Marek LEŚNIEWICZ

Military Communication Institute

Expected Entropy as a Measure and Criterion of Randomness of Binary Sequences

Abstract. *We present a new approach to modelling binary random sequences. We introduce a new concept of expected entropy which enables to explain us the problem that in practice the sample entropy never achieves its limit values. We show how to use the expected entropy to estimate the randomness of physically generated binary random sequences. Our theoretical analysis have been verified experimentally.*

Streszczenie. *Przedstawiamy nowe podejście do modelowania losowych ciągów binarnych. Wprowadzamy nowe pojęcie entropii oczekiwanej, które pozwala wyjaśnić, dlaczego entropia prób ciągów nigdy nie osiąga wartości granicznej. Pokazujemy, jak wykorzystać entropię oczekiwaną do oszacowania losowości ciągów losowych generowanych sprzętowo. Nasze analizy teoretyczne zostały potwierdzone doświadczalnie.* (***Entropia oczekiwana jako miara i kryterium losowości ciągów binarnych***)

Keywords: randomness, entropy, random binary sequence, Markov chain
Słowa kluczowe: losowość, entropia, losowy ciąg binarny, łańcuch Markowa

doi:10.12915/pe.2014.01.10

Introduction

In his seminal papers [1, 2] Shannon introduced the notion of *information entropy* which enabled him to formulate the fundamental principles of communication theory and cryptology. He defined the entropy of an N–dimensional binary random variable $\boldsymbol{X} = (X_1, \ldots X_N)$ (each variable X_i takes values 0 or 1 and hence $\boldsymbol{X}$ has values from $\{0, 1\}^N$) as a function of their probabilities $P(X_1, \ldots X_N)$, i.e.

$$H(X_1, ..., X_N) = -\frac{1}{N} \sum_{X_1,\ldots,X_N=0}^{2^N-1} P(X_1, \ldots, X_N) \log_2 P(X_1, \ldots, X_N), \tag{1}$$

which is an information entropy in the sense *entropy rate* (per each random variable X_i in the N–dimensional random variable $(X_1, \ldots X_N)$).

The information entropy in the sense of Shannon is a function of probabilities of random variables and cannot be measured like physical or numerical values. This is a consequence of the fact that the probabilities are *a priori* properties of the variables. However, we can measure a *posteriori* relative frequencies of realizations of random variables as approximations of the probabilities in the classical sense. In the following, realizations of binary random variables are named *binary random sequences* (further called a *random variables* and *random sequences*).

Let $n(X_1, \ldots X_N)$ be the number of N–element subsequences observed in a sample of n bits from a random sequence, where $(X_1, \ldots X_N)$ describes a given pattern of N bits. We define the *relative frequencies* $n(X_1, \ldots X_N)\, N / n$ and the *sample entropy* as

$$H_R(X_1, \ldots, X_N) \mid n) = -\frac{1}{N} \sum_{X_1,\ldots,X_N=0}^{2^N-1} (n(X_1, \ldots, X_N)\, N / n) \log_2 (n(X_1, \ldots, X_N)\, N / n), \tag{2}$$

We adopt the convention that $0 \cdot \log_2 0 = 0$ if in a sample $n(X_1, \ldots, X_N) = 0$.

The first experiment

In communication theory after the process of scrambling and in cryptography after encryption we use a model of *perfectly random sequences* as a realization of an N–dimensional *perfectly random variable* $(X_1, \ldots, X_N)$ with uniform probabilities $P(X_1, \ldots, X_N) = 1/2^N$. The entropy of such a random variable is equal to $H(X_1, \ldots, X_N) = 1$.

In our experiments we used a physical random number generator and produced 3 series of 10 samples of perfectly random sequences each of size n = 10 MB, 100 MB, and 1 GB, respectively.

Table 1 presents the measured and averaged values of the sample entropy for each series of samples, the averaged deviations from the value of entropy $H(X_1, \ldots, X_N) = 1$ and the scatterings around the averaged deviations.

Table 1. Sample entropy $H_R(X_1, \ldots, X_N) \mid n)$ in form 1 – (deviation) · (scattering of deviation)

N	n = 10 MB	n = 100 MB	n = 1 GB
1	$1 - 8.60 \cdot 10^{-9} \cdot$ $\cdot (0.5 \div 2)$	$1 - 8.60 \cdot 10^{-10} \cdot$ $\cdot (0.5 \div 2)$	$1 - 8.60 \cdot 10^{-11} \cdot$ $\cdot (0.5 \div 2)$
2	$1 - 2.58 \cdot 10^{-8} \cdot$ $\cdot (0.25 \div 1.75)$	$1 - 2.58 \cdot 10^{-9} \cdot$ $\cdot (0.25 \div 1.75)$	$1 - 2.58 \cdot 10^{-10} \cdot$ $\cdot (0.25 \div 1.75)$
8	$1 - 2.19 \cdot 10^{-6} \cdot$ $\cdot (0.8 \div 1.2)$	$1 - 2.19 \cdot 10^{-7} \cdot$ $\cdot (0.8 \div 1.2)$	$1 - 2.19 \cdot 10^{-8} \cdot$ $\cdot (0.8 \div 1.2)$
16	$1 - 5.64 \cdot 10^{-4} \cdot$ $\cdot (0.98 \div 1.02)$	$1 - 5.64 \cdot 10^{-5} \cdot$ $\cdot (0.98 \div 1.02)$	$1 - 5.64 \cdot 10^{-6} \cdot$ $\cdot (0.98 \div 1.02)$

In all cases 1 MB = 8 · 1048576 bit, ..., 1 GB = 1000 MB.

We can observe the following phenomena.

– For increasing size n of samples, the differences $1 - H_R(X_1, \ldots, X_N) \mid n)$ get smaller, but always the sample entropy $H_R(X_1, \ldots, X_N) \mid n) < 1$.

– For increasing dimension N of the random variable $(X_1, \ldots, X_N)$ the differences $1 - H_R(X_1, \ldots, X_N) \mid n)$ are greater, but the scatterings around the averaged values are smaller.

The Bernoulli process, binomial distribution and two important measures

We attempt to explain theoretically the experimental results observed above. We take the Bernoulli process as a basic mathematical model, describing random variables with binary values. It is characterized by the binomial distribution with probabilities

$$P(\boldsymbol{n}, \boldsymbol{k} = k) = \binom{n}{k} P(0)^k P(1)^{n-k}, \tag{3}$$

where $\boldsymbol{k}$ denotes the random variable of the binary value 0, k the number of zeros in the sample, n the size of the sample, $P(0)$ the probability of the binary value 0 occurring in the sample, $P(1)$ the probability of occurring the binary value 1.

Two important measures characterize the Bernoulli process.

– The expected value $E(\boldsymbol{k})$ of the random variable $\boldsymbol{k}$ (it is a measure of statistical convergence, or concentration). $E(\boldsymbol{k}) = nP(0)$ and for the random variable $\boldsymbol{n} = \boldsymbol{k} / n$ we have $E(\boldsymbol{n}) = P(0)$.

– The variance V ($\boldsymbol{k}$) of the random variable $\boldsymbol{k}$ (it is a measure of statistical divergence, dispersion). V ($\boldsymbol{k}$) = $nP(0)P(1)$ and for the random variable $\boldsymbol{n} = \boldsymbol{k} / n$ we have V ($\boldsymbol{n}$) = $P(0)P(1) / n$.

However, in the general case it seems practically impossible to find an analytical relation of the above process with the notion of information entropy. We propose to assume a simplified model of information entropy which enables one to relate the Bernoulli process and sample entropy.

The simplification of the information entropy formula

We suppose that the probabilities under consideration have the form

$$P(X_1, ..., X_N) = 1/2^N + \varepsilon_{(X_1,...,X_N)}, \tag{4}$$

Where $\varepsilon_{(X_1,...,X_N)} << 1/2^N$ are the biases of a non-uniform distribution

$$p(X_1, ..., X_N) = \tag{5}$$
$$= \sum_{X_1,...,X_N=0}^{2^N-1} P(X_1, ..., X_N) = \sum_{X_1,...,X_N=0}^{2^N-1} 1/2^N + \varepsilon_{(X_1,...,X_N)}$$

of an *imperfectly random variable* with respect to the uniform distribution of a perfectly random variable with all $P(X_1, \ldots, X_N) = 1/2^N$. The definition of probability $\sum_{X_1,...,X_N=0}^{2^N-1} 1/2^N + \varepsilon_{(X_1,...,X_N)} = 1$ implies that $\sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon_{(X_1,...,X_N)} = 0$.

We use the Maclaurin expansions of the logarithmic function for $0 < x < 1$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{x^m}{m}. \tag{6}$$

Thus, we get the following approximation formulae for information entropy

$$H(X_1, \ldots, X_N) = \tag{7}$$
$$= -\frac{1}{N} \sum_{X_1,...,X_N=0}^{2^N-1} P(X_1, \ldots, X_N) \log_2 P(X_1, \ldots, X_N) =$$
$$= -\frac{1}{N} \sum_{X_1,...,X_N=0}^{2^N-1} (1/2^N + \varepsilon_{(X_1,...,X_N)}) \log_2 (1/2^N + \varepsilon_{(X_1,...,X_N)}) =$$
$$= -\frac{1}{N} \sum_{X_1,...,X_N=0}^{2^N-1} (1/2^N + \varepsilon_{(X_1,...,X_N)}) \cdot$$
$$\cdot (\log_2 1/2^N + \log_2 (1 + 2^N \varepsilon_{(X_1,...,X_N)})) =$$
$$= 1 - \frac{1}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \frac{2^N \varepsilon^2_{(X_1,...,X_N)}}{2} - \frac{2^{2N} \varepsilon^3_{(X_1,...,X_N)}}{6} +$$
$$+ \frac{2^{3N} \varepsilon^4_{(X_1,...,X_N)}}{12} - \frac{2^{4N} \varepsilon^5_{(X_1,...,X_N)}}{20} - ... =$$
$$= 1 - \frac{1}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \sum_{m=1}^{\infty} (-1)^{m+1} \frac{2^{mN} \varepsilon^{m+1}_{(X_1,...,X_N)}}{m(m+1)} =$$
$$= 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^2_{(X_1,...,X_N)} + O(\varepsilon^3_{(X_1,...,X_N)}) \cong$$
$$\cong 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^2_{(X_1,...,X_N)} = H_S(X_1, \ldots, X_N).$$

Now, we see that the information entropy $H_S((X_1, \ldots, X_N)$ in the simplified form (7) can be represented as the unit entropy minus the sum of the squares of all biases $(X_1, \ldots, X_N)$ of the non-uniform distribution.

The errors of order $O(\varepsilon^3_{(X_1,...,X_N)})$, which we omitted in the final formula (7), lead to an inaccuracy of calculating entropy

$$\Delta H(X_1, \ldots, X_N) = H(X_1, \ldots, X_N) - H_S(X_1, \ldots, X_N) \cong \tag{8}$$
$$\cong -\frac{2^{2N}}{6N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^3_{(X_1,...,X_N)} +$$
$$+ \frac{2^{3N}}{12N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^4_{(X_1,...,X_N)} \cong O(\varepsilon^3_{(X_1,...,X_N)}).$$

We must always calculate the both summands in formula (8), because for the distributions with probabilities of the form $P(X_1, \ldots, X_N) = 1/2^N + \varepsilon_{(X_1,...,X_N)} = 1/2^N + \varepsilon$ we have $\sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^3_{(X_1,...,X_N)} = 0$.

Expected entropy

We introduce a new notion of expected entropy which relates the information entropy and Bernoulli process. For a perfectly random variable the expected value is $E(\boldsymbol{n}) = P(0) = 1/2$ and for the N–dimensional random variable the expected value is $E(X_1, \ldots, X_N) = P(X_1, \ldots, X_N) = 1/2^N$, respectively.

Now, let us suppose that the variance V $(X_1, \ldots, X_N)$ of the random variable $(X_1, \ldots, X_N)$ corresponds to the deviation of the relative frequency of this variable observed in a sample sequence from the expected value $E(X_1, \ldots, X_N) = 1/2^N$. Thus, we suppose that in the special case of a one-dimensional random variable the biases satisfy $\varepsilon^2_{(X_1)} = \varepsilon^2_0 = \varepsilon^2_1 = V(\boldsymbol{n}) = P(0)P(1) / n$. In the general case of a N–dimensional random variable this implies

$$\varepsilon^2_{(X_1,...,X_N)} = \tag{9}$$
$$= V(X_1, \ldots, X_N) = P(X_1, \ldots, X_N)\,(1 - P(X_1, \ldots, X_N))\,N / n.$$

We insert $\varepsilon^2_{(X_1,...,X_N)} = V(X_1, \ldots, X_N)$ from (9) to (7) to get the value which we denote

$$H_S(X_1, \ldots, X_N) = 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} V(X_1, \ldots, X_N) = \tag{10}$$
$$= 1 - \frac{2^{N-1}\,(1 - 1/2^N)}{n \ln 2} = 1 - M(N, n) = H_E(X_1, \ldots, X_N) \mid n).$$

We name $H_E(X_1, \ldots, X_N) \mid n)$ the *expected entropy* and the term $M(N, n)$ the *masking component*.

We see that the values of the expected entropy $H_E(X_1, \ldots, X_N) \mid n)$ calculated from (10) are the measured values of the sample entropy $H_R(X_1, \ldots, X_N) \mid n)$ given in Table 1. In practice, the sample entropy for any sample from a binary sequence is always exactly equal to the expected entropy.

We can also explain why the convergence as n tends to infinity for greater N is much slower. We find that the limit is

$$\lim_{n \to \infty} H_S(X_1, \ldots, X_N) = \tag{11}$$

$$= \lim_{n\to\infty} \{ 1 - \frac{2^{N-1}(1-1/2^N)}{n\ln 2} \} = 1 ,$$

which is shown in Figure 1.

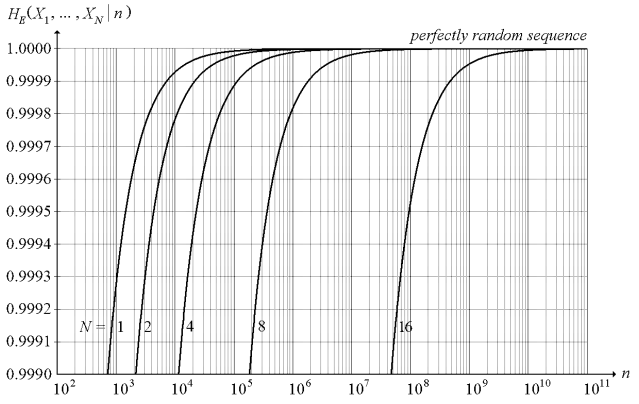

Fig. 1. Convergence of the expected entropy $H_E(X_1, \ldots, X_N) \mid n)$ for perfectly random sequences to the unit entropy for different N as a function of the size n of the sample

We see that a sample of n bits of a perfectly random sequence can achieve the value of unit entropy equal to one when the size n tends to infinity. For sequences of finite size we always get measured values of the sample entropy and expected entropy

(12) $H_R(X_1, \ldots, X_N) \mid n) = H_E(X_1, \ldots, X_N) \mid n) = 1 - M(N, n) < 1 .$

This is not a property of non-randomness of binary sequences. These are the natural features of sample entropy and expected entropy implied by the finiteness of samples of binary sequences and the Law of Large Numbers. The masking component $M(N, n)$ is an uncertainty of each estimator of the entropy $H(X_1, \ldots, X_N)$. This uncertainty constitutes a random error as a non-repeatable inaccuracy caused by a known but uncontrollable factor. The random error cannot be decreased since it depends solely on the variance of the random variable and for a given size of the tested sample always has the same value.

Let us consider the problem of decreasing the scattering of measured values of the sample entropy around the average value for increasing N. Analysis of the experimental results shows that all values of sample entropy are in the interval

(13) $H_R(X_1, \ldots, X_N) \mid n) = 1 - M(N, n)\,(1 \pm \sigma_N) ,$

where

(14) $$\sigma_N \cong \sqrt{\frac{(1-1/2^N)\,N}{2^N}} .$$

The value σ_N cannot be treated as an equivalent of the standard deviation since entropy is not a probability distribution.

The expected entropy of imperfectly random variables

We have considered perfectly random variables with the uniform probabilities $P(X_1, \ldots, X_N) = 1/2^N$ and entropy $H(X_1, \ldots, X_N) = 1$. Now, we calculate the expected entropy for imperfectly random variables.

Let us suppose that a sequence of binary random variables is modeled as a binary Markov chain of the first order (further called a *Markov chain*) which has bias s and correlation K. In this case, the probabilities $P(X_1)$ and $P(X_1 \mid X_2)$ are given by

$$P(0) = 1/2 - s ,$$
$$P(1) = 1/2 + s ,$$

$$P(0|0) = 1/2 - s + 1/2\,K ,$$
$$P(0|1) = 1/2 - s - 1/2\,K ,$$
$$P(1|0) = 1/2 + s - 1/2\,K ,$$
$$P(1|1) = 1/2 + s + 1/2\,K$$

and the probabilities $P(X_1, X_2)$ are equal to

$$P(0,0) = P(0)P(0|0) = 1/4 - s + s^2 + 1/4\,K - 1/2\,sK ,$$
$$P(0,1) = P(0)P(1|0) = 1/4 \quad - s^2 - 1/4\,K + 1/2\,sK ,$$
$$P(1,0) = P(1)P(0|1) = 1/4 \quad - s^2 - 1/4\,K - 1/2\,sK ,$$
$$P(1,1) = P(1)P(1|1) = 1/4 + s + s^2 + 1/4\,K + 1/2\,sK .$$

For $s << 1$ and $K << 1$, we have the simplified formulae

$$P(0,0) \cong 1/4 - s + 1/4\,K ,$$
$$P(0,1) \cong 1/4 \quad - 1/4\,K ,$$
$$P(1,0) \cong 1/4 \quad - 1/4\,K ,$$
$$P(1,1) \cong 1/4 + s + 1/4\,K .$$

All the probabilities above have the form $P(X_1) = 1/2 + \varepsilon_{(X_1)}$ and $P(X_1, X_2) = 1/4 + \varepsilon_{(X_1, X_2)}$, respectively, where in all cases $\varepsilon_{(X_1,\ldots,X_N)} << 1$. We insert the above values of $P(X_1)$ and $P(X_1, X_2)$ to the formula (7) for entropy to obtain

(15) $$H_S(X_1) \cong \frac{1}{2\ln 2}\,(4s^2) ,$$

(16) $$H_S(X_1, X_2) \cong \frac{1}{2\ln 2}(4s^2 + 1/2\,K^2) .$$

Information theory [3, 4, 5] tells us that the conditional entropy $H(X_2 \mid X_1)$ of two neighbouring random variables X_1 and X_2 in a Markov chain satisfies

(17) $$H(X_2 \mid X_1) = 2\,H(X_2, X_1) - H(X_1) .$$

For the Markov chain considered above, the value of the conditional entropy of the random variables in (17) is given by the relation

(18) $$H_S(X_2 \mid X_1) \cong 1 - \frac{1}{2\ln 2}(4s^2 + K^2) .$$

It is known that the entropy of the random variables in an N–dimensional random variable $(X_1, \ldots, X_N)$ for a Markov chain satisfies

(19) $$H(X_1, \ldots, X_N) = \frac{H(X_1)}{N} + \frac{N-1}{N}\,H(X_2 \mid X_1) .$$

In turn, (15) - (19) imply

(20) $$H_S(X_1, \ldots, X_N) \cong 1 - \frac{1}{2\ln 2}(4s^2 + \frac{N-1}{N}\,K^2) .$$

It is known that for independent random variables $(X_1, \ldots, X_N)$ the entropy $H(X_1, \ldots, X_N) = const.$, but for dependent random variables the entropies for various dimensions usually are different values. The most representative fact is

(21) $$\lim_{N\to\infty} H(X_1, \ldots, X_N) = H(X_2 \mid X_1) ,$$

since for $N \to \infty$ the formula (21) represents interdependence between all random variables and the entropy attains a minimal value. In our case we have

(22) $$\lim_{N\to\infty} H_S(X_1, \ldots, X_N) = H_S(X_2 \mid X_1) \cong 1 - \frac{1}{2\ln 2}(4s^2 + K^2).$$

To calculate the value of the masking component $M(N,n)$ we have to take into account that the random variables in the Markov chain are dependent and the cumulative variance of the random variable $\boldsymbol{X} = (X_1, \ldots, X_N)$ has the form

(23) $$V(\boldsymbol{X}) = E[(\boldsymbol{X} - E(\boldsymbol{X}))^2] =$$
$$= \sum_{i=1}^{N} V(X_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} E(X_i, X_j) ,$$

hence it is the sum of the variances $V(X_i)$ and covariances $E(X_i, X_j)$.

For a sequence of imperfectly random variables modeled as a Markov chain, after taking into account the masking component $M(N, n)$, the expected entropy is equal to

(24) $$H_E(X_1, \ldots, X_N \mid n) =$$

$$= 1 - \frac{1}{2\ln 2}\left(\frac{2^N(1-1/2^N)}{n}(1+2K) + 4s^2 + \frac{N-1}{N}K^2\right).$$

The term $(1 + 2K)$ is caused by the fact that dependent random variables with correlation $K > 0$ have positive covariances, hence the cumulative variance has the property of overdispersion.

We have the limit

(25) $$\lim_{n\to\infty} H_E(X_1, \ldots, X_N \mid n) \cong 1 - \frac{1}{2\ln 2}\left(4s^2 + \frac{N-1}{N}K^2\right)$$

which is depicted in Figure 2.

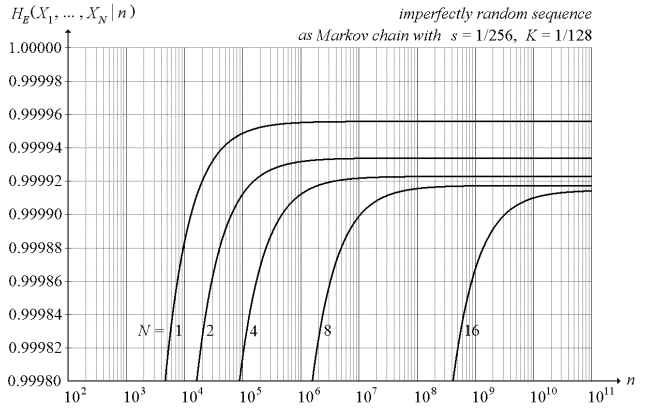

Figure 2. Convergence of expected entropy $H_E(X_1, \ldots, X_N \mid n)$ for imperfectly random sequences to non-unit entropy for different N as a function of the size n of the sample

We indicate some consequences of this analysis. The mutual information for a Markov chain is defined as

(26) $$I(X_2; X_1) = H(X_2) + H(X_1) - 2H(X_1, X_2) = H(X_2) - H(X_2 \mid X_1)$$

and (15) - (16) or (15) and (18) give

(27) $$I_S(X_2; X_1) \cong \frac{K^2}{2\ln 2} .$$

We see that the mutual information of the Markov chain with bias s and correlation K depends only on the correlation between neighbouring random variables in the chain, but it does not depend on the bias.

The second experiment

Let us suppose that we have samples of random sequences modeled as a Markov chain with bias s = 1/256 and correlation K = 1/128 (such samples can be produced by inserting in a perfectly random sequence a controlled non-deterministic bias and correlation as "errors of non-randomness"). Let us measure and average the sample entropy for 10 samples of such sequences and compare them with the theoretical values of expected entropy.

Table 2. Expected and sample entropy for samples of size n = 10 MB

Components of expected entropy	Calculated expected entropy	Measured sample entropy
unit entropy masking bias correlation	$H_E(X_1) =$ 1 $- 8.73\cdot10^{-9}$ $- 4.40\cdot10^{-5}$ $- 0 \cdot 4.40\cdot10^{-5} \cong$ $1 - 4.40\cdot10^{-5}$	$H_R(X_1) =$ $1 - 4.54\cdot10^{-5}$
unit entropy	$H_E(X_1, X_2) =$ 1	$H_R(X_1, X_2) =$
masking bias correlation	$- 2.62\cdot10^{-8}$ $- 4.40\cdot10^{-5}$ $- 1/2 \cdot 4.40\cdot10^{-5} \cong$ $1 - 6.60\cdot10^{-5}$	 $1 - 6.78\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_8) =$ $= 1 -$ $- 2.22\cdot10^{-6} -$ $- 4.40\cdot10^{-5} -$ $- 7/8 \cdot 4.40\cdot10^{-5} \cong$ $\cong 1 - 8.47\cdot10^{-5}$	$H_R(X_1, \ldots, X_8) =$ $= 1 - 8.62\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_{16}) =$ $= 1 -$ $- 5.73\cdot10^{-4} -$ $- 4.40\cdot10^{-5} -$ $- 15/16 \cdot 4.40\cdot10^{-5} \cong$ $\cong 1 - 65.8\cdot10^{-5}$	$H_R(X_1, \ldots, X_{16}) =$ $= 1 - 65.9\cdot10^{-5}$

Table 3. Expected and sample entropy for samples of size n = 100 MB

Components of expected entropy	Calculated expected entropy	Measured sample entropy
unit entropy masking bias correlation	$H_E(X_1) =$ 1 $- 8.73\cdot10^{-10}$ $- 4.40\cdot10^{-5}$ $- 0 \cdot 4.40\cdot10^{-5} \cong$ $1 - 4.40\cdot10^{-5}$	$H_R(X_1) =$ $1 - 4.45\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, X_2) =$ 1 $- 2.62\cdot10^{-9}$ $- 4.40\cdot10^{-5}$ $- 1/2 \cdot 4.40\cdot10^{-5} \cong$ $1 - 6.60\cdot10^{-5}$	$H_R(X_1, X_2) =$ $1 - 6.75\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_8) =$ 1 $- 2.22\cdot10^{-7}$ $- 4.40\cdot10^{-5}$ $- 7/8 \cdot 4.40\cdot10^{-5} \cong$ $1 - 8.27\cdot10^{-5}$	$H_R(X_1, \ldots, X_8) =$ $1 - 8.47\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_{16}) =$ 1 $- 5.73\cdot10^{-5}$ $- 4.40\cdot10^{-5}$ $- 15/16 \cdot 4.40\cdot10^{-5} \cong$ $1 - 14.3\cdot10^{-5}$	$H_R(X_1, \ldots, X_{16}) =$ $1 - 14.5\cdot10^{-5}$

Table 4. Expected and sample entropy for samples of size n = 1 GB

Components of expected entropy	Calculated expected entropy	Measured sample entropy
unit entropy masking bias correlation	$H_R(X_1) =$ 1 $- 8.73\cdot10^{-11}$ $- 4.40\cdot10^{-5}$ $- 0 \cdot 4.40\cdot10^{-5} \cong$ $1 - 4.40\cdot10^{-5}$	$H_R(X_1) =$ $1 - 4.47\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, X_2) =$ 1 $- 2.62\cdot10^{-10}$ $- 4.40\cdot10^{-5}$ $- 1/2 \cdot 4.40\cdot10^{-5} \cong$ $1 - 6.60\cdot10^{-5}$	$H_R(X_1, X_2) =$ $1 - 6.67\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_8) =$ 1 $- 2.22\cdot10^{-8}$ $- 4.40\cdot10^{-5}$ $- 7/8 \cdot 4.40\cdot10^{-5} \cong$ $1 - 8.25\cdot10^{-5}$	$H_R(X_1, \ldots, X_8) =$ $1 - 8.38\cdot10^{-5}$
unit entropy masking bias correlation	$H_E(X_1, \ldots, X_{16}) =$ 1 $- 5.73\cdot10^{-6}$ $- 4.40\cdot10^{-5}$ $- 15/16 \cdot 4.40\cdot10^{-5} \cong$ $1 - 9.10\cdot10^{-5}$	$H_R(X_1, \ldots, X_{16}) =$ $1 - 9.24\cdot10^{-5}$

We make the following observations.

– In all the cases the results of measuring of sample entropy confirm exactly the calculations which take into account the individual components of expected entropy.

– The masking component for imperfectly random sequences has the same properties as for perfectly random sequences, but for the value of K = 1/128 the influence of the factor $(1 + 2K)$ is negligible.
– The measured and averaged values of sample entropy are somewhat greater than those of expected entropy; this follows from the fact that terms of order $O(\varepsilon^3_{(X_1,\ldots,X_N)})$ were disregarded.

Expected entropy as a criterion of randomness of samples of binary sequences

The question is whether the equality (12) for the perfectly random sequences, i.e.:

$$H_R(X_1, \ldots, X_N) \mid n) = H_E(X_1, \ldots, X_N) \mid n) \tag{28}$$

implies a practical criterion of randomness of samples of binary sequences.

Let us suppose that we have a sample from a presumably imperfectly random sequence with unknown entropy $H(X_1, \ldots, X_N) < 1$. In this case, as long as we cannot check experimentally that sample entropy is not much smaller than the expected entropy

$$H_R(X_1, \ldots, X_N) \mid n) < H_E(X_1, \ldots, X_N) \mid n) \tag{29}$$

it is not possible to refute the hypothesis that it is a sample of a perfectly random sequence. To prove the inequality (29), taking into account the scatterings $(1 + \sigma_N)$ and $(1 - \sigma_N)$, we write

$$(1 + \sigma_N) \frac{2^N (1\text{-}1/2^N)}{n} < \tag{30}$$

$$< (1 - \sigma_N) \frac{2^N (1\text{-}1/2^N)}{n} + 4s^2 + \frac{N\text{-}1}{N} K^2 .$$

Hence we have to investigate a sample whose size satisfies

$$n(N)_{MIN} > \frac{2^{N/2\text{-}1}\sqrt{N(1\text{-}1/2^N)^3}}{4s^2 + \frac{N-1}{N}K^2} \tag{31}$$

and calculate the sample entropy $H_R(X_1, \ldots, X_N) \mid n)$ for a sample of size $n(N)_{MIN}$. After taking into account the scattering, the sample entropy $H_R(X_1, \ldots, X_N) \mid n)$ is comparable with the expected entropy $H_E(X_1, \ldots, X_N) \mid n)$ of a perfectly random sequence. It is necessary to test at least 3 samples with possibly large sizes to confirm their randomness.

A practical examples of the application of the criterion of randomness

Now, we present our criterion of randomness using simple examples. We assume that samples from binary sequences to test are generated by a physical random number generator which produces a sequence with the bit rate BR = 8 Mbit/s. Let us remark that it does not make sense to consider the entropy $H(X_1)$ for one-dimensional random variables since then there is no correlation. Investigation of N–dimensional random variables for $N \geq 3$ requires generating of samples of great size. Since we suppose that the random sequence under study is modeled as a first order Markov chain, it is enough to investigate a two-dimensional random variable and the corresponding entropy $H(X_1, X_2)$.

We propose to use as a measure of randomness of a binary sequence, and consequently of the random generator, the minimal time of sample generation

$$T_{MIN} > n(N)_{MIN} / BR . \tag{32}$$

Let us consider a random sequence modeled as a Markov chain with bias $s = 5 \cdot 10^{-3}$ and correlation $K = 5 \cdot 10^{-3}$. These are typical parameters for a physical random number generator with avalanche diode as a source of randomness [6]. In practice, these values cannot be significantly decreased and it is not possible to create a real source of randomness with smaller biases and correlations. For N = 2, we have $n(N)_{MIN}$ > 2000 bits, hence T_{MIN} > 22 μs. For greater size n = 10 MB we have $H_R(X_1, X_2 \mid n) = 1 - 8.12 \cdot 10^{-5}$ in the direction of the value $H_E(X_1, X_2 \mid n) = 1 - 2.58 \cdot 10^{-8}$, so the difference is unquestionable. In fact, it is a bad result and we can refute the hypothesis of perfect randomness of a sample of such a sequence. However, this does not mean that this sequence cannot be used to create a perfectly random one.

We can generate M sequences and *XOR* them. For the resulting sequence the bias and correlation decrease in proportion to $s_\oplus(M) = 1/2\ (2s)^M$ and $K_\oplus(M) = K^M$, respectively. These values are available only after non-deterministic post-processing, which is the mixing of *XOR* operations on *independent* binary random sequences. The resulting sequence inherits all properties from that modeled on a Markov chain of first order and the corresponding parameters change only to be $s_\oplus(M) << s$ and $K_\oplus(M) << K$. The expected entropy for $s_\oplus(M) = 1/2\ (2s)^M$ and $K_\oplus(M) = K^M$ is equal to

$$H_E(X_1, \ldots, X_N \mid n) \cong \tag{33}$$

$$\cong 1 - \frac{1}{2\ln 2}\left(\frac{2^N(1\text{-}1/2^N)}{n}(1 + 2K^M) + (2s)^{2M} + \frac{N\text{-}1}{N}K^{2M}\right) .$$

For a sequence having the above values of parameters we must take its size to be equal

$$n_\oplus(N,M)_{MIN} > \frac{2^{N/2\text{-}1}\sqrt{N(1\text{-}1/2^N)^3}}{(2s)^{2M} + \frac{N-1}{N}K^{2M}} . \tag{34}$$

If we take M = 8 sequences, then for $s = 10^{-2}$ and $K = 10^{-2}$ we have $s_\oplus(N{=}2, M{=}8) = 1.28 \cdot 10^{-14}$, $K_\oplus(N{=}2, M{=}8) = 10^{-16}$, $n_\oplus(N{=}2, M{=}8)_{MIN} > 1.4 \cdot 10^{27}$ and $T_{MIN} > 5.55 \cdot 10^{12}$ years, respectively. We see that the last result does not never enables one to refute the hypothesis of non-randomness of such a sequence.

Summary

We introduced the notion of expected entropy and presented a criterion of randomness and corresponding characteristics of binary sequences produced by a physical random number generator. The theoretical analysis was confirmed by experimental results. The next problems to consider for stationary and ergodic binary random variables modeled as binary Markov chains of the first order are equivalence of probability distributions and isomorphism in the measure-theoretic sense (equivalence of entropies of random variables).

REFERENCES

[1] Shannon, C. E., A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27 (1948), 379–423, 623–656.
[2] Shannon, C. E., Communication Theory of Secrecy Systems, *The Bell System Technical Journal*, 28 (1949), 656–715.
[3] Cover T.M. & Thomas J.A., Elements of Information Theory, *John Wiley and Sons*, (2006), 2nd ed.
[4] Gray R. M., Entropy and Information Theory, *Springer-Verlag*, (2010), 2nd ed.
[5] Seidler J., Nauka o informacji (Information Theory), *Wydawnictwa Naukowo-Techniczne*, (1983). (In Polish.)
[6] Leśniewicz M., Sprzętowa generacja losowych ciągów binarnych (Hardware generation of binary random sequences), *Wydawnictwo Wojskowej Akademii Technicznej (Military University of Technology)*, (2009). (In Polish.)

***Author**: dr inż. Marek Leśniewicz, Wojskowy Instytut Łączności, Zakład Kryptologii, ul. Warszawska 22A, 05-130 Zegrze, E-mail: m.lesniewicz@wil.waw.pl , marek.lesniewicz@op.pl*