Ergod. Th. & Dynam. Sys. (2008), **28**, 447–480
doi: 10.1017/S0143385707001034
© *2008 Cambridge University Press*
Printed in the United Kingdom

Shannon entropy for stationary processes and dynamical systems

D. HAMDAN, W. PARRY† and J.-P. THOUVENOT

Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Universités Paris 6 et Paris 7, Boîte Courrier 188, 4 Place Jussieu, 75252 Paris cedex 05, France
(e-mail: hamdan@ccr.jussieu.fr)

(*Received* 5 *June* 2007 *and accepted in revised form* 23 *November* 2007)

Abstract. We consider stationary ergodic processes indexed by $\mathbb{Z}$ or $\mathbb{Z}^n$ whose finite-dimensional marginals have laws which are absolutely continuous with respect to Lebesgue measure. We define an entropy theory for these continuous processes, prove an analogue of the Shannon–MacMillan–Breiman theorem and study more precisely the particular example of Gaussian processes.

1. *Introduction*

In [**17**] Shannon introduced a general theory of entropy designed to quantify the rate at which information is produced through the evolution of a stationary stochastic process. He gave two definitions: one for random variables assuming discrete values and the other for random variables which assume real values or, more generally, values in $\mathbb{R}^n$. However, these definitions represent but two aspects of a single notion.

In 1958 Kolmogorov adapted Shannon's discrete version in his definition of entropy for dynamical systems which enabled him to solve an important open problem concerning the classification of measure-preserving transformations with continuous (in fact Lebesgue) spectrum. Shortly afterwards Sinai modified and improved Kolmogorov's definition so that today one sometimes speaks of the Kolmogorov–Sinai, KS, invariant.

The KS entropy is also an invariant for stationary stochastic processes in as much as they may be represented as measure-preserving transformations. However, these processes frequently have infinite entropy, particularly when the random variables take their values in a non-discrete space.

For example, all stationary Gaussian processes with absolutely continuous spectrum have the same infinite KS entropy and indeed they are all measure-theoretically isomorphic to each other (Ornstein [**9**]).

† William Parry died on 20 August 2006, at the time when these notes were completed.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

For this reason we feel that a modification of the traditional definition of isomorphism should be considered which distinguishes between various processes with infinite KS entropy. This paper should be regarded as a move toward this end in that we produce an 'invariant' based on Shannon's second version of entropy (designed for continuous valued random variables).

Shannon's entropy was extensively investigated, especially by the Russian school, in the late 1950s and a thorough account appears in Pinsker's book [**14**]. One of our main purposes is to clarify and extend known results in this area.

The invariant investigated here (naturally referred to as Shannon entropy) is simply a normalized limit of entropy for $\mathbb{R}^n$ valued random variables, which we compute for all stationary Gaussian processes in terms of their spectral measures. In general, this entropy is finite even when the KS entropy is infinite. We show how this entropy changes when a stationary Gaussian process is subjected to a linear transformation.

This is a topic closely related to the work of Wiener [**20**] and Kolmogorov [**7**] on linear prediction theory. It is well known that stationary Gaussian processes may be regarded as nonlinear extensions of stationary linear sequences in Hilbert space (i.e. 'wide sense' processes in the language of Doob [**2**]) and entropy theory may be regarded as the prediction theory of nonlinear processes. Wiener was particularly interested in nonlinear prediction and in [**21**] he attempted to prove that all stationary processes satisfying certain mild conditions are isomorphic to independent Gaussian processes. As shown by Rosenblatt [**15**], his proof was flawed and it is interesting to note that his mistake, which concerned the behaviour of decreasing sequences of sigma-algebras, was repeated by Kolmogorov (as shown by Rokhlin) in a different context some years later. We now know through the work of Ornstein and his co-workers that Wiener's claim is actually false.

We shall be concerned almost exclusively with stationary stochastic processes $X = (X_n)$ for which the distribution measure of $(X_0, \ldots, X_n)$ is absolutely continuous with respect to $(n+1)$-dimensional Lebesgue measure for all $n = 0, 1, \ldots$. Our first theorem shows that for any ergodic measure-preserving transformation T of a probability space there exist functions F such that $(F \circ T^n)$ satisfies the above condition. We prove thereafter that an entropy 'à la Kolmogorov' can be defined using the continuous entropy definition of Shannon. We prove that these averages of entropies always converge. The theory extends naturally to $\mathbb{Z}^n$ actions. In this framework, we prove a Shannon–Breiman–MacMillan-type pointwise theorem. However, if the Kolmogorov entropy of the transformation is finite, then the limit above is always $-\infty$. In general, this limit is majorized by $(1/2)\log(2\pi)$ plus one half of the variance of the observable. In particular, the preceding inequality reduces to equality if and only if the process is Gaussian independent (this generalizes to processes a result of Shannon for the case of random variables). We also give a similar characterization for a stationary process to be Markovian and, more generally, to be with memory p. The theory applies naturally to Gaussian processes, for which we give a closed formula for this continuous entropy. It turns out that, in the Gaussian Markovian case, this entropy determines the process and, in the Gaussian case, when finite it determines the process up to unilateral isomorphism.

We also give some relationships with the rate of entropy of Pinsker and with the rate of generation of information as well.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

2. *Absolutely continuous processes based on an ergodic system*

We begin with the following.

Definition 2.1. A real valued discrete time process $X = (X_n)_{n\in I}$, indexed by a countable set I, is said to be absolutely continuous if for every finite subset K of I, with cardinality $|K|$, the joint distribution of $(X_n)_{n\in K}$ is absolutely continuous with respect to the $|K|$-dimensional Lebesgue measure.

This implies that the underlying measure space is non-atomic, a hypothesis which will hold throughout the paper.

We shall only be using $I = \mathbb{Z}$, or $I = \mathbb{Z}^2$.

In this section we prove the following.

PROPOSITION 2.2. *Let (Ω, T, μ) be an invertible ergodic dynamical system. Then there exists $F \in L^2(\mu)$ such that the process $(F \circ T^n)$ is absolutely continuous.*

Proof. Let (Y, S, γ) be the independent Gaussian dynamical system: $Y := \mathbb{R}^{\mathbb{Z}}$, S is the shift transformation $(Sy)_j = y_{j+1}$, for $y \in Y$, $j \in \mathbb{Z}$, and γ is the product measure of the measures γ_j, $j \in \mathbb{Z}$, where, for every j, $\gamma_j = (2\pi)^{-(1/2)} \exp(-(1/2)x^2)\, dl(x)$, l being the Lebesgue measure on $\mathbb{R}$. Let $Y_n : Y \to \mathbb{R}$ be the projection onto the nth coordinate.

According to Dye's theorem, [**3**], there exists an integer-valued measurable function $\tau : Y \to \mathbb{Z}$ such that if $S_1 : Y \to Y$ is the transformation defined by $S_1(y) = S^{\tau(y)}(y)$, for γ-almost all $y \in Y$, then the two dynamical systems (Ω, T, μ) and (Y, S_1, γ) are isomorphic. Let $\theta : \Omega \to Y$ be a map giving the isomorphism: $\theta \circ T = S_1 \circ \theta$ and $\gamma = \mu \circ \theta^{-1}$. Set $F = Y_0 \circ \theta$ and $Z_n = Y_0 \circ S_1^n$, for $n \in \mathbb{Z}$, so that $(F, F \circ T, \ldots, F \circ T^{n-1})$ and $(Z_0, Z_1, \ldots, Z_{n-1})$ have the same law, say α_n. We now show that α_n is absolutely continuous with respect to Lebesgue measure and this ends the proof. To do this, for every $\bar{k} = (k_1, \ldots, k_{n-1}) \in \mathbb{Z}^{n-1}$, put

$$E_{\bar{k}} = \{\tau = k_1, \tau \circ S^{k_1} = k_2, \tau \circ S^{k_1+k_2} = k_3, \ldots, \tau \circ S^{k_1+\cdots+k_{n-2}} = k_{n-1}\},$$

and let $\mathcal{F}_{\bar{k}}$ denote the sigma-algebra generated by $Y_0, Y_{k_1}, \ldots, Y_{k_1+\cdots+k_{n-1}}$. In view of the definition of (Y, S, γ), straightforward computations show then that α_n is absolutely continuous with respect to Lebesgue measure and has the density $g := \sum_{\bar{k}\in Z^n} g_{\bar{k}}$, where

$$\begin{aligned} g_{\bar{k}}(y_0, \ldots, y_{n-1}) = {} & E_\gamma[1_{E_{\bar{k}}} \mid \mathcal{F}_{\bar{k}}]\,(y_0, y_1, \ldots, y_{n-1}) \\ & \times (2\pi)^{-(n/2)} \exp(-\tfrac{1}{2}(y_0^2 + \cdots + y_{n-1}^2)), \end{aligned}$$

and $E_\gamma[1_{E_{\bar{k}}} \mid \mathcal{F}_{\bar{k}}]$ denotes the conditional expectation with respect to γ of $1_{E_{\bar{k}}}$ given the sigma-algebra $\mathcal{F}_{\bar{k}}$. □

Remark 2.3.

(1) The function F in Lemma 2 can be taken (as the proof shows) in the intersection of $\{L^p(\mu) : p \geq 1\}$.

(2) We can show that there exists an F such that, for every n, the law of $(F, F \circ T, \ldots, F \circ T^{n-1})$ is equivalent to the n-dimensional Lebesgue measure.

In the same way we have the following.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

PROPOSITION 2.4. *If T and S are measure-preserving transformations which commute on a probability space $(\Omega, \mathcal{F}, \mu)$, such that the joint action is ergodic, then there exists $F \in L^2(\mu)$ such that the process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{Z}^2}$ is absolutely continuous.*

3. *Notation, a few prerequisites and a lemma*

3.1. *Conditional entropy of probability measures.*

In this subsection we recall various definitions attached to Shannon entropy [**14**].

Definition 3.1. Let P and Q be probability measures defined on the same measurable space $(\Omega, \mathcal{F})$. Let $\mathcal{P}$ be the set of all finite measurable partitions of Ω. If Π is in $\mathcal{P}$ let

$$S_\Pi(P \mid Q) := \sum_{E\in\Pi} P(E) \log\left(\frac{P(E)}{Q(E)}\right). \tag{1}$$

Then, for $\Pi_1 \in \mathcal{P}$, Π_1 finer than Π implies that

$$S_\Pi(P \mid Q) \le S_{\Pi_1}(P \mid Q). \tag{2}$$

The entropy $H_Q(P)$ of P with respect to Q is defined by

$$H_Q(P) := \sup_{\Pi\in\mathcal{P}} S_\Pi(P \mid Q). \tag{3}$$

We list without proofs some results which we are going to use.

THEOREM A. (Gelfand *et al* [**14**]) *Let P, Q be two probability measures on the measurable space $(\Omega, \mathcal{F})$. Then, if the entropy $H_Q(P)$ is finite, P is absolutely continuous with respect to Q and*

$$H_Q(P) = \int \log\left(\frac{dP}{dQ}\right) dP.$$

(In particular, if P is not absolutely continuous with respect to Q, $H_Q(P) = +\infty$.)

This was introduced first by Shannon [**17**], for densities: if $f \in L^1_+(dx)$ is such that $\int_{\mathbb{R}} f(x)\,dx = 1$, Shannon considered $\int_{\mathbb{R}} f(x) \log f(x)\,dx$.

Let ψ be the function defined, for $x > 0$, by

$$\psi(x) = -x \log(x). \tag{4}$$

Remark 3.2.

(1) If $H_Q(P)$ is finite then

$$H_Q(P) = \int \log\left(\frac{dP}{dQ}\right) \times \frac{dP}{dQ}\, dQ = -\int \psi\left(\frac{dP}{dQ}\right) dQ. \tag{5}$$

(2) In particular, $H_Q(P) = 0$ if and only if $P = Q$.

(3) More generally, if P is absolutely continuous with respect to Q then

$$H_Q(P) < \infty \iff \int \log\left(\frac{dP}{dQ}\right) dP < \infty \iff \int (-\psi)\left(\frac{dP}{dQ}\right) dQ < \infty. \tag{6}$$

The following theorem follows from the monotonicity property (2).

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

THEOREM B. (Dobrushin [**14**]) *Let P, Q be probability measures on $(\Omega, \mathcal{F})$, $\mathcal{L}$ an algebra of sets belonging to $\mathcal{F}$, which generates the sigma-algebra $\mathcal{F}$, and let $\mathcal{R}$ be a family of finite partitions of Ω whose elements belong to $\mathcal{L}$. If every partition consisting of sets from $\mathcal{L}$ has a finer partition in $\mathcal{R}$, then*

$$H_Q(P) = \sup_{\Pi \in \mathcal{R}} S_\Pi(P \mid Q).$$

Note that, as remarked by the translator of Pinsker, in Theorem B of Dobrushin, the condition that the elements of the partitions in $\mathcal{R}$ be in $\mathcal{L}$ is not necessary.

3.2. *A lemma.* The following lemma will play an essential role in the rest of the paper.

LEMMA 3.3. *Let $(\Omega_i; \mathcal{F}_i, P_i)$ be a probability space for $i = 1, 2$, and ν a probability measure on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ with marginals ν_1 on $(\Omega_1, \mathcal{F}_1)$ and ν_2 on $(\Omega_2, \mathcal{F}_2)$. Then*

$$H_{P_1 \times P_2}(\nu) = H_{\nu_1 \times \nu_2}(\nu) + H_{P_1}(\nu_1) + H_{P_2}(\nu_2), \tag{7}$$

and in particular

$$H_{P_1 \times P_2}(\nu) \geq H_{P_1}(\nu_1) + H_{P_2}(\nu_2), \tag{8}$$

$$H_{P_1 \times P_2}(\nu) \geq H_{\nu_1 \times \nu_2}(\nu). \tag{9}$$

Proof. By Theorem B of Dobrushin, $H_{P_1 \times P_2}(\nu)$ is given by the supremum, over all finite measurable partitions Π_1 of Ω_1 and Π_2 of Ω_2, of the sums $S_{\Pi_1 \times \Pi_2}(\nu \mid P_1 \times P_2)$.

If ν is not absolutely continuous with respect to $\nu_1 \times \nu_2$, then, by Theorem A, $H_{\nu_1 \times \nu_2}(\nu) = +\infty$, and thus

$$H_{P_1 \times P_2}(\nu) \leq +\infty = H_{\nu_1 \times \nu_2}(\nu) + H_{P_1}(\nu_1) + H_{P_2}(\nu_2).$$

If ν is absolutely continuous with respect to $\nu_1 \times \nu_2$, the equalities

$$\nu(E \times F) \log \frac{\nu(E \times F)}{P_1 \times P_2(E \times F)} = \nu(E \times F) \left[\log \frac{\nu(E \times F)}{\nu_1 \times \nu_2(E \times F)} + \log \frac{\nu_1(E)}{P_1(E)} + \log \frac{\nu_2(F)}{P_2(F)} \right], \quad E \in \Pi_1, F \in \Pi_2,$$

imply

$$S_{\Pi_1 \times \Pi_2}(\nu \mid P_1 \times P_2) = S_{\Pi_1 \times \Pi_2}(\nu \mid \nu_1 \times \nu_2) + S_{\Pi_1}(\nu_1 \mid P_1) + S_{\Pi_2}(\nu_2 \mid P_2), \tag{E_1}$$

and therefore

$$S_{\Pi_1 \times \Pi_2}(\nu \mid P_1 \times P_2) \leq H_{\nu_1 \times \nu_2}(\nu) + H_{P_1}(\nu_1) + H_{P_2}(\nu_2),$$

from which it follows that

$$H_{P_1 \times P_2}(\nu) \leq H_{\nu_1 \times \nu_2}(\nu) + H_{P_1}(\nu_1) + H_{P_2}(\nu_2).$$

To prove the reverse inequality

$$H_{P_1 \times P_2}(\nu) \geq H_{\nu_1 \times \nu_2}(\nu) + H_{P_1}(\nu_1) + H_{P_2}(\nu_2), \tag{E_2}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

consider four arbitrary finite measurable partitions: Π_1, Δ_1 of Ω_1, and Π_2, Δ_2 of Ω_2; note that we can find a finite measurable partition Γ_i of Ω_i refining both Π_i and Δ_i, $i = 1, 2$. Then the partition $\Gamma_1 \times \Gamma_2 := \{M \times N \mid M \in \Gamma_1, N \in \Gamma_2\}$ also refines $\Pi_1 \times \Pi_2$. But, in view of inequality (2), we have, for $i = 1, 2$,

$$S_{\Delta_i}(\nu_i \mid P_i) \le S_{\Gamma_i}(\nu_i \mid P_i).$$

Similarly,

$$S_{\Pi_1\times\Pi_2}(\nu \mid \nu_1 \times \nu_2) \le S_{\Gamma_1\times\Gamma_2}(\nu \mid \nu_1 \times \nu_2).$$

So by summing we get, by (E_1),

$$\begin{aligned} &S_{\Delta_1}(\nu_1 \mid P_1) + S_{\Delta_2}(\nu_2 \mid P_2) + S_{\Pi_1\times\Pi_2}(\nu \mid \nu_1 \times \nu_2) \\ &\quad \le S_{\Gamma_1\times\Gamma_2}(\nu \mid P_1 \times P_2)) \le H_{P_1\times P_2}(\nu), \end{aligned}$$

from which (E_2) follows. This proves (7). As trivially (7) implies (8) and (9), the proof is finished. □

Remark 3.4. The inequality (9) implies that, for fixed ν, the infimum, over all probability measures P_1 and P_2, of $H_{P_1\times P_2}(\nu)$ is attained for $P_1 = \nu_1$ and $P_2 = \nu_2$, and formula (7) shows that it is attained only for these particular values of P_1 and P_2.

COROLLARY 3.5. *Let ν be a probability measure on a product measurable space. If the entropy $H_{P_1\times P_2}(\nu)$ of ν, with respect to a product probability measure $P_1 \times P_2$, is finite, then ν is absolutely continuous with respect to the product $\nu_1 \times \nu_2$ of its marginals, and these marginals are absolutely continuous with respect to P_1 and P_2, respectively.*

4. *Shannon entropy of absolutely continuous processes*

4.1. *Notation.* Let (Ω, T, μ) be an invertible ergodic dynamical system. Let $\mathcal{A} = \{A_1, \ldots, A_k\}$ be a finite partition of Ω. Let m on $\mathcal{A}^{\mathbb{Z}}$ be the product measure with marginals giving equal weights to the atoms of $\mathcal{A}$. Let $F = \sum_{j=0}^{k} a_j 1_{A_j}$ be discrete with $a_i \ne a_j$ for $i \ne j$, and $F_n(x) = (F(x), \ldots, F(T^{n-1}x))$, for $x \in \Omega$. Then if $\mu_n = \mu F_n^{-1}$ and m_n, respectively, are the restrictions of μ and m to $\bigvee_{j=0}^{n-1} T^j\mathcal{A}$, we obtain, with the standard definition of entropy of a partition,

$$H^{\mu}\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}\right) = -\int_{\Omega} \log \frac{d\mu F_n^{-1}}{dm_n} \circ F_n \, d\mu.$$

In this paper we are interested in the case where F is continuous valued, say real valued, with $(F \circ T^n)$ absolutely continuous (cf. Definition 2.1). In this case, denoting l^n the Lebesgue measure on $\mathbb{R}^n$, we consider the function $\mathcal{I}_n = \mathcal{I}_n(F)$ defined by

$$\mathcal{I}_n(x) := -\log\left(\frac{d\mu F_n^{-1}}{dl^n}\right) \circ F_n(x), \quad x \in \Omega, \tag{10}$$

together with its integral

$$H_n = H_n(F) := \int_{\Omega} \mathcal{I}_n \, d\mu. \tag{11}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Then, with ψ as in (4), it follows that

$$H_n(F) = \int_{\mathbb{R}^n} \psi\left(\frac{d\mu F_n^{-1}}{dl^n}\right) dl^n. \tag{12}$$

We focus on the asymptotic behavior of the sequences $(1/n)\mathcal{I}_n$ and $(1/n)H_n$, which we call, respectively, the sequences of *Shannon information* and *Shannon entropy* associated to the process $(F \circ T^n)$.

Let

$$\gamma_0 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dl(x), \quad \gamma_n = \gamma_0^{\otimes n}, \quad n \geq 1. \tag{13}$$

The following quantities are closely related to $\mathcal{I}_n$ and H_n:

$$\mathcal{I}_{n,G} = \mathcal{I}_{n,G}(F) = -\log \frac{d\mu F_n^{-1}}{d\gamma_n} \circ F_n \quad \left(H_{n,G} = H_{n,G}(F) = \int \mathcal{I}_{n,G}(F)\, d\mu\right), \tag{14}$$

$$\mathcal{I}_{n,PM} = \mathcal{I}_{n,PM}(F) = -\log \frac{d\mu F_n^{-1}}{d(\mu F^{-1})^{\otimes n}} \circ F_n \quad \left(H_{n,PM}(F) = \int \mathcal{I}_{n,PM}(F)\, d\mu\right). \tag{15}$$

The link between these quantities, which behave very much the same, is made precise in formulas (46) and (47).

We shall employ each of the above quantities as seems appropriate. It should be clear that a result formulated using one is easily transformed into a result formulated in terms of the other.

In the case of ergodic $\mathbb{Z}^2$ action we shall use the following notation.

If $F : \Omega \to \mathbb{R}$ is measurable and K is a finite subset of $\mathbb{Z}^2$, let $F_K(x) = (F(T^m S^n x))_{(m,n)\in K}$, for $x \in \Omega$. If the law of F_K is absolutely continuous with respect to $|K|$-dimensional Lebesgue measure, we denote f_K as its density. In particular, if $K = \{(i, j) \in \mathbb{Z}^2 : 0 \leq i \leq n-1, 0 \leq j \leq n-1\}$, we denote F_K by F_{n^2} and f_K by f_{n^2}. If the process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{Z}^2}$ is absolutely continuous, as in the one-dimensional case, we consider

$$h_n^{(2)} := -\log f_{n^2} \circ F_{n^2}, \tag{16}$$

and its integral

$$H_n^{(2)} := \int h_n^{(2)}\, d\mu. \tag{17}$$

As our concern is the asymptotic behavior of $(1/n^2)h_n^{(2)}$, there is no loss of generality if we suppose that $\Omega = \mathbb{R}^{\mathbb{Z}^2}$, F is the projection onto the zero coordinate, $(Tx)_g = x_{g+(0,1)}$, $(Sx)_g = x_{g+(1,0)}$, for $x \in \mathbb{R}^{\mathbb{Z}^2}$, $g \in \mathbb{Z}^2$, and that μ is a probability measure whose finite-dimensional marginals are absolutely continuous with respect to Lebesgue measure, and which is invariant by the shifts T and S. In this case all the densities f_{n^2}, for various n, will be denoted by f without subscript, so that, for every $x \in \mathbb{R}^{\mathbb{Z}^2}$, we have

$$\frac{1}{n^2} h_n^{(2)}(x) = -\frac{1}{n^2} \log f(X_{n-1,n-1}^n), \tag{18}$$

where

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

$$X^n_{i,j} := (x_{s,t})_{(s,t)\in I^n_{i,j}}, \tag{α_0}$$

and

$$I^n_{i,j} := \{(s,\, j) \mid s = 0, \ldots, i\} \cup \{(s,\, t) \mid 0 \le s \le n-1,\, 0 \le t \le j-1\}, \tag{α_1}$$

for $1 \le i,\, j \le n-1$, $I^n_{0,0} = \{(0,\, 0)\}$ and $I^n_{i,0} = \{(s,\, 0) : s = 0, \ldots, i\}$. For future use, let

$$Y^n_{i,j} := (x_{s,t})_{(s,t)\in I^n_{i,j},(s,t)\neq(0,0)}.$$

In particular,

$$X^{n+1}_{n,n} := (x_{s,t})_{s,t=0,\ldots,n}, \tag{a_0}$$

and

$$Y^{n+1}_{n,n} := (x_{s,t})_{s,t=0,\ldots,n,(s,t)\neq 0}. \tag{a_1}$$

Let

$$L := \{(i,\, j,\, n) \mid 0 \le i,\, j \le n-1,\, n \ge 1\}. \tag{a_2}$$

The set inclusion on the $I^n_{i,j}$ induces a partial order on L: we set

$$(i,\, j,\, n) \le (i',\, j',\, n') \iff I^n_{i,j} \subset I^{n'}_{i',j'} \iff (j < j',\, n \le n') \quad \text{or} \quad (j = j',\, i \le i',\, n \le n'),$$

that is, the product of the lexicographical order on $\{(j,\, i)\}$ and the usual order on $\{n\}$.

In the remark below, we single out two properties which we use later.

Remark 4.1.

(1) L is directed, and $C := \{n-1,\, n-1,\, n) \mid n \ge 1\}$ is a cofinal subset of L.

(2) If $\{(i_k,\, j_k,\, n_k) \mid k \in \mathbb{N}\}$ is an infinite subset of L, then there is an infinite subset J of $\mathbb{N}$, such that the sequence $((i_k,\, j_k,\, n_k))_{k\in J}$ is strictly increasing in L, and $\lim_{k\in J} n_k + \infty$.

For $l = (i,\, j,\, n) \in L$, denote $\mathcal{F}_l$ or $\sigma(X^n_{i,j})$, the sigma-algebra generated by $\{x_{s,t} : (s,\, t) \in I^n_{i,j}\}$. Then, for $l,\, l' \in L$, $l \le l' \iff \mathcal{F}_l \subset \mathcal{F}_{l'}$.

If m is a probability measure on $\mathbb{R}^{\mathbb{N}^2}$, $m^n_{i,j}$ will denote its restriction to $\mathcal{F}_l = \sigma(X^n_{i,j})$, and we then write

$$m^n_{i,j} := m \mid \sigma(X^n_{i,j}), \tag{r_0}$$

and, in particular, $m^n_{n-1,n-1}$ will be denoted simply by m_n.

Two particular probability measures π and ν on $\mathbb{R}^{\mathbb{N}^2}$ will be useful for our purpose. Their finite- dimensional marginals $\pi_n = \pi^n_{n-1,n-1}$ and $\nu_n = \nu^n_{n-1,n-1}$ are given by

$$\pi_n = \prod_{(s,t)\in I^n_{n-1,n-1}} f(x_{s,t})\, d\lambda, \tag{19}$$

and

$$\nu_n = f(x_{0,0}) \times f(Y^n_{n-1,n-1})\, d\lambda, \tag{20}$$

where $Y^n_{n-1,n-1}$ is as in (a_1) and $\lambda = \lambda_{n^2}$ denotes n^2-dimensional Lebesgue measure.

Recall also that

$$\mu_n = f(X^n_{n-1,n-1})\, d\lambda, \tag{21}$$

where $X^n_{n-1,n-1}$ is as in (a_0).

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

4.2. *Convergence of the Shannon entropy.* We now turn to the dynamical situation. We consider a dynamical system (Ω, T, μ) and denote by $AC(\Omega, T, \mu)$ the set of functions $F \in L^2(\mu)$ such that the process $(F \circ T^n)$ is absolutely continuous (i.e. as in Definition 2.1).

We establish the convergence of the sequence of Shannon entropy $(H_n(F)/n)$ defined by (11), and give, as in the discrete valued case, an *a priori* upper bound for this limit and a criterion implying that the process $(F \circ T^n)$ is Gaussian independent (Corollary 4.5). We also identify this limit (Lemma 4.8).

In the same way, for $\mathbb{Z}^2$ action, with T and S as generators, $AC(\Omega, T, S, \mu)$ will denote the set of $F \in L^2(\mu)$ such that the process $(F \circ T^m S^n)_{m,n\in\mathbb{Z}}$ is absolutely continuous (i.e. Definition 2.1).

The next formula gives one of the announced links.

Recall that γ_n denotes the independent Gaussian measure (cf. (13)), and ψ is as in (4).

LEMMA 4.2. *Let $F \in AC(\Omega, T, \mu)$. Then*

$$\frac{H_n(F)}{n} = \frac{1}{n}\int_{\mathbb{R}^n} \psi\left(\frac{d\mu F_n^{-1}}{d\gamma_n}\right) d\gamma_n + \frac{1}{2}\log(2\pi) + \frac{1}{2}\|F\|_2^2. \tag{22}$$

Proof. From formula (12), we get

$$H_n = \int \frac{d\mu F_n^{-1}}{d\gamma_n}\psi\left(\frac{d\gamma_n}{dl^n}\right) dl^n + \int \psi\left(\frac{d\mu F_n^{-1}}{d\gamma_n}\right) d\gamma_n.$$

But, if $h(t) := (1/\sqrt{2\pi})\exp(-(1/2)t^2)$ for $t \in \mathbb{R}$, then

$$\begin{aligned}
\int \frac{d\mu F_n^{-1}}{d\gamma_n}\psi\left(\frac{d\gamma_n}{dl^n}\right) dl^n &= \int -\log\frac{d\gamma_n}{dl^n} d\mu F_n^{-1} = -\sum_{j=0}^{n-1}\int \log(h(x_j))\, d\mu F_n^{-1}(x) \\
&= -\sum_{j=0}^{n-1}\int \log(h(F(T^j x)))\, d\mu(x) \\
&= -n\int \log(h(F(x)))\, d\mu(x) \\
&= -n\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\int F(x)^2 d\mu(x)\right] \\
&= n \times \left[\frac{1}{2}\log(2\pi) + \frac{1}{2}\|F\|_2^2\right].
\end{aligned}$$

□

Remark 4.3.

(i) The preceding formula (22) can be written as

$$H_n(F) = -H_{\gamma_n}(\mu F_n^{-1}) + \frac{n}{2}(\log 2\pi + \|F\|_2^2). \tag{23}$$

(ii) If $H_n(F)$ is infinite then $H_{n+1}(F)$ is infinite.

In fact, (ii) follows from (i), and (i) from formula (5).

LEMMA 4.4. *Let $F \in AC(\Omega, T, \mu)$. Then $(H_n(F))_{n\in N}$ is a sub-additive sequence: for $n, p \in N$,*

$$H_{n+p} \le H_n + H_p. \tag{24}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Proof. Inequality (24) follows immediately from Lemma 3.3 and formula (23). □

The following corollary is the analogue for the Shannon entropy of the fact that the Kolmogorov entropy of a countable states process is bounded by the entropy of the zeroth coordinate. The equality case is analogous to the fact that, in the Kolmogorov situation, the equality implies that the process is Bernoulli.

COROLLARY 4.5. *Let (Ω, T, μ) be a dynamical system. Then for any $F \in AC(\Omega, T, \mu)$ the sequence $(H_n(F)/n)$ of Shannon entropies converges to $Se(F, T)$, which may be infinite. Moreover,*

$$Se(F, T) \leq \tfrac{1}{2}(\log(2\pi) + \|F\|_2^2), \tag{25}$$

and the equality holds if and only if for every n the law μF_n^{-1} of $(F, F \circ T, \ldots, F \circ T^{n-1})$ is the Gaussian independent measure γ_n.

Proof. The sequence (H_n/n) converges to its infimum, since (H_n) is sub-additive by Lemma 4.4. On the other hand, formula (23) implies that $(H_n/n) \leq (1/2)\,(\log(2\pi) + \|F\|_2^2)$, which then yields the inequality (25).

To prove the other statement, note that, from formula (23) and the equality $Se(F, T) = \inf\{(H_n/n) \mid n \in N\}$, it follows that the equality in (25) is equivalent to the equalities $H_{\gamma_n}(\mu F_n^{-1}) = 0$, for all n. But this is equivalent to the equalities $\mu F_n^{-1} = \gamma_n$, for all n by Remark 3.2(2). □

Note that the preceding corollary generalizes to processes a theorem of Shannon that among random variables with fixed variance the maximum of the continuous entropy is achieved by a Gaussian variable.

A similar proof can yield an n-dimensional version of this theorem of Shannon.

PROPOSITION 4.6. *Let*

$$\mathcal{M}_1(n) = \left\{ p \in L_+^1(l^n) \,\middle|\, \int_{\mathbb{R}^n} p(x)\, dl^n(x) = 1, \int_{\mathbb{R}^n} \|x\|_2^2 p(x) dl^n(x) = n \right\}.$$

Then

$$\sup_{p \in \mathcal{M}_1(n)} \int_{\mathbb{R}^n} \psi(p(x))\, dl^n(x) = \frac{n}{2}(1 + \log(2\pi)).$$

Furthermore, this supremum is attained for the Gaussian independent density

$$p(x) = (2\pi)^{-(n/2)} \exp(-\tfrac{1}{2}(x_0^2 + \cdots + x_{n-1}^2)),$$

for $x = (x_0, \ldots, x_{n-1}) \in \mathbb{R}^n$, and this is the only one.

We come to the main definition of this section.

Definition 4.7. The Shannon entropy $Se(F, T)$ of the process $(F \circ T^n)$ is defined by the equality

$$Se(F, T) := \lim_n \frac{H_n(F)}{n}. \tag{26}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Next we identify the limit in Definition 4.7, the Shannon entropy $Se(F, T)$ of the process $(F \circ T^n)$, using conditional entropy (Lemma 4.8(b) (ii) and (iii)).

LEMMA 4.8.

(a) *Suppose that H_{n+p} is finite. Then*

$$H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1}) = H_n + H_p - H_{n+p}, \tag{27}$$

and, for fixed n, the sequence $H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1})$ is increasing in p.

(b) *If H_n is finite for all n, then:*

(i) *$Se(F, T)$ is finite if and only if for every n, or for some n, $\sup_p H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1})$ is finite;*

(ii)

$$\sup_p H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1}) = H_n - n \times Se(F, T); \tag{28}$$

(iii) *if ν is the law of the process $(F, F \circ T, \ldots)$ and ν_n is the probability measure on $\mathbb{R}^{\mathbb{N}}$, with $(n + p)$-marginal given by $\mu F_n^{-1} \times \mu F_p^{-1}$, for any $p \geq 0$, then*

$$H_{\nu_n}(\nu) = H_n - nSe(F, T).$$

In particular, if $Se(F, T)$ is finite, ν is absolutely continuous with respect to ν_n.

Proof. (a) Formula (27) follows from Lemma 3.3 and formula (23). The other property follows from the definition, since, when Π_1 and Π_2 are finite partitions of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively, we have

$$S_{\Pi_1 \times \Pi_2}(\mu F_{n+p}^{-1} \mid \mu F_n^{-1} \times \mu F_p^{-1}) \leq S_{\Pi_1 \times (\Pi_2 \times \mathbb{R})}(\mu F_{n+p+1}^{-1} \mid \mu F_n^{-1} \times \mu F_{p+1}^{-1}).$$

(b) Put $v_p^n := H_p - H_{n+p}$, and in particular, for $n = 1$,

$$u_p := -v_p^1 = H_{p+1} - H_p.$$

Then, from (a) above, the sequence (u_p) is decreasing. So, as $(1/N) \sum_{p=1}^N u_p$ converges to $Se(F, T)$, (u_n) also converges to $Se(F, T) = \inf_p u_p$. This proves (i). (ii) follows from (i) and the equality $v_p^n = -u_p - u_{p+1} - \cdots - u_{n+p-1}$. (iii) follows from (ii). □

Remark 4.9. If $Se(F, T)$ is finite then

$$\lim_n \frac{1}{n} \sup_p H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1}) = 0, \tag{29}$$

and, for fixed p, $H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1})$ is increasing in n, and

$$z_n := \sup_p H_{\mu F_n^{-1} \times \mu F_p^{-1}}(\mu F_{n+p}^{-1})$$

is sub-additive and increasing.

For the convergence of Shannon entropy in the case of $\mathbb{Z}^2$ action, for $F \in AC(\Omega, T, S, \mu)$, we have the following.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

LEMMA 4.10. *Let μ_n and π_n be as in (21) and (19), respectively. Then*

$$\lim_n \frac{1}{n^2} H_n^{(2)} = -\sup_n \frac{1}{n^2} H_{\pi_n}(\mu_n) - \int_{\mathbb{R}} f(t) \log f(t)\, dt.$$

Proof. With notation as in (a_0), we can write $f(X^n_{n-1,n-1})$ in the following form,

$$f(X^n_{n-1,n-1}) = \frac{f(X^n_{n-1,n-1})}{\prod_{(s,t)\in I^n_{n-1,n-1}} f(x_{s,t})} \times \prod_{(s,t)\in I^n_{n-1,n-1}} f(x_{s,t}),$$

from which we get

$$\int \log f(X^n_{n-1,n-1})\, d\mu = \int \log \frac{f(X^n_{n-1,n-1})}{\prod_{(s,t)\in I^n_{n-1,n-1}} f(x_{s,t})}\, d\mu + \sum_{(s,t)\in I^n_{n-1,n-1}} \int \log f(x_{s,t}) f(x_{s,t})\, dx_{s,t}.$$

That is,

$$\int \log f(X^n_{n-1,n-1})\, d\mu = H_{\pi_n}(\mu^n_{n-1,n-1}) + n^2 \int f(x_{0,0}) \log f(x_{0,0})\, dx_{0,0}. \tag{30}$$

Put $z_n := H_{\pi_n}(\mu^n_{n-1,n-1})$. (One can see that $(-z_n)$ is sub-additive.)

Let $R^n_{k,l} = \mathbb{Z}^2 \cap [kn-1, (k+1)n-1] \times [ln-1, (l+1)n-1]$ and denote by $\mu \mid R^n_{k,l}$ the restriction of μ to the coordinates in $R^n_{k,l}$ and similarly for $\pi \mid R^n_{k,l}$. Let n be fixed and $N \geq n$ be an integer. Write $N = p_N n + r_N$, where $p = p_N$, $r = r_N \in \mathbb{N}$, $0 \leq r < n$ and $p \geq 1$. Then clearly we have, by the definition of the conditional entropy,

$$z_N = H_{\pi_N}(\mu^N_{N-1,N-1}) \geq H_{\pi_{pn}}(\mu^{pn}_{pn-1,pn-1}) = z_{pn}.$$

But, by Lemma 3.3 and invariance, we obtain

$$z_{pn} \geq \sum_{k,l=0}^{p-1} H_{\pi|R^n_{k,l}}(\mu \mid R^n_{k,l}) = p^2 H_{\pi|R^n_{0,0}}(\mu \mid R^n_{0,0}) = p^2 z_n.$$

So

$$\limsup_N \left[-\frac{1}{N^2} z_N\right] \leq -\frac{z_n}{n^2},$$

which implies that

$$\limsup_N \left[-\frac{1}{N^2} z_N\right] \leq \inf_n \left[-\frac{z_n}{n^2}\right]. \qquad \square$$

As a consequence, we can now define an entropy of an absolutely continuous process indexed by $\mathbb{Z}^2$ as follows.

Definition 4.11. The Shannon entropy $Se(F, T, S)$ of the absolutely continuous process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{Z}^2}$ is given by the equality

$$Se(F, T, S) := \lim_n \left[-\frac{1}{n^2} \int \log f_{n^2} \circ F_{n^2}\, d\mu\right] = \lim_n \frac{1}{n^2} H_n^{(2)}. \tag{31}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Note that here too it follows from Lemma 4.10 that $Se(F, T, S) = -\int f(t) \log f(t)\, dt$ if and only if the process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{Z}^2}$ is independent.

In order to identify the limit in Definition 4.11, for $F \in AC(\Omega, T, S, \mu)$, we need some further notation. Let

$$g^n_{i,j} := \frac{f(X^n_{i,j})}{f(x_{0,0}) \times f(Y^n_{i,j})} = \frac{d\mu^n_{i,j}}{d\nu^n_{i,j}}, \tag{32}$$

where μ, ν are as in (21) and (20), respectively, and $\mu^n_{i,j}$, $\nu^n_{i,j}$ are as in (r_0).

Remark 4.12. With μ, ν, $g^n_{i,j}$ as in (21), (20) and (32), respectively, and L as in (a_2), the following properties hold: the family $(g^n_{i,j})_{(i,j,n)\in L}$ is a ν martingale;

$$\int g^n_{i,j} \log g^n_{i,j} d\nu \le \int g^n_{n-1,n-1} \log g^n_{n-1,n-1} d\nu; \tag{33}$$

$$\sup_{(i,j,n)\in L} \int g^n_{i,j} \log g^n_{i,j} d\nu = \sup_n \int g^n_{n-1,n-1} \log g^n_{n-1,n-1} d\nu; \tag{34}$$

$$\frac{1}{n^2} h^{(2)}_n = -\frac{1}{n^2} \sum_{i,j=0}^{n-1} \log g^n_{i,j} - \log f(x_{0,0}). \tag{E}$$

Remark 4.13. The notation is as in Remark 4.12. Suppose

$$\sup_n \int g^n_{n-1,n-1} \log g^n_{n-1,n-1} d\nu < \infty.$$

Then, if $s_k = (i_k, j_k, n_k)$ is an increasing sequence in L, the martingale $M_k := g^{n_k}_{i_k,j_k}$ is uniformly integrable and converges ν almost everywhere and in $L^1(\nu)$ to the density g_s of μ restricted to $\sigma(\{M_k \mid k \ge 1\})$ with respect to ν.

In particular, if $U_n := g^n_{n-1,n-1}$, then U_n converges ν almost everywhere to the density g of μ with respect to ν on the sigma-algebra $\sigma(\{x_{s,t} \mid s, t \ge 0\})$ and therefore $g^n_{i,j} = E_\nu(g \mid \sigma(X^n_{i,j}))$.

In fact, by (34), the hypothesis implies that $\sup_{(i,j,n)\in L} \int g^n_{i,j} |\log g^n_{i,j}|\, d\nu < \infty$, so that the family $(g^n_{i,j})_{(i,j,n)\in L}$ is uniformly integrable with respect to ν. Here we have used the following.

Remark 4.14. Let m be a finite measure and $\Phi = \{f_i \mid i \in I\}$ be a family of positive elements in $L^1(m)$ with the property

$$\sup_i \int f_i \log f_i\, dm < \infty.$$

Then:

(1) $\sup_i \int f_i |\log f_i| dm < \infty$;
(2) the family $\{f_i \mid i \in I\}$ is uniformly integrable;
(3) the family $\{(\log f_i)^+ \mid i \in I\}$ is uniformly integrable.

The following lemma identifies the Shannon entropy of $(F \circ S^m \circ T^n)_{(m,n)\in\mathbb{Z}^2}$, for $\mathbb{Z}^2$ action (the limit in Lemma 4.10).

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

LEMMA 4.15. *Let $F \in AC(\Omega, T, S)$, μ and ν be as in (21) and (20), respectively. Then*

$$Se(F, T, S) = \lim_n \frac{1}{n^2} H_n^{(2)} = -H_\nu(\mu) - \int f(t) \log f(t)\, dt, \tag{35}$$

and, in particular, $Se(F, T, S)$ is finite if and only if $H_\nu(\mu) < \infty$.

Proof. Equation (E) implies that

$$\frac{1}{n^2} H_n^{(2)} = -\frac{1}{n^2} \sum_{i,j=0}^{n-1} H_{\nu_{i,j}^n}(\mu_{i,j}^n) - \int f(t) \log f(t)\, dt,$$

so

$$\frac{1}{n^2} H_n^{(2)} \geq -H_{\nu_{n-1,n-1}^n}(\mu_{n-1,n-1}^n) - \int f(t) \log f(t)\, dt \geq -H_\nu(\mu) - \int f(t) \log f(t)\, dt,$$

proving one direction. On the other hand, for every n, let $t_n = \int g_{n-1,n-1}^n \log g_{n-1,n-1}^n d\nu = H_{\nu_{n-1,n-1}^n}(\mu_{n-1,n-1}^n)$, and $u_n = (1/n^2) \sum_{i,j=0}^{n-1} H_{\nu_{i,j}^n}(\mu_{i,j}^n)$. If $k > p$, then

$$\begin{aligned} u_k &\geq \frac{1}{k^2} \sum_{i=0}^{k-1} \sum_{j=p}^{k-1} H_{\nu_{i,j}^k}(\mu_{i,j}^k) \geq \frac{1}{k^2} \sum_{i=0}^{k-1} \sum_{j=p}^{k-1} H_{\nu_{p-1,p-1}^p}(\mu_{p-1,p-1}^p) \\ &= \frac{k(k-p)}{k^2} H_{\nu_{p-1,p-1}^p}(\mu_{p-1,p-1}^p) \\ &= \left(1 - \frac{p}{k}\right) H_{\nu_{p-1,p-1}^p}(\mu_{p-1,p-1}^p) = \left(1 - \frac{p}{k}\right) t_p. \end{aligned}$$

Hence, $\liminf_k u_k \geq t_p$, which implies that $\liminf_k u_k \geq \sup_p t_p$, and proves the other direction and also the equality $H_\nu(\mu) = \lim_n u_n$, or that $(1/n^2) H_n^{(2)}$ converges to $-H_\nu(\mu) - \int f(t) \log f(t)\, dt$. □

We give examples of non-Gaussian and non-independent processes with finite Shannon entropy.

LEMMA 4.16. *Let (Ω, μ, T) be a dynamical system. Let $F, G \in L^1(\mu)$. Let $\xi_n := (F, F \circ T, \ldots, F \circ T^{n-1})$, and $\eta_n := (G, G \circ T, \ldots, G \circ T^{n-1})$. Suppose that ξ_n is absolutely continuous and that ξ_n and η_n are independent. Then $\xi_n + \eta_n$ is absolutely continuous and*

$$H_n(F + G) \geq H_n(F). \tag{36}$$

Proof. If $H_n(F) = -\infty$, there is nothing to prove. Suppose then that $H_n(F) > -\infty$. Denote by f the density, with respect to l^n, of the law α of ξ_n, by μ the law of η_n, and by ν the law of $\xi_n + \eta_n$. Then $\nu = \alpha \star \mu = g\, dl^n$, and $g(t) = \int_{\mathbb{R}^n} f(t - y)\, d\mu(y)$, for l^n-almost all $t \in \mathbb{R}^n$. Thus we can write

$$\begin{aligned} H_n(F + G) &= \int_{\mathbb{R}^n} \psi(g)\, dl^n = \int_{\mathbb{R}^n} \psi\left(\int_{\mathbb{R}^n} f(t - y)\, d\mu(y)\right) dl^n(t) \\ &\geq \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \psi(f(t - y))\, d\mu(y)\right] dl^n(t) \\ &= \int_{\mathbb{R}^n} d\mu(y) \int_{\mathbb{R}^n} \psi(f(u))\, dl^n(u) = H_n(F). \end{aligned}$$

□

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

COROLLARY 4.17. *Let $F \in AC(\Omega, T, \mu)$. Assume that $G \in L^1(\mu)$ is such that the two processes $(F \circ T^n)_{n\geq 0}$ and $(G \circ T^n)_{n\geq 0}$ are independent. Then*

$$Se(F + G, T) \geq Se(F, T). \tag{37}$$

If, in addition, $G \in AC(\Omega, T, \mu)$, then

$$Se(F + G, T) \geq \max\{Se(F, T), Se(G, T)\} \geq \tfrac{1}{2}(Se(F, T) + Se(G, T)).$$

4.3. *Connections between Shannon entropy and Kolmogorov–Sinai entropy.* We show that if, for some $F \in AC(\Omega, T)$, the Shannon entropy $Se(F, T)$ is finite then the Kolmogorov entropy of (Ω, T, μ) is infinite (Corollary 4.21). We also give a new way to describe the Shannon entropy for $\mathbb{Z}$ action (Theorem 4.20(b)), and for $\mathbb{Z}^2$ action (Remark 4.22). Finally, we obtain a criterion of Markovianness (Proposition 4.23).

The following lemma is important for comparing the Kolmogorov–Sinai entropy to the Shannon entropy.

If $\mathcal{P}$ is a finite measurable partition of a probability space $(\Omega, \mathcal{F}, P)$ then the entropy of $\mathcal{P}$ will be denoted by $H(\mathcal{P})$, $H(\mathcal{P}, P)$ or $H^P(\mathcal{P})$.

LEMMA 4.18. *Let $\mathcal{E}$ be the set of finite measurable partitions of $\mathbb{R}$, and $F^{-1}\mathcal{E} = \{F^{-1}\mathcal{P} \mid \mathcal{P} \in \mathcal{E}\}$. Then*

$$H_{\mu F_n^{-1}\times\mu F_p^{-1}}(\mu F_{n+p}^{-1}) = \sup_{\mathcal{B}\in F^{-1}\mathcal{E}} \left[H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big) - H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B} \,\Big|\, \bigvee_{j=1}^{p} T^{-j}\mathcal{B}\Big)\right].$$

Proof. By Theorem B of Dobrushin,

$$H_{\mu F_n^{-1}\times\mu F_p^{-1}}(\mu F_{n+p}^{-1}) = \sup S_{\mathcal{P}_1\times\mathcal{P}_2}(\mu F_{n+p}^{-1} \mid \mu F_n^{-1} \times \mu F_p^{-1}),$$

the supremum being taken over all finite partitions $\mathcal{P}_1$ of $\mathbb{R}^n$ and $\mathcal{P}_2$ of $\mathbb{R}^p$, of the following particular forms: $\mathcal{P}_1 = \mathcal{P} \times \cdots \times \mathcal{P}$, n times, and $\mathcal{P}_2 = \mathcal{P} \times \cdots \times \mathcal{P}$, p times, where $\mathcal{P}$ run in $\mathcal{E}$. But

$$\begin{aligned} S_{\mathcal{P}_1\times\mathcal{P}_2}(\mu F_{n+p}^{-1} \mid \mu F_n^{-1} \times \mu F_p^{-1}) = {} & -H(\mathcal{P}_1 \times \mathcal{P}_2, \mu F_{n+p-1}^{-1}) + H(\mathcal{P}_1, \mu F_n^{-1}) \\ & + H(\mathcal{P}_2, \mu F_p^{-1}), \end{aligned}$$

holds and, for these particular forms of $\mathcal{P}_i$, $i = 1, 2$, setting $\mathcal{B} := F^{-1}\mathcal{P}$, we easily have that

$$H(\mathcal{P}_1, \mu F_n^{-1}) = H^\mu\Big(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\Big),$$

$$H(\mathcal{P}_2, \mu F_p^{-1}) = H^\mu\Big(\bigvee_{j=n}^{n+p-1} T^{-j}\mathcal{B}\Big)$$

and

$$H(\mathcal{P}_1 \times \mathcal{P}_2, \mu F_{n+p-1}^{-1}) = H^\mu\Big(\bigvee_{j=0}^{n+p-1} T^{-j}\mathcal{B}\Big).$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

It then follows that

$$S_{\mathcal{P}_1\times\mathcal{P}_2}(\mu F_{n+p}^{-1} \mid \mu F_n^{-1} \times \mu F_p^{-1}) = H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big) - H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big|\bigvee_{j=1}^{p} T^{-j}\mathcal{B}\Big). \quad \square$$

In the light of Lemma 4.18, formula (27) in Lemma 4.8 can be written as

$$\sup_{\mathcal{B}\in F^{-1}\mathcal{E}}\left[H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big) - H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big|\bigvee_{j=1}^{p} T^{-j}\mathcal{B}\Big)\right] = H_n + H_p - H_{n+p}. \quad (38)$$

Remark 4.19. Formula (38) allows one to obtain conditions which ensure that H_n will be finite.

One can prove for instance that, for all N, the following are equivalent:

(i) H_{N+1} is finite;

(ii) $\sup_{\mathcal{B}\in F^{-1}\mathcal{E}}[H^\mu(\mathcal{B}) - H^\mu(\mathcal{B} \mid \bigvee_{j=1}^{N} T^{-j}\mathcal{B})] < +\infty$;

(ii′) $H_{\mu F^{-1}\times\mu F_N^{-1}}(\mu F_{N+1}^{-1}) < +\infty$.

The following theorem gives, in particular, the announced new description of the Shannon entropy for $\mathbb{Z}$ action.

THEOREM 4.20. *Let $\mathcal{E}$ be as in Lemma 4.18. Suppose H_n are finite for all n. Then:*

(a) *the following are equivalent:*

(i) *$Se(F, T)$ is finite,*

(ii) *for every n (or for some n), $\sup_{\mathcal{B}\in F^{-1}\mathcal{E}}[H^\mu(\bigvee_{j=0}^{n-1} T^j\mathcal{B}) - nH^\mu(\mathcal{B} \mid \mathcal{B}^-)] < +\infty$,*

(iii) *$\lim_n (1/n) \sup_{\mathcal{B}\in F^{-1}\mathcal{E}}[H^\mu(\bigvee_{j=0}^{n-1} T^j\mathcal{B}) - nH^\mu(\mathcal{B} \mid \mathcal{B}^-)] = 0$;*

(b) *the following equality holds:*

$$Se(F, T) = H_1 + \inf_{\mathcal{B}\in F^{-1}\mathcal{E}}[H^\mu(\mathcal{B} \mid \mathcal{B}^-) - H^\mu(\mathcal{B})]. \quad (39)$$

Proof. Formulas (38), (27) and (28) imply that

$$z_n := \sup_{\mathcal{B}\in F^{-1}\mathcal{E}}\left[H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big) - H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B} \mid \mathcal{B}^-\Big)\right] = H_n - n \times Se(F, T),$$

in which as usual $\mathcal{B}^- := T^{-1}\mathcal{B} \vee T^{-2}\mathcal{B} \vee \cdots$. Now

$$H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B} \mid \mathcal{B}^-\Big) = nH^\mu(\mathcal{B} \mid \mathcal{B}^-),$$

so we get the equality

$$H_n - n \times Se(F, T) = \sup_{\mathcal{B}\in F^{-1}\mathcal{E}}\left[H^\mu\Big(\bigvee_{j=0}^{n-1} T^j\mathcal{B}\Big) - nH^\mu(\mathcal{B} \mid \mathcal{B}^-)\right], \quad (40)$$

which establishes the equivalence between (i) and (ii). As (i) implies (iii), by Remark 4.9, and trivially (iii) implies (ii), the proof is finished, because taking $n = 1$ in formula (40) yields equality (39). $\square$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

COROLLARY 4.21. *Let (Ω, T, μ) be an invertible dynamical system. If there exists $F \in AC(\Omega, T, \mu)$ such that the Shannon entropy $Se(F, T)$ of the process $(F \circ T^n)$ is finite then (Ω, T, μ) has infinite entropy.*

Proof. The corollary follows directly from the equivalence between (i) and (ii) in Theorem 4.20. □

Now, for $\mathbb{Z}^2$ action, we establish a formula analogous to (39) for $\mathbb{Z}$ action.

Remark 4.22. Let $\mathcal{E}$ be as in Lemma 4.18. Then, for the stationary absolutely continuous process $(F \circ S^m \circ T^n)$, indexed by $\mathbb{Z}^2$, the following holds:

$$Se(F, T, S) = -\int f(t) \log f(t)\, dt + \inf_{\mathcal{P} \in \mathcal{E}} [H^\mu(F^{-1}\mathcal{P} \mid (F^{-1}\mathcal{P})^-) - H^\mu(F^{-1}\mathcal{P})]. \quad (41)$$

In fact, we have $H_\nu(\mu) = \sup_n H_{\nu_n}(\mu_n)$. But

$$H_{\nu_n}(\mu_n) = \sup S_\Pi(\mu_n \mid \nu_n),$$

the supremum being taken over all partitions Π of the form

$$\Pi = \left\{ \bigcap_{(i,j) \in I^n_{n-1,n-1}} S^{-i} T^{-j} F^{-1} E : E \in \mathcal{P} \right\},$$

where $\mathcal{P}$ is a finite partition of $\mathbb{R}$. But then, with $E_{i,j} = S^{-i} T^{-j} F^{-1} E$, from the definitions of μ_n and ν_n (see (21) and (20)), it follows that

$$\begin{aligned} S_\Pi(\mu_n \mid \nu_n) &= \sum_{E \in \mathcal{P}} \mu_n\left(\bigcap_{(i,j) \in I^n_{n-1,n-1}} E_{i,j} \right) \log \frac{\mu_n(\bigcap_{(i,j) \in I^n_{n-1,n-1}} E_{i,j})}{\nu_n(\bigcap_{(i,j) \in I^n_{n-1,n-1}} E_{i,j})} \\ &= H^\mu(F^{-1}\mathcal{P}) - H^\mu\left(\bigvee_{(i,j) \in I^n_{n-1,n-1}} S^{-i} T^{-j} F^{-1} \mathcal{P} \right) + H^\mu \\ &\quad \times \left(\bigvee_{(i,j) \in J^n_{n-1,n-1}} S^{-i} T^{-j} F^{-1} \mathcal{P} \right) \\ &= H^\mu(F^{-1}\mathcal{P}) - H^\mu\left(F^{-1}\mathcal{P} \,\middle|\, \bigvee_{(i,j) \in J^n_{n-1,n-1}} S^{-i} T^{-j} F^{-1} \mathcal{P} \right). \end{aligned}$$

Hence

$$\sup_n H_{\nu_n}(\mu_n) = \sup_{\mathcal{P}} [H^\mu(F^{-1}\mathcal{P}) - H^\mu(F^{-1}\mathcal{P} \mid (F^{-1}\mathcal{P})^-)],$$

where $(F^{-1}\mathcal{P})^- = \bigvee_{n \geq 1} \bigvee_{(i,j) \in J^n_{n-1,n-1}} T^{-i} S^{-j} F^{-1} \mathcal{P}$.

By Lemma 4.15, we thus obtain (41).

The following proposition gives a criterion for the Markovianness of the process $(F \circ T^n)$.

PROPOSITION 4.23. *Let $F \in AC(\Omega, T, m)$ such that $Se(F, T)$ is finite. Then:*

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

(i) *the process* $(F \circ T^n)$ *is Markovian if and only if* $Se(F, T) = H_2 - H_1$*; or, more generally,*

(ii) *the process* $(F \circ T^n)$ *has memory* p *if and only if* $Se(F, T) = H_{p+1} - H_p$.

Proof. We prove (i). The proof of (ii) is similar. If $(F \circ T^n)$ is Markovian, taking $n = 1$ and $n = 2$ in formula (40), we see that $Se(F, T) = H_2 - H_1$.

For the other direction, put, for $n \geq 1$,

$$a_n := \sup_{\mathcal{A} \in F^{-1}\mathcal{E}} [H(\mathcal{A}) - H(\mathcal{A} \mid T^{-1}\mathcal{A} \vee \cdots \vee T^{-n}\mathcal{A}))],$$

$\mathcal{R}^n = \mathcal{R}^n(\mathcal{A}) := T^{-2}\mathcal{A} \vee \cdots \vee T^{-n}\mathcal{A}$ and $L = L_n = E[\cdot \mid T^{-1}\mathcal{A}] - E[\cdot \mid T^{-1}\mathcal{A} \vee \mathcal{R}^n]$. Let f be a bounded $F^{-1}\mathcal{E}$-measurable function, where, as in Lemma 7, $\mathcal{E}$ denotes the set of finite measurable partitions of $\mathbb{R}$. We shall show that $L(f) = 0$ and the proof will be finished. To do this, we use [**19**, Lemma 4.5 and Theorem 4.8], according to which, for any $\epsilon > 0$, there exists $\delta(\epsilon)$, $0 < \delta(\epsilon) < \epsilon$, such that for any probability space $(\Omega, \mathcal{F}, \mu)$ and any finite partitions $\mathcal{P}$ and $\mathcal{Q}$ of Ω, the inequality $H^{\mu}(\mathcal{P}) - H^{\mu}(\mathcal{P} \mid \mathcal{Q}) < \delta(\epsilon)$ implies that $\mathcal{P}$ and $\mathcal{Q}$ are ϵ-independent.

Recall that, following Ornstein [**19**, Definition 4.7], if $\mathcal{P}$ and $\mathcal{Q}$ are finite measurable partitions of a probability space $(\Omega, \mathcal{F}, \mu)$, then $\mathcal{P}$ is said to be ϵ-independent of $\mathcal{Q}$ if

$$\sum_{p \in \mathcal{P}} |\mu(p \mid q) - \mu(p)| < \epsilon,$$

for all atoms q except a set of atoms of $\mathcal{Q}$ the union of which has a measure less than ϵ.

Fix $\epsilon > 0$. Then there exists $\mathcal{A}_0$ such that

$$a_1 - \delta(\epsilon)^2 < H(\mathcal{A}) - H(\mathcal{A} \mid T^{-1}\mathcal{A}),$$

for any $\mathcal{A}$ which is finer than $\mathcal{A}_0$. But, by formula (15), the hypothesis $Se(F, T) = H_2 - H_1$ is equivalent to $a_1 = a_n$ for any n. Thus

$$H(\mathcal{A}|T^{-1}\mathcal{A}) - H(\mathcal{A}|T^{-1}\mathcal{A} \vee \mathcal{R}^n(\mathcal{A})) < \delta(\epsilon)^2.$$

So if we denote, respectively, by p, q and r, the generic element of $\mathcal{A}$, $T^{-1}\mathcal{A}$ and $\mathcal{R}^n$, we get

$$\sum_q m(q)\, [H^{m_q}(\mathcal{A}_q) - H^{m_q}(\mathcal{A}_q \mid \mathcal{R}^n_q)] < \delta(\epsilon)^2,$$

where $m_q(A) = (m(A \cap q)/m(q))$, $\mathcal{A}_q = \{p \cap q \mid p \in \mathcal{A}\}$ and similarly for $\mathcal{R}^n_q$.

Let $\mathcal{Q}_\epsilon := \{q \mid H^{m_q}(\mathcal{A}_q) - H^{m_q}(\mathcal{A}_q \mid \mathcal{R}^n_q) \geq \delta(\epsilon)\}$. It follows that

$$\sum_{q \in \mathcal{Q}_\epsilon} m(q)) < \delta(\epsilon), \tag{e_1}$$

and that, for $q \notin \mathcal{Q}_\epsilon$, the partitions $\mathcal{A}_q$ and $\mathcal{R}^n_q$ are ϵ-independent, under the measure m_q; that is, there exists J_q, a subfamily of $\mathcal{R}^n_q$, such that

$$\sum_{r \in J_q} m_q(r) > 1 - \epsilon, \tag{e_2}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

and

$$\sum_{p\in\mathcal{A}} |m(p \mid q \cap r) - m(p \mid q)| < \epsilon \quad \text{for all } r \in J_q. \tag{e_3}$$

Now we can find $\mathcal{A}$ finer than $\mathcal{A}_0$, and $g = g_\epsilon = \sum_{p\in\mathcal{A}} y_p 1_p$ such that $\|f - g\|_1 < \epsilon$, and $\|g\|_\infty \le 2 \times \|f\|_\infty$. Then

$$\|L(g)\|_1 = \sum_{q,r} \left| \sum_p y_p [m(p \mid q) - m(p \mid q \cap r)] \right| m(q \cap r) = \sum_{q\in\mathcal{Q}_\epsilon} + \sum_{q\notin\mathcal{Q}_\epsilon}.$$

In view of (e_1), the first sum in the above equality is bounded by $2\|g\|_\infty \delta(\epsilon)$. By (e_3) and (e_2), the second one is bounded by $3\|g\|_\infty \times \epsilon$.

Therefore,

$$\|L(g)\|_1 \le 2\|f\|_\infty [2\delta(\epsilon) + 3\epsilon].$$

It follows that

$$\|L(f)\|_1 \le 2\|f\|_\infty [2\delta(\epsilon) + 3\epsilon] + 2\epsilon.$$

This implies that $L(f) = 0$. □

Note that if $(F \circ T^n)$ is Markovian then, for any p, $H_{2p+1} = H_1 + 2p(H_2 - H_1)$ and $H_{2p+2} = H_2 + 2p(H_2 - H_1)$.

One might be tempted to introduce an isomorphism invariant $Se(T)$:

$$Se(T) := \sup_{F\in AC(\Omega, T\mu)} \{Se(F, T) - H_1(F, T)\}. \tag{42}$$

It is indeed an invariant, but it can only take on the two values $-\infty$ and 0.

5. *Entropy rate and Shannon entropy*

We shall establish some connections between Shannon entropy and some concepts developed by Pinsker such as information stability and entropy rate.

We recall some definitions from [**14**]. First recall that if Z is a random variable then its law is denoted P_Z.

Definition 5.1. Let $\xi = (\xi_n)_{n\ge1}$ and $\eta = (\eta_n)_{n\ge1}$ be discrete time stationary processes.

The entropy rate of ξ with respect to η is

$$\bar{H}_\eta(\xi) := \lim_n \frac{1}{n} H_{P_{(\eta_1,\dots,\eta_n)}}(P_{(\xi_1,\dots,\xi_n)}),$$

defined when, for all j, ξ_j and η_j take values in the same measurable space, and when the limit exists.

Remark 5.2. We can prove, using Lemma 1, that if η is the independent Gaussian process, then for any discrete time real state stationary process ξ, the entropy rate of ξ with respect to η is well defined.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

LEMMA 5.3. *Let $\eta = (X_n)$ be the independent Gaussian process, with law γ and γ_n its projection on the first n coordinates. Let (Ω, T, μ) be a dynamical system, $F \in AC(\Omega, T, \mu)$ and ν the law of the process $\xi := (F, F \circ T, \ldots)$. Then:*

(i) *$Se(F, T)$ is finite if and only if $\sup_n (1/n) H_{\gamma_n}(\mu F_n^{-1}) < \infty$;*

(ii) *$\nu = \gamma$ if and only if $\lim_n (1/n) H_{\gamma_n}(\mu F_n^{-1}) = 0$;*

(iii) *if $H_\gamma(\nu)$ is finite then $\nu = \gamma$.*

Proof. Formula (23) implies that

$$Se(F, T) = \frac{1}{2}(\log(2\pi) + \|F\|_2^2) - \lim_n \frac{1}{n} H_{\gamma_n}(\mu F_n^{-1}), \tag{43}$$

from which we see that $Se(F, T)$ is finite if and only if $\lim_n (1/n) H_{\gamma_n}(\mu F_n^{-1})$ is finite. This proves (i) because, by super-additivity,

$$\lim_n \frac{1}{n} H_{\gamma_n}(\mu F_n^{-1}) = \sup_n \frac{1}{n} H_{\gamma_n}(\mu F_n^{-1}).$$

Also this last equality together with Remark 3.2 (2) proves (ii).

To prove (iii) note that $H_{\gamma_n}(\mu F_n^{-1}) \le H_\gamma(\nu)$, and formula (23) thus implies the inequality

$$\frac{1}{n} H_n(F) \ge \frac{1}{2}(\log(2\pi) + \|F\|_2^2) - \frac{1}{n} H_\gamma(\nu),$$

which in turn implies the following one:

$$Se(F, T) \ge \tfrac{1}{2}(\log(2\pi) + \|F\|_2^2).$$

Hence, by Corollary 4.5, we have $\nu = \gamma$. □

COROLLARY 5.4. *Let the notation be exactly as in Lemma 5.3. Then:*

(i) *$Se(F, T)$ is finite if and only if the entropy rate of ξ with respect to η is finite;*

(ii) *$\nu = \gamma$ if and only if the entropy rate of ξ with respect to η vanishes;*

(iii)

$$Se(F, T) = \tfrac{1}{2}(\log(2\pi) + \|F\|_2^2) - \bar{H}_\eta(\xi). \tag{44}$$

LEMMA 5.5. *Let (Ω, T, μ) be a dynamical system, $F \in AC(\Omega, T, \mu)$ and ν the law of the process $\xi := (F, F \circ T, \ldots)$. Let P be the product measure $P := \mu F^{-1} \otimes \mu F^{-1} \otimes \cdots$ and $P(n) = (\mu F^{-1})^{\otimes n}$ its projection to the first n coordinates. Then:*

(i) *if for all $n \in \mathbb{N}$, H_n is finite (in particular, if $Se(F, T)$ is finite), then ν is locally absolutely continuous with respect to P: for all n, $\mu F_n^{-1} \ll (\mu F^{-1})^{\otimes n}$;*

(ii) *$Se(F, T)$ is finite if and only if $\sup_n (1/n) H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1})$ is finite;*

(iii) *$\nu = P \iff \lim_n (1/n) H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1}) = 0 \iff Se(F, T) = \int_{\mathbb{R}} \psi(d\mu F^{-1}/dl)\, dl$;*

(iv) *if $H_P(\nu)$ is finite then $\nu = P$;*

(v)

$$Se(F, T) = \int_{\mathbb{R}} \psi\left(\frac{d\mu F^{-1}}{dl}\right) dl - \sup_n \frac{1}{n} H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1}). \tag{45}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Proof. Recalling that $\mathcal{I}_n(F)$ and $\mathcal{I}_{n,PM}(F)$ are defined respectively by (10) and (15), we have the formula

$$\mathcal{I}_n(F) = \mathcal{I}_{n,PM}(F) - \sum_{j=0}^{n-1} \log \frac{d\mu F^{-1}}{dl} \circ F \circ T^j, \tag{46}$$

which can be written as

$$-\frac{1}{n} \log \frac{d\mu F_n^{-1}}{dl^n} \circ F_n(x) = -\frac{1}{n} \sum_{j=0}^{n-1} \log \frac{d\mu F^{-1}}{dl} \circ F \circ T^j(x) - \frac{1}{n} \log \frac{d\mu F_n^{-1}}{d(\mu F^{-1})^{\otimes n}}(F_n(x)),$$

a proof of which is as follows.

First, by taking $P_1 = (\mu F^{-1})^{\otimes n}$ and $P_2 = \mu F^{-1}$ in formula (7) in Lemma 3.3, we find

$$\begin{aligned} H_{(\mu F^{-1})^{\otimes n} \times \mu F^{-1}}(\mu F_{n+1}^{-1}) &= H_{\mu F_n^{-1} \times \mu F^{-1}}(\mu F_{n+1}^{-1}) + H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1}) + H_{\mu F^{-1}}(\mu F^{-1}) \\ &= H_{\mu F_n^{-1} \times \mu F^{-1}}(\mu F_{n+1}^{-1}) + H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1}). \end{aligned}$$

Next, by taking $p = 1$ in formula (27), we find

$$H_{(\mu F^{-1})^{\otimes n+1}}(\mu F_{n+1}^{-1}) = H_n + H_1 - H_{n+1} + H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1}),$$

from which it follows, when H_{n+1} is finite, that $H_{(\mu F^{-1})^{\otimes n+1}}(\mu F_{n+1}^{-1})$ is finite if and only if $H_{(\mu F^{-1})^{\otimes n}}(\mu F_n^{-1})$ is finite. This, using Theorem A, implies, by induction on n, that if H_m is finite for all m, so, for every n, μF_n^{-1} is absolutely continuous with respect to $(\mu F^{-1})^{\otimes n}$. Thus we can write

$$\frac{d\mu F_n^{-1}}{dl^n} = \frac{d\mu F_n^{-1}}{d(\mu F^{-1})^{\otimes n}} \times \frac{d(\mu F^{-1})^{\otimes n}}{dl^n}.$$

So, by taking logarithms, we obtain formula (46), from which follow immediately formula (45), (ii), (iii) and (iv). □

Remark 5.6. Formula (46) implies that

$$H_n = H_{n,PM} + n \times [\tfrac{1}{2} \log(2\pi) + \tfrac{1}{2} \|F\|_2^2 - H_{\gamma_0}(\mu F^{-1})].$$

In the same way we have

$$\mathcal{I}_n(F) = \mathcal{I}_{n,G}(F) + \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{j=0}^{n-1} F^2 \circ T^j, \tag{47}$$

and therefore (cf. formulas (22) and (23))

$$H_n = H_{n,G} + n \times [\tfrac{1}{2} \log(2\pi) + \tfrac{1}{2} \|F\|_2^2].$$

As a corollary, we obtain the following criterion for independence.

COROLLARY 5.7. *$Se(F, T) = H_1(F)$ if and only if the process $(F \circ T^n)$ is independent.*

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Note that Corollary 5.7 can also be proved by using formulas (7) and (23).

Corollary 5.7 together with formula (23) give the following improvement of Corollary 4.5.

COROLLARY 5.8. *Let γ_0 be the probability measure with density $(1/(2\pi)^{(1/2)}) \cdot \exp(-(1/2)x^2)$ with respect to Lebesgue measure one on $\mathbb{R}$. Let $F \in AC(\Omega, T, \mu)$. Then*

$$Se(F, T) \leq -H_{\gamma_0}(\mu F^{-1}) + \tfrac{1}{2}(\log(2\pi) + \|F\|_2^2), \tag{48}$$

and the equality

$$Se(F, T) = -H_{\gamma_0}(\mu F^{-1}) + \tfrac{1}{2}(\log(2\pi) + \|F\|_2^2),$$

holds if and only if the process $(F \circ T^n)$ is independent.

Note, once more, that we see from this corollary that $Se(F, T) = (1/2)(\log(2\pi) + \|F\|_2^2)$ if and only if the process $(F \circ T^n)$ is Gaussian independent (cf. Corollary 4.5).

We can also prove the following, which, in particular, improves the inequality (48) in the preceding corollary, and gives a link between the Shannon entropy and information stability.

LEMMA 5.9. *If $\xi = (\xi_n)_{n\geq 1}$ and $\eta = (\eta_n)_{n\geq 1}$ are discrete time stationary processes, the rate of generation of information about η by ξ or about ξ by η (following Pinsker) is*

$$\bar{I}(\xi, \eta) := \lim_n \frac{1}{n} H_{P_{(\xi_1,\ldots,\xi_n)} \times P_{(\eta_1,\ldots,\eta_n)}}(P_{(\xi_1,\ldots,\xi_n),(\eta_1,\ldots,\eta_n)}).$$

The pair (ξ, η) is said to be information stable if $\bar{I}(\xi, \eta) = 0$.

Let (Ω, T, μ) be an invertible dynamical system and $F \in AC(\Omega, T, \mu)$. Let ϕ and π be the processes defined by $\phi_n = F \circ T^{-n+1}$, and $\pi_n = F \circ T^n$, for $n = 1, 2, \ldots$. Then:

(i) *the pair (ϕ, π) is information stable if and only if $\lim_n (H_n/n - H_{2n}/2n) = 0$, in particular if $Se(F, T)$ is finite then the pair (ϕ, π) is information stable;*

(ii) *$Se(F, T)$ is finite if and only if*

$$\sum_{p=0}^{\infty} \frac{1}{2^p} H_{\mu F_{2^p}^{-1} \times \mu F_{2^p}^{-1}}(\mu F_{2^{p+1}}^{-1}) < \infty;$$

moreover,

(iii)

$$Se(F, T) = \frac{1}{2}(\log(2\pi) + \|F\|_2^2) - H_{\gamma_0}(\mu F^{-1}) - \frac{1}{2} \sum_{p=0}^{\infty} \frac{1}{2^p} H_{\mu F_{2^p}^{-1} \times \mu F_{2^p}^{-1}}(\mu F_{2^{p+1}}^{-1}).$$

We also have the following.

Remark 5.10. Let (Ω, T, μ) be a dynamical system, $F \in AC(\Omega, T, \mu)$ and ξ the process $\xi := (F, F \circ T, \ldots)$. Then the entropy rate of ξ with respect to the independent Gaussian stationary process η is given by

$$\bar{H}_\eta(\xi) = H_{\gamma_0}(\mu F^{-1}) + \frac{1}{2} \sum_{p=0}^{\infty} \frac{1}{2^p} H_{\mu F_{2^p}^{-1} \times \mu F_{2^p}^{-1}}(\mu F_{2^{p+1}}^{-1}).$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

6. *Application to Gaussian processes*

In this section we express the Shannon entropy $Se(F, T)$ in terms of the spectral measure of the Gaussian process $X = (X_n)_{n\in\mathbb{Z}}$, when $F = X_0$ is the zero coordinate function (Lemma 6.3). This enables us (1) to prove that, in the class of Gaussian Markovian processes, the Shannon entropy almost determines the process (Remark 6.4), (2) to show how this entropy changes by linear change of variable (Corollary 6.5), and (3) to prove that all unilateral Gaussian processes with finite Shannon entropy are isomorphic (Theorem 6.6).

First, we need some preliminaries.

Let $(v_n)_{n\geq 0}$ be a stationary sequence of unit vectors in the real Hilbert space H, with $(r(n))_{n\in\mathbb{Z}}$ a strictly positive definite sequence (defining $r(-n) = r(n)$), where $r(n) = \langle v_n, v_0\rangle = \langle v_{n+k}, v_k\rangle$. Let R_n be the $n \times n$ matrix $(R_n)_{ij} = r(i - j)$, for $i, j = 0, \ldots, n - 1$, and r be the vector $r = [r(1), \ldots, r(n - 1)]^t$. Then

$$R_n = \begin{pmatrix} 1 & r^t \\ r & R_{n-1} \end{pmatrix}.$$

Evidently for each n there exists a unique vector $a = [a_1, \ldots, a_n]^t$ such that the vector

$$w_n := v_0 - \sum_{i=1}^{n-1} a_i v_i$$

is orthogonal to v_i for $i = 1, \ldots, n - 1$. Using the orthogonal decomposition

$$v_0 = w_n + (a_1 v_1 + \cdots + a_{n-1} v_{n-1}),$$

we can prove, by taking scalar products $\langle v_0, v_0\rangle, \ldots, \langle v_{n-1}, v_0\rangle$, that a is given by the equation $r = R_{n-1}a$, or

$$a = R_{n-1}^{-1} r.$$

LEMMA 6.1. *For any $X^t = (x_0, \ldots, x_{n-1})$ set $Y^t = (x_1, \ldots, x_{n-1})$. Then we have*

$$X^t R_n^{-1} X - Y^t R_{n-1}^{-1} Y = \frac{(x_0 - \sum_{j=1}^{n-1} a_j x_j)^2}{\|w_n\|^2}, \tag{49}$$

and

$$\det(R_n) = \|w_n\|^2 \det(R_{n-1}). \tag{50}$$

In the same way, note first that we have the equality

$$R_n = \begin{pmatrix} R_{n-1} & r^i \\ (r^i)^t & 1 \end{pmatrix},$$

where r^i is the vector whose transpose is $(r^i)^t = (r(n - 1), \ldots, r(1))$. Moreover, for any n there exists a unique vector $b = [b_0, \ldots, b_{n-2}]^t$ such that the vector u_n defined by

$$u_n = v_{n-1} - b_0 v_0 - \cdots - b_{n-2} v_{n-2}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

is orthogonal to v_j for $j = 0, \ldots, n-2$. So $v_{n-1} - u_n$ is the projection of v_{n-1} onto the subspace spanned by $v_0, \ldots, v_{n-2}$, and u_n is the projection of v_{n-1} onto the orthogonal of the linear span of $\{v_0, \ldots, v_{n-2}\}$. Since

$$v_{n-1} = u_n + (b_0 v_0 + \cdots + b_{n-2} v_{n-2}), \tag{51}$$

an 'orthogonal decomposition', we obtain, by taking scalar products $\langle v_0, v_{n-1}\rangle, \ldots, \langle v_{n-1}, v_{n-1}\rangle$, the equality $r^i = R_{n-1} b$, or

$$b = R_{n-1}^{-1} r^i.$$

We also have the following.

LEMMA 6.2. *If* $X = [x_0, \ldots, x_{n-1}]^t$ *and* $Y = [x_0, \ldots, x_{n-2}]^t$, *then*

$$X^t R_n^{-1} X - Y^t R_{n-1}^{-1} Y = \frac{(x_{n-1} - \sum_{j=0}^{n-2} b_j x_j)^2}{\|u_n\|^2}, \tag{52}$$

and

$$\det(R_n) = \|u_n\|^2 \det(R_{n-1}). \tag{53}$$

Proof. Let

$$Q = \begin{pmatrix} I_{n-1} & 0 \\ -b^t & 1 \end{pmatrix},$$

where I_{n-1} is the identity matrix. Then

$$Q^{-1} = \begin{pmatrix} I_{n-1} & 0 \\ b^t & 1 \end{pmatrix}$$

and we have

$$(Q^{-1})^t R_n^{-1} Q^{-1} = \begin{pmatrix} R_{n-1}^{-1} & 0 \\ 0 \ldots 0 & \alpha \end{pmatrix},$$

where $\alpha = 1/\|u_n\|^2$. In fact this equality is equivalent to

$$Q R_n Q^t = \begin{pmatrix} R_{n-1} & 0 \\ 0 \ldots 0 & \alpha^{-1} \end{pmatrix},$$

which can be easily verified. □

Now let $\Omega = \mathbb{R}^{\mathbb{Z}}$, σ be the shift transformation and μ be a Gaussian σ invariant probability measure determined by a (strictly) positive definite sequence $(r(n))_{n\in\mathbb{Z}}$, with $r(-n) = r(n)$ for any n, so when $r(0) = 1$ there exists a probability measure ν on the unit circle $\mathbb{T}$ such that $\hat{\nu}(n) = r(n)$. In this case we shall call ν the spectral measure. In other words, each n-dimensional distribution has a density ρ_n given by

$$\rho_n(x_0, \ldots, x_{n-1}) = \frac{1}{(2\pi)^{n/2} (\det R_n)^{1/2}} \times \exp\left(-\frac{1}{2} \sum_{i,j=0}^{n-1} (R_n^{-1})_{ij} x_i x_j \right), \tag{54}$$

where

$$(R_n)_{ij} = r(i-j) = \int_\Omega x_i x_j \, d\mu(x), \quad i, j = 0, \ldots, n-1.$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

So, if $F(x) = x_0$ for $x \in \Omega$, then, for $i \geq j$,

$$(R_n)_{ij} = \int_\Omega F \circ \sigma^{i-j} F \, d\mu.$$

LEMMA 6.3. *Let P_n, Q_n and Q denote the orthogonal projections onto the linear span of $\{X_1, \ldots, X_n\}$, $\{X_{-n+1}, \ldots, X_{-1}\}$ and $\{X_{-1}, X_{-2}, \ldots\}$, respectively. Then, we have the following.*

(a)

$$Se(F, \sigma) = \frac{1}{2}\log(2\pi) + \log \|F - QF\|_2 + \frac{1}{2} \tag{55}$$

$$= \frac{1}{2}\log(2\pi) + \frac{1}{2}\int \log f \, d\lambda + \frac{1}{2}, \tag{56}$$

where f is the density of ν with respect to λ.

In particular, if $\|F - QF\| = 0$ then $Se(F, \sigma) = -\infty$.

(b) *If $\|F - QF\| > 0$, the following are equivalent:*

(i) *the Shannon information $(1/n)\mathcal{I}_n(F)$ converges almost everywhere (respectively in L^1);*

(ii) $(1/N)\sum_{j=1}^{N-1}(F - P_jF)^2(\sigma^{N-j})$ *converges almost everywhere (respectively in L^1);*

(iii) $(1/N)\sum_{j=1}^{N-1}(F - P_jF)^2(\sigma^{-j})$ *converges almost everywhere (respectively in L^1);*

(iv) $(1/N)\sum_{j=1}^{N-1}(F - Q_jF)^2(\sigma^{j})$ *converges almost everywhere (respectively in L^1);*

(v) $(1/N)\sum_{n=2}^{N}\log(\rho_n(x)/\rho_n(\sigma x))$ *converges almost everywhere (respectively in L^1).*

Proof. By formula (54) and Lemma 6.1, we get

$$-\log\frac{\rho_n(x_0, \ldots, x_{n-1})}{\rho_{n-1}(x_1, \ldots, x_{n-1})} = \frac{1}{2}\log(2\pi) + \log\|F - P_{n-1}F\|_2 + \frac{1}{2}\frac{(F - P_{n-1}F)^2(x)}{\|F - P_{n-1}F\|_2^2}. \tag{57}$$

Now, since the process is Gaussian, $P_{n-1}F$ converges to PF almost everywhere and in L^2, where P is the projection onto the linear span of $\{F \circ \sigma, F \circ \sigma^2, \ldots\}$, and $PF = E(F \mid \sigma^{-1}\mathcal{B})$, where $\mathcal{B}$ is the Borel sigma-algebra. So, in the case where $\|F - PF\|_2 > 0$, it follows from (57) that

$$\lim_n\left[-\log\frac{\rho_n(x)}{\rho_{n-1}(\sigma x)}\right] = \frac{1}{2}\log(2\pi) + \log\|F - PF\|_2 + \frac{1}{2}\frac{(F - PF)^2(x)}{\|F - PF\|_2^2}. \tag{58}$$

Then the equality

$$-\frac{1}{N}\sum_{n=3}^{N}\log\frac{\rho_n(x)}{\rho_{n-1}(\sigma x)} = -\frac{1}{N}\sum_{n=3}^{N-1}\log\left(\frac{\rho_n(x)}{\rho_n(\sigma x)}\right) - \frac{1}{N}\log\rho_N(x) + \frac{1}{N}\log\rho_2(\sigma x) \tag{59}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

proves that $-(1/N)\log\rho_N(x)$ converges almost surely (respectively in L^1) if and only if $-(1/N)\sum_{n=3}^{N-1}\log(\rho_n(x)/\rho_n(\sigma x))$ converges almost surely (respectively in L^1). Now, by (57) and (59), we obtain the equality (56).

In the same way, we get, by Lemma 6.2,

$$-\log\frac{\rho_n(x_0,\ldots,x_{n-1})}{\rho_{n-1}(x_0,\ldots,x_{n-2})}=\frac{1}{2}\log(2\pi)+\frac{1}{2}\log\frac{\det R_n}{\det R_{n-1}}+\frac{1}{2}\frac{(x_{n-1}-\sum_{j=0}^{n-2}b_jx_j)^2}{\|u_n\|^2}.$$

But, if L_{n-2} is the orthogonal projection onto the linear span of $X_0,\ldots,X_{n-2}$, we have $\|u_n\|^2=\|X_{n-1}-L_{n-2}X_{n-1}\|_2^2=\|F-Q_nF\|_2^2$, and thus

$$\begin{aligned}&\frac{1}{N-1}[-\log\rho_N(x_0,\ldots,x_{N-1})+\log\rho_1(x_0)]\\&\quad=\frac{1}{2}\log(2\pi)+\frac{1}{N-1}\sum_{n=2}^{N}\log\|F-Q_nF\|_2\\&\qquad+\frac{1}{2(N-1)}\sum_{n=2}^{N}\frac{(F-Q_nF)^2\circ\sigma^{n-1}(x)}{\|F-Q_nF\|_2^2}.\end{aligned}$$

Then, in the case where $\|F-QF\|>0$, the sequence of Shannon information $(1/n)\mathcal{I}_n(F)$ converges almost everywhere (respectively in L^1) if and only if $(1/N)\sum_{j=1}^{N-1}(F-Q_jF)^2(\sigma^j)$ does so.

The other statements can be proved in a similar way. □

We can deduce easily the following observation from (55).

COROLLARY 6.4. *Let (X_n) and (Y_n) be stationary centered Gaussian Markovian processes, with the same L^2 norm. Then they have the same Shannon entropy if and only if either they have the same law, or (Y_n) and $((-1)^nX_n)$ have the same law.*

Proof. Let ν be the spectral measure of X and ν' be the spectral measure of Y. Then if $P_r(t)=\sum_{n\in\mathbb{Z}}r^{|n|}e^{\mathrm{int}}$ is the Poisson kernel, we have $\nu=P_r(t)\,d\lambda(t)$ for some r and similarly $\nu'=P_{r'}(t)\,d\lambda(t)$ for some r'. On the other hand, from (55), the equality $Se(X_0,\sigma)=Se(Y_0,\sigma)$ holds if and only if $\|X_0-QX_0\|=\|Y_0-Q'Y_0\|$, where Q' denotes the projection to the negative coordinates of Y. But $QX_0=aX_{-1}$ and similarly $Q'Y_0=bY_{-1}$, for some constants a,b. Thus, the equality of the respective Shannon entropies is equivalent to $|a|=|b|$, or to $a\langle X_{-1},X_0\rangle=b\langle Y_{-1},Y_0\rangle$, that is to $ar=br'$. (Also, one can show by elementary calculus that $\int_{\mathbb{T}}\log P_r(t)\,d\lambda(t)=\int_{\mathbb{T}}\log P_{r'}(t)\,d\lambda(t)$ if and only if $|r|=|r'|$.) □

COROLLARY 6.5. *Let $(X_n)_{n\in\mathbb{Z}}$ be a Gaussian stationary process with spectral measure ν. Let $g=\sum_{n\in\mathbb{Z}}a_ne^{\mathrm{int}}\in L^2(\nu)$ and $(Y_n)_{n\in\mathbb{Z}}$ be the stationary Gaussian process such that $Y_0=\sum_{n\in\mathbb{Z}}a_nX_n$. Then*

$$Se(Y_0,\sigma)=Se(X_0,\sigma)+\int_{\mathbb{T}}\log(|g|)\,d\lambda.$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Proof. By the spectral theorem, if ν' is the spectral measure of (Y_n), we have $\nu' = |g|^2\nu$. So, by the Szego theorem

$$\|Y_0 - Q'Y_0\|^2 = \exp\left[\int \log\left(|g|^2 \frac{d\nu}{d\lambda}\right) d\lambda\right] = \exp\left[\int \log(|g|^2)\, d\lambda\right] \|X_0 - QX_0\|^2.$$

Thus, by (56), we get the result. □

Note that, if $g \in L^1(\lambda)$, then $\int_{\mathbb{T}} \log(|g|)\, d\lambda$ is finite if and only if there is an $h \in H^1$ such that $|g| = |h|$, and in this case, $\int \log(|g|)\, d\lambda \geq |h(0)| = |\int h\, d\lambda|$.

In particular, if $a_n = 0$ for $n > 0$ (or for $n < 0$) and $g \in L^1(\lambda) \cap L^2(\nu)$ then $\int \log(|g|)\, d\lambda \geq \log|\int g\, d\lambda| = \log|a_0|$.

More particularly, if $g \in \mathbb{H}^1(\mathbb{T})$ is an outer function, then $\int \log(|g|)\, d\lambda = \log|\int g\, d\lambda|$, and when, in addition, $\int g\, d\lambda = 0$, we obtain $Se(Y_0, \sigma) = Se(X_0, \sigma)$.

It is well known from Ornstein theory that a bilateral Gaussian process $X = (X_n)_{n\in\mathbb{Z}}$ with spectral measure absolutely continuous with respect to Lebesgue measure on the circle is isomorphic to the Gaussian independent process. There is interest in considering isomorphism for non-invertible transformations (endomorphisms). The first examples of such isomorphisms have been worked out by Parry [**13**] and elaborated by Hoffman and Rudolph [**6**]. (All the endomorphisms they consider are finite to one.) We consider now the unilateral transformation (endomorphism) associated to the Gaussian process with spectral measure equivalent to Lebesgue measure λ. For clarity, if $X = (X_n)_{n\geq 0}$ is a Gaussian process with spectral measure $\nu = f\,d\lambda$ we consider the endomorphism T_ν defined on $\mathbb{R}^{\mathbb{N}}$ by $(T_\nu x)_n = x_{n+1}$, for $x \in \mathbb{R}^{\mathbb{N}}$ and $n \geq 0$. Then the shift T_ν will be isomorphic to the shift T_λ if and only if $\log f$ is Lebesgue integrable.

To prove this we recall some useful properties that functions in $\mathbb{H}^1$ or in $\mathbb{H}^2$ can have. First recall that, for $p = 1, 2$, $\mathbb{H}^p$ is the closed subspace of all $f \in L^p(\mathbb{T}, (d\theta/2\pi))$ such that $\int_{-\pi}^{\pi} f(t)e^{\text{int}}dt = 0$, $n = 1, 2, \ldots$, and that if $0 \leq f \in L^1$ then $\log f$ is integrable if and only if there is an $F \in \mathbb{H}^2$ such that $f = |F|^2$ [**5**, Theorem, p. 53]. Recall also that an inner function f is an analytic function in the unit disc such that $|f(z)| \leq 1$ and $|f(e^{i\theta})| = 1$ almost everywhere on the unit circle, and an outer function F is an analytic function in the unit disc of the form

$$F(z) = \alpha \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} k(\theta)\, d\theta\right],$$

where k is a real-valued integrable function on the circle and α is a complex number with modulus 1 [**5**, p. 63]. For a function $F \in \mathbb{H}^2$ to be an outer function it is necessary and sufficient that the family $\{z^n F : n = 0, 1 \ldots\}$ span $\mathbb{H}^2$ [**5**, Corollary, p. 101]. Also any non-zero function $f \in \mathbb{H}^1$ can be written in the form $f = gF$ where g is inner and F is outer [**5**, Theorem, p. 63, **4**, Theorem 12].

In the next theorem the use of Shannon entropy is only to ensure that the logarithm of the density of the spectral measure is integrable.

THEOREM 6.6. *Let ν be a probability measure on the unit circle, equivalent to Lebesgue measure λ, with density f. Then the unilateral shifts T_ν and T_λ are isomorphic if and only if $Se(X_0, T_\nu)$ is finite, or equivalently* $\log f$ *is Lebesgue integrable.*

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Proof. Consider the two bilateral Gaussian processes X' and Y' with spectral measures λ and $f d\lambda$, respectively. Then, on the cyclic space $Z_{X_0} = \{X_0 \circ T^n : n \in \mathbb{Z}\}$, T is unitarily equivalent to the multiplication M_z by z on $L^2(\lambda)$. An isomorphism ϕ is given by $\phi(X_n) = z^n$, $n \in \mathbb{Z}$. Similarly, the same holds for T on the cyclic space $Z_{Y_0} = \{Y_0 \circ T^n : n \in \mathbb{Z}\}$ and the multiplication by z on $L^2(f d\lambda)$, with isomorphism ψ: $\psi(Y_n) = z^n$, $n \in \mathbb{Z}$. It follows that the action of T on Z_{X_0} is unitarily equivalent to the action of T on Z_{Y_0}. Suppose first that $\log f$ is integrable. Then there exists $F \in \mathbb{H}^2$ such that $\sqrt{f} = |F|$. Moreover, there exist an inner function g and an outer function $G \in \mathbb{H}^2$ such that $F = gG$ and thus $|F| = |G|$. Set $x = \phi^{-1}G$, so that x belongs to the closed linear span of $\{X_0, X_1, \ldots\}$, and we have

$$\begin{aligned}\langle T^n x, x\rangle &= \langle T^n \phi^{-1} G, \phi^{-1} G\rangle = \langle \phi^{-1} M_z^n G, \phi^{-1} G\rangle \\ &= \langle M_z^n G, G\rangle = \int z^n |G|^2 d\lambda = \int z^n f d\lambda = \langle T^n Y_0, Y_0\rangle.\end{aligned}$$

On the other hand, if $P = \sum_k a_k z^k$ is a polynomial, the equalities

$$\begin{aligned}\left\| X_0 - \sum_k a_k T^k x \right\| &= \left\| \phi(X_0) - \sum_k a_k \phi(T^k x) \right\| = \left\| 1 - \sum_k a_k M_z^k \phi(x) \right\| \\ &= \left\| 1 - \sum_k a_k M_z^k G \right\| = \left\| 1 - \sum_k a_k z^k G \right\| = \|1 - PG\|\end{aligned}$$

prove that X_0 belongs to the closed linear space generated by $\{T^n x \mid n \geq 0\}$ if and only if 1 belongs to the closed linear space $Z(G)$ generated by $\{z^n G \mid n \geq 0\}$. But, since G is outer, $Z(G)$ is equal to $\mathbb{H}^2$ and thus $X_0 \in \overline{\text{lin}}\{T^n x \mid n \geq 0\}$. This proves that T_ν and T_λ are isomorphic. The other implication follows from the Szegö theorem. □

We end this section with the following result concerning the speed of convergence in linear prediction.

PROPOSITION 6.7. *Let λ be the Lebesgue probability measure on $\mathbb{T}$, and ν the spectral measure of a stationary Gaussian process $(X_n)_{n\in\mathbb{Z}}$. Let Q denote the orthogonal projection onto the closed (in $L^2(\nu)$) linear span of the negative coordinates and let Q_n be the orthogonal projection onto the linear span of $\{X_{-n}, \ldots, X_{-1}\}$. Suppose that $X_0 \neq QX_0$, or equivalently $\log(d\nu/d\lambda)$ is Lebesgue integrable. Then the series $\sum_{n=1}^\infty \|QX_0 - Q_n X_0\|_2^2$ converges if and only if ν is absolutely continuous and $\nu = e^f d\lambda$, with $\sum_{n=1}^\infty n|\hat{f}(n)|^2 < \infty$.*

An equivalent form of Proposition 6.7 is the following.

Remark 6.8. Let ν be a probability measure on $\mathbb{T}$. Let H and H_n denote the closed subspaces of $L^2(\nu)$ spanned by $\{e^{ikt} \mid k \geq 1\}$ and $\{e^{ikt} \mid 1 \leq k \leq n\}$, respectively. Let F, F_n be the orthogonal projection of the constant function 1 onto H and H_n respectively. Suppose that 1 is not in H. Then the following are equivalent:

(i) $\sum_{n=1}^\infty \|F - F_n\|^2 < \infty$;

(ii) ν is absolutely continuous with respect to the Lebesgue probability measure λ and $\nu = e^f d\lambda$, with $\sum_{n=1}^\infty n|\hat{f}(n)|^2 < \infty$.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

7. $\mathbb{Z}^2$ *action, pointwise statement*

In this section we consider specifically absolutely continuous $\mathbb{Z}^n$ processes, for which we prove pointwise convergence of the Shannon entropy. The case where $n = 1$ has already been considered by Barron [**1**]. However, his method cannot extend to the higher-dimensional case; the idea of our proof is very close to that by Ornstein and Weiss [**10**] for the $\mathbb{Z}^n$ version of the Shannon–MacMillan–Breiman theorem. The proof is given for $n = 2$, but it can be easily generalized.

The notation is as in §4. In particular, we refer to (16) and (18) for $h_n^{(2)}$ and to Definition 4.11, Lemmas 4.10 and 4.15 for $Se(F, T, S)$. Namely, f_{n^2} is the density with respect to Lebesgue measure of the law of $F_{n^2} := (F \circ T^m \circ S^n)_{m,n=0,\dots,n-1}$, and $h_n^{(2)} = -\log f_{n^2} \circ F_{n^2}$. The aim of this section is to prove the following theorem.

THEOREM 7.1. *Let T, S be commuting measure-preserving transformations on the probability space $(\Omega, \mathcal{F}, \mu)$ with ergodic joint action. Let $F \in L^2(\mu)$ such that the process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{Z}^2}$ is absolutely continuous, with law ν. Let ν_0 be the law of the process $(F \circ T^m \circ S^n)_{(m,n)\in\mathbb{N}^2,(m,n)\neq(0,0)}$.*

Assume that $Se(F, T, S)$ is finite (which is equivalent to $H_{\mu F^{-1}\times\nu_0}(\nu) < \infty$). Then $(1/n^2)h_n^{(2)}$ converges almost everywhere and in $L^1(\mu)$ to $Se(F, T, S)$.

In the case when $Se(F, T, S) = -\infty$ the previous convergence still holds almost everywhere.

Proof. We establish first the invariance of $\liminf(1/n^2)h_n^{(2)}$. Next, with the help of a reduction, we prove that this lim inf is in fact almost everywhere a limit.

Recall that, for every n, $X_{n,n}^{n+1} = (x_{s,t})_{s,t=0,\dots,n}$ is, as in §4, formula (a_0).

CLAIM 7.2. *Let $h_* := \liminf_n (1/n^2)h_n^{(2)}$. Then h_* is invariant by each action.*

Proof. Let $K_{(0,n)\times(0,n)} := X_{n,n}^{n+1}$, $K_{(1,n)\times(0,n-1)} := \{x_{i,j} \mid 1 \le i \le n, 0 \le j \le n-1\}$ and $w_n = (1/n^2)h_n^{(2)}$. Then

$$w_n \circ S - \frac{(n+1)^2}{n^2} w_{n+1} = y_n + z_n,$$

where

$$y_n := (1/n^2) \log \frac{f(K_{(0,n)\times 0,n)})}{\prod_{j=0}^n f(x_{0,j}) \times \prod_{i=1}^n f(x_{i,n}) \times f(K_{1,n)\times 0,n-1)})},$$

and $z_n := (1/n^2)\log(\prod_{j=0}^n f(x_{0,j}) \times \prod_{i=1}^n f(x_{i,n}))$. Now z_n converges to zero almost everywhere by the pointwise ergodic theorems. For the first one y_n, define, for $\epsilon > 0$, $A_n(\epsilon) = A_n$ by

$$A_n := \left\{ X : f(X_{n,n}^{n+1}) \le e^{-n^2\epsilon} \prod_{j=0}^n f(x_{0,j}) \times \prod_{i=1}^n f(x_{i,n}) \times f(K_{(1,n)\times(0,n-1)}) \right\}.$$

Then, a simple calculation yields $\mu(A_n) \le e^{-n^2\epsilon}$ and thus, for μ almost all x, there is a p such that $-y_n(x) \le \epsilon$ for all $n \ge p$, and this implies that $h_* \le \epsilon + h_* \circ S$. Hence $h_* \le h_* \circ S$. It follows that $h_* = h_* \circ S$.

In the same way, we also have $h_* = h_* \circ T$ and this proves the claim. □

We now prove the following.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

CLAIM 7.3. *If* $\lim_n (1/n^2)H_n^{(2)}$ *is finite then the family* $\{(1/n^2)h_n^{(2)} : n \geq 1\}$ *is* μ *uniformly integrable.*

Proof. Recall that ν and μ are defined by their respective marginals ν_n and μ_n as in (20) and (21), respectively, and $g_{i,j}^n$ is as in (32). Also L is as in (a_2) in §4.1. We prove that the family $\{\log g_{i,j}^n : (i, j, n) \in L\}$ is μ uniformly integrable, and this will imply, by the equality (E) in Remark 4.12, that $((1/n^2)h_n^{(2)})$ is μ uniformly integrable. For $l = (i, j, n)$, denote $g_{i,j}^n$ by ρ_l, and let $l_k = (i_k, j_k, n_k)$ be an infinite sequence in L. We shall prove that $(\log \rho_{l_k})$ contains a weakly convergent subsequence, and this proves the claim. By Remark 4.1(2), (l_k) contains a strictly increasing subsequence which we still denote by (l_k). Let $\mathcal{F}_\infty := \bigvee_k \mathcal{F}_{l_k}$. By the formula (34) and Remark 4.13, ρ_{l_k} converges ν almost everywhere to $\rho_\infty := d\mu_\infty/d\nu_\infty$, where μ_∞ and ν_∞ are the restrictions of μ and ν to $\mathcal{F}_\infty$, respectively. Also we have

$$0 \geq \int -\log \rho_\infty \, d\mu = -H_{\nu_\infty}(\mu_\infty) \geq -H_\nu(\mu) > -\infty.$$

That is, $\log \rho_\infty$ is μ integrable.

But $\sup_k \int \rho_{l_k} \log \rho_{l_k} d\mu \leq \sup_l \int \rho_l \log \rho_l \, d\mu < \infty$. Hence, by Remark 4.14, $\{(\log \rho_{l_k})^+ : k \geq 1\}$ is uniformly integrable with respect to μ. Set $Y_k = \log \rho_{l_k}$ and $Y = \log \rho_\infty$, so that $Y_k, Y \in L^1(\mu)$, Y_k converges μ almost everywhere to Y and $\int Y_k \, d\mu$ converges to $\int Y d\mu$. It follows that Y_k^+ converges μ almost everywhere to Y^+, and thus, because (Y_k^+) is μ uniformly integrable, the convergence holds in $L^1(\mu)$ too. In particular, $\int Y_k^+ d\mu$ converges to $\int Y^+ d\mu$. So $\int Y_k^- d\mu$ converges to $\int Y^- d\mu$, and thus, since Y_k^- converges μ almost everywhere to Y^-, it converges in $L^1(\mu)$. This proves that (Y_k) converges in $L^1(\mu)$ and *a fortiori* it is μ uniformly integrable. □

Now we proceed to prove that $(1/n^2)h_n^{(2)}$ converges μ almost everywhere. We begin by showing the following.

CLAIM 7.4. *We can reduce ourselves to the case where the density of the law of the first coordinate is greater than one on its support, and also where* $\liminf_n (1/n^2)h_n^{(2)} < 0$.

Proof. Let $F : \mathbb{R}^{\mathbb{Z}\times\mathbb{Z}} \to \mathbb{R}$ be the projection to the $(0, 0)$ coordinate, with absolutely continuous law with density $f_0 = d\mu F^{-1}/d\lambda$, and $\alpha > 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be the map defined by

$$\phi(x) = \frac{1}{\alpha} \int_{-\infty}^{x} f_0(t) \, dt.$$

Put $G = \phi \circ F$. Then the law of G is absolutely continuous and has density g_0 given by $g_0 = \alpha 1_{[0,1/\alpha]}$.

Now let $\Phi : \mathbb{R}^{\mathbb{Z}\times\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}\times\mathbb{Z}}$ be the map defined by $(\Phi(x))_{i,j} = \phi(x_{i,j})$, for all $i, j \in \mathbb{Z}$, so that $\Phi S^m T^n = S^m T^n \Phi$, and let $m := \mu \circ \Phi^{-1}$. Define θ by

$$\theta(x) = \sup\{t \in \mathbb{R} \mid \phi(t) = x\}.$$

Then the finite-dimensional marginals of m are absolutely continuous with respect to Lebesgue measure and the following relationship holds between the densities f for μ and g for m:

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

$$g((u_{i,j})_{i,j=0,\dots,n-1}) = f((\theta(u_{i,j}))_{i,j=0,\dots,n-1}) \prod_{i,j=0,\dots,n-1} \theta'(u_{i,j}),$$

with the property that, for almost all t, $g_0(t) > 0 \Rightarrow g_0(t) \geq \alpha$. It follows that

$$-\frac{1}{n^2} \log g((u_{i,j})_{i,j=0,\dots,n-1}) = -\frac{1}{n^2} \log f((\theta(u_{i,j}))_{i,j=0,\dots,n-1}) - \frac{1}{n^2} \sum_{i,j=0}^{n-1} \log \theta'(u_{i,j}).$$

But it is easy to see that $\log \theta' \circ F \in L^1(m)$ if and only if $\int f_0(t) \log f_0(t)\, dt$ is finite. Also

$$\int \log \theta' \circ F\, dm = \log \alpha - \int f_0(t) \log f_0(t)\, dt.$$

So if $\int f_0(t) \log f_0(t)\, dt$ is finite, the sequence

$$\frac{1}{n^2} \sum_{i,j} \log \theta'(u_{i,j}) = \frac{1}{n^2} \sum_{i,j} \log \theta' \circ F \circ S^i T^j(u)$$

converges m almost everywhere to $\int \log \theta' \circ F\, dm$.

Therefore, $-(1/n^2) \log g((u_{i,j})_{i,j=0,\dots,n-1})$ converges m almost everywhere if and only if $-(1/n^2) \log f((\theta(u_{i,j}))_{i,j=0,\dots,n-1})$ also does so. In this case the corresponding limits (or lim inf), h_* and g_* verify

$$g_* = h_* - \log \alpha + \int f_0(t) \log f_0(t)\, dt.$$

So if h_* is at least zero, then if we take α such that

$$\log \alpha > h_* + \int f_0(t) \log f_0(t)\, dt,$$

we obtain $g_* < 0$.

Therefore, we can suppose that, for almost all t, if $f_0(t) > 0$ then $f_0(t) \geq \alpha \geq 1$ and $h_* := \liminf (1/n^2) h_n < 0$.

This finishes the proof of the announced reduction. □

As in the Ornstein–Weiss case, we prove the following.

CLAIM 7.5. $\liminf (1/n^2) h_n^{(2)}$ *is almost surely a limit.*

Proof. Put $a := h_*$. By Claim 7.4, we can and do suppose that $a < 0$. Let $\epsilon_3 > 0$, and $0 < \epsilon < \epsilon_3$. Let $l_0 \in \mathbb{N}$, $\epsilon_1, \epsilon_2 > 0$ and $\delta_1 > 0$, to be chosen later.

First, by the pointwise ergodic theorem, we find two sequences (k_l) and (m_l) of natural numbers converging to ∞ such that (m_l/k_l) converges to ∞ as fast as we wish, a set $\Omega_\epsilon^1 \subset \Omega$, with $\mu(\Omega_\epsilon^1) > 1 - \delta_1$, and an integer $N_0(\epsilon_1)$ such that:

(i) for all $x \in \Omega_\epsilon^1$, for all $N \geq N_0(\epsilon_1)$, there exists $J_N(x) \subset I_{N-1,N-1}^N$, $N^2(1 - \delta_1 - \epsilon_1) \leq \text{card } J_N(x)$; and

(ii) for all l, for all $(i, j) \in J_N(x)$, there exists $n(i, j, l) \in [k_l, m_l]$, $e^{-n^2(i,j,l)\,(a+\epsilon)} < f((i, j) + X_{n(i,j,l)-1,n(i,j,l)-1}^{n(i,j,l)})$.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Next, by repeated uses of a Vitali covering type lemma, we get an upper estimate of the (Lebesgue) size of the set Ω_ϵ^1, which enables us to majorize the measure of the set

$$\{x \mid f(X^N_{N-1,N-1}) \le e^{-N^2(a+\epsilon_3)}\} \cap \Omega_\epsilon^1,$$

which will ensure the convergence of the series

$$\sum_N \mu(\{x \mid f(X^N_{N-1,N-1}) \le e^{-N^2(a+\epsilon_3)}\} \cap \Omega_\epsilon^1),$$

in order to get $\limsup_n (1/n^2) h_n^{(2)} \le h_*$. The details are as follows.

It is easy to see, from the definition of a, that, given $\epsilon_0 > 0$, there exist a set Ω_ϵ, with $\mu(\Omega_\epsilon) \ge 1 - \epsilon_0$, and two strictly increasing sequences of natural numbers k_l, m_l, with (m_l/k_l) converging to infinity as fast as one wishes, such that for all l, for all $x \in \Omega_\epsilon$, there exists $n_l = n_l(x) \in [k_l, m_l]$, $e^{-n_l^2(a+\epsilon)} < f(x) = f(X^{n_l}_{n_l-1,n_l-1})$.

Let $J_N(x) := \{(i, j) : 0 \le i, j < N;\ T^i S^j x \in \Omega_\epsilon\}$. Then, by the pointwise ergodic theorem, for any $\delta > 0$, there is a measurable set A_δ with $\mu(A_\delta) > 1 - \delta$, and a natural number $N_0(\epsilon_1, \epsilon, \delta)$, such that for all $N \ge N_0(\epsilon_1, \epsilon, \delta)$, and for all $x \in A_\delta$, the following holds:

$$N^2(\mu(\Omega_\epsilon) - \epsilon_1) < \text{ card } J_N(x) \le N^2(\mu(\Omega_\epsilon) + \epsilon_1).$$

Put $\delta_1 = \epsilon_0 + \delta$, and $\Omega_\epsilon^1 := \Omega_\epsilon \cap A_\delta$.

It follows that $\mu(\Omega_\epsilon^1) \ge 1 - \delta_1$, and for $x \in \Omega_\epsilon^1$ and $(i, j) \in J_N(x)$ there exists $n(i, j, l) = n(i, j, l)\ (x) \in [k_l, m_l]$ such that

$$e^{-n^2(i,j,l)\ (a+\epsilon)} < f((i, j) + X^{n(i,j,l)}_{n(i,j,l)-1,n(i,j,l)-1}), \tag{59}$$

where we have denoted $f(S^i T^j x)$ by $f((i, j) + X^{n(i,j,l)}_{n(i,j,l)-1,n(i,j,l)-1})$.

Note that $R_{i,j,l}(x) := (i, j) + X^{n(i,j,l)}_{n(i,j,l)-1,n(i,j,l)-1}$ is the square with first vertex (i, j) and with side having length $n(i, j, l)$. Then, for any l, $\{R_{i,j,l}(x) : (i, j) \in J_N(x)\}$ is a finite cover of $J_N(x)$ by squares. So, due to the freedom, mentioned above, in the choice of k_l and m_l, by repeated applications of the Vitali covering lemma [**8**], for any ϵ_2 there exists $N_1(\epsilon_2, \epsilon)$ and $l_1, \ldots, l_k \ge l_0$, such that, for any $N \ge N_1(\epsilon_2)$, there exist subsets $J_{N,l_1}(x), \ldots, J_{N,l_k}(x)$ of $J_N(x)$, such that the squares $\{R_{i,j,l_s}(x) : (i, j) \in J_{N,l_s}(x), s = 1, \ldots, k\}$ are disjoint, and there is a subset $J_N^0(x) \subset J_N(x)$ which is covered by $\{R_{i,j,l_s}(x) : (i, j) \in J_{N,l_s}(x), s = 1, \ldots, k\}$ and with $(1 - \epsilon_2)$ card $J_N(x) \le$ card $J_N^0(x)$. It follows that, for $N \ge N_0(\epsilon_1, \epsilon, \delta)) \vee N_1(\epsilon_2, \epsilon)$ and $x \in \Omega_\epsilon^1$, we have

$$N^2 u \le \text{ card } J_N^0(x),$$

where $u := (1 - \epsilon_2)\ (1 - \delta_1 - \epsilon_1)$. Then

$$N^2 u \le \sum_{s=1}^{k} \sum_{(i,j) \in J_{N,l_s}} n^2(i, j, l_s)\ (x). \tag{60}$$

Thus, since $a < 0$, if $\epsilon > 0$ is chosen such that $a + \epsilon < 0$, then

$$N^2 u(a + \epsilon) \ge \sum_{s=1}^{k} \sum_{(i,j) \in J_{N,l_s}} n^2(i, j, l_s)\ (x)\ (a + \epsilon). \tag{61}$$

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

Let $J_N^1(x)$ be the intersection with $I_{N-1,N-1}^N$ of the union of the squares $\{R_{i,j,l_s}(x) : (i, j) \in J_{N,l_s}(x), s = 1, \ldots, k\}$. Then if $(i, j) \notin J_N^1(x)$ we have that $1 \leq (1/\alpha) f(x_{i,j})$. In particular,

$$\alpha \geq 1 \Rightarrow 1 \leq f(x_{i,j}) \quad \text{for all } (i, j) \notin J_N^1(x).$$

Since $\alpha \geq 1$, it follows, by (59), that

$$1 \leq \exp\left(\sum_{s=1}^{k} \sum_{(i,j)\in J_{N,l_s}(x)} n^2(i, j, l_s)\,(x)\,(a+\epsilon)\right)\left(\prod_{s=1,\ldots,k,(i,j)\in J_{N,l_s}(x)} f(R_{i,j,l_s}(x))\right) \times \prod_{(i,j)\notin J_N^1(x)} f(x_{i,j}).$$

So by (61), we get, for $x \in \Omega_\epsilon^1$,

$$1 \leq e^{N^2 u(a+\epsilon)}\left(\prod_{s=1,\ldots,k,(i,j)\in J_{N,l_s}(x)} f(R_{i,j,l_s}(x))\right) \times \prod_{(i,j)\notin J_N^1(x)} f(x_{i,j}). \tag{62}$$

But the number of all configurations of such disjoint squares is majorized by $C_{N^2}^{[N^2\beta]}$, which is majorized by $ce^{N^2 h(\beta,1-\beta)}$, where $\beta = 1/q^2$, q being the smallest k_l, and c is a constant. Then by (62), the Lebesgue measure $\lambda(\Omega_\epsilon^1)$ of Ω_ϵ^1 is, for N large enough, majorized by $ce^{N^2 h(\beta,1-\beta)} \times e^{N^2 u(a+\epsilon)}$. But it can be easily shown that

$$\mu(\{f(X_{N-1,N-1}^N) \leq e^{-N^2(a+\epsilon_3)}\} \cap \Omega_\epsilon^1) \leq e^{-N^2(a+\epsilon_3)}\lambda(\Omega_\epsilon^1).$$

So for N large enough, we then obtain $\mu(\{f(X_{N-1,N-1}^N) \leq e^{-N^2(a+\epsilon_3)}\} \cap \Omega_\epsilon^1) \leq ce^{-N^2\omega}$, where $\omega = a + \epsilon_3 - h(\beta, 1-\beta) - u(a+\epsilon)$. But there is a constant $\gamma > 0$, such that, for N large enough, the exponent ω is greater than γ. In fact, if we put $v = \epsilon_1 + \delta_1$, then $u = 1 - v - \epsilon_2(1-v)$, and thus $\omega = \epsilon_3 - \epsilon - h(\beta, 1-\beta) + (a+\epsilon)(v + \epsilon_2(1-v))$, so that, for $0 < \gamma < (\epsilon_3 - \epsilon)/2$, we can choose ϵ_1, ϵ_2, δ_1, l_0 and $N_2 \geq N_0 \vee N_1$, such that, for all $N \geq N_2$, we have the inequality

$$\epsilon_3 - \epsilon > 2\gamma > \gamma > h(\beta, 1-\beta) - (a+\epsilon)\,(v + \epsilon_2(1-v)).$$

Then the series

$$\sum_N \mu(\{f(X_{N-1,N-1}^N) \leq e^{-N^2(a+\epsilon_3)}\} \cap \Omega_\epsilon^1)$$

is convergent. Letting $\delta_1 \to 0$, we get $\limsup_N w_N(x) \leq a + \epsilon_3$, and this finishes the proof in the case when $-\infty < a := \liminf_N w_N < 0$.

If $\liminf w_N = -\infty$, taking a any negative number, the same proof gives $\limsup_N w_N \leq a$. □

This finishes the proof of Theorem 7.1. □

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press

REFERENCES

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

[2] J. L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

[3] H. A. Dye. On groups of measure preserving transformations. *I. Amer. J. Math.* **81** (1959), 119–159.

[4] H. Helson. *Harmonic Analysis*. Addison-Wesley, Reading, MA, 1983.

[5] K. Hoffman. *Banach Spaces of Analytic Functions*. Prentice-Hall, Englewood Cliffs, NJ, 1962.

[6] C. Hoffman and D. J. Rudolph. Uniform endomorphisms which are isomorphic to Bernoulli shift. *Ann. Math.* (2) **156**(1) (2002), 79–101.

[7] A. N. Kolmogorov. Stationary sequences in Hilbert space. *Bull. Math. Univ. Moscow* **2**(6) (1941), 1–4.

[8] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces, Fractals and Rectifiability (Cambridge Studies in Advanced Mathematics, 44)*. Cambridge University Press, Cambridge, 1995.

[9] D. S. Ornstein. *Ergodic Theory, Randomness, and Dynamical Systems*. Yale University Press, New Haven, 1974.

[10] D. S. Ornstein and B. Weiss. The Shannon–Mc Millan–Briman theorem for a class of amenable groups. *Israel J. Math.* **44**(3) (1983), 53–60.

[11] W. Parry. *Entropy and Generators in Ergodic Theory*. W. A. Benjamin, New York, 1969.

[12] W. Parry. *Topics in Ergodic Theory*. Cambridge University Press, Cambridge, 1981.

[13] W. Parry. Automorphisms of the Bernoulli endomorphism and a class of skew-products. *Ergod. Th. & Dynam. Sys.* **16** (1996), 519–529.

[14] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

[15] M. Rosenblatt. Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* **8** (1959), 665–681.

[16] A. N. Shiriayev. *Probability*, 2nd edn. Springer, Berlin, 1989.

[17] C. E. Shannon. The mathematical theory of communication. *Bell System Tech. J.* **27** (1948), 379–423; **27** (1948), 623–656

[18] B. Simon. The sharp form of the strong Szegö theorem. *Geometry, Spectral Theory, Groups, and Dynamics (Contemporary Mathematics, 3877)*. American Mathematical Society, Providence, RI, 2005, pp. 253–275.

[19] M. Smorodinsky. *Ergodic Theory, Entropy (Lecture Notes in Mathematics, 214)*. Springer, Berlin, 1971.

[20] N. Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Wiley, New York, 1945.

[21] N. Wiener. *Non-Linear Problems in Random Theory*. MIT Press, Cambridge, MA; Wiley, New York, 1958.

https://doi.org/10.1017/S0143385707001034 Published online by Cambridge University Press