#A10 INTEGERS 10 (2010), 111-127

PATTERN OCCURRENCE IN THE DYADIC EXPANSION OF SQUARE ROOT OF TWO AND AN ANALYSIS OF PSEUDORANDOM NUMBER GENERATORS

Koji Nuida
Research Center for Information Security, National Institute of Advanced Industrial Science and Technology, Akihabara-Daibiru Room 1003, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan
k.nuida@aist.go.jp

Received: 3/18/09, Revised: 12/10/09, Accepted: 12/15/09, Published: 3/5/10

Abstract

Recently, designs of pseudorandom number generators (PRNGs) using integer-valued variants of logistic maps and their applications to certain cryptographic schemes have been studied, due mostly to their ease of implementation and performance. However, it has been noted that this ease is reduced for some choices of the PRNGs accuracy parameters. In this article, we show that the distribution of such undesirable accuracy parameters is closely related to the occurrence of some patterns in the dyadic expansion of the square root of 2. We prove that for an arbitrary infinite binary word, the asymptotic occurrence rate of these patterns is bounded in terms of the asymptotic occurrence rate of zeroes. As a consequence, a classical conjecture on asymptotic evenness of occurrence of zeroes and ones in the dyadic expansion of the square root of 2 implies that the asymptotic rate of the undesirable accuracy parameters for the PRNGs is at least 1/6.

1. Introduction

Randomness is a ubiquitous element in our present life and can be found from a simple coin toss at the beginning of a football game, to more complex settings such as encrypted communication of governmental secrets. The provision, application and evaluation of randomness has occupied a major and attractive branch of mathematics. In particular, there exist several methods and techniques that generate a seemingly random-looking sequence by using shorter random sequences (often called a *seed*) and deterministic algorithm, better known as a *pseudorandom number generator* (*PRNG*), see for instance [3] for references therein.

In this article, we reveal a nontrivial relation between analysis of some PRNGs and properties of the dyadic expansion of $\sqrt{2} = (1.01101\cdots)_2$. Note, however, that the dyadic expansion of $\sqrt{2}$ does not appear in the construction of the PRNGs

itself. Also, it has been shown that the logistic map

$$L(x) = \mu x(1-x) \ , \ 0 < x < 1 \ ,$$

for some parameter μ can be effectively used for constructing good PRNGs ([7, 8]). In particular, when $\mu = 4$ is adopted, the logistic map shows chaotic behavior. However, those PRNGs deal with real number as outputs and can therefore not be implemented in computers due to their finite accuracy. As a result, a modified integer-valued logistic map of the form:

$$L_n(x) = \left\lfloor \frac{4x(2^n - x)}{2^n} \right\rfloor = \left\lfloor \frac{x(2^n - x)}{2^{n-2}} \right\rfloor \ , \ x \in X_n = \{1, 2, \ldots, 2^n - 1\}$$

where $\lfloor z \rfloor$ denotes the largest integer N such that $N \leq z$ and $2 \leq n \in \mathbb{Z}$ is an accuracy parameter, has been proposed and studied in [1, 5]. The definition of $L_n(x)$ is derived from $L(x)$ by expanding the bounds of the original seed $x \in (0, 1)$ to the larger interval $(0, 2^n)$ and then truncating the final value to obtain an integer. The corresponding PRNG first chooses an internal state $s_0 = s$ from the set X_n and then for each step $i \geq 1$, updates the internal state by $s_i = L_n(s_{i-1})$ and outputs some bits in the dyadic expansion of s_i.

For the above PRNG, it has been mentioned in [4] that when $s_i = 2^{n-1}$ for some i, the subsequent internal states eventually become stable, i.e., we have $s_{i+1} = 2^n$ and $s_j = 0$ for every $j \geq i + 2$. Since stable internal states are fatal for the purpose of providing good randomness, the value 2^{n-1} should not be used as an internal state. To correct the problem, it is not enough to simply exclude the value 2^{n-1} itself from the candidates of the initial internal state s_0. Namely, if there exists an $x \in X_n$ such that $L_n(x) = 2^{n-1}$, then the choice of internal state $s_0 = x$ for such an x also makes the internal states eventually stable. We call the accuracy parameter n *undesirable* if such an x exists, since in such a case an extra check is required for choosing an appropriate initial internal state. The purpose of this work is to estimate how many undesirable parameters exist among the integers $n \geq 2$.

We explain an aforementioned relation of the above PRNGs with combinatorial properties of $\sqrt{2}$. Let $b_i \in \{0, 1\}$ denote the i-th bit of the fractional part of the dyadic expansion of $\sqrt{2}$, namely

$$\sqrt{2} = (1.b_1 b_2 b_3 \cdots)_2 \ .$$

We show that a parameter $n \geq 2$ is undesirable if the $(n-1)$-th tail $b_{n-1} b_n b_{n+1} \cdots$ of the dyadic expansion of $\sqrt{2}$ begins with one of the three patterns 00, 0100, and 01010. For instance, since

$$\sqrt{2} = (1.011010100000100 \cdots)_2 \ , \tag{1}$$

we have $b_{12} = b_{14} = b_{15} = 0$ and $b_{13} = 1$, implying $n = 13$ is undesirable. As a result, the occurrence rate of these three patterns in the dyadic expansion of $\sqrt{2}$ gives a lower bound of the occurrence rate of undesirable parameters. Motivated by the observation, we study the distributions of the three patterns in *arbitrary* infinite binary words $w = w_1 w_2 w_3 \cdots$, and prove that the asymptotic occurrence rate of the three patterns in w is bounded by a function of the asymptotic occurrence rate of zeroes in w (see Theorem 6 for the precise statement). This result connects the asymptotic occurrence rate of undesirable parameters to the distribution of zeroes in the dyadic expansion of $\sqrt{2}$. For the latter, it has been conjectured that the asymptotic occurrence rate of zeroes in the dyadic expansion of $\sqrt{2}$ is $1/2$ (in other words, $\sqrt{2}$ is simply normal to the base 2). If the conjecture is true, it follows, by applying our above-mentioned general result (Theorem 6), that the asymptotic occurrence rate of undesirable parameters is lower bounded by $1/6$, which shows a disadvantage of the above PRNGs.

This article is organized as follows. In Section 2, we prove the aforementioned sufficient condition of an accuracy parameter n being undesirable, in terms of the occurrence rate of the three patterns 00, 0100, and 01010 in the dyadic expansion of $\sqrt{2}$. In Section 3, we state the main theorem (Theorem 6) on a relation between the asymptotic occurrence rates of the three patterns and of zeroes in arbitrary infinite binary words. As a result, we also estimate the asymptotic occurrence rate of undesirable parameters. Finally, Section 4 gives the proof of the main theorem.

2. Integer-Valued Logistic Maps

As mentioned in Section 1, the main aim of this article is to study the integer-valued logistic maps $L_n(x)$ with domain $X_n = \{1, 2, \ldots, 2^n - 1\}$, parameterized by an integer $n \geq 2$. These are defined by

$$L_n(x) = \left\lfloor \frac{4x(2^n - x)}{2^n} \right\rfloor = \left\lfloor \frac{x(2^n - x)}{2^{n-2}} \right\rfloor \quad , \quad x \in X_n = \{1, 2, \ldots, 2^n - 1\} \ . \tag{2}$$

Note that $L_n(x) \in X_n$ for any $x \in X_n \setminus \{2^{n-1}\}$, while $L_n(2^{n-1}) = 2^n$. We would like to estimate the asymptotic occurrence rate of accuracy parameters n, among all integers $n \geq 2$, that satisfy the following condition:

Definition 1 *We say that a parameter $2 \leq n \in \mathbb{Z}$ is* undesirable *if there exists an $x \in X_n$ such that $L_n(x) = 2^{n-1}$.*

This definition is motivated by an analysis of some pseudorandom number generators (PRNGs) using $L_n(x)$; see Section 1 for details. In the rest of this section,

we show that the occurrence rate of undesirable parameters is related to the occurrence of the patterns 00, 0100, and 01010 in the dyadic expansion of $\sqrt{2}$. For the purpose, first note that by the definition, a parameter n is undesirable if and only if there exists an $x \in X_n$ such that $2^{n-1} \leq x(2^n - x)/2^{n-2} < 2^{n-1} + 1$. By solving the inequality, it follows that this condition for x is equivalent to

$$\sqrt{2^{2n-3} - 2^{n-2}} < |2^{n-1} - x| \leq \sqrt{2^{2n-3}} \ . \tag{3}$$

Moreover, since

$$\sqrt{2^{2n-3}} - \sqrt{2^{2n-3} - 2^{n-2}} = \frac{2^{n-2}}{\sqrt{2^{2n-3}} + \sqrt{2^{2n-3} - 2^{n-2}}} > \frac{2^{n-2}}{2\sqrt{2^{2n-3}}} = \frac{\sqrt{2}}{4} \ ,$$

the condition (3) is satisfied if $2^{n-2}\sqrt{2} - \sqrt{2}/4 \leq |2^{n-1} - x| \leq 2^{n-2}\sqrt{2}$. Summarizing, we have the following lemma:

Lemma 2 *A parameter $n \geq 2$ is undesirable if $2^{n-2}\sqrt{2} - \sqrt{2}/4 \leq m \leq 2^{n-2}\sqrt{2}$ for some integer m.*

This lemma can be rephrased in terms of the dyadic expansion of $\sqrt{2}$ as follows. Let $\sqrt{2} = (1.b_1b_2b_3\cdots)_2$ be the dyadic expansion of $\sqrt{2}$. For instance, we have $b_1 = 0$, $b_2 = 1$ and $b_3 = 1$ (see (1)). Then the fractional part of the dyadic expansion of $2^{n-2}\sqrt{2}$ is $(0.b_{n-1}b_nb_{n+1}\cdots)_2$, while the dyadic expansion of $\sqrt{2}/4$ is $(0.01b_1b_2b_3\cdots)_2$. By using these expressions, Lemma 2 implies the following:

Lemma 3 *In the above setting, a parameter $n \geq 2$ is undesirable if*

$$(0.b_{n-1}b_nb_{n+1}\cdots)_2 \leq (0.01b_1b_2b_3\cdots)_2 \ . \tag{4}$$

Since $b_1b_2b_3 = 011$, the condition (4) is satisfied if $b_{n-1}b_n = 00$, $b_{n-1}b_nb_{n+1}b_{n+2} = 0100$, or $b_{n-1}b_nb_{n+1}b_{n+2}b_{n+3} = 01010$. Summarizing, we obtain the following sufficient condition for an accuracy parameter n being undesirable:

Proposition 4 *In the above setting, a parameter $n \geq 2$ is undesirable if $b_{n-1}b_n = 00$, $b_{n-1}b_nb_{n+1}b_{n+2} = 0100$, or $b_{n-1}b_nb_{n+1}b_{n+2}b_{n+3} = 01010$.*

Remark 5 In general, the sufficient condition given by Proposition 4 is not necessary for a parameter $n \geq 2$ being undesirable. More precisely, there exists a gap between the sufficient conditions in Proposition 4 and in Lemma 3. For instance, $n = 65$ satisfies the condition (4) but not the condition in Proposition 4. Thus a more precise study of the condition (4) would provide a better result. The author hopes that the condition (4) can motivate further interesting arguments by its self-referential structure.

3. Occurrence Rates of the Three Patterns

Motivated by Proposition 4, in this section we investigate the asymptotic occurrence rate of the three patterns 00, 0100, and 01010 in an infinite binary word. The result will be used to estimate the asymptotic occurrence rate of undesirable parameters.

To formulate the problem, we introduce the following notations. For a finite or infinite binary word $w = w_1w_2w_3\cdots$ ($w_i \in \{0,1\}$), let $\ell(w)$ denote the length of w. Let $P(w)$ denote the set of indices $i \geq 2$ in w such that one of the following three conditions holds:

- $\ell(w) \geq i$ and $w_{i-1}w_i = 00$;
- $\ell(w) \geq i+2$ and $w_{i-1}w_iw_{i+1}w_{i+2} = 0100$;
- $\ell(w) \geq i+3$ and $w_{i-1}w_iw_{i+1}w_{i+2}w_{i+3} = 01010$.

By Proposition 4, a parameter $n \geq 2$ is undesirable if $n \in P(b)$, where $b = b_1b_2b_3\cdots$ is the fractional part of the dyadic expansion of $\sqrt{2}$ as an infinite binary word. Let $w^{(k)}$ denote the initial subword of w of length k. Moreover, let $Z(w)$ denote the set of indices i in w such that $w_i = 0$. Then our main theorem in this section shows relations between the quantities

$$r_{\inf}(w) = \liminf_{n\to\infty} \frac{|Z(w^{(n)})|}{n} \quad \text{and} \quad R_{\inf}(w) = \liminf_{n\to\infty} \frac{|P(w^{(n)})|}{n} \ , \tag{5}$$

and relations between the quantities

$$r_{\sup}(w) = \limsup_{n\to\infty} \frac{|Z(w^{(n)})|}{n} \quad \text{and} \quad R_{\sup}(w) = \limsup_{n\to\infty} \frac{|P(w^{(n)})|}{n} \ . \tag{6}$$

By using the above notations, we state the main theorem as follows:

Theorem 6 *For any infinite binary word $w = w_1w_2w_3\cdots$, let $r_{\inf}(w)$, $r_{\sup}(w)$, $R_{\inf}(w)$, and $R_{\sup}(w)$ be defined in (5) and (6). Then we have*

$$\frac{5r_{\inf}(w)-2}{3} \leq R_{\inf}(w) \leq r_{\inf}(w) \quad \textit{and} \quad \frac{5r_{\sup}(w)-2}{3} \leq R_{\sup}(w) \leq r_{\sup}(w) \ . \tag{7}$$

Note that the lower bounds of $R_{\inf}(w)$ and $R_{\sup}(w)$ become trivial if $r_{\inf}(w) < 2/5$ and $r_{\sup}(w) < 2/5$, respectively. The proof of Theorem 6 will be given in Section 4.

Remark 7 In fact, we can further prove that the lower and upper bounds in Theorem 6 are tight (except the trivial exceptional cases $r_{\inf}(w) < 2/5$ and $r_{\sup}(w) < 2/5$, in which case values of the lower bounds become negative). More precisely, for any real number $2/5 \leq r \leq 1$, there exists an infinite binary word w such that $r_{\inf}(w) = r_{\sup}(w) = r$ and $R_{\inf}(w) = R_{\sup}(w) = (5r-2)/3$, therefore the lower bounds are achieved. Similarly, for any $0 \leq r \leq 1$, there exists an infinite binary word w such that $r_{\inf}(w) = r_{\sup}(w) = r$ and $R_{\inf}(x) = R_{\sup}(x) = r$, therefore the upper bounds are achieved. Details of these results will appear in a forthcoming article of the author.

Regarding the problem in Section 2, by applying Theorem 6 to the above word $w = b$ of the fractional part of the dyadic expansion of $\sqrt{2}$, we have the following theorem:

Theorem 8 *In the above setting, let d_N denote the number of the undesirable parameters $n \leq N$. Then we have*

$$\liminf_{N\to\infty} \frac{d_N}{N} \geq \frac{5r_{\inf}(b)-2}{3} \text{ and } \limsup_{N\to\infty} \frac{d_N}{N} \geq \frac{5r_{\sup}(b)-2}{3} \ . \tag{8}$$

In particular, if $r_{\sup}(b) > 2/5$, then there exist infinitely many undesirable parameters.

As a result, the (lower bound of the) asymptotic occurrence rate of zeroes in the dyadic expansion of $\sqrt{2}$ yields a lower bound of the asymptotic occurrence rate of undesirable parameters. Note that there has been the following long-standing conjecture:

Conjecture 9 *$\sqrt{2}$ is simply normal to base 2; that is, the asymptotic occurrence rate of zeroes in the dyadic expansion of $\sqrt{2}$ is $1/2$ (i.e., $r_{\inf}(b) = r_{\sup}(b) = 1/2$ in the above notations).*

This conjecture reflects our naive intuition that the dyadic expansion of $\sqrt{2}$ looks very random. There have been some further observations that sound positive for the conjecture. For instance, Borel [2] proved that almost every real number (in terms of Lebesgue measure) is simply normal to base 2 (more strongly, is normal to every base $q \geq 2$). By combining Conjecture 9 with Theorem 8, we obtain the following result that is very likely to show a disadvantage of the PRNGs mentioned in Section 1.

Corollary 10 *If Conjecture 9 is true, then the numbers d_N of undesirable parameters $n \leq N$ satisfy $\liminf_{N\to\infty} d_N/N \geq 1/6$.*

4. Proof of Main Theorem

In this section, we give a proof of Theorem 6 in Section 3. First, the upper bounds of $R_{\inf}(w)$ and $R_{\sup}(w)$ in (7) follow simply from the fact that the map $i \mapsto i-1$ is an injection from $P(w)$ to $Z(w)$ for any finite binary word w.

In the rest of this article we prove the lower bounds in Theorem 6, i.e., $R_{\inf}(w) \geq (5r_{\inf}(w)-2)/3$ and $R_{\sup}(w) \geq (5r_{\sup}(w)-2)/3$ for any infinite binary word w. In the proof, we use the following notations. For any (finite or infinite) word $w = w_1w_2w_3\ldots$ and indices $1 \leq i \leq j \leq \ell(w)$, let $w_{[i,j]} = w_iw_{i+1}\cdots w_{j-1}w_j$. Let $\emptyset$ denote the empty word. Let W_N denote the set of binary words of length N. Let $\prec$ denote the lexicographic order on W_n excluding equalities; for instance, we have $1011 \prec 1100$ and $0010 \not\prec 0010$. For two words w and w', we write $w \subset w'$ if $w = w'_{[i,j]}$ for some indices $i \leq j$. Let $w^j = ww\ldots w$ (j repetition of w) for any integer $j \geq 0$.

The outline of our proof is as follows. In the proof, we investigate the maximum value of the number $|Z(u)|$ of zeroes in $u \in W_N$ subject to the condition that $|P(u)|$ is bounded above by a fixed value. This will yield a relation between the quantities $|P(w^{(n)})|$ and $|Z(w^{(n)})|$ for each initial subword $w^{(n)}$ of a given infinite binary word w, from which the desired lower bounds will be derived. For this purpose, we will introduce some "elementary transformations" for the words $u \in W_N$ that preserve $\ell(u)$ and $|Z(u)|$ and do not increase $|P(u)|$. By iterating such elementary transformations, our argument will be reduced to the case of words in W_N of some "normal form" that can be dealt with by case-by-case analysis.

We start the above program. First, we introduce the following seven maps $\varphi_k : W_N \to W_N$, $1 \leq k \leq 7$, as aforementioned elementary transformations, where v and v' signify some (possibly empty) binary words. We define

$$\varphi_1(u) = \begin{cases} 1^pv0, & \text{if } u = v01^p,\ p \geq 1; \\ u, & \text{otherwise.} \end{cases}$$

(namely, φ_1 moves the ones at the tail of the word u, to the front of u; for instance, $\varphi_1(1010\underline{11}) = \underline{11}10100$ and φ_1 fixes 10100),

$$\varphi_2(u) = \begin{cases} 1^{p+1}v11v', & \text{if } u = 1^pv111v',\ p \geq 0,\ 111 \not\subset v \neq \emptyset,\ v_1 = v_{\ell(v)} = 0; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_2 picks up a one from the first block of at least three consecutive ones after a zero and moves it to the front; for instance, $\varphi_2(\underline{1^6}0110\underline{1^4}01^50) = \underline{1^7}0110\underline{1^3}01^50$ and φ_2 fixes $1^30011010$),

$$\varphi_3(u) = \begin{cases} v0110^{p-1}v', & \text{if } u = v0^p11v',\ p \geq 2,\ 0011 \not\subset v,\ v_{\ell(v)} \neq 0; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_3 focuses on the first block of the form 0^p11 with $p \geq 2$, and moves all but one zero in that block to the tail of that block; for instance, $\varphi_3(11011\underline{0^4 11}100110) = 11011\underline{0110^3}100110$ and φ_3 fixes 1011011),

$$\varphi_4(u) = \begin{cases} v01100v', & \text{if } u = v01010v',\ 01010 \not\subset v010; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_4 focuses on the first block of the form 01010, and permutes the third and the fourth bits in that block; for instance, $\varphi_4(11\underline{01010}10) = 11\underline{01100}10$ and φ_4 fixes 011010110101),

$$\varphi_5(u) = \begin{cases} v10^{p+2}v', & \text{if } u = v0^p100v',\ p \geq 1,\ 0100 \not\subset v0^p,\ v_{\ell(v)} \neq 0; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_5 focuses on the first block of the form 0^p100 with $p \geq 1$, and moves the unique one in that block to the front of that block; for instance, $\varphi_5(10011\underline{0^3100}100) = 10011\underline{10^5}100$ and φ_5 fixes 100110010),

$$\varphi_6(u) = \begin{cases} v010110^pv', & \text{if } u = v0^p10110v',\ p \geq 2,\ 0010110 \not\subset v0^p,\ v_{\ell(v)} \neq 0; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_6 focuses on the first block of the form 0^p10110 with $p \geq 2$, and moves all but one zeroes at the beginning of that block to the tail of that block; for instance, $\varphi_6(1\underline{0^4 10110}0101100) = 1\underline{010110^4}0101100$ and φ_6 fixes 1010110), and

$$\varphi_7(u) = \begin{cases} v1010110v', & \text{if } u = v0110110v',\ 0110110 \not\subset v0110; \\ u, & \text{otherwise} \end{cases}$$

(namely, φ_7 focuses on the first block of the form 0110110, and permutes the first and the second bits in that block; for instance, $\varphi_7(1^30\underline{0110110}110) = 1^30\underline{1010110}110$ and φ_7 fixes 0111011010).

Let W_N^{φ} denote the set of all $u \in W_N$ that are fixed by every φ_k, $1 \leq k \leq 7$. Note that each of the seven maps φ_k is well-defined and satisfies that $\ell(\varphi_k(u)) = \ell(u)$ and $|Z(\varphi_k(u))| = |Z(u)|$, since φ_k is just a permutation of bits in a given word. Moreover, it follows immediately from the definition that each φ_k is a weakly increasing map with respect to $\prec$, namely we have $u \preceq \varphi_k(u)$. Since W_N is a finite set, this implies that each $u \in W_N$ can be transformed to a word $\overline{u} \in W_N^{\varphi}$ by finitely many times of applications of the maps φ_k, $1 \leq k \leq 7$. Note that this $\overline{u}$ is not necessarily unique for a given $u \in W_N$ due to various choices of the order of applying the maps φ_k.

To reduce our argument to the case of the words in W_N^{φ}, we would like to show that $|P(\overline{u})| \leq |P(u)|$ for any $u \in W_N$. For this purpose, it suffices to show that $|P(\varphi_k(u))| \leq |P(u)|$ for every map φ_k, $1 \leq k \leq 7$. This is preceded by the following seven lemmas, where we use the notation:

$$P_{i,j}(u) = P(u) \setminus \{i, i+1, \ldots, j-1, j\} \text{ for any indices } i \leq j \text{ in } u \in W_N \ .$$

Before giving the lemmas, note that for any word u and any index i,

$$\text{we have } i \notin P(u) \text{ unless } i \geq 2 \text{ and } u_{i-1} = 0, \tag{9}$$

and therefore $1 \notin P(u)$. Similarly,

$$\text{if } u_i = 1, \text{ then } i \notin P(u) \text{ unless } 2 \leq i \leq \ell(u) - 1 \text{ and } u_{i-1} = u_{i+1} = 0. \tag{10}$$

Moreover, it is obvious that

$$\text{whether } i \in P(u) \text{ or not depends solely on } u_{[i-1,i+3]} \ . \tag{11}$$

Now we show the lemmas as follows, where we write $u' = \varphi_k(u)$ for the map φ_k under consideration:

Lemma 11 *If $u \in W_N$, then $|P(\varphi_1(u))| = |P(u)|$.*

Proof. It suffices to consider the case that $u' \neq u$, namely $u = v01^p$ and $u' = 1^p v0$ with $p \geq 1$, as in the former case of the definition of φ_1. Now if $x = u'_{[i-1,j]}$ is a subword in u' of one of the three forms 00, 0100, or 01010, corresponding to an index $i \in P(u')$, then x should be contained in $v0$ by the shapes of x and u', therefore $u_{[i-1-p,j-p]} = x$ and $i - p \in P(u)$. Similarly, if $x = u_{[i-1,j]}$ is a subword in u of the form 00, 0100, or 01010, corresponding to an $i \in P(u)$, then $x \subset v0$, therefore $u'_{[i-1+p,j+p]} = x$ and $i + p \in P(u')$. Thus $i \mapsto i + p$ is a bijection $P(u) \to P(u')$, therefore Lemma 11 holds. □

Lemma 12 *If $u \in W_N$, then $|P(\varphi_2(u))| = |P(u)|$.*

Proof. It suffices to consider the case that $u = 1^p v111v'$ and $u' = 1^{p+1} v11v'$ as in the former case of the definition of φ_2. Now by the shapes of u and u', any subword in u of the form 00, 0100, or 01010 is contained in either v or v', and the same also holds for u'. This implies that there exists a bijection $P(u) \to P(u')$, hence Lemma 12 holds. □

Lemma 13 *If $u \in W_N$, then $|P(\varphi_3(u))| \leq |P(u)|$.*

Proof. It suffices to consider the case that $u = v0^p11v'$ and $u' = v0110^{p-1}v'$ as in the former case of the definition of φ_3. Put $\ell = \ell(v)$. Then for any subword x in u' of the form 00, 0100, or 01010 corresponding to an $i \in P(u')$, one of the following four conditions is satisfied:

1. x is contained in the block $v0$;
2. $x = 00$ and x is contained in the block 0^{p-1} (thus $\ell + 5 \leq i \leq \ell + p + 2$);
3. $i = \ell + p + 3$, namely i is the first position of the block v';
4. x is contained in the block v'.

In the cases 1 and 4, x is also contained in u and we have $i \in P(u)$. In the case 2, x is also contained in the last $p - 1$ bits of the block 0^p in u, and we have $i - 2 \in P(u)$. Moreover, we have $\ell + 2 \in P(u)$ since $p \geq 2$. Thus there exists an injection $P(u') \to P(u)$ that maps $i \in P(u')$ to i for the cases 1 and 4, to $i - 2$ for the case 2, and to $\ell + 2$ for the case 3. Hence Lemma 13 holds. □

Lemma 14 *If $u \in W_N$, then $|P(\varphi_4(u))| \leq |P(u)|$.*

Proof. It suffices to consider the case that $u = v01010v'$ and $u' = v01100v'$ as in the former case of the definition of φ_4. Put $\ell = \ell(v)$. Then for any subword x in u' of the form 00, 0100, or 01010 corresponding to an $i \in P(u')$, one of the following three conditions is satisfied:

1. x is contained in the block $v0$;
2. $i = \ell + 5$, namely i is the last position of the block 01100;
3. x is contained in the block $0v'$ (thus $\ell + 6 \leq i$).

In the cases 1 and 3, x is also contained in u and we have $i \in P(u)$. Since $\ell + 2 \in P(u)$, there exists an injection $P(u') \to P(u)$ that maps $i \in P(u')$ to i for the cases 1 and 3, and to $\ell + 2$ for the case 2. Hence Lemma 14 holds. □

Lemma 15 *If $u \in W_N$, then $|P(\varphi_5(u))| \leq |P(u)|$.*

Proof. It suffices to consider the case that $u = v0^p100v'$ and $u' = v10^{p+2}v'$ as in the former case of the definition of φ_5. Put $\ell = \ell(v)$. Note that $p \geq 1$ and $v_\ell \neq 0$ by the definition of φ_5. Then for any subword x in u' of the form 00, 0100, or 01010 corresponding to an $i \in P(u')$, one of the following four conditions is satisfied:

1. x is contained in the block v;
2. x is contained in the first p bits of the block 0^{p+2} (thus $\ell + 3 \leq i \leq \ell + p + 1$);

3. $i = \ell + p + 2$, namely i is the second last position of the block 0^{p+2};

4. x is contained in the block $00v'$.

In Cases 1 and 4, x is also contained in u and we have $i \in P(u)$. In Case 2, x is also contained in the block 0^p in u, and we have $i - 1 \in P(u)$. Moreover, we have $\ell + p + 1 \in P(u)$ since $p \geq 1$ (namely $u_{[\ell+p,\ell+p+3]} = 0100$). Thus there exists an injection $P(u') \to P(u)$ that maps $i \in P(u')$ to i for the cases 1 and 4, to $i - 1$ for Case 2, and to $\ell + p + 1$ for the case 3. Hence Lemma 15 holds. □

Lemma 16 *If $u \in W_N$, then $|P(\varphi_6(u))| \leq |P(u)|$.*

Proof. It suffices to consider the case that $u = v0^p10110v'$ and $u' = v010110^pv'$ as in the former case of the definition of φ_6. Put $\ell = \ell(v)$. Note that $p \geq 2$ and $v_\ell \neq 0$ by the definition of φ_6. Then for any subword x in u' of the form 00, 0100, or 01010 corresponding to an $i \in P(u')$, one of the following four conditions is satisfied:

1. x is contained in the block $v0$ (thus $i \leq \ell - 1$ since $v_\ell \neq 0$);

2. $\ell \geq 2$, $v_{[\ell-1,\ell]} = 01$, $x = 01010$ and $i = \ell$;

3. x is contained in the block 0^p (thus $\ell + 7 \leq i \leq \ell + p + 5$);

4. x is contained in the block $0v'$ (thus $\ell + p + 6 \leq i$).

In the cases 1 and 4, x is also contained in u and we have $i \in P(u)$. In the case 3, x is also contained in the block 0^p in u, and we have $i - 5 \in P(u)$. Moreover, in the case 2, we have $\ell \in P(u)$ since $p \geq 2$ (namely $u_{[\ell-1,\ell+2]} = 0100$). Thus there exists an injection $P(u') \to P(u)$ that maps $i \in P(u')$ to i for the cases 1 and 4, to $i - 5$ for the case 3, and to ℓ for the case 2. Hence Lemma 16 holds. □

Lemma 17 *If $u \in W_N$, then $|P(\varphi_7(u))| \leq |P(u)|$.*

Proof. It suffices to consider the case that $u = v0110110v'$ and $u' = v1010110v'$ as in the former case of the definition of φ_7. Put $\ell = \ell(v)$. Then for any subword x in u' of the form 00, 0100, or 01010 corresponding to an $i \in P(u')$, one of the following four conditions is satisfied:

1. x is contained in the block v;

2. $\ell \geq 1$, $v_\ell = 0$, $x = 01010$ and $i = \ell + 1$;

3. x is contained in the block $0v'$ (thus $\ell + 8 \leq i$).

type	subword	type	subword	type	subword
1	$0v1)$	2	$0v111$	3	01010
4	0100	5	0110110	6	$001v)$ $(v \neq 0)$
6a	0011	6b	$0010v)$ $(v \neq \emptyset)$	6c	0010110

Here v is a (possibly empty) word, and ')' denotes the tail of the word $\overline{u}$.

Table 1: Excluded subwords for words in W_N^{φ}

In the cases 1 and 3, x is also contained in u and we have $i \in P(u)$. Moreover, in the case 2, we have $\ell + 1 \in P(u)$ (namely $u_{[\ell,\ell+1]} = 00$). Thus there exists an injection $P(u') \to P(u)$ that maps $i \in P(u')$ to i for the cases 1 and 3, and to $\ell + 1$ for the case 2. Hence Lemma 17 holds. □

Thus we have proven that $|P(\overline{u})| \leq |P(u)|$ for any $u \in W_N$ as desired. From now, we determine the possibilities of the shape of $\overline{u} \in W_N^{\varphi}$. For the purpose, first we show that any word in W_N^{φ} does not contain a subword of type 1–6 in Table 1. For instance, if $\overline{u} \in W_N^{\varphi}$, then we have $010111 \not\subset \overline{u}$ since 010111 is a word of type 2. We also include three other auxiliary subwords (that are special cases of subwords of type 6) in Table 1 since these are used in the proof of the above fact (Lemma 18 below).

Lemma 18 *Any $\overline{u} \in W_N^{\varphi}$ does not contain a subword of type 1–6 listed in Table 1.*

Proof. First, it is straightforward to show that a subword of type 1, 2, 3, 4, and 5 in Table 1 is not contained in $\overline{u}$ since $\overline{u}$ is fixed by a map φ_1, φ_2, φ_4, φ_5, and φ_7, respectively. From here, we show that $\overline{u}$ does not contain a subword u' of type 6. We prove this by classifying the subword v appearing in the definition of u'.

First, if v is an empty word, then u' is of type 1 in Table 1, and therefore $\overline{u}$ does not contain u' by the previous paragraph. If v begins with a one, then u' is of type 6a in Table 1, and this u' is not contained in $\overline{u}$ since $\overline{u}$ is fixed by φ_3. Thus it suffices to consider the case that v begins with a zero, and hence u' is of type 6b in Table 1 (where we reuse the notation v for simplicity).

For the u' of type 6b in Table 1, if v begins with a zero, then u' contains a subword of type 4 in Table 1, and therefore $\overline{u}$ does not contain u' by the first paragraph. Thus it suffices to consider the case that u' is of the form $00101v')$ where v' is an arbitrary finite word.

If v' is empty or begins with a zero, then u' contains a subword of type 1 or 3, respectively, and hence $\overline{u}$ does not contain u' by the first paragraph. Thus it suffices to consider the case that u' is of the form $001011v'')$ where v'' is an arbitrary finite word.

If v'' is empty or begins with a one, then u' contains a subword of type 1 or 2, respectively, and hence $\overline{u}$ does not contain u' by the first paragraph. Finally, if v'' begins with a zero, then u' contains a subword of type 6c, and this u' is not contained in $\overline{u}$ since $\overline{u}$ is fixed by φ_6. Hence the claim holds in any case, concluding the proof of the lemma. □

Owing to Lemma 18, we obtain the following classification of the words $\overline{u}$ in W_N^{φ} summarized in Table 4, where for each word $u = \overline{u}$, descriptions of values N, $|Z(u)|$ and $|P(u)|$ and relations for these values and parameters p, q (except for Types 6 and 7) and s (except for Type 1) are also included:

Lemma 19 *Any word $\overline{u}$ in W_N^{φ} is of one of the seven types in Table 4.*

Proof. First, note that any $\overline{u} \in W_N^{\varphi}$ can be expressed in the following form:

$$\overline{u} = 1^{p_0}0^{q_1}1^{p_1}\cdots 0^{q_k}1^{p_k}0^{q_{k+1}},$$
$$k \geq 0,\ p_0 \geq 0,\ q_{k+1} \geq 0,\ p_i \geq 1,\ q_i \geq 1\ (1 \leq i \leq k).$$

We apply Lemma 18 to this $\overline{u}$. First, the absence of a subword of type 1 in Table 1 implies that $\overline{u}$ does not end with a one unless $\overline{u}$ contains no zeroes. Thus we have $q_{k+1} \geq 1$ if $k \geq 1$. Secondly, the absence of a subword of type 2 in Table 1 implies that three consecutive ones do not appear after a zero, therefore we have $p_i \in \{1, 2\}$ for every $1 \leq i \leq k$. Moreover, the absence of a subword of type 6 in Table 1 implies that if $001 \subset \overline{u}$, then a zero should follow that subword 001 immediately and $\overline{u}$ should end with that zero. By these conditions, the possible shapes of $\overline{u}$ are classified as follows:

1. $\overline{u} = 1^{p_0}0^{q_1}$ (corresponding to the case $k = 0$);
2. $\overline{u} = 1^{p_0}01^{p_1}0^{q_2}$, $p_1 \in \{1, 2\}$, $q_2 \geq 1$ (corresponding to the case $k = 1$, $q_k = 1$);
3. $\overline{u} = 1^{p_0}0^{q_1}10$, $q_1 \geq 2$ (corresponding to the case $k = 1$, $q_k \geq 2$);
4. $\overline{u} = 1^{p_0}01^{p_1}\cdots 01^{p_{k-1}}01^{p_k}0^{q_{k+1}}$, $p_i \in \{1, 2\}$ $(1 \leq i \leq k)$, $q_{k+1} \geq 1$ (corresponding to the case $k \geq 2$, $q_k = 1$);
5. $\overline{u} = 1^{p_0}01^{p_1}\cdots 01^{p_{k-1}}0^{q_k}10$, $p_i \in \{1, 2\}$ $(1 \leq i \leq k)$, $q_k \geq 2$ (corresponding to the case $k \geq 2$, $q_k \geq 2$).

Case 1 corresponds to Type 1 in Table 4. In Case 2, a choice $p_1 = 1$ implies that $q_2 = 1$ by the absence of a subword 0100 of type 4 in Table 1, and it corresponds to Type 6 in Table 4 with parameter $s = 0$. On the other hand, the other choice

Type 1	$u = 1^p 0^q \quad (p \geq 0,\, q \geq 0)$
	$N = p + q$
	$\lvert Z(u)\rvert = q \qquad \lvert P(u)\rvert = q - 1$
	$\lvert Z(u)\rvert/N = \lvert P(u)\rvert/N + 1/N$
Type 2	$u = 1^p(01011)^s 0^q 10 \quad (p \geq 0,\, q \geq 2,\, s \geq 0)$
	$N = 5s + p + q + 2$
	$\lvert Z(u)\rvert = 2s + q + 1 \qquad \lvert P(u)\rvert = q - 1$
	$\lvert Z(u)\rvert/N = 3/5 \cdot \lvert P(u)\rvert/N + 2/5 + 4/(5N) - 2p/(5N)$
Type 3	$u = 1^p 011(01011)^s 0^q 10 \quad (p \geq 0,\, q \geq 2,\, s \geq 0)$
	$N = 5s + p + q + 5$
	$\lvert Z(u)\rvert = 2s + q + 2 \qquad \lvert P(u)\rvert = q - 1$
	$\lvert Z(u)\rvert/N = 3/5 \cdot \lvert P(u)\rvert/N + 2/5 + 3/(5N) - 2p/(5N)$
Type 4	$u = 1^p(01011)^s 0^q \quad (p \geq 0,\, q \geq 1,\, s \geq 1)$
	$N = 5s + p + q$
	$\lvert Z(u)\rvert = 2s + q \qquad \lvert P(u)\rvert = q - 1$
	$\lvert Z(u)\rvert/N = 3/5 \cdot \lvert P(u)\rvert/N + 2/5 + 3/(5N) - 2p/(5N)$
Type 5	$u = 1^p 011(01011)^s 0^q \quad (p \geq 0,\, q \geq 1,\, s \geq 0)$
	$N = 5s + p + q + 3$
	$\lvert Z(u)\rvert = 2s + q + 1 \qquad \lvert P(u)\rvert = q - 1$
	$\lvert Z(u)\rvert/N = 3/5 \cdot \lvert P(u)\rvert/N + 2/5 + 2/(5N) - 2p/(5N)$
Type 6	$u = 1^p(01011)^s 010 \quad (p \geq 0,\, s \geq 0)$
	$N = 5s + p + 3$
	$\lvert Z(u)\rvert = 2s + 2 \qquad \lvert P(u)\rvert = 0$
	$\lvert Z(u)\rvert/N = 2/5 + 4/(5N) - 2p/(5N)$
Type 7	$u = 1^p 011(01011)^s 010 \quad (p \geq 0,\, s \geq 0)$
	$N = 5s + p + 6$
	$\lvert Z(u)\rvert = 2s + 3 \qquad \lvert P(u)\rvert = 0$
	$\lvert Z(u)\rvert/N = 2/5 + 3/(5N) - 2p/(5N)$

Table 2: Classification of words u in W_N^{φ}

$p_1 = 2$ corresponds to Type 5 in Table 4 with parameter $s = 0$. Case 3 corresponds to Type 2 in Table 4 with parameter $s = 0$.

The remaining part of the proof focuses on Cases 4 and 5. In Case 4, the absence of a subword 01010 of type 3 in Table 1 and a subword 0110110 of type 5 in Table 1 implies that $(p_i, p_{i+1}) = (1, 2)$ or $(2, 1)$ for each $1 \le i \le k-1$. Thus the sequence $(p_1, p_2, \ldots, p_k)$ is of one of the four forms $(1, 2, 1, 2, \ldots, 1, 2)$, $(1, 2, 1, 2, \ldots, 2, 1)$, $(2, 1, 2, 1, \ldots, 2, 1)$, and $(2, 1, 2, 1, \ldots, 1, 2)$. The first and the fourth cases correspond to Type 4 and Type 5 in Table 4, respectively. On the other hand, the second and the third cases correspond to Type 6 and Type 7 in Table 4, respectively, since now we have $p_k = 1$ and the absence of a subword 0100 of type 4 in Table 1 implies that $q_{k+1} = 1$.

Finally, in Case 5, the fact that $q_k \ge 2$ and the absence of a subword 0100 of type 4 in Table 1 imply that $p_{k-1} = 2$. Now by the same argument as the previous paragraph, the sequence $(p_1, p_2, \ldots, p_{k-1})$ is either $(1, 2, 1, 2, \ldots, 1, 2)$, or $(2, 1, 2, 1, \ldots, 1, 2)$. These cases correspond to Type 2 and Type 3 in Table 4, respectively. Hence Lemma 19 holds. □

Now, we present the main part of the proof of lower bounds in Theorem 6. The key fact in the argument is the following:

Lemma 20 *For each initial subword $w^{(n)} \in W_n$ of any infinite binary word w, we have*

$$\frac{|P(w^{(n)})|}{n} \ge \frac{5}{3} \cdot \frac{|Z(w^{(n)})|}{n} - \frac{2}{3} - \frac{4}{3n} \ . \tag{12}$$

Proof. For each $w^{(n)}$, we associate a (not necessarily unique) word $y(n) = \overline{w^{(n)}}$ in W_n^{φ} to $w^{(n)}$ by applying the maps φ_i, $1 \le i \le 7$, repeatedly. Note that $|Z(y(n))| = |Z(w^{(n)})|$ and $|P(y(n))| \le |P(w^{(n)})|$ by Lemmas 11–17. If $y(n)$ is of type 1 in Table 4, then we have

$$\frac{|P(w^{(n)})|}{n} \ge \frac{|P(y(n))|}{n} = \frac{|Z(y(n))|}{n} - \frac{1}{n} = \frac{|Z(w^{(n)})|}{n} - \frac{1}{n} \ . \tag{13}$$

Now we have $|Z(w^{(n)})|/n \le 1$ by the definition, therefore the right-hand side of (13) is larger than the right-hand side of (12). On the other hand, if $y(n)$ is of types 2–5 in Table 4, then we have

$$\begin{aligned} \frac{|P(w^{(n)})|}{n} \ge \frac{|P(y(n))|}{n} &= \frac{5}{3} \cdot \frac{|Z(y(n))|}{n} - \frac{2}{3} - \frac{c}{3n} + \frac{2p}{3n} \\ &\ge \frac{5}{3} \cdot \frac{|Z(w^{(n)})|}{n} - \frac{2}{3} - \frac{4}{3n} \ , \end{aligned}$$

where $c = 4$, 3, 3, and 2 in the case of types 2, 3, 4, and 5, respectively. Moreover, if $y(n)$ is of types 6–7 in Table 4, then a direct calculation shows that the right-hand side of (12) is not positive (note that $|Z(w^{(n)})| = |Z(y(n))|$), therefore (12) obviously holds. □

Now the desired bounds $R_{\inf}(w) \geq (5r_{\inf}(w) - 2)/3$ and $R_{\sup}(w) \geq (5r_{\sup}(w) - 2)/3$ are derived by taking the $\liminf_{n\to\infty}$ and $\limsup_{n\to\infty}$ of both sides of (12), respectively. Hence the proof of lower bounds in Theorem 6 is concluded.

Acknowledgments. The author would like to thank Dr. Yoshio Okamoto, Dr. Kenji Kashiwabara, and Professor Masahiro Hachimori, for their significant comments. The author would also like to thank the anonymous referee for the invaluable comments. An extended abstract of this work was presented in 21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2009) [6].

References

[1] S. Araki, T. Miyazaki, and S. Uehara, Analysis for pseudorandom number generators using logistic map, in: *Proceedings of the 2006 International Symposium on Information Theory and its Applications (ISITA 2006)* (CD-ROM), Seoul, Korea, 2006.

[2] E. Borel, Les probabilités dénombrables et leurs applications arithmetiqués, *Rend. Circ. Mat. Palermo*, **27** (1909) 247–271.

[3] D.E. Knuth, *The Art of Computer Programming*, Volume 2, third edition, Addison-Wesley Professional, 1997.

[4] T. Miyazaki, S. Araki, and S. Uehara, Some properties of logistic map on integral domains (in Japanese), in: *Proceedings of the 2007 Symposium on Cryptography and Information Security (SCIS 2007)* (CD-ROM), Sasebo, Japan, 2007.

[5] T. Miyazaki, S. Araki, and S. Uehara, Period and link length of the logistic map over integers, in: *Proceedings of the 2008 International Symposium on Information Theory and its Applications (ISITA 2008)* (CD-ROM), Auckland, New Zealand, 2008.

[6] K. Nuida, Bounds of asymptotic occurrence rates of some patterns in binary words related to integer-valued logistic maps, in: *Proceedings of 21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2009)*, DMTCS Proceedings, http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAK0159

[7] S.C. Phatak, and S.S. Rao, Logistic map: A possible random number generator, *Physical Review E*, **51** (1995) 3670–3678.

[8] N.R. Wagner, The logistic lattice in random number generation, in: *Proceedings of the 30th Annual Allerton Conference on Communications, Control, and Computing*, Illinois, USA, 1993, pp.922–931.