# Correlation function inadequacy in random-sequence entropy measures

O. V. Usatenko ,[*] S. S. Melnyk , and G. M. Pritula 
*O. Ya. Usikov Institute for Radiophysics and Electronics of the National Academy of Sciences of Ukraine,*
*12 Proskura Street, 61805 Kharkiv, Ukraine*

(Received 13 February 2020; accepted 20 July 2020; published 14 August 2020)

Considering symbolic and numerical random sequences in the framework of the additive Markov chain approach, we establish a relation between their correlation functions and conditional entropies. We express the entropy by means of the two-point probability distribution functions and then evaluate the entropy for the numerical random chain in terms of the correlation function. We show that such approximation gives a satisfactory result only for special types of random sequences. In general case the conditional entropy of numerical sequences obtained in the two-point distribution function approach is lower. We derive the conditional entropy of the additive Markov chain as a sum of the Kullback–Leibler mutual information and give an example of random sequence with the exactly zero correlation function and the nonzero correlations.

DOI: 10.1103/PhysRevE.102.022119

## I. INTRODUCTION

Sequences with a finite-state space and nontrivial information content have been the focus of a wide variety of research in different fields of science for the past several decades. These sequences exist as natural ones (e.g., DNA or natural language texts) or arise as a result of coarse-grained mapping of the evolution of the chaotic dynamical system into a string of symbols [1,2]. The elements of sequence, depending on the system of interest, can be phonemes, syllables, words, DNA's base pairs, numbers, etc.

The finite-state sequences, considered as random, are the subject of study of the algorithmic (Kolmogorov-Solomonoff-Chaitin) complexity, information theory, computability, statistical inference problem and have many application aspects, such as, for example, data compression [3] and the natural language processing [4], which is an important branch of the artificial intelligence.

In the arsenal of modern science there are a lot of instruments for describing complex dynamical systems and random sequences associated with them: correlation functions, fractal dimensions, multipoint probability distribution functions, and many others. The entropy is one of the important macroscopic characteristics used for the numerical evaluation of the complexity and information content of dynamical systems [5,6]. Being a measure of the information content and redundancy in a sequence of data, it is a powerful and popular tool in examination of complexity phenomena. It is extensively used for the analysis of different dynamical systems. The importance of the entropy lies also in its fundamental connection with the compression of the random sequences of data.

Most lossless information compression methods use data about the discovered statistical properties of the sequence. The more accurately these properties are identified, the more accurately the information on the preceding elements can be used to predict the values of the subsequent ones, which increases the compression ratio of the sequence.

One of the ways to gain insight into the nature of interaction between elements of sequences with nontrivial information content consists in a possibility to construct a correlated sequence of symbols which reproduces some statistical characteristics of the initial system. Among the diversity of algorithms for generating correlated sequences, the high-order Markov chains take one of the most remarkable places. Such random chains, the method of their generation, and all their statistical properties are completely determined by the *conditional probability distribution function* (CPDF), known also as the transition probability function. One of the methods to reconstruct the CPDF of random sequence is to approximate the sequence by the high-order Markov chains with a finite alphabet.

The CPDF of the Markov chain of high order has a complex structure, as well as the statistics of the random sequence itself and, in general, its description may require a huge number of parameters. This effect is known as the combinatorial explosion and engenders the so called *curse of dimensionality problem.*

In Ref. [7] there was proposed a simplified high-order Markov model. The model appears to be useful for many practical tasks, including the CPDF numerical reconstruction. Nevertheless, there exists its more simple version, the so-called *additive* high-order Markov chain; see Eq. (17) and explanation therein. This Markov chain allows one to reduce the number of parameters in the CPDF of high-order Markov chain.

A standard method of analyzing statistical properties of a given random sequence of data is as follows. First, one has to find the joint probabilities of words occurring of the lengths $L$ which exceed the correlation length, $R_c$, and are less than the sequence length $S$,

$$R_c < L \ll S. \tag{1}$$

[*] usatenkoleg@gmail.com

At the same time, the number $M^L$ of different words of the length $L$ composed of letters of alphabet containing $M$ letters has to be much less than the number $S - L + 1$ of words in the sequence,

$$M^L \ll S - L + 1. \tag{2}$$

The next step is to express the correlation properties of sequence in terms of a CPDF of the Markov chain, see below Eq. (6). Note, the Markov chain should be of order $N$, which is not less than the correlation length,

$$R_c < N. \tag{3}$$

Here is the bottleneck because the correlation lengths of natural sequences of interest (e.g., written or DNA texts) are usually of the same order as the lengths of sequences. None of inequalities Eqs. (1)–(3) can be fulfilled. Really, the lengths of representative words that could estimate correctly the probability of words occurring are 4–5 for a real natural text of the length $10^6$ (written in an alphabet of 27–30 letters and symbols) or of order of 20 for a coarse-grained text represented by means of a binary sequence.

To overcome these impediments we proposed another method for examination of random sequences. We elaborated [8,9] a method of constructing the conditional probability function presented by means of pair correlators which makes it possible to calculate analytically the entropy of the sequence; see Eq. (18).

It is pertinent to mention here that the model of homogeneous additive high-order Markov chain was successfully used for studying different random systems and sequences, such as, for example, literary texts in English and other languages, Refs. [8,9,11–14], different DNA sequences, Refs. [10,13–15], wind generation time series, Ref. [16]. The class of additive and additive linear sequences can be considered as the first order approximation of the CPDF of a complex real system and in this way can be used to analyze a wide range of objects.

In a series of our papers we have studied possibilities of presentation of entropy by means of various parameters of additive Markov chain. In Ref. [13] we developed the evaluation for the entropy of random symbolic sequences with elements belonging to a *finite alphabet*. As a plausible model, we used the high-order additive stationary ergodic Markov chain with long-range memory. In Ref. [15], using the *bilinear* Markov chain approach, we studied statistical properties of natural random symbolic sequences with complex correlation properties. We showed that our method gives a much lower level of entropy as compared to the best archivers. In all these papers we presented estimates for the entropy of the sequences by means of the corresponding correlation functions.

There arises a natural question if it is possible to express the entropy of *numerical chain* by the conventional *numerical correlation function* [see its definition, e.g., below, Eq. (15)]. In one form or another the question—if the entropy is a good measure of correlation—was repeatedly raised in the literature (see, for example, Refs. [17–20]). Here we try to answer this question but posed in a slightly different way.

In the present paper we consider two types of random chains with *two different state spaces*. The first ones are the numerical chains for which the random variables are taken from a finite set of numbers. For the symbolic random chains the random variables are taken from an alphabet or from a formal set of elements. We establish a one-to-one correspondence between the elements of the symbolic and numerical sequences, calculate the conditional entropy of these sequences and show the quite obvious coincidence of the calculated entropies. After that we estimate the entropy of numerical random chain in terms of the numerical correlation function and show that this entropy gives a satisfactory result only for some special types of random chains: The correlation function can describe the conditional entropy correctly only in the case when the numerical random sequence has a diagonal correlation matrix. In general, the sequence entropy level obtained in the two-point distribution function approach is lower. This is the main result of the paper.

The scope of the paper is as follows. In Sec. II we explain the concept of the additive finite-state high-order Markov chains and provide definitions of the CPDF and numerical correlation functions. Section III presents the equations relating the correlation and memory functions of the symbolic and numerical Markov chains. In Sec. IV we represent the conditional entropy in terms of the conditional probability function of the Markov chain and then express the entropy by means of the two-point *probability distribution functions* (PDFs) as well as with the use of the correlation functions. The numerical illustrations of the obtained results are contained in Sec. V. The applications of the proposed general algorithms to some specific classes of chains are presented in the Appendix.

This work is a generalization of our previous papers [13,21] devoted to the studies of the entropy of the symbolic ergodic stationary finite-state additive Markov chains.

## II. FINITE-STATE HIGH-ORDER MARKOV CHAINS: MAIN DEFINITIONS

Let us consider an infinite random discrete-valued sequence of elements $X_n$,

$$\mathbb{S} = \ldots, X_{-1}, X_0, X_1, \ldots \tag{4}$$

Here we would like to note that though in theory we consider infinite random sequences, the sequence length $S$ is necessarily finite in practice.

We suppose that the random sequence $\mathbb{S}$ is a *high-order Markov chain* [7,22,23]. The sequence $\mathbb{S}$ is the *N-order* Markov chain if it has the following property: The conditional probability distribution function of random variable $X_n$ to have a certain value $x_n$ under the condition that *all* previous symbols are given depends only on $N$ previous symbols,

$$\begin{aligned}&\mathbb{P}(X_n = x_n|\ \ldots\ ; X_{n-N} = x_{n-N};\ \ldots\ ; X_{n-1} = x_{n-1})\\ &\quad = \mathbb{P}(X_n = x_n|X_{n-N} = x_{n-N};\ \ldots\ ; X_{n-1} = x_{n-1}),\end{aligned} \tag{5}$$

$\forall n \in \mathbb{Z}$. Such sequences are also referred to as multi- or $N$-step [8,11,22,24], and the number $N$ is also called the *memory length*. The subscript notation for the random variables $x_n$ is used to indicate the position of the symbol in the chain $\mathbb{S}$.

The CPDF of the sequence, providing that the probabilities $\mathbb{P}(X_{n-N} = x_{n-N};\ \ldots\ ; X_{n-1} = x_{n-1}, X_n = x_n)$ and $\mathbb{P}(X_{n-N} = x_{n-N};\ \ldots\ ; X_{n-1} = x_{n-1})$ of occurring

$(N+1)$- and $N$-subsequences are given, can be found in a standard way,

$$\mathbb{P}(X_n = x_n | X_{n-N} = x_{n-N};\, \dots;\, X_{n-1} = x_{n-1}) = \frac{\mathbb{P}(X_{n-N} = x_{n-N}; \dots; X_{n-1} = x_{n-1}, X_n = x_n)}{\mathbb{P}(X_{n-N} = x_{n-N}; \dots; X_{n-1} = x_{n-1})}. \qquad (6)$$

This probability function defines all the characteristics of the sequence, including its correlation functions and entropy.[1] Hereafter, we will use the concise notation

$$x_i^k \stackrel{\text{def}}{=} x_i, x_{i+1}, \dots, x_k.$$

The independence of the CPDF on the position $n$ of random variables in the sequence provides homogeneity. The homogeneity, in its turn, provides the stationarity of the sequence under consideration; and finiteness of $N$ together with the strict inequalities,

$$0 < \mathbb{P}\big(X_n = x \big| X_{n-N}^{n-1} = x_{n-N}^{n-1}\big) < 1, \qquad (7)$$

gives, according to the Markov theorem (see, e.g., Refs. [25,26]), ergodicity of the sequence.

The conditional probability being a probability measure has the following property,

$$\sum_x \mathbb{P}\big(X_n = x \big| X_{n-N}^{n-1} = x_{n-N}^{n-1}\big) = 1, \qquad (8)$$

for any realization of the previous $N$ elements of the chain; summation is performed over the state space of random variable.

The state space of the sequences, defined by all the values that random elements $X_n$, $n \in \mathbb{Z}$ can take on, is supposed to be a finite set $\mathcal{N}$ of real numbers,

$$\mathcal{N} = \{x^1, \dots, x^M\}, \qquad (9)$$

or a finite set $\mathcal{A}$ of symbols,

$$\mathcal{A} = \{a^1, \dots, a^M\}. \qquad (10)$$

Here the superscript notation for $x^n$ and $a^n$ is used to indicate the position of the symbols in the alphabets $\mathcal{N}$ or $\mathcal{A}$. In what follows we will consider the symbolic and numerical sequences with elements satisfying the conditions of one-to-one correspondence $x^m \leftrightarrows a^m$ or, in other words, each symbol from the state space of symbolic sequence has its assigned, corresponding number in the state space of the numerical alphabet and vice versa.

The (auto)correlation function of stationary symbolic chain, the *symbolic correlation matrix* $C_{ab}(n)$, complying with Ref. [13], we determine as

$$\begin{aligned} C_{ab}(n) &= \langle[\delta(a_i, a) - p(a)][\delta(a_{i+n}, b) - p(b)]\rangle \\ &= P_{ab}(n) - p(a)p(b), \end{aligned} \qquad (11)$$

where $\delta(.,.)$ is the Kronecker $\delta$ symbol, $P_{ab}(n) = \mathbb{P}(X_i = a, X_{i+n} = b) = \langle\delta(X_i, a)\delta(X_{i+n}, b)\rangle$ is the two-point PDF. Let us note that we use the same notation $X_n$ for the random variable taking its values on $\mathcal{N}$ or on $\mathcal{A}$. The quantity $p(b)$ is the relative number of symbols $b$ in the chain, or their probabilities of occurring,

$$p(b) = \mathbb{P}(X_i = b) = \langle\delta(X_i, b)\rangle. \qquad (12)$$

The sign $\langle\dots\rangle$ means a statistical average over an ensemble of sequences. Due to the ergodicity, the ensemble average of any function $f(a_{r_1}, a_{r_2}, \dots, a_{r_s})$ of $s$ arguments defined on the set $\mathcal{A}^s$ of symbols can be replaced by the arithmetic (Cesàro's, "temporal") average over the chain, e.g.,

$$p(b) = \lim_{S\to\infty} \frac{1}{S} \sum_{i=1}^{S} \delta(a_i, b). \qquad (13)$$

This latter property is very useful in numerical calculations since the averaging can be done over the sufficiently long sequence and the ensemble averaging can be avoided. The symbolic correlation matrix has the following properties:

$$C_{ab}(r) = C_{ba}(-r), \qquad \sum_{a\in\mathcal{A}} C_{ab}(r) = \sum_{b\in\mathcal{A}} C_{ab}(r) = 0. \qquad (14)$$

The first equality is a direct consequence of the stationarity, whereas the second is a consequence of the marginalization of a probability distribution. The correlation function $C(n)$ of the corresponding numerical chain can be expressed in terms of the symbolic correlation matrix $C_{x^i x^j}(n)$:

$$\begin{aligned} C(n) &= \sum_{(x^i, x^j)\in\mathcal{N}^2} x^i x^j P_{x^i x^j}(n) - \langle X\rangle^2 \\ &= \sum_{(x^i, x^j)\in\mathcal{N}^2} x^i x^j [P_{x^i x^j}(n) - p(x^i)p(x^j)] \\ &\equiv \sum_{(x,y)\in\mathcal{N}^2} x\, y\, C_{xy}(n), \end{aligned} \qquad (15)$$

where

$$\langle X\rangle = \sum_{i=1}^{M} x^i p(x^i) \equiv \sum_{x\in\mathcal{N}} x\, p(x) \qquad (16)$$

is the average value of random variable $X$ in the sequence. The last identical equalities in Eqs. (15) and (16) introduce a simplified notation for summations. The numerical correlation function is an even function of the distance $n$, $C(n) = C(-n)$.

To harness the high-order Markov chain to serve in the context of correlations and entropy, we need some simplified models.

## III. ADDITIVE HIGH-ORDER MARKOV CHAINS

### A. CPDF and memory functions of the symbolic chains

Earlier, in our papers [8,9], there were studied the two simplest models for the *additive* symbolic high-order Markov chain where the CPDF was assumed to be of a specific, simplified, "linear form" with respect to the random variables $X$:

$$\begin{aligned} P\big(a\big|a_1^N\big) &\equiv \mathbb{P}\big(X_{N+1} = a \big| X_1^N = a_1^N\big) \\ &= p(a) + \sum_{n=1}^{N} \sum_{b\in\mathcal{A}} F_{ab}(n)[\delta(a_{N+1-n}, b) - p(b)]. \end{aligned} \qquad (17)$$

[1] For the continuous-state Markov chain this equation defines the conditional probability *density* function [29,30].

Here $F_{ab}(n)$ is the so-called *memory function*. The Kronecker $\delta$ symbol $\delta(.,.)$ in Eq. (17) plays the role of the indicator function of the random variable $a_i$ and converts symbols to numbers. The additivity of the chain means that the "previous" symbols $a_1, \dots, a_N$ exert an independent effect on the probability of the "final," generated symbol $X_{N+1} = a$ occurring. The first term in the right-hand side of Eq. (17) is responsible for the correct reproduction of statistical properties of uncorrelated sequences, the second one takes into account and correctly reproduces correlation properties of the chain up to the second order. The high-order correlation functions are not independent here. We cannot control them and reproduce correctly by means of the memory function $F_{ab}(n)$.

For any values of $a, b \in \mathcal{A}$ and $n \geqslant 1$ the relationship between the correlation and memory functions was obtained [8,9],

$$C_{ab}(n) = \sum_{n'=1}^{N} \sum_{c\in\mathcal{A}} C_{ac}(n-n') F_{bc}(n'), \quad n \geqslant 1. \tag{18}$$

This formula provides a tool for constructing weak correlated sequences with a given pair correlation matrix and determines the value of the correlation function $C_{ab}(n)$ at $n \geqslant N$ by its $N$ previous values $C_{ac}(n-n'),\ n' = 1, \dots, N$.

### B. CPDF and memory functions of the numerical chains

By the discretization (this procedure is also called a quantization or a box coarse-graining) we can convert any random process (or numerical random sequence) into a numerical chain (also called categorical) that has a finite number of possible states at each discrete-time point.

In work [27] we considered such kind of numerical $N$-order Markov chain $\mathbb{S}_{\mathcal{N}}$ with *additive linear* CPDF,

$$P(x|x_1^N) \equiv \mathbb{P}(X_{N+1} = x | X_1^N = x_1^N) = p(x) + \sum_{n=1}^{N} f_n(x)(x_{N+1-n} - \langle X\rangle). \tag{19}$$

The additivity of the chain, presented here in the linear form, means that the "previous" values $x_1, x_2, \dots, x_N$ exert an independent linear effects on the probability of the "final," generated value $X_{N+1} = x$ occurring. The first term in the right-hand side of Eq. (19), the function $p(x)$, is responsible for the correct reproduction of statistical properties of uncorrelated sequences, the second term containing weight functions $f_n(x)$, $n = 1, \dots, N$ takes into account and correctly reproduces correlation properties of the chain up to the second order.

Property Eq. (8) together with the equality $\sum\limits_{x\in\mathcal{N}} p(x) = 1$ leads to

$$\sum_{x\in\mathcal{N}} f_n(x) = 0 \quad (n = 1, \dots, N). \tag{20}$$

For simplicity of calculations and without loss of generality, we suppose that $\langle X\rangle = 0$. The corresponding equality for the symbolic chain is

$$\langle F_a(n)\rangle = \sum_{b\in\mathcal{A}} F_{ab}\, p(b) = 0. \tag{21}$$

The equation for the correlation function of the numerical chain has the form similar to Eq. (18),

$$C(n) = \sum_{n'=1}^{N} \Phi(n') C(n-n'), \quad n \geqslant 1, \tag{22}$$

where the memory function $\Phi(n)$ is determined as

$$\Phi(n) = \sum_{x\in\mathcal{N}} f_n(x)\, x \quad (n = 1, \dots, N). \tag{23}$$

## IV. ENTROPY AND CORRELATION FUNCTIONS OF THE HIGH-ORDER MARKOV CHAIN

### A. Entropy and correlations in symbolic chains

To calculate the conditional entropy of stationary sequence $\mathbb{S}_{\mathcal{A}}$ of symbols $a_i$ one could use the Shannon definition [5] for the entropy per block of length $L$,

$$H(L) = -\sum_{a_1^L\in\mathcal{A}^L} P(a_1^L) \log_2 P(a_1^L). \tag{24}$$

Here $P(a_1^L)$ is the probability to find $L$-subsequence $a_1^L$ in the sequence $\mathbb{S}_{\mathcal{A}}$ of symbols $a_i$. The conditional entropy, or the entropy per one random element, with the use of the chain rule, see Ref. [6], can be presented as follows:

$$h(L) = H(L+1) - H(L). \tag{25}$$

This quantity specifies the degree of uncertainty of the $(L+1)$th random element occurring and measures the average information per this element if the correlations of $(L+1)$th number with preceding $L$ numbers are taken into account.

For weak correlations, when for all $n \neq 0$ the components of the *normalized* correlation function (known also as the correlation coefficient),

$$K_{ab}(n) = \frac{C_{ab}(n)}{C_{ab}(0)}, \tag{26}$$

are small compared to the $K_{ab}(0) = 1$, and some additional conditions are met, see Appendix A, it is possible to find (in the lowest approximation) the simple interrelation between the memory function $F_{ab}(n)$ and the correlators $C_{ba}(n)$,

$$F_{ab}(n) \approx C_{ba}(n)/p(b), \quad 1 \leqslant n \leqslant N. \tag{27}$$

Recovering the memory functions from the correlator equation is the purpose of the so called inverse problem—the problem of retrieving the CPDF of the sequence provided the correlation functions are given—which is important for modeling and simulation in different areas; see, for example, Refs. [16,27]. Equation (27) allows us to express the CPDF, Eq. (17), in terms of the correlation functions of the chain and then present the conditional entropy of a stationary ergodic weakly correlated random sequence via its correlators. As a result we have, see Ref. [13],

$$h_{\mathtt{Symb}}(L) \approx h_0 - \frac{1}{2\ln 2} \sum_{n=1}^{L} \sum_{(a,b)\in\mathcal{A}^2} \frac{C_{ab}^2(n)}{p(a)p(b)}, \tag{28}$$

where the first term in the right-hand side of the equation is the entropy of the uncorrelated sequence

$$h_0 = -\sum_{a\in\mathcal{A}} p(a)\log_2 p(a). \tag{29}$$

The conditional entropy $h_{\texttt{Symb}}(L)$ can be expressed via the mutual information $I(n)$ which is another measure of the dependence between two variables. The mutual information is the Kullback-Leibler divergence of the product of the one point distributions $p(a)$ and $p(b)$ from the joint distribution $P_{ab}(n)$ and quantifies the information (measured in bit) obtained about one symbol through observing the other one [6]:

$$I(n) = \sum_{(a,b)\in\mathcal{A}^2} P_{ab}(n)\log_2 \frac{P_{ab}(n)}{p(a)p(b)}. \tag{30}$$

Expanding Eq. (30) in a Taylor series, see Ref. [28], in the case of weak correlations we can represent the entropy of additive Markov chain as a sum of mutual informations $I(n)$,

$$h_{\texttt{Symb}}(L) \approx h_0 - \sum_{n=1}^{L} I(n), \tag{31}$$

which seems to be natural owing to the chain additivity.

Note the two important properties of Eq. (28): the conditional entropy is a nonincreasing function of $L$ and it remains constant at $L \geqslant N$ due to the property Eq. (27).

The result presented by Eq. (28) does not depend explicitly on the memory function $F_{ab}(n)$ and probably may be applicable wider than just to the considered model.

### B. Entropy and correlations in numerical chains

For the numerical Markov chain, the normalized correlation function is $K(n) = C(n)/C(0)$, $C(0) = \langle X^2\rangle$. For small correlations we should put the zero-order approximation for $C(n)$, $C(n) \simeq C(0)\delta(n,0)$, then in the first approximation from Eq. (22) we have

$$K(n) \approx \Phi(n) = \sum_{x\in\mathcal{N}} x f_n(x). \tag{32}$$

Repeating the similar calculations as in Ref. [13], presented in Appendix B for the numerical random sequence, we obtain the following result:

$$h_{\texttt{Num}}(L) \approx h_0 - \frac{\langle X^2\rangle}{2\ln 2}\sum_{n=1}^{L}\sum_{x\in\mathcal{N}} \frac{f_n^2(x)}{p(x)}. \tag{33}$$

The source Shannon entropy, also known as the entropy rate, is the conditional entropy at the asymptotic limit, $h = \lim_{L\to\infty} h(L)$. This quantity measures the average information per symbol if *all* correlations, in the statistical sense, are taken into account.

### C. Equivalence of the symbolic and numerical high-order Markov chains

To compare these two results, Eqs. (28) and (33), we should introduce some additional properties of the memory functions. Without loss of generality, as is mentioned above, in Eq. (19) we can put $\langle X\rangle = 0$. Then the CPDF for the numerical chain takes the form

$$P\big(x|x_{i-N}^{i-1}\big) = p(x) + \sum_{n=1}^{N} f_n(x)x_{i-n}. \tag{34}$$

In the same manner, if we choose the normalization of the memory function in the form $\sum_{b\in\mathcal{A}} F_{ab}(r)p(b) = 0$, then the CPDF of the *additive* symbolic Markov chain, Eq. (17), is transformed into the following one:

$$P\big(a|a_{i-N}^{i-1}\big) = p(a) + \sum_{n=1}^{N} F_{a,\,a_{i-n}}(n). \tag{35}$$

The only possibility to relate the numerical chain with the corresponding symbolic chain is to put

$$F_{ab}(n) = f_n(x)y, \quad a \Leftrightarrow x, \quad b = a_{i-n} \Leftrightarrow y_{i-n} = y. \tag{36}$$

After such identification the CPDFs Eqs. (34) and (35) become equivalent. Now, using Eqs. (27) and (36) and the definition $\langle X^2\rangle = \sum_{x\in\mathcal{N}} x^2\, p\,(x)$, we can immediately see the equivalence of Eqs. (28) and (33). So, the entropy of the numerical chain equals to the entropy of the corresponding symbolic chain. Note that the obtained relation indicates the equivalence of the linear and symbolic models only in the case of weak correlations. Without this restriction, Eq. (27) establishes a one-way relationship between their memory functions. For any given $f_n(x)$, we can associate it with $F_{xy}(n)$, but not conversely: not every matrix-valued function $F_{xy}(n)$ depends on the argument $y$ multiplicatively. In other words, in general case, a linear numerical additive Markov chain, defined on a discrete state space, is a special case of a random symbolic sequence.

### D. Entropy of the numerical chain and its diagonal approximation

The result of previous section can hardly be considered as unexpected, because the both entropies are based on the same two-point PDF, $P_{ab}(n)$, and the equivalent CPDFs, Eqs. (34) and (35).

Note also that the description of the correlation properties of random sequences by the two-point PDF is not the only way. In physics the numerical correlation function,

$$C(n) = \langle (x_i - \langle X\rangle)(x_{i+n} - \langle X\rangle)\rangle, \tag{37}$$

is used more often, than the two-point PDF, cf. Eq. (15). Besides, the correlation function (or correlation coefficient) is the most commonly used statistical characteristic in the description of random phenomena. Keeping this in mind, the question whether it is possible to express the conditional entropy by means of the numerical correlation function, Eq. (37), seems to be worth answering. In work [18] it is demonstrated that the entropy of statistical mechanics and of information theory may be viewed as a measure of correlation. Nearly the same question is formulated in Ref. [20]: “Is the entropy a good measure of correlation?” It is hardly possible to accept the positive answer to this question because the entropy is a macroscopic quantity which cannot determine the microscopic characteristics such as the correlation functions.

It is better to ask and understand the question, how the entropy can be described in terms of the correlation functions. If we compare the number of parameters of the symbolic and numerical correlation functions, then we can argue that for the former is $NM$ and for the latter does $N$ only. It means that the entropy evaluated via the numerical correlation function cannot be lower than the symbolic entropy or the numerical entropy obtained via the two-point PDF.

To show that, let us consider the Cauchy-Bunyakovski-Schwarz inequalities

$$(A_1B_1 + \cdots + A_LB_L)^2 \leqslant \big(A_1^2 + \cdots + A_L^2\big)\big(B_1^2 + \cdots + B_L^2\big), \tag{38}$$

for different values of $L \geqslant 1$, where we put $A_i = \sqrt{p(x_i)}x_i$ and $B_i = f_n(x_i)/\sqrt{p(x_i)}$, then we change the sign and add $h_0$ in both sides of the equation. Then the right-hand side of this equation engenders the numerical entropy, Eq. (33), and the left-hand side does the expression

$$h_{\mathtt{Appr}}(L) \approx h_0 - \frac{1}{2\ln 2}\sum_{n=1}^{L} K^2(n). \tag{39}$$

This quantity can be associated with the entropy of the numerical Markov chain as the entropy approximation obtained with the use of the normalized correlation function $K(n)$. The similar to Eq. (39) expression was obtained earlier in Ref. [21] for the *binary* (dichotomous, $M = 2$) additive Markov chain. Note that this formula is exact (in the weak correlation regime) for the binary chains, and it is an evaluation for the chains with $M \geqslant 3$, where $M$ is the state space dimension; see Eq. (9). Thus, for all $L \geqslant 1$ we have

$$h_{\mathtt{Symb}}(L) = h_{\mathtt{Num}}(L) \leqslant h_{\mathtt{Appr}}(L). \tag{40}$$

For the finite-state numerical chains, the approximation Eq. (39), being in general case a function of the values of random variable, is no longer the entropy. The quantity $h_{\mathtt{Appr}}(L)$ becomes the conditional entropy only when it equals the numerical conditional entropy $h_{\mathtt{Num}}(L)$, which in its turn equals the symbolic entropy $h_{\mathtt{Symb}}(L)$ due to the definition. In Eq. (40) the equality takes place for the collinear vectors $A_i = \sqrt{p(x_i)}x_i$ and $B_i = f_n(x_i)/\sqrt{p(x_i)}$ if

$$A_i = \lambda(n)B_i \;\; \Leftrightarrow \;\; \sqrt{p(x_i)}x_i = \lambda(n)f_n(x_i)/\sqrt{p(x_i)}. \tag{41}$$

It can be seen that quantity $h_{\mathtt{Appr}}(L)$, Eq. (39), can take values between $h_0$ and its minimal value $h_{\mathtt{Num}}(L)$, Eq. (33), when the condition of vectors collinearity, $p(x_i)x_i = \lambda(n)f_n(x_i)$, is fulfilled.

To clarify the meaning of the vectors collinearity condition in terms of the correlation function, let us compare Eqs. (32) and (37) (with $\langle X\rangle = 0$). We see that $\sum_{j\in\mathcal{N}} x^jC_{x^ix^j}(n) = C(0)f_n(x^j) = C(0)\lambda(n)p(x^i)x^i$. This equality can take place if the symbolic correlation matrix is diagonal,

$$C_{x^ix^j}(n) = C(0)\lambda(n)p(x^i)\delta(i,j). \tag{42}$$

This diagonal correlation matrix $C_{x^ix^j}(n)$ describes a class of random chains with the correlations only among equal symbols-numbers when the correlation function depends on the probabilities only.

Summarizing, we can formulate the results of our consideration as following. The numerical correlation function can describe correctly the conditional entropy only for some class of random sequences with the diagonal correlation matrix $C_{x^ix^j}(n) \propto \delta(i,j)$. For the rest of the numerical chains the diagonal approximation $h_{\mathtt{Appr}}(L)$, being a function of the conventional correlators depending on the states of random variable, is not the entropy. This quantity can be treated as an alternative measure of the "information" (in a wide sense) conveyed by the numerical Markov chain. Inequality Eq. (40) is explained by the fact that $h_{\mathtt{Num}}(L)$ is determined by the PDFs which contain more information about the chain than the numerical correlation function determining $h_{\mathtt{Appr}}(L)$. Finally, condition Eq. (42) can be interpreted as a way to compensate for the dependence of the correlation function on the state, i.e., to make it a function of the probabilities only.

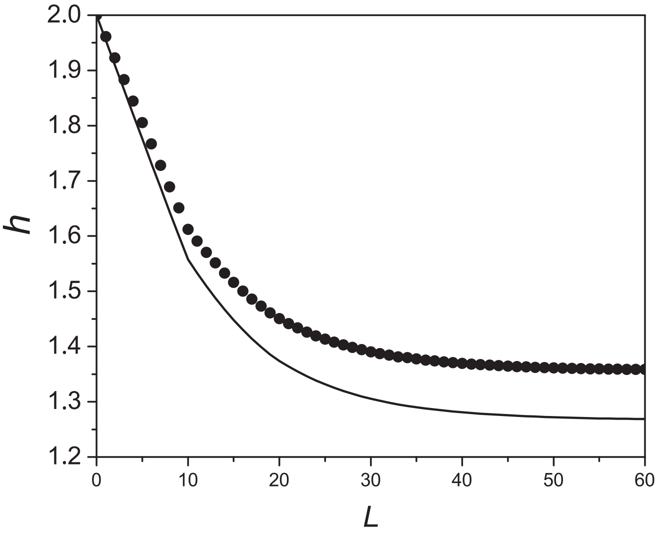

FIG. 1. The numerical conditional entropy $h_{\mathtt{Num}}(L)$(line) and its diagonal approximation $h_{\mathtt{Appr}}(L)$ (points) for the additive linear Markov chain with broken collinearity condition. The sequence size is $S = 10^6$, the memory depth is $N = 10$.

Given the redundancy of the information contained in the two-point PDF compared to the numerical correlation function, it is natural to assume that there is a whole range of sequences having different symbolic correlators for the same numerical chain. In Appendix C we demonstrate this assertion considering a linear additive Markov chain with numerical correlator strictly equal to zero while elements of the symbolic correlation matrix have nonzero values for $n \neq 0$.

## V. NUMERICAL SIMULATIONS

Figure 1 illustrates the main result of the previous section. The solid line shows the calculation for the entropy of the numerical, Eq. (33), and symbolic, Eq. (28), $h_{\mathtt{Symb}}(L) = h_{\mathtt{Num}}(L)$, additive linear Markov chains numerically generated by means of the same CPDF. The points demonstrate the calculated diagonal approximation $h_{\mathtt{Appr}}(L)$, Eq. (39). The symbolic and numerical correlators were estimated from the data by averaging over the generated chain. The alphabet consists of four characters with equal unconditional probabilities $p(a_i) = p(a^i) = p = 1/4$. To each character of the symbolic chain, the corresponding numerical value $x^i$ from the

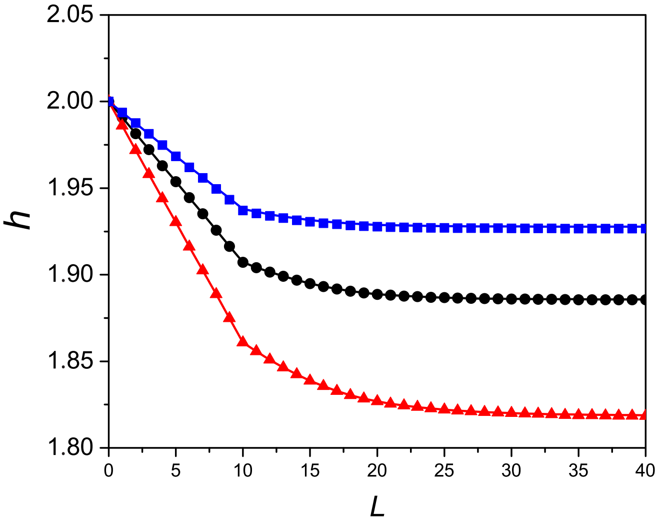

FIG. 2. The numerical conditional entropies $h_{\mathtt{Num}}(L)$ (lines) and their diagonal approximations $h_{\mathtt{Appr}}(L)$ (markers) for the three different additive linear Markov chains with collinearity condition, Eq. (41). The blue line and square symbols correspond to the coefficient $\lambda = 0.9$; $\lambda = 1$ corresponds to the black points and line; the red triangles with line present $\lambda = 1.1$.

numerical state space $\mathcal{N} = \{-0.2; -0.1; 0.1; 0.2\}$ is assigned. The sequence size is $S = 10^6$, the memory depth is $N = 10$. The weight functions $f_n$ are chosen such that the collinearity condition, Eq. (41), is violated:

$$\begin{aligned} f_n(-0.1) &= f_n(-0.2) = -0.125, \\ f_n(0.1) &= f_n(0.2) = 0.125. \end{aligned} \tag{43}$$

As expected, the conditional entropy of the numerical chain is lower than its diagonal approximation obtained with the use of the correlation function Eq. (37). This confirms the fact that the symbolic consideration of the numerical sequence contains more information than numerical.

Figure 2 shows the calculation of the numerical entropy and its diagonal approximation for sequences with fulfilled collinearity condition but different proportionality coefficients $\lambda$, Eq. (41). The blue line and square symbols correspond to the coefficient $\lambda = 0.9$; the black line and points are for $\lambda = 1$; the red line and triangles depict the calculation for $\lambda = 1.1$. For these sequences their numerical entropies and diagonal approximations are equal, $h_{\mathtt{Num}}(L) = h_{\mathtt{Appr}}(L)$. Recall, the numerical and symbolic entropies coincide.

It should be noted, that in both Figs. 1 and 2 the calculated conditional entropies are not constant beyond the order of the Markov chain $N = 10$, despite expectations. This effect is due to the two clear facts. First, the figures do not express the precise values of entropy, but their approximate ones Eqs. (28) and (39) via the correlator. Second, the correlation function in general case has more long tail, than the order of the chain: they coincide only in the case of weak correlations. The weaker the correlations, the closer to $N$ are the values $n$ where the conditional entropy becomes a constant. This is illustrated in Fig. 2: the lower red line and triangles correspond to the chain with the correlations larger and the upper blue line and squares correspond to the chain with the correlations smaller than those of the chain represented with the black line and points.

## VI. CONCLUSION

Considering symbolic and numerical random sequences in the framework of the additive Markov chain approach, we have expressed their conditional entropies by means of the two-point probability distribution functions. Then, in the assumption of weak correlations, we have evaluated the entropy of the numerical chain with the use of the numerical correlation function and have shown that this evaluation is correct only for the degenerated sequences with a diagonal correlation matrix, when the numerical correlation function is actually a function of probability. In a general case, this evaluation, due to the dependency of correlation function on the state, is no longer the proper entropy but its diagonal approximation only, which nevertheless can be considered as an alternative measure of information transmitted by the additive high-order Markov chain. The diagonal approximation exceeds the conditional entropy of the chain. Both results match in the case of the dichotomous Markov chain, when the Cauchy-Bunyakovski-Schwarz inequality turns into exact equality.

We have expressed the conditional entropy of the additive Markov chain of high order in terms of the nonlinear counterpart of the correlation function $C(n)$—the mutual information $I(n)$–which is an alternative measure of dependency of the elements in random sequences. We have presented a simple example of correlated random sequence with the numerical correlation function exactly equal to zero; see Appendix C.

The obtained results can be used for studies of the random sequences, in particular, the DNA and RNA sequences, texts and time-series of different nature.

## ACKNOWLEDGMENT

We express our special thanks to the referees for the constructive feedback and provided intellectual assistance which helped us to significantly improve the paper.

## APPENDIX A

Using the property of the symbolic correlation matrix Eq. (11),

$$C_{ab}(0) = p(a)\delta(a,b) - p(a)p(b), \tag{A1}$$

it is convenient to separate the term with $n' = n$ in Eq. (18). For $1 \leqslant n \leqslant N$,

$$C_{ab}(n) = \sum_{c\in\mathcal{A}} C_{ac}(0)F_{bc}(n) + \sum_{n'=1, n'\neq n}^{N} \sum_{c\in\mathcal{A}} C_{ac}(n-n')F_{bc}(n'). \tag{A2}$$

Simplifying the first term of this equation,

$$C_{ab}(n) = p(a)F_{ba}(n) + \sum_{n'\neq n} \sum_{c\in\mathcal{A}} C_{ac}(n-n')F_{bc}(n'), \tag{A3}$$

we obtain the recurrent relation for the memory function,

$$F_{ab}(n) = \frac{C_{ba}(n)}{p(b)} - \frac{1}{p(b)} \sum_{n' \neq n} \sum_{c \in \mathcal{A}} C_{bc}(n-n') F_{ac}(n'). \quad \text{(A4)}$$

In the case of weak correlations the second term in the right side of the equation is much smaller then the first one, then the first approximation for memory function is

$$F_{ab}(n) \approx \frac{C_{ba}(n)}{p(b)}. \quad \text{(A5)}$$

Substituting this result into the recurrent Eq. (A4) we obtain the second approximation for the $F_{ab}(n)$,

$$F_{ab}(n) \approx \frac{C_{ba}(n)}{p(b)} - \frac{1}{p(b)} \sum_{n' \neq n} \sum_{c \in \mathcal{A}} \frac{1}{p_c} C_{bc}(n-n') C_{ca}(n'). \quad \text{(A6)}$$

Now the conditions of small correlations can be formulated as the smallness of the second term in RHS of the equation compared to the first one.

## APPENDIX B

The conditional entropy $h(L)$ can be represented (e.g., using the chain rule, Ref. [23]) in terms of the conditional probability distribution function, $P(x_{L+1}|x_1^L)$,

$$h(L) = \sum_{x_1^L \in \mathcal{N}^L} P(x_1^L) h(x_{L+1}|x_1^L) = \langle h(x_{L+1}|x_1^L) \rangle, \quad \text{(B1)}$$

where $h(x_{L+1}|x_1^L)$ is the amount of information contained in the $(L+1)$th random element of the sequence conditioned on $L$ previous ones,

$$h(x_{L+1}|x_1^L) = - \sum_{x_{L+1} \in \mathcal{N}} P(x_{L+1}|x_1^L) \log_2 P(x_{L+1}|x_1^L). \quad \text{(B2)}$$

The conditional probability $P(x_{L+1}|x_1^L)$ for a subsequence of length $L < N$ can be obtained in the second approximation in the weak correlation parameter $\Delta(L, x_1^{L+1})$ from Eqs. (19) and (23) by means of a routine probabilistic reasoning, see some additional details in Ref. [13],

$$P(x_{L+1}|x_1^L) = p(x_{L+1}) + \Delta(L, x_1^{L+1}),$$
$$\Delta(L, x_1^{L+1}) \approx \sum_{n=1}^{L} f_n(x_{L+1}) x_{L+1-n}. \quad \text{(B3)}$$

Taking into account the weakness of correlations,

$$|\Delta(L, x_1^{L+1})| \ll p(x_{L+1}), \quad \text{(B4)}$$

in definition Eq. (19), expanding Eq. (B2) in Taylor series up to the second order in $\Delta(L, x_1^{L+1})$, using the evident property $\langle \Delta(L, x_1^{L+1}) \rangle = 0$, we get the conditional entropy of the numerical sequence in the form

$$h(L) \approx h_0 - \frac{1}{2 \ln 2} \sum_{x \in \mathcal{N}} \sum_{n=1}^{L} \sum_{n'=1}^{L} \frac{1}{p(x)} f_n(x) f_{n'}(x) C(n-n'). \quad \text{(B5)}$$

To obtain Eq. (33) we should replace the term $C(n-n')$ with $C(0)\delta(n, n')$, $C(0) = \langle X^2 \rangle$ when calculating the summation in Eq. (B5).

## APPENDIX C

For simplicity, we give an analytical description for the case of the first order Markov chain. A high-order example is built on the same principle, but its analytical description would be cumbersome. By analogy with the DNA nucleotide sequences, consider four-symbols chain, $M = 4$. Let all four characters $A, C, G, T$ be equally probable and the set $\mathcal{N}$ is the numbers $\{-2; -1; 1; 2\}$. In this case $\langle X \rangle = 0$, and expression for the conditional probability function Eq. (19) takes the form

$$P(x|x_1) = 1/4 + f(x) x_1. \quad \text{(C1)}$$

To complete formulating the model, it remains to determine four values $f(-2), f(-1), f(1), f(2)$ that are limited by two conditions. The first one follows from the additive linear model: $\sum_x f(x) = 0$, see Eq. (20). The second condition is introduced in such a way that numerical correlator Eq. (22) and the memory function which generates it, $\Phi(1)$, see Eq. (23), are zero:

$$\sum_x x f(x) = 0. \quad \text{(C2)}$$

It is easy to see that the both conditions can be satisfied by fixing arbitrary values of two weight functions, for example $f(-1)$ and $f(1)$, and expressing the other two through them:

$$f(-2) = -\frac{3f(-1)}{4} - \frac{f(1)}{4}, \quad f(2) = -\frac{f(-1)}{4} - \frac{3f(1)}{4}. \quad \text{(C3)}$$

So, the numerical correlator Eq. (37) of this chain is zero.

The symbolic memory function of the corresponding sequence of $(A, C, G, T)$ is defined by Eq. (36),

$$F_{x^i x^j} = - \begin{bmatrix} -2f(-2) & -f(-2) & f(-2) & 2f(-2) \\ -2f(-1) & -f(-1) & f(-1) & 2f(-1) \\ -2f(1) & -f(1) & f(1) & 2f(1) \\ -2f(2) & -f(2) & f(2) & 2f(2) \end{bmatrix}. \quad \text{(C4)}$$

The symbolic correlator in our first order case, from Eq. (18), turns out to be proportional to the memory function and, thus, it is also nonzero:

$$C_{yx} = \tfrac{1}{4} F_{xy}. \quad \text{(C5)}$$

[1] P. Ehrenfest and T. Ehrenfest, *Encyklopädie der Mathematischen Wissenschaften* (Springer, Berlin, 1911), p. 742, Bd. II.

[2] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, Cambridge, 1995).

[3] D. Salomon, *A Concise Introduction to Data Compression* (Springer, Berlin, 2008).

[4] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval* (Cambridge University Press, Cambridge, 2008).

[5] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory* 2nd ed. (Wiley, New York, 2006).

[7] A. E. Raftery, J. R. Stat. Soc. B **47**, 528 (1985).

[8] S. S. Melnyk, O. V. Usatenko, V. A. Yampol'skii, and V. A. Golick, Phys. Rev. E **72**, 026140 (2005).

[9] S. S. Melnyk, O. V. Usatenko, and V. A. Yampol'skii, Phys. A **361**, 405 (2006).

[10] S. S. Melnik and O. V. Usatenko, Comput. Biol. Chem. **53A**, 26 (2014).

[11] O. V. Usatenko, S. S. Apostolov, Z. A. Mayzelis, and S. S. Melnik, *Random Finite-Valued Dynamical Systems: Additive Markov Chain Approach* (Cambridge Scientific Publisher, Cambridge, 2010).

[12] Z. A. Mayzelis, S. S. Apostolov, S. S. Melnyk, O. V. Usatenko, and V. A. Yampol'skii, Chaos, Solitons Fractals **34**, 112 (2007).

[13] S. S. Melnik and O. V. Usatenko, Phys. Rev. E **93**, 062144 (2016).

[14] O. V. Usatenko, V. A. Yampol'skii, K. E. Kechedzhy, and S. S. Mel'nyk, Phys. Rev. E **68**, 061107 (2003).

[15] S. S. Melnik and O. V. Usatenko, Phys. Rev. E **98**, 042144 (2018).

[16] J. Weber, C. Zachow, and D. Witthaut, Phys. Rev. E **97**, 032138 (2018).

[17] T. O. Kvålseth, IEEE Trans. Syst. Man Cybernet. **17**, 517 (1987).

[18] J. H. Van Drie, arXiv:math-ph/0001024.

[19] F. Chapeau-Blondeau, Physica A **380**, 1 (2007).

[20] https://www.mimuw.edu.pl/~xliistat/slides/SkotarczakBedlewo2016.pdf.

[21] S. S. Melnik and O. V. Usatenko, Phys. Rev. E **90**, 052106 (2014).

[22] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics, Vol. 13 (American Mathematical Society, Providence, RI, 1996).

[23] M. Seifert, A. Gohr, M. Strickert, and I. Grosse, PLoS Comput. Biol. **8**, e1002286 (2012).

[24] O. V. Usatenko and V. A. Yampol'skii, Phys. Rev. Lett. **90**, 110601 (2003).

[25] A. N. Shiryaev, *Probability* (Springer, New York, 1996).

[26] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., Vol. 1 (McGraw-Hill Book Company, New York, 1991).

[27] V. E. Vekslerchik, S. S. Melnik, G. M. Pritula, and O. V. Usatenko, Phys. A **528**, 121477 (2019).

[28] H. Herzel and I. Grosse, Physica A **216**, 518 (1995).

[29] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981, 1992).

[30] N. G. van Kampen, Braz. J. Phys. **28**, 90 (1998).