On the periodic properties of self-decimated generators of pseudorandom numbers

Sergey Agievich, Oleg Solovey

National Research Center for Applied Problems of Mathematics and Informatics
Belarusian State University
Fr. Skorina av. 4, 220050 Minsk, Belarus
agievich@bsu.by, solovey@bsu.by

Abstract

We consider a self-decimated generator of pseudorandom numbers and examine the preperiod $\lambda$ and the period $\mu$ of its state sequence. We obtain the expectations and variances of $\lambda$ and $\mu$ for the case when decimation steps are chosen randomly and independently from the set $\{1, 2\}$.

1 Results

Let $\mathbb{N}$ be the set of all positive integers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $A \subset \mathbb{N}_0$ be some finite alphabet. We denote by $A^*$ the set of all finite words in the alphabet $A$ (including the empty word $\varepsilon$) and by $A^\omega$ the set of all one-way infinite words (see [1, 2] for further details). Given a word $a \in A^*$, let $l(a)$ be the length of $a$ and $w(a)$ be the sum of letters of $a$. Let $c^n$ denote $n$ successive instances of the letter $c$ and $ab$ denote the concatenation of the words $a$ and $b$.

Given $s = s_0s_1 \ldots \in A^\omega$ and $T \in \mathbb{N}$, determine the numbers $\lambda = \lambda(s, T) \in \mathbb{N}_0$ and $\mu = \mu(s, T) \in \mathbb{N}$ such that

$$T \mid (s_\lambda + s_{\lambda+1} + \ldots + s_{\lambda+\mu-1})$$

and $T$ does not divide any of the sums $s_t + s_{t+1} + \ldots + s_{\tau}$, $0 \leq t < \tau < \lambda + \mu - 1$. For example, if $s = 2212221\ldots = 2^212^31\ldots$, then

$$\lambda(1, s) = 0, \ \mu(1, s) = 1, \ \lambda(2, s) = 0, \ \mu(2, s) = 1, \ldots \ \lambda(8, s) = 2, \ \mu(8, s) = 5, \ldots$$

The characteristics $\lambda$ and $\mu$ describe periodic properties of self-decimated generators of pseudorandom numbers (see [3, 4]). Consider a generator $G$ with the set of internal states $\mathcal{S}$, $|\mathcal{S}| = T$, and the state-transition function $\varphi: \mathcal{S} \to \mathcal{S}$. Let $\varphi$ be a full cycle substitution,
The usual way of $G$ functioning, one chooses an initial state $S_0 \in S$, calculates the sequence
\[ S_{t+1} = \varphi(S_t), \quad t = 0, 1, \ldots, \]
and uses the current internal state $S_t$ to determine the current output pseudorandom number.

The self-decimation (of internal states of $G$) means using an additional function $d: S \to A$ and replacing (2) by the rule
\[ S_{t+1} = \varphi^d(S_t), \quad t = 0, 1, \ldots. \]

Now, if the word $s$ consists of the successive letters $s_t = d(S_t)$, then $\lambda(s, T)$ and $\mu(s, T)$ are respectively the preperiod and period of the state sequence (3).

If the letters of $s$ are chosen randomly, then $\lambda$ and $\mu$ become random variables. The expectations $E \lambda, E \mu$ and variances $D \lambda, D \mu$ of these variables are of interest in connection with the estimation of the preperiod and period of self-decimated generators. One of the well-known results in the theory of random mappings can be stated as follows: If the letters of $s$ are chosen randomly, independently and uniformly from the alphabet $A = \{0, 1, \ldots, T-1\}$, then $E \lambda(s, T)$ and $E \mu(s, T)$ have the form $\sqrt{\pi T/8} + O(1)$ as $T \to \infty$ (see, for example, [5]). Unfortunately, obtaining similar asymptotic (or exact) expressions for arbitrary alphabet $A$ seems to be a more complex task.

We consider the frequent choice $A = \{1, 2\}$ and obtain the following result.

**Theorem.** If the letters of the word $s = s_0 s_1 \ldots \in \{1, 2\}^\infty$ are chosen randomly, independently with the probabilities
\[ P \{s_t = 1\} = 1 - p, \quad P \{s_t = 2\} = p, \quad 0 < p < 1, \quad t = 0, 1, \ldots, \]
then as $T \to \infty$ it holds that
\[ E \lambda(s, T) = \frac{p}{(1 - p)(1 + p)^2} + O(p^T T), \quad E \mu(s, T) = \frac{T}{1 + p} + O(p^T T), \]
\[ D \lambda(s, T) = \frac{p + p^3 + p^4}{(1 - p)^2(1 + p)^4} + O(p^T T^2), \quad D \mu(s, T) = \frac{T p(1 - p)}{(1 + p)^3} + O(p^T T^2). \]

Note that (4), (5) also hold for the random word $s = s_0 s_1 \ldots \in \{q, 2q\}^\infty$, where $q \in \mathbb{N}$ is coprime to $T$ and $P \{s_t = q\} = 1 - p$, $P \{s_t = 2q\} = p$.

Note also that R. Rueppel in the paper [4] examined the case when the word $s$ is determined from a linear recurrence sequence $\sigma_0, \sigma_1, \ldots$ over the field of order 2 by the rule
\[ s_t = \begin{cases} 1, & \sigma_t = 0, \\ 2, & \sigma_t = 1, \end{cases} \quad t = 0, 1, \ldots. \]

Rueppel showed that if the characteristic polynomial of the sequence $\sigma_0, \sigma_1, \ldots$ is primitive of degree $k$, then
\[ \mu(s, T) = \left\lfloor \frac{2T}{3} \right\rfloor, \quad T = 2^k - 1, \]
where $\lfloor z \rfloor$ is the largest integer $\leq z$. As we can see from Theorem, this estimation agrees with the expectation $E \mu(s, T)$ for the random word $s$ with $p = 1/2$. 

2
Given \( s \) and \( T \), determine the words \( s_0 \ldots s_{\lambda-1} \) and \( s_{\lambda} \ldots s_{\lambda+\mu-1} \), which we call the prefix and cyclic part of \( s \), respectively. Conversely, any nonempty word \( a \in A^* \) is a cyclic part of some word \( s \). Given \( a \), we can determine the possible values of \( T = T(a) \) and the set \( B(a) \) of the possible prefixes of \( s \), using the following restrictions:

P1) \( w(a) \equiv 0 \pmod{T} \);

P2) the residues
\[
0, \ a_0, \ a_0 + a_1, \ldots, \ a_0 + a_1 + \ldots + a_{l(a)-2} \pmod{T} \tag{6}
\]
are pairwise distinct;

P3) if \( b \in B(a) \) and \( b \neq \varepsilon \), then any of the residues
\[
-b_{l(b)-1}, \ -b_{l(b)-2} - \ldots - b_0 \pmod{T}
\]
differs from the residues (6).

Indeed, the condition P1 follows from the definition of a cyclic part. If the condition P2 fails, then
\[
a_t + \ldots + a_r \equiv 0 \pmod{T}
\]
for some \( t, \tau, 0 \leq t < \tau < l(a) - 1 \) and \( a \) cannot be a cyclic part. Finally, if the condition P3 fails, then one of the congruences
\[
b_t + \ldots + b_{l(b)-1} \equiv 0 \pmod{T}, \ b_t + \ldots + b_{l(b)-1} + a_0 + \ldots + a_r \equiv 0 \pmod{T}
\]
holds for some \( t \leq l(b) - 1, \tau < l(a) - 1 \) and \( a \) cannot be a cyclic part again.

Suppose that \( A = \{1, 2\} \) and let \( x = (x_1, x_2), \ y = (y_1, y_2) \). Given \( \Omega \subseteq A^* \), define the generating function
\[
G_\Omega(x, y, z) = \sum_{n_1, n_2, m_1, m_2 \geq 0} g(n_1, n_2, m_1, m_2, t) x_1^{n_1} x_2^{n_2} y_1^{m_1} y_2^{m_2} z^t \tag{7}
\]
where \( g(n_1, n_2, m_1, m_2, t) \) is the number of the words \( s \) with the cyclic part \( a \in \Omega \) and prefix \( b \in B(a) \) such that \( a \) contains \( n_1 \) letters 1 and \( n_2 \) letters 2, \( b \) contains \( m_1 \) letters 1 and \( m_2 \) letters 2, and \( T(a) = t \).

Let us divide \( A^* \) into the subsets \( \Omega_1, \Omega_2, \Omega_3, \Omega_4 \) (which will be defined below) and determine the generating function of the form (7) for each subset.

1. \( \Omega_1 = \{2\}^* \). If \( a \in \Omega_1 \), then \( T \mid w(a) \) according to P1 and \( T \geq l(a) > w(a)/2 \) according to P2. Therefore, \( T(a) = w(a) \). Now, if \( a = 2^m \), then \( B(a) = \{2^k : k = 0, \ldots, m\} \)
\[ G_{\Omega_1}(x, y, z) = \sum_{m \geq 0} x_1^m y_2^{m+1} \sum_{k=0}^{m} y_2^k \]
\[ = \frac{x_1 z}{1 - y_2} \sum_{m \geq 0} (x_2 z^2)^m - \frac{x_1 y_2 z}{1 - y_2} \sum_{m \geq 0} (x_2 y_2 z^2)^m \]
\[ = \frac{x_1 z}{(1 - y_2)(1 - x_2 z^2)} - \frac{x_1 y_2 z}{(1 - y_2)(1 - x_2 y_2 z^2)} \]

2. \( \Omega_2 = \{1, 2\}^* \{2\}^1 \). Again, \( T(a) = w(a) \) according to P1, P2. Let \( a = \alpha 12^m 1, \alpha \in A^* \). Using P3, we obtain \( B(a) = \{2^k : k = 0, \ldots, m\} \). Therefore,
\[ G_{\Omega_2}(x, y, z) = \sum_{l \geq 0} (x_1 z + x_2 z^2)^l \sum_{m \geq 0} x_1^m x_2^{m+2} \sum_{k=0}^{m} y_2^k \]
\[ = \frac{x_1 z^2}{(1 - x_1 z - x_2 z^2)(1 - x_2 z^2)} \]

3. \( \Omega_3 = \{1, 2\}^* 2 \{2\}^* \). In this case \( T(a) = w(a) \) and \( B(a) = \{\varepsilon\} \cup \{2^k 1 : k = 0, \ldots, m\} \) for \( a = \alpha 2^m \). Therefore,
\[ G_{\Omega_3}(x, y, z) = \sum_{l \geq 0} (x_1 z + x_2 z^2)^l \sum_{m \geq 0} x_1^m x_2^{m+1} \sum_{k=0}^{m} y_1 y_2^k \]
\[ = \frac{x_1 x_2 z^3 (1 + y_1 - x_2 y_2 z^2)}{(1 - x_1 z - x_2 z^2)(1 - x_2 z^2)(1 - x_2 y_2 z^2)} \]

4. \( \Omega_4 = \{2\}^* \). If \( a = 2^m \), then \( T(a) = 2m \) and \( B(a) = \{\varepsilon\} \cup \{2^k 1 : k = 0, \ldots, m - 1\} \). In addition, \( T(a) = m \) and \( B(a) = \{\varepsilon\} \) for odd \( m \). The generating function has the form
\[ G_{\Omega_4}(x, y, z) = \sum_{m \geq 1} x_1^m z^{2m} \left( 1 + \sum_{k=0}^{m-1} y_1 y_2^k \right) + \sum_{m \geq 0} x_2^{m+1} z^{2m+1} \]
\[ = \frac{1 + x_2 y_1 z^2 - x_2 y_2 z^2}{(1 - x_2 z^2)(1 - x_2 y_2 z^2)} + \frac{x_2 z}{1 - x_2^2 z^2} \]

Finally, we get
\[ G_{A^*}(x, y, z) = G_{\Omega_1}(x, y, z) + G_{\Omega_2}(x, y, z) + G_{\Omega_3}(x, y, z) + G_{\Omega_4}(x, y, z) \]
\[ = \frac{1 + x_2 y_1 z^2 + x_1 x_2 y_2 z^3 - x_2 y_2 z^3}{(1 - x_2 y_2 z^2)(1 - x_1 z - x_2 z^2)} + \frac{x_2 z}{1 - x_2^2 z^2} \]

To find the expectation \( E \lambda(s, T) \), introduce the operator \( E_y = y_1 \frac{\partial}{\partial y_1} + y_2 \frac{\partial}{\partial y_2} \) and denote \( \pi = (1 - p, p) \). We have
\[ E \lambda(s, T) = \sum_{n_1, n_2, m_1, m_2 \geq 0} \frac{g(n_1, n_2, m_1, m_2, T)(m_1 + m_2)}{(1 - p)^{n_1 + m_1} p^{n_2 + m_2}} = [z^T] E_y G_{A^*}(\pi, y, z) \big|_{y=\pi}, \]

4
where $z_T f(z)$ is the coefficient of $z_T$ in $f(z) = f_0 + f_1 z + f_2 z^2 + \ldots$. Now the first part of (4) follows from

$$E_y G_{A^*}(\pi, y, z)|_{y=\pi} = \frac{(1-p)p z^2}{(1-z)(1-p^2 z^2)}$$

$$= \frac{p}{(1-p)(1+p)^2 (1-z)} - \frac{1}{4(1-p)z^2} - \frac{p}{4(1-p)(1-pz)}$$

$$+ \frac{1-p}{4(1+p)(1+pz)^2} - \frac{1}{4(1+p)^2 (1+pz)}$$

$$= \frac{p}{(1-p)(1+p)^2} \sum_{k \geq 0} z^k - \frac{1}{4} \sum_{k \geq 0} (k+1) p^k z^k - \frac{p}{4(1-p)} \sum_{k \geq 0} p^k z^k$$

$$+ \frac{1-p}{4(1+p)} \sum_{k \geq 0} (k+1)(-p)^k z^k - \frac{(1-p)p}{4(1+p)^2} \sum_{k \geq 0} (-p)^k z^k.$$

Similarly, introducing the operator $E_x = x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2}$ and using the equalities

$$E \mu(s, T) = [z^T] E_x G_{A^*}(x, \pi, z)|_{x=\pi},$$

$$D \lambda(s, T) = [z^T] E_y G_{A^*}(\pi, y, z)|_{y=\pi} - (E \lambda(s, T))^2,$$

$$D \mu(s, T) = [z^T] E_x G_{A^*}(x, \pi, z)|_{x=\pi} - (E \mu(s, T))^2,$$

we prove the remaining parts of (4), (5).

References

1. Goulden I. P., Jackson D. M. (1983). Combinatorial Enumeration. N.Y.: John Wiley.

2. Choffrut Ch., Karhumaki J. (1997). Combinatorics of Words, in Handbook of Formal Languages (Rozenberg G., Salomaa A. eds.). Berlin, Heidelberg: Springer-Verlag.

3. Chambers W., Gollmann D. (1998). Generators for sequences with nearmaximal linear equivalence. IEE Proc. E. Vol. 135, pp. 67-69.

4. Rueppel R. (1988). When shift registers clock themselves. Lecture Notes in Computer Science, Advances in Cryptology – Eurocrypt 87. Vol. 304. Springer-Verlag, pp. 53-64.

5. Flajolet P., Odlyzko A. (1990). Random mapping statistics. Lecture Notes in Computer Science, Advances in Cryptology – Eurocrypt 89. Vol. 434. Springer-Verlag, pp. 329-354.