SCIENTIA
Series A: *Mathematical Sciences*, Vol. 27 (2016), 1–14
Universidad Técnica Federico Santa María
Valparaíso, Chile
ISSN 0716-8446

Maximum Likelihood Estimators Under Continuity-Compactness Assumptions

Julio Saucedo-Zul and Haydée De la Garza Rodríguez

Abstract. This work concerns with the consistency property of maximum likelihood estimators in a parametric statistical model. Assuming that the parameter space is compact and that the density function is Lipschitz continuous on the parameter, it is shown that the maximum likelihood technique generates estimators that, as the sample size increases, converge to the true parameter value with probability 1. The objective of the analysis is to illustrate the application of three basic statistical and analytical results: the law of large numbers, Jensen's inequality, and the Heine-Borel property of compact sets

Key words: Consistency of estimators, Concave function, Law of large numbers, Convergence almost surely, Closed and bounded set, Open set.

1. Introduction

This note is concerned with the maximum likelihood method for parameter estimation, and the main objective is to establish the consistency of the estimators obtained from that technique. In contrast with the classical approach on the subject [7, 11], the arguments in this paper are based on the *continuity propertiest* of the underlying density, but not on its differentiability. On the other hand, it is known that a maximizer of a continuous function does not necessarily exists unless its domain is compact [5, 8], and for this reason, in this work the compactness of the parameter space is assumed. Under mild Lipschitz continuity conditions, the main result of the h paper, formulated as Theorem 3.1 of Section 3, establishes the consistency of the sequence of maximum likelihood estimator The proof of this theorem is based on three fundamental results:

(i) The law of large numbers, which states that, as the sample size increases, the average of independent random variables with common distribution converges with probability 1 to the population mean; see, for instance, [1, 3, 6, 10].

2000 *Mathematics Subject Classification.* Primary 62F12, Secondary 26A03.
Key words and phrases. Consistency of estimators, concave function, law of large numbers, almost sure convergence, closed and bounded set, open set.

(ii) Jensen's inequality, which relates the ideas of expectation and concave function [2, 10], and

(iii) The Heine-Borel theorem in an Euclidean space. This result is a fundamental property of the real number system and is also known as the finite-sub-covering theorem.

The subsequent material has been organized as follows: In section 2 the idea of consistent sequence of estimators is introduced. In Section 3, the maximum likelihood method is formulated, and it is shown that it renders consistent estimators when the parameter space is finite. Next, in Section 4, the continuity and compactness assumptions are introduced, and the consistency result under those assumptions is formulated as Theorem 5.1 in Section 5. Finally, the presentation concludes in Section 6 with a proof of this last theorem.

2. Consistency

Let $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, ...$ be a sequence of independent and identically distributed k-dimensional random vectors, and assume that their common distribution has a density $f(\mathbf{y}, \theta)$ with respect to a fixed (σ-finite) measure ρ on the Borel sets of $\mathbb{R}^k$; ρ is the the Lebesgue when the vectors $\mathbf{Y}_i$ have a continuous distribution, whereas ρ is a counting measure in the discrete case. On the other hand, θ stands for an unknown parameter that belongs to a *parameter space* Θ contained in a Euclidean space $\mathbb{R}^d$. Thus, the observer does not know exactly the distribution of the vectors $\mathbf{Y}_i$, but only knows that it has a density (with respect to ρ) that belongs to the family $\{f(\mathbf{y}; \theta) \,|\, \theta \in \Theta\}$. In what follows, $\theta_0 \in \Theta$ denotes the true value of the parameter, so that the common density of the vectors $\mathbf{Y}_i$ is $f(\mathbf{y}; \theta_0)$; however, θ_0 is not known to the observer and the main objective is to use the observed data to estimate the value of θ_0 or, more generally, the value of a function $g(\theta_0)$. An estimator of the unknown value $g(\theta_0)$ based on $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n$ is a function

$$\hat{g}_n \equiv \hat{g}_n(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n),$$

which will be used by the observer as 'an approximation' of $g(\theta_0)$. This idea is quite general, and several techniques can be used to generate an estimator $\hat{g}_n$, for instance, the methods of moments, frequency substitution, or maximum likelihood [2, 3, 6]. Regardless of the approach used to construct the estimators $\hat{g}_n$, a reasonable requirement on the sequence $\{\hat{g}_n\}$ is that, as the number n of observations increases, the values of $\hat{g}_n(\mathbf{Y}_n, ..., \mathbf{Y}_n)$ 'converge' to $g(\theta_0)$; in that case, the sequence of estimators $\{\hat{g}_n\}$ is called *consistent*. This idea can be formalized in several ways, and in this work the following formulation is used.

Definition 2.1. Let $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3...$ be a sequence of independent random vectors which are identically distributed, and suppose that their common distribution has density $f(\mathbf{y}; \theta)$, where $\theta \in \Theta$. Denote by θ_0 the true value of the parameter and let P_{θ_0} be the distribution corresponding to $f(\mathbf{y}; \theta_0)$. Given a function $g(\theta)$, a sequence $\{\hat{g}_n(\mathbf{Y}_1, ..., \mathbf{Y}_n)\}$ is consistent if, and only if,

$$P_{\theta_0}\left[\lim_{n\to\infty} \hat{g}_n(\mathbf{Y}_1, ..., \mathbf{Y}_n) = g(\theta_0)\right] = 1.$$

This idea is referred to as *strong consistency* in the literature. On the other hand, the weak consistency property (or consistency in probability), requires that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} P_{\theta_0}\left[|\hat{g}_n(\mathbf{Y}_1, ..., \mathbf{Y}_n) = g(\theta_0)| > \varepsilon\right] = 0.$$

For details about these consistency notions see [3, 7, 8]. It is known hat if a sequence $\{\hat{g}_n\}$ is consistent in the sense of Definition 2.1, then it is consistent in probability, so that the notion in the above definition is, effectively, stronger than the idea of weak consistency. A basic tool in the analysis if the idea of consistency is the the following (strong) law of large numbers, whose proof is given, for instance, in [1].

THEOREM 2.1. *Let $X_1, X_2, X_3, ...$ be independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, and suppose that $\mu = E[X_i]$ is well-defined. In this context, there exists an event Ω^* such that*

(i) $P[\Omega^*] = 1$, and

(ii) For each $\omega \in \Omega^*$, $\displaystyle\lim_{n\to\infty} \frac{X_1(\omega) + X_2(\omega) + ... + X_n(\omega)}{n} = \mu.$

3. Maximum Likelihood Estimation

In this section the maximum likelihood method is introduced. The basic conditions for the analysis below are that all of the densities in the model have the same support, and that different values of the parameter correspond to different distributions.

ASSUMPTION 3.1. For each $\theta, \theta_1 \in \Theta$ with $\theta \neq \theta_1$, the following properties (i) and (ii) hold:

(i) The densities $f(\cdot;\theta_0)$ and $f(\cdot;\theta_1)$ have the same support, that is,

$$\{\mathbf{y} \in \mathbb{R}^k \mid f(\mathbf{y}, \theta) > 0\} = \{\mathbf{y} \in \mathbb{R}^k \mid f(\mathbf{y}, \theta_1) > 0\};$$

(ii) [Identifiability.] There exist a (Borel set) $A \subset \mathbb{R}^k$ such that

$$\int_A f(\mathbf{y};\theta)\,\rho(\mathbf{dy}) \neq \int_A f(\mathbf{y};\theta_1)\,\rho(\mathbf{dy}).$$

As usual, $E_\theta[H(\mathbf{Y})]$ stands for the expected value of the random variable $H(\mathbf{Y})$ under the assumption that the density of $\mathbf{Y}$ is $f(\mathbf{Y};\theta)$. In order to obtain information about the true parameter value, the observer takes a sample $(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, ..., \mathbf{Y}_n)$, whose density belongs to the family

$$f_n(\mathbf{y}_1, ..., \mathbf{y}_n;\theta) = \prod_{i=1}^{n} f(\mathbf{y}_i;\theta), \quad \theta \in \Theta. \tag{3.1}$$

Note that the true (but unknown) density of $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n)$ is

$$f_n(\mathbf{Y}_1, ..., \mathbf{Y}_n;\theta_0) = \prod_{i=1}^{n} f(\mathbf{Y}_i;\theta_0). \tag{3.2}$$

The maximum likelihood method of estimation is motivated by the following theorem, whose proof relies on the law of large numbers in Theorem 2.1 and on the following (strict) Jensen's inequality [3, 9]: If W is a non-constant random variable taking values in the interval J, and the real-alued function φ is *strictly concave* on J, then

$$E[\varphi(W)] < \varphi\left(E[W]\right) \tag{3.3}$$

THEOREM 3.1. *(i) For each $\theta \neq \theta_0$,*

$$E_{\theta_0}\left[\log\left(\frac{f(\boldsymbol{Y};\theta)}{f(\boldsymbol{Y};\theta_0)}\right)\right] =: \nu(\theta) < 0.$$

(ii) With probability 1 respect to P_{θ_0},

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\frac{f_n(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n;\theta)}{f_n(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n;\theta_0)}\right) = \nu(\theta).$$

(iii) There exists an event Ω^* satisfying $P[\Omega^*] = 1$, for which the following property holds: For each $\omega \in \Omega^*$, there exists an integer $N(\omega;\theta)$ such that

$$f_n(\boldsymbol{Y}_1(\omega),\ldots,\boldsymbol{Y}_n(\omega);\theta) < f_n(\boldsymbol{Y}_1(\omega),\ldots,\boldsymbol{Y}_n(\omega);\theta_0), \quad n > N(\omega;\theta).$$

PROOF. (i) Observe that for each $\theta \neq \theta_0$, Assumption 3.1 implies that $f(\mathbf{Y};\theta)/f(\mathbf{Y};\theta_0)$ is a non-constant random variable, whose values belong to the interval $J = (0,\infty)$ with probability 1 with respect to P_{θ_0}. Thus, applying Jensen's inequality in (3.3) with the strictly concave function $\varphi(w) = \log(w)$ for $w > 0$, it follows that

$$E_{\theta_0}\left[\log\left(\frac{f(\mathbf{Y};\theta)}{f(\mathbf{Y};\theta_0)}\right)\right] < \log\left(E_{\theta_0}\left[\frac{f(\mathbf{Y};\theta)}{f(\mathbf{Y};\theta_0)}\right]\right). \tag{3.4}$$

Observing that

$$E_{\theta_0}\left[\frac{f(\mathbf{Y};\theta)}{f(\mathbf{y};\theta_0)}\right] = \int\left[\frac{f(\mathbf{y};\theta)}{f(\mathbf{y};\theta_0)}\right] f(\mathbf{y};\theta_0)\,d\mathbf{y} = \int f(\mathbf{y};\theta)\,d\mathbf{y} = 1,$$

it follows that the right hand side of (3.4) is zero, establishing the desired conclusion. (ii) A glance at (3.1) and (3.2) yields that

$$\log\left(\frac{f_n(\mathbf{Y}_1,\ldots,\mathbf{Y}_n;\theta)}{f_n(\mathbf{Y}_1,\ldots,\mathbf{Y}_n;\theta_0)}\right) = \log\left(\frac{\prod_{i=1}^n f(\mathbf{Y}_i;\theta)}{\prod_{i=1}^n f(\mathbf{Y}_i;\theta_0)}\right) = \sum_{i=1}^n \log\left(\frac{f(\mathbf{Y}_i;\theta)}{f(\mathbf{Y}_i;\theta_0)}\right); \tag{3.5}$$

additionally, the random variables $X_i = \log\left(\dfrac{f(\mathbf{Y}_i;\theta)}{f(\mathbf{Y}_i;\theta_0)}\right)$ are independent and identically distributed with respect to P_{θ_0} and their common expectation is

$$E_{\theta_0}[X_i] = E_{\theta_0}\left[\log\left(\frac{f(\mathbf{Y}_i;\theta)}{f(\mathbf{Y}_i;\theta_0)}\right)\right] = \nu(\theta) < 0,$$

where the inequality is due to part (i). Therefore, the law of large numbers in Theorem 2.1 implies that there exists an event Ω^* with $P_{\theta_0}[\Omega^*] = 1$, such that

$$\nu(\theta) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i(\omega) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \log\left(\frac{f(\mathbf{Y}_i(\omega);\theta)}{f(\mathbf{Y}_i(\omega);\theta_0)}\right), \quad \omega \in \Omega^*,$$

a relation that via (3.1) leads to

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\frac{f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta)}{f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta_0)}\right)=\nu(\theta),\quad \omega\in\Omega^*.$$

This completes the argument since, as already mentioned, $P[\Omega^*]=1$.

(iii) Let Ω^* be the event in part (ii). Combining the above display with the fact that $\nu(\omega)<0$, the definition of limit yields that, for each $\omega\in\Omega^*$, there exists an integer $N(\omega,\theta)$ such as

$$\frac{1}{n}\log\left(\frac{f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta)}{f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta_0)}\right)<0,\quad n>N(\omega,\theta),$$

a relation that is equivalent to

$$\log\left(f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta)\right)<\log\left(f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta_0)\right),\quad n>N(\omega,\theta),$$

that is, $f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta)<f_n(\mathbf{Y}_1(\omega),...,\mathbf{Y}_n(\omega);\theta_0)$ for all $n>N(\omega,\theta)$, completing the proof. □

Given the observed data $\mathbf{Y}_1,...,\mathbf{Y}_n$, define the corresponding *likelihood function* on Θ by

$$L_n(\theta;\mathbf{Y}_1,...,\mathbf{Y}_n)=f_n(\mathbf{Y}_1,\mathbf{Y}_2,...,\mathbf{Y}_n;\theta),\quad \theta\in\Theta. \tag{3.6}$$

If the vectors $\mathbf{Y}_i$ are discrete, the value of $L_n(\theta;\mathbf{y}_1,...,\mathbf{y}_n)$ is the probability of observing the event $[\mathbf{Y}_1=\mathbf{y}_1,\mathbf{Y}_2=\mathbf{y}_2,...,\mathbf{Y}_n=\mathbf{y}_n]$ when θ is the parameter value. The fundamental fact established in part (iii) of the previous theorem is that, for n large enough, the likelihood function $L_n(\theta;\mathbf{Y}_1,...,\mathbf{Y}_n)$ is not a maximized at $\theta\neq\theta_0$, so that it is reasonable to find good estimators of the unknown value of θ_0 among the maximizers of the likelihood function.

Definition 3.1 (Maximum Likelihood Estimators.). Let $\mathbf{Y}_1,\mathbf{Y}_2,...,\mathbf{Y}_n$ be a sample of a distribution with density $f(\mathbf{y};\theta)$, where $\theta\in\Theta$. An estimator

$$\hat\theta_n\equiv\hat\theta_n(\mathbf{Y}_1,...,\mathbf{Y}_n)\in\Theta$$

is a *maximum likelihood estimator* if $\hat\theta_n$ maximizes the mapping $\theta\mapsto L(\theta;Y_1,Y_2,...,\mathbf{Y}_n)$, $\theta\in\Theta$, that is,

$$L_n(\hat\theta_n(\mathbf{Y}_1,...,\mathbf{Y}_n);\mathbf{Y}_1,...,\mathbf{Y}_n)\geqslant L_n(\theta;\mathbf{Y}_1,...,\mathbf{Y}_n),\quad \theta\in\Theta$$

The following result shows that, if the parameter space is finite, then a sequence $\{\hat\theta_n\}$ of maximum likelihood estimators is consistent. Note that the likelihood function in (3.6) always has a maximizer when its domain Θ is finite.

Theorem 3.2. *Let $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2,...,\boldsymbol{Y}_n$ be a sample of a distribution with density $f(\boldsymbol{y};\theta)$, where $\theta\in\Theta$, where the parameter space Θ is finite. Denote by $\theta_0\in\Theta$ the true parameter value, and suppose that the Assumption 3.1 holds. In this context, there exists an event Ω^* satisfying the following properties (i) and (ii): (i) $P_{\theta_0}[\Omega^*]=1$, (ii) For each $\omega\in\Omega^*$, exists an integer $N(\omega)$ such as*

$$\hat\theta_n(\boldsymbol{Y}_1(\omega),...\boldsymbol{Y}_n(\omega))=\theta_0,\quad n>N(\omega).$$

This results shows that, with probability 1, the maximum likelihood estimator coincides with the true parameter value when the sample size n is large enough, and then the sequence $\{\hat{\theta}_n\}$ of maximum likelihood estimators converges with probability 1 to the true value of the parameter θ_0.

Proof. Write

$$\Theta = \{\theta_0, \theta_1, ..., \theta_r\}, \tag{3.7}$$

where θ_0 is the true value of the parameter.
Using Theorem 3.1 (iii) for each $i = 1, 2, ..., r$ there exists an event Ω_i^* such as
(a) $P_{\theta_0}[\Omega_i^*] = 1$, and
(b) For each $\omega \in \Omega_i^*$ exists an integer $N(\omega, \theta_i)$ such that

$$f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta_i) < f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta_0), \quad n > N(\omega, \theta_i).$$

By (3.6), this relation is equivalent to

$$L_n(\theta_i; \mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) < L_n(\theta_0; \mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)), \quad n > N(\omega, \theta_i),$$

and then Definition 3.1 yields that

$$\text{For } \omega \in \Omega_i^*, \quad \hat{\theta}_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) \neq \theta_i \quad \text{if} \quad n > N(\omega, \theta_i) \tag{3.8}$$

Now, define

$$\Omega^* : = \bigcap_{i=1}^{r} \Omega_i^*,$$

and note that $P_{\theta_0}[\Omega^*] = 1$, since $P_{\theta_0}[\Omega_i^*] = 1$ for $i = 0, 1, ..., r$. Next, recalling that $N(\omega, \theta_i) < \infty$ for $\omega \in \Omega_i^*$, set

$$N(\omega) : = \max_{i=1,2...,r} N(\omega, \theta_i) < \infty.$$

Now, select $\omega \in \Omega^*$ and take $n > N(\omega)$. In this case $\omega \in \Omega_i^*$ and $n > N(\omega, \theta_i)$ for each $i = 1, 2, ..., r$, and then (3.8) implies that $\hat{\theta}_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) \neq \theta_i$; since the maximizer $\hat{\theta}_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega))$ belongs to Θ, it follows form (3.7) that if $\omega \in \Omega^*$ and $n > N(\omega)$ then $\hat{\theta}_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) = \theta_0$.

This completes the proof since, as already observed $P_{\theta_0}[\Omega^*] = 1$, and the integer $N(\omega)$ is finite. □

By Theorem 3.2, the true parameter θ_0 is estimated consistently by the sequence $\{\hat{\theta}_n\}$ of maximum likelihood estimators. The remainder of the paper is dedicated to extend this conclusion to the case of a compact parameter space.

4. The Continuity-Compactness Framework

In this section, additional conditions on the parameter space Θ and the likelihood function $L(\cdot; \mathbf{Y}_1, ..., \mathbf{Y}_N)$ are introduced. Under the two assumptions stated below, the maximum likelihood estimator $\hat{\theta}_n$ in Definition 3.1 is always well-defined, and it is possible to establish the consistency of the sequence $\{\hat{\theta}_n\}$, Recall that a nonempty subset $C \subset \mathbb{R}^d$ is if, for every convergent sequence $\{c_n\} \subset C$, the corresponding limit also belongs to C. On the other hand, $C \subset \mathbb{R}^k$ is *bounded* if there exists a nonnegative constant B such as $\|c\| \leqslant B$ for all $c \in C$. A subset C of $\mathbb{R}^d$ is *compact* if C is both closed and bounded.

ASSUMPTION 4.1. The parameter space Θ is a compact subset of $\mathbb{R}^d$.

The main property of compact subsets of $\mathbb{R}^d$ is stated in the next theorem, where the following notation is used: For each point $\mathbf{x} \in \mathbb{R}^d$ and $\varepsilon > 0$, the ball with center $\mathbf{x}$ and radius $\varepsilon > 0$ is defined by

$$B(\mathbf{x}, \varepsilon) \colon = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y} - \mathbf{x}\| < \varepsilon\}. \tag{4.1}$$

THEOREM 4.1. *Let C be a compact subset of $\mathbb{R}^d$. Given $F \subset \mathbb{R}^d$, for each $\boldsymbol{x} \in F$ let $\varepsilon_{\boldsymbol{x}}$ be a positive number, and assume that*

$$C \subset \bigcup_{\boldsymbol{x} \in F} B(\boldsymbol{x}, \varepsilon_{\boldsymbol{x}}).$$

In this case, exists a finite subset $\{\boldsymbol{x}_1, \boldsymbol{x}_2 ..., \boldsymbol{x}_m\} \subset F$ such as

$$C \subset \bigcup_{i=1}^{m} B(\boldsymbol{x}_i, \varepsilon_{\boldsymbol{x}_i}).$$

This result, which is a fundamental property of the system of the real number, is known as the Heine-Borel theorem, or the *finite sub-covering* theorem; a proof can be found, for instance, in [4, 8]. The following additional requirement will be used to ensure the existence of maximum likelihood estimators $\hat{\theta}_n$, and to establish the consistency of the sequence $\{\hat{\theta}_n\}$, Note that the support

$$\mathcal{Y} \colon = \{y \colon f(\mathbf{y}; \theta) > 0\}$$

is the same for each $\theta \in \Theta$, by Assumption (3.1) (i).

ASSUMPTION 4.2. There exists a function $B \colon \mathcal{Y} \to [0, \infty)$ satisfying the the following properties (i) and (ii): (i) For every $y \in \mathcal{Y}$,

$$|\log f(\mathbf{y}; \theta) - \log f(\mathbf{y}; \theta_1)| \leqslant \|\theta - \theta_1\| B(\mathbf{y}), \quad \theta, \theta_1 \in \Theta. \tag{4.2}$$

(ii) The random variable $B(\mathbf{Y})$ has a finite expectation value regardless of the true parameter value, i.e.,

$$\mu_B(\theta) \colon = E_\theta[B(\mathbf{Y})] < \infty, \quad \theta \in \Theta. \tag{4.3}$$

Condition (4.2) states that $\log(f(\mathbf{y}; \theta))$ is a *Lipschitz continuous function* of θ, and $B(\mathbf{y})$ is referred to as a Lipschitz constant for $\log(f(\mathbf{y}; \cdot))$. On the other hand, condition (4.3) makes it possible to use the law of large.numbers to analyze the consistency of the sequence of maximum likelihood estimators. Examples satisfying Assumption (4.39 are discussed below.

EXAMPLE 4.1. Consider the family of densities on $\mathbb{R}^s$ given by

$$f(\mathbf{y}; \xi) = a(\xi)^{-1} h(\mathbf{y}) e^{\xi' T(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^s, \quad \xi \in \Xi$$

where $h(\cdot)$ is a nonnegative function on $\mathbb{R}^s$, $T(\mathbf{y}) = [T_1(\mathbf{y}), T_2(\mathbf{y}), ..., T_k(\mathbf{y})]'$ is a mapping from $\mathbb{R}^s$ to $\mathbb{R}^k$, $\xi \in \mathbb{R}^k$, and

$$a(\xi) := \int_{\mathbb{R}^k} h(\mathbf{y}) e^{\xi' T(\mathbf{y})} \, d\mathbf{y},$$

whereas Ξ consists of all the vectors in $\mathbb{R}^k$ for which $a(\xi)$ is positive and finite. It is assumed the Ξ is non empty, and in this case it follows that Ξ is a convex set [2, 6]; without loss of generality, it is supposed that the interior Ξ° of Ξ is non-empty. The set $\{f(\mathbf{y}; \Xi) \colon \xi \in \Xi\}$ is a k-dimensional exponential family, and the following properties are valid at each interior point ξ_0 of Ξ [6]: (i) Each component $T_i(\mathbf{Y})$ of $T(\mathbf{Y})$ has finite expectation with respect to $f(\cdot; \xi_0)$, that is,

$$E_{\xi_0}[|T_i(\mathbf{Y})|] < \infty, \quad i = 1, 2, ..., k; \tag{4.4}$$

(ii) The function $a(\cdot)$ is differentiable at ξ_0, and

$$\partial_{\xi_r} a(\xi_0) = a(\xi_0) E_{\xi_0}[T_r(Y)], \quad r = 1, 2, ..., k;$$

(iii) The partial derivatives of $a(\xi)$ are continuous at ξ_0.

Note that the common support of the densities in this example is given by

$$\mathcal{Y} = \{\mathbf{y} \colon h(\mathbf{y}) > 0\}.$$

Now, let Θ be a compact set satisfying that

$$\Theta \subset \Xi^\circ,$$

and observe that Ξ° is convex, since so is Ξ. It will be verified that Assumption 4.3 is satisfied by the family $\{f(\mathbf{y}; \xi) \colon \xi \in \Theta\}$. To achieve this goal, first note that, for $\mathbf{y} \in \mathcal{Y}\mathcal{Y}$,

$$\log f(\mathbf{y}; \xi) = \log[a(\xi)^{-1} h(\mathbf{y}) e^{\xi' T(\mathbf{y})}] = \xi' T(\mathbf{y}) + \log(h(\mathbf{y})) - \log(a(\xi)),$$

and then

$$f(\mathbf{y}; \xi) - f(\mathbf{y}; \xi_1) = (\xi - \xi_1)' T(\mathbf{y}) + \log[a(\xi)] - \log[a(\xi_1)], \quad \xi, \xi_1 \in \Xi, \quad \mathbf{y} \in \mathcal{Y}. \tag{4.5}$$

Now, observe that the Cauchy-Schwarz inequality yields that

$$|(\xi - \xi_1)' T(\mathbf{y})| \leqslant \|\xi - \xi_1\| \, \|T(\mathbf{y})\|,$$

where $\|\mathbf{w}\|$ denotes the Euclidean norm of the vector $\mathbf{w}$, so that

$$\|T(\mathbf{y})\| = \sqrt{T_1(\mathbf{y})^2 + ... + T_k(\mathbf{y})^2} \leqslant |T_1(\mathbf{y})| + ... + |T_k(\mathbf{y})|,$$

and then

$$|(\xi - \xi_1)' T(\mathbf{y})| \leqslant \|\xi - \xi_1\| \, \|T(\mathbf{y}) = \|\xi - \xi_1\| \, [|T_|(\mathbf{y})| + ... + |T_k(\mathbf{y})|]. \tag{4.6}$$

Next, a bound for the difference of logarithms in (4.5) will be established. Let $\tilde{\Theta}$ be the union of all the segments that join points of Θ, that is, $\tilde{\Theta} = \{t\theta + (1-t)\theta_1 \,|\, \theta, \theta_1 \in \Theta, \ t \in [0, 1]\}$. As Θ is a compact set contained in Ξ°, which is an open and convex set, it follows that $\tilde{\Theta}$ is also a compact subset of Ξ°. Thus, the partial derivatives of $a(\xi)$, which are defined and continuous in Ξ°, are bounded when $\xi \in \tilde{\Theta}$, i.e., exists

$\tilde{M} > 0$ such as $|\partial_{\xi_i} a(\xi)| \leqslant \tilde{M}$ for all $i = 1, 2, ..., k$ and $\xi \in \tilde{\Theta}$. Hence, the gradient vector $Da(\xi) = (\partial_{\xi_1} a(\xi), ..., \partial_{\xi_k} a(\xi))'$ satsifies that

$$\|Da(\xi)\| = \sqrt{\partial_{\xi_1} a(\xi)^2 + ... + \partial_{\xi_k} a(\xi)^2} \leqslant \sqrt{\tilde{M}^2 + ... + \tilde{M}^2} = \sqrt{k}\tilde{M} =: M, \quad \xi \in \tilde{\Theta}.$$

Select now $\xi, \xi_1 \in \Theta$ and, observe that the mean value theorem implies that there exists $t \in (0, 1)$ such that

$$a(\xi) - a(\xi_1) = Da(t\xi + (1-t)\xi_1)'(\xi - \xi_1),$$

Since $t\xi + (1-t)\xi_1 \in \tilde{\Theta}$, it follows that

$$\|Da(t\xi + (1-t)\xi_1)\| \leqslant M,$$

and the, the Cauchy-Schwarz inequality leads to

(4.7)

$$|a(\xi) - a(\xi_1)| = |Da(t\xi + (1-t)\xi_1)(\xi - \xi_1)| \leqslant \|Da(t\xi + (1-t)\xi_1)\| \; \|\xi - \xi_1\| \leqslant M\|\|\xi - \xi_1\|,$$

$$\xi, \xi_1 \in \Theta.$$

On the other hand, using that $a(\xi)$ is continuous and positive for ξ in the interior of Ξ, which contains the compact parameter space set Θ, it follows that there exists a constant b such that

$$a(\xi) \geqslant b > 0, \quad \xi \in \Theta.$$

On the other hand, since $d\log(x)/dx = 1/x$, the mean value theorem implies that exists a constant $s \in (0, 1)$ such as

$$\log a(\xi) - \log a(\xi_1) = \frac{1}{sa(\xi) + (1-s)a(\xi_1)}[a(\xi) - a(\xi_1)];$$

when ξ and ξ_1 belong to Θ, the inequalities $a(\xi), a(\xi_1) \geqslant b$ are satisfied, and then the inclusion $s \in (0, 1)$ implies that $sa(\xi) + (1-s)a(\xi_1) \geqslant b$, and this inequality and the previous display together imply that

$$|\log a(\xi) - \log a(\xi_1)| \leqslant \frac{1}{b}|a(\xi) - a(\xi_1)|, \quad \xi, \xi_1 \in \Theta;$$

via (4.7), this relation leads to

$$|\log a(\xi) - \log a(\xi_1)| \leqslant \frac{M}{b}\|\xi - \xi_1\|, \quad \xi, \xi_1 \in \Theta.$$

Combining this inequality with (4.5) and (4.6) it follows that, for every $\mathbf{y} \in \mathcal{Y}$,

$$|\log f(\mathbf{y}; \xi) - f(\mathbf{y}; \xi_1)| \leqslant \left[|T_1(\mathbf{y})| + ... + |T_k(\mathbf{y})| + \frac{M}{b}\right] \|\xi - \xi_1\|, \quad \xi, \xi_1 \in \Theta,$$

so that the first part of Assumption 4.3 holds with $B(\mathbf{Y}) : = |T_1(\mathbf{Y}) + ... + T_k(\mathbf{Y})| + \frac{M}{b}$, whereas, recalling that $\Theta \subset \xi^0$, the inequality (4.4), valid when $\xi_0 \in \xi^0$, implies that

$$E_\theta[B(\mathbf{Y})] = E_\theta[|T_1(\mathbf{Y})|] + ... + E_\theta[|T_k(\mathbf{Y})|] + \frac{M}{b} < \infty, \quad \theta \in \Theta,$$

showing that the second part of Assumption 4.3 is also valid.

The following example, concerned with a non-exponential family of densities, shows that the Assumption 4.3 can be satisfied even when $\log f(\mathbf{y};\theta)$ is not differentiable with respect to θ.

EXAMPLE 4.2. Let $f(y;\xi)$ be the Laplace's with center ξ, i.e.,

$$f(y;\xi) = \frac{1}{2}e^{-|y-\xi|}, \quad y \in \mathbb{R}, \quad \xi \in \mathbb{R}$$

In this case $\log f(y;\xi) = -|y-\xi| - \log 2$, and then

$$\log f(y;\xi) - \log f(y;\xi_1) = |y-\xi| - |y-\xi_1|.$$

Observing that $|\,|y-\xi| - |y-\xi_1|\,| \leqslant |(y-\xi)-(y-\xi_1)| = |\xi_1 - \xi|$, by the tangle inequality, it follows that $|\log f(y;\xi) - \log f(y;\xi_1)| \leqslant |\xi - \xi_1|$, and then Assumption 4.3 is satisfied with $B(y) = 1$ for every $y \in \mathbb{R}$.

5. Main Theorem

In this section, a general result on the consistency of maximum likelihood estimators is stated. First, observe that, under the Assumptions 4.2 and 4.3, the likelihood function $L_n(\theta;\mathbf{Y}_1,...,\mathbf{Y}_n)$ depends continuously on θ, and then, since the parameter space is compact, a maximizer $\hat{\theta}_n = \hat{\theta}_n(\mathbf{Y}_1,...,\mathbf{Y}_n)$ always exists [4, 8].

THEOREM 5.1. *Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3, ...$ be a sequence of independent random vectors with a common distribution, whose density belongs to the family $\{f(\boldsymbol{Y};\theta)\colon \theta \in \Theta\}$, and suppose that Assumptions 4.2 and 4.3 are valid. Denote by θ_0 the true value of the parameter and let $\hat{\theta}_n$ be an estimator of maximum likelihood based on $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_n$. In these circumstances,*

$$P_{\theta_0}\left[\lim_{n\to\infty}\hat{\theta}_n = \theta_0\right] = 1. \tag{5.1}$$

This result establishes that the true parameter θ_0 is estimated consistently by the sequence of maximum likelihood estimators; this is the classical conclusion, but it must be observed that Assumptions 4.2 and 4.3 are weaker than the usual regular conditions used, for instance, in [2, 6]. The proof of the Theorem 5.1 is somewhat technical and is presented in the following section. The argument relies on the following result, which is an extension of Theorem 3.1.

THEOREM 5.2. *Suppose that Assumptions 4.2 and 4.3 hold, and let $\theta^* \in \Theta \setminus \{\theta_0\}$ be arbitrary. In this case, there exist a real number $\varepsilon^* = \varepsilon(\theta)$ and s an event Ω^* satisfying*

$$P_{\theta_0}[\Omega^*] = 1 \text{ and } \varepsilon^* > 0,$$

as well as the following property:

For each $\omega \in \Omega^*$, there exists an integer $N(\omega;\theta^*)$ such that, for all $\theta \in B(\theta^*,\varepsilon^*)$,

$$f_n(\boldsymbol{Y}_1(\omega),...,\boldsymbol{Y}_n(\omega);\theta) < f_n(\boldsymbol{Y}_1(\omega),...,\boldsymbol{Y}_n(\omega);\theta_0), \quad n > N(\omega;\theta^*);$$

see (3.1).

This theorem implies that, if θ^* is not equal to the true parameter value θ_0, then the ball $B(\theta^*, \varepsilon^*)$ does not contain any maximizer of the likelihood function whenever the sample size n is large enough, i.e., $n > N(\omega, \theta^*)$.

PROOF. Note that

$$E_{\theta_0}\left[\log\left(\frac{f(\mathbf{Y};\theta^*)}{f(\mathbf{Y};\theta_0}\right)\right] = \nu(\theta^*) < 0,$$

byTheorem 3.1 (i), and then there exists a positive number $\varepsilon^* = \varepsilon^*(\theta^*)$ such that

$$\nu(\theta^*) + \varepsilon^* E_{\theta_0}[B(Y)] < 0, \tag{5.2}$$

where $B(\cdot)$ is the function in Assumption 4.3. Therefore,

$$E_{\theta_0}\left[\log\left(\frac{f(\mathbf{Y};\theta^*)}{f(\mathbf{Y};\theta_0)}\right) + \varepsilon^* B(\mathbf{Y})\right] = \nu(\theta^*) + \varepsilon^* E_{\theta_0}[B(Y)] < 0.$$

and combining this fact with the law of large numbers in Theorem 2.1, it follows that there exists an event Ω^* satisfying that

$$P[\Omega^*] = 1 \text{ and, for all } \omega \in \Omega^*, \text{ and}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \left[\log\left(\frac{f(\mathbf{Y}_i(\omega);\theta^*)}{f(\mathbf{Y}_i(\omega);\theta_0)}\right) + \varepsilon^* B(\mathbf{Y}_i(\omega))\right] = \nu(\theta^*) + \varepsilon^* E_{\theta_0}[B(Y)].$$

Since the right-hand side of the last equality is negative, the definition of limit yields that, for each $\omega \in \Omega^*$, there exists a positive integer $N^* = N^*(\omega)$ such that

$$\sum_{i=1}^n \left[\log\left(\frac{f(\mathbf{Y}_i(\omega);\theta^*)}{f(\mathbf{Y}_i(\omega);\theta_0)}\right) + \varepsilon^* B(\mathbf{Y}_i(\omega))\right] < 0, \quad n > N^*(\omega), \quad \omega \in \Omega^*. \tag{5.3}$$

Now, let θ in $B(\theta^*, \varepsilon^*)$ be arbitrary. In this case

$$\|\theta - \theta^*\| < \varepsilon^*, \tag{5.4}$$

and part (i) of Assumption 4.3 yields that

$$|\log f(\mathbf{y};\theta) - \log f(\mathbf{y};\theta^*)| \leqslant \|\theta - \theta^*\| B(\mathbf{y}) \leqslant \varepsilon^* B(\mathbf{y}), \quad \mathbf{y} \in \mathcal{Y},$$

and then

$$\log\left(\frac{f(\mathbf{y};\theta)}{f(\mathbf{y};\theta_0)}\right) - \log\left(\frac{f(\mathbf{y};\theta^*)}{f(\mathbf{y};\theta_0)}\right) = [\log f(\mathbf{y};\theta) - \log f(\mathbf{y};\theta_0)] - [\log f(\mathbf{y};\theta^*) - \log f(\mathbf{y};\theta_0)]$$
$$= \log f(\mathbf{y};\theta) - \log f(\mathbf{y};\theta^*) \leqslant \varepsilon^* B(\mathbf{y}),$$

so that

$$\log\left(\frac{f(\mathbf{y};\theta)}{f(\mathbf{y};\theta_0)}\right) \leqslant \log\left(\frac{f(\mathbf{y};\theta^*)}{f(\mathbf{y};\theta_0)}\right) + \varepsilon^* B(\mathbf{y}), \quad \mathbf{y} \in \mathcal{Y}.$$

This relation immediately leads to

$$\sum_{i=1}^n \log\left(\frac{f(\mathbf{y}_i;\theta)}{f(\mathbf{y}_i;\theta_0)}\right) \leqslant \sum_{i=1}^n \left[\log\left(\frac{f(\mathbf{y}_i;\theta^*)}{f(\mathbf{y}_i;\theta_0)}\right) + \varepsilon^* B(\mathbf{y}_i)\right].$$

Combining this inequality with (5.3), and recalling that $\theta \in \Theta$ is an arbitrary parameter that satisfies (5.4), it follows that

$$\sum_{i=1}^{n} \log\left(\frac{f(\mathbf{Y}_i(\omega);\theta)}{f(\mathbf{Y}_i(\omega);\theta_0)}\right) < 0, \quad \omega \in \Omega^*, \quad n > N^*(\omega), \quad \theta \in B(\theta^*, \varepsilon^*). \tag{5.5}$$

Via (3.1) , this yields that, for every $\omega \in \Omega^*$ and $\theta \in B(\theta^*, \varepsilon^*)$,

$$\log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta) < \log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta_0), \quad n > N^*(\omega),$$

and the conclusion follows, since $P[\Omega^*] = 1$, □

6. Proof of the Main Theorem

DemMainIn this section Theorem 5.1 will be established. The argument relies on the following consequence of Theorem 5.2.

Theorem 6.1. *Let $\varepsilon > 0$ be a fixed number. Under Assumptions 4.2 and 4.3, there exists an event Ω_ε with the following properties:*

$$P[\Omega_\varepsilon] = 1,$$

and For each $\omega \in \Omega_\varepsilon$ and $\theta \in \Theta \cap B(\theta_0, \varepsilon)^c$ there exists an integer $N_\varepsilon(\omega)$ such as $\log f_n(\boldsymbol{Y}_1(\omega), ..., \boldsymbol{Y}_n(\omega); \theta) < \log f_n(\boldsymbol{Y}_1(\omega), ..., \boldsymbol{Y}_n(\omega); \theta_0), \quad n > N_\varepsilon(\omega).$

This theorem implies that , with probability 1, $\log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \cdot)$ is not maximized at any point of $\Theta \cap B(\theta_0, \varepsilon)^c$ when n is large enough, so that a maximizer must belong to $B(\theta_0, \varepsilon)$.

Proof. To begin with, note that the ball $B(\theta_0, \varepsilon)$ is open, and then Assumption 4.2 yields that $\Theta \cap B(\theta_0, \varepsilon)^c$ is a compact set [4, 8]. Now, let $\theta^* \in \Theta \cap B(\theta_0, \varepsilon)^c$ be arbitrary. In this case $\|\theta^* - \theta_0\| > \varepsilon$, so that $\theta^* \neq \theta_0$. By Theorem 5.2, there exist an event $\Omega(\theta^*)$ and a positive number $\varepsilon(\theta^*)$, such that

$$P[\Omega(\theta^*)] = 1, \tag{6.1}$$

and, for each $\omega \in \Omega(\theta^*)$, the following property is satisfied by a positive integer $N(\omega, \theta^*)$: if $\|\theta - \theta^*\| < \varepsilon(\theta^*)$, then $n > N(\omega, \theta^*)$ implies that

$$\log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta) < \log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta_0). \tag{6.2}$$

Next, observe that

$$\Theta \cap B(\theta_0, \varepsilon)^c \subset \bigcup_{\theta^* \in \Theta \cap B(\theta_0, \varepsilon)} B(\theta^*, \varepsilon(\theta^*));$$

since $\Theta \cap B(\theta_0, \varepsilon)^c$ is a compact set, by the Heine-Borel property in Theorem 4.1, there exists a finite set $\{\theta_1^*, \theta_2^*, ..., \theta_r^*\}$ contained in $\Theta \cap B(\theta_0, \varepsilon)$ such that

$$\Theta \cap B(\theta_0, \varepsilon)^c \subset \bigcup_{i=1}^{r} B(\theta_i^*, \varepsilon(\theta_i^*)). \tag{6.3}$$

Now, set

$$\Omega_\varepsilon \colon = \bigcap_{i=1}^{r} \Omega(\theta_i^*), \tag{6.4}$$

and observe that $P[\Omega_\varepsilon] = 1$, by (6.1). Next, define

$$N_\varepsilon(\omega) \colon = \max_{i=1,2,\ldots,r} N(\omega, \theta_i^*), \quad \omega \in \Omega_\varepsilon. \tag{6.5}$$

It will be shown that the conclusion of the theorem is satisfied by Ω_ε and $N_\varepsilon(\omega)$ specified above. Let $\theta \in \Theta \cap B(\theta_0)^c$ and $\omega \in \Omega_\varepsilon$ be arbitrary, It follows that $\theta \in B(\theta_i^*, \varepsilon(\theta_i^*))$ for some i between 1 and r, by (6.3), whereas $\omega \in \Omega(\theta_i^*)$, by (6.4), Now, observe that $n > N_\varepsilon(\omega)$ implies that $n > N(\omega, \theta_i^*)$, by (6.5), and then (6.2) implies that

$$\log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta) < \log f_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega); \theta_0).$$

Thus, it has been shown that, if $\theta \in \Theta \cap B(\theta_0)^c$, $\omega \in \Omega_\varepsilon$ and $n > N_\varepsilon(\omega)$, then the above inequality occurs, this completes the argument since, as already noted, $P[\Omega_\varepsilon] = 1$. □

Proof of Theorem 5.1 For each integer $m = 1, 2, ...$, let $\Omega_{1/m}$ the event in the Theorem 6.1 corresponding to $\varepsilon = 1/m$, and define

$$\Omega^* = \bigcap_{m=1}^{\infty} \Omega_{1/m}. \tag{6.6}$$

In this case, $\Omega^{*c} = \bigcup_{m=1}^{\infty} \Omega_{1/m}^c$ so that $P[\Omega^{*c}] \leqslant \sum_{m=1}^{\infty} P\left[\Omega_{1/m}^c\right] = 0$, and then

$$P[\Omega^*] = 1. \tag{6.7}$$

It will be shown that, for each $\omega \in \Omega^*$, if $\hat\theta_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) \in \Theta$ is a maximizer of the likelihood function, then

$$\lim_{n\to\infty} \hat\theta_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) = \theta_0.$$

To achieve this goal, let $\omega \in \Omega^*$ and $\delta > 0$ be arbitrary, select an integer $m > 0$ such as $1/m < \delta$, and note that $\omega \in \Omega_{1/m}$ by (6.6). Now, let $N_{1/m}(\omega)$ be the integer in Theorem 6.1, so that if $n > N_{1/m}(\omega)$, then for each $\theta \in \Theta \cap B(\theta_0, 1/m)^c$, the inequality $\log f(\mathbf{Y}_1(\omega), ..., Y_n(\omega); \theta) < \log f(\mathbf{Y}_1(\omega), ..., Y_n(\omega); \theta_0)$ holds. This implies that, for $n > N_{1/m}(\omega)$, the function

$$\theta \mapsto \log f(\mathbf{Y}_1(\omega), ..., Y_n(\omega); \theta)$$

is not maximized at any point of $\theta \in \Theta \cap B(\theta_0, 1/m)^c$, so that

$$\hat\theta_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) \in \Theta \cap B(\theta_0, 1/m) \subset B(\theta_0, 1/m);$$

it follows that

$$\|\hat\theta_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) - \theta_0\| < \frac{1}{m} < \delta, \quad n > N_{1/m}(\omega),$$

a relation that, by the definition of limit, yields that

$$\lim_{n\to\infty} \hat{\theta}_n(\mathbf{Y}_1(\omega), ..., \mathbf{Y}_n(\omega)) = \theta_0.$$

Since $\omega \in \Omega^*$ is arbitrary and $P[\Omega^*] = 1$, it follows that $\{\hat{\theta}_n\}$ is a consistent sequence of estimators of θ_0, by Definition 2.1.

References

[1] R. Ash (1987). *Real Analysis and Probability*, *Academic Press*, New York.
[2] G. Casella, R. Berger (2001). *Statistical Inference*, *Duxbury Press*, New York.
[3] E. Dudewicz, N. Mishra (1988). *Modern Mathematical Statistics*, *John Wiley & Sons*, New York.
[4] J. Dugundji (1970). Topology, *Allyn & Bacon*, Boston.
[5] W. Fulks (1983). *Cálculo Avanzado*, *Limusa*, México D.F.
[6] E. L. Lehmann (2001). *Testing Statistical Hypotheses*, *Jhon Willey & Sons*, New York.
[7] A. M. Mood, F. A. Graybill, D. C. Boes (1987). *Introduction to the Theory of Statistics*, *McGraw-Hill*, New York
[8] J. R. Munkres (1974). *Topology: A First Course*, *Prentice-Hall*, New York.
[9] W. Rudin (1968). *Real and Complex Analysis*, *McGraw-Hill*, New York
[10] J. Saucedo Zul (2004). Consistencia de Estimadores de Verosimilitud Máxima Bajo Supuestos de Continuidad, Tesis de Maestría en Estadística Experimental, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo COAH, México.
[11] R. J. Serfling (1988). *Approximation Theorems of Mathematical Statistics*, *John Willey*, New York.
[12] T. A. Severini (2001). *Likelihood Methods in Statistics*, *Oxford University Press*, London.

Received 25 02 2015 revised 21 03 2016

Facultad de Ciencias Físico Matemáticas,
Universidad Autónoma de Coahuila,
Unidad Camporredondo, Edificio A. Saltillo COAH,
México.
E-mail address: juliosaucedozul@uadec.edu.mx
E-mail address: haydee.garza@uadec.edu.mx