

Strongly consistent estimations through optimization

Stime fortemente consistenti tramite ottimizzazione

Silvano Fiorin
Dipartimento di Scienze Statistiche
Università di Padova
florin@stat.unipd.it

Riassunto: La consistenza forte delle stime di massima verosimiglianza viene provata facendo uso di una tecnica che, senza alcun riferimento al rapporto di verosimiglianza, era stata messa a punto ed utilizzata dall'autore anche in ambito nonparametrico. Gli strumenti adottati, e profondamente differenti dall'armamentario classico, permettono di rivisitare in forma immediata aspetti classici e di formulare e studiare problemi finora assai difficilmente attaccabili.

Keywords: MLE estimation, Convex analysis, Uniform convergence.

1. Introduction

The strong consistency of estimates is here studied using an approach which is based on a technical argument taking its origin from convex analysis and before giving the rigorous statements we prefer to present the basic ideas under the intuitive point of view. For this reason the below example is introduced.

1.1 Example

Let us suppose that our purpose consists in estimating the unknown parameter $\theta_0 \in R^1$ on the base of a sequence of observations $\underline{x} = (x_1, x_2, \dots, x_n, \dots)$ and using the function $L_n(x_1, x_2, \dots, x_n, \theta)$ which depends on the first n observed values $x_1, x_2, \dots, x_n$ and on the real variable $\theta \in R^1$. We are very lucky because each $L_n(x_1, x_2, \dots, x_n, \cdot)$ is a strictly concave function in the argument θ and the sequence $\{L_n(x_1, x_2, \dots, x_n, \cdot)\}$ is convergent:

$$\lim_{n \rightarrow \infty} L_n(x_1, x_2, \dots, x_n, \theta) = L(\theta) \quad (1)$$

for each fixed $\theta \in R^1$.

Because of the concavity it descends immediately the concavity for the limit function $L(\cdot)$ and the uniformity of convergence over each compact subset $K \subset R^1$. It is not enough: we suppose that $L(\cdot)$ is strictly concave having θ_0 as unique global maximizer. What would we expect in a so favourable situation? It is evident: because of the strict concavity for each L_n and the uniform approximation to L over each compact K containing its unique maximizer θ_0 , we pretend that L_n has too a unique maximizer $\bar{\theta}_n$ and that

$$\lim_{n \rightarrow \infty} \bar{\theta}_n = \theta_0. \quad (2)$$

A detailed proof of the above limit is based on the following

Property 1 If $L(\cdot)$ is a strictly concave function over all the real line with global (and necessarily unique) maximizer θ_0 , for any fixed closed ball $\bar{S}(\theta_0, \epsilon)$ there exists a positive $\delta(\epsilon) > 0$ such that if

$$|L(\theta) - L(\theta_0)| \leq \delta(\epsilon) \Rightarrow |\theta - \theta_0| \leq \epsilon.$$

The above Property is a particular case of a more general statement whose proof is available in Rockafellar (1972); our purpose consists in understanding the intuitive meaning and in applying it for the strong consistency of estimates.

i) *The meaning.* The statement can be written in the following equivalent version:

for each fixed $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that for any real θ satisfying $|\theta - \theta_0| \geq \epsilon \Rightarrow |L(\theta) - L(\theta_0)| \geq \delta(\epsilon)$.

The Property states that if the distance between θ and θ_0 is made greater than ϵ , then also the corresponding value $L(\theta)$ has at least the distance $\delta(\epsilon)$ from $L(\theta_0)$; or equivalently: the distance between θ and θ_0 can be detected through the values taken by the function $L(\cdot)$. In fact if we have $|L(\theta) - L(\theta_0)| \leq \delta(\epsilon)$ then θ belongs to the ball $\bar{S}(\theta_0, \epsilon)$. Let us observe that the possibility of detecting the distance of a point θ from θ_0 through the values taken by the function $L_n(x_1, x_2, \dots, x_n, \cdot)$, which approximates L very closely, is exactly what we need for estimating θ_0 .

ii) *Applications to estimation.* The Property (1) has to be considered jointly with the uniformity of convergence for $\{L_n\}$ to L over $\bar{S}(\theta_0, \epsilon)$. By Property (1) we have that $|\theta - \theta_0| \geq \epsilon \Rightarrow |L(\theta) - L(\theta_0)| \geq \delta(\epsilon)$ and then, because of the uniform convergence, for given $\frac{\delta(\epsilon)}{4}$ there exists $n_0\left(\frac{\delta(\epsilon)}{4}\right)$ such that

$$\forall n > n_0\left(\frac{\delta(\epsilon)}{4}\right) \Rightarrow \sup_{\theta \in \bar{S}(\theta_0, \epsilon)} |L_n(x_1, x_2, \dots, x_n, \theta) - L(\theta)| < \frac{\delta(\epsilon)}{4}$$

and then $\bar{\theta}_n \in S(\theta_0, \epsilon)$ where $\bar{\theta}_n$ denotes the unique global maximizer for L_n .

Let us recall that in the above elementary proof the two key elements are the property (1) and the uniformity of convergence for $\{L_n\}$ to L ; thus our analysis will be performed on the base of such relevant tools. It is evident that the usual estimation problems lie outside the too favourable context described by the above example: for instance, the global and local concavity for L_n and L is not satisfied in general situations and, consequently, the uniformity of convergence cannot be stated using the simple pointwise convergence

$$L_n(x_1, x_2, \dots, x_n, \theta) \rightarrow L(\theta)$$

for θ fixed.

Nevertheless, the Property(1) and the uniformity of convergence for $\{L_n\}$ are not so restrictive tools to find an application only inside a context of concavity or convexity; it is in order to show the flexibility of Property (1) that the below statements are given and they will be referred in the sequel as Basic Property. The first statement is concerning the case of a concave function and was proved by Rockafellar in his book *Convex Analysis* (see theorem 27.2 page 265).

Basic Property 1 Let $f : R^k \rightarrow R^1$ be a closed proper concave function which has no direction of recession. The supremum of f is then finite and attained. Moreover, for every $\epsilon > 0$, there exists a $\delta > 0$ such that every vector x satisfying $f(x) \geq \sup f - \delta$ lies within the distance ϵ of the maximum set of f (i.e. $|z - x| < \epsilon$ for at least one z such that $f(z) = \sup f$), the maximum set here being a non empty closed bounded convex set.

Basic Property 2 Let $g : X \rightarrow R^1$ be a continuous real function defined over the metric vector space X and with the unique global maximizer θ_0 . Then the restriction of g over any given compact subset K including θ_0 satisfies the following property: for any fixed positive $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that if $\varrho(\theta, \theta_0) \geq \epsilon \Rightarrow |L(\theta) - L(\theta_0)| \geq \delta(\epsilon)$, where ϱ denotes the metric defined over X .

Let us observe that the Basic Property (1) is concerning a property for a concave function f over the whole domain R^k and furthermore, because the not strict concavity for f , the unique maximizer (of Property (1)) is replaced by a maximum set. On the other hand the Basic Property (2) is not concerning a concave context: g is a general continuous function with a unique global maximizer and the stated property is only a local one inside a fixed compact subset. Thus, even if the meaning of Property (1), Basic Properties (1) and (2) is similar, because of the different hypothesis, the three statements are perfectly fitted into different estimation problems; by Property (2) we will study the strong consistency of maximum likelihood estimates. Finally, let us observe that the proof of Basic Property (2) is available in Fiorin (2000) (see Theorem 2 page 726).

2. Strong consistency for MLE

Our purpose consists now in giving a detailed proof of the strong consistency for ML estimates using the Basic Property (2) and the uniform convergence. Furthermore the likelihood ratio method is not involved in the below proof.

Let X be an observable random variable whose probability distribution is described through a density function (a probability density function in the discrete case) $f(x, \theta_0)$, where $x \in R^d$ denotes the values taken by the r.v. X and $\theta_0 \in R^k$ the unknown parameter. The space of parameters is denoted by Θ and we suppose that it denotes any possible subsets of R^k ; nevertheless the proof of strong consistency will be given when $\Theta = R^k$ because in this case some difficulty may arise implying further assumptions with respect to the case when Θ defines a bounded subset of R^k .

On the base of n independent and identically distributed observations belonging to the sequence $\underline{x} = (x_1, x_2, \dots, x_n, \dots)$ and given the product density $f(x_1, \theta) \cdots f(x_n, \theta)$, the supremum is taken $\sup_{\theta \in R^k} f(x_1, \theta) \cdots f(x_n, \theta)$ and then

$$\bar{\theta}_n = \arg \max_{\theta \in R^k} f(x_1, \theta) \cdots f(x_n, \theta)$$

will denote the subset of points $\theta \in R^k$, if they exist, when the above supremum is attained. All the statements and proofs will be given in terms of Euclidean metric denoted by ϱ over R^k and the balls $S(\theta_0, \epsilon)$ are always thought in the ϱ metric.

2.1 The Main Result

Under the below assumptions set, it will be proved for all the sequences of observations $\underline{x} = (x_1, x_2, \dots, x_n, \dots)$ belonging to a set of probability one there exists a corresponding value $n_0(\underline{x}, \epsilon)$, depending on the sequence $\underline{x}$ of observations and on the $\epsilon > 0$ fixed, such that for all $n > n_0(\underline{x}, \epsilon)$ the subset $\bar{\theta}_n$ of the global maximizers for the product density is not empty and furthermore $\bar{\theta}_n \subseteq S(\theta_0, \epsilon)$.

It is important for all the sequel to insist on the meaning of the above statement: when the observations number n is big enough, at least one global maximizer for the corresponding product density always exists and furthermore the set $\bar{\theta}_n$ of all global maximizers, independently of its cardinality, is included in the same ball $S(\theta_0, \epsilon)$. This means that the strong consistency result is valid not only for a particular sequence of global maximizers in such a way that, for each n fixed, we have to select a particular element belonging to the set $\bar{\theta}_n$, but that all the points in $\bar{\theta}_n$ are, at the same time, strongly consistent estimates for the unknown θ_0 .

2.2 Assumptions

Observe that, in the notation adopted below, the symbol E means the expectation with respect to the distribution $F(x, \theta_0)$ and the integrals from $-\infty$ to $+\infty$ mean integration over all the space R^d .

- a) $f(\cdot, \theta)$ is a real random variable, for all θ fixed.
- b) $f(x, \cdot)$ is a continuous and real valued function defined for all $\theta \in R^k$, and this almost surely x with respect to the probability distribution $F(x, \theta_0)$.
- c) $\lim_{|\theta| \rightarrow \infty} f(x, \theta) = 0$ for all x fixed a.s. $F(x, \theta_0)$.
- d) The identifiability condition is introduced: there exists at least one x such that $F(x, \theta_0) \neq F(x, \theta)$ for each $\theta \neq \theta_0$.
- e) Denoting by $\bar{S}^c(\theta_0, r)$ the complement of the ball $\bar{S}(\theta_0, r)$ and given

$$\varphi(x, r) = \sup_{\theta \in \bar{S}^c(\theta_0, r)} \log f(x, \theta)$$

for sufficiently large r , the positive part $\varphi^+(x, r)$ is supposed to have finite expectation

$$\int_{-\infty}^{+\infty} \varphi^+(x, r) dF(x, \theta_0) < \infty.$$

- f.1) The existence is assumed for a ball $\bar{S}(\theta_0, \epsilon)$ such that the following inequalities hold true

$$\int_{-\infty}^{+\infty} \sup_{\theta \in \bar{S}(\theta_0, r)} |\log f(x, \theta)| dF(x, \theta_0) < \infty$$

- f.2) and

$$\int_{-\infty}^{+\infty} \sup_{\theta \in \bar{S}^c(\theta_0, r)} \log f(x, \theta) dF(x, \theta_0) < E \log f(X, \theta_0).$$

The meaning of the above assumptions is here discussed: a) and b) state that the trajectories of the real stochastic process $\log f(x, \cdot)$ for all $\theta \in \bar{S}(\theta_0, r)$ define a measurable random function taking values in the space $C(\bar{S}(\theta_0, r))$ of real and continuous functions defined over $\bar{S}(\theta_0, r)$ and equipped with the Borel σ -field generated by the supnorm $\|g\| = \sup_{\theta \in \bar{S}(\theta_0, r)} |g(\theta)|$, and this with probability one with respect to the underlying distribution $F(x, \theta_0)$.

Assumption d) denotes the general identifiability condition formulated by Wald (1949) and leading to the inequality $E \log f(X, \theta) < E \log f(X, \theta_0)$ for all $\theta \neq \theta_0$. A detailed proof for such an important inequality, referred as the *basic inequality* hereafter, is the argument of Lemma 1 of Wald paper. A careful description is convenient for c), e) and f) which are closely connected; e) and f) introduce the integrability conditions which are

crucial in the proof. Because of c) the limit $\lim_{r \rightarrow \infty} \varphi(x, r) = -\infty$ holds true for all x fixed (a.s. with respect $F(x, \theta_0)$) and then the request $\int_{-\infty}^{+\infty} \varphi^+(x, r) dF(x, \theta_0) < \infty$ is a reasonable one and then, applying the monotone convergence theorem to the negative parts $\varphi^-(x, r)$ it can be proved

$$\lim_{r \rightarrow \infty} \int_{-\infty}^{+\infty} \varphi(x, r) dF(x, \theta_0) = -\infty.$$

Observe that the integrability assumption is given only for the positive part $\varphi^+(x, r)$ and not for $\varphi(x, r)$: thus, for r big enough, the possibility $E\varphi(X, r) = -\infty$ is accepted.

All the conditions collected from a) to e) are in common with Wald proof and the only difference is concerning f). The inequality f.1) appears not only as a replacement but as a strengthening of the integrability condition used by Wald:

$$E|\log f(X, \theta_0)| = \int_{-\infty}^{+\infty} |\log f(x, \theta_0)| dF(x, \theta_0) < \infty.$$

2.3 The proof of the main result

The result introduced in this paragraph is thought inside the space $C(\bar{S}(\theta_0, r))$ of the real continuous functions over $\bar{S}(\theta_0, r)$. The detailed proof is available in Fiorin (2000).

Theorem 1 *Under the integrability condition f.1), the sequence of functions defined for each $\theta \in \bar{S}(\theta_0, r)$*

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta)$$

for $n = 1, 2, \dots$ is uniformly convergent to the function $L(\theta) = E \log f(X, \theta)$ for all $\theta \in \bar{S}(\theta_0, r)$, with probability one with respect to $F(x, \theta_0)$.

Proof of the main result. We begin showing that, under assumptions f.1) and f.2), when n is big enough, the set of maximizers $\bar{\theta}_n$ is not empty and is included in $\bar{S}(\theta_0, r)$. On a set with probability one we have that, by Theorem (1), the sequence $\frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta)$ is uniformly convergent to $L(\theta)$ over $\bar{S}(\theta_0, r)$ and, by the usual law of large numbers for random variables

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \bar{S}^c(\theta_0, r)} \log f(x_i, \theta) &= \int_{-\infty}^{+\infty} \sup_{\theta \in \bar{S}^c(\theta_0, r)} \log f(x, \theta) dF(x, \theta_0) \\ &< E \log f(X, \theta_0). \end{aligned}$$

Observe that, because of the inequality

$$\sup_{\theta \in \bar{S}^c(\theta_0, r)} \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta) \leq \frac{1}{n} \sum_{i=1}^n \sup_{\theta \in \bar{S}^c(\theta_0, r)} \log f(x_i, \theta)$$

the global maximizers of $\frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta)$ over R^k cannot belong to $\bar{S}^c(\theta_0, r)$ but to $\bar{S}(\theta_0, r)$; in such a case the global maximizers over R^k are given by the maximizers of a continuous function over a compact set: thus the set $\bar{\theta}_n$ is not empty. Fixed the ball $S(\theta_0, \epsilon) \subset \bar{S}(\theta_0, r)$, we apply the Basic Property (2): for $\theta \in (\bar{S}(\theta_0, r) - S(\theta_0, \epsilon))$, $L(\theta_0) - L(\theta) > \delta(\epsilon)$ and then it will be sufficient to take n satisfying $\|\frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta) - L(\theta)\| < \frac{\delta(\epsilon)}{4}$ in order that all the global maximizers $\bar{\theta}_n$ lie inside $S(\theta_0, \epsilon)$. And this completes the proof.

3. Applications

Let us suppose to have some more informations about the behaviour of $L(\cdot)$ over the ball $\bar{S}(\theta_0, r)$: for instance it is known that $L(\cdot)$ has not only the global maximizer θ_0 but also a local maximizer θ_1 as an interior point for $\bar{S}(\theta_0, r)$ in such a way that

$$L(\theta_1) > L(\theta) \forall \theta \in \bar{S}(\theta_1, \eta);$$

moreover there exist two disjoint balls $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ included into $\bar{S}(\theta_0, r)$ and the integrability assumption f.1) is satisfied. It is evident that all the procedure shown above may be applied to $\bar{S}(\theta_1, \eta)$:

- i) the Basic Property (2) holds true for $L(\cdot)$ over $\bar{S}(\theta_1, \eta)$;
- ii) the sequence $L_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta)$ is uniformly convergent to $L(\cdot)$ over $\bar{S}(\theta_1, \eta)$, with probability one.

Thus, for n big enough, the set $\bar{\theta}_n^1$ of the global maximizers for the restriction of L_n over $\bar{S}(\theta_1, \eta)$ is not empty and the inclusion is satisfied $\bar{\theta}_n^1 \subseteq S(\theta_1, \epsilon)$ for a fixed positive ϵ with $0 < \epsilon < \eta$. Of course the procedure can be applied also to the ball $\bar{S}(\theta_0, \eta)$: thus, for n big enough, and with probability one, the set $\bar{\theta}_n^0$ of global maximizers for the restriction of L_n over $\bar{S}(\theta_0, \eta)$ is not empty and $\bar{\theta}_n^0 \subseteq S(\theta_0, \epsilon)$. The Main Result of strong consistency for MLE states that, if the elements $\theta_n^1 \in \bar{\theta}_n^1$ and $\theta_n^0 \in \bar{\theta}_n^0$ are arbitrarily chosen, we, may construct two sequences $\{\theta_n^1\}$ and $\{\theta_n^0\}$ such that

$$\lim_{n \rightarrow \infty} \theta_n^1 = \theta_1 \quad \lim_{n \rightarrow \infty} \theta_n^0 = \theta_0.$$

Thus, if the differentiability is introduced for the functions $\log f(x, \cdot)$, we have at the same time the existence of two local maximizers θ_n^1 and θ_n^0 , for L_n , which are solutions to the normal equations system

$$L'_n(\theta_n^1) = 0 \quad L'_n(\theta_n^0) = 0.$$

For each fixed n (big enough) we have to decide between the two sets of solutions $\bar{\theta}_n^1 \subseteq S(\theta_1, \epsilon)$ and $\bar{\theta}_n^0 \subseteq S(\theta_0, \epsilon)$ and furthermore, when n increases, in both the cases we obtain convergent sequences, to θ_1 and to θ_0 respectively. Finally we observe that the possibility of the above situation is not so rare in the concrete problems: the key role is played by the integrability assumption f.1) which allows us to reach the continuity for $L(\cdot)$ and the uniformity of convergence for $\{L_n\}$ over $\bar{S}(\theta_0, r)$ with probability one. Unfortunately the f.1) is a more restrictive assumption if compared with the integrability conditions introduced by Wald and then somebody may suppose that f.1) is too hard to satisfy. Our purpose consists now in showing that the usual hypothesis introduced by Cramer (1946) are sufficient in order to satisfy f.1) and the below statement is concerning the case of a unidimensional parameter $\theta_0 \in R^1$.

3.1 Assumptions

- a) $f(x, \theta)$ satisfies all the conditions a), b), c), d), e) and f.2) inside assumptions (2.2);
- b) as in conditions 1), 2) of Cramer (1946) (page 501), the existence is assumed for the derivatives $\frac{\partial \log f(x, \theta)}{\partial \theta}$, $\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$, $\frac{\partial^3 \log f(x, \theta)}{\partial \theta^3}$ over $\bar{S}(\theta_0, r)$, for each fixed $x \in R^d$ and

$$E \left(\sup_{\theta \in \bar{S}(\theta_0, r)} \left| \frac{\partial^3 \log f(X, \theta)}{\partial \theta^3} \right| \right) < \infty;$$

c) the following expectations are assumed to be finite:

$$E(|\log f(X, \theta_0 - r)|), E\left(\left|\frac{\partial \log f(X, \theta_0 - r)}{\partial \theta}\right|\right), E\left(\left|\frac{\partial^2 \log f(X, \theta_0 - r)}{\partial \theta^2}\right|\right);$$

d) there exists $\eta > 0$ such that

$$\bar{S}(\theta_1, \eta) \subset \bar{S}(\theta_0, r), \bar{S}(\theta_0, \eta) \subset \bar{S}(\theta_0, r), \bar{S}(\theta_1, \eta) \cap \bar{S}(\theta_0, \eta) = \emptyset$$

and

$$\begin{aligned} L(\theta_1) &> L(\theta) \quad \forall \theta \in \bar{S}(\theta_1, \eta) \quad \text{with } \theta \neq \theta_1 \\ L(\theta_0) &> L(\theta) \quad \forall \theta \in \bar{S}(\theta_0, \eta) \quad \text{with } \theta \neq \theta_0, \end{aligned}$$

i.e. θ_1 and θ_0 are local maximizers for L over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ respectively (Moreover θ_0 is also the global maximizer for L over R^1).

Theorem 2 Under the above assumptions (3.1) the integrability condition f.1) is satisfied and, given ϵ (with $0 < \epsilon < \eta$), for all the sequences of observations $\underline{x}$ belonging to a set of probability one, there exists $n_0(\underline{x}, \epsilon)$ such that for all $n > n_0(\underline{x}, \epsilon)$ the sets $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ are not empty and satisfy the inclusions $\bar{\theta}_n^1 \subseteq S(\theta_1, \epsilon)$, $\bar{\theta}_n^0 \subseteq S(\theta_0, \epsilon)$, where $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ denote the global maximizers for the restriction of L_n over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ respectively.

Proof of the theorem. The important argument consists in proving that, under b) and c), the integrability assumption f.1) is satisfied. Such a purpose is obtained through iterative applications of Lagrange theorem for the derivatives; we adopt the simplified notation $g(\theta) = \log f(x, \theta)$ for any fixed $x \in R^d$ and then we have $g(\theta) = g'(\xi)(\theta - (\theta_0 - r)) + g(\theta_0 - r)$ where $\xi \in (\theta_0 - r, \theta)$. Thus the Lagrange theorem is applied again to g' and g'' : $g'(\xi) = g''(\gamma)(\xi - (\theta_0 - r)) + g'(\theta_0 - r)$ where $\gamma \in (\theta_0 - r, \xi)$ and $g''(\gamma) = g'''(\delta)(\gamma - (\theta_0 - r)) + g''(\theta_0 - r)$ where $\delta \in (\theta_0 - r, \gamma)$. Then:

$$\begin{aligned} |g(\theta)| &\leq |g'''(\delta)|8r^3 + |g''(\theta_0 - r)|4r^2 + |g'(\theta_0 - r)|2r + |g(\theta_0 - r)| \\ &\leq \left(\sup_{\theta \in \bar{S}(\theta_0, r)} |g'''(\theta)| \right) 8r^3 + |g''(\theta_0 - r)|4r^2 + |g'(\theta_0 - r)|2r + |g(\theta_0 - r)|. \end{aligned}$$

The last inequality holds true for any $\theta \in \bar{S}(\theta_0, r)$ in the first member and then we obtain that f.1) is satisfied under b) and c). All the assumptions are now satisfied in order to apply the Main Result to the sequence $L_n(\cdot)$ and to the limit function $L(\cdot)$ over $\bar{S}(\theta_0, r)$; the Main Result may be applied separately over the two balls $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ and the theorem is proved for each fixed $0 < \epsilon < \eta$. And this completes the proof.

3.2 Uniqueness for consistent MLE

In all the statements given above, when the optimization was considered for L_n , we had to do with sets of maximizers $\bar{\theta}_n$ and so on ... but until now we never spoke about the possibility of having a unique maximizer (local or global) for L_n over the balls $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ in such a way that $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ contain only one element. In order to study this problem the hypothesis are improved introducing the existence for the first and second derivatives $L'(\cdot)$ and $L''(\cdot)$ over $\bar{S}(\theta_0, r)$ and moreover we assume the strict concavity

for L over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$, i.e. $L''(\theta) < 0$ for each $\theta \in \bar{S}(\theta_1, \eta)$ and for each $\theta \in \bar{S}(\theta_0, \eta)$. Recalling the proofs above, the Main Result states that, for n big enough, the sets $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ of global maximizers for the restriction of L_n over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ are not empty. If we have the possibility to state the convergence for $\{L_n''\}$ to L'' then, for n big enough, $L_n'' < 0$ giving the uniqueness for $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$. Thus what we need is the uniformity of convergence for the derivatives L_n'' which is very easy to obtain applying the theorem (1) for uniformity of convergence directly to the derivatives; given $L_n' = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(x, \theta)}{\partial \theta}$ and $L_n'' = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2}$ it is well known that if $\lambda'(\cdot)$ and $\lambda''(\cdot)$ denote the uniform limit for $\{L_n'\}$ and $\{L_n''\}$ respectively over $\bar{S}(\theta_0, r)$, then $\lambda(\theta) = L'(\theta)$ and $\lambda''(\theta) = L''(\theta)$ for each $\theta \in \bar{S}(\theta_0, r)$. In order to state the below result we suppose that all the assumptions (3.1) are satisfied and that d) is improved with the following

Assumption d'). The function $L(\cdot)$ satisfies the following condition:

$$\sup_{\theta \in \bar{S}(\theta_1, \eta) \cup \bar{S}(\theta_0, \eta)} L''(\theta) < 0.$$

Theorem 3 Under the Assumptions (3.1) and d'), the following statements hold true:

- i) the function $L(\cdot)$ admits first and second derivatives satisfying $L'(\theta) = E\left(\frac{\partial \log f(X, \theta)}{\partial \theta}\right)$ and $L''(\theta) = E\left(\frac{\partial^2 \log f(X, \theta)}{\partial \theta^2}\right)$ for each $\theta \in S(\theta_0, r)$.
- ii) Given $\epsilon > 0$ (with $0 < \epsilon < \eta$) and for all the sequences belonging to a set with probability one, there exists $n_0 = (x, \epsilon)$ such that for each $n > n_0(x, \epsilon)$ the sets $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ of global maximizers for the restriction of L_n over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ respectively contain both a unique element and $\bar{\theta}_n^1 \in \bar{S}(\theta_1, \epsilon)$ and $\bar{\theta}_n^0 \in S(\theta_0, \epsilon)$.
- iii) The statement ii) above implies that there exists only one sequence $\{\bar{\theta}_n^0\}$ of strongly consistent maximum likelihood estimates for θ_0 .

A statement which is the analogous of Theorem (3) can be given for the case $\theta_0 \in R^k$ under the following assumptions.

3.3 Assumptions

- a) all the conditions a), b), c), d), e), f.2) belonging to assumptions (2.2) hold true.
- b) For each fixed $x \in R^d$ the first partial derivatives $\frac{\partial \log f(x, \theta)}{\partial \theta_j} : j = 1, \dots, k$ and the second partial derivatives $\frac{\partial^2 \log f(x, \theta)}{\partial \theta_j \partial \theta_l} : j = 1, \dots, k; l = 1, \dots, k$ are supposed to exist and define continuous functions jointly in the k real variables over $\bar{S}(\theta_0, r)$.
- c) The following expectations are supposed to be finite:

$$E\left(\sup_{\theta \in \bar{S}(\theta_0, r)} |\log f(X, \theta)|\right), \quad E\left(\sup_{\theta \in \bar{S}(\theta_0, r)} \left|\frac{\partial \log f(X, \theta)}{\partial \theta_j}\right|\right) \quad \text{for } j = 1, \dots, k,$$

$$E\left(\sup_{\theta \in \bar{S}(\theta_0, r)} \left|\frac{\partial^2 \log f(X, \theta)}{\partial \theta_j \partial \theta_l}\right|\right) \quad \text{for } j = 1, \dots, k; \quad l = 1, \dots, k.$$

d'') The condition d) of assumptions (3.1) is satisfied and furthermore

$$\inf_{\theta \in \bar{S}(\theta_1, \eta) \cup \bar{S}(\theta_0, \eta)} D(H_t(\theta)) > 0 \quad \text{with } t \text{ even}$$

$$\sup_{\theta \in \bar{S}(\theta_1, \eta) \cup \bar{S}(\theta_0, \eta)} D(H_t(\theta)) < 0 \quad \text{with } t \text{ odd.}$$

where $D(H_t)$ denotes the determinant of each Hessian submatrix for $t = 1, \dots, k$.

Theorem 4 *Under the Assumptions (3.3), the following statements hold true:*

- i) *the function $L(\cdot)$ admits first and second partial derivatives satisfying $L'_j(\theta) = E\left(\frac{\partial \log f(X, \theta)}{\partial \theta_j}\right)$ for $j = 1, \dots, k$ and $L''_{jl}(\theta) = E\left(\frac{\partial^2 \log f(X, \theta)}{\partial \theta_j \partial \theta_l}\right)$ for $j = 1, \dots, k$ and $l = 1, \dots, k$.*
- ii) *The same statement as in ii) of Theorem (3).*
- iii) *The same statement as in iii) of Theorem (3).*

Sketch of the proof for Theorems (3) and (4). The proof, in both the cases, is the result of two steps:

1. the Main Result may be applied over the balls $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ and then for each sequence of observations $\underline{x}$ belonging to a set with probability one there exists $n_1(\underline{x}, \epsilon)$ such that the sets $\bar{\theta}_n^1$ and $\bar{\theta}_n^0$ of the global maximizers for the restriction of L_n over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$ are not empty and $\bar{\theta}_n^1 \subseteq S(\theta_1, \epsilon)$, $\bar{\theta}_n^0 \subseteq S(\theta_0, \epsilon)$.
2. Because of the uniformity of convergence for the second derivatives of the function $L_n(\theta)$ over $\bar{S}(\theta_o, r)$, for each sequence of observations belonging to a set of probability one, there exists $n_2(\underline{x})$ such that if $n > n_2(\underline{x})$, then L_n defines a strictly concave function over $\bar{S}(\theta_1, \eta)$ and $\bar{S}(\theta_0, \eta)$. Taking $n_0(\underline{x}, \epsilon) = \max\{n_1(\underline{x}, \epsilon), n_2(\underline{x})\}$ the theorems are proved.

Remarks

1. The existence of a unique consistent sequence $\{\bar{\theta}_n^0\}$ of MLE is an important problem in the literature; the first result is due to Huzurbazar (1948) and is concerning the unidimensional parameter context. The extension of such a result to the multidimensional case generated very serious problems: Chanda tried an extension in 1954 and only in 1975 Tarone and Gruenhage showed that Chanda's proof was wrong; in the introduction of their paper they write: "if the parameter of interest is vector valued, such problem can be extremely difficult". A modification of the method adopted by Tarone and Gruenhage was given by Foutz. The type of proof given above by Theorems (3) and (4) differs consistently from the solutions available in the literature: in fact no difference is made between unidimensional and multidimensional parameter.
2. The existence, for the limit function $L(\theta) = E(\log f(X, \theta))$, of local maximizers, different from θ_0 , generates very serious problems. Of course it is interesting to know concrete situations where the sets of local maximizers for $L(\cdot)$ does not contain the unique element θ_0 . Such a possibility can be obtained using mixtures of densities.
3. Further Applications: the Basic Property and the uniformity of convergence allow us to derive strongly consistent estimates also in nonparametric contexts like, for instance, the Cox regression model for counting processes with time dependent coefficients as in Fiorin (1996) and Fiorin (1998).

References

- Chanda K. (1954) A note on the consistency and maxima of the roots of likelihood equations, *Biometrika*, 41, 55–61.
- Cramer H. (1946) *Mathematical methods of statistics*, Princeton University Press.
- Fiorin S. (1996) Optimization for the Cox partial likelihood functional, *C. R. Acad. Sci. Paris, Serie I*, 323, 529–534.
- Fiorin S. (1998) Strongly consistent functional estimation for a Cox regression model for counting processes, *Mathematical Methods of Statistics*, 7, 319–338.
- Fiorin S. (2000) The strong consistency for maximum likelihood estimates: a proof not based on the likelihood ratio, *C. R. Acad. Sci. Paris, Serie I*, 331, 721–726.
- Foutz R. (1977) On the unique consistent solution to the likelihood equations, *JASA*, 72, 147–148.
- Huzurbazar V. (1948) The likelihood equation, consistency and the maxima of the likelihood function, *Annals of Eugenics*, 14, 185–200.
- Rockafellar R. (1972) *Convex Analysis*, Princeton University Press.
- Tarone R. and Gruenhage G. (1975) A note on the uniqueness of roots of likelihood equations for vector-valued parameters, *JASA*, 70, 903–904.
- Wald A. (1949) Note on the consistency of the maximum likelihood estimate, *Annals of Math. Stat.*, 20, 595–601.