Bounds of asymptotic occurrence rates of some patterns in binary words related to integer-valued logistic maps

Koji Nuida

FPSAC 2009, Hagenberg, Austria DMTCS proc. **AK**, 2009, 709–720

Bounds of asymptotic occurrence rates of some patterns in binary words related to integer-valued logistic maps

Koji Nuida[1]

[1]*Research Center for Information Security (RCIS), National Institute of Advanced Industrial Science and Technology (AIST), Akihabara-Daibiru Room 1003, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan*

Abstract. In this article, we investigate the asymptotic occurrence rates of specific subwords in any infinite binary word. We prove that the asymptotic occurrence rate for the subwords is upper- and lower-bounded in the same way for every infinite binary word, in terms of the asymptotic occurrence rate of the zeros. We also show that both of the bounds are best-possible by constructing, for each bound, a concrete infinite binary word such that the bound is reached. Moreover, we apply the result to analyses of recently-proposed pseudorandom number generators that are based on integer-valued variants of logistic maps.

Résumé. Dans cet article, nous étudions les fréquences asymptotiques d'occurence de suites spécifiques dans tout mot binaire infini. Nous prouvons que la fréquence asymptotique d'occurence pour ces suites est borné supérieurement et inférieurement de la même façon pour chaque mot binaire infini, en termes des fréquences asymptotiques d'occurence de zéros. Nous montrons aussi que les deux limites sont les meilleures possibles en construisant concrètement, pour chaque limite, un mot binaire infini tel que la borne est atteinte à la limite. De plus, nous appliquons ce résultat à des analyses de générateurs de nombres pseudo-aléatoires proposés récemment qui sont basés sur des variantes des fonctions logistiques à valeurs entières.

Keywords: binary word, pattern occurrence rate, simple normality, logistic map, pseudorandom number generator

1 Introduction

In this article, we study the following very concrete problem: For infinite binary words, find relations between the sum of the asymptotic occurrence rates of three patterns 00, 0100, and 01010, and the asymptotic occurrence rates of 0s. (A motivation of this problem is explained in the next two paragraphs.) More precisely, for a finite or infinite binary word $x = x_1x_2x_3\cdots$ with $x_i \in \{0,1\}$ for every i, let $I(x)$ denote the set of indices i in x such that $x_i = 0$, and let $P(x)$ be the set of indices i in x that satisfy one of the following three conditions; (P1) $x_{i-1}x_i = 00$; (P2) $x_{i-1}x_ix_{i+1}x_{i+2} = 0100$; and (P3) $x_{i-1}x_ix_{i+1}x_{i+2}x_{i+3} = 01010$. For example, if $x = 0110101000$, that is the first 10 bits of the dyadic expansion of the fractional part $\mathrm{frac}(\sqrt{2})$ of $\sqrt{2}$ (*i.e.* $\sqrt{2}-1$), then we have $I(x) = \{1,4,6,8,9,10\}$, and $i = 9,7,5$ are examples of indices in $P(x)$ satisfying the conditions (P1), (P2), and (P3), respectively.

Let $x^{(n)}$ denote the initial subword of x of length n. In the above setting, the problem is of finding, for any infinite binary word x, relations between the quantities

$$r_{\inf}(x) = \liminf_{n\to\infty} \frac{|I(x^{(n)})|}{n} \quad \text{and} \quad R_{\inf}(x) = \liminf_{n\to\infty} \frac{|P(x^{(n)})|}{n} \ , \tag{1}$$

and those between the quantities

$$r_{\sup}(x) = \limsup_{n\to\infty} \frac{|I(x^{(n)})|}{n} \quad \text{and} \quad R_{\sup}(x) = \limsup_{n\to\infty} \frac{|P(x^{(n)})|}{n} \ . \tag{2}$$

In Section 2, we present simple upper and lower bounds of the quantities $R_{\inf}(x)$ and $R_{\sup}(x)$ in terms of $r_{\inf}(x)$ and $r_{\sup}(x)$, respectively. Moreover, we prove that these bounds are both "best possible". More precisely, for each of the lower and the upper bounds, we construct a concrete example of an infinite binary word that attains the equality in the bound. The first aim of this article is to describe the above combinatorial problem and its solution.

The problem presented in the previous paragraph, especially the specific choice of the three patterns, is motivated by analyses of pseudorandom number generators (PRNGs). To imitate random or chaotic behaviors of nature by using deterministic algorithms performed on computers is a ubiquitous and very fundamental task in several areas of science and information technology, such as computer simulation, statistics and cryptography; hence construction and analyses of PRNGs are one of the most active topics in information theory. One of the existing ideas to construct good PRNGs is to make use of the well-known chaotic behavior of the logistic maps $L(x) = \mu x(1-x)$, $0 < x < 1$, for some parameters μ (*e.g.* Wagner (1993); Phatak et al. (1995)). For example, the map $L(x)$ shows chaotic behavior by choosing a parameter $\mu = 4$. The PRNGs concerned in this article is the recently proposed schemes (see *e.g.* Araki et al. (2006)) that are based on some integer-valued variants of the map $L(x)$ with parameter $\mu = 4$. The corresponding integer-valued variant $L_n(x)$ with accuracy parameter $n \in \mathbb{Z}$, $n > 0$, is given by

$$L_n(x) = \left\lfloor \frac{4x(2^n - x)}{2^n} \right\rfloor = \left\lfloor \frac{x(2^n - x)}{2^{n-2}} \right\rfloor \text{ for } x \in X_n = \{1, 2, \ldots, 2^n - 1\} \tag{3}$$

(e.g. Kuribayashi et al. (2005)), where $\lfloor z \rfloor$ denotes the integer part of $z \in \mathbb{R}$. The description of $L_n(x)$ is derived by first expanding the domain $(0,1) \subset \mathbb{R}$ of the original logistic map $L(x)$ proportionally to a larger interval $(0, 2^n)$ and then cutting off the fractional parts of real numbers in the latter interval. Now the PRNG mentioned above, that uses the map $L_n(x)$ as the updating function of internal states, is informally described as follows:

1. Choose an initial state $s = s_0 \in X_n$.
2. Update the internal state s by applying the map $L_n(x)$ K times (with K a parameter).
3. Output bits of the binary expression of s in some suitable positions.
4. Repeat Steps 2 and 3.

In some preceding works, appropriate choices of parameters for the scheme have been investigated (*e.g.* Araki et al. (2008)).

In particular, it has been pointed out (Miyazaki et al. (2007)) that, when we start from the initial value $s_0 = 2^{n-1}$, the internal state will be pushed out the domain X_n of the map $L_n(x)$ (*i.e.* $L_n(2^{n-1}) = 2^n$), thus the value 2^{n-1} should be excluded from the candidates of the initial value. (Note that, even if we tolerate the illegal input 2^n for $L_n(x)$, we then have $L_n(2^n) = 0$ and $L_n(0) = 0$, therefore the internal state s falls eventually into a stable value. This is also a very undesirable situation since it makes the outputs of the PRNG not random at all.) Moreover, it was also pointed out that even an initial value $s_0 \in X_n$ other than 2^{n-1} may lead the internal state to the excluded value 2^{n-1}, *i.e.* we may have $L_n(s_0) = 2^{n-1}$. Thus such an undesirable initial value should also be avoided in practical use of the PRNG. However, existence conditions for such an undesirable initial value $s_0 \neq 2^{n-1}$ have not been well investigated. The problem presented in the first paragraph arises from the author's recent research (Nuida (2008)) on conditions for the accuracy parameter n such that the corresponding $L_n(x)$ possesses an undesirable initial value $s_0 \neq 2^{n-1}$ (we call such a parameter n *dangerous*). More precisely, it is shown that lower bounds of the quantities $R_{\inf}(x)$ and $R_{\sup}(x)$ for x being the fractional part $\mathrm{frac}(\sqrt{2})_2$ of the dyadic expansion of $\sqrt{2}$ give lower bounds of the asymptotic rate of the dangerous parameters n in the positive integers (in terms of the values $r_{\inf}(x)$ and $r_{\sup}(x)$ for the same x). As a consequence, our analysis of the above PRNG is also deeply related to a long-standing conjecture on the quantities $r_{\inf}(\mathrm{frac}(\sqrt{2})_2)$ and $r_{\sup}(\mathrm{frac}(\sqrt{2})_2)$. See Section 3 for details.

This article is organized as follows. In Section 2, we present a solution of the problem described in the first paragraph. The solution itself is stated as Theorem 1 that is the main result of this article. The proof of Theorem 1 is divided into the following four parts; the lower bound, its best-possibility, the upper bound, and its best-possibility. Due to the limited pages, some lemmas in Section 2 are only accompanied with a sketchy proof; for the details of the proof, see a forthcoming full version of this article. In Section 3, we describe a relation between the above problem and analyses of PRNGs of the above types. Namely, we explain how a lower bound of the asymptotic occurrence rate of the dangerous parameters in the positive integers is derived from the result on the first problem. Finally, in Section 4, we propose open problems on possible improvements or generalizations of our results in this article.

2 Results

This section shows a solution of the problem presented in the first paragraph of the Introduction. The solution, that is the main theorem of this article, is the following:

Theorem 1 *For any infinite binary word $x = x_1x_2x_3\cdots$, let $r_{\inf}(x)$, $r_{\sup}(x)$, $R_{\inf}(x)$ and $R_{\sup}(x)$ be defined as in (1) and (2). Then we have*

$$\frac{5r_{\inf}(x) - 2}{3} \leq R_{\inf}(x) \leq r_{\inf}(x) \quad \textit{and} \quad \frac{5r_{\sup}(x) - 2}{3} \leq R_{\sup}(x) \leq r_{\sup}(x) \ . \tag{4}$$

Moreover, all the bounds are best possible except trivial exceptions, in the following sense: For any $2/5 \leq r \leq 1$, there exists an infinite binary word x such that $r_{\inf}(x) = r_{\sup}(x) = r$ and $R_{\inf}(x) = R_{\sup}(x) = (5r - 2)/3$. Similarly, for any $0 \leq r \leq 1$, there exists an infinite binary word x such that $r_{\inf}(x) = r_{\sup}(x) = r$ and $R_{\inf}(x) = R_{\sup}(x) = r$.

In what follows, we give a sketch of a proof of the theorem.

2.1 Lower Bounds

We demonstrate a sketchy proof of the lower bounds in Theorem 1. For any positive integer n, let W_n denote the set of binary words of length n. Let $\prec$ denote the lexicographic order on W_n (*e.g.* $1100 \succ 1011$). Let $\ell(x)$ denote the length of a word x. Let $\emptyset$ denote the empty word. For two words y and y', we write $y \subset y'$ if y appears in y' as a consecutive subword, and let $y^j = yy\cdots y$ (j repetition) for any $j \geq 0$. Then we define the following maps $\varphi_1, \varphi_2, \ldots, \varphi_7$ from W_n to itself, where w and w' signify some binary words:

$$\begin{aligned}
\varphi_1(x) &= \begin{cases} 1^p w 0 & \text{, if } x = w01^p,\ p \geq 1\ ; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_2(x) &= \begin{cases} 1^{p+1} w 11 w' & \text{, if } x = 1^p w 111 w',\ p \geq 0,\ 111 \not\subset w \neq \emptyset,\ w_1 = w_{\ell(w)} = 0\ ; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_3(x) &= \begin{cases} w 0110^{p-1} w' & \text{, if } x = w0^p 11 w',\ p \geq 2,\ 0011 \not\subset w,\ w_{\ell(w)} \neq 0\ ; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_4(x) &= \begin{cases} w 01100 w' & \text{, if } x = w01010w',\ 01010 \not\subset w010\ ; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_5(x) &= \begin{cases} w 10^{p+2} w' & \text{, if } x = w0^p 100 w',\ p \geq 1,\ 0100 \not\subset w0^p,\ w_{\ell(w)} \neq 0\ ; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_6(x) &= \begin{cases} w 010110^p w' & \text{, if } x = w0^p 10110 w',\ p \geq 2,\ 0010110 \not\subset w0^p,\ w_{\ell(w)} \neq 0; \\ x & \text{, otherwise,} \end{cases} \\
\varphi_7(x) &= \begin{cases} w 1010110 w' & \text{, if } x = w0110110w',\ 0110110 \not\subset w0110\ ; \\ x & \text{, otherwise.} \end{cases}
\end{aligned} \tag{5}$$

Note that these seven maps are all well-defined; namely, each φ_k transforms the leftmost consecutive subword of the specified form, and leaves the word unchanged if such a subword does not exist. Moreover, each φ_k leaves the quantity $|I(x)|$ invariant and is weakly increasing with respect to $\prec$ (*i.e.* $x \preceq \varphi_k(x)$). Then a case-by-case argument shows the following property:

Lemma 1 *For each φ_k, we have $|P(\varphi_k(x))| \leq |P(x)|$ for any $x \in W_n$.*

Let W_n^{φ} be the set of the common fixed points of $\varphi_1, \ldots, \varphi_7$ in W_n. Then, since each φ_k is weakly increasing with respect to the total order $\prec$, it follows that any word in W_n can be mapped into W_n^{φ} by applying $\varphi_1, \ldots, \varphi_7$ finitely many times. Let $\varphi(x)$ denote a (not necessarily unique) word in W_n^{φ} corresponding to $x \in W_n$. Note that $|I(\varphi(x))| = |I(x)|$ and $|P(\varphi(x))| \leq |P(x)|$ for any $x \in W_n$. Moreover, we have the following property of the fixed point set W_n^{φ}:

Lemma 2 *Any $x \in W_n^{\varphi}$ involves no consecutive subword listed in Table 1.*

Proof (Sketch): The excluded subword of Type k, $1 \leq k \leq 7$, is straightforwardly derived by the condition that $\varphi_k(x) = x$. On the other hand, each excluded subword of Type k, $k \geq 8$, is deduced from the preceding ones; for example, if $x \in W_n^{\varphi}$ contains a consecutive subword 001011 of Type 8,

Tab. 1: Excluded consecutive subwords for words in W_n^{φ}
Here w is some (possibly empty) word, and the symbol ")" in Type 1 denotes the end of the word.

type	subword	type	subword	type	subword
1	$0w1)$	2	$0w111$	3	0011
4	01010	5	0100	6	0010110
7	0110110	8	001011	9	00101
10	$0010w$ $(w \neq \emptyset)$	11	$001w$ $(w \neq 0)$		

then x must contain one of the words 0010110 (Type 6), 0010111 (Type 2) and $001011)$ (Type 1) as a consecutive subword. □

Now note that any $x \in W_n$ admits an expression of the following form:

$$x = 1^{p_0}0^{q_1}1^{p_1}\cdots 0^{q_k}1^{p_k}0^{q_{k+1}} \ , \ k \geq 0,\ q_i \geq 1,\ p_i \geq 1\ (1 \leq i \leq k)\ . \tag{6}$$

Then the next lemma follows from Lemma 2:

Lemma 3 *If $x \in W_n^{\varphi}$, then the expression (6) of x satisfies the following conditions:*

1. *If $k \geq 1$, then $q_{k+1} \geq 1$.*
2. *$p_i \leq 2$ for $i \geq 1$.*
3. *$q_i = 1$ for $1 \leq i \leq k-1$.*
4. *If $q_k \geq 2$, then $p_k = 1$ and $q_{k+1} = 1$.*
5. *If $k \geq 2$ and $q_k \geq 2$, then $p_{k-1} = 2$.*
6. *$(p_i, p_{i+1}) = (1,2)$ or $(2,1)$ for $1 \leq i \leq k-2$.*
7. *If $q_k = 1$, then $p_k = 2$ or $q_{k+1} = 1$.*

Proof (Sketch): The first and the second parts follow from the absence of subwords of Types 1 and 2, respectively. The third and the fourth parts both follow from the absence of subwords of Type 11. The fifth part follows from the second part and the absence of subwords of Type 5. The sixth part follows from the second part and the absence of subwords of Types 4 and 7. Finally, the seventh part follows from the first and the second parts, and the absence of subwords of Type 5. □

By Lemma 3, we obtain a complete list of words $x \in W_n^{\varphi}$ as in Table 2. In the table, Type 1 corresponds to the case that $k = 0$ in (6). Types 2 and 3 both correspond to the case that $k \geq 1$ and $q_k \geq 2$. Types 4 and 5 both correspond to the case that $k \geq 1$, $q_k = 1$ and $p_k = 2$. Finally, types 6 and 7 both correspond to the case that $k \geq 1$, $q_k = 1$ and $p_k = 1$. Table 2 also includes the values of n, $|I(x)|$, and $|P(x)|$, and the relations between $|I(x)|/n$ and $|P(x)|/n$ that play a key role in the proof of Theorem 1.

From now, we prove the lower bound of $R_{\inf}(x)$. This is trivial if $r_{\inf}(x) \leq 2/5$, therefore we assume that $r_{\inf}(x) > 2/5$. For each n, put $m_n = |P(x^{(n)})|$ and $y_n = \varphi(x^{(n)})$. Recall that $|I(y_n)| = |I(x^{(n)})|$ and $|P(y_n)| \leq |P(x^{(n)})|$. Now we have the following:

Tab. 2: Complete list of words in W_n^{φ}

Type 1	$x = 1^p 0^q \quad (p \geq 0, q \geq 0)$
	$n = p + q$
	$\lvert I(x)\rvert = q$
	$\lvert P(x)\rvert = q - 1$
	$\lvert I(x)\rvert/n = \lvert P(x)\rvert/n + 1/n$
Type 2	$x = 1^p (01011)^s 0^q 10 \quad (p \geq 0, q \geq 2, s \geq 0)$
	$n = 5s + p + q + 2$
	$\lvert I(x)\rvert = 2s + q + 1$
	$\lvert P(x)\rvert = q - 1$
	$\lvert I(x)\rvert/n = 3\lvert P(x)\rvert/(5n) + 2/5 + 4/(5n) - 2p/(5n)$
Type 3	$x = 1^p 011 (01011)^s 0^q 10 \quad (p \geq 0, q \geq 2, s \geq 0)$
	$n = 5s + p + q + 5$
	$\lvert I(x)\rvert = 2s + q + 2$
	$\lvert P(x)\rvert = q - 1$
	$\lvert I(x)\rvert/n = 3\lvert P(x)\rvert/(5n) + 2/5 + 3/(5n) - 2p/(5n)$
Type 4	$x = 1^p (01011)^s 0^q \quad (p \geq 0, q \geq 1, s \geq 1)$
	$n = 5s + p + q$
	$\lvert I(x)\rvert = 2s + q$
	$\lvert P(x)\rvert = q - 1$
	$\lvert I(x)\rvert/n = 3\lvert P(x)\rvert/(5n) + 2/5 + 3/(5n) - 2p/(5n)$
Type 5	$x = 1^p 011 (01011)^s 0^q \quad (p \geq 0, q \geq 1, s \geq 0)$
	$n = 5s + p + q + 3$
	$\lvert I(x)\rvert = 2s + q + 1$
	$\lvert P(x)\rvert = q - 1$
	$\lvert I(x)\rvert/n = 3\lvert P(x)\rvert/(5n) + 2/5 + 2/(5n) - 2p/(5n)$
Type 6	$x = 1^p (01011)^s 010 \quad (p \geq 0, s \geq 0)$
	$n = 5s + p + 3$
	$\lvert I(x)\rvert = 2s + 2$
	$\lvert P(x)\rvert = 0$
	$\lvert I(x)\rvert/n = 2/5 + 4/(5n) - 2p/(5n)$
Type 7	$x = 1^p 011 (01011)^s 010 \quad (p \geq 0, s \geq 0)$
	$n = 5s + p + 6$
	$\lvert I(x)\rvert = 2s + 3$
	$\lvert P(x)\rvert = 0$
	$\lvert I(x)\rvert/n = 2/5 + 3/(5n) - 2p/(5n)$

Lemma 4 *If n is sufficiently large, then $y_n \in W_n^{\varphi}$ cannot be of Type 6 or 7 in Table 2.*

Proof: If y_n is of Type 6 or 7, then we have $|I(x^{(n)})|/n = |I(y_n)|/n \leq 2/5 + 4/(5n)$. Since $r_{\inf}(x) > 2/5$, there is a constant $c > 0$ such that $2/5 + 4/(5n) \leq r_{\inf}(x) - c$ and hence $|I(x^{(n)})|/n \leq r_{\inf}(x) - c$ for any sufficiently large n. Thus if y_n is of Type 6 or 7 for infinitely many n, then we have $r_{\inf}(x) \leq r_{\inf}(x) - c$, a contradiction. Hence the lemma holds. □

Now by Table 2, if y_n is of Types 1–5, then we have

$$\frac{m_n}{n} \geq \frac{|P(y_n)|}{n} \geq \frac{5|I(y_n)|}{3n} - \frac{2}{3} - \frac{4}{3n} = \frac{5|I(x^{(n)})|}{3n} - \frac{2}{3} - \frac{4}{3n} \ . \tag{7}$$

Thus Lemma 4 implies that

$$R_{\inf}(x) \geq \liminf_{n\to\infty} \left(\frac{5|I(x^{(n)})|}{3n} - \frac{2}{3} - \frac{4}{3n} \right) = \frac{5r_{\inf}(x) - 2}{3} \ , \tag{8}$$

as desired.

Secondly, we prove the lower bound of $R_{\sup}(x)$. This is trivial if $r_{\sup}(x) \leq 2/5$, therefore we assume that $r_{\sup}(x) > 2/5$. The task is to show that, for any $\varepsilon > 0$, there exist infinitely many indices n such that $m_n/n > (5r_{\sup}(x) - 2)/3 - \varepsilon$. Now take a positive ε' such that $\varepsilon' < 3\varepsilon/10$ and $\varepsilon' < r_{\sup}(x) - 2/5$. Then by the definition of $r_{\sup}(x)$, there exist infinitely many indices n such that $|I(x^{(n)})|/n > r_{\sup}(x) - \varepsilon'$. Let $\mathcal{N}$ denote the (infinite) set of the indices n with this property. Now we have the following:

Lemma 5 *$y_n \in W_n^{\varphi}$ is of Type 1–5 in Table 2 for any sufficiently large $n \in \mathcal{N}$.*

Proof: If $n \in \mathcal{N}$ and y_n is of Type 6 or 7, then $|I(y_n)|/n \leq 2/5 + 4/(5n)$ by Table 2, while $|I(y^{(n)})|/n = |I(x^{(n)})|/n > r_{\sup}(x) - \varepsilon'$ by the definition of $\mathcal{N}$. Thus we have $4/(5n) > r_{\sup}(x) - \varepsilon' - 2/5$ for such n. However, since $r_{\sup}(x) - \varepsilon' - 2/5 > 0$ by the choice of ε', the relation does not hold if n is sufficiently large. Hence the lemma holds. □

By Lemma 5 and Table 2, the inequality in (7) holds for any sufficiently large $n \in \mathcal{N}$. Thus by the definitions of $\mathcal{N}$ and ε', we have

$$\frac{m_n}{n} > \frac{5r_{\sup}(x) - 2}{3} - \frac{5\varepsilon'}{3} - \frac{4}{3n} > \frac{5r_2 - 2}{3} - \frac{\varepsilon}{2} - \frac{4}{3n} \tag{9}$$

for any sufficiently large $n \in \mathcal{N}$. Since the right-hand side of (9) is larger than $(5r_{\sup}(x) - 2)/3 - \varepsilon$ for any sufficiently large n, it follows that there exist infinitely many n with the desired property.

Hence the proof of the lower bounds in Theorem 1 is concluded.

2.2 Best-Possibility of Lower Bounds

In this subsection, we give an infinite binary word x for any $2/5 \leq r \leq 1$ such that $|I(x^{(n)})|/n$ and $|P(x^{(n)})|/n$ converge to r and $(5r-2)/3$, respectively, when $n \to \infty$. This proves the best-possibility of the lower bounds in Theorem 1. Note that we can take $x = 0101101011\cdots$ (infinite repetition of 01011) and $x = 0000\cdots$ (infinite repetition of 0) for the cases $r = 2/5$ and $r = 1$, respectively. Thus we assume that $2/5 < r < 1$.

Our construction of the word x is as follows. First, put

$$p = \left\lceil \frac{5r-2}{1-r} \right\rceil \text{ and } \alpha = p + 5 - \frac{3}{1-r} = p - \frac{5r-2}{1-r} \ , \tag{10}$$

where $\lceil z \rceil$ denotes the smallest integer m such that $z \leq m$, therefore $1 \leq p < \infty$ and $0 \leq \alpha < 1$. Let $\alpha = (0.\alpha_1\alpha_2\cdots)_2$ be the unique dyadic expansion of α with infinitely many 0s. Now we define finite binary sequences $x^{\langle 0 \rangle}, x^{\langle 1 \rangle}, \ldots$, such that $x^{\langle i \rangle}$ is a proper initial subword of $x^{\langle i+1 \rangle}$, by

$$x^{\langle 0 \rangle} = \emptyset \text{ and } x^{\langle i \rangle} = x^{\langle i-1 \rangle} x^{\langle i-1 \rangle} 010110^{p-\alpha_i} \text{ for } i \geq 1 \ . \tag{11}$$

Put $\ell_i = \ell(x^{\langle i \rangle})$, $I_i = |I(x^{\langle i \rangle})|$, and $P_i = |P(x^{\langle i \rangle})|$ for each i. Let the word x be the limit of the sequence $x^{\langle 0 \rangle}, x^{\langle 1 \rangle}, \ldots$. Then an induction on i shows the following property:

Lemma 6 *For any $i \geq 1$, we have*

$$\begin{aligned} \ell_i &= (2^i - 1)(p+5) - \sum_{j=1}^{i} 2^{i-j}\alpha_j \ , \ \ I_i = (2^i - 1)(p+2) - \sum_{j=1}^{i} 2^{i-j}\alpha_j \ , \\ P_i &= (2^i - 1)p - 1 - \sum_{j=1}^{i} 2^{i-j}\alpha_j + \delta_{p,1}\alpha_i \ , \end{aligned} \tag{12}$$

where $\delta_{i,j}$ denotes the Kronecker delta.

By Lemma 6, we have

$$5I_i - 2\ell_i = 3P_i + 3 - 3\delta_{p,1}\alpha_i \text{ for any } i \geq 1 \ . \tag{13}$$

The following property is a key ingredient of the proof:

Lemma 7 *Each finite initial subword $x^{(n)}$ of the above word x is decomposed as*

$$x^{(n)} = x^{\langle i_1 \rangle} x^{\langle i_2 \rangle} \cdots x^{\langle i_{k-1} \rangle} (x^{\langle i_k \rangle})^{\lambda+1} y \ , \tag{14}$$

where $k \geq 1$, $i_1 > i_2 > \cdots > i_k \geq 0$, $\lambda \in \{0, 1\}$, y is a (possibly empty) initial subword of $010110^{p-\alpha_{i_k+1}}$, and $i_k \geq 1$ if $k \geq 2$.

Proof: By the definition of x, it suffices to show that every initial subword x' of each $x^{\langle m \rangle}$ admits such a decomposition. We proceed the proof by induction on m. The case $m \leq 1$ is obvious (take $k = 1$, $i_1 = 0$ and $y = x'$), therefore we consider the case $m \geq 2$. Now by the construction of $x^{\langle m \rangle}$, the last position of x' is contained in the first $x^{\langle m-1 \rangle}$, in the second $x^{\langle m-1 \rangle}$, or in the remaining part $010110^{p-\alpha_m}$. In the first case, the claim follows from the induction hypothesis. In the third case, the claim follows by taking $k = 1$, $i_k = m - 1$ and $\lambda = 1$. Finally, in the second case, $x' = x^{\langle m-1 \rangle} w$ for an initial subword w of $x^{\langle m-1 \rangle}$. By the induction hypothesis, w admits a decomposition of the following form:

$$w = x^{\langle i'_1 \rangle} x^{\langle i'_2 \rangle} \cdots x^{\langle i'_{k'-1} \rangle} (x^{\langle i'_{k'} \rangle})^{\lambda'+1} y' \ , \tag{15}$$

where $m - 2 \geq i'_1 > i'_2 > \cdots > i'_{k'}$. Now the claim follows by taking $k = k' + 1$, $i_1 = m - 1$, $i_j = i'_{j-1}$ for $2 \leq j \leq k$, $\lambda = \lambda'$, and $y = y'$. Hence the lemma holds in any case. □

Now in the decomposition of $x^{(n)}$ in (14), for any $n \geq \ell_1$, we have $i_1 \geq 1$ and hence $i_k \geq 1$. Then a straightforward argument shows that, for any $n \geq \ell_1$, we have

$$\begin{aligned} n = \ell(x^{(n)}) = \sum_{j=1}^{k} \ell_{i_j} + \lambda \ell_{i_k} + \ell(y) \ , \ |I(x^{(n)})| = \sum_{j=1}^{k} I_{i_j} + \lambda I_{i_k} + |I(y)| \ , \\ |P(x^{(n)})| = \sum_{j=1}^{k} P_{i_j} + \lambda P_{i_k} + |P(y)| + k + (\lambda - \delta_{y,\emptyset})(1 - \delta_{p,1}\alpha_{i_k}) - \delta_{p,1} \sum_{j=1}^{k} \alpha_{i_j} \ . \end{aligned} \tag{16}$$

By these equalities, (12), and (13), an elementary argument shows that $|I(x^{(n)})|/n$ converges when $n \to \infty$ to $(p+2-\alpha)/(p+5-\alpha) = r$, and $|P(x^{(n)})|/n$ converges when $n \to \infty$ to $(5r-2)/3$.

Hence the lower bounds in Theorem 1 are best possible.

2.3 Upper Bounds

The proof of the upper bounds in Theorem 1 are much simpler than the case of lower bounds. In fact, for any finite binary word w, the map $i \mapsto i-1$ is an injection from $P(w)$ to $I(w)$, therefore $|P(w)| \leq |I(w)|$. Now the upper bounds are easy consequences of the inequality. Thus the nontrivial assertion on the upper bounds is only their best-possibility.

2.4 Best-Possibility of Upper Bounds

To prove the best-possibility of the upper bounds in Theorem 1, for any $0 \leq r \leq 1$, we construct an infinite binary word x such that both $|I(x)|/n$ and $|P(x)|/n$ converge to r when $n \to \infty$. Since $x = 000\cdots$ (infinite repetition of 0) satisfies the conditions when $r = 1$, we assume from now that $0 \leq r < 1$.

First, we define auxiliary values $\delta_k \in \{0, 1\}$ for $k \geq 1$ inductively by

$$\delta_k = 1 \ \text{ if } \frac{\sum_{i=1}^{k-1} \delta_i \cdot 2i + 2k}{k(k+1)} \leq r \ , \ \ \delta_k = 0 \ \text{ otherwise.} \tag{17}$$

Then we have

$$\frac{\sum_{i=1}^{k} \delta_i \cdot 2i}{k(k+1)} \leq r \text{ for any } k \ . \tag{18}$$

Now let $x^{\langle k \rangle} = (1-\delta_k)^{2k}$ be the repetition of $1 - \delta_k \in \{0, 1\}$ of length $2k$ for each k, and define $x = x^{\langle 1 \rangle} x^{\langle 2 \rangle} x^{\langle 3 \rangle} \cdots$. Then for each k, we have

$$\ell(x^{\langle 1 \rangle} \cdots x^{\langle k \rangle}) = k(k+1) \ \text{ and } |I(x^{\langle 1 \rangle} \cdots x^{\langle k \rangle})| = \sum_{i=1}^{k} \delta_i \cdot 2i \ . \tag{19}$$

Now by (18) and (19), an elementary argument shows that both $|I(x^{(n)})|/n$ and $|P(x^{(n)})|/n$ converge to r when $n \to \infty$. Thus the upper bounds in Theorem 1 are best possible.

Hence the proof of Theorem 1 is concluded.

3 Relations with PRNGs Based on Integer-Valued Logistic Maps

In this section, we explain a relation of the above problem with analyses of the PRNGs based on integer-valued logistic maps $L_n(x) = \lfloor x(2^n - x)/2^{n-2} \rfloor$ mentioned in the Introduction. A problem concerned in the analyses is to decide, for each parameter $n \in \mathbb{Z}$ with $n \geq 1$, whether there exists an initial value $s_0 \in \{1, 2, \ldots, 2^n - 1\}$ such that $L_n(s_0) = 2^{n-1}$ (note that $L_n(2^{n-1}) \neq 2^{n-1}$). Recall from the Introduction that we call a parameter n dangerous if such an undesirable initial value $s_0 \neq 2^{n-1}$ for $L_n(x)$ exists. Now the problem is rephrased as the problem on existence of dangerous parameters n in the above sense.

The condition for an accuracy parameter n to be dangerous is equivalent to that there exists an integer $1 \leq x \leq 2^n - 1$ such that $2^{n-1} \leq x(2^n - x)/2^{n-2} < 2^{n-1} + 1$ (that is equivalent to $L_n(x) = 2^{n-1}$). By solving the inequalities, the relation $L_n(x) = 2^{n-1}$ is satisfied if and only if

$$\sqrt{2^{2n-3} - 2^{n-2}} < |2^{n-1} - x| \leq \sqrt{2^{2n-3}} \ . \tag{20}$$

Moreover, since

$$\sqrt{2^{2n-3}} - \sqrt{2^{2n-3} - 2^{n-2}} = \frac{2^{n-2}}{\sqrt{2^{2n-3}} + \sqrt{2^{2n-3} - 2^{n-2}}} > \frac{2^{n-2}}{2\sqrt{2^{2n-3}}} = \frac{\sqrt{2}}{4} \ , \tag{21}$$

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

Lemma 8 *A parameter n is dangerous if $2^{n-2}\sqrt{2} - \sqrt{2}/4 \leq m \leq 2^{n-2}\sqrt{2}$ for some integer m.*

From now, we rephrase the statement of Lemma 8 in terms of the dyadic expansion of $\sqrt{2}$. Namely, let $b = b_1 b_2 b_3 \cdots$ denote an infinite binary word that is the fractional part of the dyadic expansion of $\sqrt{2}$. For example, $b^{(10)} = 0110101000$ as mentioned in the Introduction. Then the fractional part of the dyadic expansion of $2^{n-2}\sqrt{2}$ is $(0.b_{n-1}b_n b_{n+1} \cdots)_2$, while the dyadic expansion of $\sqrt{2}/4$ is $(0.01b_1 b_2 b_3 \cdots)_2$. Thus Lemma 8 implies the following:

Lemma 9 *A parameter n is dangerous if*

$$(0.b_{n-1}b_n b_{n+1} \cdots)_2 \leq (0.01b_1 b_2 b_3 \cdots)_2 \ . \tag{22}$$

In particular, since $b_1 b_2 b_3 = 011$ as above, (22) is satisfied if $b_{n-1}b_n = 00$, $b_{n-1}b_n b_{n+1}b_{n+2} = 0100$, or $b_{n-1}b_n b_{n+1}b_{n+2}b_{n+3} = 01010$; that is, $n \in P(b)$ in the sense of Sections 1 and 2. Summarizing, we have the following result:

Proposition 1 *A parameter n is dangerous if $n \in P(b)$.*

Hence by Theorem 1, we obtain the following lower bound of the asymptotic occurrence rate of the dangerous parameters in the positive integers:

Theorem 2 *Let d_N denote the number of the dangerous parameters $n \leq N$. Then we have*

$$\liminf_{N\to\infty} \frac{d_N}{N} \geq \frac{5r_{\inf}(b) - 2}{3} \text{ and } \limsup_{N\to\infty} \frac{d_N}{N} \geq \frac{5r_{\sup}(b) - 2}{3} \ . \tag{23}$$

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

Regarding the quantities $r_{\inf}(b)$ and $r_{\sup}(b)$, it has been (even implicitly) conjectured that $r_{\inf}(b) = r_{\sup}(b) = 1/2$, that is, the asymptotic occurrence rates of 0s and 1s in the dyadic expansion of $\sqrt{2}$ coincide with each other. This property for a real number is called *simple normality* to base $q = 2$, while a stronger notion is *normality* to base q meaning that for each $\ell \geq 1$, the asymptotic occurrence rate of every subword of length ℓ in a given infinite q-ary word coincides with each other (see *e.g.* Borel (1909); Kuipers et al. (1974); Hertling (2002)). A motivation of the above conjecture on the simple normality of $\sqrt{2}$ to base 2 would come from a naive intuition that the dyadic expansion of $\sqrt{2}$ (and also of other several famous irrational numbers) looks very random, and also from a theorem by Borel (Borel (1909)) that almost every (in terms of Lebesgue measure) real number is normal to every base $q \geq 2$. We mention that recently Isaac posted a preprint (Isaac (2005)) to prove that every $\sqrt{s}$ with s an integer that is not perfect square is simply normal to base 2 (however, the author could not understand that his proof is completely correct). Moreover, several computer experiments also support the conjecture.

If this conjecture is true, then Theorem 2 implies that the asymptotic occurrence rate of the dangerous parameters n is at least $1/6$. On the other hand, even if the conjecture were not true, then a weaker assumption $r_{\sup}(b) > 2/5$ still could imply that infinitely many dangerous parameters n exist. This means that, to avoid to falsely choose a dangerous parameter n in a practical use of the above PRNG, a naive countermeasure of using sufficiently large parameters does very probably not solve the problem essentially.

4 Open Problems

In the previous sections, we have given in Proposition 1 a sufficient condition for a parameter n for the PRNG to be dangerous in terms of the dyadic expansion of $\sqrt{2}$. However, it is shown that the condition is not necessary. In fact, a direct calculation based on Lemma 9 shows that $n = 65$ is a dangerous parameter that does not satisfy $n \in P(b)$. Thus it is expected that we can obtain a better bound of the asymptotic occurrence rate of dangerous parameters than Theorem 2 by investigating the condition in Lemma 9, not only in Proposition 1 that is weaker than Lemma 9.

Here we describe a rough observation of the author for this problem. Put $b_0 = 1$ and $b_{-1} = 0$ for simplicity. Let $\mathcal{C}(b)$ be the set of all binary words of the form $b_{-1}b_0b_1 \cdots b_{k-1}0$ for any $k \geq -1$ with $b_k = 1$. Then a parameter n satisfies (22) if the infinite word $b_{n-1}b_nb_{n+1} \cdots$ involves a member of $\mathcal{C}(b)$ as an initial subword. Thus for any subset $\mathcal{C}'$ of $\mathcal{C}(b)$, a lower bound for *any* infinite binary word x of the number of indices n such that $x_{n-1}x_nx_{n+1} \cdots$ involves a member of $\mathcal{C}'$ yields a lower bound of the asymptotic rate of the dangerous parameters. For example, the argument in the previous sections deals with the subset $\mathcal{C}'$ with only three members 00, 0100, and 01010. Thus an immediate improvement of the bound in Theorem 2 would be derived by applying a similar argument to a larger subset $\mathcal{C}'$ of $\mathcal{C}(b)$.

Another possible generalization is a problem of finding a lower bound of the asymptotic rate of indices n that satisfy (22), not only for the above binary word b but also for an *arbitrary* infinite binary word x. A solution of this generalized problem yields another bound of the asymptotic rate of the dangerous parameters. Since an occurrence of a bit 1 in x yields a member of the set corresponding to $\mathcal{C}(b)$ in the previous paragraph, it seems possible that some nontrivial bound holds also in the generalized setting, at least when the asymptotic rate of 1s in x is not too low. Owing to the self-referential description, the author hopes that the problem involves certain mathematically interesting structure that is worthy to investigate.

Acknowledgements

The author would like to thank Dr. Reynald Affeldt, Dr. Takayuki Miyadera, Dr. Makoto Sugita, Dr. Kentaro Imafuku, and Professor Hideki Imai, for their significant comments. Moreover, the author also would like to the anonymous referees for their precious comments.

References

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), Seoul, Korea, 2006.

S. Araki, T. Miyazaki, and S. Uehara. *A study on logistic map on integer for pseudorandom number generators* (in Japanese). In: Proceedings of The 2008 Symposium on Cryptography and Information Security (SCIS 2008), Miyazaki, Japan, 2008.

E. Borel. *Les probabilités dénombrables et leurs applications arithmetiqués*. Rend. Circ. MAt. Palermo **27** (1909) 247–271.

P. Hertling. *Simply normal numbers to different bases*. J. Univ. Comput. Sci. **8** (2002) 235–242.

R. Isaac. *On the simple normality to base 2 of the square root of s, for s not a perfect square*. arXiv:math/0512404v2, 2005.

L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Wiley, New York, 1974.

M. Kuribayashi and H. Tanaka. *Key generation scheme for broadcast encryption exploiting chaotic sequences*. In: Proceedings of The 2005 Symposium on Cryptography and Information Security (SCIS 2005), Kobe, Japan, 2005.

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), Sasebo, Japan, 2007.

K. Nuida. *On parameter choices for pseudorandom number generators based on integer-valued logistic map*. In: Proceedings of The 31st Symposium on Information Theory and its Applications (SITA 2008), Kinugawa, Japan, 2008.

S. C. Phatak and S. S. Rao. Logistic map: A possible random number generator. Phys. Rev. E **51** (1995) 3670–3678.

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.