INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Ergodic Theorems for Lower Probabilities

Simone Cerreia-Vioglio, Fabio Maccheroni,
and Massimo Marinacci

Working Paper n. 500

This Version: September, 2013

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Ergodic Theorems for Lower Probabilities*

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

Department of Decision Sciences and IGIER, Università Bocconi

September 2013

Abstract

We establish an Ergodic Theorem for lower probabilities, a generalization of standard probabilities widely used in applications. As an application, 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 capacities (that is, monotone set functions not necessarily additive) widely used in applications where standard additive probabilities turn out to be inadequate.

We consider a measurable space $(\Omega, \mathcal{F})$, endowed with an $\mathcal{F}\backslash\mathcal{F}$-measurable transformation $\tau : \Omega \to \Omega$, and a (continuous) lower probability $\nu : \mathcal{F} \to [0, 1]$. We study four different notions of invariance for lower probabilities (Definitions 1-4). 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

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

It is the weakest generalization of invariance to 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 $f : \Omega \to \mathbb{R}$ 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 in terms of upper and lower Choquet integrals for such limit.

Under the stronger notions of invariance (Definitions 2-4), 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 $(\Omega, \mathcal{F})$ is a standard measurable space we can better characterize the limit of the 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.

**AMS 2000 subject classifications.* 28A12, 28D05, 37A05, 37A30, 37A50, 60F15, 60G10. *Key words and phrases.* Capacities, Choquet Integral, Ergodic Theorems, Strong Law of Large Numbers, Invariant Measures. 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) and of the AXA Research Fund.

2 Mathematical Preliminaries

Set functions Consider a measurable space (S, Σ), where S is a nonempty set and Σ is a σ-algebra of subsets of S. Subsets of S are understood to be in Σ even where not stated explicitly.

A set function $\nu : \Sigma \to [0,1]$ 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 $\lim_{n\to\infty} \nu(A_n) = \nu(A)$ whenever either $A_n \downarrow A$ or $A_n \uparrow A$;

(v) *continuous at S* if $\lim_{n\to\infty} \nu(A_n) = 1$ 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.[1] Given $\mathcal{M} \subseteq \Delta^\sigma(S, \Sigma)$, we assume that $\mathcal{M}$ is endowed with the σ-algebra $\mathcal{A}_\mathcal{M}$ which is the smallest σ-algebra that makes the evaluations $P \mapsto P(A)$ measurable for all $A \in \Sigma$.

A set function $\nu : \Sigma \to [0,1]$ is

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

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

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

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

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

$$\bar{\nu}(A) = \max_{P \in \mathcal{M}} P(A) \qquad \forall A \in \Sigma. \tag{1}$$

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

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

That is, the core is the collection of all probability measures that setwise dominate ν.

A set function $\nu : \Sigma \to [0,1]$ is

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

If ν is a convex capacity continuous at S, then ν is exact and $\emptyset \neq \operatorname{core}(\nu) \subseteq \Delta^\sigma(S, \Sigma)$ (see [7, Lemma 2 and Theorem 1], [21, Theorem 3.2], and [18, Theorem 4.2 and Theorem 4.7]). In particular, ν is a lower probability where $\mathcal{M} = \operatorname{core}(\nu)$. Conversely, if ν is a lower probability, then ν is exact, continuous at S, and $\mathcal{M} \subseteq \operatorname{core}(\nu) \subseteq \Delta^\sigma(S, \Sigma)$. Nevertheless, being exact does not automatically imply being convex. An exact capacity continuous at S is continuous.

[1] Recall that a net $\{P_\alpha\}_{\alpha \in I}$ converges to P, in the weak* topology, if and only if $P_\alpha(A) \to P(A)$ for all $A \in \Sigma$. In other words, this is 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.

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:

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

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 $-\int_S f d\nu = \int_\Omega -f d\bar{\nu}$ for all $f \in B(S,\Sigma)$. It is well known ([7, Lemma 2] and [18, Theorem 4.7]) that if ν is a convex capacity, then

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

In the rest of the paper, we consider three measurable spaces (S,Σ). The first one is $(\Omega,\mathcal{F})$ which we interpret as the space where ultimately uncertainty lives. Given a set $\mathcal{P} \subseteq \Delta^\sigma(\Omega,\mathcal{F})$, the second space will be $(\mathcal{P},\mathcal{A}_\mathcal{P})$ which we interpret as the space of all possible probability models equipped with the natural σ-algebra discussed above. Finally, given a real valued and $\mathcal{F}$-measurable stochastic process $\{f_n\}_{n\in\mathbb{N}}$ on Ω, we will consider the space $(\mathbb{R}^\mathbb{N}, \sigma(\mathcal{C}))$, which we will interpret as the space of observations endowed with the σ-algebra generated by the algebra of cylinders $\mathcal{C}$.

Prior and Predictive Capacities Given a set $\mathcal{P} \subseteq \Delta^\sigma(\Omega,\mathcal{F})$, a prior is a capacity $\rho : \mathcal{A}_\mathcal{P} \to [0,1]$. The associated predictive is the capacity $\nu_\rho : \mathcal{F} \to [0,1]$ defined by

$$\nu_\rho(A) = \int_\mathcal{P} P(A) d\rho(P) \qquad \forall A \in \mathcal{F}.$$

If ρ is additive and continuous at $\mathcal{P}$, then ρ is a prior and ν_ρ is a predictive in the traditional sense. We denote capacities that are additive and continuous at $\mathcal{P}$ by π. Given a set $\mathcal{P}$, we denote the set of strong extreme points of $\mathcal{P}$ by $\mathcal{S}(\mathcal{P})$.[2]

3 Ergodic Theorems

3.1 Invariant Capacities

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

$$P(A) = P\left(\tau^{-1}(A)\right) \qquad \forall A \in \mathcal{F}. \tag{2}$$

We denote by $\mathcal{I}$ the set of all probability measures that satisfy (2) 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 $(\Omega,\mathcal{F})$ 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 [12]). Finally, following Dunford and Schwartz [11, pp. 723-724] (see also

[2]Recall that $P \in \mathcal{P}$ is a strong extreme point of $\mathcal{P}$ if and only if the Dirac at P (i.e., δ_P) is the only probability measure $\pi : \mathcal{A}_\mathcal{P} \to [0,1]$ such that $P(A) = \int_\mathcal{P} Q(A) d\pi(Q)$ for each $A \in \mathcal{F}$.

Dowker [9]), 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

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

We denote the set of potentially invariant measures by $\mathcal{PI}$.

Next, we propose four notions of (τ-)invariance for a capacity.

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

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

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

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

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

The fourth definition also describes a procedure in which invariant capacities can be constructed. Such a procedure is a robust Bayesian procedure (see Berger [2] and Shafer [20]).

Definition 4 *A capacity ν is robustly invariant if and only if $\nu = \nu_\rho$ for some convex capacity $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{I})} \to [0,1]$.*

It can be shown that if $(\Omega, \mathcal{F})$ is a standard measurable space and ν is robustly invariant and continuous at Ω, then it is a lower probability. In the next two results, we will clarify the connection between these four notions of invariance.

Proposition 1 *Let $(\Omega, \mathcal{F})$ be a standard measurable space and ν a lower probability. The following statements are true:*

1. *If ν is strongly invariant, then ν is functionally invariant and* $\operatorname{core}(\nu) \subseteq \mathcal{I}$.
2. *If ν is robustly invariant, then ν is functionally invariant.*
3. *If ν is functionally invariant and $\mathcal{M} \in \mathcal{A}_{\mathcal{S}(\mathcal{I})}$, then ν is robustly invariant and ergodic.*
4. *If ν is functionally invariant, then ν is invariant.*

The connection among some of these notions of invariance becomes sharper when ν is convex.

Theorem 1 *Let $(\Omega, \mathcal{F})$ be a standard measurable space and ν a convex capacity continuous at Ω. The following statements are equivalent:*

(i) *ν is strongly invariant;*

(ii) *ν is functionally invariant and* $\operatorname{core}(\nu) \subseteq \mathcal{I}$;

(iii) *ν robustly invariant and* $\operatorname{core}(\nu) \subseteq \mathcal{I}$;

(iv) $\operatorname{core}(\nu) \subseteq \mathcal{I}$.

As a corollary, we obtain that the four definitions coincide with the usual definition of invariance when ν is a probability measure. Thus, all four notions are genuine generalizations to the nonadditive case of the usual notion of invariance. Under additional assumptions on Ω and T, in the additive case, the equivalence between points (i) and (iii) follows by an application of the Choquet-Bishop-de Leeuw theorem (see Phelps [19]). In our case, the equivalence between points (i) and (iii) could be proven by developing a nonadditive version of the Choquet-Bishop-de Leeuw theorem. This can be done by using the techniques contained in Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [5]. Finally, in the next section we will prove 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 nonadditive probabilities. In light of Proposition 1 and Theorem 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, robust, and functional invariance, even in the convex case. Theorem 2 will just assume the weak form of invariance of Definition 1. On the other hand, Corollary 2 will assume strong invariance. Strong invariance, paired with the convexity of ν and $(\Omega, \mathcal{F})$ being standard, will allow us to provide a sharper version of Theorem 2. In Subsection 3.3, with these stronger assumptions, we will also show that a subadditive/superadditive ergodic theorem for nonadditive probabilities can be developed.

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

$$\lim_n \frac{1}{n} \sum_{k=1}^{n} f\left(\tau^{k-1}(\omega)\right) = f^\star(\omega) \qquad \nu - a.s.$$

Moreover, if ν is ergodic, then

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

As a corollary, we are able to show a necessary property that the $\operatorname{core}(\nu)$ of an invariant lower probability ν must satisfy (cf. Proposition 1). Clearly, it is not a characterization since it is well known that there exist probability measures that are potentially invariant but not invariant.[3]

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

As a second corollary, we discuss the ergodic theorem for convex and strongly invariant capacities. The remark following the result clarifies to which extent we can further generalize the result. Compared to Theorem 2, the following corollary assumes ν convex and a stronger form of invariance that, in turn, deliver a limit function $f^\star$ which has more properties. These properties naturally generalize the ones found in the Individual Ergodic Theorem of Birkhoff.

[3] In the case ν is convex, an alternative proof of Theorem 2 can be provided, based on techniques coming from functional analysis and first developed by Eberlein [13] (see also Krengel [15, Chapter 2, Theorem 1.5] and Aliprantis and Border [1, Theorem 20.19]).

Corollary 2 *Let $(\Omega, \mathcal{F})$ 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 $f^\star \in B(\Omega, \mathcal{G})$ such that*

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} f\left(\tau^{k-1}(\omega)\right) = f^\star(\omega) \qquad \nu - a.s. \tag{3}$$

Moreover,

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

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

Remark 1 The results in (3) and point 1 can be both obtained by just imposing that ν is a functionally invariant lower probability. Similarly, point 3. can be obtained by only requiring ν to be an ergodic lower probability which is robustly invariant.

3.3 Subadditive Ergodic Theorem

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

Definition 5 *A sequence $\{S_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}$-measurable random variables is superadditive (resp., subadditive) if and only if*

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

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 $f : \Omega \to \mathbb{R}$. If we define $\{S_n\}_{n\in\mathbb{N}}$ by

$$S_n = \sum_{k=1}^{n} f \circ \tau^{k-1} \qquad \forall n \in \mathbb{N}, \tag{4}$$

then we have that $\{S_n\}_{n\in\mathbb{N}}$ is an additive process. The opposite is also true, that is, if $\{S_n\}_{n\in\mathbb{N}}$ is additive, then it takes the form (4) for some $\mathcal{F}$-measurable real valued function f. On the other hand, if we take $\{S_n\}_{n\in\mathbb{N}}$ as in (4) and we consider $\{|S_n|\}_{n\in\mathbb{N}}$ we obtain a genuine subadditive process. Note that if $f \in B(\Omega, \mathcal{F})$, then we also have that there exists $\lambda \in \mathbb{R}$ such that

$$-\lambda n \leq S_n(\omega) \leq \lambda n \qquad \forall \omega \in \Omega. \tag{5}$$

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

Theorem 3 *Let $(\Omega, \mathcal{F})$ 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 (5) and if ν is functionally invariant, then there exists $f^\star \in B(\Omega, \mathcal{G})$ such that*

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

Moreover,

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

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

4 Strong Law of Large Numbers

4.1 Stationarity and Ergodicity

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

Definition 6 *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 $\mathbb{R}^{k+1}$*

$$\nu(\{\omega \in \Omega : (f_n(\omega), ..., f_{n+k}(\omega)) \in B\}) = \nu(\{\omega \in \Omega : (f_{n+1}(\omega), ..., f_{n+k+1}(\omega)) \in B\}). \tag{6}$$

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

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

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

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

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

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

Definition 7 *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. In fact, it can be shown that 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*

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

4.2 Independence

Here we propose two different notions of independence and show that they imply ergodicity of ν, and so, by Theorem 4, under stationarity, a Strong Law of Large Numbers holds for them.

We need a few notions. Recall that we endowed $\mathbb{R}^{\mathbb{N}}$ with the σ-algebra $\sigma(\mathcal{C})$ generated by the class $\mathcal{C}$ of cylinders, where $C \in \mathcal{C}$ if and only if there exist $k \in \mathbb{N}$ and E in the Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ of $\mathbb{R}^k$ such that

$$C = \left\{ x \in \mathbb{R}^{\mathbb{N}} : (x_1, ..., x_k) \in E \right\}. \tag{7}$$

In this case, we say that the set C is a *cylinder of length* k. We call $\mathcal{C}_k$ the σ-algebra consisting of all cylinders of length k. It is immediate to see that $\mathcal{C} = \bigcup_{k \in \mathbb{N}} \mathcal{C}_k$ is an algebra. We denote by $\mathcal{C}_{k+1}^{\infty}$ the class of cylinders such that

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

where $k' > k$ and $E \in \mathcal{B}(\mathbb{R}^{k'-k})$. Note that $\mathcal{C}_{k+1}^{\infty}$ is an algebra. Finally, we say that two cylinders $C_1, C_2 \in \mathcal{C}$ are *base disjoint* if and only if there exists $k \in \mathbb{N}$ such that $C_1 \in \mathcal{C}_k$ and $C_2 \in \mathcal{C}_{k+1}^{\infty}$.

Definition 8 *A sequence $\{f_n\}_{n \in \mathbb{N}} \subseteq B(\Omega, \mathcal{F})$ consists of independent random variables wrt a capacity ν if and only if, given any two base disjoint cylinders $C_1, C_2 \in \mathcal{C}$, it holds*

$$\nu_{\mathbf{f}}(C_1)\, \nu_{\mathbf{f}}(C_2) \leq \nu_{\mathbf{f}}(C_1 \cap C_2) \leq \nu_{\mathbf{f}}(C_1)\, \bar{\nu}_{\mathbf{f}}(C_2) \leq \bar{\nu}_{\mathbf{f}}(C_1 \cap C_2).$$

We can now state the main result of the subsection:

Proposition 2 *Let ν be a capacity and $\{f_n\}_{n \in \mathbb{N}}$ a sequence of random variables in $B(\Omega, \mathcal{F})$. If $\{f_n\}_{n \in \mathbb{N}}$ is a sequence of independent random variables wrt ν, then $\mathbf{f}$ is ergodic.*

We conclude by comparing our notion of independence with the definition contained in equation (8) below, which Marinacci [17] and Maccheroni and Marinacci [16] studied in the context of totally monotone capacities.[4] Next result clarifies their relation when ν is convex.[5]

Proposition 3 *Let ν be a convex capacity and $\{f_n\}_{n \in \mathbb{N}}$ a sequence of random variables in $B(\Omega, \mathcal{F})$. If for each two base disjoint cylinders $C_1, C_2 \in \mathcal{C}$*

$$\nu_{\mathbf{f}}(C_1 \cap C_2) = \nu_{\mathbf{f}}(C_1)\, \nu_{\mathbf{f}}(C_2), \tag{8}$$

then $\{f_n\}_{n \in \mathbb{N}}$ is a sequence of independent random variables wrt ν.

We close by observing that there are few but important differences with the nonadditive Strong Law of Large Numbers of Marinacci [17] and Maccheroni and Marinacci [16]. In terms of hypotheses, we weaken the assumption of total monotonicity of ν to convexity, while we replace the i.i.d hypothesis of [17] with stationarity and ergodicity.[6] When ν is additive, this constitutes a weakening, in the nonadditive case, the relation is not clear. Finally, compared to the main result of [16], we need to assume the continuity of ν. As a consequence of these assumptions, 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 the same of [17] and [16], that is, they are in terms of the lower and the upper Choquet integrals of the random variables.

[4] The notions of independence studied in the aforementioned two papers deal just with rectangular cylinders. Nevertheless, the generalization to all cylinders is quite natural, particularly, in light of the fact that, in the additive case, the distinction is irrelevant.

[5] When ν is additive, these different formulations of independence coincide. For, in this case $\nu_{\mathbf{f}}$ is additive and $\nu_{\mathbf{f}} = \bar{\nu}_{\mathbf{f}}$.

[6] Maccheroni and Marinacci [16] actually only assume pairwise independence.

A Dynkin Spaces and Nonadditive Probabilities

Consider a standard measurable space $(\Omega, \mathcal{F})$ and a transformation $\tau : \Omega \to \Omega$ which is $\mathcal{F}\backslash\mathcal{F}$ 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, according to the definition of Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [6].

Definition 9 (Dynkin, 1978) *Let $\mathcal{P}$ be a nonempty subset of $\Delta^\sigma(\Omega, \mathcal{F})$ where $(\Omega, \mathcal{F})$ 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*

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

such that:

(a) for each $P \in \mathcal{P}$ and $A \in \mathcal{F}$, $p(A, \cdot) : \Omega \to [0,1]$ is a version of the conditional probability of A given $\mathcal{G}$;

(b) for each $\omega \in \Omega$, $p(\cdot, \omega) : \mathcal{F} \to [0,1]$ 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 $\hat{f} : \Omega \to \mathbb{R}$, defined by

$$\hat{f}(\omega) = \int_\Omega f dp(\cdot, \omega) \qquad \forall \omega \in \Omega, \tag{9}$$

is a version of the conditional expected value of f given $\mathcal{G}$ for all $P \in \mathcal{P}$, in particular, $\hat{f} \in B(\Omega, \mathcal{G})$. When $(\Omega, \mathcal{F})$ is a standard measurable space, $(\Omega, \mathcal{F}, \mathcal{P}) = (\Omega, \mathcal{F}, \mathcal{I})$, then $\mathcal{G}$ is the set of invariant events. In particular, we can consider $W = \Omega$ (see Gray [14, Theorem 8.3]).

We prove an ancillary lemma.

Lemma 2 *Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a Dynkin space and let ν be a lower probability. The following statements are true:*

1. *For each $f \in B(\Omega, \mathcal{F})$ we have that*

$$\int_\Omega f d\nu \le \min_{P \in \text{core}(\nu)} \int_\Omega f dP \le \min_{P \in \mathcal{M}} \int_\Omega f dP \tag{10}$$

where $\mathcal{M}$ is such that $\nu(A) = \min_{P \in \mathcal{M}} P(A)$ for all $A \in \mathcal{F}$.

2. *If $\nu = \nu_\rho$ for some convex capacity $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{P})} \to [0,1]$, then*

$$\int_\Omega f d\nu \le \int_\Omega \hat{f} d\nu \qquad \forall f \in B(\Omega, \mathcal{F}). \tag{11}$$

Proof. 1. If $0 \le f \in B(\Omega, \mathcal{F})$, then $\nu(\{\omega \in \Omega : f(\omega) \ge t\}) \le P(\{\omega \in \Omega : f(\omega) \ge t\})$ for all $P \in \text{core}(\nu)$ and all $t \in [0, \infty)$. It follows that

$$\int_\Omega f d\nu = \int_0^\infty \nu(\{\omega \in \Omega : f(\omega) \ge t\}) dt \le \int_0^\infty P(\{\omega \in \Omega : f(\omega) \ge t\}) dt = \int_\Omega f dP \qquad \forall P \in \text{core}(\nu).$$

Since $\mathcal{M} \subseteq \operatorname{core}(\nu)$, this implies (10). On the other hand, if $0 \not\leq f \in B(\Omega, \mathcal{F})$, then there exists $c \in \mathbb{R}$ such that $f + c1_\Omega \geq 0$. It follows that

$$\int_\Omega f d\nu + c = \int_\Omega (f + c1_\Omega)\, d\nu \leq \min_{P \in \operatorname{core}(\nu)} \int_\Omega (f + c1_\Omega)\, dP = \min_{P \in \operatorname{core}(\nu)} \int_\Omega f dP + c \leq \min_{P \in \mathcal{M}} \int f dP + c,$$

proving the statement.

2. Since $\nu = \nu_\rho$ for some convex capacity $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{P})} \to [0,1]$, we have that

$$\nu(A) = \int_{\mathcal{S}(\mathcal{P})} P(A)\, d\rho(P) = \min_{\pi \in \operatorname{core}(\rho)} \int_{\mathcal{S}(\mathcal{P})} P(A)\, d\pi(P) \qquad \forall A \in \mathcal{F}.$$

Consider $\pi \in \operatorname{core}(\rho)$. By the proof of point 1, it follows that

$$\nu(A) = \int_{\mathcal{S}(\mathcal{P})} P(A)\, d\rho(P) \leq \int_{\mathcal{S}(\mathcal{P})} P(A)\, d\pi(P) = \nu_\pi(A) \qquad \forall A \in \mathcal{F}, \forall \pi \in \operatorname{core}(\rho).$$

Since $\nu_\pi \in \Delta(\Omega, \mathcal{F})$, it follows that

$$\mathcal{M} = \{\nu_\pi : \pi \in \operatorname{core}(\rho)\} \subseteq \operatorname{core}(\nu).$$

It is immediate to see that $\mathcal{M}$ is convex and compact. By the proof of point 1 and [6, Proposition 25] and since ρ is a convex capacity, we have that if $f \in B(\Omega, \mathcal{F})$, then

$$\begin{aligned}
\int_\Omega f d\nu \leq \min_{P \in \mathcal{M}} \int_\Omega f dP &= \min_{\pi \in \operatorname{core}(\rho)} \int_{\mathcal{S}(\mathcal{P})} \left(\int_\Omega f dP \right) d\pi(P) \\
&= \min_{\pi \in \operatorname{core}(\rho)} \int_{\mathcal{S}(\mathcal{P})} \left(\int_\Omega \hat{f} dP \right) d\pi(P) = \int_{\mathcal{S}(\mathcal{P})} \left(\int_\Omega \hat{f} dP \right) d\rho(P) \\
&= \int_\Omega \hat{f} d\nu,
\end{aligned}$$

proving the statement. ■

Lemma 3 *Let $(\Omega, \mathcal{F})$ 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*

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

Proof. We proceed by assuming that $g \geq 0$. Since ν is a capacity such that $\nu(\mathcal{G}) = \{0,1\}$ and $0 \leq g \leq \lambda$ for some $\lambda \in \mathbb{R}$, it follows that the sets

$$\begin{aligned}
&I = \{t \in [0, \infty) : \nu(\{\omega \in \Omega : g(\omega) \geq t\}) = 1\} \\
&\text{and} \\
&J = \{t \in (-\infty, 0] : \nu(\{\omega \in \Omega : -g(\omega) \geq t\}) = 1\}
\end{aligned}$$

are well defined nonempty intervals. I is bounded from above and such that $0 \in I$. J is unbounded from below and such that $-\lambda \in J$. Since ν is a lower probability, ν is continuous. We can conclude that $t^\star = \sup I \in I$ and $t_\star = \sup J \in J$. Since $\nu(\mathcal{G}) = \{0,1\}$, this implies that

$$\begin{aligned}
&\int_\Omega g d\nu = \int_0^\infty \nu(\{\omega \in \Omega : g(\omega) \geq t\})\, dt = \int_0^{\sup I} dt = t^\star \\
&\textit{and} \\
&\int_\Omega -g d\nu = \int_{-\infty}^0 [\nu(\{\omega \in \Omega : -g(\omega) \geq t\}) - \nu(\Omega)]\, dt = \int_{\sup J}^0 (-1)\, dt = t_\star.
\end{aligned}$$

It follows that $t^\star = \int_\Omega g d\nu$ and $t_\star = \int_\Omega -g d\nu$. Since $t^\star \in I$ and $t_\star \in J$, we also have that

$$\nu\left(\{\omega \in \Omega : g(\omega) \geq t^\star\}\right) = 1 = \nu\left(\{\omega \in \Omega : g(\omega) \leq -t_\star\}\right).$$

Since ν is a lower probability, this implies that

$$\nu\left(\left\{\omega \in \Omega : \int_\Omega g d\nu \leq g(\omega) \leq \int_\Omega g d\bar{\nu}\right\}\right) = \nu\left(\{\omega \in \Omega : t^\star \leq g(\omega) \leq -t_\star\}\right) = 1. \tag{12}$$

We next remove the hypothesis that $g \geq 0$. Since $g \in B(\Omega, \mathcal{G})$, it follows that there exists $c \in \mathbb{R}$ such that $g + c1_\Omega \in B_+(\Omega, \mathcal{G})$. By (12) and since the Choquet integral is constant additive, it follows that

$$\begin{aligned}
1 &= \nu\left(\left\{\omega \in \Omega : \int_\Omega (g + c1_\Omega) d\nu \leq g(\omega) + c \leq \int_\Omega (g + c1_\Omega) d\bar{\nu}\right\}\right) \\
&= \nu\left(\left\{\omega \in \Omega : \int_\Omega g d\nu + c \leq g(\omega) + c \leq \int_\Omega g d\bar{\nu} + c\right\}\right) \\
&= \nu\left(\left\{\omega \in \Omega : \int_\Omega g d\nu \leq g(\omega) \leq \int_\Omega g d\bar{\nu}\right\}\right),
\end{aligned}$$

proving the statement. ■

B Proofs of Section 3

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

$$\nu \leq P \leq \bar{\nu} \qquad \forall P \in \operatorname{core}(\nu) \subseteq \Delta^\sigma(\Omega, \mathcal{F}). \tag{13}$$

1. Pick $A \in \mathcal{F}$. Since ν is strongly invariant and $\nu \leq \bar{\nu}$, we have that

$$\bar{\nu}\left(\tau^{-1}(A) \backslash A\right) = \nu\left(A \backslash \tau^{-1}(A)\right) \leq \bar{\nu}\left(A \backslash \tau^{-1}(A)\right) = \nu\left(\tau^{-1}(A) \backslash A\right) \leq \bar{\nu}\left(\tau^{-1}(A) \backslash A\right).$$

It follows that $\nu\left(A \backslash \tau^{-1}(A)\right) = \bar{\nu}\left(A \backslash \tau^{-1}(A)\right) = \bar{\nu}\left(\tau^{-1}(A) \backslash A\right) = \nu\left(\tau^{-1}(A) \backslash A\right) = k$. By (13), we can conclude that $P\left(A \backslash \tau^{-1}(A)\right) = k = P\left(\tau^{-1}(A) \backslash A\right)$ for all $P \in \operatorname{core}(\nu)$. This implies that

$$\begin{aligned}
P(A) &= P\left(A \backslash \tau^{-1}(A)\right) + P\left(A \cap \tau^{-1}(A)\right) = \\
&= P\left(\tau^{-1}(A) \backslash A\right) + P\left(A \cap \tau^{-1}(A)\right) = P\left(\tau^{-1}(A)\right) \qquad \forall P \in \operatorname{core}(\nu),
\end{aligned}$$

proving the statement.

2. By assumption, there exists a convex capacity $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{I})} \to [0,1]$ such that

$$\nu(A) = \int_{\mathcal{S}(\mathcal{I})} P(A) d\rho(P) = \min_{\pi \in \operatorname{core}(\rho)} \int_{\mathcal{S}(\mathcal{I})} P(A) d\pi(P) \qquad \forall A \in \mathcal{F}. \tag{14}$$

Define $\mathcal{M} = \{\nu_\pi : \pi \in \operatorname{core}(\rho)\}$. By [6, Lemma 24] and (14) and since ν is continuous at Ω, we have that ρ is continuous at $\mathcal{S}(\mathcal{I})$, thus, each π in $\operatorname{core}(\rho)$ is a probability measure and $\mathcal{M}$ is a compact subset of $\Delta^\sigma(\Omega, \mathcal{F})$. Moreover, we also have that $\mathcal{M} \subseteq \mathcal{I}$. We can conclude that

$$\nu(A) = \min_{\pi \in \operatorname{core}(\rho)} \int_{\mathcal{S}(\mathcal{I})} P(A) d\pi(P) = \min_{P \in \mathcal{M}} P(A) \qquad \forall A \in \mathcal{F},$$

proving the statement.

3. Fix $\mathcal{M} \in \mathcal{A}_{\mathcal{S}(\mathcal{I})}$. Consider $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{I})} \to [0,1]$ defined by

$$\rho(F) = \begin{cases} 1 & F \supseteq \mathcal{M} \\ 0 & otherwise \end{cases} \qquad \forall F \in \mathcal{A}_{\mathcal{S}(\mathcal{I})}.$$

It is immediate to check that ρ is a convex capacity. By [18, Example 4.4] and since $\mathcal{M} \in \mathcal{A}_{\mathcal{S}(\mathcal{I})}$, we have that

$$\nu(A) = \min_{P \in \mathcal{M}} P(A) = \int_{\mathcal{S}(\mathcal{I})} P(A)\, d\rho(P) \qquad \forall A \in \mathcal{F}.$$

Since $\mathcal{M} \subseteq \mathcal{S}(\mathcal{I})$, observe that $P(A) \in \{0,1\}$ for all $P \in \mathcal{M}$ and for all $A \in \mathcal{G}$. It follows that $\nu(\mathcal{G}) = \{0,1\}$.

4. Since ν is a functionally invariant lower probability, we have that $\mathcal{M} \subseteq \mathcal{I}$ and

$$\nu(A) = \min_{P \in \mathcal{M}} P(A) = \min_{P \in \mathcal{M}} P\left(\tau^{-1}(A)\right) = \nu\left(\tau^{-1}(A)\right) \qquad \forall A \in \mathcal{F},$$

proving that ν is invariant. ■

Proof of Theorem 1. Recall that if ν is convex and continuous at Ω, then it is a lower probability.

(i) implies (ii). It follows by point 1 of Proposition 1.

(ii) implies (iii). We just need to show that ν is robustly invariant. Define $I : B(\Omega, \mathcal{F}) \to \mathbb{R}$ by

$$I(f) = \int_\Omega f d\nu \qquad \forall f \in B(\Omega, \mathcal{F}).$$

By Schmeidler [22] (see also [18]), I is comonotonic additive and supermodular. Since ν is convex, we have that $I(f) = \min_{P \in \text{core}(\nu)} \int_\Omega f dP$ for all $f \in B(\Omega, \mathcal{F})$. Since $\text{core}(\nu) \subseteq \mathcal{I}$, this implies that if $\int_\Omega f_1 dP \geq \int_\Omega f_2 dP$ for all $P \in \mathcal{I}$ then $I(f_1) \geq I(f_2)$. In particular, $I(f) = I\left(\hat{f}\right)$ for all $f \in B(\Omega, \mathcal{F})$. It is also immediate to see that $I(k1_\Omega) = k$ for all $k \in \mathbb{R}$. It follows that I restricted to $B(\Omega, \mathcal{G})$ is normalized, comonotonic additive, supermodular, and such that

$$\int_\Omega f_1 dP \geq \int_\Omega f_2 dP \quad \forall P \in \mathcal{I} \Longrightarrow I(f_1) \geq I(f_2).$$

By [6, Lemma 24 and Proposition 25] and since $(\Omega, \mathcal{F}, \mathcal{I})$ is a Dynkin space, it follows that there exists $\breve{I} : B\left(\mathcal{S}(\mathcal{I}), \mathcal{A}_{\mathcal{S}(\mathcal{I})}\right) \to \mathbb{R}$ such that $\breve{I}$ is normalized, monotone, comonotonic additive, supermodular, and such that

$$I(f) = \breve{I}(\langle f, \cdot \rangle) \qquad \forall f \in B(\Omega, \mathcal{G}).$$

By [22] (see also [18]), it follows that there exists a convex capacity $\rho : \mathcal{A}_{\mathcal{S}(\mathcal{I})} \to [0,1]$ such that

$$I(f) = \int_{\mathcal{S}(\mathcal{I})} \left(\int_\Omega f dP\right) d\rho(P) \qquad \forall f \in B(\Omega, \mathcal{G}). \tag{15}$$

Since $I(f) = I\left(\hat{f}\right)$ for all $f \in B(\Omega, \mathcal{F})$, it follows that (15) holds for all $f \in B(\Omega, \mathcal{F})$. In particular, by picking $f = 1_A$ with $A \in \mathcal{F}$, this shows that ν is robustly invariant.

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

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

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

This implies that $\nu\left(A\backslash\tau^{-1}(A)\right) = 1-\nu\left(A\cup\left(\tau^{-1}(A)\right)^{c}\right) = 1-\nu\left(\left(\tau^{-1}(A)\backslash A\right)^{c}\right) = \bar{\nu}\left(\tau^{-1}(A)\backslash A\right)$. An analogous argument delivers that $\nu\left(\tau^{-1}(A)\backslash A\right) = \bar{\nu}\left(A\backslash\tau^{-1}(A)\right)$, proving the statement. ■

Before proving Theorem 2, we provide an ancillary result.

Theorem 5 *Let $(\Omega,\mathcal{F})$ be a measurable space, ν a lower probability, and $\mathcal{I}$ a nonempty set. The following statements are equivalent:*

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

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

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

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

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

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

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

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} f\left(\tau^{k-1}(\omega)\right)=f^\star \qquad \nu-a.s.;$$

(v) 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 $f^\star:\Omega\to\mathbb{R}$ by

$$f^\star(\omega)=\limsup_n \frac{1}{n}\sum_{k=1}^{n} f\left(\tau^{k-1}(\omega)\right) \qquad \forall\omega\in\Omega.$$

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

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

By Birkhoff's Ergodic Theorem (see [3, 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 $f^\star\in B(\Omega,\mathcal{G})$

$$\lim_n \frac{1}{n}\sum_{k=1}^{n} f\left(\tau^{k-1}(\omega)\right)=f^\star \qquad P-a.s.$$

By [11, Exercises 31 and 32, pag. 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 [18, 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

$$P(A) < \delta \Longrightarrow Q(A) < \varepsilon \qquad \forall Q \in \text{core}(\nu). \tag{16}$$

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

$$P(A) = 0 \Longrightarrow Q(A) = 0 \qquad \forall Q \in \text{core}(\nu). \tag{17}$$

Since $P \in \text{core}(\nu) \subseteq \mathcal{PI}$, we have that there exists $\breve{P} \in \mathcal{I}$ such that $\breve{P}(E) = P(E)$ for all $E \in \mathcal{G}$. Consider $E \in \mathcal{F}$. Assume that $\breve{P}(E) = 1$. It follows that $\breve{P}(E^c) = 0$. 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 $\breve{P} \in \mathcal{I}$, it follows that $\breve{P}(F) = \lim_n \breve{P}(F_n) \leq \breve{P}(F_1) \leq \sum_{k=1}^{\infty} \breve{P}\left(\tau^{-k}(E^c)\right) = 0$. It follows that $\breve{P}(F) = 0$, that is, $P(F) = 0$. By (17), 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, given a sequence $\{A_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}$

$$0 \leq \limsup_k \bar{\nu}(A_k) \leq \bar{\nu}\left(\limsup_k A_k\right).$$

This implies that

$$0 \leq \liminf_k \bar{\nu}\left(\tau^{-k}(E^c)\right) \leq \limsup_k \bar{\nu}\left(\tau^{-k}(E^c)\right) \leq \bar{\nu}\left(\limsup_k \tau^{-k}(E^c)\right) = \bar{\nu}(F) = 0.$$

We can conclude that

$$\lim_k \nu\left(\tau^{-k}(E)\right) = \lim_k \left[1 - \bar{\nu}\left(\tau^{-k}(E^c)\right)\right] = 1,$$

proving the statement. ■

The proof of Theorem 2 uses some of the techniques common in Ergodic Theory (see, e.g., [8, 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, $\emptyset \neq$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]) $\phi : l^{\infty} \to \mathbb{R}$, that is, a linear functional from l^{∞} to $\mathbb{R}$ such that:

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

Observe that

$$\nu(A) \leq P(A) \leq \bar{\nu}(A) \qquad \forall P \in \text{core}(\nu), \forall A \in \mathcal{F}.$$

Fix $P \in \text{core}(\nu)$, define $P_n : \mathcal{F} \to [0,1]$ by

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

Note that $P\left(\tau^{-k}(A)\right) \leq \bar{\nu}\left(\tau^{-k}(A)\right) = \bar{\nu}(A)$ 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 $x_A = (P_1(A), P_2(A), P_3(A), ...)$. Note that $0 \leq x_A \leq 1_{\mathbb{N}}$, thus, $x_A \in l^\infty$ for all $A \in \mathcal{F}$. Define $\hat{P} : \mathcal{F} \to [0,1]$ by

$$\hat{P}(A) = \phi(x_A) \qquad \forall A \in \mathcal{F}.$$

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 $k \in \mathbb{N}$. 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

$$\hat{P}(A) = \phi(x_A) \leq \phi(\bar{\nu}(A) 1_{\mathbb{N}}) = \bar{\nu}(A) \qquad \forall A \in \mathcal{F},$$

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 $n \in \mathbb{N}$

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

Define $y = (P_2(A), P_3(A), ...)$. Define $z = x_{\tau^{-1}(A)} - y \in l^\infty$. Note that

$$|z_n| = \left|P_n\left(\tau^{-1}(A)\right) - P_{n+1}(A)\right| \leq \frac{1}{n}\left|P_{n+1}(A) - P(A)\right| \leq \frac{2}{n} \qquad \forall n \in \mathbb{N}.$$

It follows that $\lim_n z_n = 0$. Since ϕ is shift invariant, linear, and it maps convergent sequences into their limit, we have that

$$\begin{aligned}
\left|\hat{P}\left(\tau^{-1}(A)\right) - \hat{P}(A)\right| &= \left|\phi\left(x_{\tau^{-1}(A)}\right) - \phi(x_A)\right| = \left|\phi\left(x_{\tau^{-1}(A)}\right) - \phi(y)\right| \\
&= \left|\phi\left(x_{\tau^{-1}(A)} - y\right)\right| = |\phi(z)| = 0,
\end{aligned}$$

proving that $\hat{P}$ is invariant. Given the previous part of the proof, $\hat{P} \in \mathcal{I}$ and $P \in \mathcal{PI}$. Since P was arbitrarily chosen in $\text{core}(\nu)$, it follows that $\mathcal{I} \neq \emptyset$ and $\text{core}(\nu) \subseteq \mathcal{PI}$. □

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

Finally, assume that ν is further ergodic. By Lemma 3 and since $f^\star \in B(\Omega, \mathcal{G})$ and ν is an ergodic lower probability, it follows that

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

By the initial part of the proof, we also have that

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

Since ν is a lower probability, this implies that

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

proving the statement. ■

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

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. By Theorem 2, the extra assumption of $(\Omega,\mathcal{F})$ being standard yields the extra property that $f^{\star}$ can be chosen to be the regular conditional expectation of f. Convexity and strong invariance imply that $\text{core}(\nu)\subseteq\mathcal{I}$. This yields that $\int_\Omega f^{\star}d\nu=\int_\Omega fd\nu$ as well as $\int_\Omega f^{\star}d\bar{\nu}=\int_\Omega fd\bar{\nu}$. This, in turn, delivers a sharper result under the assumption of ν being ergodic.

Lemma 4 *Let $\{S_n\}_{n\in\mathbb{N}}$ be a superadditive (resp., subadditive) sequence that satisfies (5) and $\mathcal{M}$ a compact subset of invariant probability measures. If $\{a_n\}_{n\in\mathbb{N}}\subseteq\mathbb{R}$ is defined by $a_n=-\min_{P\in\mathcal{M}}\int_\Omega S_ndP$ (resp., $a_n=\max_{P\in\mathcal{M}}\int_\Omega S_ndP$) for all $n\in\mathbb{N}$, then $\{a_n\}_{n\in\mathbb{N}}$ is subadditive, that is, $a_{n+k}\le a_n+a_k$ for all $n,k\in\mathbb{N}$.*

Proof. Since $\{S_n\}_{n\in\mathbb{N}}$ satisfies (5), $\{S_n\}_{n\in\mathbb{N}}\subseteq B(\Omega,\mathcal{F})$. We just prove 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 for each $n,k\in\mathbb{N}$

$$\begin{aligned}-a_{n+k}&=\min_{P\in\mathcal{M}}\int_\Omega S_{n+k}dP\ge\min_{P\in\mathcal{M}}\int_\Omega S_n+S_k\circ\tau^ndP\\&\ge\min_{P\in\mathcal{M}}\int_\Omega S_ndP+\min_{P\in\mathcal{M}}\int_\Omega S_k\circ\tau^ndP\\&=\min_{P\in\mathcal{M}}\int_\Omega S_ndP+\min_{P\in\mathcal{M}}\int_\Omega S_kdP\\&=-a_n-a_k,\end{aligned}$$

proving the statement. ■

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 (5), it follows that there exists $\lambda\in\mathbb{R}$ such that $-\lambda\le f_n,\hat{f}_n\le\lambda$ for all $n\in\mathbb{N}$. Define $f^{\star}\in B(\Omega,\mathcal{G})$ by $f^{\star}=\sup_{n\in\mathbb{N}}\hat{f}_n$ (resp., $f^{\star}=\inf_{n\in\mathbb{N}}\hat{f}_n$). By Kingman's Subadditive Ergodic Theorem (see Dudley [10, Theorem 10.7.1] and [14, Theorem 8.4]) and since $W=\Omega$, we have that $f^{\star}=\lim_n\hat{f}_n$ and

$$P\left(\left\{\omega\in\Omega:\lim_n\frac{S_n(\omega)}{n}=f^{\star}(\omega)\right\}\right)=1\qquad\forall P\in\mathcal{M}.$$

Since ν is a lower probability, it follows that

$$\nu\left(\left\{\omega\in\Omega:\lim_n\frac{S_n(\omega)}{n}=f^{\star}(\omega)\right\}\right)=1,$$

proving the main part of the statement.

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

$$\int_\Omega f d\nu = \min_{P \in \operatorname{core}(\nu)} \int_\Omega f dP \qquad \forall f \in B(\Omega, \mathcal{F}). \tag{18}$$

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 (18) and Lemma 4, we have that $\{a_n\}_{n\in\mathbb{N}}$ is subadditive. It follows that (see [14, Lemma 8.3]) $\lim_n \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n}$, that is,

$$\lim_n \frac{-a_n}{n} = \sup_{n\in\mathbb{N}} \frac{-a_n}{n}. \tag{19}$$

Recall that $\left\{\hat{f}_n\right\}$ is uniformly bounded. By Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio [4, Theorem 22], (19), and the main part of the statement and since $\operatorname{core}(\nu) \subseteq \mathcal{I}$, we have that

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

proving point 1.

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

$$\int_\Omega f d\bar{\nu} = \max_{P \in \operatorname{core}(\nu)} \int_\Omega f dP \qquad \forall f \in B(\Omega, \mathcal{F}). \tag{20}$$

Consider the sequence $\{a_n\}_{n\in\mathbb{N}}$ defined by $a_n = \int_\Omega S_n d\bar{\nu}$. By (20) and Lemma 4, we have that $\{a_n\}_{n\in\mathbb{N}}$ is subadditive. It follows that (see [14, Lemma 8.3])

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n}. \tag{21}$$

Recall that $\left\{\hat{f}_n\right\}$ is uniformly bounded. By [4, Theorem 22], (21), and the main part of the statement and since $\operatorname{core}(\nu) \subseteq \mathcal{I}$, we have that

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

proving point 2.

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

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

By the initial part of the proof, we also have that

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

Since ν is a lower probability, this implies that

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

proving the statement. ∎

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

$$S_n = \sum_{k=1}^{n} f \circ \tau^{k-1} \qquad \forall n \in \mathbb{N}$$

is an additive sequence which satisfies (5). Since ν is convex 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

$$\lim_n \frac{S_n}{n} = \lim_n \hat{f}_n = \hat{f}, \qquad \nu - a.s.$$

proving the main statement and point 1 where $f^\star = \hat{f}$.

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. At the same time, note 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. ■

Proof of Remark 1. By the same arguments contained in the proof of Corollary 2, it follows that if ν is a functionally invariant lower probability, then we have that

$$\lim_n \frac{S_n}{n} = \lim_n \hat{f}_n = \hat{f}, \qquad \nu - a.s.$$

where $f^\star = \hat{f}$. Next, assume that ν is further ergodic and robustly invariant. By Lemma 2 and since $(\Omega, \mathcal{F}, \mathcal{I})$ is a Dynkin space, we have that $\int_\Omega f d\nu \leq \int_\Omega \hat{f} d\nu$. This implies that $\int_\Omega \hat{f} d\bar{\nu} \leq \int_\Omega f d\bar{\nu}$. By Theorem 3, we also have that

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

Since $f^\star = \hat{f}$, we can conclude 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. ■

C Proofs of Section 4

Proof of Lemma 1. Consider a capacity ν and a process $\mathbf{f}$. It is immediate to see that $\nu_{\mathbf{f}}$ is a 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 $\lim_n \nu_{\mathbf{f}}(C_n) = \lim_n \nu\left(\mathbf{f}^{-1}(C_n)\right) = \lim_n \nu(A_n) = 1$, proving that $\nu_{\mathbf{f}}$ is continuous at $\mathbb{R}^{\mathbb{N}}$. Next, consider $C_1, C_2 \in \sigma(\mathcal{C})$. Since ν is convex, we have that

$$\begin{aligned}\nu_{\mathbf{f}}(C_1 \cup C_2) + \nu_{\mathbf{f}}(C_1 \cap C_2) &= \nu\left(\mathbf{f}^{-1}(C_1 \cup C_2)\right) + \nu\left(\mathbf{f}^{-1}(C_1 \cap C_2)\right) \\ &= \nu\left(\mathbf{f}^{-1}(C_1) \cup \mathbf{f}^{-1}(C_2)\right) + \nu\left(\mathbf{f}^{-1}(C_1) \cap \mathbf{f}^{-1}(C_2)\right) \\ &\geq \nu\left(\mathbf{f}^{-1}(C_1)\right) + \nu\left(\mathbf{f}^{-1}(C_2)\right) = \nu_{\mathbf{f}}(C_1) + \nu_{\mathbf{f}}(C_2),\end{aligned}$$

proving that $\nu_{\mathbf{f}}$ is convex. Next, consider $C \in \mathcal{C}$. Then, there exist $k \in \mathbb{N}$ and $E \in \mathcal{B}\left(\mathbb{R}^k\right)$ such that $C = \left\{x \in \mathbb{R}^{\mathbb{N}} : (x_1, ..., x_k) \in E\right\}$. Note that

$$\begin{aligned}\tau^{-1}(C) &= \left\{x \in \mathbb{R}^{\mathbb{N}} : \tau(x) \in C\right\} = \left\{x \in \mathbb{R}^{\mathbb{N}} : (x_2, ..., x_{k+1}) \in E\right\} \\ &= \left\{x \in \mathbb{R}^{\mathbb{N}} : (x_1, x_2, ..., x_{k+1}) \in \mathbb{R} \times E\right\}.\end{aligned} \tag{22}$$

By (22) and since $\mathbf{f}$ is stationary, it follows that

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

Since $C \in \mathcal{C}$ was arbitrarily chosen, it follows that $\mathcal{C} \subseteq \left\{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}\left(\tau^{-1}(C)\right)\right\} \subseteq \sigma(\mathcal{C})$. Since $\nu_{\mathbf{f}}$ is convex and continuous at $\mathbb{R}^{\mathbb{N}}$, we have that $\left\{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}\left(\tau^{-1}(C)\right)\right\}$ is a monotone class. By the Monotone Class Theorem (see [3, Theorem 3.4]), it follows that $\sigma(\mathcal{C}) = \left\{C \in \sigma(\mathcal{C}) : \nu_{\mathbf{f}}(C) = \nu_{\mathbf{f}}\left(\tau^{-1}(C)\right)\right\}$, that is, $\nu_{\mathbf{f}}$ is shift invariant.

Define $\mathcal{H} = \bigcap_{k=1}^{\infty} \sigma(\mathcal{C}_k^{\infty}) \cap \sigma(\mathcal{C})$. Note that $\mathbf{f}^{-1}(\mathcal{H}) = \mathcal{T}$ and define $\mathcal{G}$ the σ-algebra of shift invariant sets. Thus, $\nu_{\mathbf{f}}(\mathcal{H}) = \{0,1\}$ if and only if $\nu(\mathcal{T}) = \{0,1\}$. 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. ■

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 $\mathbb{R}$

$$\nu\left(\{\omega \in \Omega : f_1(\omega) \in B\}\right) = \nu\left(\{\omega \in \Omega : f_2(\omega) \in B\}\right) = ... = \nu\left(\{\omega \in \Omega : f_k(\omega) \in B\}\right). \tag{23}$$

By (23), this implies that for each $k \in \mathbb{N}$ and for each Borel subset B of $\mathbb{R}$

$$\begin{aligned}\nu_{\mathbf{f}}\left(\left\{x \in \mathbb{R}^{\mathbb{N}} : x_k \in B\right\}\right) &= \nu\left(\mathbf{f}^{-1}\left(\left\{x \in \mathbb{R}^{\mathbb{N}} : x_k \in B\right\}\right)\right) \\ &= \nu\left(\{\omega \in \Omega : f_k(\omega) \in B\}\right) = \nu\left(\{\omega \in \Omega : f_1(\omega) \in B\}\right).\end{aligned}$$

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 $-m 1_{\Omega} \leq f_1 \leq m 1_{\Omega}$. If we replace B with $[-m, m]$, then we can conclude that

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

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

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

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

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

Since $\nu_{\mathbf{f}}$ is a convex capacity which is further continuous at $\mathbb{R}^{\mathbb{N}}$, it follows that

$$1=\nu_{\mathbf{f}}\left(\bigcap_{k=1}^{\infty}\left\{x\in\mathbb{R}^{\mathbb{N}}:x_k\in[-m,m]\right\}\right)\leq\nu_{\mathbf{f}}\left(\bigcap_{n=1}^{\infty}\left\{x\in\mathbb{R}^{\mathbb{N}}:\frac{1}{n}\sum_{k=1}^{n}\pi\left(\tau^{k-1}(x)\right)=\frac{1}{n}\sum_{k=1}^{n}x_k\right\}\right)\leq 1. \tag{24}$$

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

$$\nu_{\mathbf{f}}\left(\left\{x\in\mathbb{R}^{\mathbb{N}}:\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\nu_{\mathbf{f}}\leq\lim_n\frac{1}{n}\sum_{k=1}^{n}\pi\left(\tau^{k-1}(x)\right)=\pi^\star(x)\leq\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\bar{\nu}_{\mathbf{f}}\right\}\right)=1. \tag{25}$$

By (24) and (25) and since $\nu_{\mathbf{f}}$ is convex, we can conclude that

$$\nu_{\mathbf{f}}\left(\left\{x\in\mathbb{R}^{\mathbb{N}}:\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\nu_{\mathbf{f}}\leq\lim_n\frac{1}{n}\sum_{k=1}^{n}x_k=\pi^\star(x)\leq\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\bar{\nu}_{\mathbf{f}}\right\}\right)=1. \tag{26}$$

Define $E=\left\{x\in\mathbb{R}^{\mathbb{N}}:\lim_n\frac{1}{n}\sum_{k=1}^{n}\pi\left(\tau^{k-1}(x)\right)=\pi^\star(x)\right\}$ and $\pi_n=\frac{1}{n}\sum_{k=1}^{n}\pi\left(\tau^{k-1}\right)$ for all $n\in\mathbb{N}$. By (25), we have that $P(E)=1$ for all $P\in\operatorname{core}(\nu_{\mathbf{f}})$. By construction, $\{1_E\pi_n\}_{n\in\mathbb{N}}\subseteq B\left(\mathbb{R}^{\mathbb{N}},\sigma(\mathcal{C})\right)$ is a uniformly bounded sequence which converges pointwise to $1_E\pi^\star$. By [4, Theorem 22] and since $\nu_{\mathbf{f}}$ is convex and $P(E)=1$ for all $P\in\operatorname{core}(\nu_{\mathbf{f}})$, this implies that

$$\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\nu_{\mathbf{f}}=\int_{\mathbb{R}^{\mathbb{N}}}1_E\pi^\star d\nu_{\mathbf{f}}=\int_{\mathbb{R}^{\mathbb{N}}}\lim_n 1_E\pi_n d\nu_{\mathbf{f}}=\lim_n\int_{\mathbb{R}^{\mathbb{N}}}1_E\pi_n d\nu_{\mathbf{f}}=\lim_n\int_{\mathbb{R}^{\mathbb{N}}}\pi_n d\nu_{\mathbf{f}}. \tag{27}$$

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

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

By (27), it follows that $\int_{\mathbb{R}^{\mathbb{N}}}\pi^\star d\nu_{\mathbf{f}}\geq\int_{\mathbb{R}^{\mathbb{N}}}\pi d\nu_{\mathbf{f}}$. A similar argument yields that $\int_{\mathbb{R}^{\mathbb{N}}}g^\star d\bar{\nu}_{\mathbf{f}}\leq\int_{\mathbb{R}^{\mathbb{N}}}\pi d\bar{\nu}_{\mathbf{f}}$. Finally, since $\int_{\mathbb{R}^{\mathbb{N}}}\pi d\nu_{\mathbf{f}}=\int_\Omega f_1 d\nu$ and $\int_{\mathbb{R}^{\mathbb{N}}}\pi d\bar{\nu}_{\mathbf{f}}=\int_\Omega f_1 d\bar{\nu}$, by (26), we can conclude that

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

proving the statement. ■

Proof of Proposition 2. By assumption, we have that whenever C_1 and C_2 are two base disjoint cylinders then

$$\nu_{\mathbf{f}}(C_1)\,\nu_{\mathbf{f}}(C_2)\leq\nu_{\mathbf{f}}(C_1\cap C_2)\leq\nu_{\mathbf{f}}(C_1)\,\bar{\nu}_{\mathbf{f}}(C_2)\leq\bar{\nu}_{\mathbf{f}}(C_1\cap C_2). \tag{28}$$

Next, fix $C_1\in\mathcal{C}$ and $\bar{C}\in\mathcal{G}$. It follows that C_1 is of length k for some $k\in\mathbb{N}$. Consider the class

$$\Lambda_k=\left\{C\in\sigma\left(\mathcal{C}_{k+1}^{\infty}\right):(28)\text{ holds with }C_2=C\right\}.$$

Since $\nu_{\mathbf{f}}$ satisfies (28) for base disjoint cylinders, note that $\mathcal{C}_{k+1}^{\infty}\subseteq\Lambda_k$. Recall that $\mathcal{C}_{k+1}^{\infty}$ is an algebra. Second, since $\nu_{\mathbf{f}}$ is continuous at $\mathbb{R}^{\mathbb{N}}$, it follows that Λ_k is a monotone class. By the Monotone Class Theorem (see [3, Theorem 3.4]), it follows that $\Lambda_k\supseteq\sigma\left(\mathcal{C}_{k+1}^{\infty}\right)\supseteq\mathcal{G}$. This implies that (28) holds for C_1 and $\bar{C}$. Since $C_1\in\mathcal{C}$ and $\bar{C}\in\mathcal{G}$ were arbitrarily chosen, we can conclude that

$$\nu_{\mathbf{f}}(C_1)\,\nu_{\mathbf{f}}(C_2)\leq\nu_{\mathbf{f}}(C_1\cap C_2)\leq\nu_{\mathbf{f}}(C_1)\,\bar{\nu}_{\mathbf{f}}(C_2)\leq\bar{\nu}_{\mathbf{f}}(C_1\cap C_2)\qquad\forall C_1\in\mathcal{C},\forall C_2\in\mathcal{G}. \tag{29}$$

By a similar argument and another application of the Monotone Class Theorem, it follows that

$$\nu_{\mathbf{f}}(C_1)\nu_{\mathbf{f}}(C_2) \leq \nu_{\mathbf{f}}(C_1 \cap C_2) \leq \nu_{\mathbf{f}}(C_1)\bar{\nu}_{\mathbf{f}}(C_2) \leq \bar{\nu}_{\mathbf{f}}(C_1 \cap C_2) \qquad \forall C_1 \in \sigma(\mathcal{C}), \forall C_2 \in \mathcal{G}. \tag{30}$$

By (30), we can conclude that for each $C \in \mathcal{G}$

$$0 \leq \nu_{\mathbf{f}}(C)\nu_{\mathbf{f}}(C^c) \leq \nu_{\mathbf{f}}(C \cap C^c) \leq \nu_{\mathbf{f}}(C)\bar{\nu}_{\mathbf{f}}(C^c) \leq \bar{\nu}_{\mathbf{f}}(C \cap C^c) \leq 0.$$

This implies that $\nu_{\mathbf{f}}(C)\nu_{\mathbf{f}}(C^c) = 0 = \nu_{\mathbf{f}}(C)\bar{\nu}_{\mathbf{f}}(C^c)$, that is, either $\nu_{\mathbf{f}}(C) = 0$ or $\bar{\nu}_{\mathbf{f}}(C^c) = 0$. In other words, either $\nu_{\mathbf{f}}(C) = 0$ or $\nu_{\mathbf{f}}(C) = 1$. Thus, we can conclude that $\nu_{\mathbf{f}}(\mathcal{G}) = \{0, 1\}$, that is, $\mathbf{f}$ is ergodic. ■

Proof of Proposition 3. Since ν is convex, $\nu_{\mathbf{f}}$ is convex. This implies that

$$\max_{P \in \mathrm{core}(\nu_{\mathbf{f}})} P(A) = \bar{\nu}_{\mathbf{f}}(A) \geq \nu_{\mathbf{f}}(A) = \min_{P \in \mathrm{core}(\nu_{\mathbf{f}})} P(A) \qquad \forall A \in \sigma(\mathcal{C}).$$

Consider two base disjoint cylinders C_1, C_2 in $\mathcal{C}$. Since $\{f_n\}_{n\in\mathbb{N}}$ satisfies (8), it follows that

$$\nu_{\mathbf{f}}(C_1)\nu_{\mathbf{f}}(C_2) = \nu_{\mathbf{f}}(C_1 \cap C_2) = \nu_{\mathbf{f}}(C_1)\nu_{\mathbf{f}}(C_2) \leq \nu_{\mathbf{f}}(C_1)\bar{\nu}_{\mathbf{f}}(C_2). \tag{31}$$

Next, observe that C_1 and C_2^c are two base disjoint cylinders. This implies that

$$\nu_{\mathbf{f}}(C_1)\bar{\nu}_{\mathbf{f}}(C_2) = \nu_{\mathbf{f}}(C_1)\left[1 - \nu_{\mathbf{f}}\left((C_2)^c\right)\right] = \nu_{\mathbf{f}}(C_1) - \nu_{\mathbf{f}}(C_1)\nu_{\mathbf{f}}(C_2^c) = \nu_{\mathbf{f}}(C_1) - \nu_{\mathbf{f}}(C_1 \cap C_2^c). \tag{32}$$

It is immediate to see that $C_1 \cap C_2^c \subseteq C_1$. By [18, Theorem 4.7] and since $\nu_{\mathbf{f}}$ is convex, there exists $P \in \mathrm{core}(\nu_{\mathbf{f}})$ such that

$$P(C_1) = \nu_{\mathbf{f}}(C_1) \text{ and } P(C_1 \cap C_2^c) = \nu_{\mathbf{f}}(C_1 \cap C_2^c).$$

Given (32), this implies that

$$\nu_{\mathbf{f}}(C_1)\bar{\nu}(C_2) = \nu_{\mathbf{f}}(C_1) - \nu_{\mathbf{f}}(C_1 \cap C_2^c) = P(C_1) - P(C_1 \cap C_2^c) = P(C_1 \cap C_2) \leq \bar{\nu}_{\mathbf{f}}(C_1 \cap C_2). \tag{33}$$

By (31) and (33), it follows that

$$\nu_{\mathbf{f}}(C_1)\nu_{\mathbf{f}}(C_2) = \nu_{\mathbf{f}}(C_1 \cap C_2) = \nu_{\mathbf{f}}(C_1)\nu(C_2) \leq \nu_{\mathbf{f}}(C_1)\bar{\nu}_{\mathbf{f}}(C_2) \leq \bar{\nu}_{\mathbf{f}}(C_1 \cap C_2).$$

Since C_1 and C_2 were arbitrarily chosen, the statement follows. ■

References

[1] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer, New York, 2006.

[2] J. O. Berger, Robust Bayesian analysis: sensitivity to the prior, *Journal of Statistical Planning and Inference*, 25, 303-328, 1990.

[3] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley & Sons, New York, 1995.

[4] 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.

[5] 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.

[6] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio, Ambiguity and Robust Statistics, *Journal of Economic Theory*, 974-1049, 2013.

[7] F. Delbaen, Convex Games and Extreme Points, *Journal of Mathematical Analysis and Applications*, 45, 210-233, 1974.

[8] Y. N. Dowker, Invariant Measure and the Ergodic Theorems, *Duke Mathematical Journal*, 4, 1051-1061, 1947.

[9] Y. N. Dowker, Finite and σ-Finite Invariant Measures, *Annals of Mathematics*, 4, 595-608, 1951.

[10] R. M. Dudley, *Real Analysis and Probability*, 2nd ed., Cambridge University Press, Cambridge, 2002.

[11] N. Dunford and J. T. Schwartz, *Linear Operators; Part I: General Theory*, Wiley, New York, 1958.

[12] E. B. Dynkin, Sufficient Statistics and Extreme Points, *The Annals of Probability*, 6, 705-730, 1978.

[13] W. F. Eberlein, Abstract Ergodic Theorems and Weak Almost Periodic Functions, *Transactions of the American Mathematical Society*, 67, 217-240, 1949.

[14] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, 2nd ed., Springer, New York, 2009.

[15] U. Krengel, *Ergodic Theorems*, Walter de Gruyter, Berlin, 1985.

[16] F. Maccheroni and M. Marinacci, A Strong Law of Large Numbers for Capacities, *The Annals of Probability*, 33, 1171-1178, 2005.

[17] M. Marinacci, Limit Laws for Non-additive Probabilities and Their Frequentist Interpretation, *Journal of Economic Theory*, 84, 145-195, 1999.

[18] M. Marinacci and Luigi Montrucchio, Introduction to the Mathematics of Ambiguity, in *Uncertainty in Economic Theory*, Routledge, New York, 2004.

[19] R. R. Phelps, *Lectures on Choquet's Theorem*, 2nd ed., Springer, 2001.

[20] G. Shafer, Belief functions and parametric models, *Journal of the Royal Statistical Society: Series B*, 44, 322-352, 1982.

[21] D. Schmeidler, Cores of Exact Games, I, *Journal of Mathematical Analysis and Applications*, 40, 214-225, 1972.

[22] D. Schmeidler, Integral Representation without Additivity, *Proceedings of the American Mathematical Society*, 97, 255-261, 1986.