Экз. № 1

Временная методика проверки соответствия образца физического генератора случайных чисел его теоретико-вероятностной модели

Москва, 2022

Содержание

Обозначения и сокращения

$\mathbb{N}$	множество натуральных чисел;
$\mathbb{Z}$	множество целых чисел;
$\mathbb{Z}_{2^s}$	кольцо вычетов по модулю $2^s,\ s \in \mathbb{N}$;
V_n	множество всех двоичных последовательностей длины $n \in \mathbb{N}$;
$\{x\}_{i=1}^n$	последовательность $x_1, \ldots, x_n$ элементов из V_1;
$[z]$	целая часть числа z, равная $\max\{s \in \mathbb{Z} \colon s \leq z\}$;
$I\{A\}$	индикатор события A;
$wt \colon \mathbb{N} \cup \{0\} \to \mathbb{N} \cup \{0\}$	функция, сопоставляющая аргументу число единиц в его двоичном представлении;
$Vec_s \colon \mathbb{Z}_{2^s} \to V_s$	биективное отображение, сопоставляющее элементу кольца $\mathbb{Z}_{2^s}$ его двоичное представление, т.е. для любого элемента $z \in \mathbb{Z}_{2^s}$, представленного в виде $z = z_0 + 2z_1 + \ldots + 2^{s-1}z_{s-1}$, где $z_i \in \{0,1\}$, $i = 0, 1, \ldots, s-1$, выполнено равенство $Vec_s(z) = (z_0, z_1, \ldots, z_{s-1})$;
$Int_s \colon V_s \to \mathbb{Z}_{2^s}$	отображение, обратное к отображению Vec_s;
χ^2_r	функция распределения хи-квадрат с $r \in \mathbb{N}$ степенями свободы;
t_z	квантиль уровня $z \in [0;1]$ стандартного нормального распределения.

Введение

Настоящая методика предназначена для проверки соответствия образца физического генератора случайных чисел (далее − ФГСЧ) его теоретико-вероятностной модели. В рамках методики предполагается, что образец ФГСЧ функционирует по схеме, представленной на рис. 1, и формирует два типа данных: исходную и выходную последовательности. Исходная последовательность является результатом оцифровки физического сигнала, регистрируемого ФГСЧ, выходная последовательность − результатом применения алгоритмов улучшения статистических характеристик типа [1] к исходной последовательности.

Проверка образца ФГСЧ выполняется путём тестирования указанных типов последовательностей. Для исходной последовательности проверяется гипотеза о согласии её статических свойств с условиями применимости алгоритмов типа [1] − проверка гипотезы о неотличимости исходной последовательности от реализации схемы независимых одинаково распределённых испытаний Бернулли. Для выходной последовательности проверяется гипотеза о её неотличимости от реализации схемы равновероятных независимых испытаний Бернулли.

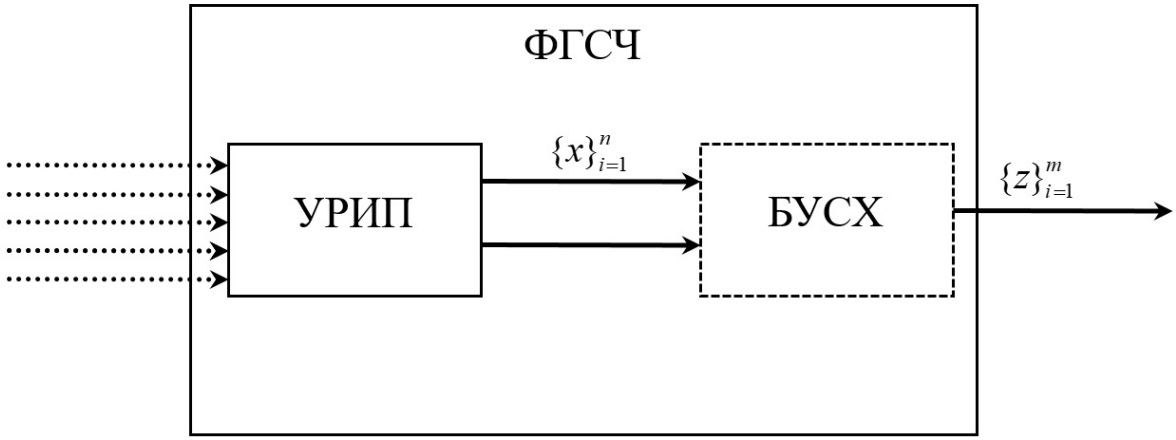

Рис. 1: Блок-схема функционирования ФГСЧ, где "УРИП" − устройство регистрации исходного процесса, "БУСХ" − блок улучшения статистических характеристик, реализующий алгоритм типа [1], $\{x\}_{i=1}^{n}$ − исходная последовательность, $\{z\}_{i=1}^{m}$ − выходная последовательность, $n, m \in \mathbb{N}$.

1. Объекты и параметры тестирования

Объектами тестирования являются:

- $\{x\}_{i=1}^{n}$ – исходная последовательность, $n \in \mathbb{N}$.
- $\{z\}_{i=1}^{m}$ – выходная последовательность, $m \in \mathbb{N}$.

Параметрами тестирования являются:

- α – уровень значимости каждого из статистических критериев, используемых в настоящей методике.
  Рекомендуемое значение параметра α равно $0,005$.
- n – длина последовательности $\{x\}_{i=1}^{n}$, определяемая соотношением

$$n \geq 10^{10}.$$

- δ – максимально допустимое отклонение частоты знака "1" в последовательности $\{z\}_{i=1}^{m}$ от $\frac{1}{2}$;
  Рекомендуемое значение параметра δ равно $0,001$.
- m – длина последовательности $\{z\}_{i=1}^{m}$, определяемая соотношением

$$m \geq \max\left\{\left(\frac{t_{1-\frac{\alpha}{2}}}{4\delta}\right)^2, 10^{10}\right\}.$$

Замечание 1 *При наличии обоснования допускается использование отличных от рекомендуемых в настоящей методике значений параметров α и δ.*

Замечание 2 *Соотношение $n \geq 10^{10}$ является оценкой снизу на длину $\{x\}_{i=1}^{n}$. В зависимости от значения частоты знака "1" в $\{x\}_{i=1}^{n}$, для прохождения этапа 2.3 тестирования может потребоваться большее значение n.*

2. Тестирование исходной последовательности

Тестирование исходной последовательности осуществляется в 3 этапа:

1. Проверка гипотезы независимости знаков в $\{x\}_{i=1}^{n}$.
2. Проверка гипотезы однородности распределения знаков в $\{x\}_{i=1}^{n}$.
3. Проверка согласия распределения числа k-грамм в $\{x\}_{i=1}^{n}$ с полиномиальным законом.

Замечание 3 *Этапы 1−3 нацелены на проверку соответствия исходной последовательности условиям применимости алгоритмов улучшения статистических характеристик типа [1].*

2.1. Проверка гипотезы независимости знаков в исходной последовательности

Для всех $j \in \{0,1\}, \vec{i} = (i_1, \ldots, i_k) \in V_k, k \in \{1, 2, \ldots, k_{max}\}$ рассчитываются величины:

- $\nu_{\vec{i},j} = \sum\limits_{t=1}^{n-k} I\{x_t = i_1, x_{t+1} = i_2, \ldots, x_{t+k-1} = i_k, x_{t+k} = j\}$;
- $\nu_{\vec{i}} = \nu_{\vec{i},0} + \nu_{\vec{i},1}$;
- $u_j = \sum\limits_{\vec{i} \in V_k} \nu_{\vec{i},j}$,

где $k_{max} = \min\left\{24, \max\left\{k \mid \nu_{\vec{i},j} \geq 20,\ \forall \vec{i} \in V_k, j \in \{0,1\}\right\}\right\}$.

Далее для каждого $k \in \{1, 2, \ldots, k_{max}\}$ вычисляется значение статистики

$$S_k^{(1)} = \frac{1}{n} \sum_{\vec{i} \in V_k} \sum_{j=0}^{1} \frac{\left(n \cdot \nu_{\vec{i},j} - \nu_{\vec{i}} \cdot u_j\right)^2}{\nu_{\vec{i}} \cdot u_j}, \tag{1}$$

которая при справедливости проверяемого свойства имеет асимптотическое распределение хи-квадрат с $2^k - 1$ степенями свободы [2, § 5.5, стр. 435].

На основе значений (1) рассчитываются *p-value*

$$P_k^{(1)} = 1 - \chi^2_{2^k-1}\left(S_k^{(1)}\right).$$

Правило принятия решения по п. 2.1. *Если для некоторого k из множества $\{1, 2, \ldots, k_{max}\}$ значение $P_k^{(1)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование по п. 2.2.*

2.2. Проверка гипотезы однородности знаков в исходной последовательности

На основе $\{x\}_{i=1}^n$ формируются последовательности вида:

$$x^{(1)} = (x_1, x_2, \ldots, x_s),$$
$$x^{(2)} = (x_{s+1}, x_{s+2}, \ldots, x_{2s}),$$
$$\vdots$$
$$x^{(\left[\frac{n}{s}\right])} = \left(x_{(\left[\frac{n}{s}\right]-1)s+1}, x_{(\left[\frac{n}{s}\right]-1)s+2}, \ldots, x_{\left[\frac{n}{s}\right]s}\right),$$

где $s \in \mathbb{N}$, $s \leq \frac{n}{2}$, и определяется

$$k = \min\left\{s \in \mathbb{N} \mid \forall j \in \left\{1, 2, \ldots, \left[\frac{n}{s}\right]\right\} : \min\left\{s - wt\left(Int_s\left(x^{(j)}\right)\right), wt\left(Int_s\left(x^{(j)}\right)\right)\right\} \geq 20\right\}.$$

Далее для всех $t \in \{1, 2, \ldots, \left[\frac{n}{k}\right]\}$, $j \in \{0, 1\}$ рассчитываются величины:

- $v_{t,j} = \sum\limits_{i=1}^{k} I\left\{x_{(t-1)k+i} = j\right\}$;
- $u_j = \sum\limits_{i=1}^{\left[\frac{n}{k}\right]} v_{i,j}$,

с использованием которых вычисляется значение статистики

$$S_k^{(2)} = \left[\frac{n}{k}\right] \sum_{i=1}^{\left[\frac{n}{k}\right]} \sum_{j=0}^{1} \frac{1}{u_j} \left(v_{i,j} - \frac{u_j}{\left[\frac{n}{k}\right]}\right)^2. \tag{2}$$

При справедливости проверяемого свойства статистика $S_k^{(2)}$ имеет асимптотическое распределение хи-квадрат с $\left[\frac{n}{k}\right] - 1$ степенями свободы [2, § 4.4, стр. 342].

На основе значения (2) рассчитывается *p-value*

$$P_k^{(2)} = 1 - \chi^2_{\left[\frac{n}{k}\right]-1}\left(S_k^{(2)}\right).$$

Правило принятия решения по п. 2.2. *Если значение $P_k^{(2)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование по п. 2.3.*

2.3. Проверка согласия распределения числа k-грамм в исходной последовательности с полиномиальным законом

На основе $\{x\}_{i=1}^{n}$ формируются последовательности вида:

$$x^{(1)} = (x_1, x_2, \ldots, x_s),$$
$$x^{(2)} = (x_{s+1}, x_{s+2}, \ldots, x_{2s}),$$
$$\vdots$$
$$x^{([\frac{n}{s}])} = \left(x_{([\frac{n}{s}]-1)s+1}, x_{([\frac{n}{s}]-1)s+2}, \ldots, x_{[\frac{n}{s}]s}\right),$$

где $s \in \mathbb{N}$, $s > 1$, и значения

$$v_i^{(s)} = \sum_{j=1}^{[\frac{n}{s}]} I\left\{Int_s(x^{(j)}) = i\right\}$$

для всех $i \in \{0, 1, \ldots, 2^s - 1\}$. Далее определяется

$$k_{max} = \min\left\{16, \max\left\{6, s_{max}\right\}\right\},$$

где

$$s_{max} = \max\left\{s \in \mathbb{N} \mid \forall i \in \{0, 1, \ldots, 2^s - 1\} \colon v_i^{(s)} \geq 20\right\}.$$

Замечание 4 *Если для некоторых $k \in \{2, 3, \ldots, 6\}$, $i \in \{0, 1, \ldots, 2^k - 1\}$ значение $v_i^{(k)} < 20$, то процедура тестирования завершается с отрицательным результатом, но образец ФГСЧ не бракуется. Допускается повторное тестирование данного образца ФГСЧ на исходной последовательности длины, достаточной для выполнения условий:*

- $n \geq 10^{10}$;
- $\forall\, k \in \{2, 3, \ldots, 6\}$, $\forall i \in \{0, 1, \ldots, 2^k - 1\}$ *значение* $v_i^{(k)} \geq 20$.

Для $k = 2, 3, \ldots, k_{max}$ при справедливости проверяемого свойства последовательность k-грамм $x^{(1)}, x^{(2)}, \ldots, x^{([\frac{n}{k}])}$ является реализацией выборки из полиномиального распределения с параметрами $(\theta^k, \theta^{k-1}(1-\theta), \ldots, (1-\theta)^k)$, где $\theta \in (0, 1)$.

Для каждого из указанных значений k на основе последовательности $x^{(1)}, x^{(2)}, \ldots, x^{([\frac{n}{k}])}$ по методу максимального правдоподобия [2, § 3.5, стр. 219]

вычисляется оценка $\widehat{\theta}_k$ для параметра θ:

$$\widehat{\theta}_k = \sum_{i=0}^{2^k-1} \frac{v_i^{(k)} \cdot wt(i)}{k\left[\frac{n}{k}\right]} = \frac{1}{k\left[\frac{n}{k}\right]} \sum_{i=1}^{k\left[\frac{n}{k}\right]} x_i.$$

Затем для всех $k \in \{2, 3, \ldots, k_{max}\}$, $i \in \{0, 1, \ldots, 2^k - 1\}$ вычисляются

$$p_i^{(k)}(\widehat{\theta}_k) = \left(\widehat{\theta}_k\right)^{wt(i)} \cdot \left(1 - \widehat{\theta}_k\right)^{k-wt(i)}$$

и значение статистики

$$S_k^{(3)} = \sum_{i=0}^{2^k-1} \frac{\left(v_i^{(k)} - p_i^{(k)}(\widehat{\theta}_k) \cdot \left[\frac{n}{k}\right]\right)^2}{p_i^{(k)}(\widehat{\theta}_k) \cdot \left[\frac{n}{k}\right]}. \tag{3}$$

При справедливости проверяемого свойства статистика $S_k^{(3)}$ имеет асимптотическое распределение хи-квадрат с $2^k - 2$ степенями свободы [2, § 4.2, стр. 329].

На основе значений (3) рассчитываются *p-value*

$$P_k^{(3)} = 1 - \chi^2_{2^k-2}\left(S_k^{(3)}\right).$$

Правило принятия решения по п. 2.3. *Если для некоторого k из множества $\{2, 3, \ldots, k_{max}\}$ значение $P_k^{(3)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование выполняется тестирование выходной последовательности по п. 3.*

3. Тестирование выходной последовательности

Тестирование выходной последовательности осуществляется в 4 этапа:

1. Проверка соответствия частот знаков в $\{z\}_{i=1}^{m}$ теоретико-вероятностной модели образца ФГСЧ.
2. Проверка гипотезы независимости знаков в $\{z\}_{i=1}^{m}$.
3. Проверка гипотезы однородности распределения знаков в $\{z\}_{i=1}^{m}$.
4. Проверка согласия распределения числа k-грамм в $\{z\}_{i=1}^{m}$ с полиномиальным законом.

3.1. Проверка соответствия частот знаков в выходной последовательности теоретико-вероятностной модели образца ФГСЧ

В соответствии с теоретико-вероятностной моделью образца ФГСЧ интервал значений для частоты знака "1" в $\{z\}_{i=1}^{m}$ при надежности $1-\alpha$ имеет вид:

$$\Delta = \left[\frac{1}{2} - \frac{t_{1-\frac{\alpha}{2}}}{2\sqrt{m}}; \frac{1}{2} + \frac{t_{1-\frac{\alpha}{2}}}{2\sqrt{m}}\right].$$

На основе $\{z\}_{i=1}^{m}$ вычисляется значение статистики

$$T^{(0)} = \frac{1}{m}\sum_{i=1}^{m} z_i.$$

Правило принятия решения по п. 3.1. *Если $T^{(0)} \notin \Delta$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование по п. 3.2.*

3.2. Проверка гипотезы независимости знаков в выходной последовательности

Для всех $j \in \{0,1\}, \vec{i} = (i_1, \ldots, i_k) \in V_k, k \in \{1, 2, \ldots, k_{max}\}$ рассчитываются величины:

- $v_{\vec{i},j} = \sum\limits_{t=1}^{m-k} I\{z_t = i_1, z_{t+1} = i_2, \ldots, z_{t+k-1} = i_k, z_{t+k} = j\}$;
- $v_{\vec{i}} = v_{\vec{i},0} + v_{\vec{i},1}$;

- $u_j = \sum\limits_{\vec{i} \in V_k} v_{\vec{i},j}$,

где $k_{max} = \min\left\{24, \max\left\{k \mid v_{\vec{i},j} \geq 20,\ \forall \vec{i} \in V_k, j \in \{0,1\}\right\}\right\}$.

Для каждого $k \in \{1, 2, \ldots, k_{max}\}$ вычисляются значения статистики

$$T_k^{(1)} = \frac{1}{m} \sum_{\vec{i} \in V_k} \sum_{j=0}^{1} \frac{\left(m \cdot v_{\vec{i},j} - v_{\vec{i}} \cdot u_j\right)^2}{v_{\vec{i}} \cdot u_j}, \tag{4}$$

которая при справедливости проверяемого свойства имеет асимптотическое распределение хи-квадрат с $2^k - 1$ степенями свободы [2, § 5.5, стр. 435].

На основе значений (4) рассчитываются *p-value*

$$R_k^{(1)} = 1 - \chi^2_{2^k-1}\left(T_k^{(1)}\right).$$

Правило принятия решения по п. 3.2. *Если для некоторого k из множества $\{1, 2, \ldots, k_{max}\}$, значение $R_k^{(1)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование по п. 3.3.*

3.3. Проверка гипотезы однородности знаков в выходной последовательности

На основе $\{z\}_{i=1}^m$ формируются последовательности вида:

$$\begin{gathered} z^{(1)} = (z_1, z_2, \ldots, z_s), \\ z^{(2)} = (z_{s+1}, z_{s+2}, \ldots, z_{2s}), \\ \vdots \\ z^{(\left[\frac{m}{s}\right])} = \left(z_{(\left[\frac{m}{s}\right]-1)s+1}, z_{(\left[\frac{m}{s}\right]-1)s+2}, \ldots, z_{\left[\frac{m}{s}\right]s}\right), \end{gathered}$$

где $s \in \mathbb{N}$, $s \leq \frac{m}{2}$, и определяется

$$k = \min\left\{s \in \mathbb{N} \mid \forall j \in \left\{1, 2, \ldots, \left[\frac{m}{s}\right]\right\} : \min\left\{s - wt\left(Int_s\left(z^{(j)}\right)\right), wt\left(Int_s\left(z^{(j)}\right)\right)\right\} \geq 100\right\}.$$

Далее для всех $t \in \{1, 2, \ldots, \left[\frac{m}{k}\right]\}$, $j \in \{0, 1\}$ рассчитываются величины:

- $v_{t,j} = \sum\limits_{i=1}^{k} I\left\{z_{(t-1)k+i} = j\right\}$;

- $u_j = \sum\limits_{i=1}^{\left[\frac{m}{k}\right]} v_{i,j}$,

с использованием которых вычисляется значение статистики

$$T_k^{(2)} = \left[\frac{m}{k}\right] \sum_{i=1}^{\left[\frac{m}{k}\right]} \sum_{j=0}^{1} \frac{1}{u_j} \left(v_{i,j} - \frac{u_j}{\left[\frac{m}{k}\right]} \right)^2 . \tag{5}$$

При справедливости проверяемого свойства статистика $T_k^{(2)}$ имеет асимптотическое распределение хи-квадрат с $\left[\frac{m}{k}\right] - 1$ степенями свободы [2, § 4.4, стр. 342].

На основе значений (5) рассчитывается *p-value*

$$R_k^{(2)} = 1 - \chi^2_{\left[\frac{m}{k}\right]-1} \left(T_k^{(2)} \right).$$

Правило принятия решения по п. 3.3. *Если значение $R_k^{(2)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае выполняется тестирование по п. 3.4.*

3.4. Проверка согласия распределения числа k-грамм в выходной последовательности с полиномиальным законом

На основе $\{z\}_{i=1}^{m}$ формируются последовательности

$$\begin{gathered} z^{(1)} = (z_1, z_2, \ldots, z_k), \\ z^{(2)} = (z_{k+1}, z_{k+2}, \ldots, z_{2k}), \\ \vdots \\ z^{(\left[\frac{m}{k}\right])} = \left(z_{(\left[\frac{m}{k}\right]-1)k+1}, z_{(\left[\frac{m}{k}\right]-1)k+2}, \ldots, z_{\left[\frac{m}{k}\right]k} \right), \end{gathered}$$

где $k = 2, 3, \ldots, 16$.

Для всех $k \in \{2, 3, \ldots, 16\}$, $i \in \{0, 1, \ldots, 2^k - 1\}$ рассчитываются

$$v_i^{(k)} = \sum_{j=1}^{\left[\frac{m}{k}\right]} I \left\{ Int_k(z^{(j)}) = i \right\},$$

с использованием которых вычисляется значения статистики

$$T_k^{(3)} = \sum_{i=0}^{2^k-1} \frac{\left(2^k \cdot v_i^{(k)} - \left[\frac{m}{k}\right] \right)^2}{2^k \left[\frac{m}{k}\right]}. \tag{6}$$

При справедливости проверяемого свойства статистика $T_k^{(3)}$ имеет асимптотическое распределение хи-квадрат с $2^k - 1$ степенями свободы [2, § 4.2, стр. 322].

На основе значений (6) расчитываются *p-value*

$$R_k^{(3)} = 1 - \chi^2_{2^k-1}\left(T_k^{(3)}\right).$$

Правило принятия решения по п. 3.4. *Если для некоторого k из множества $\{2, 3, \ldots, 16\}$, значение $R_k^{(3)} < \alpha$, то образец ФГСЧ бракуется, процедура тестирования завершается с отрицательным результатом, в противном случае образец признаётся удовлетворяющим временной методики проверки соответствия образца ФГСЧ его теоретико-вероятностной модели .*

Список литературы

[1] *Рябко Б.Я., Мачикина Е.П.* Эффективное преобразование случайных последовательностей в равновероятностные и независимые, *Проблемы передачи информации*, 1999, том 35, выпуск 2, 23–28.

[2] *Ивченко Г.И., Медведев Ю.И.* *Введение в математическую статистику*, М.: ЛЕНАНД, 2017, 606 с.