The Annals of Probability
1980, Vol. 8, No. 5, 962–973

ASYMPTOTICALLY MEAN STATIONARY MEASURES[1]

BY ROBERT M. GRAY AND J. C. KIEFFER

Stanford University and University of Missouri

Numerous properties are developed of measures that are asymptotically mean stationary with respect to a possibly nonsingular and noninvertible measurable transformation on a probability space. In particular, several necessary and sufficient conditions for the measure and transformation to satisfy the ergodic theorem are given, an asymptotic form of the Radon-Nikodym theorem for asymptotically dominated measures is developed, and the asymptotic behavior of the resulting Radon-Nikodym derivatives is described. As an application we prove a Shannon-McMillan-Breiman theorem for the case considered. Several examples are given to illustrate the results.

Introduction. Let $(\Omega, \mathfrak{F})$ be a measurable space and T: $\Omega \to \Omega$ a measurable transformation. If $(A, \mathfrak{B})$ is a measurable space, $(A_i, \mathfrak{B}_i) = (A, \mathfrak{B})$ for all integers i, and $(\Omega, \mathfrak{F})$ is the two-sided sequence space defined by the Cartesian product (Parthasarathy (1972), page 6)

$$\Pi_{i=-\infty}^{\infty}(A_i, \mathfrak{B}_i) \equiv (A^{\infty}, \mathfrak{B}^{\infty}),$$

or $(\Omega, \mathfrak{F})$ is the one-sided sequence space

$$\Pi_{i=0}^{\infty}(A_i, \mathfrak{B}_i) \equiv (A^{+}, \mathfrak{B}^{+}),$$

then T is assumed to be the shift. It is of interest in ergodic theory and its application to information theory to know under what conditions on a probability measure μ on $(\Omega, \mathfrak{F})$ the ergodic theorem and the Shannon-McMillan theorem will hold for T.

In information theory the mathematical model of a source is usually a one-sided random process $\{X_n\}_{n=0}^{\infty}$ or a two-sided random process $\{X_n\}_{n=-\infty}^{\infty}$ defined on some probability space. Hence when treating processes we will focus on the appropriate sequence space with the distribution rather than the underlying probability space and T will be the shift. The basic coding theorems of information theory are applications of the ergodic and Shannon-McMillan theorems and it is customary in the information theory literature to assume that the process distribution μ is stationary ($\mu(T^{-1}F) = \mu(F)$, all $F \in \mathfrak{F}$) in order to invoke these results. This typically involves placing restrictions on the allowed sources, coders, and channels in order to ensure that all sources arising within a communications system are stationary. As will be detailed later, several important models arising in information theory are not stationary. Examples are the outputs of codes mapping blocks of a source into encoded blocks of fixed or variable length, the output of a

[1]Research supported by NSF Grant ENG 76 02276A1.
AMS 1970 *subject classifications.* Primary 28A65; secondary 94A15, 60G10.
Key words and phrases. Asymptotically mean stationary, ergodic theorems, Shannon-McMillan-Breiman theorem.

finite state machine driven by a stationary one-sided source, and a stationary source when given side information in the form of conditioning.

More general conditions under which the ergodic and Shannon-McMillan theorems hold under additional assumptions on T have appeared in the literature. Dowker (1951) showed that if T is invertible and nonsingular (T is nonsingular if $\mu T^{-1} \ll \mu$) then the ergodic theorem holds if and only if

$$\text{(1)} \qquad \lim_{n\to\infty} n^{-1}\Sigma_{i=0}^{n-1}\mu(T^{-i}F) \text{ exists, all } F \in \mathcal{F},$$

in which case it follows from the Vitali-Hahn-Saks theorem that

$$\text{(2)} \qquad \bar{\mu}(F) = \lim_{n\to\infty} n^{-1}\Sigma_{i=0}^{n-1}\mu(T^{-1}F), \qquad F \in \mathcal{F}$$

is a probability measure. Since $\bar{\mu}$ is obviously stationary, (2) can be interpreted as saying that the measures μT^{-n} converge to the stationary measure $\bar{\mu}$ in an arithmetic mean or Cesaro mean sense. Hence we define a probability measure μ to be *asymptotically mean stationary* with respect to T if (1) holds and we call $\bar{\mu}$ *stationary mean* of $\bar{\mu}$. If T is clear from context we merely say that μ is a.m.s. A process is said to be a.m.s. if its distribution is a.m.s. If T is invertible, it follows easily that $\bar{\mu} \gg \mu$ and that μ is a.m.s. iff there exists a stationary measure η such that $\eta \gg \mu$.

Rechard (1956) generalized these ideas to nonsingular transformations that need not be invertible (such as the one-sided shift). It follows from his results that the ergodic theorem is satisfied iff μ is a.m.s. It need no longer be true, however, that $\bar{\mu} \gg \mu$. Rechard replaced this condition by an asymptotic form: a measure η on $\mathcal{F}$ is said to *asymptotically dominate* (or be asymptotically stronger than) a measure μ on $\mathcal{F}$ (with respect to T) if $F \in \mathcal{F}$ and $\eta(F) = 0$ implies that

$$\lim_{n\to\infty} \mu(T^{-n}F) = 0.$$

Rechard effectively showed that a measure μ is a.m.s. if and only if it is asymptotically dominated by a stationary measure.

Halmos (1966) argues there is no loss of generality in assuming a transformation nonsingular since similar conclusions can be drawn by replacing the original measure μ by a measure μ_0 such that T is nonsingular with respect to μ_0 and $\mu_0 \gg \mu$. This idea could be used to prove the ergodic theorem holds iff μ is a.m.s. for general measurable T without the nonsingular assumption (although this has not actually been done to our knowledge). In addition, it would likely yield generalizations of some of Rechard's other properties of a.m.s. measures to possibly singular transformations. This approach has two drawbacks. First, direct proofs without the nonsingular assumption are simpler in several cases. For example, the proof that the ergodic theorem holds iff a measure is a.m.s. is easy in the general case. Secondly, singular measures arise naturally in information theory in the study of sources that can be synchronized, that is, where an observer can find with high probability where code blocks begin by observing a long run of data (Nedoma (1964); Dobrushin (1967); Gray, Ornstein, and Dobrushin (1980); Kieffer (1977)).

When developing properties of such sources it is of interest that they describe the source itself and not an artificial dominating source. That is, the source is inherently singular and should be treated as such.

The Shannon-McMillan theorem has not been shown to hold under conditions as general as those for the ergodic theorem. Jacobs (1959) proved that if η satisfies the Shannon-McMillan theorem for T and $\eta \gg \mu$, then μ also satisfies the theorem. This immediately implies that if μ is a.m.s. and T is invertible, then the Shannon-McMillan theorem holds for T since $\mu \ll \bar{\mu}$ and $\bar{\mu}$ is stationary. If T is not invertible, however, then $\bar{\mu}$ only asymptotically dominates μ and the Shannon-McMillan theorem has not previously been proved for this case.

It is the purpose of this paper to give a survey of the properties of a.m.s. measures with respect to general measurable transformations which are not assumed invertible or nonsingular. Some of these results are generalizations of results of Dowker (1951) and Rechard (1956) and many are new. These properties are here used to prove a Shannon-McMillan-Breiman and Shannon-McMillan theorem for measures a.m.s. with respect to a general measurable transformation. They are also intended to serve as a basis for further applications.

The ergodic theorem. Let $(\Omega, \mathcal{F})$ be a measurable space and $T: \Omega \to \Omega$ a measurable transformation. If μ is a.m.s. with stationary mean, then it is clear from (2) that $\bar{\mu}(F) = \mu(F)$ for every T-invariant set $F \in \mathcal{F}$. This observation yields a simple proof of an ergodic theorem for μ.

THEOREM 1. *Let μ be a probability measure on $\mathcal{F}$. Then μ is* a.m.s. *if and only if for every bounded measurable $f: \Omega \to (-\infty, \infty)$ $\{n^{-1}\sum_{i=0}^{n-1} f T^i\}$ converges* a.e. $[\mu]$ *as $n \to \infty$.*

PROOF. If μ is a.m.s., the set on which $\{n^{-1}\sum_{i=0}^{n-1} f T^i\}$ converges is T-invariant and has $\bar{\mu}$-measure 1, and therefore has μ-measure 1. Conversely the convergence of the averages implies that μ is a.m.s.—just take f as an indicator function and integrate the sequence of averages.

Asymptotic dominance. Fix $(\Omega, \mathcal{F})$ and $T: \Omega \to \Omega$ for the rest of this section. It is sometimes difficult to determine if a measure μ on $\mathcal{F}$ is a.m.s. by using the definition. An alternate approach is provided by the concept of asymptotic dominance.

THEOREM 2. *If μ, η are probability measures on $\mathcal{F}$, where η is stationary and asymptotically dominates μ, then μ is* a.m.s.

PROOF. The stated condition implies an ergodic theorem (see the proof of Theorem 1) and hence the result follows from Theorem 1.

Since the concept of asymptotic dominance proves useful in determining whether a measure is a.m.s., it is helpful to know some conditions which imply and/or are implied by asymptotic dominance. Theorems 3 and 4 give a few conditions of this type.

First, define $\mathcal{F}_\infty = \cap_{n \geq 0} T^{-n}\mathcal{F}$. Note that if T is invertible, then $\mathcal{F}_\infty = \mathcal{F}$ and several of the following results are trivial. Note also that if Ω is a one-sided sequence space, then $\mathcal{F}_\infty$ is the tail σ-field $\cap_{n>0}\sigma(X_n, X_{n+1}, \cdots)$, where $\{X_i\}_{i=0}^\infty$ are the coordinate mappings.

THEOREM 3. *Let μ, η be probability measures on $\mathcal{F}$, where η is stationary. The following are equivalent*:

(a) *η asymptotically dominates μ;*

(b) *If $F \in \mathcal{F}$ is T-invariant and $\eta(F) = 0$, then $\mu(F) = 0$;*

(c) *If $F \in \mathcal{F}_\infty$ and $\eta(F) = 0$, then $\mu(F) = 0$.*

PROOF.

(c) $\Rightarrow$ (b) Immediate.

(b) $\Rightarrow$ (a) If $\eta(F) = 0$, then $\limsup_{n\to\infty} T^{-n}F$ is T-invariant and has η-measure 0. Hence,

$$\limsup_{n\to\infty} \mu(T^{-n}F) \leq \mu(\limsup_{n\to\infty} T^{-n}F) = 0.$$

(a) $\Rightarrow$ (c) This implication is Corollary 1 of the next section.

REMARK. Note that the equivalence of (a) and (b) implies that if μ is a.m.s., then $\bar{\mu}$ asymptotically dominates μ.

THEOREM 4. *Let μ, η be probability measures on $\mathcal{F}$, where η is invariant.*

(a)*$\mu \ll \eta$ implies η asymptotically dominates μ;*

(b)*If T is invertible and η asymptotically dominates μ, then $\mu \ll \eta$.*

PROOF.

(a) $\mu \ll \eta$ and $\eta(F) = \eta(T^{-n}F) = 0$, all n, then also $\mu(T^{-n}F) = 0$ whence $\lim_{n\to\infty} \mu(T^{-n}F) = 0$.

(b) If n asymptotically dominates μ and $\eta(F) = 0$, then $\cup_{n=-\infty}^{\infty} T^n F \in \mathcal{F}$ is T-invariant and has η measure zero and hence also μ measure zero from Theorem 3. Thus

$$\mu(F) \leq \mu(\cup_{n=-\infty}^{\infty} T^n F) = 0.$$

REMARK. Theorem 4 is a generalization of Theorem 1 of Rechard (1956) for nonsingular transformations. Rechard's proof utilizes the nonsingularity of T and is longer.

Thus asymptotic dominance and ordinary dominance are equivalent in the invertible case and $\bar{\mu} \gg \mu$.

Some asymptotic convergence theorems. Fix in this section $(\Omega, \mathcal{F})$, T, and probability measures μ, η on $\mathcal{F}$. If λ is a probability measure on $\mathcal{F}$, let λ_a denote the part of λ absolutely continuous with respect to η (that is, λ_a is the unique positive measure such that $\lambda_a \ll \eta$ and $\lambda - \lambda_a \perp \eta$, its existence being guaranteed by the Lebesgue decomposition theorem).

If μ is a.m.s. and T is invertible, then $\mu \ll \bar{\mu}$ and hence from the Radon-Nikodym theorem $\mu(F) = \int_F (d\mu/d\bar{\mu})d\bar{\mu}$. The Radon-Nikodym theorem is useful in studying a.m.s. sources with invertible shifts. In particular, Jacobs (1959) shows that if η has a Shannon-McMillan theorem and $\eta \gg \mu$, then μ also has a Shannon-McMillan theorem. Thus the Shannon-McMillan theorem is valid for all a.m.s. measures if T is invertible, e.g., for two-sided a.m.s. sources. The following theorem provides an asymptotic form of the Radon-Nikodym theorem for asymptotically dominated measures. This result is used in the next section to generalize Jacob's (1959) result to asymptotic dominance.

THEOREM 5. *Define*

$$f_n = \frac{d(\mu T^{-n})a}{d\eta}, \qquad n = 0, 1, \cdots.$$

If η asymptotically dominates μ, then

$$\lim_{n\to\infty}\{\sup_{F\in\mathcal{F}}|\mu(T^{-n}F) - \int_F f_n\,d\eta|\} = 0.$$

PROOF. From the Lebesgue decomposition theorem and the Radon-Nikodym theorem for each $n = 0, 1, 2, \cdots$ there exists a $B_n \in \mathcal{F}$ such that

$$\text{(3)} \qquad \mu T^{-n}(F) = \mu T^{-n}(F \cap B_n) + \int_F f_n\,d\eta, \qquad F \in \mathcal{F},$$
$$\mu(B_n) = 0.$$

Define $B = \cup_{n=0}^{\infty} B_n$ and we have that $\eta(B) = 0$ and hence by assumption

$$0 \leqslant \mu(T^{-n}F) - \int_F f_n\,d\eta = \mu(T^{-n}(F \cap B_n))$$
$$\leqslant \mu(T^{-n}(F \cap B)) \leqslant \mu(T^{-n}B) \to_{n\to\infty} 0.$$

Since the bound is uniform over F, the theorem is proved.

The remainder of this section is devoted to further developing the properties of the f_n. Henceforth if λ is a probability measure on $\mathcal{F}$, let λ_∞ denote the restriction of λ to $\mathcal{F}_\infty$.

COROLLARY 1. *If η is stationary and asymptotically dominates μ, then $\mu_\infty \ll \eta_\infty$.*

PROOF. Let $F \in \mathcal{F}_\infty$ satisfy $\eta(F) = 0$. Find $\{F_n\}$ so that $T^{-n}F_n = F$, $n = 1, 2, \cdots$. By Theorem 2, $\mu(F) - \int_{F_n} f_n\,d\eta \to 0$. But by stationarity of η, $\int_{F_n} f_n\,d\eta = \int_F f_n T^n\,d\eta = 0$, hence $\mu(F) = 0$.

THEOREM 6. *Let η be stationary and asymptotically dominate μ. Then*

$$f_n T^n \to_{n\to\infty} \frac{d\mu_\infty}{d\eta_\infty} \text{ a.e. } [\eta] \textit{ and } L^1(\eta).$$

PROOF. Define $f_\infty = d\mu_\infty/d\eta_\infty$ and $\mathcal{F}_n = T^{-n}\mathcal{F}$. All conditional expectations below are with respect to η. For any m

$$\text{(4)} \quad \int|f_\infty - f_n T^n|\,d\eta \leqslant \int|f_\infty - E(f_m T^m|\mathcal{F}_\infty)|\,d\eta + \int|E(f_m T^m|\mathcal{F}_\infty)$$
$$- E(f_m T^m|\mathcal{F}_n)|\,d\eta + \int|E(f_m T^m|\mathcal{F}_n) - f_n T^n|\,d\eta.$$

For all $F \in \mathscr{F}_\infty$ we have from (3) and iterated expectation that

$$\mu(F) = \int_F f_\infty \, d\eta \geqslant \int_F f_m T^m \, d\eta = \int_F E(f_m T^m | \mathscr{F}_\infty) \, d\eta$$

and hence since f_∞ and $E(f_m T^m | \mathscr{F}_\infty)$ are both measurable with respect to $\mathscr{F}_\infty$ we have from Ash (1972), page 49, that $f_\infty \geqslant E(f_m T^m | \mathscr{F}_\infty)$ a.e. $[\eta]$. Thus from Theorem 5

$$(5) \qquad \int |f_\infty - E(f_m T^m | \mathscr{F}_\infty)| \, d\eta = \int f_\infty \, d\eta - \int E(f_m T^m | \mathscr{F}_\infty) \, d\eta = 1 - \int f_m T^m \, d\eta \to_{m\to\infty} 0.$$

Therefore given $\varepsilon > 0$ we can choose M so large that $m \geqslant M$ implies

$$(6) \qquad \int |f_\infty - E(f_m T^m | \mathscr{F}_\infty)| \, d\eta \leqslant \varepsilon/2.$$

For fixed m we have from the $L^1(\eta)$ convergence of conditional expectation given decreasing σ-fields (e.g., Ash (1972), page 299) that the middle term on the right of (4) goes to zero as $n \to \infty$. From the proof of Theorem 5 and the facts that $T^{-n}B_n \in \mathscr{F}_n, \eta(B_n) = \eta(T^{-n}B_n) = 0$, we have for $n \geqslant m$ and all $F \in \mathscr{F}_n \subset \mathscr{F}_m$ that

$$\begin{aligned}\int_F E(f_m T^m | \mathscr{F}_n) \, d\eta &= \int_F f_m T^m \, d\eta = \int_{F \cap (T^{-n}B_n)^c} f_m T^m \, d\eta \leqslant \mu\big(F \cap (T^{-n}B_m)^c\big) \\ &= \int_{F \cap (T^{-n}B_n)^c} f_n T^n \, d\eta + \mu\big(F \cap (T^{-n}B_n)^c \cap T^{-n}B_n\big) \\ &= \int_F f_n T^n \, d\eta.\end{aligned}$$

Since $f_n T^n$ and $E(f_m T^m | \mathscr{F}_n)$ are both measurable with respect to $\mathscr{F}_n$, this implies from Ash (1972), page 49, that $E(f_m T^m | \mathscr{F}_n) \leqslant f_n T^n$ a.e. $[\eta]$ and hence from (6)

$$\begin{aligned}(7) \quad \int |E(f_m T^m | \mathscr{F}_n) - f_n T^n| \, d\eta &= \int f_n T^n \, d\eta - \int E(f_m T^m | \mathscr{F}_n) \, d\eta \\ &= \int f_n T^n \, d\eta - \int f_m T^m \, d\eta \to_{n\to\infty} 1 - \int f_m T^m \, d\eta \leqslant \varepsilon/2.\end{aligned}$$

Thus from (4), (6) and (7)

$$\limsup_{n\to\infty} \int |f_\infty - f_n T^n| \eta \leqslant \varepsilon$$

proving $L^1(\eta)$ convergence since ε is arbitrary. For any measure λ on $\mathscr{F}$, let λ_n denote its restriction to $\mathscr{F}_n$. From (3) we see that $f_n T^n = d(\mu_n)_a / d\eta_n$ for each n. It is an easy consequence of the Lebesgue decomposition theorem that $f_{n+1} T^{n+1}$ is the maximal $\mathscr{F}_{n+1}$- measurable function g for which $\mu(F) \geqslant \int_F g \, d\eta$, $F \in \mathscr{F}_{n+1}$. That is, if g is $\mathscr{F}_{n+1}$-measurable and $\mu(F) \geqslant \int_F g \, d\eta$, $F \in \mathscr{F}_{n+1}$, then $g \leqslant f_{n+1} T^{n+1}$ a.e. $[\eta]$. Now by (3),

$$\mu(F) \geqslant \int_F f_n T^n \, d\eta = \int_F E(f_n T^n | \mathscr{F}_{n+1}) \, d\eta, \qquad F \in \mathscr{F}_{n+1}.$$

Hence by the preceding remarks, $E(f_n T^n | \mathscr{F}_{n+1}) \leqslant f_{n+1} T^{n+1}$ a.e. $[\eta]$. This means that $\{f_n T^n\}_0^\infty$ is a reverse supermartingale so by a supermartingale convergence theorem (Neveu (1975), pages 115-119), $\{f_n T^n\}$ converges a.e. $[\eta]$.

COROLLARY 2. *If η is stationary and asymptotically dominates μ, then*

$$\lim_{n\to\infty} \sup_{F \in \mathscr{F}} |\mu(T^{-n}F) - \int_{T^{-n}F} f_\infty \, d\eta| = 0,$$

where $f_\infty = d\mu_\infty / d\eta_\infty$.

PROOF. For any m and F

$$\sup_{F\in\mathcal{F}}|\mu(T^{-n}F) - \int_{T^{-n}F} f_\infty\, d\eta| \leqslant \sup_{F\in\mathcal{F}}|\mu(T^{-n}F) - \int_{T^{-n}F} f_n T^n\, d\eta| + \sup_{F\in\mathcal{F}}|\int_{T^{-n}F} f_n T^n\, d\eta - \int_{T^{-n}F} f_\infty\, d\eta|.$$

The leftmost term on the right goes to zero as $n\to\infty$ from Theorem 5 and stationarity of η. The remaining term is bound above using Theorem 6 by

$$\sup_{F\in\mathcal{F}}\int_{T^{-n}F}|f_n T^n - f_\infty|\, d\eta \leqslant \int_\Omega |f_n T^n - f_\infty|\, d\eta \to_{n\to\infty} 0,$$

completing the proof.

The next corollary provides an (ε, δ) version of absolute dominance and generalizes Theorem 2 of Rechard (1956) to possibly nonsingular transformations.

COROLLARY 3. *If η is stationary and asymptotically dominates μ, then given $\varepsilon > 0$ there is a $\delta > 0$ and an integer N such that if $\eta(F) \leqslant \delta$, then $\mu(T^{-n}F) \leqslant \varepsilon$ for all $n \geqslant N$.*

PROOF. From the previous corollary given $\varepsilon > 0$ there is an N such that for $n \geqslant N, F \in \mathcal{F}$

$$\mu(T^{-n}F) \leqslant \int_{T^{-n}F} f_\infty\, d\eta + \varepsilon/3$$

and hence for any $r > 0$

$$\mu(T^{-n}F) \leqslant r\eta(F) + \int_{f_\infty > r} f_\infty\, d\eta + \varepsilon/3.$$

Since $f_\infty \in L^1(\eta)$ we can choose r so large that

$$\int_{f_\infty > r} f_\infty\, d\eta \leqslant \varepsilon/3$$

and hence $\delta = \varepsilon/3r$ completes the proof.

THEOREM 7. *Let η be stationary and asymptotically dominate μ, then*

$$n^{-1}\sum_{i=0}^{n-1} f_i \to_{n\to\infty} \frac{d\bar{\mu}}{d\eta} \text{ in } L^1(\eta) \text{ norm}.$$

Thus if $\eta = \bar{\mu}$,

$$n^{-1}\sum_{i=0}^{n-1} f_i \to_{n\to\infty} 1 \text{ in } L^1(\bar{\mu}) \text{ norm}.$$

PROOF. For $n = 0, 1, \cdots$ define the operator $\hat{T}_n : L^1(\Omega, \mathcal{F}, \eta) \to L^1(\Omega, \mathcal{F}_n, \eta)$ by $\hat{T}_n f = f T^n$. Then $\hat{T}_n$ is a norm-preserving isomorphism. Let U_n be the inverse of $\hat{T}_n$ and note that

$$U_n(U_m f) = U_{n+m} f, \qquad f \in L^1(\Omega, \mathcal{F}_\infty, \eta). \tag{8}$$

By Theorem 6 $f_n T^n \to f_\infty$ in $L^1(\eta)$ norm, hence since each U_n preserves $L^1(\eta)$ norm, $U_n(f_n T^n) - U_n f_\infty \to 0$ in $L^1(\eta)$ norm. Thus, since $U_n(f_n T^n) = f_n$, $n^{-1}\sum_{i=0}^{n-1} f_i$ converges in $L^1(\eta)$ norm if $n^{-1}\sum_{i=0}^{n-1} U_i f_\infty$ does. From (8) we have that $U_i f_\infty = U_1^i f_\infty$.

Since

(a) $U_1[L^1(\Omega, \mathcal{F}_\infty, \eta)] \subset L^1(\Omega, \mathcal{F}_\infty, \eta)$;

(b) $U_1[L^2(\Omega, \mathcal{F}_\infty, \eta)] \subset L^2(\Omega, \mathcal{F}_\infty, \eta)$;

(c) U_1 preserves the $L^1(\eta)$ and $L^2(\eta)$ norms, it follows from Jacobs (1962), page 94, that $n^{-1}\Sigma_{i=0}^{n-1}U_1^i f_\infty$ converges in $L^1(\eta)$ norm. By Theorem 5, if $F \in \mathcal{F}$, then

$$\int_F\{n^{-1}\Sigma_{i=0}^{n-1}f_i\}\,d\eta - n^{-1}\Sigma_{i=0}^{n-1}\mu(T^{-i}F) \to_{n\to\infty} 0.$$

Thus if g is the $L^1(\eta)$ limit of $\{n^{-1}\Sigma_{i=0}^{n-1}f_i\}$, then $\int_F g\,d\eta = \bar{\mu}(F)$, $F \in \mathcal{F}$, and so $g = d\bar{\mu}/d\eta$.

The Shannon-McMillan-Breiman theorem. In this section we prove the Shannon-McMillan theorem (L^1 convergence) and the Shannon-McMillan-Breiman theorem (a.e. convergence) for a one-sided a.m.s. process with a finite state space. (These results are immediate for 2-sided a.m.s. processes by the remark following Theorem 4 and Jacobs (1959) result.) As pointed out in Corollary 4, we can and will assume our process $\{X_n\}_0^\infty$ consists of the coordinate mappings from $A^+ \to A$, where A is a fixed finite set.

If U, V are discrete measurable functions defined on A^+, and λ is probability measure on $(A^+, \mathcal{B}^+)$, let $\lambda(U)$ and $\lambda(U|V)$ be the functions on A^+ such that

$$\begin{aligned} \lambda(U)(x) &= \lambda[U = U(x)], && x \in A^+ \\ \lambda(U|V) &= \lambda(U,V)/\lambda(V), && \lambda(V) > 0 \\ &= 0, && \text{elsewhere.} \end{aligned}$$

LEMMA. *Let η be stationary and asymptotically dominate μ. If $f: A^+ \to (-\infty, \infty)$ is a measurable tail function of $\{X_i\}$ (measurable with respect to $\mathcal{F}_\infty$), $f_n: A^+ \to (-\infty, \infty)$, $n = 1, 2, 3, \cdots$ is a sequence of measurable functions, and $k_n \to \infty$, then*

$$f_n(X_{k_n}, X_{k_n+1}, \cdots) \to f \text{ a.e. } [\eta]$$

implies

$$f_n(X_{k_n}, X_{k_n+1}, \cdots) \to f \text{ a.e. } [\mu].$$

PROOF. Assume $k_n \to \infty$ and $f_n(X_{k_n}, X_{k_n+1}, \cdots) \to f$ a.e. $[\eta]$. Then for each $\varepsilon > 0$, $\lim_{n\to\infty}\eta(E_n(\varepsilon)) = 0$, where

$$E_n(\varepsilon) = \cup_{m=n}^\infty\{|f_m(X_{k_m}, X_{k_m+1}, \cdots) - f| > \varepsilon\}.$$

By Corollary 3, it follows that $\lim_{n\to\infty}\mu(E_n(\varepsilon)) = 0$, and so

$$f_n(X_{k_n}, X_{k_n+1}, \cdots) \to f \text{ a.e. } [\mu].$$

THEOREM 8. *Let η be stationary and asymptotically dominate μ. Let h be any tail function such that*

$$\log\frac{\eta(X_1^n)}{n} \to h \text{ a.e. } [\eta],$$

then also

$$\log \frac{\mu(X_1^n)}{n} \to h \text{ a.e.} [\mu] \textit{ and in } L^1(\mu).$$

PROOF. From Jacobs (1959), $n^{-1}\log\mu(X_1^n)$ is uniformly μ-integrable and so $L^1(\mu)$ convergence will follow from a.e. convergence. First assume that $\mu(X_m^n = x) > 0, \eta(X_m^n = x) > 0$, all $m, n, x \in A^{n-m-1}$. Let k be a positive integer. If $n > k$, $\log \mu(X_1^n) = \log \mu(X_k^n) + \log \mu(X_1^{k-1}|X_k^n)$. From Parry (1969), page 14, $\sup_n|\log\mu(X_1^{k-1}|X_k^n)| \in L^1(\mu)$ whence

$$n^{-1}\log\mu(X_1^n) - n^{-1}\log\mu(X_k^n) \to_{n\to\infty} 0 \text{ a.e.} [\mu].$$

The same holds true for μ replaced by η. Hence if $k_n \to \infty$ slowly enough,

$$n^{-1}\log\mu(X_1^n) - n^{-1}\log\mu(X_{k_n}^n) \to 0 \text{ a.e.} [\mu] \tag{9}$$

$$n^{-1}\log\eta(X_1^n) - n^{-1}\log\eta(X_{k_n}^n) \to 0 \text{ a.e.} [\eta]. \tag{10}$$

If $\log = \log_2$, then

$$\eta\left(n^{-1}\log\{\mu(X_{k_n}^n)/\eta(X_{k_n}^n)\} \geqslant \varepsilon\right) \leqslant 2^{-n\varepsilon}$$

hence

$$\left(n^{-1}\log\mu(X_{k_n}^n) - n^{-1}\log\eta(X_{k_n}^n)\right)_+ \to 0 \text{ a.e.} [\eta]$$

and reversing roles yields

$$\left(n^{-1}\log\eta(X_{k_n}^n) - n^{-1}\log\mu(X_{k_n}^n)\right)_+ \to 0 \text{ a.e.} [\mu].$$

Applying the lemma,

$$n^{-1}\log\eta(X_{k_n}^n) - n^{-1}\log\mu(X_{k_n}^n) \to 0 \text{ a.e.} [\eta]. \tag{11}$$

From (10)

$$n^{-1}\log\eta(X_{k_n}^n) \to h \text{ a.e.} [\eta].$$

From the lemma,

$$n^{-1}\log\eta(X_{k_n}^n) \to h \text{ a.e.} [\mu].$$

From (11),

$$n^{-1}\log\mu(X_{k_n}^n) \to h \text{ a.e.} [\mu],$$

and hence from (9)

$$n^{-1}\log\mu(X_1^n) \to h \text{ a.e.} [\mu].$$

If μ, η are not strictly positive, pick λ stationary such that $\lambda(X_m^n = x) > 0$ for all m, n, x. Let $\mu' = (\mu + \lambda)/2$, $\eta' = (\eta + \lambda)/2$. Picking h' a tail function such that $+n^{-1}\log\eta_1'(X_1^n) \to h'$ a.e. $[\mu']$, we have by the preceding argument that $n^{-1}\log\mu'(X_1^n) \to h'$ a.e. $[\mu']$ and therefore since $\mu \ll \mu'$, $n^{-1}\log\mu(X_1^n) \to h'$ a.e. $[\mu]$ by the Jacobs (1959) result. Since $\eta \ll \eta'$, we again have by Jacobs (1959) that $n^{-1}\log\eta(X_1^n) \to h'$ a.e. $[\eta]$ and so $h = h'$ a.e. $[\eta]$. But h, h' are tail functions so by Theorem 3(c), $h = h'$ a.e. $[\mu]$.

COROLLARY 4. *The Shannon-McMillan-Breiman and Shannon-McMillan theorems hold for any one-sided* a.m.s. *processes with finite-state space.*

PROOF. If two processes have the same distribution, it is easy to see that the Shannon-McMillan theorem holds for one of the processes if and only if it holds for the other. Thus Corollary 4 follows from Theorem 8 and the observation that if $\{Y_n\}_0^\infty$ is an A-valued a.m.s. process with joint distribution μ, the process $\{X_n\}_0^\infty$ on $(A^+, \mathfrak{B}^+, \mu)$ also has joint distribution μ.

Examples. In this section several nonstationary yet a.m.s. measures are described.

EXAMPLE 1. If $T: \Omega \to \Omega$ is measurable and μ, η are probability measures on $(\Omega, \mathfrak{F})$ such that η is stationary and $\mu \ll \eta$, then μ is a.m.s. from Theorem 1. From the Radon-Nikodym theorem $\mu \ll \eta$ iff there is a measurable $f: \Omega \to [0, \infty)$ such that $\mu(F) = \int_F f\, d\eta$, $F \in \mathfrak{F}$. Thus by choosing different f one can generate many a.m.s. nonstationary μ.

EXAMPLE 2. If η is stationary and $\eta(G) > 0$, $G \in \mathfrak{F}$, then $\mu(F) = \eta(F \cap G)/\eta(G)$ is a.m.s. as in Example 1 since $\mu \ll \eta$. Furthermore, if η is a.m.s. and $\eta(G) > 0$, then $\mu(F) = \eta(F \cap G)/\eta(G)$ is a.m.s. since if $\bar{\eta}(F) = 0$, then as in the proof of Theorem 3, $\eta(\limsup_{n\to\infty} T^{-n}F) = 0$ hence $\mu(\limsup_{n\to\infty} T^{-n}F) = 0$ hence $\lim_{n\to\infty} \mu(T^{-n}F) = 0$ and μ is a.m.s. from Theorem 2. Thus the a.m.s. property is not lost by conditioning while stationarity is, in general, lost by conditioning.

EXAMPLE 3. We give an example of an a.m.s. source which cannot be generated as in Example 1. Let $\Omega = A^+, A = \{0, 1\}$. Let μ be such that $\mu(1, 0, 0, \cdots) = 1$. Then μ is a.m.s. and singular to every stationary η.

Observe that by mixing sources such as in Example 3 with a stationary measure we can model measures that are nonstationary due to transients and hence become stationary in the limit, that is, $\mu(T^{-n}F) \to \mu(F)$ as $n \to \infty$, $F \in \mathfrak{F}$.

EXAMPLE 4. Let $\{X_n\}_{n=0}^\infty$ be a one-sided process with finite alphabet A. A finite-state code consists of a finite state space S, an initial state $s_0 \in S$, and two mappings $f: A \times S \to B, g: A \times S \to S$ yielding an output (encoded) sequence Y_n and state sequence S_n, $n = 1, 2, \cdots$ where

$$\begin{aligned} Y_{n+1} &= f(X_n, S_n) \\ S_{n+1} &= g(X_n, S_n). \end{aligned}$$

Kieffer and Rahe (1979) have shown that if the process $\{X_n\}$ is stationary (or, more generally, a.m.s.), then the joint input/output process $\{X_n, Y_n\}_{n=1}^\infty$ is also a.m.s. Thus, in particular, both the ergodic theorem and Shannon-McMillan theorem hold for the output of a finite state machine driven by a stationary process.

EXAMPLE 5. If μ is stationary with respect to T^N for some integer N (or N-stationary), then μ is easily seen to be a.m.s. with respect to T with stationary mean

$$\bar{\mu}(F) = N^{-1}\Sigma_{i=0}^{N-1}\mu(T^{-i}F).$$

N-stationary processes arise naturally in information theory in the study of block codes, that is, codes mapping nonoverlapping of source N-tuples into encoded N-blocks. Furthermore, if an N-stationary source μ with alphabet A is synchronizable in the sense that given $\varepsilon > 0$ there is for sufficiently large M a function $f: A^M \to \{0, 1, \cdots, N-1\}$ such that

$$\mu T^{-n}\big(f(X_1^M) = n\big) \geqslant 1 - \varepsilon$$

(e.g., μ is a block length N encoded source and the beginning of the blocks can be determined with high probability by observing the outputs for a long time), then T is singular.

EXAMPLE 6. A generalization of N-stationary sources that arises in variable-length coding problems in information theory is a variable length shift T^* defined by

$$T^*\omega = T^{L(\omega)}\omega, \qquad \omega \in \Omega,$$

where $L: \Omega \to \{1, 2, \cdots\}$ is called a length function. If a source measure μ is stationary with respect to T^* we say it is variable-length stationary. For example, if a one-sided stationary source is encoded by mapping a source N-tuple x_1^N into output blocks of variable length $l(x^N)$, then the encoded source is stationary with respect to T^* with $L(\omega) = l(X_1^n(\omega))$. We here show that if μ is stationary with respect to T^*, then it is a.m.s. with respect to T. We assume that $EL < \infty$ and define the probability measure

$$\tilde{\mu}(f) = (EL)^{-1}\Sigma_{k=1}^{\infty}\Sigma_{i=0}^{k-1}\mu(T^{-i}F \cap L^{-1}(k)),$$

where $L^{-1}(k) = \{\omega : L(\omega) = k\}$. We have that

$$T^{*-1}F = \cup_{k=1}^{\infty}(T^{-k}F) \cap L^{-1}(k),$$

hence

$$\mu(T^{*-1}F) = \Sigma_{k=1}^{\infty}\mu((T^{-k}F) \cap L^{-1}(k)) = \mu(F)$$
$$= \Sigma_{k=1}^{\infty}\mu(F \cap L^{-1}(k)).$$

We therefore have

$$\tilde{\mu}(T^{-1}F) = (EL)^{-1}\Sigma_{k=1}^{\infty}\Sigma_{i=0}^{k-1}\mu(T^{-i-1}F \cap L^{-1}(k))$$
$$= \tilde{\mu}(F) + (EL)^{-1}\{\Sigma_{k=1}^{\infty}\mu(T^{-k}F \cap L^{-1}(k))$$
$$- \Sigma_{k=1}^{\infty}\mu(F \cap L^{-1}(k))\}$$
$$= \tilde{\mu}(F)$$

and hence $\tilde{\mu}$ is stationary with respect to T. By construction $\tilde{\mu} \gg \mu$ and hence μ is a.m.s. with respect to T from Theorems 4 and 3. Intuitively, the measure $\tilde{\mu}$ corresponds to the distribution of a process obtained from $\{X_n\}$ "randomizing" the choice of zero time, or, equivalently, inserting a random phase. From this point of view the fact that $\tilde{\mu}$ is stationary is equivalent to Theorem 3 of Cariolaro and Pierobon (1977) for the case of a bounded length function of the form $L(\omega) = l(X_1^N(\omega))$.

REFERENCES

ASH, R. B. (1972). *Real Analysis and Probability*. Academic Press, New York.

CARIOLARO, G. L. and PIEROBON, G. L. (1977). Stationary symbol sequences from variable-length word sequences. *IEEE Trans. Information Theory* **23** 243–253.

DOBRUSHIN, R. L. (1967). Shannon's theorems for channels with synchronization errors. *Problemy Peredaci Informacii* **3** 18–36.

DOWKER, Y. N. (1951). Finite and σ-finite invariant measures. *Ann. of Math.* **54** 595–608.

GRAY, R. M., ORNSTEIN, D. S. and DOBRUSHIN, R. L. (1980). Block synchronization, sliding-block coding, invulnerable sources and zero error codes for discrete noisy channels. *Ann. Probability*. **8** 639–674.

HALMOS, P. R. (1956). *Lectures on Ergodic-Theory*. Chelsea Publishing Co., New York.

JACOBS, K. (1959). Die ubertragung diskreter informationen durch periodische und fastperiodische kavale. *Math. Ann.* **137** 125–135.

JACOBS, K. (1962). *Lecture Notes on Ergodic Theory, Vol.* 1. Matematisk Institut, Aarhus Univ., Denmark.

KIEFFER, J. C. (1977). On sliding-block coding for transmission of a source over a stationary nonanticipatory channel. *Information and Control* **35** 1–19.

KIEFFER, J. C. and RAHE, M. (1979). Markov channels are asymptotically mean stationary. Unpublished manuscript.

NEDOMA, J. (1964). The synchronization for ergodic channels. *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes*. 529–539, Prague.

NEVEU, J. (1975). *Discrete-Parameter Martingales*. American Elsevier Pub. Co., New York.

PARRY, W. (1969). *Entropy and Generators in Ergodic Theory*. W. A. Benjamin Inc., New York.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

RECHARD, O. W. (1956). Invariant measures for many-one transformations. *Duke J. Math.* **23** 477–488.

INFORMATION SYSTEMS LABORATORY
DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

MATHEMATICS DEPARTMENT
UNIVERSITY OF MISSOURI
ROLLA, MISSOURI 65401