A STRONG LAW OF LARGE NUMBERS FOR CAPACITIES¹

BY FABIO MACCHERONI AND MASSIMO MARINACCI

Università Bocconi and Università di Torino

We consider a totally monotone capacity on a Polish space and a sequence of bounded p.i.i.d. random variables. We show that, on a full set, any cluster point of empirical averages lies between the lower and the upper Choquet integrals of the random variables, provided either the random variables or the capacity are continuous.

1. Introduction. In this paper we prove a strong law of large numbers for totally monotone capacities. Specifically, given a totally monotone capacity ν defined on the Borel σ -algebra \mathcal{B} of a Polish space Ω , and a sequence $\{X_n\}_{n\geq 1}$ of bounded, pairwise independent and identically distributed random variables, we show that

$$\nu\left(\left\{\omega\in\Omega:\int X_1\,d\nu\leq\liminf_n\frac{\sum_{j=1}^nX_j(\omega)}{n}\right.\\ \leq\limsup_n\frac{\sum_{j=1}^nX_j(\omega)}{n}\leq-\int-X_1\,d\nu\right\}\right)=1,$$

provided the X_n s are continuous or simple, or ν is continuous. In this way we extend earlier results of Marinacci [13].

Under different names, totally monotone capacities have been widely studied in both pure and applied mathematics. They have been introduced by Choquet [4] motivated by some problems in potential theory, and in his wake many works have studied them in both potential theory and probability theory (see, e.g., [6] and [8]).

In mathematical statistics and in mathematical economics, totally monotone capacities have been used to represent subjective prior beliefs when the information on which such beliefs are based is not good enough to represent them by a standard additive probability (see, e.g., [7, 11, 15, 20] and [19]).

Our result shows that even in a nonadditive setting, the limit behavior of empirical averages has some noteworthy properties. In particular, we show that eventually empirical averages lie, with probability one, between the lower and upper Choquet integrals associated with the given capacity. This extends to the

Received February 2004; revised June 2004.

¹Supported by the Ministero dell'Istruzione, dell'Università e della Ricerca.

AMS 2000 subject classifications. 28A12, 60F15.

Key words and phrases. Capacities, Choquet integral, strong law of large numbers, contents, measures, outer measures, strong theorems.

nonadditive setting the classic Kolmogorov limit law, to which our result reduces when ν is additive since in this case lower and upper Choquet integrals coincide.

In a subjective probability perspective, our result says that, while in the additive case a Bayesian decision maker believes that the empirical averages of a sequence of p.i.i.d. random variables tend to a given number, here he just believes that the limit behavior of the empirical averages is confined in a given interval. This reflects a possible lack of confidence in his probability assessments. This interpretation of nonadditive limit laws and its relevance in mathematical economics has been recently discussed at length in [9], to which we refer the interested reader for details and references.

2. Preliminaries. Let Ω be a Polish space and \mathcal{B} its Borel σ -algebra. A random variable (r.v.) is a (Borel) measurable function $X: \Omega \to \mathbb{R}$. A totally *monotone capacity* on \mathcal{B} is a set function $\nu : \mathcal{B} \to [0, 1]$ such that:

(c.1) $\nu(\emptyset) = 0$ and $\nu(\Omega) = 1$,

(c.2) $\nu(A) \leq \nu(B)$ for all Borel sets $A \subseteq B$,

(c.3) $\nu(B_n) \downarrow \nu(B)$ for all sequences of Borel sets $B_n \downarrow B$,

(c.4) $\nu(G_n) \uparrow \nu(G)$ for all sequences of open sets $G_n \uparrow G$, (c.5) $\nu(\bigcup_{j=1}^n B_j) \ge \sum_{\varnothing \neq J \subseteq \{1,...,n\}} (-1)^{|J|+1} \nu(\bigcap_{j \in J} B_j)$ for every collection B_1, \ldots, B_n of Borel sets.

A set function $\nu : \mathcal{B} \to [0, 1]$ such that:

(c.6) $\nu(B_n) \uparrow \nu(\Omega)$ for all sequences of Borel sets $B_n \uparrow \Omega$,

is called *continuous*. A continuous set function $\nu: \mathcal{B} \to [0, 1]$ is a totally monotone capacity if and only if (c.1), (c.2) and (c.5) hold (see [18] and [14], Theorem 10).

Let ν be a totally monotone capacity on \mathcal{B} . As in the additive case, we say that the elements of a sequence $\{X_n\}_{n\geq 1}$ of r.v.s are *pairwise independent* with respect to ν if, for each $n, m \ge 1$ and for all open subsets G_n, G_m of \mathbb{R} ,

$$\nu(\{X_n \in G_n, X_m \in G_m\}) = \nu(\{X_n \in G_n\})\nu(\{X_m \in G_m\});$$

we say that they are *identically distributed* if, for each $n, m \ge 1$ and each open subset G of \mathbb{R} ,

$$\nu(X_n \in G) = \nu(X_m \in G).$$

The Choquet integral of a bounded r.v. X with respect to a totally monotone capacity v is defined by

$$\int X \, dv \equiv \int_0^{+\infty} v(\{X > t\}) \, dt + \int_{-\infty}^0 [v(\{X > t\}) - 1] \, dt.$$

The integrals on the right-hand side are Riemann integrals and they are well defined since $v({X > t})$ is a monotone function in t. The Choquet integral is positively homogeneous, monotone and translation invariant [i.e., $\int (X + c) dv =$ $\int Xdv + c$ if c is constant]. It reduces to the standard integral when v is an additive probability measure.

In general, $\int X dv \leq -\int -X dv$. Equality holds for all r.v.s if and only if v is additive. The integrals $\int X dv$ and $-\int -X dv$ are sometimes called *lower* and *upper Choquet integrals*, respectively.

3. The law of large numbers. We can now state our main result.

THEOREM 1. Let v be a totally monotone capacity on \mathcal{B} , and $\{X_n\}_{n\geq 1}$ a sequence of bounded, pairwise independent and identically distributed random variables. Then

$$\nu\left(\left\{\omega\in\Omega:\int X_1\,d\nu\leq\liminf_n\frac{\sum_{j=1}^nX_j(\omega)}{n}\right.\\ \leq\limsup_n\frac{\sum_{j=1}^nX_j(\omega)}{n}\leq-\int-X_1\,d\nu\right\}\right)=1,$$

provided at least one the following two conditions holds:

- (i) v is continuous;
- (ii) the random variables X_n are either continuous or simple.

A few remarks are in order. First, as $\nu(B) = 1$ and $A \subseteq B^c$ imply $\nu(A) = 0$, under the assumptions of Theorem 1 we also have

$$\nu\left(\left\{\omega\in\Omega:\liminf_{n}\frac{\sum_{j=1}^{n}X_{j}(\omega)}{n}<\int X_{1}\,d\nu\right\}\right)=0$$

and

$$\nu\left(\left\{\omega\in\Omega:\limsup_n\frac{\sum_{j=1}^nX_j(\omega)}{n}>-\int-X_1\,d\nu\right\}\right)=0.$$

In other words, with zero probability empirical averages will eventually lie outside the interval $\left[\int X_1 d\nu, -\int -X_1 d\nu\right]$.

Second, when ν is additive we have $\int X_1 d\nu = -\int -X_1 d\nu$, and so in this case our result reduces to a standard Kolmogorov limit law

$$\nu\left(\left\{\omega\in\Omega:\lim_n\frac{\sum_{j=1}^nX_j(\omega)}{n}=\int X_1\,d\nu\right\}\right)=1.$$

On the other hand, when ν is not additive in some cases it may happen (see [13]) that

$$\nu\left(\left\{\omega\in\Omega:\liminf_{n}\frac{\sum_{j=1}^{n}X_{j}(\omega)}{n}<\limsup_{n}\frac{\sum_{j=1}^{n}X_{j}(\omega)}{n}\right\}\right)=1.$$

Finally, as anticipated, the closest existing theorem is due to [13]. Our result is more general since [13] assumes that Ω is compact, ν is continuous, the

r.v.s X_n are continuous, independent and that they satisfy some further technical conditions. Moreover, the proof we provide is different and much simpler. In fact, here we develop a technique that relies on the relations between totally monotone capacities and correspondences, thus making it possible to use existing laws of large numbers for correspondences. This approach might be useful in establishing further generalizations of limit laws to the framework of capacities. This will be the object of future research, along with the possibility of weakening some of the continuity conditions assumed in Theorem 1.

4. Proof and related material. Denote by \mathcal{K}_{Ω} (resp. \mathcal{G}_{Ω}) the class of all nonempty compact subsets (resp. open subsets) of Ω ; for the sake of completeness, write \mathcal{B}_{Ω} instead of \mathcal{B} . If *d* is a Polish metric on Ω , then \mathcal{K}_{Ω} is a Polish space when endowed with the Hausdorff metric

$$d_{\mathcal{H}}(K,L) \equiv \max\left(\max_{k \in K} \min_{l \in L} d(k,l), \max_{l \in L} \min_{k \in K} d(l,k)\right).$$

The Borel σ -algebra on \mathcal{K}_{Ω} is also generated by the class $\{K \in \mathcal{K}_{\Omega} : K \subseteq G\}_{G \in \mathcal{G}_{\Omega}}$.

4.1. *Measurable correspondences and totally monotone capacities.* Let $(I, \mathcal{C}, \lambda)$ be a nonatomic and complete probability space. A (*compact valued*) *correspondence* $F : I \Longrightarrow \Omega$ is a map with domain I and whose values are nonempty compact subsets of Ω . For any $A \subseteq \Omega$, we put

$$F_{-1}(A) \equiv \{s \in I : F(s) \subseteq A\}.$$

A correspondence $F : I \rightrightarrows \Omega$ is *measurable* if $F_{-1}(G) \in \mathbb{C}$ for every $G \in \mathcal{G}_{\Omega}$. As well known (see, e.g., [12]), the following facts are equivalent:

- F is measurable;
- $F_{-1}(B) \in \mathcal{C}$ for every $B \in \mathcal{B}_{\Omega}$;
- *F* is measurable as a function $F: I \to \mathcal{K}_{\Omega}$.

When a measurable correspondence F is regarded as a measurable function $F: I \to \mathcal{K}_{\Omega}$, we denote by F^{-1} its standard inverse image, that is,

$$F^{-1}(\mathscr{E}) \equiv \{ s \in I : F(s) \in \mathscr{E} \} \qquad \forall \, \mathscr{E} \subseteq \mathcal{K}_{\Omega},$$

and by $\sigma(F)$ the σ -algebra generated by F, that is,

$$\sigma(F) \equiv \left\{ F^{-1}(\mathcal{D}) : \mathcal{D} \in \mathcal{B}_{\mathcal{K}_{\Omega}} \right\}.$$

In the sequel we will need the next lemma, whose standard proof is omitted.

LEMMA 2. Let $F: I \rightrightarrows \Omega$ be a measurable correspondence. Then

$$\sigma(F) = \sigma(\{F_{-1}(G) : G \in \mathcal{G}_{\Omega}\})$$

and $\{F_{-1}(G): G \in \mathcal{G}_{\Omega}\}$ is a π -class containing I.

A measurable function $f: I \to \Omega$ induces a *probability distribution* P_f on \mathcal{B}_{Ω} defined by

$$P_f(B) \equiv \lambda (f^{-1}(B)) \qquad \forall B \in \mathcal{B}_{\Omega}.$$

In a similar way, a measurable correspondence $F: I \Rightarrow \Omega$ induces a *lower* distribution v_F on \mathcal{B}_{Ω} defined by

$$\nu_F(B) \equiv \lambda(F_{-1}(B)) \quad \forall B \in \mathcal{B}_{\Omega}.$$

The next result, which links totally monotone capacities and lower distributions, is essentially due to Choquet [4] (see also [15, 16] and [3]).

LEMMA 3. A set function $v : \mathcal{B}_{\Omega} \to [0, 1]$ is a totally monotone capacity if and only if there exists a measurable correspondence $F : I \rightrightarrows \Omega$ such that $v = v_F$.

A measurable selection of a correspondence $F: I \Rightarrow \Omega$ is a measurable function $f: I \rightarrow \Omega$ such that $f(s) \in F(s)$ for almost all $s \in I$. The set of all measurable selections of F is denoted by Sel F. The Aumann integral (see [2]) of a correspondence $F: I \Rightarrow \mathbb{R}$ with respect to λ is defined by

$$\int F \, d\lambda \equiv \left\{ \int f \, d\lambda \colon f \in \text{Sel } F \text{ and } f \text{ integrable} \right\}.$$

If $X: \Omega \to \mathbb{R}$ is continuous or simple, and F is a correspondence, then $(X \circ F)(s) \equiv X(F(s))$ is a correspondence [i.e., $X(F(s)) \in \mathcal{K}_{\mathbb{R}}$ for all $s \in I$]. Moreover, since

$$(X \circ F)_{-1}(A) = F_{-1}(X^{-1}(A)) \qquad \forall A \subseteq \mathbb{R},$$

 $X \circ F$ is measurable provided X and F are measurable.

LEMMA 4. Let $F: I \rightrightarrows \Omega$ be a measurable correspondence, and $X: \Omega \rightarrow \mathbb{R}$ be either bounded and continuous or simple and measurable. Then,

$$\int (X \circ F) d\lambda = \left[\int X d\nu_F, -\int -X d\nu_F \right].$$

This is an immediate consequence of [3], Theorem 4.1.

4.2. *Proof of Theorem* 1. Suppose first that (ii) holds. By Lemma 3, there exists a measurable correspondence $F: I \Rightarrow \Omega$ such that $v = v_F$.

Next we show that the measurable correspondences $\{X_n \circ F\}_{n\geq 1}$ are pairwise independent and identically distributed when regarded as measurable functions $X_n \circ F : I \to \mathcal{K}_{\mathbb{R}}$.

Let $n, m \ge 1$ and $G_n, G_m \in \mathcal{G}_{\mathbb{R}}$, then

$$\lambda ((X_n \circ F)_{-1}(G_n) \cap (X_m \circ F)_{-1}(G_m))$$

= $\lambda (F_{-1}(X_n^{-1}(G_n)) \cap F_{-1}(X_m^{-1}(G_m)))$
= $\lambda (F_{-1}(X_n^{-1}(G_n) \cap X_m^{-1}(G_m)))$
= $\nu (X_n^{-1}(G_n) \cap X_m^{-1}(G_m))$
= $\nu (X_n^{-1}(G_n))\nu (X_m^{-1}(G_m))$
= $\lambda (F_{-1}(X_n^{-1}(G_n)))\lambda (F_{-1}(X_m^{-1}(G_m)))$
= $\lambda ((X_n \circ F)_{-1}(G_n))\lambda ((X_m \circ F)_{-1}(G_m)))$

This proves pairwise independence, since for all j = n, m, $\{(X_j \circ F)_{-1}(G)\}_{G \in \mathcal{G}_{\mathbb{R}}}$ is a π -class containing I and generating the σ -algebra $\sigma(X_j \circ F)$ (see Lemma 2).

Moreover, for each $n, m \ge 1$, and each open subset $G \in \mathcal{G}_{\mathbb{R}}$,

$$\lambda ((X_n \circ F)^{-1} (\{K \in \mathcal{K}_{\mathbb{R}} : K \subseteq G\}))$$

= $\lambda (\{X_n \circ F \in \{K \in \mathcal{K}_{\mathbb{R}} : K \subseteq G\}\})$
= $\lambda ((X_n \circ F)_{-1}(G))$
= $\lambda (F_{-1}(X_n^{-1}(G)))$
= $\nu (X_n^{-1}(G))$
= $\nu (X_m^{-1}(G))$
= $\lambda ((X_m \circ F)^{-1} (\{K \in \mathcal{K}_{\mathbb{R}} : K \subseteq G\}))$

This proves identical distribution since $\{K \in \mathcal{K}_{\mathbb{R}} : K \subseteq G\}_{G \in \mathcal{G}_{\mathbb{R}}}$ is a π -class containing $\mathcal{K}_{\mathbb{R}}$ and generating $\mathcal{B}_{\mathcal{K}_{\mathbb{R}}}$.

Clearly, for each $n \ge 1$ and each $h \in \text{Sel } X_n \circ F$, $\int h \, d\lambda$ is finite (*h* is bounded); moreover, by Lemma 4, $\int X_n \circ F \, d\lambda \in \mathcal{K}_{\mathbb{R}}$.

In sum, $\{X_n \circ F\}_{n \ge 1}$ are pairwise independent and identically distributed measurable correspondences with $\int X_n \circ F d\lambda \in \mathcal{K}_{\mathbb{R}}$ for all $n \ge 1$. A generalization due to [10] (see also [5] and [17]) of a result of Artstein and Vitale [1] guarantees that

$$\lambda\left(\left\{s \in I : \frac{1}{n} \sum_{j=1}^{n} X_j(F(s)) \to \int X_1 \circ F \, d\lambda\right\}\right) = 1.$$

By Lemma 4,

$$\int X_1 \circ F \, d\lambda = \left[\int X_1 \, d\nu, -\int -X_1 \, d\nu \right].$$

Let
$$a_n(\omega) = \frac{\sum_{j=1}^n X_j(\omega)}{n}$$
 and set
 $S_1 \equiv \left\{ s \in I : \frac{1}{n} \sum_{j=1}^n X_j(F(s)) \rightarrow \left[\int X_1 d\nu, -\int -X_1 d\nu \right] \right\},$
 $S_2 \equiv \left\{ s \in I : \int X_1 d\nu \le \liminf_n a_n(\omega) \le \inf_n \sup_n a_n(\omega) \le \lim_n \sup_n a_n(\omega) \le -\int -X_1 d\nu \ \forall \ \omega \in F(s) \right\},$
 $\Omega_2 \equiv \left\{ \omega \in \Omega : \int X_1 d\nu \le \liminf_n a_n(\omega) \le \limsup_n a_n(\omega) \le -\int -X_1 d\nu \right\}.$

We want to show that $\nu(\Omega_2) = 1$. Notice that

$$\nu(\Omega_2) = \lambda(\{s \in I : F(s) \subseteq \Omega_2\}) = \lambda(S_2).$$

The next claim will be used to show that $S_1 \subseteq S_2$.

CLAIM 1. Let
$$\{K_n\}$$
 be a sequence in $\mathcal{K}_{\mathbb{R}}$ such that $K_n \to [\alpha, \beta]$. Then,
 $\alpha \leq \liminf_n k_n \leq \limsup_n k_n \leq \beta$

for each sequence $\{k_n\}$ in \mathbb{R} such that $k_n \in K_n$ for all $n \ge 1$.

PROOF. By definition of Hausdorff metric, K_n converges to $[\alpha, \beta]$ if and only if $\max(\max_{t_n \in K_n} \min_{r \in [\alpha, \beta]} |t_n - r|, \max_{r \in [\alpha, \beta]} \min_{t_n \in K_n} |r - t_n|) \to 0$, in particular

(1)
$$\max_{t_n \in K_n} \min_{r \in [\alpha,\beta]} |t_n - r| \to 0.$$

Let $\{k_{n_j}\}$ be a subsequence of $\{k_n\}$ such that $k_{n_j} \to \ell \in [-\infty, +\infty]$. If $\ell \notin [\alpha, \beta]$, then there exists $\varepsilon > 0$ such that eventually $|k_{n_j} - r| > \varepsilon$ for all $r \in [\alpha, \beta]$. Hence, we have eventually $\min_{r \in [\alpha, \beta]} |k_{n_j} - r| > \varepsilon$, thus contradicting (1). \Box

If $s \in S_1$, then $\frac{1}{n} \sum_{j=1}^n X_j(F(s)) \to [\int X_1 d\nu, -\int -X_1 d\nu]$. Hence, for all $\omega \in F(s)$, we have $a_n(\omega) = \frac{1}{n} \sum_{j=1}^n X_j(\omega) \in \frac{1}{n} \sum_{j=1}^n X_j(F(s))$; by Claim 1,

$$\int X_1 \, d\nu \leq \liminf_n a_n(\omega) \leq \limsup_n a_n(\omega) \leq -\int -X_1 \, d\nu.$$

Therefore, $S_1 \subseteq S_2$ and so $\nu(\Omega_2) = \lambda(S_2) \ge \lambda(S_1) = 1$. This completes the proof of the result when (ii) holds.

As to (i), denoting by τ the Polish topology on Ω , there exists a Polish topology $\tau^* \supseteq \tau$ on Ω such that $\sigma(\tau^*) = \mathcal{B}_{\Omega}$, and such that all the X_n s are τ^* -continuous (see, e.g., [21]). Since ν is continuous, then it is a totally monotone capacity with respect to the topology τ^* ; we can thus assume that (ii) holds.

Acknowledgments. We thank Steve Lalley (the Editor) and an anonymous referee for helpful suggestions.

REFERENCES

- ARTSTEIN, Z. and VITALE, R. A. (1975). A strong law of large numbers for random compact sets. Ann. Probab. 3 879–882.
- [2] AUMANN, R. J. (1965). Integrals of set-valued functions. J. Math. Anal. Appl. 12 1-12.
- [3] CASTALDO, A., MACCHERONI, F. and MARINACCI, M. (2004). Random correspondences as bundles of random variables. Sankhyā 66 409–427.
- [4] CHOQUET, G. (1954). Theory of capacities. Ann. Inst. Fourier 5 131-295.
- [5] COLUBI, A., LÓPEZ-DÍAZ, M., DOMÍNGUEZ-MENCHERO, J. S. and GIL, M. A. (1999). A generalized strong law of large numbers. *Probab. Theory Related Fields* 114 401–417.
- [6] DELLACHERIE, C. and MEYER, P.-A. (1978). Probabilities and Potential. North-Holland, Amsterdam.
- [7] DEMPSTER, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* 38 325–339.
- [8] DOOB, J. (1984). Classical Potential Theory and Its Probabilistic Counterpart. Springer, Berlin.
- [9] EPSTEIN, L. and SCHNEIDER, M. (2003). IID: Independently and indistinguishably distributed. J. Econom. Theory **113** 32–50.
- [10] HESS, C. (1984). Quelques théorèmes limites pour des ensembles aléatoires bornés ou non. In Séminaire d'Analyse Convexe 14. Univ. Sciences et Techniques de Langvedoc, Montpelier.
- [11] HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. Ann. Statist. 1 251–263.
- [12] KLEIN, E. and THOMPSON, A. C. (1984). Theory of Correspondences. Wiley, New York.
- [13] MARINACCI, M. (1999). Limit laws for non-additive probabilities and their frequentist interpretation. J. Econom. Theory 84 145–195.
- [14] MARINACCI, M. and MONTRUCCHIO, L. (2004). Introduction to the mathematics of ambiguity. In *Uncertainty in Economic Theory* (I. Gilboa, ed.) 46–107. Routledge, New York.
- [15] NGUYEN, H. T. (1978). On random sets and belief functions. J. Math. Anal. Appl. 65 531-542.
- [16] PHILIPPE, F., DEBS, G. and JAFFRAY, J.-Y. (1999). Decision making with monotone lower probabilities of infinite order. *Math. Oper. Res.* 24 767–784.
- [17] PROSKE, F. N. and PURI, M. L. (2003). A strong law of large numbers for generalized random sets from the viewpoint of empirical processes. *Proc. Amer. Math. Soc.* 131 2937–2944.
- [18] SCHMEIDLER, D. (1972). Cores of exact games. J. Math. Anal. Appl. 40 214-225.
- [19] SCHMEIDLER, D. (1989). Subjective probability and expected utility without additivity. *Econometrica* 57 571–587.
- [20] SHAFER, G. (1976). A Mathematical Theory of Evidence. Princeton Univ. Press.
- [21] SRIVASTAVA, S. M. (1998). A Course on Borel Sets. Springer, New York.

ISTITUTO DI METODI QUANTITATIVI AND IGIER UNIVERSITÀ BOCCONI VIALE ISONZO 25 20135 MILANO ITALY E-MAIL: fabio.maccheroni@unibocconi.it URL: http://web.econ.unito.it/gma/fabio.htm DIPARTIMENTO DI STATISTICA E MATEMATICA APPLICATA AND ICER UNIVERSITÀ DI TORINO PIAZZA ARBARELLO 8 10122 TORINO ITALY E-MAIL: massimo.marinacci@unito.it URL: http://web.econ.unito.it/gma/massimo.htm