Article

Entropy-Variance Curves of Binary Sequences Generated by Random Substitutions of Constant Length

Juan Carlos Nuño ^{*,†} and Francisco J. Muñoz [†]

Department of Applied Mathematics, Universidad Politécnica de Madrid, 28040 Madrid, Spain; f.j.munoz.ortega@gmail.com

* Correspondence: juancarlos.nuno@upm.es

† These authors contributed equally to this work.

Abstract: We study some properties of binary sequences generated by random substitutions of constant length. Specifically, assuming the alphabet $\{0, 1\}$, we consider the following asymmetric substitution rule of length k : $0 \rightarrow \langle 0, 0, \dots, 0 \rangle$ and $1 \rightarrow \langle Y_1, Y_2, \dots, Y_k \rangle$, where Y_i is a Bernoulli random variable with parameter $p \in [0, 1]$. We obtain by recurrence the discrete probability distribution of the stochastic variable that counts the number of ones in the sequence formed after a number i of substitutions (iterations). We derive its first two statistical moments, mean and variance, and the entropy of the generated sequences as a function of the substitution length k for any successive iteration i , and characterize the values of p where the maxima of these measures occur. Finally, we obtain the parametric curves entropy-variance for each iteration and substitution length. We find two regimes of dependence between these two variables that, to our knowledge, have not been previously described. Besides, it allows to compare sequences with the same entropy but different variance and vice versa.

Keywords: binary sequences; random substitutions; entropy-variance

Citation: Nuño, J.C.; Muñoz, F.J. Entropy-Variance Curves of Binary Sequences Generated by Random Substitutions of Constant Length. *Entropy* **2022**, *24*, 290. <https://doi.org/10.3390/e24020290>

Academic Editor: Carlo Cattani

Received: 17 January 2022

Accepted: 15 February 2022

Published: 18 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Binary sequences (strings or chains) appear naturally in physical systems for describing growing processes generated from an initial state. Examples are ubiquitous, from the spin up-down systems in physics to codification in information theory and its reflection in biomolecules as DNA and RNA (where two sets of nucleotides exist: purines and pyrimidines). They are also common objects in mathematics, representing, for example, sets of words formed from letters of an alphabet or symbolic representations of dynamical systems, where mapping can be defined to transform real trajectories into two state series [1]. Binary sequence can also codified fractal sets obtained by an iterative process, as occurs in the classical Cantor set [2,3].

In order to study the properties of systems with fractal geometry, Mandelbrot defined some random sets that are generated recursively from a given initial set as a percolation process [2,4]. These sets are formed by successive application of a defined set of rules that, either divide the initial set and discard some subsets in each partition, or enlarge the initial set by substituting the existing subsets by multiple replicas of themselves. These sets can be embedded in 1D, 2D and 3D spaces, conditioning some of their characteristics, in particular, percolation.

In this paper, we study the properties of the sequences formed by performing random substitutions of constant length (see, for instance, [5]). Specifically, we consider a binary alphabet, that we denote as $\{0, 1\}$, and substitutions of length $k = 2, 3, \dots$. As in the classical percolation process, we assume that, once a void site is formed, i.e., a 0 occurs at this site, subsequent substitutions of it yield k zeros, always. On the other hand, filled sites, i.e., 1 at this site, are substituted by random words of k letters with a (uniform) probability

of inserting a 1 defined by a parameter p . In this substitution, the probability of inserting a 0 is $q = 1 - p$.

The process defined in this way gives rise to sequences of increasing size. The number of ones (filled sites) in the sequence at iteration i is a stochastic variable that depends on the probability p and the length k . We study the probabilistic properties of these sequences. Specifically, we derive the probability distribution of the number of ones at iteration i . We also obtain the expected values of X and X^2 , the corresponding variance VAR , and obtain some properties of them. As an alternative measure of uncertainty, we calculate the mean entropy of these sequences and compare it with their variance [6,7]. Finally, we find the parametric curves that relate both measures for each iteration and substitution length. These curves present two regimes: for low values of p the dependence entropy-variance is convex (concave down), whereas for large values of p this dependence is practically lost. Therefore, for each iteration and substitution length, we can generate sequences with the same entropy but with different variance and vice versa.

2. Sequence Generation

The generation of the sequences studied in this work is a one-dimensional example of the classical well-known Mandelbrot's percolation process [4]. Contrary to the classical formulation, where an initial segment of length L is subdivided into k subsegments of length L/k and, with probability p , some of them are chosen to be further divided, the model we study here considers an initial digit, that we denote by 1, that is substituted by a word of length k formed by digits $\{0, 1\}$. As in the classical model, we assume that digits denoted by 0 are substituted by a word of k zeros (see Figure 1) and sites occupied by ones are substituted by a word of k letters with a independent probability p . This procedure is applied successively to generate a binary sequence of finite length, after i iterations or, in the limit, infinite length. When $k = 3$, this procedure can be viewed as a probabilistic version of the Cantor set for $0 < p < 1$ [2].

These procedures are called substitutions, as they are obtained by replacing a digit by a word of digits [1,8,9]. In this context, we study the following map that applies at each random substitution of length k :

$$\begin{aligned} 0 &\rightarrow \langle 00 \dots 0 \rangle \\ 1 &\rightarrow \langle Y_1 Y_2 \dots Y_k \rangle \end{aligned} \quad (1)$$

where Y_j is a Bernoulli random variable with parameter $p \in [0, 1]$, i.e., p means the (uniform) probability of inserting 1 in each position and independently for all j . We note that an alternative description of this substitution process can be defined by using the Kronecker product [10–13].

Figure 1. An example of schematic representation of the sequence generation (from top) for the cases (a) $k = 2$ and (b) $k = 3$, after six and five substitutions, respectively. For both cases, $p = 0.9$. Black and white represent 1 and 0, respectively.

The deterministic version of this procedure brings about well known sequences. For the classical Cantor set, this map reduces to:

$$\begin{aligned} 0 &\rightarrow \langle 000 \rangle \\ 1 &\rightarrow \langle 101 \rangle \end{aligned} \quad (2)$$

which, starting from a unique 1, yield the sequences:

$$\begin{aligned}s_1 &= (1, 0, 1) \\s_2 &= (1, 0, 1, 0, 0, 0, 1, 0, 1) \\s_3 &= (1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1)\end{aligned}\quad (3)$$

whose limit of the iterative process is the infinite sequence that corresponds to the classical Cantor set. For this particular case, the length of the i -th vector is $L_i = 3^i$. In general, for substitutions of constant length k , the i th-iterated vector has length $L_i = k^i$.

There are other well-known sequences generated iteratively by substitutions of different length. For instance, the Fibonacci sequence is generated applying the following map [14,15]:

$$0 \rightarrow \langle 1 \rangle \quad 1 \rightarrow \langle 10 \rangle \quad (4)$$

Starting from 0, after six iterations we obtain the binary vector: (1, 0, 1, 1, 0, 1, 0, 1). Here, the number of 1 of the i th-sequence represents the fertile population of rabbits. Thus, the number of mature pairs of rabbits after six generations is 5. Notice that, in principle, the position in the sequence has not a spatial meaning. It is also worthy to remark that this sequence can be generated by different maps [14].

Another self-similar sequence generated by a recursive process is the so-called Morse-Thue (MT), which is obtained by the following map:

$$0 \rightarrow \langle 01 \rangle \quad 1 \rightarrow \langle 10 \rangle \quad (5)$$

starting from 0 (although it could be also initiated by 1) [14,15]. Also for this example, there is an alternative way of generating this sequence, simply by take the sequence of the preceding step and add the complement. Besides, this sequence is aperiodic, although completely deterministic. On the contrary, it can be proven that there are short and long range correlations [14].

3. Main Statistical Properties

In this section, we study the statistical properties of the sequences generated by applying the rule (1) recurrently to the starting sequence (1). A classical issue to be considered is the probability distribution of each of the letters of the alphabet. In particular, in the case of binary sequences, this problem is reduced to know the probability distribution of, for instance, the number of ones of any population of sequences generated from a set of random rules, e.g. random substitutions.

As a starting point, the classical Bernoulli process could be illustrative. This process can be also viewed as a random substitution of constant length, k , where both letters, e.g., 0 and 1, are identically substituted according with the rule:

$$0 \rightarrow \langle Z_1, Z_2, \dots, Z_k \rangle; \quad 1 \rightarrow \langle Y_1, Y_2, \dots, Y_k \rangle \quad (6)$$

being Z_j and Y_j Bernoulli processes with parameter $p \in [0, 1]$. In each iteration, i , the length of the sequence is $n = k^i$, and the probability distribution of the number of ones, X , is given by the binomial distribution:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (7)$$

By applying this distribution function, the expected values of the number of ones in a sequence of length n , X_n , and its square, respectively $E^B(X_n)$ and $E^B(X_n^2)$, can be calculated:

$$E^B(X_n) = np \quad (8)$$

$$E^B(X_n^2) = np(1 + (n-1)p) \quad (9)$$

for $n = 1, 2, 3, \dots$ Using these expressions, the variance of X reads:

$$\text{VAR}^B(X_n) = E^B(X_n^2) - (E^B(X_n))^2 = n p (1 - p) \quad (10)$$

Note that for the binomial distribution, the ratio variance-mean is equal to $1 - p$ for any n .

Similarly, we can compute these moments for the sequences generated from random substitutions defined in the previous section. In this case, a recurrent formula provides this distribution for any iteration i and for any value of the substitution length k .

Let us present first the results for a substitution of length $k = 2$. As stated before, we start the generation of the binary sequences from a sequence with a unique one. Thus, the probability of having this initial sequence is assumed to be 1:

$$P(X_0 = 1) = 1 \quad (11)$$

In the first step, a sequence of length two is formed whose distribution of ones depends on the probability p (see Figure 2). To obtain the the distribution probability of ones in this first step, $P(\mathbf{X}_1)$, we multiply from left $P(X_0)$ by the substitution matrix:

$$M_{3 \times 1} = \begin{pmatrix} \binom{2}{0} (1-p)^2 \\ \binom{2}{1} p (1-p) \\ \binom{2}{2} p^2 \end{pmatrix} \quad (12)$$

giving rise to the vector:

$$\mathbf{P}(\mathbf{X}_1) = \begin{pmatrix} P_1(X=0) \\ P_1(X=1) \\ P_1(X=2) \end{pmatrix} = M_{3 \times 1} P(X_0) = M_{3 \times 1} \quad (13)$$

where $\mathbf{X}_1 = (0, 1, 2)$.

In the next substitution, a four digits sequence is formed and the probability distribution of ones can be computed from the previous one as follows:

$$\mathbf{P}(\mathbf{X}_2) = \begin{pmatrix} P(X=0) \\ P(X=1) \\ P(X=2) \\ P(X=3) \\ P(X=4) \end{pmatrix} = \begin{pmatrix} 1 & \binom{2}{0}(1-p)^2 & \binom{4}{0}(1-p)^4 \\ 0 & \binom{2}{1}p(1-p) & \binom{4}{1}p(1-p)^3 \\ 0 & \binom{2}{2}p^2 & \binom{4}{2}p^2(1-p)^2 \\ 0 & 0 & \binom{4}{3}p^3(1-p) \\ 0 & 0 & \binom{4}{4}p^4 \end{pmatrix} \begin{pmatrix} P(X=0) \\ P(X=1) \\ P(X=2) \end{pmatrix} = M_{5 \times 3} \mathbf{P}(\mathbf{X}_1) \quad (14)$$

where $\mathbf{X}_2 = (0, 1, 2, 3, 4)$. The third iteration confirms the formula (see Figure 2):

$$\mathbf{P}(\mathbf{X}_3) = M_{9 \times 5} \mathbf{P}(\mathbf{X}_2) \quad (15)$$

with

$$\mathbf{P}(\mathbf{X}_3) = \begin{pmatrix} P(X=0) \\ P(X=1) \\ P(X=2) \\ P(X=3) \\ P(X=4) \\ P(X=5) \\ P(X=6) \\ P(X=7) \\ P(X=8) \end{pmatrix} \quad (16)$$

where $\mathbf{X}_3 = (0, 1, 2, 3, 4, 5, 6, 7, 8)$, and

$$M_{9 \times 5} = \begin{pmatrix} 1 & \binom{2}{0}(1-p)^2 & \binom{4}{0}(1-p)^4 & \binom{6}{0}(1-p)^6 & \binom{8}{0}(1-p)^8 \\ 0 & \binom{2}{1}p(1-p) & \binom{4}{1}p(1-p)^3 & \binom{6}{1}p(1-p)^5 & \binom{8}{1}p(1-p)^7 \\ 0 & \binom{2}{2}p^2 & \binom{4}{2}p^2(1-p)^2 & \binom{6}{2}p^2(1-p)^4 & \binom{8}{2}p^2(1-p)^6 \\ 0 & 0 & \binom{4}{3}p^3(1-p) & \binom{6}{3}p^3(1-p)^3 & \binom{8}{3}p^3(1-p)^5 \\ 0 & 0 & \binom{4}{4}p^4 & \binom{6}{4}p^4(1-p)^2 & \binom{8}{4}p^4(1-p)^4 \\ 0 & 0 & 0 & \binom{6}{5}p^5(1-p) & \binom{8}{5}p^5(1-p)^3 \\ 0 & 0 & 0 & \binom{6}{6}p^6 & \binom{8}{6}p^6(1-p)^2 \\ 0 & 0 & 0 & 0 & \binom{8}{7}p^7(1-p) \\ 0 & 0 & 0 & 0 & \binom{8}{8}p^8 \end{pmatrix} \quad (17)$$

The general expression for the recurrent formula for the probability distribution of the number of ones in the i -th iteration is:

$$\mathbf{P}(\mathbf{X}_i) = M_{(2^i+1) \times (2^{i-1}+1)} \mathbf{P}(\mathbf{X}_{i-1}) \quad (18)$$

being the vector $\mathbf{X}_i = (0, 1, 2, \dots, 2^i)$ and the general substitution matrix, $M_{(2^i+1) \times (2^{i-1}+1)}$, whose coefficients are given by:

$$m_{ab}^i = \begin{cases} \binom{2^{(b-1)}}{a-1} p^{a-1} (1-p)^{2^{(b-1)}-(a-1)} & \text{if } a \leq 2b-1 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

for $i = 2, 3, \dots$ and $1 \leq a \leq 2^i + 1$ and $1 \leq b \leq 2^{i-1} + 1$. For $i = 1$, the coefficients of matrix $M_{3 \times 1}$ (12) are given by:

$$m_{a1}^1 = \binom{2}{a-1} p^{a-1} (1-p)^{2-(a-1)} \quad \text{for } 1 \leq a \leq 3 \quad (20)$$

By recurrence, Formula (18) can be converted into a matrix product chain:

$$\mathbf{P}(\mathbf{X}_i) = M_{(2^i+1) \times (2^{i-1}+1)} M_{(2^{i-1}+1) \times (2^{i-2}+1)} \dots M_{9 \times 5} M_{5 \times 3} \mathbf{P}(\mathbf{X}_1) \quad (21)$$

Figure 3 depicts the probability distributions for different values of p for the iteration $i = 7$. Figure 4 shows both the analytical (black continuous curves) and the corresponding numerical distributions obtained from 1000 realizations. As it can be seen, the distributions exhibit a peak at $X = 0$ whose height decreases as p increases. Besides this peak, the distributions also present other maxima whose heights depends on p , being smaller for lower values of p . This multimodality could be a consequence of the generation rule that fosters the emergence of clusters of scalable sizes. As a matter of fact, the fractal properties of these sequences are the main subject of a forthcoming paper.

The probability of generating a null sequence, i.e., formed completely by 0, after i substitutions, $P(X_i = 0)$, corresponds to the first component of $\mathbf{P}(\mathbf{X}_i)$. The following recurrence formula can be obtained [16]:

$$P(X_{i+1} = 0) = (P(X_i = 0) p + (1-p))^2 \quad (22)$$

for $i = 1, 2, \dots$. Note that $p P(X_i \neq 0)$ represents the probability of yielding a non-null sequence after a substitution that includes at least one 1 in the previous non-null sequence and then, $1 - p P(X_i \neq 0)$ represents the probability of yielding a null sequence after a substitution of length 1. This expression can be rewritten as:

$$1 - p P(X_i \neq 0) = 1 - p(1 - P(X_i = 0)) = 1 - p + p P(X_i = 0) \quad (23)$$

Because the substitution has length $k = 2$, and it is formed by independent digits, this expression must be multiplied by itself resulting in Equation (22).

Figure 2. A tree representation of the probabilities of obtaining a number of ones after $i = 3$ iterations using rule (1). As usual, $C(i, j)$ means the combinatorial numbers $C(i, j) = \binom{i}{j}$. The rightmost column represents the number of ones for each branch.

Figure 3. Discrete probability distributions of the number of ones for different values of p for the seventh iteration ($i = 7$). Concretely, $p = 0.5$ (a), $p = 0.8$ (b), $p = 0.9$ (c) and $p = 0.99$ (d). As it can be seen, the distributions for low value of p exhibit a peak at $X = 0$ that disappears when p tends to 1.

Figure 4. Analytical (black continuous curve) and the corresponding numerical distributions for $p = 0.9$ (a), $p = 0.92$ (b), $p = 0.94$ (c) and $p = 0.96$ (d). The numerical distributions are histograms (gray bars) obtained from 1000 realizations. Note the multimodal character of the distributions; for large values of p , in addition to the peak at $X = 0$, other peaks appear at $X \approx 18, 30, 54$ and 85 . The heights of these maxima vary with p and new peaks are visible as p increases (although, in the limit $p = 1$, only one remains).

The first term in this sequence is:

$$P(X_1 = 0) = (1 - p)^2 \quad (24)$$

which yields the second one:

$$P(X_2 = 0) = \left((1 - p)^2 p + (1 - p) \right)^2 \quad (25)$$

It is worthy to remark that the expansions of $P_i(X_i = 0)$ as polynomials of p and $q = 1 - p$ have the *triangle by rows* (oeis-A202019) [17] as coefficients.

Assuming that this sequence of probability distributions converges to ϕ , the following equation holds:

$$\phi = (\phi p + (1 - p))^2 \quad (26)$$

whose lowest solution:

$$\phi(p) = \min\left\{1, \left(\frac{1-p}{p}\right)^2\right\} \quad (27)$$

depends on p as depicted in Figure 5. Note that, for p below $p_c = \frac{1}{2}$ the probability of generating a null infinite sequence is one, i.e., $\phi(p) = 1$ for $p < p_c$.

Equation (22) can be generalized straightforwardly to any substitution length k using the same demonstration as above (see [16]). Now, the probability of having a null sequence at iteration $i + 1$ is obtained after multiplying k times the factor $1 - p P(X_i \neq 0)$:

$$P(X_{i+1} = 0) = (1 - p P(X_i \neq 0))^k = (P(X_i = 0)p + (1 - p))^k \quad (28)$$

As before, in the limit this equation converges to:

$$\phi = (\phi p + (1 - p))^k \quad (29)$$

that also exhibits a critical value of p :

$$p_c(k) = \frac{1}{k} \quad (30)$$

As it can be observed in Figure 5, as the substitution length k increases the probability of generating a null sequence decreases for $p > p_c(k)$; also does the range where this probability is one, i.e., $p_c(k) \rightarrow 0$ as $k \rightarrow \infty$.

Figure 5. Probability distribution of having a null sequence at the limit of the generating process, ϕ , for $k = 2$ (black), $k = 3$ (blue), $k = 4$ (red) and $k = 5$ (green), as a function of p .

Expected Values and Variance

Having already the probability distributions of ones, it is straightforward to compute its expected value after i iterations:

$$E(\mathbf{X}) = \sum_{j=0}^{2^i} j P_i(X = j) \quad (31)$$

This equation can be written in matrix form:

$$E(\mathbf{X}_i) = \mathbf{X}_i \mathbf{P}(\mathbf{X}_i) \quad (32)$$

Now, by applying Formula (18)

$$E(\mathbf{X}_i) = \mathbf{X}_i M_{(2^i+1) \times (2^{i-1}+1)} \mathbf{P}(\mathbf{X}_{i-1}) \quad (33)$$

It turns out that

$$\mathbf{X}_i M_{(2^i+1) \times (2^{i-1}+1)} = E^B(\mathbf{X}_i) \quad (34)$$

where by $E^B(\mathbf{X}_i)$ we refer to a vector whose entries are the expectation of a Binomial random variable $Y_j \sim \text{Bin}(2j, p)$, $j = 0, 1, 2, \dots, 2^{i-1}$ for successive iterations. Thus:

$$E^B(\mathbf{X}_i) = 2\mathbf{X}_i = (0, 2p, 4p, \dots, 2^i p) \quad (35)$$

Then, the following recurrence formula holds:

$$E(\mathbf{X}_{i+1}) = 2p E(\mathbf{X}_i) \quad (36)$$

for $i = 0, 1, 2, \dots$ and $E_1 = 2p$. This recurrence yields the expression:

$$E(\mathbf{X}_i) = \mu_i(p, 2) = (2p)^i \quad (37)$$

for $i = 0, 1, 2, \dots$

This formula can be straightforwardly generalized to any length k :

$$E(\mathbf{X}_i) = \mu_i(p, k) = (kp)^i \quad i = 0, 1, 2, \dots \quad (38)$$

The mean of the distribution of 0 in the sequence formed after i substitutions of length k is:

$$\mu_i^0(p, k) = k^i - (kp)^i = k^i(1 - p^i) \quad (39)$$

that yields a ratio #1/#0 at the i th-iteration:

$$\mu_i^{10}(p) = \frac{(kp)^i}{k^i(1 - p^i)} = \frac{p^i}{1 - p^i} \quad (40)$$

This ratio tends to 0 as $i \rightarrow \infty$ for all $0 \leq p < 1$ and for all k . Furthermore, the limit of $\mu_i^{10}(p)$ as $p \rightarrow 1$ is infinite for all values of $i = 1, 2, \dots$. Note that this ratio is one for the value of $p_{\mu=1} = 2^{-\frac{1}{i}}$, independently of k .

Similarly, it can be calculated the expected value of $\mathbf{X}^2$, $E(\mathbf{X}^2)$. First, for the case $k = 2$ we obtain:

$$E(\mathbf{X}_i^2) = \sum_{j=0}^i j^2 P(X = j) \quad (41)$$

For the second iteration, it is straightforward to calculate this value from this definition:

$$E(\mathbf{X}_1^2) = (0, 1, 4) \begin{pmatrix} (1-p)^2 \\ 2p(1-p) \\ p^2 \end{pmatrix} = 2p(1+p) \quad (42)$$

Using the recurrence of the probability distributions for successive iterations, the expected value of the next iteration is also computed:

$$E(\mathbf{X}_2^2) = (0, 1, 4, 9, 16) \begin{pmatrix} 1 & \binom{2}{0}(1-p)^2 & \binom{4}{0}(1-p)^4 \\ 0 & \binom{2}{1}p(1-p) & \binom{4}{1}p(1-p)^3 \\ 0 & \binom{2}{2}p^2 & \binom{4}{2}p^2(1-p)^2 \\ 0 & 0 & \binom{4}{3}p^3(1-p) \\ 0 & 0 & \binom{4}{4}p^4 \end{pmatrix} \begin{pmatrix} (1-p)^2 \\ 2p(1-p) \\ p^2 \end{pmatrix} \quad (43)$$

that simplifies to:

$$E(\mathbf{X}_2^2) = (0, 2p(1+p), 4p(1+3p)) \begin{pmatrix} (1-p)^2 \\ 2p(1-p) \\ p^2 \end{pmatrix} = 4p^2(2p^2 + p + 1) \quad (44)$$

Here we have used the expected value of X^2 for the binomial distribution for the number of ones in a sequence of length n .

In order to calculate the expected values of X^2 for successive iteration we find the following recursive formula:

$$E(\mathbf{X}_i^2) = 2p(1-p)E(\mathbf{X}_{i-1}) + (2p)^2 E(\mathbf{X}_{i-1}^2) \quad (45)$$

that can be derived using Equation (8) for the expected value of X^2 . Effectively,

$$E(\mathbf{X}_i^2) = \mathbf{X}_i^2 M_{(2^i+1) \times (2^{i-1}+1)} \mathbf{P}(\mathbf{X}_{i-1}) = E^B(\mathbf{X}_i^2) \mathbf{P}(\mathbf{X}_{i-1}) \quad (46)$$

where

$$\mathbf{X}_i^2 = (0, 1, 2^2, 3^2, \dots, (2^i)^2) \quad (47)$$

As before, $E^B(\mathbf{X}_i^2)$ represents a vector with entries $E(Y_j^2)$ where $Y_j \sim \text{Bin}(kj, p)$ for $j = 0, 1, 2, \dots, 2^{i-1}$ and, consequently, is given by:

$$E^B(\mathbf{X}_i^2) = (0, 2p(1+p), 2^2p(1+3p), \dots, 2^i p(2^i - 1)p) \quad (48)$$

Then, since this vector can be split into two vectors:

$$E^B(\mathbf{X}_i^2) = (0, 2p, 2^2p, \dots, 2^i p) + (0, 2p^2, 2^2p^2, \dots, 2^i p^2) \quad (49)$$

Equation (45) holds.

It can be proven that the expected value of $\mathbf{X}^2$ for the i -th iterations depends on p as follows:

$$E(\mathbf{X}_i^2) = 2^{i-1} p^i \left(1 + \sum_{j=0}^i (2p)^j \right) \quad (50)$$

for $n = 1, 2, 3, \dots$. If we replace the geometric sum:

$$E(\mathbf{X}_i^2) = \frac{1}{2} (2p)^i \left(1 + \frac{1 - (2p)^{i+1}}{1 - 2p} \right) \quad (51)$$

This expression can be applied to calculate the variance of the distribution of the number of ones in a sequence generated after i iterations, i.e., $\text{VAR}(\mathbf{X}_i) = E(\mathbf{X}_i^2) - (E(\mathbf{X}_i))^2$:

$$\text{VAR}(\mathbf{X}_i) = \frac{1}{2} (2p)^i \left(1 + \frac{1 - (2p)^{i+1}}{1 - 2p} \right) - (2p)^{2i} \quad (52)$$

that simplifies to:

$$\text{VAR}(\mathbf{X}_i) = \frac{1-p}{1-2p} (2p)^i (1 - (2p)^i) \quad (53)$$

This expression can be further compacted if we use the expression of the mean (37):

$$\text{VAR}(\mathbf{X}_i) = \frac{1-p}{1-2p} \mu_i(p) (1 - \mu_i(p)) \quad (54)$$

These computations can be equally performed for any length $k = 3, 4, \dots$. A general formulation for any value of k yields the following functions of p for the expected values of X and X^2 at the i -th iteration:

$$E(\mathbf{X}_i)(p, k) = (kp)^i \quad (55)$$

$$E(\mathbf{X}_i^2)(p, k) = (kp)^i \left(1 + (k-1)p \frac{1 - (kp)^i}{1 - kp} \right) \quad (56)$$

Then, the variance is straightforward calculated and yields:

$$\text{VAR}(\mathbf{X}_i) = \frac{1-p}{1-kp} (kp)^i (1 - (kp)^i) \quad (57)$$

The limits as i tends to infinity of the mean μ and the variance VAR depend on the value of p . For $0 < p < \frac{1}{k}$, these limits equal 0, whereas for $\frac{1}{k} < p < 1$, the limits are infinite.

Contrary to the Bernoulli sequences, the variance reaches a maximum for an intermediate value of p , that depends on both k and i (see Figure 6).

Figure 6. Standard deviation, $\sigma = \sqrt{\text{VAR}(\mathbf{X}_i)}$ for the substitution lengths $k = 2$ (black) and $k = 3$ (blue) after $i = 10$ substitutions as a function of p . Note the different scale at the Y-axis, likely as a consequence of the different sequence size 2^{10} vs. 3^{10} .

An interesting index that could provide information about the stochasticity of the generation process is the variance-mean ratio, also known as the dispersion index, D . For iteration i , it can be written as:

$$D_i(p, k) = \frac{\text{VAR}(\mathbf{X}_i)}{E(\mathbf{X}_i)} = \frac{1-p}{1-kp} (1 - (kp)^i) \quad (58)$$

By comparison with the Poisson distribution, that possesses a value of the dispersion equal to 1, it is said that a process that has a D -value lower than 1 is “under-dispersed”; on the contrary, having a D -value larger than 1 reflects an “over-dispersed” stochasticity. Since $D_i(0, k) = 1$ and the derivative of $D_i(p, k)$ is positive at $p = 0$ for all $i > 1$ and furthermore, $D_i(1, k) = 0$, then there exist a value of $p(k, i) = p_1$ such that $D_i(p_1, k) = 1$. This value of p is unique because $D_i(p, k)$ exhibits a unique maximum in $(0, 1)$. Moreover, it is very close to 1 for all k and i and tends to 1 as i tends to infinity, meaning that under-dispersion only occurs when the probability of inserting 0 is very low.

As observed in Figure 6, the standard deviation of X exhibits a maximum at $p = p_m$. Besides, this maximum exists for any k and i and it is achieved for the value of p such that:

$$\frac{d \text{VAR}(X_i)}{d p} = 0 \quad (59)$$

which can be written as follows:

$$k^i p^{i-1} \left(i + \sum_{j=1}^{i-1} (i+j) k^{j-1} p^j - i k^i p^i \right) = 0 \quad (60)$$

which implies that:

$$\sum_{j=1}^{i-1} (i+j) k^{j-1} p^j = i ((k p)^i - 1) \quad (61)$$

for any k and for each iteration i . Summing the finite sum of the left hand side of this equation yields to:

$$\frac{(1+i)k p - i k^2 p^2 - 2 i k^i p^i + (-1 + 2 i) k^{1+i} p^{1+i}}{k(-1 + k p)^2} = i ((k p)^i - 1) \quad (62)$$

This equation cannot be explicitly solved for p for any value of k and i . Nonetheless, the roots of the polynomial:

$$r_{k,i}(p) = i + \sum_{j=1}^{i-1} (i+j) k^{j-1} p^j - i k^i p^i \quad (63)$$

can be obtained numerically. For each substitution length, $k = 1, 2, \dots$, we are able to fit the roots of polynomial $r_{k,i}(p)$ to the function of i :

$$r_k(i) = \frac{1}{1 + \frac{\alpha(k)}{\alpha(k) + (i-1)\beta(k)}} \quad (64)$$

The values of the fitting parameters $\alpha(k)$ and $\beta(k)$, as well as the corresponding residual sum of squares, are shown in Table 1. As it can be observed, as the k -value increases, it seems that $\alpha(k) \rightarrow 0.5$ and $\beta(k) \rightarrow 1$, that is:

$$r_k(i) \rightarrow r(i) = 1 - \frac{1}{2i} \quad (65)$$

This suggests that the roots of the polynomials, i.e., the values of p where the variance of X achieves its maximum, tends to 1 as i tends to infinity. In addition, the variance of X at the maximum tends to infinity, although, by definition, this variance is null at $p = 1$. We have not a rigorous explanation of the reason behind the almost independence with k of the roots of the polynomial $r_{k,i}(p)$. Likely, it seems to be related to the self-similar properties of the generation process.

Table 1. Values of the fitting parameters for some values of the substitution length k . As it can be seen, as k increases the values of $\alpha(k)$ and $\beta(k)$ seems to converge to 0.5 and 1, respectively. The third column depicts the Residual Sum of Squares (RSS) of the fitted model.

k	α	β	RSS
2	0.6256	1.0645	1.78×10^{-5}
3	0.5832	1.0486	2.35×10^{-5}
4	0.5622	1.0383	2.10×10^{-5}
5	0.5497	1.0316	1.73×10^{-5}
10	0.5248	1.0167	6.99×10^{-6}
100	0.5025	1.0017	1.05×10^{-7}

4. Entropy

In this section, we study the mean entropy of a population of sequences generated from the substitution rules defined in the previous section. As a matter of illustration, we consider first the entropy of the Bernoulli process of probability p :

$$H(p) = -p \log_2(p) - (1-p) \log_2(1-p) \quad (66)$$

This function is symmetric and exhibits a maximum at $p = \frac{1}{2}$. If we now consider the sequences formed by n stochastic variables that follow a Bernoulli process, we can compute the entropy of a sequence of length n with a number of ones j in terms of the frequency $\frac{j}{n}$, that coincides, in average, with p :

$$H(j) = -\frac{j}{n} \log_2\left(\frac{j}{n}\right) - \left(1 - \frac{j}{n}\right) \log_2\left(1 - \frac{j}{n}\right) \quad (67)$$

Obviously, the sequences with minimum information content, $\frac{j}{n} = p = 0$ and $\frac{j}{n} = p = 1$, have null entropy, i.e., $H(0) = H(1) = 0$. In contrast, the maximum entropy is achieved for $j = \frac{n}{2}$ and takes the value: $H(\frac{n}{2}) = 1$.

For a population of Bernoulli sequences of probability p , the mean entropy can be calculated applying the binomial distribution to the entropy of the sequences generated with a probability p :

$$H_n^B(p) = \mathbf{H} \mathbf{P}(p) = \sum_{k=0}^n H(k) \binom{n}{k} p^k (1-p)^{n-k} \quad (68)$$

where $\mathbf{H}$ is the row vector entropy for a sequence of length n whose components, $H(k)$ are the corresponding entropies Equation (67) for all possible combinations of ones. For instance, for $n = 1$, this formula reduces to:

$$H_1^B(p) = \binom{2}{0} p^0 (1-p)^2 H(0) + \binom{2}{1} p^1 (1-p)^1 H(1) + \binom{2}{2} p^2 (1-p)^0 H(2) = 2p(1-p) \quad (69)$$

where $H(k)$ are the entropies of sequences with 0, 1 and 2 ones: $H(0) = 0$, $H(1) = \log_2(2) = 1$ and $H(2) = 0$. Similarly, the mean entropy for a length $n = 2$ can be calculated as follows:

$$H_2^B(p) = \sum_{k=0}^3 \binom{3}{k} p^k (1-p)^{n-k} H_2(k) = -3p(1-p) \left((1-p) \left(\frac{3}{4} \log(3) - 2 \right) + p \right) \quad (70)$$

To compute the mean entropy of a population of sequences randomly generated by substitutions according to the rules defined in Section 2, we have to know the probability of appearance of each of the frequencies of ones in order to weights adequately the sequence entropies. As an example, let us start with the substitution length $k = 2$. In this case, the

probability distributions of 1 for the i -th iteration are given by Equation (18), and this yields the mean entropy:

$$H_i(p) = \mathbf{H}_i \mathbf{P}(\mathbf{X}_i) \quad (71)$$

Here, the components of the entropy vector $\mathbf{H}_i$ are given by:

$$H_i(j) = -j 2^{-i} \log_2(j 2^{-i}) - (1 - j 2^{-i}) \log_2((1 - j 2^{-i})) \quad (72)$$

for a sequence of length $n = 2^i$, with a number of ones j , such that $0 \leq j \leq 2^i$.

Using the matrix expression for $\mathbf{P}$, (Equation (21)), this equation reduces to:

$$H_i(p) = \mathbf{H}_i M_{(2^{i+1}) \times (2^{i-1}+1)} M_{(2^{i-1}+1) \times (2^{i-2}+1)} \dots M_{5 \times 3} M_{3 \times 1} \mathbf{P}(\mathbf{X}_1) = \mathbf{H}_i M_{(2^{i+1}) \times 1} \mathbf{P}(\mathbf{X}_1) \quad (73)$$

The entropy in the first iteration, i.e., after substituting the initial 1 by applying the rule (1) with $k = 2$ is given by:

$$H_1(p) = \left(0, -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right), 0\right) \begin{pmatrix} \binom{2}{0} (1-p)^2 \\ \binom{2}{1} p(1-p) \\ \binom{2}{2} p^2 \end{pmatrix} \quad (74)$$

which yields:

$$H_1(p) = 2p(1-p) \quad (75)$$

In the next iteration, the entropy can be calculated by the formula:

$$H_2(p) = \mathbf{H}_2 M_{5 \times 3} \mathbf{P}(\mathbf{X}_1) \quad (76)$$

where

$$\mathbf{H}_2 = \left(0, -\frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right), -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right), -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right), 0\right) \quad (77)$$

Multiplying the matrices, we get:

$$H_2(p) = 10p^2(1-p) \left((p^3 - p^2 + \frac{1}{5}p + \frac{4}{5}) - \frac{3}{5}(p^3 - p^2 + \frac{1}{2}) \log_2(3) \right) \quad (78)$$

In the same way, we can calculate the entropy of the 8-digit binary sequences obtained after three substitutions:

$$\begin{aligned} H_3(p) = & 202p^3(1-p) \\ & \left[-20/101 p^3 + 4/101 p + p^{10} - 3p^9 + 285/101 p^8 + 15/101 p^7 - 173/101 p^6 + 91/101 p^5 + 17/101 p^4 - 12/101 p^2 + 12/101 + \right. \\ & + \left(-7/101 p^{10} + 21/101 p^9 - 42/101 p^8 + 35/202 p^7 + 42/101 p^6 - 42/101 p^5 - 7/101 p^4 + 21/202 p^3 + 7/101 p^2 - 7/202 \right) \log_2(7) - \\ & - \frac{35}{101} p \left(-27/35 p^7 + 9/14 p^6 + 39/35 p^5 - 33/35 p^4 - 9/35 p^3 + 3/35 p^2 + 9/70 p + 3/70 \right) \log_2(3) + \\ & \left. + \frac{35}{101} p^3 (1-p) \left(p^6 - 2p^5 + 4/7 p^4 + 13/14 p^3 - 1/2 p^2 - 3/14 p - 1/14 \right) \log_2(5) \right] \end{aligned} \quad (79)$$

In Figure 7, we depict the entropies calculated analytically for the first three iterations, as well as their corresponding numeric estimations. Note that, contrary to the Bernoulli sequences, the entropies of these sequences are not symmetric and achieve a maximum at a value of p that depends on the iteration. Furthermore, it can be shown that the entropy "per digit", i.e., $\frac{1}{2^i} H_i$, tends to 0 as $i \rightarrow \infty$ for all values of p .

The differential entropy for successive iterations:

$$h_i(p) = H_{i+1}(p) - H_i(p) \quad \text{for } i = 1, 2, \dots \quad (80)$$

presents an interesting behaviour, as it can be seen in Figure 7b. There is a critical value that separates two regimes: for $p < p_{ch}(i)$, $h_i(p)$ is negative, whereas for $p > p_{ch}(i)$, $h_i(p)$ is positive. As it can be seen, $p_{ch}(i)$ depends on the iteration: as $i \rightarrow \infty$, $p_{ch}(i) \rightarrow 1$. As a consequence, the p -interval where $h_i(p)$ is positive shrinks to 0. Besides, $h_i(p)$ tends to 0 for all p , which means that the entropy converges to the function $H_\infty(p)$, for the infinite sequence.

Figure 7. (a) Entropy as a function of p for the first 10 iterations. Note the displacement of the maximum to larger values of p as the i increases. (b) Differential entropy $h_i(p) = H_{i+1} - H_i$, for $i = 1, 2, 3, \dots, 9$. As it can be seen, $h_i(p)$ is positive for $p > p_{ch}(i)$. As the iteration increases, the interval of p where h_i is positive shrinks. As a matter of fact, in the limit, it tends to 0. Furthermore, h_i tends to 0 as $i \rightarrow \infty$, which means a convergence to a limit entropy for the infinite sequence. The substitution length is $k = 2$.

Because both variance and entropy are measures of uncertainty, it could be interesting to find a functional dependence between both (see Figures 7 and 8). The curves we see, for the case $k = 2$, can be parametrized either by p , for each iteration i , or by i , for each p . The parametric equations for each p and i are given by Equations (53) and (71), for the variance and the entropy, respectively. As it can be observed, the relationship between H and VAR depends on p for each iteration i . It is important to remark that, contrary to what is known for classical probability distributions [6,7], the dependence of H on VAR is not the graph of a function. Instead, the variation of $p \in [0, 1]$ brings about a closed curve. Certainly, for $p < p_r$, this functional dependence coincides with that of the classical probability distributions [6]. On the contrary, for $p > p_r$, where both variables decrease with p , we find an almost linear dependence between them. Figure 8 depicts the complete parametric curves for the first iterations obtained analytically.

An interesting consequence of this dependence is that, for the same variance, we can find a value of p that maximizes the entropy. This can be done for any iteration $i > 1$ (remember that for $i = 1$, $H(p) = VAR(p)$ for all $p \in [0, 1]$). Note that the dispersion index, D , of these sequences would be different because the mean would change at both values of p . As an example, we find two values of p with a similar variance, in particular,

$\text{VAR}(0.85) \approx \text{VAR}(0.99) \approx 8700$ but, with a mean entropy values $H(0.85) = 0.67$ and $H(0.99) = 0.41$. Obviously, we can also find values of p that give rise to sequences with the same entropy that have different variance. For example, a similar mean entropy is achieved for $p = 0.79$ and $p = 0.99$, whereas the variance of these populations are very distant, $\text{VAR}(0.79) \approx 3368$ and $\text{VAR}(0.99) \approx 8741$.

Figure 8. Parametric curves $H - \text{VAR}$ in terms of p for the first 10 iterations. All these curves can be split into two regimes: (i) for $0 < p < p_r$, where H has a concave (down) dependence on VAR and (ii) for $p > p_r$, when both variables decreases with p almost linearly. The value of p_r coincides with the point of the curve farthest from the origin. The dashed line connects these points. Due to the large value of the variance in comparison with the corresponding entropy values, the value of p_r tends to be the value where the maximum of VAR occurs. An interesting consequence of this dependence is that, for a given iteration i , we can find two values of p that yield sequences with the same variance but, with different entropies, having one of them the maximum value. The substitution length is $k = 2$.

5. Concluding Remarks

Motivated by the classical percolation process defined by Mandelbrot [4], we have studied a class of random substitutions of constant length that maps $0 \rightarrow \langle 0, 0, \dots, 0 \rangle$ and $1 \rightarrow \langle Y_1, Y_2, \dots, Y_k \rangle$, being Y_j a Bernoulli process with parameter $p \in [0, 1]$. Contrary to the classical Bernoulli process, this is an asymmetric rule that yields infinite sequences with an average ratio $\#1/\#0$ that tends to 0 for any $0 \leq p < 1$. By applying this substitution map, we have randomly generated binary sequences whose length increases with the substitution length k as k^i for $i = 0, 1, 2, \dots$ for iteration i . Specifically, we have analyzed the combinatorial, statistical and entropic properties of this substitution process starting from a sequence formed by the digit 1.

This asymmetry is reflected in the mass distribution of the number of ones, specifically with regards to its uncertainty. We have computed the variance of this distribution and the entropy of the ensemble of sequences generated at each iteration as a function of p . The relationship between these two magnitudes is depicted in the H - VAR curves of Figure 8. It is interesting to remark the two regimes that appear in the curves H - VAR , i.e., concavity down for $p < p_r$ and up for $p > p_r$ that, to the best of our knowledge, have not been shown before [6]. This allows to compare sequences generated with different values of H but with the same variance and, on the contrary, sequences that have the same entropy and different variance. We have left for a forthcoming paper the study of other properties of the generated sequences as, for instance, the distribution of the sizes of the sets of ones and the fractality, measured by the Hausdorff dimension.

The same kind of substitution rules can also be performed in the plane or in the space to generate sets with specific properties. In particular, random substitutions have been applied to study percolation properties of 2D fractal objects [2,4]. Using an iterated function system [3], random fractals are generated as a function of a probabilistic control parameter,

e.g. the probability of any site of being occupied, and their percolation characteristics are quantified in terms of this parameter.

Author Contributions: Conceptualization, J.C.N. and F.J.M.; Formal analysis, J.C.N. and F.J.M.; Investigation, J.C.N. and F.J.M.; Software, F.J.M.; Writing—original draft, J.C.N. and F.J.M.; Writing—review and editing, J.C.N. and F.J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially funded by the project PID2019-110330GB-C22 (Spanish Ministry of Science and Innovation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank the reviewers for their fruitful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fogg, N.P.; Berthé, V.; Ferenczi, S.; Mauduit, C.; Siegel, A. (Eds.) *Substitutions in Dynamics, Arithmetics and Combinatorics*; Springer: Berlin/Heidelberg, Germany, 2002.
2. Dekking, F.M.; Meester, R.W.J. On the Structure of Mandelbrot's Percolation Process and Other Random Cantor Sets. *J. Stat. Phys.* **1990**, *58*, 1109–1126. [[CrossRef](#)]
3. Falconer, K. *Fractal Geometry: Mathematical Foundations and Its Applications*; John Wiley & Sons Ltd.: New Jersey, NJ, USA, 1990.
4. Mandelbrot, B.B. *The Fractal Geometry of Nature*; W. H. Freeman and Company: New York, NY, USA, 1983.
5. Barbé, A.; Von Haeseler, F.; Skordev, G. Limit sets of restricted random substitutions. *Fractals* **2006**, *14*, 37–47. [[CrossRef](#)]
6. Mukherjee, D.; Ratnaparkhi, M.V. On the functional relationship between entropy and variance with related applications. *Commun. Stat.-Theory Methods* **1986**, *15*, 291–311.
7. Ebrahimi, N.; Maasoumi, E.; Soofi, E.S. Ordering univariate distributions by entropy and variance. *J. Econom.* **1999**, *90*, 317–336. [[CrossRef](#)]
8. García-Morales, V. Substitution systems and nonextensive statistics. *Phys. A Stat. Mech. Appl.* **2015**, *440*, 110–117. [[CrossRef](#)]
9. Wolfram, S. *A New Kind of Science*; Wolfram Media Champaign: Illinois, IL, USA, 2002.
10. Shallit, J.; Stolfi, J. Two methods for Generating fractals. *Comput. Graph.* **1989**, *13*, 185–191. [[CrossRef](#)]
11. Gelß, P.; Schütte, C. Tensor-generated fractals: Using tensor decompositions for creating self-similar patterns. *arXiv* **2018**, arXiv:1812.00814v1.
12. Voevodko, A.E. Fractal dimension of the kronecker product based fractals. *arXiv* **2018**, arXiv:1803.02766v1.
13. Xue, D.; Zhu, Y.; Zhu, G.X.; Yan, X. Generalized kronecker product and fractals. In *SPIE Digital Library, Proceedings of the Fourth International Conference on Computer-Aided Design and Computer Graphics, Wuhan, China, 23–25 October 1995*; International Society for Optics and Photonics: Washington, DC, USA, 1996.
14. Schroeder, M.R. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*; Courier Corporation: Massachusetts, MA, USA, 1991.
15. Baake, M.; Grimm, U.; Penrose, R. *Aperiodic Order*; Cambridge University Press: Cambridge, UK, 2013.
16. Chayes, L. Aspects of the fractal percolation process. Progress in Probability, Vol. 37. In *Fractal Geometry and Stochastics*; Springer: Berlin/Heidelberg, Germany, 1995.
17. Sloane, N.J.A. Sequence A202019 in The On-Line Encyclopedia of Integer Sequences. Available online: <http://www.oeis.org> (accessed on 16 February 2022).