THE CONSTRUCTION OF EFFICIENT STREAM CIPHERS AND CRYPTOGRAPHICALLY SECURE PSEUDO-RANDOM NUMBER GENERATORS

Dr. Jānis Buls, Dr. Jānis Valeinis, MSc. Inese Bērziņa, MSc. Līga Kuleša, MSc. Edmunds Cers

1. Introduction

While the beginnings of combinatorics on words can be traced back to the 19.th century, it is commonly held that the works of Axel Thue at the beginning of the twentieth century mark its emergence as a distinct branch of mathematics. Axel Thue studied square-free [55] and overlap-free [56] words. The first book solely concerned with combinatorics on words [36] appeared in 1983. Following a French tradition its authors wrote under a common pseudonym, in this case M. Lothaire. In a way this was a turning-point in the development of the subject. The number of publications began to rise significantly and today it is acknowledged as a separate subject in mathematics. Since 2000, combinatorics on words is included in the mathematical subject classification. Its code in the latest (MSC 2010) classification is 68R15 and it is classified as a branch of discrete mathematics in relation to computer science. A lot of interesting facts about the beginnings and history of combinatorics on words can be read in [3] and [29].

The main research object in combinatorics on words are words, i.e. finite of infinite sequences of symbols from some (usually finite) set called the alphabet of the word.

2. Machine invariant classes

Right infinite words (ω-words) are a natural generalization of finite words. Just like the notion of a line is used in geometry, despite nobody actually having seen such a thing, so the notion of infinite words is used in combinatorics on words. By using this notion the theory is made more straightforward. This creates fertile grounds for research into areas related to, for example, cryptography.

Definition 1. (also see [54]). *A three-sorted algebra $\langle \mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D} \rangle$ is called a cryptosystem if*

- *$\mathcal{P}$ is a finite set, called the set of plaintexts,*
- *$\mathcal{C}$ is a finite set, called the set of ciphertexts,*
- *$\mathcal{K}$ is a finite set, called the set of keys,*
- *$\mathcal{E}: \mathcal{P} \times \mathcal{K} \longrightarrow \mathcal{C}$ is a total map, called a cipher,*
- *$\mathcal{D}: \mathcal{P} \times \mathcal{K} \longrightarrow \mathcal{P}$ is a total map, called the deciphering map,*

and every plaintext $x \in \mathcal{P}$ and every key $k \in \mathcal{K}$ satisfies

$$\mathcal{D}(\mathcal{E}(x,k),k) = x.$$

From there a cipher is only a step away [80].

Definition 2. *A four-sorted algebra $\langle X, S, Y, K, z, f, g, h \rangle$ is called a cipher, if*

- *X is a finite alphabet of plaintexts,*
- *S is a finite set of cipher states,*
- *Y is a finite alphabet of cyphertexts,*
- *K is a finite set of keys,*

- $z: K \longrightarrow S$, $f: S \times K \times X \longrightarrow K$, $g: S \times K \times X \longrightarrow S$, $h: S \times K \times X \longrightarrow Y$ *are total maps.*

Definition 3. *A three-sorted algebra* $\langle Q, A, B; \circ, * \rangle$ *is called a Mealy machine, if* Q, A, B *- finite nonempty sets,* $Q \times A \xrightarrow{\circ} Q$ *and* $Q \times A \overset{*}{\twoheadrightarrow} B$. *The set* Q *is called the set of internal states of the machine, while* A *and* B *are called the input and output alphabets, respectively. The elements of the sets* A *and* B *are called letters. The operations* $\circ$ *and* $*$ *are called the input and output functions, respectively.*

Notice that a cipher can be considered a special case of a Mealy machine. In this way Mealy machines enter cryptography. The model of Mealy machines [39] has been studied since the 1950ies (see, e.g., [20, 27, 45, 78, 79]).

Here we will describe a single example of a symmetric cryptosystem (see Fig. 1).

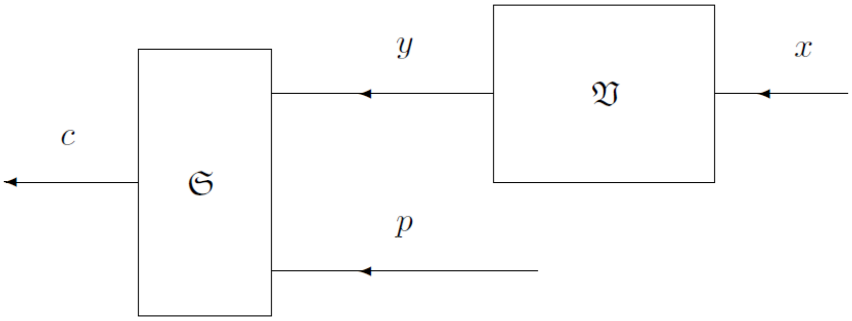


Fig. 1 Symmetric cryptosystem

Assume $\mathfrak{S}$, $\mathfrak{V}$ are devices representing addition modulo two and the corresponding Mealy machine $V = \langle Q, A, \{0,1\}, \circ, * \rangle$. All users have identical devices. Both the sets of plaintexts and ciphertexts are $\{0,1\}^*$. The persistent key is a word $x \in A^{\omega}$. At the start of each session a session key is selected, that is a pair $n \in \mathbb{N}$, $q \in Q$ Now the sender calculates

$$y = q * x[n, n + l]$$

where $l + 1$ is the length of the plaintext

$p = p_0 p_1 \dots p_l$.

The encryption is addition modulo two, i.e.,

$c_i = p_i + y_i (\text{mod } 2)$.

Clearly, the security of such a cryptosystem is significantly dependent on both the chosen Mealy machine V and the persistent key $\in A^{\omega}$. The described cryptosystem serves as additional motivation why we study transformations of infinite words by Mealy machines. If we were instead to limit ourselves to finite words, then the only thing we could conclude would be that for every pair $u, v \in A^n$ there is a Mealy machine transforming u to v. Therefore in the set of finite words A^* a classification of this sort would be trivial.

A typical area of research in combinatorics on words are classes of words (languages) and their properties. We explore a classification in the class of ω-words. This classification unites into a class all words that are “equally complex” in relation to transformations by Mealy machines. As a result the studied hierarchy classifies words by their complexity. Let us remark, that from a computing standpoint a Mealy machine is a much simpler model than a Turing machine. Consequently our proposed classification is much finer (less coarse), because a class contains only the words that are “equally complex” in relation to transformations by Mealy machines. Let's make the notion of this hierarchy precise.

A three-sorted algebra $\langle Q, A, B; q_0, \circ, * \rangle$ is called an *initial Mealy machine*, if $q_0 \in Q$ and $\langle Q, A, B; \circ, * \rangle$ is a Mealy machine. Let us assume that

$$V = \langle Q, A, B; q_0, \circ, * \rangle$$

is an initial Mealy machine and that $(x, y) \in A_0^\omega \times B_0^\omega$ where $A_0 \times B_0 \subseteq A \times B$. We will write $y = q_0 * x$ or $x \rightharpoondown_V y$ when

$$\forall n, y[0, n] = q_0 * x[0, n].$$

In this case we will say that V *transforms* x to y.

Assume that K is a set, then $\mathrm{Fin}(K) = \{G | G \subseteq K \wedge |G| < \aleph_0\}$. Assume that $\mathfrak{A} = \{a_0, a_1, \ldots, a_n, ..\}$ and $\mathfrak{Q} = \{q_1, q_2, \ldots, q_n, \ldots\}$ are some countable sets and

$\mathfrak{M} = \{\langle Q, A, B; q_0, \circ, * \rangle | Q \in \mathrm{Fin}(\mathfrak{Q}) \wedge A, B \in \mathrm{Fin}(\mathfrak{A})\}$.

Let's define a relation $\rightharpoondown$ in the set $\mathfrak{N} = \{x \in A^\omega | A \in \mathrm{Fin}(\mathfrak{A})\}$ that is true if and only if there is a Mealy machine $V = \mathfrak{M}$, such that $x \rightharpoondown_V y$.

Lemma 4. *The model* $\langle \mathfrak{N}, \rightharpoondown \rangle$ *is a preorder.*

Now let's derive the canonical order. Define the equivalence relation $(x \sim y) \Leftrightarrow (x \rightharpoondown y) \wedge (y \rightharpoondown x)$.

Proposition 5. *The model* $\langle \widetilde{\mathfrak{N}}, \rightharpoondown \rangle$ *where* $\widetilde{\mathfrak{N}} = \mathfrak{N}/\sim$ *and* $\rightharpoondown$ *come from the relation* $\rightharpoondown$ *in the set* $\mathfrak{N}$ *factorized by* $\sim$ *is an order.*

The following two theorems were demonstrated by Belovs [1].

Theorem 6. *The ordered set* $\widetilde{\mathfrak{N}}$ *is a join-semilattice.*

Theorem 7. *The ordered set* $\widetilde{\mathfrak{N}}$ *is not a meet semilattice.*

The join-semilattice $\widetilde{\mathfrak{N}}$ we have just defined allows us to introduce the concept of machine invariant set, that is, the lattice of ideals $\mathfrak{L}$ of this partial order describes the machine invariant classes.

In other terms this can be explained as follows. We say that a word $x \in A_1^\omega$ is suited for the Mealy machine $V = \langle Q, A, B; q_0, \circ, * \rangle$, if $A_1 \subseteq A$.

Definition 8. *Assume that* $\mathfrak{K} \neq \emptyset$ *is some fixed set of ω-words. The class* $\mathfrak{K}$ *is called machine invariant, if every Mealy machine transforms every suited word of the class* $\mathfrak{K}$ *into a word also of class* $\mathfrak{K}$*.*

The usual term used in the literature in such cases is that the class $\mathfrak{K}$ is closed (under transformations by Mealy machines). Closure properties of classes have been traditionally studied in combinatorics on words in relation to automata theory. We list some of the most important results.

We start with ultimately periodic words. Assume that $v \in A^+$. We denote by v^ω the ω-word $v^\omega = vv \ldots v \ldots$. The word v^ω is called a *periodic word*. A word x is called *ultimately periodic*, if there exists words $u \in A^*$ and $v \in A^+$, such that $x = uv^\omega$. In this case the length of v denoted by $|v|$ is called the *period*, while $|u|$ is called the *pre-period* of x.

Theorem 9. (Jablonskis [77]). *Any ultimately periodic word is transformed into an ultimately periodic word by every Mealy machine.*

Assume that $[n]_k$ is the notation of the number n in the numeral system

$$\Sigma_k = [0, 1, \ldots, k - 1\}$$

written with the least number from left, that is, when

$$n = \sum_{i=0}^{m} a_i k^i,$$

then

$$[n]_k = a_0 a_1 \ldots a_m.$$

Definition 10. *A word* $x = (x_n) \in A^\omega$ *is called k-automatic if there is an initial Mealy machine*

$V = \langle Q, A, B; q_0, \circ, * \rangle,$

such that $\forall n, x_n = q_0 \hat{*} [n]_k$. *A word* $y \in A^\omega$ *is called automatic if there is a* k *such that the word is k-automatic.*

Theorem 11. (Cobham [17]) *The class of automatic words is closed.*

A mapping $f: S \to S'$ is called a *morphism (homomorphism)* of the semigroup S if

$$\forall x \forall y\, f(xy) = f(x)f(y).$$

A morphism f is called an *epimorphism* if it is surjective; an injective morphism is called a *monomorphism;* a bijective morphism is called a *isomorphism.* If $S = S'$ then the morphism is called an *endomorphism.* A bijective endomorphism is called an *automorphism.*

A semigroup morphism $f: M \to M'$ is called a *monoid morphism (homomorphism)* if $f(\lambda) = \lambda'$ where λ and λ' are the respective neutral elements of the monoids M and M'.

Assume that K is some fixed arbitrary set. We define the iterations of a mapping $g: K \to K$ inductively:

1. $g^0 = \mathbb{I}$;
2. $g^{n+1} = gg^n$.

Definition 12. *A morphism* $\phi: A^* \to A^*$ *is called nonerasing if* $\phi(A^+) \subseteq B^+$.

Assume that $\phi: A^* \to A^*$ is a nonerasing morphism for which there is a letter $a \in A$ such that

$\phi(a) = au,$ where $a \in A^+$.

Then it follows that for all $n \geq 0$

$$\phi^{n+1}(a) = \phi(au) = \phi^n(a)\phi^n(u).$$

Therefore $\phi^n(a)$ is a prefix of $\phi^{n+1}(a)$ and the sequence $(\phi^n(a))$ converges to a limit denoted by $\phi^\omega(a)$, that is

$$\phi^\omega(a) = \lim_{n \to \infty} \phi^n(a).$$

In this case we say that $\phi^\omega(a)$, is the ω-word obtained by iterating the morphism ϕ on the letter a.

A morphism $\tau: A^* \to B^*$ is called a *coding* if

$$\forall a \in A\, |\tau(a)| = 1.$$

Definition 13. *An ω-word x is called a morphic word when there is a letter* a, *a morphism* $\phi: A^* \to A^*$ *and a coding* $\tau: A^* \to B^*$ *such that* $x = \tau(\phi^\omega(a))$.

Theorem 14. *(Dekking [22]) Any morphic word transformed by a Mealy machine is still a morphic word.*

Definition 15. *A word* $x \in A^\omega$ *is called recurrent if any factor of* x *enters* x *an infinite number of times. A word uy is called ultimately recurrent if y is recurrent.*

Theorem 16. (Buls [8]) *Any ultimately recurrent word transformed by a Mealy machine is still an ultimately recurrent word.*

As noted before $\widetilde{\mathfrak{N}}$ is a join-semilattice. The join in this semilattice is $[(x_i)] \vee [(y_i)] = [(x_i, y_i)]$. We now consider the algebraic properties of the semilattice $\widetilde{\mathfrak{N}}$.

Definition 17. *A join-semilattice* $\langle D, \leq \rangle$ *is called distributive if*

$$\forall xab\, (x \leq a \vee b \Rightarrow \exists a'b'(a' \leq a \,\&\, b' \leq b \,\&\, x = a' \vee b')).$$

The join-semilattice is called modular if

$$\forall □ ab\, (a \leq x \leq a \vee b \Rightarrow \exists b' \leq b(x = a \vee b')).$$

We have shown that $\widetilde{\mathfrak{N}}$ is not distributive [10] and by further developing the proof technique we showed also [11]:

Theorem 18. *The join-semilattice* $\langle \widetilde{\mathfrak{N}}, \to \rangle$ *is not modular.*

The former result directly follows from this theorem. The following proposition is the main step of the proof of Theorem 18 ([11]):

Proposition 19. *If* ${}^{f}x \rightharpoondown y$, ${}^{\zeta}x \rightharpoondown y$ *and*

$$\forall \varkappa\, \exists ak\; f(k) \leq a^{\zeta} < (a+\varkappa)^{\zeta} \leq f(k+1)$$

then y is ultimately periodic.

Here

$${}^{\zeta}x(n) = \begin{cases} 1, & \mathit{if}\ \exists k \in \mathbb{N}\; n = k^{\zeta}; \\ 0, & \mathit{otherwise}. \end{cases}$$

and

$${}^{f}x(n) = \begin{cases} 1, & \mathit{if}\ \exists k \in \mathbb{N}\; n = f(k); \\ 0, & \mathit{otherwise}. \end{cases}$$

We considered the fact that during encryption it will not always be the case that a letter will be transformed into a letter. Quite often a single letter is encrypted into a whole word. This motivated us to study transducers alongside Mealy machines as well. A transducer is a generalization of a Mealy machine.

Definition 20. *A two-sorted algebra $\langle A, B; \circ\rangle$ is called a polygon when Q, A are finite sets and $\circ$ is a mapping $Q \times A \xrightarrow{\circ} Q$.*

A three-sorted algebra $\langle Q, A, B^{*}; \circ, *\rangle$ is called a transducer if

- $\langle A, B; \circ\rangle$ is a polygon;
- B is a non-empty finite set;
- $Q \times A \xrightarrow{*} B^{*}$.

For transducers, similarly to Mealy machines, the mappings $\circ$ and $*$ are expanded to the set $Q \times A^{*}$:

$$\begin{aligned} q \circ \lambda &= q, \quad q \circ ua = (q \circ u) \circ a; \\ q * \lambda &= \lambda, \quad q * ua = (q * u) \# (q \circ u) * a; \\ q \mathbin{\hat{*}} \lambda &= \lambda, \quad q \mathbin{\hat{*}} ua = (q \circ u) * a; \end{aligned}$$

for all $(q, u, a) \in Q \times A^{*} \times A$.

It is quite obvious where to begin. We can use the same constructions that were useful in the study of words using Mealy machines in the case of transducers as well. This field of studies is developed by Līga Kuleša.

A three-sorted algebra $\langle Q, A, B^{*}; q_0, \circ, *\rangle$ is called an *initial transducer* if $q_0 \in Q$ and $\langle Q, A, B^{*}; \circ, *\rangle$ is a transducer.

Assume that

$$T = \langle Q, A, B^{*}; q_0, \circ, *\rangle,$$

is an initial transducer $(x, y) \in A_0^{\omega} \times B_0^{\omega}$, where $A_0 \times B_0 \subseteq A \times B$. We write $y = q_0 * x$ or $x \rightharpoondown_T y$ when

$$\forall n\; y[0, n] = q_0 * x[0, n].$$

In this case we say that the transducer T *transduces* the word x to y.

Assume that K is a set. Then $\mathrm{Fin}(K) = \{G | G \subseteq K \wedge |G| < \aleph_0\}$. Assume that $\mathfrak{A} = \{a_0, a_1, \dots, a_n, ..\}$ and $\mathfrak{Q} = \{q_1, q_2, \dots, q_n, \dots\}$ are some fixed countable sets and $\mathfrak{M}_T = \{\langle Q, A, B^{*}; q_0, \circ, *\rangle | Q \in \mathrm{Fin}(\mathfrak{Q}) \wedge A, B \in \mathrm{Fin}(\mathfrak{A})\}$.

We define the relation $x \rightharpoondown y$ in the set $\mathfrak{N}_T = \{x \in A^{\omega} | A \in \mathrm{Fin}(\mathfrak{A})\}$ to be true if and only if there is a transducer $T \in \mathfrak{M}_T$, such that $x \rightharpoondown_T y$.

Lemma 21. *The model $\langle \mathfrak{N}_T, \rightharpoondown\rangle$ is a preorder.*

Now we can derive a canonical order. Define the equivalence relation $(x \sim y) \Leftrightarrow (x \rightharpoondown y) \wedge (y \rightharpoondown x)$.

Proposition 22. *The model $\langle \widetilde{\mathfrak{N}}_T, \rightharpoondown\rangle$ where $\widetilde{\mathfrak{N}}_T = \mathfrak{N}_T/\sim$ and $\rightharpoondown$ derived from $\rightharpoondown$ in the set $\mathfrak{N}_T$ using a factorization by $\sim$, is an order.*

Already in this stage it turns out that it is more productive to restrict our research to morphic mappings, since every transduction $\tau: A^{\omega} \rightarrow B^{\omega}$ that can be performed by an initial transducer can be expressed as a composition of a initial Mealy machine and a morphism (see, [46]). That is to say that there exists a Mealy machine $\langle Q, A, C; q_0, \circ, * \rangle$ and a morphism $\mu: C^* \rightarrow B^*$ such that

$\tau(x) = \mu(q_0 * x)$.

Similarly as in the case of Turing machines (the correspondence we have in mind: a set X is reducible by Turing machine to set Y), we can consider the reducibility of a ω-word x to the ω-word y using a morphism. This correspondence defines a preorder in the set of ω-words. Considering the big role played by morphisms in describing ω-words we considered the partial order of ω-words. We note that finitely generated bi-ideals can be represented as morphic words as well. In lieu of this a study of morphic words is a natural development of the previous studies.

Define a relation $x \rightharpoondown y$ in the set $\mathfrak{N}_{\mu} = \{x \in A^{\omega} | A \in \mathrm{Fin}(\mathfrak{A})\}$ that is true if and only if there is a morphism μ such that $\overset{\mu}{\rightharpoondown} y$.

Lemma 23. *The model $\langle \mathfrak{N}_{\mu}, \rightharpoondown \rangle$ is a preorder.*

Now we can derive the canonical order. We define a equivalence relation $(x \sim y) \Leftrightarrow (x \rightharpoondown y) \wedge (y \rightharpoondown □)$.

Proposition 24. *The model $\langle \widetilde{\mathfrak{N}}_{\mu}, \rightharpoondown \rangle$ where $\widetilde{\mathfrak{N}}_{\mu} = \mathfrak{N}_{\mu}/\sim$ and $\rightharpoondown$ are derived from the relation*

$\rightharpoondown$ in the set $\mathfrak{N}_{\mu}$ using factorization by $\sim$, is an order.

The obtained order $\widetilde{\mathfrak{N}}_{\mu}$ is significantly different from $\widetilde{\mathfrak{N}}$.

Theorem 25. *The ordered set $\widetilde{\mathfrak{N}}_{\mu}$ is neither a join-semilattice nor a meet-semilattice.*

This theorem indirectly underscores the special role of the ordering $\widetilde{\mathfrak{N}}$ in the classification of ω-words by their computational complexity.

3. Morphisms and Simulation

The simulation was first discussed by Hartmanis [27] more than forty years ago. This concept describes the possibility on abstract level in which one machine could be replaced by another one in applications, for example, cryptography, especially, cryptanalysis of cryptographic devices. If we like to treat as it is done till now the machines by semigroups and develop the theory not only as self-sufficient discipline the connections between simulation and semigroups should be considered from every point of view too. Thus we say that a transition from machines to semigroups through some representation is *successful* if it adequately characterizes the simulation. We had demonstrated [9] how the concept of simulation can be integrated in the theory of semigroups

Definition 26. *Let $V = \langle Q, A, B \rangle$, $'V = \langle 'Q, 'A, 'B \rangle$ be Mealy machines. We say that $'V$* simulates *V by*

$$Q \xrightarrow{h_1} {}'Q, A \xrightarrow{h_2} {}'A, {}'B \xrightarrow{h_3} B$$

if the diagram

$$\begin{tikzcd}
Q \arrow[d, "h_1"'] & \times & A^* \arrow[r, "*"] \arrow[d, "h_2"] & B^* \\
'Q & \times & 'A^* \arrow[r, "*"] & 'B^* \arrow[u, "h_3"']
\end{tikzcd}$$

commutes. That is, if
$q * u = h_3(h_1(q) * h_2(u))$ *for all* $(q, u) \in Q \times A^*$.
This concept corresponds to such scheme.

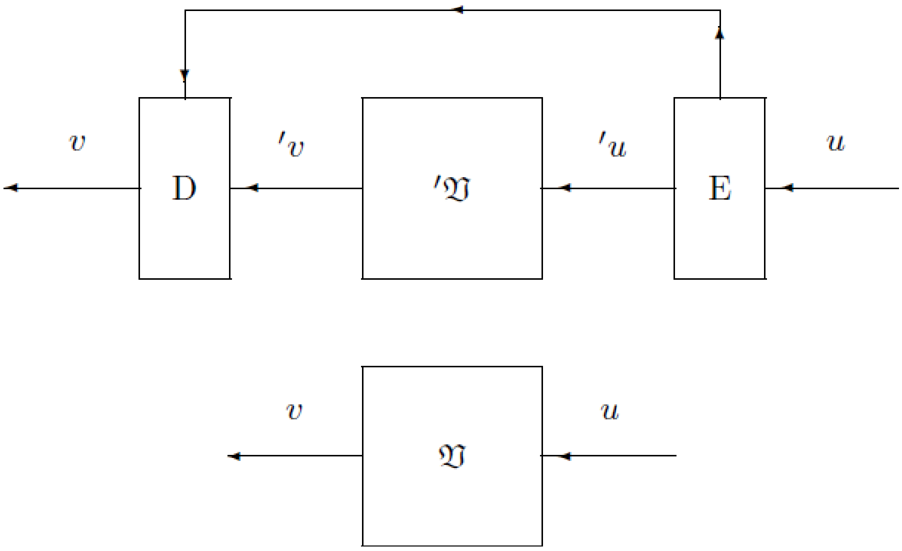


Fig. 2. Scheme of a simulation

This scheme enables to extend the notion of simulation [76].

Definition 27. *Let* $V = \langle Q, A, B \rangle$, $'V = \langle 'Q, 'A, 'B \rangle$ *be machines. We say that* $'V$ simulates V *by*

$$Q \xrightarrow{h_1} 'Q, A \xrightarrow{h_2} 'A^*, 'B^* \xrightarrow{h_3} B$$

if

$q \circ u * a = h_3(h_1(q) \circ h_2(u) * h_2(a))$ *for all* $(q, u, a) \in Q \times A^* \times A$.

Obviously, now the upper tie from encoder to decoder is necessary. Otherwise the decoder is not able to decode the word $'v$ adequately. We write $'V \geq V(h_1, h_2, h_3)$ if $'V$ simulates V by h_1, h_2, h_3. We say $'V$ *simulates* V if there exist maps h_1, h_2, h_3 such that $'V \geq V(h_1, h_2, h_3)$. We write $'V \geq V$ if $'V$ simulates V.

The two machines V and $'V$ are *incomparable* if $V \ngeq 'V$ and $'V \ngeq V$. If, on the other hand, $V \geq 'V$ and $'V \geq V$ then we say that V *mutually simulates* $'V$ and we write $V \bowtie 'V$.

This definition has an interesting consequence.

Example 28. V_1 *mutually simulates* $'V_1$.

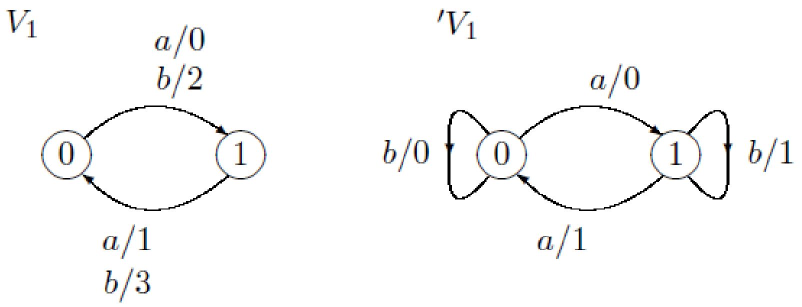


Fig. 3. $V_1 \bowtie 'V_1$

- $V_1 \geq 'V_1(h_1', h_2', h_3')$, where

$$h_1' : 0 \mapsto 0, 1 \mapsto 1;$$

$$h_2' : a \longmapsto a, b \longmapsto a^2;$$
$$h_3' : 0 \longmapsto 0, 1 \longmapsto 1, 01 \longmapsto 0, 10 \longmapsto 1 .$$

- $'V_1 \geq V_1(h_1, h_2, h_3)$, where

$$h_1 : 0 \longmapsto 0, 1 \longmapsto 1;$$
$$h_2 : a \longmapsto a, b \longmapsto a^3;$$
$$h_3 : 0 \longmapsto 0, 1 \longmapsto 1, 010 \longmapsto 2, 101 \longmapsto 3.$$

We recall the classical approach to the representation of finite machines by semigroups (see, e.g., [45]). Let $V = \langle Q, A, B; \circ, * \rangle$ be a Mealy machine, where Q, A, B are finite, nonempty sets; $Q \times A \xrightarrow{\circ} Q$ is a function and $Q \times A \xrightarrow{*} B$ is a surjective function. Let $T(Q)$ denotes the semigroup of all transformations on the set Q and let $Fun(Q, B)$ denotes the set of all maps from Q to B. On the set $S(Q, B) = T(Q) \times Fun(Q, B)$ we define the multiplication by

$$(g_1, \psi_1)(g_2, \psi_2) = (g_1 g_2, \psi_1 \psi_2);\ g_1 g_2 \in T(Q), \psi_1 \psi_2 \in Fun(Q, B).$$

Under this operation $S(Q, B)$ is easily seen to be a semigroup. Let $Q = \{q_1, q_2, \dots, q_k\}$, $A = \{a_1, a_2, \dots, a_m\}$, $B = \{b_1, b_2, \dots, b_n\}$. Define two mappings $A \xrightarrow{\alpha} T(Q)$ and $A \xrightarrow{\beta} Fun(Q, B)$ as follows. For each $a_i \in A$ define $\alpha(a_i) \in T(Q)$ and $\beta(a_i) \in Fun(Q, B)$ by

$$\alpha(a_i) = \begin{pmatrix} q_1 q_2 \dots q_k \\ q_1' q_2' \dots q_k' \end{pmatrix}$$
$$\beta(a_i) = \begin{pmatrix} q_1 q_2 \dots q_k \\ b_1' b_2' \dots b_k' \end{pmatrix}$$

where
$\forall s \ (q_s' = q_s \circ a_i \wedge b_s' = q_s * a_i.$

Now the representation $A \xrightarrow{\eta} S(Q, B)$ is defined by setting $\eta(a_i) = (\ \alpha(a_i), \beta(a_i))$. The semigroup $\langle V \rangle$ generated by $\eta(A)$ is called the *machine V semigroup*. So far [45].

We generalize the concept of similar transformation semigroups (see, e.g., [32]) to machine semigroups as follows. Let

$$\sigma = (\alpha, \beta) \in S(Q, B)$$

then we define a vector function of the machine

$\bar{\sigma}: Q \to Q \times B: q \longmapsto (\alpha(q), \beta(q)).$

The same denotation we use for a vector function $'Q \to {}'Q \times {}'B$.

Definition 29. *Let* $V = \langle Q, A, B; \circ, * \rangle$, $'V = \langle \ \ 'Q, 'A, 'B; \acute{\circ}, \acute{*} \rangle$ *be machines. We say that* $\langle V \rangle \xrightarrow{\psi} \langle 'V \rangle$ *is the* s-morphism *of machine semigroup* $\langle V \rangle$ *to* $\langle 'V \rangle$ *if there exists maps* $Q \xrightarrow{g} {}'Q$,
$B \xrightarrow{h} {}'B$, *such that the diagram*

$$\begin{tikzcd}
Q \arrow[r, "\bar{\sigma}"] \arrow[d, "g"'] & Q \arrow[d, "g"'] & \times & B \arrow[d, "h"] \\
'Q \arrow[r, "\overline{\psi(\sigma)}"] & 'Q & \times & 'B
\end{tikzcd}$$

commutes for every $\sigma \in \langle V \rangle$.

If h is an injection, then s-morphism is called the *injective s-morphism*. If g is a surjection, then s-morphism is called the *surjective s-morphism*.

Theorem 30. [13] *If there exists injective s-morphism* $\langle V \rangle \longrightarrow \langle W \rangle$ *then* W *simulates* V.

Theorem 31. [9] *Every surjective s-morphism* $\langle V \rangle \longrightarrow \langle □ \rangle$ *is a homomorphism of semigroups.*

Theorem 32. [9] *There exists the surjective s-morphism such that it does not induce the simulation.*

Mealy machines serve as models of cryptographic devices. This means that theory of simulation of Mealy machines concerns to cryptographic resistance to attack on cryptographic devices. New concept of s-morphism would lead us to better understanding of simulation.

4. Chaos and Cryptographic Generators

Random number generators are essential in many applications including stochastic simulations, statistical experiments, modeling of probabilistic algorithms and generation of secure stream chippers. Truly random sequences can only be generated by physical processes. Unfortunately, in practise it is very difficult to construct physical generators that are both fast and reliable, therefore, the most convenient and reliable way of generating the random numbers for stochastic simulations appears to be via deterministic algorithms. These algorithms are based on solid mathematical basis (see, e.g., [33]) and produce sequences that are not random, but seem to behave as if the bits were chosen independently at random.

Definition 33. *A pseudo-random generator is a two-sorted algebra* $\mathfrak{G} = \langle Q, B, q_0, T, G \rangle$, *where* Q *is a finite set of states,* $q_0 \in Q$ *is the initial state or seed, the mapping* $Q \xrightarrow{T} Q$ *is the transition function,* B *is finite set of symbols, and* $Q \xrightarrow{G} B$ *is the output function.*

In fact, this model is specialized Moore machine. The generator evolves according to the recurrence $q_n = T(q_{n-1})$, for $n = 1,2,3 \ldots.$ At the step n the generator outputs the symbol $b_n = G(q_n)$. The finiteness of state space Q implies the ultimate periodicity of sequence of states q_n, therefore this approach is limited. We are interested in generation of infinite aperiodic sequences, i.e., sequences that are not ultimately periodic.

4.1. Chaos

One method for obtaining aperiodic sequences is to use a logistic map – the simplest chaotic system. In 1982 Oishi and Inoue [47] proposed the idea of using chaos in designing a pseudo-random generator and later in 1992 Sandri [51] introduced a simple aperiodic pseudo-random generator which was based on logistic map. In 2000 Buls [7] proposed a construction of symmetric cryptosystem keys that are based on chaos. For more information of using chaotic systems in generation of pseudo random sequences, see, e.g., [42], [43]. Recently, Hu et.al. [28] introduced a true random number generator by combining congruential methods with prime numbers and higher order composition of logistic maps. It generates a 256-bit random number by computer mouse movement.

We showed [6] the construction of new chaotic maps in symbol space.

4.2. Shrinking Generator and Bi-ideals

In 1993 Coppersmith et. al. [18] introduced a new pseudo-random number generator called „shrinking generator'' (further in the text – SG). It uses two periodic pseudo-random bit-sequences (A-sequence is the source sequence and S-sequence is the „Selector”) to create a third sequence: SG deletes the term a_i from A-sequence if the i^{th} term in S-sequence equals 0, in other words, resulting sequence Z consists of terms of A-sequence which correspond to ones in S-sequence. Periodicity of A-sequence and S-sequence implies the periodicity of Z-sequence. The main idea of the construction is based on the fact that the resulting pseudo-random sequence is of a better quality – it is harder for cryptanalyst to find it. As of today, this approach is considered as one of the most perspective and secure in construction of symmetric ciphers.

The main idea of our work – how can we replace a periodic sequence in SG with an aperiodic sequence that would return an aperiodic shrunk sequence (Z-sequence) with good statistical properties, which potentially could be used as chipper in symmetric cryptography.

We considered differential sequences, residually ultimately periodic sequences, ultimately periodic sequences modulo p (for some $p \in \mathbb{N}_+$), ultimately periodic sequences threshold t (for some $t \in \mathbb{N}_+$), differentially residually ultimately periodic sequences. At last we decided to consider recurrent words.

We use less known equivalent definition of recurrent words [19]. By $\mathrm{F}^\infty(x)$ we denote the set of all factors of a word x that occurs in it infinitely many times. The fact that x is recurrent can be expressed with $\mathrm{F}^\infty(x) = \mathrm{F}(x)$. Moreover, if $x = uy$ is recurrent and y is recurrent suffix of x, then $\mathrm{F}^\infty(x) = \mathrm{F}(y)$, therefore, all recurrent suffixes of x has exactly the same factors.

Definition 34. *A sequence of finite words $v_0, v_1, \dots, v_n, \dots$ is called a bi-ideal sequence if $v_{i+1} \in v_i A^* v_i$for all $i \in \mathbb{N}$.*

The terms of a bi-ideal sequence are also known as Zimin’s words [75] and sesquipowers [52].

Lemma 35. *A sequence of finite words $v_0, v_1, \dots, v_n, \dots$ is a bi-ideal sequence if and only if there exists a sequence of finite words $u_0, u_1, \dots, u_n, \dots$ such that*

$$v_0 = u_0$$
$$v_{i+1} = \square_i u_{i+1} v_i$$

Corollary 36. *If (v_n) is a bi-ideal sequence, then*

$$\forall m \leq n \;\; v_m \in \mathrm{Pref}(v_n) \cap \mathrm{Suff}(v_n).$$

Let $(u_i)_{i \in \mathbb{N}}$ be a sequence of finite words over finite alphabet A such that u_0 is non-empty word. Then by Lemma 35 we can define inductively a bi-ideal sequence(v_i):

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

Definition 37. *The limit of the bi-ideal sequence $x = \lim_{i \to \infty} v_i$ is called a bi-ideal. In this case we say that (u_i) generates x or that x is the bi-ideal generated by a sequence (u_i). The bi-ideal x is called l-bounded if $|u_i| \leq l$ for all $i \in \mathbb{N}$. We say that bi-ideal x is bounded if there exists integer l such that x is l-bounded. The bi-ideal x is called finitely generated if*

$$\exists m \forall i \forall j \left(i \equiv j (\operatorname{mod} m) \Rightarrow u_i = u_j \right).$$

In this case we say that m-tuple $\langle u_0, u_1, \dots, u_{m-1} \rangle$ generates the bi-ideal x or that x is the bi-ideal generated by $\langle u_0, u_1, \dots, u_{m-1} \rangle$. We also say in this case that m-tuple $\langle u_0, u_1, \dots, u_{m-1} \rangle$ is the basis of the bi-ideal x.

Lemma 38. A word *$x \in A^\omega$ is recurrent if and only if it is a bi-ideal.*

It turns out almost all infinite words over finite alphabet are bi-ideals. In fact, nearly every randomly chosen infinite word is a bi-ideal, i.e., the measure of the set of all bi-ideals equals 1, which serves as additional motivation for a research on bi-ideals.

This year (15th of June) Edmunds Cers defended his Ph.D. thesis „Finitely Generated Bi-ideals and the Semilattice of Machine Invariant ω-languages”. The main result of his thesis effectively solves a decision problem: given two bases $\langle u_1, u_2, \ldots, u_n \rangle$ and $\langle u'_1, u'_2, \ldots, u'_m \rangle$, decide whether they generate the same bi-ideal.

In our construction of aperiodic SG we decided to replace periodic S-sequence with a finitely generated bi-ideal since it can be effectively generated and the sufficient condition of aperiodicity of a finitely generated bi-ideal is known [12]:

Proposition 39. *If $\bigcup_{i=0}^{m-1} \mathrm{Pref}(u_i)$ or $\bigcup_{i=0}^{m-1} \mathrm{Suff}(u_i)$ contains at least two words with the same length, then bi-ideal with basis $u_0, u_1, \ldots, u_n, \ldots$ is aperiodic.*

In fact, Proposition 39 can be conveniently used to generate aperiodic finitely generated bi-ideals: in order to obtain an aperiodic finitely generated bi-ideal it is sufficient to make simple restrictions on two terms of the basis.

We have shown ([2], Proposition 3.4.) that for each periodic infinite binary word x (which contains both ones and zeros) taken as A-sequence there exist infinitely many aperiodic finitely generated bi-ideals y that can be taken as S-sequence so that the resulting Z-sequence is aperiodic.

The structure of a finitely generated bi-ideal (i.e., periodical repetition of basis words in generation of a bi-ideal and existence of infinitely many terms of the bi-ideal sequence that are congruent modulo p, where $p \in \mathbb{N}$ is arbitrary ([2], Lemma 3.2.)) played important role in the proof of the result.

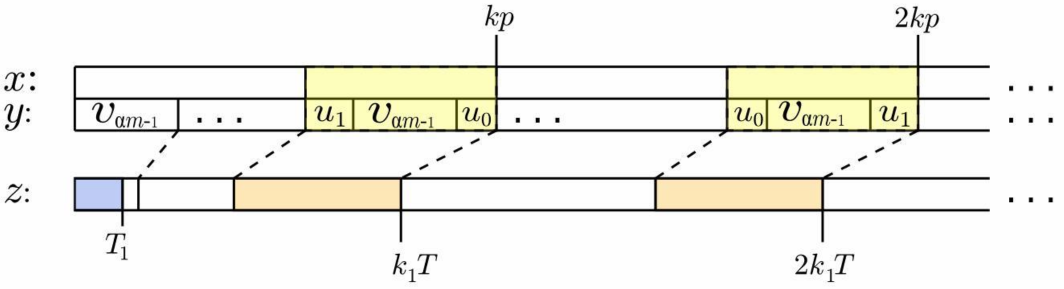


Fig. 4. The structure of SG with periodic A-sequence (□) and aperiodic S-sequence (finitely generatebi-ideal y)

On the one hand, this structure provides us possibility for each periodic A-sequence to construct infinitely many finitely generated bi-ideals so that whichever of these bi-ideals is taken as □-sequence, the resulting Z-sequence is aperiodic. On the other hand, we had to make sure that this structure does not affect statistical properties of resulting Z-sequence, therefore, we did testing with Diehard battery of tests – our modified SG passed all tests in Diehard test suit.

The construction has a shortage – the choice of a b-ideal (S-sequence) depends on randomly chosen periodic word (A-sequence). It would be more convenient if both A-sequence and S-sequence could be chosen arbitrary (randomly), therefore, considered the existence problem of universal bi-ideals.

Definition 40. *A bi-ideal y (taken as S-sequence) is called universal if for each non-trivial binary periodic word x taken as A-sequence the resulting Z-sequence is aperiodic.*

Here periodic binary word is called non-trivial if it contains both zeros and ones. We proved the existence of infinitely many universal bi-ideals [2]:

Proposition 41. *Let $m \in \mathbb{N}$, $m \geq 2$. If $u_0 = 1$, $u_1 = 10$ and $00 \notin \mathrm{F}(u_i)$ for all $i \in \{2,3,\ldots,m-1\}$, then finitely generated bi-ideal with basis $\langle u_0, u_1, \ldots, u_{m-1} \rangle$ is universal.*

One more motivation for possible use of finitely generated bi-ideals in cryptography: if characteristic sequence of a filter is an aperiodic finitely generated bi-ideal, then the filter does not preserve regularity of a language, i.e., potentially such filter can transform relatively simple language into more complex language.

4.3. Aperiodicity Problem of all Arithmetical Subsequences of a Bi-ideal

After conference SYNASC italian mathematician Marco Bodrato suggested to consider a problem that would be appreciated by cryptographers – to characterize S-sequences such that Z-sequence is strongly non-periodic.

Definition 42. *An infinite word x is called weakly periodic if there exist two integers $p > 0$ and $l \geq 0$ such that for each non-negative integer n the following condition holds*

$x[l] = x[l + n \cdot p]$.

We say that x is strongly non-periodic if it is not weakly periodic.

In particular, a sequence (word) is strongly non-periodic if all its arithmetical subsequences and the word itself are non-periodic. It was worth to analyze only infinite arithmetical subsequences since classical Van der Waerden theorem [73] states that foe each infinite word $x = x_0x_1x_2 \dots x_n \dots$ over finite alphabet Σ there exist arbitrarily long arithmetical progressions $k, k + p, \dots, k + np$ such that $x_k = x_{k+p} = \dots = x_{k+np}$.

During her doctoral studies I.Bērziņa found necessary and sufficient condition of aperiodicity of all arithmetical subsequences of a binary finitely generated bi-ideal:

Theorem 43. *All arithmetical subsequences of a finitely generated bi-ideal □ $\in \{0,1\}^{\omega}$ are aperiodic if and only if there exists basis $\langle u_0, u_1, \dots, u_{m-1} \rangle$ with $m > 1$ such that positions of zeros and ones form complete residue system modulo $k = \big||u_0| - |u_1|\big| > 0$ in the word u_0.*

In [35] Lorencs proved that each finitely generated bi-ideal has countably many bases with the same number of basis words, hence, the fact that a given basis does not satisfy the condition of Theorem 43 does not imply the existence of a periodic arithmetical subsequence of the given word. Here we use Lorencs' construction for a base change of a finitely generated bi-ideal.

Proposition 44. *If x is finitely generated bi-ideal with basis $\langle u_0, u_1, \dots, u_{m-1} \rangle$, then m-tuple $\langle u'_0, u'_1, \dots, u'_{m-1} \rangle$, where $u'_i = u_0u_s$ and $s = i + 1 \bmod m$, also is basis of x.*

In this case, we will say that the basis words of the bi-ideal x are L-prolonged or simply that the basis of the bi-ideal $x$is L-prolonged. If x is a finitely generated bi-ideal with basis $\langle u_0, u_1, \dots, u_{m-1} \rangle$, then for all $n > 0m$-tuple

$\langle u_0^{(n)}, u_1^{(n)}, \dots, u_{m-1}^{(n)} \rangle$,

where $u_i^{(n)} = u_0^{(n-1)} u_{i+1 \bmod m}^{(n-1)}$ is the basis of the basis of the bi-ideal x after n iterations of L-prolongation. The result by Cers [15], which states that each finitely generated bi-ideal has exactly one irreducible basis, implies that in our case L-prolongation is the only useful way to change the basis of the bi-ideal. Thus, if a given basis of a bi-ideal does not satisfy the conditions of Theorem 43, then after some number iterations of L-prolongation one may obtain basis that satisfies the conditions of the Theorem 43. Here the following arose: how many times do we have to L-prolong basis words to check whether all arithmetical subsequences of x are aperiodic or not? We started to work on efficiency of this algorithm after I.Bērzina rejoined the project. Up to now we have proved that the periodicity of the sequence which is formed of the lengths modulo k (for arbitrary $k \in \mathbb{N}_+$) of terms of a bi-ideal sequence which converges to a finitely generated bi-ideal:

Lemma 45. *Let $m \geq 2$and $k \in \mathbb{N}_+$. If x is finitely generated bi-ideal with basis $\langle u_0, u_1, \dots, u_{m-1}\rangle$ then there exist infinitely many numbers $\alpha, s \in \mathbb{N}_+$ such that*

$$|v_{\alpha m-1+j}| \equiv |v_{(\alpha+sn)m-1+j}| \bmod k.$$

for all $n \in \mathbb{N}$ and for all $j \in \overline{1, sm}$.

This result let us conjecture: if during a period (full cycle of length s□) the set of all positions of zeros (ones) modulo k in new obtained terms of a bi-ideal sequence remains the same, then in the next period (cycle of length sm) it will also remain the same.

Conjecture 46. *Let $m \geq 2$ and $k \in \mathbb{N}_+$. Let x be a finitely generated bi-ideal with basis $\langle u_0, u_1, \dots, u_{m-1}\rangle$ and let K_i^0 (respectively, K_i^1) be the set of all positions of zeros (respectively, ones) modulo k in the i^{th} term of the bi-ideal sequence of x. If there exist positive integers α, s such that*

1. $\forall n \in \mathbb{N}, j \in \overline{1, sm}\big(|v_{\alpha m-1+j}| \equiv |v_{(\alpha+sn)m-1+j}| \bmod k\big)$,
2. $K_{\alpha m-1}^0 = K_{(\alpha+s)m-1}^0 \wedge K_{\alpha m-1}^1 = K_{(\alpha+s)m-1}^1$,

then

$$K_{(\alpha+s)m-1}^0 = K_{(\alpha+2s)m-1}^0 \wedge K_{(\alpha+s)m-1}^1 = K_{(\alpha+2s)m-1}^1.$$

We also solved the aperiodicity problem of all arithmetical subsequences of a binary bounded bi-ideal.

Theorem 47. *All arithmetical subsequences of a bounded bi-ideal $x \in \{0,1\}^\omega$ are aperiodic if and only if there exists a basis $(u_n)_{n\geq 0}$ such that*

1. *positions of zeros and ones form complete residue system modulo $k = \big||u_0| - |u_1|\big|$ in the word u_0.*
2. *there exists infinite sequence of positive integers $i_0, i_1, \dots, i_n, \dots$ such that*
   $k = |k_{i_0}| = |k_{i_1}| = \dots = |k_{i_n}| = \cdots$, *where* $k_j = |u_j| - u_{j+1}$ *for all* $j \in \mathbb{N}$.

In case of bounded bi-ideals the algorithm is similar to the algorithm used in case of finitely generated bi-ideals. We only have to make restriction that the difference $k$occurs infinitely often, otherwise we can easily construct counterexample:

Example 48. *If x is a bi-ideal generated by the sequence*

$$01, 111, 111, 111, 111, \dots$$

Then one can see that positions of both zeros and ones form complete residue system modulo $k = |2-3| = 1$ in word $u_0 = 01$. However, if we L-prolong basis words once, then we obtain basis

$01111, 01111, 01111, 01111, \dots,$

which generates periodic word with period 5, therefore x contains arithmetical subsequence

$x_5^0 = 000\dots = 0^\omega$.

The previous example serves as explanation why it is worthless to consider efficiency of the algorithm in case of bounded bi-ideals – for arbitrary sequence (u_i) it is impossible to determine the number of iterations of L-prolongation after which all differences of finite number of occurrences will disappear. In case of finitely generated bi-ideals we do not have such problem since all differences repeats periodically infinite number of times.

During the project we also initiated the research on connection between universal bi-ideals and aperiodicity of all arithmetical subsequences of a finitely generated bi-ideal.

Proposition 49. *Universal bi-ideals do not contain arithmetical subsequence* $x_p^l = 0^\omega$.

Proposition 50. *If an universal bi-ideal satisfies the conditions of Proposition 41., then all its arithmetical subsequences are aperiodic.*

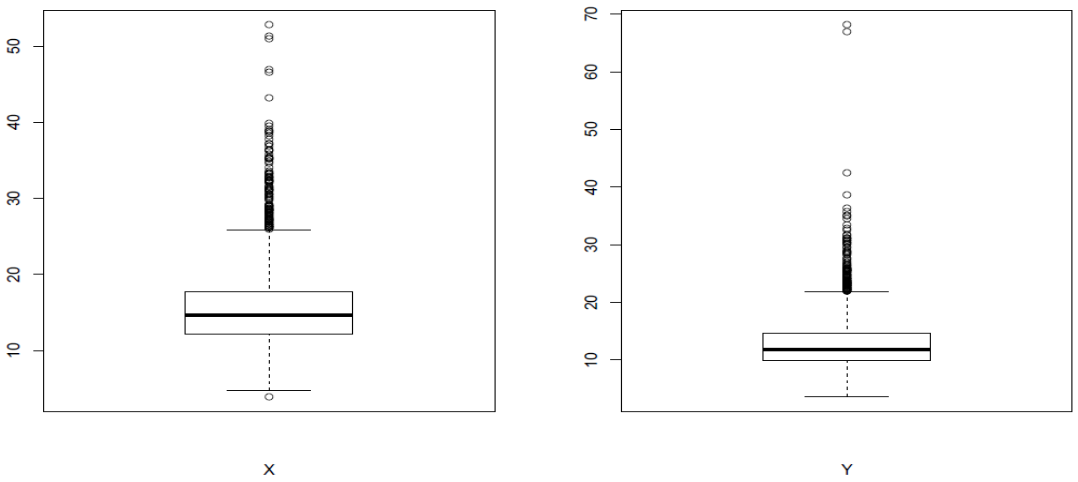
5. Empirical likelihood function

During the project one of the main goals was the analysis of the empirical likelihood (EL) method in the two-sample case. The empirical likelihood method is a nonparametric statistical method introduced by Owen [48] in 1988. This method can be used to make statistical inference such as the construction of confidence intervals, hypothesis testing, and the estimation of parameters. The EL method is especially attractive due to the asymmetry of confidence intervals and the Bartlett correctability (see [49]). In 2000 Qin and Zhao [50] introduced the EL method for the difference of two univariate parameters. Later the EL method was developed for ROC curves (Claeskens et al. [16], 2003), the quantile difference in the one-sample case (Zhou and Jing [74], 2003).

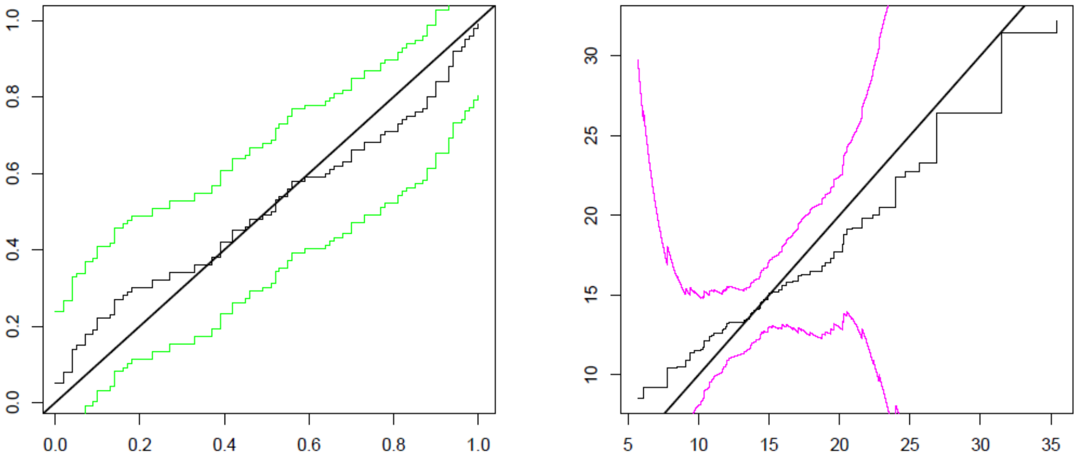
We were able to extend the two-sample empirical likelihood method for different two-sample problems in a general form. The general setup includes, for example, the well-known probability-probability (P-P) and quantile-quantile (Q-Q) plots, ROC curves, the difference of two means, distribution and quantile functions. In collaboration with Dr. Edmunds Cers we have produced several papers ([57] and [58]). We have also produced the package EL in program R, which is available online [59]. Our results have been presented in several conferences ([60], [61]). One of the most interesting applications is the construction of simultaneous confidence bands for the probability-probability and quantile-quantile plots. This is an alternative way for testing the hypothesis about the equality of two distribution functions.

For the illustration consider the data problem in [26] where the number of death due to prostate cancer has been analyzed in the USA. With the introduction of the PSA screening test in the early to mid 1990s, the number of deaths due to prostate cancer has dramatically gone down. Its effectiveness can be measured by comparing the mortality rates due to prostate cancer before and after introduction of the test. They obtained the rates of prostate cancer deaths in the USA (by county) for the two year-groups: 1990–1992 and 1999–2001 (see respective boxplots in Figure 1).

To check the simple location model in [18] the empirical process approach was used. We subtracted the estimated location parameter from the second sample. Thus we reduced the problem to the hypothesis testing of the equality of both distributions. By boxplots and confidence bands for P-P, Q-Q plots we conclude that the location model can not be rejected at the 5% significance level.



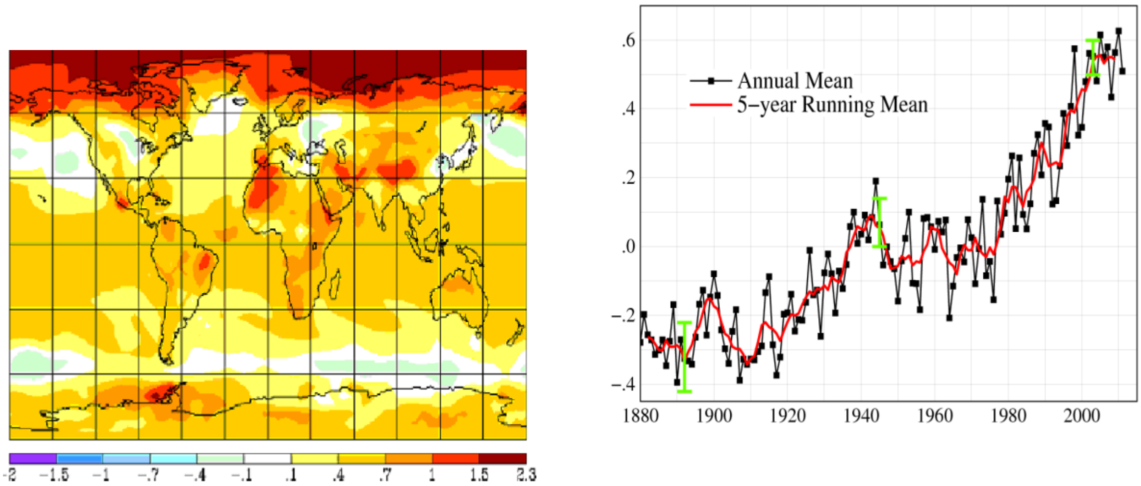
1. Figure. Boxplots comparing prostate cancer mortality rates in the USA during 1990–1992 (left plot) and 1999–2001 (right plot)



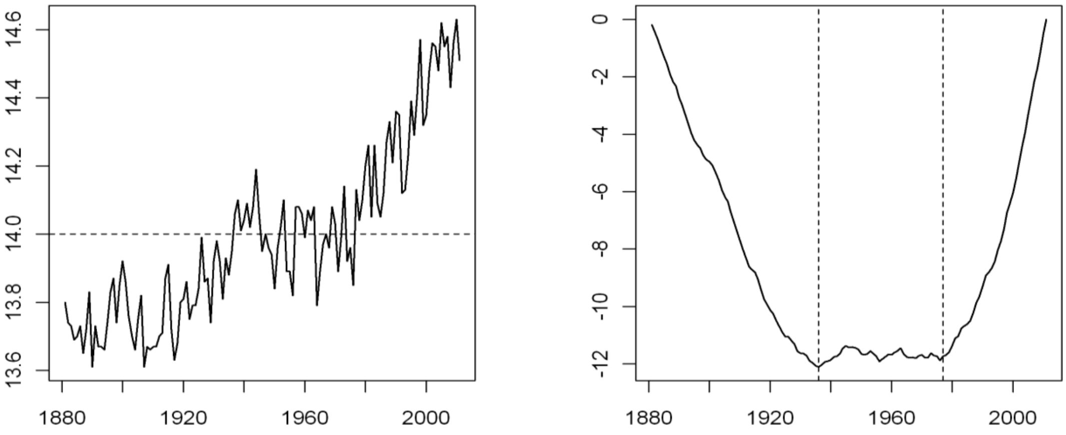
2. Figure. Probability (P-P) and quantile-quantile (Q-Q) plots with confidence bands for prostate cancer data example is analyzed in [18]. After subtracting the location parameter from the second sample, we reduced the problem of checking the location model to the hypothesis testing of equality of both distributions. The estimated location parameter was 2.7. The diagonal fits into the bands, thus the location model can not be rejected.

During the project we analyzed the EL method also for dependent observations. Mostly the financial data or data analyzed in econometric problems are dependent having short or long memory. The most popular are ARIMA and GARCH models, which are special processes from larger class of dependent processes. The general weak dependence can be described by mixing processes. Kitamura [1997] in 1997 introduced the blockwise empirical likelihood method for weakly dependent processes. The main idea is to divide observations into blocks, to use the smoothing by estimating equations and then to apply the usual EL method for the reduced sample. In the one-sample setting we compared recently introduced confidence intervals based on the Bernstein's inequality with the EL method. In collaboration with PhD student Sandra Vucāne we have produced a paper [62] and these results have been presented in several conferences ([71], [72]). For dependent observations we managed to introduce the EL method for general two-sample problems. One of the practical applications is the change-point analysis for time series data. The simplest model is the location model where the mean change is analyzed.

Consider for illustration the global warming issue and temperature change during 1890 – 2000 (see Figure 3). It seems that there is a global warming, but when these changes occur? Are there many change-points for such time series data? The Global Land-Ocean temperature index (1880-2011) characterizes global warming problem (see Figure 3). Applying the usual CUSUM algorithm we can see that there might be too significant change points – in 1936 and 1978 years (see Figure 4). On the other hand this algorithm does not tell us whether these change-points are statistically significant.



3. Figure. Left plot: Global warming data: the map of temperature change (1890 - 2000), right plot: Global Land – Ocean temperature index (1880-2011), which shows the relative change of the temperature around the global mean, which is 14.0 degrees of Celsius (corresponds to the value 0 in plot).



4. Figure. Left plot: Global Land – Ocean temperature index (1880-2011), right plot: CUSUM algorithm for detecion of change point indicates two change-points in 1936 and 1978 years.

Introducing the EL method for dependent observations in two-sample case, we have solved the change-point detection problem in the following manner. We proposed to divide the time series data into two parts with some window parameter. Then we plot the p-value graph, obtained by the two-sample hypothesis test about the equality of two means, as a function of midpoint of both time series data. We have presented our results in the 8th World Congress on Statistics and Probability theory [63].

During this project the EL method has been analyzed and applied also in robust statistics. In collaboration with prof. Dr. George Luta from USA and PhD student Mara Velina we have prepared a publication manuscript [64], that will be submitted to the international journal *Test*. At present we are working also at the two trimmed mean difference. We have presented our results in several conferences on robust statistics: ICORS 2011 and ICORS 2012 ([65] and [70]).

Finally, we have to mention the hypothesis about the two-sample location model, which also can be tested by the EL method. In medical statistics this model is of great interest when introducing a new drug or comparing the impact of the new and old drugs. General structural relationship models have been treated by Valeinis [66]. In collaboration with the master student Lidija Januševa we analyze the general shift function by the empirical processes and the EL methods (see [61] and [67]).

Next, we analyzed the Neyman and Bickel-Rosenblatt goodness-of-fit tests for dependent observations. For independent observations there exist many statistical test statistics, the most famous are the Kolmogorov-Smirnov, Anderson-Darling, chi-squared test statistics. Ledwina [34] in 1994 proposed to use the Schwarz's selection criterion for the dimension choice in the Neyman test statistic introduced already in 1937. After this result the Neyman smooth test became very popular. In collaboration with prof. Dr. Axel Munk from the University of Goettingen and prof. Dr. Jean-Pierre Stockis from the University of Kaiserslautern, we have produced a publication manuscript [41], where the Neyman test has been analyzed for dependent observations. In statistical literature one can find only one more test statistic which can be used for general dependent processes – the Bickel Rosenblatt test. This test statistic has several remarkable properties – it can be used for simple and composite hypothesis, for independent and dependent processes without any modification. On the other hand the Bickel-Rosenblatt test practically is difficult to apply due to the smoothing parameter, which has to be estimated. We have published a paper [68] in the international journal *Mathematical modeling and analysis*, where we provide some recommendations how to use this test in practice (see also [69]). Recently we have started to work also with long memory data and goodness-of-fit tests, where we plan to collaborate with prof. Dr. Donatas Surgailis from Lithuania.

Finally, we have to mention the collaboration with prof. Dr. Andrejs Cebers. First, we analyzed the smoothing of spectrum with statistical methods. It appears that the Brownian motion spectrum can be smoothed by the nonparametric statistical methods. The simultaneous confidence bands can be added to check whether the Brownian motion has the right slope. Second, we analyzed the uniform band foundation for some bacteria using some statistical testing procedures (see [14]). We wish also to mention some seminars about mathematical statistics which were offered to physicists during the project:

1. Basics in mathematical statistics and hypothesis testing (12.03.2010);
2. Goodness-of-fit tests (7.05.2010 and 14.05.2010);
3. Nonparametric statistical methods with applications to the time series forecasting, (21.05.2010, 28.05.2010).

Publications

1. **J.Buls.** (2012). *Machine Morphisms and Simulation*. World Academy of Science, Engineering and Technology, Vol.68, 1200-1203.
2. **J.Valeinis, A.Ločmelis**. (2012). *Bickel-Rosenblatt test for weakly dependent data*, Mathematical modelling and analysis, 17(3), 383-395.
3. **I.Bērziņa, R.Bēts, J.Buls, E.Cers, L.Kuleša**. (2011). *On a Non-periodic Shrinking Generator*. SYNASC 2011, 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. Proceedings. IEEE Computer Society, 348-356.
4. **A.Munk, J.P.Stockis, J.Valeinis, G.Giese**. (2011). *Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data*. Annals of the Institute of Statistical Mathematics 63(5), 939-959.
5. **J.Valeinis, E.Cers**. (2011). *EL: Two-sample Empirical likelihood*. R-package version 1.0, http://CRAN.R-project.org/package=EL.
6. **I.Bula, J.Buls, I.Rumbeniece**. (2011). *On new chaotic mappings in symbol space*, Acta Mech Sin 27(1), 114-118, DOI 10.1007/s10409~011-0408-1.
7. **J.Buls, E.Cers**. (2010). *Distributivity in the semilattice of ω −words*. Contributions to General Algebra 19, Proceedings of the Olomouc Conference 2010 (AAA79 + CYA25), Verlag Johannes Heyn, Klagenfurt, 13-22.
8. **J. Buls, E. Cers**. (2010). *Modularity in the Semilattice of ω −words*. 13th Mons Theoretical Computer Science Days, September 6 - 10, 2010. Proceedings, Université de Picardie Jules Verne, 1-10.
9. **J.Valeinis, E.Cers, J.Cielens**. (2010). *Two-sample problems in statistical data modelling*. Math. Model. Anal 15, 137-151.

Conferences

1. **J.Buls.** „ Machine Morphisms and Simulation”. World Academy of Science, Engineering and Technology. Stockholm, Sweden, July 11-12 (2012).
2. **J.Valeinis**. „Two-sample blockwise empirical likelihood for weakly dependent data with applications to change-point analysis”. 8th World Congress in Probability and Statistics, Istanbul, Turkey, July 9-14 (2012).
3. **M.Velina, J.Valeinis, G.Luta**. „Empirical Likelihood-based Methods for the Difference of Two Trimmed Means”. International conference on Robust Statistics (ICORS) 2012, Vermont, USA, August 5-10 (2012).
4. **I.Bērziņa, R.Bēts, J.Buls, E.Cers, L.Kuleša**. „On a Non-periodic Shrinking Generator”. 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. Timisoara, Romania, September 26-29 (2011).
5. **J.Valeinis**. „Two-sample plug-in empirical likelihood method with applications to structural relationship models”. 58th World Statistics Congress (ISI2011), Dublin, Ireland, August 21-26 (2011).
6. **J.Valeinis, M.Velina, G.Luta**. „Empirical likelihood-based inference for the difference of smoothed Huber estimators”. International conference on Robust Statistics (ICORS) 2011, Valladolid, Spain, 27 June – 1 July (2011).
7. **J.Buls, E.Cers**. „The semilattice of ω −words”. 79th Workshop on General Algebra, 25th Conference for Young Algebraists. Olomouc, Czech Republic, February 12 - 14 (2010).
8. **J.Buls, E.Cers**. „Modularity in the semilattice of ω −words”. 13th Mons Theoretical Computer Science Days. Amiens, France, September 6 - 10 (2010).

Bibliography

[1] A.BELOVS. (2008). *Some Algebraic Properties of Machine Poset of Infinite Words*. J. RAIRO - Theoretical Informatics and Aplications, **42**, 451 - 466.

[2] I.BĒRZIŅA, R.BĒTS, J.BULS, E.CERS, L. KULEŠA. (2011). On *a Nonperiodic Shrinking Generator*. SYNASC 2011, 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. Proceedings. IEEE Computer Society, 348 - 356.

[3] J.BERSTEL AND D.PERRIN. (2007). *The Origins of Combinatorics on Words*. European Journal of Combinatorics, **3** Vol. 28, 996 - 1022.

[4] J.BERSTEL AND J.KARHUMÄKI. (2003). *Combinatorics on Words - A Tutorial.* Bulletin of the European Association for Theoretical Computer Science, **79**, 178 - 228.

[5] J.R.BÜCHI. (1960). *Weak Second-order Arithmetic and Finite Automata.* Mathematical Logic Quarterly 6, No. 1-6, 66 - 92.

[6] I.BULA, J.BULS AND I.RUMBENIECE. (2011*). On new chaotic mappings in symbol space*, Acta Mech Sin **27**(1), 114 - 118, DOI 10.1007/s10409 011-0408-1

[7] J.BULS. (2000). *Construction of Pseuodo-random Sequence from Chaos*. 2nd International Confrenece: Control of Oscillations and Chaos. Proceedings, Vol. 3, 558 - 560.

[8] J.BULS. (2003). *Machine Invariant Classes*. Proceedings of WORDS'03, 4th International Conference on Combinatorics on Words, September 10 - 13, 2003, Turku, Finland, Tero Harju and Juhani Karhumäki (Eds.), TUCS General Publication (No 27, August), 207 - 211.

[9] J.BULS. (2012). *Machine Morphisms and Simulation*. World Academy of Science, Engineering and Technology, Vol. 68, 1200 – 1203.

[10] J.BULS AND E.CERS. (2010). *Distributivity in the semilattice of ω-words*. Contributions to General Algebra 19, Proceedings of the Olomouc Conference 2010 (AAA79 + CYA25), Verlag Johannes Heyn, Klagenfurt, 13 - 22.

[11] J.BULS AND E.CERS. (2010). *Modularity in the Semilattice of ω-words.* 13th Mons Theoretical Computer Science Days, September 6 - 10, 2010. Proceedings, Université de Picardie Jules Verne, 1 - 10.

[12] J.BULS AND A.LORENCS. (2008). *From Bi-ideals to Periodicity.* RAIRO-Theoretical Informatics and Applications, **3** Vol. 42, 467 - 475.

[13] J.BULS AND I.ZANDERE. (2004). *Injective Morphisms of the Machine Semigroups*. Contributions to General Algebra **14**, Proceedings of the Olomouc Conference 2002 (AAA64) and the Potsdam Conference 2003 (AAA65). Verlag Johannes Heyn, Klagenfurt, 15 - 19.

[14] A.CEBERS, K.ERGLIS, D.ZHULENKOVS, M. BELOVS AND J. VALEINIS. (2012). *Band formation by magnetotactic spirillum bacteria in oxygen concentration gradient* (submitted).

[15] E. CERS. (2010). *An unique basis representation of _nitely generated bi-ideals.* Proceedings of the 13th Mons theoretical computer science days (JM 2010), Universite de Picardie Jules Verne.

[16] G.CLAESKENS, B.Y. JING, L. PENG, AND W. ZHOU. (2003). *Empirical likelihood confidence regions for comparison distributions and ROC curves.* Can. J. Stat., **31**, 173 - 190.

[17] A.COBHAM. (1972). *Uniform tag sequences.* Theory of Computing Systems Vol. 6, No. 1, 164 - 192. .

[18] D.COPPERSMITH, H.KRAWCZYK AND Y.MANSOUR. (1994). *The Shrinking Generator*. Proceedings of the 13th Annual International Cryptology Conference on Advances in Cryptology, CRYPTO '93, Springer-Verlag, London, UK, 22 - 39.

[19] M.COUDRAIN AND M.P.SCHÜTZENBERGER. (1966). *Une Condition de Finitude des Monoides Finiment Engendres*. CR Acad. Sci., Paris, Ser. A Vol. 262, 1149 - 1151.

[20] J.DASSOW. (1981). *Completeness Problems in the Structural Theory of Automata*. Mathematical Research (Band 7), Akademie-Verlag, Berlin.

[21] B.A.DAVEY, H.A.PRIESTLEY. (2002). *Introduction to Lattices and Order.* Cambridge University Press.

[22] F.M.DEKKING. (1994). *Iteration of maps by an automaton*. J.Discrete Math., **126**, 81 - 86.

[23] V. DIEKERT AND M. KUEITNER. (2011). *Fragments of First-order Logic Over Infinite Word*. Theory of Computing Systems 48, 486 - 516.

[24] A.O.GELFOND. (1968). *Sur les nombres qui ont des propriétés additives et multiplicatives données.* Acta Arith. Vol.13, 259 - 265.

[25] J.A.GOGUEN. (1967). *L-fuzzy sets.* J. Math. Anal. Appl., Vol. 8, 145 - 174.

[26] K.GHOSH AND R.TIWARI. (2007). *Empirical process approach to some two-sample problems based on ranked set samples.* Annals of the Institute of Statistical Mathematics, **59**, 757 - 787.

[27] J.HARTMANIS AND R.E.STEARNS. (1966). *Algebraic Structure Theory of Sequential Machines.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[28] Y.HU, X.LIAO, K.-W. WONG AND Q.ZHOU. (2009). *A True Random Number Generator Based on Mouse Movement and Chaotic Cryptography.* Chaos Solitons and Fractals, Vol. 40, No. 5, 2286 - 2293.

[29] J.KARHUMÄKI. (2004). *Combinatorics on Words: A New Challenging Topic.* TUCS Technical Report **645**, Turku Centre for Computer Science.

[30] Y.KITAMURA. (1997*). Empirical likelihood methods with weakly dependent processes*. Annals of Statistics, **25**(5), 2084 - 2102.

[31] S.C.KLEENE AND E.L.POST. (1954). *The upper semi-lattice of degrees of recursive unsolvability*. The Annals of Mathematics 59, No. 3, 379 - 407.

[32]G.LALLEMENT. (1979). *Semigroups and Combinatorial Applications.* John Wlley & Sons, New York, Chichester, Brisbane, Toronto.

[33] P. L'ECUYER. (1998). *Random Number Generation*. Handbook on Simulation, editor Jerry Banks, Chapter 4. Wiley.

[34] T.LEDWINA. (1994). *Data-Driven Version of Neyman's Smooth Test of Fit.* Journal of the American Statistical Association, **3 89**(427), 1000 - 1005.

[35] A.LORENCS. (2012). *The Identity Problem of Finitely Generated Biideals.* Acta Informatica Vol. 49, No. 2, 105 - 115.

[36] M.LOTHAIRE. (1983). *Combinatorics on Words. Encyclopedia of Mathematics and its Applications*, Vol. 17, Addison-Wesley, Reading, Massachusetts.

[37] M.LOTHAIRE. (2002). *Algebraic combinatorics on Words.* Encyclopedia of Mathematics and its Aplications, Vol. 90, Cambridge University Press, Cambridge.

[38] A.DE LUCA AND S.VARRICCHIO. (1999). *Finiteness and Regularity in Semigroups and Formal Languages*. Springer-Verlag, Berlin, Heidelberg.

[39] G.H.MEALY. (1955). *A method for synthesizing sequential circuits.* Bell System Technical Journal 34, No. 5, 1045 - 1079.

[40] M.MORSE AND G.A.HEDLUND. (1942). *Symbolic dynamics II. Sturmian trajectories.* American Journal of Mathematics 62, No. 1, 1 - 42.

[41] A.MUNK, J.P.STOCKIS, J.VALEINIS, G.GIESE. (2011). *Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data.* Annals of the Institute of Statistical Mathematics, **63**(5), 939 - 959.

[42] V.PATIDAR, K.K.SUD AND N.K.PAREEK. (2009). *A pseudo random bit generator based on chaotic logistic map and its statistical testing.* Informatica **4** Vol. 33, Slovenian Society Informatika, 441 - 452.

[43] D.PERRIN AND J.E.PIN. (2002). *Infinite words*. Elsevier/Academic Press.

[44] S.C.PHATAK AND S.SURESH RAO. (1995). *Logistic Map: A Possible Random Number Generator*. Physical Review E, **4** Vol. 51, The American Physical Society, 3670 - 3678.

[45] B.I.PLOTKIN, I.JA.GREENGLAZ AND A.A.GVARAMIJA. (1992). *Algebraic Structures in Automata and Databases Theory*. World Scientific, Singapore, New Jersey, London, Hong Kong.

[46] Y.PRITYKIN. (2006). *Strongly Almost Periodic Sequences under Finite Automata Mappings*. http://arxiv.org/abs/cs/0605026v1, p.7.

[47] S.OISHI AND H.INOUE. (1982). *Pseudo-random number generators and chaos*. Transactions of the Institute of Electronics and Communication Engineers of Japan E, Vol. 65, 534 - 541.

[48] A.B.OWEN. (1988). *Empirical likelihood ratio confidence intervals for a single functional*. Biometrika. 75, 237 - 249.

[49] A.B.OWEN. (2011). *Empirical likelihood*. Chapman and Hall, New York.

[50] Y.QIN AND L.ZHAO. (2000). *Empirical likelihood ratio confidence intervals for various differences of two populations*. Syst. Sci. Math. Sci., 13, 23 - 30.

[51] G.H.SANDRI. (1992). *A Simple Nonperiodic Random Number Generator: A Recursive Model for the Logistic Map*. Scientific report, GL-TR- 89-1066, Boston University College of Engineering and Center for Space Physics Boston.

[52] I. SIMON. (1988). *Infinite words and a theorem of Hindman*. Rev. Mat. Apl. Vol. 9 , 97 - 104.

[53] B.SCHNEIER AND P.SUTHERLAND. (1995). *Applied cryptography: protocols, algorithms, and source code in C*. John Wiley & Sons, Inc. New York, NY, USA.

[54] D.R.STINSON. (1995). Cryptography. Theory and Practice. CRC Press. [55] A.THUE. (1906). Über unendliche Zeichenreihen. Norske Vid. Selsk. Skr. I Math-Nat. Kl. 7, 1 - 22.

[56] A.THUE. (1912). Über die gegenseitige Loge gleicher Teile gewisser Zeichenreihen. Norske Vid. Selsk. Skr. I Math-Nat. Kl. Chris. 1, 1 - 67.

[57] J.VALEINIS, E.CERS AND J.CIELENS. (2010). Two-sample problems in statistical data modelling. Math. Model. Anal., 15, 137 - 151.

[58] J.VALEINIS AND E.CERS. (2012). *Extending the two-sample empirical likelihood,* preprint

[59] J.VALEINIS AND E.CERS. (2011). *EL: Two-sample Empirical likelihood.* R-package version 1.0. URL http://CRAN.R-project.org/package=EL.

[60] J.Valeinis, A.Munk and E.Cers. (2008). *Extending the two-sample empirical likelihood method*. 7thWorld Congress in Probability and Statistics, Singapore, p. 198.

[61] J.Valeinis. (2011). *Two-sample plug-in empirical likelihood method with applications to structural relationship models*. 58th World Statistics Congress (ISI2011), Dublin, Ireland.

[62] J.Valeinis and S.Vucāne. (2012*). Tail bound inequalities and empirical likelihood for the mean*, (submitted to „Statistical methods and applications").

[63] J.Valeinis. (2012). *Two-sample blockwise empirical likelihood for weakly dependent data with applications to change-point analysis*. 8th World Congress in Probability and Statistics, Istanbul, Turkey, p. 231.

[64] J.Valeinis, M.Velina and G.Luta. (2012). *Empirical likelihoodbased inference for the difference of two smoothed Huber estimators*, (prepared preprint, planed to submit to „Test").

[65] J.Valeinis, M.Velina and G.Luta. (2011). *Empirical likelihoodbased inference for the difference of smoothed Huber estimators*, Internationa conference on Robust Statistics 2011, Valladolid, Spain.

[66] J. Valeinis. (2007*). Confidence bands for structural relationship models*. Disertācija. Gētingena, Vācija.

[67] J. Valeinis and L. Januseva. (2012). *Confidence bands for general shift function*, 17th International Conference on Mathematical Modelling and Analysis, Tallinn, Estonia.

[68] J.Valeinis and A.Ločmelis. (2012). *Bickel-Rosenblatt test for weakly dependent data*. Mathematical modelling and analysis, 17(3), 383 - 395.

[69] J. Valeinis. (2010). *Goodness-of-fit tests for weakly dependent data*. 10th International Vilnius Conference on Probability and Mathematical Statistics, Vilnius, Lithuania.

[70] M. Velina, J. Valeinis and G. Luta. (2012). *Empirical Likelihoodbased Methods for the Difference of Two Trimmed Means*, International conference on Robust Statistics 2012, Vermont, USA.

[71] S. Vucāne. (2012). *Empirical likelihood method for dependent processes*. 17th International Conference on Mathematical Modelling and Analysis, Tallinn, Estonia.

[72] S. Vucāne. (2012). *Tail bound inequalities and empirical likelihood for the mean*. 17th European Young Statisticians Meeting, Lisbon, Portugal.

[73] B.L. van der Waerden. (1927). *Beweis einer Baudet'chen Vermutung.* Nieuw. Arch. Wisk. Vol. 15, 212 - 216.

[74] W.Zhou and B.Y.Jing. (2003). *Smoothed empirical likelihood confidence intervals for the difference of quantiles*. Stat. Sinica, 13, 83 - 96.

[75] A.I. Zimin. (2003). *Blocking sets of terms.* Matematicheskii Sbornik, 3 Vol. 161, 363 - 375.

[76] Я. А. Булс (1986) *Оценка длины слова при моделировании конечных детерминированных автоматов*. Теория алгоритмов и программ. Рига: ЛГУ им. П.Стучки, С.50 - 63. (Russian)

[77] В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин. (1985). Введение в теорию автоматов. [An Introduction to the Theory of Automata.] Москва „Наука". (Russian)

[78] А. А. Курмит. (1982). Последовательная декомпозиция конечных автоматов.[Sequential Decomposition of Finite Automata.] . Рига „Зинатне". (Russian)

[79] Б. А. Трахтенброт, Я. М. Барздинь. (1970). Конечные автоматы (поведение и синтез). [Finite Automata (Behaviour and Synthesis).] Москва „Наука". (Russian)

[80] В. М. Фомичев. (2003). Дискретная математика и криптология. [Discrete Mathematics and Cryptology.] Москва. „ДИАЛОГ-МИФИ". (Russian)