

## Закон квадратного корня в задаче выявления вкраплений в цепях Маркова с неизвестной матрицей переходных вероятностей<sup>†</sup>

© 2017 г. А. В. Волгин\*

Рассматривается классическая модель вкраплений в простую двоичную цепь Маркова с неизвестной матрицей переходных вероятностей. Найдены асимптотические соотношения между длиной отрезка исходной последовательности и объемом вкраплений, при которых предложенный статистический критерий состоятелен.

**Ключевые слова:** цепь Маркова, вкрапления, статистический критерий

### 1. Введение

Задача обнаружения наличия вкраплений в случайных последовательностях рассматривается во многих работах (см., например, [2, 5–9]). При этом исходная последовательность может иметь различные типы и параметры распределения. В [2], [5] и [10], например, исследуется последовательность, полученная по полиномиальной схеме с известными вероятностями исходов, в [9] рассматривается марковская модель зависимости с известной матрицей переходных вероятностей. Для обеих моделей установлено, что гарантированно обнаружить факт наличия независимых вкраплений можно только в случае, когда длина отрезка вкрапливаемой последовательности растет по порядку быстрее корня из длины отрезка исходной последовательности. В некоторых публикациях (см., например, [9]) данный результат называют «законом квадратного корня».

В [8] анонсировано утверждение о «законе квадратного корня», которое качественно демонстрирует возможность использования статистики типа хи-квадрат для идентификации простой конечной неразложимой и ациклической цепи Маркова с неизвестной матрицей переходных вероятностей в рамках следующей модели вкраплений: к каждому элементу исходной последовательности применяется случайное преобразование. При этом рассматривается схема серий и преобразования являются независимыми.

<sup>†</sup>Статья публикуется по рекомендации Оргкомитета конференции STCrypt'2016.

\*Место работы: Московский технологический университет, e-mail: [artem.volgin@bk.ru](mailto:artem.volgin@bk.ru)

Далее рассматривается классическая модель вкрапления (см. [6]) в простые цепи Маркова с неизвестной матрицей переходных вероятностей, основанная на LSB-методе (см. [11]). Приводятся асимптотические соотношения между длиной отрезка исходной последовательности и объемом вкраплений, при которых предложенный статистический критерий состоятелен.

## 2. Обнаружение вкраплений в цепь Маркова

Пусть  $n \in \mathbb{N}$ ,  $X = \{X_0, X_1, \dots\}$  – простая конечная стационарная неразложимая и ациклическая цепь Маркова с множеством состояний  $\mathcal{A} = \{1, \dots, N\}$  и фиксированной матрицей переходных вероятностей  $\Pi = (\pi_{ab})_{N \times N}$ ,  $a, b \in \mathcal{A}$ . Стационарное распределение цепи обозначим через

$$\pi = (\pi_1, \dots, \pi_N), \quad \pi_i \in (0, 1), \quad i = 1, \dots, N, \quad \sum_{i=1}^N \pi_i = 1. \quad (1)$$

Рассмотрим последовательность процедур внесения вкраплений в цепь Маркова  $X$ , в которой используются случайные последовательности  $Z = \{Z_{0,n}, Z_{1,n}, \dots, Z_{n-1,n}\}$  и  $\Gamma = \{\gamma_{0,n}, \gamma_{1,n}, \dots, \gamma_{n-1,n}\}$ ,  $\mathbf{P}\{Z_{i,n} \in \mathcal{A}\} = \mathbf{P}\{\gamma_{i,n} \in \{0, 1\}\} = 1$ ,  $i = 0, \dots, n-1$ . При этом элементы последовательностей  $Z$  и  $\Gamma$  являются независимыми одинаково распределенными случайными величинами, и для любого  $i = 0, \dots, n-1$  верны равенства

$$\mathbf{P}\{Z_{i,n} = a\} = p_a, \quad \mathbf{P}\{\gamma_{i,n} = 0\} = \tau(n), \quad p_a, \tau(n) \in (0, 1), \quad a \in \mathcal{A}. \quad (2)$$

Далее вместо  $\tau(n)$  будем использовать обозначение  $\tau$ , подразумевая при этом, что данный параметр зависит от  $n$ . Также в обозначении элементов схем серий  $Z$  и  $\Gamma$  будем использовать только первый индекс.

В результате внесения вкраплений образуется отрезок последовательности  $Y = \{Y_0, Y_1, \dots, Y_{n-1}\}$ :

$$Y_i = \begin{cases} X_i, & \text{если } \gamma_i = 1, \\ Z_i, & \text{если } \gamma_i = 0, \end{cases} \quad i = 0, \dots, n-1. \quad (3)$$

Постановка задачи заключается в следующем. Наблюдается отрезок последовательности  $\{Y_0, \dots, Y_{n-1}\}$  длины  $n \in \mathbb{N}$ . Относительно способа образования данного отрезка выдвигаются две сложные гипотезы  $H_0: \tau = 0$  и  $H_1: \tau > 0$  об отсутствии и наличии вкраплений в цепь Маркова  $X$  соответственно. При обеих гипотезах будем предполагать, что матрица  $\Pi$ , а также величины  $\tau$ ,  $p_a$ ,  $a \in \mathcal{A}$ , и  $\pi$  неизвестны. Задача заключается в построении критерия различения гипотез  $H_0$  и  $H_1$ .

Определим вероятностную меру при гипотезе  $H_1$ . Рассмотрим вероятностное пространство  $(\Omega, \mathfrak{A}, \mathbf{P}_1)$ , где  $\Omega = \{(y_0, y_1, \dots, y_{n-1}) : y_j \in \mathfrak{A}, j = 0, \dots, n-1\}$ ,  $n \in \mathbb{N}$ ,  $\mathfrak{A}$  – совокупность всех подмножеств множества  $\Omega$ . Тогда для любого фиксированного

набора  $(y_0, \dots, y_{n-1}) \in \Omega$

$$\begin{aligned} \mathbf{P}_1\{Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}\} &= p_{y_0} \cdots p_{y_{n-1}} \tau^n + \\ + \sum_{k=1}^n \sum_{1 \leq i_0 < \dots < i_{k-1} \leq n-1} \mathbf{P}\{X_{i_1} = y_{i_1}, \dots, X_{i_k} = y_{i_k}\} & p_{y_{j_1}} \cdots p_{y_{j_{n-k}}} \\ &\times (1 - \tau)^k \tau^{n-k}, \end{aligned} \quad (4)$$

где  $\mathbf{P}\{X_{i_1} = y_{i_1}, \dots, X_{i_k} = y_{i_k}\} = \pi_{y_{i_1}}^{(i_2-i_1)} \cdots \pi_{y_{i_{k-1}} y_{i_k}}^{(i_k-i_{k-1})}$  и  $\pi_{y_{i_j} y_{i_{j+1}}}^{(i_{j+1}-i_j)}$ ,  $j = 1, \dots, k-1$ , — это вероятность перехода из состояния  $y_{i_j}$  в состояние  $y_{i_{j+1}}$  за  $i_{j+1} - i_j$  шагов  $\{j_1, \dots, j_{n-k}\} = \{0, \dots, n-1\} \setminus \{i_0, \dots, i_{k-1}\}$ . Проверим корректность определения вероятностной меры  $\mathbf{P}_1$ . Из уравнений (4) следует, что

$$\begin{aligned} \sum_{(y_0, \dots, y_{n-1}) \in \Omega} \mathbf{P}_1\{Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}\} &= \tau^n + \sum_{k=1}^{n-1} \binom{n}{k} (1 - \tau)^k \tau^{n-k} = \\ &= (1 - \tau + \tau)^n = 1. \end{aligned}$$

В [9] для цепи Маркова с известной матрицей переходных вероятностей приведены асимптотические (при  $n \rightarrow \infty$ ) условия, при выполнении которых существует процедура различения гипотез  $H_0$  и  $H_1$ . Зафиксируем  $n \in \mathbb{N}$  и обозначим через  $d_n(\tau)$  расстояние Кульбака–Лейблера между распределениями отрезка  $Y = (Y_0, \dots, Y_{n-1})$  при гипотезах  $H_0$  и  $H_1$ :

$$d_n(\tau) = \sum_{y=(y_0, \dots, y_{n-1}) \in \mathcal{A}^n} \mathbf{P}_1(Y = y | H_0) \ln \frac{\mathbf{P}_1(Y = y | H_0)}{\mathbf{P}_1(Y = y | H_1)}.$$

**Теорема 1** (см. [9]). Пусть  $n \rightarrow \infty$ . Тогда:

1) если  $\sqrt{n}\tau \rightarrow \infty$ , то существует такой критерий, который различает гипотезы  $H_0$  и  $H_1$  с вероятностями ошибок первого и второго рода  $\alpha_n \in (0, 1)$ ,  $\beta_n \in (0, 1 - \alpha_n)$  соответственно, причем для любых  $\alpha^* \in (0, 1)$ ,  $\beta^* \in (0, 1 - \alpha^*)$  существует такое число  $n_0 \in \mathbb{N}$ , что для всех  $n > n_0$  справедливы неравенства  $\alpha_n < \alpha^*$ ,  $\beta_n < \beta^*$ ,

2) если  $\sqrt{n}\tau \rightarrow 0$ , то  $d_n(\tau) \rightarrow 0$ ,

3) если  $\sqrt{n}\tau \rightarrow \kappa \in (0, \infty)$ , то  $d_n(\tau) \leq C\kappa^2$  для любого  $n \in \mathbb{N}$ , где  $C > 0$  — некоторая постоянная.

**Замечание 1.** Из теоремы 1 следует, что гипотезы  $H_0$  и  $H_1$  различимы (в терминах метрики Кульбака–Лейблера) только в том случае, когда выполнено условие  $\tau\sqrt{n} \rightarrow \infty$ . В [9] приводится состоятельный критерий различения  $H_0$  и  $H_1$ , основанный на статистике

$$\sqrt{n} \left| \frac{1}{n-1} \sum_{i=0}^{n-2} \mathbf{I}\{Y_i = a, Y_{i+1} = b\} - \pi_a \pi_{ab} \right|$$

для таких значений  $a, b \in \mathcal{A}$ , что

$$\mathbf{P}_1\{Y_0 = a, Y_1 = b | H_0\} \neq \mathbf{P}_1\{Y_0 = a, Y_1 = b | H_1\}.$$

Подчеркнём, что для применения критерия требуется, чтобы матрица переходных вероятностей цепи Маркова была известна.

Далее рассмотрим критерий различения гипотез  $H_0$  и  $H_1$  в случае, когда матрица переходных вероятностей неизвестна. В [8] анонсирован подход к различению гипотез  $H_0$  и  $H_1$  на основе статистики следующего вида:

$$S = \sum_{a,b,c \in \mathcal{A}} \frac{(\nu_{abc} - \nu_{ab}\nu_{bc}/\nu_b)^2}{\nu_{ab}\nu_{bc}/\nu_b}, \quad (5)$$

где

$$\nu_{abc} = \sum_{i=0}^{n-3} \mathbf{I}\{Y_i = a, Y_{i+1} = b, Y_{i+2} = c\},$$

$$\nu_{ab} = \sum_{i=0}^{n-2} \mathbf{I}\{Y_i = a, Y_{i+1} = b\}, \quad \nu_a = \sum_{i=0}^{n-1} \mathbf{I}\{Y_i = a\}.$$

Поясним смысл статистики (5). В [3] описан критерий согласия со сложной гипотезой  $H_0$  с использованием стандартной статистики типа хи-квадрат:

$$\chi_n^2(\theta) = \sum_{a,b,c \in \mathcal{A}} \frac{(\nu_{abc} - \nu_{ab}p_{ab:c}(\theta))^2}{\nu_{ab}p_{ab:c}(\theta)}, \quad (6)$$

где  $p_{ab:c}(\theta) = \mathbf{P}_1\{Y_i = c | Y_{i-2} = a, Y_{i-1} = b\}$ ,  $i \geq 2$ ,  $\theta$  – неизвестный параметр, от которого зависят элементы матрицы переходных вероятностей  $\Pi$ .

Статистика (6) непригодна для проведения расчетов, поскольку она зависит от неизвестного параметра  $\theta$ . Поэтому в [3] предлагается подход, который основан на замене величин  $p_{ab:c}(\theta)$  их оценками  $\hat{p}_{ab:c}$ ,  $a, b, c \in \mathcal{A}$ , полученными с использованием наблюдаемого отрезка последовательности  $Y_0, \dots, Y_{n-1}$ . Также в [3] показано, что оценки  $\hat{p}_{ab:c} = \nu_{bc}/\nu_b$ ,  $a, b, c \in \mathcal{A}$ , являются оценками максимального правдоподобия, и сформулирована теорема (ее доказательство приведено в [4]) об асимптотической (при  $n \rightarrow \infty$ ) сходимости статистики (5) к распределению хи-квадрат с  $N(N - 1)^2$  степенями свободы.

Обозначим через  $\varkappa_\alpha$  квантиль уровня  $\alpha$  распределения хи-квадрат с  $N(N - 1)^2$  степенями свободы. Рассмотрим критерий согласия с гипотезой  $H_0$ , основанный на статистике (5), размера  $\alpha$  :

$$\text{принимается гипотеза } \begin{cases} H_0, \text{ если } S < \varkappa_{1-\alpha}, \\ H_1, \text{ если } S \geq \varkappa_{1-\alpha}. \end{cases} \quad (7)$$

Ниже будет доказана теорема о состоятельности критерия (7).

**Теорема 2.** *Если существуют состояния  $a, b, c \in \mathcal{A}$ , для которых выполняется условие*

$$\sqrt{n}\tau(1 - \tau)^2 (C_1(1 - \tau) + C_2\tau) \rightarrow \infty, \quad n \rightarrow \infty, \quad (8)$$

где

$$C_1 = p_b\pi_b (\pi_{a*c} - \pi_a\pi_{bc}) + \pi_a\pi_b p_b (\pi_{bc} - \pi_c), \quad C_2 = p_b (\pi_{a*c} - \pi_a\pi_c),$$

$$\pi_{a*c} = \mathbf{P}_1\{Y_0 = a, Y_2 = c\},$$

то критерий (7) является состоятельным.

В [6] рассматривается задача выявления вкраплений в двоичных цепях Маркова с матрицами переходных вероятностей специального вида:

$$\Pi = \frac{1}{2} \begin{pmatrix} 1 + \lambda & 1 - \lambda \\ 1 - \lambda & 1 + \lambda \end{pmatrix}, \quad \lambda \in (-1, 1). \quad (9)$$

В этом случае для любого  $\lambda \in (-1, 1)$  стационарное распределение является равномерным. При  $a = b = c = 0$  величины  $C_1$  и  $C_2$  из теоремы 2 принимают следующие значения:  $C_1 = \frac{\lambda^2 p_0}{8}$ ,  $C_2 = \frac{\lambda^2 p_0}{4}$ .

**Следствие 1.** Пусть существует предел

$$\lim_{n \rightarrow \infty} \tau(n) = \text{const} \in [0, 1]$$

и матрица переходных вероятностей цепи Маркова имеет вид (9). Тогда при выполнении условия

$$\sqrt{n}\tau(1 - \tau)^2 \rightarrow \infty, \quad n \rightarrow \infty, \quad (10)$$

критерий (7) является состоятельным.

**Замечание 2.** Если в условиях теоремы 2 константы  $C_1$  и  $C_2$  положительны, то критерий (7) является состоятельным при одновременном выполнении двух условий:

$$\sqrt{n}\tau \rightarrow \infty, \quad n \rightarrow \infty, \quad (11)$$

и

$$\sqrt{n}(1 - \tau)^2 \rightarrow \infty, \quad n \rightarrow \infty. \quad (12)$$

Условие (11) совпадает с условием 1) теоремы 1 и означает, что число вкраплений должно превосходить квадратный корень из длины наблюдаемого отрезка. Условие (12) не требуется в теореме 1 и является в некотором смысле противоположным условию (11), поскольку означает, что вкраплений не должно быть «слишком много». Объясняется это тем, что при внесении большого числа вкраплений распределение отрезка последовательности  $Y$  сближается с полиномиальным распределением с вероятностями  $(p_a, a \in \mathcal{A})$ . В этом случае отрезок  $Y$  будет являться цепью Маркова с глубиной зависимости 0. Статистика (5) не позволяет различать гипотезы о цепях Маркова с глубинами зависимости 0 и 1.

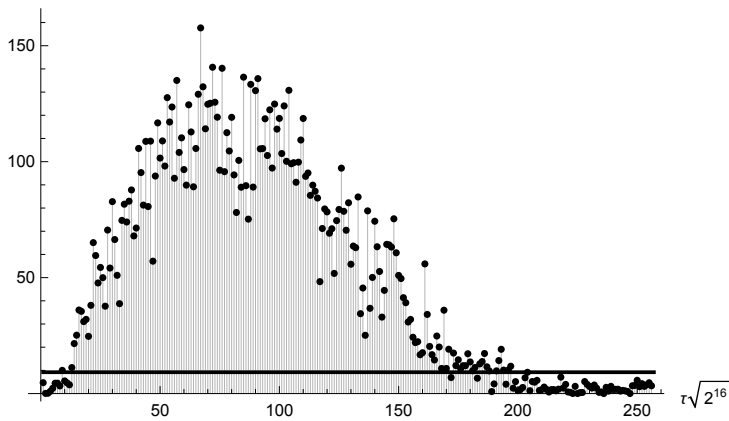
### 3. Результаты экспериментальных исследований

Теорему 2 наглядно иллюстрирует следующий эксперимент. Пусть отрезок  $Y$  имеет длину  $n = 2^{16}$ , матрица переходных вероятностей цепи Маркова  $X$  имеет вид  $\Pi = \begin{pmatrix} 1/5 & 4/5 \\ 3/5 & 2/5 \end{pmatrix}$ , стационарное распределение  $\pi = (3/7, 4/7)$ , вкрапливаемый отрезок последовательности  $Z$  имеет параметры  $p_0 = 3/8$  и  $p_1 = 5/8$ . В таблице 1 и на рисунке 1 приведены значения статистики  $S$  критерия (7) для различных  $\tau$  при пороговом значении  $\varkappa_{0,99} = 9,21$ .

Значения статистики критерия (7) для различных  $\tau$

№ испытания	$n$	$\tau$	$S$
1	$2^{16}$	0	7,7
2	$2^{16}$	$1/\sqrt{2^{16}} \approx 0,004$	0,8
3	$2^{16}$	$5/\sqrt{2^{16}} \approx 0,02$	4,4
4	$2^{16}$	$10/\sqrt{2^{16}} \approx 0,039$	7,6
5	$2^{16}$	$50/\sqrt{2^{16}} \approx 0,195$	97,6
6	$2^{16}$	$100/\sqrt{2^{16}} \approx 0,703$	101,7
7	$2^{16}$	$200/\sqrt{2^{16}} \approx 0,938$	2,6
8	$2^{16}$	1	3,9

График зависимости значений статистики  $S$  от  $\tau\sqrt{2^{16}}$   
Значение статистики  $S$



Результаты экспериментов показывают, что критерий не различает гипотезы  $H_0$  и  $H_1$  в случаях, когда вероятность вкрапления близка либо к 0, либо к 1 (испытания 1–4, а также 7–8). В остальных случаях на основании критерия принимается верное решение (испытания 5,6).

#### 4. Доказательство теоремы 2

Докажем теорему 2. Зафиксируем состояния  $a, b, c \in \mathcal{A}$ , для которых выполняется условие (8). Введем обозначения:

$$\mathbf{E}_1 = \mathbf{E} \{ \cdot | H_1 \}, \quad \mathbf{D}_1 = \mathbf{D} \{ \cdot | H_1 \},$$

$$S_{abc} = \frac{|\nu_{abc} - \nu_{ab}\nu_{bc}/\nu_b|}{\sqrt{\nu_{ab}\nu_{bc}/\nu_b}}.$$

Заметим, что для вероятности ошибки второго рода критерия (7) справедливы следующие оценки:

$$\begin{aligned} \mathbf{P}_1 \{S < \varkappa_{1-\alpha}\} &< \mathbf{P}_1 \left\{ \bigcap_{u,v,w \in \mathcal{A}} \{S_{uvw} < \sqrt{\varkappa_{1-\alpha}}\} \right\} \leq \\ &\leq \mathbf{P}_1 \{S_{abc} < \sqrt{\varkappa_{1-\alpha}}\}. \end{aligned} \quad (13)$$

Поэтому с учетом (13) для доказательства теоремы 2 достаточно показать, что

$$\mathbf{P}_1 \{S_{abc} < \sqrt{\varkappa_{1-\alpha}}\} \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Заметим, что для величины  $S_{abc}$  справедливо представление

$$\frac{S_{abc}}{\sqrt{n}} = \left| \frac{\frac{1}{n}\nu_{abc} - p_*}{\sqrt{p_*}} - \frac{\frac{1}{n}\nu_* - p_*}{\sqrt{p_*}} + \frac{\frac{1}{n}\nu_{abc} - \frac{1}{n}\nu_*}{\sqrt{\frac{1}{n}\nu_*}} \cdot \frac{\sqrt{p_*} - \sqrt{\frac{1}{n}\nu_*}}{\sqrt{p_*}} \right|, \quad (14)$$

где

$$\begin{aligned} \nu_* &= \nu_{ab}\nu_{bc}/\nu_b, \\ p_* &= ((1-\tau)^2\pi_a\pi_{ab} + \tau(1-\tau)(\pi_a p_b + \pi_b p_a) + \tau^2 p_a p_b) \times \\ &\times ((1-\tau)^2\pi_b\pi_{bc} + \tau(1-\tau)(\pi_b p_c + \pi_c p_b) + \tau^2 p_b p_c) \times \\ &\times (\pi_b(1-\tau) + p_b\tau)^{-1}. \end{aligned} \quad (15)$$

Правая часть (14) имеет громоздкий вид и сложна для дальнейших вычислений. Для ее упрощения рассмотрим следующий прием: заменим статистику  $S_{abc}$  на  $\widehat{S}_{abc}$ , где

$$\widehat{S}_{abc} = \sqrt{n} \cdot \frac{|\frac{1}{n}\nu_{abc} - p_*|}{\sqrt{p_*}},$$

и покажем, что выполнено условие

$$\mathbf{P}_1 \{\widehat{S}_{abc} < \sqrt{\varkappa_{1-\alpha}}\} \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Следующие две леммы используются для доказательства обоснованности замены  $S_{abc}$  на  $\widehat{S}_{abc}$ .

**Лемма 1.** Для любых  $a, b, c \in \mathcal{A}$  и  $\varepsilon > 0$  при  $n \rightarrow \infty$  выполнены следующие условия:

$$\mathbf{P}_1 \left\{ \frac{|\nu_*/n - p_*|}{\sqrt{p_*}} > \varepsilon \right\} \rightarrow 0, \quad (16)$$

$$\mathbf{P}_1 \left\{ \frac{|\sqrt{\nu_*/n} - \sqrt{p_*}|}{\sqrt{p_*}} > \varepsilon \right\} \rightarrow 0, \quad (17)$$

где  $\nu_*$  и  $p_*$  определены в (15).

**Доказательство.** Для доказательства леммы 1 достаточно показать, что для любых  $a, b \in \mathcal{A}$  и  $\varepsilon > 0$

$$\mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty, \quad (18)$$

и

$$\mathbf{P}_1\{|\nu_{ab}/n - \pi_{ab}^{(*,2)}| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty, \quad (19)$$

где

$$\begin{aligned} \pi_a^{(*,1)} &= \pi_a(1 - \tau) + p_a\tau, \\ \pi_{ab}^{(*,2)} &= (1 - \tau)^2\pi_a\pi_{ab} + \tau(1 - \tau)(\pi_ap_b + \pi_bp_a) + \tau^2p_ap_b. \end{aligned} \quad (20)$$

Докажем справедливость условия (18). Пусть  $a, b \in \mathcal{A}$ ; покажем, что для любого  $\varepsilon > 0$

$$\mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty.$$

В силу неравенства Чебышева

$$\mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon\} \leq \frac{\mathbf{E}_1\{(\nu_a/n - \pi_a^{(*,1)})^2\}}{\varepsilon^2}.$$

Заметим, что в соответствии с формулой полной вероятности верно следующее равенство:

$$\mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon\} = \sum_{i=1}^N \mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon | X_0 = i\} \pi_{a_i}, \quad (21)$$

где  $\{1, \dots, N\} = \mathcal{A}$ . Так как мощность алфавита  $N$  фиксирована и не изменяется с ростом  $n$ , то с учетом (21) из условия

$$\text{для любого } a_i \in \mathcal{A} \quad \mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon | X_0 = a_i\} \rightarrow 0, \quad n \rightarrow \infty,$$

следует, что

$$\mathbf{P}_1\{|\nu_a/n - \pi_a^{(*,1)}| > \varepsilon\} \rightarrow 0, \quad n \rightarrow \infty.$$

Следовательно, для доказательства (18) достаточно показать, что для любого  $b \in \mathcal{A}$  выполнено условие

$$\mathbf{E}_1\{(\nu_a/n - \pi_a^{(*,1)})^2 | X_0 = b\} \rightarrow 0, \quad n \rightarrow \infty.$$

Справедливо равенство

$$\begin{aligned} & \mathbf{E}_1\{(\nu_a/n - \pi_a^{(*,1)})^2 | X_0 = b\} = \\ &= \frac{1}{n^2} \mathbf{E}_1 \left\{ \left[ \sum_{k=0}^{n-1} (\mathbf{I}\{Y_k = a\} - \pi_a^{(*,1)}) \right]^2 | X_0 = b \right\} = \frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} m_{ab}^{(kl)}, \end{aligned}$$

где в случае  $k \neq l$

$$\begin{aligned}
m_{ab}^{(kl)} &= \mathbf{E}_1 \{ \mathbf{I}\{Y_k = a\} \mathbf{I}\{Y_l = a\} | X_0 = b \} - \pi_a^{(*,1)} \mathbf{E}_1 \{ \mathbf{I}\{Y_k = a\} | X_0 = b \} - \\
&\quad - \pi_a^{(*,1)} \mathbf{E}_1 \{ \mathbf{I}\{Y_l = a\} | X_0 = b \} + (\pi_a^{(*,1)})^2 \\
&= \pi_{ba}^{(s)} \pi_{aa}^{(t)} - \pi_a^{(*,1)} \pi_{ba}^{(k)} - \pi_a^{(*,1)} \pi_{ba}^{(l)} + (\pi_a^{(*,1)})^2 \\
&\quad - \tau \left( 2\pi_{ba}^{(s)} \pi_{aa}^{(t)} - (p_a + \pi_a^{(*,1)}) (\pi_{ba}^{(k)} + \pi_{ba}^{(l)}) + 2p_a \pi_a^{(*,1)} \right) \\
&\quad + \tau^2 \left( \pi_{ba}^{(s)} \pi_{aa}^{(t)} - p_a \pi_{ba}^{(k)} - p_a \pi_{ba}^{(l)} + p_a^2 \right), \tag{22}
\end{aligned}$$

при этом  $s = \min\{k, l\}$ ,  $t = |k - l|$ .

Для эргодических цепей Маркова с конечным числом состояний выполняется неравенство (см. [1])

$$|\pi_{ab}^{(n)} - \pi_b| \leq c_1 e^{-n}, \tag{23}$$

где  $a, b$  – состояния цепи Маркова,  $c_1 > 0$  – некоторая абсолютная постоянная,  $\pi_{ab}^{(n)}$  – вероятность перехода из состояния  $a$  в состояние  $b$  за  $n$  шагов,  $n \in \mathbb{N}$ . Из (23) следует, что существует такая абсолютная постоянная  $\rho > 0$ , что

$$|\pi_{ab}^{(n)} - \pi_b| \leq e^{-\rho n}. \tag{24}$$

Из (22) и (24) следует, что при  $k \neq l$

$$|m_{ab}^{(kl)}| \leq (e^{-\rho s} + e^{-\rho t} + e^{-\rho k} + e^{-\rho l}) (1 - \tau)^2, \tag{25}$$

где  $s = \min\{k, l\}$ ,  $t = |k - l|$ , а при  $k = l$

$$\begin{aligned}
|m_{ab}^{(kl)}| &= |m_{ab}^{(kk)}| = \mathbf{E}_1 \left( \mathbf{I}\{Y_k = a | X_0 = b\} - \pi_a^{(*,1)} \right)^2 \\
&\leq \pi_a - \pi_a^2 + \tau(\pi_a - p_a)(2\pi_a - 1) - \tau^2(\pi_a - p_a)^2 + e^{-\rho k}(1 + \tau + \tau^2). \tag{26}
\end{aligned}$$

В соответствии с (22), (25) и (26) справедливы неравенства

$$\begin{aligned}
\frac{1}{n^2} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} |m_{ab}^{(kl)}| &\leq \frac{(1 - \tau)^2}{n^2} \sum_{\substack{k, l=0, \dots, n-1, \\ k \neq l}} (e^{-\rho k} + e^{-\rho l} + e^{-\rho s} + e^{-\rho t}) \\
&+ \frac{\pi_a - \pi_a^2}{n} + \frac{\tau |\pi_a - p_a| \cdot |2\pi_a - 1|}{n} + \frac{\tau^2 (\pi_a - p_a)^2}{n} + \frac{1 + \tau + \tau^2}{n} \sum_{k=0}^{n-1} e^{-\rho k} \\
&= \frac{(1 - \tau)^2}{n} \left( \frac{4e^\rho + 2 - 2e^{\rho - \rho n}}{e^\rho - 1} - \frac{4e^{2\rho}(1 - e^{-\rho n})}{n(e^\rho - 1)^2} \right) \\
&+ \frac{\pi_a - \pi_a^2 + \tau |\pi_a - p_a| \cdot |2\pi_a - 1| + \tau^2 (\pi_a - p_a)^2}{n} \\
&+ \frac{1 + \tau + \tau^2}{n} \cdot \frac{e^\rho - e^{\rho - \rho n}}{e^\rho - 1} \rightarrow 0, \quad n \rightarrow \infty. \tag{27}
\end{aligned}$$

Следовательно, в силу (27)

$$\mathbf{E}_1\{(\nu_a/n - \pi_a^{(*,1)})^2 | X_0 = b\} \rightarrow 0, \quad n \rightarrow \infty.$$

Докажем (19). Рассмотрим состояния  $a, b, c \in \mathcal{A}$ . По аналогии с (21) достаточно показать, что

$$\mathbf{E}_1\left\{(\nu_{ab}/n - \pi_{ab}^{(*,2)})^2 | X_0 = c\right\} \rightarrow 0, \quad n \rightarrow \infty.$$

Обозначим события  $\{X_k = a\}$ ,  $\{Y_k = a\}$  и  $\{Z_k = a\}$  через  $\{X_k^a\}$ ,  $\{Y_k^a\}$  и  $\{Z_k^a\}$ ,  $k = 0, \dots, n-2$ , соответственно. Подсчет показывает, что

$$\begin{aligned} & \mathbf{E}_1\{(\nu_{ab}/n - \pi_{ab}^{(*,2)})^2 | X_0 = c\} \\ &= \frac{1}{n^2} \mathbf{E}_1 \left\{ \left[ \sum_{k=0}^{n-2} (\mathbf{I}\{Y_k = a, Y_{k+1} = b\} - \pi_{ab}^{(*,2)}) \right]^2 | X_0 = c \right\} = \frac{1}{n^2} \sum_{k=0}^{n-2} \sum_{l=0}^{n-2} m_{abc}^{(kl)}, \end{aligned} \quad (28)$$

где в случае  $|k-l| \geq 2$

$$\begin{aligned} m_{abc}^{(kl)} &= \mathbf{E}_1 \{ \mathbf{I}\{Y_k^a, Y_{k+1}^b\} \mathbf{I}\{Y_l^a, Y_{l+1}^b\} | X_0 = c \} \\ &\quad - \pi_{ab}^{(*,2)} \mathbf{E}_1 \{ \mathbf{I}\{Y_k^a, Y_{k+1}^b\} | X_0 = c \} \\ &\quad - \pi_{ab}^{(*,2)} \mathbf{E}_1 \{ \mathbf{I}\{Y_l^a, Y_{l+1}^b\} | X_0 = c \} + (\pi_{ab}^{(*,2)})^2 = \sum_{j=0}^4 A_j^{(kl)} \tau^j, \end{aligned} \quad (29)$$

при этом

$$|A_j^{(kl)}| \leq D (e^{-\rho s} + e^{-\rho t} + e^{-\rho k} + e^{-\rho l}), \quad (30)$$

где  $D$  – некоторая абсолютная постоянная,  $s = \min\{k, l\}$ ,  $t = |k-l|$ .

Из (28)–(30) следует, что

$$\begin{aligned} & \mathbf{E}_1\{(\nu_{ab}/n - \pi_{ab}^{(*,2)})^2 | X_0 = c\} \leq \frac{1}{n^2} \sum_{k=0}^{n-2} \sum_{l=0}^{n-2} |m_{abc}^{(kl)}| \\ & \leq \frac{D}{n^2} \sum_{\substack{k, l=0, \dots, n-2, \\ |k-l| \geq 2}} (e^{-\rho s} + e^{-\rho t} + e^{-\rho k} + e^{-\rho l}) \\ & = 2D \left( \frac{2e^\rho - e^{2\rho - \rho n} + e^{-\rho}}{n(e^\rho - 1)^2} + \frac{(e^{-\rho n} - e^{-\rho})(4e^{3\rho} - e^\rho) + e^{-\rho} - e^\rho}{n^2(e^\rho - 1)^2} \right) \rightarrow 0 \end{aligned} \quad (31)$$

при  $n \rightarrow \infty$ .

Также заметим, что при  $|k-l| = 1$  и при  $k = l$  сумма

$$\frac{1}{n^2} \sum_{k=0}^{n-2} \sum_{l=0}^{n-2} |m_{abc}^{(kl)}| \quad (32)$$

содержит  $3n - 5$  слагаемых. И в силу (30) каждое слагаемое не превосходит  $4D$ . Следовательно, в обоих случаях выражение (32) ограничено сверху величиной

$\frac{20D(3n-5)}{n^2}$ , которая стремится к нулю при  $n \rightarrow \infty$ . Поэтому с учетом последнего выражения имеет место сходимость

$$\mathbf{E}_1 \{ (\nu_{ab}/n - \pi_{ab}^{(*,2)})^2 | X_0 = c \} \rightarrow 0, \quad n \rightarrow \infty.$$

Лемма 1 доказана.

Далее рассмотрим вторую лемму, которую будем использовать для обоснования замены статистики  $S_{abc}$  на  $\widehat{S}_{abc}$ . Введем следующие обозначения:

$$\xi_1 = \frac{|\frac{1}{n}\nu_* - p_*|}{\sqrt{p_*}}, \quad \xi_2 = \frac{|\sqrt{p_*} - \sqrt{\frac{1}{n}\nu_*}|}{\sqrt{p_*}}. \quad (33)$$

**Лемма 2.** Для любых  $x, u, v > 0$  справедливо неравенство

$$\mathbf{P}_1 \{ S_{abc} < x \} \leq \mathbf{P}_1 \{ \widehat{S}_{abc} < x(1+v) + u\sqrt{n} \} + \mathbf{P}_1 \{ \xi_1 > u \} + \mathbf{P}_1 \{ \xi_2 > v \}.$$

**Доказательство.** Для доказательства леммы 2 воспользуемся неравенством

$$|\alpha - \beta + \gamma| \geq |\alpha| - |\beta| - |\gamma|, \quad \alpha, \beta, \gamma \in \mathbb{R}. \quad (34)$$

Заметим, что в соответствии с (14) и (34)

$$\frac{S_{abc}}{\sqrt{n}} \geq \frac{\widehat{S}_{abc}}{\sqrt{n}} - \frac{|\frac{1}{n}\nu_* - p_*|}{\sqrt{p_*}} - \frac{S_{abc}}{\sqrt{n}} \cdot \frac{|\sqrt{p_*} - \sqrt{\frac{1}{n}\nu_*}|}{\sqrt{p_*}}. \quad (35)$$

С использованием формулы полной вероятности и (35) получаем, что для любых  $x, u, v > 0$  верны неравенства

$$\begin{aligned} \mathbf{P} \{ S_{abc} < x \} &\leq \mathbf{P} \left\{ \widehat{S}_{abc} < x(1 + \xi_2) + \xi_1 \sqrt{n} \right\} \leq \\ &\leq \mathbf{P} \left\{ \widehat{S}_{abc} < x(1 + \xi_2) + u\sqrt{n} \right\} + \mathbf{P} \{ \xi_1 > u \} \leq \\ &\leq \mathbf{P} \left\{ \widehat{S}_{abc} < x(1 + v) + u\sqrt{n} \right\} + \mathbf{P} \{ \xi_1 > u \} + \mathbf{P} \{ \xi_2 > v \}. \end{aligned}$$

Лемма 2 доказана.

Рассмотрим состояния  $a, b, c$  и  $d \in \mathcal{A}$  (которые могут совпадать) и найдем выражения для математических ожиданий и дисперсий статистик  $\widehat{S}_{abc}$ . Заметим, что верно равенство

$$\mathbf{E}_1 \nu_{abc} = \sum_{k=0}^{n-3} \mathbf{P}_1 \{ Y_k = a, Y_{k+1} = b, Y_{k+2} = c \}. \quad (36)$$

В силу формулы полной вероятности

$$\mathbf{P}_1 \{ Y_k = a, Y_{k+1} = b, Y_{k+2} = c \} = \sum_{j=0}^3 B_j^{(k)} \tau^j (1 - \tau)^{3-j}, \quad (37)$$

где

$$B_0^{(k)} = \pi_{da}^{(k)} \pi_{ab} \pi_{bc}, \quad B_1^{(k)} = \pi_{da}^{(k)} \pi_{ab} p_c + \pi_{da}^{(k)} \pi_{ac}^{(2)} p_b + \pi_{db}^{(k+1)} \pi_{bc} p_a, \\ B_2^{(k)} = \pi_{da}^{(k)} p_b p_c + \pi_{db}^{(k+1)} p_a p_c + \pi_{dc}^{(k+2)} p_a p_b, \quad B_3^{(k)} = p_a p_b p_c.$$

Учитывая (36) и (37), получаем:

$$\mathbf{E}_1 \widehat{S}_{abc} = \frac{1}{\sqrt{p_*}} \mathbf{E}_1 \left| \frac{1}{\sqrt{n}} \nu_{abc} - \sqrt{n} p_* \right| \geq \frac{\frac{1}{\sqrt{n}} \mathbf{E}_1 \nu_{abc} - \sqrt{n} p_*}{\sqrt{p_*}} \geq \\ \geq \sqrt{n} \tau (1 - \tau)^2 \left( C_1 (1 - \tau) + C_2 \tau + \frac{C(1 - e^{-\varepsilon n})}{n} \right) \geq \\ \geq \sqrt{n} \tau (1 - \tau)^2 (C_1 (1 - \tau) + C_2 \tau), \quad (38)$$

где  $C_1 = p_b \pi_b (\pi_{a*c} - \pi_a \pi_{bc}) + \pi_a \pi_b p_b (\pi_{bc} - \pi_c)$ ,  $C_2 = p_b (\pi_{a*c} - \pi_a \pi_c)$ ,  $\pi_{a*c} = \mathbf{P}_1 \{Y_0 = a, Y_2 = c\}$ ,  $C, \varepsilon > 0$  – некоторые постоянные.

Из (38) следует, что если выполнено условие

$$\sqrt{n} \tau (1 - \tau)^2 (C_1 (1 - \tau) + C_2 \tau) \rightarrow \infty, \quad n \rightarrow \infty, \quad (39)$$

то

$$\mathbf{E}_1 \widehat{S}_{abc} \rightarrow \infty, \quad n \rightarrow \infty.$$

Рассмотрим дисперсию статистики  $\widehat{S}_{abc}$ . Для  $k, l = 0, \dots, n - 3$  введем события:

$$\{Y_k = a, Y_{k+1} = b, Y_{k+2} = c\} = Y_{abc}^{(k)},$$

$$\{Y_k = a, Y_{k+1} = b, Y_{k+2} = c, Y_l = a, Y_{l+1} = b, Y_{l+2} = c\} = Y_{abc}^{(kl)}.$$

Тогда

$$\mathbf{D}_1 \nu_{abc} = \frac{1}{(n-2)^2} \mathbf{E}_1 \left( \sum_{k=0}^{n-3} \left( \mathbf{I} \{Y_{abc}^{(k)} | X_0 = d\} - \mathbf{P}_1 \{Y_{abc}^{(k)} | X_0 = d\} \right) \right)^2 = \\ = \frac{1}{(n-2)^2} \sum_{k=0}^{n-3} \sum_{l=0}^{n-3} d_{abc}^{(kl)}, \quad (40)$$

где

$$d_{abc}^{(kl)} = \mathbf{E}_1 (\mathbf{I} \{Y_{abc}^{(k)} | X_0 = d\} - \mathbf{P}_1 \{Y_{abc}^{(k)} | X_0 = d\}) \times \\ \times (\mathbf{I} \{Y_{abc}^{(l)} | X_0 = d\} - \mathbf{P}_1 \{Y_{abc}^{(l)} | X_0 = d\}) = \\ = \mathbf{E}_1 \mathbf{I} \{Y_{abc}^{(k)} | X_0 = d\} \mathbf{I} \{Y_{abc}^{(l)} | X_0 = d\} - \mathbf{P}_1 \{Y_{abc}^{(k)} | X_0 = d\} \mathbf{P}_1 \{Y_{abc}^{(l)} | X_0 = d\} = \\ = \mathbf{P}_1 \{Y_{abc}^{(kl)} | X_0 = d\} - \mathbf{P}_1 \{Y_{abc}^{(k)} | X_0 = d\} \mathbf{P}_1 \{Y_{abc}^{(l)} | X_0 = d\}.$$

Заметим, что величина  $d_{abc}^{(kl)}$ ,  $k, l = 0, \dots, n - 3$ , представляется в виде

$$d_{abc}^{(kl)} = \sum_{j=0}^6 \left( \eta_j^{(kl)} - \sum_{\substack{0 \leq j_1, j_2 \leq 3: \\ j_1 + j_2 = j}} \eta_{j_1}^{(k)} \eta_{j_2}^{(l)} \right) \tau^j (1 - \tau)^{6-j}, \quad (41)$$

где

$$\eta_j^{(kl)} = \mathbf{P}_1 \{ Y_{abc}^{(kl)} \mid \text{в позициях } k, k+1, k+2, l, l+1, l+2 \text{ внесено } j \text{ вкраплений} \},$$

$$\eta_j^{(k)} = \mathbf{P}_1 \{ Y_{abc}^{(k)} \mid \text{в позициях } k, k+1, k+2 \text{ внесено } j \text{ вкраплений} \}.$$

Заметим, что для любых таких  $k, l = 0, \dots, n-3$ , что  $|k-l| \geq 3$ , для любого  $j = 0, \dots, 4$  при  $n \rightarrow \infty$

$$\left| \eta_j^{(kl)} - \sum_{\substack{0 \leq j_1, j_2 \leq 3: \\ j_1 + j_2 = j}} \eta_{j_1}^{(k)} \eta_{j_2}^{(l)} \right| \leq e^{-\rho s} + e^{-\rho t} + e^{-\rho k} + e^{-\rho l}, \quad (42)$$

где  $s = \min\{k, l\}$ ,  $t = |k-l|$ , а при  $j = 5, 6$

$$\eta_j^{(kl)} - \sum_{\substack{0 \leq j_1, j_2 \leq 3: \\ j_1 + j_2 = j}} \eta_{j_1}^{(k)} \eta_{j_2}^{(l)} = 0. \quad (43)$$

Из (40)–(43) следует, что для некоторых постоянных  $\rho, C$  и  $C_1$  справедливы неравенства

$$\begin{aligned} \mathbf{D}_1 \widehat{S}_{abc} &= \frac{1}{n P_*} \sum_{k=0}^{n-3} \sum_{l=0}^{n-3} |d_{abc}^{(kl)}| \leq \frac{C}{n} \sum_{k=0}^{n-3} \sum_{l=0}^{n-3} (e^{-\rho s} + e^{-\rho t} + e^{-\rho k} + e^{-\rho l}) = \\ &= C \cdot \frac{2 + 5e^\rho - e^{3\rho - \rho n} + 5e^{4\rho - \rho n} - 11e^{2\rho}}{(n-3)(e^\rho - 1)^2} - \\ &- C \cdot \frac{1 + 4e^\rho - 2e^{3\rho - \rho n} + 2e^{4\rho - \rho n} - 5e^{2\rho}}{(e^\rho - 1)^2} < C_1. \end{aligned} \quad (44)$$

Из (13) и леммы 2 следует, что

$$\begin{aligned} &\mathbf{P}_1 \{ S < \varkappa_{1-\alpha} \} \leq \mathbf{P}_1 \{ S_{abc} < \sqrt{\varkappa_{1-\alpha}} \} \leq \\ &\leq \mathbf{P}_1 \left\{ \widehat{S}_{abc} < \sqrt{\varkappa_{1-\alpha}}(1+v) + u\sqrt{n} \right\} + \mathbf{P}_1 \{ \xi_1 > u \} + \mathbf{P}_1 \{ \xi_2 > v \}. \end{aligned} \quad (45)$$

В силу леммы 1 для любых  $u, v > 0$  выполнены условия

$$\mathbf{P}_1 \{ \xi_1 > u \} \rightarrow 0, \quad \mathbf{P}_1 \{ \xi_2 > v \} \rightarrow 0, \quad n \rightarrow \infty.$$

Из неравенства Чебышева и (44) следует, что при выполнении условия (39) для любых  $u, v > 0$  верна оценка

$$\begin{aligned} &\mathbf{P}_1 \left\{ \widehat{S}_{abc} < \sqrt{\varkappa_{1-\alpha}}(1+v) + u\sqrt{n} \right\} \leq \\ &\leq \frac{\mathbf{D}_1 \widehat{S}_{abc}}{\left( \sqrt{\varkappa_{1-\alpha}}(1+v) + u\sqrt{n} - \mathbf{E}_1 \widehat{S}_{abc} \right)^2} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Теорема 2 доказана.

## Список литературы

1. Боровков А. А., *Теория вероятностей*, М.: Наука, 1986, 432 с.; англ. пер.: Bogovkov A. A., *Probability Theory*, New York: Gordon & Breach, 1998, 474 pp.
2. Иванов В. А., “Модели вкраплений в однородные случайные последовательности”, *Труды по дискретной математике*, **11** (2008), 18–34.
3. Ивченко Г.И., Медведев Ю.И., *Введение в математическую статистику*, ЛКИ, Москва, 2010, 600 с.
4. Крамер Г., *Математические методы статистики*, Мир, Москва, 1975, 630 с.; пер. с англ.: Cramer H., *Mathematical Methods of Statistics*, Prinseton University Press, 1962, 590 pp.
5. Пономарев К. И., “Параметрическая модель вкрапления и ее статистический анализ”, *Дискретная математика*, **21**:4 (2009), 148–157; англ. пер.: Ponomarev K. I., “A parametric model of embedding and its statistical analysis”, *Discrete Math. Appl.*, **19**:6 (2009), 587–596.
6. Харин Ю. С., Вечерко Ю. В., “Статистическое оценивание параметров модели вкраплений в двоичную цепь Маркова”, *Дискретная математика*, **25**:2 (2013), 135–148; англ. пер.: Kharin Yu. S., Vecherko E. V., “Statistical estimation of parameters for binary Markov chain models with embeddings”, *Discrete Math. Appl.*, **23**:2 (2013), 153–169.
7. Харин Ю. С., Вечерко Ю. В., “Распознавание вкраплений в двоичную цепь Маркова”, *Дискретная математика*, **27**:3 (2015), 123–144; англ. пер.: Kharin Yu. S., Vecherko E. V., “Detection of embeddings in binary Markov chains”, *Discrete Math. Appl.*, **26**:1 (2016), 13–29.
8. Шойтов А. М., “О выявлении факта зашумления конечной цепи Маркова с неизвестной матрицей переходных вероятностей”, *Прикладная дискретная математика. Приложение*, **3** (2010), 44–45.
9. Filler T., Ker A.D., Fridrich J., “The square root law of steganographic capacity for Markov covers”, *Proc. SPIE*, **7254** (2009), 31–47.
10. Ker A.D., “A capacity result for batch steganography”, *IEEE Signal Processing Letters*, **14**(8) (2007), 525–528.
11. Sharp T., “An implementation of key-based digital signal steganography”, *Proc. Information Hiding Workshop*, **2137** (2001), 13–26.

Статья поступила 05.09.2016.

Переработанный вариант поступил 20.05.2017.