УДК 004.056.5

УТОЧНЕНИЕ ТЕСТА ДЛИННЫХ СЕРИЙ ПРИ ТЕСТИРОВАНИИ ВЫХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ ЧИСЕЛ

А. Н. Гайдук

Белорусский государственный университет,
Минск, Беларусь, gaidukan@bsu.by

Тест длинных серий проверяет, соответствует ли частота максимальной длины серии из единиц (или из нулей) в двоичном фрагменте длины L наблюдаемой последовательности теоретически ожидаемому значению π_i. При больших значениях L вычисление теоретических значений π_i требует значительных вычислительных ресурсов и в опубликованных работах эти значения рассчитаны с точностью до 4 знака после запятой. В данной работе уточнены данные значения для более точного расчета статистики теста.

Ключевые слова: случайные числа; статистический тест.

REVISITING THE TEST FOR THE LONGEST RUN OF ONES IN A BLOCK IN RANDOMNESS TEST SUITE

A. N. Gaiduk

Belarusian State University,
Minsk, Belarus, gaidukan@bsu.by

The purpose of the test for the longest run of ones in a block is to determine whether the length of the longest run of ones within the tested sequence is consistent with the length of the longest run of ones that would be expected in a random sequence. For large values of block lengths, the calculation of theoretical values π_i requires significant computational resources, and in published works these values are calculated with an accuracy of 4 decimal places. In this paper, theoretical values π_i are refined for a more accurate calculation of the test statistics.

Keywords: random number; statistical test.

1. Введение

Эффективность современной криптографической защиты информации основана на применении надежных генераторов случайных числовых последовательностей (ГСЧП). ГСЧП используются при создании

криптографических ключей, векторов инициализации, выработке начальных значений параметров алгоритмов, а также для решения других задач. Основным требованием, предъявляемым к используемому двоичному ГСЧП, является требование о неотличимости выходной последовательности такого генератора от последовательности независимых испытаний Бернулли с вероятностью успеха 1/2. Для проверки качества ГСЧП используются батареи (наборы) статистических тестов, задача которых – выявление отклонений от «чистой случайности». На практике широко используются такие батареи тестов как NIST, Diehard, TestU01.

Целью данной работы является уточнение теоретических значений π_i теста длинных серий, входящего в батарею NIST [1] и в методику [2]. В опубликованных работах [1, 2] эти значения рассчитаны с точностью до 4-го знака после запятой, что приводит к возникновению значительной погрешности при вычислении результирующего *P*-значение теста [3].

2. Тест длинных серий

Приведем описание теста длинных серий согласно работе [2]. Тест длинных серий используется для выявления соответствия частоты максимальной серии из единиц (или нулей) в *L*-битном фрагменте теоретически ожидаемому значению [1].

Параметр теста: длина фрагмента *L*.

Входные данные: последовательность $X = (x_1, \ldots, x_n)$.

Шаг 1. Разбить исходную последовательность *X* на $m = [n / L]$ фрагментов длины *L* бит.

$$X_i = (x_{(i-1)L+1}, \ldots, x_{(i-1)L+L}), \quad i = \overline{1, m}.$$

Шаг 2. По *L* используя табл. 2, 4 определить разбиение множества длин серий на подмножества (классы) $D_0, \ldots, D_K$, и теоретические вероятности $\pi_0, \ldots, \pi_K$.

Шаг 3. Для $i = 1, \ldots, m$ найти длину b_i максимальной серии из единиц в фрагменте X_i.

Шаг 4. Для $j = 1, \ldots, K$ вычислить частоты попадания статистик b_i в класс D_j:

$$\nu_j = \sum_{i=1}^{m} \mathrm{I}\{b_i \in D_j\}.$$

Шаг 5. Вычислить статистику теста:

$$S = \sum_{i=1}^{m} \frac{(\nu_j - m\pi_j)^2}{m\pi_j}.$$

Шаг 6. Вычислить и возвратить *P*-значение теста:

$$P = 1 - F_{\chi^2_K}(S).$$

Условия применения. Должно выполняться условие: $m \cdot \min_{0 \leqslant i \leqslant K} \pi_i \geqslant 10$. Откуда следует, что минимальная длина последовательности для различных значений параметра должна удовлетворять значениям из табл. 1.

Таблица 1

Минимальные значения *n* теста длинных серий

L	8	128	512	1000	10000
n	427	12464	50688	100000	1490000

Примечание. Если верна гипотеза H_0 и в *i*-ом фрагменте встречается *r* единиц и $L - r$ нулей, то теоретические вероятности равны [1]:

$$\mathrm{P}\{b_i \leqslant q\} = \sum_{r=0}^{L} C_L^r \mathrm{P}\{b_i \leqslant q \mid r\} \frac{1}{2^L},$$

$$\mathrm{P}\{b_i \leqslant q \mid r\} = \frac{1}{C_L^r} \sum_{j=0}^{U} (-1)^j C_{L-r+1}^j C_{L-j(q+1)}^{L-r},$$

где $0 \leqslant q \leqslant L$, $U = \min(L - r + 1, [r/(q+1)])$, $0 \leqslant r \leqslant L$.

При больших значениях *L* вычисление вероятностей по данным формулам требует значительных вычислительных и временных ресурсов, поэтому в работах [1, 2] приведены предвычисленные значения для *L* из множества $\{8, 128, 512, 1000, 10000\}$ (табл. 2). Однако приведенные значения указаны лишь с точностью до 4 знака после запятой (табл. 3), что согласно работе [3] приводит к возникновению значительной погрешности при вычислениях статистики теста. В настоящей работе приведены более точные значения (табл. 4).

Таблица 2

Разбиение множества значений статистики b_i на классы согласно работам [1, 2]

L	8	128	512	1000	10000
K	3	5	5	5	6
D_0	$\leqslant 1$	$\leqslant 4$	$\leqslant 6$	$\leqslant 7$	$\leqslant 10$
D_1	2	5	7	8	11
D_2	3	6	8	9	12
D_3	$\geqslant 4$	7	9	10	13
D_4	—	8	10	11	14
D_5	—	$\geqslant 9$	$\geqslant 11$	$\geqslant 12$	15
D_6	—	—	—	—	$\geqslant 16$

Таблица 3

Вероятности попадания статистики в классы согласно работам [1, 2]

L	8	128	512	1000	10000
K	3	5	5	5	6
π_0	0,2148	0,1174	0,1170	0,1307	0,0882
π_1	0,3672	0,2430	0,2460	0,2437	0,2092
π_2	0,2305	0,2493	0,2523	0,2452	0,2483
π_3	0,1875	0,1752	0,1755	0,1714	0,1933
π_4	—	0,1027	0,1015	0,1002	0,1208
π_5	—	0,1124	0,1077	0,1088	0,0675
π_6	—	—	—	—	0,0727

Таблица 4

Более «точные» вероятности попадания статистики в классы

L	8	128	512	1000	10000
K	3	5	5	5	6
π_0	0,21484375	0,1174035788	0,1299334833	0,1388551909	0,0866323111
π_1	0,3671875	0,2429559593	0,2361223453	0,2369038167	0,2082006484
π_2	0,23046875	0,2493634832	0,2418343719	0,2387912400	0,2484185819
π_3	0,1875	0,1751770603	0,1729754256	0,1700180792	0,1939127867
π_4	—	0,1027010713	0,1032697631	0,1014361440	0,1214584851
π_5	—	0,1123988471	0,1158646108	0,1139955292	0,0680110893
π_6	—	—	—	—	0,0733660975

3. Заключение

В данной работе уточнены теоретические значения π_i теста длинных серий, входящего в батарею NIST [1] и в методику [2] для более точного вычисления статистики теста.

Библиографические ссылки

1. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications / L. E. Bassham [et al.]. NIST. 2010. Special Publication 800-22rev1a. URL: https://www.nist.gov/publications/statistical-test-suite-random-and-pseudorandom-number-generators-cryptographic (date of access: 08.09.2025).

2. Методика испытаний средств криптографической защиты информации на соответствие требованиям СТБ 34.101.27-2022. МИ.10127.10.01. URL: https://www.oac.gov.by/public/content/files/files/met-10127-10-01.pdf (дата обращения: 08.09.2025).

3. *Riesinger E., Fuß J.* Testing the Tests — Opportunities for Corrections and Improvements in NIST SP 800-22r1a and its Reference Code. URL: https://eprint.iacr.org/2025/856 (date of access: 08.09.2025).