

INDICES OF REGULARITY AND INDICES OF RANDOMNESS FOR m -ARY STRINGS

OSVALDO SKLIAR* RICARDO E. MONGE† GUILLERMO OVIEDO‡
VÍCTOR MEDINA§

Recibido/Received: 20 Feb 2008 — Versión revisada/Revised version: 26 Mar 2008
— Aceptado/Accepted: 25 Jul 2008

Abstract

The notions “regularity index” and “randomness index” previously introduced for binary strings (2-ary) have been modified slightly and generalized for m -ary strings ($m = 2, 3, 4, \dots$). These notions are complementary and the regular/random dichotomy has been replaced by a gradation of values of regularity and of randomness. With this approach, the more regular an m -ary string, the less random it is, and vice versa. The distributions of frequencies of different length strings —2-ary and 3-ary strings— according to their indices of randomness, are shown by histograms.

Keywords: regularity index, randomness index, m -ary strings.

Resumen

Las nociones de índice de regularidad y de índice de aleatoriedad previamente introducidas para cadenas binarias (2-arias) son modificadas ligeramente y generalizadas para cadenas m -arias ($m = 2, 3, 4, \dots$). Dichas nociones resultan complementarias y la dicotomía regular-aleatorio es sustituida por una gradación de valores de regularidad y de aleatoriedad. Con el enfoque utilizado, cuanto más regular es una cadena m -aria menos aleatoria debe ser considerada y viceversa. Las distribuciones de frecuencias de cadenas —de diversas longitudes— 2-arias y 3-arias en función de sus índices de aleatoriedad son presentadas mediante histogramas.

Palabras clave: índice de regularidad, índice de aleatoriedad, cadenas m -arias.

Mathematics Subject Classification: 49J55.

*Escuela de Informática, Universidad Nacional, Heredia, Costa Rica. E-Mail: oskliar@racsa.co.cr

†Universidad Interamericana, Heredia, Costa Rica. E-Mail: rmongeg@uinteramericana.edu

‡Universidad Latinoamericana de Ciencia y Tecnología (ULACIT), San José, Costa Rica. E-Mail: oviedogmo@gmail.com

§Escuela de Matemática, Universidad Nacional, Heredia, Costa Rica. E-Mail: vmedinabaron@yahoo.es

1 Introduction

The notions of regularity index and randomness index were previously introduced for binary strings [4]. The objective of this article is to generalize these notions for m -ary strings.

The term binary string is used to refer to a succession of digits, each of which can be a 0 (zero) or a 1 (one). Although here by convention it is accepted that each element of the binary string is a zero or a one, this choice is irrelevant; using any two types of elements whatsoever, such as “a” and “b”, would be equivalent.

The increasing order of digits in a binary string will also be established, by convention, from left to right. Thus, for instance, 0 0 1 0 1 is a binary string whose first digit is 0, the second is also 0, the third 1, the fourth 0 and the fifth 1.

An m -ary string will refer to a succession of elements each of which can be one of m different numbers. Thus, for example, if $m = 6$, then 0, 1, 2, 3, 4 and 5 will be the possible constituents of that string. In place of those numbers, other types of elements could be used, such as “a”, “b”, “c”, “d”, “e” and “f”.

Of course, a binary string can be considered a particular case of an m -ary string: that in which $m = 2$.

The increasing order of the elements in an m -ary string will be established by convention from left to right. Therefore, for example, 2 0 0 1 2 2 0 is a 3-ary string whose first digit is 2, the second is 0, the third is also 0, the fourth 1, the fifth 2, the sixth is also 2 and the seventh is 0.

The length of an m -ary string will be denoted by l_{str} (i.e., the number of elements in that string); for example, $l_{str} = 7$, for the 4-ary string 0 1 3 3 2 0 3.

Suppose that a scientist is asked how random the binary string 1 is. He might respond that in order to answer that question, he first requires a criterion to be able to assess the randomness of a binary string quantitatively. Second, he might add that this specific string – all of whose elements are “ones” – does not appear random but actually quite regular. The same answer could apply if reference is made, for example, to a binary string whose $l_{str} = 500$, and all of whose elements are “zeros”. Although to a somewhat lesser extent than in the above binary strings, others may also be considered “quite regular”, or “not too random”, such as

0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1

or

0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1.

The concepts of randomness and regularity will be considered complementary in the sense that the more regular the m -ary string, according to a criterion to be indicated below, the less random it will be. Similarly, as seen below, if an m -ary string is considered to be highly random (in naïve but intuitive terms, for now), it will not be too regular. For an alternative and clearly dichotomic approach, see [2] and [1], for example.

2 Review of the notions of regularity index and randomness index for binary strings

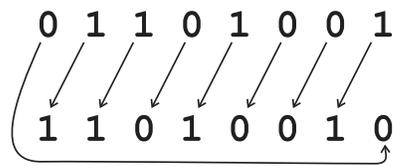
For those familiar with [4], it is important to clarify that several indices of regularity and their corresponding indices of randomness were introduced in that article. One of these regularity indices was that called “index of maximum regularity”. For the present proposal, however, it was deemed more suitable, first of all, to modify the characterization of the “index of maximum regularity”, and second, to use the term “index of regularity” to refer to this “enhanced” version of the “index of maximum regularity”.

For reasons of expository simplicity, attention will be given first to binary strings such that $l_{str} = 4n$, where $n = 1, 2, 3, \dots$; in other words, initially, the lengths of the strings considered will be multiples of 4.

Thus, let a string of the type specified above $i_1, i_2, i_3 \dots i_{l_{str}}$ (such that each element i_j , where $j = 1, 2, \dots, l_{str}$) be equal to either 0 or 1. This is the string for which the regularity index and the randomness index should be determined. In the first place, consideration will be given to another binary string written under it, obtained with the following algorithm:

- (a) for $j \neq l_{str}$, the j -th element of the new string will be the same as the $(j + 1)$ -th element of the given string, and
- (b) for $j = l_{str}$, the j -th element of the new string will be the same as the first element of the given string.

As an example, consider the following string for which $l_{str} = 8$: 0 1 1 0 1 0 0 1. Below we find the same string, and the new one obtained from the given one by applying the rule indicated above is written right under it.



The arrows illustrate again, graphically, the rule applied to obtain the new string. If the arrows (used for explanatory purposes only) are eliminated, the two strings would be specified as follows:

0 1 1 0 1 0 0 1
 1 1 0 1 0 0 1 0.

Together these two strings can be considered a set of dyads – 8 in this case – such that the first element of each dyad belongs to the given string and the second to the new string. The specific dyads to which reference was made are enclosed below.

0	1	1	0	1	0	0	1
1	1	0	1	0	0	1	0

Hence these dyads are:

(0,1), (1,1), (1,0), (0,1), (1,0), (0,0), (0,1), (1,0).

It is obvious that each of the preceding dyads must be one of the following four possible types of dyads: (0,0), (0,1), (1,0) and (1,1).

The new string was obtained from the given string by a shift (i.e., by generating a “displacement” d) of only one element in the given string ($d = 1$). What new string would have been obtained by generating a displacement of two elements ($d = 2$)? The answer is indicated below, with the first string being the given one and the second string the new one generated by displacing two elements:

```
0 1 1 0 1 0 0 1
1 0 1 0 0 1 0 1.
```

It is evident that one way of specifying the string obtained by displacing two elements is by taking the string which was previously generated from the original and shifting one element to generate another from it – the one desired – using a displacement of one more element.

Below we have, once more, the two results obtained and others which have not yet been presented:

```
0 1 1 0 1 0 0 1: given string
1 1 0 1 0 0 1 0: string generated by displacing one element (d = 1)

0 1 1 0 1 0 0 1: given string
1 0 1 0 0 1 0 1: string generated by displacing two elements (d = 2)

0 1 1 0 1 0 0 1: given string
0 1 0 0 1 0 1 1: string generated by displacing three elements (d = 3)

0 1 1 0 1 0 0 1: given string
1 0 0 1 0 1 1 0: string generated by displacing four elements (d = 4)

0 1 1 0 1 0 0 1: given string
0 0 1 0 1 1 0 1: string generated by displacing five elements (d = 5)

0 1 1 0 1 0 0 1: given string
```

0 1 0 1 1 0 1 0: string generated by displacing six elements ($d = 6$)

0 1 1 0 1 0 0 1: given string

1 0 1 1 0 1 0 0: string generated by displacing seven elements ($d = 7$)

A shift of eight elements (i.e., equal to l_{str}) would again generate the given string.

A second procedure used to obtain dyads from a given binary string will be described below.

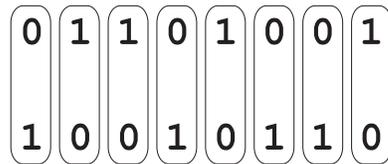
Again let the same string, $l_{str} = 8$, be that to which reference was made above: 0 1 1 0 1 0 0 1. Written below it there will be another binary string which is obtained from the first using the following algorithm:

The element i_j , where $j = 1, 2, \dots, l_{str}$, of the new string is the same as the symbol $i_{(l_{str}-j+1)}$ of the original string.

Therefore, the first symbol of the new string will be the same as the last symbol of the original string, the second symbol of the new string will be the same as the next to the last symbol of the original string, the third symbol of the new string will be the same as the third symbol from the end of the original binary string, etc. In other words, the new string has been obtained from the given string by using an operation of “spatial inversion” – or rather a “mirrored reflection” – of the given string. Hence, the original string and that obtained from it may be specified as follows:

0 1 1 0 1 0 0 1
1 0 0 1 0 1 1 0.

It will be considered that, in this situation, there is a null displacement ($d' = 0$) between the original string and that obtained from it using the procedure indicated above. Once again, these two strings (i.e., the given string and the “given inverted string”) can be considered a set of dyads:



Likewise, other sets of dyads can be obtained if the second string is subjected to successive displacements: $d' = 1, 2, 3, \dots, 7$. Thus, for example, if $d' = 1$, the following result is obtained:

0 1 1 0 1 0 0 1: given string

0 0 1 0 1 1 0 1: string generated from the “inverted” string by displacing one element ($d' = 1$).

When operating with the original string and the “inverted string”, note that:

- (a) there is no reason why only dyads like (0,0) y (1,1) can be obtained from $d' = 0$, and
- (b) for $d' = l_{str}$, the situation corresponding to $d' = 0$ can again be obtained.

With the first of the procedures described to obtain pairs of strings, seven pairs of strings were generated. With the second procedure, eight pairs of strings were obtained. For any of these fifteen pairs of binary strings, the following question may be asked: What are the probabilities that a dyad chosen at random will correspond to one of four existing types of dyads? In other words, what is the probability that for any of the fifteen pairs of strings, one dyad, chosen at random, will be (0,0)? And what is the probability that it will be (0,1)?; and so forth. These questions can be answered only if a certain amount of information is available regarding the original string. Suppose that this information leads us to accept the hypothesis that the probability that any element of that randomly chosen string will be 0 is equal to $\frac{1}{2}$, independently of the values of all the other elements in the string. In this case, the probability that the element will be 1 is also equal to $\frac{1}{2}$, without the value of each one of the other elements of the string having any relevance for that purpose. The following notation will be used:

$p(0)$: probability that the first element of any dyad will be 0

$p(1)$: probability that the first element of any dyad will be 1

$p(0,0)$, $p(0,1)$, $p(1,0)$ and $p(1,1)$: probabilities that any dyad will be (0,0), (0,1), (1,0) and (1,1) respectively

$p(0|0)$: conditional probability that the second digit of any dyad will be 0, admitting that the first is a 0

$p(0|1)$: conditional probability that the second digit of any dyad will be 0, admitting that the first is a 1

$p(1|1)$: conditional probability that the second digit of any dyad will be 1, admitting that the first is a 1

Hence, the previously formulated questions about the probabilities of the different dyads may be answered as follows:

$$\begin{aligned} p(0,0) &= p(0) \cdot p(0|0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\ p(0,1) &= p(0) \cdot p(1|0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\ p(1,0) &= p(1) \cdot p(0|1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\ p(1,1) &= p(1) \cdot p(1|1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

The essential features required to characterize the notions of regularity index and randomness index corresponding to any binary string such that $l_{str} = 4n$, where $n = 1, 2, 3, \dots$, have now been defined.

For each of the pairs of strings obtained with the first procedure specified above, $n_{0,0;d}$, $n_{0,1;d}$, $n_{1,0;d}$, $n_{1,1;d}$ will denote the numbers of dyads (0,0), (0,1), (1,0) and (1,1), respectively, with $d = 1, 2, \dots, l_{str}-1$. Likewise, for each one of the pairs of strings obtained

with the second procedure indicated above, $n_{0,0;d'}$, $n_{0,1;d'}$, $n_{1,0;d'}$, $n_{1,1;d'}$ will denote the numbers of dyads (0,0), (0,1), (1,0) and (1,1), respectively, with $d' = 0, 1, 2, \dots, l_{str} - 1$.

It is obvious that the following equalities should be satisfied:

$$n_{0,0;d} + n_{0,1;d} + n_{1,0;d} + n_{1,1;d} = l_{str}, (d = 1, 2, \dots, l_{str} - 1)$$

and

$$n_{0,0;d'} + n_{0,1;d'} + n_{1,0;d'} + n_{1,1;d'} = l_{str}, (d' = 0, 1, 2, \dots, l_{str} - 1)$$

The symbol D_d will be used to refer to the square root of the sums of the squares specified below, corresponding to the different possible displacements d ($d = 1, 2, \dots, l_{str} - 1$):

$$D_d = \left(\left(n_{0,0;d} - \frac{l_{str}}{4} \right)^2 + \left(n_{0,1;d} - \frac{l_{str}}{4} \right)^2 + \left(n_{1,0;d} - \frac{l_{str}}{4} \right)^2 + \left(n_{1,1;d} - \frac{l_{str}}{4} \right)^2 \right)^{\frac{1}{2}}.$$

Similarly, the symbol $D_{d'}^*$ will be used to refer to the square root of the sums of squares indicated below, corresponding to the different possible displacements d' ($d' = 0, 1, 2, \dots, l_{str} - 1$):

$$D_{d'}^* = \left(\left(n_{0,0;d'} - \frac{l_{str}}{4} \right)^2 + \left(n_{0,1;d'} - \frac{l_{str}}{4} \right)^2 + \left(n_{1,0;d'} - \frac{l_{str}}{4} \right)^2 + \left(n_{1,1;d'} - \frac{l_{str}}{4} \right)^2 \right)^{\frac{1}{2}}.$$

Each of the terms of the radicands of the right-hand members of the last two equalities is equal to the square of a certain expression. What exactly is this expression? It consists of the difference between a value obtained by counting and a value which is the most probable of those expected for binary strings for which the hypotheses specified are valid. In fact, for any of the displacements d or d' considered, $\frac{l_{str}}{4}$ is the most probable of the expected values for the numbers of dyads of any type. Thus, for example, $p(0,0) = \frac{1}{4}$ and given that the number of dyads for any displacement d or d' is equal to l_{str} , it is found that the most probable value for the number of dyads of the type (0,0) is equal to $p(0,0) \cdot l_{str} = \frac{1}{4} \cdot l_{str}$. And as specified above

- (a) $n_{0,0;d}$, $n_{0,1;d}$, $n_{1,0;d}$, and $n_{1,1;d}$ are the numbers of dyads effectively generated, and counted, for each of the $l_{str} - 1$ pairs of the strings obtained with the displacements d ; and
- (b) $n_{0,0;d'}$, $n_{0,1;d'}$, $n_{1,0;d'}$, and $n_{1,1;d'}$ are the numbers of dyads effectively generated, and counted, for each of the l_{str} pairs of the strings obtained with the displacements d' .

The index of regularity i_{reg} corresponding to the binary string is defined as:

$$i_{reg} = \frac{\max \{ D_1, D_2, \dots, D_{l_{str}-1}, D_0^*, D_1^*, \dots, D_{l_{str}-1}^* \}}{\left((l_{str} - \frac{l_{str}}{4})^2 + 3 \left(\frac{l_{str}}{4} \right)^2 \right)^{\frac{1}{2}}}.$$

The right-hand member of the preceding equality is a fraction. Its numerator is the maximum value of the elements belonging to the set $\{D_1, D_2, \dots, D_{l_{str}-1}, D_0^*, D_1^*, \dots, D_{l_{str}-1}^*\}$. Its denominator is any D_d , where $d = 1, 2, \dots, l_{str} - 1$, or any $D_{d'}^*$, where $d' = 0, 1, 2, \dots, l_{str} - 1$, corresponding to a binary string of length l_{str} and the maximum possible regularity. (For this string all instances of D_d and $D_{d'}^*$ are the same.) There are two of these strings:

$$\underbrace{0000\dots0}_{l_{str} \text{ zeros}} \quad \text{and} \quad \underbrace{1111\dots1}_{l_{str} \text{ ones}}.$$

If the definition given for i_{reg} is taken into account, it is obvious that its greatest possible value is 1. This value $- 1 -$ is obtained by calculating i_{reg} for any of the two strings described above.

The corresponding index of randomness $- i_{rnd}$ - is defined as follows:

$$i_{rnd} = 1 - i_{reg}.$$

Reference will be made below to how to determine the regularity index and the randomness index for any binary string such that its l_{str} is not a multiple of 4.

Suppose, for example, that one has a binary string for which $l_{str} = 4n + k$, where $n = 1, 2, \dots$ and $k = 1, 2$ or 3 . One would proceed as follows: Consideration must first be given to the binary string composed of the elements of the original string which go from the first element to that of $4n$, inclusive. The regularity index is determined for that string and is denoted as $i_{reg,1}$. Then attention is given to the binary string composed of the elements of the original string which go from the second element to that of $4n + 1$ and the regularity index is found for that string: $i_{reg,2}$. If $k = 1$, then the regularity index of the original string is computed as follows:

$$i_{reg} = \frac{i_{reg,1} + i_{reg,2}}{2}.$$

If $k = 2$, then consideration should also be given to the binary string composed of all the elements of that string from the third element to that of $4n + 2$, inclusive. The regularity index $i_{reg,3}$ is then found for this string. The index of regularity of the original string is defined, therefore, as:

$$i_{reg} = \frac{i_{reg,1} + i_{reg,2} + i_{reg,3}}{3}.$$

If $k = 3$, attention must also be given to the binary string composed of all the elements of the original string that go from the fourth element to that of $4n + 3$, inclusive. The regularity index of the original string is thus defined as:

$$i_{reg} = \frac{i_{reg,1} + i_{reg,2} + i_{reg,3} + i_{reg,4}}{4}.$$

The corresponding index of randomness of the original string i_{rnd} is defined in all cases as follows:

$$i_{rnd} = 1 - i_{reg}.$$

Several binary strings mentioned in sections 1 and 2 are presented below with their corresponding indices of randomness specified on the right.

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	$i_{rnd} = 0$
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1	$i_{rnd} = 0.4226$
0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	$i_{rnd} = 0.4226$
0 1 1 0 1 0 0 1	$i_{rnd} = 0.3578.$

3 Characterization of the notions of regularity index and randomness index for m -ary strings

As described above, binary strings are a particular case of m -ary strings, where $m = 2, 3, 4, \dots$: that in which $m = 2$.

If $m = 3$, and if the same procedures are used for 3-ary strings as those described in section 2 to obtain two $l_{str} - 1$ pairs of strings, then 3^2 , or 9, different types of dyads will be generated: (0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1) and (2,2). In general, for any value m , where $m = 2, 3, 4, \dots$, the number of different types of dyads to be generated is m^2 .

The generalization for m -ary strings arising from the notions and definitions given in section 2 is quite natural and direct. For reasons of expository clarity, consideration will be given first of all to m -ary strings such that

$$l_{str} = n \cdot m^2 \quad n = 1, 2, 3 \dots$$

The symbols D_d y $D_{d'}^*$ will be used to refer to the square root of the sums of the squares specified as follows:

$$D_d = \left(\sum_{i=1}^m \sum_{j=1}^m \left(n_{i,j;d} - \frac{l_{str}}{m^2} \right)^2 \right)^{\frac{1}{2}} \quad d = 1, 2, \dots, l_{str} - 1.$$

$$D_{d'}^* = \left(\sum_{i=1}^m \sum_{j=1}^m \left(n_{i,j;d'} - \frac{l_{str}}{m^2} \right)^2 \right)^{\frac{1}{2}} \quad d' = 0, 1, 2, \dots, l_{str} - 1.$$

The index of regularity i_{reg} of an m -ary string such that its $l_{str} = n \cdot m^2$, where $n = 1, 2, 3, \dots$, is defined as follows:

$$i_{reg} = \frac{\max \{D_1, D_2, \dots, D_{l_{str}-1}, D_0^*, D_1^*, \dots, D_{l_{str}-1}^*\}}{\left((l_{str} - \frac{l_{str}}{m^2})^2 + (m^2 - 1) (\frac{l_{str}}{m^2})^2 \right)^{\frac{1}{2}}}.$$

The corresponding index of randomness i_{rnd} is defined as follows:

$$i_{rnd} = 1 - i_{reg}.$$

It is very important to understand the following idea clearly. It may be admitted that the i_{reg} computed for an m -ary string provides a quantitative estimation of the degree of “separation” of that string with respect to another which is supposed to show the greatest lack possible of regularities, or patterns, in its structure. However, this fact does not at all imply that any hypothesis has been deemed valid with regard to the procedure used to obtain the string whose i_{reg} was computed. In effect, the i_{reg} of an m -ary string provides information only about how regular the structure of the string is and *not* about its genesis. One must recall why the hypothetical string H from which the above-mentioned shifting is estimated lacks patterns. The reason is that if any particular one of the l_{str} elements of H is chosen at random, the probability $\frac{1}{m}$ that the specific element will be one of the m possible elements composing the string is the same as the probability that it will be any other of the remaining $(m - 1)$ elements.

Indications are provided below about how to compute the i_{reg} of an m -ary string such that $l_{str} = n \cdot m^2 + k$, where $n = 1, 2, 3, \dots$ and $k = 1, 2, \dots, m^2 - 1$.

In the first place, it is necessary to compute the regularity index of the substring of that string extending from the first element to element $n \cdot m^2$, inclusive; this regularity index is denoted by $i_{reg,1}$. In the second place, the regularity index must be computed for the substring extending from the second to element $(n \cdot m^2) + 1$, inclusive; this regularity index is denoted by $i_{reg,2}$. And so forth successively, until the regularity index of the substring extending from element $(k + 1)$ to element $(n \cdot m^2) + k$, inclusive; this regularity index is denoted by $i_{reg,k+1}$.

The index of regularity of the given m -ary string is defined as:

$$i_{reg} = \frac{i_{reg,1} + i_{reg,2} + \dots + i_{reg,k+1}}{k + 1}.$$

The corresponding index of randomness i_{rnd} is defined as:

$$i_{rnd} = 1 - i_{reg}.$$

4 Numerical results of particular interest

Below we will present some of the results regarding binary strings with lengths of 8, 12, 16 and 24; and 3-ary strings with $l_{str} = 9$.

Tables 1 and 2 show, for all of the binary strings with the lengths specified, the numbers of those strings which have randomness indices between the ranges of values indicated in the first column of those tables.

Tables 3 and 4 show, for all of the 3-ary strings with $l_{str} = 9$, the numbers of those strings which have randomness indices between the ranges of values indicated in the first column of those tables.

For all of the binary strings with $l_{str} = 24$, the numbers n of those strings which have randomness indices between the ranges of values indicated on the x-axis have been specified graphically in histogram 1.

Given the relatively small numbers of binary strings with $l_{str} = 24$ which have randomness indices between the ranges of values $0 \leq i_{rnd} \leq 0.2$ or $0.8 \leq i_{rnd} \leq 1.0$, the

	$l_{str} = 8$	$l_{str} = 12$	$l_{str} = 16$	$l_{str} = 24$
$0 \leq i_{rnd} \leq 0.1$	2	2	34	50
$0.1 < i_{rnd} \leq 0.2$	16	24	240	1048
$0.2 < i_{rnd} \leq 0.3$	56	236	1048	28124
$0.3 < i_{rnd} \leq 0.4$	48	318	3376	181294
$0.4 < i_{rnd} \leq 0.5$	70	1116	13350	1477420
$0.5 < i_{rnd} \leq 0.6$	0	1920	31616	5864064
$0.6 < i_{rnd} \leq 0.7$	64	480	10752	6020832
$0.7 < i_{rnd} \leq 0.8$	0	0	5120	3148704
$0.8 < i_{rnd} \leq 0.9$	0	0	0	55680
$0.9 < i_{rnd} \leq 1$	0	0	0	0

Table 1: $m = 2$.

	$l_{str} = 8$	$l_{str} = 12$	$l_{str} = 16$	$l_{str} = 24$
$0 \leq i_{rnd} \leq 0.2$	18	26	274	1098
$0.2 < i_{rnd} \leq 0.4$	104	554	4424	209418
$0.4 < i_{rnd} \leq 0.6$	70	3036	44966	7341484
$0.6 < i_{rnd} \leq 0.8$	64	480	15872	9169536
$0.8 < i_{rnd} \leq 1$	0	0	0	55680

Table 2: $m = 2$.

existence of the strings corresponding to those ranges of values is not perceptible in the preceding histogram. Thus, to be able to observe their presence, the same information corresponding to that histogram is shown in histogram 2, in which the y-axis corresponds to $\log n$ (and not to n).

For all of the 3-ary strings with $l_{str} = 9$, the numbers n of those strings which have randomness indices between the ranges of values indicated on the x-axis have been specified graphically in histogram 3.

Given the relatively small number of 3-ary strings with $l_{str} = 9$ which have randomness indices between the range of values $0 \leq i_{rnd} \leq 0.2$, the existence of the strings corresponding to that range of values is not perceptible in the preceding histogram. Therefore, to be able to observe their presence, the same information corresponding to that histogram is shown in histogram 4, in which the y-axis corresponds to $\log n$ (and not to n).

	$l_{str} = 9$
$0 \leq i_{rnd} \leq 0.1$	3
$0.1 < i_{rnd} \leq 0.2$	54
$0.2 < i_{rnd} \leq 0.3$	612
$0.3 < i_{rnd} \leq 0.4$	864
$0.4 < i_{rnd} \leq 0.5$	7674
$0.5 < i_{rnd} \leq 0.6$	8208
$0.6 < i_{rnd} \leq 0.7$	2268
$0.7 < i_{rnd} \leq 0.8$	0
$0.8 < i_{rnd} \leq 0.9$	0
$0.9 < i_{rnd} \leq 1$	0

Table 3: $m = 3$.

	$l_{str} = 9$
$0 \leq i_{rnd} \leq 0.2$	57
$0.2 < i_{rnd} \leq 0.4$	1476
$0.4 < i_{rnd} \leq 0.6$	15882
$0.6 < i_{rnd} \leq 0.8$	2268
$0.8 < i_{rnd} \leq 1$	0

Table 4: $m = 3$.

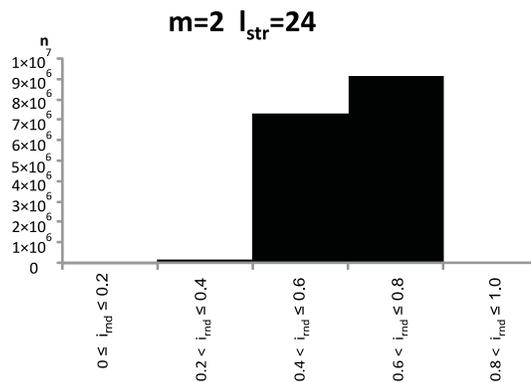


Figure 1: Histogram 1.

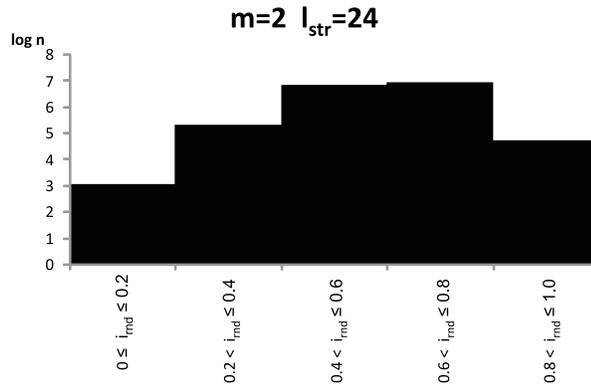


Figure 2: Histogram 2.

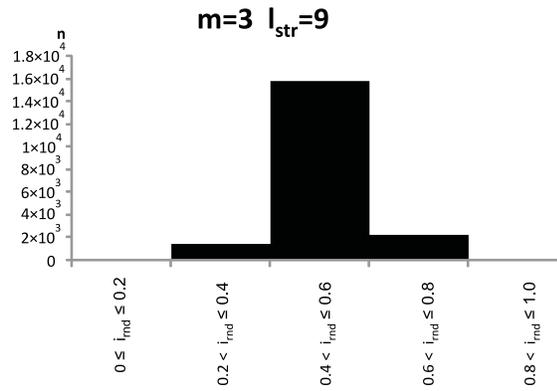


Figure 3: Histogram 3.

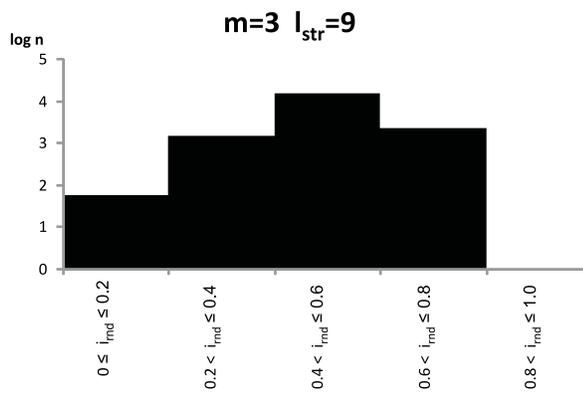


Figure 4: Histogram 4.

Note that for the 3-ary strings with $l_{str} = 9$, there are no strings whose randomness indices are within the range of values $0.8 < i_{rnd} \leq 1.0$.

Although it is true that we have just begun to study the distribution of the numbers of m -ary strings of different lengths according to the ranges of values in which the values of the randomness indices of the different strings are found, one of our most outstanding results noted so far will be described below. If a distinction is made between a “lower portion” of the ranges of values of the randomness indices such that they are less than or equal to 0.5 (i.e., such that $i_{rnd} \leq 0.5$) and an “upper portion” of ranges of values of these indices such that they are greater than 0.5 (i.e., such that $i_{rnd} > 0.5$), then there is a larger number of strings belonging to the “upper portion” than to the “lower portion”. Thus, for example, for binary strings of $l_{str} = 24$, there are 1,687,936 strings in the “lower portion” of the randomness index and 15,089,280 strings in the “upper portion” of the index.

5 Relations between approaches addressing the notions of randomness and regularity introduced by other authors and the approach presented here

Our objective is not, in the least, to make an exhaustive, comparative study between the approach described here regarding the notions of index of regularity and the index of randomness for finite-length m -ary strings and other approaches concerning the notion of randomness which can be found in the technical literature on the topic. However, the comments in this section may provide a useful initial orientation to those interested in this issue.

In S. B. Volchan’s interesting, precise text “What is a Random Sequence?”, he has stated the following [5, p. 48]:

“But now the question becomes: What does it mean to say that an individual infinite sequence of 0s and 1s is random? Historically, three main notions were proposed:

- stochasticness or frequency stability, due to von Mises, Wald, and Church;
- incompressibility, due to Solomonoff, Kolmogorov and Chaitin;
- typicality, due to Martin-Löf”.

Regarding the first of these approaches, that of “randomness as stochasticness”, the same author adds [5, p. 54]:

“Many criticisms were directed at von Mises’s proposals. Not only were his arguments based on gambling-house notions, considered inexact or at best semimathematical, but also the central notion of ‘admissible’ selections was not clarified at all”.

We believe that the serious objections made with respect to the approach developed by von Mises are correct. The approach presented in this paper is not based on that of von Mises. No use has been made, for example, of the notions of “collectives” or of “frequency stability”.

Regarding the other two approaches considered by S. B. Volchan, that author rightly states that they are equivalent [5, p. 61]:

“The notion of Martin-Löf random sequences is mathematically consistent, and unexpectedly coincides with an apparently very different notion of randomness, namely, that of incompressibility”. This equivalence is recognized by G. Chaitin, the main architect of algorithmic information theory (AIT), based on the notion of “randomness as incompressibility”, and on its applications in the field of metamathematics. Those interested in that equivalence may read, for example, the second section (pp. 129-133), “Proof that Martin-Löf randomness is equivalent to Chaitin randomness” in Part III of Chaitin’s book cited above [1].

Although it is true that Chaitin’s approach is essentially oriented toward infinite-length strings, it can be applied to finite-length m -ary strings. Informally, an idea of his approach will be given below, using binary strings for reasons of simplicity and expository clarity. Let a binary string be comprised of 10^6 digits such that each of them is equal to 1. It is obvious that this string could be generated by a very simple computer program whose objective is to print the number one 10^6 times. In computing terms, the string made up of 10^6 ones has a length of 10^6 bits. In contrast, this program designed to generate it is – also in computational terms – much shorter than 10^6 bits. The same occurs with other binary strings such as the string with a length of 10^6 composed of $5 \cdot 10^5$ repetitions of the following sequence of two digits: 0 1. Likewise in this case, a program designed to print $5 \cdot 10^5$ times the sequence 0 1 is much shorter, in bits, than 10^6 .

Any string for which there exists a program to generate it and whose length, in bits, is shorter than that string, also in bits, is considered nonrandom, according to the AIT approach. In contrast, if the shortest program that can generate a string has a length which is approximately the same as that of the string (since it is a program that should include the string considered), then it is found to be random. In other words, in the case of a random string, the shortest program that can generate it is, essentially, the same string. It can thus be noted that, according to this approach, a string may be considered random if an effect of “compression” cannot be achieved – that is, if it is not possible to generate that string using a program whose length is less than the length of the string. This approach, which is applicable not only to m -ary strings but also to other mathematical objects, is dichotomic; accordingly, an m -ary string is either random or not random. Therefore, using the AIT approach, it is not possible to make a quantitative estimate of something which may be considered the degree of randomness of an m -ary string. The approach described in the present article *does* make this estimation possible.

S. Pincus and B. H. Singer have developed an approach which varies, partially, from the dichotomic conception “random strings – nonrandom strings” [3]. According to these authors, “...there appears to be a critical need to develop a (computable) formulation of ‘closer to random’, to grade the large class of decidedly nonrandom sequences”. They accept the existence of a “large class of decidedly nonrandom sequences” and consider it appropriate to distinguish in that class between various levels of nearness to truly random sequences. This approach is different from ours, where for binary strings of a given length, there are only two of them (and not a “large class”) which can be considered definitely regular (or nonrandom): that comprised only of ones and that comprised only of zeroes. For these two strings, the index of regularity is equal to 1 and the index of randomness is equal to 0. The remaining strings are not considered to be either definitely random or

definitely regular. For each of these strings, it is feasible to compute an index of regularity (i_{reg}) and an index of randomness (i_{rnd}), such that $i_{reg} + i_{rnd} = 1$. It is obvious that, following our approach, the higher the regularity index of a binary string, the lower its randomness index, and vice versa. With this approach there are as many classes of degrees of randomness as there are different indices of randomness computed. The higher the index of randomness of a string, the more it can be considered random. Of course, for m -ary strings (with $m = 3, 4, 5, \dots$) of a given length, there are m strings for each of which $i_{reg} = 1$ and $i_{rnd} = 0$.

6 Discussion and prospects

Suppose that a scientist is consulted whether it is reasonable to accept that the binary string of $l_{str} = 20$ comprised of ten repetitions of the two elements 1 and 0 (i.e., the string 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0) has been generated by a device whose main operational characteristic can be described as follows: This device generates one digit, for example, every millisecond, and there is the same probability $-\frac{1}{2}$ – that each one of the digits generated will be a 0 or a 1. In other words, for each digit generated by that device, the following will be satisfied: $p(0) = p(1) = \frac{1}{2}$. The person consulted would undoubtedly not feel very inclined to accept the above-mentioned genesis of the particular string. On the other hand, if that person were asked the same question with regard to the binary string 1 0 0 0 1 1 1 1 0 1 0 0 1 0 0 1 0 1 1 0, it probably would be considered feasible that the above-mentioned way of generating the string had been used. Nonetheless, it must be noted that if that device was used, there is exactly the same probability $-\left(\frac{1}{2}\right)^{20}$ – that each one of them will be generated. Actually, the probability that the device will generate any of the 2^{20} possible binary strings is the same: $\left(\frac{1}{2}\right)^{20}$. The difference in the answers of the person consulted about these two situations would certainly be related to the fact that human beings have the ability to determine intuitively which strings have certain types of regularities. What types of regularities? The types that can be considered, metaphorically, to just “pop out”. Let there be a binary string with a length of 20. One can immediately distinguish, for example, a. whether that string is made up of 20 ones, or 20 zeroes; b. if it is composed of 10 reiterations of the sequence of the two elements 1 and 0 (that is, of the sequence 1 0); c. if it is comprised of 5 reiterations of the sequence 1 1 0 0; and so forth. All the strings of this type – and among them, the first binary string specified in this section – appear to be very regular; i.e., not very random. It is reasonable to suppose, then, that the first string could belong to the “lower portion” of the randomness index (according to the terms used in section 4). In contrast, the second binary string to which reference was made does not appear to be too regular; that is, it looks quite random. It is thus reasonable to suppose that this last string could belong to the “upper portion” of the randomness index. And as mentioned in section 4, the number of strings corresponding to the “upper portion” of the randomness index is greater than the number of strings pertaining to the “lower portion” of the randomness index. Therefore, whereas the probability of the device generating any particular binary string is the same, there is a greater probability of generating a string belonging to the

“upper portion” of the randomness index than there is of generating one corresponding to the “lower portion”. Thus, we can conclude that the answers given to those consultations are reasonable.

In this paper the notions of index of regularity and index of randomness have been defined for finite-length m -ary strings. These definitions make it possible to evaluate more precisely than could be done with a simple intuitive reaction to the degrees of regularity and randomness of those mathematical objects which are essential for fields such as computer science and the physical sciences.

According to the approach introduced here, the notion of “randomness” is equivalent to that of “lack of regularity”. In fact, the “index of randomness” could have been called “index of irregularity”. The notions of “regularity” and “randomness” (or “irregularity”) are, according to this approach, complementary. The more regular an m -ary string, the less random it is considered; likewise, the more random an m -ary string, the less regular it is considered. It should be stressed that the approach presented here for finite-length m -ary strings is not the same as other approaches which have been described in technical literature on the topic.

Those who need to evaluate the degree of randomness of finite-length m -ary strings to solve problems in different scientific and technological disciplines will decide which approach is the most suitable and the most pertinent for their specific requirements, as well as the most efficient and the easiest to implement.

In forthcoming work diverse types of fluctuations will be analyzed, thus making it possible to justify even more soundly and in greater depth than feasible here, the relevance and usefulness of the definitions introduced in this paper.

References

- [1] Chaitin G.J. (1998) *The Limits of Mathematics*. Springer-Verlag, New York.
- [2] Chaitin, G.J. (2001) *Exploring Randomness*. Springer-Verlag, New York.
- [3] Pincus, S.; Singer, B. H. (1996) “Randomness and degrees of irregularity”, *Proceedings of the National Academy of Sciences of the United States of America* **93**(5): 2083–2088.
- [4] Skliar, O.; Medina, V.; Oviedo G. (1995) “Un índice de aleatoriedad para cadenas binarias”, in: J. Trejos (Ed.) *Memorias del IX Simposio Métodos Matemáticos Aplicados en las Ciencias*, Turrialba: 161–172.
- [5] Volchan, S.B. (2002) “What is a random sequence?”, *The American Mathematical Monthly* **109**(1): 46–63.