

COMPACT HYPOTHESIS AND EXTREMAL SET ESTIMATORS

JOÃO TIAGO MEXIA AND PEDRO CORTE REAL

*Universidade Nova de Lisboa, Departamento de Matemática
da Faculdade de Ciências e Tecnologia*

Quinta da Torre, 2825–114 Monte de Caparica, Portugal

e-mail: parcr@mail.fct.unl.pt

Abstract

In extremal estimation theory the estimators are local or absolute extremes of functions defined on the cartesian product of the parameter by the sample space. Assuming that these functions converge uniformly, in a convenient stochastic way, to a limit function g , set estimators for the set ∇ of absolute maxima (minima) of g are obtained under the compactness assumption that ∇ is contained in a known compact U . A strongly consistent test is presented for this assumption. Moreover, when the true parameter value $\vec{\beta}_0^k$ is the sole point in ∇ , strongly consistent pointwise estimators, $\{\hat{\beta}_n^k : n \in \mathbb{N}\}$ for $\vec{\beta}_0^k$ are derived and confidence ellipsoids for $\vec{\beta}_0^k$ centered at $\hat{\beta}_n^k$ are obtained, as well as, strongly consistent tests. Lastly an application to binary data is presented.

Keywords: extremal estimators, set estimators, confidence ellipsoids, strong consistency, binary data.

2000 Mathematics Subject Classification: 60F15, 62F10.

1. INTRODUCTION

The main estimation techniques, such as least squares and maximum likelihood, are special features of extremal estimation theory (for instance, see [1] and [12]). In this theory the estimator is an absolute extreme of a function $g_n(\vec{\beta}^k, \vec{y}^n)$ defined on the Cartesian product $\Theta \times \Omega$ of the parameter by the sample spaces. When applying this theory we may, see [1], obtain a sequence

$\{\hat{\beta}_n^k : n \in \mathbb{N}\}$ of estimators that does not converge to the true value, β_0^k of the parameter.

To overcome this difficulty we show that, if $g_n(\vec{\beta}^k, \vec{y}^n)$ converges stochastically uniformly to a limit function $g(\vec{\beta}^k)$ whose set ∇ of absolute maxima (minima) is contained in a *known* compact set U , we can derive, from $g_n(\vec{\beta}^k, \vec{y}^n)$, sets Δ_n that converge, in a convenient stochastic way, to ∇ . We will present a strongly consistent test for the assumption of compactness, $\nabla \subseteq U$. These results give us the background for dealing with the case in which $\nabla = \{\beta_0^k\}$, since it is possible to show that, if $\tilde{\beta}_n^k \in \nabla_n$ then $\tilde{\beta}_n^k$ is a strongly consistent estimator of β_0^k . Special interesting is the estimator given by a generalized gravity center of ∇_n which is not only strongly consistent, but also the center of a confidence ellipsoid for β_0^k . From this ellipsoid we can obtain, using a procedure due to [19], simultaneous confidence intervals for the parameter $\vec{v}^{k'} \beta_0^k$ and, through duality, tests for hypothesis on β_0^k .

We will present two applications of our results. The first one of these will be to binary data while the second one will be to pseudo maximum likelihood estimators, these estimators are widely applicable, see [5] and guarantee that $\nabla = \{\beta_0^k\}$, thus giving good applications of our results.

The use of sets to derive estimators has been used by several authors. In particular, there is very interesting work using the epi and hypo graphs, for instance see [9] and [8]. Our approach differs from that one since our sets are limited by level curves: $\nabla_n(a) = \{\beta^k : g_n(\beta^k, \vec{y}^n) \geq a\}$. While behind the epi-graph approach we may find results on convex optimization obtained, for instance, by [21] and [17], our approach is more inspired by the original Wald (see [20]) treatment of the strong consistency of maximum-likelihood estimators.

2. ESTIMATION OF SETS

Let U be a compact set in the k -dimensional Euclidean space (\mathbb{R}^k) , $\nabla [\nabla_n]$ the set, possibly empty, of the absolute maxima of the real valued function $g(\beta)$ [$g(\beta, y) = g_n(\beta)$], with y the n -dimensional vector of observations] and β an k -dimensional parameter vector, belonging to the parameter space, Θ . We shall use *sup* as shorthand for the supremum. For a set A , $int(A)$, $front(A)$ and $cl(A)$ will be the interior, the frontier and the closure sets, respectively. The Euclidean norm of a vector v will be represented by $\|v\|$. With \mathbb{N} the set of positive integers, $a_n \rightarrow a$ indicates the convergence of the sequence $\{a_n : n \in \mathbb{N}\}$ to a . We write $a_n \downarrow a$ when the sequences decreases

to a , in the same way, $a_n \uparrow a$ when the sequences increases to a and $a_n \xrightarrow{P} a$ when it converges stochastically to a . Unless, otherwise specified, all limits will be taken as $n \rightarrow \infty$.

We will define; $\Delta(a) = \{\beta : g(\beta) \geq a\}$ and $\Delta_n(a) = \{\beta : g_n(\beta) \geq a\}$, as well as, $M = \sup_{\beta \in \Theta} g(\beta) \leq +\infty$, $M_n = \sup_{\beta \in \Theta} g_n(\beta) \leq +\infty$, $D = \sup_{\beta \notin U} g(\beta)$, $D_n = \sup_{\beta \notin U} g_n(\beta)$, $\delta = M - D$ and $\delta_n = M_n - D_n$.

From the definitions, it follows that,

$$\nabla = \bigcap_{a \leq M} \Delta(a) \quad \text{and} \quad \nabla_n = \bigcap_{a \leq M_n} \Delta_n(a).$$

With,

$$S_n = \sup |g_n(\beta) - g(\beta)|,$$

we can easily show that

$$(2.1) \quad |M_n - M| \leq S_n \quad \text{and} \quad |D_n - D| \leq S_n.$$

From (2.1) we obtain,

$$(2.2) \quad |\delta_n - \delta| \leq 2S_n.$$

We also have, with, $u, m \in \mathbb{R}$

$$(2.3) \quad \Delta(u - (m - 1)S_n) \subseteq \Delta_n(u - mS_n) \subseteq \Delta(u - (m + 1)S_n).$$

The proof of the first inclusion follows from

$$\begin{aligned} \Delta(u - (m - 1)S_n) &= \{\beta : g(\beta) \geq u - (m - 1)S_n\} \subseteq \\ &\{\beta : g_n(\beta) + S_n \geq u - (m - 1)S_n\} = \{\beta : g_n(\beta) \geq u - mS_n\} = \\ &\Delta_n(u - mS_n), \end{aligned}$$

while the proof of the second one is similar.

We now introduce the assumptions:

A1: the function $g(\beta)$ is continuous,

A2: $0 < \delta < +\infty$.

Given a set $C \subseteq \mathbb{R}^k$ and $\epsilon > 0$, let us define

$$V_\epsilon(C) = \left\{ \beta : \inf_{c \in C} \|\beta - c\| \leq \epsilon \right\}.$$

Let us establish

Proposition 2.1. *If A1 and A2 hold, then $M = \max g(\beta) = \max_{\beta \in U} g(\beta)$, $\nabla \neq \emptyset$ and an $\epsilon > 0$ exists, such that, $V_\epsilon(\nabla) \subset U$.*

Proof. By A2, we know that $M = \sup_{\beta \in U} g(\beta)$. Using A1, Weierstrass's Theorem and the compactness of U , we get, $M = \max_{\beta \in U} g(\beta)$. Moreover, we have proved that ∇ is nonempty. To prove the third part of the thesis, let us assume that no $\epsilon > 0$ exists such that $V_\epsilon(\nabla) \subset U$. Then it would be possible to find a sequence $\beta_n \notin U$, such that, $\beta_n \rightarrow \beta \in \nabla$ and, due to A1, we would have, $g(\beta_n) \rightarrow g(\beta) = M$, so that A2 would not hold. ■

Proposition 2.2. *If A1 and A2 hold, for all $\epsilon > 0$ an $\eta(\epsilon) > 0$ exists, such that, whenever, $M - \eta(\epsilon) \leq b \leq M$, then $\Delta(b) \subseteq V_\epsilon(\nabla)$. Moreover, if A1 and A2 hold and $S_n \leq \eta(\epsilon)/3$, then $\Delta_n(M_n - S_n) \subseteq V_\epsilon(\nabla)$.*

Proof. First, according to Proposition 2.1 we know that ∇ is nonempty. Now, if with $\epsilon > 0$, there was no $\eta > 0$, such that, $\Delta(M - \eta) \subseteq V_\epsilon(\nabla)$, we could take $\eta_n \downarrow 0$ such that, whichever n , $\Delta(M - \eta_n) \subseteq V_\epsilon(\nabla)$ does not hold. Moreover, there would be an $m \in \mathbb{N}$, such that, for $n > m$, $\eta_n < \delta$ and so $\Delta(M - \eta_n) \subset U$. We could now take $\beta_n \in \Delta(M - \eta_n) - V_\epsilon(\nabla)$, obtaining a sequence that, for $n > m$, lies inside U thus having a convergent subsequence $\{\beta_{v(n)}\}$. Now $\beta_{v(n)} \rightarrow \beta$ and according to A1, $g(\beta_{v(n)}) \rightarrow g(\beta)$ but $M - \eta_n \leq g(\beta_{v(n)}) \leq M$ so $g(\beta) = M$ which is impossible, because β clearly would not belong to ∇ since its distance to ∇ is at least $\epsilon > 0$. Finally, we know by (2.3) that, $\Delta_n(M_n - S_n) \subseteq \Delta(M_n - 2S_n)$. Since, $M_n - 2S_n \geq M - 3S_n \geq M - \eta(\epsilon)$ we have, by Proposition 2.2, $\Delta_n(M_n - S_n) \subseteq \Delta(M - \eta(\epsilon)) \subseteq V_\epsilon(\nabla)$. ■

We now add the assumption:

A3: $S_n \xrightarrow{P} 0$.

thus, from expressions (2.1) and (2.2), we know that

$$M_n \xrightarrow{P} M, \quad D_n \xrightarrow{P} D \quad \text{and} \quad \delta_n \xrightarrow{P} \delta.$$

This assumption implies a stochastic uniform convergence of $g_n(\beta)$ to $g(\beta)$. Let us establish

Proposition 2.3. *If A1, A2 and A3 hold, $\hat{\beta}_n \in \Delta_n(M_n - S_n)$ and $\hat{\beta}_n \xrightarrow{P} \beta$, then $P[\beta \in \nabla] = 1$.*

Proof. If $P[\hat{\beta}_n \in \Delta_n(M_n - S_n)] = 1$ and $S_n \leq \eta(\epsilon)/3$, we will have, according to Proposition 2.2, $P[\hat{\beta}_n \in V_\epsilon(\nabla)] = 1$, which establishes the thesis, since ϵ is arbitrary and $\nabla = \bigcap_{\epsilon > 0} V_\epsilon(\nabla)$. ■

Given a sequence of sets C_n in \mathbb{R}^k we will write, $C_n \xrightarrow{P_V} C \subseteq \mathbb{R}^k$, when,

$$\forall \epsilon > 0, P[(C_n \subseteq V_\epsilon(C)) \cap (C \subseteq V_\epsilon(C_n))] \rightarrow 1,$$

this being the set convergence that we consider.

We now introduce a discrete version of A3. Let $\eta_n \downarrow 0$ be a sequence, such that,

$$\mathbf{A4:} P[S_n \leq \eta_n] \rightarrow 1.$$

We now have

Proposition 2.4. *If A1, A2 and A4 hold, then $\Delta_n(M_n - \eta_n) \xrightarrow{P_V} \nabla$.*

Proof. Let $S_n \leq \eta_n < \eta(\epsilon)/6$ and consider $\beta_0 \in \Delta_n(M_n - \eta_n)$. In this case, $g(\beta_0) > M_n - 2\eta_n > M - 3\eta_n$, hence, $\beta_0 \in \Delta(M - 3\eta_n)$. Since, $M - 3\eta_n > M - \eta(\epsilon)$, we have, $\beta_0 \in \Delta(M - 3\eta_n) \subseteq \Delta(M - \eta(\epsilon))$, hence, by Proposition 2.2, $\Delta(M - \eta(\epsilon)) \subseteq V_\epsilon(\nabla)$, meaning that,

$$\Delta_n(M_n - \eta_n) \subseteq \Delta(M - 3\eta_n) \subseteq \Delta(M - \eta(\epsilon)) \subseteq V_\epsilon(\nabla).$$

Conversely, when $S_n \leq \eta_n$ we have the inclusions,

$$\nabla = \Delta(M) \subseteq \Delta_n(M_n - 2S_n) \subseteq \Delta_n(M_n - \eta(\epsilon)/3) \subseteq V_\epsilon(\Delta_n(M_n - \eta_n)).$$

So the thesis follows, since, $P[\nabla \subseteq V_\epsilon(\Delta_n(M_n - \eta_n))] \rightarrow 1$. ■

This result is interesting since it gives us a set estimator, $\Delta_n(M_n - \eta_n)$ for ∇ .

We note that, with if $C_n \xrightarrow{P_V} C$, then $cl(C_n) \xrightarrow{P_V} C$ so that we may rewrite the thesis of Proposition 2.4 as: $cl(\Delta_n(M_n - \eta_n)) \xrightarrow{P_V} \nabla$.

3. TEST FOR THE COMPACT HYPOTHESIS

When A4 holds and η_n is known, we can use δ_n as the test statistic to test that $\nabla \subseteq U$, with U a known compact set.

The test rule will be:

$$(3.1) \quad \begin{cases} \delta_n > 2\eta_n, & \text{accept A2} \\ \delta_n \leq 2\eta_n, & \text{reject A2.} \end{cases}$$

We now establish:

Proposition 3.1. *If A4 holds then the test defined by (3.1) is strongly consistent.*

Proof. From A4 we get, according to (2.2), $P[|\delta_n - \delta| \leq 2\eta_n] \rightarrow 1$. So, when $\delta > 0$ [$\delta \leq 0$] we have $P[\delta_n > 2\eta_n] \rightarrow 1$ [$P[\delta_n \leq 2\eta_n] \rightarrow 1$] and so, the probabilities of first and second type errors tend to zero when $n \rightarrow \infty$, hence (see [18]) the test is strongly consistent \blacksquare

4. POINT ESTIMATORS AND TAYLOR SERIES

We start by presenting general point estimators to be used when $\nabla = \{\beta_1\}$, that is, when there is a sole maximum. According to Proposition 2.3, when β is the limit of $\hat{\beta}_n \in \Delta_n(M_n - S_n)$, we have

$$(4.1) \quad P[\beta = \beta_1] = 1.$$

It is easy to show that, if in Proposition 2.3 we use A4 instead of A3 we can change $\hat{\beta}_n \in \Delta_n(M_n - S_n)$ into $\hat{\beta}_n \in \Delta_n(M_n - \eta_n)$. When A1 and A4 hold and η_n is known, we can use the test rule (3.1) for testing $\nabla \subseteq U$. The thesis of Proposition 3.1 and expression (4.1) will both continue to hold.

When ∇ contains a unique point we are mainly interested in the case for which that point is the true parameter β_0 . To check if this assumption holds, we may obtain the limit function $g(\cdot)$ and see if, whatever β_0 , we have $\nabla = \{\beta_0\}$.

From now on, we restrict ourselves to this case, assuming that

A5: The function $g(\cdot)$ is thrice differentiable, the set of zeros of $grad(g(\beta))$ has no accumulation point, $\nabla = \{\beta_0\}$, and

$$H = -Hessian(g(\beta_0)) \text{ is positive definite.}$$

Clearly A5 implies A1.

Then, with λ_m the minimum eigenvalue of H , we will have, $\lambda_m > 0$ and there will be an $s > 0$ such that β_0 is the sole zero of $grad(g(\beta))$ in $V_s(\beta_0)$.

Let us consider the function

$$q(\beta) = M - \frac{1}{2}(\beta - \beta_0)' H(\beta - \beta_0)$$

and the sets

$$\Delta_q(a) = \{\beta : q(\beta) \geq a\}$$

which are ellipsoids centered at β_0 .

If A5 holds, we will have

$$(4.2) \quad T(\beta) = |g(\beta) - q(\beta)| / \|\beta - \beta_0\|^2 \rightarrow 0 \text{ when } \beta \rightarrow \beta_0,$$

since the residual of a second order Taylor expansion, for thrice differentiable functions, is an infinitesimal of order higher than the second.

We can use generalized polar coordinates to define the function $Z(r) = \max_{\vec{\theta}^{k-1}} T(\beta_0 + r l(\vec{\theta}^{k-1}))$.

Proposition 4.1. *If A5 holds then $Z(r) \rightarrow 0$ as $r \rightarrow 0$.*

Proof. Let us suppose that A5 holds but that the thesis is not valid. Then, a sequence $0 < r_n \downarrow 0$ and an $c > 0$ would exist, such that, $Z_n(r_n) > c$, for all $n \in \mathbb{N}$. So, for each r_n , due to the continuity of $T(\beta)$, an $\vec{\theta}^{k-1}$ exists, such that, $T(\beta_0 + r_n l(\vec{\theta}^{k-1})) > c$, which is an impossibility since $\beta_n = \beta_0 + r_n l(\vec{\theta}_n^{k-1}) \rightarrow \beta_0$ and by (4.2), $T(\beta_n) \rightarrow 0$. ■

We can now state the following three Corollaries

Corollary 4.2. *Let $Z^+(r) = \max_{0 \leq s \leq r} Z(s)$. If A5 holds then, $Z^+(r) \rightarrow 0$ as $r \rightarrow 0$.*

Proof. Similar to the proof of Proposition 4.1. ■

Corollary 4.3. *If A5 holds then, $\sup_{\beta \in V_r(\beta_0)} |g(\beta) - q(\beta)| \leq Z^+(r)r^2 \rightarrow 0$ as $r \rightarrow 0$.*

Proof. The thesis follows directly from Corollary 4.2. ■

In the next Corollary we will consider ellipsoids putting

$$\xi(B_{k \times k}, \vec{b}^k, d) = \left\{ \vec{\beta}^k : (\vec{\beta}^k - \vec{b}^k)' B_{k \times k} (\vec{\beta}^k - \vec{b}^k) \leq d \right\},$$

where $B_{k \times k}$ is a positive definite matrix and $d \in \mathbb{R}_0^+$.

Corollary 4.4. *If A5 holds, then $\Delta_q(a) = \xi(H/2, \beta_0, M - a) = \xi(H, \beta_0, (M - a)/2)$ and, with $b_n \downarrow 0$, $\sup_{\beta \in \Delta_0(M - \alpha b_n)} |g(\beta) - q(\beta)|/b_n \rightarrow 0$, for all $\alpha \geq 0$.*

Proof. The first part of the thesis results from the definition of $q(\beta)$. Let λ_m be the smallest eigenvalue of H , so that the largest semiaxis of $\Delta_q(M - \alpha b_n)$ measures $h_{\alpha, n} = \sqrt{(M - \alpha b_n)/\lambda_m}$, hence $\Delta_q(M - \alpha b_n) \subseteq V_{h_{\alpha, n}}(\beta_0)$ and the thesis follow from Corollary 4.3. ■

We now need to state the two following Lemmas

Lemma 4.5. *If A2 and A5 hold then an $T < M$ exists such that for all a verifying $T \leq a < M$ the set $\Delta(a)$ is compact and connected.*

Proof. If $a > M - \delta$, $\Delta(a) \subset U$, so that it is closed and bounded, thus a compact. If there is no $T < M$ such that, for $a \in]T, M]$, $\Delta(a)$ is connected we could choose $a_n \uparrow a$ such that $\Delta(a_n)$ is not connected. Then, there would be pairs $O_i(a_n)$, of open sets such that $O_i(a_n) \cap \Delta(a_n) \neq \emptyset$, $i = 1, 2$; $\bigcap_{i=1}^2 (O_i(a_n) \cap \Delta(a_n)) = \emptyset$ and $\bigcup_{i=1}^2 (O_i(a_n) \cap \Delta(a_n)) = \Delta(a_n)$. There also would be an $m \in \mathbb{N}$, such that, $n > m$ implies $a_n > M - \delta$, and the closures $cl(O_i(a_n) \cap \Delta(a_n))$, $i = 1, 2$ will be contained in U so that they will be compact sets. If A5 would hold, the limit function and its gradient would be continuous, thus $g(\beta)$ would have maxima $\beta_{n,i}$ belonging to $cl(O_i(a_n) \cap \Delta(a_n))$, $i = 1, 2$. Since, $grad(g(\beta_{n,i})) = 0$, $i = 1, 2$, according to Proposition 2.2 and $\nabla = \{\beta_0\}$, β_0 would be an accumulation point of the set zeros of $grad(g(\beta))$ which is impossible whenever A5 holds. Hence, if A2 and A5 hold we cannot have $a_n \rightarrow M$ with the sets $\Delta(a_n)$ being disconnected, thus an $T < M$, fulfilling the conditions stated in the thesis, must exist. ■

Lemma 4.6. *If A is a connected and compact set, B is a compact set and $\emptyset \neq A \cap B \subset int(B)$, then $A \subset B$.*

Proof. Both $A \cap B$ and $front(B)$ are compact sets. Let us define in $A \cap B$ the distance to $front(B)$, which, according to Weierstrass's Theorem will have a minimum, say, d . Since $A \cap B$ is a closed set and $(A \cap B) \cap front(B) = \emptyset$, we know that $d > 0$.

Now let us suppose that $A - B \neq \emptyset$, then we would have open (non-empty) sets $V_{d/3}(A \cap B)$ and $V_{d/3}(A - B)$, which is impossible, due to A and B being connected and to

- $V_{d/3}(A \cap B) \cap A \neq \emptyset$, $V_{d/3}(A - B) \cap A \neq \emptyset$
- $(V_{d/3}(A \cap B) \cap A) \cap (V_{d/3}(A - B) \cap A) = \emptyset$
- $(V_{d/3}(A \cap B) \cap A) \cup (V_{d/3}(A - B) \cap A) = A$

to complete the proof we have only to point out that $A \cap B = A$ is equivalent to $A - B = \emptyset$. ■

Proposition 4.7. *If A2, A4 and A5 hold then*

$$P [\Delta_q(M_n - \eta_n) \subset \Delta(M - 2\eta_n) \subset \Delta_n(M_n - 3\eta_n) \subset \Delta(M_n - 4\eta_n) \subset \Delta_q(M_n - 5\eta_n) \subset \Delta_q(M - 6\eta_n)] \rightarrow 1.$$

Proof. If $S_n < \eta_n$ we have $\Delta(M_n - 2\eta_n) \subset \Delta_n(M_n - 3\eta_n) \subset \Delta(M_n - 4\eta_n)$, as well as, $\Delta_q(M_n - 5\eta_n) \subset \Delta_q(M - 6\eta_n)$. Now, due to Corollary 4.3, an $m \in \mathbb{N}$ exists, such that $n \geq m$ implies $z_n = \sup_{\beta \in \Delta_q(M - 7\eta_n)} |q(\beta) - g(\beta)| < \eta_n$, so that $\Delta(M_n - 4\eta_n) \cap \Delta_q(M - 7\eta_n) \subset \Delta_q(M_n - 5\eta_n)$. Moreover, if $\eta_n < (M - T)/5$, with T as defined in Lemma 4.5, and again with $S_n < \eta_n$ we will have, for n large enough, $M_n - 4\eta_n > T$, and due to Lemma 4.5, $\Delta(M_n - 4\eta_n)$ is a connected set that intersects the compact set $\Delta_q(M - 7\eta_n)$ and is contained in $\text{int}(\Delta_q(M - 7\eta_n))$, so by Lemma 4.6 we know that $\Delta_q(M - 7\eta_n) \cap \Delta(M_n - 4\eta_n) = \Delta(M_n - 4\eta_n)$. Thus we established that

$$\begin{aligned} \Delta_q(M_n - \eta_n) &\subset \Delta(M - 2\eta_n) \subset \Delta_n(M_n - 3\eta_n) \subset \\ \Delta(M_n - 4\eta_n) &\subset \Delta_q(M_n - 5\eta_n) \subset \Delta_q(M - 6\eta_n) \end{aligned}$$

the rest of the proof being obvious. \blacksquare

As a direct extension of the usual definition of gravity center (for bodies with uniform mass density), given $A \subset \mathbb{R}^k$, we take its gravity center, $GC(A)$, to be the point in \mathbb{R}^k with Cartesian coordinates given by

$$m_j = \int_A x_j \prod_{i=1}^k dx_i \quad j = 1, \dots, k.$$

Useful results on ellipsoids and gravity centers are presented in the Appendix.

We now establish

Corollary 4.8. *If A2, A4 and A5 hold then*

$$P[GC(\Delta_n(M_n - 3\eta_n)) \in \xi(H, \beta_0, 2\eta_n)] \rightarrow 1.$$

Proof. It is a direct application of Proposition 7.4, given in the Appendix. \blacksquare

5. CONFIDENCE ELLIPSOIDS AND TESTS

We now show how to derive confidence ellipsoids for β_0 from the gravity center estimator. In order to do it we replace **A4** by

A'4: There exists a sequence $\{v_n \downarrow 0 : n \in \mathbb{N}\}$ such that $\liminf P[S_n \leq zv_n] \geq G(z)$, with $G(z)$ a distribution with support in \mathbb{R}^+ .

We may now reason as for obtaining the Corollary 4.8 in order to show that

$$\liminf P \left[\hat{\beta}_n \in \xi(H, \beta_0, 2zv_n) \right] \geq G(z)$$

with $\hat{\beta}_n$ the gravity center of $\Delta_n(M_n - 3zv_n)$. Moreover, with $\hat{H}_n = -Hess(g(\hat{\beta}_n))$, since the second partial derivatives of $g(\cdot)$ are continuous and $\hat{\beta}_n$ is a strongly consistent estimator, for all $\epsilon > 0$, we have $P[\xi(H, \beta_0, 2zv_n) \subset \xi(\hat{H}_n, \beta_0, 2(z + \epsilon)v_n)] \rightarrow 1$, so that,

$$\liminf P \left[\hat{\beta}_n \in \xi \left(\hat{H}_n, \beta_0, 2(z + \epsilon)v_n \right) \right] \geq G(z).$$

Now, $\hat{\beta}_n \in \xi(\hat{H}_n, \beta_0, 2(z + \epsilon)v_n)$ if and only if $\beta_0 \in \xi(\hat{H}_n, \hat{\beta}_n, 2(z + \epsilon)v_n)$ and so $\liminf P[\beta_0 \in \xi(\hat{H}_n, \hat{\beta}_n, 2(z + \epsilon)v_n)] \geq G(z)$.

Thus $\xi(\hat{H}_n, \hat{\beta}_n, 2(z + \epsilon)v_n)$ may be considered as a confidence ellipsoid for β_0 whose limit level is not inferior to $G(z)$. We required $G(z)$ to be a distribution so that this bound may be freely chosen.

Moreover, (see [19]), there exists a pair of parallel planes orthogonal to $\vec{d}^k \neq \vec{0}^k$ and tangent to $\xi(\hat{H}_n, \hat{\beta}_n, 2(z + \epsilon)v_n)$, with β being between these planes, if and only if,

$$|d' \beta - d' \hat{\beta}_n| \leq \sqrt{2(z + \epsilon)v_n d' \hat{H}_n d}.$$

Since a point lies inside an ellipsoid, if and only if, it lies between all pairs of parallel planes tangent to that ellipsoid, we have, indicating by \bigcap_d that all vector in \mathbb{R}^k are considered,

$$P \left[\bigcap_d \left(|d' \beta - d' \hat{\beta}_n| \leq \sqrt{2(z + \epsilon)v_n d' \hat{H}_n d} \right) \right] \geq G(z),$$

thus obtaining simultaneous confidence intervals for all linear functions of the parameter. The lower bound of the joint probability for these intervals will be $G(z)$.

If $v_n/\eta_n \rightarrow 0$ we clearly see that

$$P \left[\beta_0 \in \xi \left(\hat{H}_n, \hat{\beta}_n, \eta_n \right) \right] \rightarrow 1,$$

so that we have a sequence of confidence ellipsoids with limit level 1. We point out that the largest semiaxis of these ellipsoids vanishes to zero. Let C be a subset of the parameter space. Then, if we want to test $H_0 : \beta_0 \in C$ we can use the test rule

$$\begin{cases} \xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \cap C \neq \emptyset, & \text{accept } H_0 \\ \xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \cap C = \emptyset, & \text{reject } H_0 \end{cases}$$

we can now establish

Proposition 5.1. *If C is closed the test defined above is strongly consistent.*

Proof. When H_0 holds, $\beta_0 \in \xi(\hat{H}_n, \hat{\beta}_n, \eta_n)$ will imply $\xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \cap C \neq \emptyset$, so that $P[\xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \cap C \neq \emptyset] \rightarrow 1$. If $\beta_0 \notin C$, since C is closed there will be $\epsilon > 0$, such that, $V_\epsilon(\beta_0) \cap C = \emptyset$. Since the largest semiaxis of the ellipsoids $\xi(\hat{H}_n, \hat{\beta}_n, \eta_n)$ tends to zero with probability 1 and $P[\beta_0 \in \xi(\hat{H}_n, \hat{\beta}_n, \eta_n)] \rightarrow 1$, we will have $P[\beta_0 \in \xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \subset V_\epsilon(\beta_0)] \rightarrow 1$, and so, when $\beta_0 \notin C$, $P[\xi(\hat{H}_n, \hat{\beta}_n, \eta_n) \cap C = \emptyset] \rightarrow 1$. ■

For instance, we can use this technique to test $H_0 : J\beta_0 = c$ with J an $s \times k$ matrix.

6. AN APPLICATION TO BINARY DATA

Let $Y_{n,1}, \dots, Y_{n,m}$ be independent binomial random variables, with parameters, n and $p_{i,0} = p(\bar{x}_i^h, \beta_0)$, the \bar{x}_i^h being known, $i = 1, \dots, m$, while β_0 is to be estimated.

The \ln - likelihood will be

$$\begin{aligned} L_n(\beta) &= \sum_{i=1}^m \ln \binom{n}{y_{n,i}} \\ &+ \sum_{i=1}^m y_{n,i} \ln \left(p(\bar{x}_i^h, \beta) \right) + \sum_{i=1}^m (n - y_{n,i}) \ln \left(1 - p(\bar{x}_i^h, \beta) \right) \end{aligned}$$

so that we must maximize

$$g_n(\beta) = \sum_{i=1}^m \frac{y_{n,i}}{n} \ln \left(p(\bar{x}_i^h, \beta) \right) + \sum_{i=1}^m \frac{(n - y_{n,i})}{n} \ln \left(1 - p(\bar{x}_i^h, \beta) \right).$$

We now have the limit function

$$\begin{aligned} g(\beta) &= \sum_{i=1}^m p(\bar{x}_i^h, \beta_0) \ln \left(p(\bar{x}_i^h, \beta) \right) \\ &+ \sum_{i=1}^m \left(1 - p(\bar{x}_i^h, \beta_0) \right) \ln \left(1 - p(\bar{x}_i^h, \beta) \right). \end{aligned}$$

Since, $y_{n,i}/n$ has mean value $p(\bar{x}_i^h, \beta_0)$ and variance $p(\bar{x}_i^h, \beta_0)(1-p(\bar{x}_i^h, \beta_0))/n$, $i = 1, \dots, m$, $A4$ will hold with $\eta_n = n^{-r}$, $0 < r < 1/2$. Assumption $A1$ clearly holds, and for $A2$ to be fulfilled, the statistician has only to choose a compact U large enough to contain ∇ . Moreover, if $p(\bar{x}_i^h, \beta_1) = p(\bar{x}_i^h, \beta_2)$, $i = 1, \dots, m$, implies $\beta_1 = \beta_2$, it suffices to require that there is a neighborhood of β_0 contained in U to ensure that A_1 and A_2 hold, since then, the absolute maximum of $g(\beta)$ is attained at β_0 . In this example, since η_n is known, we can test $A2$.

If we want to obtain confidence ellipsoids and or test $H_0 : J\beta_0 = c$ we must require $p(\bar{x}_i^h, \beta)$ to be thrice differentiable and check if $A5$ holds. We can then use $v_n = n^{-1/2}$, it being interesting to observe that v_n bounds the family of the $\{\eta_n\} = \{n^{-r} : r \in]0, 1/2[\}$. There is a wide range of situations involving independent binomial variables to which our results may be applied, for instance, in connection with probit analysis, see [4], when complex resistance to insecticides has been established, and, see [3] and [10], for the analysis of binary data.

7. AN APPLICATION TO PSEUDO MAXIMUM LIKELIHOOD

This estimation technique was introduced by [5] using results give by [11], [16], [7] and [2]. It is assumed that the m -dimensional observations follow the model

$$\bar{y}_i^m = \bar{f}^m(\bar{x}_i^p, \bar{\beta}_0^k) + \bar{e}_i^m, \quad \text{with } \bar{y}_i^m, \bar{e}_i^m \in \mathbb{R}^m, \bar{x}_i^p \in \mathbb{R}^p \quad \text{and } i \in \mathbb{N}.$$

The parameter space Θ will be contained in \mathbb{R}^k with $\bar{\beta}_0^m$ being the true parameter vector. The error vectors \bar{e}_i^m have null mean value and the conditional distribution of $[\bar{e}_1^m | \dots | \bar{e}_n^m]$ given $X_n = [\bar{x}_1^p, \dots, \bar{x}_n^p]$ is equal to the product of the conditional distributions \bar{e}_i^m given \bar{x}_i^p , $i = 1, 2, \dots, n$. Finally, it is assumed that if $\bar{x}_i^p = \bar{x}_j^p$ then the conditional distributions of \bar{e}_i^m and \bar{e}_j^p are equal, $i \neq j \in \{1, 2, \dots, n\}$. Thus, when X_n is a model matrix, containing the values of controlled variables, the errors are independent and if the controlled vectors are equal the errors are identically distributed.

To avoid duplication we will call generalized densities to the usual densities and to the probability functions. In deriving pseudo maximum-likelihood estimators, linear exponential generalized densities are used. A generalized density is linear exponential if can be written as:

$$l(\bar{u}^m, \bar{\mu}^m) = \exp(A(\bar{\mu}^m) + B(\bar{u}^m) + C(\bar{\mu}^m)' \bar{u}^m)$$

where $A(\vec{\mu}^m)$ and $B(\vec{u}^m)$ are scalars, $C(\vec{\mu}^m)$ is a vector and $\vec{\mu}^m$ is the mean vector of the distribution. We point out that Binomial, Poisson, Negative binomial, Gamma, Normal generalized, Normal multivariate and Multinomial are examples of distributions whose densities are linear exponential. When using one of these generalized densities to derive an estimator the logarithm of the likelihood that we would have, if the observations had that generalized density, is maximized. Since we are not maximizing a true likelihood the technique is called pseudo likelihood.

Actually, (see [5]), instead of the pseudo ln-likelihood it suffices to maximize

$$g_n(\beta) = \frac{1}{n} \sum_{i=1}^n \left[A(f(x_i, \beta)) + C(f(x_i, \beta))' y_i \right]$$

to obtain the pseudo maximum likelihood estimators.

Taking $C^0(\beta|x) = C(f(x_i, \beta))' f(x_i, \beta)$, $g_n^0(\beta) = \frac{1}{n} \sum_{i=1}^n [A(f(x_i, \beta)) + C^0(\beta|x_i)]$ and $e_n^0(\beta) = \frac{1}{n} \sum_{i=1}^n C(f(x_i, \beta))' e_i$, we get the decomposition

$$(7.1) \quad g_n(\beta) = g_n^0(\beta) + e_n^0(\beta).$$

It is important to point out that (see [5]), β_0 is the sole vector in the set ∇_n^0 of absolute maxima of $g_n^0(\beta)$ and the sole zero of $grad(g_n^0(\beta))$. Moreover $g_n(\beta)$ may also have an unique absolute maximum β_n . For instance, see [6], this happens when using the generalized Poisson density to discuss a question that we met, if we take $m = 1$ and $x_i = i$ with $i = 1, 2, \dots$, we get, (see [6])

$$g_n(\beta) = \frac{1}{n} \sum_{i=1}^n [-\exp(i\beta) + i\beta y_i]$$

so that, if $\beta > 0$, we will have

$$g_n^0(\beta) = \frac{1}{n} \sum_{i=1}^n (i\beta - 1) \exp(i\beta) \rightarrow \infty$$

and $g_n(\beta)$ will not converge almost surely to a limit function as is stated in Theorem 1 of [5], although all conditions stated there are fulfilled. Thus to use the very fine technique of pseudo-maximum likelihood we will have to introduce additional assumptions. These will be

A6: The controlled vectors \vec{x}_n^p belong to a compact set D and the parameter space Θ will also be a compact set to. Moreover, the variances $\sigma^2(x)$ and $\lambda(x)$ of e and e^2 , given x , (we restrict ourselves to the case $m = 1$ to make our

point), will be continuous functions of x , as well as, the second derivatives of $A(f(x_i, \beta))$ and $C(f(x_i, \beta))$.

Of course, the first and second order derivatives will be functions of (β, x) defined in the compact set $\Theta \times D$. According to Weierstrass's Theorem, there will be finite maxima for the second and fourth moments of $C(f(X, \beta))$.

We now establish

Proposition 7.1. *If A6 holds and $\beta \in \Theta$ then $e_n^0(\beta) \xrightarrow{a.s.} 0$ and $grad(e_n^0(\beta)) \xrightarrow{a.s.} 0$.*

Proof. The first part of the thesis may be established by proving that the components of $grad(e_n^0(\beta)) \xrightarrow{a.s.} 0$. The j -th component of the gradient will be

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial C(f(x_i, \beta))}{\partial \beta_j} e_i$$

Since, $V[\frac{\partial C(f(x_i, \beta))}{\partial \beta_j} e_i]$ is limited and $E[\frac{\partial C(f(x_i, \beta))}{\partial \beta_j} e_i] = 0$, $i = 1, \dots, n$; $j = 1, \dots, k$, according to the Law of Large Numbers for independent variables, (see [22, page 118])

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial C(f(x_i, \beta))}{\partial \beta_j} e_i \xrightarrow{a.s.} 0, \quad j = 1, \dots, k. \quad \blacksquare$$

With $M_n(\beta)$ and $M_n^0(\beta)$ the hessian matrices of, respectively, $g_n(\beta)$ and $g_n^0(\beta)$, and $\hat{\beta}_n$ a zero of the $grad(g_n(\beta))$, if $\rho(M)$ is the spectral radius of the matrix M , we will have, according to (7.1)

$$(7.2) \quad \begin{aligned} grad(g_n(\tilde{\beta}_n)) &= grad(g_n(\beta_0)) + M_n(\hat{\beta}_n)(\tilde{\beta}_n - \beta_0) \\ &= grad(e_n^0(\beta_0)) + M_n(\hat{\beta}_n)(\tilde{\beta}_n - \beta_0) = 0 \end{aligned}$$

with $\hat{\beta}_n$, such that $\|\hat{\beta}_n - \beta_0\| \leq \|\tilde{\beta}_n - \beta_0\|$.

We now follow a procedure similar to the one used in [13] to establish strong consistency for additive extremal estimators. In order to do it we introduce

A7: An upper bound exists for $\rho(M_n^{-1}(\beta))$ when $\beta \in \Theta$.

Thus, according to Proposition 7.1 and to (7.2), we get

Corollary 7.2. *If A6 and A7 hold then $\tilde{\beta}_n \xrightarrow{a.s.} \beta_0$.*

We point out that, according to (7.2), when A6 and A7 hold, the sole limit, then, thus sole accumulation point of any sequence of roots for $grad(g_n(\beta))$ is, with probability 1, the true parameter value. The zeros of $grad(g_n(\beta))$ include the local extremes of $grad(g_n(\beta))$. The accumulation points of extremes of objective functions are studied under a very general set-up in [9]. Behind the clear-cut result we now obtain is the decomposition (7.1) and the fact that $\nabla_n^0 = \{\beta_0\}$. In practical applications A7 may be difficult to verify. Thus we return to our paper main approach, replacing A7, by

A8: The vectors x_i belong to a finite set $\{\bar{u}_1^p, \dots, \bar{u}_N^p\}$, with u_i being chosen n_i times and $|n_l/n - r_l| \leq d/n$, $i = 1, \dots, N$.

Functions $|A(f(u_i, \beta))|$ and $|C(f(u_i, \beta))|$ will have, for $\beta \in \Theta$, maximums a_i and c_i , respectively. Thus, if A6 and A8 hold

$$(7.3) \quad \sup_{\beta \in \Theta} \left| g_n^0(\beta) - \sum_{j=1}^N r_j [A(f(u_j, \beta)) + C^0(\beta|u_j)] \right| \leq \frac{d}{n} \sum_{j=1}^N a_j + c_j$$

since we now have

$$\sum_{j=1}^N \frac{n_j}{n} [A(f(u_j, \beta)) + C^0(\beta|u_j)].$$

With, $\phi_l = \{i : x_i = u_l\}$ and $s_l = \sum_{i \in \phi_l} e_i$, the vector $\sqrt{n}\bar{s}^N$ with components $\sqrt{n}s_l$, $l = 1, \dots, N$, will have normal asymptotic distribution with null mean vector and covariance matrix D . The principal elements of D will be $r_l \sigma^2 u_l(\beta_0)$, $l = 1, \dots, N$. Thus, see [14], if A8 holds, with $\xi_{(1-q, N)}$ the quantile for probability $1 - q$ of the central chi-square distribution with N degrees of freedom, we will have

$$(7.4) \quad P \left[\bigcap_{(N)} |\bar{v}^{N'} \sqrt{n}\bar{s}^N| \leq \sqrt{\xi_{(1-q, N)} \bar{v}^{N'} D \bar{v}^N} \right] \rightarrow 1 - q$$

in the last expression $\bigcap_{(N)}$ indicates that all $\bar{v}^N \in \mathbb{R}^N$ are considered.

We now establish

Proposition 7.3. *If A6 and A8 hold we may take $\eta_n = n^{-s}$ with $s \in]0, 1/2[$.*

Proof. Considering (7.3) and that $g_n(\beta) - g_n^0(\beta) = e_n^0(\beta) = \sum_{i=1}^N \frac{n_i}{n} C(f(u_i, \beta))' s_i$, the thesis follows directly from (7.4), since the $C(f(u_i, \beta))$ will be bounded in Θ . ■

It is easily seen that A8 may be simplified by only requiring that

$$P \left[\bigcap_{n=m}^{\infty} \bigcap_{l=1}^N \left(\left| \frac{n_l}{n} - r_l \right| \leq \frac{d}{n} \right) \right] \rightarrow 1 - q, \quad \text{as } m \rightarrow \infty.$$

Appendix

A. Ellipsoids and gravity centers

With $B_{k \times k}$ a positive definite matrix and $d \in \mathbb{R}_0^+$, we have the ellipsoid $\xi(B, \vec{a}^k, d) = \{\vec{x}^k : (\vec{x}^k - \vec{a}^k)' B (\vec{x}^k - \vec{a}^k) \leq d\}$.

We can define a subspace of the range space of the $A_{k \times k}$ matrix by $AD = \{A\vec{x}^k : \vec{x}^k \in D\}$ with $D \subseteq \mathbb{R}^k$. If A is a regular matrix, we have $A\xi(B, \vec{a}^k, d) = \xi(A'^{-1}BA^{-1}, A\vec{a}^k, d)$.

P is the orthogonal diagonalizer of B , (about orthogonal diagonalizers, see [15]), if P is orthogonal and $PBP' = \Lambda$, with Λ the diagonal matrix whose principal elements are the eigenvalues $\{p_1, \dots, p_k\}$ of B . The row vectors of P will be the eigenvectors of B . In this case we will have $P\xi(B, \vec{a}^k, d) = \xi(P'^{-1}BP^{-1}, P\vec{a}^k, d) = \xi(\Lambda, P\vec{a}^k, d)$.

It's easy to show that $\xi(\Lambda, Pa, d) = \{\vec{x}^k : (x - Pa)' \Lambda (x - Pa) \leq d\} = \{\vec{x}^k : \sum_{i=1}^n (x_i - b_i)^2 p_i \leq d\}$, with $Pa = b = [b_i]$, $i = 1, \dots, n$.

When using generalized polar coordinates we put: $x_i = b_i + r l_i(\vec{\theta}^{k-1})$, where

$$\left\{ \begin{array}{l} l_1(\vec{\theta}^{k-1}) = \prod_{j=1}^{k-1} \cos \theta_j \\ l_2(\vec{\theta}^{k-1}) = \prod_{j=1}^{k-2} \cos \theta_j \sin \theta_{k-1} \\ \dots \\ l_i(\vec{\theta}^{k-1}) = \prod_{j=1}^{k-i} \cos \theta_j \sin \theta_{k-i+1} \quad i = 3, \dots, k-1 \\ \dots \\ l_k(\vec{\theta}^{k-1}) = \sin \theta_1 \end{array} \right.$$

with $r \in \mathbb{R}_0^+$ the radius and $\vec{\theta}^{k-1}$ the vector of center angles, verifying $\theta_i \in [-\pi/2, \pi/2]$, $i = 1, \dots, k-2$ and $\theta_{k-1} \in [0, 2\pi]$.

For each $\vec{\theta}^{k-1}$ center angles and each ellipsoid we have a semi-axis.

The difference between the lengths of semi-axis associated with $\vec{\theta}^{k-1}$ for ellipsoids $\xi(\Lambda, b, d_i)$, $i = 1, 2$, will be $\sqrt{|d_2 - d_1| / \sum_{i=1}^k p_i l_i^2(\theta)}$.

We now study the location of gravity centers assuming uniform mass densities. The gravity center $GC(D)$ of a set D is completely specified by $GC(PD)$ with P an orthogonal matrix, since these matrices are associated to rotations. Let P be an orthogonal diagonalizer of B , then, if $\xi(B, a, d_1) \subseteq D \subseteq \xi(B, a, d_2)$ with $D^0 = PD$ and $b = Pa$, $\xi(\Lambda, b, d_1) \subseteq D^0 \subseteq \xi(\Lambda, b, d_2)$. Let us decompose D^0 into truncated cylinders with axis going through b and infinitesimal rotation radius. For each $\vec{\theta}^{k-1}$ we will have such a cylinder, along its axis the positive direction will be indicated by $l(\vec{\theta}^{k-1})$, so that, there will be a right and a left part for each cylinder, with lengths $c_1(\vec{\theta}^{k-1})$ and $c_2(\vec{\theta}^{k-1})$ with masses proportional to $c_i(\vec{\theta}^{k-1}) / (c_1(\vec{\theta}^{k-1}) + c_2(\vec{\theta}^{k-1}))$ and gravity centers $b - (-1)^i c_i(\vec{\theta}^{k-1}) l(\vec{\theta}^{k-1}) / 2$, $i = 1, 2$, respectively. The gravity center of the truncated cylinder being

$$\begin{aligned} & \sum_{i=1}^2 c_i(\vec{\theta}^{k-1}) (b - (-1)^i c_i(\vec{\theta}^{k-1}) l(\vec{\theta}^{k-1}) / 2) / \sum_{i=1}^2 c_i(\vec{\theta}^{k-1}) \\ & = b - (c_2(\vec{\theta}^{k-1}) - c_1(\vec{\theta}^{k-1})) / 2. \end{aligned}$$

Now, the extremes of the truncated cylinder will lie in the shell bounded by the pair of ellipsoids, so that, for all $\vec{\theta}^{k-1}$, $d_1 \leq c_i(\vec{\theta}^{k-1}) \leq d_2$, $i = 1, 2$, and $|c_1(\vec{\theta}^{k-1}) - c_2(\vec{\theta}^{k-1})| \leq (d_2 - d_1)$, hence, $GC(\vec{\theta}^{k-1}) \in \xi(\Lambda, b, (d_2 - d_1)/2)$. Since, $GC(D^0)$ is a weighted average of the $GC(\vec{\theta}^{k-1})$ and all $GC(\vec{\theta}^{k-1})$ belong to $\xi(\Lambda, b, (d_2 - d_1)/2)$ we have, $GC(D^0)$ belonging to $\xi(\Lambda, b, (d_2 - d_1)/2)$. Inverting the rotation associated to P we establish:

Proposition 7.4. *If $\xi(B, a, d_1) \subseteq D \subseteq \xi(B, b, d_2)$, then $GC(D) \in \xi(B, b, (d_2 - d_1)/2)$.*

REFERENCES

[1] T. Amemiya, *Advanced Econometrics*, Harvard University Press 1985.
 [2] J. Burguete, R. Gallant and G. Souza, *On Unification of the Asymptotic Theory of Nonlinear Econometric Models*, *Econometric Reviews*, **1**, (1982), 151-190.
 [3] D.R. Cox and E.J. Snell, *The Analysis of Binary Data*, Chapman & Hall, London, 2nd Ed., 1989.

- [4] D. J. Finney, *Probit Analysis*, Cambridge University Press 1971.
- [5] C. Gourieroux, A. Monfort and A. Trognon, *Pseudo Maximum Likelihood Methods: Theory*, *Econometrica* **52** (3), (1984).
- [6] C. Gourieroux, A. Monfort and A. Trognon, *Pseudo Maximum Likelihood Methods: Applications To Poisson Models*, *Econometrica* **52** (3) (1984).
- [7] R. Gallant and A. Holly, *Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation*, *Econometrica* **48** (1980), 697–720.
- [8] J. Hoffman-Jorgensen, *The Theory Of Analytic Sets*, Aarhus Universitet, Matematisk Institut, Various Publications Series, No.10, March 1970.
- [9] J. Hoffman-Jorgensen, *Asymptotic Likelihood Theory*, Aarhus Universitet, Matematisk Institut, Various Publications Series, No. 40, March 1992.
- [10] D.E. Jennings, *Judging Inference Adequacy in Logist Regression*, *Journal of American Statistical Association* **81** (1986), 471–476.
- [11] R. Jennrich, *Asymptotic Properties of Nonlinear Least Squares Estimators*, *The Annals of Mathematical Statistics* **40** (1969), 633–643.
- [12] J.D. Jobson and W.A. Fuller, *Least Squares Estimation When the Covariance Matrix and Parameter Vector Are Functionally Related*, *Journal of the American Statistical Association* **75** (1980), 176–181.
- [13] J.T. Mexia and P. Corte Real, *Strong Law of Large Numbers for Additive Extremum Estimators*, *Discussiones Mathematicae - Probability and Statistics* **21** (2002), 81–88.
- [14] J.T. Mexia, *Asymptotic Chi-squared Tests, Designs, and Log-Linear Models*, *Trabalhos de Investigação No. 1*, 1992.
- [15] J.T. Mexia, *Controlled Heteroscedasticity, Quocient Vector Spaces and F Tets for Hypothesis On Mean Vectors*, *Trabalhos de Investigação No. 1*, 1991.
- [16] E. Malinvaud, *The Consistency of Nonlinear Regressions*, *The Annals of Mathematical Statistics* **41** (1970), 956–969.
- [17] U. Mosco, *Convergence Of Convex Sets Of Solutions Of Variational Inequalies*, *Advances in Mathematics* **3** (1969), 510–585.
- [18] J.T. Oliveira, *Statistical Choice of Univariate Extreme Models*, *Statistical Distributions in Scientific Work*, Taillie et al. (eds) - Reichel, Dordrecht, **6** (1980), 367–382.
- [19] H. Scheffé, *The Analysis of Variance*, John Willey & Sons, New York 1959.
- [20] A. Wald, *Note On The Consistency of The Maximum Likelihood Estiamte*, *Annals of Mathematics and Statistics* **20** (1949), 595–601.

- [21] R.A. Wijsman, *Convergence of Sequences of Convex Sets II*, Transactions of American Mathematical Society **123** (1966), 32–45.
- [22] D. Williams, *Probability with Martingales*, Cambridge Mathematical Textbooks 1991.

Received 24 November 2002

Revised 3 November 2003