LECTURE 10

Entropy

V. Berthé

Laboratoire de Mathématiques discrètes,
CNRS-UPR 9016,
Case 930, 163 avenue de Luminy,
F-13288 Marseille Cedex 9,
France

1. Introduction

The purpose of this course is to introduce some notions of entropy: entropy in information theory, entropy of a curve and topological and measure-theoretic entropies. For these last two notions, we will consider in particular dynamical symbolical systems in order to present some measures of disorder for sequences. We will allude then to the problem of classification of sequences with respect to their spectral properties thanks to the entropy. For this purpose, we will introduce the sequence of block entropies for sequences taking their values in a finite alphabet: we will then compute explicitely the block frequencies (or in other words, the measure of the associated dynamical system) for some examples of automatic sequences (Prouhet-Thue-Morse, paperfolding and Rudin-Shapiro sequences) and for Sturmian sequences (these are the sequences with minimal complexity among all non-ultimately periodic sequences; in particular, we will consider some generalized Fibonacci sequences). But, in order to understand the intuitive meaning of the notions of topological and measure-theoretic entropies, we will begin by defining the Shannon entropy of an experiment.

2. Thermodynamical entropy

The concept of entropy was fist introduced in thermodynamics by Clausius in 1854. In statistical mechanics, the entropy of a system is equal to the logarithm of the number Ω of accessible microstates corresponding to a macroscopical state of this system:

$$S = k \ln \Omega,$$

where k is Boltzmann's constant. It is assumed here that all the microstates of an isolated system are equiprobable. Hence the probability of the macroscopical state increases with the number of corresponding microstates. Entropy is thus a measure of the probability for a macroscopical state to be realized. The Second Law of Thermodynamics, i.e. the increase of the entropy of an isolated system, means that the evolution of the system is towards the state of maximal probability, which is also the state corresponding to the maximal number of microstates, that is to say the state of maximal disorder.

Let us suppose now that the isolated system is made of N indentical molecules. Let us decribe then this system by counting the number N_i of molecules in each of k states in the phase space. There are $\Omega = \frac{N!}{N_1!N_2!\cdots N_k!}$ such ways to realize this distribution. It can easily be shown, by using Stirling's formula, that the entropy is thus equal, for large N, to:

$$S = -kN \sum p_i \log p_i, \text{ where } p_i = \frac{N_i}{N}.$$

This is the kind of formula we will meet in the definitions of entropy below.

3. Information theory

Entropy is known in physics as a measure of randomness or disorder. It can also be considered as a measure of information. Namely, consider an experiment, say, the roll of a die. Randomness and uncertainty have the same measure. We thus call entropy a measure of the uncertainty about the outcome of this experiment. But the amount of uncertainty before the roll corresponds to the amount of information one receives from the outcome of this roll. Therefore, entropy is defined in information theory as a measure of the *uncertainty* about the outcome of an experiment or equivently as a measure of the *information* yielded by the happening of this experiment.

A first measure of entropy was introduced by Hartley in 1928 ([18]): he considered distributions of equiprobable events. Wiener defined then, in 1948, the entropy of a single event ([42]). Finally, Shannon introduced, also in 1948, a measure of information for finite probability distributions. This is the notion we will deal with mostly, but we will begin by the Wiener notion of entropy which is the most intuitive way of measuring information. For introductions to the notion of entropy in information theory, see for instance [1], [32], [37].

3.1. Entropy of a single event

Following Wiener ([42]), consider now the entropy H_W of a single event A with probability $p(A) \neq 0$. For instance, the event A can be "having an ace", when you roll a die. The entropy is here a measure of the information we receive when the event A occurs (and also a measure of how unexpected the event A is). We suppose that the entropy H depends only on the probability $p(A)$ of the event. The function H should hence have the following properties: it must be positive (knowing something about an event is information) and additive (the information yielded by the occurrence of two independent events is the sum of the informations obtained from each experiment). We thus have:

1. $H_W(p) \geq 0$,
2. $H_W(pq) = H_W(p) + H_W(q)$.

It is easily shown that the only functions which satisfy Properties 1 and 2 are the functions defined by:

$$f(x) = -\lambda \log(x), \text{ for all } x \in]0,1], \text{ with } \lambda \geq 0.$$

These functions are defined up to a positive multiplicative constant. We will hence normalize the entropy by assigning the unit value to the equiprobable case, hence: $H_W(1/2) = 1$.

The quantity $H_W(p) = -\log_2(p)$ is hence the only function which satisfies all the required conditions and will be called the *Wiener entropy* of a single event.

Let us note that the logarithm appears here again, in quite a natural way.

3.2. Entropy of an experiment

Consider now the entropy of an experiment E with outcomes $A_1, A_2, .., A_n$ of probabilities $p_1, p_2, \cdots, p_n$, with $\sum p_i = 1$. For instance, E is a roll of a die with n faces and $p_1, p_2, \cdots, p_n$ are the probabilities of the different faces. We will suppose here again that entropy only depends upon the probabilities.

Shannon defines in [38] the entropy H_S of the experiment E as the average value of the entropies of the single events $A_1, A_2, .., A_n$, weighted according to their probabilities. Note that Shannon has obtained this definition axiomatically by deriving it from "intuitive" properties that a measure of information should have, like for the case of a single event.

Let

$$L(x) = \begin{cases} -x \log_2(x) & \text{if } x \in]0,1], \\ 0 & \text{for } x = 0. \end{cases}$$

We then have:

$$H_S(E) = H_S(p_1, p_2, \cdots, p_n) = -\sum_{k=1}^{n} p_k \log_2(p_k) = \sum_{k=1}^{n} L(p_k).$$

The function H_S is called the *Shannon entropy* of a finite probability distribution.

One can ask the reasons for the choice of base 2 for the logarithm. Such a normalization corresponds to the choice of the unit of information: the binary entropies will be measured in *bits* as the natural entropies measure information in *natural units* or *nats*.

Remark The first measure of entropy, introduced by Hartley in 1928 ([18]) was the following: the Hartley measure h of the entropy of an experiment with n outcomes is defined as $H_H(E) = \log n$. This entropy depends only on the number of events and not upon their probabilities. For instance, Boltzmann's formula corresponds to this conception. These two notions coincide if all the states are equiprobable.

Consider now the case $n = 2$ (for instance, flipping a coin with a given coin). We have

$$H_S(E) = H(p) = -p\log_2 p - (1-p)\log_2(1-p).$$

It is easily seen that the entropy of an experiment of outcomes of probabilities 0 and 1 is equal to 0. This seems rather natural because there is no uncertainty about the outcome in this situation. The entropy is maximum in the equiprobability case ($p = 1/2$), which corresponds to the case of maximal uncertainty. We will see in the next section that this result holds generally for all finite probability distributions.

3.3. Concavity of the function L

The function L is concave: for all x_k in $[0,1]$ and all $(\lambda_1, \cdots, \lambda_n)$ such that $\lambda_i \geq 0$ and $\sum \lambda_i = 1$, Jensen's inequality

$$L(\sum_{k=1}^{n} \lambda_k x_k) \geq \sum_{k=1}^{n} \lambda_k L(x_k)$$

holds. To check this, note that the function L has a negative second derivative.

In particular, we have by putting $x_k = p_k$ and $\lambda_k = 1/n$, for all k:

$$L(\sum_{k=1}^{n} \frac{p_k}{n}) \geq \sum_{k=1}^{n} \frac{1}{n} L(p_k).$$

Note that $\sum_{k=1}^{n} p_k = 1$. We thus have the following property:

$$H_S(p_1, \cdots, p_n) \leq \log_2 n.$$

This inequality means that the entropy is always smaller than the entropy of the equiprobability case. This agrees with the fact that the uncertainty about the outcome of an experiment is maximal when all the outcomes are equally probable.

3.4. Marginal and conditional entropy

We have measured here the information obtained with a unique experiment. Suppose now that we have two experiments E and F which are not independent: let $A_1, A_2, .., A_n$ and $B_1, B_2, .., B_m$ be the possible outcomes of the experiments E and F respectively, with probabilities $p_1, p_2, \cdots, p_n$ and $q_1, q_2, \cdots, q_m$.

Suppose that we know the result of the experiment E. One can then ask what the information yielded by F would be. It would be equal to

$$H_S(E, F) - H_S(E),$$

i.e. the information obtained by the event of both experiments $(H_S(E, F))$ minus the information yielded by the first $(H_S(E))$. Now, let q_{jk} be the conditional probability of B_k under the condition A_j. It is easy to verify that:

$$H_S(E, F) - H_S(E) = \sum_{j=1}^{n} p_j H_m(q_{j1}, \cdots, q_{jm}),$$

with $H_m(q_{j1}, q_{j2}, \cdots, q_{jm}) = \sum_{k=1}^{m} L(q_{jk})$. This quantity is called the *conditional entropy* of the experiment F with respect to the experiment E. We denote it by $H_c(F/E)$. We thus have:

$$H_c(F/E) = \sum_{j=1}^{n} p_j H_m(q_{j1}, \cdots, q_{jm}),$$

and we obtain the following *chain-rule*:

$$H_S(E, F) = H_S(E) + H_c(F/E).$$

The entropy of the experiment E is called *marginal entropy* in contrast with the *conditional entropy*. The chain-rule expresses that the joint entropy of two experiments equals the marginal entropy of the first one plus the conditional entropy of the second with respect to the first.

From the concavity of the function L, we deduce the following inequality, with equality if and only if the experiments are independent:

$$H_c(F/E) \leq H_S(F),$$

or in other words that *conditioning reduces entropy*. This seems rather natural because knowledge concerning the outcome of an experiment cannot increase the uncertainty in the outcome of another experiment.

We deduce from this that the entropy of a joint event is smaller than or equal to the sum of the individual marginal entropies; this last inequality is called *independence bound on entropy*:

$$H_S(E, F) \leq H_S(E) + H_S(F).$$

Note that we have equality here if and only if $H_c(F/E) = H_S(F)$, i.e. if E and F are independent.

3.5. Entropy of a finite curve

Let us see now how to apply these notions to curves and, in the next section, to sequences.

Mendès France defines in [4] the dimension of a curve. But he also associates a notion of entropy with the curves. He defines in [22], [23], [24] or [26], for instance, the entropy of a finite curve Γ_L as:

$$H(\Gamma_L) = -\sum_{n=1}^{+\infty} p_n \log p_n,$$

where p_n is the probability that an infinite straight line cuts Γ_L in exactly n points. Thus, $H(\Gamma_L)$ is the amount of information yielded by the experiment "an infinite straight line is drawn on the plane" and the outcomes are the numbers of intersection points. This entropy is hence a measure of the complexity of plane curves.

A natural question to ask is whether this measure takes finite values, i.e. whether there is an upper bound for the entropy. Thus, Mendès France has shown the following result:

Theorem 1 *We have:*

$$H(\Gamma_L) \leq \log(\frac{2L}{C_L}) + \frac{\beta}{e^\beta - 1},$$

where $\beta = \log \frac{2L}{2L - C_L}$, L is the length of Γ_L and C_L is the "perimeter" of the curve, or in other words, the length of the convex hull of Γ_L.

The function $\hat{H}(\Gamma_L) = \log(\frac{2L}{C_L}) + \frac{\beta}{e^\beta - 1}$ is called the *maximal entropy* of the curve and corresponds in a way to a topological entropy, a notion that we will discuss later on.

Note that Mendès France gives the following thermodynamical interpretation of the coefficient β: he defines the *temperature* of a curve as $T = \frac{1}{\beta}$. Thermodynamical quantities can be defined too, for example *pressure* or *volume* of a curve, which satisfy an equation of state and even an *Heisenberg uncertainty principle* (see for instance, [24] or [26]).

Define now, still following Mendès France, the entropy of an unbounded curve Γ:

$$H(\Gamma) = \liminf_{L \to +\infty} \frac{H(\Gamma_L)}{\log L},$$

where Γ_L is a finite portion of length L of the curve with the same origin. Notice the normalization obtained by dividing by $\log L$. We obtain then the information per unit of length. Similarly, one defines the maximal entropy as:

$$\hat{H}(\Gamma) = \liminf_{L \to +\infty} \frac{\hat{H}(\Gamma_L)}{\log L}.$$

We have:

$$0 \leq H(\Gamma) \leq \hat{H}(\Gamma) \leq 1.$$

An infinite straight line, an exponential spiral of equation $\rho = e^\theta$ are of zero entropy, i.e. *deterministic*. The spiral $\rho = \theta^\alpha$, where $\alpha > 0$, has entropy $\frac{1}{1+\alpha}$. The dragon curve (see [4]) has entropy equal to $\hat{H} = \frac{1}{2}$, which is the highest value for a self-avoiding curve. The spiral $\rho = \log(\theta + 1)$ is of entropy 1, i.e. *chaotic*.

This notion of entropy is to be connected with the dimension of a curve ([4]); namely, the dimension d of a curve satisfies for a large class of them: $d \geq \frac{1}{1-H}$, where H denotes the entropy. The meaning of this inequality is that the dimension increases with the entropy, or in other words, that the entropy and the dimension increase with the disorder of curves.

3.6. The sequence of block entropies

The purpose of this section is to introduce *the block entropies* for sequences with values in a finite alphabet.

Let us recall that the frequency $P(B)$ of a block B in a given sequence is defined as follows: it is the limit, if it exists, of the number of occurrences of this block in the first N letters of the sequence divided by N.

Let u be a sequence with elements with values in the alphabet $\mathcal{A} = \{1, \cdots, d\}$. We suppose that all the block frequencies exist for u. Let

$$P(x|x_1 \cdots x_n) = \frac{P(x_1 \cdots x_n x)}{P(x_1 \cdots x_n)},$$

where $x_1 \cdots x_n$ is a block of non-zero probability and x a letter. The intuitive meaning of $P(x|x_1 \cdots x_n)$ is that it is the conditional probability that the letter x follows the block $x_1 \cdots x_n$ in the sequence u. We are going to associate with the sequence u two sequences of block entropies $(H_n)_{n \in \mathbb{N}}$ and $(V_n)_{n \in \mathbb{N}}$.

Let E_n be the experiment "choosing a factor of length n of the sequence". The outcomes are the factors of length n with probabilities $P(x_1 \cdots x_n)$. Denote the entropy of E_n by V_n. We have, for all $n \geq 1$:

$$V_n = \sum L(P(x_1 \cdots x_n)),$$

where the sum is over all the factors of length n and with $L(x) = -x \log_d(x)$, for all $x \neq 0$ and $L(0) = 0$. Note that we change here the normalization by taking the base d logarithm.

Let F be the experiment "choosing a letter of the alphabet $\mathcal{A}$" and H_n be the conditional entropy of F with respect to E_n. We have:

$$H_n = H_c(F/E_n) = \sum{}' \, P(x_1 \cdots x_n) H(x_1 \cdots x_n), \tag{1}$$

where the sum is over all the blocks of length n of non-zero probability and

$$H(x_1 \cdots x_n) = \sum_{x \in \mathcal{A}} L(P(x/x_1 \cdots x_n)).$$

Thus, H_n is the entropy of the next symbol when the preceding $(n-1)$ letters are known, i.e. H_n measures the uncertainty about what the next symbol will be, if we are told the preceding letters.

Now, let us apply to these two sequences the theorems we have seen in information theory in order to deduce some basic properties for them. Let H_0 be the marginal entropy of F. We have:

$$H_0 = \sum_{x \in \mathcal{A}} L(P(x)).$$

Obviously, $H_0 \leq 1$. Thus, we obtain: $0 \leq H_n \leq H_0 \leq 1$. Furthermore we clearly have

$$H_n = V_{n+1} - V_n,$$

for all n, by putting $V_0 = 0$. This property corresponds to the chain-rule. Thus, $(H_n)_{n \in \mathbb{N}}$ is the discrete derivative of $(V_n)_{n \in \mathbb{N}}$. Note that $(V_n)_{n \in \mathbb{N}}$ is an increasing sequence, since $H_n \geq 0$, for all n.

It can be shown that $(H_n)_{n \in \mathbb{N}}$ is a monotonic decreasing sequence of n (see, for instance [11]). The intuitive meaning of this is that the uncertainty about the choice of the next symbol decreases when the number of known preceding symbols increases; in other words, conditional entropy decreases when the conditioning increases.

From the decreasing behaviour of the positive sequence $(H_n)_{n \in \mathbb{N}}$, we deduce the existence of the limit $\lim_{n \to +\infty} H_n$. We have $H_n = V_{n+1} - V_n$. By taking Cesàro means, i.e. by considering $(\sum_{k=0}^{n-1} H_k)/n$, we obtain:

$$\lim_{n \to +\infty} H_n = \lim_{n \to +\infty} \frac{V_n}{n}.$$

We will see below, that this limit corresponds to the *measure-theoretic entropy* of the dynamical system associated to the sequence u.

Finally, let us note that the sequence $(H_n)_{n \in \mathbb{N}}$ of conditional block entropies, measures in some way the properties of predictibility of the initial sequence u. Other measures of predictibility are to be found in [12], [13], for predictions with automata, in [35] for the notion of "noise" and also in a quite similar way, in [25] and [4], for the opacity of an automaton.

These sequences have been first introduced by Shannon in [38], who wanted to measure the entropy of the English language. Namely, consider a source emitting a sequence of letters like, for example, a telegraph. If the letters are independent and with the same probability, the entropy will be:

$$H = \log_2 26.$$

But if the source emits an English text, the letters do not come with the same frequency (the letter E occurs more frequently than Q) and the probability that a U, for instance, immediately follows a Q, is larger than the probability that an X comes after a Q. Therefore, Shannon has defined the entropy of n^{th} order of the English prose by putting:

$$V_n = -\sum_{C\in\mathcal{C}_n} P(C)\log_2(P(C)) = \sum_{C\in\mathcal{C}_n} L(P(C)),$$

where $\mathcal{C}_n$ is the family of all strings of n letters and $P(C)$ denotes the probability of the block C. Such a quantity corresponds to the entropy of an artificial language approximating the natural language in the following way: the only constraint that rules this language is that the probabilities (included the conditional ones) are the same up to order n as those of the natural language. Shannon gives examples of such approximations in [38].

Shannon estimates the entropy of the English language to be about 1. The difference with the case of independent letters with same probabilities ($H = \log_2 26$) comes from the *redundancy* of English, which is according to Shannon roughly about 50%: half what we say or write is determined by the structure of the language (see [38] for more details).

Burrows and Sulston have also studied these conditional block entropies sequences in 1991. One motivation could be to find a measure of disorder for sequences which would be convenient to distinguish between sequences according to their spectral properties (see [14]).

We will see in the next section some examples of computations of these conditional block entropies for some automatic sequences and for the Sturmian sequences.

4. Topological and measure-theoretic entropies

Let us introduce now two measures of disorder for sequences with values in a finite alphabet: the *topological entropy* is defined from the notion of complexity and the definition of *measure-theoretic entropy* uses the block frequencies. We will then consider the general case of a dynamical system and give the corresponding definitions.

4.1. Topological entropy of a sequence

The *complexity* of a sequence u is defined as the function $p(n)$ which counts the number of distinct blocks of length n which occur in this sequence (see [4] or [3]). It is a combinatorial notion. The *topological entropy* ([2]) is then the exponential growth rate of the complexity as the length increases:

$$H_{top}(u) = \lim_{n\to+\infty} \frac{\log_d(p(n))}{n}.$$

It is easy to check that this limit exists because of the subadditivity of the function $\log(p(n))$:

$$\log(p(n+m)) \leq \log(p(n)) + \log(p(m)).$$

We obviously have: $0 \leq H_{top}(u) \leq \frac{\log_d}{d} = 1$.

Consider now the entropy of a substitutive sequence (see [4]). The complexity satisfies: $p(n) \leq Cn^2$, for all n, C being some constant. Thus the entropy of a of substitutive sequence is zero. As a particular case, the entropy of a periodic sequence is also zero.

Consider now a Bernoulli sequence, i.e. a sequence corresponding to the Bernoulli scheme $(p_1, p_2, \cdots, p_d)$, with $p_i \neq 0$ and $\sum_{i=1}^{d} p_i = 1$: the frequencies of the letters are independent and given by the probabilites $(p_1, p_2, \cdots, p_d)$, as for instance, a sequence obtained by tossing a coin iteratively with a given coin. Thus, the number of factors of length n is equal to d^n and the entropy equals 1.

These examples show that the topological entropy cannot distinguish between periodic and substitutive sequences nor between Bernoulli sequences.

4.2. Measure-theoretic entropy of a sequence

We will hence put a metrical structure in order to make this measure of disorder more precise. We will therefore consider block frequencies for sequences.

Let u be a sequence with values in $\mathcal{A} = \{1, \cdots, d\}$ and whose frequencies exist for all blocks. We have seen that the sequences $(H_n)_{n\in\mathbb{N}}$ and $(\frac{V_n}{n})_{n\in\mathbb{N}}$ have the same limit $H(u)$. This limit is called the *measure-theoretic entropy* of the sequence u:

$$H(u) = \lim_{n\to+\infty} \frac{V_n}{n} = \lim_{n\to+\infty} H_n.$$

It is the limit of the entropy per symbol of the choice of a block of length n, when n tends to infinity.

Some examples of computation of the measure-theoretic entropy are to be find in the following.

4.3. Measure-theoretic entropy of a partition

Let us consider in this section the general case of a dynamical system and give the corresponding definitions of metrical and topological entropies. Let (X, T, μ) be a dynamical system where X is a metrical compact set, μ a probability measure and T a continuous invertible measure-preserving tranformation. For more details, the reader is referred, for instance, to [16], [34], or [41].

Let $P = \{P_1, \cdots, P_r\}$ be a finite partition of X. For $x \in X$, let k_i be the unique integer such that $T^i(x)$ belongs to P_{k_i}. Consider now the first n points of the orbit of x under T:

$$x, Tx, \cdots, T^{n-1}(x).$$

They belong successively to

$$P_{k_0}, P_{k_1}, \cdots P_{k_{n-1}}.$$

It is convenient to define the *name* of x by: $n(x) = (k_0, \cdots, k_{n-1}, \cdots)$. Thus, it is a way of coding the trajectory of x under T.

Consider now the new partition

$$P^n = \{P_{k_0} \cap P_{k_1} \cap \cdots \cap P_{k_{n-1}};\ \text{for } (k_0, \cdots, k_{n-1})$$

$$\text{being the first terms of any name of } x \in X\}.$$

Let us define the join of two partitions $A = \{A_1, \cdots, A_r\}$ and $B = \{B_1, \cdots, B_s\}$ by:

$$A \vee B = \{A_i \cap B_j; A_i \cap B_j \neq \emptyset\ \}.$$

Let us note that $A \vee B$ is also a partition. We have obviously

$$P^n = P \vee T^{-1}P \vee \cdots \vee T^{-(n-1)}P.$$

How to define the entropy of such a partition? We recall that μ is a measure of probability. A partition can be seen as an experiment. We can therefore consider the sets P_i of the partition P as the outcomes of probability $\mu(P_i)$ of the experiment "determining to which set of the partition a point x belongs". The atoms in P and $T^{-1}P$ have the same measures. Hence the partition $T^{-1}P$ corresponds to the replication of the experiment associated to P and the partition P^n corresponds to n replications. Naturally, the experiments P and $T^{-1}P$ are not necessarily independent.

Hence, let us define the entropy of the partition P by:

$$H(P) = \sum_{i=1}^{r} L(\mu(P_i)).$$

Thus, the quotient $\frac{H(P^n)}{n}$ is the amount of information per unit of time about the name of a point x of X, or in other words, the information per replication.

It can be shown easily that $H(P^n)$ is subadditive, so $\lim\limits_{n\to+\infty} \dfrac{H(P^n)}{n}$ exists. This limit, i.e.

$$H(P,T) = \lim_{n\to+\infty} \frac{H(P \vee T^{-1}P \vee \cdots \vee T^{-(n-1)}P)}{n}$$

is called the *entropy of the transformation T with respect to the partition P.*

But the choice of P can considerably reduce the uncertainty, for instance, if $P = \{X\}$. Therefore, the measure-theoretic entropy of T is defined to be the maximal uncertainty over all finite partitions P, i.e.:

$$H(T) = \sup_P H(P,T). \tag{2}$$

Thus, the entropy $H(T)$ measures the uncertainty about the way T moves the points of X.

Now, the question is how to compute the entropy of this system. Namely, the expression (2) is not easy to use. But, a classical result shows that the supremum in (2) is obtained for the partitions, which are *generators* with respect to T: this is the Kolmogorov-Sinai Theorem (see [20] and [39]); a generator is a partition such that if $x \neq y$, then $n(x) \neq n(y)$, or in other words, such that the codings under this partition, of the orbits of two distinct points are different. Furthermore, the Krieger Generator Theorem [21] asserts the existence of a finite generator if the transformation T is ergodic and if the entropy is finite.

We consider now the particular example of a symbolical dynamical system. For more details about what follows, the reader is referred to [33] and [34]. Let u be a sequence with values in $\mathcal{A} = \{1, \cdots, d\}$ and whose frequencies exist for all blocks. Let T be the *one-sided shift*, i.e. $(Tx)_n = x_{n+1}$. Let us suppose that u is recurrent, i.e. every factor of u appears infinitely often; thus T is onto on the orbit closure $\overline{Orb(u)}$ of u under T in $\mathcal{A}^N$, endowed with the product of discrete topologies. Furthermore, we suppose that T is injective (hence invertible) on a set of full measure: this is the case, for instance, if the complexity of u is sub-affine or if $p_u(n+1)-p_u(n)$ is bounded. Let μ be the unique T-invariant measure obtained by assigning to the *cylinders* their frequencies: $\mu([w]) = P(w)$. It is easily seen that the partition $P = \{[1], \cdots, [d]\}$ is a generator. Furthermore, we have $P^n = \{[B];\ B \text{ block of length } n\}$ and we check that $H(P^n) = V_n$.

4.4. Topological entropy of an open cover

Let us now consider the topological notion corresponding to a partition, namely open covers of the compact space X, in order to define the topological entropy of a dynamical system (see [2]). Let α be any open cover of X. Let us recall that an open cover of the space X is a collection of open sets $(O_i)_{i \in I}$ such that

$$X \subset \cup_{i \in I} O_i.$$

Let $N(\alpha)$ denote the number of sets in a finite subcover of α with smallest cardinality. We denote the join of two covers as :

$$\alpha \vee \beta = \{A \cap B;\ A \in \alpha,\ B \in \beta\}.$$

The topological entropy of T with respect to α is then defined as:

$$H_\alpha(P,T) = \lim_{n \to +\infty} \frac{\log(N(\alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-(n-1)}\alpha))}{n},$$

and the topological entropy of T is given by

$$H_{top}(T) = \sup_{\alpha} H(\alpha, T), \tag{3}$$

where the supremum is taken over all open covers.

4.5. Variational principle

Let us come back to the sequences of conditional block entropies $(H_n)_{n\in N}$ and $(V_n)_{n\in N}$. What relation exists between these two sequences?

We have, for all $k \geq 0$: $H_n = V_{n+1} - V_n$ and $V_0 = 0$. We deduce from this that $\sum_{k=0}^{n-1} H_k = V_n$. The sequence $(H_n)_{n\in N}$ is decreasing. Thus we obtain that:

$$nH_n \leq \sum_{k=0}^{n-1} H_k = V_n = \sum L(P(x_1 \cdots x_n)).$$

By concavity of the function L, we have, for all $n \geq 1$: $V_n \leq \log_d p(n)$. Thus, we deduce the following proposition:

Proposition 1 *We have $H_n \leq \frac{\log_d(p(n))}{n}$, for all $n \geq 1$.*

We hence have:

$$\lim_{n\to+\infty} H_n = \lim_{n\to+\infty} \frac{V_n}{n} = H(u) \leq H_{top}(u) = \lim_{n\to+\infty} \frac{\log_d(p(n))}{n}.$$

This inequality is a particular case of a basic relationship between topological entropy and measure-theoretic entropy called the *variational principle* (see, for a proof, [27]): if T is a continuous map of a compact metric space then

$$H_{top}(T) = \sup\{H(T);\ \text{for } \mu \text{ being any measure invariant under } T\}.$$

The two limits $\lim_{n\to+\infty} H_n$ and $\lim_{n\to+\infty} \frac{\log_d(p(n))}{n}$ are distinct, in general. But, for instance, if the system $(\overline{Orb(u)}, T)$ is uniquely ergodic (see [34]), we have naturally equality between these two limits.

Here is a case where these two limits are distinct. Consider a sequence corresponding to the Bernoulli scheme $(p, 1-p)$, with $p \neq 0$ and $p \neq \frac{1}{2}$. We recall that the topological entropy equals 1. We have, for every word B: $P(B) = p^{|B|_a}(1-p)^{|B|_b}$, where $|B|_x$ denotes the number of occurrences of the letter x in the word B. We thus have, for all $n \geq 1$:

$$\frac{V_n}{n} = L(p) + L(1-p) = -p\log_2 p - (1-p)\log_2(1-p).$$

Thus, we obtain, as expected:

$$\lim_{n\to+\infty} H_n = L(p) + L(1-p) < 1 = \lim_{n\to+\infty} \frac{\log_2(p(n))}{n}.$$

The notion of measure-theoretic entropy for a sequence seems consequently to

be more precise. But we will see in the next section that in the cases we deal with mostly, we consider *deterministic* sequences, i.e. sequences with zero entropy. Determinism means that there is no uncertainty in the choice of the next letter, in other words, *past determines future.* Therefore neither metrical nor topological entropy can distinguish between these sequences. That is why we will consider in the sequel the rate of convergence of the sequence H_n towards its limit (the measure-theoretic entropy) and not only this limit.

5. Entropy and spectral properties

Let us now consider the question of classifying dynamicals systems up to isomorphism. We will not give detailed definitions here but they can be found for instance in [34]. Let us recall that roughly, a dynamical system is said to be a *factor* of another dynamical system if the first one can be "constructed" in the second one, i.e. if there exists a map from the first one into the second one preserving measure and transformation.

How does entropy react with factorisation? Intuitively, one can see that there is a loss of information. In fact, it can be shown easily that entropy decreases with factorisation (see for instance [11]). If two dynamical systems are *metrically isomorphic*, then each of them is a factor of the other. Hence entropy is an isomorphism *invariant.* Two isomorphic dynamical systems are spectrally alike, i.e. they are *spectrally isomorphic* for the structure of Hilbert space, which can be associated to a dynamical system, as explained in [34]. Hence, a spectral invariant (like ergodicity or mixing properties) is an isomorphism invariant. A natural question then arises: can we find non-spectral invariants? The measure-theoretic entropy introduced by Kolmogorov to this effect in 1958 (see [20]) allows us to answer this question: we have seen that the entropy is an isomorphism invariant; so, two Bernoulli schemes with different entropies are not isomorphic but are always spectrally isomorphic. Hence a necessary condition for two Bernoulli schemes to be isomorphic is that they have the same entropy. In fact, the converse is true. This question remained open a long time but Ornstein solved it in 1970, by giving in [30] a complete classification of the Bernoulli shifts up to isomorphism:

Ornstein's theorem *Two Bernoulli schemes are isomorphic if and only if they have the same entropy.*

In the case of *discrete spectrum*, the classification is relatively easier: we have equivalence between metrical isomorphism and spectral isomorphism. Namely, Halmos and Von Neuman showed in 1942 that the eigenvalues allow us to say whether two ergodic transformations with discrete spectrum are isomorphic or not (see [17]).

Discrete spectrum classification Theorem *Two ergodic measure-preserving transformations with discrete spectrum are spectrally isomorphic if and only if they have the same eigenvalues. Furthermore, if they are spectrally isomorphic then they are isomorphic.*

In the case of discrete spectrum, the entropy is equal to 0. Thus the entropy is not convenient to distinguish between systems with low disorder.

Note that the entropy is in general equal to 0 in the examples we deal with mostly. We have namely the following property(see for instance, [31] or [32]):

Zero entropy *If the entropy of an invertible measure-preserving transformation T is strictly positive, then T has countable Lebesgue spectrum.*

Thus, we have the following corollary:

Corollary 1 *If T is of finite multiplicity or of continuous singular spectrum or of discrete spectrum, then the entropy of T is equal to 0.*

Therefore, either low complexity sequences (like substitutive sequences, for instance) or the examples of transformations given by M. Queffélec in [34], as irrational rotations of the circle, q-odometers, or the Chacon transformation, or also, Besicovitch almost-periodic and mean almost-periodic sequences ([4]) have zero entropy.

6. Some examples of computation of block entropies

We will see in this section that one can compute explicitely the block frequencies and consequently the block entropies $(H_n)_{n\in\mathbb{N}}$, defined in the section 2.6, for some examples of automatic sequences (Prouhet-Thue-Morse, paperfolding and Rudin-Shapiro sequences) and for Sturmian sequences: these are the sequences with minimal complexity among all non-ultimately periodic sequences (see for instance, [4] or [41]); in particular we will consider some generalized Fibonacci sequences. We will finally address the following question: given a sequence u, can we deduce from the rate of convergence of the sequence of conditional block entropies whether an atomic structure associated to u is "quasicrystalline" or not? Or, in other words, can we deduce from the block entropies, spectral properties of the initial sequence?

This question was put forward by Burrows and Sulston (1991) who have introduced this measure of disorder in the study of quasiperiodic structures. By computing the first and second order entropies H_1 and H_2 for the Prouhet-Thue-Morse sequence and for some generalizations of the Fibonacci sequence, they have obtained the following comparison of disorder: among the sequences they have studied, the sequences which are quasiperiodic (or of discrete spectrum) have entropy of first and second order lower than those which are not purely discrete. But these entropies H_1 and H_2 are not sufficient, for instance, to distinguish between the Rudin-Shapiro sequence and a *normal* sequence, i.e.

a sequence such that all blocks of same length have the same frequency. Thus it is interesting to obtain entropies of all orders and to compare them. Therefore we need to compute the block frequencies.

M. Queffélec has shown how to obtain the block frequencies of a substitutive minimal sequence by using the matrix of the associate primitive substitution and the Perron-Frobenius Theorem (see [34] and for more details, [33]). We will deduce here the block frequencies of all orders from a finite number of small length block frequencies, by using recurrence formulas between the frequency of a block and the frequencies of its pre-images by the substitution. But this method does not work for Sturmian sequences, because they are generally not substitutive. The idea here, due to Dekking [15], will be to use the Rauzy graph of words [36], which we define in the following. Dekking obtains in [15], a precise description of the frequencies of the words occurring in the Fibonacci sequence.

6.1. Ultimately periodic and "random" sequences

Consider first the following two extreme cases: the case of minimal disorder, i.e. the case of *ultimately periodic sequences*, or in other words, of sequences which are periodic from some index on, and the case of maximal disorder, i.e. the case of "random" sequences. Let us note that it is the same thing, in terms of frequencies, to consider ultimately periodic sequences and purely periodic sequences.

The following result can easily be shown (see [8]):

Proposition 2 *Let u be a ultimately periodic sequence of period Ω. We have: $H_k = 0$, for all $k \geq \Omega$.*

Namely, there is no uncertainty at all in the choice of the next letter. The converse is not true. Suppose, for instance, that the frequencies of the letters are equal to 0 or 1. We then have $H_0 = 0$. The sequence $(H_n)_{n \in \mathbb{N}}$ being a decreasing sequence, we obtain $H_n = 0$, for all n.

But if the sequence is minimal, we obtain the following property:

Proposition 3 *Let u be a minimal sequence such that $H_{k_0} = 0$ for some integer k_0. Then u is a periodic sequence of period $p(k_0)$, where $p(k_0)$ denotes the complexity of order k_0.*

The proof of this statement comes from the fact that the frequencies are strictly positive in a minimal sequence.

Consider now a "random" sequence or in other words, a *normal* sequence: all the blocks of given length have the same frequency. Hence the conditional probabilities $P(x/B)$ are equal and $H_n = 1$, for all $n \geq 0$. It can easily be shown, by using (1) that the converse is true. Thus, we have the following proposition:

Proposition 4 *We have $H_n = 1$ for all n, if and only if the sequence u is a normal sequence.*

Therefore, in these two extreme cases, the sequence $(H_n)_{n\in\mathbb{N}}$ gives a characterization of the ultimately periodic and the "random" sequences.

6.2. Sturmian sequences

The *Sturmian sequences* are the sequences with minimal complexity among all non-ultimately periodic sequences (see [4]). Thus, it is rather interesting to measure the disorder of such sequences. The Fibonacci sequence and the generalized Fibonacci sequences which are defined as the fixed points of the substitutions: $\sigma(a) = a^n b$ et $\sigma(b) = a$, with $n \geq 1$ are some examples of Sturmian sequences. Let us recall that a Sturmian sequence is the itinerary of the orbit of a point ρ of the unit circle under a rotation of irrational angle α, with respect to complementary intervals of length α and $1-\alpha$ of the unit circle (see [29] and [28]). We have the following result:

Theorem 2 *Let u be a Sturmian sequence with angle α. Let m be greater than 1. Let $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ be two successive m-Farey points such that: $\frac{p_1}{q_1} < \alpha < \frac{p_2}{q_2}$.*
The frequencies of the blocks of length m are the following:

$$p_2 - \alpha q_2,\ \ \alpha q_1 - p_1,\ \ \alpha(q_1 - q_2) + p_2 - p_1.$$

More precisely, there are

- *$m - q_2 + 1$ blocks of length $p_2 - \alpha q_2$,*
- *$m - q_1 + 1$ blocks of length $\alpha q_1 - p_1$*
- *and $(q_1 + q_2) - m - 1$ blocks of length $\alpha(q_1 - q_2) + p_2 - p_1$.*

Therefore the values of the conditional block entropies satisfy:

$$H_{m-1} = L(p_2 - \alpha q_2) + L(\alpha q_1 - p_1) - L(\alpha(q_1 - q_2) + p_2 - p_1),$$

with $L(x) = -x\log_2(x)$.

Let us recall the definition of an m-Farey point: it is an element of the interval $[0,1]$ of the form $\frac{p}{q}$, where $p \geq 0$, $1 \leq q \leq m$ and $gcd(p,q) = 1$. Two m-Farey points $\frac{p_1}{q_1}$ et $\frac{p_2}{q_2}$ are successive if and only if : $p_2q_1 - p_1q_2 = 1$. The points $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ depend on m and can be given explicitly, by using the continued fraction extension of the angle α.

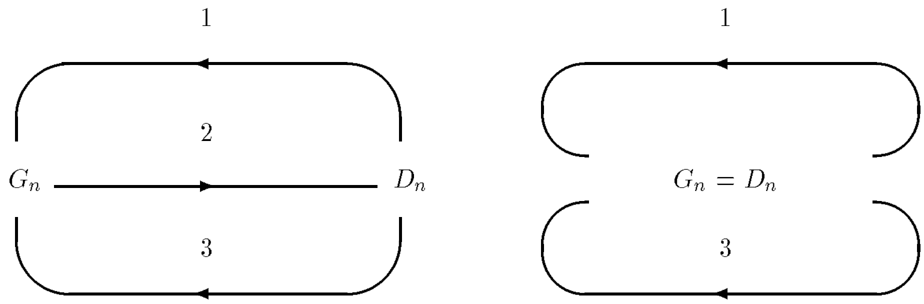
This theorem can be proved either by using the combinatorial definition of the Sturmian sequences ($p(n) = n+1$) or the dynamical characterization (Sturmian sequences are irrational rotations). It is easily seen that block frequencies correspond to intervals of the unit circle. This theorem is thus another formulation of the 3-distance Theorem (see, for instance [40]):

3-distance Theorem *Let α be any irrational number. Let us place the points $\{\alpha\}$, $\{2\alpha\}$, $\cdots, \{n\alpha\}$, where $\{\alpha\}$ denotes the integral part of α, on the line segment $[0,1]$. The $(n+1)$ segments found have at most three different lengths, one being the sum of the two others.*

The tool of the combinatorial proof here is the Rauzy graph of words. The Rauzy graph of words of length n associated to a sequence is the orientated graph whose nodes are the factors of length n of the sequence and whose arrows are defined as follows: there is an arrow from U to V, if there exists a word W such that

$U = xW$ and $V = Wy$, with $x, y \in \{a, b\}$ and xWy is a factor of the sequence.

Suppose now that we have a Sturmian sequence. From the complexity $(p(n) = n + 1)$, we deduce that there is only one factor of length n with two right extensions. Let us call this factor R_n. We define, similarly, L_n as the factor of length n with has two left extensions. Thus, the graph of words of given length has the two following forms, according as $L_n = R_n$ or $L_n \neq R_n$:



Let U be a node of the graph. Let U^+ denote the number of arrows wich come to U and U^- the number of arrows which go out U. We have the following lemma:

Lemma 1 *If $U^+ = 1$ and $V^- = 1$, then the words U and V have the same frequency.*

Namely, the factor U has only one right extension, that we denote x, and similarly, the factor V has only one left extension y. Therefore, we have the following equalities between the frequencies:

$$P(U) = P(Uy) = P(xWy) = P(xV) = P(V).$$

We deduce in particular from this lemma that all the words of the branch (1), except L_n and R_n, have the same frequency, that similarly, all the words of the branch (3), except L_n and R_n, have the same frequency and finally that all the words of the branch (2), L_n and R_n included, have the same frequency.

Thus we obtain that the frequencies of the factors of given length of a Sturmian sequence take at most three values.

More precisely, Theorem (2) can be shown by studying the lengths of the branches of the graph, whose evolution is precisely described in [7].

6.3. Block frequencies for some automatic sequences

Let us consider now some automatic sequences which are of distinct spectral types (see, for instance, [4] or [34]):

- the Prouhet-Thue-Morse sequence, which has continuous singular spectrum,
- the Rudin-Shapiro sequence, which has Lebesgue spectrum,
- the paperfolding sequence, which has discrete spectrum.

Let H^T, H^R and H^P be respectively the sequences of conditional block entropies for the Prouhet-Thue-Morse, the Rudin-Shapiro and the paperfolding sequences. We expect the following inequality between H^T, H^R and H^P:

$$H_n^P \leq H_n^T \leq H_n^R, \text{ for every } n,$$

or in other words, we expect, for instance, from the paperfolding sequence to show more order than the Prouhet-Thue-Morse sequence with respect to this particular measure of disorder.

Let us recall that these three sequences are deterministic, that is to say of zero entropy, so the sequence (H_n) converges towards 0 for each of theses sequences.

We have the following result (see [9] and also [15], for the Prouhet-Thue-Morse sequence):

Proposition 5 *The frequencies of blocks of length m, with $2^k+1 \leq m \leq 2^{k+1}$, take the following two values:*

- $\frac{1}{3.2^k}, \frac{1}{6.2^k}$, *for the Prouhet-Thue-Morse sequence, if $m \geq 2$,*
- $\frac{1}{8.2^k}, \frac{1}{16.2^k}$, *for the Rudin-Shapiro sequence, if $m \geq 8$,*
- $\frac{1}{4.2^k}, \frac{1}{8.2^k}$, *for the paperfolding sequence, if $m \geq 7$.*

We can deduce from this, the expression of the conditional block entropies:

Proposition 6 *Let H_n^T, H_n^R and H_n^P be respectively the conditional block entropies for the Prouhet-Thue-Morse, the Rudin-Shapiro and the paperfolding sequences. We have:*

- $H_n^T = \frac{4}{3.2^k}$, *for $2^k+1 \leq m \leq 3.2^{k-1}$,*
  $H_n^T = \frac{2}{3.2^k}$, *for $3.2^{k-1}+1 \leq m \leq 2^{k+1}$ and $k \geq 1$.*
- $H_n^R = \frac{1}{2^k}$, *for $2^k+1 \leq m \leq 2^{k+1}$ and $m \geq 8$,*

- $H_n^P = \frac{1}{2^k}$, *for* $2^k \leq m \leq 2^{k+1} - 1$ *and* $m \geq 7$.

We can notice the following relationship between H_n^R and H_n^P:

Proposition 7 *We have* $H_n^P = H_{n+1}^R$, *for all* n.

We will not give a detailed proof but the methods used here in the computation of the frequencies are similar to those used in the computation of the complexity (see [4]): we use for the Prouhet-Thue-Morse sequence the fact that each word of length greater than 4 comes from a unique word by the substitution (which is called the *pre-image*). For instance, the word aba comes from the words ab and ba but $baba$ comes only from the word bb. The substitution is here of length 2. Hence, the number of occurrences of a word (of length greater than 4) in the first $2n$ letters of the sequence is equal to the number of occurrences of its pre-image in the first n letters. We deduce from this that the frequency of a block is equal to half the frequency of its unique pre-image by the substitution. It is easy then to compute the frequencies by using this induction formula. The same situation occurs for the fixed point of the substitution which generates after a letter to letter projection, the Rudin-Shapiro sequence. Furthermore we have a bijection between the factors of length greater than 8 of the fixed point, and the factors of same length of the projection: namely, these two sequences have the same complexity function for n greater than 8 (see [4]). Hence, we can deduce using this bijection, the frequencies in the Rudin-Shapiro sequence from the frequencies in the fixed point. For the paperfolding sequence the same situation holds: we have a bijection between the factors of length greater than 7 of the fixed point and the factors of same length of the projection (see [4]). But there is a slight difference concerning the properties of "recognizability" of the fixed point: some words can arise from two different words by the substitution. In fact, these words which have more than one pre-image are exactly the *special* (or *expansive*) ones, that is to say the words which have at least two right extensions in the sequence. Despite this difficulty, it is possible to find here again recurrence formulas.

Remark This method works also for the generalized Rudin-Shapiro sequences which count the number of occurences of the pattern $1 \star \cdots \star 1$ in the binary expansion of every integer (see [5] and [6]). We obtain, if d is the length of the pattern $\star \cdots \star$, that the conditional block entropies are ultimately equal to 2^d times the corresponding entropies of the classical Rudin-Shapiro sequence.

6.4. Conclusion

Let us come back to the initial question of the comparison of block entropies for these sequences. It can be seen, by computing the first values of the conditional block entropies, that we have inequalities between the values of conditional block entropies of the same order for the first values. We have namely: $H^P \leq H^T \leq H^R$, for $n \leq 8$. But, for $n \geq 9$, this ordering does not hold. In particular, we have:

$$H_9^P = H_9^R = 1/8 < H_9^T = 1/6.$$

From Proposition 7, we deduce that $H_n^P \leq H_n^R$ and that for almost n, this inequality becomes an equality. More precisely,

- for $n = 2^k$, we have $H_n^P = \frac{H_n^R}{2}$,
- but for $n \neq 2^k$, we have $H_n^P = H_n^R$.

Furthermore, we see that there is a kind of shuffle between the values of H^R (and consequently of H^P) and the values of H^T:

- for $2^k + 1 \leq n \leq 3.2^{k-1}$, we have: $H_n^T = \frac{4}{3} H_n^R = \frac{4}{3} H_n^P$,
- and for $3.2^{k-1} + 1 \leq n < 2^{k+1}$, we have: $H_n^T = \frac{2}{3} H_n^R = \frac{2}{3} H_n^P$.

In particular, these three sequences of conditional block entropies converge with the same rate towards 0.

We conclude from this that this measure of disorder cannot allow us to distinguish between deterministic sequences even if they have different spectral properties.

Acknowledgments

We would like to thank J.-P. Allouche for his many useful comments and Michel Mendès France who read carefully a previous version of this paper.

References

[1] Aczél J. and Daróczy Z., On measures of information and their characterizations (Mathematics in science and engineering, Academic Press **115**, 1975).

[2] Adler R. L., Konheim A. G. and McAndrew M. H., Topological entropy, *Trans. Amer. Math. Soc.* **114** (1965) 309–319.

[3] Allouche J.-P., Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1** (1994) 133–143.

[4] Allouche J.-P. and Mendès France M., this volume.

[5] Allouche J.-P. and Liardet P., Generalized Rudin-Shapiro sequences, *Acta Arith.* **60** (1991) 1–27.

[6] Allouche J.-P. and Shallit J., Complexité des suites de Rudin-Shapiro généralisées, *Journal de Théorie des Nombres de Bordeaux* **5** (1993) 283–302.

[7] Arnoux P. and Rauzy G., Représentation géométrique de suites de complexité $2n + 1$, *Bull. Soc. math. France* **119** (1991) 199–215.

[8] Berthé V., Fonctions de Carlitz et automates. Entropies conditionnelles (Thèse, Univ. Bordeaux I, 1994).
[9] Berthé V., Conditional entropy of some automatic sequences, *J. Phys. A: Math. Gen.* **27** (1994) 7993–8006.
[10] Berthé V., Fréquences des facteurs des suites sturmiennes, *Preprint 94.37 Laboratoire de Mathématiques discrètes.*
[11] Billingsley P., Ergodic theory and information (John Wiley and Sons, New York, 1965).
[12] Broglio A., Prédiction par automates (Thèse, Aix-Marseille I, 1991).
[13] Broglio A. and Liardet P., Predictions with automata, *Contemporary Mathematics, Amer. Math. Soc.* **135** (1992).
[14] Burrows B. L. and Sulston K. W., Measures of disorder in non-periodic sequences, *J. Phys. A: Math. Gen.* **24** (1991) 3979–3987.
[15] Dekking F. M., On the Prouhet-Thue-Morse Measure, *Acta Universitatis Carolinae, Mathematica et Physica* **33** (1992) 35–40.
[16] Dekking F. M., this volume.
[17] Halmos P. R. and Von Neumann J., Operator methods in classical mechanics II, *Ann. Math.* **43** (1942) 332–350.
[18] Hartley R. V., Transmission of Information, *Bell System Tech. J.* **7** (1928) 535–563.
[19] Keane M., "Ergodic theory and subshifts of finte type", Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces, T. Bedford, M. Keane, C. Series (Eds) (Oxford University Press, 1991) pp. 35–70.
[20] Kolmogorov A.N., New metric invariants of transitive dynamical systems and automorphisms of Lebesgue spaces, *Dokl. Akad. Nauk SSSR* **119** (1958) 861–864.
[21] Krieger W., On entropy and generators of measure-preserving transformations, *Trans. Amer. Math. Soc.* **149** (1970) 453–464, Erratum: **168** (1972) 519.
[22] Mendès France M., "Chaotic curves", Rythms in Biology and Other Fields of Applications, Proc. Journ. Soc. Math. France, Luminy 1981 (Lecture Notes in Biomathematics **49**, Springer Verlag, 1983) pp.352–367.
[23] Mendès France M., "Dimension and entropy of regular curves", Fractals, non-integral dimensions and applications, G. Cherbit Ed. (John Wiley and Sons, 1990) pp. 222–230.
[24] Mendès France M., Folding paper and thermodynamics, *Phys. Rep.* **103** (1984) 161–172.
[25] Mendès France M., Opacity of an automaton. Application to the inhomogeneous Ising chain, *Comm. Math. Phys.* **139** (1991) 341–352.
[26] Mendès France M., "The Planck constant of a curve", Fractal geometry and Analysis, J. Bélair and S. Dubucs Eds. (1991) pp. 325–366.
[27] Misiurewicz M., " A short proof of the variational principle for a $\mathbb{Z}_+^n$ action on a compact space", Int. Conf. Dyn. Systems in Math. Physics, *Société Mathématique de France, Astérisque* **40** (1976) 147–158.
[28] Morse M. and Hedlund G. A., Symbolic dynamics, *Amer. J. Math.* **60** (1938) 815–866.

[29] Morse M. and Hedlund G. A., Symbolic dynamics II: Sturmian trajectories, *Amer. J. Math.* **62** (1940) 1–42.
[30] Ornstein D. S., Bernoulli shifts with the same entropy are isomorphic, *Adv. in Math.* **4** (1970) 337–352.
[31] Parry W., Topics in ergodic theory (Cambridge University Press, 1981).
[32] Petersen K., Ergodic theory (Cambridge studies in advanced mathematics; 2, 1991).
[33] Queffélec M., Substitution dynamical systems. Spectral analysis (Lecture Notes in Mathematics **1294**, Springer-Verlag, 1987).
[34] Queffélec M., this volume.
[35] Rauzy G., Nombres normaux et processus déterministes, *Acta Arithmetica* **14** (1976) 211–225.
[36] Rauzy G., Suites à termes dans un alphabet fini, *Sém. de Théorie des Nombres de Bordeaux* (1983) 25-01–25-16.
[37] Roman S., Coding and information theory (Springer-Verlag, 1992).
[38] Shannon C. E., A mathematical theory of communication, *Bell System Tech. J.* **27** (1948) 379–423, 623–656.
[39] Sinai Ya. S., Flows with finite entropy, *Dokl. Akad. Nauk SSSR* **125** (1959) 1200–1202.
[40] Slater N. B., Gaps and steps for the sequence $n\theta$ mod 1, *Proc. Cambridge Phil. Soc.* **63** (1967) 1115–1123.
[41] A. A. SütŐ, this volume.
[42] Wiener N. Cybernetics, or Control and Communication in the Animal and the Machine (Hermann, Paris; Technology Press MIT, Cambridge, Massachussets; Wiley, New York, 1948).