ISRAEL JOURNAL OF MATHEMATICS 80 (1992), 257-287

GENERIC SEQUENCES, TRANSDUCERS AND MULTIPLICATION OF NORMAL NUMBERS

BY

F. BLANCHARD

Laboratoire de Mathématiques Discrètes
Case 930, 163 av. de Luminy
13288 Marseille Cedex 9, France

AND

J. M. DUMONT

Département de Math-Info
Faculté des Sciences de Luminy
163 av. de Luminy, 13288 Marseille Cedex 9, France

AND

A. THOMAS

Université de Provence, Case F, 3 place Victor Hugo, 13331 Marseille Cedex 3, France

ABSTRACT

Existence of a topological joining between two subshifts X and X' defines a relation between points of the two. Suppose $x \in X$ is generic for an invariant measure μ on X; when is a related $x' \in X'$ also generic for some corresponding measure μ' on X'? We prove this property holds in several situations for bounded-to-one joinings: when μ and μ' are the measures with maximal entropy on intrinsically ergodic X and X', and also when μ has a unique preimage on the joining, a property for which several sufficient conditions are given. In the latter case it is also possible to prove that the nearer a point is to genericity with respect to μ, the nearer to genericity with respect to μ' a related point is.

Bounded-to-one joinings may be defined by nonambiguous rational transductions. This provides several applications, most of them in Number Theory. It is proven that transducers performing multiplication by integers have the suitable properties: this implies multiplication by a

Received May 24, 1990 and in revised form June 2, 1991

rational preserves near normality; so does addition of a rational. An application to Markov measures, and sufficient conditions for a transducer to map normality to the base p to normality to the base p', $p' \neq p$, are given.

Introduction

The purpose of this article is manifold.

There is first an arithmetical motivation. It is well known that when a normal number is multiplied or divided by an integer, the result is also normal [KN, R2]; the classical proof is based on Weyl's criterion. There operations are easily represented as actions of literal tranducers on an infinite sequence of digits; these transducers are deduced from the usual algorithm of multiplication in section 4 below. G. Rauzy asked us the following question: when does the action of a literal transducer preserve normality of such a sequence, i.e. uniform asymptotic frequency for words with the same length? Of course, in the general case, this has no particular number-theoretic meaning, because the transducer does not perform any simple operation like multiplication by a rational. But this is no idle question. So many years after Champernowne's first example [C], there remains much to understand about equidistribution.

We also asked ourselves another question: suppose the answer to Rauzy's question is yes; is the property stable for small perturbations of the uniform measure? Or else, assuming a sequence to be nearly normal (in the sense of (k, ε)-normality), under what assumptions are its transduced images almost normal? A more striking, number-theoretical aspect of this question being: given a rational number $r \neq 0$, is it true that the closer x is to normality, the closer rx will be? It is easy to give a precise meaning to these notions of closeness to normality, using the topology of weak convergence of measures. There are many instances in which, for practical reasons, one uses "random sequences" which are not strictly so in the sense of normality, but have a satisfying asymptotic frequency of words with length less than a certain bound. When such sequences are submitted to the action of a transducer, do they keep the same property, at least to some degree?

Now, let us drop the 1-torus, the Lebesgue measure and transducers for a while. Rauzy's and all other questions we asked may be given a more general abstract phrasing, having some interest of its own for ergodicians. We first introduce a few definitions. A subshift Y (i.e. a closed shift-invariant subset of some

$A^{\mathbf{Z}}$) is said to be a **joining** of X and X' if X and X' are topological factors, or continuous, onto, shift-commuting images of Y; in this case a point $x \in X$ and a point $x' \in X'$ are said to be related if they have a common preimage in Y; the corresponding definition is valid for invariant measures. Two subshifts are said to be finitely equivalent if they have a bounded-to-one joining. Finite equivalence was introduced by Parry in [P3] in order to classify subshifts of finite type according to their maximal entropy, but it is a natural idea and may serve many other purposes. Transducers with bounded-to-one input and output maps define a finite equivalence between their input and output systems; in the case of a multiplication transducer, it is a nontrivial finite equivalence between two copies of the same full shift. The idea is to investigate topological features of the problems: replace the transducer by some finite equivalence between two subshifts X and X' (not necessarily full shifts): one may ask whether if a point is generic for some invariant measure μ on X (not necessarily the one with maximal entropy, when unique), related points in X' are also generic for some corresponding measure μ'; there is no general reason for μ to be equal to μ'. This is the property we call genericity preservation. One may also ask about preservation of almost genericity. In fact, these are the questions we answer first, using mainly topological tools; solutions of the normality problems, as well as other results, are obtained as corollaries of the abstract statements.

Section 2 contains ergodic results on genericity of points related to, or transduced images of, generic points. They are obtained with elementary techniques of the theory of invariant measures on compact metric spaces. Most of them might be stated within the larger framework of Topological Dynamics. Proposition 2.1 states that when two subshifts X and X' have a joining Y, given an invariant measure μ on X, a μ-generic point $x \in X$, and an element x' of X', related to x through this joining, then the set of weak limits of Cesaro measures of x' (usually called "measures associated to x") is included in the set of invariant measures on X' related to μ. So, if we are able to prove there is only one invariant measure μ' on X' related to μ, an easy compactness argument shows that any x' related to a μ-generic x is μ'-generic. There remains to find sufficient conditions for uniqueness of measure μ'. This is achieved for the measures with maximal entropy of two finitely equivalent intrinsically ergodic subshifts (Proposition 2.2), which are necessarily related to each other only. Applying this result to transducers, one obtains an answer to Rauzy's question, and more (Corollary 2.4). Another condition, slightly more restrictive as far as joinings are concerned, but much less restrictive for measures, is the following: an ergodic measure μ on X is said to have Property T if there exists an ergodic measure ν on Y with image μ such that $\mu \circ \varphi(E) > 0$ implies $\nu(E) > 0$ for any measurable set $E \subset Y$. In Proposition 2.7 this is shown to imply uniqueness of ν, hence of μ'. Synchronizing measures possess Property T (Proposition 2.8).

Section 3 deals with preservation of almost genericity. In addition to the tools of the former section we use (k, ε, μ)-genericity, an obvious generalization of (k, ε)-normality; (k, ε)-normality was introduced by Besicovich [Bes] and further used in [CE, S1, S2, BerV]. There is a close connection between (k, ε, μ)-genericity and the metric of the set $\mathcal{M}(X)$ of measures on X, endowed with the topology of weak convergence (Lemma 3.1). A simple topological argument (Lemma 3.2) then allows to show (Proposition 3.3) that when μ is an invariant measure on X having a unique lift ν on Y, then given an integer k' and $\varepsilon' > 0$, there exist k and ε such that if x is (k, ε, μ)-generic, then any transduced image of x is (k', ε', μ')-generic. In other terms, when an invariant measure μ_1 is close enough to μ, any transduced image μ_1' of μ_1 is close to μ' (Proposition 3.4).

Most results in sections 2 and 3 might be stated in the more general framework of compact metric spaces endowed with a homeomorphism.

Section 4 is devoted to applications. The first paragraph deals with the questions at the root of this article: we prove that multiplication transducers satisfy all requirements for Propositions 2.2 and 3.4 to apply. Suitable properties are checked for input and output (i.e. for multiplication and division) for transducers corresponding to multiplication by integers: they are irreducible by Proposition 4.1, and nonambiguous for input and output in the two elementary cases: when multiplier and base are relatively prime (Lemma 4.2), when the multiplier is a divisor of the base (Lemma 4.4). All these properties belong to the folk-lore of Automata Theory; they have been included in this paper for easier understanding by people not familiar with that field. Then we exploit these results in terms of normality: Proposition 4.6 states that given a rational r, the closer a number is to normality, the closer its product by r also is.

We evoke addition of a rational, for which the methods of sections 2 and 3 work perfectly well.

Still in the same section, we address the question of preservation of genericity by transducers when the input measure μ is Markov with topological support X;

then μ possesses Property T and it is possible to give an explicit formula for the lift ν. We also give and example of a non-synchronizing automaton for which some Markov measures have property T.

Finally, Proposition 4.14 gives a condition for preservation of normality when the transducer is input-deterministic but not necessarily bounded-to-one for output, the input and output systems being two different full shifts; of course in this case "preservation of normality" only means the transducer maps normality in the first base p to normality in the second base q. This is possible only when q divides p (Proposition 4.15). To illustrate these properties, we describe a transducer transforming the 4-shift into the 2-shift, and preserving normality in that sense.

There are two fields we did not investigate at all. The first is action of general (not literal) transducers on measures; in this case the difficulty arises from the fact that the transformations do not commute with the shift. When the transducer defines a coding, some results have been obtained in [BP] about correspondence of measures; in the setting of this article it is quite likely that Property T, as a sufficient condition for uniqueness of the lift, may produce significant results. Broglio and Liardet [BrL] have also studied deterministic transducers which might be called "erasers": the output is either identical to input, or equal to the empty word; their aims do not coincide with ours, but there is some relationship: for instance, they obtain theorems on normality of subsequences closely connected with those in [K, KW], by different means. The same formalism is used in [Me] with a completely different purpose.

The second domain is joinings not defined by transducers: this is a deliberate choice. It is quite sure there are nice cases to investigate in this direction. In [KW] interested readers can find an example of an extension of $[0,1]^{\mathbf{Z}}$ by an infinite automaton: though the extension space is not compact, some tools and the nature of the result: normality preservation, are the same as in the present article. But it is not a proper joining, since the second map does not commute with the shift. Their implicit tool is in fact a kind of infinite-state non-literal transducer.

ACKNOWLEDGEMENT: We want to thank C. Frougny, G. Hansel, Y. Mathieu, C. Mauduit, D. Perrin and especially G. Rauzy for various valuable contributions. The referee of Israel Journal of Mathematics quite wisely asked us to clarify the relation with [KW].

1. Definitions

Let A be a finite set of symbols endowed with the discrete topology, A^* be the set of all finite sequences on A; a **language** on A is any subset of A^*. The set $A^{\mathbf{Z}}$ of all biinfinite sequences on A, endowed with the product topology, is a compact metric space; so is the set $A^{\mathbf{N}}$ of infinite sequences. The **shift** $\sigma : A^{\mathbf{Z}} \to A^{\mathbf{Z}}$ (or $A^{\mathbf{N}} \to A^{\mathbf{N}}$) defined by $\sigma((x_n)_{n\in\mathbf{Z},\mathbf{N}}) = ((x_{n+1})_{n\in\mathbf{Z},\mathbf{N}})$ is a homeomorphism of $A^{\mathbf{Z}}$ (a continuous transformation of $A^{\mathbf{N}}$). A **subshift** on A is any closed σ-invariant subset of $A^{\mathbf{Z}}$ (or $A^{\mathbf{N}}$). A subshift X is unambiguously determined by the language $L(X) = \{u \in A^* \mid m, n \in \mathbb{Z}$ or $\mathbb{N},\ x \in X\colon x(m,n) = u\}$; this way a language with the right, almost trivial properties defines a subshift of $A^{\mathbf{Z}}$ and a subshift of $A^{\mathbf{N}}$. When not otherwise specified, a subshift is supposed to be in $A^{\mathbf{Z}}$. For $u \in L(X)$, denote by $[u]$ the set $\{x \mid x(0, |u| - 1) = u\}$.

A **transitive subshift** is one such that for any $u, v \in L(X)$ there exists $w \in A^*$ such that $uwv \in L(X)$.

A **factor map** is a continuous, onto, shift-commuting map $\varphi : X \to X'$; in this case X' is a **factor** of X and X an **extension** of X'. A conjugacy map is a one-to-one factor map: when there exists such a map, X and X' are said to be **conjugate**. Elements of $\varphi^{-1}(x)$ are called **lifts** of x. A **bounded-to-one** factor map is such that for x in X', the number of its lifts is bounded by some k.

Automata and Transducers.

An automaton $\mathcal{A}$ consists of:

—one finite alphabet A.

—one finite set of **states** C.

—a directed graph on C, the arcs each having a label in A.

To a path in the graph, one associates its **label**, i.e. the word spelled by concatenating the labels of its arcs (usually from left to right, but the natural automata for multiplication by an integer work from right to left); the set of labels is called **the language recognized by the automaton**. When using the automaton to recognize this language, all states are initial and final. But there is another, more complicated, word one may associate to a path: to any arc, associate the couple (a, c) of its label a and the origin vertex c. Then do the same for the path by concatenating all corresponding couples. The language thus recognized may be called simply the **language of** $\mathcal{A}$.

Each of these languages defines a subshift: they are the **factor subshift** $X \subset A^{\mathbf{Z}}$ and the **automaton subshift** $Y \subset (A \times C)^{\mathbf{Z}}$ associated to $\mathcal{A}$. Elements of these subshifts correspond to infinite paths in the graph: an element of X is the label of such a one, and an element of Y is the pair consisting of the sequence of vertices and the label of an infinite path. These two subshifts have particular properties (Y is a subshift of finite type; X must be at least sofic). $\mathcal{A}$ defines a mapping φ from Y to X by projection on A of coordinates of $y \in Y$. It is of course a factor map. It is convenient, though not strictly correct, to call also φ the corresponding projection map from $(A \times C)^*$ to A^*. The same definitions are fitting for simply infinite sequences.

An automaton $\mathcal{A}$ is said to be **deterministic** if, given $a \in A$ and $c \in C$, there is at most one arc starting from state c with label a. It is said to be **nonambiguous** if there is at most one path with given label from one state to another. Deterministic automata are nonambiguous, but the reverse is not true. For a deterministic automaton, one defines a partial map from $C \times A^*$ to C: $(c, u) \to c' = c.u$, every time there is a (unique) path starting from c, ending in c', with input label u, on $\mathcal{A}$.

An automaton is said to be **irreducible** if its graph is strongly connected, i.e. if for any two states c, c', there exists a path joining c to c' in the graph. If $\mathcal{A}$ is irreducible, subshifts Y and X are transitive.

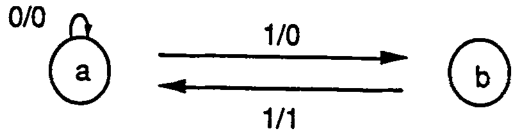
A transducer may be viewed as an automaton having two labels for each arc, or else as two automata sharing the same graph; it is often called a 2-automaton. More precisely, a **literal transducer** $\mathcal{T}$ (there are more general transducers we shall not consider in this paper) consists of:

1. two finite alphabets A and B;
2. one finite set of states C;
3. a directed graph on C,

the arcs each having an **input label** in A and an **output label** in B.

To a transducer one naturally associates two label languages, the **input** and **output languages**, and two factor subshifts, the **input** and **output subshifts** X and X'. The transducer subshift is defined as in the case of an automaton, but its symbol space is the product $A \times B \times C$. The definition of irreducibility is the same. One also defines **input-** and **output-deterministic** transducers, transducers which are nonambiguous for input and output. A **deterministic transducer** is one which is input- and output-deterministic.

Example 1.1: Here is an example of an input-deterministic transducer which is not output-deterministic, but nonambiguous for output; it is also irreducible. In this case X is the set of all sequences avoiding words $01^{2^n+1}0, n \in \mathbb{N}; X'$ is the set of all sequences avoiding word 11. ∎



An automaton $\mathcal{A}$ is said to be **bounded-to-one** (or a transducer $\mathcal{T}$ **bounded-to-one for input**) if the factor map φ is bounded-to one. The corresponding definition applies to output. When the automaton $\mathcal{A}$ is irreducible, map φ is bounded-to-one iff $\mathcal{A}$ is nonambiguous [B]; in the reducible case nonambiguousness is a sufficient condition for φ to be bounded-to-one.

Given a transducer $\mathcal{T}$, an element $x' \in X'$ is said to be a **transduced image** of $x \in X$ if they have a common lift: $\varphi^{-1}(x) \cap \psi^{-1}(x') \neq \emptyset$.

A transducer, especially when nonambiguous for input and output, defines a topologically significant relationship between its input and output subshifts. Two subshifts are said to be **finitely equivalent** [P3] if they have a common bounded-to-one extension. If X is the input and X' is the output of a bounded-to-one transducer, they are obviously finitely equivalent on account of their common extension Y. Finite equivalence creates a relation between points of X and points of X': x and x' are **related** if they have a common lift in Y; a transduced image is a special case of this situation. We shall also encounter subshifts having a common extension (not necessarily bounded-to-one): this extension is called a **joining** of the two. In this case points having a common preimage may also be called related.

Measures. Here we state all relevant facts about invariant measures on a compact metric space such as a subshift. Proofs, and other interesting properties, will be found in [DGS, K]. An invariant measure μ on subshift X is such that $\mu(\sigma^{-1}([u])) = \mu([u])$ for $u \in L(X)$. Recall the set $\mathcal{M}(X)$ of probability measures on X is a compact metric space for the topology of weak convergence, and the set $\mathcal{I}(X)$ of invariant probability measures on X is always nonempty and compact. Ergodic measures are those invariant measures for which σ-invariant sets have probability 0 or 1. The (topological) **support** of an invariant measure μ on X is the intersection of all closed invariant subsets of X having measure 1; its **entropy** is the nonnegative number

$$h_\mu = \lim_{n \to \infty} -\frac{1}{4} \sum_{c \in C_n} \mu(c) \log \mu(c)$$

where $C_n = \{[u] \mid u \in L(X),\ |u| = n\}$. The set of measures with maximal entropy on a subshift is always nonempty; when it is a singleton the subshift is said to be **intrinsically ergodic.**

For any subshift X, $x \in X$ and $f \in C(X)$, define the measure $S_n(x)$ by the formula

$$S_n(x, f) = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\sigma^i x}(f).$$

A point $x \in X$ is said to be **generic with regard to the invariant measure** μ on X, or simply μ-generic, if $S_n(x)$ converges weakly to μ as n goes to infinity; Birkhoff's pointwise convergence theorem states that if μ is an ergodic measure on X, μ-a.e. point in X is μ-generic. As genericity is an asymptotic property, depending only on nonnegative coordinates of $x \in X$, it is perfectly defined for sequences in $A^{\mathbb{N}}$. When $X = A^{\mathbb{N}}$ and μ is the Bernoulli measure λ with probability $1/\#A$ for each symbol, a generic x is called **normal**. Recall a real number $r \in [0, 1]$ is said to be **normal to the base** p if the asymptotic distribution of the sequence $rp^n (\bmod 1), n \in \mathbb{N}$, is the Lebesgue measure; this is equivalent to normality of its expansion in base p. Normal sequences are often called equidistributed in the literature.

Define $\mathcal{M}(x)$ as the set of measures **associated to** x, or limits of $(S_n(x))_{n \in \mathbb{N}}$ for the weak convergence of measure. Compactness implies it is always nonempty. Any measure in $\mathcal{M}(x)$ is invariant; x is generic for μ iff $\mathcal{M}(x) = \{\mu\}$.

Assuming φ: $Y \to X$ to be a factor map, denote by Φ: $\mathcal{M}(Y) \to \mathcal{M}(X)$ the corresponding map for measures: $\Phi(\nu) = \nu \circ \varphi^{-1}$. Map Φ is weakly continuous and shift-commuting. As a consequence, $\mathcal{M}(\varphi x) = \Phi(\mathcal{M}(x))$ and if $y \in Y$ is v-generic, $\varphi(y)$ is $\Phi(\nu)$-generic. Just as in the case of points, a **lift** of an invariant measure μ on X is an invariant measure ν of Y such that $\Phi(\nu) = \mu$.

A transducer, and more generally any kind of joining, acts on invariant measures in the same way as on points. A measure μ' on X' is said to be **related to** (when the joining is defined by a transducer, **a transduced image of**) measure μ on X if there exists $\nu \in \mathcal{M}(Y)$ such that $\Phi(\nu) = \mu$ and $\Psi(\nu) = \mu'$ (i.e. if they have a common lift).

There is no essential difference between the theory of invariant measures on subshifts of $A^{\mathbf{Z}}$ and subshifts of $A^{\mathbf{N}}$. Definitions are identical, so there is a one-to-one correspondence between invariant, or ergodic, measures on a simply infinite subshift and its doubly infinite version. Properties are the same. For instance, genericity may be considered as a property of a doubly infinite sequence, but it depends only on its restriction to positive coordinates. For the sake of convenience, results are stated for doubly infinite sequences in sections 2 and 3; they are applied to number expansions in some of the examples. ∎

2. Preservation of genericity for related points or transduced images

We first prove a general statement on genericity for two subshifts having a common extension.

PROPOSITION 2.1: *Let subshifts X and X' have a common extension Y through factor maps φ and ψ, and μ be an invariant measure on X.*

(1) *if x is μ-generic and x' is related to x, then any measure in $\mathcal{M}(x')$ is related to μ.*

(2) *if there is only one measure μ' related to μ, and x is μ-generic, then any x' related to x is μ'-generic.*

Proof:

(1) $\mu' \in \mathcal{M}(x')$ means there is an infinite subset $E \subset \mathbb{N}$ on which $S_n(x')$ converges weakly to μ'. As x' is related to x, there exists $y \in Y$ with $\varphi(y) = x$ and $\psi(y) = x'$. Consider the set of measures $\{S_n(y), n \in E\}$: since $\mathcal{M}(Y)$ is compact, there is a subset E' of E on which $S_n(y)$ converges weakly to some invariant limit ν. Continuity of Φ and Ψ, together with $\varphi(y) = x$ and $\psi(y) = x'$, imply $\Phi(\nu) = \mu$ and $\Psi(\nu) = \mu'$: so μ and μ' are related.

(2) The assumption implies $\mathcal{M}(x') \subset \{\mu'\}$. Thus by compactness of $\mathcal{M}(X')$ any subsequence of $S_n(x')$ tends to a limit which by statement (1) must be μ': hence the sequence itself tends to μ'. ∎

These results are also valid for extensions: just put $X' = Y$, $\psi = Id$.

From Proposition 2.1, 2 we shall deduce two different results, Proposition 2.2 and Corollary 2.6. Proposition 2.2 is fairly general as far as subshifts are concerned; its proof is based on an entropy argument. An answer to Rauzy's question may be deduced from it.

PROPOSITION 2.2: *Let X and X' be finitely equivalent subshifts with common extension Y and factor maps φ: $Y \to X$ and ψ: $Y \to X'$, and suppose they are intrinsically ergodic with maximal measures μ and μ'. Then if x is generic for μ, any related x' is generic for μ'.*

Proposition 2.2 might be proved in several ways, some of them certainly more satisfying than the one we chose. The following classical statement is essential for this proof (see [P2]).

PROPOSITION 2.3: *Let φ be a bounded-to-one factor map from subshift Y to subshift X, and ν be an invariant measure on Y. Then $h_\nu = h_{\Phi(\nu)}$.*

Proof of Proposition 2.2: We want to show the only measure on X' related to μ is μ'. As Y is a subshift, the set $\mathcal{M}_s(Y)$ of invariant measures with maximal entropy is nonempty. Applying Proposition 2.3 to factor maps φ and ψ, and intrinsic ergodicity of X and X', one obtains the equivalence of the three following statements:

—$\nu \in \mathcal{M}_x(Y)$;
—$\Phi(\nu) = \mu$;
—$\Psi(\nu) = \mu'$.

As a consequence, μ' is the only measure on X' related to μ. Conclusion follows from Proposition 2.1, 2. ∎

Of course when $X = X'$ this implies $\mu = \mu'$ and it is genericity with respect to this measure which is preserved. We first apply Proposition 2.2 to nonambiguous, hence bounded-to-one, transducers, and then give one counter example to illustrate the importance of the entropy hypothesis implicit in finite equivalence.

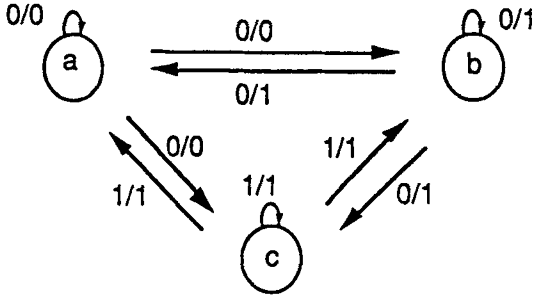
All subshifts recognized by some finite automaton (sofic systems) are known to be intrinsically ergodic provided they are transitive [F]; in particular, $A^{\mathbf{Z}}$ itself is obviously intrinsically ergodic. This remark, and Proposition 2.2, imply the following statement, which contains the answer to Rauzy's question.

COROLLARY 2.4: *Suppose transducer $\mathcal{F}$ is nonambiguous for input and output, and X and X' are transitive; let μ and μ' be the measures with maximal entropy on X and X'. Then any transduced image of a μ-generic point is μ'-generic. In particular, when $X = X' = A^{\mathbf{Z}}$, if x is normal, any transduced image of x is also normal.*

The assumption that X, X' and Y have the same topological entropy is essential in some way. When this is not the case, even supposing X and X' are intrinsically ergodic with the same maximal entropy, the lifts of measures μ and μ' may be distinct subsets of $\mathcal{M}(Y)$, so the proof no longer works.

Example 2.5: Here is a transducer with input and output $X = \{0,1\}^{\mathbf{Z}}$, but not bounded-to-one for input or output, and for which the conclusion of Corollary 2.4 does not hold. Let $C = \{a,b,c\}$, φ and ψ be the two factor maps, only depending on the states, defined by

φ: $a, b \to 0$; $c \to 1$.
ψ: $a \to 0$; $b, c \to 1$.



Suppose $x \in X$ is generic for μ: this means a on one hand, b and c on the other hand have the same frequency for $y \in \psi^{-1}(x)$, but for $y \in \varphi^{-1}(x)$, it means a and b on one hand, c on the other hand have the same frequency. Thus most transduced images of a generic x cannot be generic for μ.

In the proof of Proposition 2.2, topological entropy is a clumsy tool: the proof only works for the unique measure with maximal entropy. What about, for instance, generic points for some nonuniform Brenoulli or Markov measure on the full shift? Fortunately, by using different methods, one can answer this kind of question when the finite equivalence is defined by certain transducers, for a large class of ergodic measures having some common feature with the one with maximal entropy. We first prove an abstract statement on preservation of genericity by bounded-to-one extensions. The measures are just supposed to be invariant, but in fact our examples require ergodicity.

COROLLARY 2.6: *Let X be a factor of the subshift Y through map φ, and suppose μ is an invariant measure on X having a unique lift ν on Y. If $x \in X$ is μ-generic, then any lift y of x is ν-generic; if $X' = \psi(Y)$, $x' = \psi(y)$, then x' is $\Psi(\nu)$-generic.*

Proof: The first statement is a direct consequence of Proposition 2.1, 2. For the second, uniqueness of ν implies $\Psi(\nu)$ is the unique invariant measure on Y related to μ. ∎

And now, what about measures on X with entropy strictly less than the maximum? Proposition 2.7 below shows some instances in which Corollary 2.6 may be applied; others are to be found in section 4. For an ergodic measure ρ on a subshift, call G_ρ the measurable set of ρ-generic points. Let φ be a factor map from subshift Y onto subshift X.

Definition: An ergodic probability measure μ on X is said to have **property T** (with respect to φ) if it has an ergodic lift ν on Y such that for any measurable $F \subset Y$, $\mu(\varphi(F)) > 0$ implies $\nu(F) > 0$. ∎

Now we prove that measures having property T fulfill the assumptions of Corollary 2.6.

PROPOSITION 2.7: *Let φ be a factor map from Y to X. Suppose μ is an ergodic measure on X having property T. Then ν is the unique lift of μ.*

Proof: Birkhoff's ergodic theorem states that $\nu(G_\nu) = 1$. Suppose there exists an invariant $\nu' \neq \nu$ on Y, with $\Phi(\nu') = \mu$. The probability $\nu'(G_\nu)$ cannot have value 1: $\nu'(G_\nu) = 1$ would imply ν' is ergodic, since Cesare means converge ν'-a.s. to constants; then by Birkhoff's theorem this ergodic measure would have to be ν, contradicting the hypothesis $\nu' \neq \nu$. So $\nu'(G_\nu^c) > 0$. As $G_\nu^c \subset \varphi^{-1}(\varphi(G_\nu^c))$, this implies $\nu'(\varphi^{-1}(\varphi(G_\nu^c))) > 0$, or equivalently $\mu(\varphi(G_\nu^c)) > 0$ and, by property T, $\nu(G_\nu^c) > 0$. This contradicts ergodicity of ν. ∎

The next Proposition yields many examples of measures having Property T; actually, the hypotheses so obviously imply uniqueness of the lift that it is not necessary to use Proposition 2.7. A transducer (preferably nonambiguous for input) is said to be **(input-) synchronizing** if there exist a state c and an input word $u = u'u''$ such that any path in the graph with label u must reach vertex c at the end of prefix u'; in this case u is a **synchronizing word.** Assuming $\mathcal{F}$ to be input-synchronizing, an ergodic measure μ on X is said to be **synchronizing** if $\mu([u]) > 0$ for some synchronizing word $u \in A^*$. For instance, in Example 1.1 word 0 is synchronizing for input and 1 for output.

PROPOSITION 2.8: *Suppose transducer $\mathcal{F}$ is nonambiguous for input and input-synchronizing, and μ is a synchronizing measure on X. Then μ has property*

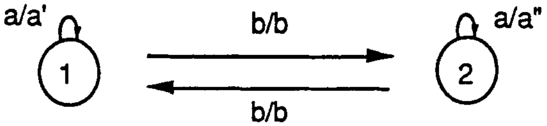
T, and there exists a measure μ' on X' such that any transduced image of a μ-generic point is μ'-generic.

Proof: One shows first that since μ is synchronizing, it has a unique lift ν on Y, possessing property T: as μ-a.s. $x \in X$ has infinitely many occurrences of some synchronizing word to the left and right, for given $n \in \mathbb{Z}$, one can find n_1, n_2, $n_1 < n < n_2$, such that the state coordinate at times n_1 and n_2 is necessarily c; as $\mathcal{F}$ is nonambiguous for input, this means all state coordinates are determined between n_1 and n_2; so, again μ-a.s., $x \in X$ has only one lift in Y. This defines a unique lift of μ on Y; but this also means that for any measurable $E \in Y$, one has $\nu(E) = \mu \circ \varphi(E)$: this equality implies Property T.

One concludes the proof by applying Corollary 2.6. ∎

Example 2.9: Here is a first, simple, example of a measure μ on X having several lifts on Y when considering transducer $\mathcal{F}$. In this case there are μ-generic points having no generic transduced images (for whatever transduced measure).

Consider the following transducer, for which X is the full shift on $\{a, b\}$ and X' is the subshift on $\{a', a'', b\}$ defined by excluding words $a'a''$, $a''a'$; $a'b^na'$, $a''b^na''$ for n odd, $a'b^na''$, $a''b^na'$ otherwise.



Let μ on X be the Dirac measure on the fixed point on letter a. This ergodic measure has two transduced images, one the Dirac measure on a', the other the Dirac measure on a''. It is obvious the sequence

$$x = a^{f(1)}ba^{f(2)}b \cdots a^{f(n-1)}ba^{f(n)}b \cdots$$

is generic for μ if $f(n)$ goes to infinity with n; but supposing

$$\frac{1}{f(n)} \sum_{i<n} f(i) \to 0,$$

neither of the two transduced images of x is generic for any measure. ∎

3. Continuity properties and (k, ϵ)-normality

In this paragraph, we address the question of preservation of almost genericity, and draw some consequences.

The notion of (k,ε)-normality was first introduced in [Bes] in a finite setting; it has a perfectly natural equivalent, for any invariant measure, in the field of infinite sequences.

Definition: Given an invariant measure μ on $A^{\mathbf{Z}}$, $k \in \mathbb{N}$ and $\varepsilon > 0$, $x \in A^{\mathbf{Z}}$ is said to be (k,ε,μ)-**generic** if there exists n_0 such that whatever $u \in A^k$, for $n > n_0$, $|S_n(x,[u]) - \mu([u])| < \varepsilon$. A (k,ε,μ)-generic word is a word $v = v_0 \cdots v_{N-1}$, $N > k$, such that whatever $u \in A^k$, if $E(u,v) = \#\{j \mid 0 \le j < N-k,\ v_j \cdots v_{j+k-1} = u\}$, then $|E(u,v)/N - \mu(u)| < \varepsilon$. An element $x \in X$ is (k,ε,μ)-generic iff the words $x(0,n)$ are (k,ε,μ)-generic for n large enough.

Before going on, we must first define the natural distance corresponding to the topology of weak convergence on $\mathcal{M}(X)$. After that we point out the relationship between this distance and (k,ε,μ)-genericity.

Let C_n be the set of cylinders corresponding to coordinates $-n$ to n. For $\mu, \mu' \subset \mathcal{M}(X)$, set

$$d(\mu,\mu') = \sum_{k=1}^{\infty} \frac{1}{2^k} \sup_{c \in C_k} |\mu(c) - \mu'(c)|.$$

The distance d endows $\mathcal{M}(X)$ with the topology of weak convergence, as is easily checked by applying the definitions. The following lemma characterizes (k,ε,μ)-genericity in terms of weak convergence.

LEMMA 3.1: *Given $\delta > 0$, there exist $k \in \mathbb{N}$ and $\epsilon > 0$ such that whenever $x \in A^{\mathbf{Z}}$ is (k,ϵ,μ)-generic, then $\mathcal{M}(x) \subset B(\mu,\delta)$. Conversely, for given k, ϵ, there exists δ such that if $\mathcal{M}(x) \subset B(\mu,\delta)$, then x must be (k,ϵ,μ)-generic.*

Proof: 1) Given δ, choose k such that for any two measures $\mu, \mu' \in \mathcal{M}(X)$ the partial sum of the series $d(\mu,\mu')$ is less than $\delta/2$ for words with length greater than k. Then choose ϵ small enough for the partial sum of the series for words with length less than or equal to k to be also less than $\delta/2$. Thus when x is (k,ϵ,μ)-generic one has $d(\mu,\mu') \le 2.\delta/2$ for any $\mu' \in \mathcal{M}(x)$.

2) Suppose x is not (k,ϵ,μ)-generic: there exist $u \in L(X)$, $|u| = k$, and an infinite subset $E \subset \mathbb{N}$ such that

$$|S_n(x,[u]) - \mu([u])| \ge \epsilon \quad \text{for } n \in E.$$

As $\mathcal{M}(X)$ is compact, there are an infinite $E' \subset E$ and an invariant measure μ_1 such that $S_n(x) \to \mu_1$ along E' in the sense of weak convergence. Thus

$|\mu_1([u]) - \mu([u])| \geq \epsilon$, and $d(\mu_1, \mu) > 2^{-k-1}\epsilon$. Setting $\delta = 2^{-k-1}\epsilon$, this means μ_1 cannot belong to $B(\mu, \delta)$. ∎

This implies in particular that a μ-generic point is one that is (k, ϵ, μ)-generic for all (k, ϵ). We now prove an elementary refinement of the classical result that a continuous bijection between two compact sets is also bicontinuous. This result is probably known and applied in some field, though we could not find it in the literature.

LEMMA 3.2: *Let K, K' be two compact spaces, $\varphi : K \to K'$ be continuous and onto, and suppose $k' \in K'$ has a unique preimage k in K. For any open set U containing k, there exists an open set U' containing k' such that if $\varphi(k_1) \in U'$, then $k_1 \in U$.*

Proof: Let $U \subset K$ be an open set containing k. On account of compactness of K, $\varphi(U^c)$ is closed; it cannot contain k' since the unique element with image k' is in U. Thus $(\varphi(U^c))^c$ is an open neighborhood of k' with preimage contained in U. ∎

Now we may prove the main result of this paragraph.

PROPOSITION 3.3: *Let subshifts X and X' have a common extension Y through factor maps φ and Ψ, ν be an invariant measure on Y, $\mu = \Phi(\nu)$, $\mu' = \Psi(\nu)$; suppose furthermore ν is the unique lift of μ. Then for any $k \in \mathbb{N}$, $\epsilon > 0$, there exist $k' \in \mathbb{N}$, $\epsilon' > 0$ such that if $x \in X$ is (k', ϵ', μ)-generic any lift y is (k, ϵ, ν)-generic. The same is true when one replaces y by $x' \in X'$, related to x, and ν by μ'.*

Proof: Lemma 3.1 states that it is equivalent for y to be (k, ϵ, ν)-generic and for the distance $d(\nu, \nu')$ to be less than some δ, whatever $\nu' \in \mathcal{M}(y)$.

Apply Lemma 3.2 to the compact metric spaces $\mathcal{J}(Y)$, $\mathcal{J}(X)$, the continuous map Φ, and measures ν and μ. Thus for any $\delta > 0$, there exists $\delta' > 0$ such that if $\varphi(\nu_1) = \mu_1$ and $d(\mu, \mu_1) < \delta'$, then $d(\nu, \nu_1) < \delta$.

To finish the proof there remains only to use the direct part of Lemma 3.1 and choose values of k' and ϵ' corresponding to δ'. The second statement is derived from the first by continuity of Ψ. ∎

To illustrate this and for further use, let us give a derived statement in the case of a transducer (certainly not the best we can get, but sufficient for our needs).

The three subshifts Y, X, and X' recognized by an irreducible transducer are known to be intrinsically ergodic [F].

COROLLARY 3.4: *Let $\mathcal{F}$ be an irreducible bounded-to-one transducer, with μ and μ' the measures with maximal entropy on X and X'. Then the conclusion of Proposition 3.3 holds. In particular, when $X = X' = A^{\mathbf{Z}}$, and λ is the uniform measure on X and X', given k' and ϵ', there exist k and ϵ such that if x is (k, ϵ, μ)-generic, then any related x' is (k', ϵ', μ')-generic.*

Proof: Intrinsic ergodicity of X and Y, together with Proposition 2.3, imply the unique lift of μ is the unique measure ν with maximal entropy on Y; the same is true for μ'. Proposition 3.3 then concludes the proof. ∎

Remark: We know a much longer, combinatorial proof of Corollary 3.4 when $X = X' = A^{\mathbf{Z}}$, and $\mu = \mu'\lambda$; it does not require irreducibility for $\mathcal{F}$, nor even intrinsic ergodicity of Y, and suggests an implicit correspondence between (k, ϵ) and (k', ϵ') which might be of some use. ∎

4. Examples and applications

A. Applications to Number Theory.

1. PROPERTIES OF MULTIPLICATION TRANSDUCERS. Now we want to investigate relevant properties of multiplication transducers, so as to be able to use results obtained in the last two sections for number-theoretical purposes. The following results do not pretend to originality; as many arithmeticians seem not to know about them, we think it is better to give them here.

The transducer $\mathcal{F}_{k,p}$ of multiplication by k in base p is just a representation of the usual algorithm. Each of the sets A and B is equal to $\{0, 1, \ldots, p-1\}$. The state set C is just the set of all possible values of the carry. Like the algorithm, the transducer acts from right to left.

Call c_n the carry at time n. The following formula yields the output b_n at time n, when one knows the carry and input a_n at the same time:

$$b_n = ka_n + c_n (\mathrm{mod}\ p) \tag{1}$$

Putting $f(n) = [n/p])$, from (1) one deduces the recurrence formula

$$c_{n-1} = f((ka_n + c_n)); \tag{2}$$

just remark the fractional part $\{(ka_n + c_n)/p\}$ is just $1/p$ times the output b_n. An immediate consequence of (2) is that the carry may always be chosen less than k. As c_{n-1} is a nondecreasing function of c_n and a_n, it is sufficient to check $c_{n-1} < k$ for $c_n = k-1$ and $a_n = p-1$. Applying (2), one gets

$$c = f((k(p-1)+k-1)) = f(kp-1) = k-1. \qquad \blacksquare$$

The graph of $\mathcal{F}_{k,p}$, together with the input labels, can be deduced from (2). Formula (1) gives the output labels. Formula (2) also testifies that $\mathcal{F}_{k,p}$ is always input-deterministic (this only means once the input and carry at time n are known, then one can deduce the carry at time $n-1$). Now, need all states from 0 to $k-1$ be considered, or will some proper subset be enough? One could do without the next Proposition, only proving the result for $k \leq p$, and restricting $\mathcal{F}_{k,p}$ to the strongly connected subgraph containing state 0 for all other cases. But this connected component is in fact equal to the graph itself.

PROPOSITION 4.1: *The graph of multiplication by k in base p, where C is the set of all carries from* 0 *to $k-1$, is always irreducible and synchronizing for input.*

Proof: Fix a base $p > 1$. We first show that given $k > 0$ there exists an integer q such that in $\mathcal{F}_{k,p}$, starting from any carry, the input word 0^q leads to carry 0. Suppose $a_n = 0$ and $c_n > 0$; then (2) implies $c_{n-1} = [c_n/p] < c_n$: input 0 strictly diminishes the carry at the next step. If $a_n = 0$ and $c_n = 0$, c_{n-1} is obviously 0. This means any path with input label 0^{k-1} is a synchronizing word, and 0 may be reached starting from any state in the graph.

The same is true for state $k-1$ and word $(p-1)^n$ for some n. It is sufficient to show that if $0 \leq c < k-1$, then $c' = f(c + k(p-1)) > c$, because if this implication is true, as we already know that $f((k(p-1)+k-1)) = k-1$, input $(p-1)^{k-1}$ necessarily leads from any state to the maximum $k-1$. So assume $0 \leq c \leq k-2$, i.e. $k \geq c+2$. Thus $c' \geq f(c+(c+2)(p-1)) = f(cp+2p-2)$, but as $p-2 \geq 0$, the last term is equal to $c + f(2p-2) \geq c + f(p) + f(p-2) \geq c+1$.

In order to prove any state in C can be connected to any state , there remains now to show any state in C can be reached starting from 0. We shall do this by induction on k. Let us call E the set of all integers k for which $\mathcal{F}_{k,p}$ has this property. Also, c is said to be k-accessible by input u if there exists a path from 0 to c with input label u in $\mathcal{F}_{k,p}$.

The initial step is already made, because $\mathcal{F}_{k,p}$ has only two states, 0 and $k-1=1$, and we just proved these are k-accessible in any base.

Now, assume that $k \in E$: to any carry $c \leq k-1$ one can associate a sequence of inputs $u = u_1u_2\cdots u_n \in \{0,\ldots,k-1\}^*$, such that c is k-accessible by u. $\mathcal{F}_{k,p}$ is deterministic, so u completely determines the state c_i of the transducer after input $u_1u_2\cdots u_i$, $i \leq n$, starting from $c_0 = 0$ and arriving to $c_n = c$. One may assume no two states among the (c_i) are identical, or else there would exist a shorter word u' having the same property (in particular, $u_i \neq 0$). But u is also an input for $\mathcal{F}_{k+1,p}$, so that one can define the state c'_i in which $\mathcal{F}_{k+1,p}$ is after input $u_1u_2\cdots u_i$. For $c' = c'_n$, we want to prove

$$c \leq c' \leq c+1. \tag{3}$$

We do this by induction on c_i and c'_i. It is evident for $c_0 = c'_0 = 0$. Now suppose it is true for i: $c_{i+1} = f(ku_i + c_i)$, and $c'_{i+1} = f(ku_i + u_i + c'_i)$ is obviously non-smaller than c_{i+1}; also $c'_{i+1} \leq f(ku_i + c_i + 1)$; as $u_i + 1 \leq p$, $c'_{i+1} \leq f(ku_i + c_i) + 1 = c_{i+1} + 1$. So (3) is true.

There remains to make use of (3) in order to prove the general step of the induction. To show that state i $(0 \leq i < k+1)$ is $k+1$-accessible, it is sufficient to check there are j and r, $0 \leq j < p$ and $0 \leq r < k+1$ such that

a) r is $k+1$-accessible by input u,

and

b) $f(j(k+1)+r) = i$.

If this is true, i is $(k+1)$-accessible by input uj.

Choosing the only value of j such that $j(k+1) \leq pi < (j+1)(k+1)$, put $c = pi - j(k+1) < k+1$. There are two cases:

—either $c < k$. So by hypothesis c is k-accessible by a certain input u. Choose $r = c'$. Applying (3), $r < k+1$ and $f(j(k+1)+r) = f(pi+d)$, $d = 0$ or 1. Thus $f(j(k+1)+r) = i$.

—or $c = k$. But we know that k is $k+1$-accessible by a certain input $(p-1)^n$. So choose $r = c$, and i is $(k+1)$-accessible by $(p-1)^n j$.

Thus $k \in E$ implies $k+1 \in E$, and the result follows. ∎

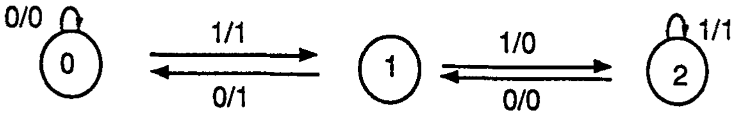
We shall examine output properties of the transducer only in two particular cases: they are sufficient for our needs.

LEMMA 4.2: *If k and p are relatively prime, $\mathcal{F}_{k,p}$ is output-deterministic and, given $b \in B$, $c \in C$, there exists exactly one arc starting from c with output label b.*

Proof: As k and p are relatively prime, Formula (1) defines a bijection on $\mathbb{Z}/p\mathbb{Z}$. So arcs starting from state c and corresponding to distinct inputs must have distinct outputs. As there are exactly p arcs starting from c, all with distinct inputs, this means all outputs are obtained, each for one arc only. ∎

It is easy to show that $\mathcal{F}_{k,p}$ is never output-synchronizing when k and p are relatively prime.

Example 4.3: This is the graph of $\mathcal{F}_{3,2}$:



The following Lemma means when k divides p, $\mathcal{F}_{k,p}$ is nondeterministic for output but, given a doubly infinite sequence of output letters, there is only one way to choose the corresponding carries.

LEMMA 4.4: *Suppose $p = kq$, $q \in \mathbb{N}$. Then:*

1) *Given $b \in B$, there is only one state c accepting b as an output.*
2) *Given $b \in B$, the corresponding state c and any $c' \in C$, there is an arc from c to c' with output b.*
3) *$\mathcal{F}_{k,p}$ is nonambiguous for output.*

Proof:

1) If a, $a' \in A$, $a' = a$ (mod$q$), then the two corresponding output letters starting from c are the same: from equality $a' = a + qr$ and Formula (1), one gets $b = ka + c(\text{mod}p)$, $b' = ka + kqr + c = b(\text{mod}p)$. This proves there are only q possible outputs for arcs starting from c, and also there are exactly k arcs starting from c with given output b. There are k states in C, q possible outputs from each state and a total number of $p = kq$ outputs: this implies there is only one state accepting b as an output.

2) Supposing $c \in C$, $a' = a + qr$ with $r > 0$, the two arcs starting from c with inputs a and a' have the same output but end in different states. Applying

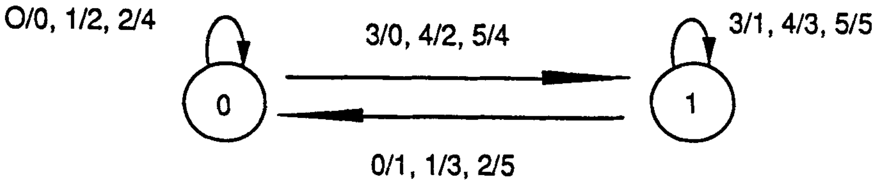
formula (2), the end vertex corresponding to input a' is

$$f(ka + kqr + c) = f(ka + c) + r > f(ka + c),$$

where the last term is the end vertex corresponding to input a. Thus the $p/q = k$ arcs with given output b starting from c end in all the k distinct states of C.

3) Given output b, the only state c accepting b as an output is completely determined: this implies given the label of a path, only the end vertex is not determined; one concludes the proof by remarking all arcs with label b end in different states. ∎

Example 4.5: Multiplication by p in base p just corresponds to the shift to the right on expansions: it is too trivial to be considered an example. Here is a less trivial one; this is the graph of $\mathcal{F}_{2,6}$:



PROPOSITION 4.6: *If r is a rational, for any k, ϵ there exist k', ϵ' such that if x is (k', ϵ', μ')-normal then rx is (k, ϵ, μ)-normal.*

Proof: Given base p and multiplier k, decompose k into the product of an integer h relatively prime to p, and of a finite sequence $\alpha, \cdots, \kappa$ of divisors of p. By Proposition 4.1 Corollary 3.4 may be applied to deterministic transducers $\mathcal{F}_{h,p}, \mathcal{F}_{\alpha,p}, \cdots, \mathcal{F}_{\kappa,p}$. This proves the result for multiplication by an integer. To get the same result for division by k, apply Lemma 4.2 to $\mathcal{F}_{h,p}$, Lemma 4.4 to the other transducers and then Corollary 3.4. Remarking this is true for any k concludes the proof. ∎

2. ADDITION OF A RATIONAL. It is also well known that addition of a rational preserves normality (see[R1]). But this operation may also be performed with the help of a transducer.

Let $r \in \mathbb{Q}$, $p \in \mathbb{N}$, $A = \{0 \ldots p-1\}$ and $c_0 \ldots c_{n-1}$ be one ultimate periodic pattern of the expansion of r in base p. The following formulas describe the graph of an input-deterministic transducer $\mathcal{S}$ acting from right to left (it represents the usual addition algorithm, applied to this particular case):

$$b_i = a_i + \alpha_i + \epsilon_i \ (\text{mod} p),$$
$$\alpha_{i-1} = c_{k-1 \bmod n} \text{ for } \alpha_i = c_k,$$
$$\epsilon_{i-1} = [(a_i + \alpha_i + \epsilon_i)/p],$$

where a_i and b_i are the input and output at time i, both belonging to A, $\alpha_i \in \{c_i,\ 0 \leq i < n\}$ marks the position of the index relative to the periodic pattern of the expansion of r at time i, and $\epsilon \in \{0,1\}$ is the carry.

It is easy to check $\mathcal{S}$ is irreducible and deterministic for output, so all results in sections 2 and 3 may be applied to its action; of course, in most cases $\mathcal{S}$ does not perform exactly addition of r, because the expansion of r is only ultimately periodic, but its output is equal to the expansion of $x+r$ but for a finite number of coordinates, so asymptotic properties like normality are preserved. In particular, addition of a rational preserves "almost normality".

B. Markov Measures.

Now let us consider the case of Markov measures. From now on assume $\mathcal{F}$ is an irreducible input-deterministic transducer with input automaton $\mathcal{A}$, and input subshift X is of finite type with excluded words of length at most 2. We suppose μ is an irreducible homogeneous (i.e. invariant) Markov measure with transition probability $(p_{aa'})$, a, $a' \in A$, and with topological support X: this means X is defined by excluding words aa' with $p_{aa'} = 0$; and we want to find a measure ν on Y such that any lift of a μ-generic point is ν-generic. Before we construct such a measure, we must prove a lemma on finite automata.

Lemma 4.8: *Suppose $\mathcal{A}$ is irreducible and deterministic, and X of finite type, defined by excluding words of length at most k. Then for any path with length $k-1$, from state c to state c', having label m, for any $a \in A$, $c'.a$ is defined iff $ma \in L(X)$.*

Proof: If the conclusion is false, there exists a path from c to c' with label m, such that $ma \in L(X)$ by $c'.a$ is undefined. We construct an infinite sequence (c_n) of distinct states and a sequence (m_n) of words, with m_{n-1} a prefix of m_n, such that, for any n, any $i < n$, $c_n.m_i$ is defined but $c_n.m_n$ is not, thus contradicting the finiteness of state space C.

Put $c_0 = c$, $m_0 = ma$. Given $c_{n-1} \in C$, $m_{n-1} \in L(X)$, choose c_n such that $c_n.m_{n-1}$ is defined, and, using irreducibility, m'_n such that $c_n.m_{n-1}m'_n = c_0$. Set $m_n = m_{n-1}m'_n ma$: its prefix $m_{n-1}m'_n m$ checks $c_n.m_{n-1}m'_n m = c'$, hence belongs to $L(X)$; as X is defined by excluding words with length less than $k+1$, and $|m| = k-1$, there remains to check ma belongs to $L(X)$: this is true by hypothesis. But $c_n.m_n = c_0.ma$ is undefined. ∎

The next Proposition is closely connected with results of [F] concerning the measure with maximal entropy on sofic systems:

PROPOSITION 4.9: *Let $\mathcal{A}$ be an irreducible deterministic automaton, with X a subshift of finite type defined by excluding words with length 2, μ an irreducible invariant Markov measure on X, with support X. Then*

(1) *There exists a unique lift ν of μ on Y. It is an ergodic Markov measure having Property T.*

(2) *The transition probability of ν is deduced from that of μ by the formula*

$$p_{\alpha\alpha'} = p_{aa'} \quad \textit{for } \alpha = (a,c), \alpha' = (a',c') \textit{ and } \alpha\alpha' \in L(Y).$$

Proof: Let D be the set of all $\alpha = (a,c) \in A \times C$ such that there is an arc starting from c with label a; Y is the subshift of finite type on D defined by excluding all words $\alpha\alpha'$ with $c.a \neq c'$.

(a) Let us first build up a Markov measure on Y satisfying (2). We must check the $|D| \times |D|$-matrix M defined by

$$M_{\alpha'\alpha} = p_{aa'} \quad \text{whenever } \alpha\alpha' \in L(Y), \ 0 \text{ otherwise},$$

is equally stochastic, i.e. for any $\alpha = (a,c) \in D$, the sum of transition probabilities $p_{aa'}$ on set $\{a' \colon \alpha' = (a',c'),\ \alpha\alpha' \in L(X)\}$ is equal to 1. For α, $\alpha' \in D$, the condition $\alpha\alpha' \in L(Y)$ is equivalent to $c.a = c'$. By Lemma 4.8, $(c.a).a'$ is defined iff $aa' \in L(X)$: as μ is a Markov measure $\Sigma_{aa' \in L(X)} p_{aa'} = 1$, so M is a stochastic matrix. As $\mathcal{A}$ is irreducible, and to any arc in the graph is associated a positive transition probability, M defines an ergodic Markov measure ν (with support Y): let $\mathbf{V} = (\nu_\alpha)_{\alpha \in D}$ be the unique eigenvector of M with norm 1 corresponding to eigenvalue 1. All its coordinates are known to be positive [G]. The formula

$$\nu([w]) = \nu(a_0, c_0) p_{a_0a_1} \cdots p_{a_{n-1}a_n} \quad \text{for } w = (a_0, c_0) \cdots (a_n, c_n) \in L(Y)$$

defines a probability ν on Y; ν is invariant because $\mathbf{V}$ is an eigenvector of matrix M, ergodic because it is possible to prove $\nu([w]) > C.\ \mu_0\varphi([w])$ for some constant C.

(b) We must check $\Phi(\nu) = \mu$,i.e. $(\Phi(\nu))([u]) = \mu([u])$ for any $u = a_0 \cdots a_n \in L(X)$. One has

$$(1) \qquad (\Phi(\nu))([u]) = (\Sigma_{c\in C}\nu(a_0,c))p_{a_0a_1}\cdots p_{a_{n-1}a_n},$$

defining $\nu(a,c)$ to be 0 whenever $(a,c) \notin D$ Since $\mathbf{V}$ is an eigenvector of M, for any $\alpha' = (a',c') \in D$,

$$\Sigma_{\alpha\in D}M_{\alpha'\alpha}\nu(\alpha) = \nu(\alpha')$$

or

$$\Sigma_{a,c:c.a=c'}p_{aa'}\nu(a,c) = \nu(a',c').$$

Now, summing up on $c' \in C$,

$$\Sigma_{a,c}p_{aa'}\nu(a,c) = \Sigma_{c'}\nu(a',c').$$

This means the vector $\mathbf{W}$ defined by $\mathbf{W}_a = \Sigma_{c\in C}\nu(a,c)$ for $a \in A$ is an eigenvector of matrix $(p_{aa'})_{a,a'\in A}$. As its norm is 1, it is equal to $(\mu(a))_{a\in A}$: effecting the summation in (1), this proves $\Phi(\nu([u])) = \mu([u])$.

(c) As $\nu(\alpha) > 0$ for $\alpha \in D$, $\mu(\varphi(E)) > 0$ implies $\mu(E) > 0$: for any word $y = y_0 \cdots y_n \in L(Y), y_i = (a_i, c_i)$, with $u = a_0 \cdots a_n$ one has

$$\nu([y]) = \mu(a_0,c_0)p_{a_0a_1}\cdots p_{a_{n-1}a_n}$$

and

$$\mu \circ \varphi([y]) \leq \mu([u]) = \mu(a_0)p_{a_0a_1}\cdots p_{a_{n-1}a_n} \leq (\sup_{a\in A,\alpha\in D} \mu(a)/\mu(\alpha)).\nu([y]).$$

This implies the corresponding inequality for all measurable sets. So Property T holds and, by Proposition 2.7 and Corollary 2.6, ν is the unique lift of μ. ∎

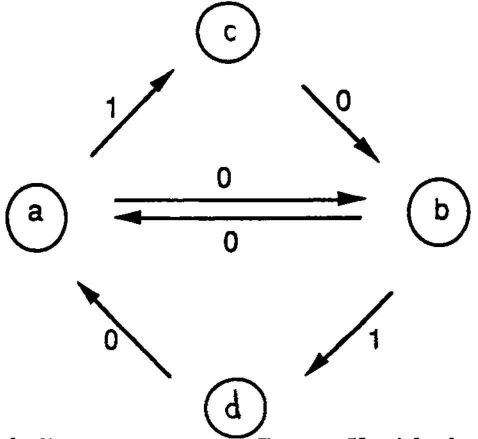
It is possible to extend this result to generalized Markov measures; the proof is practically the same, except the formulas are more complicated.

Remark 4.10: Before calculating measure ν, check on the graph whether $\mathcal{A}$ and μ have the following property: for all $(a',c') \in D$, and for all $a \in A$ such that there exists an arc with label a ending in vertex c', the value of $p_{aa'}$ does not depend on a; then denote by $p_{(a',c')}$ this common value. Let us look for the eigenvector $(\nu(a,c))_{(a,c)\in D}$ under the form $\nu(a,c) = \nu(c).p_{(a,c)}$, with $\nu(c)$ still unknown. But in this case equation (2) becomes

$$p_{(a',c')}\Sigma_c\nu(c)(\Sigma_{a:c.a=c'}p_{(a,c)}) = \nu(c')p_{(a',c')}.$$

Now, choose for $\nu(c)$ an eigenvector with norm 1 of matrix M' defined by $M'(c',c) = \Sigma_{a:c.a=c'}p_{(a,c)}$; equality $\Sigma_{(a,c)}\nu(a,c) = 1$ holds since, by Lemma 4.8, $\Sigma_{a:(a,c)\in D}p_{(a,c)} = 1$. In this case the image of ν on $C^{\mathbf{Z}}$ is Markov with transition matrix M'.

Example 4.11: Corollary 2.6 may be used in simple cases when Proposition 2.8 may not: here are a nonsynchronizing automaton $\mathcal{A}$ and a Markov measure μ on X to which it may be applied. This is the graph of $\mathcal{A}$:



X is defined by excluding the word 11. Endow X with the Markov measure defined by the transition probabilities $p_{00} = p \in [0,1[, p_{01} = q = 1-p$, and naturally $p_{10} = 1$. Then we are in the case of Remark 4.10:

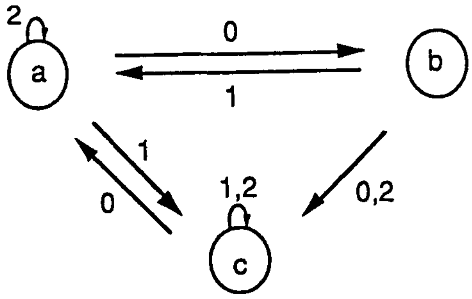
$$M' = \begin{pmatrix} 0 & p & 0 & 1 \\ p & 0 & 1 & 0 \\ q & 0 & 0 & 0 \\ 0 & q & 0 & 0 \end{pmatrix}$$

and the eigenvector with coordinates

$$\nu(a) = \nu(b) = (2+2q)^{-1}, \quad \nu(c) = \nu(d) = q.(2+2q)^{-1},$$

together with transition probabilities of μ, define an invariant Markov measure ν on Y with image μ, which is unique by Proposition 4.9.

Example 4.12: As in Example 2.9, the following automaton $\mathcal{A}$ is such that there is a Markov measure μ on X having several lifts on Y, so the lifts of μ-generic points do not behave nicely. This example is more sophisticated than Example 2.9: the support X_1 of μ is not obtained by just cutting out some arcs in the graph of $\mathcal{A}$ so as to make it reducible.



$X = \{0,1,2\}^{\mathbf{Z}}$, μ Markov with transition probability

$$P = \begin{pmatrix} 0 & 1/2 & 1 \\ 1 & 0 & 0, \\ 0 & 1/2 & 0 \end{pmatrix} \text{ hence } \begin{pmatrix} \mu(0) \\ \mu(1) \\ \mu(2) \end{pmatrix} = \begin{pmatrix} 2/5 \\ 2/5 \\ 1/5 \end{pmatrix};$$

one easily checks P^5 has positive coefficients, so μ is irreducible; nevertheless the support of μ is subshift X_1 defined by excluding all words corresponding to zero coefficients of the matrix: 00, 02, 11, 21 and 22. Suppose $x = (a_n) \in X_1$ is μ-generic. Its three preimages, corresponding to three possible choices for coordinate c_0, never have the same statistical properties:

— if $a_0 \in \{0,2\}$, choice $c_0 = a$ implies c can never be reached, whereas choice $c_0 = c_c$ implies $\#\{n < N/c_n = c\} \geq N/2$ for all N;

— if $a_0 = 1$ then $c_0 = b$ means c can never be reached; $c_0 = c$ implies $\#\{n < N/c_n = c\} \geq N/2$.

Example 4.13: When transducer $\mathcal{F}$ and Markov measure μ fulfill the assumptions of Proposition 4.9, any transduced image x' of a μ-generic x is μ'-generic with $\mu' = \Phi(\nu)$. But μ' is generally not Markov. For instance, put $\mathcal{F} = \mathcal{F}_{3,2}$ (Example 4.3) and let μ be the Bernoulli measure on $\{0,1\}$ with $\mu([0]) = p$, $\mu([1]) = 1 - p$, $p \in (0,1)$. A straightforward calculation shows that $\mu'([111])/\mu'([11]) = q$ whereas $\mu'([11])/\mu'([1]) = (p_3 + qp_2 + q_3)/(2p_2 + q_2)$; so μ' is Markov iff $p = 1/2$, or $\mu = \mu'$ is the uniform measure. In other cases genericity is preserved, but not genericity with regard to one particular measure. ∎

C. Transducers Preserving Normality While Changing the Base.

In what conditions may normality be preserved — or rather, normality to the base p be mapped to normality to the base p'— when the assumptions of Proposition 2.2 are not fulfilled? This is the question we try to answer here.

Let $\mathcal{F}$ be an irreducible input-deterministic transducer such that $X = A^{\mathbf{Z}}$, $X' = B^{\mathbf{Z}}$. As $\mathcal{F}$ is input-deterministic, Proposition 2.3 shows one cannot have $\#B \geq \#A$. If $\#B = \#A$, it is possible to show $\mathcal{F}$ must be nonambiguous for output. So the only remaining case is $\#B < \#A$. As was said in the Introduction, preservation of normality means if $x \in A_{\mathbf{Z}}$ is normal to the base $\#A$, then any transduced $y \in B_{\mathbf{z}}$ is normal to the base $\#B$.

Let φ: $A \times C \to B$ be the output function of $\mathcal{F}$: $\varphi(a, c)$ is the output of the arc starting from c with input a; by Lemma 4.8 this arc exists, and it is unique because τ is deterministic.

To express these conditions, we need to use the matrix $M(u)$ on C, associated to each output word $u \in B_*$ ($(\#A)^n M(u)_{c'c}$ being the number of paths from c to c' with output label u, $|u| = n$). Remark $u \to M(u)$ is an antimorphism, as $M(uv) = M(v)M(u)$.

The stochastic matrix $M' = \Sigma_{b \in B} M(b)$ has an eigenvector $\mathbf{V}$ with rational coordinates, for which

$$M'.\mathbf{V} = \mathbf{V}, \quad \mathbf{1}.\mathbf{V} = 1 \quad (\mathbf{1} = (1, \ldots 1)).$$

The following Proposition allows us to decide whether a suitable transducer from $A^{\mathbf{Z}}$ to $B^{\mathbf{Z}}$ preserves normality. Remark the hypotheses of the following Proposition do not depend on input labelling, provided $\mathcal{F}$ is deterministic.

Proposition 4.14: *Suppose $\mathcal{F}$ is an irreducible input deterministic transducer with $X = A^{\mathbf{Z}}$, $X' = B^{\mathbf{Z}}$.*

(i) *If $x \in A^{\mathbf{Z}}$ is normal, the limiting frequency of a word $u \in B^*$ in any transduced image $x' \in X'$ is equal to $\mathbf{1}.M(u).\mathbf{V}$.*

(ii) *A necessary and sufficient condition for x' to be normal for any normal x, is given by a finite number of conditions:*

$\Psi(\nu)([u]) = \mathbf{1}.M(u).\mathbf{V} = (\#B)^{-n}$ *for all words $u \in B^*$ with length $n = \#C$.*

Proof:

(i) The uniform measure on X is Markov with transition probability $p_{a'a} = (\#A)^{-1}$. Hence the matrix M' is the one described in Remark 4.10; by Proposition 4.9, it permits to construct measure ν on Y. One has $\mathbf{V} = (v(c))_{c\in C}$. By Corollary 2.6, for any normal $x \in X$, any transduced image $x' \in X'$ is $\Psi(\nu)$-generic.

The limiting frequency of a word $u \in B^*$ in x' is equal to

$$\Psi(\nu)([u]) = \Sigma_{\varphi(y)=u}\nu([y]).$$

As $\nu([y]) = \nu(c)(\#A)^{-n}$, with c such that $y_0 = (.,.,c)$,

$$\Psi(\nu)([u]) = \Sigma_{c\in C}\nu(c).\Sigma_{c'\in C}M(u)_{c'c} = \mathbf{1}.M(u).\mathbf{V}.$$

(ii) The conditions are obviously necessary. Conversely, suppose $\mathbf{1}.M(u).\mathbf{V} = (\#B)^{-n}$ for all $u \in B^n$ with $n = \#C$: to ensure normality of x', one need just prove the same equality for all $n \in \mathbb{N}$.

For $i \in \mathbb{N}$, define E_i as the subspace of $\mathbb{R}^{\#C}$ generated by row-vectors

$$\mathbf{V}(b,b',u)\mathbf{1}.(M(ub) - M(ub')) \quad \text{for } b,b' \in B \text{ and } u \in B^*, |u| = i-1$$

and define $E_0 = \{0\}$. The key fact is that each subspace $E'_i = \bigoplus_{k\leq i} E_k$ is strictly included in the next, up to some N after which they are identical. This results from the algebraic relationship between generators of E_{i+1} and E_i:

$$(1) \qquad \mathbf{V}(b,b',b_1u) = \mathbf{V}(b,b',u).M(b_1) \quad \text{for } b_1 \in B, |u| = i-1.$$

Indeed if $E'_{i+1} = E'_i$, then $E'_{i+2} = E'_{i+1}$: each generator of E_{i+2}, $V(b,b',b_1u)$ with $|u| = i$, is the product by $M(B_1)$ of a generator of E_{i+1}, which is a linear combination of generators of $E_k, k \leq i$; so $\mathbf{V}(b,b',b_1u)$ is a linear combination of generators of $E_{k+1}, k \leq i$, and so belongs to E'_{i+1}.

The subspaces E'_i must strictly increase up to some index N; as their dimensions are bounded by $\#C$, one has $N \leq \#C$.

Equalities $\Psi(\nu)([u]) = (\#B)^{-|u|}$ for $|u| \leq n$ are implicit in the assumptions. Finally one has also $\mathbf{V}(b,b',u).\mathbf{V} = 0$ for $|u| > \#C - 1$, since $E'_n = E'_N$ and $E'_N = E'_{\#C}$.

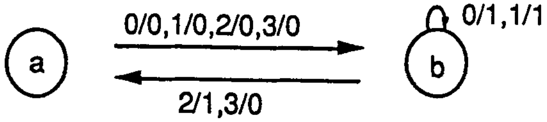
In other terms, given $u \in B^*$, all $\mathbf{1}.M(ub).\mathbf{V}$ are equal for $b \in B$; since their sum over $b \in B$ is equal to $\mathbf{1}M(u)\mathbf{V}$, they are equal to $(\#B)^{-1}\mathbf{1}.M(u).\mathbf{V}$. By induction, $\mathbf{1}.M(u).\mathbf{V} = (\#B)^{-|u|}$, whatever $u \in B^*$. ∎

The following result displays restrictive conditions for preservation of normality by a transducer from $A^{\mathbf{Z}}$ to $B^{\mathbf{Z}}$.

PROPOSITION 4.15: *Let $\mathcal{F}$ be an irreducible input -deterministic transducer with $X = A^{\mathbf{Z}}$ and $X' = B^{\mathbf{Z}}$. Preservation of normality implies $\#B$ divides $\#A$.*

Proof: *For $b \in B$ and $n \in \mathbb{N}$, one has $\mathbf{1}.M(b^n).\mathbf{V} = (\#B)^{-n}$ and, if q is the common denominator of coordinates of $\mathbf{V}$, as $(\#A)^n M(b^n)$ and $q\mathbf{V}$ have integer coordinates by definition of $M(u)$ and q, $q(\#A)^n\mathbf{1}.M(b^n).\mathbf{V}$ is an integer equal to $q(\#A/\#B)^n$. Let p'/q' be the irreducible fraction equal to $\#A/\#B$, then q'^n divides q for all $n \in \mathbb{N}$, so $q' = 1$.* ∎

Example 4.16: Consider the following input-deterministic transducer changing the 4-shift into the 2 shift (input labels do not matter at all provided $\mathcal{F}$ is input-deterministic, as there are 4 arcs starting from each vertex):



$A = \{0,1,2,3\}$; $B = \{0,1\}$; $C = \{a,b\}$; $X = A^{\mathbf{Z}}$, $X' = B^{\mathbf{Z}}$.

By Corollary 2.6, the unique measure μ' on X' related to the uniform measure λ on X is the image of the measure ν with maximal entropy on Y. Put

$$M(0) = \frac{1}{4} \cdot \begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix} \text{ and } M(1) = \frac{1}{4} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix},$$

where, for instance, $4M(0)_{i,j}$ is the number of arcs from j to i with output 0. Denoting by $\mathbf{1}$ the row 2-vector with coordinates 1, $\mathbf{V}$ the column vector with coordinates 1/3 and 2/3, $M = M(0) + M(1)$, with $M.\mathbf{V} = \mathbf{V}$ and $\mathbf{1}.M = \mathbf{1}$, by the classical results of [p1], ν is the Markov measure corresponding to distribution $\mathbf{V}$ on C and transition probability M, and one has

$$\mu'([u]) = \mathbf{1}.M(\kappa)\cdots M(\beta)M(\alpha).\mathbf{V} \quad \text{for } u = \alpha\beta\cdots\kappa \in B^*.$$

The measure μ' is uniform on $B^{\mathbf{Z}}$. To prove this, by Proposition 4.14, it is sufficient to check $\mu'([u]) = 1/4$ for $u \in B^2$. Putting $A = M(0) - M(1)$, this reduces to proving

$$\mathbf{1}.A.\mathbf{V} = \mathbf{1}.A^2.\mathbf{V} = 0,$$

which, using the characteristic polynomial of A, is done by merely showing $\mathbf{1}.A.\mathbf{V} = 0$.

References

[B] M.-P. Béal, *Codage, automates locaux et entropie,* Thèse. Publications du LITP, Université Paris 7, 1988.

[BerV] A. Bertrand-Mathis and B. Volkmann, *On (k, ϵ)-normal words in connecting dynamical systems,* Monats. Math. **107** (1989), 267–280.

[Bes] A. S. Besicovich, *The asymptotic distribution of the numerals in the decimal representation of the squares of the natural numbers,* Math. Zeitschrift **39** (1935), 146–147.

[BP] F. Blanchard and D. Perrin, *Relèvement d'une mesure ergodique par un codage,* Z. Wahrscheinlichkeitstheorie **54** (1980), 303–311.

[BrL] A. Broglio and P. Liardet, *Predictability,* preprint.

[C] D. G. Champernowne, *The construction of decimals normal in the scale of ten,* J. London Math. Soc, **8** (1933), 254–260.

[CE] A. Copeland and P. Erdős, *Note on normal numbers,* Bull. Amer. Math. Soc. **52** (1946), 852–860.

[DGS] M. Denker, C. Grillenberger and D. Sigmund, *Ergodic Theory on Compact Spaces,* Lecture Notes in Math. **527**, Springer-Verlag, Heidelberg (1975).

[DT] J.-M. Dumont and A. Thomas, *Une modification multiplicative des nombres g-normaux,* Ann. Fac. Sci. Toulouse **8** (1986—1987), 367–373.

[F] R. Fischer, *Sofic systems and graphs,* Monats. Math. **80** (1975), 179–186.

[G] F. R. Gantmacher, *Application of the Theory of Matrices,* Vol. 2, English translation, Interscience, Publishers, New York,, 1959.

[IT] S. Ito and Y. Takahashi, *Markov subshifts and realization of β-expansions,* J. Math. Soc. Japan **26** (1974), 33—55.

[K] T. Kamae, *Subsequences of normal sequences,* Israel J. Math. **16,2** (1973), 121–149.

[KW] T. Kamae and B. Weiss, *Normal numbers and selection rules,* Israel J. Math. **21** (1975), 101–110.

[Kn] D. Knuth, *The Art of Computer Programming,* Vol. 2, Addison-Wesley Reading, Ma,, 1969.

[KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences,* Wiley, New York, 1974.

[M] J. E. Maxfield, *Normal k-tuples,* Pacific J. Math. **3** (1953), 189–196.

[Me] M. Mendès France, *Opacity of an automaton, Application to the inhomogeneous Ising chain*. preprint.

[P1] W. parry, *Intrinsic Markov chains,* Trans. Amer. Math. Soc. **112** (1964), 55–65.

[P2] W. Parry, *Entropy and Generators in Ergodic Theory,* Benjamin New York, 1969.

[P3] W. Parry, *A finitary classification of topological Markov chains and sofic systems,* Bull. London Math. Soc, **9** (1977), 86–92.

[R1] G. Rauzy, *Nombres normaux et processus détrministes,* Acta. Arith. **29** (1976), 211–225.

[R2] G. Rauzy, *Propriétés statistiques des suites arithmétiques,* PUF, Paris, 1976.

[S1] R. G. Stoneham, *On the uniform ϵ-distribution of residues within the periods of rational fractions with applications to normal numbers,* Acta Arith. **22** (1973), 371–389.

[S2] R. G. Stoneham, *On (p,ϵ) normality in the rational fractions,* Acta Arith. **16** (1969), 239–254.