mathematics

MDPI

Article

Low-Entropy Stochastic Processes for Generating k-Distributed and Normal Sequences, and the Relationship of These Processes with Random Number Generators †

Boris Ryabko [1,2]

[1] Institute of Computational Technologies of the Siberian Branch of the Russian Academy of Science, 630090 Novosibirsk, Russia; boris@ryabko.net or b.riabko@g.nsu.ru

[2] Department of Information Technologies, Novosibirsk State University, 630090 Novosibirsk, Russia

† This paper is an extended version of our paper published in Proceedings of the 2016 XV International Symposium Problems of Redundancy in Information and Control Systems (REDUNDANCY), St. Petersburg, Russia, 26–29 September 2016; pp. 132–136.

Received: 12 August 2019; Accepted: 6 September 2019; Published: 10 September 2019

Abstract: An infinite sequence $x_1 x_2 ...$ of letters from some alphabet $\{0, 1, ..., b-1\}$, $b \geq 2$, is called k-distributed ($k \geq 1$) if any k-letter block of successive digits appears with the frequency b^{-k} in the long run. The sequence is called normal (or ∞-distributed) if it is k-distributed for any $k \geq 1$. We describe two classes of low-entropy processes that with probability 1 generate either k-distributed sequences or ∞-distributed sequences. Then, we show how those processes can be used for building random number generators whose outputs are either k-distributed or ∞-distributed. Thus, these generators have statistical properties that are mathematically proven.

Keywords: stochastic process; k- distributed numbers; normal numbers; random number generator; pseudorandom number generator; two-faced processes; randomness; Shannon entropy

1. Introduction

In 1909, Borel defined k-distributed and ∞-distributed sequences as follows: A sequence of digits in base b is k-distributed if for any k-letter word w over the alphabet $\{0, 1, \dots, b-1\}$

$$\lim_{t \to \infty} \nu_t(w)/(t - |w|) = b^{-|w|} \tag{1}$$

where $\nu_t(w)$ is a number of occurrences of w in the sequence $x_1 ... x_{|w|}, x_2 ... x_{|w|+1}, ..., x_{t-|w|+1} ... x_t$). The sequence is normal (or ∞-distributed) if it is k-distributed for any $k \geq 1$. Borel called normal to base b a real number from the interval $(0, 1)$ whose expansion in base b is normal sequence, and showed that almost all real numbers are normal to any base (with respect to the uniform measure) [1,2]. It is interesting that the construction of ∞-distributed sequence in an explicit form was first achieved by Champernowne in 1933 [3], who proved that the sequence

$$0\,1\,2 \ldots 9\,10\,11\,12 \ldots 99\,100\,101\,102 \ldots$$

is ∞-distributed. Later, many ∞-distributed sequences were described and investigated in numerous papers (see for review [4]). Many researchers suppose that fractional parts of π, e, and $\sqrt{2}$ and some other "mathematical" constants are normal, but it is not proven [2,5]. On the other hand, for π empirical counting over several billions of its digits suggests that this might be true (see [5,6]).

One of the reasons for interest in k- and ∞-distributed sequences is due to the fact that they are closely related to the concept of randomness. If we imagine that someone tosses a fair coin with sides marked 0 and 1, he/she obtains (almost surely) an ∞-distributed sequence [2,5]. A mathematical model of such an experiment is a sequence of an independent and identically distributed (i.i.d.) symbols from $\{0,1\}$ generated with probabilities $(1/2,1/2)$. Note that quite often this i.i.d. process and the sequences generated from them are called "true random" [2].

The true random sequences are very desirable in cryptography, simulation and modeling applications. Of course, it is practically impossible to generate them tossing a coin and nowadays there are many so-called generators of pseudo-random numbers (PRNGs), whose aim is, informally speaking, to calculate sequences which mimic the truly random (see [2,7–10]). For brevity, in what follows, we consider the case when a process generates letters from the alphabet $\{0,1\}$, but the obtained results can be extended to the case of any alphabet.

Modern PRNGs is a computer program whose input is a short word (a so-called seed), whereas its output is a long (compared to the input) word. Having taken into account that the seed is a true random word, the PRNG can be considered as an expander of randomness which stretches a short seed into a long word [2,7,10]. The output of "perfect" PRNG would have to generate true random output sequence. However, it is impossible.

To be more precise, we note that a mathematically correct definition of a random sequence was obtained in a framework of algorithmic information theory established by Kolmogorov (see [11–15]). In particular, it is shown that any algorithm (i.e., a Turing machine) can neither generate (infinite) random sequences nor stretch a short random sequence into a longer one. It means that PRNGs do not exist. The same is true in a framework of Shannon information theory. Indeed, it is known that the Shannon entropy of the true random process (i.e., i.i.d. with probabilities $(1/2,1/2)$) is one bit per letter, whereas for all other processes the entropy (per letter) is less than one (see [16]). On the other hand, any PRNG stretches a short true random sequence into a long one. The entropy of the output is not greater then the entropy of the input and, hence, the per letter entropy of the output is strictly less than 1 bit. Therefore, the demands of true randomness and low entropy are contradictory. Thus, we see that, in a framework of algorithmic information theory, as well as in a framework of Shannon information theory, "perfect" PRNGs do not exist.

In such a situation, researchers suggest and investigate PRNGs, which meet some "probabilistic" properties of true random sequences [2,17]. In particular, a property that a PRNG generates ∞-distributed sequences is very desirable (see [2]).

Another important type of random number generators (RNGs) is physical random number generators, among which the so-called quantum random number generators (QRNG) have become very popular in recent decades and are widely used in practice. By definition, the physical RNGs are devices whose output is a binary sequence that must be truly random (or at least look truly random) (see [10]). According to M. Herrero-Collantes and J.C. Garcia-Escartin [10], a physical RNG can be divided into the two following blocks: the entropy source and the post-processing stage. The output of the source of entropy is a bit string obtained by measuring a physical random process with subsequent quantization. The goal of post-processing is to translate this bit string into a true random binary sequence. Nowadays, there are many methods of post-processing [10], but, nevertheless, the statistical (probabilistic) properties of many physical RNGs and, in particular, QRNGs, are not proven mathematically and should be experimentally tested [10,18,19]. Even the so-called device-independent QNRGs guarantee only the randomness of their output, but true randomness must either be verified or obtained by post-processing [10]. Thus, transformations that transform the output into a normal sequence are desirable for all types of RNGs.

Here, we describe such random processes that their entropy is much less than 1, but Equation (1) is valid for generated sequences either for all integers or for k from a interval $1, ..., K$, where K is an integer. This shows that there exist low-entropy PRNGs which generate such sequences that Equation (1) is valid (for $b = 2$). The description of the suggested processes show that they can be used to develop PRNGs with the property in Equation (1). The described processes are generalization of so-called two-faced processes suggested in [20–22].

In detail, we propose the following two processes. First, we describe the k-order Markov chain, which is a so-called two-faced process of order k, $k \geq 1$, for which, with probability one, for any generated sequence $x_1 x_2 ...$ and all binary words $w \in \{0,1\}^k$, the frequency of occurrence of the word w in the sequence $x_1 ... x_{|w|}, x_2 ... x_{|w|+1}, ..., x_{t-|w|+1} ... x_t ...$ goes to $2^{-|w|}$. Secondly, we describe so-called normal two-faced processes for which this property is true for all k.

We also propose the so-called two-faced transformation, which translates the trajectories of any random process into the trajectories of a two-faced process. This transformation is applicable to the creation of a PRNG with proven statistical properties.

2. *K*-Distributed Sequences and Two-Faced Markov Chains

First, we consider a pair of examples in order to explain the main idea of considered Markov chains. Let a matrix of transition probabilities T' be as follows:

$$\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & \alpha_0 & \alpha_1 \\ 1 & \alpha_1 & \alpha_0 \end{array} \quad , \tag{2}$$

where α_0 and α_1 are non-negative and their sum equals 1 (i.e., $P\{x_{i+1} = 0 | x_i = 0\} = \alpha_0$, $P\{x_{i+1} = 0 | x_i = 1\} = \alpha_1, ...$).

For example, let $\alpha_0 = 0.9, \alpha_1 = 0.1$. Then, the "typical" output sequence can be as follows:

$$0000000000111111111000000000011111110 \ldots .$$

On the one hand, this sequence is clearly not true random. On the other hand, the frequencies of 1s and 0s goes to $1/2$ due to the symmetry of the matrix in Equation (2). Hence, the output is 1-distributed. Again, based on the symmetry, we can build the following matrix of the second order whose output will be 2-distributed:

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 0 & \alpha_0 & \alpha_1 & \alpha_1 & \alpha_0 \\ 1 & \alpha_1 & \alpha_0 & \alpha_0 & \alpha_1 \end{array} \tag{3}$$

(Here, $P\{x_{i+1} = 0 | x_i = 0, x_{i-1} = 0\} = \alpha_0$, $P\{x_{i+1} = 0 | x_i = 0, x_{i-1} = 1\} = \alpha_1, ...$) For $\alpha_0 = 0.9, \alpha_1 = 0.1$, the "typical" output sequence can be as follows:

$$000000000000 \;\; 110110110110110110110110110 \;\; 000 \ldots ,$$

where gaps correspond to seldom transitions. It can be easily seen that frequency of any two-letter word goes to $1/4$.

Let us give a formal definition of two-faced Markov chains. First, we define two families of random processes $T_{k,p}$ and $\hat{T}_{k,p}$, where integer k and $p \in (0,1)$ are parameters. The processes $T_{k,p}$ and $\hat{T}_{k,p}$ are Markov chains of the connectivity (memory) k, which generate letters from the binary alphabet. We define them inductively. The matrix of $T_{k,p}$ is defined as follows: $T_{1,p}(0,0) = p, T_{1,p}(0,1) = 1-p$, $T_{1,p}(1,0) = 1-p, T_{1,p}(1,1) = p$. The process $\hat{T}_{1,\pi}$ is defined by $\hat{T}_{1,p}(0,0) = 1-p, \hat{T}_{1,p}(0,1) = p$, $\hat{T}_{1,p}(1,0) = p, \hat{T}_{1,p}(1,1) = 1-p$. Let the transition matrices $T_{k,p}$ and $\bar{T}_{k,p}$ be defined, then $T_{k+1,p}$ and $\hat{T}_{k+1,p}$ are as follows

$$T_{k+1,p}(0, 0u) = T_{k,p}(0,u), \tag{4}$$

$$T_{k+1,p}(1, 0u) = T_{k,p}(1,u),$$

$$T_{k+1,p}(0, 1u) = \hat{T}_{k,p}(0,u),$$

$$T_{k+1,p}(1,1u) = \hat{T}_{k,p}(1,u),$$

Conversely,

$$\hat{T}_{k+1,p}(0,0u) = \hat{T}_{k,p}(0,u), \tag{5}$$
$$\hat{T}_{k+1,p}(1,0u) = \hat{T}_{k,p}(1,u),$$
$$\hat{T}_{k+1,p}(0,1u) = T_{k,p}(0,u),$$
$$\hat{T}_{k+1,p}(1,1u) = T_{k,p}(1,u)$$

for all $u \in \{0,1\}^k$ (vu is a concatenation of v and u). We can see that

$$T_{k+1} = (T_k, \hat{T}_k)\,. \tag{6}$$

For example,

$$T_{2,p}(0,00) = p,\ T_{2,p}(0,01) = 1-p, T_{2,p}(0,10) = 1-p,\ T_{2,p}(0,11) = p.$$

To describe the process, the initial probability distribution should be defined. We say that the initial distribution of $T_{k,\pi}$ and $\hat{T}_{k,\pi}$ is uniform, if for all $w \in \{0,1\}^k$ $P\{x_1...x_k = w\} = 2^{-k}$. Sometimes, we consider different initial distributions, which is why, in all cases, the initial distribution is mentioned.

Let μ be stationary process. Its conditional Shannon entropy of order m, $m = 1,2,...$, is defined as follows

$$h_m = -\sum_{w\in\{0,1\}^{m-1}} \mu(w) \sum_{u\in\{0,1\}} \mu(u/w)\log\mu(u/w) \tag{7}$$

and the limit entropy is as follows

$$h_\infty = \lim_{r\to\infty} h_r\ , \tag{8}$$

see [16].

The main properties of Markov chains $T_{k,p}$ and $\hat{T}_{k,p}$, $k \geq 1$, are described by the following

Theorem 1. *Let $x_1x_2...$ be generated by $T_{k,p}$ (or $\hat{T}_{k,p}$), $k \geq 1$, and $w \in \{0,1\}^k$. Then,*

(i) If the initial distribution is uniform over $\{0,1\}^k$, then

$$P(x_{j+1}...x_{j+k} = w) = 2^{-|w|} \tag{9}$$

for any $j \geq 0$.

(ii) For any initial distribution of the Markov chain $T_{k,p}$ (or $\hat{T}_{k,p}$)

$$\lim_{j\to\infty} P(x_{j+1}...x_{j+k} = w) = 2^{-|w|}. \tag{10}$$

(iii) With probability one the Markov chains $T_{k,p}$ and $\hat{T}_{k,p}$ generates k-distributed sequences.

(iv) For any $p \in (0,1)$, the k-order Shannon entropy (h_k) of the processes $T_{k,p}$ ($\hat{T}_{k,p}$) equals 1 bit per letter, whereas the limit entropy equals $-(p\log_2 p + (1-p)\log_2(1-p))$.

The proof is given in Appendix A.

Having taken into account this theorem, we give the following.

Definition 1. *If Equation (10) is valid for any $w \in \{0,1\}^k$, the process is asymptotically two-faced of order k. The process is two-faced of order k, if Equation (9) is true.*

It turns out that, in a certain sense, there are many two-faced processes. More precisely, the following theorem is true.

Theorem 2. *Let $X = x_1 x_2 \dots$, $Y = y_1 y_2 \dots$ be random processes. We define the process $Z = z_1 z_2 \dots$ by following equations $z_1 = x_1 \oplus y_1$, $z_2 = x_2 \oplus y_2, \dots$ where $a \oplus b = (a+b) \bmod 2$. If X is a k-order (asymptotically) two-faced process, then Z is also the k-order (asymptotically) two-faced process ($k \geq 1$).*

The proof is given in Appendix A.

3. Two-Faced Transformation

Now, we show that any stochastic process can be transferred into a two-faced one. For this purpose, we describe transformations which transfer random processes into two-faced ones. First, we define matrices M_k and $\hat{M}_k$, $k \geq 1$, which are based on matrices $T_{k,p}$ and $\hat{T}_{k,p}$.

Definition 2. *The matrix M_k is defined by the following equation:*

Let us define the matrix M_k as follows:

$$M_k(w, v) = \begin{cases} 0, & if \quad T_{k,p}(w,v) = p \\ 1, & if \quad T_{k,p}(w,v) = 1-p \end{cases} \tag{11}$$

for any $k \geq 1$, $v \in \{0,1\}^k$, $w \in \{0,1\}$. The matrix $\hat{M}_k$ is obtained from $\hat{T}_{k,p}$ analogously. Note that, from Equation (6), we obtain

$$M_{k+1} = (M_k \, \hat{M}_k) \,. \tag{12}$$

Definition 3. *Let k be an integer, $v_1 \dots v_k \in \{0,1\}^k$, $u \in \{0,1\}$. Define functions τ_k and $\bar{\tau}_k$ as follows:*

$$\tau_k(u, v_1 \dots v_k) = M_k(u, \;\; v_1 \dots v_k)\,, \;\; \bar{\tau}_k(u, v_1 \dots v_k) = \hat{M}_k(u, \;\; v_1 \dots v_k)\,. \tag{13}$$

Let $X = x_1 x_2 \dots$, and $v \in \{0,1\}^k$. The two-faced transformation τ_k maps a pair (X, v) into a sequence Y as follows: $y_{-k+1} y_{-k+2} \dots y_0 = v$, $\;y_i = \tau_k(x_i, \;\; y_{i-k} y_{i-k+1} \dots y_{i-1})\,,\;$ where $i = 1, 2, \dots$.

Note that, from this definition and Equation (12), we can see that for any $i \geq 1$

$$\tau_k(x, y_{i-k+1} y_{i-k+2} \dots y_i) = \begin{cases} \tau_{k-1}(x, y_{i-k+2} \dots y_i)\,, & if \quad y_{i-k+1} = 0 \\ \bar{\tau}_{k-1}(x, y_{i-k+2} \dots y_i)\,, & if \quad y_{i-k+1} = 1 \end{cases} \tag{14}$$

Theorem 3. *Let $k \geq 1$ be an integer, $X = x_1 x_2 \dots$ be generated by a stochastic process, and τ_k be a two-faced transformation. If v is uniformly distributed on $\{0,1\}^k$, then for any $u \in \{0,1\}^k$ and $r \geq 1$*

$$P\{y_{r-k}\, y_{r-k+1} \dots y_{r-1} = u\} = 2^{-k}\,, \tag{15}$$

i.e., $\tau_k(X, v)$ is two-faced of order k process. The proof is given in Appendix A.

Consider now the question of the complexity of the described transformation τ_k allowing transform any process into a two-faced. When directly implementing the transformation τ_k, one must store matrix M_k of 2 rows and 2^k columns, i.e., just 2^{k+1} numbers. Storing such matrices becomes impossible when k exceeds hundreds. Therefore, the question arises of constructing a simpler algorithm that does not require an exponential growth of memory with increasing k. It turns out that there exists an algorithm which requires $O(k)$ bits of memory and finite number of operation (per an output letter).

To describe this algorithm, we first define transformations τ_k^* and $\bar{\tau}_k^*$. Let there be an infinite word $x_1 x_2 \dots$ and a finite one $y_{-k+1} y_{-k+2} \dots y_0$. For any $r \geq 1$, denote $H_r = (\sum_{r-k+1}^{r} y_i) \mod 2$ $\bar{H} = H \oplus 1$. Then, $\tau_k^*(x_{r+1}, y_{r-k+1} y_{r-k+2} \dots y_r) = H_r \oplus x_{r+1}$ and $y_{r+1} = \tau_k^*(x_{r+1}, y_{r-k+1} y_{r-k+2} \dots y_r)$.

Similarly, $\bar{\tau}^*_k(x_{r+1}, y_{r-k+1}y_{r-k+2}...y_r) = \bar{H}_r \oplus x_{r+1}$. From those definitions and Equation (12), we can see that for any $i \geq 1$

$$\tau^*_k(x, y_{i-k+1}y_{i-k+2} \dots y_i) = \begin{cases} \tau^*_{k-1}(x, y_{i-k+2} \dots y_i), & if \quad y_{i-k+1} = 0 \\ \bar{\tau}^*_{k-1}(x, y_{i-k+2} \dots y_i), & if \quad y_{i-k+1} = 1 \end{cases} \tag{16}$$

It is important to note that there exists the simple algorithm for carrying out the transformation τ^*_k. Indeed, just store the letters $y_{r-k+1}\, y_{r-k+2}...y_r$ and the value H_r in the computer's memory. Then, read the letter x_{r+1}, calculate $y_{r+1} = H \oplus x_{r+1}$, include y_{r+1} and exclude y_{r-k+1}, i.e., store the new word $y_{r-k+2}\, y_{r-k+2}...y_{r+1}$. Then, calculate the new $H_k := H_k \oplus y_{r+1} \oplus y_{r-k+1}$, read the new letter x_{r+2} and so on.

Theorem 4. *The transformation τ^*_k equals τ_k, and, hence, the above-described algorithm performs the transformation τ_k in time $O(1)$ using memory $O(k)$ bits.*

The proof is given in Appendix A.

4. ∞-Distributed Processes

The k-order two-faced process is k distributed. Here, we describe ∞-distributed processes. We call suggested processes as normal two-faced.

Definition 4. *If, for any binary word v, Equation (9) is true, then the process is called normal two-faced, while, if Equation (10) is true, the process is asymptotically normal two-faced.*

Now, we describe a family of such processes. Suppose that $m^* = m_1, m_2, ...$ is a sequence of integers, $m_1 < m_2 < m_3....$ and $X^1 = x^1_1 x^1_2...$, $X^2 = x^2_1 x^2_2...$, $X^3 = x^3_1 x^3_2...$, ... are (asymptotically) two-faced processes of order $m_1, m_2, ...$, correspondingly. Define a process $W = w_1 w_2$ ... by the following equation:

$$w_i = \begin{cases} x^1_i & i \leq m_1, \\ x^1_i \oplus x^2_i & m_1 < i \leq m_2, \\ x^1_i \oplus x^2_i \oplus x^3_i & m_2 < i \leq m_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots & \end{cases} \tag{17}$$

and denote it as $\bigoplus_{i=1}^{\infty} X^i$.

Theorem 5. *Let all X^i, $i = 1, 2, ...$, be two-faced. Then, $\bigoplus_{i=1}^{\infty} X^i$ is normal two-faced. If X^i, $j = 1, 2, ...$ are asymptotically two-faced, then $\bigoplus_{i=1}^{\infty} X^i$ is asymptotically normal two-faced.*

The proof is given in Appendix A. From this and Theorem 2, we can derive the following.

Corollary 1. *If $X = x_1 x_2...$ and $Y = y_1 y_2...$ are stochastic processes and X is normal two-faced, then the process $Z = z_1 z_2...$, $z_1 = x_1 \oplus y_1$, $z_2 = x_2 \oplus y_2, ...$ is normal two-faced.*

Note that the entropy of the processes $X^1, X^2, ...$ can be small; hence, the entropy of the process $\bigoplus_{i=1}^{\infty} X^i$ can be arbitrary small. On the other hand, the process looks truly random.

5. Experiments

Here, we present some experiments describing the two-faced processes with different parameters. We compared obtained sequences with truly random applying the χ^2 test [23]. For this purpose, N-letter sequence $x_1 x_2...x_N$, $N = 1000$, were generated, whereas the initial part $x_{-k+1}...x_0$ was uniformly distributed. The sequence $x_1 x_2...x_N$ was presented as $x_1 x_2 x_k$, $x_{k+1} x_{k+2}...x_{2k}$, ... and the frequency of occurrence of all words from $\{0, 1\}^k$ was estimated. Then,

$$x^2 = \sum_{w \in \{0,1\}^k} (\nu(w) - (\lfloor N/k \rfloor / 2^k))^2 / ((\lfloor N/k \rfloor / 2^k)$$

was calculated, where $N = 1000$, $\nu(w)$ is the number of occurrences of w in the sequence $x_1 x_2 x_k$, $x_{k+1} x_{k+2} ... x_{2k}$, ... (Note that x^2 estimates the frequency deviation from the uniform distribution.) Then, x^2 was compared with the quantile $\chi^2_{d,\, 0.99}$, where $d = 2^k - 1$; see [23]). If $x^2 > \chi^2_{d,\, 0.99}$, we rejected H_0. Table 1 contains results of calculations. (The entropy is equal to $-(p \log_2 p + (1-p) \log_2 (1-p)))$.

Table 1. Two-faced processes testing.

π	k	Accepted	Entropy (Bits)
0.3	2	10	0.88
0.3	3	10	0.88
0.3	4	10	0.88
0.3	5	10	0.88
0.2	2	9	0.72
0.2	3	9	0.72
0.2	4	10	0.72
0.2	5	10	0.72
0.1	2	9	0.47
0.1	3	9	0.47
0.1	4	9	0.47
0.1	5	9	0.47
0.05	2	10	0.29
0.05	3	8	0.29
0.05	4	5	0.29
0.05	5	4	0.29
0.01	2	1	0.08
0.01	3	4	0.08
0.01	4	1	0.08
0.01	5	0	0.08

Thus, we can see that the two-faced processes can be obtained from low-entropic ones.

6. Conclusions

In this paper, we describe low-entropic processes which mimic truly random ones. In other words, the output is either ∞-distributed or k-distributed for some integer k. In addition, we show how those processes can be directly used in order to construct (or "improve ") PRNGs.

Funding: This work was supported by Russian Foundation for Basic Research (grant 18-29-03005).

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proofs of Theorems

Proof of Theorem 1. We prove that

$$p^*(x_1 ... x_k) = 2^{-k}, \tag{A1}$$

$(x_1 ... x_k) \in \{0,1\}^k$, is a limit (or stationary) distribution for the processes $T_{k,p}$ and $\hat{T}(k,p)$. For this purpose, we show that the system

$$p(x_1 ... x_k) = p(0 x_1 ... x_{k-1})\, T_{k,p}(x_k / 0 x_1 ... x_{k-1})$$

$$+ p(1 x_1 ... x_{k-1})\, T_{k,p}(x_k / 1 x_1 ... x_{k-1}) \,, \qquad x_1 ... x_k \in \{0,1\}^k;$$

$$\sum_{v \in \{0,1\}^k} p(v) = 1$$

has the solution $p(x_1...x_k) = 2^{-k}$, $(x_1...x_k) \in \{0,1\}^k$. Having taken into account the definitions and Equations (4) and (5), we can see that the equality $T_{k,p}(x_k/\ 0x_1\ ...x_{k-1}) + T_{k,p}(x_k/1x_1...x_{k-1}) = 1$ is valid for all $(x_1...x_k) \in \{0,1\}^k$. From the law of total probability and the latest equation, we derive Equation (A1). Taking into account that the initial distribution is uniform and, hence, is the limiting one, we derive the first claim in Equation (9). Any transition probability is either p or $1-p$, hence, they are greater than 0, thus $T_{k,p}$ is ergodic and Equation (9) is true due to ergodicity.

Let us prove Statement (iii). All transition probabilities of $T_{k,p}$ are nonzero numbers. Hence, this Markov chain is a stationary ergodic process; therefore, for any $w \in \{0,1\}^k$, the limit $\lim_{t\to\infty} \nu_t(w)/(t-|w|)$ equals $E(P\{x_{j+1}...x_{j+k} = w\})$ (see [24]). From this and Statement (ii), we obtain Statement (iii).

Let us prove Statement (iv). From Equations (4) and (5), we see that either $T_{k,p}(0/x_1...x_k) = p$, $T_{k,p}(1/x_1...x_k) = 1-p$ or $T_{k,p}(0/x_1...x_k) = 1-p$, $T_{k,p}(1/x_1...x_k) = p$. Taking into account Equation (7), we see from the latest equations that $h_{k+1} = -(p\log_2 p + (1-p)\log_2(1-p))$. From this and Equation (8), we derive $h_\infty = -(p\log_2 p + (1-p)\log_2(1-p))$. The theorem is proven. □

Proof of Theorem 2. The first claim follows from the next equations:

$$\begin{aligned} &P\{z_{j+1}...z_{j+k} = w\} = \\ &\sum_{v\in\{0,1\}^k} P\{x_{j+1}...x_{j+k} = v\}P\{y_{j+1}...y_{j+k} = v \oplus w\} \\ &= 2^{-k} \sum_{v\in\{0,1\}^k} P\{y_{j+1}...y_{j+k} = w \oplus v\} = 2^{-k} \times 1 = 2^{-k}\,. \end{aligned} \tag{A2}$$

(It follows from Equation (9) and $v \oplus w \oplus v = w$.) Note that, by definition,

$$\lim_{j\to\infty} P(x_{j+1}...x_{j+k} = w) = 2^{-|w|}$$

for all $w \in \{0,1\}^k$, see Equation (10). Thus, for any $\delta > 0$, there is such J that

$$|P(x_{j+1}...x_{j+k} = w) - 2^{-|w|}| < \delta,\ w \in \{0,1\}^k$$

if $j > J$. From this and Equation (A2), we we can see that

$$(2^{-k} - \delta) \sum_{v\in\{0,1\}^k} P\{y_{j+1}...y_{j+k} = w \oplus v\}$$

$$\le P\{z_{j+1}...z_{j+k} = w\} \le$$

$$(2^{-k} + \delta) \sum_{v\in\{0,1\}^k} P\{y_{j+1}...y_{j+k} = w \oplus v\}\,.$$

From this, we obtain that:

$$(2^{-k} - \delta) \le P\{z_{j+1}...z_{j+k} = u\} \le (2^{-k} + \delta)\,.$$

Thus, Equation (9) is true. The theorem is proven. □

Proof of Theorem 3. We prove Equation (15) by induction on r. By the condition of the theorem, y_{-k+1} y_{-k+2} $...y_0$ obeys the uniform distribution on $\{0,1\}^k$, hence Equation (15) is true for $r = 1$. Supposing this equation is proven for r, let us prove it for $r+1$. The matrix $M_k(\ ,\ )$ has 2^k columns, each of which contains 0 and 1. For any x, half of the corresponding elements of the row $M_k(x,\ )$ are 0, whereas the others are 1. By induction, $y_{r-k+1}\ y_{r--k+2}\ ...y_r$ obeys the uniform distribution; hence, with probability $1/2$, $M_k(x_{r+1}, y_{r-k+1}\ y_{r--k+2}\ ...y_r) = 0$. The theorem is proven. □

Proof of Theorem 4. We prove by induction on k. For $k = 1$, it is true by the definitions of matrices T_1 and M_1. Supposing the equation is proven for $k \geq 1$, let us prove it for $k + 1$. From this and Equation (16), we obtain

$$\tau^*_{k+1}(x, y_{i-k}y_{i-k+1} \ldots y_i) = \tau_k(x, y_{i-k+1}y_{i-k+2} \ldots y_i) \; if \; y_{i-k} = 0\,,$$

$$\tau^*_{k+1}(x, y_{i-k}y_{i-k+1} \ldots y_i) = \bar{\tau}_k(x, y_{i-k+1} \ldots y_i) \; if \; y_{i-k} = 1\,,$$

From this equation and Equation (14), we can see that $\tau^*_{k+1} = \tau_{k+1}$. □

Proof of Theorem 5. Suppose $w \in \{0, 1\}^k$. There exists such an integer n_i that $n_i \geq k$ and define $D = \bigoplus_{j=1}^{i-1} X^j \oplus \bigoplus_{j=i+1}^{\infty} X^j$. (Here, $W \oplus V = \{w_1 \oplus v_1 \; w_2 \oplus v_2 \; w_3 \oplus v_3 \ldots\}$.) Clearly, $\bigoplus_{j=1}^{\infty} X^j = X^i \oplus D$. X^i is (asymptotically) n_i-order two-faced and, from Theorem 2, we derive that $\bigoplus_{j=1}^{\infty} X^j$ is n_i-order two-faced. Taking into account that $k \leq n_i$, we can see that $\bigoplus_{j=1}^{\infty} X^j$ is k-order (asymptotically) two-faced. Thus, Equation (9) (Equation (10)) is true and the theorem is proven. □

References

1. Borel, E. Le continu mathématique et le continu physique. 1909. Available online: https://fr.wikisource.org/wiki/Le_continu_math%C3%A9matique_et_le_continu_physique (accessed on 1 August 2019)
2. L'Ecuyer, P. History of uniform random number generation. In Proceedings of the WSC 2017-Winter Simulation Conference, Las Vegas, NV, USA, 3–6 December 2017.
3. Champernowne, D.G. The construction of decimals normal in the scale of ten. *J. Lond. Math. Soc.* **1933**, *1*, 254–260. [CrossRef]
4. Bailey, D.H.; Crandall, R.E. Random generators and normal numbers. *Exp. Math.* **2002**, *11*, 527–546. [CrossRef]
5. Bailey, D.H.; Borwein, J.M.; Calude, C.S.; Dinneen, M.J.; Dumitrescu, M.; Yee, A. An empirical approach to the normality of π. *Exp. Math.* **2012**, *21*, 375–384. [CrossRef]
6. Bailey, D.H.; Borwein, J.M.; Brent, R.P.; Reisi, M. Reproducibility in computational science: A case study: Randomness of the digits of pi. *Exp. Math.* **2017**, *26*, 298–305. [CrossRef]
7. L'Ecuyer, P. *Random Number Generation and Quasi-Monte Carlo*; Wiley: Hoboken, NJ, USA, 2014.
8. Marsaglia, G. Xorshift rngs. *J. Stat. Softw.* **2003**, *8*, 1–6. [CrossRef]
9. Impagliazzo, R.; Zuckerman, D. How to recycle random bits. In Proceedings of the 30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, NC, USA, 30 October–1 November 1989; pp. 248–253.
10. Herrero-Collantes, M.; Garcia-Escartin, J.C. Quantum random number generators. *Rev. Mod. Phys.* **2017**, *89*, 015004. [CrossRef]
11. Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*; Springer-Verlag: Berlin/Heidelberg, Germany, 2008.
12. Calude, C.S. *Information and Randomness—An Algorithmic Perspective*; Springer-Verlag: Berlin/Heidelberg, Germany, 2002.
13. Downey, R.G.; Hirschfeldt, D.R. *Algorithmic Randomness and Complexity*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
14. Downey, R.; Hirschfeldt, D.R.; Nies, A.; Terwijn, S.A. Calibrating randomness. *Bull. Symb. Log.* **2006**, *12*, 411–491. [CrossRef]
15. Merkle, W.; Miller, J.S.; Nies, A.; Reimann, J.; Stephan, F. Kolmogorov–loveland randomness and stochasticity. *Ann. Pure Appl. Log.* **2006**, *138*, 183–210. [CrossRef]
16. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: New York, NY, USA, 2006.
17. L'Ecuyer, P.; Simard, R. TestU01: A C library for empirical testing of random number generators. *ACM Trans. Math. Softw. (TOMS)* **2007**, *33*, 22. [CrossRef]
18. Ma, X.; Yuan, X.; Cao, Z.; Qi, B.; Zhang, Z. Quantum random number generation. *NPJ Quantum Inf.* **2016**, *28*, 6021. [CrossRef]
19. Tamura, K.; Shikano, Y. Quantum Random Numbers generated by the Cloud Superconducting Quantum Computer. *arXiv* **2019**, arXiv:1906.04410.

20. Ryabko, B.; Suzuki, J.; Topsoe, F. Hausdorff dimension as a new dimension in source coding and predicting. In Proceedings of the 1999 IEEE Information Theory and Communications Workshop, Kruger National Park, South Africa, 25 June 1999; pp. 66–68.
21. Ryabko, B.; Monarev, V. Using information theory approach to randomness testing. *J. Stat. Plan. Inference* **2005**, *133*, 95–110. [CrossRef]
22. Ryabko, B.; Fionov, A. *Basics of Contemporary Cryptography for IT Practitioners*; World Scientific Publishing Co.: Singapore, 2005.
23. Kendall, M.; Stuart, A. *The Advanced Theory of Statistics; Vol.2: Inference and Relationship*; Hafner Publishing Company: New York, NY, USA, 1961.
24. Billingsley, P. *Ergodic Theory and Information*; John Wiley & Sons: Hoboken, NJ, USA, 1965.