УДК 519.237.3 + 519.669 + 681.51

В.И. Масол, С.В. Поперешняк

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЛОКАЛЬНЫХ УЧАСТКОВ БИТОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Ключевые слова: *s*-цепочки, битовая последовательность, случайность, локальные участки, совместное распределение.

Введение

Всесторонний статистический анализ битовой последовательности, получаемой на выходе генераторов псевдослучайных чисел, не исключает проверки гипотезы случайного расположения нулей и единиц на некоторых интервалах небольшой длины, например до ста битов (вариант классификации битовых последовательностей согласно их длине приведен в [1]). Необходимость в проверке этой гипотезы возникает также в иных областях, например в химической промышленности при производстве синтетических волокон в процессе анализа причин обрыва нити в ходе ее формирования. Здесь (0, 1)-последовательность получается по результатам взвешивания отрезков нити на небольшом участке до ее обрыва с последующим сравнением с номинальным весом этих отрезков.

В тестах для проверки гипотезы случайности расположения нулей и единиц в (0, 1)-последовательностях предполагается, что длина (0, 1)-последовательности велика (см., например, [2]). Это предположение связано с использованием асимптотических приближений для точного распределения статистики (как правило, одномерной) соответствующего теста.

В данной работе предлагаются точные совместные распределения некоторых статистик (0, 1)-последовательности длины n, $1 < n < \infty$. Для $n = 20$ (т.е. для битовой последовательности малой длины [1]) приведены таблицы, содержащие числовые значения соответствующего распределения. Эти таблицы, равно как и предлагаемые их графические представления, могут быть использованы для проверки гипотезы случайности расположения нулей и единиц.

1. Постановка задачи

Рассмотрим последовательность случайных величин

$$\gamma_1 \gamma_2 \dots \gamma_n, \tag{1}$$

где $\gamma_i = \{0, 1\}$, $i = 1, 2, \dots, n$, $n > 0$.

Подпоследовательность $\gamma_j \gamma_{j+1} \dots \gamma_{j+s-1}$ последовательности (1) будем называть s-цепочкой, $j = 1, 2, \dots, n - s + 1$, $s = 1, 2, \dots, n$. Обозначим $\eta(t_1 t_2 \dots t_s)$ число s-цепочек в последовательности (1), которые совпадают с $t_1 t_2 \dots t_s$, где $t_i = \{0, 1\}$, $i = 1, 2, \dots, s$. Сформулируем условие.

Условие. Последовательность (1) состоит из n, $n > 0$, независимых одинаково распределенных случайных величин; вероятности событий $\{\gamma_i = 1\}$, $\{\gamma_i = 0\}$ известны и равны $\mathrm{P}\{\gamma_i = 1\} = p$, $\mathrm{P}\{\gamma_i = 0\} = q$, $p + q = 1$, $i = 1, 2, \dots, n$.

В данной работе рассмотрены совместные распределения случайных величин $\eta(t_1 t_2 \dots t_s)$, $\eta(t'_1 t'_2 \dots t'_{s'})$, где $t_j, t'_i \in \{0, 1\}$, $j = 1, 2, \dots, s$, $i = 1, 2, \dots, s'$, для некоторых значений s и s'. Перейдем к точным формулировкам.

 ISSN 0572-2691

2. Совместные распределения числа 2-цепочек и числа 3-цепочек фиксированного вида

Теорема 1. Пусть выполняется сформулированное выше условие, n, k_1, k_2, k_3, t, t_1 — целые числа, такие, что $k_1 \ge 0$, $k_2 \ge 0$, $k_3 \ge 0$, $n \ge \max(2k_1, 3)$, $t, t_1 \in \{0, 1\}$. Тогда

$$\mathrm{P}\{\eta(t\,t^*) = k_1,\, \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} \sum \prod_{i=0}^{1} C_{k_1}^{\delta_1} C_{m_i-k_1}^{k_1-\delta_i}, \qquad (2)$$

где $m_0 = n - m_1$, символ $\sum$ обозначает суммирование по всем целым неотрицательным числам δ_0 и δ_1, таким, что $\delta_0 + \delta_1 = 2k_1 - k_2$, $t^* = 1 - t$;

$$\mathrm{P}\{\eta(t_1\,t_1^*) = k_1,\, \eta(t_1\,t\,t_1^*) = k_2\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} C_{k_1}^{k_2} C_{m_t-k_1}^{k_2} C_{m_{t^*}}^{k_1}; \qquad (3)$$

$$\mathrm{P}\{\eta(t\,t^*) = k_1, \eta(t1t^*) = k_2,\, \eta(t\,0\,t^*) = k_3\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} C_{k_1}^{k_2} C_{k_1}^{k_3} C_{m_1-k_1}^{k_2} C_{m_0-k_1}^{k_3}. \qquad (4)$$

Пример. Для (0, 1)-последовательности вида

0 1 1 0 0 1 1 0 1 1 1 0 0 0 0 1 0 1 0 1

значениями случайных величин $\eta(t\,t^*)$, $\eta(t1t^*)$, $\eta(t\,0\,t^*)$ и $\eta(t\,\alpha\,t^*)$, где $\alpha \in \{0, 1\}$, для $t = 1$ являются соответственно числа 5, 3, 2 и 5.

Доказательство. Проверим соотношение (3). Обозначим ν количество единиц в случайной последовательности (1). Случайная величина ν имеет биномиальное распределение с параметрами (n, p), что позволяет записать для $m = 0, 1, 2, \ldots, n$ вероятность события $\{\nu = m\}$, а именно:

$$\mathrm{P}\{\nu = m\} = C_n^m p^m q^{n-m}. \qquad (5)$$

С помощью формулы полной вероятности находим

$$\mathrm{P}\{A_1, A_2\} = \sum_{m_1=0}^{n} \mathrm{P}\{\nu = m_1\} \cdot \mathrm{P}\{A_1, A_2 / \nu = m_1\}, \qquad (6)$$

где $A_1 \overset{\text{def}}{=} \{\eta(t_1\,t_1^*) = k_1\}$, $A_2 \overset{\text{def}}{=} \{\eta(t_1\,t\,t_1^*) = k_2\}$.

Покажем, что

$$\mathrm{P}\{A_1,\, A_2 / \nu = m_1\} = (C_n^{m_1})^{-1} C_{k_1}^{k_2} C_{m_t-k_1}^{k_2} C_{m_{t^*}}^{k_1}. \qquad (7)$$

Введем следующие обозначения: $\Omega(n, m_1)$ — множество всех n-мерных (0, 1)-векторов, каждый из которых содержит m_1 единиц и m_0 нулей, $m_0 + m_1 = n$; $D(m_0, m_1, k_1, k_2; t_1)$ — подмножество множества $\Omega(n, m_1)$, которое содержит все попарно различные векторы, начинающиеся с элемента t_1, заканчивающиеся элементом t_1^*, содержащие $k_1(k_2)$ цепочек вида $t_1 t_1^*\,(t_1\,t\,t_1^*)$; Q — общее число векторов $\vec{\upsilon}$, $\vec{\upsilon} \in \Omega(n, m_1)$, каждый из которых имеет $k_1(k_2)$ цепочек вида $t_1 t_1^*\,(t_1\,t\,t_1^*)$.

Используя принятые обозначения, получаем

$$\mathrm{P}\{A_1,\ A_2 / \nu = m_1\} = Q(|\Omega(n, m_1)|)^{-1}, \tag{8}$$

$$Q = \sum_{\nu_t=0}^{m_t} \sum_{\nu_{t^*}=0}^{m_{t^*}} \left| D(m_t - \nu_t, m_{t^*} - \nu_{t^*}, k_1, k_2; t_1) \right|. \tag{9}$$

Далее убедимся в том, что

$$\left| D(m_0, m_1, k_1, k_2; t_1) \right| = C_{k_1}^{k_2} C_{m_t - k_1 - 1}^{k_2 - 1} C_{m_{t^*} - 1}^{k_1 - 1}. \tag{10}$$

Действительно, для произвольного вектора $\vec{\upsilon}$, $\vec{\upsilon} \in D(m_0, m_1, k_1, k_2; t_1)$, во-первых, событие $\{A_1, A_2\}$ имеет место тогда и только тогда, когда

$$k_2 = k_1 - \delta_1^{(t)} \geq 0, \tag{11}$$

где $\delta_1^{(t)}$ — число t-серий длины единица каждая в векторе $\vec{\upsilon}$, $t \in \{0, 1\}$; во-вторых, перестановка между собой t-серий не изменяет чисел k_1 и k_2. Поэтому вектор $\vec{\upsilon}$ определяется однозначно, если

• зафиксировано одно из $C_{k_1}^{\delta_1^{(t)}}$ возможных размещений t-серий длины единица каждая;

• зафиксировано одно из возможных разбиений $m_t - \delta_1^{(t)}$ t-элементов на $k_1 - \delta_1^{(t)}$ t-серий, длина каждой из которых равняется двум или больше двух;

• зафиксировано одно из возможных разбиений m_{t^*} t^*-элементов на k_1 t^*-серий, длина каждой из которых не меньше единицы.

Между указанными разбиениями множества t-элементов (t^*-элементов) и решениями уравнения

$$m_t - \delta_1^{(t)} = x_1^{(t)} + x_2^{(t)} + \ldots + x_{k_1 - \delta_1^{(t)}}^{(t)} \tag{12}$$

$$(m_{t^*} - \delta_1^{(t^*)} = x_1^{(t^*)} + x_2^{(t^*)} + \ldots + x_{k_1}^{(t^*)}) \tag{13}$$

в целых числах, каждое из которых не меньше двух (единицы), существует взаимно-однозначное соответствие. В свою очередь, число решений уравнения (12) или (13), таких, что $x_j^{(t)} \geq 2$, $j = 1, 2, \ldots, k_1 - \delta_1^{(t)}$ ($x_j^{(t^*)} \geq 1$, $j = 1, 2, \ldots, k_1$), равняется $C_{m_t - k_1 - 1}^{k_1 - \delta_1^{(t)} - 1}$ ($C_{m_{t^*} - 1}^{k_1 - 1}$).

Таким образом, с учетом (11) произведение $C_{k_1}^{k_2} C_{m_t - k_1 - 1}^{k_2 - 1} C_{m_{t^*} - 1}^{k_1 - 1}$ дает общее число элементов множества $D(m_0, m_1, k_1, k_2; t_1)$, что доказывает (10). Принимая во внимание (9), (10) и равенство

$$C_{n-1}^{\eta - 1} + \ldots + C_{\eta - 1}^{\eta - 1} = C_n^{\eta}, \tag{14}$$

получаем соотношение $Q = C_{k_1}^{k_2} C_{m_t - k_1}^{k_2} C_{m_t^*}^{k_1}$, которое совместно с (8) и равенством

$$\left|\Omega(n,\ m_1)\right| = C_n^{m_1} \tag{15}$$

приводит к (7). В свою очередь, из (5)–(7) непосредственно следует (3).

Соотношения (2) и (4) можно проверить, используя схему доказательства равенства (3).

Частный случай соотношения (2) установлен в [3].

3. Примеры к теореме 1

3.1. Иллюстрация использования равенства (2). В табл. 1 и на рис. 1 приведено использование соотношения (2) для $p = q = 1/2,$ малой выборки длины $n,$ $n = 20,$ и некоторых значений $k_1,\ k_2$.

В первом столбце табл. 1 помещены все возможные варианты значений k_1 и k_2, для которых вероятность $\mathrm{P}\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\} \ge 0,01.$ Во втором столбце даны вероятности (в неубывающем порядке) $\mathrm{P}\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\}$ для пар чисел $(k_1, k_2),$ указанных в первом столбце.

В каждой строке третьего столбца дана сумма накопленных вероятностей до реализации события $\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\}$ включительно, где k_1 и k_2 указаны в этой же строке в первом столбце.

Таблица 1

(k_1, k_2)	P	P_c		(k_1, k_2)	P	P_c
(7, 3)	0,012149811	0,100008011		(5, 7)	0,018882751	0,132255554
(7, 5)	0,013364792	0,113372803		(7, 4)	0,020047188	0,152302742
(5, 7)	0,018882751	0,132255554		(3, 3)	0,026035309	0,178338051
(7, 4)	0,020047188	0,152302742		(6, 3)	0,026435852	0,204773903
(3, 3)	0,026035309	0,178338051		(3, 5)	0,02863884	0,233412743
(6, 3)	0,026435852	0,204773903		(6, 6)	0,031723022	0,265135765
(3, 5)	0,02863884	0,233412743		(5, 3)	0,037765503	0,302901268
(6, 6)	0,031723022	0,265135765		(4, 3)	0,03818512	0,341086388
(5, 3)	0,037765503	0,302901268		(3, 4)	0,04295826	0,384044647
(4, 3)	0,03818512	0,341086388				

Например, для $k_1 = 7$ и $k_2 = 3$ имеем $\mathrm{P}\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\} = 0,012149811,$ $\mathrm{P}_c = \sum \mathrm{P}\{\eta(t\,t^*) = k_1, \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\},$ где знак суммирования $\sum$ распространяется на все пары $(k_1, k_2),$ для которых $\mathrm{P}\{\eta(t\,t^*) = k_1,$ $\eta(t1t^*) + \eta(t\,0\,t^*) = k_2\} \le 0,012149811.$

На рис. 1 представлена пузырьковая диаграмма [4], в которой первый параметр (горизонтальная ось) — значение $k_1,$ второй (вертикальная ось) —

значение k_2, третий (размер пузырька) — вероятность осуществления события $\{\eta(t\,t^*)=k_1,\ \ \eta(t1t^*)+\eta(t\,0\,t^*)=k_2\}$, выраженная в процентах.

Например, на рис. 1 при $k_1=5$ и $k_2=5$ вероятность осуществления события $\{\eta(t\,t^*)=k_1,\eta(t1t^*)+\eta(t\,0\,t^*)=k_2\}$ в процентах равняется 11,10 %.

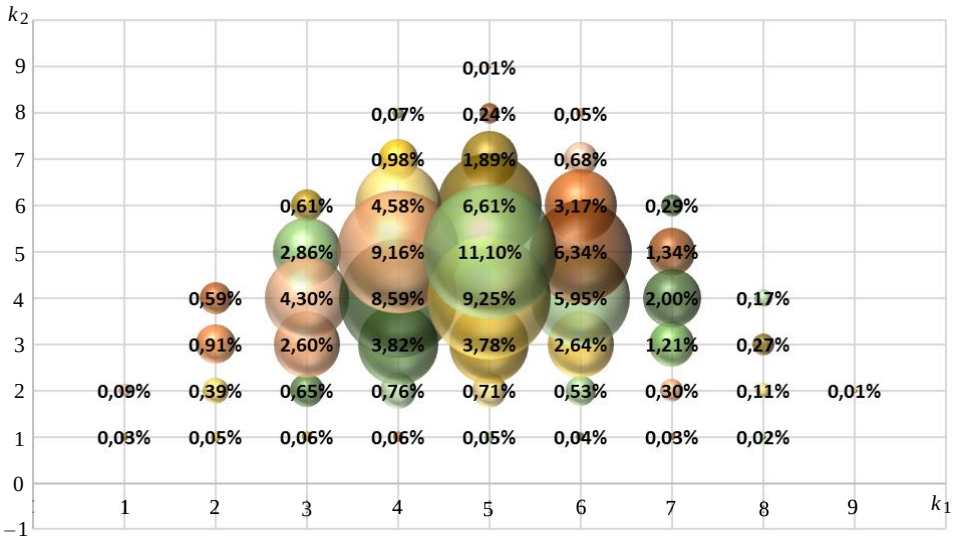

Рис. 1

3.2. Иллюстрация использования равенства (3). В табл. 2 и на рис. 2 приведено использование соотношения (3) для $p=q=1/2,$ малой выборки длины $n,\ n=20,$ и некоторых значений k_1 и k_2.

Таблица 2

(k_1, k_2)	P	P_c		(k_1, k_2)	P	P_c
(2, 2)	0,01108932	0,0622549		(4, 1)	0,04721069	0,3276072
(7, 3)	0,01214981	0,0744047		(3, 2)	0,05311203	0,3807192
(7, 1)	0,01336479	0,0877695		(5, 1)	0,05455017	0,4352694
(6, 4)	0,01952648	0,107296		(6, 3)	0,05727768	0,492547
(7, 2)	0,02004719	0,1273432		(6, 2)	0,0715971	0,5641441
(4, 4)	0,02318382	0,150527		(4, 3)	0,09273529	0,6568794
(3, 1)	0,02451324	0,1750402		(5, 3)	0,10910034	0,7659798
(3, 3)	0,03034973	0,20539		(4, 2)	0,11128235	0,8772621
(6, 1)	0,03682137	0,2422113		(5, 2)	0,12273788	1
(5, 4)	0,03818512	0,2803965				

Интерпретация табл. 2 аналогична интерпретации табл. 1.

На рис. 2 дана пузырьковая диаграмма, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2, третий (размер пузырька) — вероятность осуществления события $\mathrm{P}\{\eta(t_1\,t_1^*)=$ $=k_1,\ \ \eta(t_1\,t\,t_1^*)=k_2\}$, которая представлена в процентах.

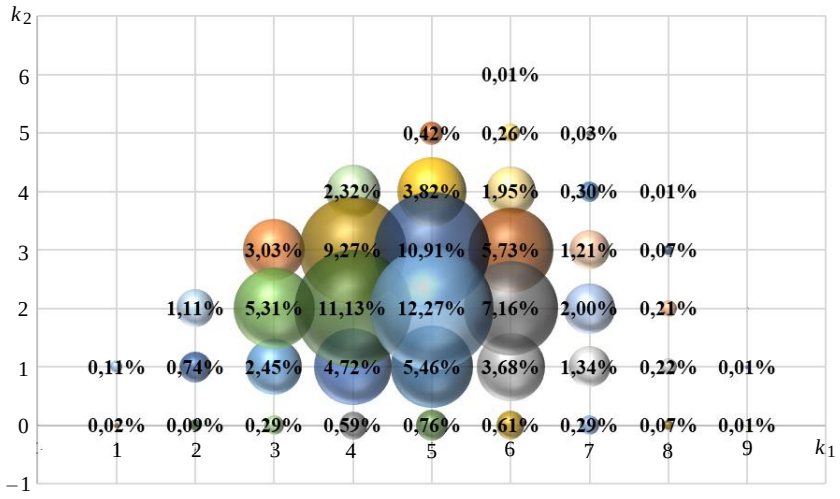

Рис. 2

3.3. Иллюстрация использования равенства (4). В табл. 3 приведено использование соотношения (4) для $p = q = 1/2$, малой выборки длины n, $n = 20$, и некоторых значений k_1, k_2 и k_3.

Таблица 3

(k_1, k_2, k_3)	P	P_c		(k_1, k_2, k_3)	P	P_c
(6, 1, 2)	0,010815	0,280055		(4, 1, 3)	0,019638	0,522853
(6, 2, 1)	0,010815	0,29087		(4, 3, 1)	0,019638	0,542491
(5, 1, 4)	0,011015	0,301885		(5, 1, 3)	0,02203	0,564521
(5, 4, 1)	0,011015	0,3129		(5, 3, 1)	0,02203	0,586551
(3, 1, 2)	0,011716	0,324615		(6, 2, 3)	0,024033	0,610583
(3, 2, 1)	0,011716	0,336331		(6, 3, 2)	0,024033	0,634616
(6, 3, 3)	0,013733	0,350064		(3, 2, 2)	0,025775	0,660391
(3, 2, 3)	0,014319	0,364384		(4, 3, 3)	0,026184	0,686575
(3, 3, 2)	0,014319	0,378703		(6, 2, 2)	0,027037	0,713612
(6, 1, 3)	0,01442	0,393123		(5, 3, 3)	0,031471	0,745083
(6, 3, 1)	0,01442	0,407542		(4, 2, 3)	0,039276	0,784359
(5, 1, 2)	0,015736	0,423278		(4, 3, 2)	0,039276	0,823635
(5, 2, 1)	0,015736	0,439013		(5, 2, 2)	0,04406	0,867695
(5, 2, 4)	0,015736	0,454749		(5, 2, 3)	0,04406	0,911755
(5, 4, 2)	0,015736	0,470485		(5, 3, 2)	0,04406	0,955814
(4, 1, 2)	0,016365	0,48685		(4, 2, 2)	0,044186	1
(4, 2, 1)	0,016365	0,503215				

Пояснения к табл. 3 аналогичны пояснениям к табл. 1.

4. Совместные распределения числа 2-цепочек и числа 3-цепочек вида $t_1 t_1^*$ и $t\alpha t^*$, $\alpha \in \{0, 1\}$

Теорема 2. Пусть выполняется сформулированное выше условие, n, k_1, k_2, k_3, t, t_1 — целые числа, такие, что $k_1 \geq 0$, $k_2 \geq 0$, $k_3 \geq 0$, $n \geq \max\{2k_1, 3\}$, t, $t_1 \in \{0, 1\}$. Тогда

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(t1t) + \eta(t0t) = k_2\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} \times$$

$$\times\left\{\sum_{i\in\{k_1,k_1+1\}} C_{i-1}^{\delta_{t^*}} C_i^{\delta_t - m_t + 2i} Z(m_t - i, m_t - i - \delta_t) C_{m_{t^*}-i+1}^{k_1-\delta_{t^*}} \chi(m_t \ge 2) + Z_t\right\}, \quad (16)$$

где $\sum$ — суммирование по всем целым неотрицательным числам δ_t и δ_{t^*}, таким, что $\delta_t + \delta_{t^*} = k_2$, $Z_t = \chi(m_t = 0)\bar{\chi}(k_1 + k_2 \ge 1) + \chi(m_t = 1)\bar{\chi}(k_2 \ge 1)\bar{\chi}(k_2 = 0, k_1 \ge 2) \times$ $\times(\chi(k_1 = k_2 = 0) + (n-1)\chi(k_1 = 1, k_2 = 0))$, $\chi(E)$ — индикатор события E, $\bar{\chi}(E) =$ $= 1 - \chi(E)$, $Z(a,b) \overset{\text{def}}{=} \begin{cases} C_{a-1}^{b-1}, & \text{если } a \ge b \ge 1; \\ 1, & \text{если } a = b = 0; \\ 0 & \text{в остальных случаях;} \end{cases}$

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(ttt) = k_2\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} \times$$

$$\times\left\{C_{m_{t^*}}^{k_1} \sum_{i\in\{k_1,k_1+1\}} C_i^{m_t - k_2 - i} Z(m_t - i, m_t - i - k_2)\chi(m_t \ge 2) + Z_t\right\}; \quad (17)$$

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(tt*t) = k_2\} =$$

$$= \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} \left\{\sum_{i\in\{k_1,k_1-1\}} C_i^{k_2} C_{m_t-1}^{i} C_{m_{t^*}-i}^{k_1-k_2} \chi(m_t \ge 2) + Z_t\right\}; \quad (18)$$

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(ttt) = k_2,\, \eta(tt*t) = k_3\} = \sum_{m_1=k_1}^{n-k_1} p^{m_1} q^{m_0} \times$$

$$\times\left\{\sum_{i\in\{k_1,k_1+1\}} C_i^{k_2 - m_t + 2i} C_{i-1}^{k_3} Z(m_t - i, m_t - i - k_2) C_{m_{t^*}-i+1}^{k_1-k_3} \chi(m_t \ge 2) + \tilde{Z}_t\right\}, \quad (19)$$

где $\tilde{Z}_t = \chi(m_t = 0)\bar{\chi}(k_1 + k_2 + k_3 \ge 1) + \chi(m_t = 1)\bar{\chi}(k_2 + k_3 \ge 1)\bar{\chi}(k_2 = k_3 = 0, k_1 \ge 2) \times$ $\times(\chi(k_1 = k_2 = k_3 = 0) + (n-1)\chi(k_1 = 1, k_2 = k_3 = 0))$.

Доказательство. Проверим соотношение (17). Аналогично (6) находим

$$\mathrm{P}\{B_1, B_2\} = \sum_{m_1=0}^{n} \mathrm{P}\{\nu = m_1\} \times \mathrm{P}\{B_1, B_2 / \nu = m_1\}, \quad (20)$$

где $B_1 \overset{\text{def}}{=} \{\eta(t_1 t_1^*) = k_1\}$, $B_2 \overset{\text{def}}{=} \{\eta(ttt) = k_2\}$.

Пусть Q — число всех векторов $\vec{\upsilon}$, $\vec{\upsilon} \in \Omega(n, m_1)$, каждый из которых имеет $k_1(k_2)$ 2-цепочек вида $t_1 t_1^*$ (3-цепочек вида ttt). Тогда с учетом (15) имеем

$$\mathrm{P}\{B_1, B_2 / \nu = m_1\} = (C_n^{m_1})^{-1} Q. \quad (21)$$

Далее установим формулу для нахождения числа Q. Рассмотрим подмножество $\left|D(m_0,\, m_1, k_1, k_2;\, t)\right| \subseteq \Omega(n, m_1)$ всех попарно отличных векторов, каж-

дый из которых начинается и заканчивается элементом t, содержит $k_1(k_2)$ 2-цепочек вида $t_1 t_1^*$ (3-цепочек вида ttt). Отметим, что произвольный вектор $\breve{\upsilon}$, $\breve{\upsilon} \in D(m_0, m_1, k_1, k_2; t)$, обладает таким свойством: перестановка в векторе $\breve{\upsilon}$ между собой α-серий, $\alpha \in \{0,1\}$, не изменяет чисел k_1 и k_2. Это позволяет записать равенство

$$Q = \sum_{\nu_{t^*}=0}^{m_{t^*}} \left| D(m_{t^*} - \nu_{t^*}, m_t, k_1, k_2; t) \right| +$$

$$+ \sum_{\nu_{t^*}=0}^{m_{t^*}} \sum_{\nu'_{t^*}=1}^{m_{t^*}} \left| D(m_{t^*} - \nu_{t^*} - \nu'_{t^*}, m_t, k_1 - 1, k_2; t) \right|. \qquad (22)$$

Покажем, что

$$\left| D(m_0, m_1, k_1, k_2; t) \right| = C_{k_1+1}^{m_t - k_2 - k_1 - 1} C_{m_{t^*}-1}^{k_1 - 1} Z(m_t - k_1 - 1, m_t - k_1 - k_2 - 1). \qquad (23)$$

Действительно, для произвольного вектора $\breve{\upsilon}$, $\breve{\upsilon} \in D(m_0, m_1, k_1, k_2; t)$, имеем

$$k_2 = \delta_1^{(t)} + m_t - 2k_1 - 2, \qquad (24)$$

где $\delta_1^{(t)}$ — число t-серий длины единица в векторе $\breve{\upsilon}$. Для $m_t \geq 2$ и $m_{t^*} \geq 1$ элемент множества $D(m_0, m_1, k_1, k_2; t)$ определяется однозначно, если

- зафиксировать одно из $C_{k_1+1}^{\delta_1^{(t)}}$ возможных размещений t-серий длины единица каждая;
- зафиксировать одно из возможных разбиений числа m_{t^*} на k_1 t^*-серий, длина каждой из которых не меньше единицы;
- зафиксировать одно из возможных разбиений числа $m_t - \delta_1^{(t)}$ на $k_1 + 1 - \delta_1^{(t)}$ t-серий, длина каждой из которых не меньше двух.

Отсюда с учетом (24) получаем (23). (Если $m_{t^*} = 0$, то в формуле (23) полагаем $C_{m_{t^*-1}}^{k_1 - 1} = \{1, \text{если } k_1 = 0; 0, \text{если } k_1 \geq 1\}$.

С помощью (14) и (23) находим для $m_t \geq 2$ и $m_{t^*} \geq 0$

$$\sum_{\nu_{t^*=0}}^{m_{t^*}} \left| D(m_{t^*} - \nu_{t^*}, m_t, k_1, k_2; t) \right| =$$

$$= C_{k_1+1}^{m_t - k_2 - k_1 - 1} C_{m_{t^*}}^{k_1} Z(m_t - k_1 - 1, m_t - k_1 - k_2 - 1), \qquad (25)$$

$$\sum_{\nu_{t^*=0}}^{m_{t^*}} \sum_{\nu'_{t^*=1}}^{m_{t^*}} \left| D(m_{t^*} - \nu_{t^*} - \nu'_{t^*}, m_t, k_1 - 1, k_2; t) \right| =$$

$$= C_{k_1}^{m_t - k_2 - k_1} C_{m_{t^*}}^{k_1} Z(m_t - k_1, m_t - k_1 - k_2). \qquad (26)$$

Из (22), (25) и (26) следует равенство

$$Q = C_{m_{t^*}}^{k_1} \sum_{i \in \{k_1, k_1+1\}} C_i^{m_t - k_2 - i} Z(m_t - i,\, m_t - i - k_2)\chi(m_t \geq 2),$$

которое вместе с (20), (21), (5) и очевидными соотношениями

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(t\, t\, t) = k_2\} = \begin{cases} (\mathrm{P}\{\gamma_1 = t^*\})^n, & \text{если } m_t = k_1 = k_2 = 0, \\ 0 & \text{в остальных случаях,} \end{cases}$$

$$\mathrm{P}\{\eta(t_1 t_1^*) = k_1,\, \eta(t\, t\, t) = k_2\} =$$

$$= \begin{cases} (\mathrm{P}\{\gamma_1 = t^*\})^{n-1}\mathrm{P}\{\gamma_1 = t\}, & \text{если } m_t = 1,\ k_1 = k_2 = 0, \\ (n-1)(\mathrm{P}\{\gamma_1 = t^*\})^{n-1}\mathrm{P}\{\gamma_1 = t\}, & \text{если } m_t = k_1 = 1,\ k_2 = 0, \\ 0 & \text{в остальных случаях} \end{cases}$$

приводит к (17).

Используя схему доказательства равенства (17), можно проверить (16), (18) и (19).

5. Примеры к теореме 2

5.1. Иллюстрация использования равенства (16). В табл. 4 и на рис. 3 приведено использование соотношения (16) для $p = q = 1/2$, малой выборки длины n, $n = 20$, и некоторых значений k_1, k_2.

Таблица 4

(k_1, k_2)	P	P_c		(k_1, k_2)	P	P_c
(3, 7)	0,010046	0,160690308		(5, 7)	0,0258274	0,415276527
(7, 4)	0,011816	0,172506332		(6, 6)	0,0259638	0,441240311
(3, 2)	0,0121012	0,184607506		(4, 6)	0,0309591	0,47219944
(3, 6)	0,0121164	0,196723938		(6, 5)	0,0358648	0,50806427
(7, 5)	0,0131178	0,209841728		(6, 3)	0,0363922	0,544456482
(3, 5)	0,013607	0,223448753		(4, 2)	0,0369673	0,581423759
(3, 3)	0,0137167	0,237165451		(4, 5)	0,0379229	0,619346619
(3, 4)	0,0142107	0,251376152		(5, 6)	0,0386162	0,657962799
(4, 8)	0,015398	0,266774178		(6, 4)	0,0412607	0,699223518
(5, 8)	0,015399	0,282173157		(4, 3)	0,0421267	0,741350174
(6, 7)	0,0161362	0,298309326		(4, 4)	0,0421772	0,783527374
(6, 2)	0,0196438	0,31795311		(5, 2)	0,0471888	0,830716133
(4, 7)	0,0229654	0,340918541		(5, 5)	0,0509062	0,881622314
(5, 1)	0,0231276	0,364046097		(5, 4)	0,0590916	0,940713882
(4, 1)	0,025403	0,38944912		(5, 3)	0,0592861	1

Табл. 4 составлена из столбцов, содержание которых аналогично содержанию столбцов табл. 1.

ISSN 0572-2691

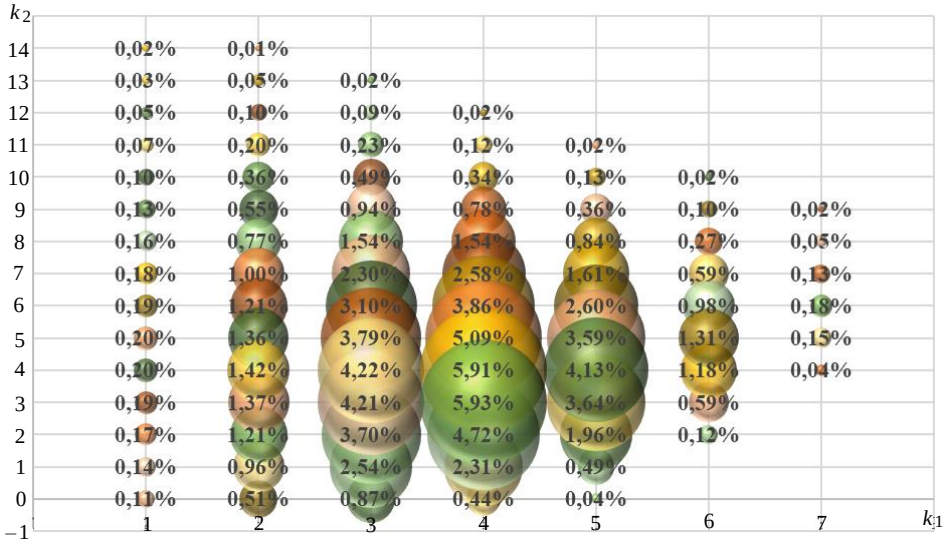

Рис. 3

На рис. 3 дана пузырьковая диаграмма, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2, третий (размер пузырька) — вероятность осуществления события $\{\eta(t_1 t_1^*) = k_1, \eta(t1t) + \eta(t0t) = k_2\}$, представленная в процентах.

5.2. Иллюстрация использования равенства (17). В табл. 5 и на рис. 4 приведено использование соотношения (17) для $p = q = 1/2$, малой выборки длины n, $n = 20$, и некоторых значений k_1, k_2.

Таблица 5

(k_1, k_2)	P	P_c		(k_1, k_2)	P	P_c
(3, 6)	0,011425	0,099241257		(4, 0)	0,0352583	0,363793373
(3, 1)	0,0120115	0,111252785		(6, 2)	0,0405884	0,404381752
(3, 5)	0,0139112	0,125164032		(4, 4)	0,040679	0,44506073
(3, 2)	0,0145435	0,139707565		(4, 1)	0,0458231	0,490883827
(5, 5)	0,0150604	0,15476799		(4, 3)	0,0493002	0,540184021
(3, 4)	0,0154448	0,170212746		(4, 2)	0,0516081	0,591792107
(3, 3)	0,0157051	0,185917854		(5, 3)	0,0564699	0,648262024
(7, 1)	0,0161695	0,202087402		(6, 1)	0,0605555	0,708817482
(4, 6)	0,0171032	0,219190598		(6, 0)	0,0657034	0,774520874
(6, 3)	0,0193596	0,238550186		(5, 0)	0,0684748	0,842995644
(7, 0)	0,027894	0,266444206		(5, 2)	0,0760345	0,91903019
(4, 5)	0,0288258	0,295269966		(5, 1)	0,0809698	1
(5, 4)	0,0332651	0,32853508				

Табл. 5 состоит из столбцов, содержание которых аналогично содержанию столбцов табл. 1.

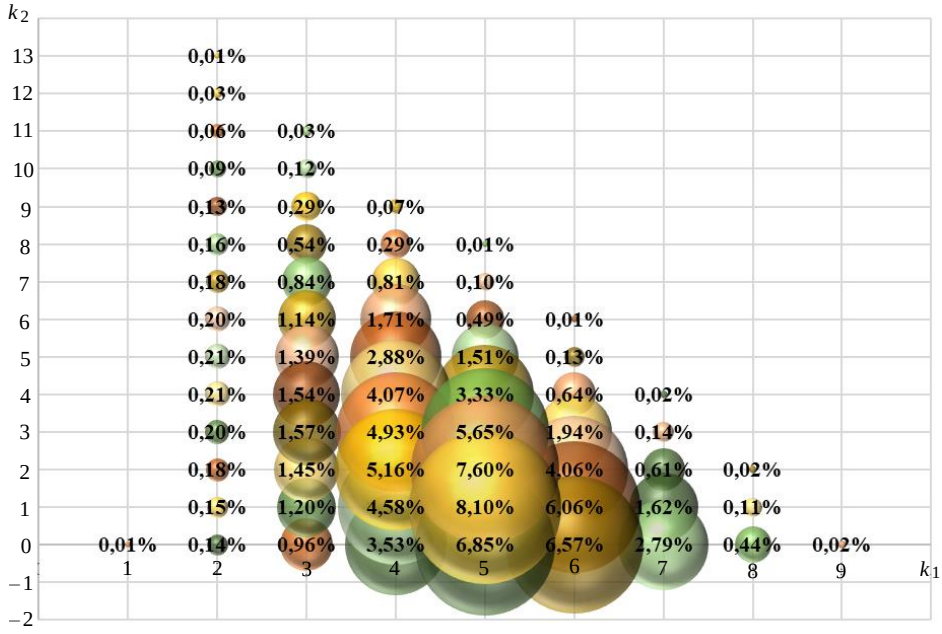

Рис. 4

На рис. 4 представлена пузырьковая диаграмма, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2, третий (размер пузырька) — вероятность осуществления события $\{\eta(t_1 t_1^*) = k_1, \eta(ttt) = k_2\}$, которая представлена в процентах.

5.3. Иллюстрация использования равенства (18). В табл. 6 и на рис. 5 приведено использование соотношения (18) для $p = q = 1/2$, малой выборки длины n, $n = 20$, и некоторых значений k_1, k_2.

Таблица 6

(k_1, k_2)	P	P_c		(k_1, k_2)	P	P_c
(2, 0)	0,0118675	0,075144768		(5, 4)	0,0383215	0,332155228
(7, 4)	0,016923	0,092067719		(3, 1)	0,0517502	0,383905411
(7, 5)	0,0171833	0,109251022		(5, 1)	0,0579596	0,441864967
(3, 2)	0,0206223	0,129873276		(6, 4)	0,0593233	0,501188278
(6, 5)	0,0233202	0,153193474		(6, 3)	0,066824	0,568012238
(6, 2)	0,0338459	0,187039375		(4, 2)	0,1008682	0,668880463
(4, 0)	0,0340939	0,221133232		(4, 1)	0,1047363	0,773616791
(3, 0)	0,0362511	0,2573843		(5, 3)	0,1050091	0,87862587
(4, 3)	0,0364494	0,293833733		(5, 2)	0,1213741	1

Табл. 6 составлена из столбцов, содержание которых аналогично содержанию столбцов табл. 1.

На рис. 5 дана пузырьковая диаграмма, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2, третий (размер пузырька) — вероятность осуществления события $\{\eta(t_1 t_1^*) = k_1, \eta(t t^* t) = k_2\}$, которая представлена в процентах.

ISSN 0572-2691

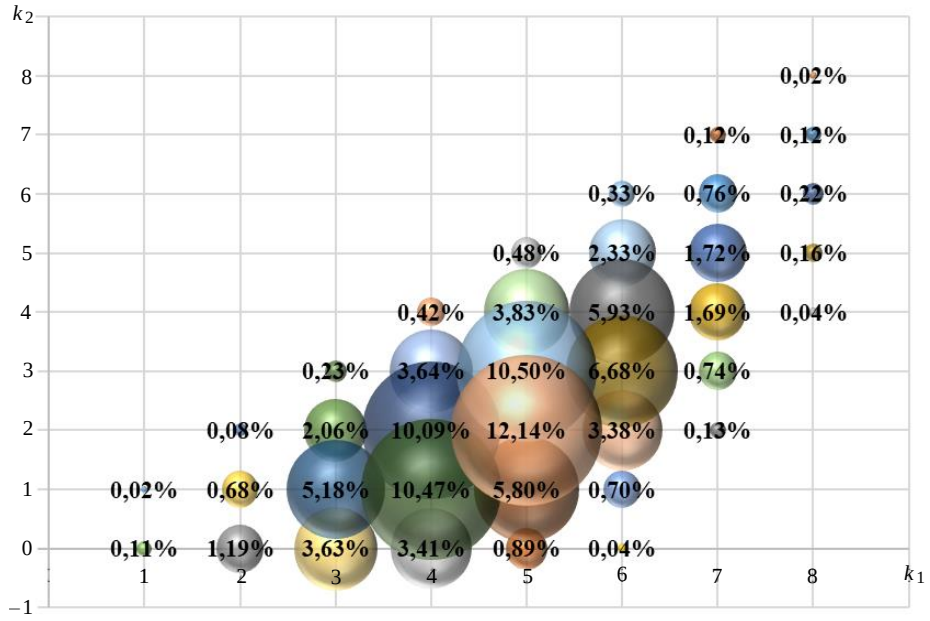

Рис. 5

5.4. Иллюстрация использования равенства (19). В табл. 7 приведено использование соотношения (19) для $p = q = 1/2,$ малой выборки длины $n,$ $n = 20,$ и некоторых значений $k_1,$ k_2 и k_3.

Таблица 7

k_1	k_2	k_3	P	P_c		k_1	k_2	k_3	P	P_c
7	0	4	0,010347	0,432714		5	3	2	0,018559	0,670737
6	1	2	0,010829	0,443543		6	1	4	0,018897	0,689634
5	2	1	0,011859	0,455402		4	3	1	0,01915	0,708784
4	5	2	0,012703	0,468105		4	3	2	0,019386	0,72817
6	2	3	0,01318	0,481285		5	0	1	0,02039	0,74856
5	4	3	0,013247	0,494532		4	1	1	0,02113	0,76969
4	4	1	0,013325	0,507856		5	3	3	0,021782	0,791471
4	1	2	0,013573	0,521429		5	1	3	0,022125	0,813597
5	0	3	0,01358	0,535009		4	2	1	0,022333	0,83593
6	0	4	0,01442	0,549429		6	1	3	0,022564	0,858494
6	2	4	0,014892	0,564321		6	0	3	0,025177	0,883671
4	0	1	0,016785	0,581105		5	2	3	0,025787	0,909458
4	4	2	0,017544	0,598649		5	0	2	0,028	0,937458
4	2	2	0,017824	0,616473		5	2	2	0,029125	0,966583
5	1	1	0,017853	0,634326		5	1	2	0,033417	1
6	0	2	0,017853	0,652179						

В табл. 7 в первом, втором и третьем столбцах приведены все возможные варианты значений $k_1,$ k_2 и $k_3,$ для которых вероятность $\mathrm{P}\{\eta(t_1\, t_1^*) = k_1, \eta(t\, t\, t) = k_2,$

$\eta(t\,t^*\,t) = k_3\} \geq 0,01$, а содержание четвертого и пятого столбцов аналогично содержанию соответственно второго и третьего столбцов табл. 1 (см. 3.1).

Заключение

В работе установлены совместные распределения числа 2-цепочек и числа 3-цепочек фиксированного вида случайной битовой последовательности. Даны примеры использования этих распределений. Возможным применением полученных формул может быть проверка гипотезы случайности расположения нулей и единиц в (0, 1)-последовательности конечной длины.

В.І. Масол, С.В. Поперешняк

СТАТИСТИЧНИЙ АНАЛІЗ ЛОКАЛЬНИХ ДІЛЯНОК БІТОВИХ ПОСЛІДОВНОСТЕЙ

Розглянуто сумісні розподіли числа 2-ланцюжків і числа 3-ланцюжків фіксованого вигляду випадкової (0, 1)-послідовності, які дозволяють здійснювати статистичний аналіз локальних ділянок цієї послідовності. Сформульовано і доведено дві теореми. У теоремах 1, 2 для числа s-ланцюжків вигляду $t\,t^*$, $t1t^*$, $t0t^*$, $t_1t_1^*$, $t_1 1t_1^*$ ($t_1t_1^*$, $t1t$, $t0t$, ttt,, tt^*t), що з'явилися у випадковій бітовій послідовності довжини n, $n > 0$, встановлено явні вирази сумісних розподілів таких подій: $\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) + \eta(t\,0\,t^*) = k_2\}$, $\{\eta(t_1\,t_1^*) = k_1,\ \eta(t_1\,t\,t_1^*) = k_2\}$, $\{\eta(t\,t^*) = k_1,\ \eta(t1t^*) = k_2,\ \eta(t\,0\,t^*) = k_3\}$, $(\{\eta(t_1\,t_1^*) = k_1,\ \eta(t1t) + \eta(t\,0\,t) = k_2\}$, $\{\eta(t_1\,t_1^*) = k_1,\ \eta(t\,t\,t) = k_2\}$, $\{\eta(t_1\,t_1^*) = k_1, \eta(t\,t^*\,t) = k_2\}$, $\{\eta(t_1\,t_1^*) = k_1, \eta(t\,t\,t) = k_2, \eta(t\,t^*t) = k_3\})$, де $\eta(t_1\,t_2 \ldots t_s)$ — число s-ланцюжків вигляду $t_1\,t_2 \ldots t_s$ у вихідній n-вимірній (0, 1)-послідовності; k_1, k_2, k_3 — відповідні цілі невід'ємні числа. Одне з основних припущень кожної теореми полягає у тому, що нулі і одиниці бітової послідовності — це незалежні однаково розподілені випадкові величини. Доведення формул для розподілів зазначених подій побудовано на підрахунку числа відповідних сприятливих подій за умови, що (0, 1)-послідовність містить фіксовану кількість нулів і одиниць. В якості прикладів використання явних виразів сумісних розподілів наведено таблиці, в яких розміщено значення ймовірностей перерахованих вище подій для випадкової (0,1)-послідовності довжини n, $n = 20$, і деяких значень параметрів k_1, k_2 і k_3 за припущенням, що нулі і одиниці з'являються рівноймовірно. Для наочності, частина таблиць проілюстрована бульбашковими діаграмами. Знайдені формули можуть становити інтерес для задач тестування локальних ділянок, які формуються на виході генераторів псевдовипадкових чисел, для деяких задач захисту інформації від несанкціонованого доступу, а також в інших сферах, де виникає необхідність в аналізі бітових послідовностей.

Ключові слова: s-ланцюжки, бітова послідовність, випадковість, локальні ділянки, сумісний розподіл.

V.I. Masol, S.V. Popereshnyak

STATISTICAL ANALYSIS OF LOCAL PLOTS OF BITS SEQUENCES

The joint distributions of the number of 2-chains and the number of 3-chains of a fixed form of a random (0, 1)-sequence, which allow a statistical analysis of local

sections of this sequence, were examined. Two theorems are formulated and proved. Consider s-chains of the form tt^*, $t1t^*$, $t0t^*$, $t_1t_1^*$, $t_11t_1^*$ ($t_1t_1^*$, $t1t$, $t0t$, ttt,, tt^*t), which appeared in random bit sequence of fixed length in Theorems 1, 2. For these *s*-chains, explicit expressions for the joint distributions of such events were established: $\{\eta(tt^*) = k_1, \eta(t1t^*) + \eta(t0t^*) = k_2\}$, $\{\eta(t_1t_1^*) = k_1, \eta(t_1tt_1^*) = k_2\}$, $\{\eta(tt^*) = k_1, \eta(t1t^*) = k_2, \eta(t0t^*) = k_3\}$, $(\{\eta(t_1t_1^*) = k_1, \eta(t1t) + \eta(t0t) = k_2\}$, $\{\eta(t_1t_1^*) = k_1, \eta(ttt) = k_2\}$, $\{\eta(t_1t_1^*) = k_1, \eta(tt^*t) = k_2\}$, $\{\eta(t_1t_1^*) = k_1, \eta(ttt) = k_2, \eta(tt^*t) = k_3\})$, where $\eta(t_1t_2...t_s)$ is the number of *s*-chains of the form $t_1t_2...t_s$ in the initial *n*-dimensional (0, 1)-sequence; k_1, k_2 and k_3 are suitable non-negative integers. One of the main assumptions of each theorem is that zeros and ones in a bit sequence are independent identically distributed random variables. The proofs of the formulas for the distributions of these events are based on counting the number of corresponding conductive events, provided that the (0, 1)-sequence contains a fixed number of zeros and ones. As examples of the use of explicit expressions of joint distributions, tables that contain the probabilities of the above events for a random (0, 1)-sequence of length n, $n = 20$, and some values of the parameters k_1, k_2 and k_3 under the assumption that zeros and units appear equally likely are given. For illustrative purposes, some of the tables are illustrated by bubble chart. The established formulas may be of interest for tasks like testing local sections formed at the output of pseudorandom number generators. Also, they may be suitable for some tasks of protecting information from unauthorized access, as well as in other areas where it becomes necessary to analyze bit sequences.

Keywords: *s*-chains, bit sequence, randomness, local segments, joint distribution.

1. Гайдышев И.П. Программное обеспечение анализа данных AtteStat. Руководство пользователя. Версия 13. 2012. 505 с.
2. A statistical test suite for random and pseudorandom number generators for cryptographic applications. Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Alan Heckert, James Dray, San Vo. *National Institute of Standards and Technology. Special Publication 800-22 revision 1a*. 2010. 131 p.
3. Масол В.И. О распределении некоторых статистик (0, 1)-вектора. *Исслед. операций и АСУ*. Вып. 29. 1987. С. 23–27.
4. Yueqi Hu, Tom Polk, Jing Yang, Ye Zhao, Shixia Liu. Spot-tracking lens: A zoomable user interface for animated bubble charts. *IEEE Pacific Visualization Symposium (PacificVis)*. 2016. P. 16–23.

Получено 22.02.2019