УДК 519.237.3+519.669+681.51

В.И. Масол, С.В. Поперешняк

ЯВНЫЙ ВИД РАСПРЕДЕЛЕНИЯ ИЗБРАННЫХ ДВУМЕРНЫХ И ТРЕХМЕРНЫХ СТАТИСТИК (0, 1)-ПОСЛЕДОВАТЕЛЬНОСТИ

Ключевые слова: 2-цепочки, 3-цепочки, совместное распределение, локальные участки, критерий согласия.

Введение

Потребность в статистическом анализе локальных участков битовых последовательностей возникает во многих прикладных задачах, в частности: тестирование генераторов псевдослучайных чисел; защита информации от несанкционированного доступа; стеганоанализ мультимедийных объектов (изображения, видео, звуки), представленных в цифровом виде, а также информации в файлах, вычислительных сетях и т.п.

В данной работе, являющейся продолжением [1], установлен явный вид распределений специально отобранных двумерных и трехмерных статистик, характеризующих взаимное расположение нулей и единиц битовой последовательности. Применение полученных формул проиллюстрировано соответствующими таблицами, пузырьковыми диаграммами, примерами подбора распределения Бернулли, при котором достигается максимальная вероятность появления заданных количеств фиксированных конфигураций битовой последовательности.

Постановка задачи

Положим

$$\gamma_1 \gamma_2 \cdots \gamma_n \tag{1}$$

— последовательность случайных величин, каждая из которых принимает два значения: ноль или единицу, $\gamma_i = \{0, 1\}$, $i = 1, 2, ..., n$, $n > 0$. Подпоследовательность $\gamma_j \gamma_{j+1} \cdots \gamma_{j+s-1}$ последовательности (1) называется s-цепочкой [2, с. 366], $j = 1, 2, ..., n - s + 1$, $s = 1, 2, ..., n$. Обозначим $\eta_n(t_1 t_2 \ldots t_s)$ число всех s-цепочек в последовательности (1), которые совпадают с фиксированным набором $t_1 t_2 \ldots t_s$, где $t_i = \{0, 1\}$, $i = 1, 2, ..., s$. Например, для последовательности вида 011010111 случайная величина $\eta_9(t\,t)$ принимает значение 3, если $t = 1$, и значение 0, если $t = 0$.

Условие. Элементы γ_1, γ_2, ..., γ_n последовательности (1), — независимые одинаково распределенные случайные величины; вероятности событий $\{\gamma_i = 1\}$ и $\{\gamma_i = 0\}$ известны: $\mathrm{P}\{\gamma_i = 1\} = p$, $\mathrm{P}\{\gamma_i = 0\} = q$, $p + q = 1$, $i = 1, 2, ..., n$.

В настоящей работе установлены (при указанном *условии*) совместные распределения пар случайных величин $\eta_n(tt)$ и $\eta_n(t^* t^* t^*)$, $\eta_n(tt)$ и $\eta_n(t^* tt^*)$,

 ISSN 1028-0979

$\eta_n(tt)$ и $\eta_n(t^*t^*t^*) + \eta_n(t^*tt^*)$, а также совместное распределение случайных величин $\eta_n(tt)$, $\eta_n(t^*t^*t^*)$ и $\eta_n(t^*tt^*)$, где $t^* = 1-t$, $t \in \{0,1\}$.

Формулировка и доказательство теоремы

Теорема. Пусть выполняется сформулированное выше условие; целые числа k_1 и k_2 такие, что $k_1 \geq 0$, $k_2 \geq 0$.

Вероятность совместных событий $\{\eta_n(tt) = k_1\}$ и $\{\eta_n(t^*t^*t^*) = k_2\}$ удовлетворяет равенству

$$P\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\} = \sum_{m_1=0}^{n} p^{m_1} q^{m_0} Z(m_t; a) \sum_{i=1}^{3} \beta_i C_{a+\alpha_i}^{\delta_i} \times$$

$$\times Z(m_{t^*} - a - \alpha_i; a + \alpha_i - \delta_i) \chi(k_1 \leq m_t - 1, k_2 \leq m_{t^*} - 2), \quad (2)$$

где $m_1 + m_0 = n,$, $a = m_t - k_1$, $\alpha_i = \{-1,$ при $i = 1; 1$, при $i = 2; 0$, при $i = 3\}$; $\beta_i = \{1,$ при $i \in \{1, 2\}$; 2 при $i = 3\}$, $\delta_i = k_2 - m_{t^*} + 2(a + \alpha_i)$, $i \in \{1, 2, 3\}$, $Z(a; b) = \{C_{a-1}^{b-1},$ при $a \geq b \geq 1; 1$, при $a = b = 1; 0$, в остальных случаях$\}$, $\chi(E)$ — индикатор события E;

- вероятность совместных событий $\{\eta_n(tt) = k_1\}$ и $\{\eta_n(t^*tt^*) = k_2\}$ удовлетворяет равенству

$$P\{\eta_n(tt) = k_1, \eta_n(t^*t\ t^*) = k_2\} = \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \{[C_{a-2}^{k_2} C_{m_{t^*}-1}^{a-2} C_{k_1+1}^{a-k_2-1} +$$

$$+ 2C_{a-1}^{k_2} C_{m_{t^*}-1}^{a-1} C_{k_1}^{a-k_2-1} + C_a^{k_2} C_{m_{t^*}-1}^{a} Z(k_1; a - k_2)] \chi(m_{t^*} \geq 1) +$$

$$+ \chi(n - k_1 - 1 = m_{t^*} = k_2 = 0)\} \chi(k_1 \leq m_t - 1, k_2 \leq m_{t^*} - k_1); \quad (3)$$

- вероятность совместных событий $\{\eta_n(tt) = k_1\}$ и $\{\eta_n(t^*t^*t^*) + \eta_n(t^*tt^*) = k_2\}$ удовлетворяет равенству

$$P\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) + \eta_n(t^*t\ t^*) = k_2\} = \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \times$$

$$\times \Big\{\Big[\sum\nolimits_1 C_{a-2}^{\delta} C_{a-1}^{\delta^*} C_{k_1+1}^{a-\delta-1} Z(m_{t^*} - a + 1; a - \delta^* - 1) + \sum\nolimits_2 C_a^{\delta} C_{a+1}^{\delta^*} Z(k_1; a - \delta) \times$$

$$\times Z(m_{t^*} - a - 1; a - \delta^* + 1) + 2\sum\nolimits_3 C_{a-1}^{\delta} C_a^{\delta^*} C_{k_1}^{a-\delta-1} Z(m_{t^*} - a; a - \delta^*)] \times$$

$$\times \chi(m_{t^*} \geq 1) + \chi(n - k_1 - 1 = m_{t^*} = k_2 = 0)\} \chi(k_1 \leq m_t - 1, k_2 \leq n - k_1 - 2), \quad (4)$$

где $\sum_i$ — суммирование по всем целым неотрицательным числам δ и δ^* таким, что $\delta + \delta^* = \delta_i$, $\delta^* \leq m_{t^*}$;

- вероятность совместных событий $\{\eta_n(tt) = k_1\}$, $\{\eta_n(t^*t^*t^*) = k_2\}$ и $\{\eta_n(t^*tt^*) = k_3\}$ (здесь целое число $k_3 \geq 0$) удовлетворяет равенству

$$P\{\eta_n(tt) = k_1,\ \eta_n(t^*t^*t^*) = k_2,\ \eta_n(t^*tt^*) = k_3\} = \sum_{m_1=0}^{n} p^{m_1} q^{m_0} \times$$

$$\times \{[C_{a-2}^{k_3} C_{a-1}^{\delta_1} C_{k_1+1}^{a-k_3-1} Z(m_{t^*} - a + 1; a - \delta_1 - 1) + C_a^{k_3} C_{a+1}^{\delta_2} Z(k_1; a - k_3) \times$$

$$\times Z(m_{t^*} - a - 1; a - \delta_2 + 1) + 2C_{a-1}^{k_3} C_a^{\delta_3} C_{k_1}^{a-k_3-1} Z(m_{t^*} - a; a - \delta_3)] \chi(m_{t^*} \geq 1) +$$

$$+ \chi(n - k_1 - 1 = m_{t^*} = k_2 = k_3 = 0)\} \chi(k_1 + k_3 \leq m_t, k_2 \leq m_{t^*} - 2). \quad (5)$$

Доказательство. Проверим соотношение (2). В силу условия количество единиц (обозначим его v) в последовательности (1) имеет биномиальное распределение с параметрами (n, p) при $m = 0, 1, 2, ..., n$

$$\mathrm{P}\{\nu = m\} = C_n^m p^m q^{n-m}. \tag{6}$$

Используя формулу полной вероятности, получаем

$$\mathrm{P}\{E_1, \mathrm{E}_2\} = \sum_{m_1=0}^{n} \mathrm{P}\{\nu = m_1\} \cdot \mathrm{P}\{E_1, E_2 / \nu = m\}, \tag{7}$$

где $E_1 = \{\eta_n(tt) = k_1\}$, $E_2 = \{\eta_n(t^*t^*t^*) = k_2\}$.

Проверим равенство

$$P\{E_1, E_2 / \nu = m_1\} = Z(m_t, a)\sum_{i=1}^{3} \beta_i C_{a+\alpha_i}^{\delta_i} Z(m_{t^*} - a - \alpha_i; a + \alpha_i - \delta_i) \times$$

$$\times \chi(k_1 \le m_t - 1; k_2 \le m_{t^*} - 2)\ (C_n^{m_1})^{-1}. \tag{8}$$

Заметим, что нарушение хотя бы одного из двух последних неравенств приводит к соотношению $\mathrm{P}\{E_1, E_2 / \nu = m_1\} = 0$. С этой целью введем обозначения:

• $\Omega(n, m_1)$ — совокупность всех n-мерных (0, 1)-векторов, каждый из которых содержит m_1 единиц и m_0 нулей, $m_0 + m_1 = n$;

• Q — число всех векторов $\vec{\upsilon}$, $\vec{\upsilon} \in \Omega(n, m_1)$, для каждого из которых имеют место события E_1 и E_2.

Число Q представим в виде

$$Q = |\Omega_1^{(tt)}| + |\Omega_2^{(t^*t^*)}| + |\Omega_3^{(tt^*)}| + |\Omega_4^{(t^*t)}|, \tag{9}$$

где $\Omega_i^{(\xi\zeta)}$, $i = 1, 2, 3, 4$, — совокупность всех векторов, начинающихся с элемента ξ, заканчивающихся элементом ζ, имеющих m_t t-элементов, m_{t*} $t*$-элементов, $m_t + m_{t*} = n$, k_1 2-цепочек вида tt, k_2 3-цепочек вида $t^*t^*t^*$, $\Omega_i^{(\xi\zeta)} \subseteq \Omega(n, m_1)$, $(\xi\,\zeta) = \{(tt)$, при $i = 1$; (t^*t^*), при $i = 2$; (tt^*), при $i = 3$; $(t*t)$, при $i = 4\}$. Далее заметим: события E_1 и E_2 позволяют утверждать, что число $\delta_{1,\,i}^*$ $t*$-серий длины 1 в векторе $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$ удовлетворяет равенству

$$\delta_{1,\,i}^* = \max(0,\ k_2 - m_{t*} + 2(a + \alpha_i)),\ \ i = 1, 2, 3, 4, \tag{10}$$

где $\alpha_4 = \alpha_3 = 0$. В свою очередь, при $i = 1, 2, 3, 4$ вектор $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$ определяется однозначно, если зафиксировать:

• одно из $R_{i,\,1}$ всех возможных размещений $t*$-серий длины 1 на всех $a + \alpha_i$ возможных позициях;

• одно из $R_{i,\,2}$ всех возможных $t*$-серий длины не менее, чем 2, которые можно сформировать из $m_{t*} - \delta_{1,\,i}^*$ $t*$-элементов и разместить на всех $m_t - k_1 + \alpha_i - \delta_i$ возможных позициях;

• одно из $R_{i,3}$ всех возможных размещений t-серий длины не менее, чем 1, которые можно сформировать из m_t t-элементов и разместить на всех $m_t - k_1$ возможных позициях.

Для чисел $R_{i,j}$, $j = 1, 2, 3$, $i = 1, 2, 3, 4,$ находим $R_{i,1} = C^{\delta_i}_{a+\alpha_i}$, $R_{i,2} = Z(m_{t^*} - - a - \alpha_i; a - \delta_i + \alpha_i)$, $R_{i,3} = Z\left(m_t; a\right)$ $R_{i,3} = Z(m_t; a),$ так что

$$|\Omega_i^{(\xi\zeta)}| = Z(m_t; a) C^{\delta_i}_{a+\alpha_i} Z(m_{t^*} - a - \alpha_i; a + \alpha_i - \delta_i), \; i \in \{1, 2, 3, 4\}. \tag{11}$$

С учетом (9)–(11) и равенств $\mathrm{P}\{E_1, E_2 / \nu = m_1\} = (|\Omega(n, m_1)|)^{-1} Q$ $\chi(k_1 \le m_t - 1; \;\; k_2 \le m_{t^*} - 2)$, $\left|\Omega(n, m_1)\right| = C_n^{m_1}$ приходим к (8). Соотношения (6)–(8) доказывают (2).

Проверим соотношение (3). Для этого воспользуемся введенными ранее обозначениями, придав некоторым из них следующую интерпретацию:

• Q — число всех векторов $\vec{\upsilon}$, $\vec{\upsilon} \in \Omega(n, m_1)$, для каждого из которых имеют место события E_1 и $\{\eta_n(t^* t t^*) = k_2\}$;

• $\Omega_i^{(\xi\zeta)}$ — совокупность всех векторов, начинающихся с элемента ξ, заканчивающихся элементом $\zeta,$, имеющих m_t t-элементов, m_{t^*} t^*-элементов, $m_t + m_{t^*} = n$, k_1 2-цепочек вида $t\,t$, k_2 3-цепочек вида $t^* t\, t^*$, $\Omega_i^{(\xi\zeta)} \subseteq \Omega(n, m_1)$, $i = 1, 2, 3, 4$.

Далее заметим: события E_1 и $\{\eta_n(t^* t t^*) = k_2\}$ позволяют утверждать, что число $\delta_{1,i}^{(t)}$ t-серий длины 1, которые расположены между t^*-сериями в векторе $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$, удовлетворяют равенству

$$\delta_{1,i}^{(t)} = k_2. \tag{12}$$

В свою очередь, число $R_{i,1}$ при $i = 1, 2, 3, 4$ всех размещений t-серий длины 1 в векторе $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$ на $a + \lambda_i$ возможных позициях (т.е. на позициях между t^*-элементами), где $\lambda_1 = 2$, $\lambda_2 = 0$, $\lambda_3 = \lambda_4 = 1,$ равняется

$$R_{i,1} = \begin{pmatrix} a - \lambda_i \\ k_2 \end{pmatrix} \tag{13}$$

при $i = 1, 2, 3, 4$. Для векторов $\vec{\upsilon} \in \Omega_1^{(tt)}$, которые не содержат t^*-элементов, т.е. представляют собою одну t-серию длины n, полагаем по определению

$$R_{1,1} = \chi(n - k_1 - 1 = k_2 = m_{t^*} = 0). \tag{14}$$

Для числа $R_{i,2}$ t^*-серий длины не менее, чем 1, которые можно сформировать из m_{t^*} t^*-элементов и расположить на всех $a - \lambda_i$ возможных позициях, находим

$$R_{1,2} = C^{a-2}_{m_{t^*}-1}, \; R_{2,2} = C^{a}_{m_{t^*}-1}, \; R_{3,2} = R_{4,2} = C^{a-1}_{m_{t^*}-1}. \tag{15}$$

Принимая во внимание, что вектор $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$, $i \in \{1, 3, 4\}$, может иметь t-серий длины 1 не обязательно между t^*-сериями, получаем, что число $R_{i,3}$ всех возможных размещений t-серий ненулевой длины в векторе $\vec{\upsilon} \in \Omega_i^{(\xi\zeta)}$, $i = 1, 2, 3, 4$, из оставшихся (в силу (12)) $m_t - k_2$ t-элементов с последующим их размещением на всех $a - k_2$ возможных позициях, равняется

$$R_{i,3} = \{C_{k_1+1}^{a-k_2-1} \text{ при } i = 1; Z(k_1; a-k_2),$$

$$\text{при } i = 2; C_{k_1}^{a-k_2-1} \text{ при } i = 3 \text{ и } i = 4\}. \quad (16)$$

Используя соотношение (13)–(16), получаем

$$|\Omega_1^{(tt)}| = C_{a-2}^{k_2} C_{m_{t^*}-1}^{a-2} C_{k_1+1}^{a-k_2-1} \chi(m_{t^*} \geq 1) + \chi(n - k_1 - 1 = m_{t^*} = k_2 = 0), \quad (17)$$

$$|\Omega_2^{(t^*t^*)}| = C_a^{k_2} C_{m_{t^*}-1}^{a} Z(k_1; a-k_2)\chi(m_{t^*} \geq 1), \quad (18)$$

$$|\Omega_3^{(tt^*)}| = |\Omega_4^{(t^*t)}| = C_{a-1}^{k_2} C_{m_{t^*}-1}^{a-1} C_{k_1}^{a-k_2-1} \chi(m_{t^*} \geq 1). \quad (19)$$

Для завершения доказательства равенства (3) осталось заметить, что

$$\mathrm{P}\{E_1, \eta_n(t^*tt^*) = k_2\} = \sum_{m_1=0}^{n} \mathrm{P}\{\nu = m_1\}\ \mathrm{P}\{E_1, \eta_n(t^*tt^*) = k_2 / \nu = m_1\} =$$

$$= \sum_{m_1=0}^{n} \mathrm{P}\{\nu = m_1\} Q\ \chi(k_1 \leq m_t - 1; k_2 \leq m_t - k_1) / C_n^{m_1},\ Q = \sum_{i=1}^{4} |\Omega_i^{(\xi\zeta)}|$$

и воспользоваться формулами (6), (17)–(19). Соотношение (3) установлено.

Аналогично доказательствам (2) и (3) проверяются равенства (4) и (5).

Теорема доказана.

Примеры к теореме

Иллюстрация использования равенства (2). В табл. 1 показано использование соотношения (2) для (0, 1)-последовательности длины n, $n = 40$, при $p = q = 0,5$ и некоторых значений k_1 и k_2.

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

В каждой строке четвертого столбца помещена сумма накопленных вероятностей до реализации события $\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\}$ включительно, где k_1 и k_2 указаны в этой же строке в первом столбце.

Например, при $k_1 = 9$ и $k_2 = 6$ имеем: $\mathrm{P}\{\eta_n(tt) = k_1,\ \eta_n(t^*t^*t^*) = k_2\} = 0,013796375$, $\mathrm{P}_c = \sum \mathrm{P}\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\} = 0,793794344$, где знак суммирования Σ распространяется на все пары (k_1, k_2), для которых $\mathrm{P}\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\} \leq 0,013796375$.

Таблица 1

k_1	k_2	P	P_C	k_1	k_2	P	P_C
7	3	0,0101101	0,5630703	12	2	0,013572	0,7662092
11	6	0,0101204	0,5731906	8	6	0,0137887	0,779998
13	4	0,0102743	0,5834649	9	6	0,0137964	0,7937943
7	7	0,0104921	0,5939571	10	2	0,0139144	0,8077087
9	7	0,0105678	0,6045248	11	2	0,0143611	0,8220698
12	5	0,0106093	0,6151341	12	3	0,0147072	0,836777
8	7	0,0111663	0,6263005	8	5	0,0154593	0,8522363
13	2	0,0118103	0,6381108	8	4	0,0154718	0,8677081
13	3	0,0120136	0,6501244	10	5	0,0155747	0,8832828
7	6	0,0122567	0,6623811	9	3	0,0159665	0,8992494
9	2	0,0122588	0,6746399	11	4	0,0159809	0,9152302
7	4	0,0123032	0,6869431	9	5	0,0163608	0,931591
10	6	0,0124132	0,6993563	11	3	0,0165612	0,9481522
7	5	0,0130005	0,7123567	10	3	0,0170595	0,9652117
12	4	0,0133672	0,725724	9	4	0,0173246	0,9825363
11	5	0,0134526	0,7391766	10	4	0,0174637	1
8	3	0,013460595	0,752637202				

Рис. 1 дает пузырьковую диаграмму, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2 третий (размер пузырька) — вероятность осуществления события $\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\}$, выраженная в процентах.

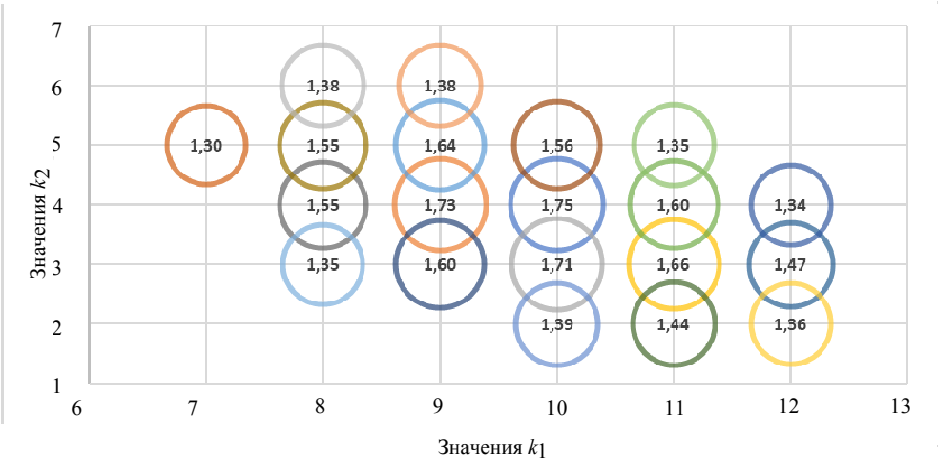

Рис. 1

Например, на рис.1 при $k_1 = 10$ и $k_2 = 2$ вероятность осуществления события $\{\eta_n(tt) = k_1, \eta_n(t^*t^*t^*) = k_2\}$ в процентах равняется 1,39 %.

Иллюстрация использования равенства (3). Использование равенства (3) для (0,1)-последовательности длины n, $n = 40$ при $p = q = 0,5$ и некоторых значений k_1 и k_2 приведено в табл. 2 и на рис. 2.

Таблица 2

k_1	k_2	P	P_C	k_1	k_2	P	P_C
11	2	0,0101099	0,3152758	10	6	0,017482	0,5824817
6	8	0,0102286	0,3255044	6	6	0,0176122	0,6000939
15	3	0,0103251	0,3358295	13	3	0,0184138	0,6185076
5	6	0,0107842	0,3466137	10	3	0,0188041	0,6373117
14	2	0,0109112	0,3575248	8	4	0,0198655	0,6571772
14	4	0,0114776	0,3690025	12	3	0,0209282	0,6781054
9	7	0,0115053	0,3805077	7	5	0,0209423	0,6990477
12	2	0,0116268	0,3921346	11	3	0,0211453	0,720193
11	6	0,0118179	0,4039524	11	5	0,0221599	0,742353
13	2	0,0118919	0,4158443	9	6	0,0224668	0,7648198
5	7	0,0120288	0,4278731	12	4	0,0228385	0,7876583
7	4	0,012981	0,4408541	7	6	0,0230041	0,8106624
6	5	0,0133541	0,4542082	8	6	0,0247609	0,8354233
14	3	0,014519	0,4687272	9	4	0,0255723	0,8609957
9	3	0,0145375	0,4832647	8	5	0,0270222	0,8880179
8	7	0,0153757	0,4986404	11	4	0,0270586	0,9150765
12	5	0,0157799	0,5144203	10	5	0,0273544	0,9424309
6	7	0,0161281	0,5305483	10	4	0,0282078	0,9706387
13	4	0,0171188	0,5476672	9	5	0,0293613	1
7	7	0,017332503	0,564999679				

Табл. 2 образована из столбцов, интерпретация которых аналогична интерпретации содержимого столбцов табл. 1.

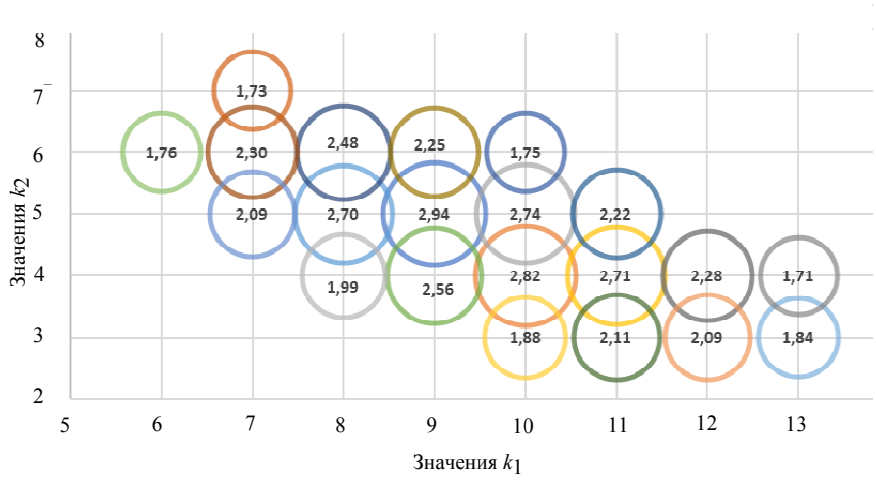

Рис. 2

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

Иллюстрация использования равенства (4). Использование равенства (4) для (0,1)-последовательности длины n, $n = 40$, при $p = q = 0,5$ и некоторых значений k_1 и k_2 приведено в табл. 3 и на рис. 3.

Таблица 3

k_1	k_2	P	P_C	k_1	k_2	P	P_C
8	8	0,010542	0,4619649	12	8	0,015998	0,6961751
14	6	0,010683	0,4726478	7	12	0,016582	0,7127572
13	8	0,010711	0,4833586	9	11	0,016816	0,7295729
9	12	0,010967	0,4943258	12	7	0,017297	0,7468699
12	9	0,011709	0,5060349	9	8	0,017318	0,7641882
10	11	0,01182	0,5178553	8	9	0,017336	0,781524
11	6	0,012002	0,5298574	11	9	0,017389	0,7989132
6	11	0,01207	0,5419274	10	10	0,017662	0,8165747
11	10	0,012085	0,5540125	7	11	0,018043	0,8346179
7	13	0,012501	0,5665131	11	7	0,018203	0,8528209
6	13	0,012788	0,5793013	8	11	0,019663	0,8724838
13	6	0,013396	0,5926978	11	8	0,02024	0,8927235
13	7	0,013655	0,6063527	8	10	0,020948	0,9136716
6	12	0,013983	0,6203357	10	8	0,021072	0,9347441
12	6	0,01407	0,6344056	9	10	0,021492	0,9562362
8	12	0,015005	0,6494103	10	9	0,021713	0,9779493
10	7	0,015358	0,6647683	9	9	0,022051	1
7	10	0,015409	0,680176894				

Табл. 3 образована из столбцов, интерпретация которых аналогична интерпретации содержимого столбцов табл. 1.

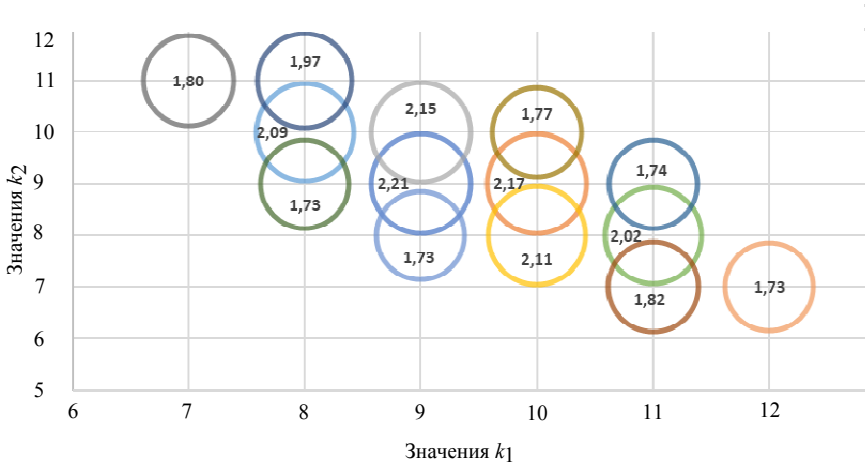

Рис. 3

Рис. 3 дает пузырьковую диаграмму, в которой первый параметр (горизонтальная ось) — значение k_1, второй (вертикальная ось) — значение k_2, третий (размер пузырька) — вероятность осуществления события $\{\eta_n(tt) = k_1,$ $\eta_n(t^*t^*t^*) + \eta_n(t^*tt^*) = k_2\}$, которая представлена в процентах.

Иллюстрация использования равенства (5). В табл. 4 приведено использование соотношения (5) для $(0,1)$-последовательности длины n, $n = 24$ при $p = q = 0,5$ и некоторых значений k_1, k_2 и k_3.

Таблица 4

k_1	k_2	k_3	P	P_C
5	1	3	0,009096	0,851162
4	4	3	0,009398	0,86056
5	1	4	0,009748	0,870309
8	1	2	0,009901	0,88021
7	1	3	0,009946	0,890155
4	3	3	0,009999	0,900154
6	3	2	0,010374	0,910529
7	1	2	0,010382	0,920911
4	2	4	0,010422	0,931332
6	2	2	0,010553	0,941885
7	2	2	0,011017	0,952902
5	3	3	0,011284	0,964186
6	2	3	0,011495	0,975681
6	1	3	0,011903	0,987584
5	2	3	0,012416	1

В первом, втором и третьем столбцах приведены все возможные варианты значений k_1, k_2 и k_3, для которых вероятность $\mathrm{P}\{\eta_n(tt)=k_1,\ \eta_n(t^*t^*t^*)=k_2,\ \eta_n(t^*tt^*)=k_3\}\ge 0,009$, а содержание четвертого и пятого столбцов аналогично содержанию третьего и четвертого столбцов табл. 1.

Замечание. Формулы (2)–(5) позволяют находить значение параметра p, $0\le p\le 1$, который максимизирует вероятность соответствующих (этим формулам) событий. Например, если $k_1=7$ и $k_2=5$ ($k_1=5$, $k_2=1$ и $k_3=3$), то вероятность $\mathrm{P}\{\eta_{40}(00)=7,\eta_{40}(111)=5\}$ ($\mathrm{P}\{\eta_{24}(00)=5,\ \eta_{24}(111)=1,\ \eta_{24}(101)=3\}$) достигает своего максимального значения 0,01569 (0,00915) при $p=0,55$ ($p=0,49$).

Заключение

Установлены совместные распределения заданного числа 2-цепочек и заданного числа 3-цепочек фиксированного вида случайной битовой последовательности конечной длины. Полученные соотношения могут применяться для проверки гипотезы случайности расположения нулей и единиц в (0,1)-последовательности, построения подходящего критерия согласия [2, с. 312], например, на основе подбора распределения Бернулли, максимизирующего вероятность одновременного появления заданных количеств указанных цепочек.

В.І. Масол, С.В. Поперешняк

ЯВНИЙ ВИГЛЯД РОЗПОДІЛУ ОБРАНИХ ДВОВИМІРНИХ ТА ТРИВИМІРНИХ СТАТИСТИК (0,1)-ПОСЛІДОВНОСТІ

Розглянуто сумісні розподіли заданого числа 2-ланцюжків та заданого числа 3-ланцюжків фіксованого вигляду випадкової бітової послідовності, які дозволяють здійснювати статистичний аналіз локальних ділянок цієї послідовності. У якості 2-ланцюжків виступають всі конфігурації, що складаються з двох поспіль або нулів, або одиниць бітової послідовності заданої довжини. У свою чергу, 3-ланцюжками являються всі конфігурації, що складаються з трьох поспіль або одиниць (за умови, що 2-ланцюжки є нульовими), або нулів (за умови, що 2-ланцюжки одиничні), а також в якості 3-ланцюжків розглядаються всі конфігурації, що складаються або з трьох поспіль цифр: один, нуль і один (за умови, що 2-ланцюжки нульові), або з трьох поспіль цифр: нуль, один і нуль (за

 ISSN 1028-0979

умови, що 2-ланцюжки одиничні). Встановлено явні вирази двовимірних і тривимірних сумісних розподілів подій, що відображають кількість деяких комбінацій зазначених ланцюжків у скінченній випадковій бітовій послідовності. Одне з основних припущень полягає у тому, що нулі та одиниці у бітової послідовності — це незалежні однаково розподілені випадкові величини. Доведення формул для розподілів зазначених подій побудовані на підрахунку числа відповідних сприятливих подій за умови, що бітова послідовність містить фіксовану кількість нулів і одиниць. Як приклади використання явних виразів сумісних розподілів наведені таблиці, в яких розміщені значення ймовірностей перерахованих вище подій для випадкової бітової послідовності довжини 40 (табл. 1–3) та довжини 24 (табл. 4) для деяких фіксованих значень числа 2-ланцюжків і числа 3-ланцюжків у припущенні, що нулі та одиниці з'являються незалежно і рівноймовірно. Табл. 1–3 проілюстровані 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: 2-ланцюжки, 3-ланцюжки, сумісний розподіл, локальні ділянки, критерій згоди.

V.I. Masol, S.V. Popereshnyak

EXPLICIT DISTRIBUTION OF SELECTED TWO-DIMENSIONAL AND THREE-DIMENSIONAL STATISTICS OF THE (0,1)-SEQUENCE

The joint distributions of the given number of 2-chains and the given number of 3-chains of a fixed form of a random bit sequence are considered, which allow performing a statistical analysis of local sections of this sequence. All configurations consisting of two consecutive zeros or ones of a bit sequence of a given length act as 2-chains. In turn, 3-chains are all configurations consisting of three consecutive either ones (provided that the 2-chains are zero) or zeros (provided that the 2-chains are one), as well as 3-chains all configurations are considered that consist either of three consecutive digits: one, zero and one (provided that the 2- chains are zero), or of three consecutive digits: zero, one and zero (provided that the 2- chains are one). The paper establishes explicit expressions for two-dimensional and three-dimensional joint distributions of events, reflecting the number of some combinations of the indicated chains in a finite random bit sequence. One of the basic assumptions is that zeros and ones in a bit sequence are independent, equally distributed random variables. The proofs of the formulas for the distributions of these events are based on counting the number of corresponding favorable events, provided that the bit sequence contains a fixed number of zeros and ones. As examples of using explicit expressions of joint distributions, tables are given in which the values of the probabilities of the events listed above for a random bit sequence of length 40 (tables 1–3) and length 24 (table 4) are given for some fixed values of the number of 2-chains and the number 3-chains under the assumption that zeros and ones appear independently and uniformly. For clarity, tables 1-3 are illustrated with bubble charts. The established formulas may be of interest for the problems of testing local sections formed at the output of pseudo-random number generators, for some problems of protecting information from unauthorized access, as well as in other areas where it becomes necessary to analyze bit sequences.

Keywords: 2-chains, 3-chains, joint distribution, local areas, goodness-of-fit test.

1. Масол В.И., Поперешняк С.В. Статистический анализ локальных участков битовых последовательностей. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2019. № 5. С. 92–105.
2. Ивченко Г.И., Медведев Ю. И. Введение в математическую статистику. М.: Изд-во ЛКИ, 2010. 600 с.

Получено 19.01.2021