Conditional entropy of some automatic sequences

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

Abstract: Burrows and Sulston have introduced conditional block entropies H_n from information theory in order to give a quantitative measure of disorder for sequences and, if possible, a characterization of quasi-crystalline sequences. We give here some properties concerning these block entropies and give an explicit formula for the sequences $(H_n)_{n\in\mathbb{N}}$ corresponding to the Thue-Morse sequence, to the Rudin-Shapiro sequence and to the paperfolding sequence. We deduce from these computations that this measure of disorder cannot allow us to distinguish between deterministic sequences even if they have different spectral properties.

1 Introduction

Burrows and Sulston have introduced a measure from information theory in [9]. Their purpose was to give a quantitative measure of disorder for sequences and to find a characterization of quasi-periodic sequences, or in other words, of unidimensional models of quasi-crystalline atomical structures.

This measure corresponds to a sequence of conditional block entropies H_n which is associated with a sequence u with values in a finite alphabet: the sequence (H_n) converges towards the metrical entropy of the dynamical symbolical system associated to the initial sequence u and the values H_n are defined in terms of conditional frequencies. More precisely, the conditional entropy H_n is a measure of the uncertainty about the next symbol, when the preceding letters are known. Thus it measures in some sense the properties of predictability of the initial sequence u.

By computing the first and second order entropies H_1 and H_2 for the Thue-Morse sequence and for some generalizations of the Fibonacci sequence, Burrows and Sulston have obtained the following comparison of their "disorder": among the sequences they study, the sequences which are quasi-periodic (or more generally, of discrete spectrum) have entropy of first and second order lower than those which have continuous component in their spectrum.

But it is easily seen that these entropies H_1 and H_2 are not sufficient to distinguish, for instance, between the Rudin-Shapiro sequence and a normal sequence, i.e. a sequence such that all blocks of the same length have the same frequency. Thus, it is interesting to obtain entropies of all orders and to compare them.

The aim of this paper is to compute and to compare the conditional block entropies of all orders for some automatic sequences which have distinct spectral types. The sequences we study here are the Thue-Morse sequence, which has continuous singular spectrum, the Rudin-Shapiro sequence and some generalizations, which have Lebesgue spectrum and the paperfolding sequence, which has discrete spectrum.

The Thue-Morse sequence $(v_n^T)_{n\in\mathbb{N}}$ gives the parity of the sum of the binary digits of the integers: if $n = \sum \varepsilon_i 2^i$, where $\varepsilon_i = 0$ or 1, then $v_n^T = \sum \varepsilon_i \ mod\ 2$. The Rudin-Shapiro sequence $(v_n^R)_{n\in\mathbb{N}}$ counts, also modulo 2, the number of occurrences of 11 in the binary representation of the integers, with overlaps: $v_n^R = \sum \varepsilon_i \varepsilon_{i+1} \ mod\ 2$. The paperfolding sequence is obtained the following way: let us fold a sheet of paper always the same way, for instance, right half over left. Let us code the "valleys" and the "mountains" that we see when the sheet is being unfolded. This process gives rise to the the paperfolding sequence (see for instance, [12]).

These sequences are automatic sequences, so we will use here the underlying substitution in order to compute the block frequencies, as in [9].

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:

$$H_n^P \leq H_n^T \leq H_n^R, \text{ for every } n, \tag{1}$$

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

But in fact, after computation of the conditional block entropies of all orders for these particular automatic sequences, we notice that there is no ordering between H^R, H^P and H^T, such as (1). In particular, we prove that these three sequences of conditional block entropies converge with the same rate towards 0. Namely, let us recall that these three automatic sequences are deterministic, that is to say of zero entropy, so the sequence (H_n) converges towards 0 for each of theses sequences.

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

We compute also in [5] or [6] block frequencies and give an explicit formula of conditional entropies H_n for Sturmian sequences. A Sturmian sequence has exactly $n+1$ factors of length n. In particular, the Fibonacci sequence is Sturmian. The Sturmian sequences are generally not substitutive, hence the method used is different: we can compute the block frequencies, either by studying the Rauzy graph of factors [15] or by considering Sturmian sequences as rotations. Namely, a Sturmian sequence is the itinerary of the orbit of a point of the unit circle under a rotation of irrational angle α, with respect to disjoint intervals of the unit circle of length α and $1-\alpha$.

2 The sequence of block entropies

The purpose of this section is to introduce the block entropies for sequences with values in a finite alphabet. These entropies have been first introduced by Shannon [16] in 1948 in information theory; he wanted in particular, to give a measure of the entropy of the English language.

The frequency $P(B)$ of a block B 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 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 to the sequence u two sequences of block entropies $(H_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$. Let us first recall that the entropy is defined in information theory as a measure both of the information yielded by the happening of an experiment and of the uncertainty about the outcome of an experiment.

We call factor of the infinite sequence u a finite block of consecutive letters, say $w = u_{n+1} \cdots u_{n+d}$; d is called the length of w and is denoted by $l(w)$. Note that this is the European terminology and that a subword consists of non-necessarily consecutive letters. In the American terminology, the terms subword and factors are synonymous.

The entropy V_n is defined as the entropy of the choice of a factor of length n of the sequence. We thus put, 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 $L(0) = 0$.

Now, let H_n be the conditional entropy of the choice of the next symbol when we know the $(n-1)$ preceding symbols. We have:

$$H_n = H_c(F/E_n) = \sum{}' P(x_1 \cdots x_n) H(x_1 \cdots x_n), \tag{2}$$

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

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

Let H_0 be the entropy of the choice of a letter:

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

Obviously, we have $H_0 \leq 1$. From the concavity of the function L, we deduce that: $0 \leq H_n \leq H_0 \leq 1$.

Furthermore, we clearly have

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

for all n, by putting $V_0 = 0$. This equality means that the conditional entropy of the choice of the next letter, when the n preceding letters are known is equal to the entropy of the choice of a factor of length $(n+1)$ minus the entropy of the choice of a factor of length n. This is a classical result in information theory (see for instance [16]).

Thus, H is the discrete derivative of V. Note that $(V_n)_{n\in\mathbb{N}}$ is an increasing sequence, since $H_n \geq 0$, for all n.

It is easily shown that $(H_n)_{n\in\mathbb{N}}$ is a monotonic decreasing sequence of n (see for instance [7]). 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. We deduce from this the existence of the limit $\lim_{n\to+\infty} H_n$. We have $V_n = H_{n+1} - H_n$. Thus, by taking Cesàro means, we obtain: $\lim_{n\to+\infty} H_n = \lim_{n\to+\infty} \frac{V_n}{n}$. In fact, this limit which we denote by $H(u)$, is equal to the measure-theoretic entropy $h_\mu(T)$ of the one-sided shift T on $\overline{\mathcal{O}(u)}$, with respect to the measure μ, where $\overline{\mathcal{O}(u)}$ is the orbit closure of u in $\mathcal{A}^{\mathbf{N}}$ and μ is the T-invariant measure defined by assigning to each cylinder the frequency of the defining factor. For more details, the reader is referred to [14], for instance.

Another consequence of the decreasing behaviour of $(H_n)_{n\in\mathbf{N}}$ is the following inequality:

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

Let $p(n)$ denote the complexity of the sequence, i.e. the function which counts the number of factors of a sequence of given length. 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 inequality:

$$H_n \leq \frac{\log_d(p(n))}{n},$$

for all $n \geq 1$. The limit $H_{top}(u)$ of the sequence $\frac{\log_d(p(n))}{n}$, which is easily seen to exist, is called topological entropy.

In particular, we 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.

The notion of metrical entropy for a sequence seems consequently to be more precise. But in the cases we deal with here, we consider deterministic sequences, i.e. sequences with zero entropy. Thus neither metrical nor topological entropy can distinguish between these sequences. That is why it is interesting to consider the rate of convergence of the sequence H_n towards its limit (the metrical entropy) and not only this limit.

3 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 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.

Proposition 1 *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. Then, we 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, i.e. if all its factors appear infinitely often and with bounded gaps, we obtain the following property.

Proposition 2 *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 (2) that the converse is true. Thus, we have the following proposition.

Proposition 3 *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 of the "random" sequences.

4 Substitutions

Now, let us recall some definitions about substitutive sequences.

4.1 Some definitions

We consider here fixed points of substitutions.

For instance, the Thue-Morse sequence is the fixed point (i.e. the infinite iteration $(\sigma^T)^\infty(0)$) of the substitution σ^T, defined on the alphabet $\{0,1\}$ by:

$$\sigma^T(0) = 01 \text{ and } \sigma^T(1) = 10;$$

its first terms are:

$$0110100110010110\cdots$$

Similarly, the Fibonacci sequence is the fixed point of the substitution:

$$\sigma^F(0) = 01 \text{ and } \sigma^F(1) = 0.$$

A substitution is called uniform or of constant length if all the images of the letters have the same length. For instance, the Thue-Morse substitution is uniform whereas the Fibonacci substitution is not of constant length.

A sequence is called automatic if it is the image by a letter to letter projection of the fixed point of a substitution of constant length. The word automatic comes from the fact that an automatic sequence is generated by a finite automaton. For more details, the reader is referred to [10].

For instance, the paperfolding and the Rudin-Shapiro sequences are automatic sequences. Namely, the Rudin-Shapiro sequence is the image of the fixed point $(\sigma^R)^\infty(a)$ of the substitution σ^R:

$$\begin{cases} \sigma^R(a) = ab \\ \sigma^R(b) = ac \\ \sigma^R(c) = db \\ \sigma^R(d) = dc \end{cases}$$

by the projection

$$\begin{cases} p^R(a) = p^R(b) = 0 \\ p^R(c) = p^R(d) = 1. \end{cases}$$

Its first terms are

$$0001001000011101...$$

We give the substitution and the projection corresponding to the paperfolding sequence in the section [7].

4.2 Special factors

We have different means to compute the conditional entropies H_n. By using Formula (2), we see that we need to know all the block frequencies of factors of given length and also conditional probabilities. We can also deduce H_n from V_{n+1} and V_n. But, we need therefore to know all the block frequencies of factors of length n and $n+1$. Hence we will see in this section a more "economical" way of computing H_n.

A factor is called right special if it has at least two right extensions in the sequence. For instance, it is easily seen that the factor 010 is a right special factor of the Thue-Morse sequence (the factors 0100 and 0101 are factors of this sequence), whereas the factor 011 is always followed by the letter 0. Similarly, a factor with more than one left extension is called a left special factor.

Let us notice here that the extension of a factor B denotes usually a factor Bx, where x is a letter which follows the block B in the sequence. But we call from now extension, by abuse of notation, the letter x itself.

It is now quite easy to compute the conditional block entropies from the frequencies of the right special factors and of their right extensions. We have namely the following lemma.

Lemma 1 *Let $\mathcal{S}_n$ be the set of right special factors of length n. We have:*

$$H_n = \sum_{B \in \mathcal{S}_n} [\sum_{x \in \mathcal{A}} L(P(Bx)) - L(P(B))].$$

Namely, if v is not a right special factor then only one of the probabilities $P(vx)$ is non-zero (and is thus equal to $P(v)$) and we have therefore: $\sum_{x \in \mathcal{A}} L(P(vx)) = L(P(vx)) = L(P(v))$.

4.3 Frequencies

M. Queffélec gives in [14] an algorithm to compute the block frequencies of all orders of a substitutive minimal sequence by using the matrix of the associate primitive substitution and the Perron-Frobenius Theorem.

Let us recall that the matrix associated to a substitution is the matrix whose entry (i, j) is the number of occurrences of the letter i in the factor $\sigma(j)$. A substitution σ is called primitive when its matrix M is primitive (a matrix is primitive if there exists an integer k such that M^k has strictly positive entries). In other words, this property means that there is an integer

k such that the image by σ^k of every letter contains all the other letters of the alphabet on which the substitution is defined.

The idea here is to give recurrence relations between the frequency of a factor and the frequencies of factors of shorter length.

From the Perron-Frobenius Theorem, it is easy to deduce that the letter frequencies of a fixed point of the substitution σ are the coordinates of the unique normalized right eigenvector associated to the largest eigenvalue of the matrix of the substitution.

In the examples we deal with here, we have a nice property of "recognizability". Namely, there is a unique unique way of cutting enough long factors: we put bars such that between two bars there is exactly the image by the substitution σ of a letter of the alphabet. For instance, in the Thue-Morse sequence, we can cut uniquely the factor 01001 as follows:

$$0|10|01| = 0|\sigma^T(1)|\sigma^T(0).$$

Although the "short" factor 010 can be cut as $01|0 = \sigma^T(0)|0$ or as $0|10 = 0\sigma^T(1)$.

4.4 Some lemmas

In what follows, the substitution σ will always denote a substitution of length 2. A preimage of a word B is a factor of smallest length such that its image by the substitution contains B.

We have the following obvious relationship between the length of a factor and the length of its preimage by σ.

Lemma 2 *If B is a factor of even length, with $l(B) = 2p$, then its preimages are of length $p+1$ or p.*

If B is a factor of odd length, with $l(B) = 2p+1$, then its preimages are of length $p+1$.

The number of occurrences of a factor in the first $2n$ letters of the sequence is equal to the number of occurrences of its preimages in the first n letters. We deduce from this remark that, if a factor has a unique preimage, the frequency of a block is equal to half the frequency of its unique preimage by the substitution. Thus, we have the following lemma.

Lemma 3 *Let B be a factor with a unique preimage B'. The frequencies of B and B' satisfy: $P(B) = \frac{P(B')}{2}$.*

Let us consider now more particularly right special factors with the lemma below.

Lemma 4 *Let B be a factor with a unique preimage B'. If the factor B is right special, then B' is also a right special factor and B is a suffix of $\sigma(B')$. In particular, if B has even length $2p$, then B' has length p.*

Proof Let us write $B = x\sigma(B')y$, where x (respectively y) is either the empty word or a letter. If y were not the empty word, then B would have as unique right extension the second letter of $\sigma(y)$. Hence, B is a suffix of $\sigma(B')$. Furthermore, if B' had a unique right extension, then B would have as unique right extension the first letter of the image of the unique extension of B'. Thus, B' is a right special factor.

5 The Thue-Morse sequence

Let us recall that the Thue-Morse sequence is the fixed point $(\sigma^T)^\infty(0)$ of the substitution: $\sigma^T(0) = 01$ and $\sigma^T(1) = 10$.

First of all, the main property, which makes everything work here, is the following one, which can easily be shown (see for instance [14]).

Lemma 5 *Each factor of length greater than 4 has a unique preimage.*

M. Dekking has shown in [11] the following result:

Theorem 1 *The frequencies of the factors of the Thue-Morse sequence of length n, with $2^k + 1 \leq n \leq 2^{k+1}$ and $n \geq 2$, take the following two values:*

$$\frac{1}{3.2^k}, \frac{1}{6.2^k}.$$

The factors of length 1 have frequency $1/2$.

More precisely, M. Dekking deduces also from the complexity function of the Thue-Morse sequence the number of blocks of given length having each of these frequencies.

We deduce from this theorem the following result.

Lemma 6 *Let B be a right special factor of length greater than 2. Let k be such that $2^k + 1 \leq l(B) \leq 2^{k+1}$. Then, B has frequency $\frac{1}{3.2^k}$ and its right extensions have frequency $\frac{1}{6.2^k}$.*

Let us note that the frequencies of the right special factors take the greatest value between the two possible ones. This seems rather natural because the special factors appear more often, because of their two possible extensions.

We prove this lemma by induction for factors of length 2^k: namely, if $l(B) = 2^{k+1}$, then its preimage B' has length 2^k (Lemma 2 and Lemma 4).

For the other factors, this lemma is a direct consequence of Theorem 1: the sum of the frequencies of the two extensions of a right special factor B is equal to the frequency of B.

The complexity of the Thue-Morse sequence satisfies the following property (see for instance [8] and [13]).

Theorem 2 *We have $p(1) = 2$, $p(2) = 4$ and $p(3) = 6$. For the following values, we have, for $k \geq 1$:*

- *if $2^k + 1 \leq m \leq 3.2^{k-1}$, then $p(n+1) - p(n) = 4$,*
- *if $3.2^{k-1} + 1 \leq m \leq 2^{k+1}$, then $p(n+1) - p(n) = 2$.*

Hence, as an immediate consequence of Lemma 6 and Theorem 2, we obtain the following expression for the conditional block entropies:

Theorem 3 *We have $H_0^T = 1$, $H_1^T = \log_2 3 - 2/3$ and $H_2^T = 2/3$. For the following values, we have, for $k \geq 1$:*

- *if $2^k + 1 \leq m \leq 3.2^{k-1}$, then $H_n^T = \frac{4}{3.2^k}$,*
- *if $3.2^{k-1} + 1 \leq m \leq 2^{k+1}$, then $H_n^T = \frac{2}{3.2^k}$.*

The first values are computed "by hand ", as in [9]. Next, let us recall that the conditional block entropies are given by:

$$H_n = \sum_{B \in \mathcal{S}_n} L(P(B0)) + L(P(B1)) - L(P(B)),$$

where $\mathcal{S}_n$ is the set of right special factors of length n and $L(x) = -x \log_2(x)$ (Lemma 1). Furthermore, the cardinal of $\mathcal{S}_n$ is given by $p(n+1) - p(n)$.

6 The Rudin-Shapiro sequence

The Rudin-Shapiro sequence v^R is the image by a projection of the fixed point u^R of a substitution of length 2. Namely, it is the image of the fixed point $u^R = (\sigma^R)^\infty(a)$ of the substitution σ^R:

$$\begin{cases} \sigma^R(a) = ab \\ \sigma^R(b) = ac \\ \sigma^R(c) = db \\ \sigma^R(d) = dc \end{cases}$$

by the projection

$$\begin{cases} p^P(a) = p^P(b) = 0 \\ p^P(c) = p^P(d) = 1. \end{cases}$$

For the fixed point u^R everything works like in the case of the Thue-Morse sequence. In particular, it is easily seen that each factor of length greater than 3 has a unique preimage.

But, the Rudin-Shapiro sequence v^R is obtained as the image of u^R by a letter to letter projection. Some natural questions arise then: does a factor have a unique antecedent by the projection? Is the image (respectively, the antecedent) of a special factor also a special one? The answer to these questions is no. Namely, let us consider for instance the factor 000100 of the sequence u^R. This factor comes from the two factors *abacab* (which has as unique extension in v^R the letter d) and *babdba* (which has as unique extension in v^R the letter b). Thus, this right special factor has two distinct antecedents which are not special.

But this kind of "perturbation" due to the projection, appears only for small length factors. Namely, we have the following property (see for instance [4]).

Theorem 4 *For all $n \geq 8$, $p_{u^R}(n) = p_{v^R}(n) = 8n - 8$. In particular, there is a bijection between the factors of the two sequences of length greater than 8.*

For the first values, we have : $p_{u^R}(1) = 4$, $p_{u^R}(2) = 8$, $p_{u^R}(n) = 8n - 8$ for $3 \leq n \leq 7$ and $p_{v^R}(1) = 2$, $p_{v^R}(2) = 4$, $p_{v^R}(3) = 8$, $p_{v^R}(4) = 16$, $p_{v^R}(5) = 24$, $p_{v^R}(6) = 36$ and $p_{v^R}(7) = 46$.

Therefore, a right special factor corresponds to a right special factor by this bijection and two factors in bijection have the same frequency. Thus, we can show the following result.

Theorem 5 *The frequencies of the factors of the Rudin-Shapiro sequence of length n, with $2^k + 1 \le n \le 2^{k+1}$ and $n \ge 7$, take the following two values:*

$$\frac{1}{8.2^k}, \frac{1}{16.2^k}.$$

The blocks of length 1 have frequency $1/2$, the blocks of length 2 have frequency $1/4$, the blocks of length 3 have frequency $1/8$, the blocks of length 4 have frequency $1/32$ or $3/32$, the blocks of length 5 have frequency $1/32$ or $1/16$, the blocks of length 6 have frequency $1/64$, $1/32$ or $3/64$.

Proof The corresponding property for the fixed point u^R is the following one: the blocks of length 1 have frequency 1/4 and the frequencies of the factors of u^R of length n, with $2^k + 1 \le n \le 2^{k+1}$ and $n \ge 2$, take the following two values

$$\frac{1}{8.2^k}, \frac{1}{16.2^k}.$$

This theorem is proved by induction. Let us note that this result is true for blocks of length 1 and 2.

The preimage of a factor of length $2^k + 1$ is of length $2^{k-1} + 1$. Furthermore, the blocks of length 2 have frequency $\frac{1}{8}$. By using Lemma 3, we obtain that the blocks of length $2^k + 1$ have frequency $\frac{1}{8.2^k}$.

Let us suppose now that the factors whose length is in $[2^{k-1} + 1, 2^k]$ have frequencies $\frac{1}{8.2^{k-1}}$ or $\frac{1}{16.2^{k-1}}$. Consider next a factor of length $m \ge 3$, with $2^k + 1 \le m \le 2^{k+1}$. From Lemma 2, we deduce that the length of its preimage is in the interval $[2^{k-1} + 1, 2^k + 1]$. Hence its preimage has frequency $\frac{1}{8.2^{k-1}}$ or $\frac{1}{16.2^{k-1}}$, by using the induction hypothesis and the result above. We conclude here again with the help of Lemma 3.

We deduce Theorem 5 from this result, by using the bijection between the fixed point u^R and the Rudin-Shapiro sequence.

We can also similarly show the following result.

Lemma 7 *Let B be a right special factor of length greater than 8. Let k be such that $2^k + 1 \le l(B) \le 2^{k+1}$. Then, B has frequency $\frac{1}{8.2^k}$ and its right extensions have frequency $\frac{1}{16.2^k}$.*

Hence we deduce the following expression for the conditional block entropies of the Rudin-Shapiro sequence, by noticing that there are 8 right special factors of given length n, for n larger than 8.

Theorem 6 *We have* $H_0^R = 1$, $H_1^R = 1$, $H_2^R = 1$, $H_3^R = 2 - \frac{3}{4}\log_2 3$, $H_4^R = -1/2 + \frac{3}{4}\log_2 3$, $H_5^R = 7/8 - \frac{3}{16}\log_2 3$, $H_6^R = 1/16 + \frac{3}{16}\log_2 3$ *and* $H_7^R = 5/16$.

We have, for $n \geq 8$ *and* $2^k + 1 \leq n \leq 2^{k+1}$*:*

$$H_n^R = \frac{1}{2^k}.$$

Remark This method works also for the generalized Rudin-Shapiro sequences which count the number of occurrences of the pattern $1 \star \cdots \star 1$ in the binary expansion of every integer (see [3] and [4]). 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.

7 The paperfolding sequence

The paperfolding sequence v^P is the image of the fixed point $u^P = (\sigma^P)^\infty(a)$ of the substitution σ^P:

$$\begin{cases} \sigma^P(a) = ab \\ \sigma^P(b) = cb \\ \sigma^P(c) = ad \\ \sigma^P(d) = cd \end{cases}$$

by the projection

$$\begin{cases} p^R(a) = p^R(b) = 1 \\ p^R(c) = p^R(d) = 0. \end{cases}$$

We have also a bijection between the factors of length greater than 7 of the fixed point u^P and the factors of the same length of the projection v^P (see [1] or [2]).

Theorem 7 *For all* $n \geq 7$, $p_{u^P}(n) = p_{v^P}(n) = 4n$.

For the first values, we have: $p_{u^P}(n) = 4n$, *for* $1 \leq n \leq 6$, *and* $p_{v^P}(1) = 2$, $p_{v^P}(2) = 4$, $p_{v^P}(3) = 8$, $p_{v^P}(4) = 12$, $p_{v^P}(5) = 18$ *and* $p_{v^P}(6) = 23$.

There is a slight difference concerning the properties of "recognizability" of the fixed point: some factors can have two preimages. But we have the following properties.

Lemma 8 *There is a unique way to put the bars for every factor of the fixed point* u^P.

Namely, the image of every letter begins with a or c and ends with b or d. Hence we put a bar after a b or a d and a bar before an a or a c.

We deduce from this that a right (respectively a left) special factor of the fixed point u^P has exactly two extensions. Hence, the last (respectively the first) letters of a right (respectively a left) special factor form also a special factor with the same extensions. Thus, by considering the extensions of the special factors of length 2, we obtain the following lemma.

Lemma 9 *The right (respectively the left) extensions of the right (respectively the left) special factors of length greater than 2 of the fixed point u^P are b and d. Furthermore, the last letter of a right special factor of length greater than 4 is a c.*

We can characterize now the factors which have only one preimage.

Lemma 10 *A factor of u^P has a unique preimage if and only if this factor is not a right special one. Furthermore, a right special factor has exactly two preimages.*

Proof Let us first note that an odd factor with a letter before the first bar of its cut has only one preimage. Namely, if such a factor B would have two preimages then they would only differ by their first letter, say x_0 and x'_0, because of Lemma 8 and because of the injectivity of σ^P on the letters. Thus, B would have as left extensions the first letter of $\sigma^P(x_0)$ and $\sigma^P(x'_0)$, i.e. a and c, which is in contradiction with Lemma 9. Hence a factor having more than one preimage has exactly two preimages and we show similarly that it is a right special factor.

Conversely, a special factor B of length greater than 4 may be written as follows (Lemma 9):

$$B = y|\sigma^P(B')|c,$$

where y denotes the empty word if B is of odd length, and a letter otherwise.

It is easily seen, by using what precedes, that $y\sigma^P(B')$ has only one preimage and hence that B has two preimages, which can be written as $zB'b$ and $zB'd$, where z is the empty word if B is of odd length. Furthermore one checks that a right special factor of length 1, 2 or 3 has also two preimages.

We can show now the following result.

Theorem 8 • *Let B be a right special factor of u^P. Let k be such that $2^k \leq l(B) \leq 2^{k+1}-1$. Then B has frequency $\frac{1}{4.2^k}$ and its right extensions have frequency $\frac{1}{8.2^k}$.*

- *More generally the frequencies of the factors of the paperfolding sequence v^P of length n, with $2^k+1 \leq n \leq 2^{k+1}$, for $n \geq 7$, take the following two values*

$$\frac{1}{4.2^k}, \frac{1}{8.2^k}.$$

The blocks of length 1 have frequency $1/2$, the blocks of length 2 have frequency $1/4$, the blocks of length 3 have frequency $1/16$ or $3/16$, the blocks of length 4 have frequency $1/16$ or $1/8$, the blocks of length 5 have frequency $1/32$, $1/16$ or $3/32$, the blocks of length 6 have frequency $1/32$ or $1/16$.

We prove this theorem by showing here again the corresponding property for the fixed point u^R.

Let us prove the first assertion of Theorem 8. It is easily shown that this result is true for right special factors of u^R of length 1,2,3. Let us suppose now that it is true for right special factors of length in $[2^{k-1}, 2^k-1]$.

Let k, greater than 2, be such that $2^k \leq l(B) \leq 2^{k+1}-1$, where B is a right special factor of u^P. Let us write B as:

$$B = y|\sigma^P(B')|c,$$

where y denotes the empty word if B is of odd length, and a letter otherwise (Lemma 9). We have seen that the two preimages of B are $xB'b$ and $xB'd$, where x is the empty word if B is of odd length.

The factor xB' is thus a right special factor of length the integral part of $l(B)/2$, which belongs to the interval $[2^{k-1}, 2^k-1]$. From the induction hypothesis, we have:

$$P(xB') = \frac{1}{4.2^{k-1}} = 2P(xB'b) = 2P(xB'd).$$

The frequencies of B, xB', $xB'b$ and $xB'd$ satisfy the following relationships:

$$P(B) = \frac{P(xB'b)+P(xB'd)}{2} = \frac{P(xB')}{2} = \frac{1}{8.2^{k-1}},$$

i.e.

$$P(B) = \frac{1}{8.2^k}.$$

Furthermore, Bb (respectively Bd) has as unique preimage $xB'b$ (respectively $xB'd$), thus we have, as expected:

$$P(Bb) = \frac{P(B'b)}{2} = \frac{1}{8.2^k} = P(Bd).$$

Let us consider now factors of u^P of length greater than 4 which are not right special. We know that these factors have a unique preimage; we thus have the usual equality between the frequencies: if B a factor of unique preimage B', then $P(B) = \frac{P(B')}{2}$.

The idea is here to show first that the blocks of length 2^k have only one frequency, i.e. $\frac{1}{4.2^k}$. It is easily shown, by induction, that the frequencies of the factors of length 2^k and 2^k+1 take the two values $\frac{1}{4.2^k}$ and $\frac{1}{8.2^k}$. We can conclude then by considering the complexity: we have $p(2^k) = 4.2^k$ factors of length 2^k (see Theorem 7); hence only one frequency is possible, namely $\frac{1}{4.2^k}$.

The rest of the proof works exactly as the one of Theorem 5.

We can deduce now from the results above the expression of the conditional block entropies, by noticing that there are 4 right special factors of given length greater than 7.

Theorem 9 *We have* $H_0^P = 1$, $H_1^P = 1$, $H_2^P = 2 - \frac{3}{4}\log_2 3$, $H_3^P = -1/2 + \frac{3}{4}\log_2 3$, $H_4^P = 7/8 - \frac{3}{16}\log_2 3$, $H_5^P = 1/16 + \frac{3}{16}\log_2 3$ *and* $H_6^P = 5/16$.

We have, for $n \geq 7$ *and* $2^k \leq n \leq 2^{k+1} - 1$*:*

$$H_n^P = \frac{1}{2^k}.$$

Remark Let us note the following relationship between H_n^R and H_n^P.

Proposition 4 *We have* $H_n^P = H_{n+1}^R$*, for all* n*.*

8 Conclusion

Let us come back to the initial question of the comparison of conditional block entropies for these sequences. We have : $H^P \leq H^T \leq H^R$, for $n \leq 8$. But, for $n \geq 9$, this ordering does not hold. In particular, we have:

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

In fact, we see that there is no relation of order between H^R, H^P and H^T. From Proposition 4, 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: $H_n^T = \frac{2}{3} H_n^R = \frac{2}{3} H_n^P$.

Acknoledgments I would like to thank the referee for many useful comments and Y. Didier whose software for exhaustive enumeration of blocks has been of great help.

References

[1] J.-P. ALLOUCHE *The number of factors in a paperfolding sequence*, Bull. Austr. Math. Soc. **46** (1992), 23–32.

[2] J.-P. ALLOUCHE and M. BOUSQUET-MÉLOU *Canonical positions for the factors in the paperfolding sequences*, Bull. Belg. Math. Soc. **1** (1994), 145–164

[3] J.-P. ALLOUCHE and P. LIARDET *Generalized Rudin-Shapiro sequences*, Acta Arith. **60** (1991), 1–27.

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

[5] V. BERTHÉ *Fonctions de Carlitz et automates. Entropies conditionnelles*, Thèse, Université Bordeaux I (1994).

[6] V. BERTHÉ *Fréquences des facteurs des suites sturmiennes*, Preprint (1994).

[7] P. BILLINGSLEY *Ergodic theory and information*, John Wiley and Sons, New York (1965).

[8] S. BRLEK *Enumeration of factors in the Thue-Morse word*, Discrete Applied Math. **24** (1989), 83–96.

[9] B. L. BURROWS and K. W. SULSTON *Measures of disorder in non-periodic sequences*, J. Phys. A: Math. Gen. **24** (1991), 3979–3987.

[10] A. COBHAM *Uniform tag sequences*, Math. Systems Theory **6** (1972), 164–192.

[11] F. M. DEKKING *On the Prouhet-Thue-Morse Measure*, Acta Universitatis Carolinae, Mathematica et Physica, **33** (1992), 35–40.

[12] F. M. DEKKING, M. MENDÈS FRANCE and A. J. van der POORTEN *FOLDS!*, Math. Intell. **4** (1982), 130–138, 173–181, 190–195.

[13] A. de LUCA and S. VARRICHIO *Some combinatorial properties of the Thue-Morse sequence*, Theoret. Comput. Sci. **63** (1989), 333–348.

[14] M. QUEFFÉLEC *Substitution dynamical systems. Spectral analysis*, Lecture Notes in Mathematics **1294**, Springer-Verlag (1987).

[15] G. RAUZY *Suites à termes dans un alphabet fini*, Sém. de Théorie des Nombres de Bordeaux (1983), 25-01–25-16.

[16] C. E. SHANNON *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656.