Tatra Mt. Math. Publ. **41** (2008), 107–117

TATRA MOUNTAINS
Mathematical Publications

THE PROPERTIES OF BI-IDEALS IN THE FREQUENCY TEST

Edmunds Cers

ABSTRACT. We explore the properties of bi-ideals in the frequency test. We prove that the properties of bi-ideals in this test are determined by the base words which generate the bi-ideal. We also show a method for base word selection, which guaranties that the generated bi-ideal will pass the frequency test.

Introduction

Periodic sequences have found a wide application in cryptography. Most notably pseudorandom bit generators and stream ciphers make heavy use of such sequences with very long periods [9].

We want to find out whether we could use non-periodic sequences instead. In particular, we explore the possibility to use some classes of bi-ideals. Bi-ideal sequences have been considered under different names in both algebra and combinatorics [1, 3, 10]. Bi-ideals are a superclass of the class of periodic sequences [5]. We are most interested in non-periodic subclasses of bi-ideals and their possible applications in cryptography.

It must be noted, that bi-ideals and their subclasses are still an actively researched topic. For some recent results, see for example [2, 4]. There is also an active interest in the statistical properties of sequences, as it is exemplified by [8].

Our approach is to look for a subclass of non-periodic bi-ideals, that would resemble independent and identically-distributed (i.i.d.) sequences. In this paper we explore the behavior of bi-ideals in the frequency test, and look for possible subclasses, that would behave "well enough".

We will prove two theorems. One shows, that bi-ideals are stable in the frequency test, in a sense, that will be evident later. After that, we show a method of selecting a subclass of bi-ideals, that is indistinguishable from i.i.d. bit-sequences

2000 Mathematics Subject Classification: 68P25, 94A60.
Keywords: frequency test, bi-ideal.

using the frequency test. We prove a theorem showing that bi-ideals generated this way are, indeed, indistinguishable from i.i.d. bit-sequences using the frequency test.

1. Preliminaries

1.1. Bi-ideals

If A is an alphabet, then let A^* denote all finite words of this alphabet, and A^ω all infinite words of this alphabet. And finally, $A^\infty = A^* \cup A^\omega$. Also, we will denote the zero-length word with λ.

Let $a\#b$ or simply ab denote word concatenation. We will also use $|a|$ to denote the length of the word a.

Definition 1.1. A word $x \in A^\omega$ is called *recurrent* if each finite sub-word occurs in the word an infinite number of times. A word uy, where $u \in A^*$ and $y \in A^\omega$, is called *almost recurrent* if y is recurrent.

Definition 1.2. A sequence of finite words $v_0, v_1, \ldots, v_n, \ldots$ is called a *bi-ideal sequence* if $\forall i\, v_{i+1} \in v_i A^* v_i$.

Or, alternatively,

Definition 1.3. A sequence of finite words $v_0, v_1, \ldots, v_n, \ldots$ is called a bi-ideal sequence, if there exists a sequence of finite words $u_0, u_1, \ldots, u_n, \ldots$, such that

$$v_0 = u_0\,, \qquad v_{i+1} = v_i u_{i+1} v_i\,.$$

Definition 1.4. Suppose, that we have a sequence of the words $(u_i)_{i\in\mathbb{N}}$, where $\forall i\ u_i \in A^*$, and $u_0 \neq \lambda$, that generates a bi-ideal sequence $(v_i)_{i\in\mathbb{N}}$ as in Definition 1.3.

The limit of the sequence $\lim_{i\to\infty} v_i = x$ is called a *bi-ideal*. We say that the sequence (u_i) generates the bi-ideal x or that x is the bi-ideal generated by the sequence (u_i). If $\forall i\ |u_i| \leq l$, then x is called an *l–restricted* bi-ideal. We call the bi-ideal x *restricted* if such a number l exists, that x is an l–restricted bi-ideal.

Alternatively, the word $x \in A^\omega$ is a bi-ideal, if and only if it is a *recurrent* word.

For a more in depth coverage of the topic see, e.g., [6].

1.2. The frequency test

Let us look at the prefix of a bit sequence $\{x_n\}$,

$$x = (x_1, x_2, \ldots, x_{N+\nu-1}).$$

This prefix has $N = |x| - \nu + 1$ overlapping sub-sequences of length ν.

Consider a specific bit sequence of the length ν,

$$s = (s_1, s_2, \ldots, s_\nu).$$

We can denote the event of the mth sub-sequence of $\{x_n\}$ being equal to s with

$$D^m A_s(x) = \big\{(x_m, x_{m+1}, \ldots, x_{m+\nu-1}) = s\big\}.$$

If $\{x_n\}$ is indistinguishable from an i.i.d. bit-sequence, then

$$E\Big(I\big(D^m A_s(x)\big)\Big) = 2^{-\nu},$$

where I is the indicator function and E denotes the expected value. Or, if we denote the number of occurrences of the sequence s in x with $|x|_s$,

$$E(|x|_s) = 2^{-\nu} N.$$

For a broader coverage see [7].

2. The properties of bi-ideals in the frequency test

2.1. The stability of bi-ideals in the frequency test

Let u and w be finite words. Then we denote:

(i) $|w|_u = \big|\big\{(u', u, u'') | u'uu'' = w\big\}\big|$. $|w|_u$ is the count of different ways u is contained in w.

(ii) We will call the number $\alpha(w, u) = \dfrac{|w|_u}{|w| - |u| + 1}$ the *relative frequency* of u in w. By this definition $0 \le \alpha(w, u) \le 1$.

Suppose, x is a bi-ideal generated by the sequence (u_i), then we can denote:

(iii) $\alpha_n(u) = \alpha(v_n, u)$, where v_n is the n-th element of the bi-ideal sequence from Definition 1.4, where $x = \lim_{i\to\infty} v_i$.

(iv) Let $Pref\, w$ denote the set of all finite prefixes of the word w.

Lemma 2.1. *If x is a restricted bi-ideal, then*

$$\forall l \in \mathbb{N}\, \forall \varepsilon > 0\, \exists \delta \in \mathbb{N}\, \forall u \in A^* \Big[|u| = l \Rightarrow \forall n \ge \delta\ |\alpha_n(u) - \alpha_\delta(u)| \le \varepsilon\Big].$$

P r o o f. Suppose, the sequence (u_i), generates an l_x–restricted bi-ideal x. Let us consider the bi-ideal sequence (v_i) generated by (u_i) as per Definition 1.3. Then from Definition 1.4 each v_i is a prefix of the bi-ideal x, and $|v_j| > |v_i|$, when $j > i$.

Let us denote:

$$\begin{array}{rclcrcll} l_i & = & |v_{\delta+i}|_u & i \geq 0, & l'_i & = & l_i - 2l_{i-1}, & i \geq 1; \\ m_i & = & |v_{\delta+i}| - l + 1, & i \geq 0, & m'_i & = & m_i - 2m_{i-1}, & i \geq 1; \end{array}$$

$$\alpha_i \quad = \quad \alpha(v_{\delta+i}, u) = \frac{|v_{\delta+i}|_u}{|v_{\delta+i}| - |u| + 1} = \frac{|v_{\delta+i}|_u}{|v_{\delta+i}| - l + 1} = \frac{l_i}{m_i}, \quad i \geq 0,$$

where $l = |u|$.

We will choose δ, such that $\frac{l_x+l-1}{m_0} < \varepsilon$, $(m_0 = |v_\delta| - l + 1)$.

Let us assess α_i, $i \geq 1$:

$$\begin{aligned} \alpha_i &= \frac{l_i}{m_i} = \frac{2l_{i-1} + l'_i}{2m_{i-1} + m'_i} = \frac{2l_{i-1}}{2m_{i-1} + m'_i} + \frac{l'_i}{2m_{i-1} + m'_i} \\ &= \alpha_{i-1} \frac{2}{2 + \frac{m'_i}{m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i} \\ &= \alpha_{i-1} \frac{1}{1 + \frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1} + m'_i}. \end{aligned}$$

From Definition 1.3

$$v_{i+1} = v_i u_{i+1} v_i, \quad i \geq 1,$$

therefore,

$$|v_{i+1}| = 2|v_i| + |u_{i+1}| \geq 2|v_i|, \tag{1}$$

and

$$|v_{\delta+i}| \geq 2^i |v_\delta|.$$

From here

$$m_{i-1} = |v_{\delta+i-1}| - l + 1 \geq 2^{i-1}|v_\delta| - l + 1 \geq 2^{i-1}(|v_\delta| - l + 1) = 2^{i-1} m_0. \tag{2}$$

Now consider,

$$\begin{aligned} m'_i &= m_i - 2m_{i-1} = |v_{\delta+i}| - l + 1 - 2(|v_{\delta+i-1}| - l + 1) \\ &= |v_{\delta+i}| - l + 1 - 2|v_{\delta+i-1}| + 2l - 2, \end{aligned}$$

from (1)

$$|v_{\delta+i}| - |v_{\delta+i-1}| = |u_{\delta+i}|,$$

therefore,

$$|v_{\delta+i}| - l + 1 - 2|v_{\delta+i-1}| + 2l - 2 = |u_{\delta+i}| + l - 1.$$

But because the bi-ideal is l_x restricted, $|u_{\delta+1}| \leq l_x$, and $m'_i \leq l_x + l - 1$ therefore, also considering (2):

$$\frac{m'_i}{2m_{i-1}} \leq \frac{l_x + l - 1}{2^i m_0} \leq \frac{\varepsilon}{2^i}\,.$$

Now we can remember that $1 \geq 1 - a^2 = (1-a)(1+a)$, and if $1 + a > 0$,

$$\frac{1}{1+a} \geq 1 - a$$

so that we can write

$$\begin{aligned}\alpha_i &= \alpha_{i-1}\frac{1}{1+\frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1}+m'_i} \geq \alpha_{i-1}\left(1 - \frac{m'_i}{2m_{i-1}}\right)\\ &\geq \alpha_{i-1}\left(1 - \frac{\varepsilon}{2^i}\right) = \alpha_{i-1} - \frac{\alpha_{i-1}\varepsilon}{2^i} \geq \alpha_{i-1} - \frac{\varepsilon}{2^i}\,.\end{aligned}$$

Now, let us look at l'_i:

$$l'_i = l_i - 2l_{i-1} = |v_{\delta+i}|_u - 2|v_{\delta+i-1}|_u.$$

But by Definition 1.3 $v_{\delta+i} = v_{\delta+i-1}u_{\delta+i}v_{\delta+i-1}$, and in $v_{\delta+i-1}$ there are l_{i-1} subsequences equal to u. We know, that of the $2m_{i-1}$ subsequences with the length l corresponding to the $v_{\delta+i-1}$ precisely $2l_{i-1}$ are equal to u. This means, that there can be at most $l_i \leq 2l_{i-1} + m_i - 2m_{i-1}$ sequences equal to u. And thus,

$$l'_i \leq m_i - 2m_{i-1} \leq m'_i \leq l_x + l - 1.$$

Also, from (2):

$$2m_{i-1} + m'_i \geq 2m_{i-1} \geq 2^i m_0.$$

From this

$$\begin{aligned}\alpha_i &= \alpha_{i-1}\frac{1}{1+\frac{m'_i}{2m_{i-1}}} + \frac{l'_i}{2m_{i-1}+m'_i}\\ &\leq \alpha_{i-1} + \frac{l_x + l - 1}{2^i m_0}\\ &\leq \alpha_{i-1} + \frac{\varepsilon}{2^i}\,.\end{aligned}$$

We have assessed

$$\alpha_{i-1} - \frac{\varepsilon}{2^i} \leq \alpha_i \leq \alpha_{i-1} + \frac{\varepsilon}{2^i}\,,$$

so that

$$\alpha_0 - \varepsilon\sum_{j=1}^{i} 2^{-j} \leq \alpha_i \leq \alpha_0 + \varepsilon\sum_{j=1}^{i} 2^{-j}.$$

And, because $\sum_{j=1}^{\infty} 2^{-j} = 1$, we can write

$$\alpha_0 - \varepsilon \leq \alpha_i \leq \alpha_0 + \varepsilon.$$

If we remember that $\alpha_0 = \alpha_\delta(u)$ and $\alpha_i = \alpha_{\delta+i}(u)$, we can conclude that the lemma is proved. □

Theorem 2.2. *If x is a restricted bi-ideal, and V_k denotes a prefix of x with length k, then*

$$\forall l \in \mathbb{N}\, \forall \varepsilon > 0\, \exists K \in \mathbb{N}\, \forall u \in A^* \, [|u| = l \Rightarrow \forall k \geq K \, |\alpha(V_K, u) - \alpha(V_k, u)| \leq \varepsilon].$$

P r o o f. We will use denotations similar to those, used in the proof of the lemma:

$$\begin{aligned} l_i &= |v_i|_u, && i \geq 0, \\ m_i &= |v_i| - l + 1, && i \geq 0, \\ \alpha_i &= \alpha(v_i, u) = \frac{l_i}{m_i}\,, && i \geq 0, \end{aligned}$$

where $l = |u|$, and (v_i) is the bi-ideal sequence associated with x as by Definition 1.4. Also, we assume that the bi-ideal is l_x restricted.

We select a parameter n, such that:

$$|\alpha_{n+i}(u) - \alpha_n(u)| < \frac{\varepsilon}{4}\,, \qquad \forall i \geq 1, \tag{3}$$

$$\frac{l_x + l}{m_n} < \frac{\varepsilon}{4}\,. \tag{4}$$

Note, that the first condition can be satisfied according to Lemma 2.1.

Then we select a parameter $g > n$, such, that

$$\frac{m_n + l_x + l}{m_g} < \frac{\varepsilon}{4}\,, \tag{5}$$

According to Lemma 2.1 and the way n and g were chosen

$$\alpha_n(u) - \frac{\varepsilon}{4} < \alpha_g(u) < \alpha_n(u) + \frac{\varepsilon}{4}\,.$$

We introduce a function $j : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, and denote

$$v_{g,h,n} = v_g u_{j(g,1)} v_n u_{j(g,2)} v_n \dots u_{j(g,h)} v_n\,,$$

such that

$$v_{g,h,n} \in Pref\, x\,.$$

Now we can introduce corresponding denotations:

$$\begin{aligned} l_{g,h,n} &= |v_{g,h,n}|_u\,, \\ m_{g,h,n} &= |v_{g,h,n}| - l + 1, \\ \alpha_{g,h,n} &= \frac{l_{g,h,n}}{m_{g,h,n}} = \alpha(v_{g,h,n}, u)\,. \end{aligned}$$

Considering the construction of $v_{g,h,n}$ and that the bi-ideal is l_x restricted,

$$m_g + h m_n \leq m_{g,h,n} \leq m_g + h m_n + h l_x + h l.$$

Let us assess $l_{g,h,n}$. It is obvious from the construction that

$$l_{g,h,n} \geq l_g + hl_n.$$

To assess the upper bound of $l_{g,h,n}$, we have to remember that the word $v_{g,h,n}$ has a total number of $m_{g,h,n}$ subsequences with length l, however the words v_g and v_n have correspondingly m_g and m_n subsequences with length l. This means, that the word $v_{g,h,n}$ has a maximum of

$$m_g + hm_n + hl_x + hl - m_g - hm_n = h(l_x + l)$$

subsequences of length l, such that we do not know if they are equal to u or not. Therefore,

$$l_{g,h,n} \leq l_g + h(l_n + l_x + l).$$

Now we can asses $\alpha_{g,h,n}$. We will start with the lower limit:

$$\alpha_{g,h,n} = \frac{l_{g,h,n}}{m_{g,h,n}} \geq \frac{l_g + hl_n}{m_g + hm_n + hl_x + hl} = \frac{\alpha_g m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl}$$

according to condition (3)

$$\begin{aligned} \frac{\alpha_g m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl} &\geq \frac{(\alpha_n - \frac{\varepsilon}{4})m_g + h\alpha_n m_n}{m_g + hm_n + hl_x + hl} \\ &= \alpha_n \frac{1}{1 + \frac{hl_x + hl}{m_g + hm_n}} - \frac{\varepsilon}{4} \frac{m_g}{m_g + hm_n + hl_x + hl}, \end{aligned}$$

according to condition (4)

$$\frac{hl_x + hl}{m_g + hm_n} \leq \frac{\varepsilon}{4},$$

therefore,

$$\alpha_n \frac{1}{1 + \frac{hl_x + hl}{m_g + hm_n}} - \frac{\varepsilon}{4} \frac{m_g}{m_g + hm_n + hl_x + hl} \geq \alpha_n \frac{1}{1 + \frac{\varepsilon}{4}} - \frac{\varepsilon}{4} \geq \alpha_n - \frac{\varepsilon}{2},$$

and

$$\alpha_{g,h,n} \geq \alpha_n - \frac{\varepsilon}{2}. \tag{6}$$

We now have to asses the upper limit of $\alpha_{g,h,n}$:

$$\alpha_{g,h,n} = \frac{l_{g,h,n}}{m_{g,h,n}} \leq \frac{l_g + hl_n + h(l_x + l)}{m_g + hm_n} = \frac{\alpha_g m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n}.$$

We can again use the condition (3)

$$\begin{aligned} \frac{\alpha_g m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n} &\leq \frac{(\alpha_n + \frac{\varepsilon}{4})m_g + h\alpha_n m_n + h(l_x + l)}{m_g + hm_n} = \\ &= \alpha_n + \frac{\varepsilon}{4} \frac{m_g}{m_g + hm_n} + \frac{h(l_x + l)}{m_g + hm_n}. \end{aligned}$$

According to the condition (4)

$$\frac{h(l_x+l)}{m_g+hm_n} \le \frac{\varepsilon}{4},$$

and, therefore

$$\alpha_{g,h,n} \le \alpha_n + \frac{\varepsilon}{2}. \tag{7}$$

Finally, let us look at V_i — a prefix of x with a length of $i > |v_g|$. We can find such an h that $V_i = v_{g,h,n}w$, where $w \in Pref(u_{j(g,h+1)}v_n)$.

Let us assess $\alpha(V_i,u) = \frac{|V_i|_u}{|V_i|-l+1}$.

It is clear from the way we selected i (and implicitly h) that

$$m_{g,h,n} \le |V_i| - l + 1 \le m_{g,h,n} + m_n + l_x + l,$$

also, it is obvious, that

$$l_{g,h,n} \le |V_i|_u \le l_{g,h,n} + m_n + l_x + l.$$

We can assess the lower limit for $\alpha(V_i,u)$:

$$\alpha(V_i,u) \ge \frac{l_{g,h,n}}{m_{g,h,n}+m_n+l_x+l} = \alpha_{g,h,n}\frac{1}{1+\frac{m_n+l_x+l}{m_{g,h,n}}}$$

according to the condition (5), and considering that $m_{g,h,n} \ge m_g$,

$$\frac{m_n+l_x+l}{m_{g,h,n}} \le \frac{\varepsilon}{4}$$

therefore,

$$\alpha_{g,h,n}\frac{1}{1+\frac{m_n+l_x+l}{m_{g,h,n}}} \ge \alpha_{g,h,n}\frac{1}{1+\frac{\varepsilon}{4}} \ge \alpha_{g,h,n} - \frac{\varepsilon}{4}.$$

Using (6) and (3)

$$\alpha(V_i,u) \ge \alpha_{g,h,n} - \frac{\varepsilon}{4} \ge \alpha_n - \frac{3\varepsilon}{4} \ge \alpha_g - \varepsilon.$$

The last thing we have to do is to assess the upper limit of $\alpha(V_i,u)$:

$$\alpha(V_i,u) \le \frac{l_{g,h,n}+m_n+l_x+l}{m_{g,h,n}} = \alpha_{g,h,n} + \frac{m_n+l_x+l}{m_{g,h,n}}.$$

According to the condition (5)

$$\frac{m_n+l_x+l}{m_{g,h,n}} \le \frac{\varepsilon}{4},$$

therefore,

$$\alpha_{g,h,n} + \frac{m_n+l_x+l}{m_{g,h,n}} \le \alpha_{g,h,n} + \frac{\varepsilon}{4},$$

and using (7) and (3)

$$\alpha(V_i, u) \leq \alpha_{g,h,n} + \frac{\varepsilon}{4} \leq \alpha_n + \frac{3\varepsilon}{4} \leq \alpha_g + \varepsilon\,.$$

So we can write:

$$\alpha_g - \varepsilon \ \leq \alpha(V_i, u) \leq \alpha_g + \varepsilon\,.$$

If we examine the conditions of the theorem, we see that it is proved, and that $K = |v_g|$. □

2.2. Making bi-ideals indistinguishable from i.i.d. bit-sequences in the frequency test

For a bit-sequence to be indistinguishable from i.i.d bit-sequences, each of the test words of a given length ν have to appear an equal number of times. Some deviations are, of course, permitted, depending on the statistical test we use to check this property.

The suggested method is as follows:

1. Choose any word of length $\nu - 1$. This word, denoted by a, will be a prefix for the generated base words.
2. The base words are found in the form ab, such that all of the test words of the length ν would appear an equal number of times in the word aba.

We will call the base words yielded by this method *good base words for a test length of* ν.

For example:

$$\underbrace{101}_{a}\underbrace{0010110000111}_{b}\underbrace{101}_{a}$$

We can see that each of the test words of the length 4 $(0000, 0001, 0010, \ldots, 1111)$ appear in the word aba exactly once. In this case $ab = 1010010110000111$ is a good base word for the test of the length 4. Although we do not currently have a precise estimate of the number of good test words, a full search reveals that there are 32 good base words with this prefix for the test of the length 4 that contain each of the test words exactly once, and 209952 that contain each of the test words exactly twice.

Lemma 2.3. *A restricted bi-ideal, generated from good base words for a test length of* ν*, will be indistinguishable from an i.i.d. bit-sequence, using test words with a length of* ν*, given a long enough bit-sequence.*

Proof. Let us consider the bi-ideal x. It can be written as:

$$x = u_0u_1u_0u_2u_0u_1\ldots$$

If we introduce a function $j : \mathbb{N} \to \mathbb{N}$, such that $j(i)$ is the index of the ith base word in the bi-ideal x. Then $j(0) = 0$, $j(1) = 1$, $j(2) = 0$, $j(3) = 2$, and so on.

Let us denote:

$$x_i = u_0 u_1 u_0 \dots u_{j(i)} \# Pref_{\nu-1}(u_{j(i+1)}) \,,$$

where $\mathrm{Pref}_n(u)$ denotes the prefix of the word u, of length n. Using the notation of the definition, if the word u would be expressed as ab, where a is a prefix of the length $\nu - 1$, it is clear, that $\forall i,\ Pref_{\nu-1}(u_i) = a$.

Let us consider x_0, x_1, and so on.

According to our definition of the good base words, and considering

$$\forall u_i, u_j \ \mathrm{Pref}_{\nu-1}\, u_i = \mathrm{Pref}_{\nu-1}\, u_j \,,$$

each test-word with a length of ν, will appear in x_0 an equal number of times.

Let us compare x_1 and x_0. It is obvious that each test word with the length ν will appear in x_1 the same number of times as it appears in x_0, plus as many times as it appears in the word $u_1 \# \mathrm{Pref}_{\nu-1}\, u_2$. But considering

$$\forall u_i, u_j \ \mathrm{Pref}_{\nu-1}\, u_i = \mathrm{Pref}_{\nu-1}\, u_j \,,$$

and that u_1 is a good base word as well. It becomes obvious that each of the test-words appears in x_1, and equal number of times as well. It is clear that this can be shown for any x_k in a similar fashion.

It is clear that any deviation from this occurs only when the prefix does not equal to one of the values of x_i. However, the maximum deviation for any given test-word will never exceed the maximum number of times the test-word appears in the longest base word of the bi-ideal. And thus, if we look at the relative frequency of the test-word, we see that it is inversely proportional to the length of the bi-ideal. This means, for a sufficiently long prefix of x, the bi-ideal will not be distinguishable from an i.i.d. bit-sequence using the frequency test with a test-word length of ν. □

Lemma 2.4. *A good base word for a test-length of ν is a good base word for all test-lengths smaller then ν.*

P r o o f. Suppose, we have a good base word for the test-length of ν. Obviously, $|u| = 2^\nu k$.

It is clear from the definition of the good base words that each of the test words with a length of ν appears in the word $u \# \mathrm{Pref}_{\nu-1}(u)$ exactly k times.

Let us consider test-words with a length of $\nu - 1$. If our assumption is correct, each of these test-words have to appear in $u \# \mathrm{Pref}_{\nu-2}(u)$ exactly $2k$ times. Let us assume the opposite, then there must be at least one test word v with a length of $\nu - 1$ that will appear in $u \# \mathrm{Pref}_{\nu-2}(u)$ at least $2k + 1$ times.

Let us examine each of the occurrences of v in $u \# \mathrm{Pref}_{\nu-2}(u)$. Considering that $u \# \mathrm{Pref}_{\nu-1}(u)$ is one bit longer then $u \# \mathrm{Pref}_{\nu-2}(u)$ we can look at each v plus the next bit. It is clear that this way we have constructed a word of the length ν, for each occurrence of v, will be a sub-word of $u \# \mathrm{Pref}_{\nu-1}(u)$.

But considering that we can only have a 0 or 1 following v, it is clear that either $v\#0$, or $v\#1$ will appear in the word $u\#\operatorname{Pref}_{\nu-1}(u)$ at least $k+1$ times. But this would mean that u is not a good base word for the length ν. Thus we have a contradiction.

It is clear, similarly, we can show the same for the lengths $\nu-2$, $\nu-3$, and so on. □

Theorem 2.5. *A restricted bi-ideal generated from good base words for the length ν will be indistinguishable from an i.i.d. bit-sequence using test words with a length of up to ν, given a long enough bit-sequence.*

P r o o f. The proof of the theorem obviously follows from the Lemmas 2.3, 2.4. □

REFERENCES

[1] BEAN, D. B.—EHRENFEUCHT, A. E.—McNULTY, G.: *Avoidable patterns in strings of symbols*, Pacific J. Math. **85** (1979), 261–294.
[2] CASSAIGNE, J.—FRID, A. E.: *On the arithmetical complexity of Sturmian words*, Theor. Comput. Sci. **380** (2007), 304–316.
[3] JACOBSON, N.: *Structure of Rings*, Amer. Math. Soc. Colloq. Publ. Vol. 37, AMS, Provicence R.I., 1956.
[4] LEVÉ, F.—RICHOMME, G. : *Quasiperiodic sturmian words and morphisms*, Theoret. Comput. Sci. **372** (2007), 15–25.
[5] LOTHAIRE, M.: *Algebraic Combinatorics on Words*, Encyclopedia Math. and Appl. Vol. 90, Cambridge University Press, Cambridge, 2002.
[6] DE LUCA, A.—VARRICCHIO, S.: *Finiteness and Regularity in Semigroups and Formal Languages*, Springer-Verlag, Berlin, 1999.
[7] NEUENSCHWANDER, D.: *Probabilistic and Statistical Methods in Cryptology*, Lecture Notes in Comput. Sci., Vol. 3028, Springer-Verlag, Berlin, 2004.
[8] NICOLAY, S.—RIGO, M.: *About frequencies of letters in generalized automatic sequences*, Theoret. Comput. Sci. **374** (2007), 25–40.
[9] RUEPPEL, R. A.: *Analysis and Design of Stream Ciphers*, Comm. Control Engrg. Ser. Vol. Xll, Springer, Berlin, 1986.
[10] SIMON, I.: *Infinite words and a theorem of Hindman*, Rev. Mat. Apl. **9** (1988), 97–104.

Received September 30, 2007

Department of Mathematical Analysis
Faculty of Physics and Mathematics
University of Latvia
Zeļļu iela 8
LV - 1002, Rīga
LATVIA
E-mail: edmunds.cers@gmail.com