УДК 519.217.2

DOI: 10.21779/2542-0321-2023-38-2-61–68

В.М. Захаров, С.В. Шалагин

Анализ псевдослучайных последовательностей заданной длины по критерию «энтропия цепей Маркова»

Казанский национальный исследовательский технический университет им. А.Н. Туполева – КАИ; Россия, 420111, г. Казань, ул. Карла Маркса, 10; Gilvv@mail.ru, SShalagin@mail.ru

Аннотация. Рассматривается задача анализа качества псевдослучайной последовательности (ПСП) заданной длины. Качество ПСП оценивается как степень приближения статистических ее свойств к свойствам случайной последовательности. Решена задача отображения закономерностей в структуре ПСП, определяемых алгоритмом формирования, в соответствующее значение параметра «энтропия цепей Маркова». Параметр вычисляется по стохастической матрице, однозначно соответствующей исследуемой ПСП, величина параметра – мера качества. Показано, что величина отклонения параметра от заданного максимального значения энтропии случайной последовательности – мера качества ПСП, оцениваемая с точностью, пропорциональной длине анализируемой ПСП.

Ключевые слова: анализ псевдослучайных последовательностей, критерий «энтропия цепей Маркова», стохастическая матрица, точность оценки.

Введение

Методы и алгоритмы анализа периодических многоразрядных псевдослучайных последовательностей (ПСП), а также подпоследовательностей заданной длины подобных ПСП, формируемых на алгоритмах, реализующих различные рекуррентные соотношения, широко отражены в публикациях [1–15]. Разнообразие алгоритмов и тестов для анализа ПСП постоянно увеличивается [10–13], что обусловлено развитием, совершенствованием моделей и алгоритмов построения генераторов ПСП с характеристиками, близкими к характеристикам случайных последовательностей [11–15], и расширением круга задач, связанных с применением ПСП в областях статистического моделирования [4; 12–17], защиты информации и др. [1–4; 6; 15].

Известно эффективное применение параметра «энтропия» [18] простой стационарной цепи Маркова (энтропия стохастической матрицы (СМ), следуя терминологии [19]), вычисляемого по заданной СМ, в качестве математического средства анализа цепей Маркова – классификация, идентификация [19–20], определение линейной сложности [21] марковских последовательностей.

Случайную последовательность с m равновероятными независимыми состояниями можно рассматривать как цепь Маркова с переходной стохастической матрицей $P\left(p_{ij}\right)$, $p_{ij}=1/m$, порядка m, каждая строка которой имеет один и тот же предельный вектор $\pi(m)=\left(\pi_0,\pi_1,...,\pi_{m-1}\right)$, $\pi_i=1/m$ [22]. Энтропия $H(P)$ данной цепи при $p_{ij}=1/m$, $\pi_i=1/m$, $i=\overline{0,m-1}$ достигает максимального значения [18]. В соответствии с [23] признаком случайности конечной последовательности символов считается отсутствие в ней закономерности. ПСП, основанные на алгоритмах, реализующих различные рекуррент-

ные соотношения, имеют строения, отражающие закономерности соответствующих алгоритмов. В связи с этим по задаче анализа качества ПСП можно поставить вопрос: каким образом существующие закономерности в структуре ПСП заданной длины, генерируемой по заданному алгоритму, отобразить в соответствующее значение параметра «энтропия цепей Маркова», вычисленное на основе заданной ПСП.

Цель работы – показать возможность применения параметра энтропии простой стационарной цепи Маркова как критерия качества статистических свойств ПСП и предложить метод анализа многоразрядных ПСП заданной длины на основе данного критерия, позволяющий с точностью, определяемой длиной ПСП, оценивать приближение статистических свойств ПСП к равномерному закону случайной последовательности.

Постановка задачи анализа ПСП

Пусть задана периодическая ПСП с периодом, равным L, формируемая конгруэнтным генератором, основанным на алгоритме, реализующем некоторое рекуррентное соотношение в цифровом (десятичное представление) алфавите $S(L)=\left(s_0, s_1, ..., s_{L-1}\right)$ [1–6; 15]. Обозначим подобные ПСП через M_L. Пусть получена некоторая подпоследовательность ПСП $M(L)$ заданной длины $N<L$ с подмножеством элементов $S(N)=\left(s_0, s_1, ..., s_{N-1}\right)$. Данные подпоследовательности можно получить, например по алгоритмам [1–6; 15]; далее их будем обозначать как ПСП M_N. Для случая $N=L$ объектом анализа является ПСП M_L с множеством $S(L)$. Множество $S(N)=\left(s_0, s_1, ..., s_{N-1}\right)$, $N \le L$, упорядочим по возрастанию элементов и разобьем на m непересекающихся подмножеств

$$\left\{A_0 \quad A_1 \quad ... \quad A_{m-1}\right\}, \tag{1}$$

мощности которых равны

$$a_i \ge 1 \text{ и } \sum_{i=0}^{m-1} a_i = N, 2 \le m \le N. \tag{2}$$

Обозначим данные подмножества соответственно символами множества $Y=\left\{y_0 \quad y_1 \quad ... \quad y_{m-1}\right\}$, $2 \le m \le N$.

Введем в рассмотрение стохастический вектор вида

$$\pi(m)=\left(\pi_0, \pi_1, ..., \pi_{m-1}\right)=\left(\alpha_0/N, \alpha_1/N, ..., \alpha_{m-1}/N\right), \tag{3}$$

где a_i, $i=\overline{0, m-1}$ удовлетворяют условию (2). Энтропия вектора (3) вычисляется по формуле [18]

$$H\left(\pi(m)\right)=-\sum_{i=0}^{m-1}\pi_i\left(\log_2 \pi_i\right). \tag{4}$$

Максимальное значение $H\left(\pi(m)\right)=\log_2 m$ достигается при $\pi_i=1/m$, $i=\overline{0, m-1}$ [18].

Пусть ПСП вида M_N при заданном векторе (3), имеющем максимальную энтропию (4), поставлена в однозначное соответствие по некоторому алгоритму стохастической эргодической [22] матрице СМ $P=P\left(p_{ij}\right)$ с рациональными элементами, обладающей предельным [22] вектором, равным заданному вектору $\pi(m)$ вида (3). Элементы СМ P имеют вид

$$p_{ij} = \left(a_{ij} / \sum_{j=0}^{m-1} a_{ij} \right),$$

$$\sum_{j=0}^{m-1} a_{ij} = a_i,\ a_i \geq 1 \text{ и } \sum_{i=0}^{m-1} a_i = N,\ i = \overline{0, m-1}. \quad (5)$$

Вычислим энтропию по матрице P (обозначение энтропии $H(P)$) по формуле [18]:

$$H(P) = -\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \pi_i p_{ij} \log 2 \left(p_{ij} \right). \quad (6)$$

Энтропия $H(P)$ данной цепи при $p_{ij} = 1/m$, $\pi_i = 1/m$, $i = \overline{0, m-1}$ достигает максимального значения [18]:

$$H(P) = H_{\max} = \log_2 m. \quad (7)$$

Замечание 1. При $m = N$ матрица P является стохастической булевой и $H(P) = 0$.

Матрицу P со свойством (7) обозначим $P(H_{\max})$. Матрица $P(H_{\max})$ обладает свойством

$$\sum_{i=0}^{m-1} \sum_{j=0}^{m-1} a_{ij} = m^2, \quad (8)$$

где $a_{ij} = 1$, $i, j = \overline{0, m-1}$ [18].

Примем ограничение, учитывая замечание 1 и свойство (8): требуемый порядок m определяемой СМ задается при фиксированном N из условия – значение величины

$$\Delta = \left| m^2 - N \right| \quad (9)$$

минимально. Параметр (9) характеризует величину разности между количествами существующих вероятностных переходов, представленных в матрице $P\left(H_{\max}\right)$ и сравниваемой P.

Введем параметр

$$\Delta_н = H_{\max} - H\left(P\right), \quad (10)$$

который будем интерпретировать как характеризующий меру отличия заданной ПСП M_N от случайной m-значной последовательности с равномерным распределением. Максимальное значение параметра (10) равно $\log_2 m$, что следует из (7) и замечания 1. Из (3), (5), (7) и замечания 1 следует, что точность вычисления параметра (10) определяется величиной N и точностью представления значений $\log_2 m$ [24] (с точностью порядка 10^{-4}).

Примем: если при значениях величин H_{max}, представленных с точностью не менее 10^{-4}, параметр $\Delta_н \leq N^{-1}$, то сравниваемые энтропии H_{max} и $H(P)$ не различаются. Ограничение вида $3 \leq N \leq 10^4$ определяется ограничением на точность представления элементов $p_{ij} = \left(a_{ij} / \sum_{j=0}^{m-1} a_{ij} \right)$, вычисляемых СМ P для экспериментальных приложений.

Модифицируем параметр (10) – приведем его к виду

$$k_{пр} = \Delta_н 100 \cdot \% / H_{\max} . \quad (11)$$

Параметр (11) (далее $k_{пр}$ – коэффициент различия, характеризующий качество ПСП) определяет меру отличимости качества заданной ПСП от случайной *m*-значной равномерно распределенной последовательности, характеризуемой матрицей $P(H_{\max})$, и позволяет отображать выраженную в процентах меру качества.

Требуется представить метод количественного анализа качества ПСП M_N, $N \le L$ на основе параметров (10), (11) при ограничении (9), позволяющий оценивать качество по (10) с точностью не менее N^{-1}.

Этапы решения задачи

Этап 1. Построение стохастической эргодической матрицы, однозначно соответствующей заданной ПСП M_N при ограничении вида (3), (5), (9).

Пусть заданы ПСП вида M_N, множество $S(N)$ элементов этой ПСП, величина *m* в соответствии с (9). Зададим разбиение вида (1), (2) множества $S(N)$ и множество $Y = \{y_0 \quad y_1 \quad ... \quad y_{m-1}\}$. Этому разбиению поставим в однозначное соответствие вектор (3). Отметим: разбиение вида (1), (2) задается таким, чтобы величина энтропии (4) вектора (3) при величинах *N*, *m* была максимальной.

Реализуем однозначное отображение $\lambda : S \to Y$. Полученную соответствующую последовательность длины *N,* состоящую из элементов множества *Y* обозначим через $\beta(N)$. Последовательности $\beta(N)$ поставим в однозначное соответствие матрице с рациональными элементами – матрице относительных частот $P' = (p'_{ij}) = (a_{ij} / a_i)$ порядка *m*, где a_i – число вхождений буквы y_i в последовательности длины *N*, $a_i \ge 1$, a_{ij} – число вхождений пары стоящих рядом букв $y_i y_j$, $i, j = \overline{0, m-1}$ (считаем, что за y_N следует y_1). Матрица P', получаемая по последовательности $\beta(N)$, – стохастическая эргодическая, обладающая свойствами [25]:

1) элементы $p'_{ij} = \left(a_{ij} / \sum_{j=0}^{m-1} a_{ij} \right)$,

$$\sum_{j=0}^{m-1} a_{ij} = \sum_{j=0}^{m-1} a_{ji} = a_i \ge 1 \text{ и } \sum_{i=0}^{m-1} a_i = N,\ i, j = \overline{0, m-1}; \quad (12)$$

2) предельный стохастический вектор матрицы P' равен вектору вида (3):

$$\overline{\pi_{пр}} = \left(\frac{a_0}{N}; \frac{a_1}{N}; ...; \frac{a_{m-1}}{N} \right). \quad (13)$$

Заметим: свойство (12) включает свойство (8).

Замечание 2. Стохастическая матрица $P(H_{\max})$ заданного порядка *m* обладает свойствами (12), (13).

Пример 1, иллюстрирующий реализацию этапа 1. Пусть построена ПСП M_L с периодом $L = 31$ на линейном регистре сдвига [2] (ЛРС), соответствующем примитивному полиному $f(x) = x^5 + x^4 + x^3 + x^2 + 1$ (*M*-последовательность [8]), задана $m = 6$ в соответствии с (9). В десятичном представлении ПСП M_L имеет вид M_{L1} =($s_1 = 31$, $s_2 = 15$,

$s_3 = 23$, $s_4 = 27$, $s_5 = 29$, $s_6 = 14$, $s_7 = 7$, $s_8 = 3$, $s_9 = 17$, $s_{10} = 8$, $s_{11} = 20$, $s_{12} = 10$, $s_{13} = 21$, $s_{14} = 26$, $s_{15} = 13$, $s_{16} = 22$, $s_{17} = 11$, $s_{18} = 5$, $s_{19} = 2$, $s_{20} = 1$, $s_{21} = 16$, $s_{22} = 24$, $s_{23} = 12$, $s_{24} = 6$, $s_{25} = 19$, $s_{26} = 9$, $s_{27} = 4$, $s_{28} = 18$, $s_{29} = 25$, $s_{30} = 28$, $s_{31} = 30$). Пусть при $m = 6$ заданный предельный вектор (3) с максимальной энтропией имеет вид (5/31, 5/31, 5/31, 5/31, 5/31, 6/31), его энтропия вычисляется по (4). Соответствующее разбиение (1) задается в виде $S = \{A_0 = \{1, ..., 5\}$, $A_1 = \{6, ..., 10\}$, $A_2 = \{11, ..., 15\}$, $A_3 = \{16, ..., 20\}$, $A_4 = \{21, ..., 25\}$, $A_5 = \{26, ..., 31\}\}$. Полученная последовательность – $\beta(L = 31) = (y_5, y_2, y_4, y_5, y_5, y_2, y_1, y_0, y_3, y_1, y_3, y_1, y_4,$ $y_5, y_2, y_4, y_2, y_0, y_0, y_0, y_3, y_4, y_2, y_1, y_3, y_1, y_0, y_3, y_4, y_5, y_5)$. По данной последовательности определяется однозначно СМ – обозначение P_1':

$$P_1' = \begin{pmatrix} 2/5 & 0 & 0 & 3/5 & 0 & 0 \\ 2/5 & 0 & 0 & 2/5 & 1/5 & 0 \\ 1/5 & 2/5 & 0 & 0 & 2/5 & 0 \\ 0 & 3/5 & 0 & 0 & 2/5 & 0 \\ 0 & 0 & 2/5 & 0 & 0 & 3/5 \\ 0 & 0 & 3/6 & 0 & 0 & 3/6 \end{pmatrix}.$$

Пусть подпоследовательность длины N = 20, с обозначением M_{N1}, есть первые 20 элементов ПСП M_{L1}, $m = 4$ (задана в соответствии с (9)); заданный предельный вектор (3) с максимальной энтропией имеет вид (5/20, 5/20, 5/29, 5/20), соответствующее разбиение (1) имеет вид: $S(N = 20) = \{A_0 = \{1, 2, 3, 5, 7\}$, $A_1 = \{8, 10, 11, 13, 14\}$, $A_2 = \{15, 17, 20, 21, 22\}$, $A_3 = \{23, 26, 27, 29, 31\}\}$.

Последовательность $\beta(N = 20) = (y_3, y_2, y_3, y_3, y_3, y_1, y_0, y_0, y_2, y_1, y_2, y_1,$ $y_2, y_3, y_1, y_2, y_1, y_0, y_0, y_0)$. По данной последовательности определяется однозначно СМ – обозначение P_2':

$$P_2' = \begin{pmatrix} 3/5 & 0 & 1/5 & 1/5 \\ 2/5 & 0 & 3/5 & 0 \\ 0 & 3/5 & 0 & 2/5 \\ 0 & 2/5 & 1/5 & 2/5 \end{pmatrix},$$ обладающая свойствами (12), (13).

Этап 2. Вычисление энтропии (6) по стохастической матрице P' и параметра (7).

Вычисляются энтропия по формуле (6), по полученным на этапе 1 СМ вида P' (для примера 1 $H(P_1') = 1{,}1543$ и $H(P_2') = 1{,}2087$) и максимальная энтропия СМ $P(H_{\max})$ заданных порядков по формуле (7) (для примера 1 для / $m = 6$ $P(H_{\max}) = 2{,}585$ и для $m = 4$ $P(H_{\max}) = 2$).

Этап 3. Вычисление по результатам этапа 2 параметров (10), (11).

Для примера 1 при $H(P_1')$ параметры (10), (11) принимают значения 1,431 и 55,3 %; при $H\left(P_2'\right)$ – значения 0,791 и 39,6 %.

Возможность оценки качества ПСП по критерию (10) с точностью $1/N$ определяет.

Утверждение 1. Пусть заданы ПСП вида M_N, число m из условия (9), вектор вида (13) с максимальной энтропией и однозначно соответствующая заданной ПСП стохастическая матрица P' порядка m. Тогда параметр (10) определяет меру различимости с точностью не менее $1/N$ заданной ПСП со случайной m-значной равномерно распределенной последовательностью, характеризуемой матрицей $P(H_{\max})$ порядка m.

Справедливость утверждения 1 следует из замечаний 1–3 и свойств (8), (12) и (13).

Заключение

Предложенный критерий (10) анализа качества ПСП при ограничении (9) определяет меру различимости с точностью не менее $1/N$ заданной ПСП вида M_N, длины $N \le L$, со случайной m-значной равномерно распределенной последовательностью. Представленные этапы решения задачи составляют следующий трехэтапный алгоритмический метод анализа качества ПСП вида M_N, $N \le L$.

Входные данные: ПСП вида M_N, $N \le L$, величина m.

Этап 1. Определение вектора (13) с максимальной энтропией при заданных N и m и построение стохастической эргодической матрицы P', однозначно соответствующей заданной ПСП.

Этап 2. Вычисление параметров (6) и (7).

Этап 3. Вычисление параметров (10), (11).

Величина параметра (11) определяет процент отличия заданной ПСП от случайной последовательности. Данная характеристика позволяет ранжировать по качеству исследуемые ПСП вида M_N с точностью $1/N$.

Литература

1. *Кнут Д.* Искусство программирования для ЭВМ: в 3 т. – 3-е изд.; пер. с анг. – М.: Мир, 1998. Т. 2.

2. *Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушкин А.В.* Основы криптографии. – М.: Гелиос АРВ, 2002. – 480 с.

3. *Шнайер Брюс*. Прикладная криптография: протоколы, алгоритмы и исходный код на языке С. – 2-е юбил. изд.: пер. с анг. – СПб.: Альфа-книга, 2017. – 1040 с.

4. *Иванов М.А. Чугунков И.В*. Теория, применение и оценка качества генераторов псевдослучайных последовательностей. – М.: КУДИЦ-ОБРАЗ, 2003. – 240 с.

5. *Delgado-Mohatar O., Fuster-Sabater A., Sierra J.M.* Performance evaluation of highly efficient techniques for software implementation of LFSR // Computers & Electrical Engineering. 2011. Vol. 37, iss. 6. – Pp. 1222–1231.

6. *Hu C.Q., Liao X.F., Cheng X.Z*. Verifiable multi-secret sharing based on LFSR sequences // Theoretical Computer Science. 2012. Vol. 445. – Pp. 52–62.

7. *Доценко В.И., Фараджев Р.Г*. Анализ и свойства последовательностей максимальной длины // Автоматика и телемеханика. 1969. № 11. – С. 119–167.

8. *Сенин А.И.* Взаимно-корреляционные свойства псевдослучайных и родственных последовательностей // ТИИЭР. 1980. Т. 68. № 5. – С. 59–90.

9. Статистическая проверка случайности двоичных последовательностей методами NIST. Habr. 2006–2023. – Режим доступа: https://habr.com/ru/company/securitycode/blog/237695/

10. Dieharder. A Random Number Test Suite / *Robert G. Brown*. 2023. – URL: http://webhome.phy.duke.edu/~rgb/General/dieharder.php

11. *Latypov R., Stolov E*. True Random Generators and Hidden Transfer of Keys // International Russian Automation Conference (RusAutoCon). – Sochi, Russia, 2019. – Pp. 1–5.

12. *Ширшова Д.В*. Метод и комплекс программ нахождения максимальной длины выборки статистически однородных двоичных последовательностей для имитационного моделирования: дис. … канд. техн. наук. – Казань, 2019.

13. *Anikin Igor V., Khaled Alnajjar*. Secure Gamma Generation for Stream Cipher based on Fuzzy Logic // International Conference on Information Technology and Nanotechnology (ITNT). – Samara, Russia, 2020. – Pp. 1–7.

14. *Dubrova E.A*. Scalable Method for Constructing Galois NLFSRs With Period 2^n-1 Using Cross-Join Pairs // IEEE TRANSACTIONS ON INFORMATION THEORY. 2013. Vol. 59, iss. 1. – Pp. 703–709.

15. *Захаров В.М., Шалагин С.В., Песошин В.А., Эминов Б.Ф.* Автоматная модель представления нелинейных псевдослучайных последовательностей с функцией выхода на основе системы инъективных преобразований // Кибернетика и программирование. 2017. № 5. – С. 64–78.

16. *Магомедова Е.С., Магомедов Р.И.* Стохастическая модель гонки вооружений // Вестник ДГУ. Сер. 1: Естественные науки. 2022. Т. 37, вып. 4. – С. 30–35.

17. *Магомедов Р.И., Магомедов И.И., Магомедова Е.С*. Моделирование изменения количества воды в водохранилище с помощью стохастического дифференциального уравнения // Вестник ДГУ. Сер. 1: Естественные науки. 2020. Т. 35, вып. 1. – С. 53–59.

18. *Хинчин А.Я*. Понятие энтропии в теории вероятностей // Успехи математических наук. 1953. № 3 (55). – С. 3–20.

19. *Захаров В.М., Нурмеев Н.Н., Салимов Ф.И., Шалагин С.В.* Классификация стохастических эргодических матриц методами кластерного и дискриминантного анализа // Исследования по информатике. – Казань: Отечество, 2000. – С. 91–106.

20. *Захаров В.М., Шалагин С.В., Эминов Б.Ф*. Многопараметрический анализ автоматных марковских моделей на основе методов кластерного и дискриминантного анализа // Вестник технологического университета. 2019. Т. 22, № 9. – С. 109–113.

21. *Захаров В.М., Эминов Б.Ф*. Статистический анализ линейной сложности регулярных цепей Маркова // Исследования по информатике. Вып. 10. ИПИ АН РТ. – Казань: Отечество, 2006. – С. 37–50.

22. *Кемени Дж., Снелл Дж*. Конечные цепи Маркова. – М.: Наука, 1970. – 272 с.

23. *Колмогоров А.Н*. Три подхода к определению понятия «количество информации» // Проблемы информации. 1965. Т. 1. – С. 3–11.

24. Математические таблицы // MathTask. 2023. – Режим доступа: http://www.mathtask.ru/0014-mathematical-tables.php

25. *Столов Е.Л*. Об одном классе генераторов псевдомарковских цепей // Исследования по прикладной математике. Вып. 8. – Казань: Изд-во Казан. ун-та, 1980. – С. 66–71.

Поступила в редакцию 2 марта 2023 г.

UDC 519.217.2

DOI: 10.21779/2542-0321-2023-38-2-61–68

The Analysis of Pseudorandom Sequences of a Given Length by «Markov Chain Entropy» Criterion

V.M. Zakharov, S.V. Shalagin

Kazan National Research Technical University named after A.N. Tupolev – KAI; Russia, 420111, Kazan, Karl Marks st., 10; Gilvv@mail.ru, SShalagin@mail.ru

Abstract. The problem of analyzing the quality of a pseudorandom sequence (PRS) of a given length is considered in the article. The quality of this PRS is estimated as the degree of approximation of its statistical properties to the properties of a random sequence. The problem of mapping regularities in the PRS structure determined by the formation algorithm into the corresponding value of the entropy parameter of Markov chain is solved. The parameter is calculated using a stochastic matrix that uniquely corresponds to the studied PRS, parameter value is a measure of quality. The research has shown that deviation of the parameter from the specified maximum value of random sequence entropy is a measure of the PRS quality, estimated with an accuracy proportional to the analyzed length of PRS.

Keywords: analysis of pseudorandom sequences, «Markov chain entropy» criterion, stochastic matrix, accuracy of assessment.

Received 2 March 2023