УДК 004.056

СТАТИСТИЧЕСКОЕ ТЕСТИРОВАНИЕ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ЗАКОНА ПОВТОРНОГО ЛОГАРИФМА

А. И. ТРУБЕЙ, канд. физ.-мат. наук, доц. В. Ю. ПАЛУХА,
магистрант П. В. ГРУДИНСКИЙ
(Белорусский государственный университет, Минск)

Аннотация. *В докладе описывается двухэтапная методика принятия решений о качестве последовательностей с применением закона повторного логарифма, в которой на первом этапе применяется тест серий, а в качестве статистического расстояния используется статистика хи-квадрат согласия. Представлены результаты экспериментов по тестированию генераторов случайных и псевдослучайных последовательностей.*

Ключевые слова: *статистическое тестирование генераторов, закон повторного логарифма.*

Введение. Идея тестирования с применением закона повторного логарифма (ЗПЛ-тестирования) была предложена в статье [1]. В работе [2] приведена двухэтапная процедура проверки гипотез с применением закона повторного логарифма для теста многомерной дискретной равномерности по непересекающимся отрезкам (МДРН) при $L=1$ – теста Монобит, а в работе [3] – для теста МДРН в общем случае. В данном докладе на первом этапе двухэтапной процедуры предлагается использовать тест серий.

Двухэтапная процедура тестирования на основе закона повторного логарифма и теста серий. Тест серий является одним из базовых критериев статистического тестирования случайных и псевдослучайных последовательностей. Он входит в состав американского стандарта обработки информации FIPS PUB 140-1, где данный тест применяется для тестирования последовательностей, используемых при выработке ключей. Целью теста серий является проверка соответствия числа серий из нулей и единиц различной длины в наблюдаемой последовательности теоретически ожидаемому числу для дискретных равномерно распределенных случайных последовательностей.

Пусть имеется двоичная последовательность:

$$X=\{x_1,x_2,...,x_n\}.$$

Фрагмент $\{x_t,x_{t+1},...,x_{t+l-1}\}$ последовательности X называется серией длины l, если $x_t=x_{t+1}=...=x_{t+l-1}$, но $x_{t-1}\neq x_{t+1}$ или $(t=1)$ и $x_{t+l-1}\neq x_{t+l}$ (или $t+l-1=1$).

Вычислим теоретические частоты серий [4]:

$$\mu_i = \frac{n-i+3}{2^{i+2}}, \quad i = \overline{1,k}.$$

где k равно наибольшему целому i, для которого $\mu_i \geq 5$.

Для $i = 1, \ldots, k$ вычислим частоты v_i^0 и v_i^1 серий соответственно из нулей и единиц длины i в последовательности X. Обозначим

$$V = \sum_{i=1}^{k} \left(v_i^0 + \nu_i^1 \right).$$

Вычислим статистику теста:

$$S(n) = \chi^2_{2k-2}(n) = \sum_{i=1}^{k} \frac{\left(\nu_i^0 - \mu_i \right)^2}{\mu_i} + \sum_{i=1}^{k} \frac{\left(\nu_i^1 - \mu_i \right)^2}{\mu_i}. \qquad (1)$$

Заметим, что серии длины i, для которых $\mu_i < 5$, при построении статистики не учитываются. Кроме того, легко проверить, что $E\{V\} \xrightarrow[n\to\infty]{} n/2$.

При $n \to \infty$ статистика $S(n)$ имеет хи-квадрат распределение $\chi^2_{2k-2}(n)$ с $2k-2$ степенями свободы.

В силу центральной предельной теоремы, при большом числе степеней свободы распределение случайной величины S(n) может быть аппроксимировано нормальным распределением. Более точно, при $2k-2 \to \infty$

$$S(n)^* = \frac{S(n) - (2k-2)}{\sqrt{2(2k-2)}} \to \mathcal{N}(0,1). \qquad (2)$$

При $2k-2 \geq 30$, то есть при $k \geq 16$, это даёт аппроксимацию, достаточную для практических целей.

Согласно закону повторного логарифма (в общем виде) справедлива формула

$$\limsup_{k\to\infty} \frac{S(n)^*}{\sqrt{2\ln\ln V}} = \limsup_{k\to\infty} \frac{\dfrac{S(n) - (2k-2)}{\sqrt{2(2k-2)}}}{\sqrt{2\ln\ln V}} = 1.$$

Следовательно, для теста серий при $k \geq 16$ можно использовать следующую статистику:

$$S_{зпл}(n) = \frac{S(n)^*}{\sqrt{2\ln\ln V}} = \frac{\dfrac{S(n) - (2k-2)}{\sqrt{2(2k-2)}}}{\sqrt{2\ln\ln V}}.$$

Для теста серий меру μ_n^U, соответствующую РРСП, можно рассчитать следующим образом:

$$\mu_n^U\left\{(-\infty, z]\right\} = \Phi\left(z\sqrt{2\ln\ln V}\right) = \sqrt{2\ln\ln V}\int_{-\infty}^{z}\phi\left(s\sqrt{2\ln\ln V}\right)ds\,, \qquad (3)$$

где $\Phi(\cdot)$ и $\phi(\cdot)$ – соответственно функция распределения и плотность стандартного нормального закона.

Однако, как было отмечено в статье [3], полученный в результате нормировки график плотности хи-квадрат распределения при $16 \leq k \leq 512$ будет недостаточно симметричным и не в полной мере похож на график плотности нормального закона. Поэтому вероятностную меру μ_n^U следует рассчитать по следующей формуле:

$$\mu_n^U\left\{(-\infty, z]\right\} = F_{\chi^2_{2k-2}}\left(2z\sqrt{(2k-2)\ln\ln V} + 2k - 2\right). \qquad (4)$$

Таким образом, чтобы оценить генератор *G* с применением закона повторного логарифма для теста серий, необходимо:

1. Осуществить генерацию набора $\mathcal{R} \in \{0,1\}^n$ из $m = 10000$ последовательностей возможно большей длины n.

2. На первом этапе двухэтапной процедуры проверки гипотез вычислить значения статистики $S_{зпл}(n)$ по всем m последовательностям.

3. На втором этапе двухэтапной процедуры проверки гипотез сравнить между собой вероятностные меры $\mu_n^{\mathcal{R}^n}$ и μ_n^U Для сравнения будем использовать следующую статистику χ^2-согласия:

$$\chi^2 = \sum_{j=1}^{|\mathcal{B}|}\frac{\left[v_n^{\mathcal{R}_n}(I_j) - mp_n^{\mathrm{U}}(I_j)\right]^2}{mp_n^{\mathrm{U}}(I_j)},$$

где $v_n^{\mathcal{R}_n}(I_j)$ – частоты попадания значений статистики $S_{зпл}(n)$ в интервал I_j разбиения $\mathcal{B}$ числовой прямой по всем *m* последовательностям;

$p_n^{\cup}(I_j)$ – теоретические вероятности попадания $S_{зпл}(n)$ в интервал I_j для РРСП, рассчитанные по формуле (4), где в качестве V используется асимптотическое математическое ожидание этой величины, равное $n/2$.

Полагаем, что генератор *G* прошел тестирование по тесту серий с применением закона повторного логарифма, если *P*-значения статистики χ^2-согласия превышают заданный уровень значимости α, то есть $P \geq \alpha$.

Результаты экспериментов. Для проверки гипотезы H_0 о том, что генератор порождает РРСП, проведено тестирование 10 000 последовательностей, выработанных соответственно линейным конгруэнтным генератором (ЛКГ) и стандартом СТБ 34.101.47-2012 (в режиме счётчика). Результаты сравнительного тестирования по тесту серий с применением закона повторного логарифма приведены в таблице.

Таблица. – Результаты тестирования по тесту серий последовательностей, выработанных ЛКГ и СТБ 34.101.47-2012 (в режиме счётчика)

Объём (GB)	ЛКГ			СТБ 34.101.47-2012		
	Степ. своб.	χ^2	P-знач.	Степ. своб.	χ^2	P-знач.
5 GB	41	131.41	2.1*10-11	41	58.89	0.0347
10 GB	41	166.53	<2.2*10-16	41	39.153	0.553

Из таблицы видно, что для последовательностей (объёмом 5 GB и 10 GB), сгенерированных в соответствии с СТБ 34.101.47-2012 (в режиме счётчика), выполняется гипотеза H_0 согласия с моделью независимых симметричных испытаний Бернулли на уровне значимости $\alpha = 0.01$. В то время как для последовательности, вырабатываемой ЛКГ, гипотеза H_0 на данном уровне значимости не выполняется.

Гистограммы частот выборок, полученных с применением ЗПЛ по тесту серий для линейного конгруэнтного генератора и алгоритма генерации псевдослучайных последовательностей в соответствии с СТБ 34.101.47-2012, приведены на рисунках 1, 2.

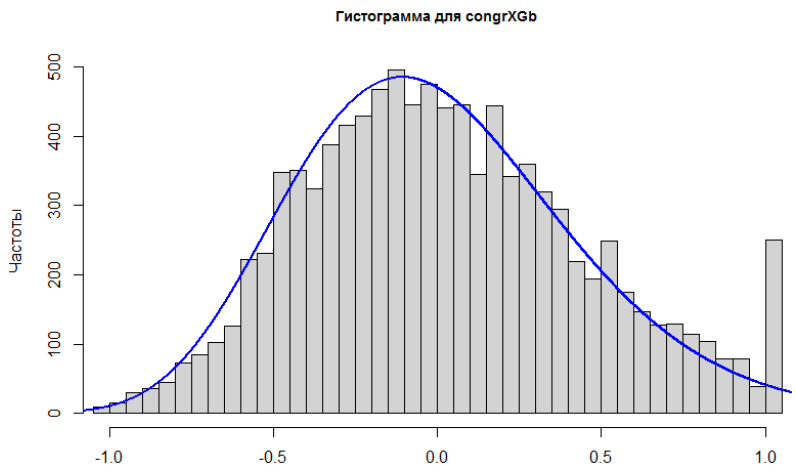

Рисунок 1. – Гистограмма частот линейного конгруэнтного генератора, 10 GB

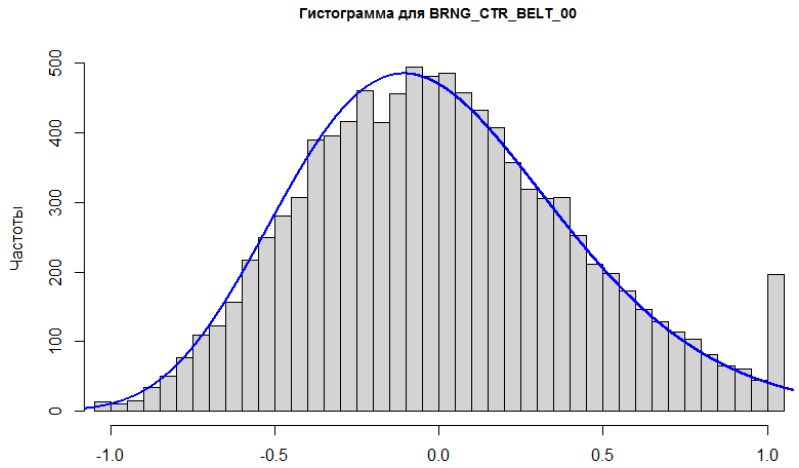

Рисунок 2. – Гистограмма частот алгоритма СТБ 34.101.47-2012, 10 GB

ЛИТЕРАТУРА

1. Wang, Y. On the Design of LIL Tests for (Pseudo) Random Generators and Some Experimental Results [Electronic resource] / Y. Wang. – 2014. – Mode of access: https://arxiv.org/abs/1401.3307. – Date of access: 01.03.2020.
2. Трубей, А. И. Разработка методики статистического тестирования псевдослучайных последовательностей с применением закона повторного логарифма / А. И. Трубей [и др.] // Теоретическая и прикладная криптография: материалы международной научной конференции, Минск, 20–21 октября 2020 г. / Белорусский государственный университет ; редколлегия: Ю. С. Харин (гл. ред.) [и др.]. – Минск : БГУ, 2020. – С. 57–67.
3. Трубей, А. И. Применение тестов на основе закона повторного логарифма для оценки качества случайных последовательностей / А. И. Трубей [и др.] // Комплексная защита информации. Материалы XXVI научно-практической конференции. Минск, 25–27 мая 2021 г. – Минск: Издатель Владимир Сивчиков, 2021. – С. 131–134.
4. Menezes, A. Handbook of Applied Cryptography / A. Menezes, P. van Oorschot, S. Vanstone. – Boca Raton : CRC Press, 1996. – 816 p.