TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 366, Number 10, October 2014, Pages 5431–5451
S 0002-9947(2014)06053-4
Article electronically published on June 6, 2014

LAWS OF LARGE NUMBERS WITHOUT ADDITIVITY

PEDRO TERÁN

Abstract. The law of large numbers is studied under a weakening of the axiomatic properties of a probability measure. Averages do not generally converge to a point, but they are asymptotically confined in a limit set for any random variable satisfying a natural 'finite first moment' condition. It is also shown that their behaviour can depart strikingly from the intuitions developed in the additive case.

1. Introduction and main results

Both the law of large numbers and non-additive probabilities appear in Bernoulli's *Ars Conjectandi* (1713), although only separately: he proved his LLN in the additive setting related to games of chance, while non-additivity pertained to epistemic problems of rational belief (see [32]). The frequentist properties of non-additive set functions have attracted attention in various fields only much more recently.

Walley and Fine [33] tried to model real-world processes whose empirical frequencies failed to stabilize, naturally leading to upper and lower probabilities. Badard [4] was concerned with knowledge representation in artificial intelligence, where possibility measures replace addition by the maximum. Molchanov [24] studied the Glivenko–Cantelli theorem for capacities induced by random sets. Quadrat gave a law of large numbers for cost measures associated to solutions of optimization problems [1]. Marinacci [21] aimed at modelling a decisor's imperfect knowledge of the limit frequencies within the multiple priors approach to economic decision theory. Puhalskii [28] was studying the role of idempotent probabilities in large deviation theory. Peng [26] was motivated by problems of stochastic differential equations and risk modelling in finance.

In any of those fields, the law of large numbers can play a foundational role akin to its role in additive probability theory. Some works use alternative 'probability' axioms where the sum and product are replaced by another set of operations [1, 4, 28], while others adopt weaker axioms making the convergence in the ordinary LLN (and maybe those alternative LLNs) a special case [21, 26, 33]. This paper contributes to the latter stream.

Received by the editors July 24, 2012 and, in revised form, December 8, 2012.

2010 *Mathematics Subject Classification.* Primary 60F15; Secondary 28A12, 60A05.

Key words and phrases. Capacity, Choquet integral, law of large numbers, non-additive probability.

This paper is dedicated to the memory of Professor Teófilo Brezmes Brezmes, an excellent lecturer and appreciated colleague.

This research was partially funded by Spain's *Ministerio de Ciencia e Innovación* (TIN2008-06796-C04-04, MTM2011-22993, ECO1022–24181).

The non-additive counterpart of Bernoulli's LLN may be the best way to introduce our result. When the probability measure is replaced by a set function ν assumed to be completely monotone (a weakened form of the union–intersection formula), the empirical frequencies $f_n(A)$ of an event A over a sequence of Bernoulli trials satisfy

$$\nu\left(\nu(A) \leq \liminf_n f_n(A) \leq \limsup_n f_n(A) \leq 1 - \nu(A^c)\right) = 1.$$

In absence of additivity the upper and lower bounds may not be the same, so in fact ν has a dual 'companion' $\overline{\nu}(A) = 1 - \nu(A^c)$; clearly, also $\nu(A) = 1 - \overline{\nu}(A^c)$.

In order to replace the event by a random variable ξ on a measurable space $(\Omega, \mathcal{A})$, one wants to calculate expectations with respect to ν and $\overline{\nu}$ in such a way that $E_\nu[I_A] = \nu(A)$, $E_{\overline{\nu}}[I_A] = 1 - \nu(A^c)$ and both expectations equal the usual definition when ν is a probability measure. That is achieved by using the Choquet integral introduced in [9].

The resulting type of limit theorem was already considered by Marinacci [21], Maccheroni and Marinacci [20] and more recent papers [7, 10, 11, 29, 30]. In those papers, the weakening of the axiomatic properties of probability has been balanced by the incorporation of extra technical assumptions on the properties of Ω and/or the random variables. For instance, that Ω is a compact or Polish topological space, that the random variables are continuous functions, or bounded, or at least have sufficiently high moments, or are independent in an unreasonably strong sense, or satisfy ad hoc regularity requirements.

The aim of this paper is to show that additivity can be replaced by complete monotony in the LLN under 'the' reasonable first moment condition that both Choquet expectations $E_\nu[\xi]$ and $E_{\overline{\nu}}[\xi]$ are finite, with no additional assumptions. Under the assumption that the sample space is endowed with a topology, also the continuity for monotone sequences can be relaxed.

In this paper, a *capacity* in a measurable space $(\Omega, \mathcal{A})$ is a set function $\nu : \mathcal{A} \to [0,1]$ such that

(i) $\nu(\varnothing) = 0, \nu(\Omega) = 1$,
(ii) $A \subset B \Longrightarrow \nu(A) \leq \nu(B)$,
(iii) $B_n \searrow B \Longrightarrow \nu(B_n) \to \nu(B)$ whenever $B_n, B \in \mathcal{A}$,
(iv) $B_n \nearrow B \Longrightarrow \nu(B_n) \to \nu(B)$ whenever $B_n, B \in \mathcal{A}$.

If Ω has a topological structure and $\mathcal{A}$ is the corresponding Borel σ-algebra $\mathcal{B}_\Omega$, it is advantageous to replace (iv) above by the following weaker assumption:

(iv′) $G_n \nearrow G \Longrightarrow \nu(G_n) \to \nu(G)$ whenever G_n, G are open.

In that case we call ν a *topological capacity.* Indeed the full continuity property of capacities is usually too strong for applications (e.g. in potential theory and random sets).

Moreover, we make the assumption that ν is *completely monotone*:

(v) $\nu(\bigcup_{i=1}^n A_i) \geq \sum_{I \subset \{1,\ldots,n\}, I \neq \varnothing} (-1)^{|I|+1} \nu(\bigcap_{i \in I} A_i)$ for any $n \in \mathbb{N}$.

The *distribution* or *law* of a random variable ξ is $\nu_\xi : A \in \mathcal{B}_\mathbb{R} \mapsto \nu(\xi \in A)$. Two random variables on $(\Omega, \mathcal{A}, \nu)$ are *identically distributed* if they have the same law. A sequence $\{\xi_n\}_n$ of random variables is called *preindependent* if

$$\nu(\bigcap_{i=1}^k \{\xi_{n_i} \in A_i\}) = \prod_{i=1}^k \nu(\xi_{n_i} \in A_i)$$

for any choice of $\{n_1, \dots, n_k\} \subset \mathbb{N}$, $A_i \in \mathcal{A}$ and $k \in \mathbb{N}$. If that property holds just for $k = 2$, the sequence is called *pairwise preindependent*.

The *Choquet integral* of a bounded random variable ξ against ν is defined to be

$$\int \xi d\nu = \int_0^\infty \nu(\{\xi \geq t\})dt - \int_{-\infty}^0 [1 - \nu(\{\xi \geq t\})]dt,$$

where both Riemann integrals exist due to the monotonicity of the function $t \mapsto \nu(\{\xi \geq t\})$. The definition extends to unbounded random variables if at least one of the (now improper) Riemann integrals is finite, in which case we say that the *lower expectation* $E_\nu[\xi] := \int \xi d\nu$ exists. Analogous definitions hold for integration against $\overline{\nu}$, with the *upper expectation* of ξ being $E_{\overline{\nu}}[\xi] := \int \xi d\overline{\nu}$ (note that $\nu \leq \overline{\nu}$ under our assumptions on ν). If both $E_\nu[\xi]$ and $E_{\overline{\nu}}[\xi]$ are finite, ξ is called *integrable*. In contrast to the additive case, this can be strictly weaker than the property $E_\nu[|\xi|] < \infty$.

Our main theorem then reads as follows.

Theorem 1.1. *Let ν be a completely monotone set function on a measurable space $(\Omega, \mathcal{A})$. Let ξ be an integrable random variable, and $\{\xi_n\}_n$ pairwise preindependent and identically distributed to ξ. Assume that one of the following holds:*

(i) *ν is a capacity,*
(ii) *$(\Omega, \mathcal{A})$ is a Polish space with its Borel σ-algebra and ν is a topological capacity,*
(iii) *$(\Omega, \mathcal{A})$ is a topological space with its Borel σ-algebra, ν is a topological capacity, and the ξ_n are continuous functions on Ω.*

Then,

(a) $\nu(E_\nu[\xi] - \varepsilon < \frac{\sum_{i=1}^n \xi_i}{n} < E_{\overline{\nu}}[\xi] + \varepsilon) \to 1$ *for every* $\varepsilon > 0$,
(b) $\nu(E_\nu[\xi] - \varepsilon < \frac{\sum_{i=1}^n \xi_i}{n} < E_{\overline{\nu}}[\xi] + \varepsilon$ *eventually*$) = 1$ *for every* $\varepsilon > 0$,
(c) $\nu(E_\nu[\xi] \leq \liminf_n \frac{\sum_{i=1}^n \xi_i}{n} \leq \limsup_n \frac{\sum_{i=1}^n \xi_i}{n} \leq E_{\overline{\nu}}[\xi]) = 1$.

To understand why separate bounds are needed instead of a single limit, it is helpful to recast the LLN as a uniform LLN over a family of probability measures: with [17, Theorem 7.1], conclusion (a) becomes

$$\inf_{P \geq \nu} P\left(E_\nu[\xi] - \varepsilon < \frac{\sum_{i=1}^n \xi_i}{n} < E_{\overline{\nu}}[\xi] + \varepsilon\right) \to 1. \tag{1}$$

The ξ_n cannot be expected to satisfy that simultaneous LLN with a unique limit.

The tools for proving these LLNs are developed in Section 3, and the laws themselves are proved in Section 4. The paper has a second part (Section 5) in which we examine at length some behaviours that the LLN allows. Such examples are conspicuously absent from previous papers; the emphasis on the formal parallels with the ordinary LLN, without testing their true reach against examples, may have caused in some readers (as was certainly my case) the wrong impression that what goes on is very similar, only with fluctuating instead of stabilizing frequencies.

The contribution to building a library of examples is, in my opinion, at least as important as the theorem-proving part of the paper. For instance, we will present a counterintuitive 'i.i.d.' sequence for which $n^{-1}\sum_{i=1}^n \xi_i \to 0$ almost surely, while $\nu(|n^{-1}\sum_{i=1}^n \xi_i| \leq 1/2) = 0$ and $n^{-1}\sum_{i=1}^n \xi_i$ converges in law to a (non-additive) distribution concentrated in $[0, 1/2]$.

A few other striking phenomena are the following. The 'weak' LLN can be stronger than the 'strong' LLN. A non-degenerate random variable can be 'independent' of itself (hence we prefer the term 'preindependence'). Two sequences of preindependent copies of a random variable can see their averages converge almost surely to different limits. The bounds in Theorem 1.1.(c) can be impossible to improve and yet can be the essential minimum and maximum of the variable.

Finally, Section 6 closes the paper with a discussion of the results and a list of open questions. The reader unfamiliar with the tools used in this paper is referred to [23, Chapter 1] for a basic overview of random sets and their interplay with capacities, and to the books [12, 14, 18] and the survey papers [2, 15] for more on capacities, Choquet integrals and some of their applications.

2. Preliminaries

The complement of a set A will be denoted A^c, its closure $\mathrm{cl}\, A$, and its convex hull $\mathrm{co}\, A$. If d is a metric, we write $d(x, A) = \inf_{y\in A} d(x, y)$.

A set function ν which satisfies (v) just for $n = 2$ is called *2-monotone*; 2-monotony already implies (ii). A capacity is called *null-additive* if $\nu(A\cup B) = \nu(A)$ whenever $\nu(B) = 0$.

If ν is completely monotone, then $\overline{\nu}$ satisfies the property dual to (iv) obtained by reversing the inequality and interchanging unions and intersections (called *complete alternation*). The set of all probability measures dominated by ν is its *core*,

$$\mathrm{core}(\nu) = \{P : \mathcal{A} \to [0, 1] \text{ probability measure} \mid P \geq \nu\}.$$

Sequences of random variables. A sequence $\{\xi_n\}_n$ *converges almost surely* to a random variable ξ if

$$\nu(\xi_n \to \xi) = 1.$$

The property that $\nu(\xi_n \not\to \xi) = 0$, or equivalently $\overline{\nu}(\xi_n \to \xi) = 1$, is strictly weaker; it may happen that $\overline{\nu}(\xi_n \to x) = 1$ and $\overline{\nu}(\xi_n \to y) = 1$ for some $x \neq y$.

The sequence *converges in probability* to ξ if, for each $\varepsilon > 0$,

$$\nu(|\xi_n - \xi| < \varepsilon) \to 1.$$

Again, the property that $\nu(|\xi_n - \xi| > \varepsilon) \to 0$, or equivalently $\overline{\nu}(|\xi_n - \xi| < \varepsilon) \to 1$, is strictly weaker.

Following Norberg [25], we say that a sequence $\{\nu_n\}_n$ of topological capacities *converges weakly* (or *in law*) to ν if both

(i) $\liminf_n \nu_n(G) \geq \nu(G)$ for all open G, and
(ii) $\limsup_n \nu_n(F) \leq \nu(F)$ for all closed F

hold. These two conditions are equivalent for probability measures, but not in general.

When a sequence $\{\xi_n\}_n$ is being considered, we denote

$$S_n = n^{-1} \sum_{i=1}^{n} \xi_i.$$

For ξ'_n, ξ''_n, we will use the notation S'_n, S''_n accordingly.

Random sets. If Ω is a topological space, we denote by $\mathcal{F}(\Omega)$ the class of all non-empty closed subsets of Ω, and by $\mathcal{K}(\Omega)$ that of all non-empty compact subsets of Ω.

A mapping Γ from a probability space $(\Omega', \mathcal{A}, \mathbb{P})$ to $\mathcal{F}(\Omega)$ is called a *random closed set* or just a *random set* if it is measurable with respect to $\mathcal{A}$ and the Effros σ-algebra generated by $\{\mathcal{F}(A) \mid A \in \mathcal{F}(\Omega)\}$. A random closed set Γ is called *strongly measurable* if the events $\{\Gamma \subset B\}$ are $\mathcal{A}$-measurable not just for closed sets but for all $B \in \mathcal{B}_\Omega$.

A *selection* $\gamma : \Omega' \to \Omega$ of Γ is a Borel measurable function such that $\gamma \in \Gamma$ $\mathbb{P}$-almost surely. The *containment functional* C_Γ is the mapping

$$A \in \mathcal{F}(\Omega) \mapsto \mathbb{P}(X \in \mathcal{F}(A)) = \mathbb{P}(X \subset A).$$

Since the family $\{\mathcal{F}(A) \mid A \in \mathcal{F}(\Omega)\}$ is a π-system, two random sets having the same containment functional are identically distributed.

3. Support results

In this section we develop the main tools for the proof of the laws of large numbers. The first result will allow us to reduce the proof in a general measurable space to the case of a Polish space with the Borel σ-algebra.

Lemma 3.1. *Let $\{\xi_n\}_n$ be a sequence of random variables defined on a space $(\Omega, \mathcal{A})$ with a capacity ν. Then, there exist a sequence $\{\eta_n\}_n$ of random variables on the Polish space $\mathbb{R}^{\mathbb{N}}$ with its Borel σ-algebra and a capacity ν_* such that*

(a) *If ν is completely monotone, then ν_* is so as well.*
(b) *For any Polish space $\mathbb{F}$, any Borel measurable function $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{F}$, and any $A \in \mathcal{B}_{\mathbb{F}}$,*
$$\nu_*(f(\eta_1, \dots, \eta_n, \dots) \in A) = \nu(f(\xi_1, \dots, \xi_n, \dots) \in A).$$
(c) *For any Polish space $\mathbb{F}$, any sequence of Borel measurable functions $f_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{F}$, and any sequence $\{A_n\}_n \subset \mathcal{B}_{\mathbb{F}}$,*
$$\nu_*(\liminf_n \{f_n(\eta_1, \dots, \eta_n, \dots) \in A_n\}) = \nu(\liminf_n \{f_n(\xi_1, \dots, \xi_n, \dots) \in A_n\})$$
and
$$\nu_*(\limsup_n \{f_n(\eta_1, \dots, \eta_n, \dots) \in A_n\}) = \nu(\limsup_n \{f_n(\xi_1, \dots, \xi_n, \dots) \in A_n\}).$$
(d) *If the ξ_n are preindependent, or pairwise preindependent, or identically distributed, then the η_n are so as well.*
(e) *$E_\nu[\xi_1] = E_{\nu_*}[\eta_1]$ and $E_{\overline{\nu}}[\xi_1] = E_{\overline{\nu_*}}[\eta_1]$.*

Proof. A generic element $(x_1, x_2, \dots)$ of $\mathbb{R}^{\mathbb{N}}$ will be denoted by x_*. The topology of pointwise convergence τ_p in $\mathbb{R}^{\mathbb{N}}$ is indeed Polish, since it is generated by the separable complete metric

$$d(x_*, y_*) = \sum_{i=1}^{n} \frac{|x_n - y_n|}{2^n(|x_n - y_n| + 1)}.$$

Its Borel σ-algebra is generated by the coordinate projections $\pi_n : x_* \mapsto x_n$, since they are τ_p-continuous and separate any two elements of $\mathbb{R}^{\mathbb{N}}$ [8, Théorème I.2.5].

Consider the mapping $\xi_* : \Omega \to \mathbb{R}^{\mathbb{N}}$ given by

$$\xi_*(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots).$$

Observe that ξ_* is Borel measurable, since each $\pi_n(\xi_*) = \xi_n$ is a random variable.

Proof of part (a). Define in $\mathbb{R}^{\mathbb{N}}$ the set function

$$\nu_* = \nu \circ \xi_*^{-1}.$$

Let us check that ν_* is a capacity and that it is completely monotone if ν is so. Properties (i, ii, v) follow from [31, p. 829] since ξ_*^{-1} is a $\bigcap$-homomorphism from $\mathcal{B}_{\mathbb{R}^{\mathbb{N}}}$ to $\mathcal{A}$. As to the continuity properties (iii, iv), they follow easily from the properties of preimages and ν.

Proof of parts (b, c). Clearly (b) is a particular case of (c). Let us prove the latter in the liminf case; the limsup is analogous.

Take the random variables $\eta_n = \pi_n : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}^{\mathbb{N}}}, \nu_*) \to \mathbb{R}$. Then

$$\begin{aligned}
&\nu_*(\liminf_n \{f_n(\eta_1, \eta_2, \ldots) \in A_n\}) \\
&= \nu_*(\{x_* \in \mathbb{R}^{\mathbb{N}} \mid \exists m \in \mathbb{N} \mid \forall n \geq m, f_n(\eta_1(x_*), \eta_2(x_*), \ldots) \in A_n\}) \\
&= \nu(\xi_*^{-1}(\{x_* \in \mathbb{R}^{\mathbb{N}} \mid \exists m \in \mathbb{N} \mid \forall n \geq m, f_n(\pi_1(x_*), \pi_2(x_*), \ldots) \in A_n\}) \\
&= \nu(\{\omega \in \Omega \mid \exists m \in \mathbb{N} \mid \forall n \geq m, f_n(\pi_1(\xi_*(\omega)), \pi_2(\xi_*(\omega)), \ldots) \in A_n\}) \\
&= \nu(\exists m \in \mathbb{N} \mid \forall n \geq m, f_n(\xi_1, \xi_2, \ldots) \in A_n) \\
&= \nu(\liminf_n \{f_n(\xi_1, \xi_2, \ldots) \in A_n\}).
\end{aligned}$$

Proof of part (d). Preindependence and equidistribution properties follow from particular choices of f and A in (b). To prove

$$\nu_*(\bigcap_{i=1}^k \{\eta_i \in B_i\}) = \prod_{i=1}^k \nu_*(\eta_i \in B_i),$$

take f the identity and $A = B_1 \times \ldots \times B_k \times \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \times \ldots$.

To prove $\nu_*(\eta_i \in B) = \nu_*(\eta_j \in B)$, take $\mathbb{F} = \mathbb{R}$ and the choices $f(x_*) = x_i, f(x_*) = x_j$.

Proof of part (e). It follows from the definition of the Choquet integral, applying part (b) with $f(x_*) = x_i$ and $A = [t, \infty)$ to each $t \geq 0$. □

An important role will be played by the following proposition, which combines results of Philippe et al. [27] and Artstein [3] in a unified probability space.

Proposition 3.2. *Let ν be a completely monotone topological capacity in a Polish space Ω. Then, there exists a complete probability space $(\Omega', \mathcal{A}', \mathbb{P})$ and a strongly measurable random set $\Gamma : \Omega' \to \mathcal{K}(\Omega)$ such that*

(a) *$\mathbb{P}(\Gamma \subset B) = \nu(B)$ for each Borel set $B \in \mathcal{B}_\Omega$.*
(b) *For each $P \in \mathrm{core}(\nu)$, there exists a selection γ of Γ for which $\mathbb{P}_\gamma = P$.*

Proof. We endow $\mathcal{K}(\Omega)$ with the hit-and-miss topology having as a subbase the sets $\{A \in \mathcal{K}(\Omega) \mid A \cap F = \varnothing\}$ for F closed and $\{A \in \mathcal{K}(\Omega) \mid A \cap G \neq \varnothing\}$ for G open (see e.g. [22, 23]). Thus $\mathcal{K}(\Omega)$ becomes a Polish space whose Borel σ-algebra is the trace Effros σ-algebra.

Theorem 2 in [27] provides a probability measure Q on $\mathcal{B}_{\mathcal{K}(\Omega)}$ whose extension $\hat{Q}$ to the σ-algebra Σ of universally measurable sets is such that

$$\nu(B) = \hat{Q}(\mathcal{K}(B))$$

for all Borel B. Consider the product $\hat{Q} \otimes \lambda_{[0,1]}$ of $\hat{Q}$ with the Lebesgue measure in $[0,1]$, and define

$$\Gamma : (\mathcal{K}(\Omega) \times [0,1], \Sigma \otimes \mathcal{B}_{[0,1]}) \to (\mathcal{K}(\Omega), \mathcal{B}_{\mathcal{K}(\Omega)})$$

by $\Gamma(K,t) = K$. For any Borel set B,

$$\{\Gamma \subset B\} = \mathcal{K}(B) \times [0,1]$$

and, by [27, Lemma 1], $\mathcal{K}(B) \in \Sigma$. Thus $\{\Gamma \subset B\} \in \Sigma \otimes \mathcal{B}_{[0,1]}$ and Γ is strongly measurable with the law $\hat{Q} \otimes \lambda_{[0,1]}$.

Now, if $P \in \mathrm{core}(\nu)$ we have

$$P(G) \geq \nu(G) = (\hat{Q} \otimes \lambda_{[0,1]})(\Gamma \subset G)$$

for open G. Since $\mathcal{K}(G) = \{A \in \mathcal{K}(\Omega) \mid A \cap G^c = \varnothing\}$ is open, the event $\{\Gamma \subset G\}$ is indeed in $\mathcal{B}_{\mathcal{K}(\Omega)} \otimes \mathcal{B}_{[0,1]} = \mathcal{B}_{\mathcal{K}(\Omega)\times[0,1]}$. Thus we write

$$P(G) \geq (Q \otimes \lambda_{[0,1]})(\Gamma \subset G)$$

and, taking complements,

$$P(F) \leq (Q \otimes \lambda_{[0,1]})(\Gamma \cap F \neq \varnothing)$$

for all closed F. Theorems 2.1 and 2.2 in [3] now yield a $\mathcal{B}_{\mathcal{K}(\Omega)\times[0,1]}$-measurable selection γ of Γ whose law $(Q \otimes \lambda_{[0,1]})_\gamma$ is P.

Finally, take $\Omega' = \mathcal{K}(\Omega) \times [0,1]$, $\mathcal{A}'$ the completion of $\Sigma \otimes \mathcal{B}_{[0,1]}$ and $\mathbb{P}$ the completion of $\hat{Q} \otimes \lambda_{[0,1]}$, and check routinely that (a) and (b) hold not just for $\hat{Q} \otimes \lambda_{[0,1]}$ but also for $\mathbb{P}$. □

The next tool, once Proposition 3.2 has provided a random set, is a Strong Law of Large Numbers for random closed sets due to Hess [16, Theorem 3.5]. Its statement requires the following notions. Let X be a random closed set in $\mathbb{R}$ (we drop the former notation Γ because in the proof certain Γ_n will be used to construct new random sets X_n to which this law of large numbers will be applied). Its *Aumann expectation* is the set

$$EX = \mathrm{cl}\{E[\xi] \mid \xi \text{ integrable selection of } X\}.$$

Operations on sets are defined as

$$\lambda A = \{\lambda x \mid x \in A\}, \quad A + B = \mathrm{cl}\{x + y \mid x \in A, y \in B\}.$$

Taking the closure is necessary to ensure that the sum is again in $\mathcal{F}(\mathbb{R})$. For instance, if $A = \mathbb{Z}$ and $B = \{n + (n-1)/(2n) \mid n \geq 2\}$, then the elementwise sum is not closed.

Lemma 3.3 (Hess). *Let $\{X_n\}_n$ be a pairwise independent sequence of identically distributed random closed sets in $\mathbb{R}$, such that $E[d(0,X_1)] < \infty$. Then there is a null set out of which, for every $x \in \mathbb{R}$,*

$$d\Big(x, n^{-1}\sum_{i=1}^{n} X_i\Big) \to d(x, E\,\mathrm{co}\,X_1).$$

The final preparations concern capacities and the expectations against them. The following properties are immediate consequences of the definitions and may be used in the sequel without explicit reference. As usual, we denote $\xi_+ = \max(\xi, 0)$ and $\xi_- = \max(-\xi, 0)$.

Lemma 3.4. *Let ν be a capacity and ξ a random variable. Then,*

$$E_{\overline{\nu}}[\xi] = E_{\overline{\nu}}[\xi_+] - E_\nu[\xi_-]$$

and

$$E_{\overline{\nu}}[\xi] = -E_\nu[-\xi],$$

in the sense that one side of the identity exists if and only if the other side exists, and then both sides are equal.

Last, we extend to random variables a result proved for events by Huber and Strassen [17, Theorem 7.1].

Proposition 3.5. *Let ν be a 2-monotone topological capacity, and ξ a random variable. If $E_\nu[\xi] \neq -\infty$, then*

$$E_\nu[\xi] = \inf_{P \in \mathrm{core}(\nu)} E_P[\xi].$$

Proof. The proof will be somewhat clearer if we focus on the equivalent claim that

$$E_{\overline{\nu}}[\xi] = \sup_{P \in \mathrm{core}(\nu)} E_P[\xi].$$

If ξ is bounded, that identity is obtained by combining parts (iii) and (iv) of Lemma A.2 in [5], so it suffices to extend it to unbounded random variables.

The inequality $E_{\overline{\nu}}[\xi] \geq \sup_{P \in \mathrm{core}(\nu)} E_P[\xi]$ is easy, since $\overline{\nu} \geq P$ for each $P \in \mathrm{core}(\nu)$ and the Choquet integral is monotone with respect to the set function, e.g. [12, Proposition 5.2.(iii)].

As regards the converse, begin by assuming $\xi \geq 0$, and set $\xi_n := \min(\xi, n)$. Then

$$E_{\overline{\nu}}[\xi] = \sup_n \int_0^n \overline{\nu}(\xi \geq t)dt = \sup_n E_{\overline{\nu}}[\xi_n].$$

By applying the identity from the bounded case, the right-hand side equals $\sup_n \sup_{P \in \mathrm{core}(\nu)} E_P[\xi_n]$, and so

$$E_{\overline{\nu}}[\xi] \leq \sup_{P \in \mathrm{core}(\nu)} E[\sup_n \xi_n] = \sup_{P \in \mathrm{core}(\nu)} E[\xi].$$

Let us drop the non-negativity requirement on ξ. Since the Choquet integral is translation invariant (e.g. [12, Proposition 5.1.(v)]), $E_{\overline{\nu}}[\xi + n] = E_{\overline{\nu}}[\xi] + n$ for any n, and Lemma 3.4 gives

$$E_{\overline{\nu}}[\xi] = E_{\overline{\nu}}[(\xi + n)_+] - n - E_\nu[(\xi + n)_-]. \tag{2}$$

Since $(\xi + n)_+$ is non-negative,

$$E_{\overline{\nu}}[\xi] = \sup_{P \in \mathrm{core}(\nu)} E_P[(\xi + n)_+] - n - E_\nu[(\xi + n)_-].$$

In turn, the analog of (2) with P replacing ν yields

$$\begin{aligned} E_{\overline{\nu}}[\xi] &= \sup_{P \in \mathrm{core}(\nu)} [E_P[\xi] + E_P[(\xi + n)_-]] - E_\nu[(\xi + n)_-] \\ &\leq \sup_{P \in \mathrm{core}(\nu)} E_P[\xi] + \sup_{P \in \mathrm{core}(\nu)} E_P[(\xi + n)_-] - E_\nu[(\xi + n)_-] \\ &= \sup_{P \in \mathrm{core}(\nu)} E_P[\xi] + E_{\overline{\nu}}[(\xi + n)_-] - E_\nu[(\xi + n)_-]. \qquad (3) \end{aligned}$$

Whenever $t \geq 0$, we have $(\xi + n)_- \geq t$ if and only if $-\xi \geq t + n$. Therefore,

$$E_{\overline{\nu}}[(\xi + n)_-] = \int_0^\infty \overline{\nu}(\{-\xi \geq t + n\})dt = \int_n^\infty \overline{\nu}(\{-\xi \geq t\})dt$$

and, analogously,

$$E_\nu[(\xi + n)_-] = \int_n^\infty \nu(\{-\xi \geq t\})dt.$$

By the hypothesis, $E_\nu[\xi] \neq -\infty$. Taking into account Lemma 3.4 and the fact that $E_{\overline{\nu}}[\xi] \geq E_\nu[\xi]$, we deduce $E_{\overline{\nu}}[-\xi] \neq \infty$ and $E_\nu[-\xi] \neq \infty$. In particular,

$$\int_n^\infty \overline{\nu}(\{-\xi \geq t\})dt \to 0, \qquad \int_n^\infty \nu(\{-\xi \geq t\})dt \to 0.$$

Passing to the limit in (3) as $n \to \infty$, we obtain

$$E_{\overline{\nu}}[\xi] \leq \sup_{P \in \operatorname{core}(\nu)} E_P[\xi].$$

□

4. Proof of the laws of large numbers

Proof of Theorem 1.1.

Proof under assumption (ii).

Step 1. Let Ω be a Polish space, $\mathcal{A}$ its Borel σ-algebra, and ν a topological capacity in $(\Omega, \mathcal{A})$. Proposition 3.2 yields a complete probability space $(\Omega', \mathcal{A}', \mathbb{P})$ and a strongly measurable random set $\Gamma : \Omega' \to \mathcal{F}(\Omega)$ such that ν is the containment functional of Γ,

$$\nu(B) = \mathbb{P}(\Gamma \subset B), \quad B \in \mathcal{B}_\Omega,$$

and, moreover, for every $P \in \operatorname{core}(\nu)$, there exists a measurable selection $\gamma : \Omega' \to \Omega$ with $\gamma \in \Gamma$ $\mathbb{P}$-almost surely and $\mathbb{P}_\gamma = P$. Therefore, for every random variable $\eta : \Omega \to \mathbb{R}$ and every $A \in \mathcal{B}_\mathbb{R}$,

$$\nu(\eta \in A) = \mathbb{P}(\Gamma \subset \eta^{-1}(A)) = \mathbb{P}(\eta(\Gamma) \subset A).$$

Step 2. For each n, set $X_n = \operatorname{cl} \xi_n(\Gamma)$. We need to check that each X_n is a random closed set. Let A be closed. Then the event

$$\{X_n \in \mathcal{F}(A)\} = \{X_n \subset A\} = \{\xi_n(\Gamma) \subset A\} = \{\Gamma \subset \xi_n^{-1}(A)\}$$

is measurable, since ξ_n is a random variable, A is Borel and Γ is strongly measurable.

The X_n are pairwise independent. Indeed, for $i \neq j$ and $A, B \in \mathcal{F}(\mathbb{R})$,

$$\mathbb{P}(X_i \in \mathcal{F}(A), X_j \in \mathcal{F}(B)) = \mathbb{P}(\Gamma \subset \xi_i^{-1}(A) \cap \xi_j^{-1}(B))$$
$$= \nu(\xi_i^{-1}(A) \cap \xi_j^{-1}(B)) = \nu(\xi_i \in A, \xi_j \in B),$$

and the pairwise preindependence of $\{\xi_n\}_n$ yields

$$\nu(\xi_i \in A, \xi_j \in B) = \nu(\xi_i \in A) \cdot \nu(\xi_j \in B) = \ldots = \mathbb{P}(X_i \in \mathcal{F}(A)) \cdot \mathbb{P}(X_j \in \mathcal{F}(B)).$$

One checks similarly that all the X_n share the same containment functional and so are identically distributed.

Since the X_n are pairwise independent and identically distributed random closed sets, Lemma 3.3 applies if we check the condition $E_\mathbb{P}[d(0, X_1)] < \infty$. Consider a probability distribution $P \in \operatorname{core}(\nu)$ and a selection γ of Γ such that $P = \mathbb{P}_\gamma$. Then

$$E_\mathbb{P}[d(0, X_1)] = E_\mathbb{P}[d(0, \xi_1(\Gamma))] \leq E_\mathbb{P}[|\xi_1(\gamma)|] = E_P[|\xi_1|] < \infty,$$

since

$$-\infty < E_\nu[\xi_1] \le E_P[\xi_1] \le E_{\overline{\nu}}[\xi_1] < \infty,$$

because ξ_1 is integrable and $\nu \le P \le \overline{\nu}$.

Therefore, Lemma 3.3 yields a null subset $N \subset \Omega'$ out of which, for every $y \in \mathbb{R}$,

$$d(y, n^{-1}\sum_{i=1}^{n} X_i) \to d(y, E \operatorname{co} X_1). \tag{4}$$

Step 3. Let us prove that, for each $\varepsilon > 0$,

$$\mathbb{P}(S_n(\Gamma) \subset (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon) \text{ eventually}) = 1. \tag{5}$$

From (4),

$$d(E_\nu[\xi] - \varepsilon/4, n^{-1}\sum_{i=1}^{n} X_i) \to d(E_\nu[\xi] - \varepsilon/4, E \operatorname{co}\operatorname{cl}\xi(\Gamma))$$

almost surely. Besides, by Propositions 3.2 and 3.5,

$$E_\nu[\xi] = \inf_{P \in \operatorname{core}(\nu)} E_P[\xi] = \inf_{\gamma \in \Gamma \text{a.s.}} E_{\mathbb{P}}[\xi(\gamma)].$$

Since $E_{\mathbb{P}}[\xi(\gamma)] \in E \operatorname{co}\operatorname{cl}\xi(\Gamma)$ and the latter is closed, we have $E_\nu[\xi] \in E \operatorname{co}\operatorname{cl}\xi(\Gamma)$, and so

$$d(E_\nu[\xi] - \varepsilon/4, n^{-1}\sum_{i=1}^{n} X_i) \le 2 \cdot \frac{\varepsilon}{4}$$

for all sufficiently large n. Therefore,

$$\inf n^{-1}\sum_{i=1} X_i \ge E_\nu[\xi] - \frac{3\varepsilon}{4} > E_\nu[\xi] - \varepsilon.$$

Combining that with an analogous reasoning for the upper part and the completeness of $\mathbb{P}$, we obtain

$$\mathbb{P}(n^{-1}\sum_{i=1}^{n} X_i \subset (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon) \text{ eventually}) \ge \mathbb{P}(N^c) = 1.$$

Finally, the inclusion

$$S_n(\Gamma) = \{n^{-1}\sum_{i=1}^{n} \xi_i(\omega) \mid \omega \in \Gamma\} \subset n^{-1}\sum_{i=1}^{n} X_i$$

yields the desired identity (5).

Step 4: *Proof of part* (a). We begin by observing

$$\begin{aligned} &\nu(S_n \in (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon)) \\ &= \mathbb{P}(\Gamma \subset \{S_n \in (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon)\}) \\ &= \mathbb{P}(S_n(\Gamma) \subset (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon)). \end{aligned}$$

Assume, by contradiction, that the latter does not converge to 1. Then, there would exist a subsequence $\{S_{n_k}\}_k$ such that, for all $k \in \mathbb{N}$,

$$\mathbb{P}(S_{n_k}(\Gamma) \subset (E_\nu[\xi] - \varepsilon, E_{\overline{\nu}}[\xi] + \varepsilon)) < 1 - \varepsilon.$$

But then

$$\begin{aligned}
&\mathbb{P}(S_n(\Gamma) \subset (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon) \text{ eventually})\\
&\quad = \sup_{m\in\mathbb{N}} \mathbb{P}(\bigcap_{n\geq m}\{S_n(\Gamma) \in (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)\})\\
&\quad \leq \sup_{m\in\mathbb{N}} \mathbb{P}(S_{n_m}(\Gamma) \in (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)) \leq 1-\varepsilon,
\end{aligned}$$

a contradiction to (5).

Step 5: *Proof of part* (b). Just observe that

$$\begin{aligned}
&\nu(S_n \in (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon) \text{ eventually})\\
&= \nu(\bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{S_n \in (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)\})\\
&= \mathbb{P}(\Gamma \subset \bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{S_n \in (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)\})\\
&= \mathbb{P}(\exists m\in\mathbb{N} \mid \forall n\geq m, S_n(\Gamma) \subset (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)) = 1.
\end{aligned}$$

Step 6: *Proof of part* (c). Write

$$A = \{E_\nu[\xi] \leq \liminf_n S_n \leq \limsup_n S_n \leq E_{\overline{\nu}}[\xi]\} = \bigcap_{k\in\mathbb{N}} A_k$$

with

$$A_k = \bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{E_\nu[\xi]-k^{-1} < S_n < E_{\overline{\nu}}[\xi]+k^{-1}\}.$$

By part (b) and the continuity of ν for decreasing sequences,

$$\nu(A) = \inf_k \nu(A_k) = 1.$$

Proof under assumption (i). Using Lemma 3.1.(a,d,e), the problem is reduced to the Polish case. Parts (a,b) of Theorem 1.1 then follow from an application of parts (b,c) of Lemma 3.1 to the functions

$$f_n(x_1,x_2,\ldots) = n^{-1}\sum_{i=1}^n x_i$$

and sets $A_n = (E_\nu[\xi]-\varepsilon, E_{\overline{\nu}}[\xi]+\varepsilon)$.

Obtaining Theorem 1.1.(c) from the transfer would be less obvious, but there is no need. Simply note that, in Step 6 above, (c) is proven directly from (b) in a way that carries over to an arbitrary measurable space.

Proof under assumption (iii). What the transfer in Lemma 3.1 needs in order to produce a topological capacity in $\mathbb{R}^{\mathbb{N}}$ is that the preimages $\xi_*^{-1}(G)$, for all $G\subset\mathbb{R}^{\mathbb{N}}$ open, be contained in the class within which ν is continuous for increasing sequences. Since each ξ_n is assumed to be continuous, ξ_* is τ_p-continuous, and therefore the $\xi_*^{-1}(G)$ are open. The assumption that ν is a topological capacity yields

$$\nu_*(G_n) = \nu(\xi_*^{-1}(G_n)) \to \nu(\xi_*^{-1}(G)) = \nu_*(G)$$

whenever G_n are open in $\mathbb{R}^{\mathbb{N}}$ and $G_n \nearrow G$. The remainder of the proof is unaltered. $\square$

5. Some phenomena absent from the additive LLN

Let us begin by presenting a natural family of set functions within which the probabilistic relationship between almost sure convergence and convergence in probability is reversed.

Assume that Ω is a topological space. An *inf-measure* on $\mathcal{B}_\Omega$ is a set function $\nu : \mathcal{B}_\Omega \to [0,1]$ defined by

$$\nu(A) = \inf \phi(A^c),$$

where ϕ is a function on Ω with $\sup \phi(\Omega) = 1$, $\inf \phi(\Omega) = 0$.

In the special case when ϕ is the indicator function of an open set G, we obtain the topological capacity

$$\nu(A) = \begin{cases} 1, & G^c \subset A, \\ 0, & G^c \not\subset A. \end{cases}$$

Since ν is called the unanimity game of G^c in cooperative game theory, and might also be termed a uniform inf-measure distribution in G^c, we will denote it by u_{G^c}.

Inf-measures are completely monotone, and under appropriate conditions on the function ϕ they become capacities or topological capacities. In particular, u_F is a completely monotone topological capacity for each closed F.

Proposition 5.1. *Let ν be an inf-measure. Then, for any random variables $\xi_n, \xi : \Omega \to \mathbb{R}$, if $\xi_n \to \xi$ in probability, then $\xi_n \to \xi$ almost surely. Moreover, if*

$$\nu(d(\xi_n, A) < \varepsilon) \to 1 \text{ for all } \varepsilon > 0$$

for some set $A \subset \mathbb{R}$, then

$$\nu(\mathrm{Clust}(\xi_n) \subset A) = 1,$$

where Clust *denotes the cluster set in* $[-\infty, \infty]$.

Proof. Write

$$A_{n,k} = \{|\xi_n - \xi| < k^{-1}\}, \quad k, n \in \mathbb{N}.$$

By its definition, an inf-measure ν has the property that $\nu(\bigcap_i B_i) = \inf_i \nu(B_i)$. By also using its monotony, we have

$$\inf_{k\in\mathbb{N}} \sup_{m\in\mathbb{N}} \inf_{n\geq m} \nu(A_{n,k}) = \inf_{k\in\mathbb{N}} \sup_{m\in\mathbb{N}} \nu(\bigcap_{n\geq m} A_{n,k})$$

$$\leq \inf_{k\in\mathbb{N}} \nu(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_{n,k}) = \nu(\bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_{n,k}).$$

Since

$$\xi_n \to \xi \text{ in probability} \iff \inf_{k\in\mathbb{N}} \sup_{m\in\mathbb{N}} \inf_{n\geq m} \nu(A_{n,k}) = 1,$$

$$\xi_n \to \xi \text{ a.s.} \iff \nu(\bigcap_{k\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_{n,k}) = 1,$$

the proof of the first part is complete. The second part is similar. ☐

We note that the usual proof that almost sure convergence implies convergence in probability does carry to the setting of capacities.

Proposition 5.2. *Let ν be a capacity. Then, for any random variables $\xi_n, \xi : \Omega \to \mathbb{R}$, if $\xi_n \to \xi$ almost surely then $\xi_n \to \xi$ in probability. Moreover, if*

$$\nu(\mathrm{Clust}(\xi_n) \subset A) = 1,$$

for some set $A \subset \mathbb{R}$, then

$$\nu(d(\xi_n, A) < \varepsilon) \to 1 \text{ for all } \varepsilon > 0.$$

Thus, for capacities which are inf-measures, 'almost surely' and 'in probability' are synonymous. But for topological capacities which are inf-measures, 'in probability' can be strictly stronger than 'almost surely', as Proposition 5.3.(b,c) will show.

We will now examine the quality of the bounds in Theorem 1.1.(c). To do so, we adopt the following definitions. For a given pairwise preindependent sequence $\{\xi_n\}_n$ of random variables identically distributed to ξ, we will call those bounds

- *sharp*, if, maybe replacing $\{\xi_n\}_n$ by other preindependent sequences $\{\xi'_n\}_n$, $\{\xi''_n\}_n$ identically distributed to ξ, the averages S'_n admit a subsequence converging a.s. to the lower bound and the S''_n admit a subsequence converging a.s. to the upper bound.
- *optimal*, if S_n admits subsequences converging a.s. to each of the bounds.
- *trivial*, if they coincide with the infimum and the supremum of the possible values of ξ.
- *subtrivial*, if the possible values of S_n are eventually confined within an interval which is a proper subset of the one given by the bounds.

Triviality and sharpness involve only unidimensional law information, while subtriviality and optimality apply to a specific choice of the whole sequence.

Proposition 5.3. *There exist a completely monotone topological capacity ν and a sequence $\{\xi_n\}_n$ of integrable random variables on $([0,1], \mathcal{B}_{[0,1]})$ such that*

(a) *The ξ_n are preindependent and identically distributed.*
(b) *$S_n \to 0$ almost surely.*
(c) *S_n fails to converge in probability; in fact, $\nu(|S_n| \leq 1/2) = 0$.*
(d) *S_n converges in law to $u_{[0,1/2]}$.*
(e) *The bounds in Theorem* 1.1.*(c) are sharp but subtrivial.*

Proof. Let ν be the uniform inf-measure distribution $u_{[0,1]}$. Set

$$\xi_n(\omega) = \begin{cases} n\omega, & \omega \in [0, n^{-1}], \\ 0, & \omega \in (n^{-1}, 1]. \end{cases}$$

Proof of part (a). Since $\{\xi_n \in A\}$ is a proper subset of $[0,1]$ whenever A does not contain $[0,1]$, the law of each ξ_n is $u_{[0,1]}$. To prove preindependence, we must check the identity

$$\nu(\xi_1 \in A_1, \ldots, \xi_n \in A_n) = \prod_{i=1}^{n} \nu(\xi_i \in A_i).$$

There are only two possibilities: either the right-hand side is 0 or it is 1. In the former case, we have $\nu(\xi_i \in A_i) = 0$ for some i, whence also the left-hand side is 0 by the monotony of ν. In the latter case, since each A_i contains $[0,1]$, we have

$$\nu(\xi_1 \in A_1, \ldots, \xi_n \in A_n) \geq \nu(\bigcap_{i=1}^{n} \{\xi_i \in [0,1]\}) = 1.$$

Proof of part (b). Clearly, $\xi_n \to 0$ pointwise. Since the Cesàro summability method is regular, $S_n \to 0$ pointwise as well.

Proof of part (c). We proceed by contradiction. Fix $\varepsilon \in (0, 1/4)$. Since $\nu = u_{[0,1]}$, it can be the case that $\nu(|S_n - \eta| < \varepsilon) \to 1$ for some random variable η only if $\{|S_n - \eta| < \varepsilon\} = [0, 1]$ for all sufficiently large n. We would then have, from the triangle inequality,

$$\{|S_n - S_m| < 2\varepsilon\} = [0, 1] \text{ for large } n, m.$$

Since $S_m(n^{-1}) = 0$ whenever $m < n$, we deduce that

$$\{|S_n| < 2\varepsilon\} = [0, 1] \text{ for large } n.$$

But

$$S_n(n^{-1}) = \frac{\sum_{i=1}^n \xi_i(n^{-1})}{n} = \frac{\sum_{i=1}^n i/n}{n} = \frac{n+1}{2n} > \frac{1}{2},$$

reaching a contradiction. Therefore, S_n does not converge in probability. The inequality $S_n(n^{-1}) > 1/2$ also yields $\nu(|S_n| \le 1/2) = 0$.

Proof of part (d). We leave it to the reader to check, using induction, that the maximum of each S_n is actually reached at n^{-1}, with value (as shown above) $(n+1)/(2n)$. From the definition of the ξ_n, the range of each S_n is $[0, (n+1)/(2n)]$. For any Borel set $A \subset [0, 1]$ such that $[0, (n+1)/(2n)] \not\subset A$, we have $(S_n)^{-1}(A) \neq [0, 1]$, and so $\nu_{S_n}(A) = 0$. Also, if $[0, (n + 1)/(2n)] \subset A$, clearly $(S_n)^{-1}(A) = [0, 1]$ and $\nu_{S_n}(A) = 1$. Thus each S_n has the $u_{[0,(n+1)/(2n)]}$ law.

There remains to prove the weak convergence of $u_{[0,(n+1)/(2n)]}$ to $u_{[0,1/2]}$. Since $u_{[0,(n+1)/(2n)]} \le u_{[0,1/2]}$,

$$\limsup_n u_{[0,(n+1)/(2n)]}(F) \le u_{[0,1/2]}(F) \text{ for all closed } F.$$

The other half,

$$\liminf_n u_{[0,(n+1)/(2n)]}(G) \ge u_{[0,1/2]}(G) \text{ for all open } G,$$

is equivalent to

'If G is open and $[0, 1/2] \subset G$, then eventually $[0, (n + 1)/(2n)] \subset G$ ',

which holds true.

Proof of part (e). Let us compute the bounds in Theorem 1.1.(c). Since ξ is non-negative,

$$E_\nu[\xi] = \int_0^1 \nu(\xi \ge t)dt = 0.$$

Analogously, with Lemma 3.4,

$$E_{\overline{\nu}}[\xi] = \int_{-1}^0 (1 - \nu(-\xi \ge t))dt = 1.$$

But, for each $\omega \in [0, 1]$,

$$0 \le S_n(\omega) \le \frac{n+1}{2n} < 1,$$

whence the bounds are subtrivial.

We already have one half of the sharpness, since part (b) gives $S_n \to 0$ a.s. We check, like we did with $\{\xi_n\}_n$, that $\{1 - \xi_n\}_n$ are preindependent and identically distributed, with the same law $u_{[0,1]}$. Taking $\xi''_n = 1 - \xi_n$, we have $S''_n \to 1$ a.s., and the proof is complete. $\square$

The next example features a different behaviour.

Proposition 5.4. *There exist a completely monotone topological capacity ν and a sequence $\{\xi_n\}_n$ of integrable random variables on $([0,1], \mathcal{B}_{[0,1]})$ such that*

(a) *The ξ_n are preindependent and identically distributed.*
(b) *S_n converges almost surely and in probability to a non-degenerate random variable.*
(c) *The bounds in Theorem* 1.1.*(c) are sharp but trivial.*
(d) $\overline{\nu}(S_n \to x) = 1$ *for every* $x \in [E_\nu[\xi], E_{\overline{\nu}}[\xi]]$.

Proof. Let $\nu = u_{[0,1]}$, and denote the identity mapping by ξ. This coincides with ξ_1 in the proof of Proposition 5.3. Now define $\xi_n = \xi$ for every $n \in \mathbb{N}$.

Proof of part (a). Trivially, the ξ_n are identically distributed; their common law is $u_{[0,1]}$ by Proposition 5.3.(a). Preindependence of $\{\xi_n\}_n$ amounts to saying that ξ is 'independent' from itself, namely

$$\nu(\xi \in \bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \nu(\xi \in A_i).$$

That can be proved like in Proposition 5.3.(a).

Proof of part (b). Trivial, since $S_n = \xi$ for all $n \in \mathbb{N}$.

Proof of part (c). The bounds were shown in Proposition 5.3.(e) to be 0 and 1, and to be sharp. Since $\xi(0) = 0$ and $\xi(1) = 1$, they are trivial.

Proof of part (d). For each $x \in \mathbb{R}$ we have $\{S_n \not\to x\} = \{x\}^c \neq [0,1]$. Therefore $\nu(S_n \not\to x) = 0$, and equivalently $\overline{\nu}(S_n \to x) = 1$. $\square$

Now we present a pairwise preindependent sequence, still with the same unidimensional marginals as above, whose cluster set is the whole interval given by the LLN bounds.

Proposition 5.5. *There exist a completely monotone topological capacity ν and a sequence $\{\eta_n\}_n$ of integrable random variables on $([0,1], \mathcal{B}_{[0,1]})$ such that*

(a) *The η_n are preindependent and identically distributed.*
(b) $\mathrm{Clust}(S_n) = [E_\nu[\eta_1], E_{\overline{\nu}}[\eta_1]]$.
(c) *The bounds in Theorem* 1.1.*(c) are optimal but trivial.*

Proof. Let $\nu = u_{[0,1]}$. With $\{\xi_n\}_n$ like in the proof of Proposition 5.3, define

$$\eta_n = \begin{cases} \xi_n, & r(n) \text{ odd}, \\ 1 - \xi_n, & r(n) \text{ even}, \end{cases}$$

where $r(1) = 1$, $r(2) = r(3) = 2$, and $r(n)$, for $n \geq 4$, is the natural number r such that

$$2^{(r-1)r/2} + 2^{r(r+1)/2} < n \leq 2^{r(r+1)/2} + 2^{(r+1)(r+2)/2}.$$

Thus,

$$\begin{gathered}\eta_1 = \xi_1,\\ \eta_2 = 1-\xi_2, \eta_3 = 1-\xi_3,\\ \eta_4 = \xi_4, \ldots, \eta_{10} = \xi_{10},\\ \eta_{11} = 1-\xi_{11}, \ldots, \eta_{72} = 1-\xi_{72},\\ \ldots\end{gathered}$$

so that the proportion of occurrences of each type is 1:2 at the end of the second row, 8:2 at the end of the third, 8:64 at the end of the fourth, and so on, each time doubling the disparity.

Denote by S_n the Cesàro averages of η_n (not ξ_n). Recall from Proposition 5.3 that $\xi_n \to 0$ almost surely, so any subsequence has Cesàro averages converging to 0. Accordingly, also any subsequence of $\{1-\xi_n\}_n$ has Cesàro averages converging to 1.

The subsequence S_{n_k} with n_k the largest index in the $(2k-1)$-th row converges almost surely to 0 since, at the end of each odd row, the proportion of indices for which $\eta_n = \xi_n$ is four times larger. Using analogously the even rows, we construct another subsequence converging almost surely to 1. To obtain a subsequence converging to $\lambda \in (0,1)$, just take the intermediate n_k which makes the proportion of indices for which $\eta_n = 1-\xi_n$ for $n \le n_k$ closest to λ (checking this formally is left to the reader).

Proof of part (a). The pairwise preindependence of $\{\eta_n\}_n$ follows easily from that of $\{\xi_n\}_n$. Also, as mentioned in the proof of Proposition 5.3.(e), the ξ_n and $1-\xi_n$ are identically distributed with law $u_{[0,1]}$; in particular, the η_n are identically distributed.

Proof of part (b). We have shown above that $\mathrm{Clust}(\eta_n) = [0,1]$. Since η_1 is ξ_1 from Proposition 5.3, the identity $[E_\nu[\eta_1], E_{\overline{\nu}}[\eta_1]] = [0,1]$ follows from the proof of Proposition 5.3.(e).

Proof of part (c). Since S_n has subsequences converging almost surely to 0 and 1, the bounds are optimal. Again Proposition 5.1.(e) shows that they are trivial. □

The former examples use the space $[0,1]$ with the topological capacity $u_{[0,1]}$. Note that $u_{[0,1]}$ is not a capacity: $A_n = [0, 1-n^{-1}) \cup \{1\}$ converges to $[0,1]$, yet $u_{[0,1]}(A_n) = 0$ for all n. Recalling that finitely additive probabilities fail to be continuous, and fail the law of large numbers as well, it is a natural suspicion that this continuity defect might be the key for such examples to exist. That is partly true in view of the sequence converging almost surely but not in probability (Proposition 5.3). But if we replace $[0,1]$ by $\{0,1\}$ in Propositions 5.4 and 5.5, those examples continue to work. Since continuity is granted in a finite space, such behaviours cannot be connected to its absence.

We mention in passing that the sequence $\{\xi_n = \xi\}_n$ on $\{0,1\}$ is an 'i.i.d.' sequence on a purely atomic space, in contrast to the fact that only non-atomic probability spaces support i.i.d. sequences. Also, $\{A_n\}_n$ above is a sequence of null sets whose union has full measure.

We conclude this section by presenting a source of examples in which the bounds in Theorem 1.1.(c) are not reached. It has been claimed that those bounds are reached with measure 1 under certain conditions. That will be disproved in Section 6 using the following proposition.

Proposition 5.6. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space, and let* $\{\xi, \xi_n\}_n$ *be random variables. Then, for any* $k \in \mathbb{N}$*:*

(a) $\mathbb{P}^k$ *is a completely monotone, null-additive capacity.*
(b) *A subset of* $\{\xi_n\}_n$ *is i.i.d. with respect to* $\mathbb{P}$ *if and only if it is preindependent i.d. with respect to* $\mathbb{P}^k$*.*
(c) ξ *is* $\mathbb{P}$*-integrable if and only if it is* $\mathbb{P}^k$*-integrable, and then*

$$E_{\mathbb{P}^k}[\xi] = E_{\mathbb{P}}[\min\{\xi^{(1)}, \ldots, \xi^{(k)}\}]$$

and

$$E_{\overline{\mathbb{P}^k}}[\xi] = E_{\mathbb{P}}[\max\{\xi^{(1)}, \ldots, \xi^{(k)}\}],$$

where the $\xi^{(i)}$ *are i.i.d. as* ξ*.*

Proof.

Proof of part (a)*.* Consider the k-fold product of copies of the space $(\Omega, \mathcal{A}, \mathbb{P})$ and take as random variables the coordinate projections $\{\pi_i\}_{i=1}^k$. Writing

$$\mathbb{P}^k(A) = \mathbb{P} \otimes \ldots \otimes \mathbb{P}(A \times \ldots \times A) = \mathbb{P} \otimes \ldots \otimes \mathbb{P}(\pi_1 \in A, \ldots, \pi_k \in A)$$
$$= \mathbb{P} \otimes \ldots \otimes \mathbb{P}(\{\pi_i\}_{i=1}^k \subset A) = C_{\{\pi_i\}_{i=1}^k}(A),$$

we see that $\mathbb{P}^k$ is a containment functional and so completely monotone. The fact that it is a capacity follows easily from the properties of $\mathbb{P}$.

As regards its null-additivity, whenever $\mathbb{P}^k(B) = 0$ we have

$$\mathbb{P}^k(A \cup B) = (\mathbb{P}(A) + \mathbb{P}(B \backslash A))^k = \mathbb{P}^k(A).$$

Proof of part (b)*.* This is obvious.

Proof of part (c)*.* Set $\xi_{\min} = \min_{1 \le i \le k} \xi^{(i)}$. Note that

$$\mathbb{P}^k(\xi \ge t) = \mathbb{P}(\{\xi_i\}_{i=1}^k \subset [t, \infty)) = \mathbb{P}(\xi_{\min} \ge t).$$

Since $|\xi_{\min}| \le \sum_{i=1}^k |\xi^{(i)}|$, indeed $\mathbb{P}$-integrability implies the finiteness of $E_{\mathbb{P}^k}[\xi]$, with

$$E_{\mathbb{P}^k}[\xi] = \int_0^\infty \mathbb{P}^k(\{\xi \ge t\})dt - \int_{-\infty}^0 [1 - \mathbb{P}^k(\{\xi \ge t\}]dt$$
$$= \int_0^\infty \mathbb{P}(\xi_{\min} \ge t)dt - \int_{-\infty}^0 [1 - \mathbb{P}(\xi_{\min} \ge t)]dt = E_{\mathbb{P}}[\xi_{\min}].$$

By applying this to $-\xi$, we obtain the corresponding identity for the upper expectation.

Conversely, if ξ is $\mathbb{P}^k$-integrable, since $\mathbb{P}^k \le \mathbb{P} \le \overline{\mathbb{P}^k}$ it follows that $E_{\mathbb{P}}[\xi]$ exists and is finite. □

Therefore, by the Strong LLN for $\mathbb{P}$, for a pairwise preindependent sequence of random variables identically distributed to a $\mathbb{P}^k$-integrable ξ, we have

$$S_n \to E_{\mathbb{P}}[\xi]$$

almost surely and in probability.

6. Discussion

6.1. **Comparison to the literature.** Part (c) of our LLNs is comparable to the LLN of Maccheroni and Marinacci [20] in terms of assumptions on the set function and core proof idea. Their Theorem 1 establishes part (c) of our conclusions in any of the following situations:

(i) ν is a completely monotone capacity in a Polish space, and the ξ_n are bounded (our assumption (i) needs neither Polishness of the space nor boundedness of the random variables).

(ii) ν is a completely monotone topological capacity in a Polish space, and the ξ_n are either simple or bounded continuous functions (our assumption (ii) needs no assumptions on the ξ_n and our assumption (iii) extends the case of continuous, maybe unbounded, functions to any topological space).

Conclusion (a) of Theorem 1.1 is a substantial improvement upon Marinacci's earlier LLN with convergence in probability [21, Theorem 13.(I)]. In that paper, the sample space is compact, the random variables are bounded continuous functions, and the sequence of random variables satisfies a restrictive assumption termed 'regularity' [21, Definition 5]. However, he did not require complete monotony but only 2-monotony, so his theorem is not subsumed by ours.

In recent years, a number of papers have contributed to the LLN for capacities. Their assumptions are too heterogeneous to discuss individually, but an LLN paralleling the classical 'first moment' condition has not been obtained.

Miranda and De Cooman [10] assumed bounded random variables. Their forward irrelevance condition is not a particular case of preindependence. The random variables in Rébillé's paper [30] are bounded continuous functions and Ω is a locally compact space. In that setting, he is able to drop the continuity for increasing sequences. Rébillé in [29] and Cozman in [11] use variance-like functionals, an approach superficially analogous to requiring a 'second moment'. We note that the supremum of all variances in the core easily becomes infinite (L^1-bounded subsets need not be L^2-bounded), so conditions involving those functionals may not be so useful. Finally, Chen [7] demands only 2-monotony but needs an '$(1+\varepsilon)$-th absolute moment', and the independence notion he uses is overly restrictive. For instance, not even $\mathbb{P}$-independent Bernoulli variables are $\mathbb{P}^k$-independent in that sense, for any $k > 1$ (whereas $\mathbb{P}$-independence is the same thing as $\mathbb{P}^k$-preindependence).

6.2. **A counterexample.** Proposition 5.6 contradicts parts (ii, iii) of [21, Theorem 13], which claim

$$\nu(S_n \geq E_\nu[\xi] + \varepsilon) \to 0$$

(and the upper bound analog) under the following assumptions: Ω is a compact topological space, ν is a 2-monotone capacity, the ξ_n are continuous, preindependent, identically distributed and *regular*.

From [21, p. 153], regularity is granted if ν is null-additive. Thus $\mathbb{P}^k$ and a sequence of non-degenerate continuous $\mathbb{P}$-i.i.d. random variables on $[0, 1]$ satisfy those requirements. But, given a sufficiently small $\varepsilon > 0$,

$$\mathbb{P}^k(S_n \geq E_{\mathbb{P}^k}[\xi] + \varepsilon) \geq \mathbb{P}^k(S_n \to E_{\mathbb{P}}[\xi]) = 1$$

by the Strong LLN applied to $\mathbb{P}$.

6.3. **Notions of independence.** The lack of a fully satisfactory generalization of probabilistic independence is a serious problem in non-additive variants of probability theory. Walley and Fine circumvented the problem of giving a definition of independence by constructing a product capacity (such constructions are not unique, see [6, 13, 19] for a few examples) and working only with the coordinate projections in a countable product space. Marinacci's definition by analogy via $\nu(A \cap B) = \nu(A) \cdot \nu(B)$ in [21] (what we have called 'preindependence') has since been often adopted and is general and pragmatic.

A limitation of the analogy definition comes from its very pragmatism, as it is weak enough to allow problematic behaviours like variables 'independent' from themselves. Our term 'preindependent' reflects the fact that a satisfactory notion might have to be stronger.

An intermediate approach to generalizing probabilistic independence would be to transfer the problem to a product space using Lemma 3.1 and then declaring the ξ_n to be independent if the distribution of the vector ξ_* is the product capacity according to some specific definition.

6.4. **Open questions.** 1. Due to the assumption of preindependence, Theorem 1.1 is valid for all possible constructions of a product capacity from the unidimensional marginals. Does choosing an independence definition along the lines suggested above, using some specific product capacity, impose interesting restrictions on the behaviour of S_n?

2. Since the 'almost sure', 'in probability' and 'in law' behaviours can all be different, is there also an LLN 'in law' whose content is not subsumed by the other two statements?

3. Do topological capacities admit a different definition of almost sure convergence which is stronger than convergence in probability, and for which the LLN holds?

4. To what extent can the assumptions on the capacity be weakened while preserving the LLN for all integrable random variables?

References

[1] Marianne Akian, Jean-Pierre Quadrat, and Michel Viot, *Duality between probability and optimization*, Idempotency (Bristol, 1994), Publ. Newton Inst., vol. 11, Cambridge Univ. Press, Cambridge, 1998, pp. 331–353, DOI 10.1017/CBO9780511662508.020. MR1608359

[2] David R. Adams, *Choquet integrals in potential theory*, Publ. Mat. **42** (1998), no. 1, 3–66, DOI 10.5565/PUBLMAT_42198_01. MR1628134 (99g:31010)

[3] Zvi Artstein, *Distributions of random sets and random selections*, Israel J. Math. **46** (1983), no. 4, 313–324, DOI 10.1007/BF02762891. MR730347 (85m:60017)

[4] Robert Badard, *The law of large numbers for fuzzy processes and the estimation problem*, Inform. Sci. **28** (1982), no. 3, 161–178, DOI 10.1016/0020-0255(82)90046-9. MR717299 (84m:60042)

[5] Adriana Castaldo, Fabio Maccheroni, and Massimo Marinacci, *Random correspondences as bundles of random variables*, Sankhyā **66** (2004), no. 3, 409–427. MR2108198 (2005i:28030)

[6] Alain Chateauneuf and Jean-Philippe Lefort, *Some Fubini theorems on product σ-algebras for non-additive measures*, Internat. J. Approx. Reason. **48** (2008), no. 3, 686–696, DOI 10.1016/j.ijar.2007.02.007. MR2437947 (2010g:28034)

[7] Z. Chen. Strong law of large numbers for capacities. Preprint, arXiv:math.PR/1006:0749.

[8] Xavier Fernique, *Processus linéaires, processus généralisés* (French), Ann. Inst. Fourier (Grenoble) **17** (1967), no. fasc. 1, 1–92. MR0221576 (36 #4628)

[9] Gustave Choquet, *Theory of capacities*, Ann. Inst. Fourier, Grenoble **5** (1953), 131–295 (1955). MR0080760 (18,295g)

[10] Gert De Cooman and Enrique Miranda, *Weak and strong laws of large numbers for coherent lower previsions*, J. Statist. Plann. Inference **138** (2008), no. 8, 2409–2432, DOI 10.1016/j.jspi.2007.10.020. MR2414259 (2009c:60065)

[11] Fabio Gagliardi Cozman, *Concentration inequalities and laws of large numbers under epistemic and regular irrelevance*, Internat. J. Approx. Reason. **51** (2010), no. 9, 1069–1084, DOI 10.1016/j.ijar.2010.08.009. MR2735798 (2011m:62028)

[12] Dieter Denneberg, *Non-additive measure and integral*, Theory and Decision Library. Series B: Mathematical and Statistical Methods, vol. 27, Kluwer Academic Publishers Group, Dordrecht, 1994. MR1320048 (96c:28017)

[13] Paolo Ghirardato, *On independence for non-additive measures, with a Fubini theorem*, J. Econom. Theory **73** (1997), no. 2, 261–291, DOI 10.1006/jeth.1996.2241. MR1449020 (98h:28001)

[14] Itzhak Gilboa, *Theory of decision under uncertainty*, Econometric Society Monographs, vol. 45, Cambridge University Press, Cambridge, 2009. MR2493167 (2010i:91002)

[15] Frank Hampel, *Nonadditive probabilities in statistics*, J. Stat. Theory Pract. **3** (2009), no. 1, 11–23, DOI 10.1080/15598608.2009.10411908. MR2667650

[16] Christian Hess, *The distribution of unbounded random sets and the multivalued strong law of large numbers in nonreflexive Banach spaces*, J. Convex Anal. **6** (1999), no. 1, 163–182. MR1713956 (2001c:60007)

[17] Peter J. Huber and Volker Strassen, *Minimax tests and the Neyman-Pearson lemma for capacities*, Ann. Statist. **1** (1973), 251–263. MR0356306 (50 #8776)

[18] Heinz König, *Measure and integration*, Springer-Verlag, Berlin, 1997. An advanced course in basic procedures and applications. MR1633615 (99h:28001)

[19] Gleb A. Koshevoy, *Distributive lattices and products of capacities*, J. Math. Anal. Appl. **219** (1998), no. 2, 427–441, DOI 10.1006/jmaa.1997.5830. MR1606354 (99i:06011)

[20] Fabio Maccheroni and Massimo Marinacci, *A strong law of large numbers for capacities*, Ann. Probab. **33** (2005), no. 3, 1171–1178, DOI 10.1214/009117904000001062. MR2135316 (2007c:28002)

[21] Massimo Marinacci, *Limit laws for non-additive probabilities and their frequentist interpretation*, J. Econom. Theory **84** (1999), no. 2, 145–195, DOI 10.1006/jeth.1998.2479. MR1669509 (2000b:60079)

[22] G. Matheron, *Random sets and integral geometry*, John Wiley & Sons, New York-London-Sydney, 1975. With a foreword by Geoffrey S. Watson; Wiley Series in Probability and Mathematical Statistics. MR0385969 (52 #6828)

[23] Ilya Molchanov, *Theory of random sets*, Probability and its Applications (New York), Springer-Verlag London Ltd., London, 2005. MR2132405 (2006b:60004)

[24] I. S. Molchanov, *Uniform laws of large numbers for the empirical accompanying functionals of random closed sets* (Russian), Teor. Veroyatnost. i Primenen. **32** (1987), no. 3, 612–616. MR914961 (89a:60035)

[25] Tommy Norberg, *Random capacities and their distributions*, Probab. Theory Relat. Fields **73** (1986), no. 2, 281–297, DOI 10.1007/BF00339941. MR855227 (88d:60016)

[26] ShiGe Peng, *Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations*, Sci. China Ser. A **52** (2009), no. 7, 1391–1411, DOI 10.1007/s11425-009-0121-8. MR2520583 (2010h:60103)

[27] Fabrice Philippe, Gabriel Debs, and Jean-Yves Jaffray, *Decision making with monotone lower probabilities of infinite order*, Math. Oper. Res. **24** (1999), no. 3, 767–784, DOI 10.1287/moor.24.3.767. MR1854252 (2002g:91050)

[28] Anatolii Puhalskii, *Large deviations and idempotent probability*, Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics, vol. 119, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR1851048 (2002i:60056)

[29] Yann Rébillé, *Law of large numbers of non-additive measures*, J. Math. Anal. Appl. **352** (2009), no. 2, 872–879, DOI 10.1016/j.jmaa.2008.11.060. MR2501933 (2010e:60014)

[30] Yann Rébillé, *Laws of large numbers for continuous belief measures on compact spaces*, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **17** (2009), no. 5, 685–704, DOI 10.1142/S0218488509006212. MR2642767 (2011c:60005)

[31] Glenn Shafer, *Allocations of probability*, Ann. Probab. **7** (1979), no. 5, 827–839. MR542132 (83d:60008)

[32] G. Shafer (2008). Non-additive probabilities in the work of Bernoulli and Lambert. In: Classic works of the Dempster–Shafer theory of belief functions, 117–182. Springer, Berlin.

[33] Peter Walley and Terrence L. Fine, *Towards a frequentist theory of upper and lower probability*, Ann. Statist. **10** (1982), no. 3, 741–761. MR663429 (84b:60003)

Departamento de Estadística e I.O. y D.M., Escuela Politécnica de Ingeniería, Universidad de Oviedo, E-33071 Gijón, Spain

E-mail address: teranpedro@uniovi.es