A BASIC RESULT ON THE CONSISTENCY OF MAXIMUM LIKELIHOOD ESTIMATORS

Rolando Cavazos-Cadena and Graciela González-Farías

Comunicación Técnica No I-07-08/07-05-2007
(PE/CIMAT)

A Basic Result on the Consistency of Maximum Likelihood Estimators*

ROLANDO CAVAZOS–CADENA, Departamento de Estadística y Cálculo, Universidad Autónoma Agraria Antonio Narro, Buenavista, Saltillo COAH, 25315, **MÉXICO**
e-mail: rcavazos@uaaan.mx

GRACIELA GONZÁLEZ–FARÍAS, Centro de Investigación en Matemáticas A. C., Apartado Postal 402, Guanajuato, GTO, 36000, **MÉXICO**
e-mail: farias@cimat.mx

August 25, 2006

Abstract. A sequence $\{X_i\}$ of independent and identically distributed random objects is considered. The common distribution of the X_i's is absolutely continuous with respect to a given measure, and the corresponding density is not completely specified but depends on an unknown parameter. Under mild topological and continuity requirements, a necessary and sufficient criterion for the consistency of a sequence of maximum likelihood estimators is obtained. When this characterization is applied to the case in which the parameter belongs to a finite dimensional Euclidean space, the conclusion is that *the sequence is consistent if and only if it is bounded* with probability 1.

Key Words: Hewitt-Savage zero-one law, Strong law of large numbers, Consistency criterion, Boundedness almost everywhere.

AMS Subject Classifications: 62H12, 62H15.

Running Head: Consistency criterion for ML estimators

1. Introduction

This work concerns the maximum likelihood estimation method, which is widely used in statistics and, under regularity conditions frequently satisfied in useful models, produces asymptotically efficient estimators (Severini, 2000, Shao, 1999, Lehmann and Casella 1998, Azzalini, 1996). The starting point is a sequence $\{X_i\}$ of independent and identically distributed (*iid*) random objects whose common distribution is absolutely continuous with respect to a given measure, and the corresponding density is not completely specified but depends on an unknown parameter θ. In this context, the most basic and desirable asymptotic property of an estimation scheme for θ is its *consistency* which, roughly, requires the convergence of the generated estimators to the true parameter value as the sample size increases; see Definition 2.1 below. In this direction, it was shown by Bahadur (1958) that the maximum likelihood procedure is not necessarily consistent (see also Lehmann and Casella, 1998, p. 445-447), and this fact provides the motivation for the main problem considered in this note: *To determine a necessary and sufficient criterion for the consistency of the maximum likelihood estimation method.*

The above problem is being analyzed, essentially, under three types of assumptions presented formally in the following sections: First, it is supposed that the parameter space Θ is *a locally compact* metric space, a weak requirement that is satisfied, for instance, when Θ is an Euclidean space $\mathbb{R}^k$. Next, it is assumed that the unknown density of the X_i's depends continuously, in a certain sense, on the parameter $\theta \in \Theta$; as usual, the discrepancy between two densities is measured in the logarithmic scale, but it is not supposed that densities corresponding to different parameters have the same support. Finally, it is assumed that, with probability 1, the maximum likelihood estimator is well-defined when the sample size is large enough. In this framework, *the main result* of the paper, stated as Theorem 4.1 in Section 4, can be roughly described as follows:

* This work was supported in part by the PSF Organization under Grant No. 2006-1, and by CONACYT–México under Grant 45974-F.

• A sequence $\{\hat{\theta}_n\}$ of maximum likelihood estimators is consistent if, and only if, there exists a compact set that, with probability 1, contains $\hat{\theta}_n$ for all n sufficiently large.

When this result is applied to the case $\Theta = \mathbb{R}^k$, the following simple characterization is obtained in Corollary 4.1:

• A sequence of (symmetric) maximum likelihood estimators is consistent if and only if the sequence is *bounded* almost surely.

This latter result is a remarkable property of the maximum likelihood estimation method, since an arbitrary bounded sequence in an Euclidean space is not necessarily convergent. The arguments used in the this paper rely heavily on two basic facts, namely, (i) the strong law of large numbers for random variables whose expectation is not necessarily finite (Ash 1972, p. 277, or Billingsley 1995, p. 284), and (ii) the Hewitt-Savage zero-one law for symmetric events (Ash 1972, p. 279, Billingsley 1995, p. 496).

The organization of the paper is as follows: First, in Section 2 the statistical model is introduced, and the basic structural assumptions are formally stated. Next, the maximum likelihood estimation procedure is briefly discussed in Section 3, and a zero-one law is established for the existence of maximum likelihood estimators $\hat{\theta}_n$ for large samples. Then, in Section 4 the criteria for the consistency of $\{\hat{\theta}_n\}$ are established; the argument in this part uses a technical tool stated as Theorem 4.2, and the exposition concludes in Section 5 with a proof of this result.

Notation. For a measurable space $(S, \mathcal{G})$ and $n = 1, 2, 3, \ldots$, S^n denotes the n-fold cartesian product of S with itself, whereas $\mathcal{G}^n$ is the σ-field generated by the sets $B_1 \times B_2 \times \cdots \times B_n$ with $B_i \in \mathcal{G}$ for $i = 1, 2, \ldots, n$. Similarly, S^∞ consists of all sequences

$$\mathbf{x} = (x_1, x_2, x_3 \ldots)$$

with $x_i \in S$ for all i, and $\mathcal{G}^\infty$ stands for the σ-field generated by the cylinders $B \times \mathcal{S}^\infty$, where $B \in \mathcal{G}^n$ for some n. On the other hand, for a positive integer m, $\mathcal{P}_m$ denotes the class of all permutations of $\{1, 2, \ldots, m\}$, and for each $\mathbf{x} \in \mathcal{S}^\infty$,

$$\mathbf{x}_\tau = (x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(m)}, x_{m+1}, x_{m+2}, \ldots), \quad \tau \in \mathcal{P}_m, \quad m = 1, 2, 3, \ldots. \tag{1.1}$$

2. Statistical Model

Let $X_1, X_2, X_3, \ldots$ be *iid* random objects defined on the probability space $(\Omega, \mathcal{F}, P)$ and taking values on the measurable space $(S, \mathcal{G})$. The common distribution of the X_i's is the measure P_X on $\mathcal{G}$ defined by

$$P_X[B] = P[X_i \in B], \quad B \in \mathcal{G}. \tag{2.1}$$

Assumption 2.1. There exists a (σ-finite) measure ν on $(S, \mathcal{G})$ such that P_X is absolutely continuous with respect to ν, i.e., for some (density) function $f_X : S \to [0, \infty)$,

$$P_X[B] = \int_B f_X(x)\, \nu(dx), \quad B \in \mathcal{G}. \tag{2.2}$$

Measure P_X—or, equivalently, density f_X—contains all probabilistic information about the sequence $\{X_i\}$. However, in practice f_X is not completely known, and the *statistical problem* consists in using the observed values of $X_1, X_2, X_3, \ldots$—the data—to approximate the unknown density f_X. Hereafter, it is supposed that *a priori* knowledge about the physical process generating the observations allows to postulate that f_X belongs to a certain restricted class of densities $\mathbb{F}$; the assertion

$$f_X \in \mathbb{F} \tag{2.3}$$

is a *statistical model*, and in this work it is assumed that the members of $\mathbb{F}$ can be indexed by the elements of a metric space Θ, i.e.,

$$\mathbb{F} = \{f(x;\theta)\colon \theta \in \Theta\} \tag{2.4}$$

where, for each $\theta \in \Theta$, the $\mathcal{G}$-measurable function $f(\cdot;\theta)\colon S \to [0,\infty)$ is a density function with respect to ν; in this case (2.3) is a *parametric model* and Θ is referred to as *the parameter space*. *The parametrization* $\theta \mapsto f(\cdot;\theta)$ from Θ onto $\mathbb{F}$ is supposed to be injective.

Assumption 2.2. For $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$,

$$\nu\,[x\colon f(x;\theta) \neq f(x;\theta')] > 0.$$

Under the assumption that $f(\cdot;\theta) = f_X(\cdot)$, the common distribution $\tilde{P}_\theta$ of the X_i's is given by

$$\tilde{P}_\theta[B] = \int_B f(x;\theta)\,\nu(dx), \quad B \in \mathcal{G} \tag{2.5}$$

(see (2.1) and (2.2)); the distribution of the whole process $\mathbf{X} = (X_1, X_2, X_3, \ldots)$ is denoted by P_θ and $E_\theta[\cdot]$ stands for the corresponding expectation operator. Notice that

$$P_\theta = \tilde{P}_\theta \times \tilde{P}_\theta \times \cdots, \tag{2.6}$$

is the countable product of measure $\tilde{P}_\theta$ with itself. The (unknown) parameter $\theta^* \in \Theta$ for which $f_X(\cdot) = f(\cdot;\theta^*)$ is the true parameter value, and the problem of looking for f_X in family $\mathbb{F}$, is the same as searching for θ^* within Θ. In general, based on a finite number of observations $X_1, X_2, \ldots, X_n$, it is not possible to determine θ^* exactly, and the given data must be used to construct an estimator $\tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ whose values are used as approximations of θ^*.

Definition 2.1. *The sequence* $\{\tilde{\theta}_n \equiv \tilde{\theta}_n(X_1, X_2, \ldots, X_n)\}$ *of estimators of* θ*—or the method used to build it—is consistent if*

$$P_\theta\left[\lim_{n\to\infty} \tilde{\theta}_n = \theta\right] = 1, \quad \theta \in \Theta.$$

In the following sections the maximum likelihood method of estimation is studied, and necessary and sufficient conditions are given for its consistency. The discussion assumes the that the following mild topological and continuity conditions hold.

Assumption 2.3. The parameter space Θ is *a locally compact* metric space.

Next, for each $\theta \in \Theta$ define the support S_θ of density $f(\cdot;\theta)$ by

$$S_\theta = \{x \in S\colon f(x;\theta) > 0\} \tag{2.7}$$

whereas, for $\varepsilon > 0$ and $\theta_1 \in S$, the ε-discrepancy function of family $\{f(\cdot;\theta)\}$ at θ_1 is given by

$$D_{\varepsilon,\theta_1}(x) = \sup_{\theta\colon d(\theta,\theta_1)<\varepsilon} \log\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right) I[x \in S_\theta \cap S_{\theta_1}], \quad x \in S, \tag{2.8}$$

where $d(\cdot,\cdot)$ is the metric in Θ.

Assumption 2.4. (a) For each $\varepsilon > 0$ and $\theta, \theta_1 \in \Theta$

(i) $D^+_{\varepsilon,\theta_1}(x) = \max\{0, D_{\varepsilon,\theta_1}(x)\}$ is $\mathcal{G}$-measurable, and

(ii) $E_\theta\left[D^+_{\varepsilon,\theta_1}(X_1)\right] \to 0$ as $\varepsilon \searrow 0$.

(b) Given θ_0 and $\theta_1 \in \Theta$, if $\nu[S_{\theta_0} \cap S_{\theta_1}^c] > 0$ then there exists $B \in \mathcal{G}$ and $\varepsilon > 0$ such that

(i) $\nu[B] > 0$, and

(ii) $B \subset S_{\theta_0} \cap S_\theta^c$ when $d(\theta_1, \theta) < \varepsilon$.

Remark 2.1. (a) Assume that there exists a set $\{\rho_1, \rho_2, \ldots\}$ dense in Θ with the following property, which is valid in all useful models.

A: For each $x \in S$ and $\theta \in \Theta$, a subsequence $\{\rho_{n_k}\}$ can be found satisfying $\rho_{n_k} \to \theta$ and $f(x, \rho_{n_k}) \to f(x, \theta)$ as $k \to \infty$.

In this case the supremum in (2.8) can be taken over the ρ_k's satisfying $d(\rho_k, \theta_1) < \varepsilon$—a countable set— and then function $D_{\varepsilon,\theta_1}(\cdot)$ is $\mathcal{G}$-measurable.

(b) In certain sense, part (b) in Assumption 2.4 guarantees that the supports S_θ do no experiment a 'sudden' growth as $\theta \to \theta_1$. When $S = \mathbb{R}^k$ endowed with the Euclidean norm, and $\nu(\cdot)$ is the corresponding Lebesgue measure, suppose that

$$S_\theta = \{(x_1, \ldots, x_k) \colon a_i(\theta) \le x_1 \le b_i(\theta), \quad i = 1, 2, \ldots, k\}$$

for certain mappings $a_i(\cdot), b_i(\cdot) \colon \Theta \to \mathbb{R}$. In this case Assumption 2.4(b) holds if, for every $i = 1, 2, \ldots, k$, $a_i(\cdot)$ is lower semi-continuous and $b_i(\cdot)$ is upper semi-continuous.

3. Maximum Likelihood Estimation

Given a positive integer n and points $x_1, x_2, \ldots, x_n \in S$, the likelihood function associated to the event

$$[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n], \tag{3.1}$$

denoted by $L_n(\cdot; x_1, x_2, \ldots, x_n) \colon \Theta \to [0, \infty)$, is defined by

$$L_n(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta), \quad \theta \in \Theta. \tag{3.2}$$

Notice that the likelihoods satisfy

$$L_n(\cdot; x_1, x_2, \ldots, x_n) = L_n(\cdot; x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(n)}), \quad \tau \in \mathcal{P}_n, \tag{3.3}$$

so that, defining the set M_n by

$$M_n = [(x_1, x_2, \ldots, x_n) \in S^n \colon L_n(\cdot; x_1, x_2, \ldots, x_n) \text{ has a maximizer}], \tag{3.4}$$

it follows that M_n is symmetric, *i.e.*,

$$(x_1, x_2, \ldots, x_n) \in M_n \iff (x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(n)}) \in M_n, \quad \tau \in \mathcal{P}_n. \tag{3.5}$$

After observing (3.1), the maximum likelihood method consists in estimating θ by a maximizer $\hat\theta(x_1, x_2, \ldots, x_n)$ of $L_n(\cdot; x_1, x_2, \ldots, x_n)$ whenever such a point exists, so that

$$\begin{aligned} &L_n(\hat\theta(x_1, x_2, \ldots, x_n); x_1, x_2, \ldots, x_n) \\ &\qquad \ge L_n(\theta; x_1, x_2, \ldots, x_n), \quad \theta \in \Theta, \quad (x_1, x_2, \ldots, x_n) \in M_n \end{aligned} \tag{3.6}$$

whereas $\hat\theta_n(x_1, x_2, \ldots, x_n)$ is defined 'arbitrily' if $L_n(\cdot; x_1, x_2, \ldots, x_n)$ does not achieve its maximum, say

$$\hat\theta(x_1, x_2, \ldots, x_n) = \theta_*, \quad (x_1, \ldots, x_n) \notin M_n \tag{3.7}$$

where θ_* is a fixed member of Θ. On the other hand, $\hat\theta_n(\cdot)$ must be a measurable function of $(x_1, x_2, \ldots, x_n)$ to ensure that $\hat\theta_n(X_1, X_2, \ldots, X_n)$ is a valid estimator, and this requires additional conditions on the mapping $(x, \theta) \mapsto f(x; \theta)$. Instead of digging into measurability topics, here it is simply *assumed* that M_n belongs to $\mathcal{G}^n$, and that $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ is a $\mathcal{G}^n$-measurable function. In these circumstances $\hat{\theta}_n \equiv \hat{\theta}(X_1, X_2, \ldots, X_n)$ is a maximum likelihood estimator of θ based on $X_1, X_2, \ldots, X_n$. Next, let $\tilde{\theta}_n(X_1, X_2, \ldots, X_n)$ be an arbitrary maximum likelihood estimator and define

$$\hat{\theta}_n(X_1, X_2, \ldots, X_n) = \tilde{\theta}_n(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$$

where the $X_{(i)}$'s are the order statistics of sample $X_1, X_2, \ldots, X_n$. In this case, form (3.3)–(3.7) it follows that $\hat{\theta}_n$ is also a maximum likelihood estimator which is symmetric, *i.e.*,

$$\hat{\theta}_n(X_1, X_2, \ldots, X_n) = \hat{\theta}_n(X_{\tau(1)}, X_{\tau(2)}, \ldots, X_{\tau(n)}), \quad \tau \in \mathcal{P}_n; \tag{3.8}$$

all 'good' estimators have this property. Observe now that, since $\hat{\theta}_n(\cdot)$ is defined arbitrarily on M_n^c, the essential characteristic of a maximum likelihood estimate is inequality (3.6), making it interesting to investigate if the inclusion $(x_1, x_2, \ldots, x_n) \in M_n$ occurs, with probability one, for n large enough; the main conclusion in this direction is the zero-one result in Lemma 2.1 below, whose statement uses the following notation: For $\mathbf{x} = (x_1, x_2, x_3 \ldots) \in S^\infty$ and a positive integer n, set

$$\mathbf{x}^n = (x_1, x_2, \ldots, x_n), \tag{3.9}$$

and let M_n' the set of trajectories $\mathbf{x} \in S^\infty$ such that, after observing $X_1, X_2, \ldots, X_n$, the corresponding likelihood L_n achieves its maximum. More precisely,

$$M_n' = [\mathbf{x} \in S^\infty : \mathbf{x}^n \in M_n] = M_n \times S^\infty. \tag{3.10}$$

With this notation, $\bigcap_{m=k}^\infty M_m'$ consists of all trajectories $\mathbf{x} \in S^\infty$ along which the likelihood corresponding to the first m observations attains its maximum for all $m \geq k$, and then

$$M^* = \bigcup_{k=1}^\infty \bigcap_{m=k}^\infty M_m' \tag{3.11}$$

is the class of all trajectories $\mathbf{x}$ for which the likelihoods $L_m(\cdot; \mathbf{x}^m)$ have maximizers when m is large enough; in the terminology of Billingsley (1995), or Shao (1999), M^* is the limit inferior of the events M_m'.

Lemma 3.1. *For each* $\theta \in \Theta$,

$$P_\theta[M^*] = 0 \quad \textit{or} \quad P_\theta[M^*] = 1.$$

Proof. Let $\theta \in \Theta$ be arbitrary but fixed, and assume that $\theta \in \Theta$ is the true parameter value, so that P_θ is the distribution of $\mathbf{X} = (X_1, X_2, \ldots)$, and then

$$P_\theta[M^*] = P[\mathbf{X} \in M^*]. \tag{3.12}$$

Let the positive integer m and $\tau \in \mathcal{P}_m$ be arbitrary, and observe that, with the notation in (1.1), for each $t \geq m$, (3.4), (3.5) and (3.10) together yield the following result: $\mathbf{x} \in M_t' \iff \mathbf{x}_\tau \in M_t'$. Therefore,

$$\mathbf{x} \in \bigcup_{k=m}^\infty \bigcap_{t=k}^\infty M_t' \iff \mathbf{x}_\tau \in \bigcup_{k=m}^\infty \bigcap_{t=k}^\infty M_t'.$$

On the other hand, observing that $\bigcap_{t=k}^\infty M_t' \subset \bigcap_{t=k_1}^\infty M_t'$ for $k \leq k_1$, it follows from (3.11) that $M^* = \bigcup_{k=m}^\infty \bigcap_{t=k}^\infty M_t'$; consequently, since the positive integer m and $\tau \in \mathcal{P}_m$ are arbitrary, the above display yields that

$$\mathbf{x} \in M^* \iff \mathbf{x}_\tau \in M^*, \quad \tau \in \mathcal{P}_m, \quad m = 1, 2, 3, \ldots, \tag{3.13}$$

i.e., M^* is symmetric. Recalling that the X_i's are *iid*, the Hewitt-Savage zero-one law for symmetric events yields that $P[\mathbf{X} \in M^*] = 1$ or $P[\mathbf{X} \in M] = 0$ (Ash 1972, Billingsley 1995), and then $P_\theta[M^*] = 0$ or $P_\theta[M^*] = 1$, by (3.12). □

Since $\hat{\theta}_n(X_1, \ldots, X_n)$ is arbitrary when $(X_1, X_2, \ldots, X_n) \notin M_n$, it is clear that a good asymptotic behavior of $\{\hat{\theta}_n\}$ can be expected only when $P_\theta[M^*] = 1$, This requirement holds for all the models usually considered in applications, and in general, can be guaranteed by imposing conditions like continuity of the mapping $\theta \mapsto f(x;\theta)$ for each $x \in S$ and (i) compactness of Θ, or (ii) $\sup_{\theta\in\Theta\setminus K_i} f(x;\theta) \to 0$ as $i \to \infty$, where $\Theta = \bigcup_{i=1}^\infty K_i$ and each set K_i is compact. Instead of giving explicit conditions to ensure that $P_\theta[M^*] = 1$, it is simply supposed that this equality holds for every $\theta \in \Theta$.

Assumption 3.1. $M_n \in \mathcal{G}^n$ for $n = 1, 2, 3, \ldots$, and $P_\theta[M^*] = 1$ for each $\theta \in \Theta$.

4. Necessary and Sufficient Conditions for Consistency

This section analyzes the consistency of a sequence of maximum likelihood estimators. As it is shown by the following simple example, under the assumptions in this work consistency of the maximum likelihood method does not necessarily hold.

Example 4.1. Let the parameter space be $\Theta = \{0, 1, 2, 3, \ldots\}$ endowed with the discrete metric, and let $\varphi(x)$ be a density on the real line such that $\varphi(x) > 0$ for every $x \in \mathbb{R} = S$. Set $f(x; 0) = \varphi(x)$ and for $k = 1, 2, 3, \ldots$

$$f(x;k) = \frac{\varphi(x)}{c_k} I[x \in [-k, k]], \quad \text{where } \ c_k = \int_{-k}^{k} \varphi(x)\, dx.$$

In this context, the assumptions in the previous sections hold and, for each $x_1, x_2, \ldots, x_n \in \mathbb{R}$, the likelihood function $L_n(\cdot; x_1, x_2, \ldots, x_n)$ achieves its maximum at the single point

$$\hat{\theta}_n(x_1, \ldots, x_n) = \min\{k \in \Theta \colon |x_i| \leq k,\ i = 1, 2, \ldots, n\},$$

and it is not difficult to see that $\hat{\theta}_n \to \infty$ with probability 1 with respect to P_0. Therefore, the condition $\hat{\theta}_n \to 0$ P_0-almost surely fails, and then $\{\hat{\theta}_n\}$ is not a consistent sequence; see Definition 2.1. (In Lehmann and Casella (1998, p.445) a more sofisticated example is presented for which the convergence $\hat{\theta}_n \to \theta$ P_θ-almost surely fails for *every* $\theta \in \Theta$.) ◁

The following theorem provides necessary and sufficient conditions for the consistency of a sequence of maximum likelihood estimators.

Theorem 4.1. *Suppose that Assumptions 2.1–2.4 as well as Assumption 3.1 hold, and let* $\{\hat{\theta}_n \equiv \hat{\theta}_n(X_1, X_2, \ldots, X_n)\}$ *be a sequence of maximum likelihood estimators. In this case, the following assertions (a) and (b) are equivalent:*

(a) $\{\hat{\theta}_n\}$ *is a consistent sequence of estimators of* θ*; see Definition 2.1.*

(b) For each $\theta \in \Theta$*, there exists a compact set* $C_\theta \subset \Theta$ *such that, with probability 1 with respect to* P_θ*,* $\hat{\theta}_n$ *belongs to* C_θ *for n large enough. More precisely,*

$$P_\theta\left[\bigcup_{k=1}^{\infty}\bigcap_{r=k}^{\infty}\left[\hat{\theta}_r \in C_\theta\right]\right] = 1, \quad \theta \in \Theta. \tag{4.1}$$

As it is shown in the following corollary, Theorem 4.1 renders a simple characterization of consistency for models with $\Theta = \mathbb{R}^k$.

Corollary 4.1. *Suppose that $\Theta = \mathbb{R}^k$ endowed with the Euclidean norm $\|\cdot\|$, and that Assumptions 2.1, 2.2, 2.4 and 3.1 hold. Let $\{\hat\theta_n\}$ be a sequence of maximum likelihood estimators satisfying the symmetry condition (3.8). In this case, $\{\hat\theta_n\}$ is consistent if and only if*

$$P_\theta\left[\limsup_{n\to\infty} \|\hat\theta_n\| < \infty\right] = 1, \quad \theta \in \Theta. \tag{4.2}$$

According to this corollary, a sequence $\{\hat\theta_n\}$ of symmetric maximum likelihood estimators is consistent if and only if, with probability 1 with respect to each distribution P_θ, $\{\hat\theta_n\}$ is *bounded.* This characterization is an interesting property of the maximum likelihood method, particulary when recalling that a general bounded sequence in $\mathbb{R}^k$ is not necessarily convergent.

Proof of Corollary 4.1. If $\{\hat\theta_n\}$ is consistent, (4.2) follows from the inclusions

$$\left[\lim_{n\to\infty} \hat\theta_n = \theta\right] \subset \left[\limsup_{n\to\infty} \|\hat\theta_n\| \le \|\theta\|\right] \subset \left[\limsup_{n\to\infty} \|\hat\theta_n\| < \infty\right].$$

Now, suppose that (4.2) holds and let $\theta \in \Theta$ be arbitrary but fixed. For each integer k, define the event

$$B_k = \left[\mathbf{x} \in S^\infty \colon \limsup_{n\to\infty} \|\hat\theta_n(\mathbf{x}^n)\| \le k\right], \tag{4.3}$$

so that $\bigcup_{k=1}^\infty B_k = \left[\limsup_{n\to\infty} \|\hat\theta_n\| < \infty\right]$, and the equality in (4.2) yield that there exists an integer $k(\theta)$ such that

$$P_\theta\left[B_{k(\theta)}\right] > 0. \tag{4.4}$$

On the other hand, from (4.3) and (3.8) it follows that each set B_k is symmetric, *i.e.*, if $\mathbf{x} \in B_k$ then $\mathbf{x}_\tau \subset B_k$ for each $\tau \in \mathcal{P}_m$ and $m = 1, 2, 3, \ldots$ (see (1.1)), so that, as in the proof of Lemma 3.1, the Hewitt-Savage zero-one law yields that $P_\theta[B_k] = 0$ or $P_\theta[B_k] = 1$, and then

$$P_\theta[B_{k(\theta)}] = 1, \tag{4.5}$$

by (4.4). To continue, notice that (4.3) yields that if $\mathbf{x} \in B_{k(\theta)}$ then there exists an integer $N(\mathbf{x})$ such that $\|\hat\theta_n(\mathbf{x}^n)\| \le k(\theta) + 1$ for $n \ge n(\mathbf{x})$, *i.e.*, $\mathbf{x} \in \bigcap_{n=N(\mathbf{x})}^\infty [\|\hat\theta_n\| \le k(\theta) + 1] \subset \bigcup_{r=1}^\infty \bigcap_{n=r}^\infty [\|\hat\theta_n\| \le k(\theta) + 1]$, that is,

$$B_{k(\theta)} \subset \bigcup_{r=1}^\infty \bigcap_{n=r}^\infty [\hat\theta_n \in C_\theta],$$

where C_θ is the compact ball $\{\theta \in \mathbb{R}^k \colon \|\theta\| \le k(\theta) + 1\}$; thus,

$$P_\theta\left[\bigcup_{r=1}^\infty \bigcap_{n=r}^\infty \left[\hat\theta_n \in C_\theta\right]\right] = 1,$$

by (4.5). Since $\theta \in \Theta$ is arbitrary, it follows that $\{\hat\theta_n\}$ is consistent, by Theorem 4.1. □

The proof of Theorem 4.1 relies on the following technical tool, which will be verified in the following section.

Theorem 4.2. *Assume that the conditions in Theorem 4.1 hold, let $\theta_0 \in \Theta$ be arbitrary but fixed, and let $K \subset \Theta$ be a compact set such that $\theta_0 \notin K$. In this context, there exists $\mathcal{U}_K \in \mathcal{G}^\infty$ with the following properties (a) and (b):*

(a) $P_{\theta_0}[\mathcal{U}_K] = 1$;

(b) For each $\mathbf{x} \in \mathcal{U}_K$, $\hat\theta_n(\mathbf{x}^n)$ *does not belong to* K *for* n *large enough. More precisely, there exists a function* $N_K{:}\,\mathcal{U}_K \to \{1, 2, 3, \ldots\}$ *such that*

$$\hat\theta_n(\mathbf{x}^n) \notin K, \quad \mathbf{x} \in \mathcal{U}_K, \quad n \geq N_K(\mathbf{x}). \tag{4.6}$$

Proof of Theorem 4.1. Assume that the sequence $\{\hat\theta_n\}$ is consistent. Let $\theta \in \Theta$ be arbitrary but fixed, and select $\varepsilon_\theta > 0$ such that the closed ball $C_\theta = \{\theta' \in \Theta{:}\, d(\theta', \theta) \leq \varepsilon_\theta\}$ is compact; see Assumption 2.3. Observing that

$$\left[\lim_{n\to\infty} \hat\theta_n = \theta\right] \subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[d(\hat\theta_n, \theta) \leq \varepsilon_\theta\right] = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[\hat\theta_n \in C_\theta\right],$$

the consistency of $\{\hat\theta_n\}$ yields (4.1).

Assume that (4.1) holds, where each set C_θ is compact. In this case let $\theta_0 \in \Theta$ be arbitrary but fixed, and observe the following facts (a)–(c):

(a) If $\mathbf{x} \in \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[\hat\theta_n \in C_{\theta_0}\right]$, there exists $k(\mathbf{x})$ such that $\mathbf{x} \in \bigcap_{n=k(\mathbf{x})}^{\infty} \left[\hat\theta_n \in C_{\theta_0}\right]$, *i.e.*,

$$\hat\theta_n(\mathbf{x}^n) \in C_{\theta_0}, \quad n \geq k(\mathbf{x}).$$

(b) Given $\varepsilon > 0$, let $B(\theta_0, \varepsilon) = \{\theta' \in \Theta{:}\, d(\theta', \theta_0) < \varepsilon\}$ be the open ball with center θ_0 and radius $\varepsilon > 0$, and define the set $K = C_{\theta_0} \cap B(\theta_0, \varepsilon)^c$, so that K is compact, $\theta_0 \notin K$, and

$$C_{\theta_0} = K \cup (C_{\theta_0} \cap B(\theta_0, \varepsilon)).$$

(c) Since $\theta_0 \notin K$, by Theorem 4.2 there exists an event $\mathcal{U}_K \in \mathcal{B}(S^\infty)$ with $P_{\theta_0}[\mathcal{U}_K] = 1$, as well as a function $N_K(\cdot){:}\,\mathcal{U}_K \to \{1, 2, 3, \ldots\}$ satisfying that, for each $\mathbf{x} \in \mathcal{U}_K$,

$$\hat\theta_n(\mathbf{x}^n) \notin K, \quad n \geq N_K(\mathbf{x});$$

see (4.6).

Setting $N(\mathbf{x}) = \max\{k(\mathbf{x}), N_K(\mathbf{x})\}$, (a)–(c) together yield

$$\begin{aligned} \mathbf{x} \in \mathcal{U}_k \cap \left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[\hat\theta_n \in C_{\theta_0}\right]\right) &\Longrightarrow \hat\theta_n(\mathbf{x}) \in C_{\theta_0} \cap B(\theta_0, \varepsilon)), \quad n \geq N(\mathbf{x}) \\ &\Longrightarrow \mathbf{x} \in \left[\hat\theta_n \in B(\theta_0, \varepsilon)\right], \quad n \geq N(\mathbf{x}) \\ &\Longrightarrow \mathbf{x} \in \bigcap_{n=N(\mathbf{x})}^{\infty} \left[d(\hat\theta_n, \theta_0) < \varepsilon\right], \end{aligned}$$

so that

$$\mathcal{U}_k \cap \left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[\hat\theta_n \in C_{\theta_0}\right]\right) \subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[d(\hat\theta_n, \theta_0) < \varepsilon\right]$$

and then $P_{\theta_0}\left[\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left[d(\hat\theta_n, \theta_0) < \varepsilon\right]\right] = 1$. Since this latter equality holds for every $\varepsilon > 0$, it follows that $P_{\theta_0}\left[\lim_{n\to\infty} \hat\theta_n = \theta_0\right] = 1$ (Billingsley, 1995, p. 70), and then $\{\hat\theta_n\}$ is a consistent sequence of estimators of θ, since $\theta_0 \in \Theta$ is arbitrary. □

5. Proof of Theorem 4.2

Throughout the remainder $\theta_0 \in \Theta$ is arbitrary but fixed, the assumptions in Sections 2 and 3 are supposed to hold without explicit reference, and $\{\hat{\theta}_n\}$ is a given sequence of maximum likelihood estimators of θ. As usual, instead of analyzing functions $L_n(\cdot; \mathbf{x}^n)$ directly, it is convenient to consider its normalized logarithm

$$\mathcal{L}_n(\theta; \mathbf{x}^n) = \frac{1}{n} \log\left(L_n(\theta; \mathbf{x}^n)\right) = \frac{1}{n} \sum_{i=1}^{n} \log(f(x_i; \theta), \quad \theta \in \Theta, \tag{5.1}$$

where the convention $\log(0) = -\infty$ is enforced; since $\log(\cdot)$ is strictly increasing on $[0, \infty)$, (3.6) is equivalent to

$$\mathcal{L}_n(\hat{\theta}(\mathbf{x}^n); \mathbf{x}^n) \geq \mathcal{L}_n(\theta; \mathbf{x}^n), \quad \theta \in \Theta, \quad \mathbf{x}^n \in M_n. \tag{5.2}$$

The proof of Theorem 4.2 is based on the following technical result.

Theorem 5.1. *Given $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$, there exist positive numbers $\varepsilon(\theta_1)$ and $\Delta(\theta_1)$, as well as an event $\mathcal{U}_{\theta_1} \in \mathcal{G}^\infty$ and a function $N_{\theta_1}: \mathcal{U}_{\theta_1} \to \{1, 2, 3, \ldots\}$ such that the following properties (a) and (b) hold:*

(a) $\mathcal{U}_{\theta_1} \subset S_{\theta_0}^\infty$ and $P_{\theta_0}[\mathcal{U}_{\theta_1}] = 1$;

(b) For each $\mathbf{x} \in \mathcal{U}_{\theta_1}$

$$\mathcal{L}_n(\theta; \mathbf{x}^n) \leq \mathcal{L}_n(\theta_0; \mathbf{x}^n) - \Delta(\theta_1), \quad \text{if} \quad n \geq N_{\theta_1}(\mathbf{x}) \quad \text{and} \quad d(\theta, \theta_1) < \varepsilon(\theta_1).$$

The argument used below to establish this result relies on the following consequence of Jensen's inequality.

Lemma 5.1. *If $\theta_1 \in \Theta \setminus \{\theta_0\}$ and $\nu(S_{\theta_0} \cap S_{\theta_1}^c) = 0$, then assertions (a) and (b) below occur.*

(a) The following inequality holds:

$$E_{\theta_0}\left[\log\left(\frac{f(X_1; \theta_1)}{f(X_1; \theta_0)}\right)\right] < 0, \tag{5.3}$$

where the expectation may be $-\infty$.

(b) There exist positive numbers $\Delta(\theta_1)$ and $\varepsilon(\theta_1)$ such that

$$E_{\theta_0}\left[D_{\varepsilon,\theta_1}^+(X_1) + \log\left(\frac{f(X_1; \theta_1)}{f(X_1; \theta_0)}\right)\right] \leq -2\Delta(\theta_1), \quad \text{if} \quad 0 < \varepsilon \leq \varepsilon(\theta_1).$$

Proof. (a) Since $\nu(S_{\theta_0} \cap S_{\theta_1}^c) = 0$, via (2.5) and (2.7) it follows that

$$1 = \tilde{P}_{\theta_0}[S_{\theta_0}] = \int_{S_{\theta_0}} f(x; \theta_0)\, \nu(dx) = \int_{S_{\theta_0} \cap S_{\theta_1}} f(x; \theta_0)\, \nu(dx) = \tilde{P}_{\theta_0}[S_{\theta_0} \cap S_{\theta_1}] \tag{5.4}$$

and

$$E_{\theta_0}\left[\log\left(\frac{f(X_1; \theta_1)}{f(X_1; \theta_0)}\right)\right] = \int_{S_{\theta_0} \cap S_{\theta_1}} \log\left(\frac{f(x; \theta_1)}{f(x; \theta_0)}\right) f(x; \theta_0)\, \nu(dx) \tag{5.5}$$

Assume now that $f(x; \theta_1)/f(x; \theta_0)$ is not constant ν-almost everywhere on $S_{\theta_0} \cap S_{\theta_1}$. In this case, using the strict concavity of $\log(\cdot)$, (5.3) follows from the two displays above via Jensen's inequality. To conclude, it is sufficient to show that (5.3) holds when, for some $c \in [0, \infty)$,

$$f(x; \theta_1)/f(x; \theta_0) = c \quad \nu\text{-almost everywhere on } S_{\theta_0} \cap S_{\theta_1}. \tag{5.6}$$

In this case, since $f(\cdot;\theta_1)$ is a density,

$$1=\int_{S_{\theta_1}} f(x;\theta_1)\nu(dx)\geq \int_{S_{\theta_0}\cap S_{\theta_1}} f(x;\theta_1)\nu(dx)=c\int_{S_{\theta_0}\cap S_{\theta_1}} f(x;\theta_0)\nu(dx)=c,$$

where (5.4) was used to set the last equality. From (5.5) it is clear that (5.3) will follow if it is shown that $c<1$, inequality that will be now established. Assume that $c=1$, so that the above displayed relation yields $1=\int_{S_{\theta_0}\cap S_{\theta_1}} f(x;\theta_1)\nu(dx)=\int_{S_{\theta_1}} f(x;\theta_1)\nu(dx)$; since $f(\cdot;\theta_1)$ is positive on S_{θ_1}, it follows that

$$\nu\left[S_{\theta_1}\cap S_{\theta_0^c}\right]=0$$

and, moreover, via (5.6), equality $c=1$ yields

$$\nu\left[[x\colon f(x;\theta_1)\neq f(x;\theta_0)]\cap(S_{\theta_0}\cap S_{\theta_1})\right]=0.$$

Observing that $[x\colon f(x;\theta_1)\neq f(x;\theta_0)]\subset S_{\theta_0}\cup S_{\theta_1}$, the condition $\nu\left[S_{\theta_0}\cap S^c_{\theta_1}\right]=0$ and the last two displays together yield $\nu\left[x\colon f(x;\theta_1)\neq f(x;\theta_0)\right]=0$, contradicting Assumption 2.2, since $\theta_1\neq\theta_0$. Therefore, $c<1$ and the proof of part (a) is complete.

(b) Using part (a), select $\Delta(\theta_1)>0$ such that $E_{\theta_0}\left[\log\left(f(X_1;\theta_1)/f(X_1;\theta_0)\right)\right]<-3\Delta(\theta_1)$. By Assumption 2.4(b) there exists $\varepsilon(\theta_1)>0$ such that $E_{\theta_0}\left[D^+_{\varepsilon,\theta_1}(X_1)\right]<\Delta(\theta_1)$ for $\varepsilon\in(0,\varepsilon(\theta_1)]$, and part (b) follows. □

The following simple result involving the discrepancy function in (2.8) will be useful.

Lemma 5.2. *Let $\theta_1\in\Theta\setminus\{\theta_0\}$ be arbitrary but fixed. For every $x\in S_{\theta_0}\cap S_{\theta_1}$ and $\varepsilon>0$*

$$\log\left(f(x;\theta)\right)\leq D^+_{\varepsilon,\theta_1}(x)+\log\left(\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)+\log\left(f(x;\theta_0)\right)\quad \textit{if}\quad d(\theta,\theta_1)<\varepsilon. \tag{5.7}$$

Proof. Let $x\in S_{\theta_0}\cap S_{\theta_1}$ be arbitrary, so that $f(x;\theta_1)f(x;\theta_0)>0$. Firstly, notice that (5.7) is valid when $x\notin S_\theta$, since the left hand side is $-\infty$. Next, assume that $x\in S_\theta$, so that $I[x\in S_\theta\cap S_{\theta_1}]=1$. From $f(x;\theta)=[f(x;\theta)/f(x;\theta_1)]\,[f(x;\theta_1)/f(x;\theta_0)]\,f(x;\theta_0)$ it follows that

$$\begin{aligned}\log\left(f(x;\theta)\right)&=\log\left(\frac{f(x;\theta)}{f(x;\theta_1)}\right)+\log\left(\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)+\log\left(f(x;\theta_0)\right)\\&=\log\left(\frac{f(x;\theta)}{f(x;\theta_1)}\right)I[x\in S_\theta\cap S_{\theta_1}]+\log\left(\frac{f(x;\theta_1)}{f(x;\theta_0)}\right)+\log\left(f(x;\theta_0)\right)\end{aligned}$$

and, via (2.8), this yields (5.7). □

Proof of Theorem 5.1. The argument is divided in two cases according to the value of $\nu[S_{\theta_0}\cap S^c_{\theta_1}]$.

Case 1: $\nu[S_{\theta_0}\cap S^c_{\theta_1}]=0$.

To start with, notice that (2.5) and (2.7) yield that $1=\tilde P_{\theta_0}[S_{\theta_0}]=\tilde P_{\theta_0}[S_{\theta_0}\cap S_{\theta_1}]$, so that

$$P_{\theta_0}\left[(S_{\theta_0}\cap S_{\theta_1})^\infty\right]=1; \tag{5.8}$$

see (2.6). Next, let $\Delta(\theta_1)$ and $\varepsilon(\theta_1)$ be as in Lemma 5.1(b), and define M as the class of all trajectories $\mathbf{x}=(x_1,x_2,x_3,\ldots)\in S^\infty$ such that

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}&\left(\sum_{i=1}^n D^+_{\varepsilon(\theta_1),\theta_1}(x_i)+\sum_{i=1}^n\log\left(\frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}\right)\right)\\&=E_{\theta_0}\left[D^+_{\varepsilon(\theta_1),\theta_1}(X_1)+\log\left(\frac{f(X_1;\theta_1)}{f(X_1;\theta_0)}\right)\right]\leq-2\Delta(\theta_1).\end{aligned}$$

Therefore,

$$P_{\theta_0}[M] = 1, \tag{5.9}$$

by the strong law of large numbers, and there exists $N_{\theta_1}(\cdot): M \to \{1, 2, 3, \ldots\}$ such that

$$\frac{1}{n}\left(\sum_{i=1}^{n} D^{+}_{\varepsilon(\theta_1),\theta_1}(x_i) + \sum_{i=1}^{n}\log\left(\frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}\right)\right) < -\Delta(\theta_1), \quad \mathbf{x} \in M, \quad n \geq N_{\theta_1}(\mathbf{x}). \tag{5.10}$$

On the other hand, for $\mathbf{x} \in (S_{\theta_0} \cap S_{\theta_1})^\infty$, it follows that $x_i \in S_{\theta_0} \cap S_{\theta_1}$ for each i, so that

$$\log\left(f(x_i;\theta)\right) \leq D^{+}_{\varepsilon(\theta_1),\theta_1}(x_i) + \log\left(\frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}\right) + \log\left(f(x_i;\theta_0)\right) \quad \text{if} \quad d(\theta,\theta_1) < \varepsilon(\theta_1),$$

by Lemma 5.2, and then (see (5.1))

$$\mathcal{L}_n(\theta;\mathbf{x}^n) \leq \frac{1}{n}\left(\sum_{i=1}^{n} D^{+}_{\varepsilon(\theta_1),\theta_1}(x_i) + \sum_{i=1}^{n}\log\left(\frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}\right)\right) + \mathcal{L}_n(\theta_0;\mathbf{x}^n)$$
$$\text{if } \mathbf{x} \in (S_{\theta_0} \cap S_{\theta_1})^\infty \text{ and } d(\theta,\theta_1) < \varepsilon(\theta_1)$$

Setting $\mathcal{U}_{\theta_1} = (S_{\theta_0} \cap S_{\theta_1})^\infty \cap M$, part (a) in Theorem 5.1 follows from (5.8) and (5.9), whereas part (b) follows combining the above relation and (5.10).

Case 2: $\nu[S_{\theta_0} \cap S^c_{\theta_1}] > 0$.

In this framework, Assumption 2.4(b) allows to select $\varepsilon(\theta_1) > 0$ and $B \in \mathcal{G}$ such that $\nu[B] > 0$ and

$$B \subset S_{\theta_0} \cap S^c_{\theta} \text{ when } d(\theta,\theta_1) < \varepsilon(\theta_1). \tag{5.11}$$

Since $f(\cdot;\theta_0)$ is positive on S_{θ_0}, it follows that

$$\tilde{P}_{\theta_0}[B] = \int_B f(x;\theta_0)\,\nu(dx) > 0$$

and defining

$$\mathcal{U}_{\theta_1} = \left[\mathbf{x} \in S^\infty_{\theta_0}\colon \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I[x_i \in B] = \tilde{P}_{\theta_0}[B]\right]$$

the strong law of large number yields that $P_{\theta_0}[\mathcal{U}_{\theta_1}] = 1$, so that $\mathcal{U}_{\theta_1}$ satisfies the conclusions in part (a) of Theorem 5.1. To conclude, part (b) of Theorem 5.1 will be verified. Define $N_{\theta_1}\colon \mathcal{U}_{\theta_1} \to \{1, 2, 3, \ldots\}$ by

$$N_{\theta_1}(\mathbf{x}) = \min\{i : x_i \in B\}, \quad \mathbf{x} \in \mathcal{U}_{\theta_1}$$

and notice that $N_{\theta_1}(\cdot)$ is always finite, since $\tilde{P}_{\theta_0}[B] > 0$. For $\mathbf{x} \in \mathcal{U}_{\theta_1}$ and $n > N_{\theta_1}(\mathbf{x})$ there exists a positive integer $i < n$ such that $x_i \in B$; in fact, $i = N_{\theta_1}(\mathbf{x})$ is such an integer. In this case $f(x_i;\theta) = 0$ if $d(\theta,\theta_1) < \varepsilon(\theta_1)$, by (5.11) and (2.7), and then

$$\mathcal{L}_n(\theta;\mathbf{x}^n) = -\infty, \quad n > N_{\theta_1}(\mathbf{x}), \quad d(\theta,\theta_1) < \varepsilon(\theta_1), \quad \mathbf{x} \in \mathcal{U}_{\theta_1};$$

see (5.1) and recall the convention $\log(0) = -\infty$. On the other hand, observe that $x_i \in S_{\theta_0}$ for all i when $\mathbf{x} \in \mathcal{U}_{\theta_1}$, and in this case $f(x_i;\theta_0) > 0$, so that $\mathcal{L}_n(\theta_0;\mathbf{x}^n) \in \mathbb{R}$, by (5.1), and then

$$\mathcal{L}_n(\theta;\mathbf{x}^n) < \mathcal{L}_n(\theta_0;\mathbf{x}^n) - 1, \quad n > N_{\theta_1}(\mathbf{x}), \quad d(\theta,\theta_1) < \varepsilon(\theta_1), \quad \mathbf{x} \in \mathcal{U}_{\theta_1},$$

completing the proof. □

Proof of Theorem 4.2. Let K be an arbitrary compact subset of Θ such that $\theta_0 \in \Theta \cap K^c$. For each $\theta' \in K$, Theorem 5.1 yields the existence of positive numbers $\varepsilon(\theta')$ and $\Delta(\theta')$ as well as an event $\mathcal{U}_{\theta'} \in \mathcal{G}^\infty$ and a function $N_{\theta'}\colon \mathcal{U}_{\theta'} \to \{1, 2, 3, \ldots\}$ such that

$$\mathcal{U}_{\theta'} \subset S^\infty_{\theta_0}, \quad P_{\theta_0}[\mathcal{U}_{\theta'}] = 1, \tag{5.12}$$

and

$$\mathcal{L}_n(\theta; \mathbf{x}^n) \leq \mathcal{L}_n(\theta_0; \mathbf{x}^n) - \Delta(\theta'), \quad \mathbf{x} \in \mathcal{U}_{\theta'}, \quad n > N_{\theta'}(\mathbf{x}), \quad \theta \in B(\theta', \varepsilon(\theta')) \tag{5.13}$$

where, as before, $B(\theta', \varepsilon(\theta'))$ is the open ball in Θ with center θ' and radius $\varepsilon(\theta')$. Observing that $K \subset \bigcup_{\theta' \in K} B(\theta', \varepsilon(\theta'))$, the compactness of K yields that for a finite set $\{\theta_1, \theta_2, \ldots, \theta_r\} \subset K$

$$K \subset \bigcup_{i=1}^{r} B(\theta_i, \varepsilon(\theta_i)). \tag{5.14}$$

Setting

$$\tilde{\mathcal{U}}_K = \bigcap_{i=1}^{r} \mathcal{U}_{\theta_i},$$

$$\tilde{N}_K(\mathbf{x}) = \max\{N_{\theta_1}(\mathbf{x}), N_{\theta_2}(\mathbf{x}), \ldots, N_{\theta_r}(\mathbf{x})\}$$

and

$$\Delta_K = \min\{\Delta(\theta_1), \Delta(\theta_2), \ldots, \Delta(\theta_r)\} > 0,$$

(5.12) yields that

$$\tilde{\mathcal{U}}_K \subset S^\infty_{\theta_0}, \quad \text{and} \quad P_{\theta_0}[\tilde{\mathcal{U}}_K] = 1, \tag{5.15}$$

whereas (5.13) and (5.14) together imply that

$$\mathcal{L}_n(\theta; \mathbf{x}^n) \leq \mathcal{L}_n(\theta_0; \mathbf{x}^n) - \Delta_K, \quad \mathbf{x} \in \tilde{\mathcal{U}}_K, \quad n > \tilde{N}_K(\mathbf{x}), \quad \theta \in K. \tag{5.16}$$

On the other hand, from the definition of M^*, there exists a function $N\colon M^* \to \{1, 2, 3, \ldots\}$ such that if $\mathbf{x} \in M^*$ then $\mathbf{x}^n \in M_n$ for $n > N(\mathbf{x})$; see (3.10) and (3.11). Combining this latter fact with (5.2), it follows that

$$\mathcal{L}_n(\hat{\theta}(\mathbf{x}^n); \mathbf{x}^n) \geq \mathcal{L}_n(\theta; \mathbf{x}^n), \quad \mathbf{x} \in M^*, \quad n > N(\mathbf{x}), \quad \theta \in \Theta. \tag{5.17}$$

To conclude, set $\mathcal{U}_K = \tilde{\mathcal{U}}_K \cap M^*$. With this notation Assumption 3.1 and (5.15) yield that $\mathcal{U}_K \subset S^\infty_{\theta_0}$ and $P_{\theta_0}[\mathcal{U}_K] = 1$, whereas setting $N_K(\mathbf{x}) = \max\{\tilde{N}_K(\mathbf{x}), N(\mathbf{x})\}$ for $\mathbf{x} \in \mathcal{U}_K$, and using that $\Delta_K > 0$, relations (5.16) and (5.17) yield that $\hat{\theta}_n(\mathbf{x}^n) \notin K$ for $\mathbf{x} \in \mathcal{U}_K$ and $n > N_K(\mathbf{x})$, completing the proof. □

References

[1]. Ash, R. B. (1972), *Real Analysis and Probability* (*Academic Press*, New York).

[2]. Azzalini, A. (1996), *Statistical Inference Based on the Likelihood* (*Chapman & Hall/CRC*, London).

[3]. Bahadur, R. R. (1958), Examples of inconsistency of maximum likelihood estimates, *Sankhya*, **18**, 211–224.

[4]. Billingsley, P. (1995) *Probability and Measure* (*Wiley*, New York, 3rd ed.).

[5]. Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation* (*Springer*, New York, 2nd ed.).

[6]. Severini, T. A. (2000), *Likelihood Methods in Statistics* (*Oxford University Press*, New York).

[7]. Shao, J. (1999) *Mathematical Statistics* (*Springer*, New York).