On the characterization of the entropy associated with infinite sequences of non-independent symbols*

Armando B. Matos[†] Luís F. Antunes[‡]

November 2001

Abstract

Entropy is widely used as a measure of the information contained in a message. It is usually assumed that the symbols of the message are statistically independent. This has as a consequence that the entropy is only a function of the symbol probabilities. However, in practice this assumption is often not justified. In this paper we present a first step towards the characterization of non-independent entropy. Assuming simple symbol Markov generators, we give closed formulas for the Shannon and Rényi entropies associated with non-independent sequences of symbols. Inspired by the seminal work of Shannon [5] we propose a definition for the Rényi entropy that takes into account the statistical dependencies between symbols. As a corollary we get a formula for the min-entropy of an infinite sequence with non-independent symbols. Larger alphabets and more complex symbol dependencies are modeled by ergodic Markov chains in order to obtain general formulas for the entropy associated with sequences of non-independent symbols. Some simulation and experimental tests are presented in the appendix.

*The work presented here has been partially supported by funds granted to LIACC through the Programa de Financiamento Plurianual, Fundação para a Ciência e Tecnologia and Programa PRAXIS.

[†] DCC–FC & LIACC, Universidade do Porto, email: `acm@ncc.up.pt`

[‡] FEP & LIACC, Universidade do Porto, email: `lfa@ncc.up.pt`

1 Introduction

Shannon entropy ([5, 4]) is widely used in various disciplines as a measure of information. An important generalization of Shannon entropy is the Rényi entropy ([2]) which has found several applications to unconditionally secure cryptosystems ([1]). An important assumption underlying the definitions of Shannon and Rényi entropies is that the symbols of the message are statistically *independent*. This has as a consequence that the entropy depends only on the symbol probabilities. However, in practice this is often not the case. In this paper we discuss the concept of *dependent entropy*, that is the entropy of *sequences of non-independent* symbols. In fact, the entropy associated with a Markov process has already been briefly considered by Shannon (see [5]) and our work is in some sense a development of that seminal paper.

The following example illustrates an erroneous application of Shannon entropy formula to a sequence of non-independent symbols.

Example 1 *Consider an infinite sequence of bits* b_1, $b_2, \cdots$ *with*

$$pr(0) = pr(1) = 0.5$$

where, with large probability, each bit is equal to the previous one. The random process that generates the sequence is the following.

$$b_i = \begin{cases} b_{i-1} & \textit{with probability } \beta \\ 0 & \textit{with probability } (1-\beta)/2 \\ 1 & \textit{with probability } (1-\beta)/2 \end{cases}$$

For $\beta = 0.8$, *a typical sequence is*

$$111110001111111110000000010000000000001111111111110$$

Using the Shannon formula, we get that the information contents of each bit is

$$H_S = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = \frac{1}{2} + \frac{1}{2} = 1$$

This result is obviously wrong because in the Shannon formula (like in the formulas for others forms of entropy) it is assumed that the symbols are independent random variables. This is clearly not the case. It seems that each symbol contains much less than one bit of information; this loss corresponds to what is usually called redundancy*.*

In this paper we extend the concept of Rényi entropy to non-independent sequences of symbols[1], deriving some closed formulas for some simple random bit generators. Based on this extension and using some know properties of Rényi entropy we obtain a formula for the *min*-entropy associated with infinite sequences of non-independent symbols, that we believe can be useful in cryptography.

The main idea behind the methodology we use to compute the entropy associated with sequences of non-independent symbols can be summarized as follows. By assuming that large blocks of bits are independent, the effects caused by the dependences between successive bits are "translated" into the probabilities associated to the block symbols (a block symbol is a sequence of n bits where n is the block size; see the definition in Section 2); large dependences will correspond to very different probabilities. We then compute the entropy that corresponds to a block alphabet $\{0,1\}^n$ and divide it by n; the entropy per bit is the limit of this fraction when $n \to \infty$.

Let us consider again Example 1. If we divide the sequence in blocks of n bits, we should obtain, if n is large enough, a value for H_s/n which is less than 1; for instance, for $n = 4$, the block alphabet has $2^4 = 16$ symbols and the sequence given in the example should be seen as

1111 1000 1111 1111 1000 0000 0100 0000 0000 0011 1111 1111 1110

In the new alphabet, the symbols are not equiprobable; for instance, 1111 is much more probable than 0101. As a consequence, we should get a value less than 1 for the entropy per bit ($H_S/4$).

The remainder of this paper is organized as follows. In the next Section some general definitions related to the concept of entropy are given. In Section 3 we characterize two simple Markov random bit generators. In Section 4 the probabilities associated with the corresponding block symbols are calculated. Closed formulas for the *dependent Shannon entropy* and *dependent Rényi entropy* are derived respectively in Sections 5 and 6. In Section 7 Markov chains are used to model larger alphabets and more complex symbol dependencies in order to obtain more general and possibly more useful results. In the concluding Section we summarize the main results of this paper and give some prospects for future work.

[1] This concept is based on the statistical dependence between the successive symbols of a message and it should not be confused with the concept of conditional entropy.

2 Definitions

Let X be a discrete random variable. Shannon entropy of X is defined as

$$H_S = \sum_{x \in X} P_X(x) \log \left(\frac{1}{P_X(x)} \right)$$

Instead of $P_X(x)$ we will write simply p_i where i is an integer corresponding to the value of x in some standard enumeration of the values associated with the random variable.

The Rényi entropy is a generalization of the Shannon entropy and is defined as

$$H_r^\alpha = \frac{1}{1-\alpha} \log \sum_{x \in X} P_X(x)^\alpha$$

where α is a nonnegative parameter. The Rényi entropy includes as particular cases: the logarithm of the alphabet size ($\alpha = 0$), the Shannon entropy, ($\alpha = 1$) and the minimum entropy ($\alpha = \infty$) which is a standard measure of randomness in several cryptographic applications (see for instance, [6] for a description of the use of minimum entropy in the theory of extractors and dispersers). Varying α from 0 to ∞ amounts to a shift on the symbol weight from a situation where all n bit blocks have the same weight to the other extreme where only the most probable one matters, by accentuating still more and more probable subsequences.

Through this paper the alphabet is $\{0,1\}$; the symbols will also be called *bits*. For a fixed integer $n \geq 1$ a *block* is a sequence of n bits.

Definition 2 (Block alphabet) *Let $\Sigma = \{0,1\}$ be the initial alphabet and let n be the size of each block. The* block alphabet *of size n is Σ^n. We denote by p_i^n, for $1 \leq i \leq |\Sigma|^n$, the probability of symbol number i. The probability of a specific block symbol $x \in \Sigma^n$ is denoted by $p^n(x)$.*

Definition 3 (Entropy for sequences of non-independent symbols) *Assume that, for some $n \geq 1$, the alphabet is $\{0,1\}^n$. The Shannon entropy per bit, h_S is defined as*

$$h_S = \frac{1}{n} H_S = \lim_{n \to \infty} \frac{1}{n} \left(\sum_{i=1}^{2^n} p_i \log \frac{1}{p_i} \right)$$

where, for $1 \leq i \leq n$, p_i denotes the probability associated with symbol number i of the block alphabet.

The Rényi entropy per bit, h_r^α is defined similarly as

$$h_r^\alpha = \frac{1}{n} H_r^\alpha = \lim_{n\to\infty} \frac{1}{n} \left(\frac{1}{1-\alpha} \log \left(\sum_{i=1}^{2^n} p_i^\alpha \right) \right)$$

We will not discuss in detail conditions for the existence of the limits mentioned in Definition 3. However, we can expect that they exist if we assume that the dependences between block symbols decrease to zero when their size increase to infinity and that the process is in some intuitive sense stationary[2]. This is certainly true for the random bit generators used in this paper.

We can also use *computed* symbol probabilities instead of probabilities. This has the advantage of being a property of a single sequence, not requiring the previous knowledge of the probabilities involved.

Definition 4 (Single sequence non-independent entropy) *The single sequence Shannon per bit entropy, h_S^s is defined as*

$$h_S^s = \lim_{n\to\infty} \frac{1}{n} \lim_{m\to\infty} \sum_{i=1}^{2^n} cp_i^{m,n} \log \frac{1}{cp_i^{m,n}}$$

where $cp_i^{m,n}$ is the relative frequency (or "computed probability") of the symbol number i in a sequence s of m block symbols with size n (that is, $s \in (\Sigma^n)^m$) and is defined as

$$cp_i^{m,n} = \frac{\textit{Number of occurrences of symbol number } i \textit{ in sequence } s}{m2^n}$$

The single sequence Rényi per bit entropy, h_r^s is defined similarly:

$$h_r^{\alpha,s} = \lim_{n\to\infty} \left(\frac{1}{n} \lim_{m\to\infty} \frac{1}{1-\alpha} \log \left(\sum_{i=1}^{2^n} (cp_i^{m,n})^\alpha \right) \right)$$

where the $cp_i^{m,n}$ are as defined above.

3 Generators

In this paper we begin by considering only two simple bit generators. More general generators seem to introduce unnecessary complications in the formulas and obscure the general ideas behind our definitions. Later, in Sectionmarkov, we will study ergodic Markov chains as symbol generators.

[2] In particular, this happens if the probabilities involved are the stationary state probabilities if an irreducible Markov chain where all states are ergodic ([3]); see also Section 7.

We are only interested in the assimptotic behavior of (infinite) sequences. In one of the generators, denoted "+", each bit, with a certain probability β, equals the previous one; in the other, each bit, with probability β, is the *opposite* of the previous one.

Both generators can be modeled by a two state Markov chain with appropriate transition probabilities. But we prefer to use models with an independent parameter (β, see below) that corresponds to the degree of dependence between successive bits.

Definition 5 (Generator $+_\beta$) *The first bit b_1 is either 0 or 1 with probability 0.5. And, for $i \geq 2$*

$$b_i = \begin{cases} b_{i-1} & \textit{with probability } \beta \\ 0 & \textit{with probability } (1-\beta)/2 \\ 1 & \textit{with probability } (1-\beta)/2 \end{cases}$$

The parameter β will be called dependence factor *as it represents the influence of the previous bit on the current one.*

Definition 6 (Generator $-_\beta$) *The first bit b_1 is either 0 or 1 with probability 0.5. And, for $i \geq 2$*

$$b_i = \begin{cases} \neg b_{i-1} & \textit{with probability } \beta \\ 0 & \textit{with probability } (1-\beta)/2 \\ 1 & \textit{with probability } (1-\beta)/2 \end{cases}$$

where $\neg 0 = 1$ and $\neg 1 = 0$.

Although these generators may be suitable for time sequences of random bits, they seem to introduce a somewhat artificial "time arrow" which may be inadequate for some cryptographic applications. Notice however that, in the stationary state, there no preferred the direction; this is a consequence of the fact that the block symbol probabilities computed in Section 4 are time-symmetrical, that is, for every $x \in \{0,1\}^n$, we get $p(x) = p(x^r)$ where x^r denotes the reversal of x.

4 Block symbol probabilities

In this section we compute the probability of each block symbol $s \in \{0,1\}^n$ where it is assumed that the symbols are generated by a $+_\beta$ or by a $-_\beta$

process and that the block size n is large so that the block symbols may be considered independent.

Let us begin with the $+_\beta$ generator. Consider a symbol of $\{0,1\}^n$:

$$b_1 b_2 \cdots b_n$$

To compute the probability of a specific symbol notice that, with probability $(1-\beta)/2$, there is a "transition" (from 0 to 1 or from 1 to 0) and that, with probability $\beta + (1-\beta)/2 = (1+\beta)/2$, there is no transition. So, if a symbol has t transitions, its probability is

$$p_+(\beta, t) = \left(\frac{1-\beta}{2}\right)^t \left(\frac{1+\beta}{2}\right)^{n-t}$$

where we are ignoring the influence of the last bit of the previous symbol on b_1; however, as we are looking for limits like $\lim_{n\to\infty} \cdots$, this approximation is justified.

Just to check the formula, we have: (i) $p(1,0) = 1$, the symbol must be 0^n or 1^n, and (ii) $p(0,n) = 2^{-n}$, if the symbols are independent, all symbols are equiprobable.

For the $-_\beta$ symbol generator we get the following probabilities.

$$p_-(\beta, t) = \left(\frac{1+\beta}{2}\right)^t \left(\frac{1-\beta}{2}\right)^{n-t}$$

Notice that, for any fixed parameter β, the probability of a block symbol depends only on the number of transitions; in particular, this has as a consequence that

$$\forall x \in \{0,1\}^n \;\; p(x) = p(x^r)$$

where x^r denotes the reversal of x.

5 Shannon entropy

We now compute the Shannon entropy per bit corresponding to a block alphabet of size n.

$$h_S = -\frac{1}{n}\sum_i p_i \log p_i = \frac{1}{n}\sum_t m(t) p(\alpha, t) \log(p(\alpha, t))$$

where

$$m(t) = \binom{n}{t}$$

is the number of symbols with t transitions.

For the $+_\beta$ generator we get

$$p_+(\beta,t) = \left(\frac{1-\beta}{2}\right)^t \left(\frac{1+\beta}{2}\right)^{n-t} = \left(\frac{1+\beta}{2}\right)^n \left(\frac{1-\beta}{1+\beta}\right)^t$$

and

$$\log(p_+(\beta,t)) = -n(1-\log(1+\beta)) + t\log\frac{1-\beta}{1+\beta}$$

so that the generalized Shannon entropy per bit is

$$\begin{array}{rcl}
h_+(\beta,n) & = & -\frac{1}{n}\sum_t m(t)p_+(\beta,t)\log p_+(\beta,t) \\
& = & -\frac{1}{n}\left(\frac{1+\beta}{2}\right)^n \sum \binom{n}{t}\left(\frac{1-\beta}{1+\beta}\right)^t \log(p_+(\beta,t)) \\
& = & -\frac{1}{n}\left(\frac{1+\beta}{2}\right)^n \sum \binom{n}{t}\left(\frac{1-\beta}{1+\beta}\right)^t \left[-n(1-\log(1+\beta)) + t\log\frac{1-\beta}{1+\beta}\right] \\
& = & (1-\log(1+\beta))\left(\frac{1+\beta}{2}\right)^n \sum \binom{n}{t}\left(\frac{1-\beta}{1+\beta}\right)^t \\
& & -\frac{1}{n}\left(\frac{1+\beta}{2}\right)^n \sum \binom{n}{t}\left(\frac{1-\beta}{1+\beta}\right)^t \left(t\log\frac{1-\beta}{1+\beta}\right) \\
& = & (1-\log(1+\beta))\left(\frac{1+\beta}{2}\right)^n \left(1+\frac{1-\beta}{1+\beta}\right)^n \\
& & -\log\frac{1+\beta}{1-\beta}\left(\frac{1+\beta}{2}\right)^n \sum \binom{n-1}{t-1}\left(\frac{1-\beta}{1+\beta}\right)^t \\
& = & (1-\log(1+\beta))(\frac{1+\beta}{2})^n(\frac{2}{1+\beta})^n - (\frac{1+\beta}{2})^n(\frac{2}{1+\beta})^{n-1}(\frac{1-\beta}{1+\beta})\log\frac{1-\beta}{1+\beta} \\
& = & 1-\log(1+\beta) - \frac{1-\beta}{2}\log\frac{1-\beta}{1+\beta}
\end{array}$$

where we have used the following identities

$$\sum_i \binom{n}{i} x^i = (1+x)^n$$

and

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$

Notice that the value of n (the block length) does not appear in the per bit entropy formula; this is due to the assumption we have done on the computation of the block symbol probabilities: the value of n is large enough so that the influence of the previous block on the current one can be discarded. In conclusion, we do not need to compute the limit mentioned in Definition 3.

We state the formula of the per bit Shannon entropy as a Theorem.

Theorem 7 *The per bit Shannon entropy of a sequence of non-independent symbols produced by with the $+_\beta$ generator is*

$$h_+(\beta) = 1 - \log(1+\beta) - \frac{1-\beta}{2}\log\frac{1-\beta}{1+\beta} \qquad (1)$$

For the $-_\beta$ generator, the Shannon entropy per bit is

$$h_-(\beta) = 1 - \log(1-\beta) - \frac{1+\beta}{2}\log\frac{1+\beta}{1-\beta} \qquad (2)$$

Notice that, if we replace in formula (1) β by $-\beta$, we get formula 2. Notice also that

$$h_+(\beta) = h_+(-\beta) = h_-(\beta) = h_-(-\beta)$$

although negative values of β (negative probabilities) do not have of course physical meaning. This symmetry is obvious from the equivalent formula

$$h_+(\beta) = h_-(\beta) = 1 - \frac{1}{2}(\log(1-\beta^2) + \beta\log(1+\beta) - \beta\log(1-\beta))$$

In Figure 1 we can see graphical representations of $h_-(\beta)$ (on the left) and $h_+(\beta)$ (on the right).

To check the formula (1) we consider two limit cases.

- When $\beta = 0$ the symbols are independent and we get that the entropy per bit is in fact 1:
$$\lim_{\beta\to 1} h_+(\beta) = 1$$
- When β is very close to 1, each symbol is, with great probability equal to the previous one and the message conveys no information:
$$\lim_{\beta\to 1} h_+(\beta) = 1 - 1 - \lim_{\beta\to 1}\frac{1-\beta}{2}\log\frac{1-\beta}{1+\beta} = 0$$

In Appendix A we give some simulation results which support formulas (1) and (2).

6 Rényi entropy

Using again the $+_\beta$ generator we can compute the non-independent Rényi entropy; the calculations are similar to those leading to Shannon entropy (formula (1)) and we present only the resulting formulas.

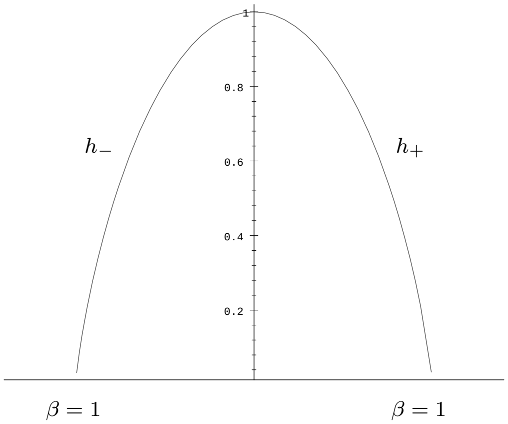


Figure 1: Shannon generalized entropy as a function of β for the generators $-_\beta$ (left) and $+_\beta$ (right).

Theorem 8 *The per bit Rényi entropy of a sequence of non-independent symbols produced by with the $+_\beta$ generator is*

$$h_+^\alpha(\beta) = \frac{\alpha}{\alpha - 1}\left(1 - \log(1+\beta) - \frac{1}{\alpha}\log\left(1 + \left(\frac{1-\beta}{1+\beta}\right)^\alpha\right)\right) \qquad (3)$$

For the $-_\beta$ generator the per bit Rényi entropy is

$$h_-^\alpha(\beta) = \frac{\alpha}{\alpha - 1}\left(1 - \log(1-\beta) - \frac{1}{\alpha}\log\left(1 + \left(\frac{1+\beta}{1-\beta}\right)^\alpha\right)\right) \qquad (4)$$

Using again the $+_\beta$ generator, let us consider some particular values of the Rényi parameter α.

First we compute the minimum entropy

$$\lim_{\alpha\to\infty} h_\alpha(\beta) = 1 - \log(1+\beta)$$

This is a particularly simple formula which can be checked for $\beta = 0$ (symbol independence) where each bit contains one bit of “information” and for $\beta = 1$ (each symbol equals the previous one) where each bit contains no “information” at all; both of these extreme values were of course expected.

In fact, the minimum entropy for any $0 < \beta \leq 1$ can be directly computed as follows. Consider, for instance the $+_\beta$ generator (for the $-_\beta$ the computations are similar). Assuming $\beta > 0$, the most probable symbols are 1^n and 0^n, both with assimptotic probabilities $((1+\beta)/2)^n$. The logarithm of the inverse of this probability divided by n is the minimum entropy per bit and it equals the formula given above, $1-\log(1+\beta)$.

The Shannon non-independent entropy (1) can be also computed by the limit

$$\lim_{\alpha \to 1} h_+^\alpha(\beta)$$

Finally, consider the entropy corresponding to the value $\alpha = 0$:

$$\lim_{\alpha \to 0} h_+^\alpha(\beta) = 1$$

This is an interesting result: the Rényi entropy with $\alpha = 0$ seems to completely ignore the symbol dependences! This case suggests that there may be a close relationship between the parameter α of the Rényi entropy and the loss of information per bit that results from the influence of neighbor bits. In Figure 2 we can see, for some fixed values of the dependence parameter β, the value h_+ as a function of α.

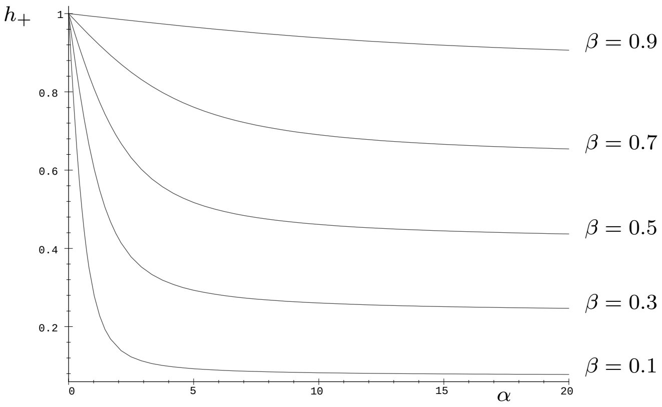


Figure 2: The value h_+ as a function of α, for some values of β

Figure 3 represents in a 3-dimensional graph the functions $h_+(\alpha, \beta)$ and $h_-(\alpha, \beta)$.

‘],tickmarks=[default,2,default])

In Appendix A we give some simulation results and comparisons with Formula 3.

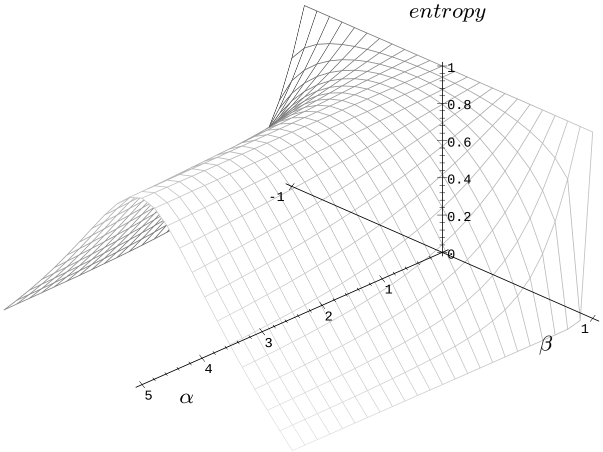


Figure 3: The relationship between symbol dependence and the parameter α of Rényi entropy.

7 Rényi entropy for any generator

We started this work considering only two simple bit generators in order to avoid obscuring the general ideas behind our definitions and unnecessary complications in the formulas. However, at this point we are able to use some well know results from Stochastic Processes so as to capture more complex forms of dependency between sets of random variables; in particular Markov Chains will be used to develop a more general formula for Rényi entropy. Like Shannon, we consider that ergodic Markov chains are an appropriate characterization of the usual stationary statistical sources:

Among the possible discrete Markoff processes there is a group with special properties of significance in communication theory. This special class consists of the "ergodic" processes and we shall call the corresponding sources ergodic sources [5]

7.1 Markov processes

For the reader convenience we include here a few well known concepts and results about Markov chains; for more information, consult for instance [3].

Definition 9 (Stochastic Process) *A stochastic process $\{X(t) : t \in \mathcal{T}\}$ is a collection of random variables defined on a common sample space and using the same probability measure P defined for all events.*

Usually t represents time, $\mathcal{T}$ a set of points in time and $X(t)$ as the value or state of the stochastic process at time t.

Definition 10 (Markov Process) *A Markov process is a Stochastic process, $\{X(t) : t \in \mathcal{T}\}$, where the probability distribution for X_{t+1} depends only of X_t, and not additionally on what occurred before time t (doesn't depend of X_s, where $s < t$).*

We will think of $\mathcal{T}$ in terms of time, and the values that $X(t)$ can assume are called *states* which are elements of a *state space* $\mathcal{S}$. Markov processes are classified by whether sets $\mathcal{T}$ and $\mathcal{S}$ are discrete (Markov chain) or continuous (Markov process).

Definition 11 (Transition probability) *The transition probability of state i to state j at time $n-1$ is given by*

$$P[X_n = j | X_{n-1} = i].$$

Definition 12 (Time homogeneous) *A Markov chain is time homogeneous if*

$$P[X_n = j|X_{n-1} = i] = P[X_{n+m} = j|X_{n+m-1} = i], m \geq 0, i, j \in S.$$

So if the chain is time homogeneous we can write

$$p_{i,j} = P[X_n = j|X_{n-1} = i]$$

i.e., if X_{n-1} takes the value i, then X_n has the distribution given by the i-th row of P. We can define the matrix of transition probability by

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,j} & \cdots \\ p_{1,0} & p_{1,1} & \cdots & p_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ p_{i,0} & p_{i,1} & \cdots & p_{i,j} & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix}$$

If S is finite then P has finite dimension. We write $i \longmapsto j$ if $p_{ij} > 0$, which means that the chain can jump directly from i to j. A Markov chain is completely defined by its one-step transition probability matrix and a specification of a probability distribution on the state of the process at time 0.

Suppose we are given an initial probability distribution π^0, where we denote $P[X_0 = i]$ by π_i^0, and a matrix of transition probability P. Then,

$$\begin{aligned}
\pi_j^{(1)} &= P[X_1 = j] \\
&= \sum_i P[X_1 = j | X_0 = i] \cdot P[X_0 = i] \\
&= \sum_{i=1}^{n} p_{ij} \cdot \pi_i^0
\end{aligned}$$

i.e.

$$\pi^{(1)} = \pi^{(0)} \cdot P.$$

So, as the process is Markovian, we can consider π^1 to be the initial probabilities for the next step of a one-step Markov chain and thus can write

$$\pi^{(2)} = \pi^{(1)} \cdot P.$$

More generally we have

$$\pi^{(n)} = \pi^{(n-1)} \cdot P, n = 1, 2, 3, ...$$

Repeating this argument, we obtain

$$\pi^{(n)} = \pi^{(0)} \cdot P^n, n = 1, 2, 3, ...$$

Here, $\pi^{(n)}$ represents the distribution after n steps. Therefore, the matrix P^n is the so called "n-step transition matrix". Its entries are

$$p_{ij}^n = P[X_{t+n} = j | X_t = i].$$

Definition 13 *A Markov chain is said to be ergodic if* $\lim_{n\to\infty} p_{ij}^n = \pi_j > 0$ *for all* j *and is independent of* i.

In this case,

$$\begin{aligned}
P^\infty &= \lim_{n\to\infty} P^n \\
&= \begin{bmatrix} \pi_1 & \dots & \pi_j & \dots & \pi_n \\ \vdots & & \vdots & & \vdots \\ \pi_1 & \dots & \pi_j & \dots & \pi_n \end{bmatrix}.
\end{aligned}$$

Hence, π is independent of the starting distribution $\pi^{(0)}$:

$$\pi = \pi^{(0)} \cdot P^\infty.$$

Any vector π which satisfies $\pi P = \pi$ and $\sum_i \pi_i = 1$ is called a *stationary distribution*. We say P is *irreducible* if for all $i, j \in \mathcal{S}, p_{ij}^n > 0$ for some n, i.e., the underlying graph is strongly connected. If P is irreducible, we say P is *aperiodic* if for some i the set of n with p_{ii}^n has greatest common divisor 1 all we need to keep in mind is that if $p_{ii} > 0$ for some i then P is aperiodic.

Theorem 14 *If $\mathcal{S}$ is finite and P is irreducible, then:*

1. *There is a unique stationary distribution π for P.*
2. *With probability 1, $\frac{|\{r \leq n : X_r = j\}|}{n} \rightarrow \pi_j$ as $n \rightarrow \infty$, regardless of X_0, i.e., π_j equals the fraction of time the Markov chain spends in state j for almost all sample paths of the chain.*
3. *If P also is aperiodic then $p_{ij}^n \rightarrow \pi_j$ as $n \rightarrow \infty$, for each i, j.*

It often happens in practice that P is "known" but π is not and solving $\pi P = \Pi$ directly is computationally infeasible. Theorem 14 shows we can simulate a variable with approximate distribution π by run a Markov chain with transition matrix P for a "long" time. This is called *Monte Carlo Markov Chain.*

Theorem 15 *A Markov chain is ergodic if and only if both of the following are true:*

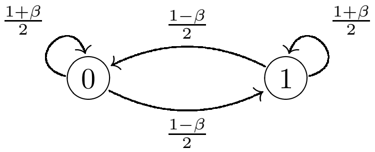
1. *it is irreducible*
2. *the chain is aperiodic.*

7.2 Rényi entropy associated with ergodic sources

As most of the sources in communication theory seem to have the ergodicity properties, we will consider only ergodic sources. First let us check that ours sources are indeed ergodic.
Consider the $+_\beta$ generator

$$b_i = \begin{cases} b_{i-1} & \text{with probability } \beta \\ 0 & \text{with probability } (1-\beta)/2 \\ 1 & \text{with probability } (1-\beta)/2 \end{cases}$$

This generator can be represented by the following Markov chain



The corresponding transition probability matrix is

$$P = \begin{bmatrix} \frac{1+\beta}{2} & \frac{1-\beta}{2} \\ \frac{1-\beta}{2} & \frac{1+\beta}{2} \end{bmatrix}$$

As we can see the underlying graph is strongly connected and

$$p_{00} > 0, \ p_{11} > 0$$

so the P is irreducible and aperiodic and by Theorem 15 the source is "ergodic". By Theorem 14 we know that $p_{ij}^n \to \pi_j$ as $n \to \infty$, for each i, j. So, the value of $\lim_{n\to\infty} P^n$ is the stationary distribution $\pi = (\frac{1}{2}, \frac{1}{2})$, i.e., for almost all sample paths of the Markov chain, the fraction of time spent in state 0 (1) equals $\frac{1}{2}$ $(\frac{1}{2})$.
In section 6 we have already computed a closed formula (3) for the Rényi entropy of this generator. We now propose a new definition for Rényi entropy of any ergodic source. We will see that, at least for the generators $+_\beta$ and $-_\beta$, this definition is compatible with (4) although at first they seem very different.

Definition 16 (Ergodic Rényi entropy) *Let S be an ergodic source with probability transition matrix $P = [p_{ij}]$, the Rényi entropy for this source is given by*

$$ER_\alpha(S) = \sum_i \pi_i H_\alpha^i$$

where π is the stationary distribution for P and H_α^i is the classical Rényi entropy for each state of the Markov chain representing S.

This definition replaces the classical Rényi formula by the weighted average of the entropies of the various states where the wheigh associated a each state is simply its probability of ocurrence.

Consider again the $+_\beta$ generator, if we now evaluate the ergodic Rényi entropy we get

$$\frac{1}{1-\alpha} \log \frac{(1+\beta)^\alpha + (1-\beta)^\alpha}{2^\alpha}$$

which is equivalent to formula (3) we proposed in Section 6.

Theorem 17 *Consider the ergodic Rényi entropy of an ergodic source then:*

1. $ER_0 = \log |X|$.

2. $ER_1 = \sum_i \pi_i H^i$,*where* H^i *is the Shannon entropy associated with state* i.
3. $ER_\infty = \sum_i \pi_i(-\log \max p_{ij})$.

Proof:
We prove the theorem using the following properties of Rényi entropy:

- $\lim_{\alpha\to 1} H^i_\alpha = H^i$
- $\lim_{\alpha\to\infty} H^i_\alpha = -\log\max p_{ij}$

1. $ER_0 = \sum_i \pi_i H^i_0 = \sum_i \pi_i \log \sum_j p^0_{ij} = \sum_i \pi_i \log \sum_j 1 = \sum_i \pi_i \log |X| = \log |X| \sum_i \pi_i = \log |X|$.
2. $ER_1 = \lim_{\alpha\to 1} \sum_i \pi_i H^i_\alpha = \sum_i \pi_i \lim_{\alpha\to 1} H^i_\alpha = \sum_i \pi_i H^i$.
3. $ER_\infty = \lim_{\alpha\to\infty} \sum_i \pi_i H^i_\alpha = \sum_i \pi_i \lim_{\alpha\to\infty} H^i_\alpha = \sum_i \pi_i(-\log\max p_{ij})$.

□

As a consequence of Theorem 17 we get a formula for the *min*-entropy of an ergodic source.

Corollary 18 (Ergodic *min*-entropy) *Let S be an ergodic source with probability transition matrix $P = [p_{ij}]$. The ergodic* min-*entropy is*

$$\sum_i \pi_i(-\log\max p_{ij})$$

8 Conclusions and future research

We gave (see Definition 3) a generalization of the concept of entropy which is appropriate for sequences of symbols that may be not statistically independent.

For simple probabilistic dependences, closed formulas (1), (2), (3), and (4) (Theorems 7 and 8) were obtained. In particular, it is interesting to look at the Rényi formulas (3) and (4) as a function of α; they seem to imply a close relationship between the parameter α associated with the (generalized) Rényi entropy and the loss of entropy due to the dependence on neighbor symbols; in particular, for $\alpha = 0$ there is no such loss, the Rényi entropy is always 1 (in the general case, the logarithm of the alphabet size).

For the general ergodic symbol generator, closed formulas were not obtained; however we think that concepts like the minimum entropy for dependent symbols (see Corollary 18) may have applications to the random sources used in Cryptography.

To conclude, we list some problems for future work in the area of information (or entropy) of sequences of non-independent symbols.

- Obtain *simple* closed formulas for the entropy associated with other, more general, forms of symbol dependence (in this paper closed formulas were only obtained for the symbol generators $+_\beta$ and $-_\beta$)
- Study of the interplay between conditional entropy and symbol dependence.
- Identify practical applications where the concept of the entropy (or information) associated with sequences of non-independent symbols is meaningful and possibly useful. A possible candidate may correspond to the repetitive use of an input random source (where each sample is considered as a symbol and successive samples may be not independent) in some cryptographic module such as an extractor([6]).

References

[1] Christian Cachin (1997). *Entropy Measures and Unconditional Security in Cryptography.* Swiss Federal Institute of Technology Zurich.

[2] Alfred Renyi (1970). *Probability Theory.* North-Holland, Amsterdam.

[3] Randolph Nelson (1995). *Probability, Stochastic Processes, and Queueing Theory.* Springer-Verlag.

[4] Claude E. Shannon (1949). *Communication Theory of Secrecy Systems.* Bell System Technical Journal, **28**, pages 656-715.

[5] Claude E. Shannon, (1948). *A Mathematical Theory of Communication.* Bell System Technical Journal, **27**, pages 379-423, 623-656.

[6] Noam Nisan, (1997). *Extracting randomness: how and why. A survey.* Institute of Computer Science, Hebrew University Jerusalem, Israel.

A Shannon and Rényi: Simulation results compared with the theoretical formulas

We compare the theoretical formulas 1, 2, 3 and 4 with approximations for several block sizes of the single sequence (Shannon and Rényi) entropy obtained by a simulation program.

For each dependence factor β and block size n, we generate 5×10^7 pseudo-random bits using the appropriate bit generator. The frequencies of each of the 2^n block symbols is computed. Then, an approximation to the value of the Shannon non-independent entropy $h_+^\alpha(\beta)$ is obtained by fixing in Definition (4) large values of m and n, instead of taking the limits.

$$\frac{1}{n}\sum_{i=1}^{2^n} cp_i^{m,n} \log \frac{1}{cp_i^{m,n}}$$

where n and m are respectively the block size and the total number of bits in the simulation (5×10^7).

For the Rényi entropy the corresponding approximation is

$$\frac{1}{n}\left(\frac{1}{1-\alpha}\log\left(\sum_{i=1}^{2^n}(cp_i^{m,n})^\alpha\right)\right)$$

The approximations to the Rényi entropy for several block sizes n are represented for the values of $\alpha = 0.5$, 1 (Shannon entropy) and 2 in Figures respectively 4, 5 and 6.

In Figure 3 we can observe a three-dimensional representation of the Rényi entropy as a function of α and β; the generators $-_\beta$ and $+_\beta$ are represented respectively on the negative and positive parts of the β axis.

In each case, it can be seen that as the dimension n of the blocks grows, the computed approximation approaches the theoretical formulas (1), (2), (3) and (4).

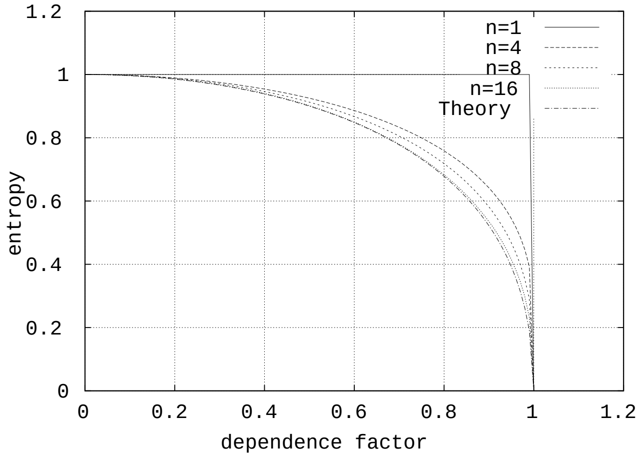


Figure 4: Experimental approximations of the Rényi entropy with parameter $\alpha = 0.5$ and block sizes 1, 4, 8 and 16. If the block size n is equal to 1, the entropy is constant and equal to 1; for larger block sizes, the entropy approaches the theoretical value (bottom curve).

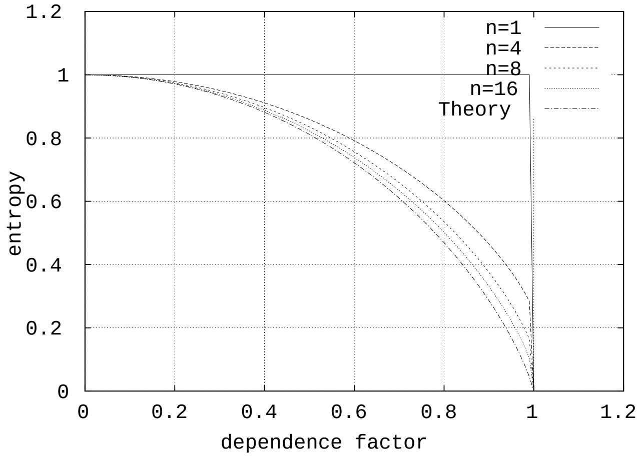


Figure 5: Experimental approximations of the Rényi entropy with parameter $\alpha = 1$ (Shannon entropy) and block sizes 1, 4, 8 and 16.

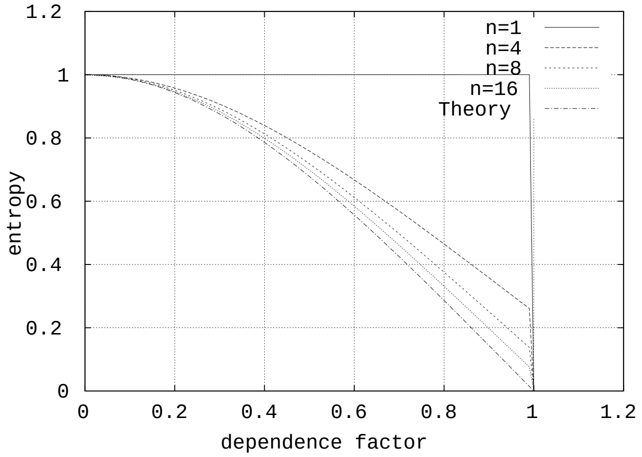


Figure 6: Experimental approximations of the Rényi entropy with parameter $\alpha = 2$ and block sizes 1, 4, 8 and 16.