ИНФОРМАТИКА

2006 **апрель-июнь** **№ 2**

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 681.32:519.2

Ю.С. Харин[1], А.Н. Ярмола[2]

ОБ ОДНОМ МЕТОДЕ ПОСТРОЕНИЯ ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Предлагается новый метод построения псевдослучайных последовательностей, использующий нелинейное комбинирование «элементарных» генераторов. Исследуются статистические свойства и свойства периодичности выходной последовательности предложенного генератора. Даются оценки вычислительной сложности генератора и результаты компьютерных экспериментов.

Введение

Генераторы псевдослучайных последовательностей являются неотъемлемой частью поточных криптосистем [1]. Такие криптосистемы используются в системах защиты информации, применяемых в компьютерных сетях, мобильной радиотелефонной связи и электронной торговле. Поэтому актуальной является задача построения псевдослучайных последовательностей (ПСП), удовлетворяющих требованию статистической близости распределений вероятностей фрагментов последовательности к равномерному распределению. Выделяются три основных подхода к построению алгоритмов генерации ПСП [1]: прямые методы построения элементарных ПСП, к которым относятся конгруэнтные генераторы, генераторы Фибоначчи, линейные рекуррентные последовательности (ЛРП) в конечном поле; методы улучшения элементарных ПСП, которые заключаются в специальных функциональных преобразованиях этих последовательностей; методы комбинирования алгоритмов генерации. Улучшение элементарных ПСП преследует следующие цели: получение последовательности с более близкими к равномерному распределениями вероятностей фрагментов; увеличение периода последовательности; увеличение криптостойкости последовательности. Одним из основных подходов к улучшению ПСП является комбинирование ЛРП [2], которое осуществляется с использованием операций сложения, умножения, прореживания. В данной работе в рамках этого подхода предлагается и исследуется новый метод построения ПСП, основанный на использовании модели INAR(m) целочисленных временных рядов [3].

1. Определение INAR-генератора

Модель INAR(m) предложена Эльзаидом и Эль-Ошем [3] в качестве целочисленного аналога модели авторегрессии порядка m для решения задач анализа целочисленных экономических временных рядов $z_t \in \mathbf{Z}_+ = \{0, 1, 2, \ldots\}$, $t = 1, 2, \ldots$:

$$z_t = \sum_{j=1}^{m} p_j \circ z_{t-j} + \eta_t \,,\ t = m+1, m+2, \ldots, \qquad (1)$$

где $\circ$ – оператор «биномиального прореживания» (binomial thinning): $p_j \circ z_{t-j} = \sum_{i=1}^{z_{t-j}} \xi_{t,i}^{(j)}$; $\{\xi_{t,i}^{(j)}$: $t = m+1, m+2, \ldots; i = 1, 2, \ldots; j = 1, \ldots, m\}$ – независимые в совокупности случайные величины Бернулли, $\mathbf{P}\{\xi_{t,i}^{(j)} = 1\} = 1 - \mathbf{P}\{\xi_{t,i}^{(j)} = 0\} = p_j$; $\{\eta_t \in \mathbf{Z}_+$: $t = m+1, m+2, \ldots\}$ – случайные величины, такие, что η_t некоррелирована с $z_{t-1}, \ldots, z_{t-m}$, $t = m+1, m+2, \ldots; z_1, \ldots, z_m$ – некоторые начальные значения. С учетом этих обозначений модель (1) принимает вид

$$z_t = \sum_{j=1}^{m} \sum_{i=1}^{z_{t-j}} \xi_{t,i}^{(t)} + \eta_t \,.$$

Используя данную идею модели INAR(m), построим генератор ПСП x_t в виде

$$x_t = (\sum_{j=1}^{m} \theta_j \sum_{i=0}^{x_{t-j}} \xi_{(t-1)N+i}^{(j)}) \bmod N\,,\ t = m+1,\ m+2,\ \ldots, \qquad (2)$$

где $\{\xi_{t,i}^{(j)} \in \{0,1\}: j = 1, \ldots, m\}$ – выходные последовательности некоторых простейших (элементарных) бинарных генераторов $G_1, \ldots, G_m$; в общем случае $\theta_j \in A_N = \{0, \ldots, N-1\}$, $\theta_m > 0$. Далее в статье исследуется специальный случай модели (2): $\theta_j \in A_2 = \{0, 1\}$, $j = 1, \ldots, m-1$, $\theta_m = 1$, где величина θ_j определяет, подключен генератор G_j в данном сеансе генерации ПСП или нет. Отличительной особенностью генератора (2) среди известных генераторов ПСП является использование в формуле (2) суммы случайного числа случайных величин $\sum_{i=0}^{x_{t-j}} \xi_{(t-1)N+i}^{(j)}$.

2. Вероятностные свойства выходной последовательности INAR-генератора

При исследовании вероятностных свойств выходной последовательности INAR-генератора в качестве модели для выходных последовательностей $\{\xi_t^{(1)}\},\ldots,\{\xi_t^{(m)}\}$ «элементарных» генераторов $G_1,\ldots, G_m$ будем использовать последовательности независимых одинаково распределенных случайных величин Бернулли: $\mathbf{P}\{\xi_t^{(j)} = 1\} = 1 - \mathbf{P}\{\xi_t^{(j)} = 0\} = p_j$, $j = 1, \ldots, m$; также будем полагать, что последовательности $\{\xi_t^{(1)}\},\ldots,\{\xi_t^{(m)}\}$ независимы. При этом уклонение $|p_j - 0{,}5| \in [0;\ 0{,}5]$ характеризует «степень несовершенства» элементарного генератора G_j ($j = 1, ..., m$). Обозначим: $r = \sum_{j=1}^{m} \theta_j \ge 1$ – вес Хэмминга вектора коэфициентов $\theta = (\theta_1, \ldots, \theta_m)'$; $1 \le k_1 < k_2 < \ldots < k_r \le m$ – различные индексы, такие, что $\theta_{k_j} = 1$, $j = 1, \ldots, r$, при этом $\theta_j = 0$, если $j \notin \{k_1, \ldots, k_r\}$; $\mathbf{I}\{B\}$ – индикаторная функция события B; $L\{\xi\}$ – закон распределения вероятностей случайной величины ξ; $Bi(k,p)$ – биномиальный закон распределения вероятностей с параметрами k, p.

Теорема 1. *Если $\{\xi_t^{(j)}\}$ – независимые в совокупности двоичные случайные величины, $\mathbf{P}\{\xi_t^{(j)} = 1\} = 1 - \mathbf{P}\{\xi_t^{(j)} = 0\} = p_j$, $j = 1, \ldots, m$, $t = m+1, m+2, ...$, то последовательность x_t, определяемая (2), – однородная цепь Маркова порядка m.*

Доказательство. Проверим марковское свойство. Учитывая, что $x_1, \ldots, x_{t-1}$ ($t = m+1, m+2, ...$) зависят только от $\{\xi_{N(k-1)+i}^{(j)}: j = 1, \ldots, m;\ k = 1, \ldots, t-1;\ i = 0, \ldots, N-1\}$ и не зависят от $\{\xi_{N(t-1)+i}^{(j)}: j = 1, \ldots, m,\ i = 0, \ldots, N-1\}$, находим

$$\mathrm{P}\{x_t = i_t \mid x_{t-1} = i_{t-1},\ldots,x_1 = i_1\} = \sum_{\substack{a_{N(t-1)}^{(1)},\ldots,a_{Nt-1}^{(1)},\\ a_{N(t-1)}^{(2)},\ldots,a_{Nt-1}^{(m)} \in A_2}} \mathrm{I}\left\{(\sum_{j=1}^{m} \theta_j \sum_{i=0}^{i_{t-j}} a_{(t-1)N+i}^{(j)}) \bmod N = i_t\right\} \times$$

$$\times \mathrm{P}\{\xi_{N(t-1)}^{(1)} = a_{N(t-1)}^{(1)},\ldots,\xi_{Nt-1}^{(m)} = a_{Nt-1}^{(m)} \mid x_{t-1} = i_{t-1},\ldots,x_1 = i_1\} =$$

$$= \sum_{a_{N(t-1)}^{(1)},\ldots,a_{Nt-1}^{(m)} \in A_2} \mathrm{I}\left\{(\sum_{j=1}^{m} \theta_j \sum_{i=0}^{i_{t-j}} a_{(t-1)N+i}^{(j)}) \bmod N = i_t\right\} \prod_{j=1}^{m} \prod_{i=0}^{N-1} (p_j a_{(t-1)N+i}^{(j)} + (1-p_j)(1-a_{(t-1)N+i}^{(j)})).$$

Таким образом, условное распределение вероятностей зависит только от m предыдущих значений $i_{t-1}, \ldots, i_{t-m}$, поэтому x_t – цепь Маркова m-го порядка. Поскольку найденное условное распределение вероятностей не зависит от t, то цепь Маркова однородная. ■

Из теоремы 1 следует, что в последовательности (2) появляются стохастические зависимости, которых не было в «элементарных» последовательностях, однако данный факт не ухудшает статистических свойств последовательности (2), что будет показано в следующем разделе.

Пусть $P=(p_{i_0,\ldots,i_m})$, $i_0, \ldots, i_m \in A_N$, – матрица вероятностей одношаговых переходов для цепи Маркова x_t, $p_{i_0,\ldots,i_m}=\mathbf{P}\{x_t=i_m \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}$, $i_0, \ldots, i_m \in A_N$, $t \ge m+1$.

Лемма 1. *Если* $\theta_1, \ldots, \theta_m \in A_2$, *то матрица вероятностей переходов* $P=(p_{i_0,\ldots,i_m})$ *имеет вид*

$$p_{i_0,\ldots,i_m}=\sum_{k=0}^{m-1}\mathbf{I}\{i_m+kN\le S(i_0,\ldots,i_{m-1})\}P(i_m+kN,i_0,\ldots,i_m),\ i_0,\ldots,i_m\in A_N; \tag{3}$$

$$S(i_0,\ldots,i_{m-1})=\sum_{j=1}^{m}\theta_j(i_{m-j}+1); \tag{4}$$

$$\begin{aligned}P(i,i_0,\ldots,i_{m-1})=&\sum_{j_1=0}^{\min(N_1,i)}\sum_{j_2=0}^{\min(N_2,i-j_1)}\cdots\sum_{j_{r-1}=0}^{\min(N_{r-1},i-j_1-\ldots-j_{r-2})}\prod_{l=1}^{r-1}\left(\mathrm{C}_{N_l}^{j_l}\,p_{k_l}^{j_l}(1-p_{k_l})^{N_l-j_l}\right)\times\\ \times\mathrm{C}_{N_r}^{i-j_1-\ldots-j_{r-1}}\,&p_{k_r}^{i-j_1-\ldots-j_{r-1}}(1-p_{k_l})^{N_r-i+j_1+\ldots+j_{r-1}},\ i=0,\ldots,S(i_0,\ldots,i_{m-1}),\ i_0,\ldots,i_{m-1}\in A_N;\end{aligned} \tag{5}$$

$$N_j=i_{m-k_j}+1,\ j=1,\ldots,r. \tag{6}$$

Доказательство. Обозначим:

$$\kappa_t^{(j)}=\sum_{i=0}^{x_{t-j}}\xi_{(t-1)N+i}^{(j)},\ j=1,\ldots,m,\ t>m;\ \eta_t=\sum_{j=1}^{m}\theta_j\sum_{i=0}^{x_{t-j}}\xi_{(t-1)N+i}^{(j)}=\sum_{j=1}^{m}\theta_j\kappa_t^{(j)}=\sum_{j=1}^{r}\kappa_t^{(k_j)},\ t>m.$$

Очевидно, что $x_t=\eta_t \bmod N$, $t>m$. Таким образом,

$$\mathbf{P}\{x_t=i_m \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}=\sum_{k=0}^{m-1}\mathbf{P}\{\eta_t=i_m+kN \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}. \tag{7}$$

Легко видеть, что при $x_{t-1}=i_{m-1}, \ldots, x_{t-m}=i_0$ случайная величина η_t с ненулевой вероятностью принимает значения только из $\{0, \ldots, S(i_0, \ldots, i_{m-1})\}$. При этом

$$\begin{aligned}\mathrm{P}\{\eta_t=i \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}&=\mathrm{P}\{\sum_{j=1}^{r}\kappa_t^{(k_j)}=i \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}=\\ =\sum_{\substack{j_k\in\{0,\ldots,N_k\},k=1,\ldots,r,\\ j_1+\ldots+j_r=i}}&\mathrm{P}\{\kappa_t^{(k_1)}=j_1,\ldots,\kappa_t^{(k_r)}=j_r \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}.\end{aligned}$$

Заметим, что в силу независимости случайных последовательностей $\{\xi_t^{(1)}\}, \{\xi_t^{(2)}\}, \ldots, \{\xi_t^{(m)}\}$

$$\mathbf{P}\{\kappa_t^{(k_1)}=j_1,\ldots,\kappa_t^{(k_r)}=j_r \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}=\prod_{l=1}^{r}\mathbf{P}\{\sum_{i=0}^{x_{t-k_j}}\xi_{(t-1)N+i}^{(k_l)}=j_l \mid x_{t-k_j}=i_{m-k_l}\}.$$

Поскольку $L\{\kappa_t^{(j)} \mid x_{t-j}=i_{m-j}\}=Bi(i_{m-j}+1, p_j)$, $j=1, \ldots, m$, то получаем

$$\mathbf{P}\{\eta_t=i \mid x_{t-1}=i_{m-1},\ldots,x_{t-m}=i_0\}=P(i,i_0,\ldots,i_{m-1}).$$

Подставляя последнее равенство в (7), приходим к утверждению леммы. ■

Следствие 1. *Если в условиях леммы 1 генераторы* $\{G_j\}$ *однородные:* $p_1=p_2=\ldots=p_m=p$, *то*

$$p_{i_0,\ldots,i_m}=\sum_{k=0}^{m-1}\mathbf{I}\{i_m+kN\le S(i_0,\ldots,i_{m-1})\}\,\mathrm{C}_{S(i_0,\ldots,i_{m-1})}^{i_m+kN}\,p^{i_m+kN}(1-p)^{S(i_0,\ldots,i_{m-1})-i_m-kN},\ i_0, \ldots, i_m \in A. \tag{8}$$

Доказательство. Заметим, что если $p_1 = p_2 = ... = p_m = p$, то $L\{\eta_t \mid x_{t-1} = i_{m-1}, \ldots, x_{t-m} = i_0\} = Bi(S(i_0, \ldots, i_{m-1}), p)$, и воспользуемся выражением (7). ■

Таким образом, вероятности переходов зависят только от i_m и величины $S(i_0, \ldots, i_{m-1}) \in \{r, r+1, \ldots, Nr\}$ и для вычисления матрицы вероятностей переходов необходимо определить только $(N-1)((N-1)r+1)$ значений вероятностей переходов.

Лемма 2. *Если $r \geq N-1$ и $0 < p_{k_j} < 1$, $j = 1, \ldots, r$, то цепь Маркова x_t эргодическая.*

Доказательство. Эргодичность сложной цепи Маркова порядка m понимается как эргодичность «векторной» цепи Маркова первого порядка для векторов $X_t = (x_{t-(m-1)}, \ldots, x_t)$ с матрицей вероятностей переходов $P_X = (p_{\bar{k},\bar{i}})$, $p_{\bar{k},\bar{i}} = \mathbf{I}\{i_0 = k_1, \ldots, i_{m-2} = k_{m-1}\} p_{k_0,\ldots,k_{m-1},i_{m-1}}$, $\bar{k} = (k_0, \ldots, k_{m-1})$, $\bar{i} = (i_0, \ldots, i_{m-1}) \in A_N^m$. Так как в условиях леммы $S(i_0, \ldots i_{m-1}) \geq N-1$ для любых $i_0, \ldots, i_m \in A_N$, то правая часть выражения (4) содержит, по крайней мере, одно слагаемое $P(i_m, i_0, \ldots, i_{m-1})$. Поскольку $0 < p_{k_j} < 1$, $j = 1, \ldots, r$, то $P(i_m, i_0, \ldots, i_{m-1}) > 0$, следовательно, $p_{i_0,\ldots,i_m} > 0$, $i_0, \ldots, i_m \in A_N$, поэтому все элементы матрицы P_X^m положительны. Таким образом, цепь Маркова x_t эргодическая [4]. ■

3. Вероятностные свойства для бинарного INAR-генератора

Исследуем наиболее важный на практике бинарный случай [2], когда $N = 2$ и генераторы $G_1, \ldots, G_m$ однородные:

$$p_j = \frac{1}{2}(1+\varepsilon), j=1,\ldots,m, \ |\varepsilon| < 1, \tag{9}$$

где $\varepsilon \in (-1; 1)$ – величина, характеризующая «степень несовершенства» генераторов $G_1, \ldots, G_m$ (если $\varepsilon = 0$, то $p_1 = \ldots = p_m = 1/2$ и все генераторы $G_1, \ldots, G_m$ порождают «чисто случайные» последовательности). Заметим, что при $N = 2$ INAR-генератор (2) допускает упрощенное представление, показывающее его нелинейность:

$$x_t = \Big(\sum_{j=1}^{m} \theta_j (\xi_{2(t-1)}^{(j)} + x_{t-j}\xi_{2(t-1)+1}^{(j)})\Big) \bmod 2, \ t > m.$$

Отметим также, что в силу (9) и леммы 2 x_t – эргодическая цепь Маркова.

Обозначим $\Pi^{(t)} = (\pi_{i_1,\ldots,i_m}^{(t)})$, $i_1, \ldots, i_m \in A_2$, $t \geq m$, m-мерное распределение вероятностей процесса x_t, $\pi_{i_1,\ldots,i_m}^{(t)} = \mathbf{P}\{x_{t-m+1} = i_1, \ldots, x_t = i_m\}$; $\Pi^* = (\pi_{i_1,\ldots,i_m}^*)$, $i_1, \ldots, i_m \in A_2$, – m-мерное стационарное распределение вероятностей [4], которое существует в силу свойства эргодичности.

Исследуем распределение вероятностей цепи Маркова m-го порядка x_t, порождаемой генератором (2), с помощью следующих функционалов, которые возникают в задаче дискриминантного анализа цепей Маркова [5]: $\Delta = 2^{-m} \sum_{i_0,\ldots,i_m \in A_2} | p_{i_0,\ldots,i_m} - 0,5 | \geq 0$ – среднее уклонение распределения вероятностей одношаговых переходов от равномерного распределения; $Z_m = \sum_{i_1,\ldots,i_m \in A_2} | \pi_{i_1,\ldots,i_m}^* - 2^{-m} | \geq 0$ – октаэдрическая норма уклонения стационарного m-мерного распределения Π^* от равномерного на A_2^m распределения. Чем Δ, Z_m меньше, тем выходная последовательность генератора (2) ближе к «чисто случайной последовательности».

В дальнейшем потребуются два вспомогательных утверждения.

Лемма 3. *Если случайная величина η имеет биномиальное распределение вероятностей с параметрами n, p, то $\mathbf{P}\{\eta \bmod 2 = 1\} = \mathbf{P}\{\eta \bmod 2 = 0\} = 1/2$ тогда и только тогда, когда $p = 1/2$.*

Доказательство. Вычислим

$$\mathbf{P}\{\eta \bmod 2 = 1\} - \mathbf{P}\{\eta \bmod 2 = 0\} = \sum_{k=0}^{n} (-1)^{k+1} C_n^k p^k (1-p)^{n-k} = -((1-p)-p)^n = -(1-2p)^n.$$

Из полученного равенства и следует утверждение леммы. ■

Лемма 4. *Если в условиях леммы 3* $p = (1 + \varepsilon)/2$, $|\varepsilon| < 1$, *то* $\mathbf{P}\{\eta \bmod 2 = 1\} = (1 - (-\varepsilon)^n)/2$.

Доказательство. Из доказательства леммы 3 следует, что $\mathbf{P}\{\eta \bmod 2 = 1\} - \mathbf{P}\{\eta \bmod 2 = 1\} = -(-\varepsilon)^n$. Учитывая, что $\mathbf{P}\{\eta \bmod 2 = 1\} + \mathbf{P}\{\eta \bmod 2 = 0\} = 1$, приходим к утверждению леммы. ■

Лемма 5. *Если* $N = 2$ *и выполнены условия (9), то для любых* $i_0, \ldots, i_m \in A_2$

$$p_{i_0,\ldots,i_{m-1},0} = p_{i_0,\ldots,i_{m-1},1} = 1/2$$

тогда и только тогда, когда p = 1/2.

Доказательство. Поскольку в условиях леммы $L\{\eta_t \mid x_{t-1} = i_{m-1}, \ldots, x_{t-m} = i_0\} = Bi(S(i_0,\ldots,i_{m-1}), p)$, то истинность следует из леммы 3. ■

Из леммы 5 следует, что предложенный метод не позволяет получить «совершенный» генератор из «несовершенных», однако следующая теорема позволяет оценить количественно «степень несовершенства» генератора ПСП (2) и выяснить факторы, от которых она зависит.

Теорема 2. *Если N=2 и выполнены условия (9), то*

$$p_{i_0,\ldots,i_m} = (1 + (-1)^{i_m}(-\varepsilon)^{S(i_0,\ldots,i_{m-1})})/2, \; i_0, \ldots, i_m \in A_N; \tag{10}$$

$$\Delta = |\varepsilon/2|^r \, (1 + |\varepsilon|)^r. \tag{11}$$

Доказательство. Первая часть теоремы следует из выражения $L\{\eta_t \mid x_{t-1} = i_{m-1}, \ldots, x_{t-m} = i_0\} = Bi(S(i_0,\ldots,i_{m-1}), p)$ и леммы 4.

Согласно (10) отклонение $p_{i_0,\ldots,i_m}$ от 1/2 зависит только от значения $S(i_0, \ldots, i_{m-1})$, которое, согласно (5), однозначно определяется значениями $i_{m-j_1}, \ldots, i_{m-j_r}$, поэтому

$$\Delta = 2^{-m} \sum_{i_0,\ldots,i_m \in A} \left| p_{i_0,\ldots,i_m} - 0{,}5 \right| = 2^{-m} \sum_{S=0}^{r} \sum_{\substack{i_0,\ldots,i_m \in A_2 \\ i_{m-k_1} + \ldots + i_{m-k_r} = S}} \left| p_{i_0,\ldots,i_m} - 0.5 \right| =$$

$$2^{-m} \sum_{S=0}^{r} 2^{m+1-r} \, C_r^S \, 2^{-1} \, |\varepsilon|^{r+S} = \frac{|\varepsilon|^r}{2^r} \sum_{S=0}^{r} C_r^S \, |\varepsilon|^S = \frac{|\varepsilon|^r}{2} (1 + |\varepsilon|)^r. \blacksquare$$

Замечание. *Поскольку* $r \le S(i_0, \ldots, i_{m-1}) \le 2r$ *для всех* $i_0, \ldots, i_m \in A_2$, *то*

$$|\varepsilon|^{2r}/2 \le |p_{i_0,\ldots,i_m} - 0{,}5| \le |\varepsilon|^r/2. \tag{12}$$

Отметим, что для «элементарной» последовательности $\Delta = |\varepsilon|$, следовательно, в силу теоремы 2, несмотря на наличие в выходной последовательности (2) стохастических зависимостей, данная последовательность (2) является менее предсказуемой, чем «элементарные» последовательности.

Обозначим $\alpha(i_1,\ldots,i_m) = (-1)^{i_m}(-\varepsilon)^{\sum_{j=1}^{m-1}\theta_j i_{m-j} + r}/2$. Заметим, что поскольку $\theta_m = 1$, то во введенных обозначениях

$$p_{i_0,\ldots,i_m} = 1/2 + (-\varepsilon)^{i_0} \alpha(i_1,\ldots,i_m), \; i_0, \ldots, i_m \in A_2.$$

Исследуем свойства стационарного распределения Π^*. В силу леммы 2 при $|\varepsilon| < 1/2$ бинарная цепь Маркова x_t эргодическая.

Теорема 3. *Если выполнены условия (9), то для стационарного распределения* Π^* *справедливо асимптотическое* ($\varepsilon \to 0$) *разложение:*

$$\pi^*_{i_1,\ldots,i_m} = \frac{1}{2^m}(1+\delta(i_1,\ldots,i_m)) + O(\varepsilon^{2r}),\ i_0,\ \ldots,\ i_m \in A_2, \tag{13}$$

где $\delta(i_1,\ldots,i_m) = (1-\varepsilon)(-\varepsilon)^r\left(0\sum_{j=1}^{m-1}\frac{(-1)^{i_{m-j}}}{2^j}(-\varepsilon)^{\sum_{k=1}^{m-j-1}\theta_k i_{m-j}}\prod_{k=m-j}^{m-1}(1+(-\varepsilon)^{\theta_k})\right)$ *– полином, содержащий члены* ε^k, $r \le k \le 2r$.

Доказательство. Если цепь Маркова эргодическая, то согласно [4] при любом начальном распределении $\Pi^{(m)}$ вектора $(x_1,\ldots,x_m)'$: $\Pi^{(t)} \to \Pi^*$, $t \to \infty$. Пусть $\Pi^{(m)}$ – равномерное на A_2^m распределение. Воспользуемся принципом математической индукции

$$\pi^{(m+1)}_{i_1,\ldots,i_m} = 2^{-m}(p_{0,i_1,\ldots,i_m} + p_{1,i_1,\ldots,i_m}) = 2^{-m}(1+\alpha(i_1,\ldots,i_m)(1-\varepsilon)).$$

Учитывая, что $\alpha(1,i_1,\ldots,i_{m-1}) = (-\varepsilon)^{\theta_{m-1}}\alpha(0,i_1,\ldots,i_{m-1})$, находим

$$\pi^{(m+2)}_{i_1,\ldots,i_m} = p_{0,i_1,\ldots,i_m}\pi^{(m+1)}_{0,i_1,\ldots,i_{m-1}} + p_{1,i_1,\ldots,i_m}\pi^{(m+1)}_{1,i_1,\ldots,i_{m-1}} = 2^{-m}(1+\alpha(i_1,\ldots,i_m)(1-\varepsilon) +$$
$$+(1-\varepsilon)\alpha(0,i_1,\ldots,i_{m-1})(1+(-\varepsilon)^{\theta_{m-1}})/2 + (1-\varepsilon)\alpha(i_1,\ldots,i_m)\alpha(0,i_1,\ldots,i_{m-1})(1+(-\varepsilon)^{\theta_{m-1}})).$$

Поскольку $|\varepsilon| < 1$, то $|\alpha(i_1,\ldots,i_m)| \le |\alpha(0,\ldots,0)| = |\varepsilon|^r/2$, $i_0,\ \ldots,\ i_m \in A_2$. Таким образом, $\alpha(i_1,\ldots,i_m)\alpha(0,i_1,\ldots,i_{m-1}) = O(\varepsilon^{2r})$. Продолжая вычисления, находим

$$\pi^{(2m)}_{i_1,\ldots,i_m} = 2^{-m}(1+\delta(i_1,\ldots,i_m)) + O(\varepsilon^{2r}).$$

Пусть для некоторого $t \ge 2m$ выполнено $\pi^{(t)}_{i_1,\ldots,i_m} = 2^{-m}(1+\delta(i_1,\ldots,i_m)) + O(\varepsilon^{2r})$, $i_0,\ \ldots,\ i_m \in A_2$, покажем что и $\pi^{(t+1)}_{i_1,\ldots,i_m}$ будет иметь такой же вид. Легко видеть, что

$$\pi^{(t+1)}_{i_1,\ldots,i_m} = 2^{-m}(1+(1-\varepsilon)\alpha(i_1,\ldots,i_m) + (\delta(0,i_1,\ldots,i_{m-1}) + \delta(1,i_1,\ldots,i_{m-1}))/2) + O(\varepsilon^{2r}).$$

Заметим, что $\alpha(0,\ldots,0,1) = -\alpha(0,\ldots,0)$. Следовательно,

$$(1-\varepsilon)\alpha(i_1,\ldots,i_m) + (\delta(0,i_1,\ldots,i_{m-1}) + \delta(1,i_1,\ldots,i_{m-1}))/2 = \delta(i_1,\ldots,i_m).$$

Таким образом, $\pi^{(t+1)}_{i_1,\ldots,i_m} = 2^{-m}(1+\delta(i_1,\ldots,i_m)) + O(\varepsilon^{2r})$, $t \ge 2m$.■

Следствие 2. *В условиях теоремы 3 справедливо неравенство* $Z_m \le Z_+$, *причем для верхней границы* Z_+ *справедливо асимптотическое* ($\varepsilon \to 0$) *разложение*

$$Z_+ = 2^{-r}(1-\varepsilon)|\varepsilon|^r(1+|\varepsilon|)^{r-1}\left(1+\sum_{j=1}^{m-1}2^{2(r_j-j)}\frac{|\varepsilon|^{r_j}}{(1+|\varepsilon|)^{r_j}}\prod_{k=1}^{j}(1+(-\varepsilon)^{\theta_{k-j}})\right) + O(\varepsilon^{2r}), \tag{14}$$

где $r_j = \sum_{k=1}^{j}\theta_{k-j}$, j = 1, …, $m-1$.

Доказательство. Из теоремы 3 следует

$$\left|\pi^*_{i_1,\ldots,i_m} - \frac{1}{2^m}\right| \le \frac{1}{2^m}(|\alpha(i_1,\ldots,i_m)|(1-\varepsilon) + (1-\varepsilon)\sum_{j=1}^{m-1}\frac{1}{2^j}|\alpha(0,\ldots,0,i_1,\ldots,i_{m-j})|\prod_{k=1}^{j}(1+(-\varepsilon)^{\theta_{k-j}}) + O(\varepsilon^{2r})).$$

Применяя подход, использованный при доказательстве теоремы 2, суммируем почленно это неравенство по $i_0, \ldots, i_m \in A_2$. ■

В силу следствия 2 $Z_m = O(|\varepsilon|^r)$; таким образом, распределение m-векторов (m-грамм) выходной последовательности INAR-генератора близко к равномерному на A_2^m распределению с погрешностью $O(|\varepsilon|^r)$. Заметим, что для «элементарной» последовательности погрешность значительно больше: $Z_m = O(|\varepsilon|)$. Следовательно, распределения вероятностей фрагментов выходной последовательности (2) ближе к равномерному, чем распределения фрагментов такой же длины для «элементарного» генератора. Чтобы оценить влияние этой близости на статистическую различимость распределений, рассмотрим следующую вспомогательную задачу. Пусть наблюдается реализация $X = (x_1, \ldots, x_T)$, $x_t \in A_2$, $t = 1, \ldots, T$, длительности T, по которой необходимо проверить гипотезу $H_0 = \{x_t$ – равномерно распределенная случайная последовательность (РРСП)$\}$ против альтернативы $H_1 = \{x_t$ – выходная последовательность INAR-генератора, удовлетворяющая (9) при $\varepsilon \neq 0\}$. Обозначим $P(X|H_i)$ распределение вероятностей наблюдений X при условии верной гипотезы H_i, $i = 0,1$. Одной из величин, характеризующих различимость гипотез H_0, H_1, является дивергенция Кульбака $J(H_0, H_1)$ распределений вероятностей $P(X|H_0)$, $P(X|H_1)$ [6].

Лемма 6. *Если «стартовые значения» $x_1,\ldots,x_m$ генератора (2) независимы и имеют равномерное на A_2 распределение вероятностей, то для дивергенции Кульбака $J(H_0,H_1)$ справедлива двухсторонняя оценка:*

$$J_- \le J(H_0, H_1) \le J_+, \tag{15}$$

где $J_- = (T-m)((1-|\varepsilon|^r)^{T-m} - 1)\ln(1-|\varepsilon|^r) = O(|\varepsilon|^r)$, $J_+ = (T-m)((1+|\varepsilon|^r)^{T-m} - 1)\ln(1+|\varepsilon|^r) = O(|\varepsilon|^r)$.

Доказательство. Заметим, что

$$J(H_0, H_1) = \sum_{\bar{i} \in A_2^T} (\mathbf{P}\{X = \bar{i} \mid H_1\} - \mathbf{P}\{X = \bar{i} \mid H_0\}) \ln \frac{\mathbf{P}\{X = \bar{i} \mid H_1\}}{\mathbf{P}\{X = \bar{i} \mid H_0\}} \le 2^T (P_{\max}(H_1) - 2^{-T}) \ln(2^T P_{\max}(H_1)),$$

где $P_{\max}(H_1) = \max_{\bar{i} \in A_2^T} \mathbf{P}\{X = \bar{i} \mid H_1\}$ $\max_{\bar{i} \in A_2^T} \mathrm{P}\{X = \bar{i} \mid H_1\}$. Поскольку в силу теоремы 2 $\max_{\bar{i} \in A_2^T} \mathbf{P}\{X = \bar{i} = (i_1, \ldots, i_T) \mid H_1\} = 2^{-T} \prod_{t=m+1}^{T} (1 + (-1)^{i_t} (-\varepsilon)^{S(i_{t-m}, \ldots, i_{t-1})})$, то используя правую часть выражения (11), имеем $P_{\max}(H_1) \le 2^{-T}(1 + |\varepsilon|^r)^{T-m}$. Таким образом, приходим к оценке для $J(H_0,H_1)$ сверху. Аналогично можно получить оценку снизу. Лемма доказана. ■

Из леммы 6 следует, что число наблюдений, необходимое для выявления в последовательности (2) отклонения от РРСП ($O(|\varepsilon|^r)$), значительно превышает число наблюдений для решения этой же задачи в случае «элементарного» генератора ($O(|\varepsilon|)$).

4. Свойства периодичности выходной последовательности INAR-генератора

При исследовании свойств периодичности выходной последовательности генератора (2) будем полагать, что выходные последовательности «элементарных» генераторов $G_1,\ldots, G_m$ – детерминированные периодические последовательности. При этом предположении (общепринятом при анализе периодичности [1, 2]) x_t – также детерминированная периодическая последовательность.

Обозначим T_j период выходной последовательности $\xi_t^{(j)}$ генератора G_j ($j = 1, \ldots, m$).

Лемма 7. *Если $N < \min\{T_{k_1}, \ldots, T_{k_r}\}$, то для периода T^* выходной последовательности INAR-генератора (2) справедлива следующая оценка сверху:*

$$T^* \le T_{\max} = \mathrm{НОК}\left(\frac{T_{k_1}}{\mathrm{НОД}(T_{k_1}, N)}, \ldots, \frac{T_{k_r}}{\mathrm{НОД}(T_{k_r}, N)} \right). \tag{16}$$

Доказательство. Пусть для некоторого $T \ge 1$ выполнено $x_t = x_{t+T}$ для всех $t \ge t^*$, следовательно,

$$(\sum_{j=1}^{m}\theta_j\sum_{i=0}^{x_{t-j}}\xi^{(j)}_{(t-1)N+i})\bmod N=(\sum_{j=1}^{m}\theta_j\sum_{i=0}^{x_{t+T-j}}\xi^{(j)}_{(t+T-1)N+i})\bmod N,\ t\ge t.$$

Последнее равенство заведомо выполнено, если $\xi^{(j)}_{(t-1)N+i}=\xi^{(j)}_{(t+T-1)N+i}$, $j = k_1, \ldots, k_r$, $i = 0, \ldots, N-1$, $t \ge t^*$. Полученное условие выполнено, если $T = kT_{\max}$, $k = 1, 2, \ldots$. Таким образом, $T^* \le T_{\max}$. ■

Будем предполагать, что выходные последовательности генераторов $\{G_j: j = 1, \ldots, m\}$ имеют нулевой предпериод. Наиболее важным для практики является случай, когда T^* достигает своего наибольшего значения.

Теорема 4. *Если* $N = 2$, $\min\{T_{k_1},\ldots,T_{k_r}\}>2$, $\{T_{k_l}: j=1,\ldots,r\}$ *нечетны и взаимно просты, то*

$$T^*=T_{\max}=T_{k_1},\ldots,T_{k_r}.$$

Доказательство. Пусть существует $l \in \{1, \ldots, r\}$, для которого $T_{k_l} = T$ не делит T^*. Обозначим T_0=НОК($T^*, T_{k_1}, \ldots, T_{k_{l-1}}, T_{k_{l+1}}, \ldots, T_{k_r}$). Очевидно, что в условиях теоремы $T_0 = kT + z$, $0 < z < T$. Так как по условию теоремы НОД(2, T) = 1, то для любого $a \in \{0, \ldots, T-1\}$ существует $s \in \{0, \ldots, T-1\}$, такое, что $2a = s \bmod T$. Зафиксируем s, рассмотрим $t = cT + a + 1$, $c \in \mathbf{N}$, $t - m \ge t^*$ и пусть $x_{t-k_j} = d_j(s)$, $j = 1, \ldots, r$. Получим

$$x_t=(\sum_{j=1}^{m}\theta_j\sum_{i=0}^{x_{t-j}}\xi^{(j)}_{2(t-1)+i})\bmod 2=(\sum_{j=1}^{r}\sum_{i=0}^{d_j(s)}\xi^{(k_j)}_{(s+i)\bmod T})\bmod 2.$$

С другой стороны, учитывая периодичность последовательностей $\xi^{(j)}_t$,

$$x_{t+T_0}=(\sum_{j=1}^{r}\sum_{i=0}^{d_j(s)}\xi^{(k_j)}_{2(t+T_0-1)+i})\bmod 2=(\sum_{\substack{j=1\\ j\ne l}}^{r}\sum_{i=0}^{d_j(s)}\xi^{(k_j)}_{(s+i)\bmod T}+\sum_{i=0}^{d_l(s)}\xi^{(k_j)}_{(s+i)\bmod T+2T_0})\bmod 2.$$

Перебирая все $s \in \{0, \ldots, T-1\}$, приходим к системе уравнений

$$\xi^{(k_l)}_s\oplus d_l(s)\xi^{(k_l)}_{(s+1)\bmod T}=\xi^{(k_l)}_{s+2T_0}\oplus d_l(s)\xi^{(k_l)}_{(s+1)\bmod T+2T_0},\ s=0,\ \ldots,\ T-1. \qquad (17)$$

Возможны два случая. Первый случай: $\exists t_0 > t^*$, для которого $x_{t_0} = 0$. Решим систему (17) относительно ξ_{s+2T_0}, $s = 0, \ldots, T-1$. Очевидно, что найдется s, для которого существует c, такое, что $t_0 + k_l = cT + a + 1$, т. е. в системе (17) найдется уравнение, в котором $d_l(s) = 0$. Легко видеть, что если существует s, для которого $d_l(s) = 0$, то определитель матрицы системы (16) равен 1. Таким образом, существует единственное решение системы $\xi_{s+2T_0} = \xi_s$, $s = 0, \ldots, T-1$, следовательно, T делит $2T_0$, что приводит к противоречию.

Второй случай: $x_t = 1$, $\forall t > t^*$. Равенство $x_t = x_{t+T_0}$, $t \ge t^*$, можно представить в виде $\xi^{(k_l)}_{2(t-1)}\oplus\xi^{(k_l)}_{2(t-1)+1}=\xi^{(k_l)}_{2(t+T_0-1)}\oplus\xi^{(k_l)}_{2(t+T_0-1)+1}$, $t \ge t^*$. Рассмотрим последовательность $\zeta_t=\xi^{(k_l)}_{2(t-1)}\oplus\xi^{(k_l)}_{2(t-1)+1}$. Очевидно, что для периода T_ζ последовательности ζ_t справедливо $T = k_1T_\zeta$, $T_0 = k_2T_\zeta$, следовательно, $T_\zeta = 1$. Поскольку НОД(2, T) = 1, то из $T_\zeta = 1$ следует, что $\xi^{(k_l)}_t = b$, $b \in \{0, 1\}$, $t \ge 0$. Это вновь приводит к противоречию. Следовательно, T_{k_l} делит T^*. ■

Теорема 4 позволяет выбирать комбинируемые генераторы так, чтобы обеспечить максимальный период выходной последовательности. Заметим, что не рекомендуется использовать в качестве «элементарных» последовательности с равными периодами, поскольку в таком случае

период выходной последовательности (2) существенно зависит от начальных значений и в некоторых случаях последовательность (2) может вырождаться в постоянную.

5. Результаты компьютерных экспериментов

Для машинного времени t_{INAR} генерации одного символа с помощью INAR-генератора (2) справедлива оценка

$$t_{\text{INAR}} \le rNt_{\text{элем}} + rNt_{\text{сум}} + t_{\text{пам}},$$

где $t_{\text{элем}}$ – время генерации одного символа «элементарным» генератором G_j; $t_{\text{сум}}$ – время операции суммирования по модулю N; $t_{\text{пам}}$ – время записи сгенерированного элемента в регистр памяти. На компьютере с процессором Athlon64 3000+ генерировалось 20×10^6 символов бинарным INAR-генератором (2) при различных значениях параметров, в качестве «элементарного» генератора использовалась сумма по модулю 2 двух мультипликативных конгруэнтных генераторов. Время, необходимое для генерации при $N = 2$, представлено в таблице.

Таблица

Машинное время, необходимое для генерации символа

Генератор	Время t, мкс
«Элементарный»	0,15
Бинарный INAR, $r = m = 2$	0,70
Бинарный INAR, $r = m = 4$	1,25
Бинарный INAR, $r = m = 8$	2,40

Результаты, полученные в разд. 3 статьи, показывают, что вероятностные характеристики выходной последовательности существенно зависят от величины r – веса Хэмминга вектора θ. Из графиков зависимостей от r среднего отклонения Δ вероятности переходов от равномерного распределения, вычисленных согласно (11) (рис. 1), и графиков зависимостей от r верхней границы Z_+ для нормы отклонения Z_m стационарного m-мерного распределения от равномерного, вычисленных согласно (14) (рис. 2), видно, что даже при достаточно больших значениях ε уклонения распределений вероятностей комбинируемых последовательностей от равномерного уже при относительно малых r распределения вероятностей выходной последовательности мало отличаются от равномерного распределения.

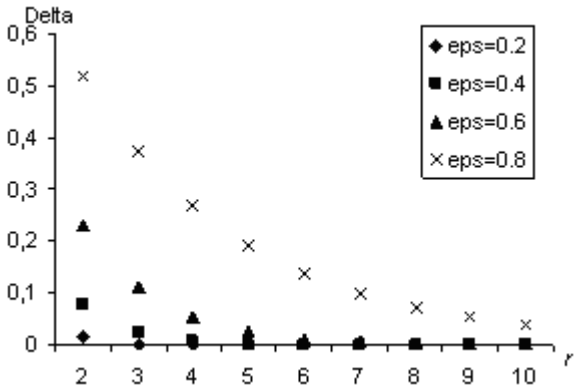

Рис. 1. Зависимость Δ от r при различных значениях ε

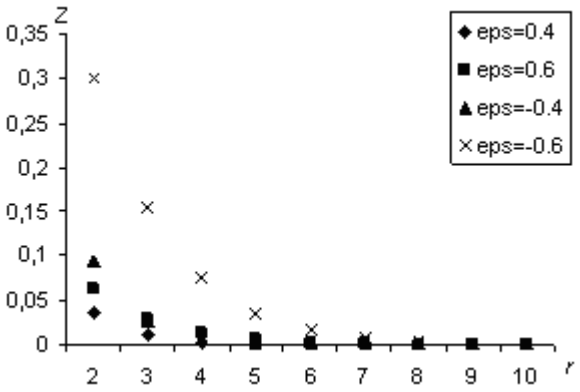

Рис. 2. Зависимость Z_+ от r при различных значениях ε

В ходе численных экспериментов исследовалась также зависимость статистических оценок для Δ и Z_m (вычисленных по методу Монте-Карло) от длительности наблюдения T и ε. В экспериментах моделировались временные ряды (2) при $N = 2$, $m = 5$, $\theta_j = 1$, $j = 1, \ldots, m$; число прогонов в методе Монте-Карло $M = 50$. Из графиков зависимости Δ при различных значениях

длительности наблюдений T и теоретических значений Δ, найденных в теореме 2 (рис. 3), а также графиков зависимости Z_m при различных значениях длительности наблюдений T и теоретических значений Z_+, найденных в следствии 2 (рис. 4), следует достаточно хорошее соответствие экспериментальных результатов теоретическим оценкам качества выходной последовательности INAR-генератора (2).

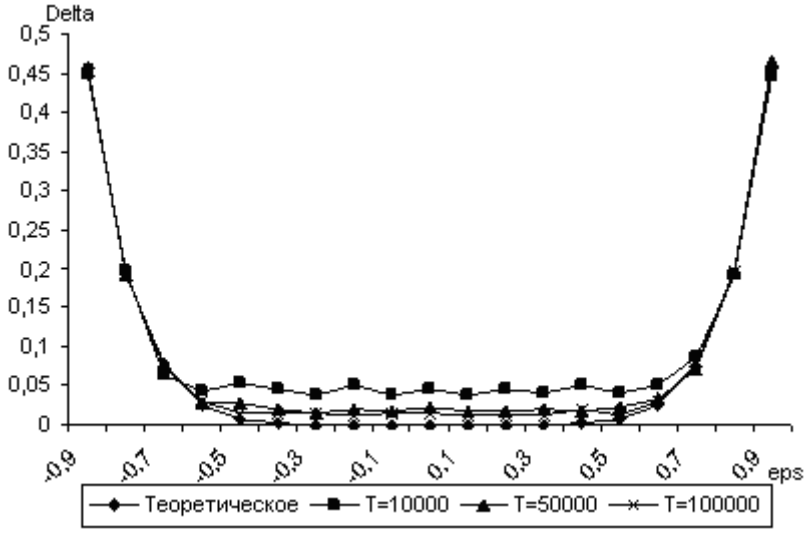

Рис. 3. Зависимость Δ от ε при различных длительностях наблюдаемых последовательностей

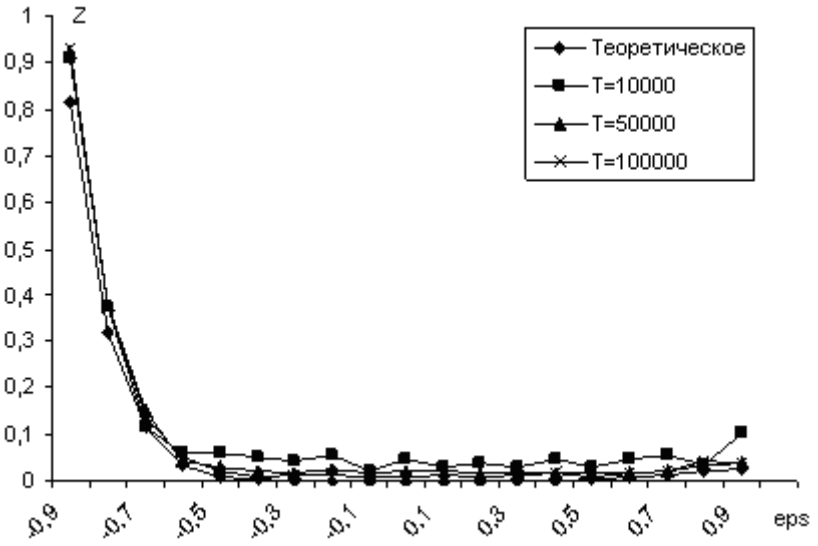

Рис. 4. Зависимость Z_m от ε при различных длительностях наблюдаемых последовательностей

Заключение

Предложен новый метод построения псевдослучайных последовательностей, основанный на модели INAR(m) временных рядов. Отличительной особенностью предложенного метода является использование в выражении (2) суммы случайного числа случайных величин. Для предложенного метода исследованы статистические свойства выходной последовательности, количественно оценен выигрыш от использования INAR-генератора по сравнению с «элементарными» генераторами. Найдена оценки периода выходной последовательности INAR-

генератора, установлено, что период является максимальным, если периоды «элементарных» генераторов взаимнопросты.

Работа частично поддержана ГПФИ «Математические модели» (проект ММ-24).

Список литературы

1. Математические и компьютерные основы криптологии / Ю.С. Харин [и др.]. – Минск: Новое знание, 2003.

2. Варфоломеев, А.А. Поточные криптосистемы. Основные свойства и методы анализа стойкости / А.А. Варфоломеев, А.Е. Жуков, М.А. Пудовкина. – М.: ПАИМС, 2000. – 272 с.

3. Alzaid, A.A. An integer-valued *p*th-order autoregressive structure (INAR(*p*)) process / A.A. Alzaid, M. Al-Osh // Journal of Applied Probability. – 1990. – № 27. – P. 314–324.

4. Боровков, А.А. Теория вероятностей / А.А. Боровков. – М.: Наука, 1986.

5. Kharin, Yu. Discriminant analysis of stationary finite Markov chains / Yu. Kharin, A. Kostevich // Math. Methods of Statistics. – 2004. – Vol. 13, № 1. – P. 235–252.

6. Кульбак, С. Теория информации и статистика / С. Кульбак. – М.: Наука, 1980.

Поступила 21.02.06

[1]*Национальный научно-исследовательский центр прикладных проблем математики и информатики, Минск, пр. Независимости, 4*
e-mail: kharin@bsu.by

[2]*Белорусский государственный университет, Минск, пр. Независимости, 4*
e-mail: and_yarmola@tut.by

Yu.S. Kharin, A.N. Yarmola

ON A METHOD FOR CONSTRUCTION OF PSEUDORANDOM SEQUENCES

A new method for construction of pseudorandom sequences is proposed; it is based on nonlinear combining of «elementary generators». Statistical properties and periodicity of output sequences of the proposed generator are studied. Estimates of the generator computational complexity and results of computer experiments are given.