ERGODICITY AND MIXING OF INVARIANT CAPACITIES AND APPLICATIONS

BY CHUNRONG FENG[1,a], WEN HUANG[2,b] CHUNLIN LIU [3,c] HUAIZHONG ZHAO[1,d]

[1]*Department of Mathematical Sciences, Durham University, DH1 3LE, United Kingdom*

[2]*School of Mathematical Sciences, University of Science and Technology of China, Hefei, Anhui, 230026, P.R. China*

[3]*School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, P.R. China*

[a]*chunrong.feng@durham.ac.uk*; [b]*wenh@mail.ustc.edu.cn*; [c]*chunlinliu@mail.ustc.edu.cn*; [d]*huaizhong.zhao@durham.ac.uk*

We introduce the notion of common conditional expectation to investigate Birkhoff's ergodic theorem and subadditive ergodic theorem for invariant upper probabilities. If, in addition, the upper probability is ergodic, we construct an invariant probability to characterize the limit of the ergodic mean. Moreover, this skeleton probability is the unique ergodic probability in the core of the upper probability, that is equal to all probabilities in the core on all invariant sets. We have the following applications of these two theorems:

- provide a strong law of large numbers for ergodic stationary sequence on upper probability spaces;
- prove the multiplicative ergodic theorem on upper probability spaces;
- establish a criterion for the ergodicity of upper probabilities in terms of independence.

Furthermore, we introduce and study weak mixing for capacity preserving systems. Using the skeleton idea, we also provide several characterizations of weak mixing for invariant upper probabilities.

Finally, we provide examples of ergodic and weakly mixing capacity preserving systems. As applications, we obtain new results in the classical ergodic theory, e.g., in characterizing dynamical properties on probability preserving systems, such as weak mixing, periodicity. Moreover, we use our results in the nonlinear theory to deduce the asymptotic independence, Birkhoff's type ergodic theorem, subadditive ergodic theorem, and multiplicative ergodic theorem for non-invariant probabilities.

CONTENTS

MSC2020 subject classifications: Primary 60A10, 28A12; secondary 28D05, 37A25.

Keywords and phrases: capacity; Choquet integral; weak mixing; law of large numbers; ergodic theorem for non-invariant probability.

1. Introduction. Ergodic theory is a branch of mathematics that focuses on studying the behaviour of a given probability preserving system $(\Omega, \mathcal{F}, P, T)$, where $(\Omega, \mathcal{F}, P)$ is a probability space (i.e., Ω is a nonempty set, $\mathcal{F}$ is a σ-algebra on Ω, and P is a probability on $\mathcal{F}$), and $T : \Omega \to \Omega$ is a measurable transformation such that $P(T^{-1}A) = P(A)$ for any $A \in \mathcal{F}$. One of the key theorems in ergodic theory is Birkhoff's Ergodic Theorem [5], which provided a rigorous mathematical framework to investigate the Boltzmann Ergodic Hypothesis, that is, for a closed system, the time averages of a physical quantity over long periods would converge to the ensemble average. Furthermore, it has many connections with other fields apart from ergodic theory, for example, number theory (c.f. Einsiedler and Ward [19] and Furstenberg [24]), stationary process (c.f. Doob [17]), and harmonic analysis (c.f. Rosenblatt and Wierdl [39]). Afterward, in order to address the conjecture for subadditive stochastic processes raised by Hammersley and Welsh [27], Kingman [29, 30] extended Birkhoff's Ergodic Theorem to subadditive sequences, which became known as the subadditive ergodic theorem. Meanwhile, it also has many other applications, for example, the study of the multiplicative ergodic theorem by Oseledec [34].

As research in ergodic theory has progressed, many generalizations and applications of these two theorems have been obtained. However, a majority of this research has focused on probability preserving systems. In real-world scenarios, it is often the case that we cannot find an ideal situation where the probability can be determined exactly. For example, it has been shown that the classical probability theory on measurable space may not be sufficient for modelling such situations, as in economics (c.f. Billot [4], Marinacci and Montrucchio[33], and Schmeidler [42]) and statistics (c.f. Walley [45]). To address this challenge, capacities (or non-additive probabilities) and nonlinear expectations are used as a tool to model heterogeneous environments, such as financial markets where biassed beliefs of future price movements drive decisions of the stock market participants and create ambiguous volatility. Moreover, the following example is well known in number theory.

EXAMPLE 1. Recall the definition of upper density for a subset A of $\mathbb{Z}$, given by

$$\bar{d}(A) = \limsup_{n\to\infty} \frac{1}{2n+1}|A \cap [-n,n]|,$$

where $|A|$ denotes the number of elements of A. Let $T:\mathbb{Z}\to\mathbb{Z}, x\mapsto x+1$ and $2^{\mathbb{Z}}$ be the family consisting of all subsets of $\mathbb{Z}$. Then $(\mathbb{Z}, 2^{\mathbb{Z}}, \bar{d}, T)$ is a capacity preserving system, but not a probability preserving system, as $\bar{d}$ is not additive.

Thus, the study of dynamical systems on a capacity space is a natural and necessary extension. Due to the loss of additivity, many classical results in probability theory and ergodic theory may fail. So far, there are only two main works on invariant capacities by Cerreia-Vioglio, Maccheroni and Marinacci [9] and Feng, Wu and Zhao [21], and one work on invariant sublinear expectations by Feng and Zhao [22]. In this paper, we focus on invariant capacities. Following ideas in classical ergodic theory, we call $(\Omega, \mathcal{F}, \mu, T)$ a capacity preserving system if $(\Omega, \mathcal{F})$ is a measurable space, $T:\Omega\to\Omega$ is a measurable transformation and μ is a (T-)invariant capacity (see Definition 2.10). In this paper, we mainly consider a special type of capacity, namely the upper probabilities (see (6) for definition).

Firstly, we introduce the concept of common conditional expectation (see Definition 2.15) to study the ergodic theory on capacity spaces. In particular, given a standard measurable space $(\Omega, \mathcal{F})$, and a measurable transformation $T:\Omega\to\Omega$, we prove that for any bounded $\mathcal{F}$-measurable function f, there exists a bounded $\mathcal{I}$-measurable function $g_f:\Omega\to\mathbb{R}$ such that for any T-invariant probability P on $(\Omega, \mathcal{F})$,

$$\mathbb{E}_P(f \mid \mathcal{I}) = g_f, \ P\text{-almost surely,} \tag{1}$$

where $\mathcal{I} = \{A\in\mathcal{F} : T^{-1}A = A\}$ and $\mathbb{E}_P(f \mid \mathcal{I})$ is the conditional expectation of f with respect to the σ-algebra $\mathcal{I}$ and the probability P. The conditional expectations for G-Brownian motion were introduced by Peng [36], and Bartl's definition of conditional sublinear expectation can be found in [3]. In addition, the usual conditional expectation with respect to a σ–algebra $\mathcal{I}$ offers an alternative way to define conditional expectations for certain upper expectations relative to a special σ–algebra. More specifically, given a subset Λ of invariant probabilities, let $\hat{\mathbb{E}} = \sup_{P\in\Lambda}\mathbb{E}_P$ be the upper expectation with respect to Λ. Then we can define the nonlinear conditional expectation for $\hat{\mathbb{E}}$ with respect to $\mathcal{I}$ by

$$\hat{\mathbb{E}}(f \mid \mathcal{I}) = g_f \text{ for any bounded } \mathcal{F}\text{-measurable function } f.$$

It is easy to verify that $\hat{\mathbb{E}}[1_A \cdot \hat{\mathbb{E}}[f|\mathcal{I}]] = \hat{\mathbb{E}}[1_A \cdot f]$ for any $A\in\mathcal{I}$. Moreover, we show that for any upper probability V, if V is T-invariant then

$$V(\Omega \setminus \{\omega\in\Omega : \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega) = g_f(\omega)\}) = 0,$$

where g_f is the common conditional expectation obtained in (1) (see Theorem 3.1).

Feng, Wu, and Zhao [21] introduced a definition of ergodicity of capacities to describe the inability to decompose the system into disjoint subsystems inspired by the ergodicity of sublinear expectations [22] as follows: Given a capacity preserving system $(\Omega, \mathcal{F}, \mu, T)$, μ is said to be ergodic (with respect to T) if for any $B\in\mathcal{I}$ the following two conditions hold:

(i) $\mu(B) = 0$ or $\mu(B) = 1$
(ii) $\mu(B) = 0$ or $\mu(\Omega\setminus B) = 0$.

They provided a number of equivalent characterizations of the ergodicity, especially, in terms of the spectral properties of transformation operator whose eigenvalue 1 being simple was proved. This leads to the result that an invariant upper probability V is ergodic with respect to T if and only if for any bounded $\mathcal{F}$-measurable function f, there exists a constant $c_f \in \mathbb{R}$ such that $V(\Omega \setminus \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega) = c_f\}) = 0$. When V is a probability, we know that $c_f = \int f dV$ by Birkhoff's ergodic theorem for probabilities. However, the uncertainty of the upper probability V results in the loss of information about this constant c_f. In this paper, we demonstrate what this constant is. In fact, one of the main results can be stated as follows:

The first main result (Theorem 3.2 and Theorem 3.3). Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is an upper probability. Then V is ergodic with respect to T if and only if there exists a unique ergodic probability Q on $\mathcal{F}$ such that $Q(A) \le V(A)$ for any $A \in \mathcal{F}$ and $Q(B) = V(B)$ for any $B \in \mathcal{I}$. Moreover, these are equivalent to that there exists an ergodic probability Q on $\mathcal{F}$ with respect to T such that for any $f \in L^1(\Omega, \mathcal{F}, Q)$,

$$V(\Omega \setminus \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i\omega) = \int f dQ\}) = 0.$$

This suggests that the above requirement for ergodicity is not redundant. Meanwhile, the definition of ergodicity that $\mu(\mathcal{I}) \in \{0,1\}$ suggested by [9] cannot imply that the uniqueness of such Q and thus the irreducibility of the dynamical system cannot hold. The invariant probability Q is called an invariant skeleton and satisfies $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$ for any $P \in \mathrm{core}(V)$ (see (7) for the definition).

Recall that the strong law of large numbers for processes on probability spaces can be obtained from Birkhoff's ergodic theorem. We extend this result to upper probability spaces, that is, for any ergodic stationary process $\{Y_n\}_{n\in\mathbb{N}}$ on an upper probability space $(\Omega, \mathcal{F}, V)$, there exists a probability Q on $(\Omega, \mathcal{F})$ such that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \int Y_1 dQ$ almost surely (**see Theorem 3.12**).

As another application of the first main result, motivated by the ergodic theory of probability preserving systems, we provide a characterization for ergodicity of upper probabilities in terms of independence as follows (**see Theorem 4.3**): Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is an upper probability. Then V is ergodic if and only if there exists an ergodic probability Q on $\mathcal{F}$ such that $V(A) = Q(A)$ for any $A \in \mathcal{I}$, and $\lim_{n\to\infty} \int \frac{1}{n}\sum_{i=0}^{n-1} f \cdot (g \circ T^i) dV = \int f dV \int g dQ$ for any bounded $\mathcal{F}$-measurable functions f, g with $g \ge 0$, where the integral with respect to the capacity is the Choquet integral (see (9) for definition).

In addition to ergodicity, the concept of mixing plays a fundamental role in understanding the behaviour of probability preserving systems. Weak mixing, a type of mixing that exhibits a certain level of randomness and unpredictability, is an important tool for understanding the properties of dynamical systems and has connections to the theory of unique ergodicity and rigidity (refer to the book by Glasner [26]). Meanwhile, the study of weak mixing has important implications for a range of research fields, including number theory and combinatorics (we refer to Einsiedler and Ward [19] and Furstenberg [24]). Therefore, studying weak mixing for capacity preserving systems is also critical for capacities. In this paper, we introduce the definition of weak mixing for capacity preserving systems and study its properties. It is well known that an invariant probability is weakly mixing if and only if the product probability of itself is ergodic. Naturally, we want to have a similar characterization for capacities. However, Carathéodory's extension theorem from an algebra to a σ-algebra is not true for capacities (see [15, Chapter 12] for example). To address this issue, we provide a means to define the product of two upper probabilities and show that the unique invariant skeleton of a weakly mixing upper probability is a weakly mixing probability (**see Lemma 5.7**). Moreover, we prove that weak mixing for an invariant upper probability is equivalent to the ergodicity of the product upper probability of itself (**see Theorem 5.10**). More specifically, let $(\Omega, \mathcal{F}, V, T)$ be an invertible capacity preserving system, where V is an upper probability. Then the following statements are equivalent:

(i) V is weakly mixing with respect to T;

(ii) for any invertible capacity preserving system $(\Omega', \mathcal{F}', V', T')$ with V' being an ergodic upper probability, $V \times V'$ is ergodic with respect to $T \times T'$;

(iii) $V \times V$ is ergodic with respect to $T \times T$.

As a corollary, we extend Birkhoff's ergodic theorem along polynomial subsequences to weakly mixing upper probability spaces (**see Corollary 5.8**). That is, let $(\Omega, \mathcal{F}, V, T)$ be a weakly mixing upper probability space, and let $p(x)$ be a polynomial with integer coefficients. Then there exists a weakly mixing probability Q on $\mathcal{F}$ such that for any $f \in L^r(\Omega, \mathcal{F}, Q)$, $r > 1$,

$$V(\Omega \setminus \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^{p(i)}\omega) = \int f dQ\}) = 0.$$

We also provide some examples of ergodic and weakly mixing capacity preserving systems by concave distortion of ergodic and weakly mixing probability preserving systems. Meanwhile, we show that a subadditive weakly mixing capacity must be ergodic, and we provide examples to show that the reverse is not true. As applications, for an ergodic probability preserving system $(\Omega, \mathcal{F}, P, T)$, we establish lower and upper bounds to the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P^{1/2}(B \cap T^{-i}C)$ for any $B, C \in \mathcal{F}$, and we utilize this limit to obtain new characterizations of the systems of

(i) weak mixing (**see Proposition 6.2**): P is weakly mixing if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P^{1/2}(B \cap T^{-i}C) = P^{1/2}(B)P^{1/2}(C) \text{ for any } B, C \in \mathcal{F};$$

(ii) periodic (**see Proposition 6.3**): there exists $B \in \mathcal{F}$ with $P(B) > 0$ such that for any $C \subset B$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P^{1/2}(B \cap T^{-i}C) = P^{1/2}(B)P(C)$$

if and only if there exist $r \in \mathbb{N}$ and distinct points $\omega_1, \ldots, \omega_r \in \Omega$ such that $P(\{\omega_i\}) = \frac{1}{r}$, $i = 1, 2, \ldots, r$.

Further applications of the nonlinear theory are for non-invariant probabilities (linear), namely, for a probability space $(\Omega, \mathcal{F}, P)$ and an invertible measurable map $T : \Omega \to \Omega$. The programme of studying an ergodic theory for non-invariant probabilities was initiated in Hurewicz [28]. The main result obtained was Birkhoff's law of large numbers on the convergence of pathwise average of a function along P almost every trajectory. In this paper, with the help of the nonlinear ergodic theory of upper probabilities, we push the study of this problem further, of which the main results are briefly described as follows. Suppose that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P \circ T^{-i}$ exists, that is, the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P(T^{-i}A)$ exists for any $A \in \mathcal{F}$. Denote the limit by $Q(A)$ for each $A \in \mathcal{F}$. By the Vitali–Hahn–Saks theorem (see Lemma 2.2), Q is a probability measure and it is easy to verify that Q is invariant. Furthermore, if Q is

(i) ergodic (**see Theorem 4.7**) then for any $f \in L^1(\Omega, \mathcal{F}, Q)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega) = \int f dQ \text{ for } P\text{-a.s. } \omega \in \Omega, \tag{2}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P(B \cap T^{-i}C) = P(B)Q(C) \text{ for any } B, C \in \mathcal{F}; \tag{3}$$

(ii) weakly mixing (**see Theorem 5.19**) then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |P(B \cap T^{-i}C) - P(B)Q(C)|^2 = 0 \text{ for any } B, C \in \mathcal{F}, \tag{4}$$

and for any $f \in B(\Omega, \mathcal{F})$, there exists a subset $J = J_f$ of $\mathbb{N}$ with $D(J) = 0$ (see Definition 5.13) such that

$$\lim_{n\notin J, n\to\infty} \int f \circ T^n dP = \int f dQ. \tag{5}$$

Moreover, we can prove that (2) and (3) are equivalent and both are equivalent to Q being ergodic; and (4) and (5) are equivalent and both are equivalent to Q being weakly mixing.

Finally, as a further extension of Birkhoff's ergodic theorem for capacities, we extend Kingman's subadditive ergodic theorem to upper probability spaces by taking advantage of the common conditional expectation.

The second main result (Theorem 7.1). Let $(\Omega, \mathcal{F})$ be a standard measurable space, $T : \Omega \to \Omega$ be a measurable transformation, and V be an invariant upper probability. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-measurable functions satisfying the following conditions:

(i) there exists $\lambda > 0$ such that $-\lambda n \le f_n(\omega) \le \lambda n$ for any $n \in \mathbb{N}$, and $\omega \in \Omega$;

(ii) for each $m, n \in \mathbb{N}$, $V(\Omega \setminus \{\omega \in \Omega : f_{n+m}(\omega) \le f_n(\omega) + f_m(T^n\omega)\}) = 0$.

Then there exists a bounded T-invariant $\mathcal{F}$-measurable function f^* such that

$$V(\Omega \setminus \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} f_n(\omega) = f^*(\omega)\}) = 0.$$

Note that f^* can be represented by the common conditional expectations of $\{f_n\}_{n\in\mathbb{N}}$.

As an application of this theorem, we extend the Furstenberg-Kesten theorem [25] from probability spaces to upper probability spaces. Moreover, we use this extension to prove the multiplicative ergodic theorem on upper probability spaces (**see Theorem 7.5**). Meanwhile, we also provide the subadditive ergodic theorem **(see Theorem 7.3)** and the multiplicative ergodic theorem **(see Theorem 7.6)** for a class of non-invariant probabilities.

The structure of the paper is as follows. In Section 2, we recall some basic notions and prove some basic properties that we use in this paper. In Section 3, we study Birkhoff's ergodic theorem for upper probabilities, and prove a strong law of large numbers for processes on capacity spaces. In Section 4, we provide characterizations for ergodicity of upper probabilities in terms of independence. In Section 5, we introduce the notion of weak mixing for capacity preserving systems, and provide a number of their characterizations. In Section 6, we will study some examples including some applications to probability preserving systems. In Section 7, we investigate subadditive ergodic theorem for upper probabilities and prove the multiplicative ergodic theorem on upper probability spaces.

2. Preliminaries. In this paper, we denote by $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{C}$ the set of all natural numbers, natural numbers with 0, integers, real numbers, non-negative real numbers and complex numbers, respectively.

Let $(\Omega,\mathcal{F})$ be a measurable space. Denote by $B(\Omega,\mathcal{F})$ the set of all bounded and $\mathcal{F}$-measurable functions from Ω to $\mathbb{R}$. For a subset A of Ω, write $\Omega\setminus A$ as A^c.

If Ω is a topological space, then we denote by $\mathcal{B}(\Omega)$ the Borel σ-algebra on Ω. A measurable space $(\Omega,\mathcal{F})$ is said to be standard if there exists a complete and separable metric space X such that $(\Omega,\mathcal{F})$ is isomorphic to $(X,\mathcal{B}(X))$, that is, there exists a bijection $f:\Omega\to X$ such that for any $E\subset\Omega$, we have $E\in\mathcal{F}$ if and only if $f(E)\in\mathcal{B}(X)$. Note that in this paper, unless stated otherwise, we do not require measurable space $(\Omega,\mathcal{F})$ is standard.

2.1. *Set functions and Choquet integrals.* Let $(\Omega,\mathcal{F})$ be a measurable space. Recall that a set function $\mu:\mathcal{F}\to[0,1]$ is

- a capacity if $\mu(\emptyset)=0,\mu(\Omega)=1$, and $\mu(A)\le\mu(B)$ for all $A,B\in\mathcal{F}$ such that $A\subset B$;
- concave if $\mu(A\cup B)+\mu(A\cap B)\le\mu(A)+\mu(B)$ for all $A,B\in\mathcal{F}$;
- subadditive if $\mu(A\cup B)\le\mu(A)+\mu(B)$ for all $A,B\in\mathcal{F}$ with $A\cap B=\emptyset$;
- additive if $\mu(A\cup B)=\mu(A)+\mu(B)$ for all $A,B\in\mathcal{F}$ with $A\cap B=\emptyset$;
- σ-additive if $\mu(\cup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}\mu(A_n)$ for all $\{A_n\}_{n\in\mathbb{N}}\subset\mathcal{F}$ with $A_i\cap A_j=\emptyset$ for any $i\neq j$;
- continuous from below if $\lim_{n\to\infty}\mu(A_n)=1$ for $A_n\uparrow\Omega$;
- continuous from above if $\lim_{n\to\infty}\mu(A_n)=0$ for $A_n\downarrow\emptyset$;
- continuous if it is both continuous from below and above;
- a probability if it is a σ-additive capacity.

Denote by $\Delta(\Omega,\mathcal{F})$ the set of all additive capacities on $(\Omega,\mathcal{F})$, and by $\Delta^\sigma(\Omega,\mathcal{F})$ the set of all probabilities on $(\Omega,\mathcal{F})$. We endow both sets with the weak* topology[1]. Given a sub-σ-algebra $\mathcal{F}'$ of $\mathcal{F}$, denote by $\mu|_{\mathcal{F}'}$ the capacity μ restricted to $\mathcal{F}'$. Given $P\in\Delta^\sigma(\Omega,\mathcal{F})$, we denote

$$L^r(\Omega,\mathcal{F},P)=\left\{f:\Omega\to\mathbb{R}: f \text{ is } \mathcal{F}\text{-measurable and } \|f\|_{r,P}:=\left(\int|f|^r dP\right)^{\frac{1}{r}}<\infty\right\},\ r\ge 1$$

and

$$L^\infty(\Omega,\mathcal{F},P)=\{f:\Omega\to\mathbb{R}: f \text{ is } \mathcal{F}\text{-measurable and } \|f\|_{\infty,P}<\infty\},$$

where $\|f\|_{\infty,P}:=\inf\{C\ge 0:|f(\omega)|\le C \text{ for } P\text{-a.s. } \omega\in\Omega\}$.

A capacity is called an upper probability if there exists a compact subset Λ of $\Delta^\sigma(\Omega,\mathcal{F})$ in the weak* topology such that

$$V(A)=\max_{P\in\Lambda}P(A),\ \text{for any } A\in\mathcal{F}. \tag{6}$$

In this case, for any $A\in\mathcal{F}$, there exists $P\in\Lambda$ such that $V(A)=P(A)$. Given a sub-σ-algebra $\mathcal{F}'$ of $\mathcal{F}$, it is easy to check that $V|_{\mathcal{F}'}$ is also an upper probability.

The core of a capacity μ is defined by

$$\mathrm{core}(\mu)=\{P\in\Delta(\Omega,\mathcal{F}):P(A)\le\mu(A) \text{ for any } A\in\mathcal{F}\}. \tag{7}$$

We remark that for a general capacity μ, $\mathrm{core}(\mu)$ may be empty. If it is not empty, then $\mathrm{core}(\mu)$ is compact in the weak* topology (see [33, Proposition 4.2] for example).

[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 $A\in\mathcal{F}$ (see [9, Page 3382] for more details).

Let $V=\sup_{P\in\Lambda}P$ be an upper probability, where $\Lambda\subset\Delta^\sigma(\Omega,\mathcal{F})$ is compact. Then by standard results (see [9, pp. 3382–3383], [21, Lemma 2.2(ii)], or as a direct corollary of Lemma 5.2), V is continuous from above, and hence $\Lambda\subset\text{core}(V)\subset\Delta^\sigma(\Omega,\mathcal{F})$. Since $\text{core}(V)$ is itself compact, we may equivalently write

$$V=\sup_{P\in\text{core}(V)}P. \tag{8}$$

Conversely, if V is a capacity that is both concave and continuous from above, then V admits the representation $V=\sup_{P\in\text{core}(V)}P$ with $\text{core}(V)\subset\Delta^\sigma(\Omega,\mathcal{F})$ compact, and hence V is an upper probability.

DEFINITION 2.1. In a capacity space $(\Omega,\mathcal{F},\mu)$, we say that a statement holds for μ-almost surely (μ-a.s. for short) if it holds on a set $A\in\mathcal{F}$ with $\mu(A^c)=0$.

The following result can be found in [18, III. 7.2, Theorem 2 and Corollary 8].

LEMMA 2.2 (Vitali-Hahn-Saks). *Let $(\Omega,\mathcal{F})$ be a measurable space, and $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probabilities. Suppose that for any $A\in\mathcal{F}$ the limit $\lim_{n\to\infty}P_n(A)=Q(A)$ exists. Then Q is a probability. If we further suppose that each P_n is absolutely continuous with respect to the probability Q, then the absolute continuity of the P_n with respect to Q is uniform in $n\in\mathbb{N}$, that is, for any $\epsilon>0$ there exists $\delta>0$ such that for any $A\in\mathcal{F}$, if $Q(A)<\delta$ then $P_n(A)<\epsilon$ for all $n\in\mathbb{N}$.*

Next, we recall Choquet integral introduced by Choquet [12]. A capacity μ induces Choquet integral, defined by

$$\int_\Omega f d\mu=\int_0^\infty \mu(\{\omega\in\Omega: f(\omega)\geq t\})dt+\int_{-\infty}^0(\mu(\{\omega\in\Omega: f(\omega)\geq t\})-1)\,dt \tag{9}$$

for all $\mathcal{F}$-measurable functions f, where the integrals on the right-hand side are the Lebesgue integrals. If μ is additive, then the Choquet integral reduces to the standard additive integral.

The following result can be checked by definitions.

PROPOSITION 2.3. Let $(\Omega,\mathcal{F},\mu)$ be a capacity space, f and g be two $\mathcal{F}$-measurable functions. Then

(i) (Positive homogeneity): $\int_\Omega \alpha f d\mu=\alpha\int_\Omega f d\mu$ for each $\alpha\geq 0$.
(ii) (Translation invariance): $\int_\Omega (f+\alpha 1_\Omega)d\mu=\int_\Omega f d\mu+\alpha$ for each $\alpha\in\mathbb{R}$.
(iii) (Monotonicity): $\int_\Omega f d\mu\geq\int_\Omega g d\mu$ if $f\geq g$.

If there is no ambiguity, we will omit Ω, and write $\int_\Omega$ as $\int$ for simplicity.

The following result provides a dominated convergence theorem in a capacity space with respect to Choquet integral, which was proved in [21, Lemma 2.2].

LEMMA 2.4. *Let μ be a continuous subadditive capacity on $(\Omega,\mathcal{F})$. For any $\mathcal{F}$-measurable functions $\{f_n\}_{n\in\mathbb{N}}$, g and h with $g\leq f_n\leq h$ for each $n\in\mathbb{N}$, and $\int g d\mu$, $\int h d\mu$ being finite, if $f_n\to f$ μ-a.s. then*

$$\lim_{n\to\infty}\int f_n d\mu=\int f d\mu.$$

2.2. *Invariant probabilities and capacities.* In this section, we fix a measurable space $(\Omega,\mathcal{F})$ and a measurable transformation T from Ω to itself. Denote by $\mathcal{M}(T)$ the set of all (T-)invariant probabilities on $(\Omega,\mathcal{F})$. An invariant probability P is said to be ergodic if and only if $P(\mathcal{I})=\{0,1\}$. We denote by $\mathcal{M}^e(T)$ the set of all ergodic probabilities on $(\Omega,\mathcal{F})$. Furthermore, if $P\times P$ is ergodic with respect to $T\times T$, then P is said to be weakly mixing with respect to T. Denote by $\mathcal{M}^{wm}(T)$ the set of all weak mixing probabilities on $(\Omega,\mathcal{F})$.

For convenience of use, we list the following well-known results.

THEOREM 2.5 (Birkhoff's ergodic theorem [5]). *Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system. For any $f\in L^1(\Omega,\mathcal{F},P)$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_P(f\mid\mathcal{I})(\omega)$$

for P-a.s. $\omega\in\Omega$, and $\mathbb{E}_P(f\mid\mathcal{I})$ is the conditional expectation of f with respect to the σ-algebra $\mathcal{I}$. If in addition, P is ergodic then $\mathbb{E}_P(f\mid\mathcal{I})=\mathbb{E}_P(f)$, P-a.s.

THEOREM 2.6 (Subadditive ergodic theorem [29, 30]). *Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-measurable functions satisfying the following conditions:*

(i) $\int|f_1|dP<\infty$;

(ii) *for each $k,n\in\mathbb{N}$, $f_{n+k}\le f_n+f_k\circ T^n$, P-a.s.*

Then there exists a T-invariant function $\phi:\Omega\to\mathbb{R}$ such that

$$\lim_{n\to\infty}\frac{1}{n}f_n=\phi,\ P\text{-a.s.}$$

Moreover, the function ϕ is given by

$$\phi(\omega)=\inf_{n\in\mathbb{N}}\frac{1}{n}\mathbb{E}_P(f_n\mid\mathcal{I})(\omega)=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_P(f_n\mid\mathcal{I})(\omega)\text{ for each }\omega\in\Omega.$$

The following result is obtained by Bourgain [6, Theorem 1].

THEOREM 2.7. *Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system, and let $p(x)$ be a polynomial with integer coefficients. If $f\in L^r(\Omega,\mathcal{F},P)$, $r>1$, then the limit*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f(T^{p(i)}\omega)\text{ exists for }P\text{-a.s. }\omega\in\Omega.$$

Furthermore, if T is weakly mixing, then the limit is equal to $\int f dP$, for P-a.s. $\omega\in\Omega$.

The following result should be standard, but we were unable to locate a clear reference to it. Thus, we provide a proof.

LEMMA 2.8. *Let $(\Omega,\mathcal{F})$ be a measurable space, and let $T:\Omega\to\Omega$ be a measurable transformation. Given $P,Q\in\mathcal{M}(T)$ if $P(A)=Q(A)$ for any $A\in\mathcal{I}$, then $P=Q$.*

PROOF. For any bounded measurable function f, by Theorem 2.5, one has that

$$P(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_P(f\mid\mathcal{I})(\omega)\})=1$$

and

$$Q(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_Q(f\mid\mathcal{I})(\omega)\})=1.$$

Since $\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_P(f\mid\mathcal{I})(\omega)\}\in\mathcal{I}$ and $P|_{\mathcal{I}}=Q|_{\mathcal{I}}$, it follows that

$$Q(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_P(f\mid\mathcal{I})(\omega)\})=1.$$

Thus, $\mathbb{E}_P(f\mid\mathcal{I})=\mathbb{E}_Q(f\mid\mathcal{I})$, Q-a.s. Since $P|_{\mathcal{I}}=Q|_{\mathcal{I}}$ and $\mathbb{E}_P(f\mid\mathcal{I})$ are $\mathcal{I}$-measurable, it follows that

$$\int fdP=\int\mathbb{E}_P(f\mid\mathcal{I})dP=\int\mathbb{E}_P(f\mid\mathcal{I})dQ=\int\mathbb{E}_Q(f\mid\mathcal{I})dQ=\int fdQ.$$

The proof is complete as f is arbitrary. □

REMARK 2.9. Given two probabilities P and Q, they are said to be singular if there exist two disjoint subsets $B,C\in\mathcal{F}$ with $\Omega=B\cup C$ such that $P(C)=0$ and $Q(B)=0$. By Lemma 2.8, it is easy to check that for any $P,Q\in\mathcal{M}^e(T)$ with $P\neq Q$, P is singular with Q.

DEFINITION 2.10. Let $(\Omega,\mathcal{F},\mu)$ be a capacity space and T be a measurable transformation from Ω to itself. Then $(\Omega,\mathcal{F},\mu,T)$ is called a capacity preserving system if μ is T-invariant, that is, for each $A\in\mathcal{F}$, $\mu(A)=\mu(T^{-1}A)$.

Recall that a T-invariant capacity μ is ergodic if for any $A\in\mathcal{I}$, $\mu(A)\in\{0,1\}$ and $\mu(A)=0$ or $\mu(A^c)=0$. If in addition, we suppose that μ is subadditive, then it is ergodic if and only if for any $A\in\mathcal{I}$, $\mu(A)=0$ or $\mu(A^c)=0$.

The following result is a characterization of ergodicity for subadditive capacities via measurable functions.

LEMMA 2.11 (Theorem 4.4 in [21]). *Let $(\Omega,\mathcal{F},\mu,T)$ be a capacity preserving system. If μ is subadditive and continuous then the following three statements are equivalent:*

(i) *μ is ergodic;*
(ii) *if $f\in B(\Omega,\mathcal{F})$ is T-invariant then f is constant μ-a.s.;*
(iii) *if $f:\Omega\to\mathbb{R}$ is $\mathcal{F}$-measurable and T-invariant μ-a.s. then f is constant μ-a.s.*

Note that (iii) ⇒ (ii) ⇒ (i) does not need the continuity of μ.

We recall a version of Birkhoff's ergodic theorem for ergodic upper probabilities [21, Theorem 4.5].

THEOREM 2.12. *Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system, where V is an upper probability. Then V is ergodic with respect to T if and only if for any $f\in B(\Omega,\mathcal{F})$, there exists a unique $c_f\in\mathbb{R}$ such that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=c_f \text{ for } V\text{-a.s. } \omega\in\Omega.$$

LEMMA 2.13. *Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system, where V is an upper probability. Then*

(i) *for any $P\in\text{core}(V)$, there exists a unique $\hat{P}\in\mathcal{M}(T)\cap\text{core}(V)$ such that $\hat{P}|_{\mathcal{I}}=P|_{\mathcal{I}}$;*

(ii) *given $A\in\mathcal{F}$, if for any $P\in\text{core}(V)$, $\hat{P}(A)=0$, then $V(A)=0$.*

PROOF. The existence of $\hat{P}$ in the first statement was proved in Corollary 1 of [9], and the uniqueness is obtained in Lemma 2.8.

Fix any $A\in\mathcal{F}$ satisfying for any $P\in\text{core}(V)$, $\hat{P}(A)=0$. Let $\bar{A}=\cap_{n=1}^{\infty}\cup_{k=n}^{\infty}T^{-k}A$. Then by the invariance of $\hat{P}$, one deduces that

$$\hat{P}(\bar{A})\le\sum_{i=1}^{\infty}\hat{P}(T^{-i}A)=0$$

for any $P\in\text{core}(V)$. Since $\cup_{k=n}^{\infty}T^{-k}A$ decreases to $\bar{A}$ as $n\to\infty$, it follows that $T^{-1}\bar{A}=\bar{A}$, that is, $\bar{A}\in\mathcal{I}$. This, together with (i), implies that $P(\bar{A})=0$ for any $P\in\text{core}(V)$. Thus, it follows from (8) that $V(\bar{A})=0$ holds. Then, by the invariance of V,

$$V(A)=\lim_{n\to\infty}V(T^{-n}A)=\limsup_{n\to\infty}\int 1_{T^{-n}A}dV\le\int\limsup_{n\to\infty}1_{T^{-n}A}dV=V(\bar{A})=0,$$

where the inequality is obtained from [47, Corollary 9.5]. The proof is complete. □

REMARK 2.14. In the following, the invariant probability $\hat{P}$ given in the above lemma is called the invariant skeleton of $P\in\text{core}(V)$.

2.3. *Common conditional expectations.* In this subsection, we introduce the notion of common conditional expectation as follows.

DEFINITION 2.15. Let $(\Omega,\mathcal{F})$ be a measurable space, and $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$. Given a subset $\Lambda\subset\Delta^{\sigma}(\Omega,\mathcal{F})$ and $f\in B(\Omega,\mathcal{F})$, an $\mathcal{H}$-measurable function g_f is said to be the common conditional expectation of f with respect to $\mathcal{H}$ for Λ if

$$\mathbb{E}_P(f\mid\mathcal{H})=g_f,\ P\text{-a.s.}$$

for any $P\in\Lambda$, where $\mathbb{E}_P(f\mid\mathcal{H})$ is the conditional expectation of f with respect to $\mathcal{H}$ for P.

The following result shows that there exists a common conditional expectation with respect to $\mathcal{I}$ for all invariant probabilities on a standard measurable space. That is,

LEMMA 2.16. *Let $(\Omega,\mathcal{F})$ be a standard measurable space, and let $T:\Omega\to\Omega$ be measurable. Then for any $f\in B(\Omega,\mathcal{F})$, there exists an $\mathcal{I}$-measurable function $g_f\in B(\Omega,\mathcal{F})$ such that for any $P\in\mathcal{M}(T)$, $g_f\in L^1(\Omega,\mathcal{I},P)$, and*

$$\mathbb{E}_P(f\mid\mathcal{I})=g_f,\ P\text{-a.s.}$$

PROOF. Since $(\Omega,\mathcal{F})$ is a standard measurable space, let $\hat{\mathcal{F}}$ be a countable generating algebra of $\mathcal{F}$. For any $F\in\hat{\mathcal{F}}$, let

$$G(F)=\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_F(T^i\omega)\text{ exists}\}$$

and let

$$G(\hat{\mathcal{F}}) = \cap_{F \in \hat{\mathcal{F}}} G(F).$$

For $\omega \in G(\hat{\mathcal{F}})$ and $F \in \hat{\mathcal{F}}$, define

$$\mathfrak{p}_\omega(F) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 1_F(T^i\omega).$$

It is easy to check that $\mathfrak{p}_\omega$ is nonnegative, normalized (i.e., $\mathfrak{p}_\omega(\Omega) = 1$), and finitely additive on $\hat{\mathcal{F}}$. Since the space is standard, by Carathéodory's extension theorem, for any $\omega \in G(\hat{\mathcal{F}})$, $\mathfrak{p}_\omega$ can be extended to a probability on $\mathcal{F}$, uniquely. We still denote them by $\mathfrak{p}_\omega$.

Note that $\mathfrak{p}_\omega(F)$ is independent of $P \in \mathcal{M}(T)$, since it is a pointwise limit. Fix any $P \in \mathcal{M}(T)$. By Theorem 2.5, one has that $P(G(\hat{\mathcal{F}})) = 1$, and for any $F \in \hat{\mathcal{F}}$,

$$\mathfrak{p}_\omega(F) = \mathbb{E}_P(1_F \mid \mathcal{I})(\omega) \text{ for } P\text{-a.s. } \omega \in G(\hat{\mathcal{F}}).$$

Thus, for any $F \in \hat{\mathcal{F}}$,

$$\mathfrak{p}_\omega(F) = \mathbb{E}_P(1_F \mid \mathcal{I})(\omega) \text{ for } P\text{-a.s. } \omega \in \Omega.$$

Note that for P-a.s. $\omega \in \Omega$, $\mathbb{E}_P(1_F \mid \mathcal{I})(\omega)$ is also nonnegative, normalized, and finitely additive on $\hat{\mathcal{F}}$. By the uniqueness of the extension, one has for P-a.s. $\omega \in \Omega$,

$$\mathfrak{p}_\omega(F) = \mathbb{E}_P(1_F \mid \mathcal{I})(\omega) \text{ for any } F \in \mathcal{F}.$$

Furthermore, by the standard argument for approximating a measurable function by simple functions, one has that for P-a.s. $\omega \in \Omega$,

$$\int_\Omega f d\mathfrak{p}_\omega = \mathbb{E}_P(f \mid \mathcal{I})(\omega) \text{ for any } f \in B(\Omega, \mathcal{F}). \tag{10}$$

Let $g_f(\omega) = \int_\Omega f d\mathfrak{p}_\omega$ if $\omega \in G(\hat{\mathcal{F}})$, and otherwise $g_f(\omega) = 0$. By (10) and the arbitrariness of $P \in \mathcal{M}(T)$, g_f is desired. □

REMARK 2.17. The above result can be deduced from the results in [35, 44]. We give our proof for the sake of readability of this paper.

3. Birkhoff's ergodic theorem for capacities. In the study of probability preserving systems, Birkhoff's ergodic theorem plays a crucial role in understanding complex systems. It characterizes the ergodicity of a system and connects with the time average of a test function along the dynamical system with the invariant probability. Upper probabilities are a natural generalization of probabilities and have important applications in areas such as decision theory and risk analysis. In this section, we continue the work of Feng, Wu and Zhao [21] in extending Birkhoff's ergodic theorem to upper probabilities. By studying it, we gain a deeper understanding of the dynamics of these more general systems. As an application, we provide a strong law of large numbers for stationary and ergodic sequences on upper probability spaces.

3.1. *Birkhoff's ergodic theorem for invariant upper probabilities.* In this section, we use common conditional expectations to represent the limit of ergodic mean for invariant upper probabilities.

THEOREM 3.1. *Let $(\Omega,\mathcal{F})$ be a standard measurable space, $T:\Omega\to\Omega$ be a measurable transformation, and V be an invariant upper probability. Then for any $f\in B(\Omega,\mathcal{F})$,*

$$V(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}^c)=0,$$

where g_f is the common conditional expectation of f given as in Lemma 2.16.

PROOF. For any $P\in\text{core}(V)$, let $\hat{P}$ be the invariant skeleton of P. Since $\hat{P}\in\mathcal{M}(T)$, it follows from classical Birkhoff's ergodic theorem that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\mathbb{E}_{\hat{P}}(f\mid\mathcal{I})(\omega)\text{ for }\hat{P}\text{-a.s. }\omega\in\Omega.$$

By Lemma 2.16, we have that $\mathbb{E}_{\hat{P}}(f\mid\mathcal{I})(\omega)=g_f(\omega)$ for $\hat{P}$-a.s. $\omega\in\Omega$. Since the set $\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}\in\mathcal{I}$, for any $P\in\text{core}(V)$,

$$\begin{aligned}&P(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}^c)\\&\qquad=\hat{P}(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}^c)=0.\end{aligned}$$

Thus, from (8),

$$\begin{aligned}&V(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}^c)\\&\qquad=\max_{P\in\text{core}(V)}P(\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=g_f(\omega)\}^c)=0,\end{aligned}$$

proving this result. □

3.2. *Birkhoff's ergodic theorem for ergodic upper probabilities.* In this section, we finish the proof of our first main result, whose proof is divided into Theorems 3.2 and 3.3.

For the sake of presentation in the following proofs of this section, we denote

$$\Omega_{f,\alpha}=\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\omega)=\alpha\}\text{ for }f\in B(\Omega,\mathcal{F}),\alpha\in\mathbb{R}.$$

Now we prove the first statement in our first main result as follows.

THEOREM 3.2. *Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system. If V is an upper probability, then V is ergodic if and only if there exists a (unique) $Q\in\mathcal{M}^e(T)\cap\text{core}(V)$ such that for any $P\in\text{core}(V)$,*

$$P(A)=Q(A)\text{ for any }A\in\mathcal{I}.$$

PROOF. ($\Rightarrow$) Let $\hat{P}\in\text{core}(V)\cap\mathcal{M}(T)$ be the invariant skeleton of $P\in\text{core}(V)$ obtained from Lemma 2.13.

Step 1. Prove that $\hat{P}\in\mathcal{M}^e(T)$ for any $P\in\text{core}(V)$.

Fix any $A \in \mathcal{I}$. It suffices to prove $\hat{P}(A) = 0$ or 1. As $\hat{P} \in \text{core}(V)$, one has $\hat{P}(A) \leq V(A)$. Since V is ergodic and subadditive, it follows that either

$$V(A) = 0 \text{ or } V(A^c) = 0.$$

If $V(A) = 0$, then $\hat{P}(A) = 0$; if $V(A^c) = 0$, then $\hat{P}(A^c) \leq V(A^c) = 0$ and hence $\hat{P}(A) = 1$. This shows that $\hat{P}(A) = 0$ or 1, proving Step 1.

Step 2. Prove the existence of Q.

Given any $A \in \mathcal{F}$, by Theorem 2.12, there exists a unique $c_A \in \mathbb{R}$ such that $V((\Omega_{1_A, c_A})^c) = 0$. Thus,

$$\hat{P}(\Omega_{1_A, c_A}) = 1 \text{ for any } P \in \text{core}(V). \tag{11}$$

On the other hand, by Step 1, $\hat{P} \in \mathcal{M}^e(T)$, which together with the classical Birkhoff's ergodic theorem implies that

$$\hat{P}(\Omega_{1_A, \hat{P}(A)}) = 1 \text{ for any } P \in \text{core}(V). \tag{12}$$

Comparing (11) and (12), one has

$$\hat{P}(A) = c_A \text{ for any } P \in \text{core}(V).$$

Define $Q(A) = c_A$ for any $A \in \mathcal{F}$. Thus, for any $P \in \text{core}(V)$,

$$Q(A) = c_A = \hat{P}(A) \text{ for any } A \in \mathcal{F},$$

that is, $Q = \hat{P}$ for any $P \in \text{core}(V)$. This together with Step 1, implies that $Q \in \mathcal{M}^e(T) \cap \text{core}(V)$. The existence of Q has been proved.

Step 3. Prove that Q is the unique invariant probability in $\text{core}(V)$.

We assume that $Q' \in \mathcal{M}(T) \cap \text{core}(V)$. By Step 2, it follows that for any $P \in \text{core}(V)$, $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$. In particular, $Q(A) = Q'(A)$ for any $A \in \mathcal{I}$, which together with Lemma 2.8 implies that $Q = Q'$.

($\Leftarrow$) For any $A \in \mathcal{I}$, we have $Q(A) = 0$ or 1, as Q is ergodic. This implies that either $P(A) = 0$ for any $P \in \text{core}(V)$, or $P(A) = 1$ for any $P \in \text{core}(V)$. In the first case, $V(A) = 0$. In the second case, $P(A^c) = 0$ for any $P \in \text{core}(V)$, and hence $V(A) = 1$ and $V(A^c) = 0$. As $A \in \mathcal{I}$ is arbitrary, we finish the proof. □

Now we prove the second statement in our first main result, which provides more information about the constant c_f in Theorem 2.12 to overcome the uncertainty caused by upper probabilities.

THEOREM 3.3. *Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is an upper probability. Then V is ergodic with respect to T if and only if there exists a unique $Q \in \mathcal{M}^e(T) \cap \text{core}(V)$ such that for any $f \in L^1(\Omega, \mathcal{F}, Q)$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega) = \int f dQ \textit{ for } V\textit{-a.s. } \omega \in \Omega. \tag{13}$$

PROOF. ($\Leftarrow$) Consider $A \in \mathcal{I}$, then $Q(A) = 0$ or $Q(A) = 1$, as $Q \in \mathcal{M}^e(T)$. Note that if $Q(A) = 0$ then

$$\Omega_{1_A, Q(A)} = \{\omega \in \Omega : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i \omega) = 0\} = \{\omega \in \Omega : 1_A(\omega) = 0\} = A^c.$$

By (13), one has $V(A) = V((\Omega_{1_A,Q(A)})^c) = 0$. Meanwhile, if $Q(A) = 1$ then

$$\Omega_{1_A,Q(A)} = \{\omega \in \Omega : 1_A(\omega) = 1\} = A.$$

Similarly, by (13), we have $V(A^c) = V((\Omega_{1_A,Q(A)})^c) = 0$. Thus, we find that $V(A) = 0$ or $V(A^c) = 0$ for any $A \in \mathcal{I}$. Hence, V is ergodic.

($\Rightarrow$) Let $Q \in \mathcal{M}^e(T) \cap \mathrm{core}(V)$ be the ergodic probability obtained by Theorem 3.2. Since Q is ergodic, we have that for any $f \in L^1(\Omega, \mathcal{F}, Q)$, $Q((\Omega_{f,\int f dQ})^c) = 0$. Since the set $\Omega_{f,\int f dQ} \in \mathcal{I}$, it follows that $P((\Omega_{f,\int f dQ})^c) = 0$ for any $P \in \mathrm{core}(V)$. Thus, $V((\Omega_{f,\int f dQ})^c) = 0$. The proof is completed, as the uniqueness can be established by an argument similar to that of Theorem 3.2. □

In classical ergodic theory, probability invariance seems to play an essential role. To break this restriction, in his seminal paper [28], Hurewicz initiated an interesting problem of investigating Birkhoff's ergodic theorem under non-invariant probabilities, and provided a sufficient condition involving non-wandering sets. But progress along this line did not go very far and only a limited number of results have been achieved since then.

As a consequence of Theorem 3.3, in the next result, we provide a condition in terms of an upper probability for Birkhoff's ergodic theorem to hold for a non-invariant probability. Beginning with this preliminary result, with the help of further results on ergodic theory of upper probabilities that we obtain in this paper later, we are able to push the ergodic theory of non-invariant probability a very big step by considering types of ergodic theorems. Moreover, we will be able to construct an invariant upper probability and invariant skeleton from the non-invariant probability and measurable transformation. Consequently, we obtain a number of results, including averaging asymptotic independence, long-time independence, subadditive ergodic theorem and multiplicative ergodic theorem, with no need of referring to the upper probability. See Remark 4.6, Theorem 4.7, Theorem 5.19, Theorem 7.3 and Theorem 7.6 for details.

COROLLARY 3.4. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be a measurable transformation. If there exists an upper probability V on $(\Omega, \mathcal{F})$ such that V is ergodic with respect to T, and $P \in \mathrm{core}(V)$, then there exists a unique ergodic probability $Q \in \mathcal{M}^e(T) \cap \mathrm{core}(V)$ on $(\Omega, \mathcal{F})$ such that for any $f \in L^1(\Omega, \mathcal{F}, Q)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega) = \int f dQ, \text{ for } P\text{-a.s. } \omega \in \Omega.$$

REMARK 3.5. (i) The limit $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega)$ is a pathwise limit for $f \in B(\Omega, \mathcal{F})$, so it does not depend on the choice of the upper probabilities. In fact, if V' is another ergodic upper probability such that $P \in \mathrm{core}(V')$, and Q' is the corresponding ergodic probability in $\mathcal{M}^e(T) \cap \mathrm{core}(V')$, then $Q' = Q$, as by Corollary 3.4, we have $\int f dQ = \int f dQ'$ for any $f \in B(\Omega, \mathcal{F})$.

(ii) We will provide further results in constructing an appropriate upper probability V and the ergodic probability Q for a class of probabilities on $(\Omega, \mathcal{F})$.

Let us see a concrete example of Birkhoff's ergodic theorem for non-invariant probability. More examples can be found in Section 6.

EXAMPLE 2. Let $\Omega = [0,2)$, where $[0,2)$ is regarded as a representative set of the compact space $\mathbb{R}/(2\mathbb{Z})$, and $\mathcal{B}(\Omega)$ be the Borel σ-algebra on Ω. Given an irrational number α, define $T_\alpha : \Omega \to \Omega$ by

$$T_\alpha(x) = \begin{cases} ((x+\alpha) \bmod 1) + 1, & x \in [0,1), \\ x - 1, & x \in [1,2). \end{cases}$$

For each $i = 1, 2$, define

$$\bar{P}_i(A) := P_i\left(A \cap [i-1, i)\right), \text{ for any } A \in \mathcal{B}(\Omega),$$

where P_i is the Lebesgue measure on $[i-1,i)$, and the upper probability is defined by

$$V = \max\{\bar{P}_1, \bar{P}_2\}. \tag{14}$$

First, it is not difficult to check that $\bar{P}_1, \bar{P}_2 \in \mathrm{core}(V)$, and V is T-invariant. Note that the probability $Q := \frac{1}{2}(\bar{P}_1 + \bar{P}_2)$ is ergodic with respect to T_α, by the observation that for any $A \in \mathcal{I}$, $T_\alpha^{-2}(A \cap [i-1,i)) = A \cap [i-1,i)$ for $i = 1,2$, and the ergodicity of the irrational rotation on the torus with respect to the Lebesgue measure. Note that for any $A \in \mathcal{F}$, if $Q(A) = 0$ then $\bar{P}_i(A) = 0$, $i = 1,2$, and hence $V(A) = 0$. In particular, for any $P \in \mathrm{core}(V)$, $P|_\mathcal{I} = Q|_\mathcal{I}$, as $Q(\mathcal{I}) = \{0,1\}$. By Theorem 3.2, it follows that V is ergodic. The ergodicity of V was obtained in Example 4.6 of [20] by a different argument. By Corollary 3.4, we have Birkhoff's ergodic theorem for the probabilities $\bar{P}_1$ and $\bar{P}_2$, but both $\bar{P}_1$ and $\bar{P}_2$ are not invariant with respect to T_α.

The following corollary shows the core structure of any ergodic upper probability. The first statement strengthens the result in Theorem 3.2.

COROLLARY 3.6. If V is an ergodic upper probability on a measurable space $(\Omega, \mathcal{F})$ with respect to a measurable transformation $T : \Omega \to \Omega$, then $\mathrm{core}(V) \cap \mathcal{M}(T) = \mathrm{core}(V) \cap \mathcal{M}^e(T)$ has only one element, denoted by Q. Moreover, for any $A \in \mathcal{F}$,

$$Q(A) = 0 \text{ if and only if } V(A) = 0.$$

In particular, $V(A) = Q(A)$ for any $A \in \mathcal{I}$.

PROOF. From Theorem 3.2, there exists a unique probability $Q \in \mathrm{core}(V) \cap \mathcal{M}^e(T)$ that satisfies $Q|_\mathcal{I} = P|_\mathcal{I}$ for any $P \in \mathrm{core}(V)$. If there exists $Q' \in \mathrm{core}(V) \cap \mathcal{M}(T)$, then Q is the invariant skeleton of Q', which together with Lemma 2.8 implies that $Q = Q'$, proving the first statement.

Now we prove the second statement. Given $A \in \mathcal{F}$, if $V(A) = 0$, then $Q(A) = 0$, as $Q \in \mathrm{core}(V)$. Conversely, suppose that $Q(A) = 0$. By Theorem 3.2, it follows that for any $P \in \mathrm{core}(V)$, $P|_\mathcal{I} = Q|_\mathcal{I}$, which, together with (i) of Lemma 2.13, implies

$$\hat{P} = Q \text{ for any } P \in \mathrm{core}(V).$$

In particular, $\hat{P}(A) = Q(A) = 0$ for any $P \in \mathrm{core}(V)$. Thus, by (ii) of Lemma 2.13, we have $V(A) = 0$. □

Finally, we consider invariant upper probabilities on a class of special measurable spaces.

COROLLARY 3.7. Let $(\Omega, \mathcal{F})$ be a measurable space, and let $T : \Omega \to \Omega$ be a measurable transformation. If $(\Omega, \mathcal{F})$ is uniquely ergodic with respect to T, that is, $\mathcal{M}(T) = \{Q\}$, then each invariant upper probability is ergodic.

PROOF. Fix an invariant upper probability V on $(\Omega,\mathcal{F})$. Let $\hat{P}$ be the invariant skeleton of $P\in\mathrm{core}(V)$. Since $(\Omega,\mathcal{F})$ is uniquely ergodic with respect to T, it follows that $\hat{P}=Q$ for any $P\in\mathrm{core}(V)$. It is well known that if $\mathcal{M}(T)$ has only one element, then $\mathcal{M}(T)=\mathcal{M}^e(T)$, and hence, $Q\in\mathcal{M}^e(T)$. By Theorem 3.2, it is easy to see that V is ergodic. □

In the closing part of this subsection, we extend the Krylov–Bogolyubov existence theorem from probabilities to upper probabilities and show that the class of invariant upper probabilities is strictly larger.

Let (Ω,d) be a compact metric space and $T:\Omega\to\Omega$ a continuous map. By the Krylov–Bogolyubov theorem (see, for instance, [46, Corollary 6.9.1]) there exists at least one T-invariant ergodic probability P on Ω. Because every probability is, in particular, an upper probability, this yields an ergodic upper probability. Hence, the Krylov–Bogolyubov theorem also holds for upper probability. Moreover, Theorem 3.2 shows that the existence of an ergodic upper probability already guarantees the existence of an ergodic probability (in its core). Therefore, the existence of an ergodic upper probability and the existence of an ergodic probability are equivalent. But we have the following nontrivial extension to Krylov–Bogolyubov theorem.

COROLLARY 3.8. If $T:\Omega\to\Omega$ is a continuous transformation of a compact metric space Ω, then the set of ergodic upper probabilities is nonempty.

Moreover, if (Ω,T) is nontrivial[2], then the set of ergodic upper probabilities is uncountable.

PROOF. We have already shown that the set of ergodic probabilities is nonempty.

We now prove that for nontrivial systems, the class of ergodic upper probabilities is uncountable. Fix an ergodic probability measure P that is not a Dirac measure, and consider a concave, strictly increasing, continuous function $f:[0,1]\to[0,1]$ with $f(0)=0$ and $f(1)=1$. Define

$$V_f(A):=f\big(P(A)\big),\qquad A\in\mathcal{F}.$$

Then V_f is an ergodic upper probability that is not a probability measure (see Example 4 for details). Since the collection of such functions f has uncountable cardinality, so does the corresponding collection of upper probabilities. This completes the proof. □

The following example shows that even if a system admits only one ergodic probability measure, it can still possess uncountably many ergodic upper probabilities.

EXAMPLE 3. Let $\Omega=[0,1)$, where $[0,1)$ is regarded as a representative set of the compact space $\mathbb{R}/\mathbb{Z}$. Given an irrational number α, define $R_\alpha:\Omega\to\Omega$ by

$$T_\alpha(x)=(x+\alpha)\bmod 1.$$

It is well known that the system (Ω,R_α) admits a unique ergodic measure m, namely the Lebesgue measure. However, by Corollary 3.8, it has uncountably many ergodic upper probabilities.

[2] A system (Ω,T) is said to be trivial if every invariant set consists of fixed points of T.

3.3. *Ergodicity of stationary processes on capacity spaces.* In classical probability theory, the notion of stationary stochastic process is one possible generalization of independent identically distributed random variables. Recently, it was defined on capacity spaces by Cerreia-Vioglio, Maccheroni and Marinacci [9] as follows.

DEFINITION 3.9. Given a capacity space $(\Omega, \mathcal{F}, \mu)$, a stochastic process $\{Y_n\}_{n\in\mathbb{N}}$ is called to be stationary if for each $n \in \mathbb{N}, k \in \mathbb{Z}_+$ and Borel subset A of $\mathbb{R}^{k+1}$,

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

In classical probability theory, independent identically distributed random variables form a stationary sequence. However, in the context of capacity theory, this result does not hold in general. A counterexample can be found in Example 2.1 of [21].

A stationary ergodic process is a type of stochastic process that has been extensively studied in the fields of probability theory, statistics, and information theory. As it conforms to the ergodic theorem, there exists a strong law of large numbers for stationary ergodic stochastic sequences on probability spaces (see [17, Theorem 2.1 of Chapter X] for example). More recently, Feng, Wu and Zhao extended this result to upper probability spaces [21, Theorem 5.1].

In this section, we use Theorem 3.3 to obtain a stronger result. Let $\left(\mathbb{R}^{\mathbb{N}}, \sigma(\mathcal{C})\right)$ denote the space of sequences endowed with the σ-algebra generated by the set of all cylinders $\mathcal{C}$. Any set $C \in \mathcal{C}$ is called a cylinder, which has the following form

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

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 left shift transformation $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ defined by

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

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

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

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

$$\mu_{\mathbf{Y}}(C) = \mu\left(\mathbf{Y}^{-1}(C)\right) \text{ for any } C \in \sigma(\mathcal{C}). \tag{15}$$

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 $\{C_n\}_{n\in\mathbb{N}} \subseteq \sigma(\mathcal{C})$.

The following result was obtained in Proposition 5.1 of [21] to establish the relation between dynamical systems and stationary stochastic processes on a capacity space.

PROPOSITION 3.10. Let $\mathbf{Y} = \{Y_n\}_{n\in\mathbb{N}}$ be a stochastic process on the capacity space $(\Omega, \mathcal{F}, \mu)$, where μ is continuous. Then $\mathbf{Y}$ is stationary if and only if $\mu_{\mathbf{Y}}$ is T-invariant.

Similar to the ergodicity of stochastic processes on probability spaces, Feng, Wu and Zhao [21] introduced the ergodicity of stochastic processes on capacity spaces as follows.

DEFINITION 3.11. The stationary stochastic process $\mathbf{Y}$ on a capacity space $(\Omega, \mathcal{F}, \mu)$ is called ergodic if the left shift transformation T is ergodic with respect to $\mu_{\mathbf{Y}}$.

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

THEOREM 3.12. *Let V be an upper probability on a measurable space $(\Omega, \mathcal{F})$. Given a bounded stationary process $\boldsymbol{Y} = \{Y_n\}_{n\in\mathbb{N}}$ on the capacity space $(\Omega, \mathcal{F}, V)$, if $\boldsymbol{Y}$ is ergodic, then there exists $Q \in \mathrm{core}(V)$ such that*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \int Y_1 dQ,\ V\text{-a.s.}$$

REMARK 3.13. There are many references on strong law of large numbers for capacities under different definitions, see, for example [9–11, 14, 32, 37]. In particular, Feng, Wu and Zhao [21] replaced the independent identically distributed hypothesis by stationarity and ergodicity. It was obtained in [21] that there exists a constant c such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = c,\ V\text{-a.s.}$$

By strengthening Birkhoff's ergodic theorem for capacity preserving systems, in the following, we obtain a probability $Q \in \mathrm{core}(V)$ such that $c = \int Y_1 dQ$. Moreover, if V is a probability, then $\mathrm{core}(V) = \{V\}$, and so this result is the classical strong law of large numbers in ergodic theory for probabilities (see [31, Page 24] for example).

Before proving Theorem 3.12, we prove a lemma.

LEMMA 3.14. *Let V be an upper probability on a measurable space $(\Omega, \mathcal{F})$, and let $\boldsymbol{Y} = \{Y_n\}_{n\in\mathbb{N}}$ be a bounded stationary process on a capacity space $(\Omega, \mathcal{F}, V)$. Then for any $\widetilde{P} \in \mathrm{core}(V_{\mathbf{Y}})$, there exists $P' \in \mathrm{core}(V)$ such that*

$$\widetilde{P}(C) = P'(\mathbf{Y}^{-1}(C)) \text{ for any } C \in \sigma(\mathcal{C}). \tag{16}$$

PROOF. Let $\mathcal{Y}^\infty = \mathbf{Y}^{-1}(\sigma(\mathcal{C}))$. It can be verified directly from the definition that $\mathcal{Y}^\infty$ is a σ-algebra. As the argument is standard, we omit the proof.

Now let $V_{\mathbf{Y}}$ be defined as in (15). Then $V_{\mathbf{Y}}$ is an upper probability as well. Indeed, 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 $\{C_n\}_{n\in\mathbb{N}} \subseteq \sigma(\mathcal{C})$, we have that $V_{\mathbf{Y}}$ is continuous, which together with the fact that $V_{\mathbf{Y}}(C) = \max_{P\in\mathrm{core}(V)} P(\mathbf{Y}^{-1}C)$ for any $C \in \mathcal{F}$, implies that $V_{\mathbf{Y}}$ is an upper probability.

Fix $\widetilde{P} \in \mathrm{core}(V_{\mathbf{Y}})$. As $V_{\mathbf{Y}}$ is an upper probability, it follows that $\widetilde{P}$ is a probability on $\sigma(\mathcal{C})$. For any $A \in \mathcal{Y}^\infty$, let $C_A \in \sigma(\mathcal{C})$ such that $A = \mathbf{Y}^{-1}(C_A)$. Define a set function $P^* : \mathcal{Y}^\infty \to [0,1]$ via

$$P^*(A) = \widetilde{P}(C_A) \text{ for any } A \in \mathcal{Y}^\infty.$$

First, we prove that the notion is well-defined, i.e., for any $A \in \mathcal{Y}^\infty$, if there exist $C_A, C'_A \in \sigma(\mathcal{C})$ such that $A = \mathbf{Y}^{-1}(C_A) = \mathbf{Y}^{-1}(C'_A)$ then $\widetilde{P}(C_A \Delta C'_A) = 0$, where $C_A \Delta C'_A = (C_A \setminus C'_A) \cup (C'_A \setminus C_A)$. Indeed, since $\widetilde{P} \in \mathrm{core}(V_{\mathbf{Y}})$, one has that

$$\begin{aligned}
\widetilde{P}(C_A \Delta C'_A) \le & V_{\mathbf{Y}}(C_A \Delta C'_A) \le V_{\mathbf{Y}}(C_A \cap (C'_A)^c) + V_{\mathbf{Y}}(C'_A \cap (C_A)^c) \\
= & V((\mathbf{Y}^{-1}C_A) \cap (\mathbf{Y}^{-1}(C'_A)^c)) + V((\mathbf{Y}^{-1}C'_A) \cap (\mathbf{Y}^{-1}(C_A)^c)) \\
= & V(A \cap A^c) + V(A \cap A^c) = 0,
\end{aligned}$$

proving this definition is well-defined.

Now we prove that P^* is a probability on $\mathcal{Y}^\infty$: First, it is easy to check that $P^*(\emptyset)=0$ and $P^*(\Omega)=1$. Next we only need to prove the σ-additivity of P^*. Fix any $\{A_n\}_{n\in\mathbb{N}}\subset\mathcal{Y}^\infty$ with $A_i\cap A_j=\emptyset$ for any $i\neq j$. Then

$$\widetilde{P}(C_{A_i}\cap C_{A_j})\le V_{\mathbf{Y}}(C_{A_i}\cap C_{A_j})=V(\mathbf{Y}^{-1}C_{A_i}\cap\mathbf{Y}^{-1}C_{A_j})=V(A_j\cap A_j)=0,\ \text{for any } i\neq j,$$

which together with the fact that $\cup_{i=1}^\infty A_i=\mathbf{Y}^{-1}(\cup_{i=1}^\infty C_{A_i})$ implies that

$$P^*(\cup_{i=1}^\infty A_i)=\widetilde{P}(\cup_{i=1}^\infty C_{A_i})=\sum_{i=1}^\infty\widetilde{P}(C_{A_i})=\sum_{i=1}^\infty P^*(A_i).$$

Moreover, by the construction, it is easy to check that $\widetilde{P}(C)=P(\mathbf{Y}^{-1}(C))$ for any $C\in\sigma(\mathcal{C})$.

To finish the proof, now we only need to prove P^* can be extended to a probability P' on $\mathcal{F}$ such that $P'|_{\mathcal{Y}^\infty}=P^*$ and $P'\in\text{core}(V)$. Define a functional I on $B(\Omega,\mathcal{F})$ by

$$I(f)=\sup_{P\in\text{core}(V)}\int fdP,\ \text{for any } f\in B(\Omega,\mathcal{F}).$$

By the linearity of integrals of probabilities $P\in\text{core}(V)$, it is easy to check that $I(f+g)\le I(f)+I(g)$ for any $f,g\in B(\Omega,\mathcal{F})$ and $I(\lambda f)=\lambda I(f)$ for any $\lambda\ge 0$ and $f\in B(\Omega,\mathcal{F})$, i.e., I is a sublinear functional on $B(\Omega,\mathcal{F})$. As $J(f):=\int fdP^*$ is a linear functional on $B(\Omega,\mathcal{Y}^\infty)$ with $J(f)\le I(f)$ for any $f\in B(\Omega,\mathcal{Y}^\infty)$ and $B(\Omega,\mathcal{Y}^\infty)$ is a linear subspace of $B(\Omega,\mathcal{F})$, it follows from Hahn-Banach dominated extension theorem that there exists a bounded linear functional J' on $B(\Omega,\mathcal{F})$ such that

$$J'(f)=J(f),\ \text{for any } B(\Omega,\mathcal{Y}^\infty), \tag{17}$$

and

$$J'(f)\le I(f),\ \text{for any } f\in B(\Omega,\mathcal{F}). \tag{18}$$

By Riesz representation theorem (see [1, Lemma 14.3] for example), there exists $P'\in\Delta(\Omega,\mathcal{F})$ such that $J'(f)=\int fdP'$ for any $f\in B(\Omega,\mathcal{F})$.

Now we prove that P' is desired. By (18), we have that for any $A\in\mathcal{F}$, $P'(A)\le I(1_A)=V(A)$, and hence $P'\in\text{core}(V)$. In particular, P' is a probability on $(\Omega,\mathcal{F})$, as V is an upper probability. From (17), for any $A\in\mathcal{Y}^\infty$, one deduces that $P'(A)=P^*(A)$, and hence $\widetilde{P}(C)=P^*(\mathbf{Y}^{-1}(C))=P'(\mathbf{Y}^{-1}(C))$ for any $C\in\sigma(\mathcal{C})$. The proof is completed. □

With the help of the above lemma, we are able to prove Theorem 3.12.

PROOF OF THEOREM 3.12. By Proposition 3.10, $V_{\mathbf{Y}}$ is the left shift transformation T-invariant. Define $f:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$ by $f(\mathbf{x})=f(x_1,x_2,x_3,\dots)=x_1$, for any $\mathbf{x}=(x_1,x_2,x_3,\dots)\in\mathbb{R}^{\mathbb{N}}$. Since $V_{\mathbf{Y}}$ is ergodic with respect to T, then by Theorem 3.3 we deduce that there exists $\widetilde{Q}\in\text{core}(V_{\mathbf{Y}})\cap\mathcal{M}^e(T)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\mathbf{x})=\int fd\widetilde{Q}\ \text{for } V_{\mathbf{Y}}\text{-a.s. } \mathbf{x}\in\mathbb{R}^{\mathbb{N}}.$$

Notice that $\frac{1}{n}\sum_{i=1}^n x_i=\frac{1}{n}\sum_{i=0}^{n-1}f(T^i\mathbf{x})$ for each $n\in\mathbb{N}$, we have

$$\mathbf{Y}^{-1}\left(\left\{\mathbf{x}\in\mathbb{R}^{\mathbb{N}}:\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i=\int fd\widetilde{Q}\right\}\right)=\left\{\omega\in\Omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n Y_i(\omega)=\int fd\widetilde{Q}\right\}.$$

Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Y_i(\omega)=\int fd\widetilde{Q}\text{ for }V\text{-a.s. }\omega\in\Omega.$$

By Lemma 3.14, there exists $Q\in\text{core}(V)$ such that $\widetilde{Q}(A)=Q(\mathbf{Y}^{-1}A)$ for any $A\in\mathcal{B}(\mathbb{R}^{\mathbb{N}})$.Thus,

$$\int_{\mathbb{R}^N}fd\widetilde{Q}=\int_{\mathbb{R}^{\mathbb{N}}}fdQ\left(\mathbf{Y}^{-1}\right)=\int_{\Omega}f(\mathbf{Y})dQ=\int_{\Omega}Y_1dQ,$$

proving the theorem. □

4. Ergodicity in terms of independence. In this section, we characterize the ergodicity of upper probabilities in terms of independence. As a reminder, the ergodicity of an invariant probability measure can be characterized as follows (see, for example, (2.31) in [19]).

PROPOSITION 4.1. Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system. Then P is ergodic if and only if

$$\lim_{n\to\infty}\int\frac{1}{n}\sum_{i=0}^{n-1}1_B\cdot(1_C\circ T^i)dP=P(B)P(C)\text{ for any }B,C\in\mathcal{F}.$$

4.1. *Upper probabilities versus probabilities in terms of independence.* The following proposition shows that for an ergodic upper probability, it satisfies a corresponding result in Proposition 4.1 if and only if it is a probability.

PROPOSITION 4.2. Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system, where V is an ergodic upper probability. Then the following two statements are equivalent:

(i) $\lim_{n\to\infty}\int\frac{1}{n}\sum_{i=0}^{n-1}1_B\cdot(1_C\circ T^i)dV=V(B)V(C)$ for any $B,C\in\mathcal{F}$;

(ii) V is a probability.

PROOF. (ii) ⇒ (i). This is obtained by Proposition 4.1.
(i) ⇒ (ii). Let $Q\in\mathcal{M}^e(T)\cap\text{core}(V)$ be the measure obtained in Theorem 3.2. Fix any $B,C\in\mathcal{F}$. By Theorem 3.3, one has that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_B\cdot(1_C\circ T^i)=Q(C)\cdot 1_B,\ V\text{-a.s.}$$

This combined with Lemma 2.4 implies that

$$\lim_{n\to\infty}\int\frac{1}{n}\sum_{i=0}^{n-1}1_B\cdot(1_C\circ T^i)dV=V(B)Q(C).$$

Taking $B=\Omega$, then $V(B)=1$, and hence by (i), $Q(C)=V(C)$. The proof is completed, as $C\in\mathcal{F}$ is arbitrary. □

4.2. *Characterizations of ergodicity of upper probabilities by weak independence.* In this subsection, we provide the following characterization of ergodicity in terms of "weak" independence.

THEOREM 4.3. *Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is an upper probability. Then the following four statements are equivalent:*

(i) *V is ergodic;*

(ii) *there exists $Q \in \mathrm{core}(V) \cap \mathcal{M}^e(T)$ such that $V(A) = Q(A)$ for any $A \in \mathcal{I}$, and*

$$\lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} f \cdot (g \circ T^i) dV = \left(\int f dV\right)\left(\int g dQ\right)$$

for any $f, g \in B(\Omega, \mathcal{F})$ such that $g \geq 0$;

(iii) *there exists $Q \in \mathrm{core}(V) \cap \mathcal{M}(T)$ such that*

$$\lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} 1_B \cdot (1_C \circ T^i) dV = V(B)Q(C)$$

for any $B, C \in \mathcal{F}$;

(iv) *there exists $Q \in \Delta^\sigma(\Omega, \mathcal{F})$ such that $V|_\mathcal{I} \ll Q|_\mathcal{I}$ (i.e., for any $A \in \mathcal{I}$ if $Q(A) = 0$ then $V(A) = 0$) and*

$$\liminf_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} 1_B \cdot (1_C \circ T^i) dV \geq V(B)Q(C) \text{ for any } B, C \in \mathcal{F}. \tag{19}$$

PROOF. (i) $\Rightarrow$ (ii). Suppose that V is ergodic. Given $f, g \in B(\Omega, \mathcal{F})$ with $g \geq 0$, as V is an upper probability, and there exists $M > 0$ such that

$$-M \leq \frac{1}{n} \sum_{i=0}^{n-1} f(\omega) g(T^i \omega) \leq M \text{ for each } \omega \in \Omega \text{ and } n \in \mathbb{N},$$

it follows from Lemma 2.4 and Theorem 3.3 that there exists $Q \in \mathcal{M}^e(T) \cap \mathrm{core}(V)$ such that

$$\lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} f(\omega) g(T^i \omega) dV(\omega) = \int \left(f(\omega) \int g dQ\right) dV(\omega) = \left(\int f dV\right)\left(\int g dQ\right),$$

which finishes the proof of (ii).

(ii) $\Rightarrow$ (iii). This is obtained by taking $f = 1_B$ and $g = 1_C$.

(iii) $\Rightarrow$ (iv). It suffices to prove that $V|_\mathcal{I} \ll Q|_\mathcal{I}$. In fact, for any $A \in \mathcal{I}$ with $Q(A) = 0$, we have

$$0 = V(\Omega)Q(A) = \lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} 1_\Omega \cdot (1_A \circ T^i) dV = V(A),$$

proving (iv).

(iv) $\Rightarrow$ (i). Given $A \in \mathcal{I}$, it suffices to prove that $V(A) = 0$ or $V(A^c) = 0$. Indeed, let $B = A$ and $C = A^c$ in (19). Note that $A^c \in \mathcal{I}$, so $1_{A^c} \circ T^i = 1_{A^c}$, then we obtain that

$$Q(A^c)V(A) \leq 0.$$

If $V(A) \neq 0$, then $Q(A^c) = 0$, which, with the assumption that $V|_\mathcal{I} \ll Q|_\mathcal{I}$, implies that $V(A^c) = 0$. Thus, V is ergodic. □

Moreover, we provide a characterization of ergodicity of an upper probability via the asymptotic independence of probabilities in its core.

THEOREM 4.4. *Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system, where V is an upper probability. Then the following statements are equivalent:*

(i) *V is ergodic;*

(ii) *there exists $Q\in\mathrm{core}(V)\cap\mathcal{M}^e(T)$ such that for any $P\in\mathrm{core}(V)$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)=P(B)Q(C)\text{ for any }B,C\in\mathcal{F};$$

(iii) *there exists $Q\in\mathrm{core}(V)\cap\mathcal{M}(T)$ such that for any $P\in\mathrm{core}(V)$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)=P(B)Q(C)\text{ for any }B,C\in\mathcal{F}. \tag{20}$$

PROOF. (i) $\Rightarrow$ (ii). This is a direct consequence of Theorem 3.3 and dominated convergence theorem.

(ii) $\Rightarrow$ (iii). It is trivial.

(iii) $\Rightarrow$ (i). By (20), one has for any $P\in\mathrm{core}(V)$,

$$P(A)=P(A)Q(A),\text{ for any }A\in\mathcal{I}. \tag{21}$$

Firstly, we prove that Q is ergodic. Indeed, for any $A\in\mathcal{I}$, if $V(A)=0$ then $Q(A)=0$. If $V(A)>0$, then there exists $P\in\mathrm{core}(V)$ such that $P(A)>0$. It follows from (21) that $Q(A)=1$. Therefore, Q is ergodic.

Next, we prove that $Q|_{\mathcal{I}}=P|_{\mathcal{I}}$ for any $P\in\mathrm{core}(V)$. Fix any $P\in\mathrm{core}(V)$ and $A\in\mathcal{I}$. As Q is ergodic, $Q(A)=0$ or 1. By (21), if $Q(A)=0$ then $P(A)=0$. If $Q(A)=1$, then by applying (21) on A^c, we have $P(A^c)=0$, therefore $P(A)=1=Q(A)$. Thus, $Q|_{\mathcal{I}}=P|_{\mathcal{I}}$ for any $P\in\mathrm{core}(V)$, which together with Theorem 3.2, implies that V is ergodic. □

4.3. *Applications: weak independence for non-invariant probabilities in terms of the invariant skeleton.* In Corollary 3.4, Birkhoff's ergodic theorem for non-invariant probabilities was discussed. The following corollary as a direct consequence of the proof of Theorem 4.4 provides a further result in this direction as the asymptotic independence on large time average.

COROLLARY 4.5. Let $(\Omega,\mathcal{F},P)$ be a probability space, and let $T:\Omega\to\Omega$ be a measurable transformation. If there exists an upper probability V on $(\Omega,\mathcal{F})$ such that V is ergodic with respect to T, and $P\in\mathrm{core}(V)$, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)=P(B)Q(C)\text{ for any }B,C\in\mathcal{F}. \tag{22}$$

In particular,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P\circ T^{-i}=Q. \tag{23}$$

REMARK 4.6. (i) Formula (23) gives the construction of the ergodic probability $Q \in \mathcal{M}^e(T) \cap \mathrm{core}(V)$ as the average of $\{P \circ T^{-i}\}_{i=0}^{\infty}$ under the assumption that there exists an ergodic upper probability V with $P \in \mathrm{core}(V)$.

(ii) From the proof of Proposition 4.2, under the assumption that an upper probability mentioned in (i) exists, then the large time average $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} P' \circ T^{-i} = Q$ for all $P' \in \mathrm{core}(V)$ independent of what P' is used as long as $P' \in \mathrm{core}(V)$.

(iii) If T is invertible, then the σ-algebra of invariant subsets with respect to T is equal to that with respect to T^{-1}. If Q defined by (23) is ergodic with respect to T, then Q is also ergodic with respect to T^{-1}. Thus, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} P \circ T^{i} = Q$ on $\mathcal{F}$, which together with (23) implies that

$$\lim_{\substack{n+m\to\infty\\ m,n\geq 0}} \frac{1}{m+n+1}\sum_{i=-m}^{n} P\circ T^i = Q. \tag{24}$$

(iv) Instead of starting with a given ergodic upper probability, let us consider a probability space $(\Omega, \mathcal{F}, P)$ and an invertible measurable transformation $T : \Omega \to \Omega$. Suppose that the limit (23) exists, denoted by Q. Then by Vitali-Hahn-Saks's theorem (Lemma 2.2), we have that Q is a probability. It is obvious that $Q \in \mathcal{M}(T)$ and $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$. If we further assume that Q is ergodic, by the classical Birkhoff's ergodic theorem for invertible transformations, we have that

$$\lim_{\substack{n+m\to\infty\\ m,n\geq 0}} \frac{1}{n+m+1}\sum_{i=-m}^{n} 1_A(T^i\omega) = Q(A), \text{ for } Q\text{-a.s. } \omega\in\Omega.$$

Denote

$$\widetilde{\Omega} = \{\omega \in \Omega : \lim_{\substack{n+m\to\infty\\ m,n\geq 0}} \frac{1}{n+m+1}\sum_{i=-m}^{n} 1_A(T^i\omega) = Q(A)\}.$$

Then $\widetilde{\Omega} \in \mathcal{I}$, and so $P(\widetilde{\Omega}) = Q(\widetilde{\Omega}) = 1$. It follows that (24) holds true without referring to an upper probability beforehand.

However, we can construct an upper probability under above assumptions. Let us begin with a claim that $P \circ T^i$ is absolutely continuous with respect to Q for each $i \in \mathbb{Z}$. Indeed, for any $A \in \mathcal{F}$ with $Q(A) = 0$, then $Q(A^\infty) = 0$, where $A^\infty = \cup_{i=-\infty}^{\infty} T^i A$. As $A^\infty \in \mathcal{I}$, it follows that for any $i \in \mathbb{Z}$,

$$P(T^iA) \leq P(T^iA^\infty) = P(A^\infty) = Q(A^\infty) = 0,$$

proving this claim. Define

$$\lambda_{m,n} = \frac{1}{m+n+1}\sum_{i=-m}^{n} P\circ T^i, \ m,n\geq 0$$

and

$$V(A) = \sup_{m,n\in\mathbb{Z}_+}\{\lambda_{m,n}(A)\} \text{ for each } A\in\mathcal{F}. \tag{25}$$

Now by Vitali-Hahn-Saks's theorem again, for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $A \in \mathcal{F}$ if $Q(A) < \delta$ then $\lambda_{m,n}(A) < \epsilon$ for all $m, n \in \mathbb{Z}$. Thus, it is easy to check that V is continuous, so V is an upper probability. From the definition of V, it follows that V is T-invariant. Then $(\Omega, \mathcal{F}, V, T)$ is an upper probability preserving system. Moreover, if Q is ergodic, then V is also ergodic.

(v) Note that if we have no additional condition on Q then the supremum of a family of probabilities defined as (25) may not be an upper probability. For example, we consider $P=\delta_x$ for some non-periodic point $x\in\Omega$, and let

$$A_k=\{T^l x : |l|\geq k\} \text{ for each } k\in\mathbb{N}.$$

Then A_k decreases to $\emptyset$, as $k\to\infty$, and $V(A_k)=\max_{i\in\mathbb{Z}}\delta_x(T^iA_k)=1$ for each $k\in\mathbb{N}$. Thus, V is not continuous, and hence it is not an upper probability.

THEOREM 4.7. *Let $(\Omega,\mathcal{F},P)$ be a probability space, and let $T:\Omega\to\Omega$ be an invertible measurable transformation. Suppose that the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P\circ T^i$ exists, denoted by Q. Then $Q\in\mathcal{M}(T)$. Moreover, if Q is ergodic, then*

$$\lim_{\substack{n+m\to\infty\\ m,n\geq 0}}\frac{1}{m+n+1}\sum_{i=-m}^{n}P(B\cap T^iC)=P(B)Q(C) \text{ for any } B,C\in\mathcal{F} \tag{26}$$

and for any $f\in L^1(\Omega,\mathcal{F},Q)$,

$$\lim_{\substack{n+m\to\infty\\ m,n\geq 0}}\frac{1}{m+n+1}\sum_{i=-m}^{n}f(T^i\omega)=\int f dQ \text{ for } P\text{-a.s. } \omega\in\Omega. \tag{27}$$

Conversely, if (26) or (27) holds, then Q is ergodic. In particular, (26) and (27) are equivalent.

PROOF. Let V be defined by (25). Then V is an upper probability by the argument in Remark 4.6 (iv). By the definition of Q and the assumption that Q is ergodic, we can see that $Q\in\mathrm{core}(V)\cap\mathcal{M}^e(T)$ and $P|_{\mathcal{I}}=Q|_{\mathcal{I}}$. For any $A\in\mathcal{I}$, as Q is ergodic, $Q(A)\in\{0,1\}$. So $P(T^iA)=Q(A)\in\{0,1\}$ for any $i\in\mathbb{Z}$. Thus, either

$$P(T^iA)=0 \text{ for all } i\in\mathbb{Z} \text{ or } P(T^iA^c)=0 \text{ for all } i\in\mathbb{Z}.$$

Therefore, $V(A)=0$ or $V(A^c)=0$. This implies that V is ergodic. Then applying Corollary 4.5 and Theorem 3.3 on T and T^{-1}, we finish the proof of the first part of this theorem, as $P\in\mathrm{core}(V)$.

Now we prove the converse part. If (26) holds, then for any $A\in\mathcal{I}$, when $P(A)>0$, we consider $0=P(A\cap A^c)=P(A)Q(A^c)$, and hence $Q(A^c)=0$; when $P(A)=0$, we consider $0=P(A^c\cap A)=P(A^c)Q(A)=Q(A)$. Thus, Q is ergodic.

If (27) holds, then for any $A\in\mathcal{I}$, one has that $\mathbf{1}_A(\omega)=Q(A)$ for P-a.s. $\omega\in\Omega$. On the other hand, since $\mathbf{1}_A(\omega)=1$ for all $\omega\in A$, it follows that if $Q(A)>0$, then $Q(A)=1$. Therefore, $Q(A)\in\{0,1\}$ for all $A\in\mathcal{I}$, and hence Q is ergodic. □

REMARK 4.8. In the special case that P is a T-invariant probability, we have $V=Q=P$ and $P\circ T^i=P$, $i\in\mathbb{Z}$, then the results of Theorem 4.7 are results in classical ergodic theory without any extra condition imposed (e.g. see (1.1.4) in Da Prato and Zabczyk [13] corresponding to (26) and Birkhoff's law of large numbers corresponding to (27)). Our results are sharp in the classical ergodic theory and hold true for possibly non-invariant probabilities.

4.4. *Characterizations of ergodicity of continuous concave capacities.* In this subsection, we provide more characterizations of ergodicity of a special type of upper probabilities, namely, continuous concave capacities. Let us recall an important property of concave capacities.

LEMMA 4.9 (Proposition 3 in [41]). *If μ is a concave capacity on a measurable space $(\Omega, \mathcal{F})$, then*

$$\int f d\mu = \max_{P \in \mathrm{core}(\mu)} \int f dP \text{ for any } f \in B(\Omega, \mathcal{F}).$$

THEOREM 4.10. *Let $(\Omega, \mathcal{F}, \mu, T)$ be a capacity preserving system. If μ is continuous from above and concave then the following four statements are equivalent:*

(i) *μ is ergodic;*

(ii) *there exists $Q \in \mathrm{core}(\mu) \cap \mathcal{M}^e(T)$ such that $\mu(A) = Q(A)$ for any $A \in \mathcal{I}$, and*

$$\lim_{n\to\infty} \max_{P \in \mathrm{core}(\mu)} \frac{1}{n} \sum_{i=0}^{n-1} \int f \cdot (g \circ T^i) dP$$
$$= \max_{P \in \mathrm{core}(\mu)} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int f \cdot (g \circ T^i) dP = \int f d\mu \int g dQ$$

for any $f, g \in B(\Omega, \mathcal{F})$ such that $g \geq 0$;

(iii) *there exists $Q \in \mathrm{core}(\mu) \cap \mathcal{M}(T)$ such that*

$$\lim_{n\to\infty} \max_{P \in \mathrm{core}(\mu)} \frac{1}{n} \sum_{i=0}^{n-1} P(B \cap T^{-i}C) = \max_{P \in \mathrm{core}(\mu)} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P(B \cap T^{-i}C) = \mu(B)Q(C)$$

for any $B, C \in \mathcal{F}$;

(iv) *there exists $Q \in \Delta^\sigma(\Omega, \mathcal{F})$ such that $\mu|_{\mathcal{I}} \ll Q|_{\mathcal{I}}$ and*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(B \cap T^{-i}C) \geq \mu(B)Q(C) \text{ for any } B, C \in \mathcal{F}.$$

PROOF. (i) $\Rightarrow$ (ii). From [9, Page 3382 and 3383], since μ is concave and continuous from above, it follows that μ is an upper probability. Thus, by Theorem 3.2, there exists $Q \in \mathcal{M}^e(T) \cap \mathrm{core}(\mu)$ such that for any $A \in \mathcal{I}$, $Q(A) = P(A)$ for any $P \in \mathrm{core}(\mu)$. Given any $f, g \in B(\Omega, \mathcal{F})$ with $g \geq 0$, let

$$\mathcal{A}_g = \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i\omega) = \int g dQ\}.$$

Applying Birkhoff's ergodic theorem on the ergodic probability Q, one has $Q(\mathcal{A}_g) = 1$, which together with the invariance of $\mathcal{A}_g$, implies that

$$P(\mathcal{A}_g) = 1 \text{ for any } P \in \mathrm{core}(\mu). \tag{28}$$

By (28) and the dominated convergence theorem, one deduces that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int f \cdot (g \circ T^i) dP = \int f dP \int g dQ \text{ for any } P \in \mathrm{core}(\mu). \tag{29}$$

Using (28) again, one has $\mu(\mathcal{A}_g^c) = 0$, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i\omega) = \int g dQ \text{ for } \mu\text{-a.s. } \omega \in \Omega.$$

This implies that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(\omega)g(T^i\omega)=f(\omega)\int gdQ \text{ for } \mu\text{-a.s. } \omega\in\Omega.$$

By Lemma 2.4, we deduce that

$$\lim_{n\to\infty}\int\frac{1}{n}\sum_{i=0}^{n-1}f(\omega)g(T^i\omega)d\mu(\omega)=\int fd\mu\int gdQ.$$

Meanwhile, by Lemma 4.9, one has that

$$\begin{aligned}\int\frac{1}{n}\sum_{i=0}^{n-1}f(\omega)g(T^i\omega)d\mu(\omega)&=\max_{P\in\text{core}(\mu)}\int\frac{1}{n}\sum_{i=0}^{n-1}f(\omega)g(T^i\omega)dP(\omega)\\&=\max_{P\in\text{core}(\mu)}\frac{1}{n}\sum_{i=0}^{n-1}\int f(\omega)g(T^i\omega)dP(\omega).\end{aligned}$$

Thus, by (29),

$$\begin{aligned}\int fd\mu\int gdQ&=\lim_{n\to\infty}\max_{P\in\text{core}(\mu)}\frac{1}{n}\sum_{i=0}^{n-1}\int f(\omega)g(T^i\omega)dP(\omega)\\&\geq\max_{P\in\text{core}(\mu)}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\int f(\omega)g(T^i\omega)dP(\omega)\\&=\max_{P\in\text{core}(\mu)}\int fdP\int gdQ=\int fd\mu\int gdQ,\end{aligned}$$

which finishes the proof of (ii).

(ii) ⇒ (iii). This is obtained by taking $f=1_B$ and $g=1_C$.

(iii) ⇒ (iv). Since

$$\max_{P\in\text{core}(\mu)}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)\leq\frac{1}{n}\sum_{i=0}^{n-1}\mu(B\cap T^{-i}C) \text{ for any } B,C\in\mathcal{F},$$

it follows from (iii) that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(B\cap T^{-i}C)\geq\liminf_{n\to\infty}\max_{P\in\text{core}(\mu)}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)=\mu(B)Q(C) \text{ for any } B,C\in\mathcal{F}.$$

Now we prove $\mu|_{\mathcal{I}}\ll Q|_{\mathcal{I}}$. Indeed, for any $A\in\mathcal{I}$ with $Q(A)=0$, by (iii) for $B=\Omega$, $C=A$, one has $\mu(A)=\max_{P\in\text{core}(\mu)}P(A)=Q(A)=0$. This completes the proof of (iv).

(iv) ⇒ (i). It can be proved by the same arguments as in the proof of (iv) ⇒ (i) in Theorem 4.3. □

5. Weak mixing for capacities. In this section, we provide a formal definition of weak mixing for capacity preserving systems. In classical ergodic theory, weak mixing has been extensively studied for probability preserving systems. Naturally, the concept of weak mixing should play a similar role in characterizing the level of randomness or disorder in capacity preserving systems.

5.1. *Definition of weak mixing of invariant capacities.* In classical ergodic theory, weak mixing can be characterized by measurable eigenfunctions (see [19] for example). Namely, for a probability preserving system $(\Omega, \mathcal{F}, P, T)$, P is weakly mixing if and only if each $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{C}$ that satisfies $f \circ T = \lambda f$, P-a.s. for some $\lambda \in \mathbb{C}$, is constant, P-a.s. It follows that $\lambda = 1$ is the unique eigenvalue of the transformation operator $f \mapsto f \circ T$ and the eigenvalue is simple. Motivated by this characterization, we state a similar definition for capacities.

DEFINITION 5.1. Let $(\Omega, \mathcal{F}, \mu, T)$ be a capacity preserving system. The capacity μ is called weakly mixing (with respect to T) if each $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{C}$ satisfying that $f \circ T = \lambda f$, μ-a.s. for some $\lambda \in \mathbb{C}$, is constant, μ-a.s.

By Lemma 2.11, it is easy to see that for a subadditive capacity μ, weak mixing implies ergodicity. Examples of weakly mixing capacity preserving systems can be found in Section 6. It is easy to check that given any $\mathcal{F}$-measurable function f that is not zero μ-a.s., if there exists $\lambda \in \mathbb{C}$ such that $f \circ T = \lambda f$, μ-a.s., then $|\lambda| = 1$. This can be seen from the fact that the Choquet integral satisfies $\int |f|^2 d\mu = \int |f \circ T|^2 d\mu = |\lambda|^2 \int |f|^2 d\mu$.

5.2. *Weak mixing and ergodicity on the product space of upper probabilities.* Recall that for a probability preserving system $(\Omega, \mathcal{F}, P, T)$, P is weakly mixing if and only if $P \times P$ is ergodic with respect to $T \times T$ (see [19, Theorem 2.36]). This raises a natural question: is there a similar result for capacity preserving systems? Before answering this question, we must define the product system of two capacity preserving systems.

We note that from [15, Chapter 12], the Carathéodory's extension theorem from an algebra to a σ-algebra is not true for capacities. However, we have a natural method to define the product of two upper probabilities as follows: Let $V_i = \max_{P_i \in \text{core}(V_i)} P_i$ be two upper probabilities defined on measurable spaces $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, and define

$$V_1 \times V_2 = \sup_{(P_1, P_2) \in \text{core}(V_1) \times \text{core}(V_2)} P_1 \times P_2, \tag{30}$$

where $P_1 \times P_2$ is the product measure of P_1 and P_2. It is easy to check that for any $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$,

$$V_1 \times V_2(A \times B) = V_1(A) \cdot V_2(B).$$

Thus, the notion is well defined. We now prove that $V_1 \times V_2$ is an upper probability on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ by beginning with the following lemma.

LEMMA 5.2. *Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathcal{P} \subset \Delta^\sigma(\Omega, \mathcal{F})$. Then $\overline{\mathcal{P}}$ is a (weak*-)compact subset of $\Delta^\sigma(\Omega, \mathcal{F})$ if and only if there exists $\lambda \in \Delta^\sigma(\Omega, \mathcal{F})$ with $\lambda(A) \leq \sup_{P \in \mathcal{P}} P(A)$ for all $A \in \mathcal{F}$ such that*

$$\lim_{\lambda(A) \to 0} \sup_{P \in \mathcal{P}} P(A) = 0.$$

PROOF. By Theorem IV.9.2 and Corollary IV.9.3 of [18], we immediately obtain that $\overline{\mathcal{P}}$ is a sequentially weakly compact subset of $\Delta^\sigma(\Omega, \mathcal{F})$ if and only if there exists $\lambda \in \Delta^\sigma(\Omega, \mathcal{F})$ such that

$$\lambda(A) \leq \sup_{P \in \mathcal{P}} P(A) \quad \text{for all } A \in \mathcal{F},$$

and

$$\lim_{\lambda(A) \to 0} \sup_{P \in \mathcal{P}} P(A) = 0.$$

According to the Eberlein–Šmulian theorem (see [1, p. 241]), the sequential weak compactness of $\overline{\mathcal{P}}$ is equivalent to its weak compactness. Moreover, by [33, Lemma 4.1], the weak compactness of $\overline{\mathcal{P}}$ is also equivalent to its weak* compactness. This completes the proof. $\square$

Applying Lemma 5.2 we are able to prove that $V_1 \times V_2$ is an upper probability.

PROPOSITION 5.3. The set $\overline{\mathrm{core}(V_1) \times \mathrm{core}(V_2)}$ is a compact subset of $\Delta^\sigma(\Omega, \mathcal{F})$. In particular, $V_1 \times V_2$ defined as in (30) is an upper probability.

PROOF. For $i = 1, 2$, since V_i is an upper probability, $\mathrm{core}(V_i)$ is compact. Combined with Lemma 5.2, this implies that there exists $\lambda_i \in \mathrm{core}(V_i)$ such that

$$\lim_{\lambda_i(A_i)\to 0} \sup_{P_i \in \mathrm{core}(V_i)} P_i(A_i) = 0.$$

Namely, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\text{for any } A_i \in \mathcal{F}_i, \quad \lambda_i(A_i) < \delta \implies V_i(A_i) < \epsilon/2, \ i = 1, 2. \tag{31}$$

Using Lemma 5.2 again, to prove that $\overline{\mathrm{core}(V_1) \times \mathrm{core}(V_2)}$ is a compact subset of $\Delta^\sigma(\Omega, \mathcal{F})$, it suffices to show that for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\text{for any } \tilde{A} \in \mathcal{F}_1 \times \mathcal{F}_2, \quad \lambda_1 \times \lambda_2(\tilde{A}) < \delta^2 \implies V_1 \times V_2(\tilde{A}) < \epsilon. \tag{32}$$

To this end, for any $x_1 \in \Omega_1$, denote

$$\tilde{A}_{x_1} := \{x_2 \in \Omega_2 : (x_1, x_2) \in \tilde{A}\}.$$

By Fubini's theorem, one has

$$\lambda_1 \times \lambda_2(\tilde{A}) = \int_{\Omega_1} \lambda_2(\tilde{A}_{x_1}) d\lambda_1(x_1).$$

Thus, if $\lambda_1 \times \lambda_2(\tilde{A}) < \delta^2$, then

$$\lambda_1(\{x_1 \in \Omega_1 : \lambda_2(\tilde{A}_{x_1}) \geq \delta\}) \leq \frac{1}{\delta} \int_{\Omega_1} \lambda_2(\tilde{A}_{x_1}) d\lambda_1(x_1) < \delta.$$

By (31), we deduce that

$$V_1(\{x_1 \in \Omega_1 : \lambda_2(\tilde{A}_{x_1}) \geq \delta\}) < \epsilon/2. \tag{33}$$

Thus, for any $P_1 \times P_2 \in \mathrm{core}(V_1) \times \mathrm{core}(V_2)$, using Fubini's again,

$$\begin{aligned}
P_1 \times P_2(\tilde{A}) &= \int_{\Omega_1} P_2(\tilde{A}_{x_1}) dP_1(x_1) \\
&= \int_{\{x_1 \in \Omega_1 : P_2(\tilde{A}_{x_1}) < \epsilon/2\}} P_2(\tilde{A}_{x_1}) dP_1(x_1) + \int_{\{x_1 \in \Omega_1 : P_2(\tilde{A}_{x_1}) \geq \epsilon/2\}} P_2(\tilde{A}_{x_1}) dP_1(x_1) \\
&\overset{(31)}{\leq} \epsilon/2 + \int_{\{x_1 \in \Omega_1 : \lambda_2(\tilde{A}_{x_1}) \geq \delta\}} P_2(\tilde{A}_{x_1}) dP_1(x_1) \\
&\leq \epsilon/2 + V_1(\{x_1 \in \Omega_1 : \lambda_2(\tilde{A}_{x_1}) \geq \delta\}) \\
&\overset{(33)}{<} \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

By the arbitrariness of $P_1 \times P_2 \in \mathrm{core}(V_1) \times \mathrm{core}(V_2)$, (32) holds true.

Moreover, as

$$V_1 \times V_2 = \sup_{(P_1, P_2) \in \mathrm{core}(V_1) \times \mathrm{core}(V_2)} P_1 \times P_2 = \max_{(P_1, P_2) \in \overline{\mathrm{core}(V_1) \times \mathrm{core}(V_2)}} P_1 \times P_2,$$

it follows that $V_1 \times V_2$ is an upper probability. $\square$

It follows from $V_1 \times V_2$ being an upper probability that

$$V_1 \times V_2 = \max_{\widetilde{P} \in \mathrm{core}(V_1 \times V_2)} \widetilde{P}.$$

Note that, generally, $\overline{\mathrm{core}(V_1) \times \mathrm{core}(V_2)} \subsetneq \mathrm{core}(V_1 \times V_2)$. For example, in Example 2, $\frac{1}{2}(\bar{P}_1 \times \bar{P}_2 + \bar{P}_2 \times \bar{P}_1) \in \mathrm{core}(V_1 \times V_2)$, but not in $\overline{\mathrm{core}(V_1) \times \mathrm{core}(V_2)}$.

Consider two capacity preserving systems $(\Omega_1, \mathcal{F}_1, V_1, T_1)$ and $(\Omega_2, \mathcal{F}_2, V_2, T_2)$, where V_1 and V_2 are upper probabilities. Let $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, V_1 \times V_2)$ be their product space defined as above and $T_1 \times T_2 : \Omega_1 \times \Omega_2 \to \Omega_1 \times \Omega_2$, by $(\omega_1, \omega_2) \mapsto (T_1\omega_1, T_2\omega_2)$.

In general, when both V_1 is T_1-invariant and V_2 is T_2-invariant, we cannot expect their product $V_1 \times V_2$ to be $T_1 \times T_2$-invariant as well. This is because the product of upper probabilities does not satisfy Fubini's theorem (see, e.g. [15, Proposition 12.1]).

To ensure the $(T_1 \times T_2)$-invariance of $V_1 \times V_2$, it suffices that both $(\Omega_1, \mathcal{F}_1, V_1, T_1)$ and $(\Omega_2, \mathcal{F}_2, V_2, T_2)$ satisfy at least one of the following two conditions (not necessarily the same for each):

(a) T is invertible and V is T^{-1}-invariant;
(b) V is concave.

Indeed, we will prove that both cases satisfy condition (34), as stated in the following lemma.

LEMMA 5.4. *Let $(\Omega_i, \mathcal{F}_i, V_i, T_i)$, for $i = 1, 2$, be two capacity preserving systems, where V_1 and V_2 are upper probabilities. Suppose that for each $i = 1, 2$,*

$$(T_i)_* \,\mathrm{core}(V_i) = \mathrm{core}(V_i), \tag{34}$$

where $(T_i)_* P_i := P_i \circ T_i^{-1}$. Then the product upper probability $V_1 \times V_2$ is $(T_1 \times T_2)$-invariant.

PROOF. First, for any $i = 1, 2$, as V_i is invariant, for any $P_i \in \mathrm{core}(V_i)$ and $A_i \in \mathcal{F}_i$,

$$P_i(T_i^{-1}A_i) \le V_i(T_i^{-1}A_i) = V_i(A_i),$$

which implies that $(T_i)_* P_i \in \mathrm{core}(V_i)$. Therefore,

$$(T_i)_*(\mathrm{core}(V_i)) \subset \mathrm{core}(V_i), \tag{35}$$

hence for any $P_i \in \mathrm{core}(V_i)$, one has for any $\tilde{A} \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$\begin{aligned} V_1 \times V_2(\tilde{A}) &\ge \sup_{(P_1,P_2) \in \mathrm{core}(V_1) \times \mathrm{core}(V_2)} (T_1)_* P_1 \times (T_2)_* P_2(\tilde{A}) \\ &= \sup_{(P_1,P_2) \in \mathrm{core}(V_1) \times \mathrm{core}(V_2)} P_1 \times P_2((T_1 \times T_2)^{-1}\tilde{A}) \\ &= V_1 \times V_2((T_1 \times T_2)^{-1}\tilde{A}). \end{aligned}$$

where the first equality holds due to the fact that P_1 and P_2 are probabilities.

Conversely, for any fixed $\tilde{A} \in \mathcal{F}_1 \times \mathcal{F}_2$ and $\epsilon > 0$, there exist $P_i \in \mathrm{core}(V_i)$, $i = 1, 2$ such that $P_1 \times P_2(\tilde{A}) \ge V_1 \times V_2(\tilde{A}) - \epsilon$. According to (34), for each $i = 1, 2$, there exists $Q_i \in \mathrm{core}(V_i)$ such that

$$P_i = (T_i)_* Q_i. \tag{36}$$

By Fubini's theorem for probabilities, one has

$$
\begin{aligned}
V_1 \times V_2(\tilde{A}) \leq & P_1 \times P_2(\tilde{A}) + \epsilon \\
& = \int P_2(\tilde{A}_{x_1}) dP_1(x_1) + \epsilon \\
& \overset{(36)}{=} \int P_2(\tilde{A}_{x_1}) d((T_1)_* Q_1)(x_1) + \epsilon \\
& = \int P_2(\tilde{A}_{T_1(x_1)}) dQ_1(x_1) + \epsilon \\
& \overset{(36)}{=} \int (T_2)_* Q_2(\tilde{A}_{T_1(x_1)}) dQ_1(x_1) + \epsilon, \\
& = \int Q_2(((T_1 \times T_2)^{-1}\tilde{A})_{x_1}) dQ_1(x_1) + \epsilon \\
& = Q_1 \times Q_2((T_1 \times T_2)^{-1}\tilde{A}) + \epsilon,
\end{aligned} \tag{37}
$$

where the second-to-last equality holds due to

$$x_2 \in T_2^{-1}\tilde{A}_{T_1(x_1)} \Leftrightarrow (T_1(x_1), T_2(x_2)) \in \tilde{A} \Leftrightarrow (x_1, x_2) \in (T_1 \times T_2)^{-1}\tilde{A}.$$

Combining this with the arbitrariness of $\epsilon > 0$, one deduces that

$$V_1 \times V_2(\tilde{A}) \leq V_1 \times V_2((T_1 \times T_2)^{-1}\tilde{A}).$$

Hence, the proof is complete. □

The following result is a direct consequence of [40, Lemma 3.9] together with the Riesz representation theorem. For the sake of completeness, we include a proof below.

LEMMA 5.5. *Let $I : \mathrm{ba}(S, \Sigma) \to \mathbb{R}$ be a linear functional, where*

$$\mathrm{ba}(\Omega, \mathcal{F}) := \left\{ P : \mathcal{F} \to \mathbb{R} \;\middle|\; P \textit{ is finitely additive and of bounded variation} \right\}.$$

Then the following two statements are equivalent:

(i) *I is continuous with respect to weak*-topology;*
(ii) *there is $f \in B(\Omega, \mathcal{F})$ such that for all $P \in \mathrm{ba}(\Omega, \mathcal{F})$ we have $I(P) = \int f dP$.*

PROOF. (ii) $\Rightarrow$ (i) is clear. Now we prove (i) $\Rightarrow$ (ii). Given $f_1, \dots, f_n \in B(\Omega, \mathcal{F})$ and $r_1, \dots, r_n > 0$, let

$$U(f_1, \dots, f_n; r_1, \dots, r_n) := \{P \in \mathrm{ba}(\Omega, \mathcal{F}) : |\int f_i dP| < r_i \text{ for } 1 \leq i \leq n\}.$$

Since the dual space of $B(\Omega, \mathcal{F})$ is $\mathrm{ba}(\Omega, \mathcal{F})$, the Riesz representation theorem (see [18, Theorem IV.5.1]) implies the collection of sets of the form $U(f_1, \dots, f_n; r_1, \dots, r_n)$ forms a local base of the weak* topology at the origin in $\mathrm{ba}(\Omega, \mathcal{F})$; see [16, p. 12].

By the weak*-continuity of I, there exists $U := U(f_1, \dots, f_n; r_1, \dots, r_n)$ such that for any $P \in U$, one has

$$|I(P)| < 1. \tag{38}$$

In particular, the functional I satisfies condition (b) in [40, Lemma 3.9], which, in turn, implies condition (c) in the same lemma. Therefore, there exist coefficients $a_1, \ldots, a_n \in \mathbb{R}$ such that

$$I(P) = \sum_{i=1}^{n} a_i \int f_i \, dP, \quad \text{for all } P \in \mathrm{ba}(\Omega, \mathcal{F}).$$

Letting $f := \sum_{i=1}^{n} a_i f_i$, the proof is complete. □

We now verify that both Case (a) and Case (b) satisfy condition (34).

PROPOSITION 5.6. Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is an upper probability. Suppose that the system satisfies either Condition (a) or Condition (b). Then condition (34) holds.

PROOF. **Condition (a).** Assume that T is invertible and V is T^{-1}-invariant. Applying an argument similar to that in (35) to both T and T^{-1}, we obtain

$$T_* \, \mathrm{core}(V) \subset \mathrm{core}(V) \quad \text{and} \quad (T^{-1})_* \, \mathrm{core}(V) \subset \mathrm{core}(V),$$

which together imply

$$\mathrm{core}(V) = T_* \, \mathrm{core}(V),$$

hence (34) holds.

Condition (b). Assume that V is concave. By Lemma 4.9, one has

$$J(f) := \int f dV = \max_{P \in \mathrm{core}(V)} \int f dP.$$

In particular,

$$J(f \circ T) = J(f), \text{ for any } f \in B(\Omega, \mathcal{F}). \tag{39}$$

Suppose, for a contradiction, that there exists $Q \in \mathrm{core}(V) \setminus (T_* \, \mathrm{core}(V))$. It is easy to see that $T_* \, \mathrm{core}(V)$ is a convex compact subset of $\mathrm{core}(V)$. Then by Hahn–Banach separation theorem, there exists a continuous linear functional I such that

$$I(Q) > \sup_{P \in T_* \, \mathrm{core}(V)} I(P).$$

By Lemma 5.5, there exists $f \in B(\Omega, \mathcal{F})$ such that

$$\int f dQ > \sup_{P \in T_* \, \mathrm{core}(V)} \int f dP = \sup_{P \in \mathrm{core}(V)} \int f \circ T dP.$$

In particular,

$$J(f) \geq \int f dQ > J(f \circ T),$$

which contradicts with (39). Thus, it holds (34). □

In what follows, we focus on the study of weak mixing of upper probabilities. Recall that $\mathcal{M}^{wm}(T)$ is the set of all weakly mixing probabilities on $(\Omega, \mathcal{F})$. Similarly to Theorem 3.2, the following lemma provides a characterization of the cores of weakly mixing upper probabilities.

LEMMA 5.7. *Let* $(\Omega, \mathcal{F}, V, T)$ *be a capacity preserving system, where* V *is an upper probability. Then* V *is weakly mixing if and only if there exists a (unique)* $Q \in \mathcal{M}^{wm}(T) \cap \text{core}(V)$ *such that for any* $P \in \text{core}(V)$*,* $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$*.*

PROOF. ($\Rightarrow$) By Theorem 3.2, one has proved that there exists a unique $Q \in \mathcal{M}^e(T) \cap \text{core}(V)$ such that for any $P \in \text{core}(V)$, $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$. Thus, we only need to prove that Q is weakly mixing. To see this, consider an $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{C}$ with $f \circ T = \lambda f$, Q-a.s. for some $\lambda \in \mathbb{C}$. Since $Q(\{\omega \in \Omega : f(T\omega) = \lambda f(\omega)\}^c) = 0$ and V is ergodic, it follows from Corollary 3.6 that

$$V(\{\omega \in \Omega : f(T\omega) = \lambda f(\omega)\}^c) = 0.$$

Since V is weakly mixing, there exists a constant c_f such that $V(\{\omega \in \Omega : f(\omega) = c_f\}^c) = 0$. Since $Q \in \text{core}(V)$, one has $Q(\{\omega \in \Omega : f(\omega) = c_f\}^c) = 0$, i.e., f is constant, Q-a.s. This implies $Q \in \mathcal{M}^{wm}(T)$.

($\Leftarrow$) Consider an $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{C}$ with $f \circ T = \lambda f$, V-a.s. for some constant $\lambda \in \mathbb{C}$. Since $Q \in \text{core}(V)$, it follows that $Q(\{\omega \in \Omega : f(T\omega) = \lambda f(\omega)\}^c) = 0$, which together with the assumption $Q \in \mathcal{M}^{wm}(T)$, implies that there exists a constant c_f such that $Q(\{\omega \in \Omega : f(\omega) = c_f\}^c) = 0$. By Theorem 3.2, then V is ergodic. Thus, it follows from Corollary 3.6 that $V(\{\omega \in \Omega : f(\omega) = c_f\}^c) = 0$, i.e., f is constant V-a.s. Thus, V is weakly mixing. □

As a corollary of Lemma 5.7, we provide a Birkhoff's ergodic theorem along polynomials on weakly mixing upper probability spaces.

COROLLARY 5.8. Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where V is a weakly mixing upper probability, and let $p(x)$ be a polynomial with integer coefficients. Let Q be the unique weakly mixing probability given in Lemma 5.7. Then for $f \in L^r(\Omega, \mathcal{F}, Q)$, $r > 1$, one has that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^{p(i)}\omega) = \int f dQ \text{ for } V\text{-a.s. } \omega \in \Omega.$$

PROOF. Since Q is weakly mixing, it follows from Theorem 2.7 that

$$Q(\{\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^{p(i)}\omega) = \int f dQ\}^c) = 0.$$

By Corollary 3.6, we have

$$V(\{\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^{p(i)}\omega) = \int f dQ\}^c) = 0,$$

proving this theorem. □

REMARK 5.9. When $r = 1$ and $p(x) = x^2$, even if V is a weakly mixing probability, Corollary 5.8 may not hold. In fact, Buczolich and Mauldin [7] proved that it is not true that given a probability preserving system $(\Omega, \mathcal{F}, P, T)$ and $f \in L^1(\Omega, \mathcal{F}, P)$, the ergodicity mean $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f \circ T^{n^2}$ converges, P-a.s.

We now proceed to establish a correspondence between the weak mixing of a capacity preserving system and the ergodicity of its product system.

THEOREM 5.10. *Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving dynamical system, where the upper probability V satisfies condition (34). Then the following statements are equivalent:*

(i) *V is weakly mixing with respect to T;*

(ii) *For any capacity preserving system $(\Omega', \mathcal{F}', V', T')$ with V' being an ergodic upper probability and satisfying condition (34), $V \times V'$ is ergodic with respect to $T \times T'$;*

(iii) *$V \times V$ is ergodic with respect to $T \times T$.*

PROOF. (i) $\Rightarrow$ (ii). Denote $\mathcal{I}' = \{A' \in \mathcal{F}' : T'^{-1}A' = A'\}$ and $\widetilde{\mathcal{I}} = \{\widetilde{A} \in \mathcal{F} \times \mathcal{F}' : (T \times T')^{-1}\widetilde{A} = \widetilde{A}\}$. Applying Lemma 5.7 and Theorem 3.2 on $(\Omega, \mathcal{F}, V, T)$ and $(\Omega', \mathcal{F}', V', T')$, respectively, we have that there exists a unique $Q \in \mathcal{M}^{wm}(T) \cap \mathrm{core}(V)$ such that for any $P \in \mathrm{core}(V)$, $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$, and a unique $Q' \in \mathcal{M}^e(T') \cap \mathrm{core}(V')$ such that for any $P' \in \mathrm{core}(V')$, $P'|_{\mathcal{I}'} = Q'|_{\mathcal{I}'}$.

By contradiction, we assume that $V \times V'$ is not ergodic. Then there exists $F \in B(\Omega \times \Omega', \widetilde{\mathcal{I}})$ which is not constant, $V \times V'$-a.s. On the other hand, since $Q \in \mathcal{M}^{wm}(T)$ and $Q' \in \mathcal{M}^e(T)$, it follows that $Q \times Q'$ is ergodic (see Theorem 2.36 in [19]). Thus, there exists a constant $c_F \in \mathbb{C}$ such that

$$F = c_F, \quad Q \times Q'\text{-a.s.}$$

Let

$$\widetilde{A} = \{(\omega, \omega') \in \Omega \times \Omega' : F(\omega, \omega') = c_F\}^c \in \mathcal{F} \times \mathcal{F}'.$$

Then

$$V \times V'(\widetilde{A}) > 0 \text{ and } Q \times Q'(\widetilde{A}) = 0. \tag{40}$$

By Fubini's theorem (a suitable version can be found in [23, Theorem 2.36]), one has that

$$0 = Q \times Q'(\widetilde{A}) = \int 1_{\widetilde{A}} d(Q \times Q') = \int Q'(\widetilde{A}_\omega) dQ(\omega),$$

where $\widetilde{A}_\omega = \{\omega' \in \Omega' : (\omega, \omega') \in \Omega \times \Omega'\}$ for each $\omega \in \Omega$. Thus,

$$Q'(\widetilde{A}_\omega) = 0 \text{ for } Q\text{-a.s. } \omega \in \Omega.$$

By Corollary 3.6, we obtain that for Q-a.s. $\omega \in \Omega$, $V'(\widetilde{A}_\omega) = 0$. Thus, for Q-a.s. $\omega \in \Omega$, $P'(\widetilde{A}_\omega) = 0$ for any $P' \in \mathrm{core}(V')$. Note that $\{\omega \in \Omega : P'(\widetilde{A}_\omega) = 0\}$ is an $\mathcal{F}$-measurable set, as P' is a probability. Then we deduce that

$$Q(\{\omega \in \Omega : P'(\widetilde{A}_\omega) = 0\}^c) = 0, \text{ for any } P' \in \mathrm{core}(V').$$

Using Corollary 3.6 again, one has that

$$V(\{\omega \in \Omega : P'(\widetilde{A}_\omega) = 0\}^c) = 0, \text{ for any } P' \in \mathrm{core}(V').$$

Therefore, for any $P \in \mathrm{core}(V)$,

$$P(\{\omega \in \Omega : P'(\widetilde{A}_\omega) = 0\}^c) = 0.$$

Note that every $P \in \mathrm{core}(V)$ is a σ-additive probability. Hence, for any $P, P' \in \mathrm{core}(V)$, the product measure $P \times P'$ is also σ-additive. This allows us to apply the standard Fubini's theorem, yielding

$$P \times P'(\widetilde{A}) = \int P'(\widetilde{A}_\omega)\, dP(\omega) = 0.$$

As $(P,P') \in \mathrm{core}(V) \times \mathrm{core}(V')$ is arbitrary, it follows that

$$V \times V'(\widetilde{A}) = \sup_{(P,P') \in \mathrm{core}(V) \times \mathrm{core}(V')} P \times P'(\widetilde{A}) = 0,$$

which is a contradiction with (40). Thus, $V \times V'$ is ergodic.

(ii) $\Rightarrow$ (iii). This is trivial.

(iii) $\Rightarrow$ (i). Let f be an $\mathcal{F}$-measurable function with $f \circ T = \lambda f$, V-a.s., for some $\lambda \in \mathbb{C}$. If $f = 0$, V-a.s., there is nothing to prove. So we suppose that

$$V(\{\omega \in \Omega : f(\omega) = 0\}^c) > 0. \tag{41}$$

Let

$$F(\omega_1, \omega_2) = f(\omega_1)\overline{f(\omega_2)} \text{ for any } (\omega_1, \omega_2) \in \Omega \times \Omega.$$

Then $F(T\omega_1, T\omega_2) = |\lambda|^2 F(\omega_1, \omega_2)$ for $V \times V$-a.s. $(\omega_1, \omega_2) \in \Omega \times \Omega$. Since $f \circ T = \lambda f$, V-a.s., it follows that $\int |f \circ T| dV = |\lambda| \int |f| dV$. By (41), $\int |f| dV > 0$, and thus, by the T-invariance of V, we have that $|\lambda| = 1$. So $F(T\omega_1, T\omega_2) = F(\omega_1, \omega_2)$ for $V \times V$-a.s. $(\omega_1, \omega_2) \in \Omega \times \Omega$. Applying Lemma 2.11 on the ergodic upper probability $V \times V$, one deduces that there exists $a \in \mathbb{C}$ such that $F = a$, $V \times V$-a.s. Denote

$$\widetilde{\Omega} = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : f(\omega_1)\overline{f(\omega_2)} = a\}.$$

Then

$$V \times V(\widetilde{\Omega}^c) = 0. \tag{42}$$

By (41), there exists $P_0 \in \mathrm{core}(V)$ such that

$$P_0(\{\omega \in \Omega : f(\omega) \neq 0\}) > 0. \tag{43}$$

Now we prove that $f = \int f dP_0$, V-a.s. By contradiction, we assume that $V(\{\omega \in \Omega : f(\omega) = \int f dP_0\}^c) > 0$. In particular, there exists $P_1 \in \mathrm{core}(V)$ such that

$$P_1(\{\omega \in \Omega : f(\omega) \neq \int f dP_0\}) > 0. \tag{44}$$

Let $P = \frac{1}{2}P_0 + \frac{1}{2}P_1$. Since $\mathrm{core}(V)$ is convex, $P \in \mathrm{core}(V)$. According to (42), we have $P \times P(\widetilde{\Omega}) = 1$. By Fubini's theorem, for P-a.s. $\omega_2 \in \Omega$, $P(\widetilde{\Omega}_{\omega_2}) = 1$, where $\widetilde{\Omega}_{\omega_2} = \{\omega_1 \in \Omega : (\omega_1, \omega_2) \in \widetilde{\Omega}\}$. By (43), there exists $\omega_2 \in \Omega$ such that $f(\omega_2) \neq 0$ and $P(\widetilde{\Omega}_{\omega_2}) = 1$. Thus, for P-a.s. $\omega_1 \in \Omega$, $f(\omega_1) = a/\overline{f(\omega_2)}$. As $P = \frac{1}{2}P_0 + \frac{1}{2}P_1$, it follows that for P_0-a.s. $\omega_1 \in \Omega$, $f(\omega_1) = a/\overline{f(\omega_2)}$, which implies that

$$\int f(\omega_1) dP_0(\omega_1) = \int (a/\overline{f(\omega_2)}) dP_0(\omega_1) = a/\overline{f(\omega_2)}.$$

Therefore, $f(\omega_1) = \int f dP_0$ for P-a.s. $\omega_1 \in \Omega$, and hence it also holds P_1-a.s. This is a contradiction with (44). So $f = \int f dP_0$, V-a.s., which shows that V is weakly mixing. □

From Theorem 5.10, we have that in this situation the upper probability V is weakly mixing, and $V \times V$ is ergodic. Applying Lemma 5.7 to V, we obtain a weakly mixing probability Q on $(\Omega, \mathcal{F})$. It follows that the product probability $Q \times Q$ is ergodic on $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$. Moreover, Theorem 3.2 applied to $V \times V$ yields an ergodic probability $\widetilde{Q}$ on the same product space. The following result implies that these two probabilities are the same.

COROLLARY 5.11. Let $(\Omega,\mathcal{F},V,T)$ be a weakly mixing capacity preserving system, where V is an upper probability and satisfies condition (34), and $Q\in\mathcal{M}^{wm}(T)\cap\text{core}(V)$ be as in Lemma 5.7. Then for any $\widetilde{P}\in\text{core}(V\times V)$,

$$\widetilde{P}(\widetilde{A})=Q\times Q(\widetilde{A})\text{ for any }\widetilde{A}\in\widetilde{\mathcal{I}}.$$

Furthermore, $\mathcal{M}(T\times T)\cap\text{core}(V\times V)=\mathcal{M}^e(T\times T)\cap\text{core}(V\times V)=\mathcal{M}^{wm}(T)\times\mathcal{M}^{wm}(T)\cap\text{core}(V\times V)=\{Q\times Q\}$.

PROOF. By Theorem 3.2, there exists a unique $\widetilde{Q}\in\mathcal{M}^e(T\times T)\cap\text{core}(V\times V)$ such that for any $\widetilde{P}\in\text{core}(V\times V)$, $\widetilde{P}|_{\widetilde{\mathcal{I}}}=\widetilde{Q}|_{\widetilde{\mathcal{I}}}$. It is well known that $Q\times Q\in\mathcal{M}^e(T\times T)$, as Q is weakly mixing. Due to Lemma 2.8, to prove $\widetilde{Q}=Q\times Q$, we only need to prove that $Q\times Q|_{\mathcal{I}}=\widetilde{Q}|_{\mathcal{I}}$. Consider any $\widetilde{A}\in\widetilde{\mathcal{I}}$, then $\widetilde{Q}(\widetilde{A})=0$ or 1, as $\widetilde{Q}$ is ergodic. If $\widetilde{Q}(\widetilde{A})=0$, by Corollary 3.6, one has that $V\times V(\widetilde{A})=0$, which implies that $Q\times Q(\widetilde{A})=0$, as $Q\times Q\in\text{core}(V\times V)$. Similarly, if $\widetilde{Q}(\widetilde{A})=1$, one can prove that $Q\times Q(\widetilde{A})=1$. Thus, $Q\times Q|_{\mathcal{I}}=\widetilde{Q}|_{\mathcal{I}}$, hence this corollary is proved by applying Theorem 3.2. □

5.3. *Asymptotic independence and long time convergence in laws.* The following result shows that each element in the core of a weakly mixing upper probability has a kind of asymptotic independence.

THEOREM 5.12. *Let $(\Omega,\mathcal{F},V,T)$ be a capacity preserving system, where V is an upper probability and satisfies condition (34). Then the following three statements are equivalent:*

(i) *V is weakly mixing;*

(ii) *there exists $Q\in\text{core}(V)\cap\mathcal{M}^{wm}(T)$ such that for any $P\in\text{core}(V)$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|P(B\cap T^{-i}C)-P(B)Q(C)|^2=0\textit{ for any }B,C\in\mathcal{F}.$$

(iii) *there exists $Q\in\text{core}(V)\cap\mathcal{M}(T)$ such that for any $P\in\text{core}(V)$,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|P(B\cap T^{-i}C)-P(B)Q(C)|^2=0\textit{ for any }B,C\in\mathcal{F}. \tag{45}$$

PROOF. (i) ⇒ (ii). Combining Corollary 4.5 and Corollary 5.11, one deduces that there exists $Q\in\text{core}(V)\cap\mathcal{M}^{wm}(T)$ such that for any $P\in\text{core}(V)$,

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)^2\\&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P\times P((B\times B)\cap(T\times T)^{-i}(C\times C))=P(B)^2Q(C)^2\end{aligned}\tag{46}$$

for any $B,C\in\mathcal{F}$. Thus, using Corollary 4.5 again, we have that for any $B,C\in\mathcal{F}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|P(B\cap T^{-i}C)-P(B)Q(C)|^2$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} P(B\cap T^{-i}C)^2 - \lim_{n\to\infty} \frac{2}{n} \sum_{i=0}^{n-1} P(B\cap T^{-i}C)P(B)Q(C) + P(B)^2Q(C)^2$$
$$= P(B)^2Q(C)^2 - 2P(B)^2Q(C)^2 + P(B)^2Q(C)^2 = 0,$$

where the last equation is obtained by (46).

(ii) ⇒ (iii). This is trivial.

(iii) ⇒ (i). Since $Q\in \mathrm{core}(V)$, it follows from (45) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |Q(B\cap T^{-i}C) - Q(B)Q(C)|^2 = 0 \text{ for any } B, C \in \mathcal{F}.$$

By [46, Theorem 1.21], we have that Q is weakly mixing. For any $P\in \mathrm{core}(V)$ and $A\in\mathcal{I}$, taking $B=\Omega$ and $C=A$, we have $Q(C)=P(C)$. This shows that

$$P|_{\mathcal{I}} = Q|_{\mathcal{I}} \text{ for any } P\in \mathrm{core}(V).$$

By Lemma 5.7, we know that V is weakly mixing. The proof is completed. □

DEFINITION 5.13. A subset J of $\mathbb{N}$, we define the upper density of J by

$$\overline{D}(J) = \limsup_{n\to\infty} \frac{|J\cap\{1,2,\dots,n\}|}{n}$$

and the lower density by

$$\underline{D}(J) = \liminf_{n\to\infty} \frac{|J\cap\{1,2,\dots,n\}|}{n}.$$

In particular, if $\overline{D}(J)=\underline{D}(J)$, we call the set J has natural density, denoted by $D(J)$.

Finally, we use an equivalent characterization to end this section. We recall that from Theorems 1.20 and 1.23 in [46], given any weakly mixing probability preserving system $(\Omega,\mathcal{F},Q,T)$, for any $f,g\in L^2(\Omega,\mathcal{F},Q)$, there exists a subset $J_{f,g}$ of $\mathbb{N}$ with $D(J_{f,g})=0$ such that

$$\lim_{n\notin J_{f,g}, n\to\infty} \int (f\circ T^n)\cdot g dQ = \int f dQ \int g dQ. \tag{47}$$

We change this slightly for our needs.

LEMMA 5.14. *Let $(\Omega,\mathcal{F},Q,T)$ be a weakly mixing probability preserving system, where $(\Omega,\mathcal{F})$ is standard. Then for any $f\in L^\infty(\Omega,\mathcal{F},Q)$, there exists a subset $J=J_f$ of $\mathbb{N}$ such that*

$$\lim_{n\notin J, n\to\infty} \int (f\circ T^n)\cdot g dQ = \int f dQ \int g dQ, \text{ for any } g\in L^1(\Omega,\mathcal{F},Q).$$

PROOF. Fix $f\in L^\infty(\Omega,\mathcal{F},Q)$. Since $(\Omega,\mathcal{F})$ is standard, we can find $\{g_m\}_{m\in\mathbb{N}}\subset L^2(\Omega,\mathcal{F},Q)$ such that it is a dense subset of $L^1(\Omega,\mathcal{F},Q)$. By (47), for each $m\in\mathbb{N}$, there exists $J_m\subset\mathbb{N}$ with $D(J_m)=0$ such that

$$\lim_{n\notin J_m, n\to\infty} \int (f\circ T^n)\cdot g_m dQ = \int f dQ \int g_m dQ. \tag{48}$$

Let $J'_m = \cup_{i=1}^m J_i$ for each $m\in\mathbb{N}$. Then $D(J'_m)=0$ and $J'_m\subset J'_{m+1}$ for each $m\in\mathbb{N}$. Thus, for any $m\in\mathbb{N}$, there exists $N_m\in\mathbb{N}$ such that for any $N\geq N_m$, $|J'_m\cap[1,N]|/N<1/m$. Let

$$J = \cup_{m=1}^\infty (J'_m \cap [N_{m-1}+1, N_m]).$$

Then for any $K \in \mathbb{N}$, there exists $m \in \mathbb{N}$ such that $N_m + 1 \le K < N_{m+1}$, and hence $J \cap [1,K] \subset J'_m \cap [1,K]$.Thus, $\frac{|J\cap[1,K]|}{K} < 1/m$, which shows that $D(J) = 0$. Meanwhile, by the construction of J and (48), one has for any $m \in \mathbb{N}$,

$$\lim_{n\notin J, n\to\infty} \int (f \circ T^n) \cdot g_m dQ = \int f dQ \int g_m dQ.$$

Since $f \in L^\infty(\Omega, \mathcal{F}, Q)$, $L := \max\{\|f\|_{\infty,Q}, 1\} < \infty$. For any $g \in L^1(\Omega, \mathcal{F}, Q)$, there exists a subsequence $\{q_m\}_{m\in\mathbb{N}}$ of $\{g_m\}_{m\in\mathbb{N}}$ such that $\lim_{m\to\infty} \|q_m - g\|_{1,Q} = 0$. Thus, for any $\epsilon > 0$, there exists $M > 0$ such that for any $m \ge M$, $\|q_m - g\|_{1,Q} < \epsilon/(4L)$. Therefore,

$$\begin{aligned}
&|\int (f\circ T^n)\cdot g dQ - \int f dQ \int g dQ| \\
\le& |\int (f\circ T^n)\cdot (g - q_m) dQ| + |\int (f\circ T^n)\cdot q_m dQ - \int f dQ \int q_m dQ| \\
&+ |\int f dQ \int (q_m - g) dQ| \\
\le& L\cdot \frac{\epsilon}{4L} + |\int (f\circ T^n)\cdot q_m dQ - \int f dQ \int q_m dQ| + L\cdot\frac{\epsilon}{4L} \\
\le& \epsilon/2 + |\int (f\circ T^n)\cdot q_m dQ - \int f dQ \int q_m dQ|,
\end{aligned}$$

which shows that

$$\limsup_{n\notin J, n\to\infty} |\int (f\circ T^n)\cdot g dQ - \int f dQ \int g dQ| \le \epsilon/2.$$

Letting $\epsilon \to 0$, one has that

$$\lim_{n\notin J, n\to\infty} \int (f\circ T^n)\cdot g dQ = \int f dQ \int g dQ.$$

The proof is completed, as $g \in L^1(\Omega, \mathcal{F}, Q)$ is arbitrary. □

REMARK 5.15. If the measurable space $(\Omega, \mathcal{F})$ is not standard, then by a similar argument, we can prove a weaker result: for any $f \in L^\infty(\Omega, \mathcal{F}, Q)$ and $g \in L^1(\Omega, \mathcal{F}, Q)$, there exists a subset $J = J_{f,g}$ of $\mathbb{N}$ such that

$$\lim_{n\notin J, n\to\infty} \int (f\circ T^n)\cdot g dQ = \int f dQ \int g dQ.$$

THEOREM 5.16. *Let $(\Omega, \mathcal{F}, V, T)$ be a capacity preserving system, where $(\Omega, \mathcal{F})$ is standard and V is an upper probability. Then V is weakly mixing if and only if there exists $Q \in \mathcal{M}^{wm}(T) \cap \mathrm{core}(V)$ such that for any $f \in L^\infty(\Omega, \mathcal{F}, Q)$, there exists a subset $J = J_f$ of $\mathbb{N}$ with $D(J) = 0$ such that for any $g \in L^\infty(\Omega, \mathcal{F}, Q)$,*

$$\lim_{n\notin J, n\to\infty} \int (f\circ T^n)\cdot g dP = \int f dQ \int g dP \text{ for any } P \in \mathrm{core}(V). \tag{49}$$

PROOF. (⇒) Since V is weakly mixing, it follows from Lemma 5.7, there exists $Q \in \mathcal{M}^{wm}(T) \cap \mathrm{core}(V)$ such that $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$ for all $P \in \mathrm{core}(V)$. By Corollary 3.6, we have that for any $A \in \mathcal{F}$ with $Q(A) = 0$, one has that $V(A) = 0$, and hence $P(A) = 0$ for any $P \in \mathrm{core}(V)$. This means that $P \ll Q$ for any $P \in \mathrm{core}(V)$.

Fix any $P \in \mathrm{core}(V)$. As $P \ll Q$, let $h = \frac{dP}{dQ} \in L^1(\Omega, \mathcal{F}, Q)$ be the Radon-Nikodym derivative of P with respect to Q. Note that for any $g \in L^\infty(\Omega, \mathcal{F}, Q)$, $g \cdot h \in L^1(\Omega, \mathcal{F}, Q)$. By Lemma 5.14, one has that for any $f \in L^\infty(\Omega, \mathcal{F}, Q)$, there exists a subset $J = J_f$ of $\mathbb{N}$ with $D(J) = 0$ such that for any $g \in L^\infty(\Omega, \mathcal{F}, Q)$,

$$\begin{aligned}\lim_{n\notin J, n\to\infty} \int (f \circ T^n) \cdot g dP &= \lim_{n\notin J, n\to\infty} \int (f \circ T^n) \cdot (g \cdot h) dQ \\ &= \int f dQ \int g \cdot h dQ = \int f dQ \int g dP = \int f dQ \int g dP.\end{aligned}$$

($\Leftarrow$) For any $A \in \mathcal{F}$, consider $f = 1_A$ and $g = 1$ in (49). We have that for any $P \in \mathrm{core}(V)$,

$$Q(A) = \lim_{n\notin J,\ n\to\infty} P(T^{-n}A) \leq \limsup_{n\notin J,\ n\to\infty} V(T^{-n}A) = V(A).$$

This shows that $Q \in \mathrm{core}(V)$. From Lemma 5.7, it suffices to prove that for any $P \in \mathrm{core}(V)$, $P|_{\mathcal{I}} = Q|_{\mathcal{I}}$. This is clear by taking $f = 1_A$ for any $A \in \mathcal{I}$ and $g = 1$ in (49). □

REMARK 5.17. According to Remark 5.15, if the measurable space $(\Omega, \mathcal{F})$ is not standard, we have the following result: V is weakly mixing if and only if there exists $Q \in \mathcal{M}^{wm}(T) \cap \mathrm{core}(V)$ such that for any $f, g \in L^\infty(\Omega, \mathcal{F}, Q)$ and $P \in \mathrm{core}(V)$, there exists a subset $J = J_{f,g,P}$ of $\mathbb{N}$ with $D(J) = 0$ such that

$$\lim_{n\notin J, n\to\infty} \int (f \circ T^n) \cdot g dP = \int f dQ \int g dP.$$

5.4. *Applications: asymptotic independence and convergence in laws for non-invariant probabilities.* In this subsection, we investigate recurrent and ergodic theorems for non-invariant probabilities. Let us begin with the following lemma.

LEMMA 5.18. *Under the same conditions as in Theorem 4.7, for any* $\epsilon > 0$, *there exists* $\delta > 0$ *such that for any* $f, g \in L^\infty(\Omega, \mathcal{F}, Q)$ *if* $\|f - g\|_{1,Q} < \delta$ *then*

$$\|f - g\|_{1, P\circ T^{-m}} < \epsilon, \textit{ for any } m \in \mathbb{Z}_+.$$

PROOF. By the construction of Q, we have $P \circ T^{-m} \ll Q$ for every $m \in \mathbb{Z}_+$. Hence, if $h \in L^\infty(\Omega, \mathcal{F}, Q)$ satisfies

$$Q\big(\{\omega : |h(\omega)| > M\}\big) = 0$$

for some $M > 0$, then also

$$P\big(T^{-m}\{\omega : |h(\omega)| > M\}\big) = (P \circ T^{-m})\big(\{\omega : |h(\omega)| > M\}\big) = 0.$$

Choosing $M = \|h\|_{\infty,Q} + \delta$ with an arbitrary $\delta > 0$ gives

$$|h(\omega)| \leq \|h\|_{\infty,Q} + \delta,\ P \circ T^{-m}\text{-a.s.}$$

Because $\delta > 0$ is arbitrary, letting $\delta \to 0$ yields

$$\|h\|_{\infty, P\circ T^{-m}} \leq \|h\|_{\infty,Q}, \text{ for every } m \in \mathbb{Z}_+. \tag{50}$$

By Remark 4.6 (iv), for any $\epsilon > 0$ there exists $\delta' > 0$ such that for any $A \in \mathcal{F}$ with $Q(A) < \delta'$,

$$P(T^{-m}A) < \frac{\epsilon}{2(\|f\|_{\infty,Q} + \|g\|_{\infty,Q})}, \text{ for any } m \in \mathbb{Z}_+.$$

Let $\delta = \frac{\delta' \epsilon}{4}$. If $\|f-g\|_{1,Q} < \delta$, then

$$Q(A_\epsilon) \le (2/\epsilon) \cdot \int_{A_\epsilon} |f-g| dQ \le (2/\epsilon) \cdot \|f-g\|_{1,Q} < (2/\epsilon) \cdot \delta < \delta',$$

where $A_\epsilon = \{\omega \in \Omega : |f(\omega) - g(\omega)| > \epsilon/2\}$. Thus,

$$P(T^{-m} A_\epsilon) < \frac{\epsilon}{2(\|f\|_{\infty,Q} + \|g\|_{\infty,Q})}, \text{ for any } m \in \mathbb{Z}_+.$$

This, together with (50), implies that for each $m \in \mathbb{Z}_+$,

$$\begin{aligned}\|f-g\|_{1,P\circ T^{-m}} &= \int_{A_\epsilon} |f-g| d(P \circ T^{-m}) + \int_{A_\epsilon^c} |f-g| d(P \circ T^{-m}) \\ &< (\|f\|_{\infty,Q} + \|g\|_{\infty,Q}) \cdot \frac{\epsilon}{2(\|f\|_{\infty,Q} + \|g\|_{\infty,Q})} + \epsilon/2 \\ &\le \epsilon,\end{aligned}$$

proving the lemma. □

THEOREM 5.19. *Under the same conditions in Theorem 4.7, the following statements are equivalent:*

(i) *Q is weakly mixing;*

(ii) $\lim_{\substack{n+m\to\infty \\ m,n\ge 0}} \frac{1}{m+n+1} \sum_{i=-m}^{n} |P(B \cap T^{-i}C) - P(B)Q(C)|^2 = 0$ *for any* $B, C \in \mathcal{F}$;

(iii) *for any $f, g \in L^\infty(\Omega, \mathcal{F}, Q)$, there exists a subset $J = J_{f,g}$ of $\mathbb{N}$ with $D(J) = 0$ such that*

$$\lim_{n\notin J, n\to\infty} \int (f \circ T^n) \cdot g dP = \int f dQ \int g dP. \tag{51}$$

Moreover, these equivalent statements can imply that letting $p(x)$ be a polynomial with integer coefficients, then for any $f \in L^r(\Omega, \mathcal{F}, Q)$, $r > 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^{p(i)} \omega) = \int f dQ \text{ for } P\text{-a.s. } \omega \in \Omega.$$

PROOF. (i) $\Rightarrow$ (ii). Let V be defined by (25). Under the assumptions of this theorem, $Q \in \mathcal{M}^{wm}(T) \cap \mathrm{core}(V)$ such that $Q|_{\mathcal{I}} = V|_{\mathcal{I}}$, and so V is also weakly mixing. Applying a similar argument of (iii) in Remark 4.6 on Theorem 5.12, we obtain (ii).

(ii) $\Rightarrow$ (iii). It follows from [46, Theorem 1.20] that (ii) is equivalent to that for any $B, C \in \mathcal{F}$, there exists a subset $J_{B,C}$ of $\mathbb{N}$ such that $\lim_{n\notin J_{B,C}, n\to\infty} P(B \cap T^{-n}C) = P(B)Q(C)$. It turns out that for any simple functions f, g there exists a subset $J_{f,g}$ of $\mathbb{N}$ such that

$$\lim_{n\notin J_{f,g}, n\to\infty} \int (f \circ T^n) \cdot g dP = \int f dQ \int g dP.$$

Now we prove the above equation holds for any $f, g \in L^\infty(\Omega, \mathcal{F}, Q)$. Given $f, g \in L^\infty(\Omega, \mathcal{F}, Q)$, there exists two increasing sequences of simple functions $\{f_k\}_{k\in\mathbb{N}}$ and $\{g_k\}_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \|f_k - f\|_{1,Q} = \lim_{k\to\infty} \|g_k - g\|_{1,Q} = 0.$$

By Lemma 5.18, for any $\epsilon > 0$, there exists $K > 0$ such that

$$\|f_K - f\|_{1,P\circ T^{-m}} \le \frac{\epsilon}{8\|g\|_{\infty,Q}},\ \|g_K - g\|_{1,P} \le \frac{\epsilon}{8\|f\|_{\infty,Q}} \text{ for any } m \in \mathbb{Z}_+, \tag{52}$$

and

$$\|f_K - f\|_{1,Q} \le \frac{\epsilon}{8\|g\|_{\infty,Q}},\ \|g_K - g\|_{1,Q} \le \frac{\epsilon}{8\|f\|_{\infty,Q}}. \tag{53}$$

Since f_K and g_K are simple functions, we have that

$$\lim_{n\notin J_{f_K,g_K}, n\to\infty} \int (f_K \circ T^n)\cdot g_K dP = \int f_K dQ \int g_K dP.$$

Thus, there exists $N_K > 0$ such that for any $n \notin J_{f_K,g_K}$ and $n \ge N_K$

$$|\int (f_K \circ T^n)\cdot g_K dP - \int f_K dQ \cdot \int g_K dP| < \epsilon/8. \tag{54}$$

Since $P \ll Q$, one has that $\|h\|_{\infty,P} \le \|h\|_{\infty,Q}$ for any $\mathcal{F}$-measurable function h. Indeed, as $Q(\{\omega\in\Omega : h(\omega) > \|h\|_{\infty,Q}\}) = 0$, it follows that $P(\{\omega\in\Omega : h(\omega) > \|h\|_{\infty,Q}\}) = 0$, and hence $\|h\|_{\infty,P} \le \|h\|_{\infty,Q}$. Thus, we have that for any $n \notin J_{f_K,g_K}$ and $n \ge N_K$,

$$\begin{aligned}
&|\int (f\circ T^n)\cdot g dP - \int f dQ \cdot \int g dP| \\
\le& |\int (f\circ T^n)\cdot g dP - \int (f\circ T^n)\cdot g_K dP| \\
&+ |\int (f\circ T^n)\cdot g_K dP - \int (f_K\circ T^n)\cdot g_K dP| \\
&+ |\int (f_K\circ T^n)\cdot g_K dP - \int f_K dQ \cdot \int g_K dP| \\
&+ |\int f_K dQ \cdot \int g_K dP - \int f_K dQ \cdot \int g dP| \\
&+ |\int f_K dQ \cdot \int g dP - \int f dQ \cdot \int g dP| \\
\le& \|f\|_{\infty,Q}\cdot \|g - g_K\|_{1,P} + \|g\|_{\infty,Q}\cdot\|f - f_K\|_{1,P\circ T^{-n}} \\
&+ \epsilon/8 + \|f\|_{\infty,Q}\cdot\|g - g_K\|_{1,P} + \|g\|_{\infty,Q}\cdot \|f - f_K\|_{1,Q} \\
\le& \epsilon/8 + \epsilon/8 + \epsilon/8 + \epsilon/8 + \epsilon/8 < \epsilon.
\end{aligned}$$

Similar to the construction of the subset with zero density in Lemma 5.14, we can prove (iii).

(iii) ⇒ (i). Firstly, we prove that Q is ergodic. Indeed, for any $A \in \mathcal{I}$, if $P(A) > 0$, letting $f = 1_{A^c}$ and $g = 1_A$ in (51), we have that $0 = P(A\cap A^c) = Q(A^c)P(A)$, which implies that $Q(A^c) = 0$; if $P(A) = 0$, letting $f = 1_A$ and $g = 1_{A^c}$ in (51), we have that $Q(A) = 0$. Thus, $Q(A) \in \{0, 1\}$, and hence Q is ergodic.

Now we prove Q is weakly mixing. Fix any $\mathcal{F}$-measurable function h with $\int |h|dQ > 0$ such that $h\circ T = \lambda h$, Q-a.s. for some $\lambda \in \mathbb{C}\setminus\{0\}$. Since Q is ergodic and $|\lambda| = 1$, it follows that $|h|$ is constant, Q-a.s. In particular, $h \in L^\infty(\Omega, \mathcal{F}, Q)$. Taking $f = \frac{\bar h}{|h|^2}$ and $g = h$ in (51), one has that

$$\lim_{n\to\infty, n\notin J_{f,g}} \lambda^n = \int \frac{\bar h}{|h|^2} dQ \int h dP. \tag{55}$$

Since $|\lambda|=1$, we write $\lambda=e^{i\theta}$ for some $\theta\in\mathbb{R}$. If $\theta\in\mathbb{Q}\setminus\{0\}$, then $\{\lambda^n\}_{n\in\mathbb{Z}_+}$ is a periodic sequence with a period of $t>1$, and hence for any constant $c\in\mathbb{C}$ there is no subset J of $\mathbb{N}$ with $D(J)=0$ such that $\lim_{n\to\infty,n\notin J}\lambda^n=c$, which is a contradiction with (55). If $\theta\notin\mathbb{Q}$, it is well known that $\{\lambda^n\}_{n\in\mathbb{Z}_+}$ is equidistributed in $[0,1)$ (see for example [19, Theorem 1.4]). Thus, for any constant $c\in\mathbb{C}$ there is no subset J of $\mathbb{N}$ with $D(J)=0$ such that $\lim_{n\to\infty,n\notin J}\lambda^n=c$, which is also a contradiction with (55). Thus, $\lambda=1$, and hence $h\circ T=f$, Q-a.s., which implies that f is constant, Q-a.s., as Q is ergodic. Therefore, Q is weakly mixing.

The last statement is a direct corollary of Corollary 5.8. □

REMARK 5.20. Mirroring Remark 4.8, in the case that P is an invariant probability, all the three results in the above theorem are exactly the same as the results in classical ergodic theory without extra condition. So our results in the case that P is an invariant probability are sharp and hold true for a class of non-invariant probabilities.

6. Further applications to invariant probabilities. In this section, we see examples of ergodic and weakly mixing capacity preserving systems, and provide applications for invariant probabilities.

6.1. *Distortions of invariant probabilities and their applications in characterization of weakly mixing and periodic probabilities.* The first example shows that each ergodic (resp. weakly mixing) probability preserving system can give rise to a non-trivial ergodic (resp. weakly mixing) upper probability. Then the ergodic theory of upper probabilities leads to some new results on the classical ergodic theory for probability preserving systems.

EXAMPLE 4. Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system. Given a strictly increasing continuous concave function $f:[0,1]\to[0,1]$ with $f(0)=0$ and $f(1)=1$, define a concave distortion of P with respect to f by $V_f(A):=f(P(A))$ for any $A\in\mathcal{F}$ (see [8] for more properties about this type of capacities). Note that for any $A,B\in\mathcal{F}$, $P(A\cap B)\le P(A),P(B)\le P(A\cup B)$. By the property of concave functions,

$$f(P(A\cap B))+f(P(A\cup B))\le f(P(A))+f(P(B)),$$

which shows that V_f is concave capacity. It is easy to see that V_f is also T-invariant. Meanwhile, it is continuous by the continuity of the function f and the probability P. By the choice of f, we have that $f(P(A))\ge P(A)$ for any $A\in\mathcal{F}$, hence $P\in\operatorname{core}(V_f)$. If the system $(\Omega,\mathcal{F},P,T)$ is ergodic (resp. weakly mixing), it is easy to see from the definition of V_f that $V_f|_{\mathcal{I}}=P|_{\mathcal{I}}$. So by Theorem 3.2 (resp. Lemma 5.7), we have that V_f is also ergodic (resp. weakly mixing).

Thus, all results can apply to the capacity preserving system $(\Omega,\mathcal{F},V_f,T)$. Some of them are new and striking to the classical theory. We can see some of them in the following.

In particular, we suppose that P is an ergodic probability and consider $f(x)=\sqrt{x}$. Applying (iv) of Theorem 4.10 on V_f, we have

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)\ge P^{1/2}(B)P(C)\text{ for any }B,C\in\mathcal{F}.$$

Meanwhile, by the concavity of the function $f(x)=x^{1/2}$, $x\ge 0$, one has that for any $B,C\in\mathcal{F}$,

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)\le\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)\right)^{1/2}=P^{1/2}(B)P^{1/2}(C).$$

Thus, for any $B, C \in \mathcal{F}$,

$$
\begin{aligned}
P^{1/2}(B)P(C) &\leq \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} P^{1/2}(B\cap T^{-i}C) \\
&\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} P^{1/2}(B\cap T^{-i}C) \leq P^{1/2}(B)P^{1/2}(C).
\end{aligned} \tag{56}
$$

REMARK 6.1. Under the assumption that P is ergodic, the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)$ exists, but the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)$ may not exist. In fact, for a general sequence $\{a_i\}_{i\in\mathbb{N}}\subset[0,1]$, even if the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i$ exists, we may not be able to obtain the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i^{1/2}$ exists. For example, we consider the sequence $\{a_i\}_{i\in\mathbb{N}}$ defined via $a_i=1/4$ if $i\in(2^{2k-1},2^{2k}]$; $a_i=1/2$ if $i\in(2^{2k},2^{2k+1}]$ is even; $a_i=0$ if $i\in(2^{2k},2^{2k+1}]$ is odd for each $k\in\mathbb{N}$. In the following, we consider $\frac{1}{n}\sum_{i=1}^{n}a_i$ and $\frac{1}{n}\sum_{i=1}^{n}a_i^{1/2}$ instead for convenience of notation in this special example. By computation, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i=1/4$, but the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_i^{1/2}$ does not exist, as

$$
\begin{aligned}
\lim_{k\to\infty}\frac{1}{2^{2k}}\sum_{i=1}^{2^{2k}}a_i^{1/2} &= \lim_{k\to\infty}\frac{1}{2^{2k}}\left(\sum_{i=1}^{k}\frac{1}{2}(2^{2i}-2^{2i-1})+\frac{1}{2}\sum_{i=1}^{k}\frac{1}{\sqrt{2}}(2^{2i-1}-2^{2i-2})\right)\\
&=\frac{1}{3}+\frac{1}{6\sqrt{2}},
\end{aligned}
$$

but

$$
\begin{aligned}
\lim_{k\to\infty}\frac{1}{2^{2k+1}}\sum_{i=1}^{2^{2k+1}}a_i^{1/2} &= \lim_{k\to\infty}\frac{1}{2^{2k+1}}\left(\sum_{i=1}^{k}\frac{1}{2}(2^{2i}-2^{2i-1})+\frac{1}{2}\sum_{i=0}^{k}\frac{1}{\sqrt{2}}(2^{2i+1}-2^{2i})\right)\\
&=\frac{1}{6}+\frac{1}{3\sqrt{2}}.
\end{aligned}
$$

Thus, the $\liminf$ and the $\limsup$ in (56) may not be equal.

The following two propositions show that the limit of Cesàro summation $\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)$, for $n\in\mathbb{N}, B, C\in\mathcal{F}$, is closely related to the complexity of the dynamical system. Firstly, we characterize the weak mixing probability preserving systems by the limit is equal to the upper bound in (56).

PROPOSITION 6.2. Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system. Then the following three statements are equivalent:

(i) P is weakly mixing;

(ii) for any $r>0$, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^r(B\cap T^{-i}C)=P^r(B)P^r(C)$ for any $B,C\in\mathcal{F}$;

(iii) for any $B,C\in\mathcal{F}$, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)=P(B)P(C)$, and there exists $r=r_{B,C}\in(0,1/2]$ such that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^r(B\cap T^{-i}C)=P^r(B)P^r(C)$.

In particular, P is weakly mixing if and only if P is ergodic and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)=P^{1/2}(B)P^{1/2}(C)$ for any $B,C\in\mathcal{F}$.

PROOF. (i) $\Rightarrow$ (ii). Given any $B,C\in\mathcal{F}$, it is well known that (see [46, Theorem 1.21] for example) there exists a subset $J=J_{B,C}\subset\mathbb{N}$ with nature density $D(J)=0$ such that

$$\lim_{n\notin J, n\to\infty} P(B\cap T^{-n}C)=P(B)P(C),$$

which implies that for any $r>0$,

$$\lim_{n\notin J, n\to\infty} P^r(B\cap T^{-n}C)=P^r(B)P^r(C),$$

which implies (ii).

(ii) $\Rightarrow$ (iii). This is trivial.

(iii) $\Rightarrow$ (i). Given any $B,C\in\mathcal{F}$, suppose that there exists $r=r_{B,C}\in(0,1/2]$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^r(B\cap T^{-i}C)=P^r(B)P^r(C).$$

Then we have the following claim.

CLAIM 1. $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)=P^{1/2}(B)P^{1/2}(C).$

PROOF OF THE CLAIM. By Hölder inequality and the first statement of (iii), we have

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)\le\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)\right)^{1/2}=P^{1/2}(B)P^{1/2}(C).$$

Using Hölder inequality again, we have that

$$\begin{aligned}P^r(B)P^r(C)&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^r(B\cap T^{-i}C)\\&\le\liminf_{n\to\infty}\frac{1}{n}\left(\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)\right)^{2r}\cdot n^{1-2r},\end{aligned}$$

which implies that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)\ge P^{1/2}(B)P^{1/2}(C).$$

Now we finish the proof of the claim. □

The above claim, together with the first statement of (iii), implies that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(P^{1/2}(B\cap T^{-i}C)-P^{1/2}(B)P^{1/2}(C))^2\\=&\lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=0}^{n-1}P(B\cap T^{-i}C)-2\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)P^{1/2}(B)P^{1/2}(C)+P(B)P(C)\right]\\=&0.\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(P(B\cap T^{-i}C)-P(B)P(C))^2$$
$$\leq 4\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(P^{1/2}(B\cap T^{-i}C)-P^{1/2}(B)P^{1/2}(C))^2=0.$$

As $B,C\in\mathcal{F}$ are arbitrary, by [46, Theorem 1.17], we finish the proof. □

Conversely, if the Cesàro summation above-mentioned converges to the lower bound in (56), then the system is simple. That is,

PROPOSITION 6.3. Let $(\Omega,\mathcal{F},P,T)$ be an ergodic probability preserving system, where $(\Omega,\mathcal{F})$ is a standard measurable space. Then the following two statements are equivalent:

(i) there exists $B\in\mathcal{F}$ with $P(B)>0$ such that for any $C\in\mathcal{F}$ with $C\subset B$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}C)=P^{1/2}(B)P(C); \tag{57}$$

(ii) P is a periodic probability, i.e., there exist $r\in\mathbb{N}$ and distinct points $\omega_1,\dots,\omega_r\in\Omega$ such that $P(\{\omega_i\})=\frac{1}{r}$, $i=1,2,\dots,r$.

In this case, $r=\frac{1}{P(B)}\in\mathbb{N}$.

PROOF. (ii) ⇒ (i). Let $B=\{\omega_1\}$. Then we only need to check (57). Note that $P^{1/2}(B\cap T^{-i}B)=1/\sqrt{r}$, if $i=kr$ for some $k\in\mathbb{N}$, otherwise $P^{1/2}(B\cap T^{-i}B)=0$. Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^{1/2}(B\cap T^{-i}B)=\frac{1}{r}\cdot\frac{1}{\sqrt{r}}=P^{1/2}(B)P(B).$$

(i) ⇒ (ii). Let B be as in assumption (i). Fix any $C\in\mathcal{F}$ with $C\subset B$. Let

$$a_i=\left(\frac{P(B\cap T^{-i}C)}{P(B)}\right)^{1/2}\quad\text{for each } i\in\mathbb{Z}_+.$$

Then $0\leq a_i\leq 1$ for any $i\in\mathbb{Z}_+$. From the assumption (57) and P being ergodic,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}a_i=P(C)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}a_i^2.$$

So $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}a_i(1-a_i)=0$. Thus, by Theorem 1.20 in [46], there exists a subset $J\subset\mathbb{Z}_+$ with natural density $D(J)=1$ such that

$$\lim_{n\in J,n\to\infty}a_n(1-a_n)=0.$$

If $P(C)>0$ then there exists a subset $J_1\subset J$ with positive natural density such that $\lim\limits_{n\in J_1,n\to\infty}a_n=1$, that is,

$$\lim_{n\in J_1,n\to\infty}P(B\cap T^{-n}C)=P(B).$$

This implies $P(C)=P(B)$. If this is not true, note that

$$P(B\cap T^{-n}C)\le P(C)<P(B).$$

This is a contradiction. Thus, $P(C)=P(B)$. As $C\in\mathcal{F}$ with $C\subset B$ is arbitrary, it follows that B is an atom. Since $(\Omega,\mathcal{F})$ is standard, it follows that each atom is a singleton, denoted by $B=\{\omega\}$ for some $\omega\in\Omega$. Since $P(\cup_{i=0}^{\infty}T^i\{\omega\})\le 1$, and T is measure-preserving, so there exists a smallest integer $r\in\mathbb{N}$ such that $T^r\omega=\omega$. Since P is ergodic, and $\{\omega,T\omega,\dots,T^{r-1}\omega\}$ is an invariant set, it follows that

$$P(\{\omega,T\omega,\dots,T^{r-1}\omega\})=1.$$

Let $\omega_i=T^{i-1}\omega$ for $i=1,2,\dots,r$. Then $P(\omega_i)=\frac{1}{r}=P(\omega)=P(B)$ for $i=1,2,\dots,r$. In particular, $r=\frac{1}{P(B)}$. □

More generally, we have the following. The proof is similar to that of (56) using the concavity of the function f, so it is omitted.

PROPOSITION 6.4. Let $(\Omega,\mathcal{F},P,T)$ be a probability preserving system and $f:[0,1]\to[0,1]$ be an increasing concave continuous function with $f(0)=0$ and $f(1)=1$. Then for any $B,C\in\mathcal{F}$,

$$\begin{aligned}f(P(B))P(C)&\le \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(P(B\cap T^{-i}C))\\&\le \limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(P(B\cap T^{-i}C))\le f(P(B)P(C)).\end{aligned}$$

6.2. *Counter example of ergodic capacity preserving system in number theory without Birkhoff's law of large numbers.* In this subsection, we will prove that the law of large numbers does not hold for any invariant capacity on $(\mathbb{Z},2^{\mathbb{Z}})$ with respect to $T:\mathbb{Z}\to\mathbb{Z},n\to n+1$. For convenience, we denote by $[m,n]=\{m,m+1,\dots,n\}$, for any $m<n\in\mathbb{Z}$.

Given any capacity preserving system $(\mathbb{Z},2^{\mathbb{Z}},\mu,T)$, we prove that Birkhoff's ergodic theorem does not hold for μ. By contradiction, if it holds, then for any $A\subseteq\mathbb{Z}$, there exist $B\subset\mathbb{Z}$ with $\mu(B^c)=0$ (i.e., $B\neq\emptyset$) and $c\ge 0$ such that for any $m\in B$,

$$\lim_{n\to\infty}\frac{1}{2n+1}|A\cap[m-n,m+n]|=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_A(T^im)=c.$$

Thus,

$$\lim_{n\to\infty}\frac{1}{2n+1}|A\cap[-n,n]|=c.$$

This means that the natural density of any subset A of $\mathbb{Z}$ exists. However, it is not true, for example, that the set $A=\bigcup_{n=0}^{\infty}[2^{2n},2^{2n+1}]$ does not have a natural density. This can be seen as

$$\limsup_{n\to\infty}\frac{|A\cap[-n,n]|}{2n+1}\ge\lim_{n\to\infty}\frac{|A\cap[-2^{2n+1},2^{2n+1}]|}{2^{2n+2}+1}>\lim_{n\to\infty}\frac{2^{2n+1}-2^{2n}}{2^{2n+2}+1}=\frac{1}{4},$$

but

$$\liminf_{n\to\infty}\frac{|A\cap[-n,n]|}{2n+1}\le\lim_{n\to\infty}\frac{|A\cap[-2^{2n},2^{2n}]|}{2^{2n+1}+1}\le\lim_{n\to\infty}\frac{2^{2n-1}}{2^{2n+1}+1}=\frac{1}{4}.$$

REMARK 6.5. As a corollary of Theorem 3.3 and the above consequence, there is no T-ergodic upper probability on $(\mathbb{Z}, 2^{\mathbb{Z}})$.

Now we study some well-known concrete examples of capacities on $(\mathbb{Z}, 2^{\mathbb{Z}})$, but they are in fact not upper probabilities. Firstly, we prove $\bar{d}$ in Example 1, introduced in the Introduction, is an ergodic subadditive capacity.

EXAMPLE 5. We recall the capacity preserving system $(\mathbb{Z}, 2^{\mathbb{Z}}, \bar{d}, T)$ defined in Example 1. Now we prove that $\bar{d}$ is ergodic with respect to T. Note that $T^{-1}A = \{n \in \mathbb{Z} : n+1 \in A\}$. There are only two possible sets $A = \emptyset$ and $A = \mathbb{Z}$ satisfying $T^{-1}A = A$. In particular, $\bar{d}(A) = 0$ or $\bar{d}(A) = 1$ and $\bar{d}(A^c) = 0$, proving the ergodicity of $\bar{d}$. But consider $A_k = [k, \infty)$ for all $k \in \mathbb{N}$. Then $A_k \downarrow \emptyset$, but for any fixed $k \in \mathbb{N}$, $\bar{d}(A_k) = 1/2$. Thus, $\bar{d}$ is not continuous from above. Meanwhile, we consider $B_k = [-k, k]$ for each $k \in \mathbb{N}$. Then $A_k \uparrow \mathbb{Z}$, but for any fixed $k \in \mathbb{N}$, $\bar{d}(A_k) = 0$. Thus, $\bar{d}$ is not continuous from below. So the subadditive capacity $\bar{d}$ is not an upper probability.

Recall that a continuous concave capacity must be an upper probability, and hence if it is ergodic then Birkhoff's ergodic theorem holds. The following example shows that there exists an ergodic concave capacity which is continuous from below such that Birkhoff's ergodic theorem does not hold.

EXAMPLE 6. Let $T : \mathbb{Z} \to \mathbb{Z}$, $x \mapsto x+1$, and $2^{\mathbb{Z}}$ be the family consisting of all subsets of $\mathbb{Z}$. Define

$$\mu = \max_{n \in \mathbb{Z}} \delta_n,$$

where δ_n is the Dirac measure for $n \in \mathbb{Z}$. Then it is easy to check that μ is an invariant concave capacity continuous from below, and in particular $(\mathbb{Z}, 2^{\mathbb{Z}}, \mu, T)$ is a capacity preserving system. The ergodicity of μ can be obtained by the same argument in Example 5. However, by the same argument as in Example 5, μ is not continuous from above, and hence μ is also not an upper probability.

6.3. *Examples: ergodicity but not weak mixing.* We recall that a weakly mixing subadditive capacity must be ergodic. In the case for probabilities, there are many examples showing that an ergodic probability may not be weakly mixing (for example, irrational rotations on the torus). Now we provide some examples for more general capacities.

EXAMPLE 7. Let $(\mathbb{Z}, 2^{\mathbb{Z}}, \mu, T)$ be the ergodic system in Example 6. We claim that μ is not weakly mixing. To see this, we consider the measurable function $f(n) = \lambda^n$ on $\mathbb{Z}$, where $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. Then $f(Tn) = f(n+1) = \lambda f(n)$ for each $n \in \mathbb{Z}$. Since f is not constant, it follows that μ is not weakly mixing.

EXAMPLE 8. Given $N \in \mathbb{N}$, let $\big(\Omega_i, \mathcal{F}_i, P_i\big)$ for $i = 1, 2, \ldots, N$ be probability spaces, and let

$$\pi_i : (\Omega_i, \mathcal{F}_i) \to \big(\Omega_{i+1(\bmod N)}, \mathcal{F}_{i+1(\bmod N)}\big)$$

be invertible measurable maps satisfying

$$P_i \circ \pi_i^{-1} = P_{i+1(\bmod N)}.$$

We then define a measurable transformation $S : \Omega_N \to \Omega_N$ so that $\big(\Omega_N, \mathcal{F}_N, P_N, S\big)$ is a weakly mixing probability preserving system (see [38] for examples).

Let $\Omega = \bigcup_{n=1}^{N} \Omega_i$. We equip Ω with the σ-algebra

$$\mathcal{F} \;=\; \Big\{ A \subset \Omega : A \cap \Omega_i \in \mathcal{F}_i \text{ for each } i = 1, \dots, N \Big\},$$

and define $T : \Omega \to \Omega$ by

$$T(\omega) = \begin{cases} \pi_i(\omega), & \omega \in \Omega_i,\ i = 1, 2, \dots, N-1, \\ \pi_N(S(\omega)), & \omega \in \Omega_N. \end{cases}$$

For each $i = 1, 2, \dots, N$, let

$$\bar{P}_i(A) := P_i\left(A \cap \Omega_i\right), \qquad \text{for any } A \in \mathcal{F},$$

and define the upper probability

$$V = \max_{1 \le i \le N} \bar{P}_i.$$

Note that for any $A \in \mathcal{F}$,

$$T^{-1}A \cap \Omega_i = \begin{cases} \pi_i^{-1}(A \cap \Omega_{i+1}), & i = 1, 2, \dots, N-1, \\ S^{-1}\pi_N^{-1}(A \cap \Omega_1), & i = N. \end{cases} \tag{58}$$

Thus, V is T-invariant. Indeed, for any $A \in \mathcal{F}$,

$$\begin{aligned} V(T^{-1}A) &= \max_{1 \le i \le N} P_i(T^{-1}A \cap \Omega_i) \\ &= \max\{ \max_{1 \le i \le N-1} P_i(\pi_i^{-1}(A \cap \Omega_{i+1})), P_N(S^{-1}\pi_N^{-1}(A \cap \Omega_1)) \} \\ &= \max\{ \max_{1 \le i \le N-1} P_{i+1}(A \cap \Omega_{i+1}), P_1(A \cap \Omega_1) \} \\ &= V(A). \end{aligned}$$

Next, we prove that V is ergodic. Using (58) again, we know that for $A \in \mathcal{F}$ with $T^{-1}A = A$,

$$A \cap \Omega_i = \begin{cases} \pi_i^{-1}(A \cap \Omega_{i+1}), & i = 1, 2, \dots, N-1, \\ S^{-1}\pi_N^{-1}(A \cap \Omega_1), & i = N. \end{cases} \tag{59}$$

Since each π_i is invertible, it follows from (59) that

$$A \cap \Omega_N = S^{-1}\big(\pi_N^{-1}(A \cap \Omega_1)\big) = S^{-1}\big(\pi_N^{-1}\pi_1^{-1} \cdots \pi_{N-1}^{-1}(A \cap \Omega_N)\big) = S^{-1}(A \cap \Omega_N).$$

Because $\big(\Omega_N, \mathcal{F}_N, P_N, S\big)$ is weakly mixing, any S-invariant set in $\mathcal{F}_N$ has P_N-measure 0 or 1. Hence

$$\bar{P}_N(A) = P_N(A \cap \Omega_N) \;\in\; \{0, 1\}.$$

On the other hand, using $P_i \circ \pi_i^{-1} = P_{i+1}$, for $i = 1, 2, \dots, N-1$ together with (59), one shows

$$P_1(A \cap \Omega_1) = P_1\big(\pi_1^{-1}(A \cap \Omega_2)\big) = P_2(A \cap \Omega_2) = \cdots = P_N(A \cap \Omega_N).$$

Therefore, one of the following holds:

$$\bar{P}_i(A) = 0 \quad \text{for all } i = 1, 2, \dots, N,$$

or

$$\bar{P}_i(A^c) = 0 \quad \text{for all } i = 1, 2, \dots, N.$$

It follows that either $V(A) = 0$ or $V(A^c) = 0$.

Finally, we prove it is not weakly mixing. Let

$$f(\omega) = e^{2\pi i \frac{\ell}{N}} \text{ for } \omega \in X_\ell,\ \ell \in \{1, 2, \dots, N\}.$$

Then $f \circ T = e^{2\pi i \frac{\ell}{N}} f$, and f is not constant, V-a.s., which shows that V is not weakly mixing.

7. Subadditive ergodic theorem for capacities. In the last section of this paper, we apply the common conditional expectation and invariant skeleton to study the subadditive ergodic theorem on upper probability spaces, which provides a way to understand the long-term behaviour of subadditive functions. Subadditive functions have applications in a variety of fields, including probability theory, information theory, and statistical physics.

7.1. *Proof of subadditive ergodic theorem for invariant upper probabilities.* We recall that a sequence of $\mathcal{F}$-measurable functions $\{f_n\}_{n\in\mathbb{N}}$ on the capacity space $(\Omega,\mathcal{F},\mu)$ is said to be subadditive (resp. superadditive) if for each $k,n\in\mathbb{N}$, $f_{n+k}\le f_n+f_k\circ T^n$ (resp. $f_{n+k}\ge f_n+f_k\circ T^n$), μ-a.s. If a sequence is subadditive and superadditive, then it is said to be additive. Given an $\mathcal{F}$-measurable function g, let $f_n=\sum_{i=0}^{n-1} g\circ T^i$ for each $n\in\mathbb{N}$. Then $f_{n+k}=f_n+f_k\circ T^n$ for each $k,n\in\mathbb{N}$, and hence it is additive. However, for example, $\{|f_n|\}_{n\in\mathbb{N}}$ is only subadditive. Thus, subadditive ergodic theorem can be viewed as an extension of Birkhoff's ergodic theorem. We remark that, as the proofs of subadditive and superadditive sequences are similar, we only state and prove the results for subadditive sequences.

In the following, a standard setup is a capacity preserving system $(\Omega,\mathcal{F},V,T)$, where $(\Omega,\mathcal{F})$ is a standard measurable space, $T:\Omega\to\Omega$ is a measurable transformation, and V is a T-invariant upper probability.

THEOREM 7.1. *Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-measurable functions satisfying the following conditions:*

(i) *there exists $\lambda>0$ such that $-\lambda n\le f_n(\omega)\le\lambda n$ for any $n\in\mathbb{N}$, and $\omega\in\Omega$;*

(ii) *for each $k,n\in\mathbb{N}$, $f_{n+k}\le f_n+f_k\circ T^n$, V-a.s.*

Then there exists $f^*\in B(\Omega,\mathcal{I})$ such that

$$\lim_{n\to\infty}\frac{1}{n}f_n(\omega)=f^*(\omega)\text{ for }V\text{-a.s. }\omega\in\Omega.$$

If, in addition, V is ergodic, then f^* is a constant V-a.s.

PROOF. Define

$$S_n=\frac{1}{n}f_n\text{ for each }n\in\mathbb{N}.$$

By Assumption (i), one has $S_n\in B(\Omega,\mathcal{F})$ for each $n\in\mathbb{N}$. By Lemma 2.16, for each $n\in\mathbb{N}$, there exists $\hat{S}_n\in B(\Omega,\mathcal{I})$ such that for any $\eta\in\mathcal{M}(T)$,

$$\hat{S}_n=\mathbb{E}_\eta(S_n\,|\,\mathcal{I}),\quad \eta\text{-a.s.}$$

Let $f^*=\inf_{n\in\mathbb{N}}\hat{S}_n$ and $\Omega^*=\{\omega\in\Omega:\lim_{n\to\infty}S_n(\omega)=f^*(\omega)\}$. Then by Theorem 2.6, one has $\eta(\Omega^*)=1$ for any $\eta\in\mathcal{M}(T)\cap\mathrm{core}(V)$. In particular, for any $P\in\mathrm{core}(V)$, its invariant skeleton satisfies that $\hat{P}((\Omega^*)^c)=0$. By Lemma 2.13 (ii), we deduce that $V((\Omega^*)^c)=0$.

Suppose that V is ergodic. Applying Lemma 2.11 on the T-invariant function f^*, we have f^* is constant, V-a.s. □

REMARK 7.2. (i) Cerreia-Vioglio, Maccheroni and Marinacci [9] considered a subadditive ergodic theorem on the upper probability space $(\Omega,\mathcal{F},V)$. However, they imposed the additional condition that there exists a compact subset Λ of $\mathcal{M}(T)$ such that $V=\max_{P\in\Lambda}P$. We recall Example 2, in which V does not satisfy this condition.

(ii) When V is ergodic, $\mathrm{core}(V)\cap\mathcal{M}(T)$ has only one element. So we do not need to use the common conditional expectation. Therefore, in this case, the subadditive ergodic theorem holds for general probability spaces, but not necessarily standard ones.

Continuing with the viewpoint of Theorem 4.7, the following result provides the subadditive ergodic theorem for a class of non-invariant probabilities.

THEOREM 7.3. *Let $(\Omega,\mathcal{F},P)$ be a probability space (not necessarily standard), and $T:\Omega\to\Omega$ be an invertible measurable transformation. Suppose that the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P\circ T^i$ exists, denoted by Q. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-measurable functions satisfying the following conditions:*

(i) *there exists $\lambda>0$ such that $-\lambda n\le f_n(\omega)\le\lambda n$ for any $n\in\mathbb{N}$, and $\omega\in\Omega$;*

(ii) *for each $k,n\in\mathbb{N}$, $f_{n+k}\le f_n+f_k\circ T^n$.*

If Q is ergodic, then there exists a constant $c\in\mathbb{R}$ such that

$$\lim_{n\to\infty}\frac{1}{n}f_n=c,\ P\text{-a.s.}$$

7.2. *The multiplicative ergodic theorem for capacities.* In this subsection, we apply Theorem 7.1 to obtain an extension of Furstenberg-Kesten theorem ([25], see also [46, Corollary 10.1.1] or [2, Theorem 3.3.3]) to an upper probability space and further use it to prove the multiplicative ergodic theorem [34] on upper probability spaces. Let us begin with the notation. Denote by $\mathbb{R}^{d\times d}$ the space of all linear operators on $\mathbb{R}^d$. By choosing a basis of $\mathbb{R}^d$, we can view $\mathbb{R}^{d\times d}$ as the space of all $d\times d$ matrices. For any $A\in\mathbb{R}^{d\times d}$, denote by A^* its transpose matrix.

Let $(\Omega,\mathcal{F},\mu,T)$ be a capacity preserving system and $L:\Omega\to\mathbb{R}^{d\times d}$ be a measurable map such that for each $\omega\in\Omega$, $L(\omega)$ is a $d\times d$ matrix. Define

$$\Phi(n,\omega):=L(T^{n-1}\omega)L(T^{n-2}\omega)\cdots L(\omega),\ \text{for } n\ge 1 \text{ and } \omega\in\Omega, \tag{60}$$

where $\Phi(1,\omega):=L(\omega)\in\mathbb{R}^{d\times d}$ is the generator of Φ. Let $\wedge^k\mathbb{R}^d$, $1\le k\le d$, be the k-fold exterior power of $\mathbb{R}^d$ (see [43, Chapter V] for more details). The cocycle property

$$\Phi(n+m,\omega)=\Phi(m,T^n\omega)\,\Phi(n,\omega)$$

lifts to $\wedge^k\mathbb{R}^d$, $1\le k\le d$ (see [2, Lemma 3.2.6])

$$\wedge^k\Phi(n+m,\omega)=(\wedge^k\Phi)(m,T^n\omega)(\wedge^k\Phi)(n,\omega). \tag{61}$$

In the following, we always suppose that $\|\cdot\|$ is a matrix norm on $\mathbb{R}^{d\times d}$, i.e., a norm on $\mathbb{R}^{d\times d}$ with additional property $\|AB\|\le\|A\|\|B\|$.

Now we prove the Furstenberg-Kesten theorem on upper probability spaces.

THEOREM 7.4 (Furstenberg-Kesten theorem for upper probabilities). *Let Φ be defined, by (60), on the capacity preserving system $(\Omega,\mathcal{F},V,T)$, where V is an upper probability. If the generator $L:\Omega\to\mathbb{R}^{d\times d}$ of Φ is a measurable function such that $\log\|L(\omega)\|\in B(\Omega,\mathcal{F})$, then*

(i) *for each $k=1,\dots,d$ the sequence $\{f_n^{(k)}(\omega)\}_{n\in\mathbb{N}}$ defined by*

$$f_n^{(k)}(\omega):=\log\left\|\wedge^k\Phi(n,\omega)\right\|\ \text{for each } n\in\mathbb{N}$$

is subadditive;

(ii) *there exist* $\gamma^{(k)} \in B(\Omega, \mathcal{F})$ *for* $k = 1, 2, \dots, d$ *such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| \wedge^k \Phi(n, \cdot) \right\| = \gamma^{(k)},\ V\text{-a.s.} \tag{62}$$

PROOF. Fix $k \in \{1, 2, \dots, d\}$. Since $\|(\wedge^k L_1) \cdot (\wedge^k L_2)\| \leq \| \wedge^k L_1\| \cdot \| \wedge^k L_2\|$ for any $L_1, L_2 \in \mathbb{R}^{d\times d}$ (see [2, Lemma 3.2.6 (vi)]) it follows from (61) that $\{f_n^{(k)}(\omega)\}$ is subadditive. Meanwhile, since $\log \|L(\omega)\| \in B(\Omega, \mathcal{F})$, there exists $M > 0$ such that

$$|\log \|L(\omega)\|| \leq M \text{ for any } \omega \in \Omega,$$

which shows that for each $n \in \mathbb{N}$,

$$|f_n^{(k)}(\omega)| \leq knM \text{ for any } n \in \mathbb{N}. \tag{63}$$

The proof of (ii) is completed by applying Theorem 7.1 on the sequence $\{f_n^{(k)}(\omega)\}_{n\in\mathbb{N}}$ for each $k = 1, 2, \dots, d$. □

THEOREM 7.5 (Multiplicative ergodic theorem for upper probabilities). *Under the same conditions as in Theorem 7.4, there exists* $\tilde{\Omega} \in \mathcal{I}$ *with* $V(\tilde{\Omega}^c) = 0$ *such that for any* $\omega \in \tilde{\Omega}$,

(i) *The limit* $\lim_{n\to\infty}(\Phi(n,\omega)^*\Phi(n,\omega))^{1/2n} := \Psi(\omega)$ *exists.*

(ii) *Let* $e^{\lambda_{p(\omega)}(\omega)} < \dots < e^{\lambda_1(\omega)}$ *be the different eigenvalues of* $\Psi(\omega)$, *where* $p(\omega)$ *is the number of the different eigenvalues of* $\Psi(\omega)$, *and let* $U_{p(\omega)}(\omega), \dots, U_1(\omega)$ *be the corresponding eigenspaces with multiplicities* $d_i(\omega) := \dim U_i(\omega)$. *Then*

$$p(T\omega) = p(\omega), \lambda_i(T\omega) = \lambda_i(\omega), \text{ and } d_i(T\omega) = d_i(\omega) \text{ for all } i \in \{1, 2, \dots, p(\omega)\}.$$

(iii) *Put* $V_{p(\omega)+1}(\omega) := \{0\}$ *and for* $i = 1, 2, \dots, p(\omega)$

$$V_i(\omega) := U_{p(\omega)}(\omega) \oplus \dots \oplus U_i(\omega)$$

such that

$$V_{p(\omega)}(\omega) \subset \dots \subset V_i(\omega) \subset \dots \subset V_1(\omega) = \mathbb{R}^d.$$

Then for each $x \in \mathbb{R}^d \backslash \{0\}$ *the Lyapenov exponent*

$$\lambda(\omega, x) := \lim_{n\to\infty} \frac{1}{n} \log \|\Phi(n, \omega)x\|$$

exists as a limit, and

$$\lambda(\omega, x) = \lambda_i(\omega) \Leftrightarrow x \in V_i(\omega) \backslash V_{i+1}(\omega)$$

equivalently

$$V_i(\omega) = \{x \in \mathbb{R}^d : \lambda(\omega, x) \leq \lambda_i(\omega)\}.$$

(iv) *For all* $x \in \mathbb{R}^d \setminus \{0\}$

$$\lambda(T\omega, L(\omega)x) = \lambda(\omega, x),$$

whence

$$L(\omega)V_i(\omega) \subset V_i(T\omega) \text{ for all } i \in \{1, \dots, p(\omega)\}.$$

(v) *If* V *is ergodic, then the function* p *is constant on* $\tilde{\Omega}$, *and the functions* λ_i *and* d_i *are constant on* $\{\omega \in \tilde{\Omega} : p(\omega) \geq i\}$, $i = 1, 2, \dots, d$.

PROOF. Under the assumption that $\log\|L(\omega)\| \in B(\Omega,\mathcal{F})$, it is easy to see that for any $\omega\in\Omega$, $\limsup_{n\to\infty}\frac{1}{n}\log\|L(T^n\omega)\|\le 0$. Moreover, Φ_n defined as in (60) also satisfies (62) on a subset Ω_1 of Ω with $V(\Omega_1^c)=0$. Thus, (i), (ii) and (iii) are true on Ω_1 from Oseledec's deterministic multiplicative ergodic theorem (see [2, Proposition 3.4.2]).

Now we check that (iv) holds for any $\omega\in\Omega_1$. Note that $\Phi(n,T\omega)L(\omega)=\Phi(n+1,\omega)$ for any $n\in\mathbb{N}$ and $\omega\in\Omega$. Thus, for any $\omega\in\Omega_1$, by (iii), one has

$$\lambda(T\omega,L(\omega)x)=\lim_{n\to\infty}\frac{1}{n}\log\|\Phi(n,T\omega)L(\omega)x\|=\lim_{n\to\infty}\frac{1}{n}\log\|\Phi(n+1,\omega)x\|=\lambda(\omega,x).$$

Using (iii) again, for any $x\in V_i(\omega)$, $\lambda(\omega,x)\le\lambda_i(\omega)$, which shows that

$$\lambda(T\omega,L(\omega)x)=\lambda(\omega,x)\le\lambda_i(\omega)\overset{(ii)}{=}\lambda_i(T\omega).$$

Therefore, $L(\omega)x\in V_i(T\omega)$, which by the arbitrariness of $x\in V_i(\omega)$, implies that $L(\omega)V_i(\omega)\subset V_i(T\omega)$.

Since p is T-invariant on Ω_1 and $V(\Omega_1^c)=0$, it follows from Lemma 2.11 that p is constant on some measurable set $\Omega_2\subset\Omega_1$ with $V(\Omega_2^c)=0$. Using Lemma 2.11 again, there exists $\widetilde{\Omega}\subset\Omega_2$ with $V(\widetilde{\Omega}^c)=0$ such that the functions λ_i and d_i are constant on $\{\omega\in\tilde{\Omega}:p(\omega)\ge i\}$, $i=1,2,\dots,d$. The proof is complete. □

As a direct corollary of Theorems 7.3 and 7.5, one has that multiplicative ergodic theorem holds for a class of non-invariant probabilities.

THEOREM 7.6. *Let $(\Omega,\mathcal{F},P)$ be a probability space, and $T:\Omega\to\Omega$ be an invertible measurable transformation. Suppose that the limit $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P\circ T^i$ exists, denoted by Q. Then (i)-(v) in Theorem 7.5 holds, P-a.s.*

Acknowledgments. CL gratefully acknowledges Durham University and the University of Science and Technology of China; the bulk of this work was completed during his visit to Durham University and his graduate study at the University of Science and Technology of China.

The authors are very grateful to the anonymous referee for their constructive comments and suggestions, which have significantly improved the clarity and quality of this paper.

Funding. WH was supported by NNSF of China (12090012, 12031019, 12090010).

CL was supported by CSC of China (No. 202206340035), the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation under Grant Number BX20250067, and the China Postdoctoral Science Foundation under Grant Number 2025M773074.

HZ was supported by the Royal Society Newton Fund (ref. NIF \R1\221003) and the EPSRC (ref. EP/S005293/2).

REFERENCES

[1] ALIPRANTIS, C. D., AND BORDER, K. C. *Infinite-dimensional analysis. A hitchhiker's guide.* Second edition. Springer Berlin, 1999.

[2] ARNOLD, L. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[3] BARTL, D. Conditional nonlinear expectations *Stoch. Proc. Appl.* vol. 130(2), (2020), 785-805.

[4] BILLOT, A. From fuzzy set theory to non-additive probabilities: How have economists reacted? *Fuzzy Sets and Systems 49*, 1 (1992), 75–90.

[5] BIRKHOFF, G. D. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. U.S.A. 17*, 12 (1931), 656–660.

[6] BOURGAIN, J. Pointwise ergodic theorems for arithmetic sets. *Inst. Hautes Études Sci. Publ. Math.*, 69 (1989), 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[7] BUCZOLICH, Z., AND MAULDIN, R. D. Divergent square averages. *Ann. of Math. (2) 171*, 3 (2010), 1479–1530.

[8] CARLIER, G., AND DANA, R. A. Core of convex distortions of a probability. *J. Econom. Theory 113*, 2 (2003), 199–222.

[9] CERREIA-VIOGLIO, S., MACCHERONI, F., AND MARINACCI, M. Ergodic theorems for lower probabilities. *Proc. Amer. Math. Soc. 144*, 8 (2016), 3381–3396.

[10] CHEN, Z. Strong laws of large numbers for sub-linear expectations. *Sci. China Math. 59*, 5 (2016), 945–954.

[11] CHEN, Z., WU, P., AND LI, B. A strong law of large numbers for non-additive probabilities. *Internat. J. Approx. Reason. 54*, 3 (2013), 365–377.

[12] CHOQUET, G. Theory of capacities. *Ann. Inst. Fourier (Grenoble) 5* (1953/54), 131–295 (1955).

[13] DA PRATO, G., AND ZABCZYK, J. *Ergodicity for infinite-dimensional systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[14] DE COOMAN, G., AND MIRANDA, E. Weak and strong laws of large numbers for coherent lower previsions. *J. Statist. Plann. Inference 138*, 8 (2008), 2409–2432.

[15] DENNEBERG, D. Non-additive measure and integral, basic concepts and their role for applications. In *Fuzzy measures and integrals*, vol. 40 of *Stud. Fuzziness Soft Comput.* Physica, Heidelberg, 2000, pp. 42–69.

[16] DIESTEL, J. *Sequences and series in Banach spaces.* Graduate Texts in Mathematics, 92. Springer-Verlag, New York, 1984.

[17] DOOB, J. L. *Stochastic processes*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953.

[18] DUNFORD, N., AND SCHWARTZ, J. T. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[19] EINSIEDLER, M., AND WARD, T. *Ergodic theory with a view towards number theory*, vol. 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

[20] FENG, C., QU, B., AND ZHAO, H. A sufficient and necessary condition of PS-ergodicity of periodic measures and generated ergodic upper expectations. *Nonlinearity 33*, 10 (2020), 5324–5354.

[21] FENG, C., WU, P., AND ZHAO, H. Ergodicity of invariant capacities. *Stochastic Process. Appl. 130*, 8 (2020), 5037–5059.

[22] FENG, C., AND ZHAO, H. Ergodicity of sublinear Markovian semigroups. *SIAM J. Math. Anal. 53*, 5 (2021), 5646–5681.

[23] FOLLAND, G. B. *Real analysis*, second ed. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[24] FURSTENBERG, H. *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures.

[25] FURSTENBERG, H., AND KESTEN, H. Products of random matrices. *Ann. Math. Statist. 31* (1960), 457–469.

[26] GLASNER, E. *Ergodic theory via joinings*, vol. 101 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2003.

[27] HAMMERSLEY, J. M., AND WELSH, D. J. A. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In *Proc. Internat. Res. Semin., Stat., Univ. California, Berkeley, Calif., 1963* (1965), Springer-Verlag, New York, pp. 61–110.

[28] HUREWICZ, W. Ergodic theorem without invariant measure. *Ann. of Math. (2) 45* (1944), 192–206.

[29] KINGMAN, J. F. C. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B 30* (1968), 499–510.

[30] KINGMAN, J. F. C. Subadditive ergodic theory. *Ann. Probab. 1* (1973), 883–909.

[31] KRENGEL, U. *Ergodic Theorems*. De Gruyter, Berlin, New York, 1985.

[32] MACCHERONI, F., AND MARINACCI, M. A strong law of large numbers for capacities. *Ann. Probab. 33*, 3 (2005), 1171–1178.

[33] MARINACCI, M., AND MONTRUCCHIO, L. *Introduction to the mathematics of ambiguity, in: I. Gilboa (Ed.), Uncertainty in Economic Theory: Essays in Honor of David Schmeidler's 65th Birthday*. Routledge, New York, 2004.

[34] OSELEDEC, V. I. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskov. Mat. Obšč. 19* (1968), 179–210.

[35] OXTOBY, J. C. Ergodic sets. *Bull. Amer. Math. Soc. 58* (1952), 116–136.

[36] PENG, S. G-expectation, G-Brownian motion and related stochastic calculus of Itô type. In *Stochastic analysis and applications*, vol. 2 of *Abel Symp.* Springer, Berlin, 2007, pp. 541–567.

[37] PENG, S. Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations. *Sci. China Ser. A 52*, 7 (2009), 1391–1411.

[38] QUAS, A. *Ergodicity and Mixing Properties*. Springer New York, New York, NY, 2011, pp. 225–240.

[39] ROSENBLATT, J. M., AND WIERDL, M. Pointwise ergodic theorems via harmonic analysis. In *Ergodic theory and its connections with harmonic analysis (Alexandria, 1993)*, vol. 205 of *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, Cambridge, 1995, pp. 3–151.

[40] RUDIN, W. *Functional analysis*, Second edition. *International Series in Pure and Applied Mathematics.* McGraw-Hill, Inc., New York, 1991. Cambridge Univ. Press, Cambridge, 1995, pp. 3–151.

[41] SCHMEIDLER, D. Integral representation without additivity. *Proc. Amer. Math. Soc. 97*, 2 (1986), 255–261.

[42] SCHMEIDLER, D. Subjective probability and expected utility without additivity. *Econometrica 57*, 3 (1989), 571–587.

[43] TEMAM, R. *Infinite-dimensional dynamical systems in mechanics and physics*, vol. 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1988.

[44] VOLNÝ, D. Martingale decompositions of stationary processes. *Yokohama Math. J. 3*5 (1987), no. 1-2, 113–121.

[45] WALLEY, P. *Statistical reasoning with imprecise probabilities*, vol. 42 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London, 1991.

[46] WALTERS, P. *An introduction to ergodic theory*, vol. 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982. ix+250 pp.

[47] WANG, Z., AND KLIR, G. J. *Generalized measure theory*, vol. 25. Springer, 2009.