

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Ergodicity of invariant capacities

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1016/j.spa.2020.02.010>

PUBLISHER

Elsevier

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This paper was accepted for publication in the journal Stochastic Processes and their Applications and the definitive published version is available at <https://doi.org/10.1016/j.spa.2020.02.010>.

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Feng, Chunrong, Panyu Wu, and Huaizhong Zhao. 2020. "Ergodicity of Invariant Capacities". Loughborough University. <https://hdl.handle.net/2134/11889675.v1>.

Ergodicity of Invariant Capacities

Chunrong Feng^a, Panyu Wu^{b,*}, Huaizhong Zhao^a

^a*Department of Mathematical Sciences, Loughborough University, LE11 3TU, UK*

^b*Zhongtai Institute for Financial Studies, Shandong University, 250100, China*

Abstract

In this paper, we investigate capacity preserving transformations and their ergodicity. We obtain some limit properties under capacity spaces and then give the concept of ergodicity for a capacity preserving transformation. Based on this definition, we give several characterizations of ergodicity. In particular, we obtain a type of Birkhoff's ergodic theorem and prove that the ergodicity of a transformation with respect to an upper probability is equivalent to a type of strong law of large numbers.

Keywords: Capacity; Ergodicity; Invariant set; Strong law of large numbers; Choquet integral

1. Introduction

In this paper, we investigate capacity preserving transformations and their ergodicity. Capacities (or nonadditive probabilities) arise in modelling heterogeneous environments, for example, a financial market where biased beliefs of future price movements drives the decision of stock-market participants and creates ambiguous volatility. It was pointed out that the additive probability theory might not be adequate in either economics (see [1], [18] and [25]) or statistics (see [27]). Dynamical systems on a capacity space concern transformations from the capacity space to itself. It is vitally important to study the dynamics of such transformations, of which little is known. When additivity ceases to be valid, many classical results turn out to be invalid and

*Corresponding author.

Email addresses: C.Feng@lboro.ac.uk (Chunrong Feng), wupanyu@sdu.edu.cn (Panyu Wu), H.Zhao@lboro.ac.uk (Huaizhong Zhao)

the situations become more complicated. In particular, irreducibility has not been defined for capacity.

The classical ergodic theory deals with a probability preserving map θ from Ω to Ω on a probability space (Ω, \mathcal{F}, P) . Let \mathcal{G} denote all the invariant sets with respect to θ . Then θ is called ergodic if any invariant set $B \in \mathcal{G}$ has either $P(B) = 0$ or $P(B) = 1$. This is equivalent to for any $B \in \mathcal{G}$, $P(B) = 0$ or $P(B^c) = 1$. This means that the dynamical system cannot be decomposed into different dynamical systems (see [20] or [28]). However, in the capacity theory, the equivalence is no longer true. To see this, let $(\mu, \bar{\mu})$ denote a pair of conjugate capacities on \mathcal{F} . Then for any $B \in \mathcal{G}$, $\mu(B) = 0$ or 1 is equivalent to that for any $B \in \mathcal{G}$, $\mu(B) = 0$ or $\bar{\mu}(B^c) = 0$. But it is not equivalent to either that for any $B \in \mathcal{G}$, $\mu(B) = 0$ or $\mu(B^c) = 0$, or that for any $B \in \mathcal{G}$, $\bar{\mu}(B) = 0$ or $\bar{\mu}(B^c) = 0$. So how to define an ergodic transformation is an issue worthy of discussions.

Cerreia-Vioglio, Maccheroni and Marinacci called a capacity μ ergodic if $\mu(\mathcal{G}) = \{0, 1\}$ and then established an ergodic theorem for lower probabilities in [3]. But with the requirement of $\mu(\mathcal{G}) = \{0, 1\}$ in a capacity space, it is still possible that $\mu(B^c) = 1$ when $\mu(B) = 1$. That means if θ is “ergodic” in the sense of [3], the space Ω may still be divided into two sets B and B^c , each having full capacity but $\theta(B) = B$ and $\theta(B^c) = B^c$, that is, θ is reducible (see Example 4.2 in this paper).

The irreducibility condition is arguably important even in the sublinear expectation ([16]) or the capacity settings as in the classical ergodic theory situation. It was proved that the G -Brownian motion on a circle has an invariant expectation and it is ergodic in [16]. Inspired by the results of [16] for the ergodicity on sublinear expectation spaces, we strengthen the ergodicity concept by adding another condition that is for any θ -invariant set B , either $\mu(B) = 0$ or $\mu(B^c) = 0$. Under this case, if θ is ergodic with respect to a capacity μ then the space Ω cannot be decomposed into two sets B and B^c , each having positive capacity but $\theta(B) = B$ and $\theta(B^c) = B^c$. In other words, θ is irreducible. Subsequently, we obtain two equivalent characterizations of our ergodicity with respect to a continuous subadditive capacity μ : recurrence (Theorem 4.3) and the shift invariant random variable being a constant μ -almost surely (Theorem 4.4).

In the classical additive probability case, the result that any shift invariant random variable must be constant, as stated in Theorem 4.4 for capacities, provides an important characteristics of ergodicity in terms of the spectral structure of the corresponding transformation operator on the space of mea-

surable functions. For a Markov semigroup with an invariant measure, this suggests that 0 is a simple eigenvalue of the infinitesimal generator of the Markov semigroup if and only if the invariant measure is ergodic. A well-known case in literature is that of mixing stationary processes. In this case the Koopman-von Neumann theorem implies that the generator has only one eigenvalue 0 on the imaginary axis which is simple. Recently it was observed in [17] that random periodic process is another ergodic regime of which the spectral structure of the generator is distinct from that of the mixing regime. In the ergodic random periodic regime the infinitesimal generator has infinite number of simple eigenvalues including 0 equally placed on the imaginary axis.

In [3], Cerreia-Vioglio, Maccheroni and Marinacci obtained an important result that for bounded random variable ξ , if the lower probability v is θ -invariant, then

$$v \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \text{ exists} \right\} \right) = 1.$$

Based on this result, we obtain that θ is ergodic with respect to a continuous upper probability in the sense defined in this paper if and only if $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega)$ is a constant V -almost surely, where V is the corresponding upper probability (Theorem 4.5). This is achieved by the irreducibility given in the ergodicity definition of this paper. However, this is not the case under the framework of [3] as demonstrated in Example 4.7. Moreover, if the upper probability is concave, we show that this constant is bounded by the Choquet integrals with respect to the upper probability and the conjugate lower probability (Theorem 4.6). This is due to the property that θ preserving capacity can infer θ preserving the corresponding Choquet integral (Proposition 3.1).

So far, no other limit theory about capacity provided a condition for the limit of $\frac{1}{n} \sum_{i=1}^n X_i$ being a constant V -almost surely. Therefore, the ergodic theory of this paper can provide new insight to the study of capacity. On the other hand, there are also other papers attempting to investigate the ergodicity in capacity spaces or sublinear expectation spaces from different angles, one can see [10] and [19] and the references therein. None of these papers dealt with dynamical property of processes especially the non-decomposable property.

The Kolmogorov 0-1 Law plays an important role in the limit theory

under the classical probability framework, which implies that the tail event happens with probability 0 or 1. In Section 2, we investigate the Kolmogorov 0-1 Law for the independence in the capacity space setting (Definition 2.1). We give an example to show that a sequence of independent random variables with respect to capacity μ cannot deduce $\sigma(Y_k, k \leq n)$ and $\sigma(Y_k, k \geq n+1)$ being independent with respect to μ , for all $n \in \mathbb{N}$. Here and in the sequel, we use \mathbb{N} to denote the set of all the positive integers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Then we investigate some properties of Choquet integral which will be used in this paper. In Section 3, we study the properties of θ -invariant capacities. In Section 4, we firstly investigate the limit properties under the scenario that $\mu(\mathcal{G}) = \{0, 1\}$ and then give the definition of a transformation to be ergodic. Based on our definition, we give several characterisations of ergodicity and a type of Birkhoff's ergodic theorem. In Section 5, we give a strong law of large numbers for stationary and ergodic sequences in upper probability spaces.

2. Basic concepts and independence on capacity space

Let (Ω, \mathcal{F}) be a measurable space. Recall a set function $\mu : \mathcal{F} \rightarrow [0, 1]$ is

- a capacity/nonadditive probability if $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$, and $\mu(A) \leq \mu(B)$ for all $A, B \in \mathcal{F}$ such that $A \subseteq B$;
- concave/submodular/2-alternating if $\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$ for all $A, B \in \mathcal{F}$;
- convex/supermodular/2-monotonic if $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$ for all $A, B \in \mathcal{F}$;
- subadditive if $\mu(A \cup B) \leq \mu(A) + \mu(B)$ for all $A, B \in \mathcal{F}$ with $A \cap B = \emptyset$;
- superadditive if $\mu(A \cup B) \geq \mu(A) + \mu(B)$ for all $A, B \in \mathcal{F}$ with $A \cap B = \emptyset$;
- continuous from below/inner continuous if $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$ for $A_n \uparrow A$;
- continuous from above/outer continuous if $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$ for $A_n \downarrow A$;
- continuous if it is both continuous from below and above.

For a capacity μ on \mathcal{F} , $(\Omega, \mathcal{F}, \mu)$ is called a capacity space. The conjugate capacity $\bar{\mu}$ on \mathcal{F} is defined by

$$\bar{\mu}(A) = 1 - \mu(A^c), \quad \text{for any } A \in \mathcal{F},$$

where A^c is the complementary set of A . Notice that if μ is additive, then $\bar{\mu} = \mu$. Capacity μ is continuous at Ω if and only if $\bar{\mu}$ is continuous at \emptyset . Moreover, capacity μ is convex if and only if $\bar{\mu}$ is concave. However, the superadditivity and the subadditivity do not have such a conjugate relation.

Let $\Delta(\Omega, \mathcal{F})$ denote the set of all finitely additive probabilities on \mathcal{F} and $\Delta^\sigma(\Omega, \mathcal{F})$ denote the set of all probabilities (σ -additive) on \mathcal{F} . The widely studied conjugate capacities which satisfy the subadditivity and superadditivity respectively are upper and lower probabilities. A pair of capacities (V, v) is called the upper and lower probabilities on (Ω, \mathcal{F}) (generated by \mathcal{P}) if

$$V(A) = \sup_{P \in \mathcal{P}} P(A) \text{ and } v(A) = \inf_{P \in \mathcal{P}} P(A), \text{ for any } A \in \mathcal{F},$$

where \mathcal{P} is a nonempty set of $\Delta(\Omega, \mathcal{F})$.

Before we establish the Kolmogorov 0-1 Law in a capacity space, we give the following definitions which are natural extensions of the corresponding concepts in the classical probability theory.

Definition 2.1. Let $(\Omega, \mathcal{F}, \mu)$ be a capacity space and J be an index set.

Events A and B are called independent with respect to μ if $\mu(A \cap B) = \mu(A)\mu(B)$.

Events set $\{A_t, t \in J\}$ are called (mutually) independent with respect to μ if for any finite subset $I \subseteq J$

$$\mu \left(\bigcap_{t \in I} A_t \right) = \prod_{t \in I} \mu(A_t).$$

Let $\{\mathcal{D}_t, t \in J\}$ be a set of subclasses of \mathcal{F} . If for any finite subset $I \subseteq J$

$$\mu \left(\bigcap_{t \in I} A_t \right) = \prod_{t \in I} \mu(A_t), \quad \text{for any } A_t \in \mathcal{D}_t, t \in I,$$

then $\{\mathcal{D}_t, t \in J\}$ are called (mutually) independent subclasses with respect to μ .

Lemma 2.1. *If $\{\mathcal{D}_t, t \in J\}$ are independent subclasses with respect to a continuous capacity μ and \mathcal{D}_t is an algebra for every $t \in J$, then $\{\sigma(\mathcal{D}_t), t \in J\}$ are independent σ -algebra with respect to the capacity μ .*

Proof. This lemma can be deduced from the monotone class theorem. We omit the details. \square

Definition 2.2. *Random variables $\{Y_t, t \in J\}$ on $(\Omega, \mathcal{F}, \mu)$ are said to be independent with respect to capacity μ if the σ -algebras $\{\sigma(Y_t), t \in J\}$ are independent with respect to μ .*

For a sequence of random variables $\{Y_n\}_{n \in \mathbb{N}}$, Definition 2.2 is equivalent to the definition of independent random variables sequence with respect to μ given by [22] (Definition 4). But it is worth noting that $\{Y_n\}_{n \in \mathbb{N}}$ being independent with respect to μ cannot deduce $\sigma(Y_k, k \leq n)$ and $\sigma(Y_k, k \geq n+1)$ being independent with respect to μ , for all $n \in \mathbb{N}$. We give the following example to illustrate this.

Example 2.1. *Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, \mathcal{F} be all subsets of Ω and μ be a capacity on \mathcal{F} , with $\mu(A) = 0$ if $|A| \leq 3$, and $\mu(A) = 1$ if $|A| \geq 4$, where $|A|$ denotes the number of elements in A . It is easy to see that μ is superadditive but is not convex. However, the conjugate capacity of μ is neither subadditive nor superadditive. Let*

$$Y_1(\omega) = \begin{cases} 0 & \omega = \omega_1, \omega_2 \\ 1 & \omega = \omega_3, \omega_4, \omega_5 \end{cases}, \quad Y_2(\omega) = \begin{cases} 0 & \omega = \omega_1, \omega_4, \omega_5 \\ 1 & \omega = \omega_2, \omega_3 \end{cases},$$

$$Y_3(\omega) = \begin{cases} 0 & \omega = \omega_3, \omega_4 \\ 1 & \omega = \omega_1, \omega_2, \omega_5 \end{cases}, \quad Y_4(\omega) = \begin{cases} 0 & \omega = \omega_1, \omega_4 \\ 1 & \omega = \omega_2, \omega_3, \omega_5 \end{cases}.$$

It is easy to check that Y_1, Y_2, Y_3, Y_4 are independent and identically distributed with respect to μ . However, $\sigma(Y_1, Y_2)$ and $\sigma(Y_3, Y_4)$ are not independent with respect to μ since

$$\mu((\{Y_1 = 0\} \cup \{Y_2 = 0\}) \cap (\{Y_3 = 0\} \cup \{Y_4 = 1\})) = \mu(\{\omega_2, \omega_4, \omega_5\}) = 0 \neq 1 = \mu(\{Y_1 = 0\} \cup \{Y_2 = 0\}) \cdot \mu(\{Y_3 = 0\} \cup \{Y_4 = 1\}).$$

Moreover, notice that

$$\mu(\{\omega : (Y_1, Y_2)(\omega) \in \{(0, 0), (0, 1), (1, 1)\}\}) = \mu(\{\omega_1, \omega_2, \omega_3\}) = 0$$

but

$$\mu(\{\omega : (Y_3, Y_4)(\omega) \in \{(0, 0), (0, 1), (1, 1)\}\}) = \mu(\{\omega_2, \omega_3, \omega_4, \omega_5\}) = 1.$$

So (Y_1, Y_2) and (Y_3, Y_4) are not identically distributed with respect to μ .

Proposition 2.1. *An upper/lower probability must be a subadditive/superadditive capacity. The converse is not true.*

Proof. It is obvious that an upper/lower probability must be a subadditive/superadditive capacity.

We disprove the converse by giving the following counterexample. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, \mathcal{F} be all subsets of Ω and μ be a capacity on \mathcal{F} , with $\mu(A) = 1/2$ if $1 \leq |A| \leq 2$, where $|A|$ denotes the number of elements in A . Then μ is subadditive. If μ is an upper probability, then $\bar{\mu}$ (the conjugate capacity of μ) is lower probability, thus $\bar{\mu}$ is superadditive. Actually, $\bar{\mu} = \mu$ and $\bar{\mu}(\{\omega_1, \omega_2\}) = 1/2 < 1 = \bar{\mu}(\{\omega_1\}) + \bar{\mu}(\{\omega_2\})$ which contradicts with the superadditivity of $\bar{\mu}$. Therefore, μ cannot be an upper probability.

Now we turn to the superadditive capacity μ in Example 2.1. This μ cannot be a lower probability since the conjugate capacity of μ is not subadditive. \square

Next we give the Kolmogorov 0-1 Law in capacity spaces which will be used in the proof of Corollary 5.1.

Theorem 2.1. *Let $\{Y_n\}_{n \in \mathbb{N}}$ be random variables such that for any $n \in \mathbb{N}$, $\sigma(Y_k, k \leq n)$ and $\sigma(Y_k, k \geq n+1)$ are independent with respect to a continuous capacity μ . By \mathcal{T} we denote the tail σ -algebra of $\{Y_n\}_{n \in \mathbb{N}}$, that is $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(Y_k, k \geq n)$. Then for any $A \in \mathcal{T}$, the following two statements are true:*

- (i) $\mu(A) = 0$ or 1 ;
- (ii) $\mu(A) = 0$ or $\mu(A^c) = 0$.

Proof. It is obvious that \mathcal{T} is independent of $\mathcal{A} = \bigcup_{n \geq 1} \sigma(Y_k, k \leq n)$ with respect to μ since $\mathcal{T} \subseteq \sigma(Y_k, k \geq n+1)$, for all $n \in \mathbb{N}$. Notice that \mathcal{A} is an algebra, then by Lemma 2.1, \mathcal{T} is independent of $\sigma(\mathcal{A})$. On the other hand, $\mathcal{T} \subseteq \sigma(\mathcal{A})$. Hence, \mathcal{T} is independent of itself with respect to μ . So for any $A \in \mathcal{T}$, $\mu(A) = \mu(A \cap A) = \mu(A) \cdot \mu(A)$, which implies $\mu(A) = 0$ or 1 . Meantime, $\mu(\emptyset) = \mu(A \cap A^c) = \mu(A) \cdot \mu(A^c)$, which deduces $\mu(A) = 0$ or $\mu(A^c) = 0$. \square

Remark 2.1. *Although the proof of the Kolmogorov 0-1 Law in capacity spaces is no big difference with the proof of the classical Kolmogorov 0-1 Law, we still present here since the two statements (i) and (ii) in Theorem 2.1 are equivalent when μ is additive, but not equivalent when μ is nonadditive.*

At the end of this section, we recall the Choquet integral/expectation of a random variable, introduced by Choquet in [7]. For any \mathcal{F} -measurable real valued random variable ξ , the Choquet integral/expectation of ξ with respect to μ is defined by

$$\int_{\Omega} \xi(\omega) d\mu = \int_0^{\infty} \mu(\{\omega : \xi(\omega) \geq t\}) dt + \int_{-\infty}^0 [\mu(\{\omega : \xi(\omega) \geq t\}) - 1] dt,$$

if the right side is not $\infty - \infty$. In this paper, we always consider the random variables taking real values. The asymmetry is one of the most important properties of Choquet integral (see Proposition 5.1 in [12]), which means that

$$\int_{\Omega} -\xi(\omega) d\mu = - \int_{\Omega} \xi(\omega) d\bar{\mu}.$$

It is well known that a capacity is concave/convex if and only if the related Choquet integral is subadditive/superadditive (see Exercise 5.1 and Theorem 6.3 in [12]). A concave/convex capacity must be an upper/lower probability (Proposition 10.3 in [12]). However, the converse is not true. It is easy to check that the upper probability given in Example 4.4 is not concave.

Definition 2.3. *In a capacity space $(\Omega, \mathcal{F}, \mu)$, we call that a statement holds μ -almost surely (μ -a.s. for short) if it holds outside a set A with $\mu(A) = 0$.*

The following lemma provides a dominated convergence theorem in a capacity space with respect to the Choquet integral which has been studied in [24] (Theorem 3.3) for bounded random variables. Suppose that the capacity is continuous first, [29] established a similar result (Theorem 11.10) as this lemma which will be used in the proof of Theorem 4.6.

Lemma 2.2. *Let μ be a subadditive capacity on \mathcal{F} and continuous from above, we have*

(i) for random variables $\{X_n\}_{n \in \mathbb{N}}$, Y and Z with $Y \leq X_n \leq Z$ and $\int_{\Omega} Y d\mu$, $\int_{\Omega} Z d\mu$ being finite, if $X_n \rightarrow X$ μ -a.s. then

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n d\mu = \int_{\Omega} X d\mu; \tag{1}$$

(ii) capacity μ is continuous.

Proof. (i) For any $\epsilon > 0$, we get

$$\mu \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| \geq \epsilon\} \right) = 0$$

from

$$\begin{aligned} 0 = \mu(\{\omega : X_n(\omega) \not\rightarrow X(\omega)\}) &= \mu \left(\bigcup_{\epsilon > 0} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| \geq \epsilon\} \right) \\ &\geq \mu \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| \geq \epsilon\} \right). \end{aligned}$$

Therefore, by the continuity from above of μ , we have

$$\begin{aligned} 0 &\leq \limsup_{n \rightarrow \infty} \mu(\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) \\ &\leq \limsup_{n \rightarrow \infty} \mu \left(\bigcup_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| \geq \epsilon\} \right) \\ &= \mu \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| \geq \epsilon\} \right) = 0, \end{aligned}$$

which means X_n converges to X μ -stochastically (p97 in [12]). Since μ is subadditive, then by Proposition 8.5 and Theorem 8.9 in [12], we can get (1). Claim (ii) is a direct corollary of (i) if we consider $X_n = I_{A_n}$ and $X = I_A$. \square

3. Properties of invariant capacities

We consider a \mathcal{F}/\mathcal{F} -measurable transformation $\theta : \Omega \rightarrow \Omega$. A set A is called invariant set with respect to θ if $\theta^{-1}A = A$. It is easy to check that A^c is an invariant set if and only if A is an invariant set. Let \mathcal{G} denote the set of all invariant sets with respect to θ , it is easy to check that \mathcal{G} is a sub- σ -algebra of \mathcal{F} (Exercise 7.1.1 in [13]). Corresponding to the θ -invariant probability, Cerreia-Vioglio, Maccheroni and Marinacci introduced the definition of θ -invariant capacity (Definition 1 in [3]) as follows.

Definition 3.1. A capacity μ is θ -invariant if for each $A \in \mathcal{F}$, $\mu(A) = \mu(\theta^{-1}A)$. We also say that θ preserves μ if μ is θ -invariant.

Remark 3.1. Let \mathcal{N} denote the set of all θ -invariant capacities, $\overline{\mathcal{N}}$ denote the set of all subadditive θ -invariant capacities and $\underline{\mathcal{N}}$ denote the set of all superadditive θ -invariant capacities. Using the construction given in the proof of the Claim in [3] (p3390), one can show that for any \mathcal{F}/\mathcal{F} -measurable transformation θ , there exists a θ -invariant finitely additive probability. Then \mathcal{N} , $\overline{\mathcal{N}}$ and $\underline{\mathcal{N}}$ are non-empty. Actually, we can define a pair of conjugate capacities $(\mu, \bar{\mu})$ by

$$\mu(A) = \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(\theta^k \omega), \quad \bar{\mu}(A) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(\theta^k \omega),$$

which belongs to $\underline{\mathcal{N}}$ and $\overline{\mathcal{N}}$ respectively. It is easy to check that the sets \mathcal{N} , $\overline{\mathcal{N}}$ and $\underline{\mathcal{N}}$ are convex.

Proposition 3.1. (i) A capacity μ is θ -invariant if and only if its conjugate capacity $\bar{\mu}$ is θ -invariant.

(ii) A capacity μ is θ -invariant if and only if θ preserves the Choquet integral with respect to μ , that is, for any random variable ξ , the Choquet integrals of ξ and $\xi(\theta)$ with respect to μ exist at the same time and

$$\int_{\Omega} \xi(\omega) d\mu = \int_{\Omega} \xi(\theta\omega) d\mu$$

when Choquet integral of ξ or $\xi(\theta)$ with respect to μ exists.

Proof. (i) is easy to check by the definition of θ -invariant and conjugate capacity.

(ii) If capacity μ is θ -invariant, then for any $t \in \mathbb{R}$,

$$\begin{aligned} \mu(\{\omega : \xi(\theta\omega) \geq t\}) &= \mu(\{\omega : \omega \in \theta^{-1}(\xi^{-1}[t, \infty))\}) \\ &= \mu(\{\omega : \omega \in \xi^{-1}[t, \infty)\}) = \mu(\{\omega : \xi(\omega) \geq t\}). \end{aligned}$$

Thus, the Choquet integrals of ξ and $\xi(\theta)$ with respect to μ exist at the same time and

$$\int_{\Omega} \xi(\omega) d\mu = \int_{\Omega} \xi(\theta\omega) d\mu$$

when Choquet integral of ξ or $\xi(\theta)$ with respect to μ exists.

On the other hand, we consider $\xi = I_A$, for any $A \in \mathcal{F}$. Then $\mu(A) = \int_{\Omega} I_A(\omega) d\mu = \int_{\Omega} I_A(\theta\omega) d\mu = \int_{\Omega} I_{\theta^{-1}A}(\omega) d\mu = \mu(\theta^{-1}A)$. \square

The following lemma can be deduced from the Theorem 2 obtained by Cerreia-Vioglio, Maccheroni and Marinacci in [3] which will be useful in our characterization about the ergodicity in an upper probability space.

Lemma 3.1. *Let v be a continuous lower probability on (Ω, \mathcal{F}) . If v is θ -invariant, then for any bounded \mathcal{F} -measurable random variable ξ ,*

$$v \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) \text{ exists} \right\} \right) = 1.$$

4. Ergodicity under capacity spaces

Before studying the ergodicity under a capacity space, we first study the following properties of random variables which are measurable with respect to a sub- σ -algebra of \mathcal{F} with capacity only 0 or 1.

Theorem 4.1. *Let μ be a continuous capacity on \mathcal{F} and \mathcal{F}_0 be any sub- σ -algebra of \mathcal{F} with $\mu(\mathcal{F}_0) = \{0, 1\}$. Then, for any \mathcal{F}_0 -measurable random variable ξ , we have*

(i) *the Choquet integrals of ξ with respect to μ and $\bar{\mu}$ are finite and*

$$\mu \left(\left\{ \omega : \xi(\omega) \geq \int_{\Omega} \xi d\mu \right\} \right) = 1, \quad (2)$$

and

$$\mu \left(\left\{ \omega : \xi(\omega) \leq \int_{\Omega} \xi d\bar{\mu} \right\} \right) = 1; \quad (3)$$

(ii) *if further $\mu(A \cap B) = 1$ for any $A, B \in \mathcal{F}_0$ with $\mu(A) = 1$ and $\mu(B) = 1$, then*

$$\mu \left(\left\{ \omega : \int_{\Omega} \xi d\mu \leq \xi(\omega) \leq \int_{\Omega} \xi d\bar{\mu} \right\} \right) = 1. \quad (4)$$

Proof. (i) Since μ is continuous from below, we have

$$\begin{aligned} 1 = \mu(\{\omega : \xi(\omega) \in (-\infty, \infty)\}) &= \mu \left(\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) \in [-n, n]\} \right) \\ &= \lim_{n \rightarrow \infty} \mu(\{\omega : \xi(\omega) \in [-n, n]\}). \end{aligned}$$

Notice that $\mu(\mathcal{F}_0) = \{0, 1\}$ and ξ is \mathcal{F}_0 -measurable, there exists $n \in \mathbb{N}$, such that

$$\mu(\{\omega : \xi(\omega) \in [-n, n]\}) = 1.$$

It turns out that the set I is not empty, where

$$I = \{t \in \mathbb{R} : \mu(\{\omega : \xi(\omega) \geq t\}) = 1\}.$$

Define $t^* = \sup I$. Since μ is continuous from above, we get $\mu(\{\omega : \xi(\omega) \geq t^*\}) = 1$, so $t^* \in I$. Due to $\mu(\mathcal{F}_0) = \{0, 1\}$, for any $t > t^*$, $\mu(\{\omega : \xi(\omega) \geq t\}) = 0$, and for any $t \leq t^*$, $\mu(\{\omega : \xi(\omega) \geq t\}) = 1$. The above observations lead to that, if $t^* \geq 0$ then

$$\int_{\Omega} \xi d\mu = \int_0^{\infty} \mu(\{\omega : \xi(\omega) \geq t\}) dt + \int_{-\infty}^0 [\mu(\{\omega : \xi(\omega) \geq t\}) - 1] dt = \int_0^{t^*} 1 dt = t^*;$$

and if $t^* < 0$ then

$$\int_{\Omega} \xi d\mu = \int_0^{\infty} \mu(\{\omega : \xi(\omega) \geq t\}) dt + \int_{-\infty}^0 [\mu(\{\omega : \xi(\omega) \geq t\}) - 1] dt = \int_{t^*}^0 (-1) dt = t^*.$$

That is, $t^* = \int_{\Omega} \xi d\mu \in I$. Therefore, the Choquet integral of ξ with respect to μ is finite and the equality (2) holds.

Considering random variable $-\xi$ in (2), we get $\int_{\Omega} -\xi d\mu$ is finite and

$$\mu\left(\left\{\omega : -\xi(\omega) \geq \int_{\Omega} -\xi d\mu\right\}\right) = 1.$$

Thus $\int_{\Omega} \xi d\bar{\mu}$ is finite and the equality (3) holds since $\int_{\Omega} -\xi d\mu = -\int_{\Omega} \xi d\bar{\mu}$.

(ii) Under the additional assumption, equality (4) can be deduced directly from equalities (2) and (3). \square

Remark 4.1. (i) If ξ is bounded, then we can replace the continuity of μ by the continuity from above of μ , the conclusions in Theorem 4.1 still hold.

(ii) When ξ is bounded and μ is lower probability, Theorem 4.1 degenerates to Lemma 2 in [3].

If μ is a lower probability, it obviously satisfies that $\mu(A \cap B) = 1$ for any $\mu(A) = 1$ and $\mu(B) = 1$. There are also other capacities rather than lower probabilities satisfying this condition. See the following example and proposition.

Example 4.1. Let (Ω, \mathcal{F}, P) be a probability space and $f : [0, 1] \rightarrow [0, 1]$ be an increasing function with $f(0) = 0$, $f(1) = 1$. Then $\mu = f(P)$ is a capacity on \mathcal{F} , called a distorted probability. Especially if f is left (right) continuous then μ is continuous from below (above); if f is strictly increasing on point 1, then μ satisfies $\mu(A \cap B) = 1$ for $\mu(A) = 1$ and $\mu(B) = 1$.

Proposition 4.1. Let $(\mu, \bar{\mu})$ is a pair of conjugate capacities on (Ω, \mathcal{F}) . If μ is subadditive, then $\bar{\mu}$ satisfies $\bar{\mu}(A \cap B) = 1$ for $\bar{\mu}(A) = 1$ and $\bar{\mu}(B) = 1$.

Proof. It follows from the conjugate relation and the subadditivity of μ that $\bar{\mu}(A \cap B) = 1 - \mu(A^c \cup B^c) \geq 1 - (\mu(A^c) + \mu(B^c)) = 1 - (1 - \bar{\mu}(A) + 1 - \bar{\mu}(B)) = 1$. \square

Corollary 4.1. Let (V, v) be a pair of continuous upper and lower probabilities on \mathcal{F} with $v(\mathcal{F}_0) = \{0, 1\}$, then for any \mathcal{F}_0 -measurable random variable ξ , $\int_{\Omega} \xi dV$ and $\int_{\Omega} \xi dv$ are finite and

$$v \left(\left\{ \omega : \int_{\Omega} \xi dv \leq \xi(\omega) \leq \int_{\Omega} \xi dV \right\} \right) = 1, \quad (5)$$

$$V \left(\left\{ \omega : \xi(\omega) = \int_{\Omega} \xi dv \right\} \right) = 1, \quad (6)$$

$$V \left(\left\{ \omega : \xi(\omega) = \int_{\Omega} \xi dV \right\} \right) = 1. \quad (7)$$

Proof. The finiteness of $\int_{\Omega} \xi dV$ and $\int_{\Omega} \xi dv$ and equality (5) are directly from Theorem 4.1 since the lower probability v satisfies $v(A \cap B) = 1$ if $v(A) = 1$ and $v(B) = 1$.

Applying the result of Theorem 4.1 (i) to v and V , we can get the following four equalities

$$v \left(\left\{ \omega : \xi(\omega) \geq \int_{\Omega} \xi dv \right\} \right) = 1, \quad (8)$$

$$v \left(\left\{ \omega : \xi(\omega) \leq \int_{\Omega} \xi dV \right\} \right) = 1, \quad (9)$$

$$V \left(\left\{ \omega : \xi(\omega) \geq \int_{\Omega} \xi dV \right\} \right) = 1, \quad (10)$$

$$V \left(\left\{ \omega : \xi(\omega) \leq \int_{\Omega} \xi dv \right\} \right) = 1. \quad (11)$$

Since (V, v) is a pair of upper and lower probabilities, we have $V(A \cap B) = 1$ if $V(A) = 1$ and $v(B) = 1$. So (6) can be deduced from (8) and (11) while (7) can be deduced from (9) and (10). \square

Theorem 4.2. *Let μ be a continuous capacity on \mathcal{F} with $\mu(\mathcal{G}) = \{0, 1\}$, where \mathcal{G} is the set of all invariant sets under θ . For any bounded random variables ξ , there exist \mathcal{G} -measurable random variables ξ^* and ξ_* such that*

$$\mu \left(\left\{ \omega : \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) \geq \int_{\Omega} \xi_* d\mu \right\} \right) = 1, \quad (12)$$

and

$$\mu \left(\left\{ \omega : \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) \leq \int_{\Omega} \xi^* d\bar{\mu} \right\} \right) = 1. \quad (13)$$

Moreover, if μ satisfies $\mu(A \cap B) = 1$ for any $A, B \in \mathcal{G}$ with $\mu(A) = 1$ and $\mu(B) = 1$, then

$$\mu \left(\left\{ \omega : \int_{\Omega} \xi_* d\mu \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) \leq \int_{\Omega} \xi^* d\bar{\mu} \right\} \right) = 1. \quad (14)$$

Proof. Let $\xi_* = \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega))$, $\xi^* = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega))$. Notice that

$$\begin{aligned} \xi_*(\theta(\omega)) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^{k+1}(\omega)) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \left[\sum_{k=0}^n \xi(\theta^k(\omega)) - \xi(\omega) \right] \\ &= \liminf_{n \rightarrow \infty} \left[\frac{n+1}{n} \frac{1}{n+1} \sum_{k=0}^n \xi(\theta^k(\omega)) - \frac{1}{n} \xi(\omega) \right] = \xi_*(\omega), \end{aligned}$$

we have ξ_* is \mathcal{G} -measurable. Similarly, ξ^* is \mathcal{G} -measurable. Therefore (12) (13) and (14) can be derived from Theorem 4.1. \square

As mentioned in the introduction, we think $\mu(\mathcal{G}) = \{0, 1\}$ used in [3] is too permissive to define the ergodicity in a capacity space. We will use the following example to illustrate the reason.

Example 4.2. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, \mathcal{F} be all subsets of Ω . Define $\theta : \Omega \rightarrow \Omega$ by

$$\theta(\omega_1) = \omega_2, \quad \theta(\omega_2) = \omega_1, \quad \theta(\omega_3) = \omega_4, \quad \theta(\omega_4) = \omega_3.$$

Let μ_1 be a capacity on \mathcal{F} , with $\mu_1(A) = 0$ if $|A| \leq 1$, $\mu_1(A) = 1$ if $|A| \geq 2$, where $|A|$ denotes the number of elements in A . Then μ_1 is neither subadditive nor superadditive.

Let P_1 and P_2 be probabilities on \mathcal{F} with $P_1(\omega_1) = P_1(\omega_2) = \frac{1}{2}$, $P_1(\omega_3) = P_1(\omega_4) = 0$, $P_2(\omega_1) = P_2(\omega_2) = 0$, $P_2(\omega_3) = P_2(\omega_4) = \frac{1}{2}$. For any $A \in \mathcal{F}$, let $\mu_2(A) = \max_{i=1,2} P_i(A)$ be an upper probability.

Then it is easy to check that θ preserves both μ_1 and μ_2 and the set of all invariant sets is $\mathcal{G} = \{\Omega, \emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$. Here $\mu_i(\mathcal{G}) = \{0, 1\}$, $i = 1, 2$, so they together with their conjugate capacities satisfy the ergodicity definition in [3]. Note the conjugate capacity of μ_2 is a lower probability. But under these two different capacities, Ω can be split into two invariant sets $\{\omega_1, \omega_2\}$ and $\{\omega_3, \omega_4\}$ with $\mu_i(\{\omega_1, \omega_2\}) = 1$ and $\mu_i(\{\omega_3, \omega_4\}) = 1$, $i = 1, 2$. That is to say Ω is decomposable under θ .

Now we give our definition of an ergodic transformation in a capacity space.

Definition 4.1. A measurable capacity preserving transformation θ on the capacity space $(\Omega, \mathcal{F}, \mu)$ is said to be ergodic (with respect to μ) if for any θ -invariant set B the following two conditions hold:

- (i) $\mu(B) = 0$ or $\mu(B) = 1$,
- (ii) $\mu(B) = 0$ or $\mu(B^c) = 0$.

Remark 4.2. If θ is not ergodic with respect to capacity μ then the space Ω can be split into two θ -invariant sets B and B^c either each having positive capacity or one of them having positive capacity which is less than 1. This is to say θ is not “irreducible”.

The following example shows why we do not only consider (ii) in Definition 4.1 to define ergodicity.

Example 4.3. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, \mathcal{F} be all subsets of Ω and P_1, P_2, P_3 be probabilities on \mathcal{F} with $P_1(\omega_1) = 0$, $P_1(\omega_2) = P_1(\omega_3) = \frac{1}{2}$, $P_2(\omega_2) = 0$, $P_2(\omega_1) = P_2(\omega_3) = \frac{1}{2}$, $P_3(\omega_3) = 0$, $P_3(\omega_1) = P_3(\omega_2) = \frac{1}{2}$. Let $v(A) =$

$\min_{i=1,2,3} P_i(A)$, for any $A \in \mathcal{F}$, be a lower probability. We consider the following transformation θ with

$$\theta(\omega_1) = \omega_2, \quad \theta(\omega_2) = \omega_1, \quad \theta(\omega_3) = \omega_3.$$

It is easy to check that v is θ -invariant and the set of all invariant sets with respect to θ is $\mathcal{G} = \{\Omega, \emptyset, \{\omega_1, \omega_2\}, \{\omega_3\}\}$. So for any set $B \in \mathcal{G}$, $v(B) = 0$ or $v(B^c) = 0$. But $v(\mathcal{G}) = \{0, \frac{1}{2}, 1\}$, and Ω can be split into two invariant sets $\{\omega_1, \omega_2\}$ and $\{\omega_3\}$ with $v(\{\omega_1, \omega_2\}) = \frac{1}{2}$.

It is easy to check that the following proposition holds.

Proposition 4.2. *Let $(\mu, \bar{\mu})$ be a pair of conjugate capacities on (Ω, \mathcal{F}) . If μ is subadditive, then*

- (i) *for any $A \in \mathcal{F}$, $\bar{\mu}(A) \leq \mu(A)$;*
- (ii) *a transformation θ being ergodic with respect to μ is equivalent to for any θ -invariant set B , either $\mu(B) = 0$ or $\mu(B^c) = 0$;*
- (iii) *a transformation θ being ergodic with respect to μ implies that μ is additive on the θ -invariant σ -algebra;*
- (iv) *the ergodicity of θ with respect to μ implies the ergodicity of θ with respect to $\bar{\mu}$.*

Remark 4.3. *For a lower probability v , it is easy to see that $v(B) = 0$ or $v(B) = 1$ implies $v(B) = 0$ or $v(B^c) = 0$. Thus, condition (i) in Definition 4.1 is adequate to guarantee the ergodicity of v . Therefore, the definition of ergodicity given in [3] agrees with our definition in the case of lower probability. However, as $v(B) = 0$ does not imply $V(B) = 0$, so the ergodicity of θ under the lower probability does not imply the ergodicity of θ with respect to the upper probability.*

Motivated by Theorems 2.6 and 2.7 in [16], we derive the following Theorems 4.3 and 4.4 as the characterizations of ergodicity in a subadditive capacity space.

Theorem 4.3. *Let μ be a subadditive capacity on (Ω, \mathcal{F}) and θ be a measurable transformation from Ω to Ω preserving μ . Then the following four statements:*

- (i) *the transformation θ is ergodic;*
- (ii) *if every $B \in \mathcal{F}$ with $\mu(\theta^{-1}B \triangle B) = 0$, then $\mu(B) = 0$ or $\mu(B^c) = 0$;*

(iii) for every $A \in \mathcal{F}$ with $\mu(A) > 0$, we have $\mu\left(\left(\bigcup_{n=1}^{\infty} \theta^{-n}A\right)^c\right) = 0$;
(iv) for every $A, B \in \mathcal{F}$ with $\mu(A) > 0$ and $\mu(B) > 0$, there exists $n \in \mathbb{N}$ such that $\mu(B \cap \theta^{-n}A) > 0$,
have the following relations: (ii) implies (i); (iv) implies (i). If μ is continuous from below, then (i) implies (ii); (iii) implies (iv). If μ is continuous from above, then (i) implies (iii). Moreover, if μ is continuous, then all the above four statements are equivalent.

Proof. By Proposition 4.2, (ii) \Rightarrow (i) is obvious.

(iv) \Rightarrow (i). Let B be any invariant set. If $\mu(B) > 0$ and $\mu(B^c) > 0$, then by (iv) and invariant assumption of B , there exists $n \in \mathbb{N}$ such that

$$0 < \mu(B^c \cap \theta^{-n}B) = \mu(B^c \cap B) = 0$$

which derives a contradiction. Hence $\mu(B) = 0$ or $\mu(B^c) = 0$. Therefore by Proposition 4.2, (i) is proved.

(i) \Rightarrow (ii) under the continuity from below assumption of μ . Let $B \in \mathcal{F}$ with $\mu(\theta^{-1}B \triangle B) = 0$. Since for any $n \in \mathbb{N}$

$$\theta^{-n}B \triangle B \subseteq \bigcup_{k=0}^{n-1} (\theta^{-(k+1)}B \triangle \theta^{-k}B) = \bigcup_{k=0}^{n-1} \theta^{-k}(\theta^{-1}B \triangle B)$$

then by the monotonicity, subadditivity and θ -invariance of μ ,

$$\begin{aligned} \mu(\theta^{-n}B \triangle B) &\leq \mu\left(\bigcup_{k=0}^{n-1} \theta^{-k}(\theta^{-1}B \triangle B)\right) \\ &\leq \sum_{k=0}^{n-1} \mu(\theta^{-k}(\theta^{-1}B \triangle B)) = \sum_{k=0}^{n-1} \mu(\theta^{-1}B \triangle B). \end{aligned}$$

Because of $\mu(\theta^{-1}B \triangle B) = 0$, we have

$$\mu(\theta^{-n}B \triangle B) = 0. \tag{15}$$

Moreover,

$$\left(\bigcup_{k=0}^{\infty} \theta^{-k}B\right) \triangle B \subseteq \bigcup_{k=0}^{\infty} (\theta^{-k}B \triangle B),$$

thus from the monotonicity, continuity from below and subadditivity of μ and (15) we have

$$\begin{aligned} \mu \left(\left(\bigcup_{k=n}^{\infty} \theta^{-k} B \right) \triangle B \right) &\leq \mu \left(\bigcup_{k=0}^{\infty} (\theta^{-k} B \triangle B) \right) \\ &= \lim_{n \rightarrow \infty} \mu \left(\bigcup_{k=0}^n (\theta^{-k} B \triangle B) \right) \leq \sum_{k=0}^{\infty} \mu (\theta^{-k} B \triangle B) = 0. \end{aligned}$$

Immediately, we have

$$\mu \left(\left(\bigcup_{k=n}^{\infty} \theta^{-k} B \right) \setminus B \right) = 0, \quad (16)$$

and

$$\mu \left(B \setminus \left(\bigcup_{k=n}^{\infty} \theta^{-k} B \right) \right) = 0. \quad (17)$$

Define $B_{\infty} = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} \theta^{-k} B$. Combining (16), it is direct from $B_{\infty} \setminus B \subseteq (\bigcup_{k=n}^{\infty} \theta^{-k} B) \setminus B$ and the monotonicity of μ that

$$\mu(B_{\infty} \setminus B) = 0. \quad (18)$$

Meanwhile,

$$B \setminus \left(\bigcup_{k=n}^{\infty} \theta^{-k} B \right) \uparrow B \setminus \bigcap_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} \theta^{-k} B \right) = B \setminus B_{\infty},$$

together with the continuity from below of μ and (17), we have

$$\mu(B \setminus B_{\infty}) = 0. \quad (19)$$

On the other hand, B_{∞} is an invariant set since

$$\theta^{-1} B_{\infty} = \bigcap_{n=0}^{\infty} \bigcup_{k=n+1}^{\infty} \theta^{-k} B = B_{\infty}.$$

By the ergodicity assumption of μ , we have $\mu(B_{\infty}) = 0$ or $\mu(B_{\infty}^c) = 0$.

If $\mu(B_\infty) = 0$, then by the subadditivity of μ and (19), we have $\mu(B) = 0$ since

$$\mu(B) = \mu(B) - \mu(B_\infty) \leq \mu(B) - \mu(B \cap B_\infty) \leq \mu(B \setminus (B \cap B_\infty)) = \mu(B \setminus B_\infty) = 0.$$

If $\mu(B_\infty^c) = 0$, then similarly by the subadditivity of μ and (18), we have $\mu(B^c) = 0$ since

$$\begin{aligned} \mu(B^c) = \mu(B^c) - \mu(B_\infty^c) &\leq \mu(B^c) - \mu(B^c \cap B_\infty^c) \\ &\leq \mu(B^c \setminus (B^c \cap B_\infty^c)) = \mu(B^c \setminus B_\infty^c) = \mu(B_\infty \setminus B) = 0. \end{aligned}$$

Hence, the statement (ii) is proved.

(iii) \Rightarrow (iv) under the continuity from below assumption of μ . Let $A, B \in \mathcal{F}$ with $\mu(A) > 0$ and $\mu(B) > 0$. From (iii), we know $\mu\left(\left(\bigcup_{n=1}^{\infty} \theta^{-n} A\right)^c\right) = 0$. By the subadditivity, monotonicity and continuity from below of μ , we have

$$\begin{aligned} 0 < \mu(B) &\leq \mu\left(B \cap \left(\bigcup_{n=1}^{\infty} \theta^{-n} A\right)\right) + \mu\left(B \cap \left(\bigcup_{n=1}^{\infty} \theta^{-n} A\right)^c\right) \\ &\leq \mu\left(\bigcup_{n=1}^{\infty} (B \cap \theta^{-n} A)\right) + \mu\left(\left(\bigcup_{n=1}^{\infty} \theta^{-n} A\right)^c\right) \\ &= \mu\left(\bigcup_{n=1}^{\infty} (B \cap \theta^{-n} A)\right) = \lim_{k \rightarrow \infty} \mu\left(\bigcup_{n=1}^k (B \cap \theta^{-n} A)\right) \\ &\leq \sum_{n=1}^{\infty} \mu(B \cap \theta^{-n} A). \end{aligned}$$

Thus there exists $n \in \mathbb{N}$ such that $\mu(B \cap \theta^{-n} A) > 0$. Therefore the assertion (iv) is proved.

(i) \Rightarrow (iii) under the continuity from above assumption of μ . Let $A \in \mathcal{F}$ with $\mu(A) > 0$. Define

$$A_1 = \bigcup_{k=1}^{\infty} \theta^{-k} A \quad \text{and} \quad A_\infty = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \theta^{-k} A.$$

It is easy to see that

$$\theta^{-n} A_1 = \bigcup_{k=n+1}^{\infty} \theta^{-k} A \downarrow A_\infty.$$

It follows from the continuity from above and θ -invariance of μ that

$$\mu(A_\infty) = \lim_{n \rightarrow \infty} \mu(\theta^{-n} A_1) = \mu(A_1) \geq \mu(\theta^{-1} A) = \mu(A) > 0.$$

Since $\theta^{-1} A_\infty = A_\infty$, by (i), we have $\mu(A_\infty^c) = 0$. Notice that $A_\infty \subseteq \bigcup_{n=1}^{\infty} \theta^{-n} A$, therefore $\mu((\bigcup_{n=1}^{\infty} \theta^{-n} A)^c) = 0$ and (iii) is proved.

It is then obvious that all the four statements are equivalent under the continuity assumption of μ . \square

Remark 4.4. *If the upper probability V is generated by \mathcal{P} which is a subset of $\Delta^\sigma(\Omega, \mathcal{F})$, then V is continuous from below (see Lemma 2.1 in [6]). Thus, the upper probability related to G -expectation is continuous from below (see section 3.2 in [11]).*

Remark 4.5. *In Theorem 4.3, from the equivalence between (i) and (ii) when μ is continuous from below, we can define the invariant sets in a wider sense on a subadditive capacity space as almost invariant set in classical ergodic theory (see p330 in [13]). A set B is said to be μ -almost invariant with respect to θ in a subadditive capacity space $(\Omega, \mathcal{F}, \mu)$ if $\mu(\theta^{-1} B \triangle B) = 0$. If the set B is μ -almost invariant, then $\mu(\theta^{-1} B \cap B) = \mu(B) = \mu(\theta^{-1} B)$. Thus we can consider the μ -almost invariant sets when we study the ergodicity.*

The statement (iii) means that if $\mu(A) > 0$ then $\theta^k \omega$ will lie into A in finite steps μ -a.s.

The statement (iv) means that if $\mu(A) > 0$ and $\mu(B) > 0$, then those ω starting from B , with the flow $\theta^k \omega$ arriving into A in finite steps have positive capacity.

Theorem 4.4. *Let μ be a subadditive capacity on (Ω, \mathcal{F}) and θ be a measurable transformation from Ω to Ω preserving μ . Then the following three statements:*

- (i) the transformation θ is ergodic;*
- (ii) if $\xi : \Omega \rightarrow \mathbb{R}$ is bounded measurable and $\xi(\theta \cdot) = \xi(\cdot)$, then ξ is a constant μ -a.s.;*
- (iii) if $\xi : \Omega \rightarrow \mathbb{R}$ is measurable and $\xi(\theta \cdot) = \xi(\cdot)$ μ -a.s., then ξ is a constant μ -a.s.,*

have the following relations: (iii) implies (ii); (ii) implies (i). Moreover, if μ is continuous from below, then (ii) is equivalent to (i). If further μ is continuous, then all three statements are equivalent.

Proof. The proof of (iii) \Rightarrow (ii) is trivial.

We now prove (ii) \Rightarrow (i). For any invariant set A , $I_A(\theta\omega) = I_A(\omega)$. Thus, by (ii), I_A is a constant μ -a.s. So $I_A = 0$ or 1 μ -a.s. If $I_A = 0$ μ -a.s., then $\mu(A) = 0$. If $I_A = 1$ μ -a.s., then $\mu(A^c) = 0$. Thus θ is ergodic.

Next we prove (i) \Rightarrow (iii) under the assumption that μ is continuous.

For any $t \in \mathbb{R}$, let $A_t = \{\omega : \xi(\omega) > t\}$ and $A_t^c = \{\omega : \xi(\omega) \leq t\}$. Notice that $\theta^{-1}A_t \triangle A_t \subseteq \{\omega : \xi(\theta\omega) \neq \xi(\omega)\}$, we have $\mu(\theta^{-1}A_t \triangle A_t) = 0$ due to $\xi(\theta \cdot) = \xi(\cdot)$ μ -a.s. Since θ is ergodic and μ is continuous from below, by Theorem 4.3, we know that $\mu(A_t) = 0$ or $\mu(A_t^c) = 0$. Thus, $\mu(A_t) = 0$ or 1 .

Let $I = \{t : \mu(A_t) = 0\}$. By the continuity from above of μ , we have

$$0 = \mu(\{\omega : \xi(\omega) = \infty\}) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

Thus there exists $n \in \mathbb{N}$ such that $\mu(A_n) = 0$, that is $n \in I \neq \emptyset$. So set $t_* = \inf I$, and immediately $t_* \in I$ since μ is continuous from below. Hence, for any $t \geq t_*$, we have $\mu(A_t) = 0$ and for any $t < t_*$, we have $\mu(A_t) = 1$ and $\mu(A_t^c) = 0$. Due to the continuity from below of μ , we have $\mu(\{\omega : \xi(\omega) < t_*\}) = 0$. Combining $\mu(\{\omega : \xi(\omega) > t_*\}) = 0$ and the subadditivity of μ , we get $\mu(\{\omega : \xi(\omega) \neq t_*\}) = 0$. Thus ξ is a constant t_* μ -a.s.

From the proof of (i) \Rightarrow (iii) we can see the continuity from above of μ is used to prove $I \neq \emptyset$. But if ξ is bounded, of course $I \neq \emptyset$, so we do not need the continuity from above of μ assumption. Therefore, if μ is continuous from below, we can prove (i) \Rightarrow (ii) in the same way as (i) \Rightarrow (iii). \square

Remark 4.6. *The result can be also presented in the language of transformation operator U defined by $U(\xi)(\omega) = \xi(\theta\omega)$ on the space of measurable functions on Ω . Theorem 4.4 says that a map is ergodic to a subadditive continuous capacity if and only if the transformation operator U has eigenvalue 1 which is simple. In classical probability space case, see Da Prato and Zabczyk [8] or Walters [28].*

Example 4.4. *Consider Example 4.1.1 introduced by Feng, Qu and Zhao in [15]. Let $\Omega_1 = [0, 1)$, $\theta_\alpha : \Omega_1 \rightarrow \Omega_1$, $\theta_\alpha(x) = (x + \alpha) \bmod 1$, where α is a irrational number. Consider $\hat{\Omega} = [0, 2) = \Omega_1 \cup \Omega_2$, where $\Omega_2 = [1, 2)$. Define $\hat{\theta}_\alpha : \hat{\Omega} \rightarrow \hat{\Omega}$ by*

$$\hat{\theta}_\alpha(x) = \begin{cases} \theta_\alpha(x) + 1, & x \in \Omega_1, \\ x - 1, & x \in \Omega_2. \end{cases}$$

Let P_i be the Lebesgue measure on Ω_i , $i = 1, 2$. From the classical ergodic theory, we know that $(\Omega_i, \mathcal{B}(\Omega_i), P_i, (\hat{\theta}_\alpha^{2n}|_{\Omega_i})_{n \geq 1})$, $i = 1, 2$ are ergodic dynamical systems.

Define

$$\bar{P}_i(A) := P_i(A \cap \Omega_i), \text{ for any } A \in \mathcal{B}(\hat{\Omega}),$$

$$\bar{P} := \frac{1}{2}(\bar{P}_1 + \bar{P}_2), \quad V := \bar{P}_1 \vee \bar{P}_2, \quad v := \bar{P}_1 \wedge \bar{P}_2,$$

and for all random variable X such that $\mathbb{E}_{\bar{P}_1}[X] \vee \mathbb{E}_{\bar{P}_2}[X] < \infty$, define

$$\hat{\mathbb{E}}[X] := \mathbb{E}_{\bar{P}_1}[X] \vee \mathbb{E}_{\bar{P}_2}[X].$$

Then, by the ergodic Theorem of periodic measure in [17] (Theorem 2.20), \bar{P} is an invariant measure with respect to $\hat{\theta}_\alpha$ and ergodic. By classical Birkhoff's ergodic theorem (see [2]), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\hat{\theta}_\alpha^k \cdot) = \mathbb{E}_{\bar{P}}[\xi] = \frac{1}{2}(\mathbb{E}_{\bar{P}_1}[\xi] + \mathbb{E}_{\bar{P}_2}[\xi]), \quad \bar{P}\text{-a.s.}$$

It is easy to check that \bar{P} -a.s. and V -a.s. are equivalent. Hence,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\hat{\theta}_\alpha^k \cdot) = c, \quad V\text{-a.s.},$$

and $\int_{\hat{\Omega}} \xi dv \leq c = \mathbb{E}_{\bar{P}}[\xi] \leq \int_{\hat{\Omega}} \xi dV$.

On the other hand, by Proposition 4.5 in [15], we know that $(\hat{\Omega}, \mathcal{B}(\hat{\Omega}), \hat{\mathbb{E}}, (\hat{\theta}_\alpha^n)_{n \geq 1})$ is an ergodic sublinear dynamical system. By the definition of ergodicity of dynamical systems under sublinear expectation (Definition 4.2 in [15]), if B is an invariant set, then $\hat{\mathbb{E}}[I_B] = 0$ and $\hat{\mathbb{E}}[I_{B^c}] = 0$. Thus $V(B) = 0$ or $V(B^c) = 0$. So, $\hat{\theta}_\alpha$ is ergodic with respect to V .

Example 4.5. Consider the distorted probability $f(P)$ (Example 4.1). If f is a concave function then $V := f(P)$ is a concave capacity (Example 2.1 in [12]), and hence V is an upper probability (Proposition 10.3 in [12]) and its conjugate capacity v is a lower probability. If further f is strict increasing on points 0 and 1, then for $A \in \mathcal{F}$

$$P(A) = 0 \Leftrightarrow V(A) = 0 \Leftrightarrow v(A) = 0, \quad P(A) = 1 \Leftrightarrow V(A) = 1 \Leftrightarrow v(A) = 1.$$

Let θ be an ergodic transformation with respect to P . By the classical Birkhoff's ergodic theorem (see [2]), for any bounded \mathcal{F} -measurable random variable ξ ,

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot) = E_P[\xi] \right) = 1,$$

therefore,

$$v \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot) = E_P[\xi] \right) = 1,$$

that is $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ is a constant V -a.s. and this constant lies in $[\int_{\Omega} \xi dv, \int_{\Omega} \xi dV]$. Notice that θ is also ergodic with respect to both v and V .

Motivated by above two examples, we will show that for a continuous upper probability V , the ergodic average is a constant V -a.s. Moreover, we give a characterization of ergodicity through strong law of large numbers which is a type of Birkhoff's ergodic theorem in an upper probability space. If further V is concave, then we will give an estimate of the ergodic average constant.

Theorem 4.5. *Let V be a continuous upper probability on \mathcal{F} and θ be a measurable transformation from Ω to Ω preserving V . Then θ is ergodic with respect to V if and only if for any bounded \mathcal{F} -measurable random variable ξ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$$

is a constant V -a.s.

Proof. For any bounded \mathcal{F} -measurable random variable ξ , let v be the conjugate lower probability of V ,

$$A = \left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \text{ exists} \right\} \quad (20)$$

and

$$\tilde{\xi}(\omega) = \begin{cases} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega), & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}.$$

By Proposition 3.1 and Lemma 3.1, $v(A) = 1$. Therefore $\tilde{\xi}$ is a bounded \mathcal{F} -measurable random variable with $\tilde{\xi}(\theta \cdot) = \tilde{\xi}(\cdot)$ V -a.s. and $\tilde{\xi}(\cdot) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ V -a.s.

If θ is ergodic with respect to V , then by Theorem 4.4, $\tilde{\xi}$ is a constant V -a.s. Therefore $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ is a constant V -a.s.

If for any bounded \mathcal{F} -measurable random variable ξ , $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ is a constant V -a.s. We consider ξ with $\xi(\theta \cdot) = \xi(\cdot)$, then $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) = \xi(\omega)$. Hence ξ is a constant V -a.s. It follows from Theorem 4.4 that θ is ergodic with respect to V . \square

Example 4.6. Let (V, v) denote the g -probabilities generated by the backward stochastic differential equations (or g -expectations) with generators $g(z) = k|z|$ and $g = -k|z|$ respectively, where k is any fixed number in \mathbb{R}^+ . Then (V, v) is a pair of continuous upper-lower probabilities (see Example 1 in [5]). Thus, our results in Theorems 4.3, 4.4 and 4.5 apply immediately to this case.

The following example will show that if θ is ergodic with respect to v not to V , then we can not get the ergodic average is a constant V -a.s.

Example 4.7. Consider $(\Omega, \mathcal{F}, \theta)$ in Example 4.2 and let V be the upper probability μ_2 and v be the corresponding lower probability of V . Define $\xi(\omega) = I_{\{\omega_1, \omega_2\}}(\omega)$, then for all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) = \xi(\omega).$$

In this example, θ is ergodic with respect to v but not to V and for any constant c ,

$$v \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) = c \right\} \right) = 0.$$

Theorem 4.6. Let V be a continuous and concave capacity on \mathcal{F} , and v be the conjugate capacity to V . If θ is ergodic with respect to V and ξ is a bounded \mathcal{F} -measurable random variable, then there exists a constant c such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot) = c, \quad V\text{-a.s. and } c \in \left[\int_{\Omega} \xi dv, \int_{\Omega} \xi dV \right].$$

Proof. It follows from Proposition 10.3 in [12] that (V, v) is a pair of upper and lower probabilities. Then, by Theorem 4.5, there exists a constant c such that

$$v \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) = c \right\} \right) = 1.$$

We only need to prove $c \in [\int_{\Omega} \xi dV, \int_{\Omega} \xi dV]$. Since θ is ergodic with respect to V , we have $v(\mathcal{G}) = \{0, 1\}$. It follows from Theorem 4.2 that

$$v \left(\left\{ \omega : \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \leq \int_{\Omega} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) dV \right\} \right) = 1.$$

Thus, $c \leq \int_{\Omega} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) dV$. Let A be the set defined by (20) in the proof of Theorem 4.5, then A is a θ -invariant set and $V(A^c) = 0$. Combining above two equalities, by the subadditivity of the Choquet integral with respect to V (Theorem 6.3 in [12]), we get

$$\begin{aligned} c &\leq \int_{\Omega} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \right) I_A(\omega) dV + \int_{\Omega} \left(\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \right) I_{A^c}(\omega) dV \\ &= \int_{\Omega} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \right) I_A(\omega) dV \\ &= \lim_{n \rightarrow \infty} \int_{\Omega} \left(\frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \omega) \right) I_A(\omega) dV \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \int_{\Omega} \xi(\theta^k \omega) I_A(\theta^k \omega) dV \\ &= \int_{\Omega} \xi(\omega) I_A(\omega) dV \\ &= \int_{\Omega} \xi(\omega) dV, \end{aligned}$$

where the second equality is due to the dominated convergence theorem (Lemma 2.2) and the penultimate equality is come from θ preserving the Choquet integral with respect to V (Proposition 3.1). Now we consider $-\xi$ and $-c \leq \int_{\Omega} -\xi(\omega) dV$, that is $c \geq \int_{\Omega} \xi(\omega) dv$. The proof of this theorem is completed. \square

Remark 4.7. *Cerreia-Vioglio, Maccheroni and Marinacci obtained that for a θ -invariant continuous lower probability v and any bounded \mathcal{F} -measurable random variable ξ , $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ exists V -a.s. If θ is ergodic in their sense, they showed that the random variable $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ lies in the interval*

$$\left[\int_{\Omega} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) dV, \int_{\Omega} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k(\omega)) dV \right]$$

V -a.s. (Theorem 2 in [3]). And then, for a continuous convex capacity v , if it is further strongly invariant that requires all the probabilities in $\text{core}(v)$ need to be θ -invariant, they confirmed the interval is $[\int_{\Omega} \xi dV, \int_{\Omega} \xi dV]$ (Corollary 2 in [3]). Our result in Theorem 4.5 says that the ergodicity of θ with respect to V is equivalent to $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(\theta^k \cdot)$ being a constant V -a.s. This means Birkhoff's strong law of large numbers is a necessary and sufficient condition for ergodicity. Moreover, for a convex continuous capacity v , Theorem 4.6 shows that the constant lies in $[\int_{\Omega} \xi dV, \int_{\Omega} \xi dV]$. For this result we do not need the assumption of v being strongly invariant.

5. Ergodicity of stationary processes on capacity spaces

The notion of stationary stochastic process on a capacity space generalizing the usual notion of stationary stochastic process in classical probability space, was given in [3] as follows.

Definition 5.1. *Given a capacity space $(\Omega, \mathcal{F}, \mu)$, we say that stochastic process $\{Y_n\}_{n \in \mathbb{N}}$ is stationary if and only if for each $n \in \mathbb{N}$, $k \in \mathbb{N}_0$ and Borel subset A of \mathbb{R}^{k+1} ,*

$$\mu(\{\omega : (Y_n(\omega), \dots, Y_{n+k}(\omega)) \in A\}) = \mu(\{\omega : (Y_{n+1}(\omega), \dots, Y_{n+1+k}(\omega)) \in A\}).$$

It is easy to see that $\{Y_n\}_{n \in \mathbb{N}}$ is stationary on $(\Omega, \mathcal{F}, \mu)$ if and only if $\{Y_n\}_{n \in \mathbb{N}}$ is stationary on the conjugate capacity space $(\Omega, \mathcal{F}, \bar{\mu})$. In classical probability theory, independent identically distributed random variables sequence must be stationary. However, such a result will not be valid in the capacity theory, we can find such a case in Example 2.1.

Let $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ denote the space of sequences endowed with the σ -algebra generated by the set of all cylinders \mathcal{C} . We denote a generic element of $\mathbb{R}^{\mathbb{N}}$ by \mathbf{x} . Any set C in \mathcal{C} called cylinder, has the following form

$$C = \{\mathbf{x} = (x_1, x_2, x_3, \dots) : (x_1, \dots, x_n) \in H\} \quad (21)$$

where $n \in \mathbb{N}$ and $H \in \mathcal{B}(\mathbb{R}^n)$. It is well known that \mathcal{C} is an algebra. We consider the shift transformation $\tau : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}}$ defined by

$$\tau(\mathbf{x}) = \tau(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots), \quad \text{for any } \mathbf{x} = (x_1, x_2, x_3, \dots) \in \mathbb{R}^{\mathbb{N}}.$$

The stochastic process $\{Y_n\}_{n \in \mathbb{N}}$ induces a measurable map from (Ω, \mathcal{F}) to $(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C}))$ by

$$\omega \mapsto \mathbf{Y}(\omega) = (Y_1(\omega), Y_2(\omega), Y_3(\omega), \dots), \quad \text{for any } \omega \in \Omega.$$

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

$$\mu_{\mathbf{Y}}(C) = \mu(\mathbf{Y}^{-1}(C)), \quad \text{for any } C \in \sigma(\mathcal{C}).$$

It is easy to check that $\mu_{\mathbf{Y}}$ is a capacity on $\sigma(\mathcal{C})$ and $\mu_{\mathbf{Y}}$ is continuous/convex/concave if μ is continuous/convex/concave respectively, as $\mathbf{Y}^{-1}\left(\bigcup_{n=1}^{\infty} C_n\right) = \bigcup_{n=1}^{\infty} \mathbf{Y}^{-1}(C_n)$ and $\mathbf{Y}^{-1}\left(\bigcap_{n=1}^{\infty} C_n\right) = \bigcap_{n=1}^{\infty} \mathbf{Y}^{-1}(C_n)$, for any $\{C_n\}_{n \in \mathbb{N}} \subseteq \sigma(\mathcal{C})$. Moreover, $\overline{\mu_{\mathbf{Y}}} = \bar{\mu}_{\mathbf{Y}}$.

Proposition 5.1. *Let $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}}$ be a stochastic process on the capacity space $(\Omega, \mathcal{F}, \mu)$ and μ be continuous. Then $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}}$ is stationary if and only if $\mu_{\mathbf{Y}}$ is the shift transformation τ -invariant.*

Proof. On one hand, assume $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}}$ is stationary and let

$$\mathcal{M} = \{A : A \in \sigma(\mathcal{C}), \mu_{\mathbf{Y}}(\tau^{-1}(A)) = \mu_{\mathbf{Y}}(A)\}.$$

Then by the stationarity of \mathbf{Y} , for any $C \in \mathcal{C}$ with H given in (21) corresponding to set C , we have

$$\begin{aligned} \mu_{\mathbf{Y}}(\tau^{-1}(C)) &= \mu(\{\omega : \tau \mathbf{Y}(\omega) \in C\}) = \mu(\{\omega : (Y_2(\omega), \dots, Y_{n+1}(\omega)) \in H\}) \\ &= \mu(\{\omega : (Y_1(\omega), \dots, Y_n(\omega)) \in H\}) \\ &= \mu(\{\omega : \mathbf{Y}(\omega) \in C\}) = \mu_{\mathbf{Y}}(C), \end{aligned}$$

which implies that $\mathcal{C} \subseteq \mathcal{M}$. Because of the continuity of μ , $\mu_{\mathbf{Y}}$ is continuous and then \mathcal{M} is a monotone class. Therefore, by the monotone class theorem we can get $\mathcal{M} = \sigma(\mathcal{C})$ which means $\mu_{\mathbf{Y}}$ is the shift transformation τ -invariant.

On the other hand, assume that $\mu_{\mathbf{Y}}$ is the shift transformation τ -invariant, then for each $n \in \mathbb{N}$, $k \in \mathbb{N}_0$ and $A \in \mathcal{B}(\mathbb{R}^{k+1})$,

$$\begin{aligned} \mu(\{\omega : (Y_n(\omega), \dots, Y_{n+k}(\omega)) \in A\}) &= \mu_{\mathbf{Y}}(\mathbb{R}^{n-1} \times A \times \mathbb{R}^{\mathbb{N}-n-k}) \\ &= \mu_{\mathbf{Y}}(\tau^{-1}(\mathbb{R}^{n-1} \times A \times \mathbb{R}^{\mathbb{N}-n-k})) \\ &= \mu_{\mathbf{Y}}(\mathbb{R}^n \times A \times \mathbb{R}^{\mathbb{N}-n-k-1}) \\ &= \mu(\{\omega : (Y_{n+1}(\omega), \dots, Y_{n+1+k}(\omega)) \in A\}). \end{aligned}$$

That is to say $\{Y_n\}_{n \in \mathbb{N}}$ is stationary. \square

Definition 5.2. *The stochastic process $\{Y_n\}_{n \in \mathbb{N}}$ on capacity space $(\Omega, \mathcal{F}, \mu)$ is called ergodic if the shift transformation τ is ergodic with respect to $\mu_{\mathbf{Y}}$.*

Now we give the strong law of large numbers for stationary and ergodic stochastic sequences on an upper probability space.

Theorem 5.1. *Let (V, v) be a pair of continuous upper and lower probabilities on \mathcal{F} . If a bounded stationary process $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}}$ on capacity space (Ω, \mathcal{F}, V) is ergodic, then there exists a constant c such that*

$$v \left(\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k(\omega) = c \right\} \right) = 1.$$

If further V is concave, then $c \in [\int_{\Omega} Y_1 dv, \int_{\Omega} Y_1 dV]$.

Proof. It is easy to check that $(V_{\mathbf{Y}}, v_{\mathbf{Y}})$ is a pair of continuous upper and lower probabilities on $\sigma(\mathcal{C})$ since (V, v) is a pair of continuous upper and lower probabilities on \mathcal{F} . By Proposition 5.1, $v_{\mathbf{Y}}$ and $V_{\mathbf{Y}}$ are the shift transformation τ -invariant. Define $\xi : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ by $\xi(\mathbf{x}) = \xi(x_1, x_2, x_3, \dots) = x_1$, for any $\mathbf{x} = (x_1, x_2, x_3, \dots) \in \mathbb{R}^{\mathbb{N}}$. Since τ is ergodic with respect to $V_{\mathbf{Y}}$, then we can get the following equality by Theorem 4.5 that there exists a constant c such that

$$v_{\mathbf{Y}} \left(\left\{ \mathbf{x} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi(\tau^{k-1} \mathbf{x}) = c \right\} \right) = 1.$$

Notice that $\frac{1}{n} \sum_{k=1}^n x_k = \frac{1}{n} \sum_{k=1}^n \xi(\tau^{k-1} \mathbf{x})$, we have

$$1 = v_{\mathbf{Y}} \left(\left\{ \mathbf{x} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n x_k = c \right\} \right) = v \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k(\omega) = c \right\} \right).$$

If further V is concave, then $V_{\mathbf{Y}}$ is concave while v and $v_{\mathbf{Y}}$ are convex. Thus, it follows from Theorem 4.6 that $c \in [\int_{\mathbb{R}^N} \xi dv_{\mathbf{Y}}, \int_{\mathbb{R}^N} \xi dV_{\mathbf{Y}}]$. By the transformation rule of Choquet integral (see Proposition 5.1 in [12]), we have

$$\int_{\mathbb{R}^N} \xi dv_{\mathbf{Y}} = \int_{\mathbb{R}^N} \xi dv(\mathbf{Y}^{-1}) = \int_{\Omega} \xi(\mathbf{Y}) dv = \int_{\Omega} Y_1 dv$$

and similarly

$$\int_{\mathbb{R}^N} \xi dV_{\mathbf{Y}} = \int_{\Omega} Y_1 dV.$$

As a consequence, we completed the proof of Theorem 5.1. \square

Corollary 5.1. *Let (V, v) be a pair of continuous upper and lower probabilities on \mathcal{F} . If a bounded stochastic process $\mathbf{Y} = \{Y_n\}_{n \in \mathbb{N}}$ on the capacity space (Ω, \mathcal{F}, V) is stationary and for any $n \in \mathbb{N}$, $\sigma(Y_k, k \leq n)$ and $\sigma(Y_k, k \geq n+1)$ are independent with respect to V . Then there exists a constant c such that*

$$v \left(\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k(\omega) = c \right\} \right) = 1.$$

If further V is concave, then $c \in [\int_{\Omega} Y_1 dv, \int_{\Omega} Y_1 dV]$.

Proof. From Theorem 5.1, we only need to prove that the shift transformation τ is ergodic with respect to $V_{\mathbf{Y}}$. Let B be any τ -invariant set, then for any $n \in \mathbb{N}$

$$\begin{aligned} \mathbf{Y}^{-1}(B) &= \mathbf{Y}^{-1}(\tau^{-1}(B)) = \mathbf{Y}^{-1}(\tau^{-n}(B)) = \{\omega : \tau^n \mathbf{Y}(\omega) \in B\} \\ &= \{\omega : (Y_{n+1}(\omega), Y_{n+2}(\omega), \dots) \in B\} \in \sigma(Y_k, k \geq n+1). \end{aligned}$$

So $\mathbf{Y}^{-1}(B) \in \mathcal{T}$. By the assumption that for any $n \in \mathbb{N}$, $\sigma(Y_k, k \leq n)$ and $\sigma(Y_k, k \geq n+1)$ are independent with respect to V , it follows from the Kolmogorov 0-1 Law in capacity spaces (Theorem 2.1) that $V(\mathbf{Y}^{-1}(B)) = 0$ or $V(\mathbf{Y}^{-1}(B)^c) = 0$. Therefore $V_{\mathbf{Y}}(B) = 0$ or $V_{\mathbf{Y}}(B^c) = 0$, that is τ is ergodic with respect to $V_{\mathbf{Y}}$ by Proposition 4.2. \square

Remark 5.1. *There are many references on strong law of large numbers for capacities under different definitions of independence and identical distributions, see for example [4], [6], [9], [14], [21], [22], [23], [26], [30] and references therein. Comparing with these papers, in this paper, we replace the independent identically distributed hypothesis by the stationarity and ergodicity as in [3]. It was obtained in [3] that the empirical average exists V -a.s. By strengthening the definition of ergodicity, we obtain that the empirical average is constant V -a.s. This is a property that was not present in previous work.*

Acknowledgements

The authors would like to thank the anonymous referees whose comments led to improvements of the paper. The work of the first and the third authors were supported by the Royal Society Newton Fund (Ref No. NA150344), with the third author being supported by EPSRC grant of UK (Ref No. EP/S005293/1). The work of the second author was supported by the National Key R&D Program of China (Ref No. 2018YFA0703900) and the National Natural Science Foundation of China (Ref Nos. 11601280, 11971266).

References

References

- [1] A. Billot, From fuzzy set theory to non-additive probabilities: How have economists reacted?, *Fuzzy Set. Syst.* 49 (1992) 75–90.
- [2] G. D. Birkhoff, Proof of the ergodic theorem, *Proc. Natl. Acad. Sci. USA.* 17 (1931) 656–660.
- [3] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, Ergodic theorems for lower probabilities, *Proc. Amer. Math. Soc.* 144 (2016) 3381–3396.
- [4] Z. Chen, Strong laws of large numbers for sub-linear expectations, *Sci. China Math.* 59 (2016) 945–954.
- [5] Z. Chen, W. Huang, P. Wu, Extension of the strong law of large numbers for capacities, *Math. Control Relat. F.* 9(2019) 175–190.
- [6] Z. Chen, P. Wu, B. Li, A strong law of large numbers for non-additive probabilities, *Internat. J. Approx. Reason.* 54 (2013) 365–377.

- [7] G. Choquet, Theory of capacities, *Ann. Inst. Fourier, Grenoble* 5 (1953) 131-295.
- [8] G. Da Prato, J. Zabczyk, *Ergodicity for infinite dimensional systems*, London Mathematical Society Lecture Note Series, 229, Cambridge University Press, Cambridge, 1996.
- [9] G. De Cooman, E. Miranda, Weak and strong laws of large numbers for coherent lower previsions, *J. Statist. Plann. Inference* 138 (2008) 2409–2432.
- [10] G. De Cooman, J. De Bock, S. Lopatatzidis, Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems, *Internat. J. Approx. Reason.* 76 (2016) 18–46.
- [11] L. Denis, M. Hu, S. Peng, Function spaces and capacity related to a sublinear expectation: application to G -Brownian motion paths, *Potential Anal.* 34 (2011) 139–161.
- [12] D. Denneberg, *Non-additive Measure and Integral*, Theory and Decision Library. Series B: Mathematical and Statistical Methods, 27. Kluwer Academic Publishers Group, Dordrecht 1994.
- [13] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge Series in Statistical and Probabilistic Mathematics, 31, Cambridge University Press, Cambridge, 2010.
- [14] L. Epstein, and Schneider, D. IID: Independently and indistinguishably distributed, *J. Econom. Theory* 113 (2003) 32–50.
- [15] C. Feng, B. Qu, H. Zhao, A sufficient and necessary condition of PS-ergodicity of periodic measures and generated ergodic upper expectations, 2018, available at [arXiv:1806.03680v1](https://arxiv.org/abs/1806.03680v1).
- [16] C. Feng, H. Zhao, Ergodicity on sublinear expectation spaces, 2017, available at [arXiv:1705.03549](https://arxiv.org/abs/1705.03549).
- [17] C. Feng, H. Zhao, Random periodic processes, periodic measures and ergodicity, 2019, available at [arXiv:1408.1897v6](https://arxiv.org/abs/1408.1897v6).

- [18] I. Gilboa, *Uncertainty in Economic Theory: Essays in Honor of David Schmeidler's 65th Birthday*, Routledge Frontiers of Political Economy, 63. Routledge, New York, 2004.
- [19] M. Hu, H. Li, F. Wang, G. Zheng, Invariant and ergodic nonlinear expectations for G-diffusion processes, *Electron. Commun. Probab.* 20 (2015) 1-15.
- [20] Y. Kifer, *Ergodic theory of random transformations*, Progress in Probability and Statistics, 10, Birkäuser Boston, Inc., Boston, 1986.
- [21] F. Maccheroni, M. Marinacci, A strong law of large number for capacities, *Ann. Probab.* 33 (2005) 1171–1178.
- [22] M. Marinacci, Limit laws for non-additive probabilities and their frequentist interpretation, *J. Econom. Theory* 84 (1999) 145–195.
- [23] S. Peng, Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations, *Sci. China Ser. A* 52 (2009) 1391–1411.
- [24] Y. Rébillé, Autocontinuity and convergence theorems for the Choquet integral, *Fuzzy Set. Syst.* 194 (2012) 52–65.
- [25] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica* 57 (1989) 571–587.
- [26] P. Terán, Laws of large numbers without additivity, *Trans. Amer. Math. Soc.* 366 (2014) 5431–5451.
- [27] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Monographs on Statistics and Applied Probability, 42, Chapman and Hall, Ltd., London, 1991.
- [28] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, 79, Springer-Verlag, New York-Berlin, 1982.
- [29] Z. Wang, J. G. Klir, *Generalized Measure Theory*, IFSR International Series on Systems Science and Engineering 25, Springer, New York, 2009.
- [30] L. Zhang, Rosenthal's inequalities for independent and negatively dependent random variables under sub-linear expectations with applications, *Sci. China Math.* 59 (2016) 751–768.