Math-Net.Ru

Общероссийский математический портал

В. И. Круглов, Пуассоновская аппроксимация для распределения числа “параллелограммов” в случайной выборке из $\mathbb{Z}_N^q$, *Матем. вопр. криптогр.*, 2012, том 3, выпуск 2, 63–78

DOI: <https://doi.org/10.4213/mvk54>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.197.217.227

26 апреля 2023 г., 10:46:34

**Пуассоновская аппроксимация
для распределения числа
«параллелограммов»
в случайной выборке из $\mathbb{Z}_N^q$**

В. И. Круглов

Математический институт им. В. А. Стеклова РАН, Москва

Получено 20.V.2011

Для случайной выборки с возвращением $\xi_1, \dots, \xi_T$ из группы $\mathbb{Z}_N^q$, $N \geq 4$, изучается распределение числа ζ четырехэлементных подмножеств, удовлетворяющих соотношению вида $\xi_{i_1} - \xi_{i_2} = \xi_{i_3} - \xi_{i_4}$ и дополнительному условию, сформулированному в терминах вводимой на группе метрики. Получена оценка точности пуассоновской аппроксимации распределения случайной величины ζ и указаны условия, при которых распределение ζ сходится к распределению Пуассона.

Ключевые слова: случайные элементы группы, совпадение разностей, пуассоновская аппроксимация

**Poisson approximation for the distribution of the number
of “parallelograms” in a random sample from $\mathbb{Z}_N^q$**

V. I. Kruglov

Steklov Mathematical Institute of RAS, Moscow

Abstract. For a random sample with replacement $\xi_1, \dots, \xi_T$ from a group $\mathbb{Z}_N^q$, $N \geq 4$, we consider the distribution of the number ζ of 4-element subsets satisfying the relation of type $\xi_{i_1} - \xi_{i_2} = \xi_{i_3} - \xi_{i_4}$ and additional condition given in terms of a metric on this group. Estimates of the accuracy of Poisson approximation for the distribution of ζ are obtained and conditions of the weak convergence of ζ to the Poisson law are established.

Key words: random elements of a group, coincidence of differences, Poisson approximation

Citation: *Mathematical Aspects of Cryptography*, 2012, vol. 3, no. 2, pp. 63–78 (Russian).

1. Введение

Дж. Марсалья (см. [3, 4]) предложил использовать для проверки качества датчиков случайных чисел статистику

$$\mu(N, T) = \sum_{1 \leq i < j \leq T-1} I \{X_{(i+1)} - X_{(i)} = X_{(j+1)} - X_{(j)}\},$$

где $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(T)}$ — вариационный ряд, построенный по числам $X_1, \dots, X_T$. Ранее было показано (см. [1, 2]), что если $X_1, \dots, X_T$ — независимые случайные величины, имеющие равномерное распределение на множестве $\{0, \dots, N-1\}$, то при $\lim_{T, N \rightarrow \infty} T^3/(4N) = \lambda$ распределение случайной величины $\mu(N, T)$ сходится к распределению Пуассона с параметром λ .

Случайную величину $\mu(N, T)$ можно интерпретировать как количество пар равных отрезков $\nu_i = X_{(i+1)} - X_{(i)}$. В настоящей работе рассматривается аналогичная задача, а именно, исследуется распределение суммы индикаторов выполнения линейных соотношений вида $X_{i_1} - X_{i_2} = X_{i_3} - X_{i_4}$ и дополнительного условия (которое будет сформулировано ниже) между случайными элементами группы $\mathbb{Z}_N^q$. На основе полученных результатов предложен статистический критерий проверки гипотезы о независимости элементов случайной выборки и равновероятности их распределения.

Отметим, что линейные соотношения вида $X_{i_1} - X_{i_2} = X_{i_3} - X_{i_4}$ представляют интерес с точки зрения криптографии. Например, разностный криптоанализ основан на том, что условная вероятность выполнения соотношения $B_{i_1} - B_{i_2} = B_{i_3} - B_{i_4}$ для шифртекстов $B_{i_1}, \dots, B_{i_4}$ при условии, что для соответствующих открытых текстов $A_{i_1}, \dots, A_{i_4}$ разности $A_{i_1} - A_{i_2}$ и $A_{i_3} - A_{i_4}$ совпадают, может быть больше, чем безусловная вероятность этого события. Тогда, если векторы X_i образованы путем объединения векторов A_i и B_i , $X_i = \begin{pmatrix} A_i \\ B_i \end{pmatrix}$, вероятность выполнения соотношения $X_{i_1} - X_{i_2} = X_{i_3} - X_{i_4}$ больше, чем при равновероятном распределении векторов X_i , что и используется в разностном криптоанализе. Другим примером является так называемая «атака методом бумеранга» ([6]), основанная на использовании таких наборов пар (P, C) , (P', C') , (Q, D) , (Q', D') , каждая из которых состоит из блока открытого текста и соответствующего ему блока шифрованного текста, что $P - P' = C - C' = D - D' = Q - Q'$.

Результаты и методы данной работы могут быть полезны при оценке минимального объема выборки, требуемого для наличия в ней пар элементов с одинаковыми разностями.

2. Постановка задачи

Будем рассматривать группу $\mathbb{Z}_N^q$, где $N \geq 4$, как q -мерный дискретный тор $\{0, \dots, N-1\}^q$. Введем на $\mathbb{Z}_N^q$ метрику: для

$$\begin{aligned} x &= (x_1, \dots, x_q) \in \mathbb{Z}_N^q, \\ y &= (y_1, \dots, y_q) \in \mathbb{Z}_N^q \end{aligned}$$

положим

$$\rho(x, y) = \max_{i=1, \dots, q} (|x_i - y_i|_N),$$

где

$$|x_i - y_i|_N = \min(|x_i - y_i|, N - |x_i - y_i|), \quad i = 1, \dots, q.$$

Пусть в данном торе равномерно и независимо выбраны случайные точки $\xi_1, \dots, \xi_T$, $T \geq 4$. Зададим величину $\rho_{\max}$, $1 \leq \rho_{\max} < \frac{N-1}{2}$, и для каждой упорядоченной пары чисел $1 \leq i, j \leq T$ построим вектор

$$\Delta_{i,j} = (\xi_j - \xi_i)I\{\rho(\xi_i, \xi_j) \leq \rho_{\max}\}.$$

Будем говорить, что точки $\xi_h, \xi_f, \xi_l, \xi_m$, где $1 \leq h < f < l < m \leq T$, составляют параллелограмм, если имеет место событие

$$B(h, f, l, m) = \bigcup_{\sigma} \{\Delta_{\sigma(h), \sigma(f)} = \Delta_{\sigma(l), \sigma(m)} \neq 0\},$$

т.е. хотя бы для одной перестановки $\sigma: \{h, f, l, m\} \rightarrow \{h, f, l, m\}$ векторы $\Delta_{\sigma(h), \sigma(f)}$ и $\Delta_{\sigma(l), \sigma(m)}$ являются ненулевыми и равными.

Введем случайную величину

$$\zeta = \zeta(N, q, T, \rho_{\max}) = \sum_{1 \leq h < f < l < m \leq T} I_{B(h, f, l, m)},$$

равную количеству таких параллелограммов, образованных элементами выборки $\xi_1, \dots, \xi_T$. В работе изучаются условия, при которых распределение случайной величины ζ можно аппроксимировать пуассоновским распределением.

3. Технические леммы

3.1. Леммы о распределении индикаторов

Так как случайные величины $\xi_1, \dots, \xi_T$ независимы и имеют одинаковые распределения, то распределение индикатора $I_{B(h,f,l,m)}$ не зависит от конкретного значения набора (h, f, l, m) , $1 \leq h < f < l < m \leq T$. Будем обозначать

$$p = \mathbf{P} \{I_{B(1,2,3,4)} = 1\}.$$

Лемма 1. При $N, T \geq 4$

$$\begin{aligned} \mathbf{P} \{\Delta_{1,2} = \Delta_{3,4} \neq 0\} &= \frac{(2\rho_{\max} + 1)^q - 1}{N^{2q}}, \\ \mathbf{P} \{\Delta_{1,2} = \Delta_{3,4} \neq 0, \Delta_{1,3} = \Delta_{2,4} \neq 0\} &= \frac{((2\rho_{\max} + 1)^q - 1)^2}{N^{3q}}. \end{aligned}$$

Доказательство. Так как $\xi_1, \dots, \xi_T$ независимы и имеют равномерное распределение на $\mathbb{Z}_N^q$, то

$$\begin{aligned} \mathbf{P} \{\Delta_{1,2} = \Delta_{3,4} \neq 0\} &= \mathbf{P} \{\xi_2 - \xi_1 = \xi_4 - \xi_3, 0 < \rho(\xi_1, \xi_2) \leq \rho_{\max}\} = \\ &= \mathbf{P} \{\xi_4 = \xi_3 + \xi_2 - \xi_1\} \mathbf{P} \{0 < \rho(\xi_1, \xi_2) \leq \rho_{\max}\} = \frac{1}{N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q}. \end{aligned}$$

Аналогично,

$$\begin{aligned} \mathbf{P} \{\Delta_{1,2} = \Delta_{3,4} \neq 0, \Delta_{1,3} = \Delta_{2,4} \neq 0\} &= \\ &= \mathbf{P} \{\xi_2 - \xi_1 = \xi_4 - \xi_3, 0 < \rho(\xi_1, \xi_2) \leq \rho_{\max}, 0 < \rho(\xi_1, \xi_3) \leq \rho_{\max}\} = \\ &= \frac{1}{N^q} \left(\frac{(2\rho_{\max} + 1)^q - 1}{N^q} \right)^2. \quad \blacksquare \end{aligned}$$

Лемма 2. При $N, T \geq 4$ справедливо равенство

$$p = \frac{3}{N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right).$$

Доказательство. а) Событие $B = B(1, 2, 3, 4) = \bigcup_{\sigma \in S_4} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\}$ равносильно тому, что произошло хотя бы одно из событий B^1, B^2, B^3 , схематично иллюстрируемых следующим образом:

В формальной записи события B^1 , B^2 , B^3 строятся путем разбиения множества перестановок $\sigma \in S_4$ в события B на три непересекающихся множества перестановок, в соответствии с которыми

$$B^1 = B^1(1, 2, 3, 4) = \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 2, 3, 4), (2, 1, 4, 3), (1, 3, 2, 4), (3, 1, 4, 2), \\ (3, 4, 1, 2), (4, 3, 2, 1), (2, 4, 1, 3), (4, 2, 3, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\},$$

$$B^2 = B^2(1, 2, 3, 4) = \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 2, 4, 3), (2, 1, 3, 4), (1, 4, 2, 3), (4, 1, 3, 2), \\ (4, 3, 1, 2), (3, 4, 2, 1), (2, 3, 1, 4), (3, 2, 4, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\},$$

$$B^3 = B^3(1, 2, 3, 4) = \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 4, 3, 2), (3, 1, 2, 4), (1, 3, 4, 2), (4, 1, 2, 3), \\ (3, 2, 1, 4), (2, 4, 3, 1), (4, 2, 1, 3), (2, 3, 4, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\}.$$

Таким образом, $B = \bigcup_{j=1, \dots, 3} B^j$ и

$$p = \mathbf{P}(B) = \mathbf{P}(B^1) + \mathbf{P}(B^2) + \mathbf{P}(B^3) - \\ - \mathbf{P}(B^1 B^2) - \mathbf{P}(B^1 B^3) - \mathbf{P}(B^2 B^3) + \mathbf{P}(B^1 B^2 B^3),$$

здесь и далее набор индексов $(1, 2, 3, 4)$ опускается для краткости записи.

б) Так как случайные величины $\xi_1, \dots, \xi_4$ имеют одинаковые распределения, то $\mathbf{P}(B^1) = \mathbf{P}(B^2) = \mathbf{P}(B^3)$, и из леммы 1 следует, что

$$\mathbf{P}(B^1) = \mathbf{P}(\{\Delta_{1,2} = \Delta_{3,4} \neq 0\} \cup \{\Delta_{1,3} = \Delta_{2,4} \neq 0\}) = \\ = \mathbf{P}\{\Delta_{1,2} = \Delta_{3,4} \neq 0\} + \mathbf{P}\{\Delta_{1,3} = \Delta_{2,4} \neq 0\} - \\ - \mathbf{P}\{\Delta_{1,2} = \Delta_{3,4} \neq 0, \Delta_{1,3} = \Delta_{2,4} \neq 0\} = \\ = \frac{(2\rho_{\max} + 1)^q - 1}{N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \right).$$

в) Вычислим $\mathbf{P}(B^1 B^2)$. Заметим, что

$$\begin{aligned} B^1 \cap B^2 &\Rightarrow \begin{cases} \xi_2 - \xi_1 = \xi_4 - \xi_3, \\ \xi_2 - \xi_1 = \xi_3 - \xi_4 \end{cases} \Rightarrow \begin{cases} \xi_3 - \xi_4 = \xi_4 - \xi_3, \\ \xi_2 - \xi_1 = \xi_3 - \xi_4 \end{cases} \Rightarrow \\ &\Rightarrow \begin{cases} \xi_3 - \xi_4 = 0, \\ \xi_2 - \xi_1 = 0 \end{cases} \Rightarrow \begin{cases} \xi_1 = \xi_2, \\ \xi_3 = \xi_4. \end{cases} \end{aligned}$$

Поэтому, если события B^1 и B^2 произошли одновременно, $\Delta_{1,2} = \Delta_{2,1} = \Delta_{3,4} = \Delta_{4,3} = 0$ и, с учетом того, что для любых h, f, l, m

$$\Delta_{h,f} = \Delta_{l,m} \neq 0 \Leftrightarrow \Delta_{f,h} = \Delta_{m,l} \neq 0 \Leftrightarrow \Delta_{l,m} = \Delta_{h,f} \neq 0 \Leftrightarrow \Delta_{m,l} = \Delta_{f,h} \neq 0,$$

в этом случае

$$\begin{aligned} B^1 &= \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 2, 3, 4), (2, 1, 4, 3), (1, 3, 2, 4), (3, 1, 4, 2), \\ (3, 4, 1, 2), (4, 3, 2, 1), (2, 4, 1, 3), (4, 2, 3, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\} = \\ &= \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 3, 2, 4), (3, 1, 4, 2), \\ (2, 4, 1, 3), (4, 2, 3, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\} = \{\Delta_{1,3} = \Delta_{2,4} \neq 0\}, \end{aligned}$$

$$\begin{aligned} B^2 &= \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 2, 4, 3), (2, 1, 3, 4), (1, 4, 2, 3), (4, 1, 3, 2), \\ (4, 3, 1, 2), (3, 4, 2, 1), (2, 3, 1, 4), (3, 2, 4, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\} = \\ &= \bigcup_{\substack{(\sigma(1), \sigma(2), \sigma(3), \sigma(4)) = \\ (1, 4, 2, 3), (4, 1, 3, 2), \\ (2, 3, 1, 4), (3, 2, 4, 1)}} \{\Delta_{\sigma(1), \sigma(2)} = \Delta_{\sigma(3), \sigma(4)} \neq 0\} = \{\Delta_{1,4} = \Delta_{2,3} \neq 0\}. \end{aligned}$$

Следовательно,

$$\begin{aligned} B^1 \cap B^2 &\Leftrightarrow \left\{ \begin{array}{l} \xi_1 = \xi_2, \xi_3 = \xi_4, \\ \Delta_{1,3} = \Delta_{2,4} \neq 0, \Delta_{1,4} = \Delta_{2,3} \neq 0 \end{array} \right\} \Leftrightarrow \\ &\Leftrightarrow \{\xi_1 = \xi_2, \xi_3 = \xi_4, \Delta_{1,3} \neq 0\} \Leftrightarrow \{\xi_1 = \xi_2, \xi_3 = \xi_4, 0 < \rho(\xi_1, \xi_3) \leq \rho_{\max}\}. \end{aligned}$$

Аналогично,

$$\begin{aligned} B^1 \cap B^3 &\Leftrightarrow \{\xi_1 = \xi_3, \xi_2 = \xi_4, 0 < \rho(\xi_1, \xi_4) \leq \rho_{\max}\}, \\ B^2 \cap B^3 &\Leftrightarrow \{\xi_1 = \xi_4, \xi_2 = \xi_3, 0 < \rho(\xi_1, \xi_2) \leq \rho_{\max}\}, \end{aligned}$$

и, так как случайные величины $\xi_1, \dots, \xi_4$ независимы и имеют равномерное распределение на $\mathbb{Z}_N^q$,

$$\mathbf{P}(B^1 B^2) = \mathbf{P}(B^1 B^3) = \mathbf{P}(B^2 B^3) = \frac{1}{N^{2q}} \frac{(2\rho_{\max} + 1)^q - 1}{N^q}.$$

г) Покажем, что $\mathbf{P}(B^1 B^2 B^3) = 0$. Так как

$$B^1 \cap B^3 \Rightarrow \begin{cases} \xi_1 = \xi_2, \\ \xi_3 = \xi_4, \end{cases} B^1 \cap B^3 \Rightarrow \begin{cases} \xi_1 = \xi_3, \\ \xi_2 = \xi_4, \end{cases} B^2 \cap B^3 \Rightarrow \begin{cases} \xi_1 = \xi_4, \\ \xi_2 = \xi_3, \end{cases}$$

то

$$B^1 \cap B^2 \cap B^3 \Rightarrow \xi_1 = \xi_2 = \xi_3 = \xi_4 \Rightarrow \Delta_{i,j} = 0 \quad \forall 1 \leq i, j \leq 4,$$

т. е. события B^1, B^2, B^3 в совокупности несовместны.

д) Окончательно,

$$\begin{aligned} p &= 3\mathbf{P}(B^1) - 3\mathbf{P}(B^1 B^2) = \\ &= 3 \frac{(2\rho_{\max} + 1)^q - 1}{N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \right) - 3 \frac{1}{N^{2q}} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} = \\ &= 3 \frac{(2\rho_{\max} + 1)^q - 1}{N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q - 1}{N^q} - \frac{1}{N^q} \right) = \\ &= 3 \frac{(2\rho_{\max} + 1)^q - 1}{N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right). \quad \blacksquare \end{aligned}$$

Следствие 1. При $N, T \geq 4$

$$\mathbf{E}\zeta = C_T^4 p = \frac{T^{[4]}}{4!} p = \frac{T^{[4]}}{8N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right),$$

где $T^{[4]} = T(T-1)(T-2)(T-3)$.

3.2. Леммы о распределениях пар индикаторов

Лемма 3. Пусть $N, T \geq 4$ и

$$v = |\{h', f', l', m'\} \setminus \{h, f, l, m\}| \geq 1.$$

Тогда

$$\mathbf{P}(B(h, f, l, m), B(h', f', l', m')) \leq \frac{6p}{N^q}.$$

Доказательство. Пусть, например, $m' \notin \{h, f, l, m\}$. Тогда в силу независимости случайных величин $\xi_1, \dots, \xi_T$

$$\begin{aligned} \mathbf{P}(B(h, f, l, m), \Delta_{h', f'} = \Delta_{l', m'} \neq 0) &\leq \\ &\leq \mathbf{P}(B(h, f, l, m), \xi_{f'} - \xi_{h'} = \xi_{m'} - \xi_{l'}) = \\ &= \mathbf{P}(B(h, f, l, m), \xi_{m'} = \xi_{l'} + \xi_{f'} - \xi_{h'}) = \frac{1}{N^q} \mathbf{P}(B(h, f, l, m)), \end{aligned}$$

следовательно,

$$\begin{aligned} \mathbf{P}(B(h, f, l, m), B(h', f', l', m')) &= \\ &= \mathbf{P}\left(B(h, f, l, m), \bigcup_{j=1}^3 B^j(h', f', l', m')\right) \leq \\ &\leq \sum_{j=1}^3 \mathbf{P}(B(h, f, l, m), B^j(h', f', l', m')) = \\ &= 3\mathbf{P}(B(h, f, l, m), B^1(h', f', l', m')) = \\ &= 3\mathbf{P}\left(B(h, f, l, m), \{\Delta_{h', f'} = \Delta_{l', m'} \neq 0\} \cup \{\Delta_{h', l'} = \Delta_{f', m'} \neq 0\}\right) \leq \\ &\leq 3\left(\mathbf{P}(B(h, f, l, m), \{\Delta_{h', f'} = \Delta_{l', m'} \neq 0\}) + \right. \\ &\quad \left. + \mathbf{P}(B(h, f, l, m), \{\Delta_{h', l'} = \Delta_{f', m'} \neq 0\})\right) \leq \\ &\leq \frac{6}{N^q} \mathbf{P}(B(h, f, l, m)) = \frac{6p}{N^q}. \quad \blacksquare \end{aligned}$$

Лемма 4.

Пусть $N, T \geq 4$ и

$$v = |\{h', f', l', m'\} \setminus \{h, f, l, m\}| = 3.$$

Тогда события $B(h, f, l, m)$ и $B(h', f', l', m')$ независимы.

Доказательство. Для простоты выкладок будем считать, что

$$\{h, f, l, m\} \cap \{h', f', l', m'\} = \{1\},$$

т.е. только случайная величина ξ_1 встречается одновременно среди случайных величин $\xi_h, \xi_f, \xi_l, \xi_m$ и $\xi_{h'}, \xi_{f'}, \xi_{l'}, \xi_{m'}$. Докажем, что события $B_1 = B(h, f, l, m)$ и $B_2 = B(h', f', l', m')$ независимы.

Зафиксируем произвольное $g_0 \in \mathbb{Z}_N^q$. Структура группы $\mathbb{Z}_N^q$ такова, что для любого $g \in \mathbb{Z}_N^q$

$$\mathbf{P}(B_i|\xi_1 = g) = \mathbf{P}(B_i|\xi_1 = g_0) = \mathbf{P}(B_i), \quad i = 1, 2,$$

следовательно, по формуле полной вероятности

$$\begin{aligned} \mathbf{P}(B_i) &= \sum_{g \in \mathbb{Z}_N^q} \mathbf{P}(B_i|\xi_1 = g)\mathbf{P}(\xi_1 = g) = \\ &= \sum_{g \in \mathbb{Z}_N^q} \mathbf{P}(B_i|\xi_1 = g_0)\mathbf{P}(\xi_1 = g) = \mathbf{P}(B_i|\xi_1 = g_0), \quad i = 1, 2. \end{aligned}$$

Таким образом,

$$\begin{aligned} \mathbf{P}(B_1, B_2) &= \sum_{g \in \mathbb{Z}_N^q} \mathbf{P}(B_1, B_2|\xi_1 = g)\mathbf{P}(\xi_1 = g) = \\ &= \sum_{g \in \mathbb{Z}_N^q} \mathbf{P}(B_1, B_2|\xi_1 = g_0)\mathbf{P}(\xi_1 = g) = \mathbf{P}(B_1, B_2|\xi_1 = g_0) = \\ &= \mathbf{P}(B_1|\xi_1 = g_0)\mathbf{P}(B_2|\xi_1 = g_0) = \mathbf{P}(B_1)\mathbf{P}(B_2). \quad \blacksquare \end{aligned}$$

4. Пуассоновская аппроксимация

Как мы уже показали ранее, для произвольных $1 \leq h < f < l < m \leq T$

$$p = \mathbf{P}\{I_{B(h,f,l,m)} = 1\} = \frac{3}{N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q}\right),$$

и

$$\lambda = \mathbf{E}\zeta = \frac{T^{[4]}}{8N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q}\right).$$

Рассмотрим расстояние полной вариации

$$d_{TV}(\zeta, \pi_\lambda) = \sup_{A \subset \mathbb{Z}_+} |\mathbf{P}\{\zeta \in A\} - \mathbf{P}\{\pi_\lambda \in A\}|$$

между случайной величиной ζ и случайной величиной π_λ , имеющей пуассоновское распределение с параметром $\lambda = \mathbf{E}\zeta$.

Теорема 1. При $N, T \geq 4$ справедлива оценка

$$d_{TV}(\zeta, \pi_\lambda) < d(N, q, T, \rho_{\max}) = \\ = \left(1 - e^{-\lambda}\right) \left(\frac{4}{3}T^3p + 3T^2p + 4Tp + p + \frac{18T^2}{N^q} + \frac{24T}{N^q}\right).$$

Доказательство. Обозначим

$$\Gamma = \{(h, f, l, m) | 1 \leq h < f < l < m \leq T\}, \\ X_{h,f,l,m} = I_{B(h,f,l,m)}$$

и для каждого $(h, f, l, m) \in \Gamma$ представим множество Γ в виде объединения непересекающихся подмножеств $\Gamma_{h,f,l,m}^\omega$ и $\Gamma_{h,f,l,m}^s$: будем считать, что $(h', f', l', m') \in \Gamma_{h,f,l,m}^\omega$, если

$$\{h, f, l, m\} \cap \{h', f', l', m'\} = \emptyset,$$

и $(h', f', l', m') \in \Gamma_{h,f,l,m}^s$ — в противном случае. Положим

$$Z_{h,f,l,m} = \sum_{(h',f',l',m') \in \Gamma_{h,f,l,m}^s} X_{h',f',l',m'}, \quad W_{h,f,l,m} = \sum_{(h',f',l',m') \in \Gamma_{h,f,l,m}^\omega} X_{h',f',l',m'}.$$

Тогда (см. [5, § 2.3])

$$d_{TV}(\zeta, \pi_\lambda) \leq k_1(\lambda) \sum_{(h,f,l,m) \in \Gamma} \mathbf{E} |\mathbf{E} X_{h,f,l,m} - \mathbf{E}(X_{h,f,l,m} | W_{h,f,l,m})| + \\ + k_2(\lambda) \sum_{(h,f,l,m) \in \Gamma} (\mathbf{E} X_{h,f,l,m} \mathbf{E}(X_{h,f,l,m} + Z_{h,f,l,m}) + \mathbf{E} X_{h,f,l,m} Z_{h,f,l,m}),$$

$$\text{где } k_1(\lambda) = \min \left(1, \sqrt{\frac{2}{e\lambda}}\right) \text{ и } k_2(\lambda) = \frac{1 - e^{-\lambda}}{\lambda}.$$

По построению множества $\Gamma_{h,f,l,m}^\omega$ в формулу для случайной величины $W_{h,f,l,m}$ случайные величины $\xi_h, \xi_f, \xi_l, \xi_m$ не входят, следовательно, в силу независимости $\xi_1, \dots, \xi_T$ случайные величины $X_{h,f,l,m}$ и $W_{h,f,l,m}$ независимы, поэтому

$$\mathbf{E} |\mathbf{E} X_{h,f,l,m} - \mathbf{E}(X_{h,f,l,m} | W_{h,f,l,m})| \equiv 0$$

для любого набора $(h, f, l, m) \in \Gamma$. Значит,

$$d_{TV}(\zeta, \pi_\lambda) \leq k_2(\lambda) \sum_{(h,f,l,m) \in \Gamma} (p \mathbf{E}(X_{h,f,l,m} + Z_{h,f,l,m}) + \mathbf{E} X_{h,f,l,m} Z_{h,f,l,m}).$$

Оценим величины $p\mathbf{E}(X_{h,f,l,m} + Z_{h,f,l,m})$ и $\mathbf{E}X_{h,f,l,m}Z_{h,f,l,m}$. Отметим, что для всякого $(h', f', l', m') \in \Gamma_{h,f,l,m}^s$

$$1 \leq |\{h', f', l', m'\} \setminus \{h, f, l, m\}| \leq 3,$$

и разобьем множество $\Gamma_{h,f,l,m}^s$ на непересекающиеся подмножества $\Gamma_{h,f,l,m}^{s,v}$, где $v = 1, 2, 3$, следующим образом:

$$(h', f', l', m') \in \Gamma_{h,f,l,m}^{s,v} \Leftrightarrow |\{h', f', l', m'\} \setminus \{h, f, l, m\}| = v.$$

Легко видеть, что $|\Gamma_{h,f,l,m}^{s,v}| = C_4^{4-v} C_{T-4}^v$, т. е.

$$|\Gamma_{h,f,l,m}^{s,1}| = C_4^3 C_{T-4}^1 < 4T, \quad |\Gamma_{h,f,l,m}^{s,2}| = C_4^2 C_{T-4}^2 < 3T^2,$$

$$|\Gamma_{h,f,l,m}^{s,3}| = C_4^1 C_{T-4}^3 < \frac{2}{3}T^3.$$

Следовательно,

$$|\Gamma_{h,f,l,m}^s| = |\Gamma_{h,f,l,m}^{s,1}| + |\Gamma_{h,f,l,m}^{s,2}| + |\Gamma_{h,f,l,m}^{s,3}| < \frac{2}{3}T^3 + 3T^2 + 4T$$

и

$$\begin{aligned} p\mathbf{E}(X_{h,f,l,m} + Z_{h,f,l,m}) &= p\mathbf{E}\left(X_{h,f,l,m} + \sum_{(h', f', l', m') \in \Gamma_{h,f,l,m}^s} X_{h', f', l', m'}\right) = \\ &= p^2 (|\Gamma_{h,f,l,m}^s| + 1) < p^2 \left(\frac{2}{3}T^3 + 3T^2 + 4T + 1\right). \end{aligned}$$

В силу леммы 3 для любого $(h', f', l', m') \in \Gamma_{h,f,l,m}^{s,1}, \Gamma_{h,f,l,m}^{s,2}$

$$\mathbf{E}X_{h,f,l,m}X_{h', f', l', m'} \leq \frac{6p}{Nq},$$

и в то же время из леммы 4 следует, что для любого $(h', f', l', m') \in \Gamma_{h,f,l,m}^{s,3}$

$$\mathbf{E}X_{h,f,l,m}X_{h', f', l', m'} = \mathbf{E}X_{h,f,l,m}\mathbf{E}X_{h', f', l', m'} = p^2.$$

Поэтому

$$\begin{aligned}
 \mathbf{E}X_{h,f,l,m}Z_{h,f,l,m} &= \sum_{(h',f',l',m') \in \Gamma_{h,f,l,m}^s} \mathbf{E}X_{h,f,l,m}X_{h',f',l',m'} = \\
 &= \sum_{v=1}^3 \sum_{(h',f',l',m') \in \Gamma_{h,f,l,m}^{s,v}} \mathbf{E}X_{h,f,l,m}X_{h',f',l',m'} \leq \\
 &\leq |\Gamma_{h,f,l,m}^{s,1}| \frac{6p}{Nq} + |\Gamma_{h,f,l,m}^{s,2}| \frac{6p}{Nq} + |\Gamma_{h,f,l,m}^{s,3}| p^2 < 4T \frac{6p}{Nq} + 3T^2 \frac{6p}{Nq} + \frac{2}{3} T^3 p^2.
 \end{aligned}$$

Итак, с учетом равенства $\lambda = C_T^4 p$

$$\begin{aligned}
 d_{TV}(\zeta, \pi_\lambda) &\leq k_2(\lambda) \sum_{(h,f,l,m) \in \Gamma} (p \mathbf{E}(X_{h,f,l,m} + Z_{h,f,l,m}) + \mathbf{E}X_{h,f,l,m}Z_{h,f,l,m}) < \\
 &< k_2(\lambda) \sum_{(h,f,l,m) \in \Gamma} \left(p^2 \left(\frac{2}{3} T^3 + 3T^2 + 4T + 1 \right) + \right. \\
 &\quad \left. + \left(4T \frac{6p}{Nq} + 3T^2 \frac{6p}{Nq} + \frac{2}{3} T^3 p^2 \right) \right) = \\
 &= k_2(\lambda) C_T^4 \left(p^2 \left(\frac{2}{3} T^3 + 3T^2 + 4T + 1 \right) + \left(4T \frac{6p}{Nq} + 3T^2 \frac{6p}{Nq} + \frac{2}{3} T^3 p^2 \right) \right) = \\
 &= k_2(\lambda) \lambda \left(\frac{4}{3} T^3 p + 3T^2 p + 4Tp + p + \frac{18T^2}{Nq} + \frac{24T}{Nq} \right) = \\
 &= (1 - e^{-\lambda}) \left(\frac{4}{3} T^3 p + 3T^2 p + 4Tp + p + \frac{18T^2}{Nq} + \frac{24T}{Nq} \right). \quad \blacksquare
 \end{aligned}$$

Теорема 2. Если $T \rightarrow \infty, N \rightarrow \infty, \rho_{\max}^q \rightarrow \infty$ так, что

$$\lambda \rightarrow \lambda_0 = \text{const} \in (0, +\infty),$$

то $d_{TV}(\zeta, \pi_\lambda) \rightarrow 0$.

Доказательство. Как было доказано в теореме 1,

$$\begin{aligned}
 d_{TV}(\zeta, \pi_\lambda) &< d(N, q, T, \rho_{\max}) = \\
 &= (1 - e^{-\lambda}) \left(\frac{4}{3} T^3 p + 3T^2 p + 4Tp + p + \frac{18T^2}{Nq} + \frac{24T}{Nq} \right).
 \end{aligned}$$

Легко видеть, что при $T \rightarrow \infty$ равномерно по N , q и $\rho_{\max}$

$$d(N, q, T, \rho_{\max}) = \left(1 - e^{-\lambda}\right) \left(\frac{4}{3}T^3p + \frac{18T^2}{N^q}\right) (1 + o(1)), \quad (1)$$

и так как $\lambda = \mathbf{E}\zeta = C_T^4 p$, то

$$\frac{4}{3}T^3p = \frac{4T^3}{3C_T^4}\lambda = \frac{32}{T}\lambda (1 + o(1)). \quad (2)$$

Кроме того,

$$\lambda = \frac{T^4}{8} \frac{(2\rho_{\max} + 1)^q}{N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q}\right) (1 + o(1)),$$

откуда, с учетом $\frac{(2\rho_{\max} + 1)^q}{N^q} = \left(\frac{2\rho_{\max} + 1}{N}\right)^q < 1$, мы получаем

$$\begin{aligned} \frac{T^2}{N^q} &= \sqrt{\frac{8\lambda}{(2\rho_{\max} + 1)^q \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q}\right)}} (1 + o(1)) < \\ &< \sqrt{\frac{8\lambda}{(2\rho_{\max} + 1)^q}} (1 + o(1)). \end{aligned}$$

Итак, поскольку $T \rightarrow \infty$ и $\rho_{\max}^q \rightarrow \infty$,

$$\begin{aligned} d_{TV}(\zeta, \pi_\lambda) &< d(N, q, T, \rho_{\max}) = \\ &= \left(1 - e^{-\lambda}\right) \left(\frac{4}{3}T^3p + \frac{18T^2}{N^q}\right) (1 + o(1)) < \\ &< \left(1 - e^{-\lambda}\right) \left(\frac{32\lambda}{T} + 36\sqrt{\frac{2\lambda}{(2\rho_{\max} + 1)^q}}\right) (1 + o(1)) \rightarrow 0. \quad \blacksquare \end{aligned}$$

Теорема 3. Если параметры N и $\rho_{\max}$ фиксированы, $N \geq 4$, и $T \rightarrow \infty, q \rightarrow \infty$ так, что

$$\lambda \rightarrow \lambda_0 = \text{const} \in (0, +\infty),$$

то $d_{TV}(\zeta, \pi_\lambda) \rightarrow 0$.

Доказательство. Так как $\frac{2\rho_{\max} + 1}{N} < 1$, то

$$\begin{aligned}\lambda &= \frac{T^4 (2\rho_{\max} + 1)^q}{8 N^{2q}} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right) (1 + o(1)) = \\ &= \frac{T^4 (2\rho_{\max} + 1)^q}{4 N^{2q}} (1 + o(1)),\end{aligned}$$

поэтому

$$\frac{T^2}{N^q} = \sqrt{\frac{4\lambda}{(2\rho_{\max} + 1)^q}} (1 + o(1)),$$

и из теоремы 1 и соотношений (1), (2) получаем

$$\begin{aligned}d_{TV}(\zeta, \pi_\lambda) &< d(N, q, T, \rho_{\max}) = \\ &= (1 - e^{-\lambda}) \left(\frac{4}{3} T^3 p + \frac{18 T^2}{N^q} \right) (1 + o(1)) = \\ &= (1 - e^{-\lambda}) \left(\frac{32\lambda}{T} + 36 \sqrt{\frac{\lambda}{(2\rho_{\max} + 1)^q}} \right) (1 + o(1)) \rightarrow 0. \quad \blacksquare\end{aligned}$$

5. Момент появления первого «параллелограмма» в выборке

Результат теоремы 1 позволяет оценить вероятность того, что при заданном объеме T выборка $\xi_1, \dots, \xi_T$ содержит не более m параллелограммов:

$$\left| \mathbf{P} \{ \zeta \leq m \} - e^{-\lambda} \sum_{k=0}^m \frac{\lambda^k}{k!} \right| < d(N, q, T, \rho_{\max}),$$

где

$$\lambda = \mathbf{E}\zeta = \frac{T^{[4]} (2\rho_{\max} + 1)^q - 1}{8N^q N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right).$$

В частности, для вероятности того, что выборка $\xi_1, \dots, \xi_T$ не содержит параллелограммов, справедливо неравенство

$$\left| \mathbf{P} \{ \zeta = 0 \} - e^{-\lambda} \right| < d(N, q, T, \rho_{\max}).$$

Сформулируем предельную теорему для распределения момента $\tau_0 = \min\{T : \zeta > 0\}$ первого появления параллелограмма в выборке.

Теорема 4. В условиях теорем 2 и 3 для любого $t \geq 0$

$$\mathbf{P} \left\{ \tau_0 > \sqrt[4]{t \cdot \frac{8N^q}{\alpha^q (2 - \alpha^q)}} \right\} \rightarrow e^{-t},$$

$$\text{где } 0 < \alpha = \frac{2\rho_{\max} + 1}{N} < 1.$$

Доказательство. Отметим, что при выполнении условий теорем 2 и 3 справедливы соотношения $T \rightarrow \infty$, $(2\rho_{\max} + 1)^q \rightarrow \infty$ и $d_{TV}(\zeta, \pi_\lambda) \rightarrow 0$. В силу первых двух соотношений

$$\begin{aligned} e^{-\lambda} &= \exp \left\{ -\frac{T^{[4]} (2\rho_{\max} + 1)^q - 1}{8N^q N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right) \right\} = \\ &= \exp \left\{ -\frac{T^4 (2\rho_{\max} + 1)^q}{8N^q N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right) (1 + o(1)) \right\} = \\ &= \exp \left\{ -\frac{T^4}{8N^q} \alpha^q (2 - \alpha^q) (1 + o(1)) \right\}. \end{aligned}$$

Поскольку $\{\tau_0 > T\} \Leftrightarrow \{\zeta = 0\}$, то, подставляя $T = \sqrt[4]{t \cdot \frac{8N^q}{\alpha^q (2 - \alpha^q)}}$, мы получаем неравенство

$$\begin{aligned} d_{TV}(\zeta, \pi_\lambda) &\geq \left| \mathbf{P} \{\zeta = 0\} - e^{-\lambda} \right| = \left| \mathbf{P} \{\tau_0 > T\} - e^{-\left(-\frac{T^4}{8N^q} \alpha^q (2 - \alpha^q) (1 + o(1))\right)} \right| = \\ &= \left| \mathbf{P} \left\{ \tau_0 > \sqrt[4]{t \cdot \frac{8N^q}{\alpha^q (2 - \alpha^q)}} \right\} - e^{-t(1+o(1))} \right|, \end{aligned}$$

которое с учетом $d_{TV}(\zeta, \pi_\lambda) \rightarrow 0$ доказывает утверждение теоремы 4. ■

6. Статистический критерий

На основе полученных результатов построим критерий согласия с гипотезой H_0 , которая заключается в том, что наблюдаемые случайные величины $\xi_1, \dots, \xi_T, T \geq 4$, независимы и имеют равномерное распределение на группе $\mathbb{Z}_N^q, N \geq 4$. Вычислим значение статистики

$$\zeta = \zeta(\xi_1, \dots, \xi_T) = \sum_{1 \leq h < f < l < m \leq T} I_{B(h,f,l,m)}$$

и величину

$$\lambda = \frac{T^{[4]}}{8N^q} \frac{(2\rho_{\max} + 1)^q - 1}{N^q} \left(2 - \frac{(2\rho_{\max} + 1)^q}{N^q} \right).$$

Обозначим $\alpha_m = \alpha_m(\lambda) = e^{-\lambda} \sum_{k=m+1}^{\infty} \frac{\lambda^k}{k!}$, $m = 0, 1, 2, \dots$, и зададим критический уровень $l = 0, 1, 2, \dots$.

Критерий 1. Гипотеза H_0 принимается, если $\zeta \leq l$, и отвергается, если $\zeta > l$.

Для уровня значимости $\alpha = \mathbf{P}\{\zeta > l | H_0\}$ в силу теоремы 1 справедлива оценка

$$|\alpha - \alpha_l| \leq (1 - e^{-\lambda}) \left(\frac{4}{3} T^3 p + 3T^2 p + 4Tp + p + \frac{18T^2}{N^q} + \frac{24T}{N^q} \right),$$

где, напомним, $p = \lambda/C_T^4$.

Список литературы

1. Клыкова Н. В. Предельное распределение числа совпадающих промежутков // Обозр. прикл. и промысл. матем. — 2000. — Т. 7. Вып. 1. — С. 109.
2. Клыкова Н. В. Предельное распределение числа совпадающих промежутков // Теория вероятностей и ее применения. — 2002. — Т. 47. Вып. 1. — С. 147–152.
3. Кнут Д. Искусство программирования. Т. 2. Получисленные методы. — М.: ООО «И. Д. Вильямс», 2007.
4. A statistical test suite for random and pseudorandom number generators for cryptographic applications. — NIST Special Publication 800-22.
<http://csrc.nist.gov/groups/ST/toolkit/rng/documents/>
5. Barbour A. D., Chen L. H. Y. An introduction to Stein's method. — World Scientific, 2005.
6. Wagner D. The boomerang attack // In: FSE-99, Lect. Notes in Comput. Sci. — 1999. — Vol. 1636. — P. 156–170.