The Annals of Statistics
1996, Vol. 24, No. 3, 1298–1315

EPI-CONVERGENCE OF SEQUENCES OF NORMAL INTEGRANDS AND STRONG CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

BY CHRISTIAN HESS

Université Paris Dauphine

Using the variational properties of epi-convergence together with suitable results on the measurability of multifunctions and integrands, we prove a strong law of large numbers for sequences of integrands from which we deduce a general theorem of almost sure convergence (strong consistency) for the maximum likelihood estimator.

1. Introduction. In mathematical statistics the maximum likelihood principle is one of the most popular in order to construct estimators. On the other hand, an important property of estimators is the strong consistency, that is, the almost sure convergence toward the true value of the unknown parameter, as the sample size tends to ∞. Since Wald [33], who first gave a rigorous and general proof of the consistency for the maximum likelihood estimator (MLE), several other proofs of this fact have appeared, under more or less restrictive hypotheses; for example, a more recent proof can be found in the book by Ibragimov and Has'minski [21] (Theorem 4.3 and Remark 4.1).

The main objective of this paper is to present an approach, based on epi-convergence, for proving the almost sure convergence of the MLE. This is done under weaker hypotheses than those usually assumed. Indeed, in the present paper, the space V of parameters is only assumed to be a Suslin metric space, the space E of observations is an arbitrary measurable space and the observations are pairwise independent (instead of mutually independent) identically distributed E-valued random variables. Further, the function f from the product space $E \times V$ into the positive reals, which defines the family of densities, is not assumed to be continuous with respect to the parameter v. It only satisfies a sup-compactness assumption with respect to v. This hypothesis ensures the existence of MLE's, but we also consider the convergence of approximate MLE's which are more realistic in concrete situations. Our method, using epi-convergence, can also apply to convergence results concerning other kinds of estimators whose construction involves an optimization procedure (this is typically the case of M-estimators).

The principle of our approach consists of regarding the consistency of the MLE as the approximation of a deterministic optimization problem whose solution is the true value of the unknown parameter, by a suitable sequence

Received April 1992; revised July 1995.
AMS 1991 *subject classifications*. 28B20, 49K45, 54D35, 62F12, 62H12, 62L20.
Key words and phrases. Maximum likelihood estimator, strong consistency, measurable multifunctions, normal integrands, set convergence, epi-convergence.

of stochastic optimization problems associated with the samples. This allows us to apply a method for approximating optimization problems, which appeared in the 1960s in the papers of Wijsman [34] and Mosco [26]. It is based on a notion of convergence for sequences of functions often called "epi-convergence" which means the Painlevé–Kuratowski convergence of the epigraphs (upper graphs) of these functions, viewed as subsets of the product space $V \times \mathbf{R}$. Our proof of the strong consistency of the MLE precisely relies upon an epi-convergence result for the averaged sum of integrands which can be viewed as a parametrized version of the strong law of large numbers (Theorem 5.1). This type of result is also useful in stochastic optimization (see [23]), in the calculus of variations (see [10]) and in certain problems of mechanics involving nonhomogeneous materials (see [9]). Moreover, let us mention that this method, for studying the consistency of estimators (and also the rate of convergence), has already been used by Dupacova and Wets [12] under more restrictive hypotheses.

On the other hand, the measurability properties of functions (single-valued as well as multivalued) play an important part in our work, since they are needed to apply the strong law of large numbers to suitable sequences of functions, and to prove the existence of exact and approximate MLE's. This is why we also provide briefly the needed results, following mainly [6] and [8].

The paper is organized as follows: in Section 2 we present our main result which is Theorem 2.1. In Section 3 we provide a short description of epi-convergence and we present convenient expressions of the lower and upper epi-limits. Section 4 is devoted to the measurability of multifunctions and integrands. In Section 5 we prove a strong law of large numbers for integrands, which involves epi-convergence and is one of the main arguments in the proof of Theorem 2.1. Finally, in Section 6, after having shown how our method uses epi-convergence, the proof of the main result is provided as well as some comments and examples.

2. Main result. Let $(\Omega, \mathscr{A}, P)$ be a probability space, X a random variable defined on Ω with values in a measurable space $(E, \mathscr{E})$ and μ a positive σ-finite measure defined on $(E, \mathscr{E})$. Further, let (V, d) be a Suslin metric space whose Borel σ-field is denoted by $\mathscr{B}(V)$; without restricting the generality we can assume $d \le 1$ on V. Also let f be a function from $E \times V$ into $\mathbf{R}_+$, the set of positive reals, which satisfies the following *hypotheses*:

(a) f is $\mathscr{E} \otimes \mathscr{B}(V)$-measurable.

(b) For every $u \in V$, $f(\cdot, u)$ is a probability density function relative to μ.

(c) For μ-almost every $x \in E$, $f(x, \cdot)$ is *sup-compact* in the following sense: for each strictly positive real r, the subset $\{u \in V : f(x, u) \ge r\}$ is compact; but the subset $\{u \in V : f(x, u) = 0\}$ is not assumed to be compact.

(d) $u_1 \ne u_2$ implies $f(\cdot, u_1) \ne f(\cdot, u_2)$, that is, $\mu\{x \in E : f(x, u_1) \ne f(x, u_2)\} > 0$.

(e) For some $\bar{u} \in V$, $f(\cdot, \bar{u})$ is "the" density function of the random variable X with respect to μ.

(f) Let $(X_n)_{n\geq 1}$ be a sequence of E-valued random variables defined on Ω, being pairwise independent and having the same distribution as X; for any integer $n \geq 1$, let $(x_1, \dots, x_n)$ be an n-tuple of possible values of the sample $(X_1, \dots, X_n)$.

In mathematical statistics, Ω is called the set of elementary outcomes, $\mathscr{A}$ is the σ-field of events and P is the probability measure. Further, $(E, \mathscr{E})$ is the space of *observations* and V the space of *parameters*. The random variable X represents the outcome of any observation; it is an E-valued random variable whose density relative to the σ-finite dominating measure μ is assumed to be a member of the family $\{f(\cdot, u): u \in V\}$. From condition (d), it is clear that $\bar{u}$ is unique; it is the *unknown* (*or true*) *value* of the parameter. The problem of *statistical estimation* consists of obtaining an approximate value of $\bar{u}$ after having observed values $x_1, \dots, x_n$ of a random sample $X_1, \dots, X_n$ of X, where n, the *sample size*, is large enough. The approximate value of $\bar{u}$ is provided by a suitable function of $X_1, \dots, X_n$ which is called an *estimator* of the parameter u.

We begin with some definitions and notation. First, define the function b on E by

$$b(x) := \sup\{f(x, u) : u \in V\}. \tag{2.1}$$

The *likelihood function* L_n is defined, for any $u \in V$ and $(x_1, \dots, x_n) \in E^n$, by

$$L_n(x_1, \dots, x_n, u) := \prod_{i=1}^{n} f(x_i, u). \tag{2.2}$$

For each $n \geq 1$, we consider the n-fold product measure space $(E^n, \mathscr{E}^n, \mu_n)$ associated with the measure space $(E, \mathscr{E}, \mu)$. So, $\mathscr{E}^n$ is the product σ-field on E^n and μ_n the n-fold product measure on the measurable space $(E^n, \mathscr{E}^n)$. We also define the function B_n by

$$B_n(x_1, \dots, x_n) := \sup\{L_n(x_1, \dots, x_n, u) : u \in V\}. \tag{2.3}$$

Our main result is as follows.

THEOREM 2.1. *Let* $(\Omega, \mathscr{A}, P)$ *be a probability space*, X *a random variable defined on* Ω *with values in a measurable space* $(E, \mathscr{E})$ *and* μ *a positive* σ*-finite measure defined on* $(E, \mathscr{E})$. *Further*, *let* V *be a Suslin metrizable space and* f *a function from* $E \times V$ *into* $\mathbf{R}_+$ *which satisfies hypotheses* (a) *to* (e) *above*. *Also assume the integrability condition*

$$\int_E f(x, \bar{u}) \log\left(\frac{b(x)}{f(x, \bar{u})}\right) \mu(dx) < +\infty, \tag{H}$$

where $b(\cdot)$ *is defined by* (2.1). *Then for every decreasing sequence* $(\alpha_n)_{n\geq 1}$ *of nonnegative numbers verifying* $\lim_n \alpha_n = 0$, *the two following statements hold true*:

(A) *There exists a sequence* $(t_{n,\alpha_n})_{n\geq 1}$ [*also denoted by* $(t_n)_{n\geq 1}$] *of* α_n-*approximate MLE's, namely, a sequence of maps from* E^n *into* V *satisfying the two following properties*:

(j) *for every* $n \geq 1$, t_n *is* $\mathscr{E}_n$-*measurable*;

(jj) *for every* $(x_1,\dots,x_n) \in E^n$, $L_n(x_1,\dots,x_n,t_n(x_1,\dots,x_n)) \geq B_n(x_1,\dots,x_n) - \alpha_n$.

(B) *For every sequence* (t_n) *as above, one has for almost all* $\omega \in \Omega$,

$$\lim_n t_n(X_1(\omega),\dots,X_n(\omega)) = \bar{u}.$$

REMARK 2.2. (i) In case the X_n's are not mutually independent, it would be better to say that t_n is an M-estimator relative to (2.2) rather than an MLE. (ii) In the foregoing result it can be observed that the (exact or approximate) MLE, whose existence is provided by statement (A), need not be unique and that statement (B) is valid for any sequence of MLE's. Moreover, condition (c) yields the existence of exact MLE's as well.

REMARK 2.3. In [25], Le Cam presented miscellaneous examples where estimators constructed by the maximum likelihood method misbehave. In particular, he studied one, adapted from Bahadur [4], where the MLE is not consistent. In view of Theorem 2.1, it is interesting to note that in this example ([25], pages 157–158) the integrability condition (H) does not hold. However, we shall see in Remark 6.5 that condition (H) is not necessary for the MLE to be consistent.

REMARK 2.4. Moreover, we shall see that one of the main arguments in the proof of Theorem 2.1 is related to the possibility of approximating a normal integrand (i.e., a measurable function defined on $E \times V$, lower semicontinuous with respect to the second variable), by an increasing sequence of Lipschitz integrands (Proposition 4.4). In [24], Lanéry proved a result comparable to Theorem 2.1 for exact MLE's. However, his approach is based on a different approximation scheme of integrands and does not use epi-convergence. On the other hand, see Proposition 3.9 in [12] for a related result, but in a much more special setting.

REMARK 2.5. Let us also mention the important monograph of Hoffmann-Jørgensen [20] (I was not aware of this reference when I wrote the first version of the present paper) in which the general problem of studying the accumulation and limit points of "sequences of approximating maximums" (see Remark 6.4 for the definition) is addressed. This includes the study of the consistency of the MLE, but in [20] "consistency" is used in a different meaning. For example, the above accumulation points are allowed to be outside the given parameter space V. Although the method is different from ours, one can observe that epi-convergence (in fact hypo-convergence) is implicitly used.

3. Epi-convergence of sequences of functions defined on a metric space. In this section we present some needed facts about epi-convergence, but we refer the reader to the monographs [2] and [9]. Let (V, d) be a metric space and f a function from V into $\overline{\mathbf{R}}$, the set of extended reals. f is said to be *proper* if it is not identically $+\infty$ and if it does not take the value $-\infty$. The *epigraph* of f, denoted $\mathrm{epi}(f)$, is defined by

$$\mathrm{epi}(f) := \{(u, \lambda) \in V \times \mathbf{R} : f(u) \leq \lambda\}.$$

The epigraph is also called the "upper graph." A function f is lower semicontinuous (lsc) if and only if $\mathrm{epi}(f)$ is a closed subset of $V \times \mathbf{R}$. Now, let us recall the definition of epi-convergence. In the present section $(f_n)_{n \geq 1}$ [or (f_n) for short] denotes a sequence of functions from V into $\overline{\mathbf{R}}$. For any $u \in V$, the lsc functions $\mathrm{li}_e f_n$ and $\mathrm{ls}_e f_n$ are defined by

$$\mathrm{li}_e f_n(u) := \sup_{k \geq 1} \liminf_{n \to \infty} \inf_{v \in B(u, 1/k)} f_n(v) \tag{3.1}$$

and

$$\mathrm{ls}_e f_n(u) := \sup_{k \geq 1} \limsup_{n \to \infty} \inf_{v \in B(u, 1/k)} f_n(v), \tag{3.2}$$

where $B(u, 1/k)$ denotes the open ball of radius $1/k$ centered at u. The function $\mathrm{li}_e f_n(u)$ is the *epi-limit inferior* and $\mathrm{ls}_e f_n(u)$ is the *epi-limit superior* of the sequence (f_n). The sequence (f_n) is said to be *epi-convergent* at u, if the following equality holds true:

$$\mathrm{li}_e f_n(u) = \mathrm{ls}_e f_n(u). \tag{3.3}$$

If (3.3) is satisfied for any $u \in V$, we simply say that (f_n) *epi-converges*. In this case, the common value of (3.3) defines an lsc function which is called the *epi-limit* of the sequence (f_n) and will be denoted by $\lim_e f_n$. Now we recall the sequential characterization of epi-convergence (Proposition 1.14 in [2]).

PROPOSITION 3.1. *A sequence of functions (f_n) from V into the extended reals epi-converges to a function f at $v \in V$, if and only if the two following properties hold*:

(a) *For each sequence (v_n) converging to v in V, one has* $f(v) \leq \liminf_n f_n(v_n)$.

(b) *There exists a sequence (v_n) in V converging to v and such that* $f(v) \geq \limsup_n f_n(v_n)$.

The term "epi-convergence" comes from the fact that $f = \lim_e f_n$ if and only if the sequence of subsets $(\mathrm{epi}(f_n))$ in $V \times \mathbf{R}$ converges to $\mathrm{epi}(f)$ in the sense of Painlevé and Kuratowski.

Now, let us recall the variational properties of epi-convergence, which play a crucial role in the present paper. For this purpose, some notation is useful.

For any extended real-valued function f defined on V, we define the *set of (exact) minimizers* of f on V by setting

$$\operatorname{Argmin}(f) := \Big\{u \in V : f(u) = \inf_{v \in V} f(v)\Big\}.$$

More generally, for a function f such that the infimum on V is different from $-\infty$, we define, for any $\alpha \geq 0$, the *set of α-approximate minimizers* by

$$\alpha\text{-}\operatorname{Argmin}(f) := \Big\{u \in V : f(u) \leq \inf_{v \in V} f(v) + \alpha\Big\}.$$

Whenever $\alpha > 0$, the set α-approximate minimizers is nonempty (unlike the set of exact minimizers which is obtained for $\alpha = 0$). The variational properties of epi-convergence that we need are stated in the following result (see [2], Corollary 2.10, page 131).

PROPOSITION 3.2. *Let (α_n) be a sequence of positive reals converging to* 0. *Assume that (f_n) is epi-convergent to f, that is, $f = \lim_e f_n$. Then, the following inequality holds*:

$$\inf_{v \in V} f(v) \geq \limsup_{n \to \infty} \Big(\inf_{v \in V} f_n(v)\Big). \tag{3.4}$$

Further, for any $n \geq 1$, let v_n be an α_n-approximate minimizer of f_n. If the sequence (v_n) admits a subsequence converging to some $u \in V$, then u belongs to Argmin(f) *and* (3.4) *becomes*

$$\min_{v \in V} f(v) = \limsup_{n \to \infty} \Big(\inf_{v \in V} f_n(v)\Big). \tag{3.5}$$

Now, let us recall the following procedure of approximation for a function which goes back to Hausdorff. For any $u \in V$ and any integer $k \geq 1$, we set

$$f^k(u) = \inf\{f(v) + k\,d(u,v) : v \in V\}.$$

The function f^k is the *Lipschitz approximation* of f, *of order* k. Its main properties are listed in the following proposition.

PROPOSITION 3.3. *Let f be a lower semicontinuous function from V into $\overline{\mathbf{R}}$, not identically $+\infty$ and satisfying the following condition*: *there exist $a > 0$, $b \in \mathbf{R}$ and $u_0 \in V$ such that, for each $u \in V$, $f(u) + a\,d(u, u_0) + b \geq 0$. Then, the following properties hold*:

(i) *for any $k > a$ and $u \in V$, one has $f^k(u) + a\,d(u, u_0) + b \geq 0$*;

(ii) *for every integer $k \geq 1$, f^k is finite valued and Lipschitzian with Lipschitz constant k*;

(iii) *for any $u \in V$, the sequence $(f^k(u))$ is increasing and $f(u) = \sup_k f^k(u)$.*

The next proposition, which will serve in the proof of Theorem 5.1, provides expressions of $\mathrm{li}_e f_n$ and $\mathrm{ls}_e f_n$ in terms of the Lipschitz approximations of functions f_n. Although this type of result is not entirely new, it extends several previous ones, and the method of proof presented here can be generalized to the case of functions with values in an ordered vector space (see [22]). Similar formulas for the Moreau–Yosida approximation were obtained by Attouch [2], Theorem 2.65, page 232 (see also [27]).

PROPOSITION 3.4. *Assume the following hypothesis*:

(3.6) *there exist* $a > 0$, $b \in \mathbf{R}$ *and* $u_0 \in V$ *such that*, *for any integer* $n \geq 1$ *and for any* $u \in V$, $f_n(u) + a\, d(u, u_0) + b \geq 0$.

Then, *for any* $u \in V$, *the two following equalities hold*:

$$\mathrm{li}_e f_n(u) = \sup_{k \geq 1} \liminf_{n \to \infty} f_n^k(u), \tag{3.7}$$

$$\mathrm{ls}_e f_n(u) = \sup_{k \geq 1} \limsup_{n \to \infty} f_n^k(u). \tag{3.8}$$

PROOF. Let us prove (3.7). Fix $u \in V$. First, we shall show that the left-hand side is greater than or equal to the right-hand side. Using the definitions of the epi-limit inferior and the Lipschitz approximation, we easily obtain

$$\mathrm{li}_e f_n(u) \geq \sup_{p \geq 1} \left[\left(\liminf_{n \to \infty} f_n^k(u) \right) - k/p \right] = \liminf_{n \to \infty} f_n^k(u).$$

This being true for any $k \geq 1$, we obtain the desired inequality. In order to show the reverse inequality, set, for any $u \in V$,

$$\psi(u) := \sup_{k \geq 1} \liminf_{n \to \infty} f_n^k(u). \tag{3.9}$$

We ought to prove $\psi(u) \geq \mathrm{li}_e f_n(u)$. If $\psi(u)$ equals $+\infty$, there is nothing to prove; otherwise, fix an integer $p \geq 1$ and $\alpha \in (0,1)$. For every $n, k \geq 1$, consider the two following conditions:

$$f_n(v) + k\, d(u, v) \geq f_n^k(u) + \alpha \qquad \forall\, v \notin B(u, 1/p) \tag{3.10}$$

and

$$f_n^k(u) = \inf\{ f_n(v) + k\, d(u, v) : v \in B(u, 1/p) \}. \tag{3.11}$$

Clearly, (3.10) implies (3.11). On the other hand, looking at (3.9), we see that, for any $k \geq 1$, one can find a strictly increasing sequence of integers $(n(k, m))_{m \geq 1}$ verifying

$$\psi(u) + \alpha \geq f_{n(k,m)}^k(u). \tag{3.12}$$

Our goal is to obtain condition (3.10) with $n = n(k, m)$, for any $m \geq 1$ and for any $k \geq k_0$, where k_0 will be chosen below. So, for every $v \notin B(u, 1/p)$ we want the following inequality to hold:

$$f_{n(k,m)}(v) + k\, d(u, v) \geq f_{n(k,m)}^k(u) + \alpha, \tag{3.13}$$

which, in view of (3.12), is a consequence of

$$f_{n(k,m)}(v) + k\,d(u,v) \geq \psi(u) + 2. \tag{3.14}$$

Now, using condition (3.6), it is readily seen that (3.14) is implied by the following choice of k:

$$k \geq k_0 := \text{integer part of } \{p(\psi(u) + 2 + a\,d(u,u_0) + b) + a + 1\}.$$

Thus, we can deduce that, if $k \geq k_0$, inequality (3.13) holds for any $m \geq 1$ and $v \notin B(u, 1/p)$ which, as already noticed, implies

$$f^k_{n(k,m)}(u) = \{f_{n(k,m)}(v) + k\,d(u,v) : v \in B(u, 1/p)\}. \tag{3.15}$$

Consequently, using (3.12), we see that, for any $k \geq k_0$, the following relationships are valid:

$$\psi(u) + \alpha \geq \liminf_{m\to\infty} f^k_{n(k,m)}(u) = \liminf_{m\to\infty} \inf_{v\in B(u,1/p)} \left[f_{n(k,m)}(v) + k\,d(u,v)\right],$$

whence

$$\psi(u) + \alpha \geq \liminf_{n\to\infty} \inf_{v\in B(u,1/p)} f_n(v).$$

This last inequality, being true for any $p \geq 1$ and $\alpha \in (0,1)$, yields $\psi(u) \geq \mathrm{li}_e f_n(u)$, which is the desired inequality and, thus, completes the proof of (3.7). The proof of (3.8) is similar. □

REMARK 3.5. It can be observed that condition (3.6) is of a global nature. An inspection of the above proof shows that (3.6) can be replaced by the following *local condition*: for any $u \in V$ there exist a neighborhood W of u and $\beta > 0$ such that, for every $n \geq 1$ and $v \in W$, one has $f_n(v) \geq \beta$.

REMARK 3.6. Epi-convergence is a one-sided notion which has been devised for analyzing the behavior of sequences of minimization problems. The corresponding symmetric notion for maximization problems is that of "hypo-convergence" defined by considering the Painlevé–Kuratowski convergence of the hypographs (lower graphs) of functions.

REMARK 3.7. For a detailed treatment of epi-convergence, set convergence and miscellaneous applications, we refer the reader to [2], [3], [5], [9], [10], [12], [14], [15], [23], [26], [29] and [34].

4. Some facts on the measurability of multifunctions and integrands. In this section we present several measurability properties of functions (or integrands) defined on a product space $S \times V$. These properties concern the measurability of several operations to be used later, infimum, argmin multifunction, sum and Lipschitz approximation. They will also serve to prove the existence of exact and approximate MLE's. Most of the results of this section are known (see, e.g., [6], [8], [17], [18] and [28]).

Let $(S, \mathscr{S})$ be a measurable space and (V, d) a separable metric space whose Borel σ-field is denoted by $\mathscr{B}(V)$. A map Γ defined on S which assigns to each $s \in S$ the set $\Gamma(s)$ in V is called a *multifunction* with values in V. The *graph* $\mathrm{Gr}(\Gamma)$ and the *domain* $\mathrm{dom}(\Gamma)$ of Γ are, respectively, defined by

$$\mathrm{Gr}(\Gamma) := \{(s, u) \in S \times V / u \in \Gamma(s)\} \quad \text{and} \quad \mathrm{dom}(\Gamma) := \{s \in S / \Gamma(s) \neq \varnothing\}.$$

Consider the σ-field $\hat{\mathscr{S}}$ on S defined by $\hat{\mathscr{S}} = \bigcap_{\lambda} \mathscr{S}_{\lambda}$, where λ ranges over the set of all positive σ-finite measures on $(S, \mathscr{S})$ and where $\mathscr{S}_{\lambda}$ denotes the λ-completion of $\mathscr{S}$. $\hat{\mathscr{S}}$ is the σ-field of *universally measurable* (or absolutely measurable) subsets of S, relative to $\mathscr{S}$.

A *selector* of Γ is a function γ from $\mathrm{dom}\,\Gamma$ into V such that, for any $s \in \mathrm{dom}(\Gamma)$, $\gamma(s) \in \Gamma(s)$. Concerning the existence of *measurable selectors*, we recall the following fact (Theorem 3.22 in [8]): if V is a Suslin space, then any multifunction Γ whose graph is $\mathscr{S} \otimes \mathscr{B}(V)$-measurable admits at least one selector which is measurable relative to the σ-fields $\hat{\mathscr{S}}$ and $\mathscr{B}(V)$.

DEFINITION 4.1. Let h be an extended real-valued function defined on $S \times V$ and ψ an extended real-valued function on S. Define the *infimum function* $m(\cdot)$ and the *level set multifunction* Γ associated with h and ψ, by setting, for every $s \in S$.

$$m(s) := \inf\{h(s, v) : v \in V\} \quad \text{and} \quad \Gamma(s) := \{s \in S : h(s, v) \le \psi(s)\}.$$

In the special case where $\psi(\cdot) = m(\cdot)$, the multifunction Γ is called the (exact) *Argmin multifunction* of h. Further, assume that $m(s) > -\infty$ and consider, for any positive real α, the function ψ defined on S by $\psi(s) := m(s) + \alpha$. For such a function ψ, the multifunction Γ defined above is the *approximate Argmin multifunction* of order α.

Now, we provide a result concerning the measurability of the infimum function and the level set multifunction of a function defined on $S \times V$. It is a slight extension of Theorem 2 of [6], page 908.

PROPOSITION 4.2. *Let V be a Suslin space and h an $\mathscr{S} \otimes \mathscr{B}(V)$-measurable function defined on $S \times V$. Then*:

(a) *The infimum function $m(\cdot)$ is $\hat{\mathscr{S}}$-measurable.*

(b) *For any real-valued $\mathscr{S}$-measurable function ψ, the graph of the level set multifunction Γ, associated with h and ψ, is $\hat{\mathscr{S}} \otimes \mathscr{B}(V)$-measurable.*

(c) *The exact and approximate Argmin multifunctions of h have $\hat{\mathscr{S}} \otimes \mathscr{B}(V)$-measurable graphs. Consequently, each of them admits at least one $\hat{\mathscr{S}}$-measurable selector.*

PROOF. To prove (a), it suffices to observe that, for any $\alpha \in \mathbf{R}$, the following equality holds true:

$$\{s \in S : m(s) < \alpha\} = \mathrm{proj}_S(\{(s, v) \in S \times V : f(s, v) < \alpha\})$$

and to invoke the measurable projection theorem (see, e.g., Theorem 3.23 in [8]).

As for (b), since the function $(s, v) \to h(s, v) - \psi(s)$ is $\mathscr{S} \otimes \mathscr{B}(V)$-measurable, $\mathrm{Gr}(\Gamma)$ is a member of this product σ-field. Statement (c) follows easily. □

REMARK 4.3. The version of the measurable projection theorem invoked in the above proof is due to Sainte-Beuve [30] and extends previous results of Aumann and Von Neumann. On the other hand, Brown and Purves also proved a variant of Proposition 4.2 (Corollary 1 in [6], page 904). Their proof is based on another projection theorem due to Arsenin and Novikov ([1], [11] and [32]). This *second projection theorem* states that if B is a nonempty member of the product σ-field $\mathscr{S} \otimes \mathscr{B}(V)$ and satisfies the condition: "for any $s \in S$, the set $B(s) := \{v \in V/(s, v) \in B\}$ is the countable union of compact subsets of V," then $\mathrm{proj}_S(B)$ is a member of $\mathscr{S}$. Further, this theorem implies the existence of an $\mathscr{S}$-measurable selector for multifunction $B(\cdot)$. Clearly, this could provide an alternate setting for the measurability issues of the present paper. See [18] and [31] for more recent developments in this direction and [16] for another variant of Proposition 4.2. The following result concerns the measurability of the Lipschitz approximation.

PROPOSITION 4.4. *Let V be a metric Suslin space, h an $\mathscr{S} \otimes \mathscr{B}(V)$-measurable function defined on $S \times V$, $u_0 \in V$ and $k \geq 1$. In addition, assume that there exists a positive constant a and a function $\alpha(\cdot)$ such that for every $(s, v) \in S \times V$, one has $h(s, v) + a\,d(v, u_0) + \alpha(s) \geq 0$. Then, the Lipschitz approximation of order k of h, denoted by h^k, is $\hat{\mathscr{S}} \otimes \mathscr{B}(V)$-measurable.*

PROOF. For every $u \in V$ the function $(s, v) \to h(s, v) + k\,d(u, v)$ is $\mathscr{S} \otimes \mathscr{B}(V)$-measurable. Hence, using Proposition 4.2(a), we can see that

$$s \to h^k(s, u) := \inf\{h(s, v) + k\,d(u, v) : v \in V\}$$

is an $\hat{\mathscr{S}}$-measurable function. Thus, it suffices to show that, for any $s \in S$, $h^k(s, \cdot)$ is continuous. For this purpose define $S_0 = \{s \in S : h(s, \cdot)$ is identically $+\infty\}$. Clearly, on S_0 the function $h^k(s, \cdot)$ is equal to the constant $+\infty$, whence it is continuous, whereas for every $s \in S \setminus S_0$ the function $h^k(s, \cdot)$ is finite valued and Lipschitzian with Lipschitz constant k. □

5. A strong law of large numbers for integrands. Let $(\Omega, \mathscr{A}, P)$ be a probability space, $(E, \mathscr{E})$ a measurable space, (V, d) a metrizable Suslin space and g a *normal integrand* defined on the product space $E \times V$, that is, an $\mathscr{E} \otimes \mathscr{B}(V)$-measurable function which is lsc with respect to the second variable. Also consider an E-valued random variable X defined on Ω. In the proof of the main result of the section, we shall use Etemadi's SLLN (see [13]). It extends the Kolmogorov SLLN in that the real random variables are only supposed to be pairwise independent, instead of mutually independent. For positive random variables the SLLN holds even if the expectation is equal to $+\infty$.

So, let $(X_n)_{n\geq 1}$ be a sequence of E-valued random variables defined on Ω, pairwise independent and having the same distribution as X. The following theorem, already contained in [16], will be the main argument in the proof of the almost sure convergence of the MLE (Section 6). It also provides an SLLN for integrands which may serve in other situations.

THEOREM 5.1. *Let V be a Suslin metric space and g a normal integrand on $E \times V$ with values in $\overline{\mathbf{R}}_+$. For any $n \geq 1$, define h_n on $\Omega \times V$ by*

$$h_n(\omega, u) := \frac{1}{n} \sum_{i=1}^{n} g(X_i(\omega), u)$$

and also assume that the function $\phi(\cdot) := \mathbf{E}g(X, \cdot)$ is not identically $+\infty$ on V. Then, there exists a P-negligible subset N of Ω such that, for every $u \in V$ and $\omega \in \Omega \setminus N$, one has

$$\mathbf{E}g(X, u) = \lim_{n\to\infty}{}_e h_n(\omega, u), \tag{5.1}$$

where $\mathbf{E}$ stands for the expectation of a random variable.

PROOF. In order to show (5.1), it suffices to prove the two following inequalities:

$$\mathrm{li}_e h_n(\omega, u) \geq \mathbf{E}g(X, u) \qquad \forall\, \omega \in \Omega \setminus N_1, \forall\, u \in V, \tag{5.2}$$

$$\mathrm{ls}_e h_n(\omega, u) \leq \mathbf{E}g(X, u) \qquad \forall\, \omega \in \Omega \setminus N_2, \forall\, u \in V, \tag{5.3}$$

where N_1 and N_2 are some negligible subsets of Ω. First, recall that, for any $\omega \in \Omega$ and for any fixed $k \geq 1$, the Lipschitz approximation, of order k, of $h_n(\cdot, \omega)$ is defined by

$$h_n^k(\omega, u) := \inf\{h_n(\omega, v) + k\, d(u, v) : v \in V\} \qquad \forall\, u \in V.$$

Now, using the super-additivity of the infimum operation, we easily obtain

$$h_n^k(\omega, u) \geq \frac{1}{n} \sum_{i=1}^{n} g^k(X_i(\omega), u).$$

An appeal to Proposition 4.4 shows that g^k and h_n^k are $\hat{\mathscr{A}} \otimes \mathscr{B}(V)$-measurable. Consequently, for any $u \in V$ and $k \geq 1$, we can apply Etemadi's SLLN to the sequence $(g^k(X_n, u))_{n\geq 1}$. This proves the existence of a negligible subset $N_1(u, k)$ such that, for any $\omega \in \Omega \setminus N_1(u, k)$,

$$\liminf_n h_h^k(\omega, u) \geq \mathbf{E}g^k(X, u) \tag{5.4}$$

Set $N_1 := \bigcup_{u\in D} \bigcup_{k\geq 1} N_1(u, k)$, where D is a dense countable subset of V. Inequality (5.4) is valid for $\omega \in \Omega \setminus N_1$, $k \geq 1$ and $u \in D$; moreover, it remains valid for any $u \in V$, because each side of (5.4) defines a Lipschitzian function of u, with Lipschitz constant k. Then, taking the supremum, with respect to k, in both sides of (5.4) and using formula (3.7) together with the monotone convergence theorem, we obtain (5.2).

To prove (5.3), it is useful to put, for any $k > 1$ and $u \in V$, $\phi^k(u) := \inf\{\phi(v) + k\,d(u,v)/v \in V\}$. First, observe that, due to the properness of ϕ, ϕ^k is finite on V. On the other hand, an appeal to Fatou's lemma shows that ϕ is lsc on V. Further, for any $u \in D$, $p \ge 1$ and $k \ge 1$, one can find $v' = v'(u,p,k) \in V$ such that $\phi(v') + k\,d(u,v') \le \phi^k(u) + 1/p$. Hence, for each $u \in D$ and $k \ge 1$, the following equality holds true:

$$\phi^k(u) = \inf\{\phi(v'(u,p,k)) + k\,d(u, v'(u,p,k)) : p \ge 1\}. \tag{5.5}$$

Further, applying Etemadi's strong law of large numbers to the sequence $\{g(X_n, v'(u,p,k))\}_{n \ge 1}$, we can see that, for every $u \in D$, $k \ge 1$ and $p \ge 1$, there exists a negligible subset $N_2(u,p,k)$ such that, for every $\omega \in \Omega \setminus N_2(u,p,k)$,

$$\phi(v'(u,p,k)) = \lim_n h_n(\omega, v'(u,p,k)). \tag{5.6}$$

Put $N_2 := \bigcup_{u \in D} \bigcup_{p \ge 1} \bigcup_{k \ge 1} N_2(u,p,k)$ and consider $\omega \in \Omega \setminus N_2$. For any $u \in D$ and $k \ge 1$, we have

$$\limsup_n h_n^k(\omega, u) \le \inf_{v \in V} \limsup_{n \to \infty} [h_n(\omega, v) + k\,d(u,v)].$$

Restricting the infimum to the subset $\{v'(u,p,k) : p \ge 1\}$ and using (5.6) and (5.5), we obtain

$$\limsup_n h_n^k(\omega, u) \le \inf_{p \ge 1} [\phi(v'(u,p,k), \omega) + k\,d(u, v'(u,p,k))] = \phi^k(u).$$

So, we have proved, for each $k \ge 1$ and $\omega \in \Omega \setminus N_2$,

$$\limsup_n h_n^k(\omega, u) \le \phi^k(u) \qquad \forall\, u \in D. \tag{5.7}$$

Then, invoking once more the Lipschitz property, we conclude that (5.7) remains valid for all $u \in V$. Finally, taking the supremum on k in both sides of (5.7) and using (3.8), we get (5.3). □

REMARK 5.2. Let us mention some extensions of Theorem 5.1.

(a) It is not difficult to check that the conclusion of this theorem still holds under the following less restrictive hypotheses:

(i) g is a normal integrand on $E \times V$.
(ii) There exist an integrable function $\beta(\cdot)$, a constant $a > 0$ and $u_0 \in V$ such that, for any (x,u) in the product space $E \times V$, $g(x,u) + \beta(x) + a\,d(u,u_0) \ge 0$.

(b) Returning to Remark 3.5, it is readily seen that the conclusion of Theorem 5.1 also holds under the following *local condition*: for any $u \in V$ there exist a neighborhood W of u and an integrable function $\beta(\cdot)$ such that, for every $(x,v) \in E \times V$, $g(x,v) \ge \beta(x)$.

(c) More generally, an inspection of the above proof shows that if g is not assumed to be lsc on V, but only $\mathscr{E} \otimes \mathscr{B}(V)$-measurable, the conclusion continues to hold at each point u such that $g(x, \cdot)$ is lsc at u, for almost all x in E.

REMARK 5.3. Results similar to Theorem 5.1 have been obtained by King and Wets [23] when V is a reflexive separable Banach space and by Attouch and Wets [3] when V is a separable Banach space. On the other hand, the measurably parametrized Lipschitz approximation studied in Section 4 has already been used by Castaing [7] in order to prove epi-convergence results for certain sequences of integrands.

6. Proof of the main result. Observe first that, by condition (c) in Section 2 and Proposition 4.2(a), the function b defined by

$$b(x) := \sup\{f(x,u) : u \in V\}$$

is finite valued and $\hat{\mathscr{E}}$-measurable. Also define the function F from V^2 into the extended reals by

$$F(u,v) := \int_E f(x,u)\log\left(\frac{b(x)}{f(x,v)}\right)\mu(dx). \tag{6.1}$$

REMARK 6.1. In (6.1), it is possible to replace the measure μ by $b\mu$, which is still σ-finite, and the function f by f/b, so that F can be rewritten as

$$F(u,v) := \int_E f(x,u)\log\frac{1}{f(x,v)}\mu(dx).$$

Therefore, in view of the definition of F, f can be assumed to take its values in $[0,1]$. This will be assumed from now on.

Further, as in Section 5, define the function ϕ on V by $\phi(v) := \mathbf{E}g(X,v) = F(\bar{u},v)$, where $g(x,v) := -\log f(x,v)$. Recall that $\bar{u}$ denotes the unknown value of the parameter to be estimated.

REMARK 6.2. (i) Recall that, using the convexity of the function $t \to -\log(t)$ and Jensen's inequality, it is readily shown that $\phi(\bar{u}) = \inf[\phi(v) : v \in V]$. (ii) In addition, because the convexity is strict, the inequality $\phi(u) \le \phi(\bar{u})$ implies $u = \bar{u}$.

PROOF OF THEOREM 2.1(A). Here, the sample size n is assumed to be fixed. From condition (c) it follows that the function B_n, defined on E^n by

$$B_n(x_1,\dots,x_n) := \sup\{L_n(x_1,\dots,x_n,u) : u \in V\}, \qquad x_1,\dots,x_n \in E^n, \tag{6.2}$$

is finite valued. For any fixed n and $\alpha > 0$, define the multifunction $\Gamma_{n,\alpha}$ on E^n, with closed values in V, by

$$\Gamma_{n,\alpha}(x_1,\dots,x_n) := \{u \in V : L_n(x_1,\dots,x_n,u) \ge B_n(x_1,\dots,x_n) - \alpha\}$$

and also, the multifunction Γ_n such that

$$\Gamma_n(x_1,\dots,x_n) := \{u \in V : L_n(x_1,\dots,x_n,u) = B_n(x_1,\dots,x_n)\}.$$

Now consider the n-fold product measure μ_n on the n-fold product space $(E^n, \mathscr{E}^n)$. Using Proposition 4.2, it is not hard to prove the measurability of function B_n, and of multifunctions Γ_n and $\Gamma_{n,\alpha}$, relative to the σ-field $(\mathscr{E}^n)\mu_n$ which denotes the μ_n-completion of the product σ-field $\mathscr{E}^n$ on E^n. Then, denoting by t_n a selector of Γ_n and, for any $\alpha > 0$, by $t_{n,\alpha}$ a selector of $\Gamma_{n,\alpha}$, which are measurable with respect to $(\mathscr{E}^n)\mu_n$, we obtain, for any $(x_1, \dots, x_n) \in E^n$,

$$L_n(x_1, \dots, x_n, t_n(x_1, \dots, x_n)) = B_n(x_1, \dots, x_n) \tag{6.3}$$

and

$$L_n(x_1, \dots, x_n, t_{n,\alpha}(x_1, \dots, x_n)) \geq B_n(x_1, \dots, x_n) - \alpha. \tag{6.4}$$

Clearly, t_n is an MLE and $t_{n,\alpha}$ an α-approximate MLE for the parameter $u \in V$. Thus part (A) of Theorem 2.1 is proved. □

Now, it will be convenient to introduce the following notation:

$$g(x, u) := -\log f(x, u) \qquad \forall\, (x, u) \in E \times V$$

and, for every $n \geq 1$ and $\omega \in \Omega$,

$$h_n(\omega, u) := \frac{1}{n} \sum_{i=1}^{n} g(X_i(\omega), u) = -\frac{1}{n} \log L_n(X_1(\omega), \dots, X_n(\omega), u).$$

By the convention of Remark 6.1, g takes its values in the interval $[0, +\infty]$. Further, it is clear that, for each $n \geq 1$ and $\omega \in \Omega$, the maximization of L_n with respect to u is equivalent to the minimization of $h_n(\omega, \cdot)$.

Now, let us explain how epi-convergence is involved in our approach to the convergence of MLEs (or more generally of M-estimators). For each $n \geq 1$ and for almost all ω in Ω, we have to consider the minimization problem

$$(\mathscr{P}_n) \qquad \min\{h_n(\omega, u) : u \in V\}.$$

The SLLN for positive random variables implies that, for any $u \in V$, $\mathbf{E}g(X, u) = \lim_n h_n(\omega, u)$ for all ω in $\Omega \setminus N(u)$, where $N(u)$ is a negligible subset of Ω (depending generally of u). Thus it is natural to consider also the following minimization problem:

$$(\mathscr{P}_\infty) \qquad \min\{\mathbf{E}g(X, u) : u \in V\}.$$

Note that $(\mathscr{P}_n)$ is a stochastic minimization problem, in that the variable ω appears in the objective function, whereas $(\mathscr{P}_\infty)$ is a deterministic minimization problem in which ω no longer appears. In Theorem 5.1, we have shown that, for almost all $\omega \in \Omega$, the sequence $(h_n(\omega, \cdot))$ epi-converges on V to $\mathbf{E}g(X, \cdot)$. We shall deduce, from this and from the sup-compactness assumption on f, the almost sure convergence of the MLE, as the sample size n tends to ∞.

PROOF OF THEOREM 2.1(B).

Step 1. V is assumed to be compact.

As already noted, we may assume that $b(x) = 1$ on E. Further, observe that hypothesis (H) implies that $\mathbf{E}g(X, \cdot)$ is proper, which allows us to apply Theorem 5.1. Thus, there exists a negligible subset N of Ω verifying, as n tends to ∞.

$$\mathbf{E}g(X, u) = \lim_e h_n(\omega, u) \qquad \forall\, u \in V, \forall\, \omega \in \Omega \setminus N.$$

Now, let $(\alpha_n)_{n \ge 1}$ be a decreasing sequence of positive reals converging to 0 and, for each $n \ge 1$, let t_n be an α_n-approximate MLE. Thus, for each $\omega \in \Omega \setminus N$, $t_n(\omega) := t_n(X_1(\omega), \dots, X_n(\omega))$ satisfies

$$h_n(\omega, t_n(\omega)) \le \inf_{v \in V} h_n(\omega, v) + \alpha_n.$$

Consequently, using Proposition 3.2, we get

$$\limsup_n h_n(\omega, t_n(\omega)) \le \limsup_n \inf_{v \in V} h_n(\omega, v) \le \inf_{v \in V} \mathbf{E}g(X, v), \tag{6.5}$$

whence from Remark 6.2(i)

$$\limsup_n h_n(\omega, t_n(\omega)) \le \mathbf{E}g(X, \bar{u}). \tag{6.6}$$

On the other hand, for any $\omega \in \Omega \setminus N$, one can extract a subsequence $(t_{n(k)}(\omega))_{k \ge 1}$ converging to some $v_0 \in V$ (the subsequence and the limit point generally depend upon ω). Using Proposition 3.1, we obtain

$$\liminf_k h_{n(k)}\big(\omega, t_{n(k)}(\omega)\big) \ge \mathbf{E}g(X, v_0),$$

whence by (6.6) $\mathbf{E}g(X, v_0) \le \mathbf{E}g(X, \bar{u})$ which, by Remark 6.2(ii), implies $v_0 = \bar{u}$. Therefore, $\bar{u}$ is the unique cluster point of the sequence $(t_n(\omega))_{n \ge 1}$ so that $\bar{u} = \lim_n t_n(\omega)$.

Step 2. The general case.

Consider the embedding Ψ from V into the Hilbert cube W, and define g', h'_n, for $n \ge 1$, and ϕ' on $E \times W$, $\Omega \times W$ and W, respectively, by setting, for any $\omega \in \Omega$, $x \in E$ and $w \in \Psi(V)$,

$$g'(x, w) := g(x, \Psi^{-1}(w)), \qquad h'_n(\omega, w) := h_n(\omega, \Psi^{-1}(w)),$$
$$\phi'(w) := \phi(\Psi^{-1}(w)).$$

These functions are taken to be equal to $+\infty$ if $w \notin \Psi(V)$. From these definitions, we deduce that, for any $(\omega, w) \in \Omega \times W$, one has

$$h'_n(\omega, w) = \frac{1}{n} \sum_{i=1}^{n} g'(X_i(\omega), w) \quad \text{and} \quad \phi'(w) = \mathbf{E}g'(X, w).$$

For any $x \in E$, the sup-compactness of $f(x, \cdot)$ implies the inf-compactness of $g(x, \cdot)$ on V, which, in turn, implies the lower semicontinuity of $g'(x, \cdot)$ on W. Indeed, for any $r \in \mathbf{R}$ and $x \in E$, we have $\{w \in W : g'(x, w) \le r\} = \Psi(\{v \in V / g(x, v) \le r\})$. Thus, for each $n \ge 1$, h'_n is inf-compact, too. Further, applying Fatou's lemma, it is easily seen that ϕ' is lsc on W. Now, let $(t_n)_{n\geq 1}$ be a sequence of an α_n-approximate MLE and consider the sequence $(t'_n)_{n\geq 1}$ defined by $t'_n(\omega) := \Psi(t_n(\omega))$. As in the first step, we can show that the sequence $(t'_n(\omega))_{n\geq 1}$ converges to some $\overline{w} \in W$. Then, using (3.5), it is not hard to see that $\overline{w} = \Psi(\overline{u})$ for $\overline{u} \in V$, which is a solution of $(\mathscr{P}_\infty)$. Finally, the continuity of Ψ^{-1} yields $\overline{u} = \lim_n t_n(\omega)$ a.s. □

REMARK 6.3. Several variants of Theorem 2.1 could be given. For instance, using Remark 5.2(a) instead of Theorem 5.1 in the proof of part (B), it is easily seen that condition (H) can be replaced by the two following conditions:

(a) There exist $u_0 \in V$, $a > 0$ and $\beta(\cdot) \in L^1$ such that, for any $(x, v) \in E \times V$,

$$f(x,v) \leq \exp\{a\, d(v,u_0) + \beta(x)\}.$$

(b) $$\int_E f(x,\overline{u})\left\{\log f(x,\overline{u})^{-1}\right\}^{-} \mu(dx) < +\infty,$$

where, for any real u, $u^{-} := \max(-u, 0)$.

REMARK 6.4. It is interesting to make some additional remarks comparing our results with those obtained recently by Hoffmann-Jørgensen in the first chapter of his monograph [20].

(i) In [20], in order to study the consistency of estimators obtained by a maximization procedure, Hoffman-Jørgensen has introduced the notion of *sequence of approximating maximums* (SAMs). In our setting, in terms of minimization (due to the use of epi-convergence instead of hypo-convergence), it can be observed that the sequence (t_n) appearing in inequality (6.6) precisely corresponds to a SAM. In fact, the above proof shows that, in the presence of epi-convergence, a sequence (t_n) of α_n-approximate MLE's, such that $\lim_n \alpha_n = 0$, is a SAM.

(ii) On the other hand, in [20], no sup-compactness condition like (c) is explicitly assumed. In our approach, this condition yields the finiteness of function $b(\cdot)$ and ensures that any limit point of the sequence $(t_n(\omega))$ is a member of V. In [20], there is no similar requirement, since the accumulation points of sequences of MLE's are allowed to go outside the given parameter space V.

REMARK 6.5. Finally, it is important to point out that, in spite of its interest in explaining some pathological behaviors of the MLE (see Remark 2.3), condition (H) is not necessary for the MLE to be consistent, as the following example shows. Take $E = \mathbf{N}^*$ = the set of strictly positive integers, V = the set of all probability measures on E (V is endowed with the l^1 metric) and μ the measure on E which assigns the value 1 to each integer. Then we have $f(x,v) = v_x$ and $b(x) \equiv 1$. Using the SLLN and Lebesgue's dominated convergence theorem, it is readily seen that the MLE is consistent, but hypothesis (H) need not be satisfied. Indeed, it suffices to assume that the true value of the parameter is a sequence $\bar{u} = (\bar{u}_n)$ in V such that $\sum_{n \geq 1} \bar{u}_n \log(\bar{u}_n)^{-1} = +\infty$.

Acknowledgments. I am indebted to Professors E. Lanéry and M. Valadier and Doctor V. Jalby for valuable discussions. I also thank Professor L. D. Brown as well as the referees for their numerous remarks which helped me to improve the preliminary versions of this paper.

REFERENCES

[1] ARSENIN, W. J., LJAPUNOV, A. A. and CEGOLKOV, E. A. (1955). *Arbeiten zur deskriptiven mengenlehre*. Deutscher Verlag der Wissenschaften, Berlin.

[2] ATTOUCH, H. (1984). *Variational Convergence for Functions and Operators*. Pitman, Boston.

[3] ATTOUCH, H. and WETS, R. (1991). Epigraphical processes: laws or large numbers for random lsc functions. Exposé No. 13, Séminaire d'Analyse Convexe, Univ. Montpellier, France.

[4] BAHADUR, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā Ser. A* **20** 207–210.

[5] BEER, G. (1993). *Topologies on Closed and Closed Convex Sets*, *Mathematics and Its Applications*. Kluwer, Dordrecht.

[6] BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 902–912.

[7] CASTAING, C. (1976). À propos de l'existence des sections séparément mesurables et séparément continues d'une multiapplication séparément semi continue inférieurement. *Travaux Sém. Anal. Convexe* **6**.

[8] CASTAING, C. and VALADIER, M. (1977). *Convex Analysis and Measurable Multifunctions*. *Lecture Notes in Math.* **580**. Springer, New York.

[9] DAL MASO, G. (1993). *An Introduction to Γ-Convergence*. Birkhäuser, Boston.

[10] DE GIORGI, E. (1979). Convergence problems for functions and operators. *Proceedings of the International Meeting on Recent Methods in Nonlinear Analysis*, *Roma 1978* (E. De Giorgi, E. Magenes, J. Mosco, Pitagora and Bologna, eds.) 131–188.

[11] DELLACHERIE, C. (1975). Ensembles analytiques: théorèmes de séparation et applications. *Séminaire de Probabilites IX. Lecture Notes in Math.* **465**. Springer, Berlin.

[12] DUPACOVA, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.

[13] ETEMADI, N. (1981). An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete* **55** 119–122.

[14] HESS, C. (1991). Convergence of conditional expectations for unbounded random sets. *Math. Oper. Res.* **16** 627–649.

[15] HESS, C. (1991). On multivalued martingales whose values may be unbounded: martingale selectors and Mosco convergence. *J. Multivariate Anal.* **39** 175–201.

[16] HESS, C. (1991). Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. Cahier de Mathematique de la Decision, Univ. Paris Dauphine. **9121**.

[17] HIMMELBERG, C. J. (1975). Measurable relations. *Fund. Math.* **87** 53–72.

[18] HIMMELBERG, C. J., PARTHASARATHY, T. and VAN VLECK, F. S. (1981). On measurable relations. *Fund. Math.* **111** 161–167.

[19] HOFFMANN-JØRGENSEN, J. (1970). *The Theory of Analytic Sets*. Various Publications Series 10. Mat. Inst., Aarhus Univ.

[20] HOFFMAN-JØRGENSEN, J. (1992). *Asymptotic Likelihood Theory*. *Functional Analysis III*. Various Publications Series 40. Matematisk Institute, Aarhus Univ.

[21] IBRAGIMOV, I. A. and HAS'MINSKI, R. Z. (1981). *Statistical Estimation*, *Asymptotic Theory*. Springer, New York.

[22] JALBY, V. (1993). Semi-continuité, convergence et approximation des applications vectorielles, Loi des grands nombres. Preprint. Laboratoire d'Analyse Convexe, Univ. Montpellier.

[23] KING, A. J. and WETS, R. (1991). Epi-consistency of convex stochastic programs. *Stochastics Stochastics Rep.* **34** 83–92.

[24] LANÉRY, E. (1991). Convergence presque sûre du maximum de vraisemblance. *C. R. Acad. Sci. Paris Sér. I Math.* **313** 301–303.

[25] LE CAM, L. (1990). Maximum likelihood: an introduction. *Internat. Statist. Rev.* **58** 153–171.

[26] MOSCO, U. (1969). Convergence of convex sets and of solutions of variational inequalities. *Adv. in Math.* **3** 510–585.

[27] PENOT, J. P. and POMMELET, A. (1985). Toward minimal assumptions for the infimal convolution regularization. Technical Report 23, AVAMAC, Univ. Perpignan.

[28] ROCKAFELLAR, R. T. (1976). Integral functionals, normal integrands and measurable selections. *Non-linear Operators and Calculus of Variations. Lecture Notes in Math.* **543**. Springer, Berlin.

[29] ROCKAFELLAR, R. T. and WETS, R. (1984). Variational systems, an introduction. In *Multifunctions and Integrands, Stochastic Analysis, Approximation and Optimization* (G. Salinetti, ed.). *Lecture Notes in Math.* **1091** 1–54. Springer, Berlin.

[30] SAINTE-BEUVE, M. F. (1974). On the extension of Von Neumann–Aumann's theorem. *J. Funct. Anal.* **17** 112–129.

[31] NOUGUÈS-SAINTE-BEUVE, M. F. (1981). Une extension des théorèmes de Novikov et d'Arsenin. *Sém. Anal. Convexe* **18**.

[32] SAINT-RAYMOND, J. (1976). Boréliens à coupes K_σ. *Bull. Soc. Math. France* **104** 389–400.

[33] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[34] WIJSMAN, R. A. (1966). Convergence of sequences of convex sets. II. *Trans. Amer. Math. Soc.* **123** 32–45.

CENTRE DE RECHERCHE DE MATHEMATIQUES
DE LA DECISION (CEREMADE)
URA CNRS NO. 749
UNIVERSITÉ PARIS DAUPHINE
75775 PARIS CEDEX 16
FRANCE
E-MAIL: hess@paris9.dauphine.fr