Randomness and determinism, is it possible to quantify these notions?

PETRU CARDEI
Computer Engineering
INMA
6, Ion Ionescu de la Brad Blvd., Sector 1, Bucharest, 013813
ROMANIA

Abstract: - The article presents results obtained in attempts to quantify randomness characteristics for real numerical sequences or strings, using relative entropy. The characterization of the randomness of a series of real numbers is proposed to guide researchers in investigating phenomena towards deterministic or stochastic models. A numerical string's relative entropy is calculated using the histograms corresponding to the analysed strings, compared to the maximum entropy for the same histogram. It is shown that the entropy values have an asymptotic behaviour, but the relative entropy decreases with the increase in the number of histogram classes. Compared to other methods of characterizing the randomness of strings, which are not many, most of them being based on statistical tests, the method proposed in this article determines a better resolution for the classification of strings and, in addition, it can designate them as belonging to a class of randomness similar to that of some known strings, such as finite substrings of prime numbers, pseudorandom strings generated by common programs, trigonometric strings, etc.
The attempt to quantify the randomness of real numerical strings, the results of which are presented in this article, is a first step in characterizing the randomness of experimental numerical strings, this being the final goal of the investigations.

Key-Words: - Randomness, Determinism, Quantification, Relative Entropy, random sequences, random strings

Received: March 29, 2022. Revised: October 26, 2022. Accepted: November 21, 2022. Published: December 31, 2022.

1 Introduction

We started this research from a concrete problem, namely, given a discrete signal (of theoretical or experimental origin, a string or a sequence of real numerical data), we must decide whether it is random or deterministic. Also, a problem very close to the one mentioned is to decide if, between two signals designated to be random, one of them can be characterized by a "higher random intensity". In other words, is it possible to quantify the notion of randomness? This is the main problem of our research. The proposed method is tested on examples of strings or sequences recognized as random through the "vague" characterization of this notion, common in current scientific expositions.

The purpose of elaborating such quantification of numerical strings is first of all the classification as a deterministic or random string, in order to motivate subsequent theoretical approaches (dynamic modelling, optimization) in a deterministic style (using differential or algebraic models) or random using the theory of random functions and their statistical dynamics, possibly optimizations with random functions. Secondly, such a quantification of the degree of uncertainty (randomness, [12]) is a desideratum of knowledge and offers an argument for the theoretical space in which we fit the analysis of physical, social, biological or other phenomena.

In [5] it is shown that deterministic signals are a special category of frequency stationary signals and relatively constant amplitude over a long period of time. These can be expressed by an exact analytical relationship (formula), which leads to the precise determination of their value at any time. Such signals are not information bearers, they "do not say anything new", being absolutely predictable.

Also in [5], it was stated that random or non-deterministic signals are those whose evolution cannot be anticipated with certainty, such as vocal, video, seismic signals, etc. The unpredictability of the signals is positively correlated with the amount of energy transported. For example, the signal received during the transmission of news to a radio station is listened to with interest, due to its novelty. In the case of non-deterministic signals, so that the information can be received, the one who transmits it and the one who receives it uses the same language (code, alphabet, etc.). As shown in [5], the non-deterministic signal has specific characteristics, namely media, dispersion, global media, global dispersion, histogram, spectral power density, etc. The signal can have a certain degree of predictability of its evolution over time. Depending on certain characteristics of it, the non-deterministic signal can be:

• Stationary - media and dispersion do not depend on time, but are constant,

• Ergodic - the media on portions does not differ from the global average,

• White noise - has a constant spectral density throughout the frequency band.

The idea of quantifying randomness is not new. In 2017, the author of [6] stated, "Given the impossibility of the random true, the effort is directed to study the random degrees". Also, the same author shows that it can be proved that there is an infinite hierarchy (in terms of quality or power) of the forms of random.

According to [13], „a randomness test (or test for randomness), in data evaluation, is a test used to analyse the distribution of a set of data to see if it can be described as random (pattern less). In stochastic modelling, as in some computer simulations, the hoped-for randomness of potential input data can be verified, by a formal test for randomness, to show that the data are valid for use in simulation runs. In some cases, data reveals an obvious non-random pattern, as with so-called "runs in the data" (such as expecting random 0–9 but finding "4 3 2 1 0 4 3 2 1..." and rarely going above 4). "Also in [13], it is shown that the issue of randomness is an important philosophical and theoretical question. Tests for randomness can be used to determine whether a data set has a recognisable pattern, which would indicate that the process that generated it is significantly non-random. For the most part, statistical analysis has, in practice, been much more concerned with finding regularities in data as opposed to testing for randomness. Many "random number generators" in use today are defined by algorithms, and so are actually pseudo-random number generators. The sequences they produce are called pseudo-random sequences. These generators do not always generate sequences which are sufficiently random but instead can produce sequences which contain patterns. Stephen Wolfram used randomness tests on the output of Rule 30 to examine its potential for generating random numbers, [14] though it was shown to have an effective key size far smaller than its actual size [15] and to perform poorly on a chi-squared test [16]. The use of an ill-conceived random number generator can put the validity of an experiment in doubt by violating statistical assumptions. Though there are commonly used statistical testing techniques such as NIST standards, Yongge Wang showed that NIST standards are not sufficient. Furthermore, Yongge Wang [17] designed statistical–distance–based and law–of–the–iterated–logarithm–based testing techniques. Using this technique, Yongge Wang and Tony Nicol [18] detected the weakness in commonly used pseudorandom generators such as the well-known Debian version of the OpenSSL pseudorandom generator which was fixed in 2008.

Also [13] show that there have been a fairly small number of different types of (pseudo-)random number generators used in practice. They can be found in the list of random number generators, and have included: Linear congruential generators and Linear-feedback shift registers, Generalized Fibonacci generators, Cryptographic generators, Quadratic congruential generators, Cellular automaton generators, Pseudo-random binary sequences. These different generators have varying degrees of success in passing the accepted test suites. There are many practical measures of randomness for a binary sequence. These include measures based on statistical tests, transforms, and complexity or a mixture of these. A well-known and widely used collection of tests was the Diehard Battery of Tests, introduced by Marsaglia; this was extended to the TestU01 suite by L'Ecuyer and Simard. The use of the Hadamard transform to measure randomness was proposed by S. Kak and developed further by Phillips, Yuen, Hopkins, Beth and Dai, Mund, Marsaglia and Zaman, [19]. Several of these tests, which are of linear complexity, provide spectral measures of randomness. T. Beth and Z-D. Dai purported to show that Kolmogorov complexity and linear complexity are practically the same, [20], although Y. Wang later showed that their claims are incorrect, [21]. Nevertheless, Wang also demonstrated that for Martin-Löf random sequences, the Kolmogorov complexity is essentially the same as linear complexity.

The need to quantify an important characteristic of phenomena in the field of physics or in the field of numerical calculation has imposed the appearance of an important scientific discipline: quantifying uncertainty. According to [29], uncertainty quantification (UQ) is the science of quantitative characterization and reduction of uncertainties in both computational and real-world applications. It tries to determine how likely certain outcomes are if some aspects of the system are not exactly known. Many problems in the natural sciences and engineering are also rife with sources of uncertainty. Computer experiments by computer simulations are

the most common approach to studying problems in uncertainty quantification, [26], [27], and [28].
We must understand today that computer simulation has taken unacceptable proportions in research, which is felt in the quality of this activity. The enormous costs of the experiments required in serious research are more and more inaccessible to universities, research institutes and other institutions whose activity is research and design. At least in the academic field, in the scientific literature, the simulation tends to try at any cost to create the illusion that physical reality is very close to the virtual world. One can go so far as to violate reality in order to (hypothetically) adopt behaviour as (false) virtual reality. These are the reasons why I believe that in research and design, maximum caution is necessary for the use of simulation to solve concrete problems.
Uncertainty is sometimes classified into two categories, [31], and [32]. Aleatoric uncertainty is also known as stochastic uncertainty and is representative of unknowns that differ each time we run the same experiment, [33]. Also from [33], epistemic uncertainty is also known as systematic uncertainty, and is due to things one could in principle know but does not in practice. This may be because the measurement is not accurate, because the model neglects certain effects, or because particular data have been deliberately hidden. According to [33], in real-life applications, both kinds of uncertainties are present. Uncertainty quantification intends to explicitly express both types of uncertainty separately. The quantification for the aleatoric uncertainties can be relatively straightforward, where traditional (frequentist) probability is the most basic form. Techniques such as the Monte Carlo method are frequently used. A probability distribution can be represented by its moments (in the Gaussian case, the mean and covariance suffice, although, in general, even knowledge of all moments to arbitrarily high order still does not specify the distribution function uniquely), or more recently, by techniques such as Karhunen–Loève and polynomial chaos expansions. To evaluate epistemic uncertainties, efforts are made to understand the (lack of) knowledge of the system, process or mechanism. Epistemic uncertainty is generally understood through the lens of Bayesian probability, where probabilities are interpreted as indicating how certain a rational person could be regarding a specific claim. And also [33] summarizes the mathematical point of view: in mathematics, uncertainty is often characterized in terms of a probability distribution. From that perspective, epistemic uncertainty means not being certain of the relevant probability distribution, and aleatoric uncertainty means not being certain what a random sample drawn from a probability distribution will be.
Two important types of problems of the quantification of uncertainty are deeply involved in the particular field of the experimental measurement activity of the traction forces: the propagation of uncertainty, which is the quantification of the uncertainties at the outputs of the system, resulting from uncertainties and, the reverse quantification of the uncertainty, which involves calibration parameters or simply calibration.
A direct approach to the random or deterministic character for numerical or alphanumeric strings is found on the web page [22]. In 4.4 we used the program from [22] to compare its decision with the values of relative entropy for several strings of small length, only because the introduction of data into the program is difficult and the maximum length of the strings is limited. Essential is the fact that [22] makes a characterization of strings on classes of suspicion of random, but entropy gives values that characterize the string or lead it to the vicinity of a series with known behaviour, with whose randomness or determinism, the analysed string can be assimilated.
Considering the impressive theoretical and applied developments that addressed the notions of randomness, uncertainty and others from the same family of words, the only novelty in this attempt is the possible introduction of entropy as a measure of randomness or uncertainty, for now for a narrow category of some mathematical objects (real numerical strings, very common in experimental and theoretical-empirical techniques). In relation to other methods of studying the random sequences or strings, the one proposed in this article through relative entropy, although it gives a result which is obtained also by other methods (on classes of suspicion of random), the resolution is better or the random clusters can be narrowed. This means that the interval [0, 1], where the relative entropy varies, can be divided by those who perform the analysis. In addition, they have at their disposal special classes of random strings with which to make a comparison: the string of prime numbers, pseudo-random generated by various programs, and original strings, with the desired length. As I showed above, in the end, the distribution of a string analysed in random classes (sets), provides an orientation of the mathematical model of the phenomenon that generates the analysed string to the deterministic or stochastic approach.

2 A possible measure of random degree-the relative entropy

Next, the definitions of the estimators of the numerical strings that we will use in this research, and their mathematical formulas will be introduced: entropy and relative entropy.

2.1 Relative entropy and entropy

According to [1], in information theory, Shannon entropy or information entropy measures the uncertainty associated with a random variable. This measure also indicates the amount of information contained in the message. It is usually expressed in bits or in bits on the symbol. When expressed in bits, it represents the minimum length that a message must have to communicate the information.

It also represents an absolute limit of the best compression without loss applicable to communicated data: treating a message as a series of symbols, the shortest possible representation of the message has a length equal to the Shannon entropy on the symbol multiplied by the number of symbols of the original message.

For now, we will use, for the aim of this article, only the most common and usual definition of informational entropy, for example, in [3]:

$$E = -\sum_{i=1}^{n} p_i \log p_i \tag{1}$$

where p_i is the probability of the event, n is the volume of data of the random variable, and E is the entropy of the random string considered. So far we consider random strings of finite length and do not specify the basis of the logarithm except in the case of numerical results. The maximum value of entropy (1) is obtained in case the probabilities of all events are equal:

$$p_i = \frac{1}{n} \tag{2}$$

the maximum value of entropy being:

$$E_{max} = \log n \tag{3}$$

The relative entropy of a random variable is defined, as a percentage value, according to formula (4).

$$E_r = \frac{100E}{E_{max}} = \frac{-100\sum_{i=1}^{n} p_i \log p_i}{\log n} \tag{4}$$

It is important to note that:

e1) entropy is zero if at least one of the messages is reliable (i.e. has a probability $p_i = 1$);

e2) the entropy value is a real value, always greater than or equal to zero;

e3) the entropy of a source with two alternative events can vary from 0 to 1;

e4) entropy is an additive quantity: the entropy of a source whose messages consist of messages from several statistically independent sources is equal to the sum of the entropies of these sources;

e5) entropy will be maximum if all messages are equally likely.

The greater the entropy, the greater the uncertainty of the message transmitted by the random variable. According to [4], randomness is the characteristic of an uncertain event, which depends on future conditions, themselves uncertain. The same source characterizes uncertainty as something uncertain, and doubtful. Going along this line of vague human language, we can extrapolate the notions to the statement: entropy will be maximum at maximum uncertainty or random intensity. Small informational entropy (for example, in relation to the maximum value for the same volume of data) can be associated with reduced uncertainty and randomness. In other words, a random variable is "more random", the higher the entropy values.

If you work with finite strings of numbers, the probabilities involved in calculating the relative entropy are calculated starting from the histograms of the values of the strings analysed. For the calculation of the number of intervals of the string histograms, we adopted some rules used in many works, [8-11]. Numerical studies have shown that relative entropy depends on the number of classes considered when constructing the diagrams. I used, for some reference strings (table 1), six of the most well-known formulas for calculating the number of histogram classes. The results differed depending on the number of classes of the histograms, but not very much, not so much as to make a series with a random behaviour, to change to one with a deterministic character or vice versa. In order to give an image of how these numerical analyses and the relative entropy can be used in the decision of the random or deterministic character of some numerical strings, we averaged the relative entropy values obtained from the six results calculated using the specified formulas for the number of classes of histograms.

3 Assessment Tests

In order to understand the characterization that relative entropy gives to numerical strings, in this chapter we will comparatively analyse several cases of strings, as well as the effects of choosing the number of classes of the histograms of the tested strings.

3.1 Particular cases

In order to understand the behaviour of the relative entropy in describing the intensity of the randomness of the numerical data strings, some examples of relative entropy values for finite numerical strings from various sources are given in this chapter. Due to the large number of formulas proposed for calculating the number of classes of histograms, we calculated the relative entropy as an average of six of the most common formulas for the number of classes of histograms, which are found in table 1 from [11]. The strings for which the calculations were performed are listed in table 1.

Table 1 The average value of the relative entropy (E_r), for the six formulas for calculating the number of classes of the histograms, the coefficient of variation (CV) and the amplitude of the standard deviation above the average (ASDAM), for experimentally obtained strings (1, 2, 3), sequences obtained using generation programs for pseudorandom sequences (4, 5, 6), [7] and sequences of sinusoidal type (7, 8, 9) or exponential (13, 14, 15) and finite sub-sequences of prime numbers.

Index	E_r ,%	CV	ASDAM
1	91.943	0.234	2.384
2	90.320	-0.216	2.545
3	88.836	0.250	2.894
4	97.578	0.581	1.747
5	96.810	0.543	1.831
6	97.450	0.587	1.705
7*	94.746	0.566	1.414
8**	40.497	0.625	1.176
9***	81.202	1.529	2.441
10	97.900	0.649	1.841
11	98.152	0.641	1.813
12	98.287	0.636	1.805
13[1]	20.914	0.019	4.767
14[2]	37.670	0.026	3.327
15[3]	6.986	0.011	8.395

*Function $f(t) = A\sin(2\pi\nu t) + 1,\ A = 1, \nu = 1, t \in [0,30]$, with a sampling frequency of 10 samples per second, $t$being the time.

**Function $f(t) = [A\sin(2\nu t) + 1.6],\ A = 1.2\ , \nu = 1, t \in [0,30]$, with a sampling frequency of 10 samples per second, $[x]$being the whole part of the number x.

***The sum of five sinusoids with the amplitudes: 1.2, 2.3, -1.9, -0.31, 0.71, frequencies: 1, 2, 10.57, 7.0, 11.0 Hz, and the phases: 0, 0.1, 0.37, -0.53, -0.73. The sum of sinusoids is sampled with a sampling frequency of 10 samples per second.

[1]Function $f(t) = 0.1e^{-100(t-1)^2} + 1$.

[2]Function $f(t) = 0.1e^{-100(t-1)^2} + 0.1e^{-100(t-2)^2} + 1$

[3]Function $f(t) = 0.1e^{-1000(t-1.5)^2} + 1$.

Table 1 lists the average values of the relative entropy (for the six formulas for calculating the number of histogram classes), (E_r), the coefficient of variation (CV) and the amplitude of the standard deviation above the average, for experimentally obtained sequences (1, 2, 3), fig. 1, strings generated with generation programs for pseudorandom strings (4, 5, 6), [7], fig. 2, and sinusoidal strings (7, 8, 9), defined by formulas below the table, and finite substrings of prime numbers.

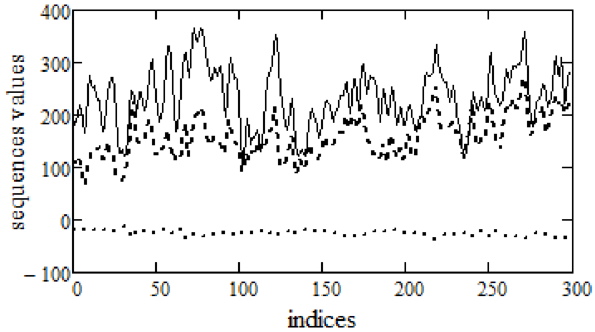

Fig. 1 The graphs of the three experimental records.

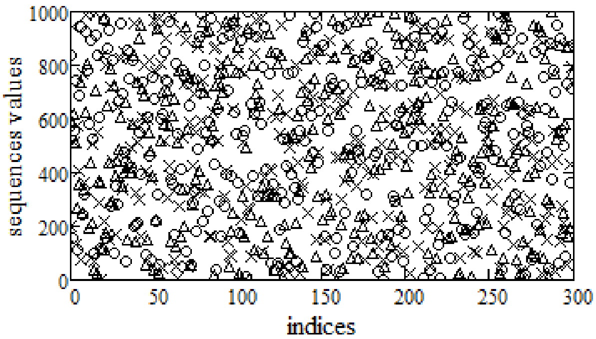

Fig. 2 Three pseudo-random sequences were generated with the program [7].

Also, classic random strings obtained from the string of prime numbers or from the discretization of some Gauss curves were introduced in the tests. The formulas for calculating the number of histogram classes, used in this study, were, according to [11]: Mosteller and Tukey's formula, 1977, Sturges' formula, 1926, Velleman's formula, 1976, and Scott's formula, 1979 and two control formulas of the behaviour of entropy relative, having the number of classes equal to the number of elements of the string, respectively with its half.

The ranking of the strings according to the value of the relative entropy is given in table 2. The indexation of the strings has been preserved as in table 1, and the observations with lowercase written after table 1 remain valid for table 2.

3.2 Comments on the relative entropy assessment

The results listed in tables 1 and 2 suggest some observations:

Table 2 The average value of the relative entropy (E_r), for the six formulas for calculating the number of classes of histograms, the variation coefficient (CV) and the amplitude of the standard over average, for the set of strings in table 1, sorted by relative entropy values.

Index	E_r ,%	CV	ASDAM
12	98.287	0.636	1.805
11	98.152	0.641	1.813
10	97.900	0.649	1.841
4	97.578	0.581	1.747
6	97.450	0.587	1.705
5	96.810	0.543	1.831
7*	94.746	0.566	1.414
1	91.943	0.234	2.384
2	90.320	-0.216	2.545
3	88.836	0.250	2.894
9***	81.202	1.529	2.441
8**	40.497	0.625	1.176
14[2]	37.670	0.026	3.327
13[1]	20.914	0.019	4.767
15[3]	6.986	0.011	8.395

O1) The highest values of the relative entropy in table 1 correspond to the finished sequences of prime numbers, and the relative entropy increases with the length of the sequences of prime numbers (it is easier to see in table 2);
O2) Values close to the entropy relative to those corresponding to the finished sequences of prime numbers, are obtained for the three strings generated with the help of the program access to [7], but between them and the sequences of prime numbers there is a noticeable difference;
O3) The sequences obtained from experimental records are characterized by values of relative entropy over 85% but at a remarkable distance to the pseudo-random strings obtained using [7];
O4) Between the strings given by sinusoidal formulas it is observed that the sinusoid with a single component, composed with the floor function, (the set of values of this function has only three elements), has the lowest value of relative entropy, 40.497. The sinusoidal sequence that comes from a pure sinusoid has a value maximum value of entropy, even higher than any of the strings from experimental measurements. The sinusoidal sequence with five components is positioned immediately after the pseudo-random strings.
O5) The most ordinary laws containing the smallest quantity of information are those obtained from the discretization of the Gauss curves, obviously from the collection of curves examined in this study.
O6) The relative entropy produces a strict order relationship on the set of the strings in table 1, which can be used for the purpose of ranking the intensity of the random character or of the uncertainty of the random strings described by such strings, as follows:
O6.1) It can be appreciated, at a first evaluation, that random variables or strings with relative entropy greater than 50% can be considered as suspects of random character, respectively those with entropy less or equal to 50% can be considered determinists;
O6.2) A higher resolution separation can be obtained by appreciating that strings characterized by relative entropy values lower than or equal to 33%, respectively, can be considered deterministic strings, and strings whose relative entropy is strictly higher than 66%, can be considered suspicious of intense randomness. Strings whose relative entropy is between 33% and 66% can be considered as having an undecidable character.
O6.3) Another criterion for appreciating the intensity of the random or deterministic character of a string of numerical data can be obtained by comparison with standard strings whose relative entropy is known and does not vary much with the number of classes of histograms used for the calculation of it. Thus, for example, the strings of pseudo-random numbers can be considered to be close to the random intensity of the finished sequences of prime numbers. The experimental strings are in the category of random but less random than the finished rows of prime numbers considered. The sequences obtained by discretization of the Gauss curves are characterized by a high degree of determinism. The sinusoidal sequences can be located in the area of random sequences, in the undecidable or deterministic area, according to the values of amplitudes, frequencies, or the way of generation (composition with the floor function, for example, or with generators of pseudo-random numbers).

3.3 Variation of relative entropy with the number of the histogram's classes

The variation of entropy relative to the number of classes of histograms poses difficult problems for analysis. First of all, we wonder if the number of classes of histograms increases, it can be reached in the situation that the random string changes the characteristic, from random to determinist, or vice versa (intuitive, the last option is unlikely)? The denominator of the fraction that defines the relative entropy (4), tends to infinity with the number of histogram classes. Starting from a certain number of classes, n^* begin to appear intervals that do not contain values of the string, so they have a zero probability. The existence of some classes of the histogram with zero elements implies belonging to the sum that defines the relative entropy of some meaningless terms, but which, at the limit, tend to zero. Therefore, intuitively, starting with the number of classes greater than or equal to n^*, the denominator of the relative entropy (4) increases, while the numerator should have an asymptotic behaviour, to a certain value, characteristic of the analysed string. As a result of this reasoning, I sought to stop the process of growing the number of classes of the histograms to a number near n^*. If $\{x_i\}_{i=1...N}$, it is the random string, the number of samples in each class of the histogram, either $m = \min_{i=1,...,N} x_i$, respectively $M = \max_{i=1,...,N} x_i$ (they can be equal, if the string is constant), then, assuming a number of the histogram, with equal intervals, we obtain the size or length of the histogram range: $\delta_n = \frac{M-m}{n}$. Suppose that the slightest distance of two terms of the string is:

$$\delta_{min} = \min_{i=1,...1N-1} \{|x_{i+1} - x_i|\} \quad (5)$$

where the symbol | |represents the magnitude operator or the module of a real number (or absolute value).Then one can be considered the number n^*, given, approximately, by the formula:

$$n^* = \left[\frac{M-m}{\delta_{min}}\right] + 1 \quad (6)$$

Details and examples are given in 4.2.

4 The general working algorithm for estimating the relative entropy of a string

Although the method of calculating the relative entropy of a string (random variable) is quite simple at first glance, there are important stages that require discussions and, very likely, choices that can influence the result.

4.1 Steps of the relative entropy calculation algorithm

A formulation as short as possible of the calculation algorithm for the relative entropy of a string is given in this sub-session.

E1. A finite numerical string with real components, $\{x_i\}_{i=1,...,N}$, is entered, where N is the volume of data or the length of the string or, again, the number of components. In addition, the descriptive statistics of the string x are calculated (average value, average standard deviation, coefficient of variation, and possibly other characteristics).

E2. A maximum number $n^*$of classes are chosen for the histograms used to evaluate the relative entropy (4), for example, N or the floor of a fraction of N, or a number of classes calculated according to N and, possibly, certain descriptive statistical characteristics of the string x, according to [11] for example. Alternatively, the procedure for determining a maximum number of classes can be used as in (5) and (6) from 2.3.

E3. Entropy and relative entropy are calculated for each $k = 2, ..., n^*$.

E4. A selection criterion of a number k^* is applied, which satisfying this criterion, designates the histogram to which the relative entropy of the string x will correspond. This criterion is based on the entropy curve - the number of classes of the histogram, the discrete curve obtained in step E3. In 2.3 it was explained why this curve is taken and not the discrete curve relative entropy –the number of classes of the histogram. The criterion for obtaining the number of classes k^*can be formulated in various ways:

E4.1 It is chosen for k^*, that value of the number of classes for which the module or the magnitude of the average value of one or more consecutive differences of the entropy series $\{E_k\}_{k=1,...,N}$ does not exceed an arbitrarily chosen limit, possibly a fraction of the corresponding maximum entropy, for example. The minimum value of the index at which this criterion is met is chosen for k^*. This criterion was used to obtain the results from Tables 1 and 2. It is a criterion that must be assisted by the operator on the computer because the entropy variation is not strictly monotonous (at least at the current level of the algorithm). For example, for the results presented in Tables 1 and 2, we worked with three consecutive differences between the terms of the entropy string, and for the results in 4.3, we worked with a single consecutive difference in the terms of the same string.

E4.2 A criterion that avoids the small monotony variations of the discrete curve entropy - number of classes of the histograms, uses the interpolation of this curve through a continuous exponential curve with horizontal asymptote towards infinity. On this continuous curve, the criterion for determining the number k^* is set by imposing an arbitrarily chosen limit on the slope of the interpolation curve. This criterion was used to obtain the results presented in 4.3. It should be noted that not in all cases the continuous curve with the horizontal asymptote at plus infinity, succeeds in interpolating the entropy series. In cases with random strings that are formed with few values, for example, or are concentrated on narrow ranges of numbers, the interpolation can be done with rational power type curves or with discontinuous curves and for this reason, the proposed algorithms must be assisted by the human operator in thing, for now.

4.2 A way of choosing the minimum number of classes for the histogram used in the evaluation of relative entropy, for example, the string of the first 1000 prime numbers

In order to study the influence of the number of classes of the histogram with which the entropy of a random string is calculated, I took as an example the series of 1000 primary numbers. Since the first prime number is two, and the 1000th is 7919, the minimum distance between two of the elements of this string is 1 (the distance between 2 and 3). The criterion of the number of classes where each class contain only one element explained in the relations (5) and (6), leads to a number of 7917 classes. According to the method presented in 3.1, the histogram, probabilities, and finally, entropy, relative entropy and other characteristics of the string of the first 1000 prime numbers are calculated.

According to the criterion for calculating the number of classes of the histograms, we calculated the entropy, the maximum entropy and the relative entropy for all numbers from 2 to 7917. Additionally, in order to provide some statistical complements, is calculated: the coefficient of variation and the amplitude of the standard deviation above the average. In fig. 3 and 4, it can be observed that the entropy increases monotonically asymptotically, while the relative entropy decreases monotonically with an unclear asymptotic trend. For this reason, the selection criterion of the "optimal" number of classes (intervals) necessary for a "good enough" assessment of the relative entropy was based on the variation of the entropy and not on the relative entropy with the number of classes of the histogram. Thus, the stopping criterion is given by the condition that the optimal number of classes is the lowest number of classes for which the sum of three (a higher or lower number can be used) consecutive differences between the entropy values are lower than a number which is arbitrarily chosen number. In the case of this study, we chose:

$$\varepsilon = \frac{\max(E)}{0.1 \cdot N_{hmax}} \tag{7}$$

where E is the entropy of the string, and N_{hmax} is the maximum number of the classes of tested histograms, 7917. In the case of this numerical test, we use the value ε=0.013 which is obtained, for the maximum value of the entropy, $\max(E) = 9.966$. The stopping value is a bit exaggerated (below 0.126% of the maximum entropy value), but this numerical study is only an example. For the histogram with 131 classes, $E_{\max 131}$=7.033 and $E_{r\,131} = 99.269\%$, were obtained, 1% higher than the value in tables 1 and 2.

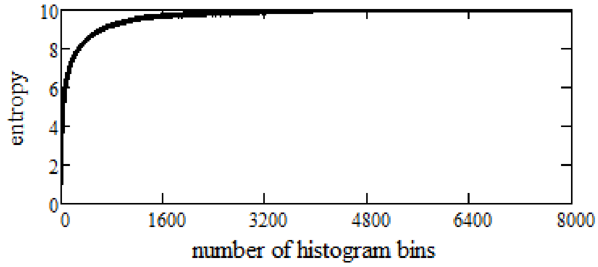

Fig. 3 Dependence of the entropy of the sequence on the number of classes of the histograms.

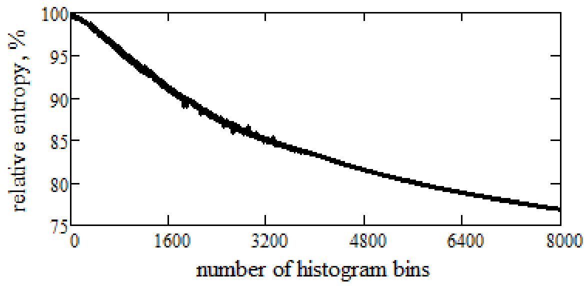

Fig. 4 Dependence of the relative entropy of the sequence on the number of classes of the histograms.

4.3 A procedure for choosing the number of classes of the histograms used to estimate the relative entropy based on the interpolation of the dependence curve of entropy by the number of classes of the histogram

This appendix gives the results of the relative entropy evaluation for the strings in table 1, obtained using a criterion for determining the number of classes of the histogram, type E4.2, presented in chapter 3.

The algorithm used to obtain the results presented in 4.3 uses the discrete entropy curve - the number of classes as described in chapter 3, E4.2, combined in certain cases (the only concrete one among those included in the collection of evaluated strings) with the selection algorithm of the number of histogram classes given in E4.1. A curve of dependence between entropy and the number of classes of the histogram is given in fig. 6. This curve corresponds to the experimental data (tensometry records) from fig. 5, included in the analysed collection listed in table 1, at position 1.

By putting the condition that the minimum number of classes of the histogram to be the smallest abscissa, for which the slope has less than 1°, on the curve in fig. 5, are obtained the next results $n^* = 74$, $E = 5.15$, $E_{max} = 6.209$, $E_r = 93.641\%$ are obtained.

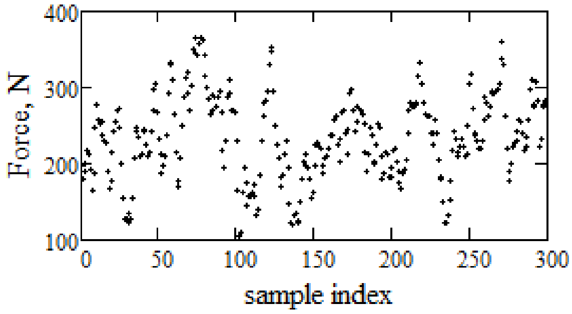

Fig. 5 Graphical representation of the sequences of experimental data described by the series in table 1, position 1.

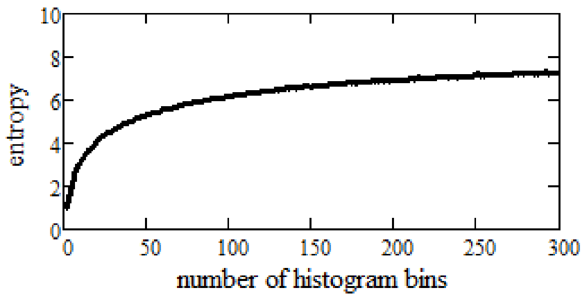

Fig. 6 The discrete dependence curve, entropy - number of histogram classes, for the string in table 1, position 1.

The solution which uses the selection algorithm from E4.1, with the same constant (the tangent of 1-degree angle), leads to the solution: $n^* = 156$, $E = 6.675$, $E_{max} = 7.304$, $E_r = 91.395\%$, if used for the number of classes of histogram the criterion of the average of three consecutive differences of entropy. If the selection is made using the simple successive differences between the elements of the entropy string, the solution obtained is the next: $n^* = 45$, $E = 5.183$, $E_{max} = 5.555$, $E_r = 93.315\%$. The limit value used is the same approximation of the tangent of the angle of 1° (0.017).The results of the running of the calculation programs based on the criteria described in E4.1 and E4.2, with the additional clarifications, are given in Table 3 and 4.

The two calculation algorithms for the relative entropy of the strings belonging to the collection taken as an example produce the same ranking of randomness, [12]. Only the characteristic values differ, but not significantly, see Table 5.

Table 3 Results of entropy calculation for the set of sequences from table 1, using criteria E4.1.

Sequence Index, table 1	n_h	E	E_{max}	E_r
1	45	5.18	5.55	93.31
2	51	5.28	5.73	92.19
3	65	5.52	6.07	90.96
4	55	5.72	5.83	98.02
5	50	5.55	5.70	97.45
6	60	5.83	5.95	97.94
7	75	5.97	6.27	95.29
8	3	1.58	2.32	68.26
9	63	5.59	6.02	92.88
10	62	5.91	6.00	98.49
11	85	6.39	6.44	99.11
12	76	6.25	6.28	99.39
13	13	0.86	3.91	21.96
14	30	1.89	5.00	37.82
15	4	0.23	2.58	8.93

Table 4 Results of entropy calculation for the set of strings from table 1, using criteria E4.2.

Sequence Index, table 1	n_h	E	E_{max}	E_r
1	74	5.81	6.21	93.64
2	73	5.72	6.19	92.40
3	73	5.64	6.19	91.05
4	75	6.10	6.23	97.95
5	72	5.98	6.17	96.97
6	73	6.05	6.19	97.83
7	81	6.04	6.34	95.31
8*	5	1.58	2.32	68.26
9*	74	5.80	6.21	93.46
10	77	6.18	6.27	98.55
11	82	6.31	6.36	99.24
12	85	6.37	6.41	99.36
13	5	0.59	2.32	25.49
14	27	1.86	4.75	39.17
15	6	0.23	2.58	8.93

*Random strings for which, in the case of the E4.2 algorithm, the interpolation was done by discontinuous functions, second-degree polynomial for the case of histograms with two and three classes and constant ceiling for more than three classes.

It can be observed that, compared to the randomness ranking of the strings in table 2, the changes are small, namely, the string of five sinusoids is inserted between the first two strings of experimental origin, increasing in relative entropy value. From the point of view of the category, there are no transitions from the class of random strings to that of deterministic strings or vice versa.

Table 5 The randomness ranking for the fifteen analysed sequences.

Sequence rank	E_r, E4.1	E_r, E4.2
10	99.387	99.365
11	99.115	99.244
12	98.487	98.554
4	98.017	97.947
5	97.942	97.829
6	97.447	96.966
7	95.287	95.306
1	93.315	93.641
9	92.884	93.457
2	92.189	92.405
3	90.963	91.052
8	68.261	68.261
14	37.819	39.17
13	21.963	25.488
15	8.926	8.926

4.4 The relationship between the characterization of strings by relative entropy and the characterization of the same strings using randomness statistical tests

An important opinion on the random or deterministic character of the numerical strings (including alphanumeric strings, using numerical encoding) also expresses the statistics, through random tests, [13]. In order to compare the results of testing randomness (uncertainty or determinism) of some strings by relative entropy with the results of testing by statistical tests of randomness, in this appendix an example of characterization is given for 11 strings, some of them from the list given in table 1, others elaborated to highlight the differences of viewpoints and cover all interesting cases of the randomness test used. The randomness test used is a free online test, [22]. The theoretical foundations of the randomness statistical test are presented in [22]. This test requires manual data entry, so we limited the length of the strings to 20 elements. From the strings of the collection given in table 1, we took only the first 20 elements. I rescaled the sinusoidal or Gaussian series so that the discrete curves keep the visual identity of the curve. I introduced three new strings: two to highlight the characterization of some reference strings (the constant string, the triangular string), and the third, a string "as random as possible", created by the author, in order to cover the limit characterizations of the statistical test, fig. 7.

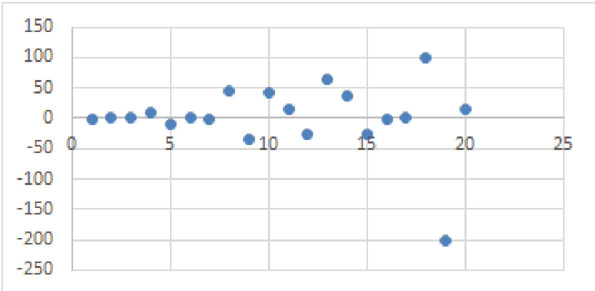

Fig. 7 Random string used to cover the "Little or no real evidence against randomness" characterization case of the statistical test.

The characterizations given by the statistical test program of the random character of the strings, parallel to the relative entropy value, are shown in table 6.

Table 6 Characterizations of the randomness of some numerical strings, using statistical tests and using relative entropy.

Sequence	Decision	E_r
the sequence of the first 20 prime numbers	**a**	96.749
the sequence of the first 40 prime numbers	**e**	98.692
the sequence of the first 60 prime numbers	**e**	98.361
sequence of 20 pseudo randomly generated numbers	**b**	99.636
sinusoidal string resolution 20 /s in, 1 s	**a**	93.498
Gauss bell position 13, table 1, 20 points, time between 1.375 s and 1625 s	**a**	96.749
discrete sinusoid with sampling frequency 20/s, for 1 s	**b**	67.657
constant string, all terms are 0.05, 20 components	**e**	0
string of 20 points in which all terms are 0.05 except terms 9, 10 and 11, which have the values 0.5, 1 and 0.5 (triangular)	**a**	32.197
the first 20 elements of the first experimental sequence - position 1, table 1	**a**	90.835
some sequence - see fig. A3.1	**d**	67.555

It is observed that the decisions of the program used in [22] for the randomness test of numerical or alphanumeric strings are divided into five classes: **a-** Very strong evidence against randomness (trend or seasonality), **b**-Moderate evidence against randomness, **c**-Suggestive evidence against randomness, **d**-Little or no real evidence against randomness, **e**-Strong evidence against randomness.

5 Conclusion

The numerical investigations carried out so far in the problem of quantifying the random or deterministic character (uncertainty or determinism), show that the research deserves to be continued, using the entropy of random strings, calculated on the histograms of these strings and the probabilities calculated with their help.
The relative entropy of strings can produce a ranking on the set of sequences and a sequence can be designated as belonging to a class of random or deterministic strings, possibly undecidable. I described this classification in the observations in chapter 2.2.
Another possibility of describing the random or deterministic character is a relative one, by comparing or associating an analysed string with an already studied string having both close relative entropy values. For example, in the rankings produced in this article in the collection of considered strings, the sinusoidal string is located in the immediate vicinity of the random strings generated with pseudo-random string generation programs.
I repeat the importance of designating a series of data as having a random or deterministic character consists in obtaining an argument for which the model of the phenomena characterized by such series will be oriented, towards the approach with random models (description within the theory of random functions) or towards deterministic models (the classical framework of the majority of the usual models in classical mechanics).
Finally, the answer to the question of the title of the article seems, at least for now, to be affirmative or at least promising.
Obviously, many problems remain to be studied in order to clarify the problem of quantifying the random or deterministic nature of numerical strings. Among them, first of all, there are those related to the calculation algorithms used, the criteria for choosing the number of classes of the histograms, and the way to perform the selection (discrete or continuous). The most severe limitation of operator intervention in the numerical schemes of these algorithms (solving nonlinear equations and/or nonlinear minimization of some selection functions) is an important objective.
The completion of the classification of the degree of randomness of the strings is also related to the development of other estimators that we suggested and calculated in the developed algorithms: the coefficient of variation and the amplitude of the standard deviation above the average. They introduce a direct connection between the problems of autocorrelation and the correlation of signal fragments, respectively with the theory of signal coding and decoding. For these reasons, our direction of research on the issue remains current.
A slightly more distant objective, which I have already approached in the beginning, is the complete transition of the problem of discrete strings to the study based on continuous functions, starting from the interpolating of histograms. It is important to note how high the precision of this alternative can be in relation to the purely discrete method, that was used first.
This study can now be used to estimate the randomness of some physical phenomena, for example, the tensile strength of some tillage machines. This could give precise recommendations for the terms in which the answer to experimental and theoretical research must be formulated.
Continuations of the investigations, in any of the ways exposed, or others, will be done only to the extent that there is interest in this problem, especially considering that the development of a mathematical model of some phenomena within the theory of random functions could be received with some reservations by the specialists involved in the related engineering field.

References:

[1] https://ro.wikipedia.org/wiki/Entropie_informa%C8%9Bional%C4%83, last access 20.12.2022
[2] Shannon, C. E., A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol. 27, No. 3, 1948, pp. 379–423.
[3] Iosifescu M., Moineagu C., Trebici V., Ursianu E., *Mica enciclopedie de statistica*, Editura Stiintifica si Enciclopedica, Bucuresti, 1985.
[4] https://dexonline.ro/definitie/aleatoriu, last access 20.11.2022.
[5] Mobil Industrial AG, www.mobilindustrial.ro/current_version/online_docs/COMPENDIU/semnale_deterministe_si_nedeterministe_aleatorii_htm, last access 22.11.2022.
[6] Calude C.S., Quantum Randomness: From Practice to Theory and Back, *Springer International Publishing AG, S.B. Cooper, M.I. Soskova (eds.), The Incompatibile, Theory and Application*, 2017, pp. 169–181.
[7] Random.org, www.random.org/strings/?num=100&len=3&di

gits=on&unique=off&format=html&rnd=new, last access 28.11.2022.
[8] ro.mcfairbanks.com/1350-histogram-formula, last access 24.11.2022.
[9] www.umfcv.ro/files/b/i/Biostatistica%20MG%20-%20Cursul%20IV.pdf, last access 21.12.2022.
[10] invatatiafaceri.ro/uncategorized-ro/formula-histogramei/, last access 22.12.2022.
[11] Doğan N., Doğan I., Determination of the number of bins/classes used in histograms and frequency tables: a short bibliography, *Journal of Statistical Research*, vol. 7, No. 2, 2010, pp 77-86.
[12] dexonline.ro/definitie/aleatorism , last access 19.12.2022.
[13] en.wikipedia.org/wiki/Randomness_test, last access 19.12.2022.
[14] Wolfram, S., A New Kind of Science, *Wolfram Media, Inc.*, 2002, pp. 975–976.
[15] Meier W., Staffelbach O., Analysis of pseudo random sequences generated by cellular automata, *Advances in Cryptology: Proc. Workshop on the Theory and Application of Cryptographic Techniques, EUROCRYPT '91. Lecture Notes in Computer Science*, 1991, pp. 547-186.
[16] Sipper M.; Tomassini M., (1996), Generating parallel random number generators by cellular programming, *International Journal of Modern Physics C*, Vol. 7, No. 2, 1991, pp. 181–190.
[17] Wang Y., On the Design of LIL Tests for (Pseudo) Random Generators and Some Experimental Results, http://webpages.uncc.edu/yonwang/ , 2014.
[18] Wang Y., Nicol T. (2014), Statistical Properties of Pseudo Random Sequences and Experiments with PHP and Debian OpenSSL, *Esorics 2014*, LNCS 8712,2014, pp. 454–471.
[19] Ritter T., Randomness tests: a literature survey, webpage: CBR-rand, last access 27.12.2022.
[20] Beth, T, Dai, Z-D., On the Complexity of Pseudo-Random Sequences or: If You Can Describe a Sequence It Can't be Random, *Advances in Cryptology EUROCRYPT '89, Springer-Verlag*, 1989, pp. 533-543.
[21] Wang Y., Linear complexity versus pseudorandomness: on Beth and Dai's result, *Proc. Asiacrypt 99 LNCS 1716, Springer Verlag*, 1999, pp. 288-298.
[22] home.ubalt.edu/ntsbarsh/business-stat/otherapplets/Randomness.htm
[23] en.wikipedia.org/wiki/NaN
[24] Kenneth B., An Introduction to Programming with IDL: Interactive Data Language, *Academic Press.*, 2006, pp. 26.
[25] Press W.H., Teukolski S. A.; Vetterling, W. T.; Flannery, B. P., Numerical Recipes: The Art of Scientific Computing, *Cambridge University Press.*, 2007, p. 34.
[26] Sacks, J., Welch, W. J.; Mitchell, T. J.; Wynn, H. P., Design and Analysis of Computer Experiments, *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–423.
[27] Iman, R. L., Helton, J. C., An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models, *Risk Analysis, Wiley*, Vol. 8, No.1, 1988, pp. 71–90.
[28] Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P.; van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated Assessment. Swets&Zeitlinger Publishers*, Vol. 4, No. 1, 2003, pp. 5–17.
[29] Saouma V., Hariri_Ardebili M., A., Uncertainty Quantification, Aging, Shaking, and Cracking of Infrastructures, From Mechanics to Concrete Dams and Nuclear Structures, 2021.
[30] en.wikipedia.org/wiki/Uncertainty_Quantification, last access 25.12.2022.
[31] Der Kiureghian, A., Ditlevsen, O., (2009). "Aleatory or epistemic? Does it matter?", *Structural Safety*, Vol. 31, No. 2, 2009, pp. 105–112.
[32] Matthies, H. G., Quantifying Uncertainty: Modern Computational Representation of Probability and Applications, Extreme Man-Made and Natural Hazards in Dynamics of Structures, *NATO Security through Science Series*, 2007, pp. 105–135.
[33] en.wikipedia.org/wiki/Uncertainty_quantification#Aleatoric_and_epistemic, last access 23.12.2022