The Annals of Probability
2003, Vol. 31, No. 1, 63–92

A FUNCTIONAL VERSION OF THE BIRKHOFF ERGODIC THEOREM FOR A NORMAL INTEGRAND: A VARIATIONAL APPROACH

BY CHRISTINE CHOIRAT, CHRISTIAN HESS AND RAFFAELLO SERI

Université Paris Dauphine, Université Paris Dauphine and CREST-LFA

In this paper, we prove a new version of the Birkhoff ergodic theorem (BET) for random variables depending on a parameter (alias integrands). This involves variational convergences, namely epigraphical, hypographical and uniform convergence and requires a suitable definition of the conditional expectation of integrands. We also have to establish the measurability of the epigraphical lower and upper limits with respect to the σ-field of invariant subsets. From the main result, applications to uniform versions of the BET to sequences of random sets and to the strong consistency of estimators are briefly derived.

1. Introduction. In ergodic theory, the Birkhoff ergodic theorem is certainly one of the central results and starting point for further generalizations. It has for a long time proved to be a useful tool in several areas, such as mechanics, statistics and mathematical physics. In many fields, it seems interesting to have a corresponding result holding for random variables depending on parameters, that is, for random functions. For such objects, several types of convergence can be looked for. In the literature some results have been shown to hold for uniform convergence [see, e.g., Burke (1965) in connection with the Glivenko–Cantelli problem].

In the present paper, we shall focus our attention on epigraphical convergence for sequences of stochastic functions defined on a metric space. Epigraphical convergence (epiconvergence, for short) is weaker than uniform convergence, but it is well suited to approximate minimization problems. Indeed, under suitable compactness assumptions, it entails the convergence of infima and minimizers [see, e.g., Attouch (1984b) or Dal Maso (1993)]. A symmetric notion, called hypographical convergence, enjoys similar properties with respect to maximization problems. Further, as it is known, this type of convergence is closely related to the Painlevé–Kuratowski convergence for sequences of subsets,

Received January 2001; revised November 2001.

AMS 2000 subject classifications. Primary 60F17; secondary 28D05, 60G10, 37A30, 62F12, 49J35, 26E25, 28B20, 52A20.

Key words and phrases. Birkhoff ergodic theorem, stationary sequences, normal integrands, measurable set-valued maps, epigraphical convergence, set convergence, strong consistency of estimators.

so that it has an interesting geometric interpretation allowing for connections with the theory of random sets.

Our main result (Theorem 2.3) consists of a version of the Birkhoff ergodic theorem (BET) for random variables depending on a parameter. More precisely, given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a separable metric space (E, d) and an $\mathcal{A} \otimes \mathcal{B}(E)$-measurable function f from $\Omega \times E$ into the extended reals, we prove that for almost all $\omega \in \Omega$, one has

$$\operatorname*{lim_e}_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega, \cdot) = (\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot) \tag{1.1}$$

on E, where T denotes a given measure-preserving transformation and $\mathcal{I}$ is the σ-field of invariant sets (see Section 2.2). On the right-hand side, $\mathbb{E}^{\mathcal{I}} f$ denotes the conditional expectation of the integrand f whose precise definition and main properties will be given in Section 2.1. Moreover, the subscript "e" indicates that the limit is an epigraphical one. This kind of result can be useful in many applied or theoretical situations that are briefly described in the following.

In stochastic programming, one has often to solve an optimization problem of the form (1.2) [see Birge and Louveaux (1997), page 332]

$$\inf_{x\in E} \mathbb{E} g(Y, x), \tag{1.2}$$

where $E \subset \mathbb{R}^p$ and Y is a $\mathbb{R}^q$-valued random variable. Most of the time the integral functional (also called the mean functional) $\mathbb{E} g(Y, x)$ cannot be explicitly calculated, but can be approximated through sampling methods [see Birge and Louveaux (1997), Chapter 10]. Note that much more general stochastic programming problems can be dealt with in this way, such as multistage stochastic programs with recourse and stochastic programs with chance constraints. Suppose that a sample of realizations of the random variable Y, say $(Y_i)_{i=1,\dots,n}$, is available, we would like to find conditions under which the solution of the approximated problem

$$\inf_{x\in E} \frac{1}{n} \sum_{i=1}^{n} g(Y_i, x)$$

converges almost surely to the solution of the original problem (1.2). It is common to assume that $(Y_i)_{i=1,\dots,n}$ is a sample of independent and identically distributed realizations of the random variable Y and that the approximated objective function converges uniformly on E almost surely to the original one,

$$\frac{1}{n} \sum_{i=1}^{n} g(Y_i, x) \to \mathbb{E} g(Y, x) \qquad \text{a.s.};$$

this can be cast in the framework of the Glivenko–Cantelli problem and implies the convergence of the minimizers. Under a suitable compactness assumption, our main theorem allows us to extend the previous results to the case in which $(Y_i)_{i=1,\dots,n}$ is a realization of a stationary ergodic time series and the convergence is only epigraphical.

The previous framework can be extended to encompass also the so-called M-estimation used in statistics and econometrics, that is, estimation obtained by optimizing a function with respect to some parameters (Section 2.5). We remark that in this case great advantage can be gained from considering a more general metric space E, in order to allow for nonparametric and semiparametric estimation, and that epigraphical convergence allows for dealing with discontinuous objective functions, which can be of interest in robust statistics and in constrained estimation.

As explained more deeply in Section 2.4, duality theory allows us to derive some useful results for random sets, such as a Painlevé–Kuratowski ergodic theorem for the sum of integrable closed convex random sets and for the essential intersection of integrable random sets; moreover, it is shown how these results could be easily extended to generalize some results appearing in the literature.

Finally, our main theorem can be used to derive useful results in the theory of homogenization of composite materials. In these materials, the physical properties such as conductivity, elasticity and so on, vary randomly with the location: the objective is to find a homogeneous material whose macroscopic characteristics are similar to the properties of the inhomogeneous one. In these cases epigraphical convergence has proved to be very useful since the equilibrium relation can often be written as a minimization problem [see Attouch (1984a) and Dal Maso and Modica (1986a, b)].

However, the epigraphical version of the BET in the nonergodic case is more difficult than standard results, because one has to define, and to handle with care, the conditional expectation for random functions such as function f in (1.1). Indeed, the proof of our main theorem relies heavily on the definition and on appropriate results dealing with the conditional expectation of a random variable depending on a parameter, that is, of a map f from $\Omega \times E$ into the extended reals. In the sequel, it will be convenient to use the name integrand for this kind of object, in the same spirit as in Castaing and Valadier (1977), Rockafellar (1976) and Rockafellar and Wets (1984). We shall also make precise the notions of equality and inequality between two integrands. These matters are closely examined in Section 3.2, especially when f is lower semicontinuous with respect to the second variable and satisfies a local minorization condition. The Lipschitz continuity is also considered in connection with an useful approximation scheme which is involved in our main result.

The next section contains our main result and a few applications; the proofs are deferred to Section 4. Section 3 contains auxiliary results. First, we recall some properties of epiconvergence and, in particular, its connection with uniform convergence. Section 3.2 is devoted to the comparison between two integrands, from which the uniqueness of the conditional expectation is derived. In Section 3.3, we prove a result asserting that the epigraphical lower and upper limits are measurable with respect to the invariant σ-field $\mathcal{I}$. In Section 3.4 we state a version of the BET for extended real-valued positive random variables and we deduce Lemma 3.9, which plays a crucial part in the proof of our main result. In Section 5, we compare our results with already existing ones. At last, in the Appendix, we provide a proof of the existence and uniqueness of the conditional expectation of an integrand for the convenience of the reader.

2. The main results. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (E, d) be a Suslin metric space whose Borel σ-field is $\mathcal{B}(E)$. (A Suslin topological space is the continuous image of a Polish space.)

We denote by $\mathcal{L}^0(\Omega, \mathcal{A})$ [resp. $\mathcal{L}^0(\Omega, \mathcal{A}; E)$] the set of all $\mathcal{A}$-measurable functions with values in $\overline{\mathbb{R}}$ (resp. in E). The quotient space with respect to the $\mathbb{P}$-almost sure equality is denoted by $L^0(\Omega, \mathcal{A}, \mathbb{P})$ [resp. $L^0(\Omega, \mathcal{A}, \mathbb{P}; E)$].

We say that an extended function $f : \Omega \times E \to \overline{\mathbb{R}}$ is an integrand if it is $\mathcal{A} \otimes \mathcal{B}(E)$-measurable. [For real-valued functions, this corresponds exactly to the Neveu (1964), Definition III-4-4, page 86 and Gikhman and Skorokhod (1969), Definition 1, page 157 definitions of a measurable random real function.] Furthermore, f is called a normal integrand if $f(\omega, \cdot)$ is a lower semicontinuous (lsc) function for almost all $\omega \in \Omega$. f is said to be k-Lipschitz on E if for almost all $\omega \in \Omega$ and for all $x, y \in E$,

$$|f(\omega, x) - f(\omega, y)| \le k d(x, y).$$

An integrand is said to be positive if, for almost every $\omega \in \Omega$, $f(\omega, \cdot)$ takes on its values in $[0, +\infty]$.

We say that the integrands f_1 and f_2 are equal, and we write $f_1 = f_2$, if $f_1(\omega, x) = f_2(\omega, x)$ for all $(\omega, x) \in (\Omega \setminus N) \times E$, where N is a suitable negligible subset of $(\Omega, \mathcal{A}, \mathbb{P})$. To use the terminology of stochastic processes, f_1 and f_2 are indistinguishable. The inequality $f_1 \le f_2$ is defined in the same way.

Given a sequence $(B_i)_{i \ge 1}$ in $\mathcal{B}(E)$ and a sequence $(m_i)_{i \ge 1}$ of real-valued integrable functions, we say that an integrand f satisfies the condition denoted by $\mathcal{C}[(B_i), (m_i), i \ge 1]$, or simply $(\mathcal{C})$, if the following properties hold:

$(\mathcal{C})$
(a) E is covered by $(B_i)_{i \ge 1}$; namely $E = \bigcup_{i \ge 1} B_i$;
(b) for all $i \ge 1$, one has $f(\omega, x) \ge m_i(\omega)$ for all $(\omega, x) \in \Omega \times B_i$.

Further, we say that f satisfies condition $(\mathcal{C}_0)$ if the Borel subsets B_i are assumed to be open.

Observe that condition $(\mathcal{C})$ implies that, for every $i \ge 1$ and every $v \in L^0(\Omega, \mathcal{A}, \mathbb{P}; B_i)$, the following inequality holds:

$$\int_\Omega f(\omega, v(\omega)) \mathbb{P}(d\omega) \ge \int_\Omega m_i(\omega) \mathbb{P}(d\omega)$$

so that the left-hand side is well defined and does not take the value $-\infty$.

2.1. *Conditional expectation of an integrand.* The Birkhoff ergodic theorem involves the conditional expectation of a random variable. Generalizing this result to integrands (i.e., to random variables depending on parameters) requires a suitable extension of the concept of conditional expectation.

THEOREM 2.1. *Let f be an $\mathcal{A} \otimes \mathcal{B}(E)$-measurable integrand satisfying condition $\mathcal{C}[(B_i), (m_i), i \geq 1]$ and $\mathcal{B}$ be a sub-σ-field of $\mathcal{A}$. Then, there exists a $\mathcal{B} \otimes \mathcal{B}(E)$-measurable integrand g satisfying condition $\mathcal{C}[(B_i), (\mathbb{E}^{\mathcal{B}}(m_i)), i \geq 1]$ and*

$$\int_B f(\omega, v(\omega))\mathbb{P}(d\omega) = \int_B g(\omega, v(\omega))\mathbb{P}(d\omega)$$

for all $B \in \mathcal{B}$, for all $i \geq 1$ and for all $v \in L^0(\Omega, \mathcal{B}, \mathbb{P}; B_i)$. Moreover, the integrand g is unique up to indistinguishability. It is also denoted by $\mathbb{E}^{\mathcal{B}} f$.

REMARK 2.1. (i) The presence of the function v in the above equality is justified by Proposition 3.5. Equivalently, for every $i \geq 1$ and $v \in L^0(\Omega, \mathcal{B}, \mathbb{P}; B_i)$, one has $(\mathbb{E}^{\mathcal{B}} f)(\cdot, v(\cdot)) = \mathbb{E}^{\mathcal{B}}[f(\cdot, v(\cdot))]$ a.s., where $f(\cdot, v(\cdot))$ stands for the measurable function $\omega \mapsto f(\omega, v(\omega))$. In this almost sure equality, the negligible set may depend on v. We could say that $(\mathbb{E}^{\mathcal{B}} f)(\cdot, v(\cdot))$ is a modification of $\mathbb{E}^{\mathcal{B}}[f(\cdot, v(\cdot))]$. It is known that this concept is weaker than indistinguishability.

(ii) As explained in Remark 3.2(i) when $\mathcal{B}$ is different from $\{\Omega, \varnothing\}$, it is not possible to replace the class of $\mathcal{B}$-measurable functions by that of singletons $\{x\}$ where x ranges over E.

(iii) Consider the special case where a regular conditional probability Q with respect to $\mathcal{B}$ exists [see, e.g., Breiman (1992)]. Then, for every $\mathcal{B}$-measurable function $v : \Omega \to E$, there exists a negligible subset N such that

$$g(\omega, v(\omega)) = \int_\Omega f(\omega', v(\omega'))\, Q(d\omega', \omega)$$

for every $\omega \in \Omega \backslash N$. Conversely this property characterizes the integrand g.

COROLLARY 2.2. *Under the same hypotheses as in Theorem* 2.1, *the following two statements hold.*

(i) *If f is k-Lipschitz on each B_i, for $i \geq 1$, so is $g = \mathbb{E}^{\mathcal{B}}(f)$.*

(ii) *Assume in addition that f satisfies condition $(\mathcal{C}_0)$, that is, that the B_i's of condition $(\mathcal{C})$ are open. If f is a normal integrand, then $g = \mathbb{E}^{\mathcal{B}}(f)$ is a $\mathcal{B} \otimes \mathcal{B}(E)$-measurable normal integrand.*

REMARK 2.2. (i) In particular, condition $(\mathcal{C})$ is satisfied when f is positive. On the other hand, condition $(\mathcal{C})$ can be replaced by a similar one involving majorization, instead of minorization.

(ii) In Theorem 2.1, what is really needed to prove the existence of $\mathbb{E}^{\mathcal{B}} f$ is an abstract measurable space $(E, \mathcal{E})$. The metric structure of E (and the Suslin property) is only used for proving uniqueness and Corollary 2.2.

(iii) When the integrand f is k-Lipschitz on E, condition $(\mathcal{C})$ can be weakened: it is enough to assume the existence of some $x_0 \in E$, such that the function $\omega \mapsto f(\omega, x_0)$ is integrable.

(iv) Theorem 2.1 extends results of Bismut (1973) and of Castaing and Valadier (1977), where the space E was supposed to be a Banach space and the integrand to be convex with respect to the second variable. Later, Thibault (1981) considered the case of a normal integrand (i.e., lower semicontinuous with respect to the second variable) in connection with another integrability condition which is stronger than ours. In view of the local character of condition $(\mathcal{C})$, our results are noncomparable variants of those of Evstigneev (1986), Truffert (1991) and Castaing and Ezzaki (1993).

2.2. *An epigraphical Birkhoff ergodic theorem.* Since epiconvergence is present in our main result, we provide a short presentation. More details are given in Section 3.1. Let $h : E \to \overline{\mathbb{R}}$ be a function from E into the extended reals. Its *epigraph* is defined by:

$$\mathrm{Epi}(h) = \{(x, \lambda) \in E \times \mathbb{R} : h(x) \le \lambda\}.$$

The *hypograph* of h, denoted by $\mathrm{Hypo}(h)$, is defined by reversing the inequality. Let $(h_n)_{n\ge 1}$ [or (h_n) for short] be a sequence of functions from E into $\overline{\mathbb{R}}$. For any $x \in E$, we introduce the quantities

$$\begin{aligned} \mathrm{li}_e\, h_n(x) &\triangleq \sup_{k\ge 1} \liminf_{n\to\infty} \inf_{y \in \mathsf{B}(x,1/k)} h_n(y), \\ \mathrm{ls}_e\, h_n(x) &\triangleq \sup_{k\ge 1} \limsup_{n\to\infty} \inf_{y \in \mathsf{B}(x,1/k)} h_n(y), \end{aligned} \tag{2.1}$$

where $\mathsf{B}(x, 1/k)$ denotes the open ball of radius $1/k$ centered at x. The function $x \mapsto \mathrm{li}_e\, h_n(x)$ [resp. $x \mapsto \mathrm{ls}_e\, h_n(x)$] is called the lower (resp. upper) epilimit of the sequence (h_n). These functions are lsc. If $\mathrm{li}_e\, h_n(x) = \mathrm{ls}_e\, h_n(x)$, then (h_n) is said to be epiconvergent at x. If this is true for all $x \in E$, then the sequence (h_n) epiconverges. Its epilimit is denoted by $\lim_e h_n$.

Equalities (2.1) have a geometric counterpart involving the Painlevé–Kuratowski convergence of epigraphs on the space of closed sets of $E \times \mathbb{R}$ [see, e.g., Attouch (1984b) or Dal Maso (1993)]. The Painlevé–Kuratowski convergence is defined as follows. Given a sequence $(C_n)_{n\ge 1}$ of sets in E, we define

$$\begin{aligned} \mathrm{Li}\, C_n &\triangleq \{x \in E : x = \lim x_n,\ x_n \in C_n,\ \forall n \ge 1\}, \\ \mathrm{Ls}\, C_n &\triangleq \{x \in E : x = \lim x_i,\ x_i \in C_{n(i)},\ \forall i \ge 1\}, \end{aligned}$$

where $(C_{n(i)})_{i\geq 1}$ is a subsequence of $(C_n)_{n\geq 1}$. The subsets $\mathrm{Li}\, C_n$ and $\mathrm{Ls}\, C_n$ are the *lower limit* and the *upper limit* of $(C_n)_{n\geq 1}$. It is not difficult to check that they are both closed and that they satisfy $\mathrm{Li}\, C_n \subset \mathrm{Ls}\, C_n$. A sequence $(C_n)_{n\geq 1}$ is said to *converge* to C, in the sense of Painlevé–Kuratowski, if $C = \mathrm{Li}\, C_n = \mathrm{Ls}\, C_n$. This is denoted by $C = \text{PK-lim}_{n\to\infty} C_n$. As mentioned above, this notion is strongly connected with epiconvergence: a sequence of functions $h_n : E \to \overline{\mathbb{R}}$ epiconverges to h if and only if the sequence $(\mathrm{Epi}(h_n))_{n\geq 1}$ PK-converges to $\mathrm{Epi}(h)$, in $E \times \mathbb{R}$.

Let $T : \Omega \to \Omega$ be an $\mathcal{A}$-measurable transformation. We assume that T is measure-preserving; that is, $\mathbb{P}(T^{-1}A) = \mathbb{P}(A)$ for all $A \in \mathcal{A}$.

The sets $A \in \mathcal{A}$ that satisfy $T^{-1}A = A$ are called invariant sets and constitute a sub-σ-field $\mathcal{I}$ of $\mathcal{A}$. We shall use repeatedly the following basic result: a random variable X is $\mathcal{I}$-measurable iff $X(\omega) = X(T\omega)$ for all $\omega \in \Omega$. X is said to be an invariant random variable.

Here is the main result of the present paper. It is an epigraphical version of the Birkhoff ergodic theorem for random variables depending on a parameter.

THEOREM 2.3. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $T : \Omega \to \Omega$ be a measure-preserving transformation and (E, d) be a Suslin metric space. Further, let $f : \Omega \times E \to \overline{\mathbb{R}}$ be an $\mathcal{A} \otimes \mathcal{B}(E)$-measurable normal integrand satisfying condition $\mathcal{C}_0[(B_i), (m_i), i \geq 1]$. Under these conditions, $\mathbb{E}^{\mathcal{I}} f$ is a $\mathcal{I} \otimes \mathcal{B}(E)$-measurable normal integrand and one has for almost all $\omega \in \Omega$,*

$$(\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot) = \operatorname*{lim_e}_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega, \cdot). \tag{2.2}$$

REMARK 2.3. (i) A similar minorization condition has been considered by Hess (1991, 1996) and Artstein and Wets (1996) in their results on the epigraphical SLLN. This assumption is weaker than the one in Korf and Wets (2001): indeed, the latter assume that for every $x \in E$ there exists a closed neighborhood such that the infimum on the neighborhood is integrable. Clearly the second hypothesis constrains the infimum of the integrand $f(\omega, x)$ to be integrable, while the first one does not. For example, if E reduces to a singleton $\{x\}$, Korf and Wets' condition requires that $\omega \mapsto f(\omega, x)$ is integrable, whereas our condition only requires that $\omega \mapsto \max\{-f(\omega, x), 0\}$ is integrable.

(ii) A straightforward extension of Theorem 2.3 is worth being mentioned. Following Gray and Kieffer (1980), let us recall that a probability measure μ on $(\Omega, \mathcal{A})$ is said to be asymptotically mean stationary (ams) with respect to T if the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A)$$

exists for every $A \in \mathcal{A}$. Further, Gray and Kieffer's Theorem 1 shows that a probability measure μ on $(\Omega, \mathcal{A})$ is ams if and only if for every bounded measurable function $f : \Omega \to \mathbb{R}$, $\frac{1}{n}\sum_{i=0}^{n-1} fT^i$ converges μ-a.s., as $n \to \infty$. A quick inspection of the proof of Theorem 2.3 shows that its conclusion remains valid assuming only that $\mathbb{P}$ is ams with respect to T.

Our results also allow for recovering epigraphical convergence in the ergodic case. The measurable and measure-preserving transformation T is said to be ergodic if $\mathbb{P}(A) = 0$ or 1 for all invariant sets A. In this case, the invariant σ-field reduces to $\{\Omega, \varnothing\}$ and the conditional expectation of the integrand f, denoted here by $\mathbb{E}f$, is the deterministic function defined on the space E by

$$(\mathbb{E}f)(x) = \int_\Omega f(\omega, x)\mathbb{P}(d\omega), \qquad x \in E.$$

COROLLARY 2.4. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$, T *and* (E, d) *be as in Theorem* 2.3, *and* $f : \Omega \times E \to \overline{\mathbb{R}}$ *be an* $\mathcal{A} \otimes \mathcal{B}(E)$-*measurable normal integrand satisfying condition* $(\mathcal{C}_0)$. *If in addition we assume that T is ergodic then the following equality holds for almost every* $\omega \in \Omega$:

$$(\mathbb{E}f)(\cdot) = \lim_{n\to\infty}{}_e \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega, \cdot).$$

REMARK 2.4. In the statement of the BET, we have used a measurable and measure-preserving transformation $T : \Omega \to \Omega$. An almost surely equivalent formulation can be given in terms of stationary sequences. Recall that a sequence $X_1, X_2, \ldots$ is said to be *stationary* if

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_{k+1} \le x_1, \ldots, X_{k+n} \le x_n)$$

for all integers $n, k \ge 1$ and for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Indeed, any stationary sequence $X_1, X_2, \ldots$ can almost surely be rewritten using a measurable and measure-preserving transformation T [see, e.g., Proposition 6.11 in Breiman (1992)]. In particular, rewriting Corollary 2.4 in terms of stationary sequences yields a version of the SLLN for iid sequences, in the sense of epigraphical convergence.

2.3. *A uniform ergodic theorem.* Theorem 2.3 allows for proving a uniform version of the Birkhoff ergodic theorem. We need the following condition $(\mathcal{C}_0')$, which can be viewed as a bilateral version of condition $(\mathcal{C}_0)$:

There exist a sequence $(B_i)_{i\ge1}$ of open subsets of E and a sequence $(m_i)_{i\ge1}$ of real-valued integrable functions satisfying the conditions:

$(\mathcal{C}_0')$ (a) E is covered by the above sequence of open subsets;
(b) for all $i \ge 1$, one has $|f(\omega, x)| \le m_i(\omega)$ for all $(\omega, x) \in \Omega \times B_i$;
(c) for almost every $\omega \in \Omega$, $f(\omega, \cdot)$ is continuous.

THEOREM 2.5. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$, T *and* (E, d) *be as in Theorem* 2.3, *and* $f : \Omega \times E \to \mathbb{R} \cup \{+\infty\}$ *be an* $\mathcal{A} \otimes \mathcal{B}(E)$*-measurable normal integrand satisfying condition* $(\mathfrak{C}_0')$.

Under these hypotheses, for every compact subset K *of* E, *one has for almost all* $\omega \in \Omega$,

$$\lim_{n\to\infty} \sup_{x\in K} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega, x) - (\mathbb{E}^{\mathcal{I}} f)(\omega, x) \right| = 0.$$

2.4. *Application to the convergence of random sets.* In the following, we assume that E is a finite-dimensional Banach space and we denote by $\mathfrak{C}$ (resp. $\mathfrak{C}_c$) the set of all nonempty closed (resp. closed convex) subsets of E. On 2^E we consider the Minkowski addition, denoted by "+", and the scalar multiplication respectively defined by

$$C + C' \triangleq \{x + x' : x \in C,\ x' \in C'\},$$
$$\alpha C \triangleq \{\alpha x : x \in C\},$$

where $C, C' \in 2^E$ and $\alpha \in \mathbb{R}$. Consider a set-valued map (alias multifunction, correspondence) F from Ω to $\mathfrak{C}$. A function f from Ω into E is called a *selection* of F if, for almost all $\omega \in \Omega$, one has $f(\omega) \in F(\omega)$. By $L^1(E) \triangleq L^1(\Omega, \mathcal{A}, \mathbb{P}; E)$, we denote the Banach space of (equivalence classes of) measurable functions $f : \Omega \to E$ such that the integral $\int_\Omega \|f(\omega)\| \mathbb{P}(d\omega)$ is finite. A map F from Ω into $\mathfrak{C}$ is said to be $\mathcal{A}$-measurable if for every open subset U of E, the set $\{\omega \in \Omega : F(\omega) \cap U \neq \varnothing\}$ is a member of $\mathcal{A}$. A measurable set-valued map is also called a random set. For any measurable set-valued F we define

$$S^1(F, \mathcal{A}) \triangleq \{f \in L^1(\Omega, \mathcal{A}, \mathbb{P}; E) : f(\omega) \in F(\omega) \text{ a.s.}\}.$$

In this definition, $\mathcal{A}$ can be replaced by any sub-σ-field $\mathcal{B}$ of $\mathcal{A}$. $S^1(F, \mathcal{A})$ is a $L^1(E)$-closed set. It is nonempty if and only if the function $d(0, F(\cdot)) \in L^1$. If $S^1(F, \mathcal{A}) \neq \varnothing$, F is said to be integrable. Given two $\mathcal{A}$-measurable closed valued random sets F and G, the following equivalence holds [see Hiaï and Umegaki (1977)]:

$$S^1(F, \mathcal{A}) = S^1(G, \mathcal{A}) \iff F(\omega) = G(\omega) \qquad \text{a.s.}$$

The set-valued integral of an integrable multifunction F is defined by

$$I(F) \triangleq \left\{ \int_\Omega f \, d\mathbb{P} : f \in S^1(F, \mathcal{A}) \right\},$$

where $\int_\Omega f \, d\mathbb{P}$ is the integral (or expectation) of f, also denoted by $\mathbb{E}(f)$. This set-valued integral, originally introduced by Aumann (1965), was defined with respect to the interval [0, 1] endowed with the Lebesgue measure.

Given a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ and an integrable $\mathcal{A}$-measurable random set F, Hiaï and Umegaki (1977) showed the existence of a $\mathcal{B}$-measurable integrable random set G such that

$$S^1(G, \mathcal{B}) \triangleq \mathrm{cl}\{\mathbb{E}^{\mathcal{B}} f : f \in S^1(F, \mathcal{A})\},$$

the closure being taken in $L^1(E)$. G is unique up to a null set. It is called the (set-valued) conditional expectation of F given $\mathcal{B}$ and is denoted by $\mathbb{E}^{\mathcal{B}} F$.

The following result is a version of the BET for closed convex-valued random sets. Another approach for proving it can be found in Krupa [(1998), Theorem 8.2.4].

THEOREM 2.6. *Let $T : \Omega \to \Omega$ be a measure-preserving transformation and $\mathcal{I}$ be the σ-field of invariant sets. If F is a closed convex-valued integrable random set, then for almost every $\omega \in \Omega$, one has*

$$\mathbb{E}^{\mathcal{I}}(F)(\omega) = \text{PK-}\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} F(T^i \omega).$$

Another application can be given to the essential intersection [see, e.g., Hiriart-Urruty (1977)]. Here, we assume that $\mathcal{I} = \{\Omega, \varnothing\}$. The *essential intersection* (also called continuous intersection) of $F : \Omega \to \mathfrak{C}$ is denoted by $\mathrm{Int}(F)$ and defined by

$$\mathrm{Int}(F) = \bigcup_{N \in \mathcal{N}} \bigcap_{\omega \in \Omega \setminus N} F(\omega),$$

where $\mathcal{N}$ denotes the set of all null sets of $(\Omega, \mathcal{A}, \mathbb{P})$. Given a member C of $\mathfrak{C}$, the *support function* and the *indicator function* of C are, respectively, denoted by $s(\cdot, C)$ and $\chi(\cdot, C)$. (Here, it is the convex analysis indicator function.) They are defined by

$$s(y, C) \triangleq \sup\{\langle y, x\rangle : x \in C\}, \qquad y \in E,$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product, and for $x \in E$,

$$\chi(x, C) \triangleq \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{if } x \notin C. \end{cases}$$

Moreover, when $C \in \mathfrak{C}_c$ these functions are conjugate to each other [see, e.g., Aubin (1998), page 48]. Indeed, the following equalities hold:

$$s(y, C) = \sup\{\langle y, x\rangle - \chi(x, C) : x \in E\},$$

$$\chi(x, C) = \sup\{\langle y, x\rangle - s(y, C) : y \in E\}.$$

By a result of Hiriart-Urruty [(1977), Proposition 21, page IV.34], the indicator function of $\mathrm{Int}(F)$ is given by the integral

$$\chi(x, \mathrm{Int}(F)) = \int_\Omega \chi(x, F(\omega))\mathbb{P}(d\omega), \qquad x \in E.$$

On the other hand, the indicator function of a finite intersection of subsets is the sum of the indicator functions. Using this fact and Theorem 2.3, we can prove the following results, which shows that the essential intersection is almost surely the PK-limit of a finite intersection.

THEOREM 2.7. *Let $T : \Omega \to \Omega$ be a measure preserving transformation. If T is ergodic and if F is an integrable random set then for almost every $\omega \in \Omega$, one has*

$$\mathrm{Int}(F) = \text{PK-}\lim_{n\to\infty} \bigcap_{i=0}^{n-1} F(T^i\omega).$$

REMARK 2.5. (i) Following a remark in the monograph by Krengel [(1985), page 10], continuous versions of Theorem 2.3 could be easily proved. Thus, Theorem 2.6 can be viewed as an extension to the case of random sets with unbounded values of Theorems 3.14 and 3.15 of Wang and Wang (1997). There, the random sets are assumed to be integrably bounded, whence almost surely bounded valued (see Section 4). However, as mentioned in Krengel (1985), the case of local continuous ergodic theorems would require more care.

(ii) Several versions of the BET have already been proved for random sets. Let us mention the works of Hess (1979, 1984), of Schürger (1983) and of Krupa (1998). On the other hand, for a related result dealing with set-valued versions of the Lebesgue derivation theorem, see Hess (1992).

(iii) In view of Remark 2.4, it can be observed that Theorem 2.6 implies the set-valued version of the SLLN proved by Artstein and Hart (1981) for closed convex random sets.

(iv) Since, on the space of nonempty compact subsets of $\mathbb{R}^d$, the PK-convergence coincides with the convergence induced by the Hausdorff distance, Theorem 2.6 also implies the set-valued version of the SLLN proved by Artstein and Vitale (1975).

2.5. *Application to statistical estimation.* As explained in the Introduction, our Theorem 2.3 can be most helpful in establishing consistency and measurability results for M-estimators and for solutions of stochastic programming problems. In the following, we provide an application to a problem of statistical estimation. Further extensions will be deferred to a forthcoming paper.

Consider a stochastic process $(Y_t)_{t\in\mathbb{N}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ such that any Y_t takes on its values in a measurable space $(V, \mathcal{V})$: for convenience, we consider the coordinate-variable process, obtained defining $\Omega \triangleq V^{\mathbb{N}}$, $\omega \triangleq (y_t)_{t\in\mathbb{N}}$ and identifying any random variable Y_t with the projection operator $Y_t(\omega) \triangleq y_t$. Moreover, under stationarity of $(Y_t)_{t\in\mathbb{N}}$, the process can be represented through the measure-preserving shift operator T as $Y_t(\omega) = Y_0(T^t\omega)$: then ergodicity properties of the stochastic process $(Y_t)_{t\in\mathbb{N}}$ can be expressed in terms of ergodicity of the shift transformation T.

Then we consider a statistical estimation problem in which the estimator is the maximizer of an objective function defined as the product of the marginal probability density functions $L_n(\omega, x) \triangleq \prod_{t=1}^{n} g(Y_t, x)$, where x belongs to the parameter space E; clearly, L_n is not a likelihood function. It is, however, a *pseudolikelihood* according to econometrics jargon, and it is very useful when the complete density function of the observed data $(y_t)_{t=1,\dots,n}$ is too complicated. Through the maximization of the pseudolikelihood, it is possible to obtain estimators of the parameters appearing in the marginal density function: other parameters can be estimated using ad hoc procedures [see Gouriéroux, Monfort and Trognon (1985)]. The M-estimator $(t_n)_{n\geq 1}$ defined through the maximization of $L_n(\omega, x)$ with respect to the parameter x is often called pseudomaximum-likelihood estimator, or PMLE. In particular, Theorem 2.8 allows us to deal with the following statistical estimation problems:

1. In cross-sectional estimation, that is, when observations are drawn at a fixed time from a population, we obtain consistency when random variables are not independent.
2. In time series estimation, we obtain a consistent estimator even when the temporal dependence structure of the process $(Y_t)_{t\in\mathbb{N}}$ is unknown and is neglected: remark, however, that the allowed range of dependence is restricted by the ergodicity assumption.

Similar pseudolikelihoods have been studied by Levine (1983) and Gouriéroux, Monfort and Trognon (1985) in the time series case, but our assumptions on the density functions are much weaker than theirs. In the cross-section case, we give a version of Theorem 2.1 of Hess (1996) in the case of stationary ergodic observations.

We state Theorem 2.8 without proof: It can be established mimicking Theorem 2.1 of Hess (1996) and substituting our Theorem 2.3 to his Theorem 5.1. See Hess (1996) for a discussion of the hypotheses.

THEOREM 2.8. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space*, $(V, \mathcal{V})$ *a measurable space and* μ *a positive* σ*-finite measure defined on* $(V, \mathcal{V})$. *Further, let* E *be a Suslin metrizable space and* g *a function from* $V \times E$ *into* $\mathbb{R}_+$ *which satisfies the following hypotheses*:

(i) g *is* $\mathcal{V} \otimes \mathcal{B}(E)$*-measurable*;

(ii) *for every* $x \in E$, $g(\cdot, x)$ *is a probability density function relative to* μ;

(iii) *for* μ*-almost every* $y \in V$, $g(y, \cdot)$ *is* sup-compact *in the following sense*: *for each strictly positive real* r, *the subset* $\{x \in E : g(y, x) \geq r\}$ *is compact*, *but the subset* $\{x \in E : g(y, x) = 0\}$ *is not assumed to be compact*;

(iv) $x_1 \neq x_2$ *implies* $g(\cdot, x_1) \neq g(\cdot, x_2)$, *that is*, $\mu\{\omega \in \Omega : g(y, x_1) \neq g(y, x_2)\} > 0$;

(v) *for some* $x_0 \in E$, $g(\cdot, x_0)$ *is a version of the true marginal density of* Y_0;

(vi) *let* $(Y_n)_{n\geq 1}$ *be a sequence of* V*-valued random variables defined on* Ω, *forming a stationary ergodic process*; *for any integer* $n \geq 1$, *let* $(y_1, \dots, y_n)$ *be an* n*-tuple of possible values of the sample* $(Y_1, \dots, Y_n)$;

(vii) *the integrability condition* $\mathbb{E}[\ln \sup_{x\in E} g(Y, x) - \ln g(Y, x_0)] < +\infty$ *holds, where* $\mathbb{E}$ *represents integration with respect to* $\mathbb{P}$.

Then for every decreasing sequence $(\alpha_n)_{n\geq 1}$ *of nonnegative numbers verifying* $\lim_n \alpha_n = 0$, *the following two 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 PMLEs*, *namely*, *a sequence of maps from* V^n *into* E *satisfying the two following properties*:

   (i) *for every* $n \geq 1$, t_n *is* $\mathcal{V}^{\otimes n}$*-measurable*;

   (ii) *for every* $(y_1, \dots, y_n) \in V^n$,

$$L_n\big(y_1, \dots, y_n, t_n(y_1, \dots, y_n)\big) \geq \sup\{L_n(y_1, \dots, y_n, x) \mid x \in E\} - \alpha_n.$$

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

$$\lim_n t_n\big(Y_1(\omega), \dots, Y_n(\omega)\big) = x_0.$$

3. Auxiliary results.

3.1. *Epiconvergence and uniform convergence.* Given a lsc function $h : E \to \overline{\mathbb{R}}$ and an integer $k \geq 1$, the Lipschitz approximation of order k of h is defined by:

$$h^k(x) \triangleq \inf_{y\in E}\{h(y) + kd(x, y)\}, \qquad k \geq 1.$$

Its main properties are listed in the following proposition.

PROPOSITION 3.1. *Let* $h : E \to \overline{\mathbb{R}}$ *be a lsc function nonidentically equal to* $+\infty$. *Suppose that there exists* $a > 0$, $b \in \mathbb{R}$ *and* $x_0 \in E$ *such that*, *for all* $x \in E$, $h(x) + ad(x, x_0) + b \geq 0$. *Then*:

(i) $\forall\, k > a$ *and* $\forall\, x \in E$, $h^k(x) + ad(x, x_0) + b \geq 0$;

(ii) $\forall\, k \geq 1$, $h^k < +\infty$ *and* h^k *is Lipschitz of constant* k;

(iii) $\forall\, x \in E$, *the sequence* $(h^k(x))_{k\geq 1}$ *is increasing and* $h(x) = \sup_{k\geq 1} h^k(x)$.

Moreover, the Lipschitz approximations provide a useful characterization of the lower and upper epilimits defined in (2.1) [see Hess (1996), Proposition 3.4].

PROPOSITION 3.2. *Let* (h_n) *be a sequence of functions from* E *to* $\overline{\mathbb{R}}$ *satisfying: there exist* $a > 0$, $b \in \mathbb{R}$ *and* $x_0 \in E$ *such that, for every* $n \geq 1$ *and* $x \in E$, $h_n(x) + ad(x, x_0) + b \geq 0$. *Then for all* $x \in E$,

$$\mathrm{li}_e\, h_n(x) = \sup_{k\geq 1} \liminf_{n\to\infty} h_n^k(x),$$

$$\mathrm{ls}_e\, h_n(x) = \sup_{k\geq 1} \limsup_{n\to\infty} h_n^k(x).$$

The lower (resp. upper) hypolimits, as well as the hypoconvergence of a sequence (h_n), can be obtained in a symmetric way. Indeed, the sequence (h_n) hypoconverges to h iff $(-h_n)$ epiconverges to $-h$. In order to present the relation between epiconvergence, hypoconvergence and uniform convergence the following characterizations of epi- and hypoconvergences are needed [see Attouch (1984b) or Dal Maso (1993)].

PROPOSITION 3.3. *A sequence* (h_n) *of functions from* E *to* $\overline{\mathbb{R}}$ *epiconverges to* h *at* $x \in E$ *iff*:

(i) *for each sequence* (x_n) *converging to* x, $h(x) \leq \liminf_{n\to\infty} h_n(x_n)$;
(ii) *there exists a sequence* (x_n) *converging to* x *such that* $h(x) \geq \limsup_{n\to\infty} h_n(x_n)$.

REMARK 3.1. Properties (i) and (ii) are equivalent to (i) and (ii′) where:

(ii′) There exists a sequence (x_n) converging to x such that $h(x) = \lim_{n\to\infty} h_n(x_n)$.

Replacing (h_n) with $(-h_n)$ and h with $-h$, we get similar characterizations of hypoconvergence. Consequently, a sequence (h_n) is both epi- and hypoconvergent to h if and only if the following property holds:

$$\forall x \in E,\ \forall (x_n) \to x, \qquad h(x) = \lim_{n\to\infty} h_n(x_n). \tag{3.1}$$

The following simple result shows the connection with uniform convergence.

PROPOSITION 3.4. *If* h *and* (h_n) *are real-valued and satisfy* (3.1), *then* h *is continuous and* (h_n) *converges uniformly to* h *on all compact sets*.

PROOF. First observe that Remark 3.1 shows that (h_n) is both epi- and hypoconvergent to h. Thus h is both lower and upper semicontinuous, hence continuous on E. Further, consider a compact subset K of E and suppose that (h_n) does not converge uniformly to h on K. It is therefore possible to find $\alpha > 0$ and a subsequence (h_m) of (h_n) satisfying

$$\|h_m - h\|_{u,K} = \sup_{y\in K} |h_m(y) - h(y)| \geq \alpha > 0 \qquad \text{for all } m \geq 1.$$

But for all $m \geq 1$, there exists $y_m \in K$ such that

$$|h_m(y_m) - h(y_m)| \geq \| h_m - h\|_{u,K} - \frac{1}{m}. \tag{3.2}$$

Moreover, by extracting a subsequence (denoted similarly) converging to some $y \in K$, we have

$$|h_m(y_m) - h(y)| \geq |h_m(y_m) - h(y_m)| - | h(y_m) - h(y)|. \tag{3.3}$$

From (3.2) and (3.3), we get

$$|h_m(y_m) - h(y)| \geq \|h_m - h\|_{u,K} - \frac{1}{m} - |h(y_m) - h(y)|.$$

Thus $\liminf_{m\to\infty} |h_m(y_m) - h(y)| \geq \alpha > 0$, which contradicts property (3.1). □

3.2. *Comparison between two integrands.* The following result gives a characterization of the inequality between two integrands, which was defined at the beginning of Section 2.

PROPOSITION 3.5. (i) *If f_1 and f_2 are two integrands satisfying condition $\mathcal{C}[(B_i), (m_i), i \geq 1]$ [As can be easily checked, assuming that the (B_i) and (m_i) are the same for f_1 and f_2 does not restrict the generality], then the following three statements are equivalent*:

(a) *$f_1 \leq f_2$ as defined above*;
(b) *for every $i \geq 1$ and every $v \in L^0(\Omega, \mathcal{A}, \mathbb{P}; B_i)$, one has*

$$f_1(\omega, v(\omega)) \leq f_2(\omega, v(\omega)) \qquad \text{for almost all } \omega \in \Omega;$$

(c) *for every $i \geq 1$, every $v \in L^0(\Omega, \mathcal{A}, \mathbb{P}; B_i)$ and every $A \in \mathcal{A}$, the following inequality holds*

$$\int_A f_1(\omega, v(\omega))\mathbb{P}(d\omega) \leq \int_A f_2(\omega, v(\omega))\mathbb{P}(d\omega).$$

(ii) *A similar property holds for the equality.*

PROOF. As to part (i), observe that implication (a) ⇒ (b) is clear since for all $i \geq 1$ and $\omega \in \Omega \setminus N$ we can take $x = v(\omega)$, where v satisfies the required properties. Implication (b) ⇒ (c) follows from the very definition of $\mathbb{P}$-almost sure inequality. To prove implication (c) ⇒ (a), let us prove *not* (a) ⇒ *not* (c). For every $i \geq 1$ define the set $A_i = \{\omega \in \Omega : f_1(\omega, x) > f_2(\omega, x) \text{ for some } x \in B_i\}$. It is $\widehat{\mathcal{A}}$-measurable, because it is the projection on Ω of the $\mathcal{A} \otimes \mathcal{B}(E)$-measurable set $G_i = \{(\omega, x) \in \Omega \times E : f_1(\omega, x) > f_2(\omega, x)\}$. $\widehat{\mathcal{A}}$ is the σ-field of universally measurable sets [see Definition III.21 in Castaing and Valadier (1977) or Hess (1996)]. Moreover, since (a) does not hold, condition $(\mathcal{C})$ implies the existence of an integer $k \geq 1$ such that $\mathbb{P}(A_k) > 0$. Let k be fixed and consider the multifunction Γ defined by

$$\Gamma(\omega) = \{x \in B_k : f_1(\omega, x) > f_2(\omega, x)\}.$$

We have $A_k = \{\omega \in \Omega : \Gamma(\omega) \neq \varnothing\}$. Further, the graph of the multifunction Γ, namely $\mathrm{Gr}(\Gamma) = \{(\omega, x) \in \Omega \times B_k : x \in \Gamma(\omega)\}$, is equal to G_k, which is $\mathcal{A} \otimes \mathcal{B}(E)$-measurable. Consequently [see Theorem III.22 in Castaing and Valadier (1977)], there exists an $\widehat{\mathcal{A}}$-measurable selection v_0 of Γ, that is, $v_0(\omega) \in \Gamma(\omega)$ for all $\omega \in A_k$. By construction, v_0 satisfies $f_1(\omega, v_0(\omega)) > f_2(\omega, v_0(\omega))$ for almost all $\omega \in A_k$. At this stage, observe that the integrals of $f_1(\cdot, v_0(\cdot))$ and $f_2(\cdot, v_0(\cdot))$ over A_k may both take on the value $+\infty$. However, it is not difficult to find two rationals r and s, and an $\mathcal{A}$-measurable set C of positive measure, contained in A_k and satisfying $f_1(\omega, v_0(\omega)) > r > s > f_2(\omega, v_0(\omega))$ for all $\omega \in C$. This implies

$$\int_C f_1(\omega, v_0(\omega))\mathbb{P}(d\omega) \geq r\mathbb{P}(C) > s\mathbb{P}(C) \geq \int_C f_2(\omega, v_0(\omega))\mathbb{P}(d\omega)$$

which yields *not* (c). Statement (ii) is an immediate consequence of (i). □

REMARK 3.2. (i) When $\mathcal{A}$ does not reduce to the trivial σ-field $\{\Omega, \varnothing\}$, it is not possible to replace the class of $\mathcal{A}$-measurable functions v involved in statement (b) and (c) by that of the constant functions $v(\cdot) = x$ from Ω to E. In other words, the following condition

$$f_1(\omega, x) \leq f_2(\omega, x) \qquad \text{a.s. } \forall x \in E, \tag{3.4}$$

which is implied by statement (b) of Proposition 3.5, does not imply it. Indeed, consider the special case where $\Omega = E = [0, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbb{P}$ is the Lebesgue measure and $f_2(\omega, x) = 0$ for every $(\omega, x) \in \Omega \times E$. Further, define the integrand f_1 by $f_1(\omega, x) = 1$ if $\omega = x$ and $f_1(\omega, x) = 0$ if $\omega \neq x$. Clearly, condition $(\mathcal{C})$ and condition (3.4) are satisfied, but if we define the $\mathcal{A}$-measurable function $v : \Omega \to E$ by $v(\omega) = \omega$, we have $1 = f_1(\omega, v(\omega)) > f_2(\omega, v(\omega)) = 0$ for all ω.

(ii) In particular, Proposition 3.5 holds when f_1 and f_2 are positive.

3.3. *Invariance of the epigraphical limit.* It is easy to prove that the lim inf and lim sup of the Cesaro means involved in the BET are invariant if the random variables are finite-valued [see, e.g., the proof of Theorem 13.10 in Davidson (1994)]. However, we need this result for extended real-valued random variables. Since it has not been possible to find this result in the literature for random variables which may take on the value $+\infty$ on a set of strictly positive measure, we provide a short proof based on Poincaré's recurrence theorem.

Let $v : \Omega \to [0, +\infty]$ be an $\mathcal{A}$-measurable function. For all $\omega \in \Omega$, we set

$$u_n(\omega) = \frac{1}{n} \sum_{i=0}^{n-1} v(T^i \omega).$$

PROPOSITION 3.6. *If* $v:\Omega\to[0,+\infty]$ *is* $\mathcal{A}$*-measurable*, *then* $\liminf_{n\to\infty} u_n$ *and* $\limsup_{n\to\infty} u_n$ *are* $\mathcal{I}$*-measurable*.

PROOF. Consider first the $\mathcal{A}$-measurable function $u=\limsup_{n\to\infty} u_n$. To prove that u is also $\mathcal{I}$-measurable, let us prove that it is invariant, that is, that $u(T\omega)\overset{\text{a.s.}}{=}u(\omega)$. For all $n\geq 1$ and for all $\omega\in\Omega$, we can write

$$u_{n+1}(\omega)=\frac{v(\omega)}{n+1}+\frac{n}{n+1}u_n(T\omega). \tag{3.5}$$

Now consider the $\mathcal{A}$-measurable subset $A=\{\omega\in\Omega: v(\omega)=+\infty\}$. If $\omega\in A^c$ then we get the result taking the lim sup in both sides of (3.5),

$$u(\omega)=\limsup_{n\to\infty} u_{n+1}(\omega)=\limsup_{n\to\infty} u_n(T\omega)=u(T\omega).$$

If $\mathbb{P}(A)=0$, the proof is complete. If $\mathbb{P}(A)>0$, we apply Poincaré's recurrence theorem [see, e.g., Petersen (1989), Theorem 3.2, page 34] to A: almost every point of A is recurrent with respect to A; that is, for such a point ω, there exists $k\geq 1$ such that $T^k\omega\in A$; that is, $v(T^k\omega)=+\infty$. Therefore, as soon as $n>k$, we have $u_{n+1}(\omega)=u_n(\omega)=+\infty$, which implies $u(\omega)=u(T\omega)=+\infty$.

A similar proof holds for the inferior limit. □

The following result that deals with the measurability of the epigraphical limits is necessary to derive the main result of our paper (see Remark 4.1).

PROPOSITION 3.7. *Let* f *be an integrand on* $\Omega\times E$ *with* E *metrizable and separable space*. *Then the epigraphical limits*

$$\operatorname*{li_e}_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega,\cdot)\quad\textit{and}\quad \operatorname*{ls_e}_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega,\cdot)$$

are $\mathcal{I}\otimes\mathcal{B}(E)$*-measurable*.

PROOF. We consider first an integrand of the form $f(\omega,x)=\mathbb{1}_A(\omega)\cdot\mathbb{1}_F(x)$ for $(\omega,x)\in\Omega\times E$, $A\in\mathcal{A}$ and $F\in\mathcal{B}(E)$. For every $n\geq 1$, let

$$g_n(\omega,x)=\left(\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_A(T^i\omega)\right)\cdot\mathbb{1}_F(x).$$

Applying the BET to $\mathbb{1}_A$ and using the definition of $\mathrm{li}_e\, g_n(\omega,x)$, it is readily checked that

$$\operatorname*{li_e}_{n\to\infty} g_n(\omega,x)=\mathbb{1}_{\mathring{F}}(x)\mathbb{E}^{\mathcal{I}}(\mathbb{1}_A)(\omega),$$

where $\mathring{F}$ denotes the interior of F. This shows the $\mathcal{I}\otimes\mathcal{B}(E)$-measurability of $\mathrm{li}_e\, g_n$.

Now, consider the case where f is a linear combination of indicators of pairwise disjoint rectangles $A_j \times F_j$, $(A_j \in \mathcal{A}, F_j \in \mathcal{B}(E))$ with real weights α_j, $j = 1, \dots, m$. One has

$$\operatorname*{li_e}_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega, x) = \operatorname*{li_e}_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}\left[\sum_{j=1}^{m} \alpha_j \cdot \mathbb{1}_{A_j}(T^i\omega)\cdot \mathbb{1}_{F_j}(x)\right]$$
$$= \sum_{j=1}^{m} \alpha_j \cdot \mathbb{1}_{\mathring{F}_j}(x)\mathbb{E}^{\mathcal{I}}(\mathbb{1}_{A_j})(\omega),$$

which proves the $\mathcal{I} \otimes \mathcal{B}(E)$-me asurability of $\operatorname{li_e}_{n\to\infty} g_n$. When f is positive, the proof is completed by approximating f by a nondecreasing sequence of integrands of the previous type. In the general case, it suffices to apply the result of the previous step on each $\Omega \times B_i$ to the integrand $(\omega, x) \mapsto f(\omega, x) - m_i(\omega)$. The same holds true for the integrand $\operatorname{ls_e}_{n\to\infty} g_n$; we only have to replace the lim inf over n by the lim sup. □

3.4. *Ergodic theorem in L^0.* We state a version of the BET for positive random variables in $L^0(\Omega, \mathcal{A}, \mathbb{P})$. The proof is available from the authors upon request since it was not possible to find it in the literature.

PROPOSITION 3.8. *For every extended real-valued positive function $v \in L^0(\Omega, \mathcal{A}, \mathbb{P})$, one has*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} v(T^i\omega) \overset{a.s.}{=} \mathbb{E}^{\mathcal{I}}(v)(\omega)$$

(*where both sides can be equal to* $+\infty$).

Clearly, this result also holds for quasi-integrable random variables in the sense of Neveu [(1964), page 40], that is, satisfying the hypothesis that either $\mathbb{E}\max\{v, 0\}$ or $\mathbb{E}\min\{v, 0\}$ is finite.

The following simple consequence of Proposition 3.8 will be needed twice in the proof of the main result.

LEMMA 3.9. *If f is an $\mathcal{A} \otimes \mathcal{B}(E)$-measurable normal integrand and if $v \in L^0(\Omega, \mathcal{I}, \mathbb{P})$ is such that $f(\cdot, v(\cdot))$ is quasi-integrable, then, for almost all $\omega \in \Omega$, the following equality holds*:

$$(\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i\omega, v(\omega)\big).$$

PROOF. Since v is $\mathcal{I}$-measurable, it is invariant and we can write

$$\frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i\omega, v(\omega)\big) = \frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i\omega, v(T^i\omega)\big) = \frac{1}{n}\sum_{i=0}^{n-1} h(T^i\omega),$$

where h is defined by $h(\omega) = f(\omega, v(\omega))$. Appealing to Proposition 3.8 we can take the limit and from the definition of the conditional expectation of an integrand, we have

$$(\mathbb{E}^{\mathcal{I}} h)(\omega) = \mathbb{E}^{\mathcal{I}}\big[f(\cdot, v(\cdot))\big](\omega) = (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)). \qquad \square$$

4. Proofs of the main results.

PROOF OF THEOREM 2.3. The statement concerning the $\mathcal{I}\otimes\mathcal{B}(E)$-measurability of $\mathbb{E}^{\mathcal{I}} f$ immediately follows from Proposition 3.7. To simplify our notations, we introduce

$$g_n(\omega, \cdot) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega, \cdot).$$

To prove that the Cesaro sum epiconverges to the conditional expectation, we have to prove that, for almost all $\omega \in \Omega$, the following inequalities hold:

$$\operatorname*{li_e}_{n\to\infty} g_n(\omega, \cdot) \geq (\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot),$$

$$\operatorname*{ls_e}_{n\to\infty} g_n(\omega, \cdot) \leq (\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot).$$

Proposition 3.7 entails that $\mathrm{li}_{e n\to\infty} g_n$ and $\mathrm{ls}_{e n\to\infty} g_n$ are $\mathcal{I}\otimes\mathcal{B}(E)$-measurable. Moreover, according to Proposition 3.5, it is enough to show that for all $v \in L^0(\Omega, \mathcal{I}, \mathbb{P}; E)$ the following inequalities hold, for almost all $\omega \in \Omega$:

$$\operatorname*{li_e}_{n\to\infty} g_n(\omega, v(\omega)) \geq (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)), \tag{4.1}$$

$$\operatorname*{ls_e}_{n\to\infty} g_n(\omega, v(\omega)) \leq (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)). \tag{4.2}$$

To prove inequalities (4.1) and (4.2), we shall proceed in three steps.

Step 1. The integrand f is assumed to be positive. Let us prove inequality (4.1). Let $v \in L^0(\Omega, \mathcal{I}, \mathbb{P}; E)$ be fixed. For every $k \geq 1$ and every $\omega \in \Omega$, we have

$$\begin{aligned} g_n^k(\omega, v(\omega)) &= \inf_{y\in E}\{g_n(\omega, y) + kd(y, v(\omega))\} \\ &\geq \frac{1}{n}\sum_{i=0}^{n-1}\Big\{\inf_{y\in E}\big[f(T^i\omega, y) + kd(y, v(\omega))\big]\Big\}, \end{aligned}$$

whence

$$g_n^k(\omega, v(\omega)) \geq \frac{1}{n}\sum_{i=0}^{n-1} f^k\big(T^i\omega, v(\omega)\big). \tag{4.3}$$

Further, applying Lemma 3.9 to the positive normal integrand f^k, we get

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f^k\big(T^i\omega, v(\omega)\big) = (\mathbb{E}^{\mathcal{I}} f^k)(\omega, v(\omega)) \qquad \text{a.s.}$$

More precisely, the above equalities hold for any $\omega \in \Omega \backslash N^0(k)$ where $N^0(k)$ is a suitable negligible set (depending also on v). If we take the lim inf over n in both sides of (4.3), we get

$$\liminf_{n\to\infty} g_n^k(\omega, v(\omega)) \geq (\mathbb{E}^{\mathcal{I}} f^k)(\omega, v(\omega))$$

for every $k \geq 1$ and every ω outside some negligible subset N^0 containing $\bigcup_{k\geq 1} N^0(k)$. Now, taking the supremum over k and appealing to the monotone convergence theorem for the conditional expectation yield the existence of a negligible set N^1 containing N^0 such that, for every $\omega \in \Omega \setminus N^1$,

$$\mathop{\mathrm{li}_e}_{n\to\infty} g_n(\omega, v(\omega)) \geq \sup_{k\geq 1} \mathbb{E}^{\mathcal{I}}\big(f^k(\cdot, v(\cdot))\big)(\omega) = (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)).$$

So inequality (4.1) has been proved. To prove inequality (4.2), consider as before a fixed member v of $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}; E)$ and the $\mathcal{I}$-measurable random variable ϕ defined by

$$\phi(\omega) = (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)).$$

Clearly, it suffices to consider only the case where $\phi(\omega)$ is finite. For every $k \geq 1$, we also define ϕ_k by

$$\phi_k(\omega) = \inf_{y\in E} \big\{(\mathbb{E}^{\mathcal{I}} f)(\omega, y) + kd(y, v(\omega))\big\}.$$

For any $k, p \geq 1$, consider the multifunction $\Gamma_{k,p}$ such that

$$\Gamma_{k,p}(\omega) = \left\{ y \in E : (\mathbb{E}^{\mathcal{I}} f)(\omega, y) + kd(y, v(\omega)) \leq \phi_k(\omega) + \frac{1}{p} \right\}.$$

It is readily seen that $\Gamma_{k,p}$ is nonempty valued and that its graph is $\mathcal{I} \otimes \mathcal{B}(E)$-measurable. Therefore, there exists an $\widehat{\mathcal{I}}$-measurable selection $v_{k,p}$ of $\Gamma_{k,p}$, that can be modified on a negligible set $N^2(k, p)$ so as to be $\mathcal{I}$-measurable. Hence, for any $k \geq 1$, ϕ_k is also given by

$$\phi_k(\omega) = \inf_{p\geq 1} \big\{(\mathbb{E}^{\mathcal{I}} f)\big(\omega, v_{k,p}(\omega)\big) + kd\big(v_{k,p}(\omega), v(\omega)\big)\big\} \tag{4.4}$$

for each ω in the negligible set $N^2(k) = \bigcup_{p\geq 1} N^2(k, p)$. On the other hand, if we denote by $g_n^k(\omega, \cdot)$ the Lipschitz approximation of order k of $g_n(\omega, \cdot)$, we can write

$$g_n^k(\omega, v(\omega)) = \inf_{y\in E} \{kd(y, v(\omega)) + g_n(\omega, y)\},$$

whence, by taking the lim sup,

$$\begin{aligned}\limsup_{n\to\infty} g_n^k(\omega, v(\omega)) &= \limsup_{n\to\infty} \inf_{y\in E}\{kd(y, v(\omega)) + g_n(\omega, y)\}\\ &\le \inf_{y\in E} \limsup_{n\to\infty}\{kd(y, v(\omega)) + g_n(\omega, y)\}\\ &\le \inf_{p\ge 1} \limsup_{n\to\infty}\{kd(v_{k,p}(\omega), v(\omega)) + g_n(\omega, v_{k,p}(\omega))\}\\ &= \inf_{p\ge 1}\left\{kd(v_{k,p}(\omega), v(\omega)) + \limsup_{n\to\infty} g_n(\omega, v_{k,p}(\omega))\right\}.\end{aligned}$$

By Lemma 3.9, for every $k, p \ge 1$, we can write

$$\lim_{n\to\infty} g_n(\omega, v_{k,p}(\omega)) = (\mathbb{E}^{\mathcal{I}} f)(\omega, v_{k,p}(\omega))$$

for all $\omega \in \Omega \setminus N^3(k, p)$, where $N^3(k, p)$ is a suitable negligible set. Now, for every $k \ge 1$, define the negligible subset $N^3(k)$ by $\bigcup_{p\ge 1} N^3(k, p)$. Without loss of generality we can assume that $N^3(k)$ contains $N^2(k)$. By (4.4) we have

$$\limsup_{n\to\infty} g_n^k(\omega, v(\omega)) \le \inf_{p\ge 1}\{kd(v_{k,p}(\omega), v(\omega)) + (\mathbb{E}^{\mathcal{I}} f)(\omega, v_{k,p}(\omega))\} = \phi_k(\omega).$$

Consequently, taking the supremum over k on both sides of the above inequality gives

$$\begin{aligned}\operatorname*{ls_e}_{n\to\infty} g_n(\omega, v(\omega)) &= \sup_{k\ge 1} \limsup_{n\to\infty} g_n^k(\omega, v(\omega))\\ &\le \sup_{k\ge 1} \phi_k(\omega) = \phi(\omega) = (\mathbb{E}^{\mathcal{I}} f)(\omega, v(\omega)).\end{aligned}$$

The inequality holds for every $\omega \in N^3$, where the negligible subset N^3 is defined by $N^3 = \bigcup_{k\ge 1} N^3(k)$. The proof is over by noting that equality (2.2) holds for every $\omega \in \Omega \setminus (N^1 \cup N^3)$.

Step 2. We assume the existence of a real-valued integrable function m such that $f(\omega, x) \ge m(\omega)$ for all $(\omega, x) \in \Omega \times E$. In this case, we consider the positive integrand g defined on $\Omega \times E$ by

$$g(\omega, x) = f(\omega, x) - m(\omega).$$

We can apply the result of the first step to g and Proposition 3.8 to m. Further, it is readily checked that the sum of an epiconvergent sequence of functions and of a convergent sequence of real numbers is epiconvergent to the sum of the limits. This entails equality (2.2).

Step 3. Now we pass to the general case. The desired result is a consequence of the local character of epiconvergence [see Remark 4.3 in Dal Maso (1993)]. Indeed, given a sequence of functions $f_n : E \to \overline{\mathbb{R}}$, consider its lower epilimit at $x \in E$, which is defined by

$$\operatorname*{li_e}_{n\to\infty} f_n(x) = \sup_{k\ge 1} \liminf_{n\to\infty} \inf_{y\in \mathsf{B}(x,1/k)} f_n(y).$$

Due to the monotonicity of the infimum over k, one also has

$$\operatorname{li}_{e \atop n\to\infty} f_n(x) = \sup_{k\geq 1/r} \liminf_{n\to\infty} \inf_{y\in \mathsf{B}(x,1/k)} f_n(y).$$

A similar equality holds true for the upper epilimit. Due to condition $(\mathcal{C}_0)$, E can be covered by a countable collection of open subsets B_i. Thus, applying the result of the second step for each $i \geq 1$ with E replaced with B_i and extracting appropriate negligible subsets yield the desired conclusion. □

REMARK 4.1. (i) According to Proposition 3.5 and to Remark 3.2, the $\mathcal{I} \otimes \mathcal{B}(E)$-measurability of $\operatorname{li}_{e n\to\infty} g_n$ and $\operatorname{ls}_{e n\to\infty} g_n$ is needed in the proof of Theorem 2.3. Indeed, if these two functions were only known to be $\mathcal{A} \otimes \mathcal{B}(E)$-measurable, one could only deduce the inequalities

$$\mathbb{E}^{\mathcal{I}}\Big(\operatorname{li}_{e \atop n\to\infty} g_n\Big)(\omega,\cdot) \geq (\mathbb{E}^{\mathcal{I}} f)(\omega,\cdot) \geq \mathbb{E}^{\mathcal{I}}\Big(\operatorname{ls}_{e \atop n\to\infty} g_n\Big)(\omega,\cdot) \qquad \text{a.s.}$$

(ii) Apart from the use of the conditional expectation for integrands, the above proof follows the same lines as in Hess (1991, 1996), where a strong law of large numbers was proved for sequences of pairwise independent identically distributed normal integrands.

PROOF OF THEOREM 2.5. Condition $(\mathcal{C}_0')$ allows for applying Corollary 2.2(ii) to f and to $-f$, which yields the continuity of $(\mathbb{E}^{\mathcal{I}} f)(\omega,\cdot)$, for almost all $\omega \in \Omega$. Further, from Theorem 2.3 applied to f and $-f$, it follows that for almost every $\omega \in \Omega$ the sequence

$$\left(\frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega,\cdot)\right)_{n\geq 1}$$

is both epiconvergent and hypoconvergent to $(\mathbb{E}^{\mathcal{I}} f)(\omega,\cdot)$. The proof is complete by appealing to Proposition 3.4. □

PROOF OF THEOREM 2.6. Given an integrable set-valued map F, we consider the integrand $f : \Omega \times E \to \overline{\mathbb{R}}$ defined by

$$f(\omega, y) = s(y, F(\omega)), \qquad (\omega, y) \in \Omega \times E. \tag{4.5}$$

If we denote by g the conditional expectation of f with respect to $\mathcal{B}$ (as defined in Section 2.1), we have for any $\mathcal{B}$-measurable function $v : \Omega \to E$ and for almost any $\omega \in \Omega$,

$$g(\omega, v(\omega)) = \mathbb{E}^{\mathcal{B}}\big[s\big(v(\cdot), F(\cdot)\big)\big](\omega) = s\big(v(\omega), \mathbb{E}^{\mathcal{B}}(F)(\omega)\big). \tag{4.6}$$

The last equality is a consequence of the definition of the set-valued conditional expectation [see Hiaï (1985)].

Let w be an integrable selection of F. For every $\omega \in \Omega$ and y in the closed unit ball of E, one has

$$-\|w(\omega)\| \leq \langle y, w(\omega)\rangle \leq s(y, F(\omega))$$

so that the minorization condition in Theorem 2.3 is satisfied. Further, consider the integrand defined by (4.5). From Theorem 2.3 and equalities (4.6), we can deduce that

$$(\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot) = \lim_{n\to\infty}{}_e \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega, \cdot) \qquad \text{a.s.}$$

which yields

$$s\big(\cdot, \mathbb{E}^{\mathcal{I}}(F)(\omega)\big) = \lim_{n\to\infty}{}_e \frac{1}{n} \sum_{i=0}^{n-1} s\big(\cdot, F(T^i \omega)\big) \qquad \text{a.s.}$$

Taking the conjugate (also called the Young–Fenchel transform) of both sides and using the continuity of this operation with respect to epiconvergence [see Attouch (1984b) or Dal Maso (1993)], we obtain

$$\chi\big(\cdot, \mathbb{E}^{\mathcal{I}}(F)(\omega)\big) = \lim_{n\to\infty}{}_e \frac{1}{n} \sum_{i=0}^{n-1} \chi\big(\cdot, F(T^i \omega)\big) \qquad \text{a.s.}$$

It is not difficult to check that this yields the Painlevé–Kuratowski convergence as claimed. □

5. Comparison with the literature. The above stated theorems are related to other results in statistics, econometrics and convex analysis. In particular, the ergodic theorem (Theorem 2.3) that we state appears as a new result in an active field of research.

Indeed, results such as those expressed in (1.1) first appeared in a paper by King and Wets (1991) who proved a version of the SLLN for an i.i.d. sequence of normal convex integrands; that is, the function f appearing in (1.1) was assumed to be lower semicontinuous and convex with respect to the second variable. In this work, E was assumed to be a reflexive Banach space, and the proof was based on a set-valued version of the SLLN for closed convex random sets and on the use of the continuity of the Young–Fenchel transform. Later, a new approach was initiated independently by Attouch and Wets [(1991), Theorem 3.3] and by Hess [(1991), Theorem 4.3] with different proofs. The important difference with King and Wets's results is that the consideration of the Young–Fenchel transform was no longer necessary. In Attouch and Wets (1991), a version of the SLLN was proved assuming that E is a separable Banach space and that f is bounded from below by a square integrable random variable. In Hess (1991), the SLLN was shown to hold when E is a Suslin metric space without linear structure and assuming only the minorization of f by an integrable random variable. Moreover, the integrands were only assumed to be pairwise independent. The latter result was presented in several conferences (e.g., at the XXIV Journées de Statistique, Bruxelles 18–22 May, 1992) and appeared eventually in 1996 [see Hess (1996)]. A partial extension of this kind of SLLN in the framework of ordered vector spaces was also proved by Jalby (1993). Afterwards, several authors became interested in extensions or variants of this version of the SLLN. For example, the BET was examined by Castaing and Ezzaki (1993), who obtained partial results, and by Licht and Michaille (1994) who proved a version of the BET in the ergodic case, but in the more general case of subadditive processes. More recently, Artstein and Wets (1996) gave a proof of the SLLN very similar to that of Hess (1991) and provided applications to stochastic optimization.

Following the same lines, Korf and Wets [(2000a), Theorem 7.2 and (2001), Theorem 6.2] state an epigraphical ergodic theorem under quite restrictive conditions using the method of scalarization of Korf and Wets (2000b); indeed, they take E to be a Polish space and they suppose that the random set $\omega \mapsto \mathrm{Epi}[(\mathbb{E}^{\mathcal{I}} f)(\omega, \cdot)]$ has a dense countable subset. Valadier (1999, 2000a, b) has taken further the work of Castaing and Ezzaki (1993), and has given a result similar to Theorem 2.3 for a positive integrand. However, Valadier's proof seems to work only in the ergodic case. This is due to repeated appeals to the monotone convergence theorem for the conditional expectation, which yields a noncountable family of negligible subsets [this argument was contained in Castaing and Ezzaki (1993)]. This has been corrected in Valadier (2002).

Moreover, as shown in Section 2.5, under the hypothesis that the transformation T is ergodic, the ergodic theorem can be turned into an epigraphical SLLN that can be used, in statistical and econometric applications, to derive consistency of estimators or of stochastic programming problems. The approach based on epi-convergence has been pursued by Dupačová and Wets (1988), Geyer (1994) and Hess (1991, 1996). In this sense, epigraphical SLLN should be perceived as a substitute of the uniform laws of large numbers (ULLN) that, starting from the seminal papers of Huber [(1967), pages 224–226] and Jennrich [(1969), Theorem 2, page 636], are considered as a cornerstone of the modern theory of statistical inference. In the standard case considered in the statistical and econometric literature, $(\Omega, \mathcal{A})$ is a complete measurable space, and x is a parameter varying in E, a compact subset of a Euclidean space. However, our result holds for much more general spaces and can be easily extended to take into account nonparametric and semiparametric estimation.

The real constraint is given by the $\mathcal{A} \otimes \mathcal{B}(E)$-measurability of the integrand; we remark that if, for any $x \in E$, $f(\cdot, x)$ is $\mathcal{A}$-measurable and, for any $\omega \in \Omega$, $f(\omega, \cdot)$ is continuous in x, then by Lemma III.14 of Castaing and Valadier (1977), f is $\mathcal{A} \otimes \mathcal{B}(E)$-measurable. This joint measurability assumption is not uncommon in asymptotic theory [see, e.g., Gouriéroux and Monfort (1995), Chapter 24, or Hess (1996)], but it is often replaced by the following weaker condition: For any

$x \in E$, $f(\cdot, x)$ is $\mathcal{A}$-measurable and for any $x \in E$, the set $N(x)$ of continuous trajectories at x has $\mathbb{P}$-measure 1. Indeed, our more restrictive hypothesis implies measurability of the supremum [see Brown and Purves (1973)], while the latter guarantees only outer measurability.

The Jennrich [(1969), Theorem 2, page 636] ULLN follows as a corollary of our Theorem 2.5 under the hypothesis that the random variables are independent and identically distributed. [See also Amemiya (1985), Theorem 4.2.1, page 116 and Theorem 4.2.3, page 118.] Gallant and Holly's [(1980), Theorem 2, page 702] ULLN use the concept of Cesaro summable sequence: under the requirement of stationarity, it corresponds to an ergodic sequence. Tauchen's [(1985), Lemma 1, page 423] ULLN is a special case of ours, too.

Clearly, our Theorem 2.5 is derived under the restrictive hypothesis of ergodicity. In this sense our law of large numbers complements Andrews (1987) and Pötscher and Prucha (1989, 1996), since they relax the conditions of ergodicity but assume stronger hypotheses on the behavior of the functions. In a forthcoming paper we shall investigate further epigraphical SLLN for more general stochastic processes.

APPENDIX

Conditional expectation of an integrand.

PROOF OF THEOREM 2.1. Uniqueness is an immediate consequence of Proposition 3.5. As to the proof of existence, we shall proceed in five steps.

Step 1. Consider $G = A \times F$, where $A \in \mathcal{A}$ and $F \in \mathcal{B}(E)$, and assume that the integrand f is defined as $f(\omega, x) = \mathbb{1}_G(\omega, x)$. The conditional expectation $g = \mathbb{E}^{\mathcal{B}}(f)$ is well defined. Indeed, for all $B \in \mathcal{B}$ and for all $(\mathcal{B}, \mathcal{B}(E))$-measurable function $v : \Omega \to E$, we have

$$\begin{aligned}\int_B \mathbb{1}_A(\omega) \cdot \mathbb{1}_F(v(\omega))\mathbb{P}(d\omega) &= \int_B \mathbb{1}_A(\omega) \cdot \mathbb{1}_{v^{-1}(F)}(\omega)\mathbb{P}(d\omega)\\ &= \int_B \mathbb{E}^{\mathcal{B}}(\mathbb{1}_A \cdot \mathbb{1}_{v^{-1}(F)})(\omega)\mathbb{P}(d\omega)\\ &= \int_B \mathbb{1}_{v^{-1}(F)}(\omega)\mathbb{E}^{\mathcal{B}}(\mathbb{1}_A)(\omega)\mathbb{P}(d\omega).\end{aligned}$$

The last equality follows from the $\mathcal{B}$-measurability of v. This shows that g is defined by $g(\omega, x) = \mathbb{1}_F(x)\mathbb{E}^{\mathcal{B}}(\mathbb{1}_A)(\omega)$.

Step 2. To prove that $g = \mathbb{E}^{\mathcal{B}}(f)$ can be defined for every $f = \mathbb{1}_G$, where G is an arbitrary member of $\mathcal{A} \otimes \mathcal{B}(E)$, it suffices to use a monotone class argument and the monotone convergence theorem for the conditional expectation [this method is similar to the construction of a product measure; see, e.g., Neveu (1964), Proposition III-2-1].

Step 3. When f is a positive integrand, it can be written as the supremum of a sequence of positive *simple integrands*, that is, integrands of the form $\sum_{i=1}^{n} \alpha_i \cdot \mathbb{1}_{G_i}$, where $\alpha_i \in \mathbb{R}_+$, $G_i \in \mathcal{G}$, $n \in \mathbb{N}^*$. The additivity of the integral and an appeal to the monotone convergence theorem for the conditional expectation yields the existence of $g = \mathbb{E}^{\mathcal{B}} f$, which is positive from Proposition 3.5.

Step 4. When f is minorized on E by an integrable function $m : \Omega \to \mathbb{R}$, it is sufficient to apply the result of the third step to $(\omega, x) \mapsto f(\omega, x) - m(\omega)$.

Step 5. We prove the result in the general case, that is, when f satisfies condition $\mathcal{C}[(B_i), (m_i),\ i \geq 1]$. According to the result of the fourth step, for each $i \geq 1$, it is possible to define a unique $\mathcal{B} \otimes \mathcal{B}(E)$-measurable integrand g_i on $\Omega \times B_i$ satisfying as $\mathbb{E}^{\mathcal{B}} f(\cdot, v(\cdot)) = g_i(\cdot, v(\cdot))$ for every $v \in L^0(\Omega, \mathcal{A}, \mathbb{P}; B_i)$. Now, let us define the integrand g on $\Omega \times E$ by $g(\omega, x) = g_i(\omega, x)$ for all $(\omega, x) \in \Omega \times B_i$. It remains to show that this makes sense. More precisely, consider two distinct integers $i, j \geq 1$ such that $B_i \cap B_j \neq \varnothing$ and a $\mathcal{B}$-measurable function $v : \Omega \to B_i \cap B_j$. In view of condition $(\mathcal{C})$ one has

$$f(\omega, v(\omega)) \geq \max\bigl(m_i(\omega), m_j(\omega)\bigr),$$

whence, for every $B \in \mathcal{B}$,

$$\int_B f(\omega, v(\omega))\mathbb{P}(d\omega) = \int_B g_i(\omega, v(\omega))\mathbb{P}(d\omega) = \int_B g_j(\omega, v(\omega))\mathbb{P}(d\omega).$$

Proposition 3.5 implies $g_i = g_j$ on $\Omega \times (B_i \cap B_j)$; namely, there exists a negligible subset N such that

$$g_i(\omega, x) = g_j(\omega, x) \qquad \forall (\omega, x) \in (\Omega \setminus N) \times (B_i \cap B_j). \qquad \square$$

PROOF OF COROLLARY 2.2. As to part (i), consider an integrand satisfying condition $(\mathcal{C})$ and which is k-Lipschitz on each B_i. We have already proved that the conditional expectation of f exists and is characterized by

$$g(\omega, v(\omega)) = \mathbb{E}^{\mathcal{B}}\bigl[f(\cdot, v(\cdot))\bigr](\omega) \qquad \text{a.s.}$$

for every $i \geq 1$ and every $v \in L^0(\Omega, \mathcal{B}, \mathbb{P}; B_i)$. Let $i \geq 1$ be fixed in the rest of the proof of the present statement. Since E is separable, there exists a set D that is countable and dense in B_i. For all $x \in B_i$, there exists a negligible set N_x such that: $g(\omega, x) = \mathbb{E}^{\mathcal{B}}(f(\cdot, x))(\omega)$ for all $\omega \notin N_x$. Now, define $N = \bigcup_{x \in D} N_x$. For all $x, y \in D$ and for all $\omega \notin N$, we have

$$\begin{aligned} |g(\omega, x) - g(\omega, y)| &\leq \mathbb{E}^{\mathcal{B}}\bigl(|f(\cdot, x) - f(\cdot, y)|\bigr)(\omega) \\ &\leq \mathbb{E}^{\mathcal{B}}(kd(x, y)) = kd(x, y). \end{aligned}$$

From the Lipschitz version of the extension theorem [see, e.g., Aliprantis and Border (1999), Lemma 3.8], there is a unique function $\widetilde{g} : (\Omega \setminus N) \times B_i \to \overline{\mathbb{R}}$

satisfying

$$\begin{aligned}|\widetilde{g}(\omega,x)-\widetilde{g}(\omega,y)|&\le kd(x,y) \qquad \forall\,(x,y)\in B_i^2,\\ \widetilde{g}(\omega,\cdot)|_D&=g(\omega,\cdot),\\ \widetilde{g}(\omega,x)&\ge \mathbb{E}^{\mathcal{B}}m_i \qquad \forall\,(\omega,x)\in(\Omega\setminus N)\times B_i.\end{aligned}$$

It remains to show that, for every $x\in B_i$, $\widetilde{g}(\cdot,x)$ satisfies

$$\widetilde{g}(\omega,x)=\mathbb{E}^{\mathcal{B}}(f(\cdot,x))(\omega) \qquad \text{a.s.}$$

For every $x\in B_i\backslash D$, there exists a sequence $(x_n)_n$ in D converging to x. Further, for every $B\in\mathcal{B}$ and every $n\ge 1$, we have

$$\int_B f(\omega,x_n)\mathbb{P}(d\omega)=\int_B g(\omega,x_n)\mathbb{P}(d\omega)=\int_B \widetilde{g}(\omega,x_n)\mathbb{P}(d\omega).$$

The following two inequalities are readily obtained:

$$\left|\int_B f(\omega,x_n)\mathbb{P}(d\omega)-\int_B f(\omega,x)\mathbb{P}(d\omega)\right|\le k\mathbb{P}(B)d(x,x_n),$$

$$\left|\int_B \widetilde{g}(\omega,x_n)\mathbb{P}(d\omega)-\int_B \widetilde{g}(\omega,x)\mathbb{P}(d\omega)\right|\le k\mathbb{P}(B)d(x,x_n).$$

Letting n tend to infinity, we immediately deduce

$$\int_B f(\omega,x)\mathbb{P}(d\omega)=\int_B \widetilde{g}(\omega,x)\mathbb{P}(d\omega),$$

which yields the desired conclusion.

As to part (ii), suppose now that f is a normal integrand on each B_i (it is lsc with respect to x). Assume that the integer i is fixed. From Proposition 3.1, we know that on each B_i f can be written as the supremum of the Lipschitz approximations $f^k(\omega,\cdot)$ defined as

$$\forall\,k\ge 1,\ \forall\,x\in B_i,\ \forall\,\omega\in\Omega, \qquad f^k(\omega,x)\triangleq\inf_{y\in B_i}\{f(\omega,y)+kd(x,y)\};$$

moreover, f^k is $\mathcal{A}\otimes\mathcal{B}(E)$-measurable [see Hess (1996)], and $\mathbb{E}^{\mathcal{B}}f^k(\omega,\cdot)$ is also Lipschitz of constant k on each B_i. For each $i\ge 1$, we can apply the monotone convergence theorem for conditional expectation in L^0 [see, e.g., Theorem 10.5 in Davidson (1994)] on each B_i. Thus, for almost all $\omega\in\Omega$, the restriction of $\mathbb{E}^{\mathcal{B}}f(\omega,\cdot)$ to each B_i is lsc. Since the B_i are assumed to be open, $\mathbb{E}^{\mathcal{B}}f(\omega,\cdot)$ is lsc on each B_i, which yields the desired result. □

Acknowledgments. We thank G. Dal Maso, H. Doss, S. Johansen, R. Lucchetti and M. Valadier for valuable discussions. The Editor and a referee helped us considerably to improve a previous version of this paper.

REFERENCES

ALIPRANTIS, C. D. and BORDER, K. C. (1999). *Infinite Dimensional Analysis*, 2nd ed. Springer, Berlin.

AMEMIYA, T. (1985). *Advanced Econometric Methods*. Harvard Univ. Press.

ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica* **55** 1465–1471.

ARTSTEIN, Z. and HART, S. (1981). Law of large numbers for random sets and allocation processes. *Math. Oper. Res.* **6** 485–492.

ARTSTEIN, Z. and VITALE, R. A. (1975). A strong law of large numbers for random compact sets. *Ann. Probab.* **3** 879–882.

ARTSTEIN, Z. and WETS, R. J.-B. (1996). Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2** 1–17.

ATTOUCH, H. (1984a). Variational properties of epi-convergence. Applications to limit analysis problems in mechanics and duality theory. *Multifunctions and Integrands*. *Lecture Notes in Math.* **1091** 80–104. Springer, Berlin.

ATTOUCH, H. (1984b). *Variational Convergence for Functions and Operators*. Pitman, New York.

ATTOUCH, H. and WETS, R. J.-B. (1991). Epigraphical processes: Laws of large numbers for random lsc functions. Semin. Anal. Convexe Univ. Sci. Tech. Languedoc, Exp. 13.

AUBIN, J.-P. (1998). *Optima and Equilibria*: *An Introduction to Nonlinear Analysis*. Springer, Berlin.

AUMANN, R. J. (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* **12** 1–12.

BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York.

BIRGE, J. R. and LOUVEAUX, F. (1997). *Introduction to Stochastic Programming*. Springer, Berlin.

BISMUT, J.-M. (1973). Intégrales convexes et probabilités. *J. Math. Anal. Appl.* **42** 639–673.

BREIMAN, L. (1992). *Probability*, 2nd ed. SIAM, Philadelphia.

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

BURKE, G. (1965). A uniform ergodic theorem. *Ann. Math. Statist.* **36** 1853–1858.

CASTAING, C. and EZZAKI, F. (1993). Variational inequalities for integrand martingales and additive random sequences. *Acta Math. Vietnam* **18** 137–171.

CASTAING, C. and VALADIER, M. (1977). *Convex Analysis and Measurable Multifunctions*. *Lecture Notes in Math.* **580**. Springer, Berlin.

DAL MASO, G. (1993). *An Introduction to* Γ*-convergence*. Birkhäuser, Boston.

DAL MASO, G. and MODICA, L. (1986a). Nonlinear stochastic homogenization. *Ann. Mat. Pura Appl. Ser. IV* **144** 347–389.

DAL MASO, G. and MODICA, L. (1986b). Nonlinear stochastic homogenization and ergodic theory. *J. Reine Angew. Math.* **368** 28–42.

DAVIDSON, J. (1994). *Stochastic Limit Theory*. Oxford Univ. Press.

DUPAČOVÁ J. and WETS, R. J.-B. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.

DYNKIN, E. B. and EVSTIGNEEV, I. V. (1976). Regular conditional expectations of correspondences. *Theory Probab. Appl.* **21** 334–347.

EVSTIGNEEV, I. V. (1986). Regular conditional expectations of random variables depending on parameters. *Theory Probab. Appl.* **31** 515–518.

GALLANT, A. R. and HOLLY, A. (1980). Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation. *Econometrica* **48** 697–720.

GEYER, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. Roy. Statist. Soc. Ser. B* **56** 261–274.

GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes*. W. B. Saunders, Philadelphia (reprinted by Dover).

GOURIÉROUX, C. and MONFORT, A. (1995). *Statistics and Econometric Models*. Cambridge Univ. Press.

GOURIÉROUX, C., MONFORT, A. and TROGNON, A. (1985). A general approach to serial correlation. *Econom. Theory* **1** 315–340.

GRAY, R. M. and KIEFFER, J. C. (1980). Asymptotically mean stationary measures. *Ann. Probab.* **8** 962–973.

HESS, C. (1979). Théorème ergodique et loi forte des grands nombres pour des ensembles aléatoires. *C. R. Acad. Sci. Paris* **288** 519–522.

HESS, C. (1984). Quelques théorèmes limites pour des ensembles aléatoires bornés ou non. Semin. Anal. Convexe. Univ. Sci. Tech. Languedoc, Exp. 12.

HESS, C. (1991). Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. Cahiers de Mathématiques de la Décision 9121. Univ. Paris Dauphine.

HESS, C. (1992). Quelques versions multivoques du théorème de derivation de Lebesgue. Semin. Anal. Convexe. Univ. Sci. Tech. Languedoc, Exp. 18.

HESS, C. (1996). Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. *Ann. Statist.* **24** 1298–1315.

HIAÏ, F. (1985). Convergence of conditional expectations and strong laws of large numbers for multivalued random variables. *Trans. Amer. Math. Soc.* **291** 613–627.

HIAÏ, F. and UMEGAKI, H. (1977). Integrals, conditional expectations and martingales of multivalued functions. *J. Multivariate Anal.* **7** 149–182.

HIRIART-URRUTY, J.-R. (1977). Contribution à la programmation mathématique: cas déterministe et stochastique. Ph.D. thesis, Univ. de Clermont-Ferrand II.

HUBER, P. J. (1967). The behaviour of the maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* 221–233. California Univ. Press, Berkeley.

JALBY, V. (1993). Semi-continuité, convergence et approximation des applications vectorielles. Preprint, Laboratoire d'Analyse Convexe, Univ. Montpellier II.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

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

KORF, L. A. and WETS, R. J.-B. (2000a). An ergodic theorem for stochastic programming problems. *Optimization. Lecture Notes in Math. Econom.* **481** 203–217. Springer, Berlin.

KORF, L. A. and WETS, R. J.-B. (2000b). Random lsc functions: Scalarization. *Stochastic Programming E-Print Ser.* 2000-7.

KORF, L. A. and WETS, R. J.-B. (2001). Random lsc functions: An ergodic theorem. *Math. Oper. Res.* **26** 421–445.

KRENGEL, U. (1985). *Ergodic Theorems*. de Gruyter, Berlin.

KRUPA, G. (1998). Limit theorems for random sets. Ph.D. thesis, Univ. Utrecht.

LEVINE, D. (1983). A remark on serial correlation in maximum likelihood. *J. Econometrics* **23** 337–342.

LICHT, C. and MICHAILLE, G. (1994). Convergence results for subadditive processes. Preprint, Laboratoire d'Analyse Convexe, 28, Univ. Montpellier II.

NEVEU, J. (1964). *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris.

PETERSEN, K. E. (1989). *Ergodic Theory*. Cambridge Univ. Press.

PÖTSCHER, B.M. and PRUCHA, I. R. (1989). A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica* **57** 675–683.

PÖTSCHER, B. M. and PRUCHA, I. R. (1996). *Dynamic Nonlinear Econometric Models*. Springer, Berlin.

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.

ROCKAFELLAR, R. T. and WETS, R. J.-B. (1984). Variational systems. *Multifunctions and Integrands. Lecture Notes in Math.* **1091** 1–54. Springer, Berlin.

SCHÜRGER, K. (1983). Ergodic theorems for subadditive superstationary families of convex compact random sets. *Z. Wahrsch. Verw. Gebiete* **62** 125–135.

TAUCHEN, G. E. (1985). Diagnostic testing and evaluation of maximum likelihood models. *J. Econometrics* **30** 415–443.

THIBAULT, L. (1981). Espérances conditionnelles d'intégrandes semicontinus. *Ann. Inst. H. Poincaré B* **17** 337–350.

TRUFFERT, A. (1991). Conditional expectation of integrands and random sets. *Ann. Oper. Res.* **30** 117–156.

VALADIER, M. (1999). Théorème ergodique pour un intégrande positif. Working Paper, Univ. Montpellier.

VALADIER, M. (2000a). Survey on some questions about stationary and ergodic processes. Working Paper, Univ. Montpellier.

VALADIER, M. (2000b). Conditional expectation and ergodic theorem for a positive integrand. *J. Nonlinear Convex Anal.* **1** 233–244.

VALADIER, M. (2002). Corrigendum to: Conditional expectation and ergodic theorem for a positive integrand. *J. Nonlinear Convex Anal.* **3** 123.

WANG, R. and WANG, Z. (1997). Set-valued stationary processes. *J. Multivariate Anal.* **63** 180–198.

C. CHOIRAT
C. HESS
CENTRE DE RECHERCHE VIABILITÉ
JEUX, CONTRÔLE
UNIVERSITÉ PARIS DAUPHINE
75775 PARIS, CEDEX 16
FRANCE
E-MAIL: choirat@viab.ufrmd.dauphine.fr
Christian.Hess@dauphine.fr

R. SERI
CREST-LFA
TIMBRE J320
15 BD GABRIEL PÉRI
92245 MALAKOFF CEDEX
FRANCE
E.MAIL: seri@ensae.fr