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 analyses have been verified by measuring experiments.

Streszczenie. W artykułe przedstawiono nowe podejście do modelowania losowych ciągów binarnych. Wszechstronnie przeanalizowano problemy ich stażomątności i ergodyczności, postaci wielowymiarowych rozkładów prawdopodobieństw i istotę izomorfizmu. Analizy teoretyczne zostały potwierdzone doświadczalnie. (Analizy i pomiary sprzętu 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 sequence sample 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 or 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 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 an 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 $P(\chi^t | \chi^{t-1})$, that can be summarized into matrix of transition probabilities (simply – transition matrix) as

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

If for all $P(\chi^t | \chi^{t-1})$ the conditions $0 \leq P(\chi^t | \chi^{t-1}) \leq 1$ and $P(0|0) + P(1|0) = P(0|1) + P(1|1) = 1$ are fulfilled, the matrix $M$ is referred to as the stochastic matrix.

The probabilities for subsequent elements of the sequence can be described as the total probabilities
formally indicates the characteristic roots (8), where the Kolmogorov equation adopts the form (12). But anyway, even if and is referred to as the

(12)

and for subsequent m elements

(13)

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

(14)

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

(15)

where

(16)

and

(17)

The M* 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 ergodic and irreducible properties of the matrix. The second root \( \varphi_2 = 1 - P(0|1) - P(1|0) < 1 \) shows the non-cyclic property of the matrix.

The M* matrix meets the convergence provision

(18)

Due to the foregoing property the M* is referred to as the ergodic matrix with its \( \pi(0) \) and \( \pi(1) \) 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

(19)

It turns out that there is only one \( P \) vector that fits the foregoing equation. The vector is

(20)

and is referred to as the stationary distribution whilst its terms are equal to terms in rows of the ergodic matrix M*, 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 \( m \)

(21)

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

(22)

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

(23)

where \( || \cdot ||_v \) 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 \( P_{n|0} \) for the \( n \)th element and the stationary probabilistic vector of \( P \):

(24)

\( C(0 < \infty \) is the constant that depends on the vector of initial conditions, i.e. the probabilistic vector of \( P_{0|0}(\cdot) \); \( \rho < 1 \) is the constant that corresponds to the second characteristic root of the matrix \( M \), 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

(25)

to (15) leads to the following equation

(26)

where \( C(0 = 1/2 \) is the constant that depends on the constant \( \kappa^n \) and \( \rho^n \).

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 \( C < \infty \) to be independent on initial conditions

(27)

However, it is possible to find out that the uniform ergodicity is characteristic for stationary Markov chains since for \( P(0|0) = \pi(0) \) and \( P(1|0) = \pi(1) \) the provision

(28)

is fulfilled.

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

(29)
since $C \leq 1$, $\pi(z) = 1/2$, the $p^n = K^n$ 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 $E(n)$ and the variance $V(n)$ for the Gaussian distribution of the random variable $n = k / n$, 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]

\begin{equation}
E(n) = \pi(0) + \frac{P(0)(0) - \pi(0)}{n} 1 - K^n
\end{equation}

and

\begin{equation}
V(n) = \pi(0)\pi(1) 1 + K
\end{equation}

where $k$ 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 $n = k / n$.

It is easily to see that for a stationary Markov chain, i.e. $P(0)(0) = \pi(0)$ the mathematical expectation is immediately $E(n) = \pi(0)$ whilst the variance $V(n)$ 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 $M$ 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 correlation $K$ 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

\begin{equation}P(X_1, ..., X_N) = P(X_1 | X_1, ..., X_{N-1}) P(X_1, ..., X_{N-1}) \end{equation}

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

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

\begin{equation}
P(X_1, ..., X_N) = P(X_1 | X_1) P(X_2 | X_1) \cdots P(X_N | X_{N-1}) .
\end{equation}

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

\begin{equation}
P(X_1, ..., X_N) = P(X_1 | X_1) P(X_2 | X_2) \cdots P(X_N | X_{N-1}) .
\end{equation}

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

\begin{equation}
\mathbf{M} = \begin{pmatrix} P(0)(0) & P(1)(1) & \cdots & P(N-1)(N-1) \\ P(0)(1) & P(1)(1) & \cdots & P(N-2)(N-1) \\ & \vdots & & \vdots \\ P(0)(N-1) & P(1)(N-1) & \cdots & P(1)(N-1) \end{pmatrix}
\end{equation}

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 $s < s_1$ and $K < K_1$. By rejection of entries with significantly less values and with consideration to the Kolmogorov axiom the following equation can be achieved

\begin{equation}
\sum_{X_1 = 0}^{X_1, ..., X_N = 0} P(X_1, ..., X_N) = 1 ,
\end{equation}

that is correct for $N$-dimensional distributions:

A. one-dimensional distribution

\begin{equation}P(0) = 1/2 - s 
\end{equation}

\begin{equation}P(1) = 1/2 + s
\end{equation}

B. two-dimensional distribution

\begin{equation}P(0,0) = 1/4 - s + 1/4 K = 1/4 (1 - 4 s + K)
\end{equation}

\begin{equation}P(0,1) = 1/4 - 1/4 K = 1/4 (1 - K)
\end{equation}

\begin{equation}P(1,0) = 1/4 - 1/4 K = 1/4 (1 - K)
\end{equation}

\begin{equation}P(1,1) = 1/4 + s + 1/4 K = 1/4 (1 + 4 s + K)
\end{equation}

C. three-dimensional distribution

\begin{equation}P(0,0,0) = 1/8 - 3/4 s + 1/4 K = 1/8 (1 - 6 s + 2 K)
\end{equation}

\begin{equation}P(0,0,1) = 1/8 - 1/4 s = 1/8 (1 - 2 s)
\end{equation}

\begin{equation}P(0,1,0) = 1/8 - 1/4 s - 1/4 K = 1/8 (1 - 2 s - 2 K)
\end{equation}

\begin{equation}P(1,0,1) = 1/8 + 1/4 s = 1/8 (1 + 2 s)
\end{equation}

\begin{equation}P(1,1,0) = 1/8 + 1/4 s = 1/8 (1 + 2 s)
\end{equation}

\begin{equation}P(1,1,1) = 1/8 + 3/4 s + 1/4 K = 1/8 (1 + 6 s + 2 K)
\end{equation}

D. four-dimensional distribution

\begin{equation}P(0,0,0,0) = 1/16 - 1/2 s + 1/3 K = 1/16 (1 - 8 s + 3 K)
\end{equation}

\begin{equation}P(0,0,0,1) = 1/16 - 1/4 s + 1/4 K = 1/16 (1 - 14 s + K)
\end{equation}

\begin{equation}P(0,0,1,0) = 1/16 - 1/4 s - 1/4 K = 1/16 (1 - 14 s - K)
\end{equation}

\begin{equation}P(0,1,0,1) = 1/16 + 1/4 K = 1/16 (1 + K)
\end{equation}

\begin{equation}P(0,1,1,0) = 1/16 - 1/4 K = 1/16 (1 - K)
\end{equation}

\begin{equation}P(1,0,1,1) = 1/16 + 1/4 s + 1/4 K = 1/16 (1 + 4 s + K)
\end{equation}

\begin{equation}P(1,0,0,1) = 1/16 - 1/4 s + 1/4 K = 1/16 (1 - 4 s + K)
\end{equation}

\begin{equation}P(1,0,1,0) = 1/16 - 1/4 s = 1/16 (1 - 2 s)
\end{equation}

\begin{equation}P(1,1,0,1) = 1/16 + 1/4 s = 1/16 (1 + 2 s)
\end{equation}

\begin{equation}P(1,1,1,0) = 1/16 + 1/4 s = 1/16 (1 + 2 s)
\end{equation}

\begin{equation}P(1,1,1,1) = 1/16 + 3/4 s + 1/4 K = 1/16 (1 + 6 s + 2 K)
\end{equation}
\[ P(1,1,0,0) \approx 1/16 + 1/16 K = 1/16 (1 + K) \]
\[ P(1,1,0,1) \approx 1/16 + 1/4 s - 1/16 K = 1/16 (1 + 4 s - K) \]
\[ P(1,1,1,0) \approx 1/16 + 1/4 s + 1/16 K = 1/16 (1 + 4 s + K) \]
\[ P(1,1,1,1) \approx 1/16 + 1/2 s + 3/16 K = 1/16 (1 + 8 s + 3 K) \]

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 \( 2^N \) 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

\[ P(X_1, ..., X_N) = \frac{1}{2^N} \left[ 1 + 2 s \{ L(1) - L(0) \} + K \left[ L(0,0) + L(1,1) - L(0,1) - L(1,0) \right] \right], \]

where

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

\[ \varepsilon_{(X_1, ..., X_N)} \ll 1 \]

is following

\[ H(X_1, ..., X_N) = \frac{1}{N} \sum_{x_1, ..., x_N} P(x_1, ..., x_N) \log_2 P(x_1, ..., x_N) \approx \]

\[ \approx 1 - \frac{N^2}{N \ln 2} \sum_{x_1, ..., x_N} \varepsilon_{(x_1, ..., x_N)}, \]

The transformation of (29) to the form

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

\[ = 1/2^N + \varepsilon_{(X_1, ..., X_N)}, \]

and the substitution

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

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

\[ H(X_1, ..., X_N) \approx 1 - \frac{1}{2 \ln 2} \left( 4 s^2 + \frac{N-1}{N} K^2 \right). \]

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 \( s = 5 \cdot 10^{-5} \) and correlation \( K = 5 \cdot 10^{-5} \). 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 \( s_{(M)} = 1/2 (2s)^M \) and \( K_{(M)} = K^M \),

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 \( s_{(M)} < s \) and \( K_{(M)} < K \).

Hence, the probabilities for any \( N \)-dimensional random variable can be rewritten in the form of the recurrence relationship

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

\[ = 1/2^N + \varepsilon_{(X_1, ..., X_N)}, \]

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. invariance 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 \( N \rightarrow \infty \) 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 \( k \) the said measure is defined by the theoretical relationship [3]
For the maximum excess of ones over zeros and the maximum correlation between the sequence of zeros the
expected relative frequencies shall amount to \( \frac{1}{2} \) relative frequency is \( \frac{1}{2} \) respectively +0.007% and –0.0054%.
For the equality between (1) and (0) and the minimum correlation between the alternated zeros and ones probability amounts
\( P(11111111_{\text{dec}}) = 1.256 \{ 1 + 2 (s = 1/256) \ (L(1) - L(0)) = +8 \} + \\
+ (K = 1/128) \ [ (L(0,0) + L(1,1) - L(0,0) - L(1,1)) = +7 ] = \\
= 0.0043640 \\
– relative frequency (255_{\text{dec}}) = 0.0043646, \\
– relative difference between the probability and the relative frequency –0.014%.

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
\[ H_S(X_1, \ldots, X_N | n) = \sum_{i=1}^{2^n-1} \log_2(\frac{n(X_i, \ldots, X_N | n)}{n(X_i, \ldots, X_N | n)}), \]
where for \( N = 8 \) and \( n = 1 \) GB the result is \( H_S(X_1, \ldots, X_N | n) = 1 \ \text{–} \ 8.38 \times 10^{-18}. \)
Practically equal result can be obtained from the simplified formula for entropy [2]
\[ H_S(X_1, \ldots, X_N | n) \approx \frac{1}{N} \sum_{i=1}^{2^n-1} \log_2(\frac{n(X_i, \ldots, X_N | n)}{n(X_i, \ldots, X_N | n)}), \]
\[ \approx 1 - \frac{2^{N-1}}{N \ln 2} \sum_{X_1, \ldots, X_N = 0}^{2^n-1} \left[ \frac{1}{2^n} - n(X_1, \ldots, X_N | n) \right]^2. \]
Let us check where such values appeared from. One has to keep in mind that the formula for entropy is the following [2]
\[ H_S(X_1, \ldots, X_N | n) \approx \frac{1}{2 \ln 2} \sum_{X_1, \ldots, X_N = 0}^{2^n-1} \left( \frac{2^n - 1}{n} \right) (1 + 2K) + 3x^2 + \frac{N-1}{N} K^2. \]
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 \( H_S(X_1, \ldots, X_N | n) = 1 \ \text{–} \ 2.22 \times 10^{-8} - 4.40 \times 10^{-8} - 7/8 \cdot 4.40 \times 10^{-8} = 1 \ \text{–} \ 8.25 \times 10^{-8}. \)
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 \( M(n, n) = 2^n \ (1 - 2^{-n}) / n \) is
\[ 2.22 \times 10^{-8}. \]
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 \( x_s(M) = 1/2 (2)^M = 6.94 \times 10^{-18} \)
and \( K_s(M) = K^M = 1.39 \times 10^{-17} \) 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 = 8388608000 bits).
The table shows a set of relative frequencies with their average values very close to 1/256 = 3.90625·10⁻⁸, but with the module

\[
\alpha_{\text{av}}(N) | = \frac{1}{256} \sum_{X_{1}=0}^{255} | 1/256 - m(X_{1}, ..., X_{N}) (N = 8) / (n = 100 MB) \mid = 4.83·10^{-6}.
\]

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

\[
\alpha_{\text{av}}(N) | \equiv (16\pi (n = 100 MB))^{-1/2} = 4.87·10^{-6}.
\]

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 periodic oscillations of the bias and correlation into the sequence under examination or implementation of additional correlations that may convert the sequence into a higher order 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 requirements to generate the modeled sequence as the Markov chain of the first order. It may also 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

\[
\alpha_{\text{av}}(N=8) | = 1.54 \cdot 10^{-7} \text{ for } n = 1 \text{ GB and } | \alpha_{\text{av}}(N=8) | = 8.9 \cdot 10^{-7} \text{ for } n = 3 \text{ GB.}
\]

**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 \( h(X_2 \mid X_1) \geq 1 - 1.2 \cdot 10^{-2} \) is the same for all sequences.
- do all these sequences exhibit the same mutual information? Yes, all exhibit the same mutual information \( h(X_2, X_1) \geq K/2 \cdot \ln 2 = 7.21 \cdot 10^{-5} \).
- 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 \( \| P(n) - P \|_n \leq C(0) K^n < 0 \), thus the contradiction occurs \( \| P(n) - P \|_n < 0 \).
- 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 \( n = 10 \text{ MB}, 100 \text{ MB} \) and \( 1 \text{ GB} \) (minimum 3 samples for each size), for each case at least for \( N = (1, \ldots, 8) \) 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 \( n = 10 \text{ MB}, 100 \text{ MB} \) and \( 1 \text{ GB} \) (minimum 3 samples for each size), for each case for \( N = (1, \ldots, 8) \) 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 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 Instytutu Łączności, Zakład Kryptologii, ul. Warszawska 22A, 05-130 Zegrze, E-mail: m.lesnieicz@wil.waw.pl, m.lesnieicz@op.pl

REFERENCES


