

Ergodic Theorems for Lower Probabilities

S.Cerreia-Vioglio, F. Maccheroni, and M. Marinacci

ABSTRACT. We establish an Ergodic Theorem for lower probabilities, a generalization of standard probabilities widely used in applications. As a by-product, we provide a version for lower probabilities of the Strong Law of Large Numbers.

1. Introduction

The purpose of this paper is to state and prove an Ergodic Theorem for lower probabilities: a class of monotone set functions that are not necessarily additive and are widely used in applications where standard additive probabilities turn out to be inadequate (for applications in Economics see Marinacci and Montrucchio [16], for applications in Statistics see Walley [19]).

We consider a measurable space (Ω, \mathcal{F}) , endowed with an $\mathcal{F} \setminus \mathcal{F}$ -measurable transformation $\tau : \Omega \rightarrow \Omega$, and a (continuous) lower probability $\nu : \mathcal{F} \rightarrow [0, 1]$. We study three different notions of invariance for lower probabilities (Definitions 1-3).¹ They are equivalent in the additive case, and so are genuine generalizations to the nonadditive setting of the usual concept of invariance.

The most natural definition of invariance for a lower probability ν (Definition 1) requires that

$$\nu(A) = \nu(\tau^{-1}(A)) \quad \forall A \in \mathcal{F}.$$

It is the weakest form of invariance for the nonadditive case. Nevertheless, it is still possible to derive a version of the Ergodic Theorem (Theorem 2). In other words, if ν is an invariant lower probability, then for each real valued, bounded, and measurable function $f : \Omega \rightarrow \mathbb{R}$ the limit

$$\lim_n \frac{1}{n} \sum_{k=1}^n f \circ \tau^{k-1}(\omega)$$

exists on a set that has measure 1 with respect to ν . If, in addition, ν is ergodic, we are able to provide bounds for such limit in terms of lower and upper Choquet integrals.

Corresponding Author: Fabio Maccheroni <fabio.maccheroni@unibocconi.it>, U. Bocconi, via Sarfatti 25, 20136, Milano, ITALY. The authors gratefully acknowledge the financial support of MIUR (PRIN grant 20103S5RN3_005).

¹In the arXiv working paper version (arXiv:1508.01329), using the techniques of [4] and [5], we study a fourth notion of invariance arising naturally from the literature on Robust Statistics.

Under the stronger notions of invariance (Definitions 2 and 3), the previous result can be strengthened in several ways. First, we develop a nonadditive version of Kingman's super-subadditive ergodic theorem (Theorem 3). Second, when (Ω, \mathcal{F}) is a standard measurable space we can better characterize the limit of time averages (Corollary 2).

As an application of our main result, we establish a nonadditive version of the Strong Law of Large Numbers (Theorem 4) for stationary and ergodic processes.

2. Mathematical Preliminaries

2.1. Set functions. Consider a measurable space (S, Σ) , where S is a non-empty set and Σ is a σ -algebra of subsets of S . Subsets of S are understood to be in Σ even when not stated explicitly. A set function $\nu : \Sigma \rightarrow [0, 1]$ is

- (i) a *capacity* if $\nu(\emptyset) = 0$, $\nu(S) = 1$, and $\nu(A) \leq \nu(B)$ for all A and B such that $A \subseteq B$;
- (ii) *convex* if $\nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B)$ for all A and B ;
- (iii) *additive* if $\nu(A \cup B) = \nu(A) + \nu(B)$ for all disjoint A and B ;
- (iv) *continuous* if $\lim_{n \rightarrow \infty} \nu(A_n) = \nu(A)$ whenever either $A_n \downarrow A$ or $A_n \uparrow A$;
- (v) *continuous at S* if $\lim_{n \rightarrow \infty} \nu(A_n) = \nu(S)$ whenever $A_n \uparrow S$;
- (vi) a *probability* if it is an additive capacity;
- (vii) a *probability measure* if it is a probability which is continuous at S .

We denote by $\Delta(S, \Sigma)$ the set of all probabilities on Σ and by $\Delta^\sigma(S, \Sigma)$ the set of all probability measures on Σ . We endow both sets with the relative topology induced by the weak* topology.² A set function $\nu : \Sigma \rightarrow [0, 1]$ is

- (viii) a *lower probability (measure)* if there exists a compact set $\mathcal{M} \subseteq \Delta^\sigma(S, \Sigma)$ such that

$$\nu(A) = \min_{P \in \mathcal{M}} P(A) \quad \forall A \in \Sigma.$$

Given a capacity ν , its conjugate $\bar{\nu} : \Sigma \rightarrow [0, 1]$ is given by

$$\bar{\nu}(A) = 1 - \nu(A^c) \quad \forall A \in \Sigma.$$

It is immediate to verify that if ν is a lower probability, then

$$\bar{\nu}(A) = \max_{P \in \mathcal{M}} P(A) \quad \forall A \in \Sigma.$$

The core of a capacity ν is the weak* compact set defined by

$$\text{core}(\nu) = \{P \in \Delta(S, \Sigma) : P \geq \nu\},$$

that is, the core is the collection of all probabilities that setwise dominate ν . A capacity $\nu : \Sigma \rightarrow [0, 1]$ is

- (ix) *exact* if $\text{core}(\nu) \neq \emptyset$ and $\nu(A) = \min_{P \in \text{core}(\nu)} P(A)$ for each A .

²Recall that a net $\{P_\alpha\}_{\alpha \in I}$ converges to P , in the weak* topology, if and only if $P_\alpha(A) \rightarrow P(A)$ for all $A \in \Sigma$. The weak* topology is thus the restriction to $\Delta(S, \Sigma)$ of the topology $\sigma(ba(S, \Sigma), B(S, \Sigma))$ where $B(S, \Sigma)$ is the space of all real valued, bounded, and Σ -measurable functions on S and $ba(S, \Sigma)$ is the set of all bounded and finitely additive set functions on Σ . In the case of S being a Polish space and Σ the Borel σ -algebra, the above topology should not be confused with the topology generated by real valued, bounded, and *continuous* functions on S .

If ν is a convex capacity continuous at S , then ν is exact and $\emptyset \neq \text{core}(\nu) \subseteq \Delta^\sigma(S, \Sigma)$ (see [6, Lemma 2 and Theorem 1], [17, Theorem 3.2], and [16, Theorem 4.2 and Theorem 4.7]). In particular, ν is a lower probability where $\mathcal{M} = \text{core}(\nu)$. Conversely, if ν is a lower probability, then ν is exact, continuous at S , and $\mathcal{M} \subseteq \text{core}(\nu) \subseteq \Delta^\sigma(S, \Sigma)$. Nevertheless, being exact does not automatically imply being convex. An exact capacity continuous at S is continuous. Finally, we say that a statement about a random element holds ν -*a.s.* if and only if there exists an event A such that $\nu(A) = 1$ and the statement holds for all $s \in A$.

2.2. Integrals. We denote by $B(S, \Sigma)$ the set of all bounded and Σ -measurable functions from S to \mathbb{R} . A capacity ν induces a functional on $B(S, \Sigma)$ via the Choquet integral, defined for all $f \in B(S, \Sigma)$ by:

$$\int_S f d\nu = \int_0^\infty \nu(\{s \in S : f(s) \geq t\}) dt + \int_{-\infty}^0 [\nu(\{s \in S : f(s) \geq t\}) - \nu(S)] dt$$

where the right hand side integrals are (improper) Riemann integrals. If ν is additive, then the Choquet integral reduces to the standard additive integral. It is also routine to check that $-\int_S f d\nu = \int_S -f d\bar{\nu}$ for all $f \in B(S, \Sigma)$. It is well known (see [6, Lemma 2], [18, Proposition 3], and [16, Theorem 4.7]) that if ν is a convex capacity, then

$$\int_S f d\nu = \min_{P \in \text{core}(\nu)} \int_S f dP \quad \text{and} \quad \int_S f d\bar{\nu} = \max_{P \in \text{core}(\nu)} \int_S f dP \quad \forall f \in B(S, \Sigma).$$

In the rest of the paper, we consider two measurable spaces (S, Σ) . The first one is (Ω, \mathcal{F}) which we interpret as the space where ultimately uncertainty lives. As for the second one, given a real valued and \mathcal{F} -measurable stochastic process $\{f_n\}_{n \in \mathbb{N}}$ on Ω , we consider the space $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$, which we interpret as the space of observations endowed with the σ -algebra generated by the algebra of cylinders \mathcal{C} .

3. Ergodic Theorems

3.1. Invariant Capacities. In this section, we consider a measurable space (Ω, \mathcal{F}) . We also consider a transformation $\tau : \Omega \rightarrow \Omega$ which is \mathcal{F}/\mathcal{F} -measurable. Recall that a probability measure P is (τ) -invariant if and only if

$$(3.1) \quad P(A) = P(\tau^{-1}(A)) \quad \forall A \in \mathcal{F}.$$

We denote by \mathcal{I} the set of all probability measures that satisfy (3.1) and by \mathcal{G} the set of all invariant events of \mathcal{F} , that is, $A \in \mathcal{G}$ if and only if $A \in \mathcal{F}$ and $\tau^{-1}(A) = A$. An invariant probability measure P is said to be ergodic if and only if $P(\mathcal{G}) = \{0, 1\}$. Similarly, we say that a capacity ν is ergodic if and only if $\nu(\mathcal{G}) = \{0, 1\}$. We denote by $\mathcal{S}(\mathcal{I})$ the subset of \mathcal{I} such that

$$\mathcal{S}(\mathcal{I}) = \{P \in \mathcal{I} : P(\mathcal{G}) = \{0, 1\}\}.$$

If (Ω, \mathcal{F}) is a standard measurable space, then it can be checked that $\mathcal{S}(\mathcal{I})$ is the set of strong extreme points of \mathcal{I} (see Dynkin [11]). Finally, following Dunford and Schwartz [10, pp. 723-724] (see also Dowker [8]), we say that a probability measure P is potentially (τ) -invariant if and only if there exists a probability measure $\hat{P} \in \mathcal{I}$ such that

$$P(E) = \hat{P}(E) \quad \forall E \in \mathcal{G}.$$

We denote the set of potentially invariant probability measures by \mathcal{PI} . Next, we propose three notions of (τ -)invariance for a capacity.

DEFINITION 1. *A capacity ν is invariant if and only if for each $A \in \mathcal{F}$*

$$\nu(A) = \nu(\tau^{-1}(A)).$$

DEFINITION 2. *A capacity ν is strongly invariant if and only if for each $A \in \mathcal{F}$*

$$\nu(A \setminus \tau^{-1}(A)) = \bar{\nu}(\tau^{-1}(A) \setminus A) \text{ and } \nu(\tau^{-1}(A) \setminus A) = \bar{\nu}(A \setminus \tau^{-1}(A)).$$

DEFINITION 3. *A lower probability ν is functionally invariant if and only if $\mathcal{M} \subseteq \mathcal{I}$.*

In what follows, we clarify the connection between these notions of invariance.

THEOREM 1. *Let (Ω, \mathcal{F}) be a measurable space and ν a lower probability. Then,*

(i) \implies (ii) \implies (iii) *where*

- (i) ν *is strongly invariant;*
- (ii) ν *is functionally invariant and $\text{core}(\nu) \subseteq \mathcal{I}$;*
- (iii) $\text{core}(\nu) \subseteq \mathcal{I}$.

Moreover, if ν is convex and continuous at Ω , then statements (i), (ii), and (iii) are equivalent.

As a corollary, we obtain that the three definitions coincide with the usual definition of invariance when ν is a probability measure. Finally, if ν is a functionally invariant *lower probability*, then it is immediate to see that ν is invariant.³ In the next section, we show that, if ν is an invariant lower probability, then its core must be contained in \mathcal{PI} .

3.2. Ergodic Theorem. Given the notions of invariance previously discussed, we could then ask ourselves if suitable ergodic theorems can be developed for non-additive probabilities. Given Subsection 3.1, an immediate dichotomy presents. In fact, the notion of invariance of Definition 1 stands separate from, and it is actually weaker than, the other notions of strong and functional invariance, even in the convex case. Theorem 2 only assumes the weak form of invariance of Definition 1. On the other hand, Corollary 2 assumes strong invariance. Strong invariance, paired with the convexity of ν and (Ω, \mathcal{F}) being standard, allows us to provide a sharper version of Theorem 2.

THEOREM 2. *Let (Ω, \mathcal{F}) be a measurable space and ν a lower probability. If ν is invariant, then for each $f \in B(\Omega, \mathcal{F})$ there exists $f^* \in B(\Omega, \mathcal{G})$ such that*

$$\lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) = f^*(\omega) \quad \nu - a.s.$$

Moreover, if ν is ergodic, then

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \leq \int_{\Omega} f^* d\bar{\nu} \right\} \right) = 1.$$

³Since ν is a functionally invariant lower probability, we have that $\mathcal{M} \subseteq \mathcal{I}$ and $\nu(A) = \min_{P \in \mathcal{M}} P(A) = \min_{P \in \mathcal{M}} P(\tau^{-1}(A)) = \nu(\tau^{-1}(A))$ for all $A \in \mathcal{F}$.

As a corollary, we are able to show a necessary property that $\text{core}(\nu)$ of an invariant lower probability ν must satisfy (cf. Theorem 1). Clearly, it is not a characterization since it is well known that there are probability measures that are potentially invariant, but not invariant.

COROLLARY 1. *If a lower probability ν is invariant, then $\text{core}(\nu) \subseteq \mathcal{PI}$.*

As a second corollary, we discuss the ergodic theorem for convex and strongly invariant capacities. Compared to Theorem 2, the following corollary assumes ν convex and a stronger form of invariance that, in turn, yield a limit function f^* which has more properties. These properties naturally generalize the ones found in the Individual Ergodic Theorem of Birkhoff. In this case, convergence of empirical averages is a simple consequence of Birkhoff's theorem applied to each probability in $\text{core}(\nu)$. Nevertheless, the relation between f and f^* in terms of Choquet expectations is not immediate at first sight. A similar comment applies to Theorem 3.

COROLLARY 2. *Let (Ω, \mathcal{F}) be a standard measurable space and ν a convex capacity continuous at Ω . If ν is strongly invariant, then for each $f \in B(\Omega, \mathcal{F})$ there exists $f^* \in B(\Omega, \mathcal{G})$ such that*

$$(3.2) \quad \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) = f^*(\omega) \quad \nu - a.s.$$

Moreover,

- (1) For each $P \in \mathcal{I}$, f^* is a version of the conditional expectation of f given \mathcal{G} .
- (2) $\int_{\Omega} f^* d\nu = \int_{\Omega} f d\nu$.
- (3) If ν is ergodic, then

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f d\nu \leq \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \leq \int_{\Omega} f d\nu \right\} \right) = 1.$$

3.3. Subadditive Ergodic Theorem. Next we turn to a Subadditive Ergodic Theorem for lower probabilities.

DEFINITION 4. *A sequence $\{S_n\}_{n \in \mathbb{N}}$ of \mathcal{F} -measurable random variables is superadditive (resp., subadditive) if and only if*

$$S_{n+k} \geq S_n + S_k \circ \tau^n \quad (\text{resp., } \leq) \quad \forall n, k \in \mathbb{N}.$$

The sequence $\{S_n\}_{n \in \mathbb{N}}$ is additive if and only if it is superadditive and subadditive.

Consider an \mathcal{F} -measurable function $f : \Omega \rightarrow \mathbb{R}$. If we define $\{S_n\}_{n \in \mathbb{N}}$ by

$$(3.3) \quad S_n = \sum_{k=1}^n f \circ \tau^{k-1} \quad \forall n \in \mathbb{N},$$

then we have that $\{S_n\}_{n \in \mathbb{N}}$ is an additive sequence. The opposite is also true, that is, if $\{S_n\}_{n \in \mathbb{N}}$ is additive, then it takes the form (3.3) for some \mathcal{F} -measurable real valued function f . On the other hand, if we take $\{S_n\}_{n \in \mathbb{N}}$ as in (3.3) and we consider $\{|S_n|\}_{n \in \mathbb{N}}$ we obtain a genuine subadditive sequence. Note that if $f \in B(\Omega, \mathcal{F})$, then we also have that there exists $\lambda \in \mathbb{R}$ such that

$$(3.4) \quad -\lambda n \leq S_n(\omega) \leq \lambda n \quad \forall \omega \in \Omega.$$

Similarly, we have that $-\lambda n \leq |S_n| \leq \lambda n$ for all $n \in \mathbb{N}$.

THEOREM 3. *Let (Ω, \mathcal{F}) be a standard measurable space and ν a lower probability. If $\{S_n\}_{n \in \mathbb{N}}$ is either a superadditive or a subadditive sequence that satisfies (3.4) and if ν is functionally invariant, then there exists $f^* \in B(\Omega, \mathcal{G})$ such that*

$$\lim_n \frac{S_n}{n} = f^* \quad \nu - a.s.$$

Moreover,

- (1) *If ν is convex and strongly invariant and $\{S_n\}_{n \in \mathbb{N}}$ superadditive, then $\int_{\Omega} f^* d\nu = \sup_{n \in \mathbb{N}} \int_{\Omega} \frac{S_n}{n} d\nu$.*
- (2) *If ν is convex and strongly invariant and $\{S_n\}_{n \in \mathbb{N}}$ subadditive, then $\int_{\Omega} f^* d\nu = \inf_n \int_{\Omega} \frac{S_n}{n} d\nu$.*
- (3) *If ν is ergodic and $\{S_n\}_{n \in \mathbb{N}}$ is either subadditive or superadditive, then*

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq \lim_n \frac{S_n(\omega)}{n} \leq \int_{\Omega} f^* d\nu \right\} \right) = 1.$$

4. Strong Law of Large Numbers

As an application of Theorem 2, we provide a nonadditive version of the Strong Law of Large Numbers. Before doing so, we need to introduce some notation and terminology. Consider a sequence of real valued, bounded, and measurable random variables $\mathbf{f} = \{f_n\}_{n \in \mathbb{N}} \subseteq B(\Omega, \mathcal{F})$. We denote by \mathcal{T} the tail σ -algebra

$$\bigcap_{k \in \mathbb{N}} \sigma(f_k, f_{k+1}, \dots).$$

DEFINITION 5. *Given a capacity ν , we say that $\mathbf{f} = \{f_n\}_{n \in \mathbb{N}}$ is stationary if and only if for each $n \in \mathbb{N}$, for each $k \in \mathbb{N}_0$, and for each Borel subset B of \mathbb{R}^{k+1}*

$$(4.1) \quad \nu(\{\omega \in \Omega : (f_n(\omega), \dots, f_{n+k}(\omega)) \in B\}) = \nu(\{\omega \in \Omega : (f_{n+1}(\omega), \dots, f_{n+k+1}(\omega)) \in B\}).$$

This notion generalizes the usual notion of stationary stochastic process by allowing for the nonadditivity of the underlying probability measure. Recall that $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ denotes the space of sequences endowed with the σ -algebra generated by the algebra of cylinders. We denote a generic element of $\mathbb{R}^{\mathbb{N}}$ by x . We also consider the shift transformation $\tau : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}}$ defined by

$$\tau(x) = (x_2, x_3, x_4, \dots) \quad \forall x \in \mathbb{R}^{\mathbb{N}}.$$

The sequence $\{f_n\}_{n \in \mathbb{N}}$ induces a natural (measurable) map between (Ω, \mathcal{F}) and $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$, defined by

$$\omega \mapsto \mathbf{f}(\omega) = (f_1(\omega), \dots, f_k(\omega), \dots) \quad \forall \omega \in \Omega.$$

Define $\nu_{\mathbf{f}} : \sigma(\mathcal{C}) \rightarrow [0, 1]$ by

$$\nu_{\mathbf{f}}(C) = \nu(\mathbf{f}^{-1}(C)) \quad \forall C \in \sigma(\mathcal{C}).$$

DEFINITION 6. *Given a capacity ν , we say that $\mathbf{f} = \{f_n\}_{n \in \mathbb{N}}$ is ergodic if and only if $\nu_{\mathbf{f}}$ is ergodic with respect to the shift transformation.*

LEMMA 1. *If ν is a convex capacity continuous at Ω and \mathbf{f} is stationary, then $\nu_{\mathbf{f}}$ is a convex capacity continuous at $\mathbb{R}^{\mathbb{N}}$ which is shift invariant. Moreover, \mathbf{f} is ergodic if $\nu(\mathcal{T}) = \{0, 1\}$.*

This observation is a first step to deduce the Strong Law of Large Numbers as a corollary of Theorem 2 applied to $\nu_{\mathbf{f}}$. In a nutshell, the assumption of stationarity yields that the limit

$$\lim_n \frac{1}{n} \sum_{k=1}^n f_k$$

exists ν -a.s. In order to obtain also a characterization of the limit in terms of the (Choquet) expected value, we further need $\nu_{\mathbf{f}}$ to be ergodic.

THEOREM 4. *Let ν be a convex capacity continuous at Ω . If $\mathbf{f} = \{f_n\}_{n \in \mathbb{N}}$ is stationary and ergodic, then*

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f_1 d\nu \leq \lim_n \frac{1}{n} \sum_{k=1}^n f_k(\omega) \leq \int_{\Omega} f_1 d\bar{\nu} \right\} \right) = 1.$$

We close by observing that there are few but important differences with the non-additive Strong Law of Large Numbers of Marinacci [15] and Maccheroni and Marinacci [14]. In terms of hypotheses, we weaken the assumption of total monotonicity of ν to convexity, while we replace the i.i.d hypothesis of [15] with stationarity and ergodicity. Finally, compared to the main result of [14], we need to assume the continuity of ν . In turn, we obtain that empirical averages exist ν -a.s., a property that was not present in previous works. The bounds for these empirical averages are in terms of the lower and the upper Choquet integrals of the random variable f_1 , as in [15] and [14].

Appendix A. Dynkin Spaces and Nonadditive Probabilities

Consider a standard measurable space (Ω, \mathcal{F}) and a transformation $\tau : \Omega \rightarrow \Omega$ which is $\mathcal{F} \setminus \mathcal{F}$ measurable. Recall that we denote by \mathcal{I} the set of all invariant probability measures. If \mathcal{I} is a nonempty set, then the triple $(\Omega, \mathcal{F}, \mathcal{I})$ forms a Dynkin space.

DEFINITION 7 (Dynkin, 1978). *Let \mathcal{P} be a nonempty subset of $\Delta^{\sigma}(\Omega, \mathcal{F})$ where (Ω, \mathcal{F}) is a separable measurable space. The triple $(\Omega, \mathcal{F}, \mathcal{P})$ is a Dynkin space if and only if there exist a sub- σ -algebra $\mathcal{G} \subseteq \mathcal{F}$, a set $W \in \mathcal{F}$, and a function*

$$\begin{aligned} p : \mathcal{F} \times \Omega &\rightarrow [0, 1] \\ (A, \omega) &\mapsto p(A, \omega) \end{aligned}$$

such that:

- (a) for each $P \in \mathcal{P}$ and $A \in \mathcal{F}$, $p(A, \cdot) : \Omega \rightarrow [0, 1]$ is a version of the conditional probability of A given \mathcal{G} ;
- (b) for each $\omega \in \Omega$, $p(\cdot, \omega) : \mathcal{F} \rightarrow [0, 1]$ is a probability measure;
- (c) $P(W) = 1$ for all $P \in \mathcal{P}$ and $p(\cdot, \omega) \in \mathcal{P}$ for all $\omega \in W$.

It is not hard to check that, given $f \in B(\Omega, \mathcal{F})$, the function $\hat{f} : \Omega \rightarrow \mathbb{R}$, defined by

$$(A.1) \quad \hat{f}(\omega) = \int_{\Omega} f dp(\cdot, \omega) \quad \forall \omega \in \Omega,$$

is a version of the conditional expected value of f given \mathcal{G} for all $P \in \mathcal{P}$, in particular, $\hat{f} \in B(\Omega, \mathcal{G})$ (see also [5, Remark 13]). When (Ω, \mathcal{F}) is a standard measurable space,

if $(\Omega, \mathcal{F}, \mathcal{P}) = (\Omega, \mathcal{F}, \mathcal{I})$, then \mathcal{G} is the set of invariant events. In particular, we can consider $W = \Omega$ (see Gray [12, Theorem 8.3]). The next lemma is ancillary.⁴

LEMMA 2. *Let (Ω, \mathcal{F}) be a measurable space and \mathcal{G} a sub- σ -algebra of \mathcal{F} . If ν is a lower probability such that $\nu(\mathcal{G}) = \{0, 1\}$ and $g \in B(\Omega, \mathcal{G})$, then*

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} g d\nu \leq g(\omega) \leq \int_{\Omega} g d\bar{\nu} \right\} \right) = 1.$$

Appendix B. Proofs

Proof of Theorem 1. Recall that if ν is a lower probability, we have that

$$(B.1) \quad \nu \leq P \leq \bar{\nu} \quad \forall P \in \text{core}(\nu) \subseteq \Delta^{\sigma}(\Omega, \mathcal{F}).$$

(i) implies (ii). Pick $A \in \mathcal{F}$. Since ν is strongly invariant and $\nu \leq \bar{\nu}$, we have $\bar{\nu}(\tau^{-1}(A) \setminus A) = \nu(A \setminus \tau^{-1}(A)) \leq \bar{\nu}(A \setminus \tau^{-1}(A)) = \nu(\tau^{-1}(A) \setminus A) \leq \bar{\nu}(\tau^{-1}(A) \setminus A)$. It follows that $\nu(A \setminus \tau^{-1}(A)) = \bar{\nu}(A \setminus \tau^{-1}(A)) = \bar{\nu}(\tau^{-1}(A) \setminus A) = \nu(\tau^{-1}(A) \setminus A) = k$. By (B.1), we can conclude that $P(A \setminus \tau^{-1}(A)) = k = P(\tau^{-1}(A) \setminus A)$ for all $P \in \text{core}(\nu)$. This implies that $P(A) = P(A \setminus \tau^{-1}(A)) + P(A \cap \tau^{-1}(A)) = P(\tau^{-1}(A) \setminus A) + P(A \cap \tau^{-1}(A)) = P(\tau^{-1}(A))$ for all $P \in \text{core}(\nu)$.

(ii) implies (iii). It is trivial.

Assume that ν is further convex and continuous at Ω . Then, it is a lower probability. We are left to show:

(iii) implies (i). Since ν is convex and $\text{core}(\nu) \subseteq \mathcal{I}$, it follows that

$$\begin{aligned} \nu(A \setminus \tau^{-1}(A)) + \nu(A \cup (\tau^{-1}(A))^c) &= \int_{\Omega} (1_{\Omega} + 1_A - 1_{\tau^{-1}(A)}) d\nu \\ &= \min_{P \in \text{core}(\nu)} \int_{\Omega} (1_{\Omega} + 1_A - 1_{\tau^{-1}(A)}) dP = 1. \end{aligned}$$

Thus, we have that

$$\nu(A \setminus \tau^{-1}(A)) = 1 - \nu(A \cup (\tau^{-1}(A))^c) = 1 - \nu((\tau^{-1}(A) \setminus A)^c) = \bar{\nu}(\tau^{-1}(A) \setminus A).$$

An analogous argument yields that $\nu(\tau^{-1}(A) \setminus A) = \bar{\nu}(A \setminus \tau^{-1}(A))$, proving the statement. \blacksquare

Before proving Theorem 2, we provide an ancillary key result.

THEOREM 5. *Let (Ω, \mathcal{F}) be a measurable space, ν a lower probability, and assume that the family \mathcal{I} of invariant probability measures is not empty. The following statements are equivalent:*

(i) *There exists $\check{P} \in \mathcal{I}$ such that for each $E \in \mathcal{F}$*

$$\check{P}(E) = 1 \implies \lim_k \nu(\tau^{-k}(E)) = 1;$$

(ii) *There exists $\check{P} \in \mathcal{I}$ such that for each $E \in \mathcal{G}$*

$$\check{P}(E) = 1 \implies \nu(E) = 1;$$

(iii) *For each $E \in \mathcal{G}$*

$$P(E) = 1 \quad \forall P \in \mathcal{I} \implies \nu(E) = 1;$$

⁴The proof is standard and available in the arXiv working paper version of this work (arXiv:1508.01329).

(iv) For each $f \in B(\Omega, \mathcal{F})$ there exists $f^* \in B(\Omega, \mathcal{G})$ such that

$$\lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) = f^*(\omega) \quad \nu - a.s.;$$

(v) $\text{core}(\nu) \subseteq \mathcal{PI}$.

Proof. (i) implies (ii). If $E \in \mathcal{G}$, then $\tau^{-k}(E) = E$ for all $k \in \mathbb{N}$, yielding the statement.

(ii) implies (iii). It is trivial.

(iii) implies (iv). Consider $f \in B(\Omega, \mathcal{F})$. Define $f^* : \Omega \rightarrow \mathbb{R}$ by

$$f^*(\omega) = \limsup_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \quad \forall \omega \in \Omega.$$

Define $f_* : \Omega \rightarrow \mathbb{R}$ by considering the lim inf. Since $f \in B(\Omega, \mathcal{F})$, it can be shown that $f^*, f_* \in B(\Omega, \mathcal{G})$. Consider the event

$$\begin{aligned} E &= \left\{ \omega \in \Omega : \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \text{ exists} \right\} = \{ \omega \in \Omega : f^*(\omega) = f_*(\omega) \} \\ &= \left\{ \omega \in \Omega : f^*(\omega) = \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) = f_*(\omega) \right\}. \end{aligned}$$

By Birkhoff's Ergodic Theorem (see [2, Theorem 24.1]), we have that $P(E) = 1$ for all $P \in \mathcal{I}$. By assumption, this yields that $\nu(E) = 1$. Since f was chosen to be generic, the statement follows.

(iv) implies (v). Recall that for each $P \in \text{core}(\nu)$, $P(A) \geq \nu(A)$ for all $A \in \mathcal{F}$. By assumption, we can conclude that for each $P \in \text{core}(\nu)$, for each $f \in B(\Omega, \mathcal{F})$ there exists $f^* \in B(\Omega, \mathcal{G})$ such that

$$\lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) = f^*(\omega) \quad P - a.s.$$

By [13, p. 964] (see also [10, Exercises 31 and 32, pp. 723–724]), it follows that $P \in \mathcal{PI}$.

(v) implies (i). Since ν is a lower probability, it is continuous at Ω and exact. By [16, Theorem 4.2], it follows that there exists a measure $P \in \text{core}(\nu)$ such that for each $A \in \mathcal{F}$, for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(B.2) \quad P(A) < \delta \implies Q(A) < \varepsilon \quad \forall Q \in \text{core}(\nu).$$

It is immediate to show that P is such that for each $A \in \mathcal{F}$

$$(B.3) \quad P(A) = 0 \implies Q(A) = 0 \quad \forall Q \in \text{core}(\nu).$$

Since $P \in \text{core}(\nu) \subseteq \mathcal{PI}$, we have that there exists $\check{P} \in \mathcal{I}$ such that $\check{P}(E) = P(E)$ for all $E \in \mathcal{G}$. Consider $E \in \mathcal{F}$. Assume that $\check{P}(E) = 1$. It follows that $\check{P}(E^c) = 0$. At the same time, define $F_n = \cup_{k=n}^{\infty} \tau^{-k}(E^c)$. Note that $F_n \downarrow F \in \mathcal{G}$. Since $\check{P} \in \mathcal{I}$, it follows that $\check{P}(F) = \lim_n \check{P}(F_n) \leq \check{P}(F_1) \leq \sum_{k=1}^{\infty} \check{P}(\tau^{-k}(E^c)) = 0$. It follows that $\check{P}(F) = 0$, that is, $P(F) = 0$. By (B.3), we have that $Q(F) = 0$ for all $Q \in \text{core}(\nu)$, that is, $\bar{\nu}(F) = 0$. Since ν is a lower probability, $\bar{\nu}$ satisfies the Fatou's property, that is, $0 \leq \limsup_k \bar{\nu}(A_k) \leq \bar{\nu}(\limsup_k A_k)$ for each sequence $\{A_k\}_{k \in \mathbb{N}} \subseteq \mathcal{F}$. This implies that $0 \leq \liminf_k \bar{\nu}(\tau^{-k}(E^c)) \leq$

$\limsup_k \bar{\nu}(\tau^{-k}(E^c)) \leq \bar{\nu}(\limsup_k \tau^{-k}(E^c)) = \bar{\nu}(F) = 0$. We can conclude that $\lim_k \nu(\tau^{-k}(E)) = \lim_k [1 - \bar{\nu}(\tau^{-k}(E^c))] = 1$, proving the statement. \blacksquare

The proof of Theorem 2 uses some of the techniques common in Ergodic Theory (see, e.g., [7, Theorem 7]). Also, note that, given a capacity ν , we have that

$$\text{core}(\nu) = \{P \in \Delta(\Omega, \mathcal{F}) : \bar{\nu} \geq P \geq \nu\} = \{P \in \Delta(\Omega, \mathcal{F}) : \bar{\nu} \geq P\}.$$

Proof of Theorem 2. We first prove that, given the assumptions, $\emptyset \neq \text{core}(\nu) \subseteq \mathcal{PI}$. In particular, this shows that $\mathcal{I} \neq \emptyset$.

Claim: Let ν be a lower probability. If ν is invariant, then $\text{core}(\nu) \subseteq \mathcal{PI}$. In particular, $\mathcal{I} \neq \emptyset$.

Proof of the Claim. Since ν is invariant, $\bar{\nu}$ is invariant. Since ν is a lower probability, ν is continuous at Ω and, in particular, $\emptyset \neq \text{core}(\nu) \subseteq \Delta^\sigma(\Omega, \mathcal{F})$. Fix a Banach-Mazur limit (see [1, pag. 550]) $\phi : l^\infty \rightarrow \mathbb{R}$, that is, a functional from l^∞ to \mathbb{R} such that:

- (1) ϕ is linear;
- (2) ϕ is positive;
- (3) $\phi(x_1, x_2, \dots) = \phi(x_2, x_3, \dots)$ for all $x \in l^\infty$;
- (4) $\phi(x_1, x_2, \dots) = \lim_n x_n$ for all $x \in c$.

Observe that $\nu(A) \leq P(A) \leq \bar{\nu}(A)$ for all $P \in \text{core}(\nu)$ and all $A \in \mathcal{F}$. Fix $P \in \text{core}(\nu)$, define $P_n : \mathcal{F} \rightarrow [0, 1]$ by

$$P_n(A) = \frac{1}{n} \sum_{k=0}^{n-1} P(\tau^{-k}(A)) \quad \forall A \in \mathcal{F}.$$

Note that $P(\tau^{-k}(A)) \leq \bar{\nu}(\tau^{-k}(A)) = \bar{\nu}(A)$ for all $A \in \mathcal{F}$ and for all $k \in \mathbb{N}_0$. Since $\text{core}(\nu)$ is convex, this implies that $\{P_n\}_{n \in \mathbb{N}} \subseteq \text{core}(\nu)$. For each $A \in \mathcal{F}$, define $x_A = (P_1(A), P_2(A), P_3(A), \dots)$. Note that $0 \leq x_A \leq 1_{\mathbb{N}}$, thus, $x_A \in l^\infty$ for all $A \in \mathcal{F}$. Define $\hat{P} : \mathcal{F} \rightarrow [0, 1]$ by $\hat{P}(A) = \phi(x_A)$ for all $A \in \mathcal{F}$. Since ϕ is positive, note that \hat{P} is a well defined positive set function. Next, consider $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$. Since $\{P_n\}_{n \in \mathbb{N}} \subseteq \Delta(\Omega, \mathcal{F})$, it follows that $P_n(A \cup B) = P_n(A) + P_n(B)$ for all $n \in \mathbb{N}$. Since ϕ is linear, this implies that $\hat{P}(A \cup B) = \phi(x_{A \cup B}) = \phi(x_A + x_B) = \phi(x_A) + \phi(x_B) = \hat{P}(A) + \hat{P}(B)$, proving that \hat{P} is additive. Next, consider $A \in \mathcal{G}$. Since $\tau^{-k}(A) = A$ for all $k \in \mathbb{N}_0$, it follows that $P_n(A) = P(A)$ for all $n \in \mathbb{N}$. Since ϕ maps convergent sequences into their limit, we have that $\hat{P}(A) = \phi(x_A) = P(A)$. In particular, this implies that $\hat{P}(\Omega) = 1$ and $\hat{P}(\emptyset) = 0$. Up to now, we have proved that $\hat{P} \in \Delta(\Omega, \mathcal{F})$ and $\hat{P}(A) = P(A)$ for all $A \in \mathcal{G}$. Since $\{P_n\}_{n \in \mathbb{N}} \subseteq \text{core}(\nu)$, we have that $x_A \leq \bar{\nu}(A) 1_{\mathbb{N}}$. Since ϕ is linear and positive, it follows that $\hat{P}(A) = \phi(x_A) \leq \phi(\bar{\nu}(A) 1_{\mathbb{N}}) = \bar{\nu}(A)$ for all $A \in \mathcal{F}$, that is, $\hat{P} \in \text{core}(\nu)$. Since $\text{core}(\nu) \subseteq \Delta^\sigma(\Omega, \mathcal{F})$, we can conclude that $\hat{P} \in \Delta^\sigma(\Omega, \mathcal{F})$. We next show that \hat{P} is invariant. Note that for each $A \in \mathcal{F}$ and for each $n \in \mathbb{N}$

$$\begin{aligned} P_n(\tau^{-1}(A)) &= \frac{1}{n} \sum_{k=0}^{n-1} P(\tau^{-k-1}(A)) = \frac{n+1}{n} \cdot \frac{1}{n+1} \sum_{k=0}^n P(\tau^{-k}(A)) - \frac{1}{n} P(A) \\ &= \frac{n+1}{n} P_{n+1}(A) - \frac{1}{n} P(A). \end{aligned}$$

Define $y = (P_2(A), P_3(A), \dots)$. Define $z = x_{\tau^{-1}(A)} - y \in l^\infty$. Note that $|z_n| = |P_n(\tau^{-1}(A)) - P_{n+1}(A)| \leq \frac{1}{n} |P_{n+1}(A) - P(A)| \leq \frac{2}{n}$ for all $n \in \mathbb{N}$. It follows that $\lim_n z_n = 0$. Since ϕ satisfies properties 3, 1, and 4, we have that $|\hat{P}(\tau^{-1}(A)) - \hat{P}(A)| = |\phi(x_{\tau^{-1}(A)}) - \phi(x_A)| = |\phi(x_{\tau^{-1}(A)}) - \phi(y)| = |\phi(z)| = 0$, proving that \hat{P} is invariant. Given the previous part of the proof, $\hat{P} \in \mathcal{I}$ and $P \in \mathcal{PT}$. Since P was arbitrarily chosen in $\text{core}(\nu)$, it follows that $\mathcal{I} \neq \emptyset$ and $\text{core}(\nu) \subseteq \mathcal{PT}$. \square

By the previous claim and Theorem 5, the main statement follows.

Finally, assume that ν is further ergodic. By Lemma 2 and since $f^* \in B(\Omega, \mathcal{G})$ and ν is an ergodic lower probability, it follows that

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq f^*(\omega) \leq \int_{\Omega} f^* d\bar{\nu} \right\} \right) = 1.$$

Since $\nu \left(\left\{ \omega \in \Omega : f^*(\omega) = \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \right\} \right) = 1$ and ν is a lower probability, this implies that

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq \lim_n \frac{1}{n} \sum_{k=1}^n f(\tau^{k-1}(\omega)) \leq \int_{\Omega} f^* d\bar{\nu} \right\} \right) = 1,$$

proving the statement. \blacksquare

Proof of Corollary 1. It is the proof of the claim contained in the proof of Theorem 2. \blacksquare

We next proceed by proving Theorem 3 and obtaining Corollary 2 as a corollary of this former result. It is also possible to provide a proof of Corollary 2 as a consequence of Theorem 2.

LEMMA 3. *Let $\{S_n\}_{n \in \mathbb{N}}$ be a superadditive (resp., subadditive) sequence that satisfies (3.4) and \mathcal{M} a compact subset of invariant probability measures. If $\{a_n\}_{n \in \mathbb{N}}$ in \mathbb{R} is defined by $a_n = -\min_{P \in \mathcal{M}} \int_{\Omega} S_n dP$ (resp., $a_n = \max_{P \in \mathcal{M}} \int_{\Omega} S_n dP$) for all $n \in \mathbb{N}$, then $\{a_n\}_{n \in \mathbb{N}}$ is subadditive, that is, $a_{n+k} \leq a_n + a_k$ for all $n, k \in \mathbb{N}$.*

Proof. Since $\{S_n\}_{n \in \mathbb{N}}$ satisfies (3.4), $\{S_n\}_{n \in \mathbb{N}} \subseteq B(\Omega, \mathcal{F})$. We just prove the statement for the superadditive case, being the subadditive one similarly proven. If $\{S_n\}_{n \in \mathbb{N}}$ is superadditive and \mathcal{M} is a compact subset of invariant probability measures, then we have that $-a_{n+k} = \min_{P \in \mathcal{M}} \int_{\Omega} S_{n+k} dP \geq \min_{P \in \mathcal{M}} \int_{\Omega} S_n + S_k \circ \tau^n dP \geq \min_{P \in \mathcal{M}} \int_{\Omega} S_n dP + \min_{P \in \mathcal{M}} \int_{\Omega} S_k \circ \tau^n dP = \min_{P \in \mathcal{M}} \int_{\Omega} S_n dP + \min_{P \in \mathcal{M}} \int_{\Omega} S_k dP = -a_n - a_k$ for all $n, k \in \mathbb{N}$, proving the statement. \blacksquare

Proof of Theorem 3. Since ν is a functionally invariant lower probability, we have that $\mathcal{M} \subseteq \mathcal{I}$. Define $\{f_n\}_{n \in \mathbb{N}} \subseteq B(\Omega, \mathcal{F})$ by $f_n = S_n/n$ for all $n \in \mathbb{N}$. It follows that $\hat{f}_n \in B(\Omega, \mathcal{G})$ for all $n \in \mathbb{N}$. Since $\{S_n\}_{n \in \mathbb{N}}$ satisfies (3.4), it follows that there exists $\lambda \in \mathbb{R}$ such that $-\lambda \leq f_n, \hat{f}_n \leq \lambda$ for all $n \in \mathbb{N}$. Define $f^* \in B(\Omega, \mathcal{G})$ by $f^* = \sup_{n \in \mathbb{N}} \hat{f}_n$ (resp., $f^* = \inf_{n \in \mathbb{N}} \hat{f}_n$). By Kingman's Subadditive Ergodic Theorem (see Dudley [9, Theorem 10.7.1] and [12, Theorem 8.4]) and since $W = \Omega$, we have that $f^* = \lim_n \hat{f}_n$ and $P \left(\left\{ \omega \in \Omega : \lim_n \frac{S_n(\omega)}{n} = f^*(\omega) \right\} \right) = 1$ for all $P \in \mathcal{M}$. Since ν is a lower probability, it follows that $\nu \left(\left\{ \omega \in \Omega : \lim_n \frac{S_n(\omega)}{n} = f^*(\omega) \right\} \right) = 1$, proving the main part of the statement.

1. If ν is convex and strongly invariant, then we have that $\text{core}(\nu) \subseteq \mathcal{I}$ and

$$(B.4) \quad \int_{\Omega} f d\nu = \min_{P \in \text{core}(\nu)} \int_{\Omega} f dP \quad \forall f \in B(\Omega, \mathcal{F}).$$

Consider the sequence $\{a_n\}_{n \in \mathbb{N}}$ defined by $a_n = -\int_{\Omega} S_n d\nu$ for all $n \in \mathbb{N}$. By (B.4) and Lemma 3, we have that $\{a_n\}_{n \in \mathbb{N}}$ is subadditive. It follows that (see [12, Lemma 8.3]) $\lim_n \frac{a_n}{n} = \inf_{n \in \mathbb{N}} \frac{a_n}{n}$, that is,

$$(B.5) \quad \lim_n \frac{-a_n}{n} = \sup_{n \in \mathbb{N}} \frac{-a_n}{n}.$$

Recall that $\{\hat{f}_n\}_{n \in \mathbb{N}}$ is uniformly bounded. By Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [3, Theorem 22], (B.5), and the main part of the statement and since $\text{core}(\nu) \subseteq \mathcal{I}$, we have that

$$\begin{aligned} \int_{\Omega} f^* d\nu &= \int_{\Omega} \lim_n \hat{f}_n d\nu = \lim_n \int_{\Omega} \hat{f}_n d\nu = \lim_n \left[\min_{P \in \text{core}(\nu)} \int_{\Omega} \hat{f}_n dP \right] \\ &= \lim_n \left[\min_{P \in \text{core}(\nu)} \int_{\Omega} f_n dP \right] = \lim_n \int_{\Omega} f_n d\nu = \lim_n \frac{\int_{\Omega} S_n d\nu}{n} \\ &= \lim_n \frac{-a_n}{n} = \sup_{n \in \mathbb{N}} \frac{-a_n}{n} = \sup_n \frac{\int_{\Omega} S_n d\nu}{n} = \sup_{n \in \mathbb{N}} \int_{\Omega} f_n d\nu, \end{aligned}$$

proving point 1.

2. If ν is convex and strongly invariant, then we have that $\text{core}(\nu) \subseteq \mathcal{I}$ and

$$(B.6) \quad \int_{\Omega} f d\bar{\nu} = \max_{P \in \text{core}(\nu)} \int_{\Omega} f dP \quad \forall f \in B(\Omega, \mathcal{F}).$$

Consider the sequence $\{a_n\}_{n \in \mathbb{N}}$ defined by $a_n = \int_{\Omega} S_n d\bar{\nu}$. By (B.6) and Lemma 3, we have that $\{a_n\}_{n \in \mathbb{N}}$ is subadditive. It follows that (see [12, Lemma 8.3])

$$(B.7) \quad \lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n}.$$

Recall that $\{\hat{f}_n\}_{n \in \mathbb{N}}$ is uniformly bounded. By [3, Theorem 22], (B.7), and the main part of the statement and since $\text{core}(\nu) \subseteq \mathcal{I}$, we have that

$$\begin{aligned} \int_{\Omega} f^* d\bar{\nu} &= \int_{\Omega} \lim_n \hat{f}_n d\bar{\nu} = \lim_n \int_{\Omega} \hat{f}_n d\bar{\nu} = \lim_n \left[\max_{P \in \text{core}(\nu)} \int_{\Omega} \hat{f}_n dP \right] \\ &= \lim_n \left[\max_{P \in \text{core}(\nu)} \int_{\Omega} f_n dP \right] = \lim_n \int_{\Omega} f_n d\bar{\nu} = \lim_n \frac{\int_{\Omega} S_n d\bar{\nu}}{n} \\ &= \lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} = \inf_n \frac{\int_{\Omega} S_n d\bar{\nu}}{n} = \inf_{n \in \mathbb{N}} \int_{\Omega} f_n d\bar{\nu}, \end{aligned}$$

proving point 2.

3. By Lemma 2 and since ν is ergodic, it follows that

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq f^*(\omega) \leq \int_{\Omega} f^* d\bar{\nu} \right\} \right) = 1.$$

By the initial part of the proof, we have that $\nu \left(\left\{ \omega \in \Omega : f^*(\omega) = \lim_n \frac{S_n(\omega)}{n} \right\} \right) = 1$. Since ν is a lower probability, this implies that

$$\nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f^* d\nu \leq \lim_n \frac{S_n(\omega)}{n} \leq \int_{\Omega} f^* d\bar{\nu} \right\} \right) = 1,$$

proving the statement. \blacksquare

Proof of Corollary 2. Pick $f \in B(\Omega, \mathcal{F})$. It is immediate to see that $\{S_n\}_{n \in \mathbb{N}}$, defined by $S_n = \sum_{k=1}^n f \circ \tau^{k-1}$ for all $n \in \mathbb{N}$, is an additive sequence which satisfies (3.4). Since ν is convex, continuous at Ω , and strongly invariant, it is a functionally invariant lower probability. Define $\{f_n\}_{n \in \mathbb{N}}$ by $f_n = S_n/n$ for all $n \in \mathbb{N}$. Note that $\hat{f}_n = \hat{f}$ for all $n \in \mathbb{N}$. By the proof of Theorem 3, we have that $\lim_n \frac{S_n}{n} = \lim_n \hat{f}_n = \hat{f}$, $\nu - a.s.$, proving the main statement and point 1 where $f^* = \hat{f}$.

2. Since ν is convex and strongly invariant, then we have that $\text{core}(\nu) \subseteq \mathcal{I}$ and $\int_{\Omega} f d\nu = \min_{P \in \text{core}(\nu)} \int_{\Omega} f dP$. By point 1 and since $\text{core}(\nu) \subseteq \mathcal{I}$, we have that $\int_{\Omega} f d\nu = \min_{P \in \text{core}(\nu)} \int_{\Omega} f dP = \min_{P \in \text{core}(\nu)} \int_{\Omega} \hat{f} dP = \int_{\Omega} \hat{f} d\nu$, proving point 2. Note also that $\int_{\Omega} f d\bar{\nu} = \max_{P \in \text{core}(\nu)} \int_{\Omega} f dP = \max_{P \in \text{core}(\nu)} \int_{\Omega} \hat{f} dP = \int_{\Omega} \hat{f} d\bar{\nu}$.

3. By point 3 of Theorem 3 and the proof of point 2, the statement follows. \blacksquare

Proof of Lemma 1. Consider a convex capacity ν and a process \mathbf{f} . It is immediate to see that $\nu_{\mathbf{f}}$ is a convex capacity. Next, consider $\{C_n\}_{n \in \mathbb{N}} \subseteq \sigma(\mathcal{C})$ such that $C_n \uparrow \mathbb{R}^{\mathbb{N}}$. It follows that the sequence $\{A_n\}_{n \in \mathbb{N}}$, defined by $A_n = \mathbf{f}^{-1}(C_n)$ for all $n \in \mathbb{N}$, is such that $A_n \uparrow \Omega$. Since ν is continuous at Ω , we have that $\lim_n \nu_{\mathbf{f}}(C_n) = \lim_n \nu(\mathbf{f}^{-1}(C_n)) = \lim_n \nu(A_n) = 1$, proving that $\nu_{\mathbf{f}}$ is continuous at $\mathbb{R}^{\mathbb{N}}$. Next, consider $C \in \mathcal{C}$. Then, there exist $k \in \mathbb{N}$ and $E \in \mathcal{B}(\mathbb{R}^k)$ such that $C = \{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \dots, x_k) \in E\}$. Note that $\tau^{-1}(C) = \{x \in \mathbb{R}^{\mathbb{N}} : (x_1, x_2, \dots, x_{k+1}) \in \mathbb{R} \times E\}$. Since \mathbf{f} is stationary, it follows that

$$\begin{aligned} \nu_{\mathbf{f}}(C) &= \nu(\mathbf{f}^{-1}(C)) = \nu(\{\omega \in \Omega : (f_1(\omega), \dots, f_k(\omega)) \in E\}) \\ &= \nu(\{\omega \in \Omega : (f_2(\omega), \dots, f_{k+1}(\omega)) \in E\}) \\ &= \nu(\{\omega \in \Omega : (f_1(\omega), f_2(\omega), \dots, f_{k+1}(\omega)) \in \mathbb{R} \times E\}) \\ &= \nu(\mathbf{f}^{-1}(\tau^{-1}(C))) = \nu_{\mathbf{f}}(\tau^{-1}(C)). \end{aligned}$$

Since $C \in \mathcal{C}$ was arbitrarily chosen, it follows that $\mathcal{C} \subseteq \{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}(\tau^{-1}(C))\} \subseteq \sigma(\mathcal{C})$. Since $\nu_{\mathbf{f}}$ is convex and continuous at $\mathbb{R}^{\mathbb{N}}$, we have that $\{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}(\tau^{-1}(C))\}$ is a monotone class. By the Monotone Class Theorem (see [2, Theorem 3.4]), it follows that $\sigma(\mathcal{C}) = \{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}(\tau^{-1}(C))\}$,

that is, $\nu_{\mathbf{f}}$ is shift invariant. Define $\mathcal{H} = \bigcap_{k=1}^{\infty} \sigma(\mathcal{C}_{k+1}^{\infty}) \cap \sigma(\mathcal{C})$.⁵ Note that $\mathbf{f}^{-1}(\mathcal{H}) \subseteq$

\mathcal{T} . Thus, $\nu_{\mathbf{f}}(\mathcal{H}) = \{0, 1\}$ if $\nu(\mathcal{T}) = \{0, 1\}$. Let \mathcal{G} be the σ -algebra of shift invariant events. It is well known that $\mathcal{G} \subseteq \mathcal{H}$. In light of these observations, it is immediate to see that if $\nu(\mathcal{T}) = \{0, 1\}$, then $\nu_{\mathbf{f}}(\mathcal{G}) = \{0, 1\}$, that is, $\nu_{\mathbf{f}}$ is ergodic. \blacksquare

⁵ $\mathcal{C}_{k+1}^{\infty}$ is the class of cylinders such that

$$C = \{x \in \mathbb{R}^{\mathbb{N}} : (x_1, \dots, x_k, x_{k+1}, \dots, x_{k'}) \in \mathbb{R}^k \times E\}$$

where $k' > k$ and $E \in \mathcal{B}(\mathbb{R}^{k'-k})$.

Proof of Theorem 4. By induction and since \mathbf{f} is stationary, it follows that for each $k \in \mathbb{N}$ and for each Borel subset B of \mathbb{R}

$$(B.8) \quad \nu(\{\omega \in \Omega : f_1(\omega) \in B\}) = \nu(\{\omega \in \Omega : f_k(\omega) \in B\}).$$

By (B.8), this implies that for each $k \in \mathbb{N}$ and for each Borel subset B of \mathbb{R}

$$\nu_{\mathbf{f}}(\{x \in \mathbb{R}^{\mathbb{N}} : x_k \in B\}) = \nu(\{\omega \in \Omega : f_k(\omega) \in B\}) = \nu(\{\omega \in \Omega : f_1(\omega) \in B\}).$$

In particular, since $\{f_n\}_{n \in \mathbb{N}} \subseteq B(\Omega, \mathcal{F})$, it follows that there exists $m \in \mathbb{R}$ such that $-m1_{\Omega} \leq f_1 \leq m1_{\Omega}$. If we replace B with $[-m, m]$, then we can conclude that

$$(B.9) \quad \nu_{\mathbf{f}}(\{x \in \mathbb{R}^{\mathbb{N}} : x_k \in [-m, m]\}) = \nu(\{\omega \in \Omega : f_1(\omega) \in [-m, m]\}) = 1 \quad \forall k \in \mathbb{N}.$$

Define $\pi : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ by

$$\pi(x) = \begin{cases} x_1 & \text{if } x_1 \in [-m, m] \\ 0 & \text{otherwise} \end{cases} \quad \forall x \in \mathbb{R}^{\mathbb{N}}.$$

It is immediate to see that $\pi \in B(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$. Note also that

$$(B.10) \quad \bigcap_{k=1}^{\infty} \{x \in \mathbb{R}^{\mathbb{N}} : x_k \in [-m, m]\} \subseteq \bigcap_{n=1}^{\infty} \left\{ x \in \mathbb{R}^{\mathbb{N}} : \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1}(x)) = \frac{1}{n} \sum_{k=1}^n x_k \right\}.$$

By (B.9) and (B.10) and since $\nu_{\mathbf{f}}$ is a convex capacity which is further continuous at $\mathbb{R}^{\mathbb{N}}$, it follows that

$$(B.11) \quad \nu_{\mathbf{f}} \left(\bigcap_{n=1}^{\infty} \left\{ x \in \mathbb{R}^{\mathbb{N}} : \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1}(x)) = \frac{1}{n} \sum_{k=1}^n x_k \right\} \right) = 1.$$

By Theorem 2 and since $\nu_{\mathbf{f}}$ is shift invariant and ergodic, we have that there exists $\pi^* \in B(\mathbb{R}^{\mathbb{N}}, \mathcal{G})$ such that

$$(B.12) \quad \nu_{\mathbf{f}} \left(\left\{ x \in \mathbb{R}^{\mathbb{N}} : \int_{\mathbb{R}^{\mathbb{N}}} \pi^* d\nu_{\mathbf{f}} \leq \liminf_n \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1}(x)) = \pi^*(x) \leq \int_{\mathbb{R}^{\mathbb{N}}} \pi^* d\bar{\nu}_{\mathbf{f}} \right\} \right) = 1.$$

By (B.11) and (B.12) and since $\nu_{\mathbf{f}}$ is convex, we can conclude that

$$(B.13) \quad \nu_{\mathbf{f}} \left(\left\{ x \in \mathbb{R}^{\mathbb{N}} : \int_{\mathbb{R}^{\mathbb{N}}} \pi^* d\nu_{\mathbf{f}} \leq \liminf_n \frac{1}{n} \sum_{k=1}^n x_k = \pi^*(x) \leq \int_{\mathbb{R}^{\mathbb{N}}} \pi^* d\bar{\nu}_{\mathbf{f}} \right\} \right) = 1.$$

Let $E = \{x \in \mathbb{R}^{\mathbb{N}} : \liminf_n \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1}(x)) = \pi^*(x)\}$ and $\pi_n = \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1})$ for all $n \in \mathbb{N}$. By (B.12), we have that $P(E) = 1$ for all $P \in \text{core}(\nu_{\mathbf{f}})$. By construction, $\{1_E \pi_n\}_{n \in \mathbb{N}} \subseteq B(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ is a uniformly bounded sequence which converges pointwise to $1_E \pi^*$. By [3, Theorem 22] and since $\nu_{\mathbf{f}}$ is convex and $P(E) = 1$ for all $P \in \text{core}(\nu_{\mathbf{f}})$, this implies that

$$(B.14) \quad \int_{\mathbb{R}^{\mathbb{N}}} \pi^* d\nu_{\mathbf{f}} = \int_{\mathbb{R}^{\mathbb{N}}} 1_E \pi^* d\nu_{\mathbf{f}} = \int_{\mathbb{R}^{\mathbb{N}}} \liminf_n 1_E \pi_n d\nu_{\mathbf{f}} = \liminf_n \int_{\mathbb{R}^{\mathbb{N}}} 1_E \pi_n d\nu_{\mathbf{f}} = \liminf_n \int_{\mathbb{R}^{\mathbb{N}}} \pi_n d\nu_{\mathbf{f}}.$$

Next, since $\nu_{\mathbf{f}}$ is convex and shift invariant, note that for each $n \in \mathbb{N}$

$$\int_{\mathbb{R}^{\mathbb{N}}} \pi_n d\nu_{\mathbf{f}} = \int_{\mathbb{R}^{\mathbb{N}}} \frac{1}{n} \sum_{k=1}^n \pi(\tau^{k-1}) d\nu_{\mathbf{f}} \geq \frac{1}{n} \sum_{k=1}^n \int_{\mathbb{R}^{\mathbb{N}}} \pi(\tau^{k-1}) d\nu_{\mathbf{f}} = \int_{\mathbb{R}^{\mathbb{N}}} \pi d\nu_{\mathbf{f}}.$$

By (B.14), it follows that $\int_{\mathbb{R}^N} \pi^* d\nu_{\mathbf{f}} \geq \int_{\mathbb{R}^N} \pi d\nu_{\mathbf{f}}$. A similar argument yields that $\int_{\mathbb{R}^N} \pi^* d\bar{\nu}_{\mathbf{f}} \leq \int_{\mathbb{R}^N} \pi d\bar{\nu}_{\mathbf{f}}$. Finally, since $\int_{\mathbb{R}^N} \pi d\nu_{\mathbf{f}} = \int_{\Omega} f_1 d\nu$ and $\int_{\mathbb{R}^N} \pi d\bar{\nu}_{\mathbf{f}} = \int_{\Omega} f_1 d\bar{\nu}$, by (B.13), we can conclude that

$$\begin{aligned} 1 &= \nu_{\mathbf{f}} \left(\left\{ x \in \mathbb{R}^N : \int_{\mathbb{R}^N} \pi d\nu_{\mathbf{f}} \leq \lim_n \frac{1}{n} \sum_{k=1}^n x_k \leq \int_{\mathbb{R}^N} \pi d\bar{\nu}_{\mathbf{f}} \right\} \right) \\ &= \nu \left(\left\{ \omega \in \Omega : \int_{\Omega} f_1 d\nu \leq \lim_n \frac{1}{n} \sum_{k=1}^n f_k(\omega) \leq \int_{\Omega} f_1 d\bar{\nu} \right\} \right), \end{aligned}$$

proving the statement. ■

References

- [1] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer, New York, 2006.
- [2] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley & Sons, New York, 1995.
- [3] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio, Signed Integral Representations of Comonotonic Additive Functionals, *Journal of Mathematical Analysis and Applications*, 385, 895-912, 2012.
- [4] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio, Choquet Integration on Riesz Spaces and Dual Comonotonicity, *Transactions of the American Mathematical Society*, forthcoming.
- [5] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio, Ambiguity and Robust Statistics, *Journal of Economic Theory*, 974-1049, 2013.
- [6] F. Delbaen, Convex Games and Extreme Points, *Journal of Mathematical Analysis and Applications*, 45, 210-233, 1974.
- [7] Y. N. Dowker, Invariant Measure and the Ergodic Theorems, *Duke Mathematical Journal*, 4, 1051-1061, 1947.
- [8] Y. N. Dowker, Finite and σ -Finite Invariant Measures, *Annals of Mathematics*, 4, 595-608, 1951.
- [9] R. M. Dudley, *Real Analysis and Probability*, 2nd ed., Cambridge University Press, Cambridge, 2002.
- [10] N. Dunford and J. T. Schwartz, *Linear Operators; Part I: General Theory*, Wiley, New York, 1958.
- [11] E. B. Dynkin, Sufficient Statistics and Extreme Points, *The Annals of Probability*, 6, 705-730, 1978.
- [12] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 2nd ed., Springer, New York, 2009.
- [13] R. M. Gray and J. C. Kieffer, Asymptotically Mean Stationary Measures, *The Annals of Probability*, 8, 962-973, 1980.
- [14] F. Maccheroni and M. Marinacci, A Strong Law of Large Numbers for Capacities, *The Annals of Probability*, 33, 1171-1178, 2005.
- [15] M. Marinacci, Limit Laws for Non-additive Probabilities and Their Frequentist Interpretation, *Journal of Economic Theory*, 84, 145-195, 1999.
- [16] M. Marinacci and L. Montrucchio, Introduction to the Mathematics of Ambiguity, in *Uncertainty in Economic Theory*, Routledge, New York, 2004.
- [17] D. Schmeidler, Cores of Exact Games, I, *Journal of Mathematical Analysis and Applications*, 40, 214-225, 1972.
- [18] D. Schmeidler, Integral Representation without Additivity, *Proceedings of the American Mathematical Society*, 97, 255-261, 1986.
- [19] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.