М. П. Савелов, Предельное совместное распределение статистик критериев «Monobit test», «Frequency Test within a Block» и «Binary Matrix Rank Test», *Дискрет. матем.*, 2022, том 34, выпуск 4, 84–98

DOI: 10.4213/dm1739

Дискретная математика

том 34 выпуск 4 * 2022

УДК 519.248 DOI https://doi.org/10.4213/dm1739

Предельное совместное распределение статистик критериев «Monobit test», «Frequency Test within a Block» и «Binary Matrix Rank Test»

© 2022 г. М. П. Савелов*

Найдено предельное совместное распределение статистик T_1, T_2, T_3 следующих трех критериев пакета NIST: «Monobit Test», «Frequency Test within a Block» и «Binary Matrix Rank Test» в ситуации, когда исследуемая последовательность состоит из независимых случайных величин, имеющих распределение Бернулли с параметром $p = \frac{1}{2}$. Установлены необходимые и достаточные условия асимптотической некоррелированности, а также необходимые и достаточные условия асимптотической независимости данных статистик. Доказано, что ковариационная матрица $C = \|C_{ij}\|$ предельного распределения вектора (T_1, T_2, T_3) удовлетворяет соотношениям $C_{12} = C_{21} = C_{13} = C_{31} = 0$, $C_{23} = C_{32} \geqslant 0$. Для широкого класса значений $p \neq \frac{1}{2}$ описано предельное поведение вектора (T_1, T_2, T_3).

Ключевые слова: совместные распределения статистик, статистический пакет NIST, критерии согласия, критерий частот, критерий частот в блоках, критерий рангов бинарных матриц, асимптотически некоррелированные статистики, асимптотически независимые статистики

1. Введение

Для проверки качества генераторов случайных чисел используются различные пакеты статистических критериев: NIST [1], Diehard, TestU01, Crypt-X и др. Одним из наиболее известных является пакет NIST, который используется для проверки генераторов двоичных последовательностей. Этому пакету и смежным вопросам посвящено большое количество работ, см., например, [2–18].

Пусть $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ — последовательность независимых случайных величин, имеющих распределение Бернулли $\mathrm{Bern}(p)$, $p \in (0, 1)$. Через H_{p_0} обозначим гипотезу, в соответствии с которой $p = p_0$. Для проверки гипотезы $H_{\frac{1}{2}}$ в пакете NIST предлагается использовать 15 статистических критериев. Мы рассмотрим статистики следующих трех из них: статистику критерия частот («Monobit Test»), критерия

*Место работы: МГУ им. М. В. Ломоносова, e-mail: savelovmp@gmail.com

частот в блоках («Frequency Test within a Block») и критерия рангов бинарных матриц («Binary Matrix Rank Test»). Будем предполагать, что в критерии «Frequency Test within a Block» используется N блоков, а в критерии «Binary Matrix Rank Test» используются матрицы размера $V_1 \times V_2$. Нас будет интересовать случай, когда V_1, V_2 и N фиксированы и n стремится к бесконечности.

Положим

$$L = V_1 V_2, \quad M = \left[\frac{n}{N}\right], \quad Q = \left[\frac{n}{L}\right].$$

Если из исходной последовательности случайных величин $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ отбросить последние $n - NM$ элементов, то оставшиеся элементы разбиваются на N «больших» непересекающихся блоков длины M: первый блок состоит из случайных величин $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M$, второй — из случайных величин $\varepsilon_{M+1}, \varepsilon_{M+2}, \ldots, \varepsilon_{2M}$, и т. д. Аналогичным образом можно отбросить последние $n - LQ$ элементов и разбить оставшиеся на Q «малых» блоков длины L. Понятия «большой блок» и «малый блок» потребуются в дальнейшем. Далее, положим

$$\pi_j = \frac{1}{M} \sum_{i=(j-1)M+1}^{jM} \varepsilon_i \quad (1 \leqslant j \leqslant N), \qquad S_k = \sum_{i=1}^{k} \varepsilon_i \quad (0 \leqslant k \leqslant n).$$

Малый блок под номером j $(1 \leqslant j \leqslant Q)$ имеет вид $(\varepsilon_{(j-1)L+1}, \varepsilon_{(j-1)L+2}, \ldots, \varepsilon_{jL})$. Построим из его элементов матрицу A_j размера $V_1 \times V_2$ следующим образом:

$$A_j = \begin{pmatrix} \varepsilon_{(j-1)L+1} & \varepsilon_{(j-1)L+2} & \cdots & \varepsilon_{(j-1)L+V_2} \\ \varepsilon_{(j-1)L+V_2+1} & \varepsilon_{(j-1)L+V_2+2} & \cdots & \varepsilon_{(j-1)L+2V_2} \\ \cdots & \cdots & \cdots & \cdots \\ \varepsilon_{(j-1)L+V_2(V_1-1)+1} & \varepsilon_{(j-1)L+V_2(V_1-1)+2} & \cdots & \varepsilon_{jL} \end{pmatrix}.$$

Обозначим через ζ_j ранг матрицы A_j над полем $GF(2)$. Множество значений ζ_j имеет вид $\{0, 1, \ldots, \min(V_1, V_2)\}$. Распределение ζ_j известно в случае, когда верна $H_{\frac{1}{2}}$ (см. §3.5 в [1] и формулу (1) в [21]). А именно,

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = r) = 2^{-(V_1-r)(V_2-r)} \prod_{i=0}^{r-1} \frac{(1-2^{i-V_1})(1-2^{i-V_2})}{1-2^{i-r}} \tag{1}$$

при $r = 0, 1, 2, \ldots, \min(V_1, V_2)$. Здесь через $\mathbf{P}_{\frac{1}{2}}$ обозначается вероятность, вычисляемая в предположении о том, что верна гипотеза $H_{\frac{1}{2}}$. При этом предполагается, что произведение по пустому множеству индексов равно единице. Отметим также, что асимптотические свойства $\mathbf{P}_{\frac{1}{2}}(\zeta_1 = r)$ установлены в теореме 3.2.1 [22].

Пусть η_j — количество единиц в j-м малом блоке, т. е.

$$\eta_j = \sum_{i=(j-1)L+1}^{jL} \varepsilon_i.$$

Таким образом, j-му малому блоку соответствуют случайные величины η_j и ζ_j, причем векторы (η_j, ζ_j), $1 \leqslant j \leqslant Q$, независимы и одинаково распределены. Кроме того, очевидно, что $0 \leqslant \zeta_j \leqslant \eta_j \leqslant L$ и $\zeta_j \leqslant \min(V_1, V_2)$.

Далее, фиксируем $K \geqslant 1$. Разобьем множество $\{0, 1, \ldots, \min(V_1, V_2)\}$ на $K+1$ непустых непересекающихся подмножеств $\alpha_0, \alpha_1, \ldots, \alpha_K$: $\{0, 1, \ldots, \min(V_1, V_2)\} = \bigsqcup_{k=0}^{K} \alpha_k$. При $0 \leqslant k \leqslant K$ и $0 \leqslant s \leqslant L$ положим

$$\nu_k = \sum_{i=1}^{Q} I_{\zeta_i \in \alpha_k}, \quad w_{k,s}(p) = \mathbf{P}(\zeta_1 \in \alpha_k, \eta_1 = s), \quad w_k(p) = \mathbf{P}(\zeta_1 \in \alpha_k) = \sum_{s=0}^{L} w_{k,s}(p).$$

Несложно видеть, что величина $\mathbf{P}(\zeta_1 \leqslant r \,|\, \eta_1 = s)$ не зависит от p, поэтому

$$w_{k,s}(p) = 2^L p^s (1-p)^{L-s} w_{k,s}\Big(\frac{1}{2}\Big).$$

Случайная величина ζ_1 с положительной вероятностью попадает в каждую из точек множества $\{0, 1, \ldots, \min(V_1, V_2)\}$, поэтому $w_k(p) > 0$ при $0 \leqslant k \leqslant K$. Рассмотрим следующие статистики:

$$T_1 = \frac{2S_n - n}{\sqrt{n}}, \quad T_2 = 4M \sum_{j=1}^{N} \Big(\pi_j - \frac{1}{2}\Big)^2, \quad T_3 = \sum_{k=0}^{K} \frac{(\nu_k - Q w_k(\frac{1}{2}))^2}{Q w_k(\frac{1}{2})}.$$

В выражении для T_3 фигурирует величина $w_k(\frac{1}{2})$, явную формулу для которой легко получить с помощью (1).

Статистики T_1, T_2 и T_3 используются в критериях «Monobit Test», «Frequency Test within a Block» и «Binary Matrix Rank Test» соответственно. Каждая из этих статистик позволяет построить критерий согласия с гипотезой $H_{\frac{1}{2}}$ (см. [1]). При этом для статистики T_3 в [1] используются значения $V_1 = V_2 = 32$, $K = 2$ и множества α_k^{NIST}, $0 \leqslant k \leqslant 2$, определенные следующим образом:

$$\alpha_0^{\mathrm{NIST}} = \{32\}, \quad \alpha_1^{\mathrm{NIST}} = \{31\}, \quad \alpha_2^{\mathrm{NIST}} = \{j \geqslant 0 \colon j \leqslant 30\}.$$

Выбор множеств α_k^{NIST} основан на следующей идее: если верна гипотеза $H_{\frac{1}{2}}$, то с большой вероятностью ранг случайной квадратной бинарной матрицы «почти максимален» в том смысле, что он отличается от максимально возможного не более чем на 2. Данный факт иллюстрируется приведенной ниже леммой 5, в соответствии с которой не только при $V_1 = V_2 = 32$, но и при любых $V_1 = V_2 \geqslant 2$ естественно будет выбрать значение $K = 2$ и множества $\alpha_0^* = \{\min(V_1, V_2)\}$, $\alpha_1^* = \{\min(V_1, V_2) - 1\}$, $\alpha_2^* = \{j \geqslant 0 : j \leqslant \min(V_1, V_2) - 2\}$.

Для построения критерия согласия с гипотезой H_p при произвольном $p \in (0, 1)$ можно использовать следующие статистики:

$$T_1^{(p)} = \frac{S_n - np}{\sqrt{np(1-p)}}, \quad T_2^{(p)} = \frac{M}{p(1-p)} \sum_{j=1}^{N} (\pi_j - p)^2, \quad T_3^{(p)} = \sum_{k=0}^{K} \frac{(\nu_k - Q w_k(p))^2}{Q w_k(p)}.$$

Данные статистики являются модификациями статистик T_i, $1 \leqslant i \leqslant 3$, причем $T_i = T_i^{(\frac{1}{2})}$ при $1 \leqslant i \leqslant 3$. Из построения $T_i^{(p)}$ следует, что если верна гипотеза H_p и числа N, V_1, V_2, K и множества $\alpha_0, \ldots, \alpha_K$ фиксированы, то

$$\mathcal{L}(T_1^{(p)}) \xrightarrow{d} \mathcal{N}(0, 1), \quad \mathcal{L}(T_2^{(p)}) \xrightarrow{d} \chi^2_N, \quad \mathcal{L}(T_3^{(p)}) \xrightarrow{d} \chi^2_K \tag{2}$$

при $n \to \infty$. В частности, при $p = \frac{1}{2}$ из (2) получаются предельные распределения каждой из статистик T_1, T_2, T_3 в случае, когда верна $H_{\frac{1}{2}}$.

Нам потребуются следующие определения. Всюду далее под векторами мы будем понимать векторы-строки. Пусть $\vec{\xi} = (\xi_1, \ldots, \xi_{d_1})$ и $\vec{\eta} = (\eta_1, \ldots, \eta_{d_2})$ — случайные векторы. Ковариацией случайных векторов $\vec{\xi}$ и $\vec{\eta}$ будем называть матрицу $\operatorname{cov}(\vec{\xi}, \vec{\eta}) = \mathbf{E}(\vec{\xi} - \mathbf{E}\vec{\xi})^{\mathrm{T}}(\vec{\eta} - \mathbf{E}\vec{\eta})$ размера $d_1 \times d_2$. По аналогии с одномерным случаем положим $\mathbf{D}\vec{\xi} = \operatorname{cov}(\vec{\xi}, \vec{\xi})$.

Предположим, что многомерные статистики $\vec{T} \in \mathbb{R}^{d_1}$ и $\vec{T}^* \in \mathbb{R}^{d_2}$ построены по выборке $(\varepsilon_1, \ldots, \varepsilon_n)$. Будем говорить, что статистики $\vec{T} \in \mathbb{R}^{d_1}$ и $\vec{T}^* \in \mathbb{R}^{d_2}$ асимптотически независимы, если существуют такие независимые случайные векторы $\vec{\zeta} \in \mathbb{R}^{d_1}$ и $\vec{\zeta}^* \in \mathbb{R}^{d_2}$, что $(\vec{T}, \vec{T}^*) \xrightarrow{d} (\vec{\zeta}, \vec{\zeta}^*)$ при $n \to \infty$. Будем говорить, что статистики $\vec{T} = (T_1, \ldots, T_{d_1}) \in \mathbb{R}^{d_1}$ и $\vec{T}^* = (T_1^*, \ldots, T_{d_2}^*) \in \mathbb{R}^{d_2}$ асимптотически некоррелированы (асимптотически неотрицательно коррелированы), если при всех $1 \leqslant i \leqslant d_1$, $1 \leqslant j \leqslant d_2$ выполнено равенство $\lim_{n\to\infty} \operatorname{cov}(T_i, T_j^*) = 0$ (соответственно, $\underline{\lim}_{n\to\infty} \operatorname{cov}(T_i, T_j^*) \geqslant 0$). Асимптотическая положительная коррелированность определяется аналогично с помощью неравенства $\underline{\lim}_{n\to\infty} \operatorname{cov}(T_i, T_j^*) > 0$. Отметим, что лемма 4 работы [18] (см. также [18, следствие 3]) позволяет при выполнении некоторых дополнительных условий свести вопрос об асимптотической некоррелированности (а также об асимптотической неотрицательной/положительной коррелированности) статистик к вопросу о том, являются ли нулевыми (соответственно, неотрицательными/положительными) соответствующие элементы ковариационной матрицы предельного распределения этих статистик.

Будем называть значение параметра $p \in (0, 1)$ «типичным», если выполнено одно из двух условий: либо $p = \frac{1}{2}$, либо существует такое $0 \leqslant j \leqslant K$, что $w_j(p) \neq w_j(\frac{1}{2})$.

Пример 1. Пусть $V_1 = V_2 = 2$, $L = 4$, $K = 1$, $\alpha_0 = \{2\}$ и $\alpha_1 = \{0, 1\}$. Выпишем все матрицы размера 2×2, ранг которых равен двум:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Заметим, что $w_0(p) = \mathbf{P}(\zeta_1 \in \alpha_0) = 2p^2(1-p^2)$, $w_1(p) = \mathbf{P}(\zeta_1 \in \alpha_1) = 1 - \mathbf{P}(\zeta_1 \in \alpha_0) = 1 - 2p^2(1-p^2)$. Поэтому значение $p_0 = \frac{\sqrt{3}}{2}$ не является типичным.

В теореме 1 настоящей работы при фиксированных $N, V_1, V_2, K, \alpha_0, \ldots, \alpha_K$ указано предельное (при $n \to \infty$) совместное распределение статистик $\sqrt{N}T_1^{(p)}$, $T_2^{(p)}$, $NT_3^{(p)}$ в случае, когда верна гипотеза H_p. Кроме того, при всех «типичных» значениях p описано предельное поведение вектора (T_1, T_2, T_3) в случае, когда верна гипотеза H_p — см. следствие 1 и лемму 2 ниже. Ковариационная матрица предельного распределения вектора (T_1, T_2, T_3) в случае, когда верна $H_{\frac{1}{2}}$, указана в следствии 1.

Кроме того, установлен следующий результат. Выберем любые два не пересекающихся подмножества из множества $\{T_1, T_2, T_3\}$ и образуем из них два вектора $\vec{T}$ и $\vec{T}^*$. Например, можно выбрать подмножества $\{T_1, T_2\}$ и $\{T_3\}$ и получить из них два вектора $\vec{T} = (T_1, T_2)$ и $\vec{T}^* = T_3$. В силу равенства $\left(T_1^{(\frac{1}{2})}, T_2^{(\frac{1}{2})}, T_3^{(\frac{1}{2})}\right) = (T_1, T_2, T_3)$ сформулированное ниже утверждение 1 позволяет получить необходимые и достаточные условия асимптотической независимости, а также необходимые и достаточные условия асимптотической некоррелированности для произвольной пары $\vec{T}$ и $\vec{T}^*$ в случае, когда верна гипотеза $H_{\frac{1}{2}}$.

Замечание 1. Пусть T_{runs}, T_{template}, T_{cusum}, $T_{\text{longest run}}$ — статистики критериев «Runs Test», «Non-overlapping Template Matching Test», «Cumulative Sums Test» и «Test for the Longest Run of Ones in a Block» пакета NIST. Совместные распределения статистик $T_1, T_2, T_{\text{runs}}$ изучались В. Г. Михайловым в 2019–2020 гг. Совместным распределениям статистик $T_1, T_2, T_{\text{cusum}}$ посвящена работа [16]. В [15] найдены предельные совместные распределения статистик $T_1, T_2, T_{\text{runs}}, T_{\text{template}}$. Настоящая работа наиболее близка к работе [18], в которой найдены предельные совместные распределения статистик $T_1, T_2, T_{\text{longest run}}$.

2. Основные результаты

Рассмотрим матрицу $C(p) = ||C_{ij}(p)||_{i,j=1}^{K+2}$ со следующими элементами:

$$C_{ij}(p) = \begin{cases} I_{i=j} - \sqrt{w_{i-1}(p)w_{j-1}(p)} & \text{при } 1 \leqslant i,j \leqslant K+1, \\ D_j^{-1}\big(\sum_{r=0}^{L} r w_{j-1,r}(p) - Lp \cdot w_{j-1}(p)\big), & \text{если } i = K+2, 1 \leqslant j \leqslant K+1, \\ D_i^{-1}\big(\sum_{r=0}^{L} r w_{i-1,r}(p) - Lp \cdot w_{i-1}(p)\big), & \text{если } j = K+2, 1 \leqslant i \leqslant K+1, \\ 1, & \text{если } i = j = K+2, \end{cases}$$

где $D_j = D_j(p) = \sqrt{Lp(1-p)w_{j-1}(p)}$.

Данная матрица симметрична и неотрицательно определена, так как является ковариационной. В самом деле, положим

$$\vec{X} = \left(\frac{I_{\zeta_1 \in \alpha_0} - w_0(p)}{\sqrt{w_0(p)}}, \frac{I_{\zeta_1 \in \alpha_1} - w_1(p)}{\sqrt{w_1(p)}}, \ldots, \frac{I_{\zeta_1 \in \alpha_K} - w_K(p)}{\sqrt{w_K(p)}}, \frac{\eta_1 - Lp}{\sqrt{Lp(1-p)}} \right)^{\top}. \tag{3}$$

Тогда $\mathbf{E}\vec{X} = \vec{0}$, и так как $\operatorname{cov}(\eta_1, I_{\zeta_1 \in \alpha_k}) = \sum_{r=0}^{L} r w_{k,r}(p) - Lp \cdot w_k(p)$, то матрица $C(p)$ является ковариационной матрицей вектора $\vec{X}$.

Теорема 1. *Пусть $\varepsilon_1, \varepsilon_2, \ldots$ — последовательность Бернулли с параметром $p \in (0,1)$. Пусть независимые случайные векторы $\vec{Y}^{(i)} = (Y_1^{(i)}, \ldots, Y_{K+2}^{(i)})$, $1 \leqslant i \leqslant N$, имеют распределение $\mathcal{N}(0, C(p))$. Если числа N, V_1, V_2, K и множества $\alpha_0, \ldots, \alpha_K$ фиксированы, то*

$$\big(\sqrt{N}T_1^{(p)}, T_2^{(p)}, NT_3^{(p)}\big) \xrightarrow{d} \left(\sum_{i=1}^{N} Y_{K+2}^{(i)}, \sum_{i=1}^{N} \big(Y_{K+2}^{(i)}\big)^2, \sum_{j=1}^{K+1} \left(\sum_{i=1}^{N} Y_j^{(i)} \right)^2 \right) \tag{4}$$

при $n \to \infty$. Ковариационная матрица F предельного вектора из соотношения (4) *имеет следующий вид*:

$$F = N \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2\sum_{j=1}^{K+1} C_{j,K+2}^2(p) \\ 0 & 2\sum_{j=1}^{K+1} C_{j,K+2}^2(p) & 2NK \end{pmatrix}. \tag{5}$$

Если верна гипотеза $H_{\frac{1}{2}}$, то ковариационную матрицу предельного распределения вектора $(\sqrt{N}T_1, T_2, NT_3)$ можно получить с помощью теоремы 1. Значит, можно получить и ковариационную матрицу предельного распределения вектора (T_1, T_2, T_3).

Следствие 1. *Пусть* $\varepsilon_1, \varepsilon_2, \dots$ *— последовательность Бернулли с параметром* $\frac{1}{2}$. *Пусть независимые случайные векторы* $\vec{Y}^{(i)} = (Y_1^{(i)}, Y_2^{(i)}, \dots, Y_{K+2}^{(i)})$, $1 \leqslant i \leqslant N$, *имеют распределение* $\mathcal{N}(0, C(\frac{1}{2}))$. *Если числа* N, V_1, V_2, K *и множества* $\alpha_0, \dots, \alpha_K$ *фиксированы, то*

$$(\sqrt{N}T_1, T_2, NT_3) \xrightarrow{d} \left(\sum_{i=1}^{N} Y_{K+2}^{(i)}, \sum_{i=1}^{N} \left(Y_{K+2}^{(i)}\right)^2, \sum_{j=1}^{K+1} \left(\sum_{i=1}^{N} Y_j^{(i)} \right)^2 \right)$$

при $n \to \infty$. *Ковариационная матрица предельного распределения вектора* (T_1, T_2, T_3) *имеет вид*

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2N & 2\sum_{j=1}^{K+1} C_{j,K+2}^2(\frac{1}{2}) \\ 0 & 2\sum_{j=1}^{K+1} C_{j,K+2}^2(\frac{1}{2}) & 2K \end{pmatrix}.$$

Лемма 1. *Если выполнены условия теоремы* 1, *то имеет место поэлементная сходимость ковариационных матриц* $\mathbf{D}(\sqrt{N}T_1^{(p)}, T_2^{(p)}, NT_3^{(p)}) \to F$, $n \to \infty$, *где* F *определена в* (5).

Далее, выберем любые два непересекающиеся подмножества из множества $\{T_1^{(p)}, T_2^{(p)}, T_3^{(p)}\}$ и образуем из них два вектора $\vec{T}$ и $\vec{T}^*$. Например, можно выбрать подмножества $\{T_1^{(p)}, T_2^{(p)}\}$ и $\{T_3^{(p)}\}$ и получить из них два вектора $\vec{T} = (T_1^{(p)}, T_2^{(p)})$ и $\vec{T}^* = T_3^{(p)}$. Следующее утверждение опирается на лемму 1 и позволяет установить необходимые и достаточные условия асимптотической независимости, а также необходимые и достаточные условия асимптотической некоррелированности для произвольной пары $\vec{T}$ и $\vec{T}^*$.

Утверждение 1. *Пусть* $\varepsilon_1, \varepsilon_2, \dots$ *— последовательность Бернулли с параметром* $p \in (0,1)$, *числа* N, V_1, V_2, K *и множества* $\alpha_0, \dots, \alpha_K$ *фиксированы и* $n \to \infty$.

1. *Статистики* $T_1^{(p)}$ *и* $(T_2^{(p)}, T_3^{(p)})$ *асимптотически некоррелированы, статистики* $T_2^{(p)}$ *и* $T_3^{(p)}$ *асимптотически неотрицательно коррелированы.*
2. $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$ *тогда и только тогда, когда* η_1 *и* $I_{\zeta_1 \in \alpha_j}$ *некоррелированы при всех* $0 \leqslant j \leqslant K$.
3. *Статистики* $T_2^{(p)}$ *и* $T_3^{(p)}$ *асимптотически некоррелированы* $\iff$ $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$.
4. *Статистики* $T_2^{(p)}$ *и* $T_3^{(p)}$ *асимптотически положительно коррелированы* $\iff \sum_{j=1}^{K+1} C_{j,K+2}^2(p) \neq 0$.
5. *Статистики* $T_1^{(p)}$ *и* $T_2^{(p)}$ *асимптотически зависимы.*
6. *Если хотя бы одна из статистик* $T_1^{(p)}$, $T_2^{(p)}$ *асимптотически независима с* $T_3^{(p)}$, *то* $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$.
7. *Если* $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$, *то вектор* $(T_1^{(p)}, T_2^{(p)})$ *и статистика* $T_3^{(p)}$ *асимптотически независимы.*

Пример 2. Пусть выполнены условия следствия 1. Как показывает непосредственная проверка, если $2 \leqslant V_1, V_2 \leqslant 5$, $K = 2$, $\alpha_0 = \{\min(V_1, V_2)\}$, $\alpha_1 = \{\min(V_1, V_2) - 1\}$, $\alpha_2 = \{j \geqslant 0 : j \leqslant \min(V_1, V_2) - 2\}$, то $C_{1,4}(\frac{1}{2}) \neq 0$, поэтому $\sum_{j=1}^{3} C_{j,4}^2(\frac{1}{2}) > 0$ и статистики T_1, T_2, T_3 попарно асимптотически зависимы.

Далее, следствие 1 описывает предельное распределение вектора (T_1, T_2, T_3) в случае, когда верна гипотеза $H_{\frac{1}{2}}$, фиксированы числа N, V_1, V_2, K и множества $\alpha_0, \dots, \alpha_K$ и $n \to \infty$. Этот результат дополняет следующая лемма.

Лемма 2. *Пусть $\varepsilon_1, \varepsilon_2, \dots$ — последовательность Бернулли с параметром $p \in (0,1)$, числа N, V_1, V_2, K и множества $\alpha_0, \dots, \alpha_K$ фиксированы и $n \to \infty$. Если $p \neq \frac{1}{2}$, то $(T_1, T_2) \xrightarrow{\text{п.н.}} (\operatorname{sgn}(2p-1) \cdot (+\infty), +\infty)$. Если существует такое $0 \leqslant k \leqslant K$, что $w_k(p) \neq w_k(\frac{1}{2})$, то*

$$(T_1, T_2, T_3) \xrightarrow{\text{п.н.}} (\operatorname{sgn}(2p-1) \cdot (+\infty), +\infty, +\infty).$$

3. Доказательства

Доказательства теоремы 1, следствия 1, утверждения 1 и лемм 1–2 аналогичны соответственно доказательствам теоремы 1, следствия 1, утверждения 1 и лемм 1–2 из [18]. Приведем их для полноты изложения.

3.1. Доказательство теоремы 1. Положим

$$R = \left[\frac{M}{L}\right] - 1.$$

Без ограничения общности будем предполагать, что n достаточно велико для того, чтобы выполнялось неравенство $R \geqslant 1$. Каждый из N больших блоков имеет длину M, поэтому в нем целиком содержится R или $R+1$ малых «подблоков». Например, если $n = 22$, $N = 3$ и $L = 3$, то $M = 7$, первый большой блок имеет вид $(\varepsilon_1, \dots, \varepsilon_7)$ и в нем целиком содержатся 2 малых блока $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ и $(\varepsilon_4, \varepsilon_5, \varepsilon_6)$, а второй большой блок имеет вид $(\varepsilon_8, \dots, \varepsilon_{14})$ и в нем целиком содержится только один малый блок $(\varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12})$.

Далее, для каждого n в каждом большом блоке фиксируем какие-нибудь R малых блоков, которые целиком в нем содержатся. Будем называть эти малые блоки «основными». Например, если $n = 22$, $N = 3$ и $L = 3$, то в качестве основных блоков можно взять следующие три малых блока: $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, $(\varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12})$ и $(\varepsilon_{16}, \varepsilon_{17}, \varepsilon_{18})$. Идея доказательства теоремы 1 состоит в том, что блоки, не являющиеся основными, не дают вклад в предельное распределение вектора (T_1, T_2, T_3), а по основным блокам строятся независимые случайные величины, к которым удобно применять центральную предельную теорему.

У каждого малого блока есть свой номер, принадлежащий множеству $\{1, 2, \dots, Q\}$. Множество номеров малых блоков, которые являются основными, обозначим через $Z = \{z_1, z_2, \dots\} \subset \{1, 2, \dots, Q\}$. Будем считать, что $z_1 < z_2 < \cdots$ Например, если $n = 22$, $N = 3$, $L = 3$, то $R = 1$, и если в качестве основных блоков взяли малые блоки $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, $(\varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12})$ и $(\varepsilon_{16}, \varepsilon_{17}, \varepsilon_{18})$, то $z_1 = 1$, $z_2 = 4$, $z_3 = 6$. Так как в каждом из N больших блоков содержится R основных малых блоков, то $|Z| = RN$. При этом $z_{R(j-1)+1}, z_{R(j-1)+2}, \dots, z_{Rj}$ — номера основных блоков, содержащихся в j-м большом блоке, $1 \leqslant j \leqslant N$.

Далее, содержащиеся в j-м большом блоке R малых основных блоков занимают LR мест из M, поэтому количества единиц в них удовлетворяют неравенству

$$0 \leqslant \sum_{i=(j-1)M+1}^{jM} \varepsilon_i - (\eta_{z_{R(j-1)+1}} + \eta_{z_{R(j-1)+2}} + \cdots + \eta_{z_{Rj}}) \leqslant M - LR \tag{6}$$

при всех $1 \leqslant j \leqslant N$. При $1 \leqslant i \leqslant RN$ и $0 \leqslant k \leqslant K$ положим

$$\tilde{\eta}_i = \eta_{z_i}, \quad \tilde{\zeta}_i = \zeta_{z_i}, \quad \tilde{\nu}_k = \sum_{i \in Z} I_{\zeta_i \in \alpha_k}.$$

Тогда $\tilde{\nu}_k = \sum_{i=1}^{RN} I_{\tilde{\zeta}_i \in \alpha_k}$. Так как $R = \left[\frac{M}{L}\right] - 1$, то $M - (R+1)L \leqslant L - 1$, откуда, учитывая (6), получаем, что

$$0 \leqslant M\pi_j - \sum_{s=1}^{R} \tilde{\eta}_{R(j-1)+s} \leqslant M - LR \leqslant 2L - 1. \tag{7}$$

Кроме того, векторы $(\tilde{\eta}_j, \tilde{\zeta}_j)$ независимы и имеют такое же распределение, как и вектор (η_1, ζ_1).

Так как $[x] \leqslant x < [x]+1$ при всех $x \in \mathbb{R}$, то $Q = \left[\frac{n}{L}\right] \leqslant \frac{n}{L}$ и $R+1 = \left[\frac{M}{L}\right] \geqslant \frac{M}{L} - 1 = \frac{1}{L}\left[\frac{n}{N}\right] - 1 \geqslant \frac{1}{L}\left(\frac{n}{N} - 1\right) - 1$, поэтому

$$Q - (R+1)N \leqslant \frac{n}{L} - N\left(\frac{1}{L}\left(\frac{n}{N} - 1\right) - 1\right) = \frac{N}{L} + N \leqslant 2N. \tag{8}$$

Далее, $\nu_k = \sum_{i=1}^{Q} I_{\zeta_i \in \alpha_k}$ и $Z \subset \{1, 2, \ldots, Q\}$, поэтому

$$Q \geqslant |Z| = RN \tag{9}$$

и, кроме того,

$$0 \leqslant \sum_{i=1}^{Q} I_{\zeta_i \in \alpha_k} - \sum_{i \in Z} I_{\zeta_i \in \alpha_k} \leqslant Q - |Z| = Q - RN \leqslant 3N$$

в силу (8), откуда следует, что

$$0 \leqslant \nu_k - \tilde{\nu}_k \leqslant 3N. \tag{10}$$

Положим

$$\tilde{\pi}_j = \frac{1}{M} \sum_{s=1}^{R} \tilde{\eta}_{R(j-1)+s}, \quad 1 \leqslant j \leqslant N, \quad \tilde{S}_n = \sum_{i=1}^{RN} \tilde{\eta}_i.$$

Тогда $\tilde{S}_n = M \sum_{j=1}^{N} \tilde{\pi}_j$. Кроме того, в силу (7) выполнено неравенство

$$0 \leqslant M\pi_j - M\tilde{\pi}_j \leqslant 2L - 1, \quad 1 \leqslant j \leqslant N. \tag{11}$$

Далее, $S_n = \sum_{i=1}^{NM} \varepsilon_i + \sum_{i=NM+1}^{n} \varepsilon_i$, причем в последней сумме $n - MN$ слагаемых и $n - MN < N$. Значит, $M \sum_{j=1}^{N} \pi_j = \sum_{i=1}^{NM} \varepsilon_i \in [S_n - N, S_n]$, откуда в силу (11) и равенства $\tilde{S}_n = M \sum_{j=1}^{N} \tilde{\pi}_j$ получаем, что

$$0 \leqslant S_n - \tilde{S}_n \leqslant 2LN. \tag{12}$$

Положим

$$\vec{X}^{(j)} = \left(\frac{I_{\tilde{\zeta}_j \in \alpha_0} - w_0(p)}{\sqrt{w_0(p)}}, \frac{I_{\tilde{\zeta}_j \in \alpha_1} - w_1(p)}{\sqrt{w_1(p)}}, \dots, \frac{I_{\tilde{\zeta}_j \in \alpha_K} - w_K(p)}{\sqrt{w_K(p)}}, \frac{\tilde{\eta}_j - Lp}{\sqrt{Lp(1-p)}} \right).$$

Так как векторы $(\tilde{\eta}_j, \tilde{\zeta}_j)$ независимы и имеют такое же распределение, как и вектор (η_1, ζ_1), то векторы $\vec{X}^{(j)}$ независимы и имеют такое же распределение, как вектор $\vec{X}$ из формулы (3). Следовательно, векторы $\sum_{j=1}^{R} \vec{X}^{(j)}$, $\sum_{j=R+1}^{2R} \vec{X}^{(j)}, \dots, \sum_{j=(N-1)R+1}^{NR} \vec{X}^{(j)}$ независимы и одинаково распределены. В силу многомерной центральной предельной теоремы $R^{-\frac{1}{2}} \sum_{j=1}^{R} \vec{X}^{(j)} \xrightarrow{d} \mathcal{N}(0, C(p))$. Следовательно,

$$R^{-\frac{1}{2}} \cdot \left(\sum_{j=1}^{R} \vec{X}^{(j)}, \sum_{j=R+1}^{2R} \vec{X}^{(j)}, \dots, \sum_{j=(N-1)R+1}^{NR} \vec{X}^{(j)} \right) \xrightarrow{d} (\vec{Y}^{(1)}, \vec{Y}^{(2)}, \dots, \vec{Y}^{(N)}), \qquad (13)$$

где под вектором $(\vec{Y}^{(1)}, \vec{Y}^{(2)}, \dots, \vec{Y}^{(N)})$ подразумевается $N(K+2)$-мерный вектор, первые $K+2$ координат которого совпадают с соответствующими координатами вектора $\vec{Y}^{(1)}$, вторые $K+2$ координат — с координатами вектора $\vec{Y}^{(2)}$ и т. д. Рассмотрим следующее преобразование вектора $(\vec{Y}^{(1)}, \vec{Y}^{(2)}, \dots, \vec{Y}^{(N)})$:

$$f((\vec{Y}^{(1)}, \vec{Y}^{(2)}, \dots, \vec{Y}^{(N)})) = \left(\sum_{i=1}^{N} Y_{K+2}^{(i)}, Y_{K+2}^{(1)}, Y_{K+2}^{(2)}, \dots, Y_{K+2}^{(N)}, \sum_{i=1}^{N} Y_1^{(i)}, \dots, \sum_{i=1}^{N} Y_{K+1}^{(i)} \right).$$

Из (13) следует, что

$$f\left(\left(\frac{\sum_{j=1}^{R} \vec{X}^{(j)}}{\sqrt{R}}, \frac{\sum_{j=R+1}^{2R} \vec{X}^{(j)}}{\sqrt{R}}, \dots, \frac{\sum_{j=(N-1)R+1}^{NR} \vec{X}^{(j)}}{\sqrt{R}} \right) \right) \xrightarrow{d} f((\vec{Y}^{(1)}, \vec{Y}^{(2)}, \dots, \vec{Y}^{(N)})),$$

т. е.

$$\left(\frac{\tilde{S}_n - NLRp}{\sqrt{Lp(1-p)R}}, \frac{M\tilde{\pi}_1 - LRp}{\sqrt{Lp(1-p)R}}, \dots, \frac{M\tilde{\pi}_N - LRp}{\sqrt{Lp(1-p)R}}, \frac{\tilde{\nu}_0 - NRw_0(p)}{\sqrt{Rw_0(p)}}, \dots, \frac{\tilde{\nu}_K - NRw_K(p)}{\sqrt{Rw_K(p)}} \right)$$
$$\xrightarrow{d} \left(\sum_{i=1}^{N} Y_{K+2}^{(i)}, Y_{K+2}^{(1)}, Y_{K+2}^{(2)}, \dots, Y_{K+2}^{(N)}, \sum_{i=1}^{N} Y_1^{(i)}, \dots, \sum_{i=1}^{N} Y_{K+1}^{(i)} \right). \qquad (14)$$

Из (8) и (9) следует, что $RN \leqslant Q \leqslant RN + 3N$, т. е. $|Q - RN| = O(1)$, $n \to \infty$. Аналогично можно показать, что $|M - LR| = O(1)$ и $|n - NLR| = O(1)$ при $n \to \infty$. Значит, из (10)–(12), (14) и леммы Слуцкого (см. [19], глава 1) следует, что

$$\left(\frac{S_n - np}{\sqrt{Lp(1-p)R}}, \frac{M\pi_1 - Mp}{\sqrt{Lp(1-p)R}}, \dots, \frac{M\pi_N - Mp}{\sqrt{Lp(1-p)R}}, \frac{\nu_0 - Qw_0(p)}{\sqrt{Rw_0(p)}}, \dots, \frac{\nu_K - Qw_K(p)}{\sqrt{Rw_K(p)}} \right)$$
$$\xrightarrow{d} \left(\sum_{i=1}^{N} Y_{K+2}^{(i)}, Y_{K+2}^{(1)}, Y_{K+2}^{(2)}, \dots, Y_{K+2}^{(N)}, \sum_{i=1}^{N} Y_1^{(i)}, \dots, \sum_{i=1}^{N} Y_{K+1}^{(i)} \right).$$

Таким образом,

$$\left(\frac{S_n - np}{\sqrt{Lp(1-p)R}}, \frac{M^2\sum_{i=1}^N(\pi_i - p)^2}{Lp(1-p)R}, \sum_{j=0}^K \frac{(\nu_j - Qw_j(p))^2}{Rw_j(p)}\right)$$
$$\xrightarrow{d} \left(\sum_{i=1}^N Y_{K+2}^{(i)}, \sum_{i=1}^N (Y_{K+2}^{(i)})^2, \left(\sum_{i=1}^N Y_1^{(i)}\right)^2 + \left(\sum_{i=1}^N Y_2^{(i)}\right)^2 + \cdots + \left(\sum_{i=1}^N Y_{K+1}^{(i)}\right)^2\right).$$

Значит,

$$\left(\sqrt{\frac{n}{LR}}T_1^{(p)}, \frac{M}{LR}T_2^{(p)}, \frac{Q}{R}T_3^{(p)}\right) \xrightarrow{d} \left(\sum_{i=1}^N Y_{K+2}^{(i)}, \sum_{i=1}^N (Y_{K+2}^{(i)})^2, \sum_{j=1}^{K+1}\left(\sum_{i=1}^N Y_j^{(i)}\right)^2\right),$$

откуда следует (4).

Далее, в силу (2) диагональные элементы ковариационной матрицы предельного распределения вектора $(\sqrt{N}T_1^{(p)}, T_2^{(p)}, NT_3^{(p)})$ имеют вид

$$F_{11} = N, \quad F_{22} = 2N, \quad F_{33} = 2N^2K.$$

Лемма 3. *Пусть (ξ_1, ξ_2) — двумерный случайный вектор, имеющий нормальное распределение с нулевым средним. Тогда* $\operatorname{cov}(\xi_1, \xi_2^2) = 0$ *и* $\operatorname{cov}(\xi_1^2, \xi_2^2) = 2\operatorname{cov}^2(\xi_1, \xi_2)$.

Применяя лемму 3 к векторам $Y_{K+2}^{(s)}$ и $\sum_{i=1}^N Y_j^{(i)}$ при $1 \leqslant j \leqslant K+1$, $1 \leqslant s \leqslant N$, получаем, что

$$F_{13} = \operatorname{cov}\left(\sum_{s=1}^N Y_{K+2}^{(s)}, \sum_{j=1}^{K+1}\left(\sum_{i=1}^N Y_j^{(i)}\right)^2\right) = \sum_{s=1}^N\sum_{j=1}^{K+1}\operatorname{cov}\left(Y_{K+2}^{(s)}, \left(\sum_{i=1}^N Y_j^{(i)}\right)^2\right) = 0.$$

Аналогичным образом получаем, что $F_{12} = 0$. Далее,

$$F_{23} = \operatorname{cov}\left(\sum_{s=1}^N \left(Y_{K+2}^{(s)}\right)^2, \sum_{j=1}^{K+1}\left(\sum_{i=1}^N Y_j^{(i)}\right)^2\right) = \sum_{s=1}^N\sum_{j=1}^{K+1}\operatorname{cov}\left(\left(Y_{K+2}^{(s)}\right)^2, \left(\sum_{i=1}^N Y_j^{(i)}\right)^2\right)$$
$$= 2\sum_{s=1}^N\sum_{j=1}^{K+1}\operatorname{cov}^2\left(Y_{K+2}^{(s)}, \sum_{i=1}^N Y_j^{(i)}\right) = 2\sum_{s=1}^N\sum_{j=1}^{K+1}\operatorname{cov}^2\left(Y_{K+2}^{(s)}, Y_j^{(s)}\right) = 2N\sum_{j=1}^{K+1} C_{j,K+2}^2(p).$$

Тем самым теорема 1 доказана.

3.2. Доказательство следствия 1. Так как $\left(T_1^{(\frac{1}{2})}, T_2^{(\frac{1}{2})}, T_3^{(\frac{1}{2})}\right) = (T_1, T_2, T_3)$, то следствие 1 — частный случай теоремы 1.

3.3. Доказательство леммы 1. Следующее утверждение доказано в [18].

Лемма 4. *Пусть последовательность двумерных случайных векторов (α_n, β_n) сходится по распределению к вектору (α, β) при $n \to \infty$. Если* $\sup_{n \geqslant n_0} \mathbf{E}\alpha_n^4 < \infty$ *и* $\sup_{n \geqslant n_0} \mathbf{E}\beta_n^4 < \infty$, *то существуют ковариационные матрицы $\mathbf{D}(\alpha, \beta)$ и $\mathbf{D}(\alpha_n, \beta_n)$, $n \geqslant n_0$, и имеет место поэлементная сходимость этих матриц: $\mathbf{D}(\alpha_n, \beta_n) \to \mathbf{D}(\alpha, \beta)$ при $n \to \infty$.*

Статистика $T_3^{(p)}$ является частным случаем статистики критерия хи-квадрат, поэтому из (2.9) [20] следует, что при $n \to \infty$ выполнены соотношения $\mathbf{E}T_3^{(p)} = O(1)$ и $\mathbf{E}\big(T_3^{(p)} - \mathbf{E}T_3^{(p)}\big)^k = O(1)$ при $2 \leqslant k \leqslant 4$. Следовательно,

$$\sup_{n \geqslant n_0} \mathbf{E}\big(T_3^{(p)}\big)^4 < \infty, \tag{15}$$

где $n_0 = \max(N, L)$. Отметим, что все три статистики $T_1^{(p)}, T_2^{(p)}, T_3^{(p)}$ определены при $n \geqslant n_0$.

Из соотношений (2.3) и (2.5) в [20] следует, что $\mathbf{E}(S_n - np)^4 = O(n^2)$ и $\mathbf{E}(S_n - np)^8 = O(n^4)$ при $n \to \infty$. Значит,

$$\sup_{n \geqslant n_0} \mathbf{E}\big(T_1^{(p)}\big)^4 < \infty \tag{16}$$

и $\sup_{n \geqslant n_0} \mathbf{E}\big(\frac{S_n - np}{\sqrt{n}}\big)^8 < \infty$ Так как $\sqrt{M}(\pi_1 - p) = \frac{S_M - Mp}{\sqrt{M}}$, то аналогично получаем, что $\mathbf{E}M^4(\pi_1 - p)^8 = O(1), n \to \infty$. Положим $\gamma_i(n) = M(\pi_i - p)^2$, $1 \leqslant i \leqslant N$. Случайные величины $\gamma_i(n)$, $1 \leqslant i \leqslant N$, независимы, одинаково распределены (при фиксированном n) и $\sup_{n \geqslant n_0} \mathbf{E}\gamma_i^4(n) = C < \infty$, $1 \leqslant i \leqslant N$. В силу неравенства Минковского

$$\left(\mathbf{E}\Big|\sum_{i=1}^N \gamma_i(n)\Big|^4\right)^{\frac{1}{4}} \leqslant \sum_{i=1}^N \left(\mathbf{E}\Big|\gamma_i(n)\Big|^4\right)^{\frac{1}{4}} \leqslant NC^{\frac{1}{4}},$$

поэтому $\sup_{n \geqslant n_0} \mathbf{E}\big(\sum_{i=1}^N \gamma_i(n)\big)^4 < \infty$. Таким образом, $\sup_{n \geqslant n_0} \mathbf{E}\big(T_2^{(p)}\big)^4 < \infty$, откуда, учитывая (15), (16), теорему 1 и лемму 4, получаем утверждение леммы 1.

3.4. Доказательство утверждения 1. П. 2 следует из определения матрицы $C(p)$. Из теоремы 1 и леммы 1 сразу следуют пп. 1, 3 и 4. Кроме того, из теоремы 1 следует п. 7, так как если $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$, то у каждого из фигурирующих в теореме 1 векторов $\vec{Y}^{(i)}$ последняя координата не зависит от первых $K+1$ координат.

П. 5 был доказан в [15] при условии, что n делится на N, однако доказательство проводится аналогично и без этого условия. Покажем это.

Из теоремы 1 следует, что

$$\big(\sqrt{N}T_1^{(p)}, T_2^{(p)}\big) \xrightarrow{d} \left(\sum_{i=1}^N Y_{K+2}^{(i)}, \sum_{i=1}^N \big(Y_{K+2}^{(i)}\big)^2\right), \quad n \to \infty, \tag{17}$$

где независимые случайные векторы $\vec{Y}^{(i)} = (Y_1^{(i)}, Y_2^{(i)}, \ldots, Y_{K+2}^{(i)})$, $1 \leqslant i \leqslant N$, имеют распределение $\mathcal{N}(0, C(p))$. Из доказательства леммы 3 работы [15] следует, что если $\gamma_1, \gamma_2, \ldots$ — независимые одинаково распределенные случайные величины с нормальным распределением и если $\sum_{i=1}^N \gamma_i$ и $\sum_{i=1}^N \gamma_i^2$ независимы, то $\mathbf{D}\gamma_1 = 0$. Отсюда и из (17) следует, что если бы статистики $\sqrt{N}T_1^{(p)}$ и $T_2^{(p)}$ были асимптотически независимыми, то векторы $\sum_{i=1}^N Y_{K+2}^{(i)}$ и $\sum_{i=1}^N \big(Y_{K+2}^{(i)}\big)^2$ были бы независимыми и выполнялось бы условие $\mathbf{D}Y_{K+2}^{(1)} = 0$. Однако $\mathbf{D}Y_{K+2}^{(i)} = C_{K+2,K+2}(p) = 1$. Тем самым п. 5 доказан.

Если $T_2^{(p)}$ асимптотически независима с $T_3^{(p)}$, то $F_{23} = 0$ и, как следствие, $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$.

Предположим теперь, что $T_1^{(p)}$ асимптотически независима с $T_3^{(p)}$. Нам осталось доказать, что в этом случае $\sum_{j=1}^{K+1} C_{j,K+2}^2(p) = 0$.

Из теоремы 1 следует, что

$$\bigl(\sqrt{N}T_1^{(p)}, NT_3^{(p)}\bigr) \xrightarrow{d} \biggl(\sum_{s=1}^{N} Y_{K+2}^{(s)}, \sum_{j=1}^{K+1}\biggl(\sum_{i=1}^{N} Y_j^{(i)}\biggr)^2\biggr). \tag{18}$$

Из (18) и асимптотической независимости $T_1^{(p)}$ и $T_3^{(p)}$ следует, что случайные величины $\sum_{s=1}^{N} Y_{K+2}^{(s)}$ и $\sum_{j=1}^{K+1}\bigl(\sum_{i=1}^{N} Y_j^{(i)}\bigr)^2$ независимы. Значит,

$$0 = \operatorname{cov}\biggl(\biggl(\sum_{s=1}^{N} Y_{K+2}^{(s)}\biggr)^2,\ \sum_{j=1}^{K+1}\biggl(\sum_{i=1}^{N} Y_j^{(i)}\biggr)^2\biggr) = \sum_{j=1}^{K+1} A_j,$$

где $A_j = \operatorname{cov}\bigl(\bigl(\sum_{s=1}^{N} Y_{K+2}^{(s)}\bigr)^2, \bigl(\sum_{i=1}^{N} Y_j^{(i)}\bigr)^2\bigr)$. Учитывая лемму 3, получаем, что

$$\begin{aligned} A_j = 2\operatorname{cov}^2\biggl(\sum_{s=1}^{N} Y_{K+2}^{(s)},\ \sum_{i=1}^{N} Y_j^{(i)}\biggr) &= 2\biggl(\sum_{s,i=1}^{N} \operatorname{cov}\bigl(Y_{K+2}^{(s)},\ Y_j^{(i)}\bigr)\biggr)^2 \\ &= 2\biggl(\sum_{s=1}^{N} \operatorname{cov}\bigl(Y_{K+2}^{(s)},\ Y_j^{(s)}\bigr)\biggr)^2 = 2(NC_{j,K+2}(p))^2. \end{aligned}$$

Следовательно,

$$0 = \operatorname{cov}\biggl(\biggl(\sum_{s=1}^{N} Y_{K+2}^{(s)}\biggr)^2,\ \sum_{j=1}^{K+1}\biggl(\sum_{i=1}^{N} Y_j^{(i)}\biggr)^2\biggr) = \sum_{j=1}^{K+1} A_j = \sum_{j=1}^{K+1} 2N^2 C_{j,K+2}^2(p).$$

Утверждение 1 доказано.

3.5. Доказательство леммы 2. Первое утверждение леммы 2 установлено в лемме 4 работы [15] при дополнительном предположении о том, что n делится на M, однако доказательство проводится аналогично и без этого предположения. Согласно усиленному закону больших чисел $\frac{\nu_k}{Q} \xrightarrow{\text{п.н.}} \mathbf{E}I_{\zeta_1\in\alpha_k} = w_k(p)$ при $0 \leqslant k \leqslant K+1$, поэтому

$$\frac{T_3}{Q} = \sum_{k=0}^{K} \frac{1}{w_k(\frac12)}\biggl(\frac{\nu_k - Qw_k(\frac12)}{Q}\biggr)^2 \xrightarrow{\text{п.н.}} \sum_{k=0}^{K} \frac{1}{w_k(\frac12)}\biggl(w_k(p) - w_k\Bigl(\frac12\Bigr)\biggr)^2 > 0,$$

откуда следует, что $T_3 \xrightarrow{\text{п.н.}} +\infty$. Лемма 2 доказана.

4. Дополнение

Лемма 5. *Пусть верна гипотеза* $H_{\frac{1}{2}}$ *и* $\min(V_1, V_2) = k \geqslant 2$. *Тогда* $\mathbf{P}_{\frac{1}{2}}(\zeta_1 \geqslant \min(V_1, V_2) - 2) \geqslant 0.9947 - 2^{-k}$. *Если, к тому же,* $V_1 = V_2$, *то*

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2)) \in [0.288, 0.375], \qquad \mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2) - 1) \in [0.562, 0.579],$$
$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2) - 2) \in [0.062, 0.129], \quad \mathbf{P}_{\frac{1}{2}}(\zeta_1 < \min(V_1, V_2) - 2) \leqslant 0.007.$$

Доказательство. Заметим, что если к любой матрице A размера $k \times k$ (над полем $GF(2)$) дописать $V_2 - k$ столбцов справа, а затем к полученной матрице дописать $V_1 - k$ строк снизу, то ранг матрицы A не уменьшится. Поэтому далее будем без ограничения общности предполагать, что $k = V_1 = V_2$.

Имеет место следующее неравенство:

$$1 - 2 \cdot 2^{-s} \leqslant \prod_{j=s}^{k}(1 - 2^{-i}) \leqslant 1 - 2^{-s}. \tag{19}$$

Правое неравенство в (19) очевидно, а левое выполнено в силу леммы 2.3 [21]. Положим

$$\gamma = \prod_{i=1}^{\infty}(1 - 2^{-i}) = 0.288788\ldots .$$

В силу (1) и (19) при некотором $\delta \in [0, 1]$ имеем:

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2)) = \prod_{i=0}^{k-1}(1 - 2^{i-k}) = \frac{\prod_{j=1}^{\infty}(1 - 2^{-j})}{\prod_{j=k+1}^{\infty}(1 - 2^{-j})}$$
$$= \frac{\gamma}{1 - (1 + \delta)2^{-k-1}} \in \left[\frac{\gamma}{1 - 2^{-k-1}}, \frac{\gamma}{1 - 2^{-k}}\right] \subset \left[\gamma, \gamma(1 + 2 \cdot 2^{-k})\right]. \tag{20}$$

Аналогично получаем, что

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2) - 1) = \frac{2(1 - 2^{-k})\prod_{j=1}^{\infty}(1 - 2^{-j})}{\prod_{j=k+1}^{\infty}(1 - 2^{-j})} =$$
$$= \frac{2\gamma(1 - 2^{-k})}{1 - (1 + \delta)2^{-k-1}} \in \left[\frac{2\gamma(1 - 2^{-k})}{1 - 2^{-k-1}}, 2\gamma\right] \subset \left[2\gamma(1 - 2^{-k}), 2\gamma\right], \tag{21}$$

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = \min(V_1, V_2) - 2) \in \left[\frac{4\gamma(1 - 2^{-k+1})(1 - 2^{-k})}{9(1 - 2^{-k-1})}, \frac{4\gamma}{9}\right] \subset \left[\frac{4\gamma(1 - 3 \cdot 2^{-k})}{9}, \frac{4\gamma}{9}\right]. \tag{22}$$

Следовательно,

$$\mathbf{P}_{\frac{1}{2}}(\zeta_1 \geqslant \min(V_1, V_2) - 2) \geqslant \gamma + 2\gamma(1 - 2^{-k}) + \frac{4\gamma(1 - 3 \cdot 2^{-k})}{9} = \left(\frac{31}{9} - \frac{10}{3}2^{-k}\right)\gamma,$$

откуда следует первое утверждение леммы.

Следующая таблица получается с помощью формулы (1) и содержит значения вероятностей, округленные до 3-го знака после запятой.

k	2	3	4	5	6	7	8	9
$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = k)$	0.375	0.328	0.308	0.298	0.293	0.291	0.290	0.289
$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = k-1)$	0.563	0.574	0.577	0.577	0.578	0.578	0.578	0.578
$\mathbf{P}_{\frac{1}{2}}(\zeta_1 = k-2)$	0.063	0.096	0.112	0.120	0.124	0.126	0.127	0.128
$\mathbf{P}_{\frac{1}{2}}(\zeta_1 \leqslant k-3)$	0.000	0.002	0.003	0.04	0.005	0.005	0.005	0.005

Таким образом, при $2 \leqslant k \leqslant 9$ утверждение леммы доказано, а при $k \geqslant 10$ оно следует из неравенств (20)–(22). □

Автор выражает благодарность А. М. Зубкову за постоянное внимание к работе.

Список литературы

1. Rukhin A., Soto J., Nechvatal J., Smid M., Barker E., Leigh S., Levenson M., Vangel M., Banks D., Heckert A., Dray J., Vo S., "A statistical test suite for random and pseudorandom number generators for cryptographic applications", *NIST Special Publication* 800-22 *Revision* 1*a*, *ed. L. E. Bassham III*, *NIST*, April 2010.
2. Серов А. А., "Формулы для чисел последовательностей, содержащих заданный шаблон заданное число раз", *Дискретная математика*, **32**:4 (2020), 120–136.
3. Zubkov A. M., Serov A. A., "A natural approach to the experimental study of dependence between statistical tests", *Матем. вопр. криптогр.*, **12**:1 (2021), 131–142.
4. Zubkov A. M., Serov A. A., "Testing the NIST Statistical Test Suite on artificial pseudorandom sequences", *Матем. вопр. криптогр.*, **10**:2 (2019), 89–96.
5. Zaman J. K. M. S. , Ghosh R., "Review on fifteen statistical tests proposed by NIST", *J. Theor. Phys. Cryptography*, **1** (2012), 18–31.
6. Sulak F., Doğanaksoy A., Uğuz M., Koçak O., "Periodic template tests: A family of statistical randomness tests for a collection of binary sequences", *Discrete Applied Mathematics*, **271** (2019), 191–204.
7. Soto J., Bassham L., "Randomness testing of the Advanced Encryption Standard finalist candidates", *NIST IR* 6483, *National Institute of Standards and Technology*, 2000.
8. Sulak F., Uğuz M., Koçak O., Doğanaksoy A., "On the independence of statistical randomness tests included in the NIST test suite", *Turkish J. Electr. Eng. & Comput. Sci., T.*, **25**:5 (2017), 3673–3683.
9. Georgescu C., Simion E., "New results concerning the power of NIST randomness tests", *Proc. Romanian acad., ser. A.*, **18** (2017), 381–388.
10. Rukhin A. L., "Testing randomness: a suite of statistical procedures", *Teor. Veroyatnost. i Primenen.*, **45**:1 (2000), 137—162.
11. Рябко Б. Я., Пестунов А. И., "«Стопка книг» как новый статический тест для случайных чисел", *Пробл. передачи информ.*, **40**:1 (2004), 73–78.
12. Мальцев М. В., Харин Ю. С., "О тестировании выходных последовательностей криптографических генераторов на основе цепей Маркова условного порядка", *Информатика*, **4** (2013), 104–111.
13. Burciu P., Simion E., "A systematic approach of NIST statistical tests dependencies", *J. Electr. Eng., Electronics, Control and Comput. Sci.*, **5**:15 (2019), 1–6.
14. Marsaglia G., Tsay L. H., "Matrices and the structure of random number sequences", *Linear Algebra Appl.*, **67** (1985), 147–156.
15. Савелов М. П., "Предельные совместные распределения статистик четырех критериев пакета NIST", *Дискретная математика*, **33**:2 (2021), 141–154.

16. Савелов М. П., “Предельные совместные распределения статистик трех критериев пакета NIST”, *Дискретная математика*, **33**:3 (2021), 92–106.
17. Савелов М. П., “Предельная теорема для сглаженного варианта спектрального критерия равновероятности двоичной последовательности”, *Дискретная математика*, **33**:4 (2021), 132–140.
18. Савелов М. П., “Предельное совместное распределение критериев «Monobit test», «Frequency Test within a Block» и «Test for the Longest Run of Ones in a Block»”, *Дискретная математика*, **34**:3 (2022), 70–84.
19. Serfling R., *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, 1980, 398 pp.
20. Туманян С. Х., “Асимптотическое распределение критерия χ^2 при одновременном возрастании объема наблюдений и числа групп”, *Теория вероятн. и ее примен.*, **1**:1 (1956), 131—145.
21. Fulman J., Goldstein L., “Stein’s method and the rank distribution of random matrices over finite fields”, *Ann. Probab.*, **43**:3 (2015), 1274–1314.
22. Колчин В. Ф., *Случайные графы*, 2-е изд., ФИЗМАТЛИТ, М., 2004, 256 с.

Статья поступила 14.06.2022.