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

FOR

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 WINKER

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 $\nabla f(P_o) / f(P_o)$  
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:

\[ I(\theta) = \left[ \frac{2}{3\theta} \ln f(x - \theta) \right]^2 f(x - \theta) \, dx = \]

\[ = \left[ \frac{3}{3e} \ln f(x - \theta) \right] f(x - \theta) \, dx. \]

Later, Dmitriev-Tarasenko (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 Vineze, 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

\[
\frac{\mathcal{V}f(P_0)}{f(P_0)} = \begin{pmatrix} f_1(P_0) & f_2(P_0) & \cdots & f_n(P_0) \\ \frac{\partial}{\partial \theta} f(P_0) & \frac{\partial}{\partial \theta} f(P_0) & \cdots & \frac{\partial}{\partial \theta} f(P_0) \end{pmatrix}
\]

will be given too, where \( f_i(P_0) = \frac{\partial f}{\partial \theta_i}(P_0) \).
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 \( \Delta x_1, \Delta x_2, \ldots, \Delta x_p \) we can denote by \( |C| = \Delta x_1 \cdot \Delta x_2 \cdots \Delta x_p \) the volume of \( C \).

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

\[
\frac{\Delta x_i}{\sqrt[p]{\Delta x_j^2}} \rightarrow a_i^2, \quad i=1,\ldots,p \quad (1.1)
\]

with \( 0 < a_i < 1 \) and \( \sum_i a_i^2 = 1 \), then

\[
\frac{(M \cdot S)_j}{|C|^{2/p}} \rightarrow \frac{a_j^2}{\prod_{k=1}^{p} a_k^{2/p}} \cdot \frac{1}{|C|^{2/p}} \cdot \frac{f'(P_0)}{f(P_0)} , \quad j = 1, \ldots, p \quad (1.2)
\]

where \( (M \cdot S)_j \) is the \( j \)-th component of the vector \( (M \cdot S) \).
PROOF. Due to the assumption concerning $f$ we can write
\[ f(x) = f(x_0) + (x - x_0, \nabla f(x_0)) + o(|x - x_0|) \quad (1.3) \]

where $x$ is the position vector of a point $P$ in $C$ and $x_0$ is the position vector of $P_0$, $(\ldots)$ stands for the inner product.

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

If $m$ is the position vector of $M$, then
\[
\frac{\int_C (x - m) f(x) \, dV}{\int_C f(x) \, dV} = \frac{\int_C (x - m) [f(0) + (x, \nabla f(0)) + o(|x|)] \, dV}{\int_C [f(0) + (x, \nabla f(0)) + o(|x|)] \, dV}.
\]

Integration in the denominator gives:
\[
\int_C f(0) \, dV = f(0) |C|,
\]
\[
\int_C (x, \nabla f(0)) \, dV = \left( \int_C x \, dV, \nabla f(0) \right) = |C|(m, \nabla f(0)) \text{ and,}
\]

according to (1.1)
\[
\int_C o(|x|) \, dV \leq c \left( \sqrt{(dx_1)^2 + \ldots + (dx_d)^2} |C| = o(|C|^{1/d}) |C| \right).
\]
in the numerator, we have:

\[ \int_C (x - m)f(x) \, dv = 0, \text{ by symmetry,} \]

\[ \int_C (m - x) (x, \nabla f(0)) \, dv = \]

\[ \int_C (m - x) (x - m, \nabla f(0)) \, dv + \int_C (m - x) (m, \nabla f(0)) \, dv. \]

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

\[ \int_C (x_j - m_j) \sum_{j=1}^{p} (x_j - m_j) f_j'(0) \, dv = \]

\[ = \left[ \sum_{i \neq j} (x_j - m_j) (x_i - m_i) f_i'(0) \, dv + \int_C (x_j - m_j)^2 f_j'(0) \, dv \right] \]

\[ = \sum_{i \neq j} \frac{(x_j - m_j)^2}{2} \left| \begin{array}{c} \Delta x_j^2 \\ \Delta x_i^2 \\ \Delta x_k^1 \\ \Delta x_k^1 \\ \Delta x_k^1 \end{array} \right| f_j'(0) \left| \begin{array}{c} \Delta x_i^1 \\ \Delta x_k^1 \end{array} \right| \]

\[ + \frac{(x_j - m_j)^3}{3} \left| \begin{array}{c} \Delta x_j^2 \\ \Delta x_i^1 \\ \Delta x_k^1 \\ \Delta x_k^1 \end{array} \right| f_j'(0). \]

Here, \(\Delta x_j^1, \Delta x_j^2\) 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:

\[ f_j(0) \frac{\Delta x_j^3}{12} \sum_{i \neq j}^{p-1} \Delta x_k = |c| \frac{f_j(0)}{12} \Delta x_j^2. \]

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

\[ \left| \int_{C} (x_j - m_j) \circ \left| x \right| \, dv \right| \leq \Delta x_j \circ \left( \sqrt{\Delta x_1^2 + \ldots + \Delta x_p^2} \right) |c|. \]

So, we have:

\[ \frac{\Delta x_j}{|c|^{2/p}} = \frac{1}{12} \frac{|c|^2 f_j(0) \Delta x_j^2 + \left( \int_{C} (x - m) \circ \left( |x| \right) dv \right)}{|c|^{2/p} \left( |c| f(0) + |c| \int_{C} \circ \left( |x| \right) dv \right)}. \]

The quotient

\[ \frac{\Delta x_j^2}{|c|^{2/p}} = \frac{\phi \Delta x_j^2}{i=1} \frac{\Delta x_j^2}{|c|^{2/p}} \rightarrow \frac{s_j^2}{p} \]

by (1.1).

For the second term in the numerator, we consider

\[ \frac{\int_{C} (x_j - m_j) \circ \left| x \right| \, dv}{|c|^{(2/p) + 1}} \leq \frac{\Delta x_j \circ \left( \sqrt{\Delta x_1^2 + \ldots + \Delta x_p^2} \right)}{|c|^{2/p}} = \]

\[ = \frac{p}{n} \left( \frac{\Delta x_k^2}{\|x_j\|^2} \right)^{2/p}. \]
and this tends to zero.

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

**REMARK.** If \( C \) is a \( p \)-dimensional cube, i.e.

\[
\Delta x_i = \Delta x_j \quad i, j = 1, \ldots, p,
\]

then

\[
\frac{\sum_{i=1}^{p} a_i^2}{2/p} = 1,
\]

since, in this case, \( a_i = 1 / \sqrt{p} \), \( i = 1, 2, \ldots, p \).

Therefore

\[
\frac{\mathbb{E}^p}{|C|^{2/p}} \to \frac{1}{12} \frac{V^p(p_0)}{z(p_0)}
\]

For \( p = 1 \), this result agrees with that of Vincze.
2. THE ESTIMATION OF $V f (p_o) / f (p_o)$.

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 $V f / f$ is a point $P_0 = p_0(x_1^0, x_2^0, \ldots, x_p^0)$, where the conditions of the lemma are fulfilled for $f(p_0)$.

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

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

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

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 $n \to \infty$. The exact way of dependence will be made precise in the following theorem.

If $a = (a_1, a_2, \ldots, a_p)$ and $b = (b_1, b_2, \ldots, b_p)$ are two $p$-dimensional vectors, we put
\[ a \cdot b = (a_1b_1, a_2b_2, \ldots, a_nb_n). \]

Let us define the vector
\[ b = \left( \frac{a_1^{2/p} }{a_1^{2}}, \frac{a_2^{2/p} }{a_2^{2}}, \ldots, \frac{a_n^{2/p} }{a_n^{2}} \right) \]
where \( a_1, a_2, \ldots, a_n \) are the constants defined in (1.1).

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

The estimator
\[ T_n = \frac{X_n - m}{\frac{1}{n} |C|^{2/p}} \]
a) converges stochastically to \( \nabla f(P_0)/f(P_0) \) if
\[ |C_n|^{(2/p)+1} \xrightarrow{n \to \infty} \text{ when } n \to \infty. \]

b) converges with probability 1 to
\[ \nabla f(P_0)/f(P_0) \]
if
\[ |C_n| = n^{\frac{EP}{P^2}} \text{ when } n \to \infty, \quad 0 < \varepsilon < 1. \]

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 $\Delta$. If $\mu = E[X]$ and $X_1, X_2, \ldots, X_v$ is a random sample of $X$, then

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

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

PROOF.

$$(\bar{X} - \mu)^s = \frac{\sum_{i=1}^v (X_i - \mu)^s}{v^s} =$$

$$= v^{-s} \sum_{S} \left(\prod_{i=1}^v (X_i - \mu)^{t_i}\right)$$

where $S = \{(t_1, \ldots, t_v): \sum t_i = s, t_i \text{ nonnegative integers}\}$.

Since the $X_i$'s are independent, we have

$$E[(\bar{X} - \mu)^s] = v^{-s} \sum_{S} \left(\prod_{i=1}^v E(X_i - \mu)^{t_i}\right).$$

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

$$\sum_{i=1}^v E(X_i - \mu)^{t_i} = o(\Delta^s).$$
The multinomial coefficient

\[ \binom{s}{t_1 \ldots t_v} = \frac{s!}{t_1! \ldots t_v!} , \]

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

\[ A_s = \{ (t_1, \ldots, t_v) : \sum_{i=1}^v t_i = s, t_i \neq \lambda, t_i \text{ nonnegative integers} \} . \]

If we consider the function

\[ g(x) = (1 + x^2 + x^3 + \ldots)^v = \sum_{k=0}^\infty a_k x^k , \]

the coefficients \( a_k \) are the cardinals of the sets \( A_k \).

But

\[ g(x) = (1 + \frac{x}{1-x})^v = \sum_{k=0}^\infty \binom{v}{k} \frac{1}{(1-x)^k} x^k = \]

\[ = \sum_{k=0}^\infty \binom{v}{k} x^{2k} \sum_{r=0}^\infty (-1)^r \frac{r^k}{k!} , \]

therefore

\[ a_s = \sum_{k=0}^v \binom{v}{k} (-1)^r \binom{-k}{r} . \]
where the sum is taken over all the values of \( k \) and \( r \) such that: \( 0 \leq k \leq \nu, \ r \geq 0 \) and \( 2k + r = s \).

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

\[
a_s = \sum_{k=0}^{s'} (-1)^{2(s' - k)} \binom{\nu}{k} \binom{-k}{2(s' - k)} = \sum_{k=0}^{s'} \binom{\nu}{k} \binom{-k}{2(s' - k)}.
\]

The coefficient

\[
\binom{\nu}{k} = 0 \ (\nu^{n/2}), \text{ for } k = 0, 1, \ldots, s'.
\]

Furthermore

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

does not depend on \( \nu \).

Then

\[
a_s = 0 \ (\nu^{n/2}).
\]

and the proof of the lemma is complete.

**PROOF:** (of theorem (2.1)).

\[
\left| \frac{\nu f(p_0)}{f(p_0)} \right| = \frac{\nu c_j - m_j}{\frac{1}{12} - c_j^{2/p}} b_j = \frac{f_j(p_0)}{f(p_0)}.
\]
\[
\begin{align*}
\frac{\bar{X}_j - E_c X_j}{\frac{1}{12} |c|^{2/p}} b_j + \frac{E_c X_j - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_o)}{f(p_o)} &< \\
\frac{\bar{X}_j - E_c X_j}{\frac{1}{12} |c|^{2/p}} b_j + \frac{E_c X_j - m_j}{\frac{1}{12} |c|^{2/p}} b_j - \frac{f_j(P_o)}{f(p_o)}.
\end{align*}
\] (3.3)

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
\[
\delta_c \left[ \left( \frac{\bar{X}_j - E_c X_j}{\frac{1}{12} |c|^{2/p}} \right)^2 \right] v = v_o \Rightarrow 0 \leq \delta_o \left( \frac{\Delta^2}{v_o |c|^{4/p}} \right).
\] (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 \( np = EY \approx \).
PROOF:

\[ E \left( \frac{1}{Y^n} \right) = \frac{\sum_{k=1}^{n} \frac{n}{n} \left( \frac{n}{k} \right) \frac{1}{k^n} p^k (1 - p)^{n-k}}{1 - (1 - p)^n} < \]

\[
\text{const.} \frac{\sum_{k=1}^{n} \frac{n}{k} \left( \frac{1}{k+1} \right) \left( \frac{1}{k+2} \right) \cdots \left( \frac{1}{k+s} \right)}{1 - (1 - p)^n} p^k (1-p)^{n-k}
\]

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

The sum in (2.5) is equal to:

\[
\frac{\sum_{k=1}^{n+s} \frac{n+s}{k+s} \frac{1}{(k+1)(k+2) \cdots (k+s)} p^k (1-p)^{n-k}}{n^{n-p}} \frac{1}{(EX)^n} < \frac{1}{n^{n-p}} \frac{1}{(EX)^n}.
\]

On the other hand:

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

Therefore

\[ E \left( \frac{1}{Y^n} \right) < \text{const.} \frac{1}{(EX)^n}. \]

Expectation with respect to \(v\) in (2.4) gives

\[ E_C \left[ \left( \frac{E_C^1 - E_C Y^1}{\frac{1}{12} C^{2/p}} \right)^2 \right] = 0 \left( \frac{\Delta^2}{n |C|^{1+4/p}} \right) = 0 \left( \frac{1}{n |C|^{1+2/p}} \right) \]

since \(v\) has a binomial distribution with parameter \(n\).
and \( P[\{X \in C\} = f(P_o) | C| + o(|C|) \) and \( n \cdot P[\{X \in C\} = + \) 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 \( n \to \infty \).

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

\[
P \left( \left| \frac{X_{C1} - \mathbb{E}_C X_{C1}}{\frac{1}{12} |C|^{3/p}} \right| > \beta \right) \leq \mathbb{E}_C \left[ \left( \frac{X_{C1} - \mathbb{E}_C X_{C1}}{\frac{1}{12} |C|^{3/p}} \right)^{s} \right] \left( \frac{1}{\beta^s} \right)
\]

for any even positive integer \( s \).

But the right member is

\[
0 \left( \frac{1}{n^{s/2} |C|^{2p/2p}} \right)
\]

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

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

where $f$ is the density function of $X$ and $\psi$ is a function defined on $\mathbb{R}^p$.

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 equal with sides $\Delta x_j$, $j=1,2,\ldots,p$, for each $i$.

b) If $N \rightarrow \infty$

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

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 $v_1$ be the number of sample
elements which lie in $C_1$, $\bar{X}_1$ the sample mean of the elements in $C_1$, $M_1$ the middle point of $C_1$ and $E_1 = E(\bar{X}/X \in C_1)$.

**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

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

Then:

$$\sum_{i=1}^{n} \psi \left( \frac{x_i - M_1}{\frac{1}{12} |C_1|^{2/p}} - b \right) = E \left[ \psi \left( \frac{\varphi}{T} \right) \right] \quad (3.2)$$

stochastically if $n$ and $N$ tend to infinity and

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

Here, $b$ is the vector with the components $b_i = \frac{\rho}{\sum_{j=1}^{p} a_j^{2/p}} a_j^2$ and the notation $a * b = (a_1 b_1, a_2 b_2, \ldots, a_p b_p)$ is used.

**Proof.** If $\varphi \psi$ denotes the matrix of the second derivatives of $\psi$ we may write:

$$\psi(r) = \psi(r_0) + (\varphi(r_0) \cdot \Delta r) + \Delta r^T \psi' \psi(r_0 + \Theta \Delta r) \Delta r \quad (3.3)$$

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

We apply (3.3) to the following vectors:

\[
\begin{align*}
    r &= r_i = \frac{\bar{X}_i - \bar{e}_i}{\frac{1}{2} |C_i|^{2/p}} * b \\
    r &= r_{oi} = \frac{E_i - E_i}{\frac{1}{2} |C_i|^{2/p}} * b, \quad \text{and} \\
    \Delta r &= \Delta r_i = \frac{\bar{X}_i - E_i}{\frac{1}{3} |C_i|^{2/p}} * b.
\end{align*}
\]

If \( p_i = \int_{C_i} f(x) \, dy \), we know by the definition of the integral and Lemma 1.1 that

\[
\sum_{i=1}^{N} \psi \left( \frac{E_i - E_i}{\frac{1}{2} |C_i|^{2/p}} * b \right) p_i \longrightarrow E \left[ \psi \left( \frac{r_o}{\bar{v}} \right) \right]
\]

if \( N \to \infty \) and (3.1) holds.

Therefore, we have to prove that

\[
\begin{align*}
    \sum_{i=1}^{N} \psi(r_{oi}) \left( \frac{\bar{v}}{n} - p_i \right) + \sum_{i=1}^{N} \left( \nabla \psi(r_{oi}) \cdot \Delta r_i \right) \frac{\bar{v}_i}{n} + \\
    \sum_{i=1}^{N} \Delta r_i \frac{\bar{v}}{n} \psi(r_{oi} + \Theta \Delta r_i) \frac{\bar{v}_i}{n} \Delta r_i
\end{align*}
\]

\((3.4)\)
tends to zero.

The vector $r_{01}$ is bounded (Lemma 1.1), therefore $|\psi(r_{01})| < \alpha$ for $N$ large enough.

Using the Cauchy inequality:

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

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

$\psi(r_{01})$ is also a bounded vector for $N$ large enough, because of the continuity of $\psi/ax_1$, $j=1, \ldots, p$ and the boundedness of $r_{01}$. Therefore:

$$\left| \sum_{i=1}^{N} \left( \psi(r_{01}) \cdot \Delta x_i \right) \frac{v_i}{n} \right| \leq \beta \sum_{i=1}^{N} \sum_{j=1}^{p} |\Delta x_{ij}| \frac{v_i}{n}$$

where $\Delta x_{ij}$ is the $j$-th component of $\Delta x_i$.

Again, the Cauchy's inequality leads to:
\[
\left( \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{p} |\Delta r_{ij}| \right) \left( \frac{\sqrt{v_{i1}}}{\sqrt{n}} \right)^2 \leq \frac{N}{n} \sum_{i=1}^{N} \left( \frac{1}{n} \sum_{j=1}^{p} |\Delta r_{ij}| \right) \frac{v_{i1}}{n} \leq \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{p} (\Delta r_{ij})^2 \frac{v_{i1}}{n} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n} \sum_{j=1}^{p} \left( \frac{v_{i1} - E_{ij}}{\frac{1}{12} |C_i|^{2/p}} \right)^2 \frac{v_{i1}}{n}.
\]

The expectation of this expression is

\[
P \sum_{i=1}^{N} \sum_{j=1}^{p} b_j^2 \frac{\sigma_{ij}^2}{|C_i|^{4/p}} \frac{v_{i1}}{n} \quad \text{(3.6)}
\]

where \( \sigma_{ij}^2 \) 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 \( C \):

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

therefore, (3.6) is

\[
0 \left( \frac{p+2}{N \frac{P}{n}} \right).
\]

Then the second term of (3.4) tends to zero in probability.
if \( N = o\left(n^{p/(2+p)}\right) \) by the Markov inequality.

The assumptions about \( \psi \) makes the absolute value of the third term less then or equal to

\[
\sum_{i=1}^{N} \left( \frac{p}{j=1} |\Delta r_{ij}|^2 \right) \frac{\nu_i}{n} \leq \sum_{i=1}^{N} \frac{p}{j=1} \frac{(\Delta r_{ij})^2 \nu_i}{n} \frac{p}{j=1} \frac{1}{1}
\]

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 \( f(x_1 - \theta, \ldots, x_p - \theta) \) positive on a \( p \)-dimensional cube \( C \). Let \( X_1, \ldots , X_n \) be a random sample of \( X \). If \( f \) verifies the hypothesis of the theorem, the estimate of the information

\[
I(\theta) = E \left( \frac{\partial \log f(x)}{\partial \theta} \right)^2
\]

can be

\[
\sum_{i=1}^{N} \left( \frac{p}{j=1} \frac{x_{ij} - m_{ij}}{\Gamma_j |c_j|^{2/p}} \right)^2 \frac{\nu_i}{n} .
\]

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 \( \psi (\theta) = \left( \frac{\partial}{\partial \theta_1} x_1 \right)^2 \) because the information is the expectation of

\[
\left( \frac{\partial \log f}{\partial \theta} \right)^2 = \left( \sum_j \frac{\partial \log f}{\partial x_j} \frac{dx_j}{d\theta} \right)^2 = \left( \sum_j \frac{f}{f'} \right)^2.
\]

The result now follows from the theorem.

In the case of a multidimensional location parameter

\[ \theta = (\theta_1, \ldots, \theta_p) \]

the information is a matrix with general term

\[ I_{ij} = \int_{\mathbb{R}^p} \frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j} f \, dv \quad i, j = 1, \ldots, p. \]

**Corollary 3.2.** Let \( X \) be a \( p \)-dimensional random vector with density function

\[ f(x_1 - \theta_1, x_2 - \theta_2, \ldots, x_p - \theta_p). \]

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

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

This estimation is asymptotically consistent in probability, if

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

**Proof.** Taking $\psi(x) = x_j x_k$ 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

\[ X_1, X_2, \ldots, X_n \]

and the corresponding values of a parameter

\[ \theta_1, \theta_2, \ldots, \theta_n. \]

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 \( \theta \). The information of \( X \) for \( \theta \) is:

\[
I(\theta) = \int \left( \frac{3 \log f(x|\theta)}{3\theta} \right)^2 f(x|\theta) \, dx = \\
= \int \left( \frac{f_k(x|\theta)}{f(x|\theta)} \right)^2 \varphi(x|\theta) \, dx. \tag{4.1}
\]

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

\[(X_1, \theta_1), (X_2, \theta_2), \ldots, (X_n, \theta_n).\]
We denote by \( f(x, \theta) \) and \( \phi(\theta) \) the joint density of \((X, \theta)\) and the marginal density of \( \theta \) (or a priori density) respectively.

If we write
\[
f(x, \theta) = f(x/\theta) \phi(\theta)
\]
differentiating with respect to \( \theta \), 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
\[
\left( \frac{f'_{\theta}(x/\theta)}{f(x/\theta)} \right)^2 = \left( \frac{f'_{\theta}(x, \theta)}{f(x, \theta)} \right)^2 + \left( \frac{\phi'(\theta)}{\phi(\theta)} \right)^2 - 2 \frac{f'_{\theta}(x, \theta)}{f(x, \theta)} \frac{\phi'(\theta)}{\phi(\theta)}
\]

Therefore, an estimation of the information function (4.1) can be obtained by estimating each of the following terms:
\[
\int \left( \frac{f'_{\theta}(x, \theta)}{f(x, \theta)} \right)^2 f(x/\theta) \, dx + \left( \frac{\phi'(\theta)}{\phi(\theta)} \right)^2 -
\]
\[-2 \frac{\phi'(\theta)}{\phi(\theta)} \int \frac{\phi(x, \theta)}{f(x, \theta)} f(x, \theta) \, dx. \tag{4.2}\]

Let us take a fixed \( \theta_0 \) 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 \( I_1, I_2, \ldots, I_N \) be a partition of \((a, b)\) with length

\[ \chi_i = \frac{b - a}{N} = \xi. \]

\( \nu_{ij} \) will stand for the number of elements of the learning model in \( I_i \times J \).

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

\( \bar{\theta}_{ij} \) denote the sample mean of the \( \theta \)'s in \( I_i \times J \) and \( \bar{\theta}_J \) is the sample mean of the \( \theta \)'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
This estimation is asymptotically consistent in probability if \( N = o(n^{1/4}) \) and \( \delta \sim c/N \) and with probability one if

\[
N = k n^{(1/4) - \rho}, \quad 0 < \rho < 1/4.
\]

Here,

\[
b_\theta = \frac{a_\theta^2}{a_x a_8^2}, \quad \text{where} \quad a_\theta = \lim_{\epsilon, \delta \to 0} \frac{\delta}{\sqrt{\epsilon^2 + \delta^2}} \quad \text{and} \quad a_8 = \lim_{\epsilon, \delta \to 0} \frac{c}{\sqrt{\epsilon^2 + \delta^2}} \quad \text{and} \quad a_\theta^2 + a_x^2 = 1.
\]

PROOF. First we consider the term

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

and its estimator

\[
\frac{N}{\sum_{i=1}^{\infty} \left( \frac{\bar{y}_i - \theta_0}{\frac{1}{12} e \cdot \delta b_\theta} \right)^2 \frac{\nu_{ij}}{\nu_j}}.
\]
As before, we write:

$$(\tilde{V}_{iJ} - \theta_0)^2 = (\tilde{V}_{iJ} - E_{iJ} \theta + E_{iJ} \theta - \theta_0)^2 =$$

$$= (\tilde{V}_{iJ} - E_{iJ} \theta)^2 + (E_{iJ} \theta - \theta_0)^2 +$$

$$+ z(\tilde{V}_{iJ} - E_{iJ} \theta)(E_{iJ} \theta - \theta_0).$$

If the notation

$$P_{i/o} = p[x \in I_{i/o}] = \int f(x/\theta_0) \, dx$$

$$E_{iJ} \theta = E(\theta | (x, \theta) \in I_{iJ})$$

is used, we know that

$$\frac{N}{i=1} \left( \frac{E_{iJ} \theta - \theta_0}{\frac{1}{L^2} \in \delta \theta} \right)^2 P_{i/o} \rightarrow \int \left( \frac{f_i(x, \theta_0)}{f(x/\theta_0)} \right)^2 f(x/\theta_0) \, dx.$$

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

$$\frac{N}{i=1} \left( \frac{E_{iJ} \theta - \theta_0}{\frac{1}{L^2} \in \delta \theta} \right)^2 \left( \frac{v_{iJ}}{v_{iJ} - P_{i/o}} \right)$$

(4.5)
\[
\sum_{i=1}^{N} \left( \frac{T_{ij} - E_{ij}}{\frac{1}{2} \in \delta b_{\theta}} \right)^2 \frac{\nu_{ij}}{\nu_j} \quad (4.6)
\]

\[
\sum_{i=1}^{N} \left( \frac{(\bar{E}_{ij} - E_{ij})}{(E_{ij} \theta_{0} - \theta_{0})} \right)^2 \frac{\nu_{ij}}{\nu_j}. \quad (4.7)
\]

The expression
\[
\left( \frac{(E_{ij} \theta - \theta_{0})}{\frac{1}{2} \in \delta b_{\theta}} \right)^2
\]

is bounded for \( \epsilon \) and \( \delta \) small enough because of Lemma 1.1.

and

\[
\sum_{i=1}^{N} \left| \frac{\nu_{ij}}{\nu_j} - \frac{P_{1/o}}{P_{4/J}} \right| = \sum_{i=1}^{N} \left| \frac{\nu_{ij}}{\nu_j} - P_{1/J} + P_{4/J} - P_{1/o} \right| \leq \sum_{i=1}^{N} \left| \frac{\nu_{ij}}{\nu_j} - P_{4/J} \right| + \sum_{i=1}^{N} \left| P_{1/J} - P_{1/o} \right| \quad (4.8)
\]

with \( P_{1/J} = P \left[ X \in I_{1/o} \cap J \right] \).

For the first sum, we have:

\[
\left( \sum_{i=1}^{N} \left| \frac{\nu_{ij}}{\nu_j} - P_{1/J} \right| \right)^2 \leq \frac{1}{\nu_j} \sum_{i=1}^{N} \frac{(\nu_{ij} - \nu_j P_{1/J})^2}{\nu_j P_{1/J}} \sum_{i=1}^{N} P_{1/J}
\]
\[ = \frac{1}{N} X_{\nu_j}^2, \]

But

\[ E\left[ \left( \frac{1}{N} X_{\nu_j}^2 \right)^2 \right] \nu_j = t = 0 \left( \frac{N^2}{t^2} \right). \]

then, the expectation with respect to \( \nu_j \), lemma (2.2) and the hypothesis about \( \delta \) yield

\[ E\left[ \left( \frac{1}{N} X_{\nu_j}^2 \right)^2 \right] = 0 \left( \frac{N^2}{n^2 \delta^2} \right) = 0 \left( \frac{N^4}{n^4} \right). \]

Therefore \( \frac{1}{N} X_{\nu_j}^2 N \) 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).

\[
P_{I/J} = \frac{P \left[ \int I_k \mid \theta \in J \right]}{P \left[ \theta \in J \right]} = \frac{\int_{I_k \cap J} f(x, \theta) \, dx \, d\theta}{\int_J \Phi(\theta) \, d\theta}
= \frac{\int_{I_k \cap J} \left[ f(x, \theta) + f_\theta(x, \theta') \frac{\theta - \theta'}{\delta} + o \left( (\theta - \theta')^2 \right) \right] \, dx \, d\theta}{\int_J \left[ \Phi(\theta) + \Phi'(\theta) \frac{\theta - \theta'}{\delta} + o \left( (\theta - \theta')^2 \right) \right] \, d\theta}
\]
\[
\int \left[ f(x/\theta_0) \phi(\theta_0) \delta + o(\delta^2) \right] dx = \phi(\theta) \delta + o(\delta^2)
\]

\[
\int f(x/\theta_0) \ dx + o(\delta \varepsilon) \over 1 + o(\delta)
\]

Therefore, \( |P_{4/J} - P_{4/O} | = o(\delta \varepsilon) \), and

\[
\frac{1}{N} \sum_{i=1}^{N} |P_{4/J} - P_{4/O} | = N o(\delta \varepsilon) = o(\delta).
\]

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 \( \nu \neq 0 \) and get:

\[
\sum_{i=1}^{N} \frac{a_{ij}^2}{n(\delta \varepsilon b_0)^2} \frac{1}{\nu_j}
\]

where \( a_{ij}^2 = \text{Var} (\theta | (X, \theta) \in I_{4/Xj} ) = 0 (\delta) \).

By lemma (2.2)

\[
E^*(\frac{1}{\nu_j}) = 0 (\frac{1}{n^8})
\]
since \( n\delta \sim c n/N \rightarrow \alpha \) as \( n \rightarrow \infty \) and \( N = o(n^{1/4}) \).

Therefore, the expected value of (4.6) is

\[
0 \left( \frac{N\delta}{c^2 \delta^2 n\delta} \right) = 0 \left( \frac{n^4}{n} \right)
\]

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

\[
\left| \sum_{i=1}^{n} \left( \frac{E_{iJ} - E_{iJ} \theta}{\varepsilon \delta b_\theta} \right)^2 \right| \leq \sum_{i=1}^{n} \left| \frac{E_{iJ} - E_{iJ} \theta}{\varepsilon \delta b_\theta} \right| \left( \frac{1}{\varepsilon \delta b_\theta} \right)
\]

and the Cauchy's inequality gives:

\[
\left( \sum_{i=1}^{n} \left( \frac{E_{iJ} - E_{iJ} \theta}{\varepsilon \delta b_\theta} \right)^2 \right)^{1/2} \leq \sum_{i=1}^{n} \left| \frac{E_{iJ} - E_{iJ} \theta}{\varepsilon \delta b_\theta} \right| \left( \frac{1}{\varepsilon \delta b_\theta} \right)^{1/2}
\]

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
\[
\frac{\tilde{\theta}_J - \theta_0}{\frac{1}{12} \delta^2}.
\]

But
\[
\frac{\tilde{\theta}_J - \theta_0}{\frac{1}{12} \delta^2} = \frac{\tilde{\theta}_J - E_J \theta + E_J \theta - \theta_0}{\frac{1}{12} \delta^2} = \frac{\tilde{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} + \frac{E_J \theta - \theta_0}{\frac{1}{12} \delta^2}.
\]

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

As before, we can consider the following expectation, for fixed \( \nu_J \):
\[
E_{\nu_J} \left( \frac{\tilde{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} \right)^2 = \frac{\nu_J^2}{\delta^4} = \frac{\eta(\delta^2)}{\nu_J} = \frac{1}{\nu_J} \left( \frac{1}{4 \delta^2} \right).
\]

Taking again expectation, with respect to \( \nu_J \) this time, and applying lemma 2.2 we get:
\[
E \left( \frac{\tilde{\theta}_J - E_J \theta}{\frac{1}{12} \delta^2} \right)^2 = 0 \left( \frac{1}{n \delta^2} \right) = 0 \left( \frac{N^3}{n} \right).
\]
and the stochastic convergence follows easily.

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

\[
\sum_{i=1}^{N} \frac{\bar{\theta}_{ij} - \theta_0}{1 \mathcal{I} \epsilon \in \mathcal{B}_\theta} \frac{v_i}{v_J} \quad (4.10)
\]

is a consistent estimator of

\[
\int \frac{f_i(x, \theta)}{f(x, \theta)} f(x/\theta) dx.
\]

Writing \( \bar{\theta}_{ij} - \theta_0 = \bar{\theta}_{ij} - \bar{\theta} - \bar{\theta} + \bar{\theta}_0 - \theta_0 \), (4.10) can be written in the following way:

\[
\sum_{i=1}^{N} \frac{E_{ij} \theta - \theta_0}{1 \mathcal{I} \epsilon \in \mathcal{B}_\theta} P_{1/0} + \sum_{i=1}^{N} \frac{E_{ij} \theta - \theta_0}{1 \mathcal{I} \epsilon \in \mathcal{B}_\theta} \left( \frac{v_i}{v_J} - P_{1/0} \right)
\]

\[
+ \sum_{i=1}^{N} \frac{\bar{\theta}_{ij} - E_{ij} \theta}{1 \mathcal{I} \epsilon \in \mathcal{B}_\theta} \frac{v_i}{v_J}.
\]

The first sum converges to

\[
\int \frac{f_i(x, \theta)}{f(x, \theta)} f(x/\theta) dx.
\]
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 $i_{1J}$, $i=1,\ldots,N$

\[
E\left[ \sum_{i=1}^{N} \left( \frac{1}{\theta_{1J}} - E\theta_{1J} \right)^{2} \nu_{iJ} \right] \leq \\
E\left[ \sum_{i=1}^{N} \left( \sum_{l=1}^{s} \left( \frac{1}{\theta_{1J}} - E\theta_{1J} \right)^{2} \nu_{iJ} \right) \right] \\
\cdot \left( \sum_{l=1}^{2s} \left( \frac{1}{\theta_{2J}} - E\theta_{2J} \right)^{2} \nu_{2J} \right) \left( \sum_{l=1}^{2s} \left( \frac{1}{\theta_{NJ}} - E\theta_{NJ} \right)^{2} \nu_{NJ} \right)
\]

Since $\theta_{iJ}$'s are stochastically independent, this is equal to

\[
\sum_{i=1}^{N} E_{\nu} \left[ \left( \sum_{l=1}^{s} \left( \frac{1}{\theta_{iJ}} - E\theta_{iJ} \right)^{2} \nu_{iJ} \right) \right] = \\
\sum_{i=1}^{N} E_{\nu} \left[ \left( \sum_{l=1}^{s} \left( \frac{1}{\theta_{iJ}} - E\theta_{iJ} \right)^{2} \nu_{iJ} \right) \right]
\]
\[ \sum_{i=1}^{N} t_i = s \left( t_1 \cdots t_N \right) \cdot 0 \left( \frac{2t_i}{\nu_{ij}} \right) \cdot \nu_{ij} = 0 \left( 4^{2s}n^s \right). \]

Then

\[ E_{\nu}(A)^s = \left[ \frac{\prod_{i=1}^{N} (\delta_{ij} - \delta_{i0})^2 \cdot \nu_{ij}}{\nu_{ii} \cdot \delta_{i0} \cdot \delta_{ij}^2} \right]^{s} = \frac{1}{\nu_{ij}} \left( \delta_{ii}^s \right) = \frac{1}{\nu_{ij}} \left( \frac{N^s}{n^s} \right) = \frac{1}{\nu_{ij}} (N^3s). \]

Therefore

\[ E\left( E_{\nu}(A)^s \right) = 0 \left( \frac{N^3s}{n^3\delta^s} \right) = 0 \left( \frac{N^4s}{n^s} \right), \]

by lemma 2.2.

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

\[ P\left( |A| > k \right) = P\left( \lambda^s > k^s \right) \leq \frac{E(\lambda^s)}{k^s} = 0 \left( \frac{N^4s}{n^s} \right) = 0 \left( n^{-4\rho s} \right) = 0 \left( \frac{1}{1+\alpha} \right). \]
REFERENCES


