Marek LEŚNIEWICZ

Military Institute of Communication

doi:10.15199/48.2016.11.64

Analyses and Measurements of Hardware Generated Random Binary Sequences Modeled as Markov Chains

Abstract*. The paper presents a new approach to modelling of binary random sequences, where versatile models that reflect properties of these sequences are revealed and discussed. In particular, the analysis covers the problems of stationarity and ergodicity of random sequences, the forms of their multi-dimensional distributions of probability and the essence of their isomorphism. The theoretical analyzes have been verified by measuring experiments.*

Streszczenie. *W artykule przedstawiono nowe podejście do modelowania losowych ciągów binarnych. Wszechstronnie przeanalizowano problemy ich stacjonarności i ergodyczności, postaci wielowymiarowych rozkładów prawdopodobieństw i istotę izomorfizmu. Analizy teoretyczne zostały potwierdzone doświadczalnie.* (***Analizy i pomiary sprzętowo generowanych losowych ciągów binarnych modelowanych jako łańcuchy Markowa.***)

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

Introduction

Investigations and evaluation of physically generated random binary sequences can be carried out by means of various methods.

The first approach that is rather historical and nowadays is commonly put in question is the method of investigations by means of a set of statistic tests [1]. The method assumes that the sample sequence is examined *a posteriori,* i.e. upon having the sequence generated with no consideration to the mechanism of generation, thus the properties and parameters of the sequence are *a priori* unknown. The method is merely sufficient to check whether of not the examined sequence meet requirements of the specific statistical test.

The second method that is described in the associated paper [2] consists in verification how much the sample sequence subjects to the random pattern of distribution, where the verification is carried out on the basis of the *expected entropy* that is theoretically determined beforehand and the *sample entropy* calculated from practical measurements. The method makes it possible to *a priori* predict the expected entropy and then calculate the sample entropy on the basis of *a posteriori* measurements. Next, the both entropies are mutually compared to find out whether the examined sample sequence meets the *entropy criterion of randomness*.

The third approach, the most recent one and considered as the most reliable is the method where the primary statistical properties of the sequence are proved prior to its generation since these properties result from analysis of the mechanism that is applied to generation of that sequence. Such sequences are referred to as implementations with the *a priori complete statistical information,* but with unknown and unpredictable values of subsequent elements within a sequence made up of zeros and ones. Obviously, such properties can be also confirmed *a posteriori* by means of statistical analysis and should be in line with the assumed statistics according to the *Laws of Large Numbers*.

In a general case the interest is focused on all significant properties and criteria for randomness of the sequence since such properties and criteria are explicitly entailed by primary mechanisms applied to the sequence generation. The analysis of those properties is only possible when various branches of mathematics are applied, i.e. – theory of probability, theory of information and theory of dynamical systems.

Firstly, from the viewpoint of the theory of probability, the basic criteria of randomness are *stationarity* and *ergodicity* as well as the *stochastic equivalence of probability distributions*.

Secondly, in terms of the theory of information and theory of dynamical systems, the key criterion is *equivalence of entropy* for various sequences and their mutual *isomorphism in the measure-theoretic sense*. Since entropy is the function of probabilities, the equivalence of entropy is a secondary criterion. However in practice, in particular for multi-dimensional random variables, it is more convenient to deal with the single-dimensional entropy.

These studies assume that random binary sequences are generated with use of a hardware generator with an *avalanche diode* as a source of randomness. The analyses shall be based on the model of these sequences where the actual sequences shall be considered as binary Markov chains of the first order (further referred to as *Markov chains*). Such an assumption results from the fact that the adopted model is a sufficiently accurate reproduction of properties and parameters for random sequences supplied by the aforementioned generator, the analysis of such sequences is pretty easily and results of the analysis can be construed in a uncomplicated manner.

Stationarity and ergodicity

Let us initially consider stationarity and ergodicity of a random binary sequence being modeled as a Markov chain. Let us also assume that the sequence has the bias s and the correlation K, whilst the sequence elements are designated by means of subsequent numbers n, n+1, etc. If the bias and correlations are *constant* and *independent* on the element number, the sequence is referred to as *homogeneous*.

Such a chain is described by *conditional probabilities* $P(X|Y)$, that can be summarized into *matrix of transition probabilities* (simply – *transition matrix*) as

(1)
$$\mathbf{M}=\begin{bmatrix} P(0\,|\,0) & P(1\,|\,0) \\ P(0\,|\,1) & P(1\,|\,1) \end{bmatrix} = \begin{bmatrix} 1/2-s+1/2K & 1/2+s-1/2K \\ 1/2-s-1/2K & 1/2+s+1/2K \end{bmatrix}.$$

If for all $P(X|Y)$ the conditions $0 \le P(X|Y) \le 1$ and $P(0|0) + P(1|0) = P(0|1) + P(1|1) = 1$ are fulfilled, the matrix $\mathbf{M}$ is referred to as the *stochastic matrix*.

The probabilities for subsequent elements of the sequence can be described as the *total probabilities*

(2) $$P(0)_{(n+1)} = P(0)_{(n)}\, P(0|0) + P(1)_{(n)}\, P(0|1)$$

and

(3) $$P(1)_{(n+1)} = P(0)_{(n)}\, P(1|0) + P(1)_{(n)}\, P(1|1)$$

Under the assumption that the *vectors of probabilities* $P(0)_{(n)}$ and $P(1)_{(n)}$ are defined as the matrix $\mathbf{P}_{(n)} = [\ P(0)_{(n)}\ P(1)_{(n)}\ ]$, the following notation is true

(4) $$\mathbf{P}_{(n+1)} = \mathbf{P}_{(n)}\,\mathbf{M}\ ,$$

and for subsequent m elements

(5) $$\mathbf{P}_{(n+m)} = \mathbf{P}_{(n)}\,\mathbf{M}^m\ .$$

In the general case the notation leads to the *Chapman-Kolmogorov equation* in its matrix version

(6) $$\mathbf{P}_{(n+m)} = \mathbf{P}_{(n+i+j)} = \mathbf{P}_{(n)}\,\mathbf{M}^m = \mathbf{P}_{(n)}\,\mathbf{M}^{i+j} = \mathbf{P}_{(n)}\,\mathbf{M}^i\,\mathbf{M}^j\ .$$

Then the question arises what the form of the $\mathbf{M}^n$ matrix should be if the matrix $\mathbf{M}$ is defined as in (1). The literature references provide various forms of the $\mathbf{M}^n$ matrix, usually with really sophisticated and useless forms. By use of the spectral theory of matrices one can bring the $\mathbf{M}^n$ matrix to the *spectral matrix* in form [3]

(7) $$\mathbf{M}^n = \begin{bmatrix} \pi(0) & \pi(1) \\ \pi(0) & \pi(1) \end{bmatrix} + K^n \begin{bmatrix} \pi(1) & -\pi(1) \\ -\pi(0) & \pi(0) \end{bmatrix}$$

where

(8) $$\pi(0) = \frac{P(0\,|\,1)}{P(0\,|\,1) + P(1\,|\,0)} = 1/2 - \frac{s}{1-K}$$

and

(9) $$\pi(1) = \frac{P(1\,|\,0)}{P(0\,|\,1) + P(1\,|\,0)} = 1/2 + \frac{s}{1-K}\ .$$

The $\mathbf{M}^n$ matrix is also a stochastic one and has two single *characteristic roots* that make up its *spectrum*. The first root $\varphi_1 = 1$ referred to as the *spectral radius*, is equal to 1, thus it formally indicates the *ergodice and irreducible properties* of the matrix. The second root $\varphi_2 = 1 - P(0|1) - P(1|0) < 1$ shows of the *non-cyclic* property of the matrix.

The $\mathbf{M}^n$ matrix meets the convergence provision

(10) $$\lim_{n\to\infty} \mathbf{M}^n = \mathbf{M}^\infty = \begin{bmatrix} \pi(0) & \pi(1) \\ \pi(0) & \pi(1) \end{bmatrix}.$$

Due to the foregoing property the $\mathbf{M}^\infty$ is referred to as the *ergodic matrix* with its $\pi(0)$ and $\pi(0)$ elements that are called *ergodic probabilities* whilst the Markov chain itself is an *ergodic chain*, i.e. demonstrating the *ergodic property*.

Let us ask the question what the vector of probabilities must be to satisfy the equation

(11) $$\mathbf{P} = \mathbf{P}\,\mathbf{M}\ .$$

It turns out that there is only one $\mathbf{P}$ vector that fits the foregoing equation. The vector is

(12) $$\mathbf{P} = [\ \pi(0)\ \ \pi(1)\ ]$$

and is referred to as the *stationary distribution* whilst its terms are equal to terms in rows of the ergodic matrix $\mathbf{M}^\infty$, i.e. the $\pi(0)$ and $\pi(1)$ ergodic probabilities.

By substitution of the relationship (7) to the equation (5) the following provision is fulfilled for any n and m

(13) $$\mathbf{P}_{(n+m)} = \mathbf{P}_{(n)}\,\mathbf{M}^m = \\ = [\ P(0)_{(n+m)}\ \ P(1)_{(n+m)}\ ] = \\ = [\ \pi(0) + (\ P(0)_{(n)} - \pi(0)\ )\,K^m\ \ \ \pi(1) + (\ P(1)_{(n)} - \pi(1)\ )\,K^m\ ]\ .$$

If $P(0)_{(n)} = \pi(0)$ and $P(1)_{(n)} = \pi(1)$ the $\mathbf{P}$ vector immediately adopts the form (12). But anyway, even if $P(0)_{(n)} \neq \pi(0)$ and $P(1)_{(n)} \neq \pi(1)$, the $\mathbf{P}$ vector is quickly brought to the form (12) due to the K^m factor. It is easy to notice that whenever the $\mathbf{P}$ adopts the form (12), the for shall be then *reproduced* for subsequent elements in the unaltered form, because

(14) $$\mathbf{P}_{(m)} = [\ \pi(0)\ \ \pi(1)\ ]\,\mathbf{M}^m = [\ \pi(0)\ \ \pi(1)\ ] = \mathbf{P}$$

for any m. It is why the Markov chain can be referred to as the *stationary chain* i.e. demonstrating the *stationary property*.

It can be also noticed that the stationary property is the secondary feature that can be derived from ergodicity since each ergodic sequence is also a stationary one, which also results from the foregoing analysis.

Such stationary properties is only the *stationarity in weak-sense*. In the subsequent part of this paper also the *stationarity in strict-sense* shall be the subject of further deliberations.

Also the conditions for ergodicity can be considered in the more detailed sense since other interesting understanding of ergodicity also exist, i.e. *geometric ergodicity* and *uniform ergodicity*.

Each *homogeneous*, *irreducible* and *non-cyclic* Markov chain is *ergodic in the geometric sense* when the following provision is fulfilled

(15) $$\|\ \mathbf{P}_{(n)} - \mathbf{P}\ \|_{\mathrm{tv}} = \frac{1}{2}\sum_{i=0}^{1} |\ P(X_i)_{(n)} - \pi(X_i)\ | = \\ = \frac{1}{2}\{\,|\ P(0)_{(n)} - \pi(0)\ | + |\ P(1)_{(n)} - \pi(1)\ |\,\} \le \mathrm{C}(0)\ \rho^n\ ,$$

where

$\|\cdot\|_{\mathrm{tv}}$ is the measure of the *variation distance* in the sense of an extremum for the *total variation norm,* where the distance is determined between the probabilistic vector $\mathbf{P}_{(n)}$ for the n^{th} element and the stationary probabilistic vector of $\mathbf{P}$;

$\mathrm{C}(0) < \infty$ is the constant that depends on the vector of initial conditions, i.e. the probabilistic vector of $\mathbf{P}_{(0)}$;

$\rho < 1$ is the constant that corresponds to the second characteristic root of the matrix $\mathbf{M}$, i.e. $\rho = \varphi_2 = 1 - P(0|1) - P(1|0) = K < 1$.

Substitution of the probabilistic vector for the n element (13) in the form of

(16) $$\mathbf{P}_{(n)} = \mathbf{P}_{(0)}\,\mathbf{M}^n = \\ = [\ P(0)_{(n)}\ \ P(1)_{(n)}\ ] = \\ = [\ \pi(0) + (\ P(0)_{(0)} - \pi(0)\ )\,K^n\ \ \ \pi(1) + (\ P(1)_{(0)} - \pi(1)\ )\,K^n\ ]\ ,$$

to (15) leads to the following equation

(17) $$\|\ \mathbf{P}_{(n)} - \mathbf{P}\ \|_{\mathrm{tv}} = \\ = \frac{1}{2}\{\,|\ P(0)_{(0)} - \pi(0)\ | + |\ P(1)_{(0)} - \pi(1)\ |\,\}\,K^n \le \mathrm{C}(0)\ \rho^n\ ,$$

where $\mathrm{C}(0) = 1/2\ \{\ |\ P(0)_{(0)} - \pi(0)\ | + |\ P(1)_{(0)} - \pi(1)\ |\ \} \le 1$ and $\rho^n = K^n$.

As one can see, the Markov chain in question is *ergodic* in the *geometric* sense.

However, the foregoing provision also demonstrate that the Markov chain under consideration *is not uniformly ergodic* since it fails to fulfill the following, more strict provision that requires from the constant $\mathrm{C} < \infty$ to be *independent on initial conditions*

(18) $$\|\ \mathbf{P}^n - \mathbf{P}\ \|_{\mathrm{tv}} \le \mathrm{C}\ \rho^n\ .$$

However, it is possible to find out that the *uniform ergodicity* is characteristic for *stationary Markov chains* since for $P(0)_{(n)} = \pi(0)$ and $P(1)_{(n)} = \pi(1)$ the provision $\|\ \mathbf{P}^n - \mathbf{P}\ \|_{\mathrm{tv}} = 0$ is fulfilled.

Obviously, for the both cases the convergence exists for the function of the number of elements

(19) $$\lim_{n\to\infty} \|\ \mathbf{P}^n - \mathbf{P}\ \|_{\mathrm{tv}} = 0,$$

since $C \le 1$, $\pi(X) \approx 1/2$, the $\rho^m = K^m$ factor very quickly brings the limit to zero. Sometimes the foregoing convergence property is understood as the general term of ergodicity.

The problem of ergodicity is considered not only from the viewpoint of the ergodic properties themselves but also as the provision for other properties to come true, e.g. *the convergence of the Markov chain to the normal density* (the central limit theorem (CLT)) [4]. It turns out that such a convergence takes place when the chain demonstrates the property of geometric ergodicity [4]. Beside proving the fact of convergence it is also possible to calculate the *expected value* $E(\boldsymbol{n})$ and the *variance* $V(\boldsymbol{n})$ for the Gaussian distribution of the random variable $\boldsymbol{n} = \boldsymbol{k} / n$, and these distribution parameters depend on the ergodic probabilities $\pi(0)$ and $\pi(1)$, the K correlation and the number of elements n of the sequence. For the first approximation these parameters are the following [3]

$$E(\boldsymbol{n}) = \pi(0) + \frac{P(0)_{(0)} - \pi(0)}{n} \frac{1-K^n}{1-K} \tag{20}$$

and

$$V(\boldsymbol{n}) \cong \frac{\pi(0)\pi(1)}{n} \frac{1+K}{1-K} , \tag{21}$$

where $\boldsymbol{k}$ stands for the random variable of the binary value 0, k the number of zeros in the sample, n the size of the sample and $\boldsymbol{n} = \boldsymbol{k} / n$.

It is easily to see that for a stationary Markov chain, i.e. $P(0)_{(0)} = \pi(0)$ the mathematical expectation is immediately $E(\boldsymbol{n}) = \pi(0)$ whilst the variance $V(\boldsymbol{n})$ has the property of *overdispersion* [2]. Regardless the fact that it is the simplest case of a one-dimensional random variable the analysis of the case is really sophisticated and extends scope of this paper. For multi-dimensional random variables any satisfactory analytic solutions are actually non existent and the expected values $E(X_1, ..., X_N)$ and variances $V(X_1, ..., X_N)$ can be calculated exclusively by means of estimation methods.

To recapitulate the foregoing, the sequence under consideration shall demonstrate all the foregoing properties if such a sequence is defined by the stochastic matrix $\mathbf{M}$ in the form (1). Obviously, it is infeasible to verify the foregoing analyses by experiments since no probability can be measured *a priori* and only properties of implementations are measurable as relative frequencies [2] that are inapplicable for direct verification whether the foregoing analyses are true. However, the correctness of them can be verified on the indirect way. Initially – by measurements of the bias s and corellation K parameters and then by providing the proof that the examined sequence corresponds to the model of the first order Markov chain. The secondary proof is possible by experimental confirmation of stationary and ergodic properties of the chain samples.

Stochastic equivalence of probability distributions

Determination of probability distribution is a sophisticated task, in particular for multi-dimensional random variables. Although it is not difficult to find out distribution for independent random variables since they are characterized exclusively by bias and not encumbered by correlation, the real random variables that are used for modeling of actual random sequences are never independent in practice and their mutual correlations must be also taken into consideration. The difficulty results from the need to find out *joint probabilities* for N–dimensional random variables from the defining relationship

$$P(X_1, ..., X_N) = P(X_N \mid X_1, ..., X_{N-1})\, P(X_1, ..., X_{N-1}) . \tag{24}$$

For this case, even the probabilities for $(N\text{-}1)$–dimensional random variables $P(X_1, ..., X_{N-1})$ are known, no information is usually available for the *conditional probabilities* $P(X_N \mid X_1, ..., X_{N-1})$.

The problem can be simplified when it is reduced to determination of distributions for multi-dimensional random variables that are modeled as first-order Markov chains. Let us benefit from the *Markov property*

$$P(X_N \mid X_1, ..., X_{N-1}) = P(X_N \mid X_{N-1}) . \tag{25}$$

Such an approach makes it possible to rewrite the relationship (24) in the form of a recurrence formula

$$P(X_1, ..., X_N) = P(X_N \mid X_{N-1})\, P(X_{N-1} \mid X_{N-2}) \ldots P(X_2 \mid X_1)\, P(X_1). \tag{26}$$

One to have also to keep on mind that individual terms within the formula (26) are known since they are entries of the stochastic matrix

$$\mathbf{M} = \begin{bmatrix} P(0 \mid 0) & P(1 \mid 0) \\ P(0 \mid 1) & P(1 \mid 1) \end{bmatrix} = \begin{bmatrix} 1/2 - s + 1/2K & 1/2 + s - 1/2K \\ 1/2 - s - 1/2K & 1/2 + s + 1/2K \end{bmatrix}. \tag{27}$$

Calculation of probabilities from the relationship (26) is simple, although quite burdensome [3]. The calculations lead to really vast polynomials but they can be substantially simplified under the assumptions that $s << 1$ and $K << 1$. By rejection of entries with significantly less values and with consideration to the *Kolmogorov axiom* the following equation can be achieved

$$\sum_{X_1,\ldots,X_N=0}^{2^N-1} P(X_1, \ldots, X_N) = 1 , \tag{28}$$

that is correct for N–dimensional distributions:

A. one-dimensional distribution

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

B. two- dimensional distribution

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

C. three-dimensional distribution

$$P(0,0,0) \cong 1/8 - 3/4\, s + 1/4\, K = 1/8\, (1 - 6\, s + 2\, K)$$
$$P(0,0,1) \cong 1/8 - 1/4\, s = 1/8\, (1 - 2\, s)$$
$$P(0,1,0) \cong 1/8 - 1/4\, s - 1/4\, K = 1/8\, (1 - 2\, s - 2\, K)$$
$$P(0,1,1) \cong 1/8 + 1/4\, s = 1/8\, (1 + 2\, s)$$
$$P(1,0,0) \cong 1/8 - 1/4\, s = 1/8\, (1 - 2\, s)$$
$$P(1,0,1) \cong 1/8 + 1/4\, s - 1/4\, K = 1/8\, (1 + 2\, s - 2\, K)$$
$$P(1,1,0) \cong 1/8 + 1/4\, s = 1/8\, (1 + 2\, s)$$
$$P(1,1,1) \cong 1/8 + 3/4\, s + 1/4\, K = 1/8\, (1 + 6\, s + 2\, K)$$

D. four-dimensional distribution

$$P(0,0,0,0) \cong 1/16 - 1/2\, s + 3/16\, K = 1/16\, (1 - 8\, s + 3\, K)$$
$$P(0,0,0,1) \cong 1/16 - 1/4\, s + 1/16\, K = 1/16\, (1 - 4\, s + K)$$
$$P(0,0,1,0) \cong 1/16 - 1/4\, s - 1/16\, K = 1/16\, (1 - 4\, s - K)$$
$$P(0,0,1,1) \cong 1/16 + 1/16\, K = 1/16\, (1 + K)$$
$$P(0,1,0,0) \cong 1/16 - 1/4\, s - 1/16\, K = 1/16\, (1 - 4\, s - K)$$
$$P(0,1,0,1) \cong 1/16 - 3/16\, K = 1/16\, (1 - 3\, K)$$
$$P(0,1,1,0) \cong 1/16 - 1/16\, K = 1/16\, (1 - K)$$
$$P(0,1,1,1) \cong 1/16 + 1/4\, s + 1/16\, K = 1/16\, (1 + 4\, s + K)$$
$$P(1,0,0,0) \cong 1/16 - 1/4\, s + 1/16\, K = 1/16\, (1 - 4\, s + K)$$
$$P(1,0,0,1) \cong 1/16 - 1/16\, K = 1/16\, (1 - K)$$
$$P(1,0,1,0) \cong 1/16 - 3/16\, K = 1/16\, (1 - 3\, K)$$
$$P(1,0,1,1) \cong 1/16 + 1/4\, s - 1/16\, K = 1/16\, (1 + 4\, s - K)$$

$$P(1,1,0,0) \cong 1/16 + 1/16\ \mathrm{K} = 1/16\ (1 + \mathrm{K})$$
$$P(1,1,0,1) \cong 1/16 + 1/4\ \mathrm{s} - 1/16\ \mathrm{K} = 1/16\ (1 + 4\ \mathrm{s} - \mathrm{K})$$
$$P(1,1,1,0) \cong 1/16 + 1/4\ \mathrm{s} + 1/16\ \mathrm{K} = 1/16\ (1 + 4\ \mathrm{s} + \mathrm{K})$$
$$P(1,1,1,1) \cong 1/16 + 1/2\ \mathrm{s} + 3/16\ \mathrm{K} = 1/16\ (1 + 8\ \mathrm{s} + 3\ \mathrm{K})$$

How is has already been noted, determination of distributions for random variables of higher rank is really burdensome. Let us then ask a question whether any recurrent rule can be derived from the foregoing relationships to enable determination of probability distributions for random variables of finite rank but for any, however finite dimension of N? When to consider right-hand sides of all relationships one can spot that beside the $1/2^N$ factor the equations comprise sums of biases and correlations with the values that depend on mutual arrangement of zeros and ones in the structure of the random variable $(X_1, ..., X_N)$.

The detailed analysis of structures for all variables makes it possible to find out that the following recurrence formula is valid for any N–dimensional random variable

$$P(X_1, ..., X_N) = \\ = 1/2^N \{ 1 + 2s [L(1) - L(0)] + \\ + K [L(0,0) + L(1,1) - L(0,1) - L(1,0)] \}, \tag{29}$$

where
$L(1)$ – total number of entries with the value of (1),
$L(0)$ – total number of entries with the value of (0),
$L(0,0)$ – total number of pairs made up by adjacent entries with the same values of (0,0),
$L(1,1)$ – total number of pairs made up by adjacent entries with the same values of (1,1),
$L(0,1)$ – total number of pairs made up by adjacent entries with the opposite values, i.e. (0,1),
$L(1,0)$ – total number of pairs made up by adjacent entries with the opposite values, i.e. (1,0).

It is easily to see that the probabilities described by the relationship (29) fulfill the Kolmogorov axiom (28) by the nature of matters.

Let us recall that the simplified formula for entropy [2] for probabilities defined as $P(X_1, ..., X_N) = 1/2^N + \varepsilon_{(X_1,...,X_N)}$ and $\varepsilon_{(X_1,...,X_N)} << 1$ is following

$$H(X_1, ..., X_N) = \\ = -\frac{1}{N} \sum_{X_1,...,X_N=0}^{2^N-1} P(X_1, ..., X_N)\ \log_2 P(X_1, ..., X_N) \cong \\ \cong 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^2_{(X_1,...,X_N)}\ . \tag{30}$$

The transformation of (29) to the form

$$P(X_1, ..., X_N) = \\ = 1/2^N + 1/2^N \{ 2s [L(1) - L(0)] + \\ + [L(0,0) + L(1,1) - L(0,1) - L(1,0)] \} = \\ = 1/2^N + \varepsilon_{(X_1,...,X_N)}, \tag{31}$$

and the substitution

$$\varepsilon_{(X_1,...,X_N)} = 1/2^N \{ 2s\ [L(1) - L(0)] + [L(0,0) + L(1,1) - L(0,1) - L(1,0)] \}$$

to (30) with simultaneous discarding of negligible elements with their substantially less values leads to the same results of each N as in case when the relationship [2] is applied

$$H(X_1, \ldots, X_N) \cong 1 - \frac{1}{2 \ln 2} (4s^2 + \frac{N-1}{N} K^2) \tag{32}$$

But anyway, calculations of that type are really burdensome. However, it is possible to notice that various approaches may lead to the same results and correctness of these results can be additionally verified by measurements [2].

In practice, the random binary sequences generated with use of hardware means demonstrate poor properties of randomness that are exhibited by large values of bias $s = 5 \cdot 10^{-3}$ and correlation $K = 5 \cdot 10^{-3}$. However, it is demonstrated in [2] that such large values can be reduced on a simple way.

Upon generation of M independent random binary sequences and then these sequences are *XORed*, the bias and correlation for the resulting sequence shall decrease in the proportion to $s_{\oplus}(M) = 1/2\ (2s)^M$ and $K_{\oplus}(M) = K^M$, respectively.

The resulting sequence inherits all properties from the one that is modeled by means of a Markov chain of the first order and only the corresponding parameters subject to changes and become $s_{\oplus}(M) << s$ and $K_{\oplus}(M) << K$.

Hence, the probabilities for any N–dimensional random variable can be rewritten in the form of the recurrence relationship

$$P(X_1, ..., X_N) = \\ = 1/2^N \{ 1 + (-1)^{M+1} (2s)^M [L(1) - L(0)] + \\ + K^M [L(0,0) + L(1,1) - L(0,1) - L(1,0)] \}, \tag{33}$$

where $L(X)$ and $L(X,Y)$ have the same meanings as in case of the (29).

When are known forms of probability distributions for random variables with a whichever dimension of N, where distributions depend only on bias and correlation and are irrelevant to other factors, is true the thesis on the *stationarity in the strict-sense*, i.e. *invariability* of all probability distributions with the finite dimensions.

The experiments

Let us try to confirm correctness of the foregoing analyzes by several experiments.

For needs of our experiments have been used a hardware generator to produce a sample of the size n = 1 GB. Let us suppose that the samples of random binary sequences are available and these samples are modeled as a Markov chain with the bias s = 1/256 and the correlation K = 1/128.

It is then possible to investigate distributions of random variables with any dimension of N but with consideration to the fact that for $N \to \infty$ the probabilities $P(X_1, ..., X_N)$ represent interdependence between all random variables and such a representation is more exhaustive, in contrary to e.g. one-dimensional distribution that takes account only of the bias.

However, the dimension of the random variable must not be too high since the analysis of 2^N points for the distribution may prove too difficult. For instance, when N = 8 the number of points is only 256, but for N = 16 as much as 65536 and the analysis of such huge number of points can be carried out exclusively by means of numerical methods. Moreover, for large dimensions of random variable the accuracy for measurements of relative frequencies is worsened.

Let us then consider how the size of a sample sequence affects accuracy of measurements for relative frequencies. Obviously, the analysis can be based on the variance and the standard deviation but these measures are not practicable. It is much more convenient to employ the module of the *mean relative deviation from the expected value*.

For the sample sequence with the size of n elements and taken for the N–dimensional random variable $\boldsymbol{k}$ the said measure is defined by the theoretical relationship [3]

(34) $$| \alpha_{av\,(N)} | = E(| \boldsymbol{k} - E(\boldsymbol{k}) |)\, N / n = \sqrt{\frac{(1-1/2^N)\, N}{2^{N-1} \pi n}} ,$$

but for the practice of measurements

(35) $$| \alpha_{av\,(N)} | = 1/2^N \sum_{X_1,...,X_N=0}^{2^N-1} | 1/2^N - n(X_1, ..., X_N)\, N / n | .$$

For N = 8 and n = 1 GB we have $| \alpha_{av\,(N=8)} | \cong (16\pi\ (n$ = 1 GB))$^{-1/2}$ = 1.54·10^{-6}, but for N = 16 and n = 1 GB the measure is $|\alpha_{av\,(N=16)}| \cong (2048\ \pi\ (n$ = 1 GB))$^{-1/2}$ = 1.36·10^{-7}.

The expected relative frequencies shall amount to 1/2$^{N=8}$ = 1/256 = 3.90625·10^{-3} and 1/2$^{N=16}$ = 1/65536 ≈ 1.52588·10^{-5}. As one can see it is possible to expect accuracies in the sense $|\alpha_{av\,(N)}|$ / 1/2^N respectively equal to 3.9·10^{-4} and 8.9·10^{-3}. Since the assumption was made that s = 3.90625·10^{-3} and K = 7.8125·10^{-3} the foregoing accuracy would be definitely too low for N = 16. Thus, to investigate 16-dimensional distributions with sufficient accuracy it would be necessary to generate samples with the size of many tens of GB. The generation process itself would take a long time and the calculations would be also extremely time-consuming. Moreover, the practice of measurements confirms that 16-dimensional distributions are merely a recurrence extension of 8-bit distributions and bring no important, additional information.

Thus, it is sufficient to find out relative frequencies $n(X_1, ..., X_8)\ N\ /\ n$ for all 8–element subsequences incorporated into a sample with the size of n = 1 GB (1000 · 8 · 1048576 bits = 8388608000 bits).

Table 1. Relative frequencies $n(X_1, ..., X_N)\ N\ /\ n$ for N = 8 and the sample with the size of n = 1 GB

0: 0.0038793	1: 0.0038847	2: 0.0038191	3: 0.0039392	4: 0.0038216	5: 0.0038198	6: 0.0038813	7: 0.0039968
8: 0.0038196	9: 0.0038214	10: 0.0037624	11: 0.0038808	12: 0.0038779	13: 0.0038790	14: 0.0039396	15: 0.0040615
16: 0.0038174	17: 0.0038205	18: 0.0037623	19: 0.0038778	20: 0.0037598	21: 0.0037592	22: 0.0038167	23: 0.0039367
24: 0.0038792	25: 0.0038791	26: 0.0038185	27: 0.0039342	28: 0.0039392	29: 0.0039362	30: 0.0039970	31: 0.0041168
32: 0.0038212	33: 0.0038225	34: 0.0037607	35: 0.0038790	36: 0.0037619	37: 0.0037605	38: 0.0038183	39: 0.0039375
40: 0.0037593	41: 0.0037604	42: 0.0037016	43: 0.0038149	44: 0.0038218	45: 0.0038121	46: 0.0038739	47: 0.0039944
48: 0.0038800	49: 0.0038789	50: 0.0038158	51: 0.0039384	52: 0.0038194	53: 0.0038167	54: 0.0038756	55: 0.0039962
56: 0.0039400	57: 0.0039359	58: 0.0038715	59: 0.0039952	60: 0.0039948	61: 0.0039941	62: 0.0040539	63: 0.0041779
64: 0.0038191	65: 0.0038228	66: 0.0037615	67: 0.0038771	68: 0.0037632	69: 0.0037631	70: 0.0038172	71: 0.0039354
72: 0.0037600	73: 0.0037573	74: 0.0036976	75: 0.0038112	76: 0.0038177	77: 0.0038139	78: 0.0038757	79: 0.0039959
80: 0.0037605	81: 0.0037588	82: 0.0036967	83: 0.0038132	84: 0.0036954	85: 0.0036903	86: 0.0037515	87: 0.0038681
88: 0.0038206	89: 0.0038146	90: 0.0037522	91: 0.0038737	92: 0.0038751	93: 0.0038714	94: 0.0039334	95: 0.0040535
96: 0.0038785	97: 0.0038799	98: 0.0038196	99: 0.0039349	100: 0.0038221	101: 0.0038196	102: 0.0038758	103: 0.0039923
104: 0.0038202	105: 0.0038140	106: 0.0037528	107: 0.0038720	108: 0.0038707	109: 0.0038736	110: 0.0039342	111: 0.0040523
112: 0.0039397	113: 0.0039355	114: 0.0038769	115: 0.0039929	116: 0.0038770	117: 0.0038697	118: 0.0039341	119: 0.0040518
120: 0.0039946	121: 0.0039963	122: 0.0039306	123: 0.0040520	124: 0.0040596	125: 0.0040520	126: 0.0041157	127: 0.0042361
128: 0.0038832	129: 0.0038824	130: 0.0038187	131: 0.0039400	132: 0.0038238	133: 0.0038207	134: 0.0038787	135: 0.0040013
136: 0.0038203	137: 0.0038215	138: 0.0037580	139: 0.0038758	140: 0.0038818	141: 0.0038753	142: 0.0039379	143: 0.0040570
144: 0.0038199	145: 0.0038179	146: 0.0037576	147: 0.0038713	148: 0.0037591	149: 0.0037552	150: 0.0038189	151: 0.0039339
152: 0.0038788	153: 0.0038791	154: 0.0038138	155: 0.0039331	156: 0.0039389	157: 0.0039338	158: 0.0039922	159: 0.0041161
160: 0.0038204	161: 0.0038196	162: 0.0037611	163: 0.0038778	164: 0.0037584	165: 0.0037549	166: 0.0038144	167: 0.0039318
168: 0.0037598	169: 0.0037560	170: 0.0036931	171: 0.0038123	172: 0.0038135	173: 0.0038101	174: 0.0038732	175: 0.0039899
176: 0.0038781	177: 0.0038755	178: 0.0038145	179: 0.0039340	180: 0.0038148	181: 0.0038086	182: 0.0038688	183: 0.0039893
184: 0.0039370	185: 0.0039308	186: 0.0038688	187: 0.0039873	188: 0.0039960	189: 0.0039869	190: 0.0040521	191: 0.0041722
192: 0.0039369	193: 0.0039423	194: 0.0038807	195: 0.0039966	196: 0.0038781	197: 0.0038762	198: 0.0039348	199: 0.0040605
200: 0.0038762	201: 0.0038750	202: 0.0038119	203: 0.0039311	204: 0.0039375	205: 0.0039347	206: 0.0039991	207: 0.0041171
208: 0.0038797	209: 0.0038787	210: 0.0038179	211: 0.0039346	212: 0.0038116	213: 0.0038082	214: 0.0038686	215: 0.0039935
216: 0.0039372	217: 0.0039373	218: 0.0038682	219: 0.0039930	220: 0.0039952	221: 0.0039894	222: 0.0040530	223: 0.0041758
224: 0.0039969	225: 0.0039990	226: 0.0039356	227: 0.0040584	228: 0.0039359	229: 0.0039333	230: 0.0039912	231: 0.0041151
232: 0.0039346	233: 0.0039323	234: 0.0038697	235: 0.0039938	236: 0.0039944	237: 0.0039883	238: 0.0040515	239: 0.0041705
240: 0.0040575	241: 0.0040604	242: 0.0039957	243: 0.0041194	244: 0.0039942	245: 0.0039921	246: 0.0040480	247: 0.0041721
248: 0.0041181	249: 0.0041144	250: 0.0040497	251: 0.0041727	252: 0.0041774	253: 0.0041779	254: 0.0042343	255: 0.0043646

Form – $(X_1, ..., X_8)_{DECIMAL}$: relative frequency for $(X_1, ..., X_8)$

Let us now estimate probabilities for several selected, characteristic random variables and compare them against the already measured relative frequencies. The process shall be carried out for the assumed values of bias s = 1/256 and correlation K = 1/128.

For the maximum excess of zeros over ones and the maximum correlation between the sequence of zeros the probability is calculated as

(36) $$P(00000000_{BIN}) = \\ = 1/256\ \{\ 1 + 2\ (s = 1/256)\ [\ L(1) - L(0) = -8\ ] + \\ + (K = 1/128)\ [\ L(0,0) + L(1,1) - L(0,1) - L(1,0) = +7\ ]\ \} = \\ = 0.0038757$$

– relative frequency (0$_{DEC}$) = 0.0038793,
– relative difference between the probability and the relative frequency –0.093%

For the maximum excess of ones over zeros and the maximum correlation between the sequence of ones the probability is calculated as

(37) $$P(11111111_{BIN}) = \\ = 1/256\ \{\ 1 + 2\ (s = 1/256)\ [\ L(1) - L(0) = +8\ ] + \\ + (K = 1/128)\ [\ L(0,0) + L(1,1) - L(0,1) - L(1,0) = +7\ ]\ \} = \\ = 0.0043640$$

– relative frequency (255$_{DEC}$) = 0.0043646,
– relative difference between the probability and the relative frequency –0.014%.

For the equality between (1) and (0) and the minimum correlation between the alternated zeros and ones probability amounts to

(38) $$P(01010101_{BIN}) = P(10101010_{BIN}) = \\ = 1/256\ \{\ 1 + 2\ (s = 41/256)\ [\ L(1) - L(0) = 0\ ] + \\ + (K = 1/128)\ [\ L(0,0) + L(1,1) - L(0,1) - L(1,0) = -7\ ]\ \} = \\ = 0.0036926$$

– relative frequency (85$_{DEC}$) = 0.0036903,
– relative frequency (170$_{DEC}$) = 0.0036931,
– relative difference between the probability and the relative frequency is respectively +0.07% and –0.0054%.

The foregoing estimations demonstrate that the relative differences between the calculated probability values and the measured values of relative frequencies are really low and never exceed ±0.1%.

Let us also check the sample entropy and compare it against the expected entropy. Sample entropy can be calculated from the following defining relationship [2]

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

where for N = 8 and n = 1 GB the result is $H_R(X_1, ..., X_8 \mid n$ = 1 GB) = 1 – 8.38·10^{-5}.
Practically equal result can be obtained from the simplified formula for entropy [2]

(40) $$H_R(X_1, ..., X_N) \cong \\ \cong 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} \varepsilon^2_{(X_1,...,X_N)} = \\ = 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1,...,X_N=0}^{2^N-1} [\ 1/2^N - n(X_1, ..., X_N)\, N / n\ ]^2 .$$

Let us check where such values appeared from. One has to keep in mind that the formula for entropy is the following [2]

(41) $$H_S(X_1, ..., X_N \mid n) \cong \\ \cong 1 - \frac{1}{2 \ln 2} \left(\frac{2^N (1-1/2^N)}{n} (1 + 2K) + 4s^2 + \frac{N-1}{N} K^2 \right) .$$

By substitution of the aforementioned values, i.e. n = 1 GB, s = 1/256 and K = 1/128 the final result for N = 8 is $H_S(X_1, ..., X_8 \mid n$ = 1 GB) = 1 – 2.22·10^{-8} – 4.40·10^{-5} – 7/8 · 4.40·10^{-5} = 1 – 8.25·10^{-5}. It is the value that is very close to the figure for sample entropy. It can also be seen that for n = 1 GB and N = 8 the value of masking component is relatively insignificant and negligible $M(N, n) = 2^N\ (1 - 1/2^N)\ /\ n$ = 2.22·10^{-8}.

Let us then investigate samples of the random sequences after the *XOR* operation. Let the bias is s = 1/256 and the correlation K = 1/128, M = 8. Hence $s_{\oplus}(M) = 1/2\ (2s)^M$ = 6.94·10^{-18} and $K_{\oplus}(M) = K^M$ = 1,39·10^{-17} respectively. Obviously, these values are so low that in practice they are not identifiable when the relative frequencies are measure. For comparison of results from examination of a sample with the size n = 1 GB let us then investigate a sample with the size 10 times less, i.e. n = 100 MB (100 · 8 · 1048576 bits = 838860800 bits).

Table 2. Relative frequencies $n(X_1, ..., X_N)$ N / n for N = 8 and the sample with the size of n = 100 MB

0: 0.0039065	1: 0.0038929	2: 0.0039251	3: 0.0039051	4: 0.0039048	5: 0.0039036	6: 0.0039008	7: 0.0039002
8: 0.0039052	9: 0.0039061	10: 0.0039109	11: 0.0039079	12: 0.0039075	13: 0.0039115	14: 0.0039111	15: 0.0039023
16: 0.0039041	17: 0.0039089	18: 0.0039106	19: 0.0039134	20: 0.0039050	21: 0.0039008	22: 0.0039063	23: 0.0039082
24: 0.0038973	25: 0.0039128	26: 0.0039137	27: 0.0039219	28: 0.0039130	29: 0.0039079	30: 0.0039178	31: 0.0039097
32: 0.0039064	33: 0.0038988	34: 0.0039124	35: 0.0039075	36: 0.0039026	37: 0.0039108	38: 0.0039050	39: 0.0039028
40: 0.0039051	41: 0.0038911	42: 0.0039081	43: 0.0039060	44: 0.0039111	45: 0.0038964	46: 0.0038992	47: 0.0039043
48: 0.0038977	49: 0.0039079	50: 0.0039117	51: 0.0039061	52: 0.0039013	53: 0.0038952	54: 0.0039105	55: 0.0039125
56: 0.0038952	57: 0.0038995	58: 0.0039106	59: 0.0039112	60: 0.0039089	61: 0.0039003	62: 0.0039012	63: 0.0039083
64: 0.0039020	65: 0.0039116	66: 0.0039057	67: 0.0039176	68: 0.0039098	69: 0.0039044	70: 0.0038985	71: 0.0039206
72: 0.0038925	73: 0.0039062	74: 0.0039068	75: 0.0039035	76: 0.0038979	77: 0.0039082	78: 0.0038974	79: 0.0039027
80: 0.0039122	81: 0.0038994	82: 0.0039148	83: 0.0039092	84: 0.0039075	85: 0.0039035	86: 0.0038982	87: 0.0039081
88: 0.0039111	89: 0.0038943	90: 0.0039058	91: 0.0039038	92: 0.0039099	93: 0.0039126	94: 0.0039053	95: 0.0039124
96: 0.0038939	97: 0.0039069	98: 0.0039014	99: 0.0039068	100: 0.0039104	101: 0.0039104	102: 0.0039172	103: 0.0039115
104: 0.0039053	105: 0.0038971	106: 0.0039039	107: 0.0039086	108: 0.0038994	109: 0.0039132	110: 0.0039052	111: 0.0039050
112: 0.0039082	113: 0.0039035	114: 0.0039122	115: 0.0038974	116: 0.0039205	117: 0.0039023	118: 0.0039047	119: 0.0038988
120: 0.0039115	121: 0.0039042	122: 0.0038978	123: 0.0039052	124: 0.0039077	125: 0.0039047	126: 0.0039101	127: 0.0039087
128: 0.0039093	129: 0.0038954	130: 0.0039054	131: 0.0039057	132: 0.0039164	133: 0.0039026	134: 0.0039066	135: 0.0039185
136: 0.0039165	137: 0.0039156	138: 0.0039060	139: 0.0039098	140: 0.0039120	141: 0.0039066	142: 0.0039081	143: 0.0039023
144: 0.0039089	145: 0.0039057	146: 0.0038993	147: 0.0039138	148: 0.0039120	149: 0.0039115	150: 0.0039126	151: 0.0039090
152: 0.0039120	153: 0.0038959	154: 0.0038989	155: 0.0039005	156: 0.0039068	157: 0.0039019	158: 0.0039043	159: 0.0039023
160: 0.0038961	161: 0.0039067	162: 0.0039056	163: 0.0039085	164: 0.0039087	165: 0.0039088	166: 0.0039054	167: 0.0039029
168: 0.0039030	169: 0.0039002	170: 0.0039038	171: 0.0039084	172: 0.0039010	173: 0.0039020	174: 0.0039098	175: 0.0039117
176: 0.0039138	177: 0.0038946	178: 0.0039079	179: 0.0039081	180: 0.0039118	181: 0.0039095	182: 0.0039043	183: 0.0039130
184: 0.0039022	185: 0.0038976	186: 0.0039111	187: 0.0039012	188: 0.0039055	189: 0.0039082	190: 0.0039182	191: 0.0039078
192: 0.0039149	193: 0.0038966	194: 0.0039100	195: 0.0039069	196: 0.0039073	197: 0.0039028	198: 0.0039050	199: 0.0039115
200: 0.0038980	201: 0.0039051	202: 0.0039011	203: 0.0039059	204: 0.0038981	205: 0.0039053	206: 0.0039123	207: 0.0039137
208: 0.0039065	209: 0.0039038	210: 0.0039042	211: 0.0039044	212: 0.0038973	213: 0.0039131	214: 0.0038990	215: 0.0039106
216: 0.0039092	217: 0.0039113	218: 0.0039005	219: 0.0039150	220: 0.0038986	221: 0.0038979	222: 0.0039034	223: 0.0039223
224: 0.0039087	225: 0.0039096	226: 0.0039135	227: 0.0039048	228: 0.0039044	229: 0.0039123	230: 0.0038994	231: 0.0039092
232: 0.0038949	233: 0.0038969	234: 0.0038972	235: 0.0039150	236: 0.0039064	237: 0.0039011	238: 0.0039111	239: 0.0038969
240: 0.0039105	241: 0.0039117	242: 0.0039003	243: 0.0039055	244: 0.0039098	245: 0.0039079	246: 0.0039126	247: 0.0039006
248: 0.0039080	249: 0.0039077	250: 0.0039062	251: 0.0039215	252: 0.0038998	253: 0.0039069	254: 0.0038964	255: 0.0039057

Form – $(X_1, ..., X_8)_{\mathrm{DECIMAL}}$: relative frequency for $(X_1, ..., X_8)$

The table exhibits a set of relative frequencies with their average values very close to 1/256 = $3.90625 \cdot 10^{-3}$, but with the module

$$(42) \qquad |\alpha_{\mathrm{av}\,(N=8)}| = \\ = 1/256 \sum_{X_1,...,X_N=0}^{255} |1/256 - n(X_1, ..., X_8)\,(N=8) / (n = 100\ \mathrm{MB})| \\ = 4.83 \cdot 10^{-6},$$

Hence, the experimental result is very close to the theoretical one

$$(43) \qquad |\alpha_{\mathrm{av}\,(N=8)}| \cong (16\pi\,(n = 100\ \mathrm{MB}))^{-1/2} = 4.87 \cdot 10^{-6}.$$

The fact that the values of the sample entropy and the expected entropy are different from one (1) result merely from the non-zero value of the the masking component. The mentioned values of entropy amount to $H_{\mathrm{R}}(X_{1_2}\ \ldots,\ X_{\ 8}) = 1 - 2.18 \cdot 10^{-7}$ and $H_{\mathrm{S}}(X_1, \ldots, X_8) = 1 - 2.19 \cdot 10^{-7}$.

The question appears whether such a coincidence between the theoretical results and the results obtained from measurements of examined sequences finally confirms adherence of their model to the model of a first order Markov chain. The answer is affirmative also because the model is really accurate and sensitive. It can be easily verified by introducing even insignificant but periodical oscillations of the bias and correlation into the sequence under examination or implementation of additional correlations that may convert the sequence into a Markov chain of higher order. For such a case the probability distributions for higher orders and entropies shall no longer correspond to the model and the measurement results shall be non-compliant and become practically unidentifiable. In the practice associated with hardware generation of random sequences with use of avalanche diodes it frequently happens that the generated sequence is degenerated, e.g. with the negative correlation factors K < 0 or with strong correlations that correspond to the models of Markov chains of the second and third order. Fortunately, the measurement results enable immediate identification of such sequences and the diode can be replaced with a new one, i.e. the one that meets the requirement to generate the modeled sequence as the Markov chain of the first order. It also may happen that such an avalanche diode fails to sustain its properties during long-term operation. The practice related to hardware generation of random binary sequences adheres to the rule and the mechanism that properties and parameters of the output sequence are subjected to permanent monitoring and verification whether they are in line with the adopted model, i.e. the conformity of probability distribution and the entropy values for the sequence in question [3].

The experimental evidences for stationary and ergodic properties are very simple. It is necessary to generate at least 3 sample sequences with the size of n = 1 GB and check statistics of them for random variables with the dimensions N = (1, …, 8). When the same relative frequencies with the values corresponding to (31) is obtained for each sample one can assume that the samples are the outcome of a process that is stationary in strict-sense since their distributions are invariant. It is also possible to examine higher dimensions (ranks) N = (9, …, 12) but if results for N = (1, …, 8) confirm the stationarity in the strict sense it will also occurs for higher dimensions. The ergodicity is also examined by concatenation of 3 samples into a single one with the size of 3 GB and examination of its statistics that should be identical as for the samples of 1 GB size. The perfectly random binary sequences must also lead to uniform distributions but the statistics for relative frequencies must indicate appropriate and strict values of modules – $|\alpha_{\mathrm{av}\,(N=8)}| = 1.54 \cdot 10^{-7}$ for n = 1 GB and $|\alpha_{\mathrm{av}\,(N=8)}| = 8.9 \cdot 10^{-7}$ for n = 3 GB.

Isomorphism of the random binary sequences in the in a measure-theoretic sense

Dealing with the problem of isomorphism attributable to dynamical systems Kolmogorov as early as in 50's of 20th century introduced the new understanding of isomorphism *in a measure-theoretic sense*. Such an approach makes it possible to compare and classify such systems, including sequences, based on the values of their *entropy*. He was the man who found out much deeper and more general features in relatively simple Shannon's presumptions that lead to rather heuristic relationship. Kolmogorov highly appreciated scientific knowledge and engineering intuition of Shannon and wrote about him „*In the ages of increasing specialization in science, C. Shannon emerges as an outstanding talent combining the deep mathematical thinking with wide, but concrete reasoning of the current technology. He can be considered both as the great mathematician and the gifted engineer of the XX century".* Although this paper is far away from diving deeply into the areas where the theory of dynamical systems can be applied it is reasonable to benefit from Shannon's output in the scope of analysis and classification of random sequences.

The fundamental theorems related to the theory of dynamical systems with regard to entropy read the following:

The Kolmogorov-Sinai theorem: „*if two finite state, discrete Bernoulli or Markov processes have different entropies, then they are not isomorphic in the measure-theoretic sense*". This theorem refers to independent sequences and dependent random variables.

Remark – the sequences with different entropies are not isomorphic in the measure-theoretic sense by the nature of matters.

Ornstein Theorem: „*if two finite state, discrete Bernoulli processes have the same entropy, then they are isomorphic in the measure-theoretic sense*" (simply – „*Independent processes with the same entropy are isomorphic*" [4]). The theorem indicates the isomorphism in the measure-theoretic sense for any sequences of independent random variables with the same entropies.

Remark – the sequences generated as a result of Bernoulli process comprise all possible sub-sequences but such sub-sequences may occur with different probabilities. In the very specific case the entropies of two different sequences may be the same even if probabilities of such incorporated sub-sequences are different.

Adler, Shields and Smorodinsky Theorem (I): „*any two irreducible, stationary, finite state, discrete Markov processes are isomorphic in the measure-theoretic sense if and only if they have the same periodicity and the same entropy*". This theorem refers to sequences of dependent random variables where the sequences can be modeled as periodical Markov chains. A typical example of such a chain is a sequence generated by a stream cipher.
Remark – the nature of the requirement '*the same periodicity*' results from the fact that only the sequences with the same period may comprise a finite number of subsequences with the same lengths.

Adler, Shields and Smorodinsky Theorem (II): „*an irreducible, stationary, finite state, discrete Markov process is isomorphic in the measure-theoretic sense to a finite state, discrete Bernoulli process of the same entropy rate if and only if the Markov process is aperiodic*".
The theorem refers to independent sequences and dependent random variables.
Remark – the nature of the requirement '*aperiodicity*' results from the fact that any periodical sequence comprises a finite number of sub-sequences with the limited length. On the contrary, the non-periodical sequence of dependent random variables comprises any sub-sequences with unlimited length, similarly to the sequences of independent random variables.

One can see that all foregoing theorems assume that only the sequences with equal entropies can be considered as isomorphic ones with some additional provisions in case of sequences made up of dependent random variables and modeled as the Markov chains.

Let us now assume that consistency with the model of the first-order Markov chain was found out for four samples of random chains and these sequences have biases and correlations with the following values:
– first sequence: $s = +10^{-2}$, $K = +10^{-2}$,
– second sequence: $s = +10^{-2}$, $K = -10^{-2}$,
– third sequence: $s = -10^{-2}$, $K = +10^{-2}$,
– fourth sequence: $s = -10^{-2}$, $K = -10^{-2}$.

Let us now ask the following questions
– are all sequences isomorphic in the measure-theoretic sense? Yes, since the conditional entropy $H(X_2 \mid X_1) \cong 1 - 1/2 \cdot \ln 2 \ (4s^2 + K^2) = 1 - 1.77 \cdot 10^{-4}$ is the same for all sequences,
– do all these sequences exhibit the same mutual information? Yes, all exhibit the same mutual information $I(X_2 ; X_1) \cong K^2/2 \cdot \ln 2 = 7.21 \cdot 10^{-5}$.
– are all the sequences stationary in the strict-sense and ergodic? Yes.
– are all sequences ergodic in the sense of geometric ergodicity and uniform ergodicity? No, since for $K < 0$ and each odd number n the provision takes place $\| \mathbf{P}_{(n)} - \mathbf{P} \|_{\text{tv}} \le \text{C}(0) \ K^n < 0$, thus the contradiction occurs $\| \mathbf{P}_{(n)} - \mathbf{P} \|_{\text{tv}} < 0$.
– do all sequences have equivalent distributions of whichever dimension? No. Only one-dimensional distributions for the first/second sequence and for the third/fourth sequence are the same.

Therefore it is possible to conclude that for practical applications the entropy is a synthetic and convenient indicator of the sequence randomness, in particular for large values of N, since it is the merely one-dimensional parameter that is able to confirm the level of the sequence randomness assumed by its parameters. Therefore it is interesting to find out how the entropy can be applied to investigation of the sequence randomness.

For perfectly random binary sequences the entropy can confirm the property of perfect randomness and demonstrate that the sample entropy adopts nearly the same value as the expected entropy that depends merely on the sample size. The examination must be carried out for independent samples with the sizes n = 10 MB, 100 MB and 1 GB (minimum 3 samples for each size), for each case at least for N = (1, …, 8) at least.

For imperfectly random binary sequences the entropy can confirm the property of satisfactory randomness and demonstrate that the sample entropy adopts the value close to the expected entropy and these values are strictly associated with the parameters of bias and correlation as well as the sample size. The examination should be carried out for independent samples with the sizes n = 10 MB, 100 MB and 1 GB (minimum 3 samples for each size), for each case for N = (1, …, 8) at least. If the values of sample entropy are not close to the values of expected entropy it means that the sequence fails to correspond to the model of a first order Markov chain.

Since *Adler, Shields and Smorodinsky theorems* assume stationarity of sequences, the stationarity can be verified only by comparison between properties and parameters of all samples with the same size. To check ergodicity of sequences these samples must be concatenated into a single sequence with further investigation of its statistics.

Summary

The analysis carried out in this paper covered the most important properties of random binary sequences modeled as binary Markov chains. It was demonstrated that assessment whether and how much such sequences can be considered as random ones is only possible upon analysis of many properties since any incomplete set of properties is insufficient to make a trustworthy conclusion on the sequence randomness. The proof is provided that the most important property of such sequences are probability distributions of random variables carried out for a finite number of dimensions. These distributions can be then verified by measuring relative frequencies of specific sub-sequences incorporated into samples of random binary sequences. It was also demonstrated that the XOR function of M independent imperfectly random binary sequences may lead to composition of resulting sequences that can be considered as perfectly random binary sequences. Our theoretical analysis was confirmed by experimental results.

The next problem to consider consists in construction of a hardware random number generator (HRNG) that shall be a tangible generator of random binary sequences suitable for practical applications.

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

REFERENCES

[1] Rukhin A. et al.: NIST Special Publication 800-22. Revision 1a. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. *NIST*, April 2010.
[2] Leśniewicz M., Expected Entropy as a Measure and Criterion of Randomness of Binary Sequences, *Przegląd Elektrotechniczny*, 90 (2014), nr 1, 42-46.
[3] Leśniewicz M., Sprzętowa generacja losowych ciągów binarnych (Hardware generation of binary random sequences), *Wydawnictwo Wojskowej Akademii Technicznej (Military University of Technology)*, (2009). (In Polish.)
[4] Meyn S.P., Tweedie R.L, Markov Chains and Stochastic Stability. *Springer-Verlag* (1993, version compiled - 2005).