Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 120 (2010) 1908–1919
stochastic processes and their applications
www.elsevier.com/locate/spa

Ergodic theorems for extended real-valued random variables

Christian Hess[a,*], Raffaello Seri[b], Christine Choirat[c]

[a] *Département MIDO, Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France*

[b] *Dipartimento di Economia, Università degli Studi dell'Insubria, via Monte Generoso 71, 21100 Varese, Italy*

[c] *Department of Quantitative Methods, School of Economics and Business Management, Universidad de Navarra, 31080 Pamplona, Spain*

Received 26 October 2008; received in revised form 12 May 2010; accepted 15 May 2010
Available online 24 May 2010

Abstract

We first establish a general version of the Birkhoff Ergodic Theorem for quasi-integrable extended real-valued random variables without assuming ergodicity. The key argument involves the Poincaré Recurrence Theorem. Our extension of the Birkhoff Ergodic Theorem is also shown to hold for asymptotic mean stationary sequences. This is formulated in terms of necessary and sufficient conditions. In particular, we examine the case where the probability space is endowed with a metric and we discuss the validity of the Birkhoff Ergodic Theorem for continuous random variables. The interest of our results is illustrated by an application to the convergence of statistical transforms, such as the moment generating function or the characteristic function, to their theoretical counterparts.

Keywords: Birkhoff's Ergodic Theorem; Asymptotic mean stationarity; Extended real-valued random variables; Non-integrable random variables; Cesaro convergence; Conditional expectation

1. Introduction

As is well-known, the Birkhoff Ergodic Theorem (BET) is one of the most important and beautiful results of probability theory. It has found a lot of applications in various areas such as

* Corresponding author. Tel.: +33 1 44 05 46 44; fax: +33 1 44 05 40 36.
E-mail addresses: Christian.Hess@dauphine.fr (C. Hess), raffaello.seri@uninsubria.it (R. Seri), cchoirat@unav.es (C. Choirat).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2010.05.008

dynamical systems, stochastic optimization, statistical mechanics and economics. In its classical form, it asserts that the Cesaro average of a sequence of real-valued measurements on a system evolving according to the law of motion T, on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, converges to a function having an explicit form in terms of the original process. The same result can be restated in terms of random variables, which is a major generalization of the Strong Law of Large Numbers. The extension of the BET to non-integrable and/or extended real-valued random variables, especially to random variables taking on the value $+\infty$ on a set of positive measure, is of interest. This is useful when dealing with random variables depending on a parameter, for example in statistical estimation, in the theory of random sets and in stochastic optimization (see e.g. [12]). More generally, the extension of the BET to asymptotic mean stationary sequences (see e.g. [22] or [30]) of possibly non-integrable, extended real-valued random variables appears as a natural and useful objective. In this framework, it is desirable to derive necessary and sufficient conditions for the convergence of the Cesaro mean as the sample size tends to infinity. Although one can find some related results or remarks in the literature, general statements and proofs seem to be lacking. In fact, some papers present results that rely on a version of the BET for extended real-valued random variables, but no proof of this theorem is known to the authors. This point will be discussed in detail in Section 5.

The aim of the present paper is threefold. We first prove a new version of the Birkhoff Ergodic Theorem for a stationary sequence of quasi-integrable random variables taking their values in $\overline{\mathbb{R}}$. Secondly, we extend this result to asymptotically mean stationary (ams) sequences of random variables in terms of necessary and sufficient conditions. Indeed, we establish that a probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{A})$ is ams with respect to a measurable transformation T if and only if the BET holds for any extended real-valued quasi-integrable random variable defined on $(\Omega, \mathcal{A})$. Moreover, it is shown that this equivalence remains true for particular classes of random variables, e.g. lower semicontinuous ones (when Ω is assumed to be a metric space). In particular, the case of continuous random variables is examined and a counterexample shows that even if the BET holds for all continuous random variables, it may fail for lower semicontinuous ones. Finally, an application to the theory of statistical transforms is sketched. It shows how some results of the literature on estimation through statistical transforms can be simplified and extended using our results.

The paper is organized as follows. The main results are stated in Section 2. First, we present an extension of the Birkhoff Ergodic Theorem for extended real-valued random variables, without assuming ergodicity. Section 2 also contains generalizations of this result to the case of asymptotic mean stationary sequences. Formulations in terms of necessary and sufficient conditions are provided, and the special case where Ω is a metric space is examined. Examples and remarks make more precise the domain of applicability of the results and show connections with related properties. In Section 3 we show how the results of Section 2 can provide some limit theorems for statistical transforms to be used for estimation of stochastic processes. The proofs of the main results are developed in Section 4. Section 5 contains additional comments on the literature and final remarks.

In the rest of this introduction, we set our notation and terminology, and we compile some basic facts. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, an $\mathcal{A}$-measurable transformation $T : \Omega \to \Omega$ is said to be *nonsingular* if the probability $\mathbb{P}T^{-1}$ is absolutely continuous with respect to $\mathbb{P}$. The transformation T is said to be *measure-preserving* if $\mathbb{P}\left(T^{-1}A\right) = \mathbb{P}(A)$ for all $A \in \mathcal{A}$. We also say that T preserves the $\mathbb{P}$-measure. Equivalently, $\mathbb{P}$ is said to be *stationary* with respect to T. The sets $A \in \mathcal{A}$ that satisfy $T^{-1}A = A$ are called *invariant sets* and constitute a sub-σ-field $\mathcal{I}$ of $\mathcal{A}$. More generally, a random variable X such that $X(\omega) = X(T\omega)$ for all $\omega \in \Omega$ is said to be *invariant*. It is known that X is invariant if and only if it is $\mathcal{I}$-measurable. The notion of a $\mathbb{P}$-*almost sure invariant* set is also useful. The class of these sets constitutes a σ-field which is equal to the $\mathbb{P}$-completion of $\mathcal{I}$. A measurable and measure-preserving transformation T is said to be *ergodic* if $\mathbb{P}(A) = 0$ or 1 for all invariant sets A. Equivalently, the sub-σ-field $\mathcal{I}$ reduces to the trivial σ-field $\{\Omega, \varnothing\}$ (up to the $\mathbb{P}$-null sets). Another well-known formulation is also used: a sequence of random variables $X_1, X_2, \ldots$ is said to be *stationary* if the random vectors $(X_1, \ldots, X_n)$ and $(X_{k+1}, \ldots, X_{n+k})$ have the same distribution for all integers $n, k \geq 1$. Any stationary sequence $X_1, X_2, \ldots$ can almost surely be rewritten using a measurable and measure-preserving transformation T as $X_t(\omega) = X\left(T^t \omega\right)$ (see e.g. [10, Proposition 6.11]). The transformation $T : \Omega \to \Omega$ on $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be *asymptotically mean stationary* (ams) if the sequence $\frac{1}{n}\sum_{j=0}^{n-1} \mathbb{P}\left(T^{-j} A\right)$ is convergent for all $A \in \mathcal{A}$. From the Vitali–Hahn–Saks Theorem, it is known that $\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{P}T^{-j}$ is a probability measure that we shall denote by $\mathbb{P}^\star$ and call the *asymptotic mean of* $\mathbb{P}$. It is not difficult to see that the transformation T is ams on $(\Omega, \mathcal{A}, \mathbb{P})$ if and only if for each bounded real-valued random variable X the limit

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i \omega\right)$$

exists for $\mathbb{P}$-almost all $\omega \in \Omega$ (see e.g. [30, Theorem 4.10]). In the sequel, when this limit exists, without assuming that T is stationary, it will be convenient to say that the *generalized* BET holds for the random variable X. The random variable X is said to be *quasi-integrable* in the sense of Neveu (see [34, p. 40]) if either $\mathbb{E}X^+$ or $\mathbb{E}X^-$ is finite, where $X^+ = \max\{X, 0\}$ (resp. $X^- = \max\{-X, 0\}$) stands for the positive (resp. the negative) part of X. Given a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$, the *conditional expectation* of X with respect to $\mathcal{B}$ is denoted by $\mathbb{E}(X|\mathcal{B})$ or $\mathbb{E}^{\mathcal{B}}(X)$. For any $A \in \mathcal{A}$, the *(probabilistic) indicator function* of A is denoted by 1_A and defined by $1_A(\omega) = 1$ if $\omega \in A$, 0 otherwise.

2. Main results

We first present a version of the BET for quasi-integrable extended real-valued random variables under stationary, but not necessarily ergodic, transformations. The random variables may even take infinite values on a set of positive measure. As we shall see in the proof (Section 4), this can be handled by an appropriate application of the Poincaré Recurrence Theorem.

Theorem 1. *Let $(\Omega, \mathcal{A})$ be a measurable space and $T : \Omega \to \Omega$ be a measurable transformation. Suppose moreover that $\mathbb{P}$ is a stationary probability measure with respect to T on $(\Omega, \mathcal{A})$. Then, for every quasi-integrable extended real-valued random variable X, one has*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i \omega\right) = \mathbb{E}\left(X\,|\mathcal{I}\right)(\omega) \quad \mathbb{P}\text{-almost surely}$$

(where both sides can be equal to $+\infty$ or $-\infty$).

Remark 2. The extensions to finite measure spaces and two-sided sequences are standard. As for continuous time averages, the reader may consult [30, p. 10] and [26, p. 183, Corollary 10.9].

We now examine the extension of the Birkhoff Ergodic Theorem for quasi-integrable extended real-valued random variables in the case of ams transformations.

Theorem 3. *Let $(\Omega, \mathcal{A})$ be a measurable space, $T : \Omega \to \Omega$ be a measurable transformation and X be an extended real-valued random variable defined on $(\Omega, \mathcal{A})$. In addition, suppose that $\mathbb{P}$ is an* ams *probability measure with respect to T on $(\Omega, \mathcal{A})$ with stationary mean $\mathbb{P}^\star$. Also assume that X is $\mathbb{P}^\star$-quasi-integrable. Then, for $\mathbb{P}$ (and $\mathbb{P}^\star$)-almost every $\omega \in \Omega$, one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X\left(T^i \omega\right) = \mathbb{E}^\star (X \,|\mathcal{I})(\omega),$$

where each side can be equal to $+\infty$ or $-\infty$, and where $\mathbb{E}^\star (X\,|\mathcal{I})$ denotes the conditional expectation taken on $(\Omega, \mathcal{A}, \mathbb{P}^\star)$.

This kind of result can be stated in quite a different form. Gray (see [21, p. 174]) introduces the following definition: a dynamical system $(\Omega, \mathcal{A}, \mathbb{P}, T)$ is said to have the *ergodic property with respect to the measurement X* if the BET holds for X. More generally, a dynamical system is said to have the *ergodic property with respect to the class of measurements $\mathcal{X}$* if it has the ergodic property with respect to any $X \in \mathcal{X}$. Thus, Theorem 3 shows that an ams dynamical system $(\Omega, \mathcal{A}, \mathbb{P}, T)$ has the ergodic property with respect to the class of $\mathbb{P}^\star$-quasi-integrable functions.[1] The following result presents necessary and sufficient conditions for the generalized BET to hold.

Theorem 4. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $T : \Omega \to \Omega$ be a $\mathbb{P}$-measurable transformation. Then the following statements are equivalent.*

1. *$\mathbb{P}$ is an* ams *probability measure with respect to T on $(\Omega, \mathcal{A})$.*
2. *The generalized* BET *holds for every nonnegative extended real-valued random variable X, that is $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X\left(T^i\omega\right)$ exists for $\mathbb{P}$-almost all $\omega \in \Omega$.*
3. *For every nonnegative $\mathbb{P}$-integrable or $\mathbb{P}^\star$-integrable real-valued random variable X, the generalized* BET *holds.*
4. *For every bounded real-valued random variable X, the generalized* BET *holds.*
5. *For every $A \in \mathcal{A}$ the generalized* BET *holds for 1_A.*

The following corollary deals with the case where Ω is a Polish space endowed with its Borel σ-field. Recall that a topological space Ω is said to be *Polish* if it admits an equivalent metric d making (Ω, d) separable and complete (see e.g. [3]).

Corollary 5. *Assume that Ω is a Polish space endowed with its Borel σ-field $\mathcal{A} = \mathcal{B}(\Omega)$. Then the following statements are equivalent.*

1. *$\mathbb{P}$ is an* ams *probability measure with respect to T on $(\Omega, \mathcal{A})$.*
2. *For every open subset G of Ω the generalized* BET *holds for 1_G.*
3. *For every nonnegative lower semicontinuous function $X : \Omega \to \mathbb{R}$ the generalized* BET *holds.*

Remark 6. Obviously, if a necessary and sufficient condition has to be used to verify ams through testing that the BET holds over a class of functions, it is preferable that the class of functions be as small as possible. On the other hand, it is important to display the strongest consequences of the ams property. These two opposite aspects have been taken into account in the statement of the above two results. Clearly, several other equivalent properties could have been given as well.

---

[1] However, uniform convergence over classes of measurements can fail (see Nobel [35]).

Remark 7. Also observe that, when $\mathbb{P}$ is ams with respect to T, but T does not preserve the $\mathbb{P}$-measure, the Cesaro limit appearing in Theorem 3 is not necessarily equal to $\mathbb{E}(X|\mathcal{I})$, the conditional expectation taken on $(\Omega, \mathcal{A}, \mathbb{P})$. More precisely, if X is $\mathbb{P}^\star$-quasi-integrable, the Cesaro limit is equal to $\mathbb{E}^\star(X|\mathcal{I})$, the conditional expectation taken on $(\Omega, \mathcal{A}, \mathbb{P}^\star)$, but not necessarily to $\mathbb{E}(X|\mathcal{I})$. For example, consider the set $\Omega = \{0, 1\}$ endowed with the σ-field $\mathcal{A} = \{\varnothing, \{0\}, \{1\}, \Omega\}$, the identity map $X : \Omega \to \Omega$ and the transformation $T : \Omega \to \Omega$ defined by $T(0) = 0$, $T(1) = 0$. Thus, the invariant σ-field reduces to $\mathcal{I} = \{\Omega, \varnothing\}$. Moreover, consider the probability $\mathbb{P}$ on $(\Omega, \mathcal{A})$ such that $\mathbb{P} = \{p_0, p_1\}$ with $p_0 > 0$ and $p_1 > 0$ summing to one. Clearly, T does not preserve $\mathbb{P}$. Further, for every integer $n \geq 1$ and $A \in \mathcal{A}$ we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{P}\left(\omega \in T^{-i}A\right) = \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{P}\left(T^{i}\omega \in A\right)$$

and this sequence converges to $\mathbb{P}(0 \in A)$. Consequently, $\mathbb{P}^\star(A)$ exists and satisfies $\mathbb{P}^\star(A) = 1$ if $0 \in A$ and $\mathbb{P}^\star(A) = 0$ otherwise. Hence, we can see that T is ams with respect to $\mathbb{P}$ and that the corresponding stationary mean $\mathbb{P}^\star$ is the Dirac measure at 0, denoted by δ_0. On the other hand, for every $n \geq 1$ the Cesaro mean

$$\frac{1}{n}\sum_{i=0}^{n-1}X\left(T^{i}\omega\right) = \frac{1}{n}\sum_{i=0}^{n-1}T^{i}\omega$$

is equal to 0 if $\omega = 0$ and to $\frac{1}{n}$ if $\omega = 1$. Thus, the limit exists and is equal to 0 for all $\omega \in \Omega$. As to the conditional expectation of X with respect to $\mathcal{I}$, it is immediately checked that $\mathbb{E}(X|\mathcal{I}) = \mathbb{E}(X) = p_1$ and $\mathbb{E}^\star(X|\mathcal{I}) = \mathbb{E}^\star(X) = 0$.

The above results show the relevance of asymptotic mean stationarity to ergodic theory and in particular to the convergence of Cesaro means. Similar conditions have appeared throughout the literature, especially in statistics, econometrics and information theory. In these domains, they are intended to allow for a certain freedom in the behavior of the processes that are to be modeled, nevertheless retaining the asymptotic properties of stationarity. In information theory, [18,27] study ams channels, and [22,6,2] derive versions of the Shannon–McMillan–Breiman Theorem for ams processes. In queueing theory, ams processes have proved to be a valuable tool as shown by [41, pp. 120–121] and [33, p. 424] where several applications and references are provided. Several alternative definitions of asymptotic stationarity are considered in [42,43], one of which (called *strong asymptotic stationarity in mean*; see [42, p. 821]) coincides with ams. In econometrics, an assumption similar to asymptotic mean stationarity is used in [37, p. 677, Assumption 5B] and [38, p. 42, Assumption D']. In economics, Kurz's theory of rationality is based on a variation of asymptotic mean stationarity (see [31]).

Remark 8. Clearly, statement 3 of Corollary 5 implies that for every positive continuous function $X : \Omega \to \mathbf{R}$ the generalized BET holds. This is connected with the concept of Cesaro summable sequences introduced in [19, pp. 77–78] and [20, Theorem 2, p. 702]. Assume that V is a Borel subset of some Polish space. A sequence $(s_i)_{i\geq 1}$ in V is said to *generate Cesaro summable sequences* with respect to the probability measure ν defined on the Borel subsets of V if, for every real-valued continuous function f with $\int_V |f(s)|\,\nu(\mathrm{d}s) < +\infty$, one has

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f(s_i) = \int_V f(s)\,\nu(\mathrm{d}s)\,. \tag{1}$$

If we take $f = X$ with $X : V \to V$, a Borel measurable function, and $s_i = T^i\omega$, it follows from Corollary 5 that (1) is implied by asymptotic mean stationarity, with ergodic mean $\mathbb{P}^\star = \nu$. In other words, this definition entails that the sequence $\left(T^i\omega\right)_{i\geq 1}$ generates a Cesaro summable sequence with respect to ν.

Remark 9. As already observed, statement 3 of Corollary 5 implies that for every positive continuous function $X : \Omega \to \mathbb{R}$ the generalized BET holds. However, the following counterexample shows that the converse implication is not true. By Corollary 5, it suffices to show that, even if the generalized BET holds for every continuous function, it may fail for some lower semicontinuous function. Consider a sequence $(\omega_n)_{n\in\mathbb{N}}$ of reals that converges to some point ω_∞. We set $\Omega = \{\omega_n : n \in \mathbb{N} \cup \{\infty\}\}$ and we assume that the ω_n's are pairwise distinct. We define the measurable transformation T by $T\omega_n = \omega_{n+1}$ for all $n \in \mathbb{N}$ and $T\omega_\infty = \omega_\infty$. The probability measure $\mathbb{P}$ on Ω is only required to satisfy $\mathbb{P}(\{\omega_n\}) > 0$ for some $n \in \mathbb{N}$. Clearly, for all $\omega \in \Omega$, $T^n\omega$ converges to ω_∞ as n goes to infinity. Consequently, for every continuous function f defined on Ω, one has $f(\omega_\infty) = \lim_{n\to\infty} f(T^n\omega)$, which, as is well-known, entails

$$f(\omega_\infty) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f\left(T^i\omega\right).$$

Now, consider the function g defined on Ω by $g(\omega_\infty) = 0$ and for each $n \in \mathbb{N}$

$$g(\omega_n) = \begin{cases} 2^k & \text{if } n = 2^k \text{ with } k \in \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, g is lower semicontinuous on Ω, but not continuous at ω_∞, since

$$g(\omega_\infty) = 0 = \liminf_{n\to\infty} g(\omega_n) < \limsup_{n\to\infty} g(\omega_n) = +\infty.$$

On the other hand, it is not difficult to check that, for each $\omega \in \Omega \setminus \{\omega_\infty\}$, one has

$$1 = \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} g\left(T^i\omega\right) < \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} g\left(T^i\omega\right) = 2.$$

This shows that the BET does not hold for g.

3. Application: statistical transforms

In this section we show how the previous results can provide some limit theorems for statistical transforms to be used for estimation of the distribution of stochastic processes (see e.g. [17,15]). Instead of providing a unified theory, we prefer to focus on two simple examples that provide an illustration of the power of the results and of the extent to which they simplify the derivation of limit theorems. In both examples we show the convergence of an empirical statistical transform to a limiting transform computed according to the probability $\mathbb{P}^\star$. Further comments are contained in the remarks.

The first example concerns the *Laplace transform* or *moment generating function* and shows that the absence of integrability requirements can be useful whenever it is not known whether integrability holds. Define the Laplace transform of a random vector $X \in \mathbb{R}^k$ with probability $\mathbb{P}$ as $m(\theta) = \mathbb{E}e^{\theta' X} = \int_{\mathbb{R}^k} e^{\theta' x}\mathbb{P}(dx)$ for $\theta \in \mathbb{R}^k$. Suppose now that we observe a vector time series $\{X_t\}_{t=0,\dots,(T-1)}$ extracted from a stochastic process $\{X_t\}_{t\in\mathbb{N}}$ and build the empirical measure $\mathbb{P}_T$ on the basis of the observed sequence. We define the empirical Laplace transform

$m_T(\theta) = \frac{1}{T}\sum_{t=0}^{T-1} e^{\theta' X_t} = \int_{\mathbb{R}^k} e^{\theta' x}\mathbb{P}_T(dx)$. We show that this function converges to $m(\theta) = \mathbb{E}^\star e^{\theta' X} = \int_{\mathbb{R}^k} e^{\theta' x}\mathbb{P}^\star(dx)$, yielding therefore a (pointwise) consistent estimator of $m(\theta)$. The following theorem is a version of Theorem 2 in [25] in this special context.

Theorem 10. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ is a $\mathbb{R}^k$-valued* ams *process with stationary mean $\mathbb{P}^\star$ that is also ergodic. Then $m_T(\theta) \to m(\theta)$ $\mathbb{P}$ and $\mathbb{P}^\star$-almost surely for any $\theta \in \mathbb{R}^k$.*

Remark 11. (i) No integrability requirement is necessary, in the sense that if $m(\theta) = +\infty$, $m_T(\theta)$ will almost surely diverge to $+\infty$.
(ii) If $\{X_t\}_{t\in\mathbb{N}}$ is stationary ergodic, then $m(\theta) = \mathbb{E}e^{\theta' X}$. As an example, the result holds for iid processes and for ergodic Markov chains without any further requirement. This implies that the result is relevant for estimation of the stationary distribution of ergodic Markov chains.
(iii) Statistical estimation based on the Laplace transform is considered, among other references, in [17].

Proof. The function $e^{\theta' x}$ is bounded from below by 0 and therefore quasi-integrable. The other conditions of Theorem 3 are immediately checked. □

On the other hand, our main theorem can also be used to extend the class of stochastic processes covered by classical limit theorems. Consider for simplicity a scalar time series $\{X_t\}_{t=0,\ldots,(T-1)}$ extracted from a stochastic process $\{X_t\}_{t\in\mathbb{N}}$. As in Section 5 of [16] and in [15], build the vector $X_t^p = \left(X_t, X_{t-1}, \ldots, X_{t-p}\right)'$. If the process $\{X_t\}_{t\in\mathbb{N}}$ (and the associated probability measure $\mathbb{P}$) were stationary, we could have defined the *poly-characteristic function* (pcf) $c^p(\theta) = \mathbb{E}e^{i\theta' X_t^p}$ where $\mathbb{E}$ is the average under $\mathbb{P}$; since this is not the case, we let $c^p(\theta) = \mathbb{E}^\star e^{i\theta' X_t^p}$. We define the empirical pcf (or epcf) as $c_T^p(\theta, \omega) = \frac{1}{T}\sum_{t=p}^{T-1} e^{i\theta' X_t^p}$. In [15] a consistency result is provided under ergodicity and strong mixing of $\{X_t\}_{t\in\mathbb{N}}$ (or under the stricter hypothesis 2.1); remark however that the paper also investigates asymptotic normality and efficiency and more stringent requirements than asymptotic mean stationarity are required in these cases. Here we show that, as regards consistency, asymptotic mean stationarity with an ergodic mean is enough.

Theorem 12. *Suppose $\{X_t\}_{t\in\mathbb{N}}$ is a $\mathbb{R}$-valued* ams *process with stationary mean $\mathbb{P}^\star$ that is also ergodic. Then $c_T^p(\theta, \omega) \to c^p(\theta)$ $\mathbb{P}$ and $\mathbb{P}^\star$-almost surely for any $\theta \in \mathbb{R}^{p+1}$. Moreover the convergence is uniform for $\theta \in [-T, +T]^{p+1}$ for any fixed $0 < T < \infty$.*

Remark 13. (i) The extension to vectorial stochastic processes is straightforward and left to the reader.
(ii) Also Theorem 10 could be extended to give a limit theorem for the empirical Laplace transform of X_t^p.
(iii) This result can be used whenever the observed time series has a transient and the object of interest is its steady state synthesized in the probability $\mathbb{P}^\star$ (see e.g. [39] for an economic example).

Proof. If the original stochastic process $\{X_t\}_{t\in\mathbb{N}}$ is ams with ergodic mean, then also $\left\{X_t^p\right\}_{t\geq p}$ is. As regards $e^{i\theta' X_t^p}$, we recall that $e^{i\theta' X_t^p} = \cos\left(\theta' X_t^p\right) + i\cdot\sin\left(\theta' X_t^p\right)$. Since the two functions are bounded, we apply Theorem 3 separately to the real and the imaginary part. The rest of the proof follows the pattern of that of Theorem 2.1 in [15]. □

4. Proofs of the main results

The present section contains the proofs of the results stated in Section 2. Given an $\mathcal{A}$-measurable transformation $T : \Omega \to \Omega$ and a real-valued or extended real-valued random variable X, the corresponding Cesaro mean is denoted by

$$u_n(\omega) = \frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i\omega\right) \quad n \geq 1,\ \omega \in \Omega.$$

For proving Theorem 1, we need the following lemma on the $\mathcal{I}$-measurability of $\liminf_{n\to\infty} u_n$ and $\limsup_{n\to\infty} u_n$ when X is an extended real-valued random variable. It is more precise than [12, Proposition 3.6], and we include the short proof for the convenience of the reader.

Lemma 14. *Let $(\Omega, \mathcal{A})$ be a measurable space, $T : \Omega \to \Omega$ a measurable transformation, $\mathbb{P}$ a stationary probability measure (with respect to T) on $(\Omega, \mathcal{A})$ and $X : \Omega \to [0, +\infty]$ an $\mathcal{A}$-measurable random variable.*
(i) *If X is finite valued, then $\liminf_{n\to\infty} u_n$ and $\limsup_{n\to\infty} u_n$ are $\mathcal{I}$-measurable.*
(ii) *If X can take on the value $+\infty$, the above random variables are $\mathcal{I}_{\mathbb{P}}$-measurable, where $\mathcal{I}_{\mathbb{P}}$ denotes the $\mathbb{P}$-completion of $\mathcal{I}$.*

Proof. As to statement (i), consider the $\mathcal{A}$-measurable function $u = \limsup_{n\to\infty} u_n$. In order to prove that u is also $\mathcal{I}$-measurable, let us show that it is invariant, i.e. that $u(T\omega) = u(\omega)$ for all $\omega \in \Omega$. For this purpose, observe that for all $n \geq 1$ and for all $\omega \in \Omega$, we can write

$$u_{n+1}(\omega) = \frac{X(\omega)}{n+1} + \frac{n}{n+1} u_n(T\omega). \tag{2}$$

The result immediately follows by taking the lim sup on both sides of (2). To prove statement (ii), consider the $\mathcal{A}$-measurable subset

$$A_\infty = \{\omega \in \Omega : X(\omega) = +\infty\}.$$

If $\mathbb{P}(A_\infty) = 0$, the proof is over. In the case where $\mathbb{P}(A_\infty) > 0$, we appeal to the Poincaré Recurrence Theorem (see e.g. [36, Theorem 3.2, p. 34]) applied to A_∞. This result asserts that $\mathbb{P}$-almost every point of A_∞ is recurrent with respect to A_∞, that is, for such a point ω, there exists $k \geq 1$ such that $T^k\omega \in A_\infty$ or, equivalently, $X\left(T^k\omega\right) = +\infty$. Therefore, as soon as $n > k$, we have $u_{n+1}(\omega) = u_n(\omega) = +\infty$, which in turn implies $u(\omega) = u(T\omega) = +\infty$ for $\mathbb{P}$-almost all $\omega \in \Omega$. This shows the $\mathcal{I}_{\mathbb{P}}$-measurability of u. Similar proofs hold for the inferior limit. □

Proof of Theorem 1. Since X is quasi-integrable, either $\mathbb{E}\max(X, 0)$ or $\mathbb{E}\min(X, 0)$ is finite, so it suffices to consider the non-integrable part of X, say the positive one. Therefore, we can restrict our analysis to a positive random variable X. It is enough to establish the following two inequalities:

$$\liminf_{n\to\infty} u_n(\omega) \geq \mathbb{E}(X|\mathcal{I})(\omega) \quad \mathbb{P}\text{-a.s.} \tag{3}$$

$$\limsup_{n\to\infty} u_n(\omega) \leq \mathbb{E}(X|\mathcal{I})(\omega) \quad \mathbb{P}\text{-a.s.} \tag{4}$$

because, according to Lemma 14, the corresponding events are invariant. As to (3), for every integer $m \geq 1$, consider the random variable X_m defined by

$$X_m(\omega) = \begin{cases} X(\omega) & \text{if } X(\omega) \le m \\ 0 & \text{otherwise.} \end{cases}$$

Since X_m is integrable, the BET implies

$$\liminf_{n\to\infty} u_n(\omega) \ge \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} X_m\left(T^i\omega\right) = \mathbb{E}\left(X_m \,|\mathcal{I}\right)(\omega).$$

Letting $m \to \infty$ and invoking the Monotone Convergence Theorem for conditional expectation (see e.g. [13, Theorem 10.15]) we get (3). Now, let us prove (4). Since, by Lemma 14, both sides of (4) are $\mathcal{I}$-measurable, this inequality is equivalent to

$$\int_B \limsup_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega) \le \int_B X(\omega)\,\mathbb{P}(\mathrm{d}\omega) \tag{5}$$

for any $B \in \mathcal{I}$. If $\int_B X(\omega)\,\mathbb{P}(\mathrm{d}\omega) = +\infty$, then (5) is trivially satisfied. If $\int_B X(\omega)\,\mathbb{P}(\mathrm{d}\omega)$ is finite, then the random variable $1_B X$ is integrable. Since 1_B is $\mathcal{I}$-measurable, it is invariant, which entails for almost all $\omega \in \Omega$

$$\begin{aligned}\int_B \left(\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega) &= \int_\Omega \left(\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} (1_B X)\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega)\\ &= \int_\Omega \left(\liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} (1_B X)\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega).\end{aligned}$$

The second equality follows from the equality between the lim sup and the lim inf. Indeed, the limit of the sequence $\frac{1}{n}\sum_{i=0}^{n-1}(1_B X)\left(T^i\omega\right)$ exists for almost all $\omega \in \Omega$ (and, by the BET, is equal to $\mathbb{E}\left(1_B X \,|\mathcal{I}\right)(\omega)$). Therefore, Fatou's Lemma yields

$$\int_B \left(\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} X\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega) \le \liminf_{n\to\infty}\int_\Omega \left(\frac{1}{n}\sum_{i=0}^{n-1} (1_B X)\left(T^i\omega\right)\right)\mathbb{P}(\mathrm{d}\omega). \tag{6}$$

Now, applying the BET in L^1-mean to $1_B X$ shows that the sequence $\left(\frac{1}{n}\sum_{i=0}^{n-1}(1_B X)\right)_{n\ge 1}$ converges in L^1 to $\mathbb{E}\left(1_B X \,|\mathcal{I}\right)$, that is

$$\lim_{n\to\infty}\int_\Omega \left|\frac{1}{n}\sum_{i=0}^{n-1}(1_B X)\left(T^i\omega\right) - \mathbb{E}\left(1_B X\,|\mathcal{I}\right)(\omega)\right|\mathbb{P}(\mathrm{d}\omega) = 0.$$

Since $B \in \mathcal{I}$, this equality implies

$$\lim_{n\to\infty}\int_\Omega \frac{1}{n}\sum_{i=0}^{n-1}(1_B X)\left(T^i\omega\right)\mathbb{P}(\mathrm{d}\omega) = \int_\Omega \mathbb{E}\left(1_B X\,|\mathcal{I}\right)(\omega)\,\mathbb{P}(\mathrm{d}\omega) = \int_B X(\omega)\,\mathbb{P}(\mathrm{d}\omega),$$

which, together with (6), entails inequality (4). □

Proof of Theorem 3. We use the same notation as in the previous proof. Since $\mathbb{P}$ is ams with stationary mean $\mathbb{P}^\star$, the transformation T preserves the $\mathbb{P}^\star$-measure. Thus, Lemma 14(ii) yields the $\mathcal{I}_{\mathbb{P}^\star}$-measurability of $\liminf_{n\to\infty} u_n$ and $\limsup_{n\to\infty} u_n$. This allows applying the same arguments as in the proof of Theorem 1, replacing $\mathbb{P}$ with $\mathbb{P}^\star$, and yields the desired result. □

Proof of Theorem 4. The implication 1 $\Rightarrow$ 2 immediately follows from Theorem 3. Indeed, every nonnegative random variable is quasi-integrable with respect to any probability measure. Implication 2 $\Rightarrow$ 3 is trivial since 3 is a special case of 2. To prove implication 3 $\Rightarrow$ 4, simply observe that any bounded random variable X can be written as $X = X^+ - X^-$ and that statement 3 applies to X^+ and X^-. Implication 4 $\Rightarrow$ 5 is trivial. Finally, implication 5 $\Rightarrow$ 1 is well-known and follows from Lebesgue's Dominated Convergence Theorem (see e.g. the proof in [30, Theorem 4.10]). $\square$

Proof of Corollary 5. By the equivalence 1 $\Leftrightarrow$ 2 of Theorem 4 and by the fact that a lower semicontinuous function is Borel measurable, we deduce the implication 1 $\Rightarrow$ 3. Implication 3 $\Rightarrow$ 2 is clear because the indicator function of an open subset is lower semicontinuous. It remains to prove implication 2 $\Rightarrow$ 1. For this purpose, it is enough to prove that 2 implies statement 5 of Theorem 4. Thus, consider $A \in \mathcal{A} = \mathcal{B}(\Omega)$ and $\alpha > 0$. Since every probability measure on a metric space is regular (see e.g. [9, Theorem 1.1]), there exist a closed set F and an open set G such that $F \subseteq A \subseteq G$ and $\mathbb{P}^\star(G \setminus F) \leq \alpha$. At this point, it is convenient to introduce the notation

$$u_n(\omega, A) = \frac{1}{n} \sum_{i=0}^{n-1} 1_A\left(T^i \omega\right) \quad n \geq 1,\ \omega \in \Omega,\ A \in \mathcal{A},$$
$$v(\omega, A) = \liminf_{n\to\infty} u_n(\omega, A) \quad \text{and} \quad w(\omega, A) = \limsup_{n\to\infty} u_n(\omega, A).$$

From Lemma 14 we know that $v(\cdot, A)$ and $w(\cdot, A)$ are $\mathcal{I}_{\mathbb{P}^\star}$-measurable. Moreover, the hypothesis implies that the following two equalities hold $\mathbb{P}^\star$-almost surely:

$$\mathbb{E}^\star(1_F|\mathcal{I})(\omega) = \lim_{n\to\infty} u_n(\omega, F) \quad \text{and} \quad \mathbb{E}^\star(1_G|\mathcal{I})(\omega) = \lim_{n\to\infty} u_n(\omega, G).$$

Consequently, one has for $\mathbb{P}^\star$-almost all $\omega \in \Omega$

$$\mathbb{E}^\star(1_F|\mathcal{I})(\omega) \leq v(\omega, A) \leq w(\omega, A) \leq \mathbb{E}^\star(1_G|\mathcal{I})(\omega).$$

In turn, this entails

$$\begin{aligned}\int_\Omega (w(\omega, A) - v(\omega, A))\,\mathbb{P}^\star(\mathrm{d}\omega) &\leq \int_\Omega \mathbb{E}^\star(1_G - 1_F|\mathcal{I})(\omega)\,\mathbb{P}^\star(\mathrm{d}\omega) \\ &= \int_\Omega (1_G - 1_F)(\omega)\,\mathbb{P}^\star(\mathrm{d}\omega) = \mathbb{P}^\star(G \setminus F) \leq \alpha.\end{aligned}$$

Since α is arbitrary this proves that $v(\omega, A) = w(\omega, A)$ for $\mathbb{P}^\star$-almost (and $\mathbb{P}$-almost) all $\omega \in \Omega$ and finishes the proof. $\square$

5. Additional comments on the literature

In this final section, we compare the results of Section 2 with already existing ones. The result in Theorem 1 was briefly mentioned without proof and reference by Arnold [4, p. 539], Korf and Wets [28, p. 443] and Valadier [44, p. 238]. Recently, the result has been used in the companion paper [12] (see [11] for a sketch of the proof in a special case). Versions of the BET under weak integrability requirements are given by Halmos [23, p. 32], Loève [32, Section 34.2], Krengel [30, p. 15], Breiman [10, p. 116], Valadier [44, pp. 238–239] and Kallenberg [26, Theorem 10.6]. However, they suffer from some restrictions in their applicability. Indeed, Loève's result is applicable only to real-valued quasi-integrable random variables and to transformations

T respecting the conditions stated on his page 96 (this appears to have been overlooked in subsequent references). The result of Halmos, Krengel, Breiman and Valadier, which is in fact the same, is restricted to real-valued quasi-integrable random variables for ergodic T. In the book by Kallenberg, one can find a version of the BET for quasi-integrable real-valued random variables for stationary T, but the random variables are not allowed to take on the value $+\infty$ on a set of positive measure. As to the more general case of measure-preserving T for real-valued quasi-integrable random variables, some results in the literature could be adapted to prove other versions of the BET, namely the BET for superstationary processes of Krengel [29], the subadditive ergodic theorems of Abid [1] and [30, p. 38, Theorem 5.4] (see [30, p. 49] for the definition of superstationarity). Nonetheless, the latter result is not comparable to the BET, because the superstationarity condition is stronger than stationarity.

As we have seen in Section 2, the BET can be extended to the case where T is only assumed to be ams. Results along this line have been proved by Dowker [14, Theorems III and V], Rechard [40, p. 483] and Loève [32, Section 34.2], but without explicit reference to ams. On the other hand, the ams property was explicitly considered by Gray and Kieffer [22] and, more recently, by Becker [7,8]. Dowker's result holds for invertible nonsingular T and bounded random variables, and that of Rechard is valid for nonsingular T, but for integrable random variables. As to Loève's result, it suffers from the limitations previously exposed. Gray and Kieffer's Theorem 1 in [22] (see also [30, p. 33, Theorem 4.10], [21, p. 217, Theorem 7.2.1]) is valid only for bounded random variables. Finally, Becker gives necessary and sufficient conditions for a measure to be ams, among which is an ergodic theorem for integrable functions. In spite of their diversity, each paper stresses the fact that asymptotic mean stationarity turns out to be a necessary and sufficient condition for the BET in a certain class of random variables. This shows the relevance of this question which has motivated our results in Section 2.

To conclude, we would like to point out that other classes of problems, different from the ones of statistical transforms discussed in Section 3, can be addressed with our theorems. In particular, in a companion paper, we exploit the full force of our results to study the essential intersection of random sets (see e.g. [24]) with applications to robust optimization (see e.g. [5]).

References

[1] M. Abid, Un théorème ergodique pour des processus sous-additifs et sur-stationnaires, C. R. Acad. Sci. Paris Sér. A-B 287 (3) (1978) A149–A152.

[2] P. Algoet, T. Cover, A sandwich proof of the Shannon–McMillan–Breiman theorem, Ann. Probab. 16 (2) (1988) 899–909.

[3] C. Aliprantis, K. Border, Infinite-dimensional analysis, Springer-Verlag, Berlin, 1999.

[4] L. Arnold, Random dynamical systems, in: Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998.

[5] D. Bai, T. Carpenter, J. Mulvey, Making a case for robust models, Manage. Sci. 43 (1997) 895–907.

[6] A. Barron, The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem, Ann. Probab. 13 (4) (1985) 1292–1303.

[7] M. Becker, Measures with stationary mean, Rev. Un. Mat. Argentina 39 (3-4) (1995) 205–208.

[8] M. Becker, A pointwise ergodic theorem, Rev. Un. Mat. Argentina 41 (3) (1999) 35–38.

[9] P. Billingsley, Convergence of probability measures, John Wiley & Sons Inc., New York, 1968.

[10] L. Breiman, Probability, in: Classics in Applied Mathematics, 7, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, corrected reprint of the 1968 original.

[11] C. Choirat, C. Hess, R. Seri, A functional version of the Birkhoff ergodic theorem for a normal integrand: a variational approach, Tech. Rep. No. 0101, Cahiers du Centre de Recherche Viabilité–Jeux–Contrôle, Université Paris Dauphine, 2001.

[12] C. Choirat, C. Hess, R. Seri, A functional version of the Birkhoff ergodic theorem for a normal integrand: a variational approach, Ann. Probab. 31 (1) (2003) 63–92.

[13] J. Davidson, Stochastic limit theory, in: Advanced Texts in Econometrics, The Clarendon Press Oxford University Press, New York, 1994.
[14] Y. Dowker, Finite and σ-finite invariant measures, Ann. of Math. 54 (2) (1951) 595–608.
[15] A. Feuerverger, An efficiency result for the empirical characteristic function in stationary time-series models, Canad. J. Statist. 18 (2) (1990) 155–161.
[16] A. Feuerverger, P. McDunnough, On some Fourier methods for inference, J. Amer. Statist. Assoc. 76 (374) (1981) 379–387.
[17] A. Feuerverger, P. McDunnough, On statistical transform methods and their efficiency, Canad. J. Statist. 12 (4) (1984) 303–317.
[18] R. Fontana, R. Gray, J. Kieffer, Asymptotically mean stationary channels, IEEE Trans. Inform. Theory 27 (3) (1981) 308–316.
[19] A. Gallant, Three-stage least-squares estimation for a system of simultaneous, nonlinear, implicit equations, J. Econometrics 5 (1) (1977) 71–88.
[20] A. Gallant, A. Holly, Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation, Econometrica 48 (3) (1980) 697–720.
[21] R. Gray, Probability, Random Processes, and Ergodic Properties, Springer-Verlag, New York, 1988.
[22] R. Gray, J. Kieffer, Asymptotically mean stationary measures, Ann. Probab. 8 (5) (1980) 962–973.
[23] P. Halmos, Lectures on ergodic theory, in: Publications of the Mathematical Society of Japan, vol. 3, The Mathematical Society of Japan, 1956.
[24] J.-B. Hiriart-Urruty, Contribution à la programmation mathématique : cas déterministe et stochastique, Ph.D. thesis, Université de Clermont-Ferrand II, 1977.
[25] S.T. Jensen, B. Nielsen, On convergence of multivariate Laplace transforms, Statist. Probab. Lett. 33 (2) (1997) 125–128.
[26] O. Kallenberg, Foundations of Modern Probability, Probability and its Applications (New York), Springer-Verlag, New York, 2002.
[27] J. Kieffer, M. Rahe, Markov channels are asymptotically mean stationary, SIAM J. Math. Anal. 12 (3) (1981) 293–305.
[28] L. Korf, R.-B. Wets, Random-lsc functions: an ergodic theorem, Math. Oper. Res. 26 (2) (2001) 421–445.
[29] U. Krengel, Un théorème ergodique pour les processus sur-stationnaires, C. R. Acad. Sci. Paris Sér. A-B 282 (17) (1976) A1019–A1021. Aiii.
[30] U. Krengel, Ergodic theorems, in: de Gruyter Studies in Mathematics, vol. 6, Walter de Gruyter & Co., Berlin, 1985.
[31] M. Kurz, On the structure and diversity of rational beliefs, Econom. Theory 4 (6) (1994) 877–900.
[32] M. Loève, Probability Theory. II, Springer-Verlag, New York, 1978.
[33] T. Nakatsuka, Absorbing process in recursive stochastic equations, J. Appl. Probab. 35 (2) (1998) 418–426.
[34] J. Neveu, Bases mathématiques du calcul des probabilités, Masson et Cie, Éditeurs, Paris, 1964.
[35] A. Nobel, A counterexample concerning uniform ergodic theorems for a class of functions, Statist. Probab. Lett. 24 (2) (1995) 165–168.
[36] K. Petersen, Ergodic theory, in: Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989.
[37] B. Pötscher, I. Prucha, A uniform law of large numbers for dependent and heterogeneous data processes, Econometrica 57 (3) (1989) 675–683.
[38] B. Pötscher, I. Prucha, Dynamic Nonlinear Econometric Models, Springer-Verlag, Berlin, 1997.
[39] D.T. Quah, Empirics for growth and distribution: Stratification, polarization, and convergence clubs, J. Econom. Growth 2 (1997) 27–59.
[40] O. Rechard, Invariant measures for many–one transformations, Duke Math. J. 23 (1956) 477–488.
[41] T. Rolski, Queues with nonstationary inputs, Queueing Syst. Theory Appl. 5 (1–3) (1989) 113–129.
[42] W. Szczotka, Stationary representation of queues. I, Adv. Appl. Probab. 18 (3) (1986) 815–848.
[43] W. Szczotka, Stationary representation of queues. II, Adv. Appl. Probab. 18 (3) (1986) 849–859.
[44] M. Valadier, Conditional expectation and ergodic theorem for a positive integrand, J. Nonlinear Convex Anal. 1 (3) (2000) 233–244.