b}{a}\right)^b
\]
and the equality holds if and only if \( q_1 = \ldots = q_k. \)

The proof is easy and therefore omitted.

**Proof of the Theorem.**

Write \( x^{(n)} = (x_{n+1}, \ldots, x_{2n}), \quad x_{2n} = (x_1, \ldots, x_{2n}), \quad \mu^*_n \) the empirical measure of \( x^{(n)} \), and
\[
D_n(x^{(n)}, \phi) = \sum_{A \in \Phi} |\mu_n(A) - \mu(A)|
\]
\[
D^*_n(x^{(n)}, \phi) = \sum_{A \in \Phi} |\mu_n^*(A) - \mu(A)|
\]
\[
G_n(x^{2n}, \phi) = \sum_{A \in \Phi} |\mu_n(A) - \mu_n^*(A)|
\]
\[
G_n = G_n(x^{2n}) = \sup(G_n(x^{2n}, \phi^*): \phi \in \Phi)
\]

Since \( \{G_n > \varepsilon/2\} = \bigcup_{i=1}^{\infty} \{G_n(\phi_i) > \varepsilon/2\} \cup \bigcup_{i=1}^{\infty} \{D_n(\phi_i) > \varepsilon\} \cap \{D^*_n(\phi_i) < \varepsilon/2\} \)
and \( \{D_n(\phi_i): i = 1, 2, \ldots, \} \), \( \{D^*_n(\phi_i): i = 1, 2, \ldots, \} \) are independent, it is well known that
\[
P(\varepsilon/2) \geq \inf_i P(D^*_n(\phi_i) < \varepsilon/2) P(\bigcup_{i=1}^{\infty} \{D_n(\phi_i) > \varepsilon\})
\]
\[
= \inf_i P(D^*_n(\phi_i) < \varepsilon/2) P(D_n > \varepsilon)
\]
Suppose that \( n \) satisfies the conditions indicated in Theorem A, then \( k/n < \epsilon^2/80 \), and by Lemma 1 we have, simultaneously for all \( x^n \):

\[
P(D^*_n(x^{(n)}, \phi_i) \geq \epsilon/2|x^n) \leq 3\exp(-\eta \epsilon^2/100) \leq 1/2, \quad i = 1, 2, \ldots
\]

Therefore, \( P(D^*_n(\phi_i) < \epsilon/2, \quad i = 1, 2, \ldots, \) and

\[
P(G_n > \frac{\epsilon}{2}) \leq \frac{1}{2}P(D_n > \epsilon) \quad (13)
\]

From (13), it is seen that the proof of Theorem A reduces to the problem of finding an upper bound for \( P(G_n > \epsilon/2) \). For this purpose, denote by \( T \) a permutation \( (j_1, j_2, \ldots, j_{2n}) \) of \( (1, 2, \ldots, 2n) \), so that \( TX^{2n} = (x_{j_1}, x_{j_2}, \ldots, x_{j_{2n}}) \).

Further, denote by \( u_n(T) \) and \( u_n(T)^* \) the empirical measures generated by \( (x_{j_1}, \ldots, x_{j_n}) \) and \( (x_{j_{n+1}}, \ldots, x_{j_{2n}}) \), respectively. Then it is readily seen that

\[
P(G_n > \frac{\epsilon}{2}) = \int_{R^{2nd}} \frac{1}{(2n)!} \sum_{t=1}^{T} \sup_{\phi \in F} |u_n(T)(A) - u_n(T)^*(A)| > \frac{\epsilon}{2} \, dP
\]

\[
\leq \int_{R^{2nd}} \frac{1}{(2n)!} \sum_{t=1}^{T} \sup_{\phi \in F} |u_n(T)(A) - u_{2n}(A)| > \frac{\epsilon}{4} \, dP \quad (14)
\]

where the summation \( \sum \) is taken over all \( (2n)! \) permutations of \( (1, 2, \ldots, 2n) \), and \( P = \mu^m \).

Now fix \( x^{2n} \), and denote by \( U \) the set with elements \( x_1, \ldots, x_{2n} \). Each \( \phi \in F \) induces a partition of the set \( U \). Denote by \( m_n(U) \) the number of different partitions induced by all \( \phi \in F \). We have

\[
m_n(U) \leq \frac{(2n+k-1)!}{k-1} \leq \left( \frac{3n}{k} \right)^d \leq \left( \frac{3en}{k} \right)^d \quad (15)
\]
Let $F^*$ be a subset of $F$ having $m_n(U)$ members, such that if $\phi_i \in F^*, i = 1, 2$ and $\phi_1 \neq \phi_2$, then $\phi_1$ and $\phi_2$ induce different partitions of $U$. We have

$$\frac{1}{(2n)!} \sum_T \sup_{\phi \in F} \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

$$= \frac{1}{(2n)!} \sum_T \sup_{\phi \in F^*} \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

$$= \frac{1}{(2n)!} \sum_T \sup_{\phi \in F^*} \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

$$\leq \sum_{\phi \in F^*} \frac{1}{(2n)!} \sum_T \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

$$\leq m_n(U) \sup_{\phi \in F} \frac{1}{(2n)!} \sum_T \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

(16)

Fix $\phi = \{A_1, \ldots, A_k\} \in F$. Denote by $Y_1, \ldots, Y_n$ a random sample taken from $U$ without replacement, and $\{Z_i, i \geq 1\}$ be a sequence of random samples taken from $U$ with replacement. Write $\tilde{P}(\cdot) = P(\cdot | \chi^{2n})$, $\tilde{E}(\cdot)$ $= E(\cdot | \chi^{2n})$, and

$$\tilde{p}_\ell = \mu_{2n}(A_\ell), \quad N_\ell = 2n \tilde{p}_\ell, \quad \ell = 1, \ldots, k$$

$$V_n = \sum_{\ell=1}^k |I(\sum_{i=1}^n I(Y_i \in A_\ell) - n \tilde{p}_\ell)|, \quad W_n = \sum_{\ell=1}^k \sum_{i=1}^n I(Z_i \in A_\ell) - n \tilde{p}_\ell|$$

(17)

Then we have

$$\frac{1}{(2n)!} \sum_T \sum_{A \in \phi} |\mu_n^T(A) - \mu_{2n}(A)| > \frac{\varepsilon}{4}$$

$$= \tilde{E}(I(\sum_{\ell=1}^k \frac{1}{n} \sum_{i=1}^n I(Y_i \in A_\ell) - \tilde{p}_\ell > \frac{\varepsilon}{4})) = \tilde{P}(V_n > \varepsilon n/4)$$

(18)
Now we proceed to show that

$$\tilde{E}\{\exp(tV_n)\} \leq (4\pi e^{1/6} n/k)^{k/2} \tilde{E}\{\exp(tW_n)\}$$

(19)

for any $t > 0$. In fact,

$$\tilde{E}\exp(tW_n)$$

$$= \sum' \frac{n!}{n! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k} \exp(t \sum_{\xi=1}^k |n_\xi - np_\xi|),$$

(20)

where the summation $\sum'$ is taken over all integer-valued vectors $(n_1, \ldots, n_k)$ satisfying

$$n_1 \geq 0, \ldots, n_k \geq 0 \quad \text{and} \quad \sum_{\xi=1}^k n_\xi = n.$$

In the same way, we have

$$\tilde{E}\{\exp(tV_n)\} = \sum' C(n_1, \ldots, n_k) \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k} \exp(t \sum_{\xi=1}^k |n_\xi - np_\xi|).$$

(21)

Here, as usual, we put $\binom{n}{m} = 0$ for $m > n$. Also,

$$C(n_1, \ldots, n_k) = \frac{n!(2n)^n}{(2n)!} \prod_{j=1}^k \frac{N_j!}{(N_j-n_j)!} \frac{N_j!}{N_j-n_j!} \frac{N_j!}{N_j!}$$

$$= \frac{n!(2n)^n}{(2n)!} \prod_{(I)} (N_j!)(-N_j!) \frac{N_j!}{(N_j-n_j)!} \frac{N_j!}{(N_j-n_j)!} \frac{N_j!}{N_j!},$$

where $\prod_{(I)}$ is taken over all $j$'s satisfying $N_j = n_j$ and $\prod_{(II)}$ is taken over all $j$'s satisfying $0 < n_j < N_j$. Using Stirling's formula

$$\sqrt{2\pi n} \ n^n e^{-n} < n! < \sqrt{2\pi n} \ n^n e^{-n+1/(12n)}$$
and the fact that \( \sum_{j=1}^{k} n_j = n \), we get

\[
C(n_1, \ldots, n_k) \leq 2^{n-1/2} e^{n+1/12} (2\pi)^{k/2} \exp(k/12 - \sum_{j=1}^{k} n_j) \prod_{j=1}^{k} \left( \frac{N_j - n_j}{N_j} \right)^{N_j - n_j - 1/2} \]

\[
\leq 2^{n-1/2} (2\pi)^{k/2} e^{(k+1)/12} \prod_{j=1}^{k} \left( \frac{N_j - n_j}{N_j} \right)^{N_j - n_j} \prod_{j=1}^{k} \sqrt{N_j},
\]

(22)

and the summation \( \xi \) appearing below, are taken over all \( (\text{III})' \), \( (\text{III}) \) j's satisfying \( 0 \leq n_j < N_j \). Putting \( q_j = (N_j - n_j)/N_j \), \( \lambda_j = N_j \) in Lemma 2, we get

\[
ad \△ \sum_{j=1}^{k} \lambda_j = \sum_{j=1}^{k} N_j \leq 2n,
\]

\[
b \△ \sum_{j=1}^{k} \lambda_j q_j = \sum_{j=1}^{k} (N_j - n_j) = \sum_{j=1}^{k} (N_j - n_j) = n,
\]

and

\[
\prod_{j=1}^{k} \left( \frac{N_j - n_j}{N_j} \right)^{N_j - n_j} = \prod_{j=1}^{k} \lambda_j q_j \geq (b/a)^b \geq 2^{-n}.
\]

(23)

On the other hand,

\[
k \prod_{j=1}^{k} \sqrt{N_j} \leq \left( \frac{1}{k} \sum_{j=1}^{k} N_j \right)^{k/2} = (2n/k)^{k/2}.
\]

(24)

By (22) - (24),

\[
C(n_1, \ldots, n_k) \leq 2^{-1/2} e^{(k+1)/12} (4\pi n/k)^{k/2} \leq (4\pi n/k)^{k/2} e^{k/12},
\]

(25)

and (19) follows from (20), (21) and (25).

Let \( N \) be a Poisson random variable, \( E(N) = n \), and \( N, (Y_1, \ldots, Y_n, Z_1, Z_2, \ldots) \) are independent. Since \( \sum_{i=1}^{N} I(Z_i \in A_\ell), \ell = 1, \ldots, k, \) are
independent Poisson variables with mean \( np_1, \ldots, np_k \) respectively, it follows from (19) that for any \( t > 0 \),

\[
\tilde{P}(V_n > n\epsilon/4) \\
\leq \tilde{P}(|N-n| > n\epsilon/8) + e^{-tn\epsilon/4}E(e^{tV_nI(|N-n| \leq n\epsilon/8)}) \\
= \tilde{P}(|N-n| > n\epsilon/8) + e^{-tn\epsilon/4}tV_n\tilde{P}(|N-n| \leq n\epsilon/8) \\
\leq \tilde{P}(|N-n| > n\epsilon/8) + (4\pi e^{1/6} n/k) k/2 e^{-tn\epsilon/4}E(e^{tV_nI(|N-n| \leq n\epsilon/8)}) \\
= \tilde{P}(|N-n| > n\epsilon/8) + (4\pi e^{1/6} n/k) k/2 e^{-tn\epsilon/4}E(e^{tV_nI(|N-n| \leq n\epsilon/8)}).
\]

From the independence mentioned above and

\[
e^{tW_nI(|N-n| \leq n\epsilon/8)} \leq \exp\left(t \sum_{i=1}^k \sum_{\ell=1}^N I(Z_{i\ell}eA_{\ell}) - np_{\ell}\right) + t\epsilon/8),
\]

it follows that

\[
\tilde{P}(V_n > n\epsilon/4) \\
\leq \tilde{P}(|N-n| > n\epsilon/8) \\
+ (4\pi e^{1/6} n/k) k/2 e^{-tn\epsilon/4}E(\exp(t \sum_{i=1}^k \sum_{\ell=1}^N I(Z_{i\ell}eA_{\ell}) - np_{\ell})). \tag{26}
\]

Now suppose that \( V \) is a Poisson variable and \( EV = \lambda \). From

\[
e^{-t} + t \leq e^{t - t} = e^t - t \text{ for } t > 0, \text{ it follows that} \\
E(e^{t|V-\lambda|}) \leq E(e^{t(V-\lambda) + e^{-t(V-\lambda)})} \\
= \exp(\lambda(e^{t} - 1 - t)) + \exp(\lambda(e^{-t} - 1 + t)) \leq 2\exp(\lambda(e^{t} - 1 - t)).
\]

So we have

\[
P(|V-\lambda| \geq \lambda\epsilon) \leq E(\exp(t|V-\lambda| - t\lambda\epsilon))
\]
Take \( t = \log(1+\epsilon) \), we get

\[
P(\{|V-\lambda| \geq \lambda \epsilon\} \leq 2\exp(\lambda(1+\epsilon)\log(1+\epsilon)))
\]

\[
\leq 2\exp(-\lambda \epsilon^2/(2+2\epsilon)) \leq 2\exp(-\lambda \epsilon^2/4)
\]

for \( \epsilon \in (0,1) \). Repeat this argument and take \( t = \log(1+\epsilon/8) \), by (26) we have

\[
\overline{P}(V_n > n \epsilon/4)
\]

\[
\leq 2\exp(-n\epsilon^2/256) + (4\pi e^{1/6} n/k)^{k/2} e^{-t\epsilon/8} \prod_{\ell=1}^{k} \{2\exp(nP_{\ell}(e^{t-1-t}))\}
\]

\[
\leq 2\exp(-n\epsilon^2/256) + (4\pi e^{1/6} n/k)^{k/2} 2^k \exp(n(e^{-1-t-t\epsilon/8}))
\]

\[
\leq 2\exp(-n\epsilon^2/256) + (16\pi e^{1/6} n/k)^{k/2} \exp(-n\epsilon^2/256)
\]

\[
\leq 3(16\pi e^{1/6} n/k)^{k/2} \exp(-n\epsilon^2/256)
\]

From (14) - (18) and (27), it follows that

\[
P(G_n > \epsilon/2) \leq 3(3en/k)^{kd}(16\pi e^{1/6} n/k)^{k/2} \exp(-n\epsilon^2/256)
\]

\[
= 3 \exp(-n\epsilon^2/256 + kd \log (3en/k) + \frac{k}{2}\log(16\pi e^{1/6} n/k)).
\]

Under the conditions of Theorem A, \( n/k > 16 e^{1/6}/(9\epsilon^2) \) and

\( k(d+1)\log(3en/k) < n\epsilon^2/2^9 \). Hence,

\[
P(G_n > \epsilon/2) \leq 3 \exp(-n\epsilon^2/256 + k(d+1)\log(3en/k))
\]

\[
\leq 3 \exp(-n\epsilon^2/2^9).
\]

From this and (13), Theorem A follows.
3. PROOF OF THEOREM 1

Define
\[ \tilde{f}_n(x) = \int_{I_n(x)} f(u) \frac{du}{\lambda(I_n(x))}. \] (28)

It is enough to show that for any \( t > 0 \),
\[ \lim_{n \to \infty} \int_{Q_t} |f(x) - \tilde{f}_n(x)| \, dx = 0 \quad \text{a.s.} \] (29)

and
\[ \lim_{n \to \infty} \int_{Q_t} |f_n(x) - \tilde{f}_n(x)| \, dx = 0 \quad \text{a.s.} \] (30)

For any \( \varepsilon > 0 \), we can find a function \( g(x) \geq 0 \) which is continuous on \( \mathbb{R}^d \) and has a bounded support, such that \( \int |f-g| \, dx < \varepsilon \). Define
\[ \tilde{g}_n(x) = \int_{I_n(x)} g(u) \frac{du}{\lambda(I_n(x))}. \]

Then
\[
\int_{Q_t} |f-\tilde{f}_n| \, dx \leq \int |f-g| \, dx + \int |\tilde{f}_n-\tilde{g}_n| \, dx + \int_{Q_t} |\tilde{g}_n-g| \, dx
\leq 2\int |f-g| \, dx + \int_{Q_t} |\tilde{g}_n-g| \, dx
< 2\varepsilon + \int_{Q_t} |\tilde{g}_n-g| \, dx. \] (31)

By (7*), there exists a set \( B_0 \subseteq \mathbb{R}^d \) such that \( P(B_0) = 0 \) and for
\( \omega \in (X_1, X_2, \ldots) \in B_0 \), we have \( \lim_{n \to \infty} D(I_n(x)) = 0 \) for \( x \in \mathbb{R}^d \), a.e. \( \lambda \), and in turn it follows that \( \lim_{n \to \infty} \tilde{g}_n(x) = g(x) \) for \( x \in \mathbb{R}^d \), a.e. \( \lambda \).

By the dominated convergence theorem,
\[ \lim_{n \to \infty} \int_{Q_t} |\tilde{g}_n(x)-g(x)| \, dx = 0 \quad \text{a.s.}, \] (32)

and (29) follows from (31) and (32).
From (9) it can be shown that there exists a sequence \( \{ \delta_n \} \) of positive numbers such that \( \lim_{n \to \infty} \delta_n = 0 \) and
\[
\lim_{n \to \infty} \frac{C_{nt}}{[n\delta_n]} = 0 \quad \text{a.s.,}
\]
where \( [n\delta_n] \) denotes the integer part of \( n\delta_n \). For any \( \varepsilon \in (0,1) \), there exists a set \( B_{1-\varepsilon/2} \subset \mathbb{R}^d \) such that \( \mathbb{P}(B_{1-\varepsilon/2}) > 1 - \varepsilon/2 \) and
\[
\lim_{n \to \infty} \frac{C_{nt}}{[n\delta_n]} = 0 \quad \text{uniformly for } (x_1, x_2, \ldots) \in B_{1-\varepsilon/2}.
\]
So there exists a positive integer \( N \) such that
\[
C_{nt} < [n\delta_n], \quad \text{for } n \geq N \text{ and } (x_1, x_2, \ldots) \in B_{1-\varepsilon/2}.
\]
Now we recall \( \phi_n \equiv \phi_n(x^n) = \{ I(\ell, x^n), \; \ell = 1, 2, \ldots \} \) is the partition of \( \mathbb{R}^d \) which defines the data-based histogram \( f_n \). It is easy to see that we can find \( k \leq 3^d C_{nt} \) and \( \phi \in F_k \) such that
\[
(\{ I : I \in \phi, I \cap Q_t \neq \emptyset \}) = (\{ I : I \in \phi_n, I \cap Q_n \neq \emptyset \}).
\]
Hence, for \( (x_1, x_2, \ldots) \in B_{1-\varepsilon/2}, \; n \geq N \) and \( k = 3^d [n\delta_n] \), we have
\[
\int_{Q_t} |f_n(x) - \hat{f}_n(x)| \, dx \leq \sum_{I \in \phi_n} \mathbb{I}_{I \cap Q_t \neq \emptyset} |F_n(I) - F(I)| \leq \sup_{\phi \in F_k} \sum_{A \in \phi} |F_n(A) - F(A)| \triangleq D_n. \quad (33)
\]
Since \( k/n = 3^d [n\delta_n]/n \leq 3^d \delta_n \to 0 \) as \( n \to \infty \), from Theorem A, we have
\[
\lim_{n \to \infty} D_n = 0 \quad \text{a.s.} \quad (34)
\]
By (33) and (34), there exists a set \( B_{1-\varepsilon} \subset \mathbb{R}^d \) such that \( B_{1-\varepsilon} \subset B_{1-\varepsilon/2} \), \( \mathbb{P}(B_{1-\varepsilon}) > 1 - \varepsilon \) and
\[ \lim_{n \to \infty} \int |f_n(x) - \tilde{f}_n(x)| \, dx = 0 \quad \text{uniformly for} \quad (X_1, X_2, \ldots) \in B_{1-\varepsilon}. \]

Since \( \varepsilon > 0 \) is arbitrarily given, (30) is proved, and the proof of Theorem 1 is completed.
4. PROOF OF THEOREM 2

Define \( \tilde{f}_n(x) \) as before by (28). Find a nonnegative function \( g \), continuous everywhere on \( \mathbb{R}^d \) and with a bounded support, such that

\[
\int |f - g|^r dx < \varepsilon^r.
\]

Put

\[
\tilde{g}_n(x) = \int_{I_n(x)} g(u) du / \lambda(I_n(x)).
\]

Then

\[
\left( \int |f - \tilde{f}_n|^r dx \right)^{1/r} \leq \left( \int |f - g|^r dx \right)^{1/r} + \left( \int |\tilde{f}_n - \tilde{g}_n|^r dx \right)^{1/r} + \left( \int |\tilde{g}_n - g|^r dx \right)^{1/r}
\]

\[
\leq 2 \left( \int |f - g|^r dx \right)^{1/r} + \left( \int |\tilde{g}_n - g|^r dx \right)^{1/r}
\]

\[
< 2\varepsilon + \left( \int |\tilde{g}_n - g|^r dx \right)^{1/r}.
\]

By (11), for any \( x \in \mathbb{R}^d \), we have

\[
D(I_n(x)) \to 0 \quad \text{a.s.}
\]

There exists a set \( B_0 \subset (\mathbb{R}^d)_{\infty} \) such that \( P(B_0) = 0 \) and for \( \omega \equiv (x_1, x_2, \ldots) \in B_0 \) we have \( \lim_{n \to \infty} D(I_n(x)) = 0 \) for \( x \in \mathbb{R}^d \), a.e. \( \lambda \), and in turn it follows that \( \lim_{n \to \infty} \tilde{g}_n(x) = g(x) \) for \( x \in \mathbb{R}^d \), a.e. \( \lambda \). By the dominated convergence theorem,

\[
\lim_{n \to \infty} \int |\tilde{g}_n - g|^r dx = 0 \quad \text{for} \quad \omega \in B_0.
\]

By (35) and (36), we have

\[
\lim_{n \to \infty} \int |f(x) - \tilde{f}_n(x)|^r dx = 0 \quad \text{a.s.}
\]

Now we proceed to prove that

\[
\int |f_n - \tilde{f}_n|^r dx = \sum_{I_{\ell} \in \mathcal{P}_n} |F_n(I_{\ell}) - F(I_{\ell})|^r / \lambda(I_{\ell})^{r-1} \to 0, \quad \text{a.s.}
\]
Put $H = \inf_{I \in \phi_n} a(I)$. Since $nH^d \to \infty$ a.s., it can be shown that there exists a sequence $C_n \to \infty$ such that $\lim_{n \to \infty} nH^d/C_n^d = \infty$ a.s.. Without loss of generality, we can assume that $C_n^d/n \to 0$. Take $h = h_n = C_n/n^{1/d}$, then $h_n \to 0$, $nh_n^d \to \infty$ and $H/h_n \to \infty$, a.s.

Construct a partition of $\mathbb{R}^d$ into disjoint finite intervals, say $\psi_n = \{\Delta_1, \Delta_2, \ldots\}$, where $\Delta_m's$ are all cubes with the same edge length $h$. Define

$$\xi_n(x) = F_n(\Delta_m)/h^d \quad \text{for} \quad x \in \Delta_m, \ m = 1, 2, \ldots$$

and

$$\tilde{\xi}_n(x) = F(\Delta_m)/h^d \quad \text{for} \quad x \in \Delta_m, \ m = 1, 2, \ldots$$

By the theorem of [7],

$$\lim_{n \to \infty} \int |\xi_n(x) - f(x)|^r dx = 0. \ a.s.$$

An argument similar to that leading to (37) gives

$$\lim_{n \to \infty} \int |\tilde{\xi}_n(x) - f(x)|^r dx = 0.$$

So we have

$$\int |\xi_n(x) - \tilde{\xi}_n(x)|^r dx = \sum_{\Delta_m \in \psi_n} |F_n(\Delta_m) - F(\Delta_m)|^r/(h^d)^{r-1} \to 0 \ a.s. \ as \ n \to \infty. \ (39)$$

For $I \in \phi_n$, denote by $H_1\ell, \ldots, H_d\ell$ the lengths of the edges of $I \ell$, and write

$$M_\ell = \{m : \Delta_m \in \psi_n, \Delta_m \subset I \ell\},$$

$$\tilde{M}_\ell = \{m : \Delta_m \in \psi_n, \Delta_m \cap I \ell \neq \emptyset, \Delta_m \setminus I \ell \neq \emptyset\}. \ (40)$$
Since $H/h_n \rightarrow \infty$ a.s., we can find $B_* \subseteq \mathbb{R}^d$ such that $P(B_*) = 0$ and $H/h_n \rightarrow \infty$ for $\omega \in B_*$. In the sequel we always keep $\omega \in B_*$. Thus, for $n$ large, $H_{i,s} > 2h$ for all $i$ and $\ell$. We have

$$d \pi (H_{i,s}/h - 2) \leq \#(M_{i,s}) \leq \pi (H_{i,s}/h),$$

(41)

and

$$\#(M_{i,s}) \leq \pi (H_{i,s}/h - 2) \leq \pi (H_{i,s}/h - 2)$$

$$\leq \lambda (I_{i,s})h^{-d}\{\pi (1+2h/H_{i,s}) - \pi (1-2h/H_{i,s})\}$$

$$\leq \lambda (I_{i,s})h^{-d}\{(1+2h/H)^d - (1-2h/H)^d\}$$

$$\leq h^{-d}\lambda(I_{i,s})C(d)h/H,$$

(42)

where

$$C(d) = 2^d(2^{d-1}+1).$$

Now, by (41) and (39) we have

$$\rho_n \Delta \sum_{i=1}^{\pi (H_{i,s}/h - 2)} \sum_{m \in \Delta_{i,s}} \left| F_n\left( \sum_{m \in \Delta_{i,s}} \Delta_m \right) - F\left( \sum_{m \in \Delta_{i,s}} \Delta_m \right) \right| r/\lambda(I_{i,s})^{r-1}$$

$$\leq \sum_{i=1}^{\pi (H_{i,s}/h - 2)} \sum_{m \in \Delta_{i,s}} \left| F_n(\Delta_m) - F(\Delta_m) \right| r/\lambda(I_{i,s})^{r-1}$$

$$\leq \sum_{\Delta_m \in \Delta_{i,s}} \left| F_n(\Delta_m) - F(\Delta_m) \right| r/(h^d)^{r-1} + O, \quad a.s.$$  

(43)

On the other hand, by (42) we have

$$\rho_n \Delta \sum_{i=1}^{\pi (H_{i,s}/h - 2)} \sum_{m \in \Delta_{i,s}} \left| F_n\left( \sum_{m \in \Delta_{i,s}} (I_{i,s} \Delta_m) \right) - F\left( \sum_{m \in \Delta_{i,s}} (I_{i,s} \Delta_m) \right) \right| r/\lambda(I_{i,s})^{r-1}$$

$$\leq \sum_{i=1}^{\pi (H_{i,s}/h - 2)} \sum_{m \in \Delta_{i,s}} \left| F_n(I_{i,s} \Delta_m) - F(I_{i,s} \Delta_m) \right| r/\lambda(I_{i,s})^{r-1}$$

$$\leq \sum_{i=1}^{\pi (H_{i,s}/h - 2)} \sum_{m \in \Delta_{i,s}} \left| F_n(I_{i,s} \Delta_m) - F(I_{i,s} \Delta_m) \right| r/\lambda(I_{i,s})^{r-1}$$

$$\leq h^{-d}\lambda(I_{i,s})C(d)h/H.$$
\[(hC(d)H^{-1})^{r-1} \sum_{I_\xi \in \Psi_n} \sum_{I_\eta \in M_\xi} |F_n(I_\xi \Delta_m) - F(I_\xi \Delta_m)|^{r/(h^d)^{r-1}}. \quad (44)\]

For each \( \Delta_m \in \Psi_n \), define

\[N_m = \{ I_\xi \in \Psi_n, I_\eta \cap \Delta_m \neq \emptyset \}. \quad (45)\]

Since \( H_{i\xi} > 2h \) for all \( i \) and \( \xi \), for any \( m \) the set \( N_m \) contains at most \( 2^d \) elements. By (44),

\[\rho_n \leq (C(d)h/H)^{r-1} \sum_{\Delta_m \in \Psi_n} \sum_{I_\xi \in N_m} |F_n(I_\xi \Delta_m) - F(I_\xi \Delta_m)|^{r/(h^d)^{r-1}}\]

\[\leq (C(d)h/H)^{r-1} \sum_{\Delta_m \in \Psi_n} \#(N_m)2^{r-1}(F_n(\Delta_m)^r + F(\Delta_m)^r/(h^d)^{r-1})\]

\[\leq 2^{d+r-1}(C(d)h/H)^{r-1} \sum_{\Delta_m \in \Psi_n} 2^{r-1}F_n(\Delta_m)^r - F(\Delta_m)^r/(h^d)^{r-1}\]

\[+ 2^{d+r-1}(C(d)h/H)^{r-1} \sum_{\Delta_m \in \Psi_n} (2^{r-1} + 1)F(\Delta_m)^r/(h^d)^{r-1}\]

\[\Delta \sim \rho_{n1} + \rho_{n2}. \quad (46)\]

By (43),

\[\lim_{n \to \infty} \rho_{n1} = 0 \quad \text{a.s.} \quad (47)\]

By Jensen's inequality,

\[\sum_{\Delta_m \in \Psi_n} F(\Delta_m)^r/(h^d)^{r-1} = \sum_{\Delta_m \in \Psi_n} h^{-d} \int_{\Delta_m} f(x) dx = h^d \int f(x) dx,\]

which implies that

\[\lim_{n \to \infty} \rho_{n2} = 0 \quad \text{a.s.} \quad (48)\]
From (46) - (48), we obtain

\[ \lim_{n \to \infty} \rho_n = 0 \quad \text{a.s.} \quad (49) \]

By (43) and (49), we have

\[
\int |f_n(x) - \tilde{f}_n(x)|^r dx = \sum_{I_N \in \mathcal{P}_n} |F_n(I_N) - F(I_N)|^{\frac{r}{\lambda}}(I_N)^{r-1} \\
\leq 2^{r-1}(\rho_n + \tilde{\rho}_n) \to 0 \quad \text{a.s.}
\]

Thus, (38) is proved, and Theorem 2 follows from (37) and (38).
REFERENCES


END
DATE
FILMED
5-88
DTC