

NOTAS DE MATEMATICA

Nº 40

NON PARAMETRIC ESTIMATION OF THE GRADIENT OF
THE DENSITY FUNCTION IN THE MULTIVARIATE CASE

POR

NELIDA WINZER

UNIVERSIDAD DE LOS ANDES
FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMATICAS
MERIDA-VENEZUELA

1980

NON PARAMETRIC ESTIMATION OF THE GRADIENT OF
THE DENSITY FUNCTION IN THE MULTIVARIATE CASE.

BY

NELIDA WINZER

DOCTORAL THESIS PRESENTED AT
THE FACULTY OF SCIENCES OF
THE ROLAND EÖTVÖS UNIVERSITY

BUDAPEST, 1979

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. István Vincze for his valuable guidance and encouragement.

I am also indebted to Dr. Endre Csáki, who read the originals, for his constructive criticism.

I gratefully acknowledge the financial support given by the Departamento de Matemática of the Universidad de Los Andes, Mérida, Venezuela, that made this work possible and to the Hungarian Academy of Sciences which allowed me to study in Hungary.

Budapest, october 1979.

TABLE OF CONTENTS

INTRODUCTION	I-II
1.- The multivariate lemma	1
2.- The estimation of $\nabla f(P_0) / f(P_0)$	6
3.- Estimation of the information	14
4.- The learning model	22
5.- References	35

INTRODUCTION

Following the pioneering works by Rosenblatt and Parzen, several authors have studied the estimation of density functions and their derivatives and some functionals related with them.

Wegman has given a survey of the methods and of the most important results up to 1972.

Using the kernel's method, Bhattacharyya (1967), gave for the first time an estimation of the Fisher's information function:

$$\begin{aligned} I(\theta) &= \int \left[\frac{\partial}{\partial \theta} \ln f(x - \theta) \right]^2 f(x - \theta) dx = \\ &= \int \left[\frac{\partial^2}{\partial \theta^2} \ln f(x - \theta) \right] f(x - \theta) dx . \end{aligned}$$

Later, Dmitriev-Tarassenko (1973) obtained some new results for $I(\theta)$ with a similar estimation, and also studied an estimation for the entropy

$$\int f(y) \ln f(y) dy.$$

Singh (1976) worked with multidimensional distribution functions and found kernel estimations for the density function and their derivatives.

Csaki and Vincze, using a lemma for real function due to the second author, obtained an estimator of the information function in the unidimensional case with location parameter.

In the present work, a similar lemma for the n-dimensional case will be proved. On the basis of this lemma we are able to estimate:

- a) the Fisher's information function for unidimensional location parameter;
- b) the information matrix for the case of location parameter in the multivariate case;
- c) Fisher's information function for the general case which can be done in the frame of the learning model only.

For a), b) and c) conditions will be given in order to obtain consistency in probability and consistency with probability one.

An estimation for the quotient

$$\frac{\nabla f(P_0)}{f(P_0)} = \left(\frac{f_1(P_0)}{f(P_0)}, \frac{f_2(P_0)}{f(P_0)}, \dots, \frac{f_n(P_0)}{f(P_0)} \right)$$

will be given too, where $f_i(P_0) = \frac{\partial f}{\partial x_i}(P_0)$.

1. THE MULTIVARIATE LEMMA.

We begin with the proof of the following

LEMMA 1.1. Let $f: \mathbb{R}^p \rightarrow \mathbb{R}$ be a positive function with continuous first partial derivatives in a neighbourhood of a point P_0 . Let us suppose that this point belongs to a p -dimensional parallelepiped C , contained in the same neighbourhood of P_0 .

S and M stand for the mass center of C with respect to the given function and the middle point of C respectively.

If the sides of C are $\Delta x_1, \Delta x_2, \dots, \Delta x_p$ we can denote by $|C| = \Delta x_1 \cdot \Delta x_2 \cdot \dots \cdot \Delta x_p$ the volume of C .

If we let C tending to P_0 in such a way that

$$\frac{\Delta x_i}{\sqrt{\sum_{j=1}^p \Delta x_j^2}} \rightarrow a_i, \quad i=1, \dots, p \quad (1.1)$$

with $0 < a_i < 1$ and $\sum a_i^2 = 1$, then

$$\frac{\overrightarrow{(MS)}_j}{|C|^{2/p}} \rightarrow \frac{a_j^2}{\prod_{k=1}^p a_k^{2/p}} \cdot \frac{1}{12} \cdot \frac{f'_j(P_0)}{f(P_0)}, \quad j = 1, \dots, p \quad (1.2)$$

where $\overrightarrow{(MS)}_j$ is the j -th component of the vector $\overrightarrow{(MS)}$.

PROOF. Due to the assumption concerning f we can write

$$f(\underline{r}) = f(\underline{r}_0) + (\underline{r} - \underline{r}_0, \nabla f(\underline{r}_0)) + o(|\underline{r} - \underline{r}_0|) \quad (1.3)$$

where \underline{r} is the position vector of a point P in C and \underline{r}_0 is the position vector of P_0 , (\dots) stands for the inner product.

For the sake of simplicity and without loss of generality we may take $P_0 = (0, 0, \dots, 0)$.

If \underline{m} is the position vector of M , then

$$\begin{aligned} \vec{MS} &= \frac{\int_C (\underline{r} - \underline{m}) f(\underline{r}) dV}{\int_C f(\underline{r}) dV} = \\ &= \frac{\int_C (\underline{r} - \underline{m}) \left[f(\underline{0}) + (\underline{r}, \nabla f(\underline{0})) + o(|\underline{r}|) \right] dV}{\int_C \left[f(\underline{0}) + (\underline{r}, \nabla f(\underline{0})) + o(|\underline{r}|) \right] dV} \end{aligned}$$

Integration in the denominator gives:

$$\int_C f(\underline{0}) dV = f(\underline{0}) |C|,$$

$$\int_C (\underline{r}, \nabla f(\underline{0})) dV = \left(\int_C \underline{r} dV, \nabla f(\underline{0}) \right) = |C| (\underline{m}, \nabla f(\underline{0})) \text{ and,}$$

according to (1.1)

$$\int_C o(|\underline{r}|) dV \leq o\left(\sqrt{(\Delta x_1)^2 + \dots + (\Delta x_p)^2}\right) |C| = o(|C|^{1/p}) |C|.$$

In the numerator, we have:

$$\int_C (\underline{r} - \underline{m}) f(\underline{0}) dV = 0, \text{ by symmetry,}$$

$$\int_C (\underline{m} - \underline{r}) (\underline{r}, \nabla f(\underline{0})) dV =$$

$$\int_C (\underline{m} - \underline{r}) (\underline{r} - \underline{m}, \nabla f(\underline{0})) dV + \int_C (\underline{m} - \underline{r}) (\underline{m}, \nabla f(\underline{0})) dV.$$

The second summand is again zero. For the first, we consider the j -th component and get

$$\begin{aligned} & \int_C (x_j - m_j) \sum_{i=1}^p (x_i - m_i) f'_i(\underline{0}) dV = \\ & = \int_C \sum_{i \neq j} (x_j - m_j) (x_i - m_i) f'_i(\underline{0}) dV + \int_C (x_j - m_j)^2 f'_j(\underline{0}) dV \\ & = \sum_{i \neq j} \frac{(x_j - m_j)^2}{2} \left| \begin{array}{c} \Delta x_j^2 \\ \Delta x_j^1 \end{array} \right| \frac{(x_i - m_i)^2}{2} \left| \begin{array}{c} \Delta x_i^2 \\ \Delta x_i^1 \end{array} \right| f'_i(\underline{0}) \prod_{k \neq i, j} \Delta x_k \\ & + \frac{(x_j - m_j)^3}{3} \left| \begin{array}{c} \Delta x_j^2 \\ \Delta x_j^1 \end{array} \right| \prod_{k \neq j} \Delta x_k f'_j(\underline{0}). \end{aligned}$$

Here, Δx_j^1 (Δx_j^2) stands for the lower (upper) extremes of the j -th side of C .

The first sum is zero and the last term is equal to:

$$f'_j(\underline{0}) \frac{\Delta x_j^3}{12} \prod_{k \neq j} \Delta x_k = |c| \frac{f'_j(\underline{0})}{12} \Delta x_j^2.$$

The absolute value of the j-th component of the third term is

$$\left| \int_C (x_j - m_j) \circ(|\underline{x}|) dV \right| \leq \Delta x_j \circ \left(\sqrt{(\Delta x_1)^2 + \dots + (\Delta x_p)^2} \right) |c|.$$

So, we have:

$$\frac{\overset{\rightarrow}{(MS)}_j}{|c|^{2/p}} = \frac{\frac{1}{12} |c| f'_j(\underline{0}) \Delta x_j^2 + \left(\int_C (\underline{x} - \underline{m}) \circ(|\underline{x}|) dV \right)_j}{|c|^{2/p} \left[|c| f(\underline{0}) + |c| (m, \nabla f(\underline{0})) + \int_C \circ(|\underline{x}|) dV \right]}.$$

The quotient

$$\frac{\Delta x_j^2}{|c|^{2/p}} = \frac{\Delta x_j^2 / \prod_{i=1}^p \Delta x_i^2}{\prod_{i=1}^p \left(\Delta x_i / \sqrt{\sum_{k=1}^p \Delta x_k^2} \right)^{2/p}} \longrightarrow \frac{a_j^2}{\prod_{i=1}^p a_i^{2/p}}$$

by (1.1).

For the second term in the numerator, we consider

$$\frac{\left| \int_C (x_j - m_j) \circ(|\underline{x}|) dV \right|}{|c|^{(2/p) + 1}} \leq \frac{\Delta x_j \circ \left(\sqrt{\Delta x_1^2 + \dots + \Delta x_p^2} \right)}{|c|^{2/p}} =$$

$$\frac{\frac{\Delta x_j}{\sqrt{\sum \Delta x_i^2}} \cdot O(1)}{\prod_{k=1}^p \left(\frac{\Delta x_k}{\sqrt{\sum x_i^2}} \right)^{2/p}}.$$

and this tends to zero.

In the denominator, the second term tends to zero since $m \rightarrow 0$ and the last term tends obviously to zero.

REMARK. If C is a p -dimensional cube, i.e.

$$\Delta x_i = \Delta x_j \quad i, j = 1, \dots, p,$$

then

$$\frac{a_j^2}{\prod_{i=1}^p a_i^{2/p}} = 1,$$

since, in this case, $a_i = 1 / \sqrt{p}$, $i = 1, 2, \dots, p$.

Therefore

$$\frac{\overline{MS}}{|C|^{2/p}} \longrightarrow \frac{1}{12} \frac{\nabla f(P_0)}{f(P_0)}$$

For $p = 1$, this result agrees with that of Vincze.

2. THE ESTIMATION OF $\nabla f(p_0) / f(p_0)$.

Let now $f(P)$ be a density function in the p -dimensional euclidean space.

As a first application of Lemma 1.1 we will show how to estimate the quotient $\nabla f / f$ in a point $P_0 = P_0(x_1^0, x_2^0, \dots, x_p^0)$, where the conditions of the lemma are fulfilled for $f(P_0)$.

Let us suppose that X_1, X_2, \dots, X_n is a random sample taken from a population with density function $f(P)$.

If $C = C_n$ is a parallelepiped containing P_0 , with \underline{m} as its middle point, we denote by $X_{1C}, X_{2C}, \dots, X_{vC}$ the elements of the sample lying in C_n and define by

$$\bar{X}_C = \frac{\sum_{i=1}^v X_{iC}}{v}$$

their sample mean.

The parallelepiped C depends on the sample size n in such a way that the volume of C , $|C|$ tends to zero as $n \rightarrow \infty$. The exact way of dependence will be made precise in the following theorem.

If $\underline{a} = (a_1, a_2, \dots, a_p)$ and $\underline{b} = (b_1, b_2, \dots, b_p)$ are two p -dimensional vectors, we put

$$\underline{a} * \underline{b} = (a_1 b_1, a_2 b_2, \dots, a_p b_p).$$

Let us define the vector

$$\underline{b} = \left(\frac{\prod a_i^{2/p}}{a_1^{2/p}}, \frac{\prod a_i^{2/p}}{a_2^{2/p}}, \dots, \frac{\prod a_i^{2/p}}{a_p^{2/p}} \right)$$

where a_1, a_2, \dots, a_p are the constants defined in (1.1).

THEOREM 2.1. Let X be a random vector with density function $f(x_1, x_2, \dots, x_p)$ positive and differentiable at the point P_0 , and let X_1, X_2, \dots, X_n be a random sample for X .

The estimator

$$T_n = \frac{\bar{X}_C - m}{\frac{1}{12} |C|^{2/p}} * \underline{b}$$

a) converges stochastically to $\nabla f(P_0) / f(P_0)$ if

$$|C_n|^{(2/p)+1} n \rightarrow \infty \text{ when } n \rightarrow \infty.$$

b) converges with probability 1 to

$$\nabla f(P_0) / f(P_0)$$

if

$$|C_n| = n^{-\frac{\epsilon p}{p+2}} \text{ when } n \rightarrow \infty, 0 < \epsilon < 1.$$

For the proof of this theorem we will need the following

LEMMA 2.1. Let f be the density function of a random variable X with values on an interval of length Δ . If $\mu = EX$ and X_1, X_2, \dots, X_ν is a random sample of X , then

$$E[(\bar{X} - \mu)^s] = O(\Delta^s / \nu^{s/2}) \quad (2.2)$$

For even s , ν tending to infinity and $\Delta \rightarrow 0$.

PROOF.

$$\begin{aligned} (\bar{X} - \mu)^s &= \frac{\left[\sum_{i=1}^{\nu} (X_i - \mu) \right]^s}{\nu^s} = \\ &= \nu^{-s} \sum_S \binom{s}{t_1 \dots t_\nu} \prod_{i=1}^{\nu} (X_i - \mu)^{t_i} \end{aligned}$$

where $S = \{(t_1, \dots, t_\nu) : \sum t_i = s, t_i \text{ nonnegative integers}\}$.

Since the X_i 's are independent, we have

$$E[(\bar{X} - \mu)^s] = \nu^{-s} \sum_S \binom{s}{t_1 \dots t_\nu} \prod_{i=1}^{\nu} E(X_i - \mu)^{t_i}$$

If $t_i = 1$, $E(X_i - \mu)^{t_i} = 0$ and for $t_i \neq 1$

$$E(X_i - \mu)^{t_i} = O(\Delta^{t_i}).$$

Therefore

$$\prod_{i=1}^{\nu} E(X_i - \mu)^{t_i} = O(\Delta^s).$$

The multinomial coefficient

$$\binom{s}{t_1 \dots t_v} \leq s! ,$$

so it remains to be calculated the number of terms in the sum which are different from zero, i.e. the cardinal of the set

$$A_s = \{(t_1, \dots, t_v) : \sum_{i=1}^v t_i = s, t_i \neq 1, t_i \text{ nonnegative integers}\}.$$

If we consider the function

$$g(x) = (1 + x^2 + x^3 + \dots)^v = \sum_{k=0}^{\infty} a_k x^k$$

the coefficients a_k are the cardinals of the sets A_k .

But

$$\begin{aligned} g(x) &= \left(1 + \frac{x^2}{1-x}\right)^v = \sum_{k=0}^v \binom{v}{k} x^{2k} \frac{1}{(1-x)^k} = \\ &= \sum_{k=0}^v \binom{v}{k} x^{2k} \sum_{r=0}^{\infty} \binom{-k}{r} (-1)^r x^r = \\ &= \sum_{k=0}^v \sum_{r=0}^{\infty} \binom{v}{k} \binom{-k}{r} (-1)^r x^{r+2k} , \end{aligned}$$

therefore

$$a_s = \sum \binom{v}{k} \binom{-k}{r} (-1)^r$$

where the sum is taken over all the values of k and r such that: $0 \leq k \leq v$, $r \geq 0$ and $2k + r = s$.

Calling $s' = s/2$ we can write $r = 2(s' - k)$, and

$$a_s = \sum_{k=0}^{s'} (-1)^{2(s'-k)} \binom{v}{k} \binom{-k}{2(s'-k)} = \sum_{k=0}^{s'} \binom{v}{k} \binom{-k}{2(s'-k)}.$$

The coefficient

$$\binom{v}{k} = 0 \binom{v^{s/2}}{k}, \text{ for } k = 0, 1, \dots, s'.$$

Furthermore

$$\sum_{k=0}^{s'} \binom{-k}{2(s'-k)} = \sum_{k=0}^{s'} \binom{2s'-k-1}{k-1}$$

does not depend on v .

Then

$$a_s = 0 \binom{v^{s/2}}{k}.$$

and the proof of the lemma is complete.

PROOF: (of theorem (2.1)).

$$\left| \left(T_n - \frac{Vf(P_0)}{f(P_0)} \right)_j \right| = \left| \frac{\bar{X}_{C_j} - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_0)}{f(P_0)} \right| =$$

$$\left| \frac{\bar{X}_{Cj} - E_C X_j}{\frac{1}{12} |c|^{2/p}} b_j + \frac{E_C X_j - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_0)}{f(P_0)} \right| < \frac{1}{12} |c|^{2/p} b_j + \frac{E_C X_j - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_0)}{f(P_0)}$$

$$\left| \frac{\bar{X}_{Cj} - E_C X_j}{\frac{1}{12} |c|^{2/p}} b_j \right| + \left| \frac{E_C X_j - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_0)}{f(P_0)} \right|. \quad (2.3)$$

Here, the second term tends to zero, by the lemma (1.1).

Let $E_C U$ denote the conditional expectation of U , given $U \in C$.

By lemma 2.1

$$E_C \left[\left(\frac{\bar{X}_{Cj} - E_C X_j}{\frac{1}{12} |c|^{2/p}} \right)^2 \middle| v = v_0 \right] = 0 \left(\frac{\Delta_j^2}{v_0 |c|^{4/p}} \right). \quad (2.4)$$

If we take now the expectation with respect to v we need

$$E^* \left(\frac{1}{v} \right) = E \left(\frac{1}{v} \mid v \neq 0 \right).$$

This can be estimated with the following result:

LEMMA 2.2. If Y is a binomial random variable with parameters p and n then, for a positive integer s

$$E^* \left(\frac{1}{Y^s} \right) = E \left(\frac{1}{Y^s} \mid Y \neq 0 \right) = 0 \left(\frac{1}{(EY)^s} \right)$$

for $np = EY \rightarrow \infty$.

PROOF:

$$E^* \left(\frac{1}{Y^s} \right) = \frac{\sum_{k=1}^n \binom{n}{k} \frac{1}{k^s} p^k (1-p)^{n-k}}{1 - (1-p)^n} <$$

$$< \frac{\text{const.} \sum_{k=1}^n \binom{n}{k} \frac{1}{(k+1)(k+2)\dots(k+s)} p^k (1-p)^{n-k}}{1 - (1-p)^n} \quad (2.5)$$

since $1/k < C_1/(k+1)$ for some constant C_1 .

The sum in (2.5) is equal to:

$$\sum_{k=1}^n \binom{n+s}{k+s} \frac{p^{k+s} (1-p)^{n-k}}{(n+1)(n+2)\dots(n+s)p^s} < \frac{1}{n^s p^s} = \frac{1}{(EY)^s} .$$

On the other hand:

$$(1-p)^n \approx e^{-np} \rightarrow 0 \text{ as } np \rightarrow \infty$$

Therefore

$$E^* \left(\frac{1}{Y^s} \right) < (\text{const.})' \frac{1}{(EY)^s} .$$

Expectation with respect to v in (2.4) gives

$$E_C \left[\left(\frac{\bar{X}_{C_1} - E_C X_1}{\frac{1}{12} C^{2/p}} \right)^2 \right] = 0 \left(\frac{\Delta_1^2}{n |C|^{1+(4/p)}} \right) = 0 \left(\frac{1}{n |C|^{1+(2/p)}} \right)$$

since v has a binomial distribution with parameter n

and $P[X \in C] = f(P_0) |C| + o(|C|)$ and $n \cdot P[X \in C] \rightarrow \infty$ for both hypothesis in the statement of the theorem.

By the Markov's inequality the first summand in (2.3) tends stochastically to zero as $n \rightarrow \infty$.

The convergence with probability one can be achieved using lemma (2.1):

$$P\left(\left|\frac{\bar{X}_{Cj} - E_C X_{1j}}{\frac{1}{\sqrt{12}} |C|^{2/p}} b_j\right| > \beta\right) \leq E_C \left[\left(\frac{\bar{X}_{Cj} - E_C X_{1j}}{\frac{1}{\sqrt{12}} |C|^{2/p}} b_j\right)^s \right] \frac{1}{\beta^s}$$

for any even positive integer s .

But the right member is

$$O\left(\frac{1}{n^{s/2} |C|^s \frac{2+p}{2p}}\right)$$

by lemmas (2.1) and (2.2).

Using the statement b/ of the theorem, we get a power $\frac{s}{2} (1-\epsilon)$ of n . Choosing s in such a way that $\frac{s}{2} (1-\epsilon) = 1 + \alpha$ for $\alpha > 0$ the convergence with probability one follows from the Borel-Cantelli lemma.

3. ESTIMATION OF THE INFORMATION

In the case of a p -dimensional random vector X we can use Lemma 1.1 to estimate the information matrix or, more generally, to estimate

$$E \psi \left(\frac{\nabla f}{f} \right)$$

where f is the density function of X and ψ is a function defined on R^p .

We will consider the case in which f has a finite rectangular support C .

Let us take a partition of size N of C as follows:

- a) the elements C_i of the partition are all equals with sides Δx_j , $j=1, 2, \dots, p$, for each i .
- b) If $N \rightarrow \infty$

$$\frac{\Delta x_j}{(\Delta x_1^2 + \dots + \Delta x_p^2)^{1/2}} \rightarrow a_j, \quad j=1, \dots, p \quad (3.1)$$

where the a_j are real constants such that $0 < a_j < 1$

and

$$\sum_{j=1}^p a_j^2 = 1.$$

If the sample size is n , let v_i be the number of sample

elements which lie in C_i , \bar{X}_i the sample mean of the elements in C_i , m_i the middle point of C_i and $E_i = E(X / X \in C_i)$.

THEOREM 3.1. Let f be a p -dimensional density function differentiable in C . The function $\psi : (R^p \rightarrow R)$ will be assumed twice differentiable with

$$\left| \frac{\partial^2 \psi(r)}{\partial x_i \partial x_j} \right| < T \quad i, j = 1, \dots, p.$$

Then:

$$\sum_{i=1}^n \psi \left(\frac{x_i - m_i}{\frac{1}{\sqrt{2}} |C_i|^{2/p}} * \underline{b} \right) \rightarrow E \left[\psi \left(\frac{\nabla f}{f} \right) \right] \quad (3.2)$$

stochastically if n and N tend to infinity and

$$N = o(n^{p/(p+2)}).$$

Here, \underline{b} is the vector with the components $b_i = \frac{P}{\prod_{j=1}^p a_j^{2/p}} / a_i^2$

and the notation $\underline{a} * \underline{b} = (a_1 b_1, a_2 b_2, \dots, a_p b_p)$ is used.

PROOF. If $\nabla^2 \psi$ denotes the matrix of the second derivatives of ψ we may write:

$$\psi(r) = \psi(r_0) + (\nabla \psi(r_0) \cdot \Delta r) + \Delta r^T \nabla^2 \psi(r_0 + \theta \Delta r) \Delta r \quad (3.3)$$

where a^T means the transpose of a , $\Delta r = r - r_0$, $0 < \theta < 1$.

We apply (3.3) to the following vectors:

$$r = r_i = \frac{\bar{X}_i - \frac{m_i}{n}}{\frac{1}{n} |C_i|^{2/p}} * \underline{b}$$

$$r = r_{oi} = \frac{E_i - \frac{m_i}{n}}{\frac{1}{n} |C_i|^{2/p}} * \underline{b}, \text{ and}$$

$$\Delta r = \Delta r_i = \frac{\bar{X}_i - E_i}{\frac{1}{n} |C_i|^{2/p}} * \underline{b}.$$

If $P_i = \int_{C_i} f(r) dV$, we know by the definition of the integral and Lemma 1.1 that

$$\sum_{i=1}^N \psi \left(\frac{E_i - \frac{m_i}{n}}{\frac{1}{n} |C_i|^{2/p}} * \underline{b} \right) P_i \longrightarrow E \left[\psi \left(\frac{\nabla f}{f} \right) \right]$$

if $N \rightarrow \infty$ and (3.1) holds.

Therefore, we have to prove that

$$\begin{aligned} & \sum_{i=1}^N \psi(r_{oi}) \left(\frac{v_i}{n} - P_i \right) + \sum_{i=1}^N \left(\nabla \psi(r_{oi}) \cdot \Delta r_i \right) \frac{v_i}{n} + \\ & \sum_{i=1}^N \Delta r_i^T \nabla^2 \psi(r_{oi} + \theta \Delta r_i) \frac{v_i}{n} \Delta r_i \end{aligned} \quad (3.4)$$

tends to zero.

The vector r_{oi} is bounded (Lemma 1.1), therefore $|\psi(r_{oi})| < \alpha$ for N large enough.

Using the Cauchy inequality:

$$\left[\sum_{i=1}^N \left(\frac{v_i}{n} - p_i \right) \right]^2 = \left(\sum_{i=1}^N \frac{v_i - np_i}{n\sqrt{p_i}} \sqrt{p_i} \right)^2 \leq$$

$$\leq \frac{1}{n} \sum_{i=1}^N \frac{(v_i - np_i)^2}{np_i} \quad \sum p_i = \frac{1}{n} \chi_{n,N}^2$$

and this tends stochastically to zero, if n and N tends to infinity.

$\nabla\psi(r_{oi})$ is also a bounded vector for N large enough, because of the continuity of $\partial\psi/\partial x_j$, $j=1, \dots, p$ and the boundedness of r_{oi} . Therefore:

$$\left| \sum_{i=1}^N (\nabla\psi(r_{oi}) \cdot \Delta r_i) \frac{v_i}{n} \right| \leq \beta \sum_{i=1}^N \sum_{j=1}^p |\Delta r_{ij}| \frac{v_i}{n}$$

where Δr_{ij} is the j -th component of Δr_i .

Again, the Cauchy's inequality leads to:

$$\begin{aligned}
& \left(\sum_{i=1}^N \sum_{j=1}^p |\Delta r_{ij}| \sqrt{\frac{v_i}{n}} \sqrt{\frac{v_i}{n}} \right)^2 \leq \sum_{i=1}^N \left(\sum_{j=1}^p |\Delta r_{ij}| \right)^2 \frac{v_i}{n} \sum \frac{v_i}{n} \leq \\
& \leq \sum_{i=1}^N \sum_{j=1}^p \left(\Delta r_{ij} \right)^2 \sum_{j=1}^p 1 \cdot \frac{v_i}{n} = p \sum_{i=1}^N \sum_{j=1}^p \left(\frac{\bar{X}_{ij} - E_{ij}}{\frac{1}{12} |C_i|^{2/p} b_j} \right)^2 \frac{v_i}{n} .
\end{aligned}
\tag{3.5}$$

The expectation of this expression is

$$p \sum_{i=1}^N \sum_{j=1}^p b_j^2 \frac{\sigma_{ij}^2 / v_i}{\frac{1}{12^2} |C_i|^{4/p}} \frac{v_i}{n}
\tag{3.6}$$

where σ_{ij}^2 denotes the conditional variance of the j -th component of X given that $x \in C_i$.

So

$$\sigma_{ij}^2 = 0 \left(\Delta x_j^2 \right) .$$

By the assumptions on the partition of C :

$$|C_i| = \frac{|C|}{N} \text{ and } \Delta x_j \approx |C_i|^{1/p} = \left(\frac{|C|}{N} \right)^{1/p} ,$$

therefore, (3.6) is

$$0 \left(\frac{\frac{p+2}{N^p}}{n} \right) .$$

Then the second term of (3.4) tends to zero in probability

if $N = o(n^{p/(2+p)})$ by the Markov inequality.

The assumptions about ψ makes the absolute value of the third term less than or equal to

$$T \sum_{i=1}^N \left(\sum_{j=1}^p |\Delta r_{ij}|^2 \right) \frac{v_i}{n} \leq T \sum_{i=1}^N \sum_{j=1}^p (\Delta r_{ij})^2 \frac{v_i}{n} \sum_{j=1}^p 1$$

and this is, except for the constants, equal to (3.5).

NOTE. For $p=1$, this result agrees with that of Csáki and Vincze.

COROLLARY 3.1. Let X be a p -dimensional random vector with density function $f(x_1 - \theta, \dots, x_p - \theta)$ positive on a p -dimensional cube C . Let X_1, \dots, X_n be a random sample of X . If f verifies the hypothesis of the theorem, the estimate of the information

$$I(\theta) = E \left(\frac{\partial \lg f(x)}{\partial \theta} \right)^2$$

can be

$$\sum_{i=1}^N \left(\sum_{j=1}^p \frac{\bar{x}_{ij} - m_{ij}}{\frac{1}{12} |c_i|^{2/p} b_j} \right)^2 \frac{v_i}{n} .$$

This estimator is asymptotically consistent in probability if

$$N = o(n^{p/(p+2)}).$$

(Here, we use the same notation as in the theorem).

PROOF. It suffices to take $\psi(\underline{x}) = \left(\sum_{j=1}^p x_j \right)^2$ because the information is the expectation of

$$\left(\frac{\partial \lg f}{\partial \theta} \right)^2 = \left(\sum \frac{\partial \lg f}{\partial x_j} \frac{dx_j}{d\theta} \right)^2 = \left(\sum_j \frac{f'_j}{f} \right)^2.$$

The result now follows from the theorem.

In the case of a multidimensional location parameter

$$\theta = (\theta_1, \dots, \theta_p)$$

the information is a matrix with general term

$$I_{ij} = \int \dots \int_{R^p} \frac{\partial \lg f}{\partial \theta_i} \frac{\partial \lg f}{\partial \theta_j} f \, dV \quad i, j=1, \dots, p.$$

COROLLARY 3.2. Let X be a p -dimensional random vector with density function

$$f(x_1 - \theta_1, x_2 - \theta_2, \dots, x_p - \theta_p).$$

Using the same notation as in the theorem, if f verifies

the above hypothesis, then an estimator of the general term I_{jk} in the matrix information can be

$$\sum_{i=1}^N \left(\frac{\bar{X}_{ij} - m_{ij}}{\frac{1}{12} |c_i|^{2/p}} b_j \cdot \frac{\bar{X}_{ik} - m_{ik}}{\frac{1}{12} |c_i|^{2/p}} b_k \right) \frac{v_i}{n}$$

This estimation is asymptotically consistent in probability, if

$$N = o(n^{p/(2+p)}).$$

PROOF. Taking $\psi(\underline{x}) = x_j x_k$ the corollary follows easily from the theorem.

4. THE LEARNING MODEL.

We consider now a learning model, i.e. we repeat an experiment n -times independently and get the values of a random variable

$$X_1, X_2, \dots, X_n$$

and the corresponding values of a parameter

$$\theta_1, \theta_2, \dots, \theta_n,$$

The X_i 's may be the values of a random variable as well as a statistic obtained in the i -th experiment.

Let $f(x/\theta)$ be the density function of X given the parameter θ . The information of X for θ is:

$$\begin{aligned} I(\theta) &= \int \left(\frac{\partial \lg f(x/\theta)}{\partial \theta} \right)^2 f(x/\theta) dx = \\ &= \int \left(\frac{f'_\theta(x/\theta)}{f(x/\theta)} \right)^2 f(x/\theta) dx \quad (4.1) \end{aligned}$$

Our aim is to estimate this information if we have the values of the learning model:

$$(X_1, \theta_1), (X_2, \theta_2), \dots, (X_n, \theta_n).$$

We denote by $f(x, \theta)$ and $\phi(\theta)$ the joint density of (X, θ) and the marginal density of θ (or a priori density) respectively.

If we write

$$f(x, \theta) = f(x/\theta)\phi(\theta)$$

differentiating with respect to θ , we obtain:

$$f'_\theta(x, \theta) = f'_\theta(x/\theta)\phi(\theta) + f(x/\theta)\phi'(\theta)$$

and

$$\frac{f'_\theta(x, \theta)}{f(x, \theta)} = \frac{f'_\theta(x/\theta)}{f(x/\theta)} + \frac{\phi'(\theta)}{\phi(\theta)}$$

Finally

$$\begin{aligned} \left(\frac{f'_\theta(x, \theta)}{f(x, \theta)} \right)^2 &= \left(\frac{f'_\theta(x/\theta)}{f(x/\theta)} \right)^2 + \left(\frac{\phi'(\theta)}{\phi(\theta)} \right)^2 - \\ &\quad - 2 \frac{f'_\theta(x, \theta)}{f(x, \theta)} \frac{\phi'(\theta)}{\phi(\theta)}. \end{aligned}$$

Therefore, an estimation of the information function (4.1) can be obtained by estimating each of the following terms:

$$\int \left(\frac{f'_\theta(x, \theta)}{f(x, \theta)} \right)^2 f(x/\theta) dx + \left(\frac{\phi'(\theta)}{\phi(\theta)} \right)^2 -$$

$$- 2 \frac{\phi'(\theta)}{\phi(\theta)} \int \frac{f_{\theta}(x, \theta)}{f(x, \theta)} f(x/\theta) dx . \quad (4.2)$$

Let us take a fixed θ_0 and a neighbourhood of it

$$J = (\theta_0 - \frac{\delta}{2}, \theta_0 + \frac{\delta}{2}) .$$

We assume that $f(x, \theta)$ and $f(x/\theta)$ are positive only on a finite interval (a, b) .

Let I_1, I_2, \dots, I_N be a partition of (a, b) with length

$$I_1 = \frac{b - a}{N} = \epsilon .$$

v_{1J} will stand for the number of elements of the learning model in $I_1 \times J$.

v_J will be the number of elements of the learning model in $(a, b) \times J$.

$\bar{\theta}_{1J}$ denote the sample mean of the θ 's in $I_1 \times J$ and $\bar{\theta}_J$ is the sample mean of the θ 's in J .

THEOREM 4.1. Let $f(x, \theta)$, $f(x/\theta)$ and $\phi(\theta)$ be unknown densities with continuous first derivatives $\forall x \in (a, b)$ and $\theta \in J$.

Then the information function defined in (4.2) can be estimated by

$$\sum \left(\frac{\bar{\theta}_{1J} - \theta_0}{\frac{1}{12} \epsilon \delta b_\theta} \right)^2 \frac{v_{1J}}{v_J} + \left(\frac{\bar{\theta}_J - \theta_0}{\frac{1}{12} \delta^2} \right)^2 -$$

$$- 2 \frac{\bar{\theta}_J - \theta_0}{\frac{1}{12} \delta^2} \sum_{i=1}^N \frac{\bar{\theta}_{1J} - \theta_0}{\frac{1}{12} \delta \epsilon b_\theta} \frac{v_{1J}}{v_J} \quad (4.3)$$

This estimation is asymptotically consistent in probability if $N = o(n^{1/4})$ and $\delta \sim c/N$ and with probability one if

$$N \approx k n^{(1/4)-\rho}, \quad 0 < \rho < 1/4. \quad \text{Here,}$$

$$b_\theta = \frac{a_\theta^2}{a_x a_\theta} \quad \text{where } a_\theta = \lim_{\epsilon, \delta \rightarrow 0} \frac{\delta}{\sqrt{\delta^2 + \epsilon^2}} \quad \text{and}$$

$$a_x = \lim_{\epsilon, \delta \rightarrow 0} \frac{\epsilon}{\sqrt{\epsilon^2 + \delta^2}} \quad \text{and } a_\theta^2 + a_x^2 = 1. \quad (4.4)$$

PROOF. First we consider the term

$$\int \left(\frac{f_\theta(x, \theta_0)}{f(x, \theta_0)} \right)^2 f(x/\theta_0) dx \quad \text{of } (4.2)$$

and its estimator

$$\sum_{i=1}^N \left(\frac{\bar{\theta}_{1J} - \theta_0}{\frac{1}{12} \epsilon \delta b_\theta} \right)^2 \frac{v_{1J}}{v_J}.$$

As before, we write:

$$\begin{aligned} (\bar{\theta}_{iJ} - \theta_o)^2 &= (\bar{\theta}_{iJ} - E_{iJ}^{\theta} + E_{iJ}^{\theta} - \theta_o)^2 = \\ &= (\bar{\theta}_{iJ} - E_{iJ}^{\theta})^2 + (E_{iJ}^{\theta} - \theta_o)^2 + \\ &+ 2(\bar{\theta}_{iJ} - E_{iJ}^{\theta})(E_{iJ}^{\theta} - \theta_o). \end{aligned}$$

If the notation

$$P_{i/o} = P[X \in I_i / \theta_o] = \int_{I_i} f(x/\theta_o) dx$$

$$E_{iJ}^{\theta} = E(\theta | (X, \theta) \in I_i \times J)$$

is used, we know that

$$\sum_{i=1}^N \left(\frac{E_{iJ}^{\theta} - \theta_o}{\frac{1}{\sqrt{2}} \in \delta b_{\theta}} \right)^2 P_{i/o} \rightarrow \int \left(\frac{f'_{\theta}(x, \theta_o)}{f(x, \theta_o)} \right)^2 f(x/\theta_o) dx.$$

Therefore, we have to prove that the following terms tend to zero:

$$\sum_{i=1}^N \left(\frac{E_{iJ}^{\theta} - \theta_o}{\frac{1}{\sqrt{2}} \in \delta b_{\theta}} \right)^2 \left(\frac{v_{iJ}}{v_J} - P_{i/o} \right) \quad (4.5)$$

$$\sum_{i=1}^N \left(\frac{\bar{\theta}_{iJ} - E\theta_{iJ}}{\frac{1}{12} \in \delta b_{\theta}} \right)^2 \frac{v_{iJ}}{v_J} \quad (4.6)$$

$$\sum_{i=1}^N \frac{(\bar{\theta}_{iJ} - E_{iJ}\theta) (E_{iJ}\theta - \theta_0)}{\left(\frac{1}{12} \in \delta b_{\theta} \right)^2} \frac{v_{iJ}}{v_J} \quad (4.7)$$

The expression

$$\left(\frac{E_{iJ}\theta - \theta_0}{\frac{1}{12} \in \delta b_{\theta}} \right)^2$$

is bounded for \in and δ small enough because of Lemma 1.1.

and

$$\begin{aligned} \sum_{i=1}^N \left| \frac{v_{iJ}}{v_J} - p_{i/o} \right| &= \sum_{i=1}^N \left| \frac{v_{iJ}}{v_J} - p_{i/J} + p_{i/J} - p_{i/o} \right| \leq \\ &\leq \sum_{i=1}^N \left| \frac{v_{iJ}}{v_J} - p_{i/J} \right| + \sum_{i=1}^N \left| p_{i/J} - p_{i/o} \right| \end{aligned} \quad (4.8)$$

with $p_{i/J} = P [X \in I_i / \theta \in J]$.

For the first sum, we have:

$$\left(\sum_{i=1}^N \left| \frac{v_{iJ}}{v_J} - p_{i/J} \right| \right)^2 \leq \frac{1}{v_J} \sum_{i=1}^N \frac{(v_{iJ} - v_J p_{i/J})^2}{v_J p_{i/J}} \sum_{i=1}^N p_{i/J}$$

$$= \frac{1}{v_J} \chi_{v_J, N}^2.$$

But

$$E \left[\left(\frac{1}{v_J} \chi_{v_J, N}^2 \right)^2 \mid v_J = t \right] = 0 \left(\frac{N^2}{t^2} \right).$$

then, the expectation with respect to v_J , lemma (2.2) and the hypothesis about δ yield

$$E^* \left[\left(\frac{1}{v_J} \chi_{v_J, N}^2 \right)^2 \right] = 0 \left(\frac{N^2}{n^2 \delta^2} \right) = 0 \left(\frac{N^4}{n^2} \right).$$

Therefore $\frac{1}{v_J} \chi_{v_J, N}^2$ tends to zero with probability one by Markov's inequality and Borel Cantelli lemma for both values of N in the hypothesis.

For the second sum in (4.8).

$$P_{1/J} = P[X \in I_1 \mid \theta \in J] = \frac{P[X \in I_1, \theta \in J]}{P[\theta \in J]} =$$

$$= \frac{\iint_{I_1 \times J} f(x, \theta) dx d\theta}{\int_J \phi(\theta) d\theta} =$$

$$= \frac{\iint_{I_1 \times J} [f(x, \theta_0) + f_\theta(x, \theta_0)(\theta - \theta_0) + o(|\theta - \theta_0|)] dx d\theta}{\int_J [\phi(\theta_0) + \phi'(\theta_0)(\theta - \theta_0) + o(|\theta - \theta_0|)] d\theta}$$

$$= \frac{\int_{I_1} \left[f(x/\theta_0) \phi(\theta_0) \delta + o(\delta^2) \right] dx}{\phi(\theta) \delta + o(\delta^2)} =$$

$$= \frac{\int_{I_1} f(x/\theta_0) dx + o(\delta \epsilon)}{1 + o(\delta)} .$$

Therefore, $|p_{i/J} - p_{i/O}| = o(\delta \epsilon)$, and

$$\sum_{i=1}^N |p_{i/J} - p_{i/O}| = N o(\delta \epsilon) = o(\delta).$$

So, (4.5) tends to zero with probability one if N and n tends to infinity.

In (4.6) we take expectation for a fixed value of $v_J \neq 0$ and get:

$$\sum_{i=1}^N \frac{\sigma_{iJ}^2}{\left(\frac{1}{i^2} \epsilon \delta b_\theta\right)^2} \frac{1}{v_J} \quad (4.9)$$

where $\sigma_{iJ}^2 = \text{Var}(\theta | (X, \theta) \in I_i XJ) = o(\delta)$.

By lemma (2.2)

$$E^* \left(\frac{1}{v_J} \right) = o \left(\frac{1}{n\delta} \right)$$

since $n\delta \sim cn/N \rightarrow \infty$ as $n \rightarrow \infty$ and $N = o(n^{1/4})$.

Therefore, the expected value of (4.6) is

$$O\left(\frac{N\delta}{\epsilon^2 \delta^2 n\delta}\right) = O\left(\frac{N^4}{n}\right)$$

and the stochastic convergence to zero of this term follows from Markov's inequality and the assumption on N .

For the third term (4.7) we can apply an analogous argument, because

$$\left| \sum_{i=1}^N \frac{(\bar{\theta}_{iJ} - E_{iJ}\theta)(E_{iJ}\theta - \theta_0)}{\left(\frac{1}{12} \epsilon \delta b_\theta\right)^2} \frac{v_{iJ}}{v_J} \right| \leq k \sum_{i=1}^N \frac{|\bar{\theta}_{iJ} - E_{iJ}\theta|}{\left(\frac{1}{12} \epsilon \delta b_\theta\right)} \frac{v_{iJ}}{v_J}$$

and the Cauchy's inequality gives:

$$\left(\sum_{i=1}^N \frac{|\bar{\theta}_{iJ} - E_{iJ}\theta|}{\frac{1}{12} \epsilon \delta b_\theta} \sqrt{\frac{v_{iJ}}{v_J}} \sqrt{\frac{v_{iJ}}{v_J}} \right)^2 \leq$$

$$\frac{\sum_{i=1}^N (\bar{\theta}_{iJ} - E_{iJ}\theta)^2}{\left(\frac{1}{12} \epsilon \delta b_\theta\right)^2} \frac{v_{iJ}}{v_J} \sum_{i=1}^N \frac{v_{iJ}}{v_J}$$

and now we follow as in (4.6).

For the second term in (4.2) it suffices to estimate $\phi'(\theta)/\phi(\theta)$.

Let us use the statistic

$$\frac{\bar{\theta}_J - \theta_0}{\frac{1}{12} \delta^2}.$$

But

$$\frac{\bar{\theta}_J - \theta_0}{\frac{1}{12} \delta^2} = \frac{\bar{\theta}_J - E_J \theta + E_J \theta - \theta_0}{\frac{1}{12} \delta^2} = \frac{\bar{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} + \frac{E_J \theta - \theta_0}{\frac{1}{12} \delta^2}$$

The second term converges to $\phi'(\theta_0)/\phi(\theta_0)$ if $\delta \rightarrow 0$, so we only need to prove the quotient $(\bar{\theta}_J - E_J \theta) / \frac{1}{12} \delta^2$ tends to zero.

As before, we can consider the following expectation, for fixed v_J :

$$\begin{aligned} E_{v_J} \left(\frac{\bar{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} \right)^2 &= \frac{\sigma_J^2 \theta}{\frac{1}{12^2} \delta^4 v_J} = \frac{O(\delta^2)}{\frac{1}{12^2} \delta^4 v_J} = \\ &= \frac{1}{v_J} O(1/\delta^2). \end{aligned}$$

Taking again expectation, with respect to v_J this time, and applying lemma 2.2 we get:

$$E \left(\frac{\bar{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} \right)^2 = O \left(\frac{1}{n \delta^3} \right) = O \left(\frac{N^3}{n} \right)$$

and the stochastic convergence follows easily.

For the last term in (4.2) we only need to prove that

$$\sum_{i=1}^N \frac{\bar{\theta}_{iJ} - \theta_0}{\frac{1}{12} \delta \in b_\theta} \frac{v_{iJ}}{v_J} \quad (4.10)$$

is a consistent estimator of

$$\int \frac{f'_\theta(x, \theta)}{f(x, \theta)} f(x/\theta) dx.$$

Writing $\bar{\theta}_{iJ} - \theta_0 = \bar{\theta}_{iJ} - E_{iJ}\theta + E_{iJ}\theta - \theta_0$, (4.10) can be

written in the following way:

$$\begin{aligned} & \sum_{i=1}^N \frac{E_{iJ}\theta - \theta_0}{\frac{1}{12} \delta \in \delta b_\theta} p_{i/0} + \sum_{i=1}^N \frac{E_{iJ}\theta - \theta_0}{\frac{1}{12} \delta \in b_\theta} \left(\frac{v_{iJ}}{v_J} - p_{i/0} \right) \\ & + \sum_{i=1}^N \frac{\bar{\theta}_{iJ} - E_{iJ}\theta}{\frac{1}{12} \delta \in b_\theta} \frac{v_{iJ}}{v_J}. \end{aligned}$$

The first sum converges to

$$\int \frac{f'_\theta(x, \theta)}{f(x, \theta)} f(x/\theta) dx$$

by the definition of the integral and lemma 1.1.

The second sum tends to zero with probability one for the same reasons as the term (4.5) does.

For the last term, the application of the Cauchy's inequality leads to (4.6).

This completes the proof of the stochastic convergence.

In order to obtain the convergence with probability one, we will use lemma 2.1.

Let s be an integer. Then: for fixed v_{iJ} , $i=1, \dots, N$

$$E_v \left(\sum_{i=1}^N (\bar{\theta}_{iJ} - E\theta_{iJ})^2 v_{iJ} \right)^2 =$$

$$E_v \left[\sum_{\substack{N \\ \sum_{i=1}^N t_i = s}} \binom{s}{t_1 \dots t_N} (\bar{\theta}_{1J} - E\theta_{1J})^{2t_1} v_{1J}^{t_1} \cdot \right.$$

$$\left. \cdot (\bar{\theta}_{2J} - E\theta_{2J})^{2t_2} v_{2J}^{t_2} \dots (\bar{\theta}_{NJ} - E\theta_{NJ})^{2t_N} v_{NJ}^{t_N} \right].$$

Since θ_{iJ} 's are stochastically independent, this is equal to

$$\sum_{\sum t_i = s} \binom{s}{t_1 \dots t_N} \prod_{i=1}^N E_v \left(\bar{\theta}_{iJ} - E\theta_{iJ} \right)^{2t_i} v_{iJ}^{t_i} =$$

$$= \sum_{t_i=s} \binom{s}{t_1 \dots t_N} \prod_{i=1}^N O\left(\frac{\delta^{2t_i}}{v_{iJ}^{t_i}}\right) v_{iJ}^{t_i} = O\left(\delta^{2s} N^s\right).$$

Then

$$\begin{aligned} E_V(A)^s &= E_V \left[\frac{\sum_{i=1}^N (\bar{\theta}_{iJ} - E\theta_{iJ})^2 v_{iJ}}{\left(\frac{1}{12} \epsilon \delta b_\theta\right)^2 v_J} \right]^s = \\ &= \frac{1}{v_J^s} O\left(\frac{\delta^{2s} N^s}{(\epsilon \delta)^{2s}}\right) = \frac{1}{v_J^s} O\left(\frac{N^s}{\epsilon^{2s}}\right) = \\ &= \frac{1}{v_J^s} O(N^{3s}). \end{aligned}$$

Therefore

$$E\left(E_V(A)^s\right) = O\left(\frac{N^{3s}}{n^s \delta^s}\right) = O\left(\frac{N^{4s}}{n^s}\right),$$

by lemma 2.2.

Therefore, if we choose an even s such that $(1+\alpha) = 4s\rho$ for a given $1/4 > \rho > 0$ and $\alpha > 0$, we have, by Markov's inequality:

$$\begin{aligned} P\left\{|A| > k\right\} &= P\left\{A^s > k^s\right\} \leq \frac{E(A^s)}{k^s} = O\left(\frac{N^{4s}}{n^s}\right) = \\ &= O\left(n^{-4\rho s}\right) = O\left(\frac{1}{n^{1+\alpha}}\right). \end{aligned}$$

REFERENCES

- 1) BHATTACHARYYA, P.K.: Estimation of a probability density function and its derivatives. Sankhya A 29 (1967), p. 373-382.
- 2) CSAKI, E., and VINCZE, I.: A lemma on real functions and its application in the nonparametric theory. Proceedings of the IASPS Symposium to the honour of Jerzy Neyman, Warsaw, (1975), p.83-92.
- 3) DMITRIEV, Yu., and TARASENKO, F.: On estimation of functionals of the probability density function and its derivatives. Teor. Verojatn. Primen. 18 (1973), p. 662-668.
- 4) PARZEN, E.: Estimation of density and mode. AMS 33 (1962), p. 1065-1076.
- 5) ROSENBLATT, M.: Remarks on some non-parametric estimation of a density function. AMS 27 (1956), p.832-837.
- 6) SINGH, R.S.: Nonparametric estimation of mixed partial derivatives of a multivariate density. Jour. Mult. Analysis 6 (1976), p. 111-122.
- 7) VINCZE, I.: Determination of distributions by means of their conditional expectations. (Hungarian). MTA III Oszt. Közl. 4 (1954), p. 513-523.