The Annals of Probability
1994, Vol. 22, No. 3, 1662–1677

CHAPMAN–KOLMOGOROV EQUATION FOR NON-MARKOVIAN SHIFT-INVARIANT MEASURES

BY M. COURBAGE AND D. HAMDAN

Université Pierre et Marie Curie

We study the class C_π of probability measures invariant with respect to the shift transformation on $K^{\mathbb{Z}}$ (where K is a finite set of integers) which satisfies the Chapman–Kolmogorov equation for a given stochastic matrix Π.

We construct a dense subset of measures in C_π distinct from the Markov measure. When Π is irreducible and aperiodic, these measures are ergodic but not weakly mixing. We show that the set of measures with infinite memory is G_δ dense in C_π and that the Markov measure is the unique measure which maximizes the Kolmogorov–Sinai (K–S) entropy in C_π. We give examples of ergodic measures in C_π with zero entropy.

1. Introduction. This work has been motivated by the description of the irreversible approach to statistical equilibrium and the construction of Boltzmann entropy in a sense explained in [3] and [4]. This problem has been studied for unstable dynamical systems such as the K-systems in [6] and [7]. Here we consider systems which are not necessarily K-systems.

This work is also related to a problem studied by Lévy and Feller [5] concerning the construction of non-Markovian stochastic processes which satisfy the Chapman–Kolmogorov equation.

We formulate the problem as follows: Let $(\Omega, \mathcal{A}, S, \mu)$ be a dynamical system; that is, $(\Omega, \mathcal{A})$ is a measurable space and S is a one-to-one measurable transformation on $(\Omega, \mathcal{A})$ preserving the measure μ. Consider a finite partition $\mathcal{P} = \{P_0, \dots, P_{k-1}\}$ of Ω, where $P_i \in \mathcal{A}$ and $\mu(P_i) > 0$ for all i. We say that $\mathcal{P}$ satisfies the Chapman–Kolmogorov equation if the family of matrices Π_n, $n \geq 0$, defined by

$$(\Pi_n)_{ij} = \mu(S^{-n}P_j \mid P_i)$$

is a semigroup: $\Pi_{n+n'} = \Pi_n \Pi_{n'}$, $\forall n, n' \geq 0$.

The systems which are isomorphic to Markov shifts possess such a partition. In this case the last equation is nothing but the Chapman–Kolmogorov equation satisfied by the Markov chain.

The existence of such partitions entails a necessary spectral property of the dynamical system; namely, it must have a spectrum with Lebesgue component when the matrix Π_1 (denoted from now on as Π) is aperiodic and irreducible. To see this, recall that, in this case, the matrix Π has an eigenvalue λ such that

Received July 1991.

AMS 1991 *subject classifications*. 60G10, 28D05.

Keywords and phrases: Chapman–Kolmogorov, stationary Markov chain, ergodic non–weakly mixing dynamical systems, entropy, skew-product.

$|\lambda| < 1$. Let $(\alpha_i) \in \mathbb{C}^k$ be an eigenvector of Π for this eigenvalue, and

$$f = \sum_{i=0}^{k-1} \alpha_i 1_{P_i}.$$

Let μ_f be the measure on $[0, 2\pi[$ whose Fourier coefficients are

$$\widehat{\mu}_f(n) = \langle f, U^n f \rangle = \int_0^{2\pi} e^{inx}\, d\mu_f(x),$$

where $(Uf)(x) = f(S^{-1}x)$. Let us denote $\lambda = re^{i\theta}$. Then

$$\begin{aligned}\langle f, U^n f \rangle &= \sum_{i,j} \overline{\alpha}_i \alpha_j (\Pi^n)_{ij} \mu(P_i)\\ &= r^{|n|} e^{in\theta} \left(\sum_i |\alpha_i|^2 \mu(P_i) \right).\end{aligned}$$

If, for simplicity, we take $\sum_{i=0}^{k-1} |\alpha_i|^2 \mu(P_i) = 1$, then μ_f has the following density with respect to the Lebesgue measure:

$$\begin{aligned}g(x) &= \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} r^{|n|} e^{in(\theta - x)}\\ &= \frac{1}{2\pi} \frac{1 - r^2}{\big(r - \cos(\theta - x)\big)^2 + \sin^2(\theta - x)}.\end{aligned}$$

This shows that μ_f is equivalent to the Lebesgue measure.

Here we investigate the existence of shift dynamical systems for which the Chapman–Kolmogorov equation is satisfied for invariant measures that are not Markov measures (see [11]). We consider the set C_π of all invariant probability measures on the shift space $K^{\mathbb{Z}}$ (where $K = \{0, 1, \dots, k-1\}$) which satisfy the Chapman–Kolmogorov equation with respect to the partition $P_i = \{\omega \in K^{\mathbb{Z}} : \omega_0 = i\}$ for a given stochastic matrix Π. In general, Π has to satisfy some properties of aperiodicity or irreducibility.

In Section 2, we give a procedure to construct a dense class of measures in C_π, denoted by S_π. For any strictly positive Π, S_π contains measures which are distinct from the Markov measure. When Π is irreducible and aperiodic, all the non-Markovian measures of S_π are ergodic and not weakly mixing.

In Section 3 we show that the set of all measures with infinite memory is G_δ dense in C_π.

In Section 4, we show that the entropy of the previously constructed class is positive and that the Markov measure is the unique one which maximizes the entropy in C_π.

Finally, we give examples of invariant ergodic measures satisfying the Chapman–Kolmogorov equation with zero entropy. All these examples come from skew product dynamical systems, and the partition at different times is independent, that is, $\mu(S^{-n}P_i \cap P_j) = \mu(P_i)\mu(P_j)$ for $n \geq 1$.

It is an interesting question whether there exist partitions satisfying the Chapman–Kolmogorov equation in the nonindependent case for dynamical systems with zero entropy.

In a paper published in 1982, Alexeyev [1] raised the question of a possible realization of any given spectral type of a dynamical system by a bounded function. The dynamical systems which we construct are examples of systems (some of them are of zero entropy) whose Lebesgue spectral type is realized by a function taking a finite number of values.

It would be interesting to answer the following question: for a dynamical system having a Lebesgue component in its spectrum, is there a partition satisfying the Chapman–Kolmogorov equation?

2. A class of invariant measures verifying the Chapman–Kolmogorov equation. Let $K = \{0, 1, \ldots, k-1\}$, $k \geq 2$. Let $\Omega = K^{\mathbb{Z}}$ be the set of all double sequences $\omega = (\omega_i), \omega_i \in K, i \in \mathbb{Z}$. Let σ be the shift transformation $(\sigma\omega)_i = \omega_{i+1}$ and $\mathcal{A}$ the σ-algebra generated by the cylinder sets $\{\omega: \omega_{i_1} = j_1, \ldots, \omega_{i_n} = j_n\}$, also denoted by $\{\omega_{i_1} = j_1, \ldots, \omega_{i_n} = j_n\}$ if no more precision is necessary. Let Π be a $k \times k$ stochastic matrix and $p = (p_i)$ be a row probability vector invariant under Π. Denote by μ_π the Markov measure on $(\Omega, \mathcal{A})$:

$$\mu_\pi\big(\{\omega_n = x_0, \ldots, \omega_{n+p} = x_p\}\big) = p_{x_0}\Pi_{x_0,x_1}\cdots\Pi_{x_{p-1},x_p}, \tag{1}$$

which we simply denote by $\mu_\pi(x_0, \ldots, x_p)$. This measure is also called the (p, Π) Markov chain.

Denote by C_π the set of all σ-invariant probability measures ν on $(\Omega, \mathcal{A})$ that satisfy the Chapman–Kolmogorov equation, which we write as

$$\nu\big(\{\omega_n = j \mid \omega_0 = i\}\big) = (\Pi^n)_{ij}. \tag{2}$$

The set C_π is a compact convex subspace of the set $\mathcal{M}(\Omega, \sigma)$ of all σ-invariant probability measures on $(\Omega, \mathcal{A})$ for the weak* topology.

Now, we shall give a procedure to construct measures in C_π which may be distinct from μ_π.

Let $C_{\pi,n}$ be the set of all probability measures on K^{n+1} that are invariant under the shift and satisfying (2), that is,

(i) $$\sum_y \mu\Big(\{\omega \in K^{n+1}: \omega_0 = y, \omega_1 = x_0, \ldots, \omega_{p+1} = x_p\}\Big) = \mu\Big(\{\omega \in K^{n+1}: \omega_0 = x_0, \ldots, \omega_p = x_p\}\Big), \qquad 0 \leq p \leq n-1,$$

(ii) $$\mu\Big(\{\omega \in K^{n+1}: \omega_0 = x_0, \omega_p = x_p\}\Big) = \mu_\pi\big(\{\omega_0 = x_0, \omega_p = x_p\}\big), \qquad 1 \leq p \leq n.$$

We have shown in [3] that there are many such measures.

For simplicity, we denote

$$\mu\Big(\{\omega \in K^{n+1}: \omega_0 = x_0, \ldots, \omega_k = x_k\}\Big) = \mu(x_0, \ldots, x_k), \qquad k \leq n.$$

We also use this notation for any invariant measure on $K^{\mathbb{N}}$ or $K^{\mathbb{Z}}$.

If (p_i) is a strictly increasing sequence of integers with $p_0 = 0$, and (μ_i) a sequence of measures such that $\mu_i \in C_{\pi, p_{i+1} - p_i}$, we define a measure ν_0 on $K^{\mathbb{N}}$ by

$$\begin{aligned}(3)\quad &\nu_0(\{\omega_0 = x_0, \dots, \omega_{p_1} = x_{p_1}, \dots, \omega_{p_r} = x_{p_r}\})\\ &\quad = \mu_0(x_0, \dots, x_{p_1})\mu_1(x_{p_1+1}, \dots, x_{p_2} \mid x_{p_1}) \cdots \mu_{r-1}(x_{p_{r-1}+1}, \dots, x_{p_r} \mid x_{p_{r-1}})\end{aligned}$$

for any positive integer r. Consider now the sequence of measures

$$\nu_m = \frac{1}{m} \sum_{i=0}^{m-1} \sigma^i \nu_0.$$

Then any limit point ν of this sequence for the weak* topology is a σ-invariant measure which satisfies the Chapman–Kolmogorov equation. To show this, it is sufficient to satisfy

$$\nu_0(\{\omega_s = u,\ \omega_t = v\}) = \mu_\pi(\{\omega_s = u,\ \omega_t = v\}), \qquad t \geq s.$$

Let $p_i \leq s < p_{i+1}$, $p_j \leq t < p_{j+1}$. Then

$$\begin{aligned}&\nu_0(\{\omega_s = u,\ \omega_t = v\})\\ &\quad = \sum_x \mu_i(\omega_{s-p_i} = u,\ \omega_{p_{i+1}} = x_{p_{i+1}})\\ &\qquad \times \mu_{i+1}(\omega_{p_{i+2}-p_{i+1}} = x_{p_{i+2}} \mid \omega_0 = x_{p_{i+1}}) \cdots \mu_j(\omega_{t-p_j} = v \mid \omega_0 = x_{p_j})\\ &\quad = \sum_x \mu_\pi(u)(\Pi^{p_{i+1}-s})_{u, x_{p_{i+1}}}(\Pi^{p_{i+2}-p_{i+1}})_{x_{p_{i+1}}, x_{p_{i+2}}} \cdots (\Pi^{t-p_j})_{x_{p_j}, v}\\ &\quad = \mu_\pi(u)(\Pi^{t-s})_{u,v} = \mu_\pi(\{\omega_s = u, \omega_t = v\}).\end{aligned}$$

From now on, we shall be concerned with the case: $p_i = ni$ and $\mu_i = \mu$ for all $i \in \mathbb{N}$, where $n \geq 1$ is a fixed integer and μ is a fixed measure in $C_{\pi,n}$. In this case, ν_0 is given by

$$\begin{aligned}(3')\quad &\nu_0(\{\omega_0 = x_0, \omega_1 = x_1, \dots, \omega_{rn} = x_{rn}\})\\ &\quad = \mu(x_0, \dots, x_n)\mu(x_{n+1}, \dots, x_{2n} \mid x_n) \cdots \mu(x_{(r-1)n+1}, \dots, x_{rn} \mid x_{(r-1)n}).\end{aligned}$$

It can be easily seen that ν_0 is σ^n-invariant. Then ν is given by

$$(4)\qquad \nu = \frac{1}{n} \sum_{i=0}^{n-1} \sigma^i \nu_0.$$

LEMMA 1. *If Π is irreducible and aperiodic, the following conditions are equivalent*:

(i) *ν is strongly mixing.*

(ii) ν *is weakly mixing.*
(iii) ν_0 *is* σ*-invariant.*
(iv) μ *is the projection of* μ_π *on* K^{n+1} *denoted by* $\mu_\pi\,|_{K^{n+1}}$.
(v) $\nu = \mu_\pi$.

PROOF. (ii) $\Rightarrow$ (iii) If ν is weakly mixing for σ, it is also weakly mixing for σ^n, hence ergodic for σ^n. Then (4) implies that $\sigma^i\nu_0 = \nu_0$ for any i.
(iii) $\Rightarrow$ (iv) We have

$$\sigma^i\nu_0(x_0,\dots,x_{n-i+1}) = \mu(\omega_i = x_0,\dots,\omega_n = x_{n-i})\mu\bigl(x_{n-i+1}\mid x_{n-i}\bigr),$$
$$\nu_0(x_0,\dots,x_{n-i+1}) = \mu(x_0,\dots,x_{n-i+1}) \quad \text{for } i = 1,\dots,n.$$

Since $\sigma^i\nu_0 = \nu_0$ for $i = 1,\dots,n$, it follows that, for any p,

$$\mu\bigl(x_p \mid x_0,\dots,x_{p-1}\bigr) = \mu\bigl(x_p \mid x_{p-1}\bigr),$$

that is, $\mu = \mu_\pi$. The other implications (i) $\Rightarrow$ (ii), (iv) $\Rightarrow$ (v) and (v) $\Rightarrow$ (i) are evident. □

Thus our construction yields a measure ν distinct from μ_π if μ is distinct from $\mu_\pi|_{K^{n+1}}$. It has been shown in [3] that there are many such μ when Π is strictly positive.
In what follows we make explicit, when necessary, the dependence of ν_0 with respect to μ. We also denote by ϕ_n the mapping which associates to each $\mu \in C_{\pi,n}$ the probability measure $\nu \in C_\pi$.

THEOREM 1. *If* Π^n *is an irreducible matrix, then, for any* $\mu \in C_{\pi,n}$, $\nu = \phi_n\mu$ *is ergodic.*

PROOF. It is sufficient to prove that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{s=0}^{N-1}\nu(A\cap\sigma^{-s}B) = \nu(A)\nu(B) \tag{5}$$

for any A, B such that

$$A = \{\omega_0 = x_0,\dots,\omega_{ln} = x_{ln}\},$$
$$B = \{\omega_0 = y_0,\dots,\omega_{pn} = y_{pn}\}.$$

Let N be given by $N = Mn + j$, $j \in \{0,\dots,n-1\}$. Then

$$\begin{aligned}\frac{1}{N}\sum_{0}^{N-1}\nu(A\cap\sigma^{-s}B) &= \frac{1}{Mn+j}\sum_{t=0}^{M-1}\sum_{r=0}^{n-1}\nu(A\cap\sigma^{-tn-r}B)\\ &\quad+\frac{1}{Mn+j}\sum_{s=Mn}^{Mn+j-1}\nu(A\cap\sigma^{-s}B).\end{aligned} \tag{6}$$

The second term of the right-hand side of (6) tends to 0 as $M \to \infty$. The first one can be written by using the definition (4) of ν

$$\begin{aligned}(7)\qquad &\frac{1}{Mn+j}\sum_{t=0}^{M-1}\sum_{r=0}^{n-1}\nu(A\cap\sigma^{-tn-r}B)\\ &=\frac{M}{Mn+j}\sum_{r=0}^{n-1}\frac{1}{n}\sum_{q=0}^{n-1}\frac{1}{M}\sum_{t=0}^{M-1}\sigma^q\nu_0(A\cap\sigma^{-tn-r}B).\end{aligned}$$

By the definition (3′), $\sigma\nu_0$ is a σ^n-invariant nth order Markov chain. Then it is isomorphic to the Markov shift on $(K^n)^{\mathbb{Z}}$ under the isomorphism

$$x=(x_i)\in K^{\mathbb{Z}}\to\varphi(x)\in(K^n)^{\mathbb{Z}},$$
$$\big(\varphi(x)\big)_k=\big(x_{nk},x_{nk+1},\dots,x_{(n+1)k-1}\big).$$

It is easy to see that this Markov chain has the following transition matrix:

$$(8)\qquad W_{x_i\dots x_n,\,y_1\dots y_n}=\mu\big(y_1,\dots,y_n\mid x_n\big).$$

The Chapman–Kolmogorov property and the irreducibility of Π^n imply that this chain is ergodic. Therefore, the measures $\sigma^q\nu_0$ are σ^n-ergodic for all q. This leads to

$$(9)\qquad \lim_{M\to\infty}\frac{1}{M}\sum_{t=0}^{M-1}\sigma^q\nu_0(A\cap\sigma^{-tn-r}B)=\sigma^q\nu_0(A)\sigma^r\nu_0(\sigma^{-q}B)$$

Hence we obtain by substituting (9) in (7) and (6):

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N}\sum_{s=0}^{N-1}\nu(A\cap\sigma^{-s}B)\\ &=\frac{1}{n}\sum_{r=0}^{n-1}\frac{1}{n}\sum_{q=0}^{n-1}\sigma^q\nu_0(A)\cdot\sigma^r\nu_0(\sigma^{-q}B)\\ &=\frac{1}{n}\sum_{q=0}^{n-1}\sigma^q\nu_0(A)\nu(\sigma^{-q}B)\\ &=\frac{1}{n}\sum_{q=0}^{n-1}\sigma^q\nu_0(A)\cdot\nu(B)\quad(\sigma\text{-invariance of }\nu)\\ &=\nu(A)\nu(B).\end{aligned}$$

□

REMARK. The construction of the measures $\phi_n\mu$ can be extended to a larger class of μ than $C_{\pi,n}$. In fact, let $D_{\pi,n}$ be the set of probability measures on K^{n+1} such that

$$\mu\Big(\big\{\omega\in K^{n+1}:\omega_i=x_0,\ \omega_{i+p}=x_p\big\}\Big)=\mu_\pi\big(\{\omega_0=x_0,\ \omega_p=x_p\}\big)$$

and such that the measure ν_0 associated with it according to (3′) is σ^n-invariant. It is easy to see from the preceding proof that the theorem holds for the measures $\phi_n\mu$ associated with this larger class of μ.

Additional examples of the ergodic measures of C_π have been proposed by an associate editor of *The Annals of Probability*. They are constructed according to (3) where one takes $p_1 = n - 1$, $p_2 = n$, $\mu_0 = \mu$, $\mu_1 = \mu_\pi$ and such that ν_0 is σ^n-invariant, that is,

$$\begin{aligned}
\nu_0(\,\omega_0 &= x_0, \dots, \omega_{pn} = x_{pn}) \\
&= \mu(x_0, \dots, x_{n-1})\mu(x_n \mid x_{n-1}) \\
&\quad \times \mu(x_{n+1}, \dots, x_{2n-1} \mid x_n)\mu(x_{2n} \mid x_{2n-1}) \times \cdots \\
&\quad \times \mu(x_{(p-1)n+1}, \dots, x_{pn-1} \mid x_{(p-1)n})\mu(x_{pn} \mid x_{pn-1}) \\
&= \mu(x_0)\mu(x_1 \mid x_0) \dots \mu(x_{n-1} \mid x_0, \dots, x_{n-2})\mu(x_n \mid x_{n-1}) \\
&\quad \times \mu(x_{n+1} \mid x_n) \times \cdots \times \mu(x_{2n-1} \mid x_n, \dots, x_{2n-2})\mu(x_{2n} \mid x_{2n-1}) \times \cdots \\
&\quad \times \mu(x_{(p-1)n+1} \mid x_{(p-1)n}) \times \cdots \\
&\quad \times \mu(x_{pn-1} \mid x_{(p-1)n}, \dots, x_{pn-2})\mu(x_{pn} \mid x_{pn-1}).
\end{aligned}$$

Since ν_0 is σ^n-invariant, let ν be defined by

$$\nu = \frac{1}{n}\sum_{i=0}^{n-1} \sigma^i \nu_0.$$

It can be seen from the preceding discussion that this is a function of a Markov chain. The state space of this Markov chain is $\bigcup_{m=1}^{n} K^m$. The allowable transitions are transitions from $(x_1, \dots, x_m)$ to $(x_1, \dots, x_{m+1})$, $m \le n - 1$, with probability

$$P_{(x_1, \dots, x_m), (x_1, \dots, x_m, x_{m+1})} = \mu(x_{m+1} \mid x_1, \dots, x_m)$$

and from $(x_1, \dots, x_n)$ to $y \in K$ with probability $P_{(x_1, \dots, x_n), y)} = \mu(y \mid x_n)$.

The stationary probability row vector is $\{(1/n)\mu(x_0, \dots, x_m)\}$, $0 \le m \le n - 1$. The function is the projection onto the last coordinate. It follows easily from the irreducibility of Π^n and the Chapman–Kolmogorov property of μ that the preceding chain is irreducible. Now a function of an ergodic process is ergodic, so the ergodicity of ν follows.

THEOREM 2. *Let $R_n\nu$ denote the restriction of the measure $\nu \in C_\pi$ to K^{n+1}. Then $\phi_n R_n \nu$ converges to ν in the weak* topology.*

PROOF. It follows from the definition of ν_0 [formula (3′)] that we have, for any fixed set $(\{\omega_0 = x_0, \dots, \omega_s = x_s\})$ and for $i + s \le n$,

$$\sigma^i \nu_0(R_n\nu)(\{\omega_0 = x_0, \dots, \omega_s = x_s\}) = \nu(\{\omega_0 = x_0, \dots, \omega_s = x_s\}).$$

Thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1}\sigma^i\nu_0(R_n\nu)\big(\{\omega_0 = x_0,\dots,\omega_s = x_s\}\big)$$
$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\nu\big(\{\omega_0 = x_0,\dots,\omega_s = x_s\}\big)$$
$$= \nu\big(\{\omega_0 = x_0,\dots,\omega_s = x_s\}\big). \qquad \square$$

COROLLARY 1. *The set* $S_\pi = \bigcup_n \phi_n C_{\pi,n}$ *is dense in* C_π.

COROLLARY 2. *If* Π *is irreducible and aperiodic, the set of ergodic measures is dense in* C_π.

It can be useful to compute the fixed points of $\phi_n R_n$. A trivial fixed point is μ_π. The following theorem gives a complete characterization.

THEOREM 3. *The only fixed point of* $\phi_n R_n$ *is* μ_π.

PROOF. Let $\nu = \phi_n R_n \nu$. It is sufficient to show that, for any $s \le n$,

$$\nu(x_0,\dots,x_s) = \mu_\pi(x_0,\dots,x_s). \tag{10}$$

We show this by induction. In fact, (10) is true for $s = 1$. Suppose it is true for $s < n$. Then, we compute $\phi_n R_n \nu(x_0,\dots,x_{s+1})$ as in the proof of the Theorem 2: for any $i < n - s$ we have

$$\sigma^i\nu_0(R_n\nu)\big(\{\omega_0 = x_0,\dots,\omega_{s+1} = x_{s+1}\}\big) = \nu(x_0,\dots,x_{s+1}).$$

For $n - s \le i \le n - 1$, we have

$$\sigma^i\nu_0(R_n\nu)\big(\{\omega_0 = x_0,\dots,\omega_{s+1} = x_{s+1}\}\big)$$
$$= \nu(x_0,\dots,x_{n-i})\nu\big(x_{n-i+1},\dots,x_{s+1} \mid x_{n-i}\big)$$
$$= \mu_\pi(x_0,\dots,x_{s+1}).$$

Here, we used the hypothesis according to which ν coincides with μ_π on the coordinates $(\omega_0,\dots,\omega_s)$. This entails that

$$\nu(x_0,\dots,x_{s+1}) = \phi_n R_n \nu(x_0,\dots,x_{s+1})$$
$$= \frac{n-s}{n}\nu(x_0,\dots,x_{s+1}) + \frac{s}{n}\mu_\pi(x_0,\dots,x_{s+1})$$

and completes the proof. $\square$

3. Chains with Infinite Memory in C_π. An invariant measure ν on $K^{\mathbb{Z}}$ is called a p-order Markov chain if, for any $n > p$,

$$\begin{aligned}\nu\big(\omega_n = x_n \mid \omega_0 = x_0,\dots,\omega_{n-1} = x_{n-1}\big)\\ = \nu\big(\omega_n = x_n \mid \omega_{n-p} = x_{n-p},\dots,\omega_{n-1} = x_{n-1}\big).\end{aligned} \tag{11}$$

In what follows, all Markov chains are stationary.

An invariant measure ν is called a chain of infinite memory if it is not a Markov chain of any finite order.

We shall investigate the existence and the abundance of such infinite memory measures in C_π. It will be shown that they are generic.

LEMMA 2. *Let μ_i, $i = 1,\dots,r$, be a family of distinct irreducible Markov chains on the same state space K. If, for any $i \neq j$ and any $y \in K$ there $x \in K$ such that $\mu_i(x,y)\mu_j(x,y) > 0$, then the measure*

$$\nu = \sum_{i=1}^{r} \alpha_i \mu_i, \qquad 0 < \alpha_i < 1, \ \sum \alpha_i = 1,$$

has infinite memory.

PROOF. Suppose that ν is a p-order Markov chain. Then, by (11), we have

$$\nu(\omega_0,\dots,\omega_n)\nu(\omega_{n-p},\dots,\omega_{n-1}) = \nu(\omega_0,\dots,\omega_{n-1})\nu(\omega_{n-p},\dots,\omega_n) \tag{12}$$

for any $n \geq p$. It follows, using the definition of ν,

$$\begin{aligned}\sum_{i,j} \alpha_i\alpha_j \big[\mu_i(\omega_0,\dots,\omega_n)\mu_j(\omega_{n-p},\dots,\omega_{n-1}) \\ - \mu_i(\omega_0,\dots,\omega_{n-1})\mu_j(\omega_{n-p},\dots,\omega_n)\big] = 0.\end{aligned} \tag{13}$$

Let $A = \{\omega_0 = x_0,\dots,\omega_l = x_l\}$, $B = \{\omega_0 = y_0,\dots,\omega_p = x_p\}$ and $C = \{\omega_0 = y_0,\dots,\omega_{p-1} = y_{p-1}\}$. Then, for $l < n-p$, (13) implies

$$\sum_{i,j} \alpha_i\alpha_j \Big[\mu_i\big(A \cap \sigma^{(n-p)}B\big)\mu_j(C) - \mu_i\big(A \cap \sigma^{-(n-p)}C\big)\mu_j(B)\Big] = 0. \tag{14}$$

Now, we take the Césaro mean over $n > p + l$ and we use the ergodicity of the measures μ_i to get

$$\sum_{i,j} \alpha_i\alpha_j\mu_i(A)\big[\mu_i(B)\mu_j(C) - \mu_i(C)\mu_j(B)\big] = 0. \tag{15}$$

Equation (15) also holds for any $A \in \mathcal{A}$. Let A^i be the support of μ_i. The measures μ_i, being ergodic, have disjoint supports. Therefore, $\mu_j(A^i) = 0$ for $i \neq j$. Hence (15) implies

$$\alpha_i\mu_i(A^i)\big[\mu_i(B)\nu(C) - \mu_i(C)\nu(B)\big] = 0.$$

Consequently,

$$\mu_i(B)\nu(C) - \mu_i(C)\nu(B) = 0. \tag{16}$$

Now, for the Markov chains μ_i and μ_j, the condition, for any y there exists $x \in K$ such that $\mu_i(x,y)\mu_j(x,y) > 0$, is equivalent to the condition, for any y and

$p \in \mathbb{N}$ there is $(y_0, \ldots, y_{p-2})$ such that $\mu_i(y_0, \ldots, y_{p-2}, y)\mu_j(y_0, \ldots, y_{p-2}, y) > 0$. Then, by using (16), we obtain, for any y_{p-1} and y_p,

$$\begin{aligned}\mu_i(y_p \mid y_{p-1}) = \mu_i(y_p \mid y_0, \ldots, y_{p-1}) &= \nu(y_p \mid y_0, \ldots, y_{p-1}) \\ &= \mu_j(y_p \mid y_0, \ldots, y_{p-1}) \\ &= \mu_j(y_p \mid y_{p-1}),\end{aligned}$$

that is, $\mu_i = \mu_j$, a contradiction. □

The fact that a convex combination of invariant probability measures may be a process with infinite memory can be obtained under hypotheses different from those of Lemma 2. The following lemma, which results from a question of the referee, gives such hypotheses.

LEMMA 3. *Let μ_i, $i = 1, \ldots, r$, be a family of distinct stationary measures on $K^{\mathbb{Z}}$. If μ_1 is an irreducible and aperiodic Markov chain, then every nontrivial convex combination ν of μ_i, $i = 1, \ldots, r$, has infinite memory.*

PROOF. Suppose that ν is a pth order Markov chain. By passing to a state space whose elements are blocks of p-coordinates, $(K^{\mathbb{Z}}, \nu, \sigma^p)$ and $(K^{\mathbb{Z}}, \mu_1, \sigma^p)$ are isomorphic to first-order Markov chains whose transition probabilities from $x = (x_1, \ldots, x_p)$ to $y = (y_1, \ldots, y_p)$ are, respectively, given by

$$W_{x,y} = \nu(y_1 \mid x_1, \ldots, x_p)\nu(y_2 \mid x_2, \ldots, x_p, y_1) \times \cdots \times \nu(y_p \mid x_p, y_1, \ldots, y_{p-1})$$

and

$$W^{(1)}_{x,y} = \overset{(1)}{\pi}_{x_p y_1} \times \overset{(1)}{\pi}_{y_1 y_2} \times \cdots \times \overset{(1)}{\pi}_{y_{p-1}, y_p},$$

where $\overset{(1)}{\pi}$ is the stochastic matrix associated with μ_1.

It is easy to see that $W^{(1)}$ is irreducible if and only if $(\overset{(1)}{\pi})^p$ is irreducible, and $(\overset{(1)}{\pi})^p$ is irreducible for all p if and only if $\overset{(1)}{\pi}$ is irreducible and aperiodic. Thus, under the hypotheses of the lemma, $W^{(1)}$ is irreducible for any p, and therefore the same is true for W, which implies that $((K^p)^{\mathbb{Z}}, \nu, \sigma^p)$ is ergodic. This contradicts the fact that ν is a barycenter of μ_i's. □

THEOREM 4. *If Π^2 is irreducible, then, for any $\mu \in C_{\pi,2}$ such that $\mu \neq \mu_\pi|_{K^3}$, $\phi_2\mu$ is a chain of infinite memory.*

PROOF. By the definition, ν_0 is a σ^2-invariant second-order Markov chain. Then $(K^{\mathbb{Z}}, \nu_0, \sigma^2)$ is isomorphic to the Markov shift $((K^2)^{\mathbb{Z}}, \varphi\nu_0, \sigma)$ where the isomorphism φ is given by

$$x = (x_i) \in K^{\mathbb{Z}} \to \varphi(x) \in (\mathbb{K}^2)^{\mathbb{Z}},$$
$$(\varphi(x))_k = (x_{2k}, x_{2k+1}).$$

The Markov chain on the space $\mathbb{K}^2$ has the transition matrix W^1 given by

$$W^1_{(x_0,x_1),(x_2,x_3)} = \frac{\mu(x_0,x_1,x_2)\mu(x_3 \mid x_2)}{\mu(x_0,x_1)}$$

and the invariant row probability vector for this chain is $\mu(x_0,x_1)$.

The same is true for $(K^{\mathbb{Z}},\sigma\nu_0,\sigma^2)$ which is isomorphic to a Markov shift on $(K^2)^{\mathbb{Z}}$. Here, the transition matrix W^2 is given by

$$W^2_{(x_0,x_1),(x_2,x_3)} = \mu(x_2,x_3 \mid x_1)$$

and the invariant row probability vector is $\mu(x_0,x_1)$.

If Π^2 is irreducible, then W^1 and W^2 are irreducible and $(K^{\mathbb{Z}},\nu_0,\sigma^2)$ and $(K^{\mathbb{Z}},\sigma\nu_0,\sigma^2)$ are ergodic. Now, we have to show that W^1 and W^2 satisfy the hypothesis of Lemma 2, namely, for any $(i,j) \in K^2$ there exist (x_0,x_1) such that

$$W_{(x_0x_1),(i,j)} > 0,$$

where $W_{(kl),(mn)} = W^1_{(kl),(mn)}W^2_{(kl,mn)}$. From the expression of W^1 and W^2 we see that

$$W_{(kl),(mn)} > 0 \quad \text{iff } \mu(k,l,m)\mu(l,m,n) > 0.$$

But, for any (m,n) such that $\mu(m,n) > 0$, there exist l such that $\mu(l,m,n) > 0$. As $\mu(l,m) > 0$, there also exist k such that $\mu(k,l,m) > 0$. So we may apply Lemma 2 to deduce that $\varphi(\nu_0 + \sigma\nu_0)/2$ is a chain of infinite memory and so is $\frac{1}{2}(\nu_0 + \sigma\nu_0)$. This completes the proof. □

THEOREM 5. *If Π^2 is irreducible, then the set of all measures in C_π with infinite memory is a G_δ dense set in C_π.*

PROOF. Let us denote by $\mathcal{C}_p$ the set of all measures of C_π that are pth order Markov chains [satisfying (11)]. We shall prove that $C_\pi \backslash \mathcal{C}_p$ is dense in C_π. For this, it is enough to show that for any $\lambda \in \mathcal{C}_p$ there exists a sequence of measures in $C_\pi \backslash \mathcal{C}_p$ which converges to λ. Let $\mu \in C_{\pi,2}$, $\mu \neq \mu_\pi|_{K^3}$ and $\nu = \phi_2\mu$. By the preceding theorem, $\nu \in C_\pi \backslash \mathcal{C}_p$ for any p. For $t \in [0,1]$ put $\lambda_t = t\nu + (1-t)\lambda$. As C_π is convex, $\lambda_t \in C_\pi$. There exists a sequence $t_n \searrow 0$, such that $\lambda_{t_n} \in C_\pi \backslash \mathcal{C}_p$. For, on the contrary, we can find $\varepsilon > 0$ such that $\lambda_t \in \mathcal{C}_p$ for any $t \in]0,\varepsilon]$. Then the Markov property (11), applied to λ_t, implies a polynomial relationship of the form

$$At^2 + Bt + C = 0 \quad \forall\, t \in]0,\varepsilon].$$

Here, the coefficients A, B and C depend on λ and ν. Their nullity implies that the Markov property (11) also holds for ν, a contradiction. This shows that $C_\pi \backslash C_p$ is dense in C_π. Since $C_\pi \backslash C_p$ is open, for any p by Baire's theorem $\bigcap_{p=1}^{\infty}(C_\pi \backslash C_p)$ is dense in C_π. □

REMARK. The proof of Theorem 5 can be repeated for the set of all stationary measures $M(K^{\mathbb{Z}}, \sigma)$ and shows that the set of all invariant measures with infinite memory is generic in $M(K^{\mathbb{Z}}, \sigma)$. Theorem 5 tells us that this property is preserved even when we restrict to measures satisfying the Chapman–Kolmogorov equation.

4. Kolmogorov–Sinai entropy for C_π measures. We first compute the K–S entropy $h(\nu, \sigma)$ where $\nu = \phi_n \mu$. By using the affinity of entropy, we have

$$nh(\nu, \sigma) = h(\nu, \sigma^n) = \frac{1}{n} \sum_{i=0}^{n-1} h\left(\sigma^i \nu_0, \sigma^n\right) = h(\sigma \nu_0, \sigma^n).$$

In the last equality we used the fact that all the systems $(K^{\mathbb{Z}}, \sigma^i \nu_0, \sigma^n)$ are isomorphic and have therefore the same entropy. Now, the system $(K^{\mathbb{Z}}, \sigma^1 \nu_0, \sigma^n)$ is isomorphic to the Markov chain on the state space $\mathbb{K}^n$ with the transition matrix

$$W_{(x_0, \dots, x_{n-1}), (x_n, \dots, x_{2n-1})} = \mu\left(\omega_1 = x_n, \dots, \omega_n = x_{2n-1} \mid \omega_0 = x_{n-1}\right)$$

and the invariant row probability vector $\mu(x_0, \dots, x_{n-1})$. Thus we obtain

$$\begin{aligned} nh(\nu, \sigma) = h(\sigma \nu_0, \sigma^n) = &- \sum_{x_0 \cdots x_n} \mu(x_0, \dots, x_n) \log \mu(x_0, \dots, x_n) \\ &+ \sum_{x_0} \mu(x_0) \log \mu(x_0). \end{aligned}$$

Let us denote by $H_\mu(x_0, \dots, x_n)$ and $H_\mu(x_0 \mid x_1 \dots x_k)$ the entropies

$$\begin{aligned} H_\mu(x_0, \dots, x_n) &= - \sum_{x_0 \cdots x_n} \mu(x_0, \dots, x_n) \log \mu(x_0, \dots, x_n), \\ H_\mu\left(x_0 \mid x_1, \dots, x_k\right) &= - \sum_{x_0 \cdots x_k} \mu(x_0, \dots, x_k) \log \mu\left(x_0 \mid x_1, \dots, x_k\right). \end{aligned}$$

$h(\nu, \sigma)$ can also be written as

$$h(\nu, \sigma) = \frac{1}{n} H_\mu\left(x_1, \dots, x_n \mid x_0\right), \tag{16'}$$

where

$$H_\mu\left(x_1, \dots, x_n \mid x_0\right) = \sum_{x_0 \cdots x_n} \mu(x_0, x_1, \dots, x_n) \log \mu\left(x_1, \dots, x_n \mid x_0\right).$$

It follows from the relationship

$$H_\mu(x_0, \dots, x_n) = H_\mu(x_0) + H_\mu\left(x_0 \mid x_1\right) + \cdots + H_\mu\left(x_0 \mid x_1 \dots x_n\right)$$

and from (16′) that

$$h(\nu, \sigma) = \frac{1}{n}\Big(H_\mu\left(x_0 \mid x_1\right) + \cdots + H_\mu\left(x_0 \mid x_1, \dots, x_n\right)\Big). \tag{17}$$

Now, $H_\mu(x_0 \mid x_1) = h(\mu_\pi)$ and $H_\mu(x_0 \mid x_1 \dots x_k) \le H_\mu(x_0 \mid x_1)$. We shall use these relationships to prove the following result.

THEOREM 6. (i) *For any* $\nu \in C_\pi$, $h(\nu) \le h(\mu_\pi)$.

(ii) *If* $h(\mu_\pi) > 0$ *and* $\mu \in C_{\pi, n}$, *then* $h(\phi_n \mu) > 0$ *for any* n.

(iii) *If* $\nu \in C_\pi$ *and* $R_n(\nu)$ *is the restriction of* ν *to* K^{n+1}, *then* $h(\phi_n R_n(\mu)) \searrow h(\nu)$ *as* $n \to \infty$.

(iv) μ_π *is the unique measure which maximizes the entropy in* C_π.

PROOF. (i) It is well known that $h(\nu) \le H_\nu(x_0 \mid x_1)$; on the other hand, H_ν $(x_0 \mid x_1)$ is equal to $h(\mu_\pi)$.

(ii) This can be seen from (17).

(iii) We apply (16′) to $\mu = R_n \nu$ and we use the monotonicity property of $(1/n) H_\mu(x_1, \dots, x_n \mid x_0)$ ([11]) to see that $h(\phi_n R_n(\nu))$ decreases to $h(\nu)$ as $n \to \infty$.

(iv) Let $\nu \in C_\pi$ such that $h(\nu) = h(\mu_\pi)$. It follows from (17) applied to $\mu = R_n(\nu)$ that

$$H_\nu(x_0 \mid x_1, \dots, x_n) = H(x_0 \mid x_1)$$

for any $n \ge 1$. Now, the invariance of ν implies that we have, for any p,

$$H_\nu(x_0 \mid x_1, \dots, x_p) = H_\nu(x_p \mid x_0, \dots, x_{p-1}).$$

By using these two relationships we obtain

$$H_\nu(x_n \mid x_0, \dots, x_{n-1}) = H_\nu(x_n \mid x_{n-1}).$$

It is well known [10] that this is equivalent to

$$\nu(x_n \mid x_0, \dots, x_{n-1}) = \nu(x_n \mid x_0).$$

That is, ν is a Markov chain, hence $\nu = \mu_\pi$. □

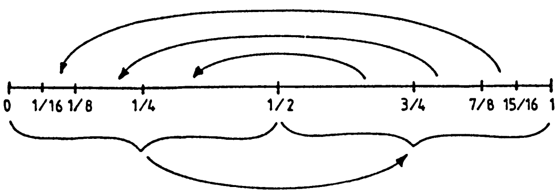
REMARK. It can be seen from the proof of Theorem 6 that it may hold in a more general setting. Let us denote by C_π^k the set of all invariant measures ν such that

$$\nu(\omega_0 = i,\ \omega_r = j) = \mu_\pi(\omega_0 = i,\ \omega_r = j) \quad \text{for any } r \le k.$$

Hence $C_\pi = \bigcap_k C_\pi^k$. The proof of Theorem 6 can be repeated for C_π^1 instead of C_π, and all the statements of the theorem hold. This theorem is similar to the Parry theorem for the topological Markov chains ([11], page 194) and the variational principle for Gibbs states ([2], Chapter 1).

In Section 1 we have observed that the existence of partition satisfying the Chapman–Kolmogorov equation entails that the system has a Lebesgue spectral component. Yet, it does not necessarily entail that the system has a strictly positive Kolmogorov-Sinai entropy. Examples of pairwise independent stationary processes $\{X_n\}$ with zero entropy can be exhibited by using the examples of Janson [8]. These examples correspond to the following dynamical system $T: (x, y) \in S^1 \times S^1) \to (x, \varphi(x) + y)$, where S^1 is the unit circle and φ is a mapping preserving the Lebesgue measure on the circle. The invariant measure for this transformation is the Lebesgue measure on $S^1 \times S^1$. It is well known that this dynamical system has zero entropy. The stationary process X_n is the projection on the second variable of $T^n(x, y)$, that is, $X_n = n\varphi(x) + y$. However, this is not an ergodic process, for the function $X_1 - X_0 = \varphi(x)$ is invariant under the shift but not almost everywhere constant. A similar example of an ergodic system may be considered, with T defined on $S^1 \times S^1$ by $T(x, y) = (\tau x, \varphi(x) + y)$, where $\tau x = x + a$ and $\varphi(x) = px, p \in \mathbb{Z}$. Here, we shall give another example of an ergodic process with pairwise independent variables, having zero entropy. The following transformation, introduced by Mathew and Nadkarni [9] to display a spectrum with a Lebesgue component of multiplicity 2, turns out to be an example of a system with zero K–S entropy having a partition satisfying the Chapman–Kolmogorov equation.

Consider first the Von Neumann transformation τ on $[0, 1]$ defined as follows: The interval $F_1 = [0, 1/2]$ is translated to the interval $[1/2, 1]$, the interval $F_2 = [1/2, 1/2 + (1/2)^2]$ is translated to the interval $[1/2^2, 1/2]$ and so on. This is schematized as follows:



Let F_{k1} be the first half of F_k and F_{k2} its second half. Define the function $\phi(x)$

$$\phi(x) = \begin{cases} 1, & \text{if } x \in F_{k1}, \\ -1, & \text{if } x \in F_{k2}. \end{cases}$$

A new transformation T is built acting on the space $\Omega = [0, 1] \times \{-1, 1\}$ equipped with the measure μ which is the product of the Lebesgue measure on the interval $[0, 1]$ and the measure giving the probability 1/2 to each of -1 and $+1$. T is defined by

$$T(x, \sigma) = \big(\tau x, \phi(x)\sigma\big), \quad x \in [0, 1] \quad \text{and} \quad \sigma \in \{-1, +1\}.$$

Let $\mathcal{P}$ be the partition of Ω into two cells P_0 and P_1 defined by

$$P_0 = \big[0, 1/2\big[\times \{-1\} \cup \big[1/2, 1\big] \times \{1\},$$
$$P_1 = \big[1/2, 1\big] \times \{-1\} \cup \big[0, 1/2\big[\times \{1\}.$$

Let $\chi(x)$ be the function

$$\chi(x) = 1_{F_1}(x) - 1/2, \qquad x \in [0, 1].$$

To any function $f(x)$ defined on $[0,1]$, associate its "odd extension" $\widehat{f}(x,\sigma)$ on Ω defined by

$$\widehat{f}(x,\sigma)=\sigma f(x)=\begin{cases}-f(x), & \sigma=-1,\\ f(x), & \sigma=1.\end{cases}$$

Clearly, χ is orthogonal to 1. Let U be the operator defined on $L_2(\Omega)$ by

$$Uf=f\circ T.$$

We have, for any function $\psi(x)$,

$$U\widehat{\psi}(x,\sigma)=\widehat{\psi}\big(\tau x,\phi(x)\sigma\big)=\phi(x)\sigma\psi(\tau x).$$

If one introduces the operator V on $L^2([0,1])$ defined by

$$V\psi(x)=\phi(x)\psi(\tau x),$$

then the above relationship is written as

$$U\widehat{\psi}=\widehat{V\psi}. \tag{18}$$

Mathew and Nadkarni [9] have proved that

$$\langle V^n\chi,\chi\rangle=0 \quad \forall\, n\neq 0.$$

Then, by using (18), we obtain

$$\begin{aligned}\langle U^n\widehat{\chi},\widehat{\chi}\rangle &= \langle \sigma V^n\chi,\sigma\chi\rangle\\ &= \langle V^n\chi,\chi\rangle=0\end{aligned}$$

and this implies that

$$\mu\big(T^{-n}P_i\cap P_j\big)=1/4 \quad \forall\, n\neq 0.$$

Therefore, the Chapman–Kolmogorov equation is satisfied for the matrix

$$\Pi=\begin{bmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{bmatrix}.$$

Acknowledgments. We are grateful to F. Ledrappier and J. P. Thouvenot for valuable discussions. We are also indebted to J. P. Thouvenot for bringing to our attention the example of Mathew and Nadkarni [9]. We thank the referees for their helpful remarks and suggestions.

REFERENCES

[1] ALEXEYEV, V. M. (1982). Existence of a bounded function of the maximal spectral type. *Ergodic Theory Dynamical Systems* **2** 259–261.

[2] BOWEN, R. (1975). Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Math.* **470**. Springer, New York.

[3] COURBAGE, M. and HAMDAN, D. (1991). A class of non-mixing dynamical systems with monotonic semi-group property. *Lett. Math. Phys.* **22** 101–106.

[4] COURBAGE, M. and NICOLIS, G. (1990). Markov evolution and H-theorem under finite coarse graining in conservative dynamical systems. *Europhys. Lett.* **11** 1–6.

[5] FELLER, W. (1959). Non markovian process with semi-group property. *Ann. Math. Statist.* **30** 1252–1253.

[6] GOLDSTEIN, S., MISRA, B. and COURBAGE, M. (1981). On intrinsic randomness of dynamical systems. *J. Statist. Phys.* **25** 111–126.

[7] GOLDSTEIN, S. and PENROSE, O. (1981). On non-equilibrium entropy for dynamical systems. *J. Statist. Phys.* **24** 325–343.

[8] JANSON, S. (1988). Some pairwise independent sequences for which the central limit theorem fails. *Stochastics* **23** 439–448.

[9] MATHEW, J. and NADKARNI, M. G. (1984). A measure preserving transformation whose spectrum has Lebesgue component with multiplicity two. *Bull. London Math. Soc.* **11** 402–406.

[10] PARRY, W. (1981). *Topics in Ergodic Theory.* Cambridge Univ. Press.

[11] WALTERS, P. (1982). *An Introduction to Ergodic Theory.* Springer, New York.

LABORATOIRE DE PROBABILITÉS
UNIVERSITÉ PIERRE ET MARIE CURIE
4 PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE