УДК 519.246.2:004.415.538:004.421.5

Э.В. ФАУРЕ, А.И. ЩЕРБА, В.Н. РУДНИЦКИЙ

МЕТОД И КРИТЕРИЙ ОЦЕНИВАНИЯ КАЧЕСТВА ПОСЛЕДОВАТЕЛЬНОСТЕЙ СЛУЧАЙНЫХ ЧИСЕЛ

Аннотация. Изучена зависимость равномерности распределения знаков эмпирической автокорреляционной функции относительно количества перекрывающихся символов отрезков, на которые разбивается последовательность случайных чисел. Установлен допустимый «порог» перекрытия, ниже которого наблюдается равномерное распределение знаков автокорреляционной функции. Определено понятие барьерной функции. На ее основании разработан критерий оценивания качества генераторов случайных чисел. Представлены методика его применения и ее реализация для нескольких известных генераторов.

Ключевые слова: случайные числа, корреляция, оценка качества, статистический критерий.

ВВЕДЕНИЕ

Использование последовательностей случайных чисел является неотъемлемой частью процессов имитационного моделирования, проведения статистических экспериментов, передачи и преобразования информации. В практических приложениях часто применяют искусственно сгенерированные последовательности псевдослучайных чисел (ПСЧ). В частности, для криптографического преобразования в поточном режиме шифрующая последовательность должна иметь свойство воспроизводимости, что ограничивает в данном случае использование последовательностей естественных случайных чисел. Вместе с тем необходимо, чтобы применяемая последовательность ПСЧ была статистически неотличимой от последовательности случайных чисел, порождаемой естественным источником дискретного белого шума (генератором случайных чисел (ГСЧ)). Отметим, что толкование термина «белый шум» неоднозначно [1, 2]. В настоящей статье используется определение из [2], и под дискретным белым шумом понимается стационарный дискретный случайный процесс, выборки которого некоррелированы.

Для анализа последовательностей ПСЧ применяется широкий спектр тестов и критериев [3–6], многие из которых реализованы в специализированных прикладных пакетах тестов (DIEHARD [7], NIST [8], TestU01 [9] и др.). Заметим, что все разработанные критерии определяют только некоторые условия, необходимые для обеспечения неотличимости последовательностей ПСЧ от последовательностей случайных чисел, и в настоящее время известны алгоритмы, которые проходят приведенные пакеты тестов [9–12]. Кроме того, статистические критерии и тесты для последовательностей ПСЧ можно также применять для анализа ГСЧ в целях выявления неоптимальных процедур преобразования физических процессов в последовательности случайных чисел. Таким образом, актуальна проблема создания новых критериев, с помощью которых можно выявлять свойства, характерные только для естественного дискретного белого шума.

Наиболее распространенным для анализа последовательности случайных чисел (аналога дискретного белого шума) является автокорреляционный тест. Для его реализации предложены различные подходы [3, 4, 13, 14]. Вместе с тем основная и общая для всех подходов задача автокорреляционного анализа — определение соответствия оценки автокорреляционной функции (АКФ) изучаемой последовательности дельта-функции Дирака [15], что, как правило, проводится с помощью t-критерия Стьюдента [16].

В работе [17] выполнен сравнительный анализ оценок АКФ последовательностей ПСЧ комбинационного генератора [18] и генератора MT19937 [10] с оценками АКФ случайных последовательностей чисел атмосферного шума [19] и квантового ГСЧ [20]. Для формирования оценок АКФ использован подход, изложенный в [13, 14]. Полученные результаты свидетельствуют об однородности оценок АКФ для всех рассмотренных последовательностей. Данная статья является продолжением начатого в [17] исследования знаков боковых лепестков эмпирической АКФ — точек на коррелограмме, соответствующих коэффициентам корреляции ненулевого порядка.

Цель данной работы — выявление корреляционных свойств, присущих последовательностям, порожденным естественными источниками дискретного белого шума, и не присущих искусственно сгенерированным последовательностям ПСЧ, а также разработка метода и критерия оценивания качества последовательностей равномерно распределенных случайных и псевдослучайных чисел, позволяющих выявлять в них не обнаруженные до настоящего времени статистические нерегулярности.

Дальнейший анализ проведем для последовательностей чисел, синтезированных естественным источником случайных чисел [19], генерирующим случайность с помощью атмосферного шума и выбранным в качестве эталонного, квантовым ГСЧ [20], основанным на измерении времени регистрации фотонов, а также генератором «Вихрь Мерсенна» MT19937 [10].

ВЫЧИСЛЕНИЕ ОЦЕНОК НОРМИРОВАННЫХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ

Корреляционный анализ исследуемых последовательностей проведем на основе вычисления оценок коэффициентов корреляции между последовательными отрезками (подпоследовательностями) фиксированной длины. Для этого последовательность чисел (выборку) $\{x(n)\}$ длины L_{seq}, $0 \le n \le L_{seq} - 1$, разобьем на следующие один за другим отрезки длины L_{segm}: $\{x(k), x(k+1), \ldots, x(k+L_{segm}-1)\}$, которые могут перекрываться или отстоять один от другого на некоторое количество символов. Расстояние между соседними отрезками $\{x(k), x(k+1), \ldots, x(k+L_{segm}-1)\}$ и $\{x(l), x(l+1), \ldots, x(l+L_{segm}-1)\}$ при $k < l$ определяется величиной $\tau = l - k - L_{segm}$. Отрицательное значение τ характеризует перекрытие (наложение, зацепление) отрезков на величину абсолютного значения τ. Таким образом, $\tau \ge 1 - L_{segm}$ и может принимать любые целые значения из этой области.

Пронумеруем отрезки в соответствии с номерами их первых элементов. Например, отрезок $\{x(k), x(k+1), \ldots, x(k+L_{segm}-1)\}$ имеет номер k.

Описанная схема разбиения выборки на отрезки представлена на рис. 1, где области перекрытия соседних отрезков обозначены черным цветом, а области между соседними отрезками — серым.

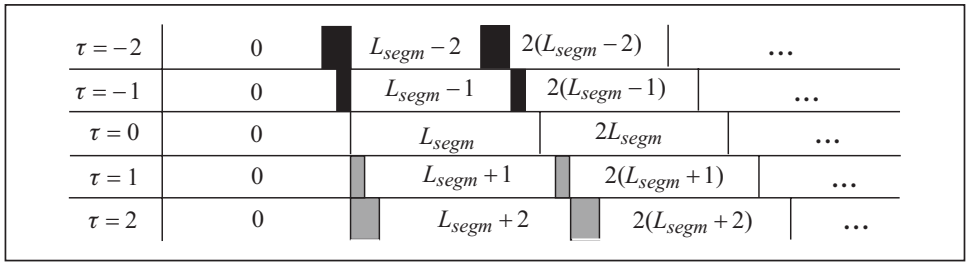

Рис. 1. Схема разбиения выборки на отрезки с перекрытием и растягиванием

Оценка нормированного коэффициента корреляции $r_i(\tau)$ вычисляется для отрезков с номерами $i\cdot(L_{segm}+\tau)$ и $(i+1)\cdot(L_{segm}+\tau)$ следующим образом:

$$r_i(\tau)=\frac{\frac{1}{L_{segm}}\sum\limits_{l=0}^{L_{segm}-1}(\overset{\circ}{x}(i\cdot(L_{segm}+\tau)+l)\cdot\overset{\circ}{x}((i+1)\cdot(L_{segm}+\tau)+l))}{\sqrt{D(X(i\cdot(L_{segm}+\tau)))}\cdot\sqrt{D(X((i+1)\cdot(L_{segm}+\tau)))}},\qquad(1)$$

где $\overset{\circ}{x}(j+l)=x(j+l)-M(X(j))$ — реализация центрированной случайной величины (с.в.), находящаяся на позиции $(j+l)$, $0\le l\le L_{segm}-1$; $x(j+l)$ — $(j+l)$-й элемент выборки, $0\le(j+l)\le L_{seq}-1$; $M(X(j))$ и $D(X(j))$ — соответственно математическое ожидание и дисперсия с.в. $X(j)$, реализациями которой являются элементы j-го отрезка $\{x(j),x(j+1),\dots,x(j+L_{segm}-1)\}$.

Таким образом, для любой выборки длины L_{seq} можно вычислить

$$i_{\max}(\tau)=entier((L_{seq}+\tau)/(L_{segm}+\tau))-1\qquad(2)$$

оценок нормированных коэффициентов корреляции $r_i(\tau)$ в соответствии с (1) (функция $entier(a)$ определяет целую часть (антье) числа a), а $i\in[0,\,i_{\max}(\tau)-1]$.

В данной работе ограничимся рассмотрением равномерно распределенных последовательностей со значениями на отрезке $[0,\,M-1]$, где M — мощность алфавита. Тогда $\forall j\in[0,\,L_{seq}-L_{segm}+1]$: $M(X(j))=(M-1)/2$, $D(X(j))=(M^2-1)/12$.

Пусть для всех используемых источников $M=256$, $L_{seq}=256256$, $L_{segm}=256$. Такой выбор при $\tau=0$ позволяет вычислять $i_{\max}(0)=1000$ оценок $r_i(0)$ по (1). Заметим, что при обозначенных параметрах $\forall j\in[0,256001]$: $M(X(j))=127.5$, $D(X(j))=5461.25$.

АНАЛИЗ k-ГРАММ ЗНАКОВ ОЦЕНОК НОРМИРОВАННЫХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ

Для анализа последовательности знаков $\{z_i(\tau)\}$ оценок нормированных коэффициентов корреляции $r_i(\tau)$ выполним преобразование $z_i(\tau)=\operatorname{sign}(\operatorname{sign}(r_i(\tau))+0.5)$, которое позволит получить множество значений $\{z_i(\tau)\}$, состоящее из двух элементов: -1 и $+1$.

Поскольку знаменатель в (1) положительный, вместо оценки коэффициента корреляции $r_i(\tau)$ можно использовать оценку коэффициента ковариации $\operatorname{cov}_i(\tau)$ или $C_i(\tau)=L_{segm}\cdot\operatorname{cov}_i(\tau)$:

$$C_i(\tau)=\sum_{l=0}^{L_{segm}-1}[\overset{\circ}{x}(i\cdot(L_{segm}+\tau)+l)\cdot\overset{\circ}{x}((i+1)\cdot(L_{segm}+\tau)+l)].\qquad(3)$$

Тогда последовательность знаков

$$\{z_i(\tau)\}=\operatorname{sign}(\operatorname{sign}\{C_i(\tau)\}+0.5)=\operatorname{sign}(\operatorname{sign}\{r_i(\tau)\}+0.5).\qquad(4)$$

Таким образом, при анализе каждой последовательности получим множество значений $\{z_i(\tau)\}$, где $i\in[0,\,i_{\max}(\tau)-1]$, $i_{\max}(\tau)$ вычисляется по (2), а $\tau\ge 1-L_{segm}$.

Для каждой последовательности $\{z_i(\tau)\}$ с фиксированным τ проверим гипотезу о равномерности распределения k-грамм знаков для $k\ge 1$ с помощью статистического критерия χ^2 [13]. Заметим, что в соответствии с рекомендациями Кнута [4] данный анализ требует разбиения последовательности $\{z_i(\tau)\}$ на неперекрывающиеся отрезки из k знаков (k-граммы) вида $\{z_{kj}(\tau),z_{kj+1}(\tau),\dots,z_{kj+k-1}(\tau)\}$. В результате проверки равномерности распределения k-грамм знаков получим множество значений статистики $\{\chi^2(k,\tau)\}$, $k\in[1,K]$.

Подобный анализ выполняется для большого количества N выборок генератора случайных или псевдослучайных чисел. В результате формируется набор значений $\{\chi_n^2(k,\tau)\}$, где $n \in [1, N]$.

Для каждой пары (k,τ) вычислим относительную частоту попадания статистики $\chi_n^2(k,\tau)$ в критическую область $(\chi^2_{1-\alpha,m};\infty)$ при заданном уровне значимости α и $m = 2^k - 1$:

$$W(k,\tau,\alpha) = N_B(k,\tau,\alpha)/N, \tag{5}$$

где $N_B(k,\tau,\alpha)$ — число появлений события $B = \{\chi_n^2(k,\tau) > \chi^2_{1-\alpha,m}\}$ в N испытаниях $(n \in [1, N])$.

Величина $N_B(k,\tau,\alpha)$, а значит и $W(k,\tau,\alpha)$, является случайной. Определим интервал $(W_{\min}(\alpha,\gamma), W_{\max}(\alpha,\gamma))$, симметричный относительно математического ожидания с.в. $W(k,\tau,\alpha)$, в который она попадает с вероятностью $P\{W_{\min}(\alpha,\gamma) < W(k,\tau,\alpha) < W_{\max}(\alpha,\gamma)\} = \gamma$.

Для решения данной задачи установим закон распределения относительной частоты попадания статистики $\chi_n^2(k,\tau)$ в критическую область $(\chi^2_{1-\alpha,m};\infty)$. Для этого вычислим значения $W(k,\tau,\alpha)$ при растягивании отрезков $(0 \le \tau \le 255)$ для $N = 1000$ последовательностей атмосферных шумов. Поскольку непересекающиеся отрезки стационарного белого шума некоррелированы между собой, распределение $W(k,\tau,\alpha)$ инвариантно по отношению к величине $\tau \ge 0$. Графики полигонов частот значений $W(k,\tau,\alpha)$, вычисленных для $\tau \in [0, 255]$, в зависимости от некоторых параметров k и α представлены на рис. 2, где штриховой линией выполнены графики при нормальном законе распределения с математическим ожиданием α и дисперсией $\alpha(1-\alpha)/N$.

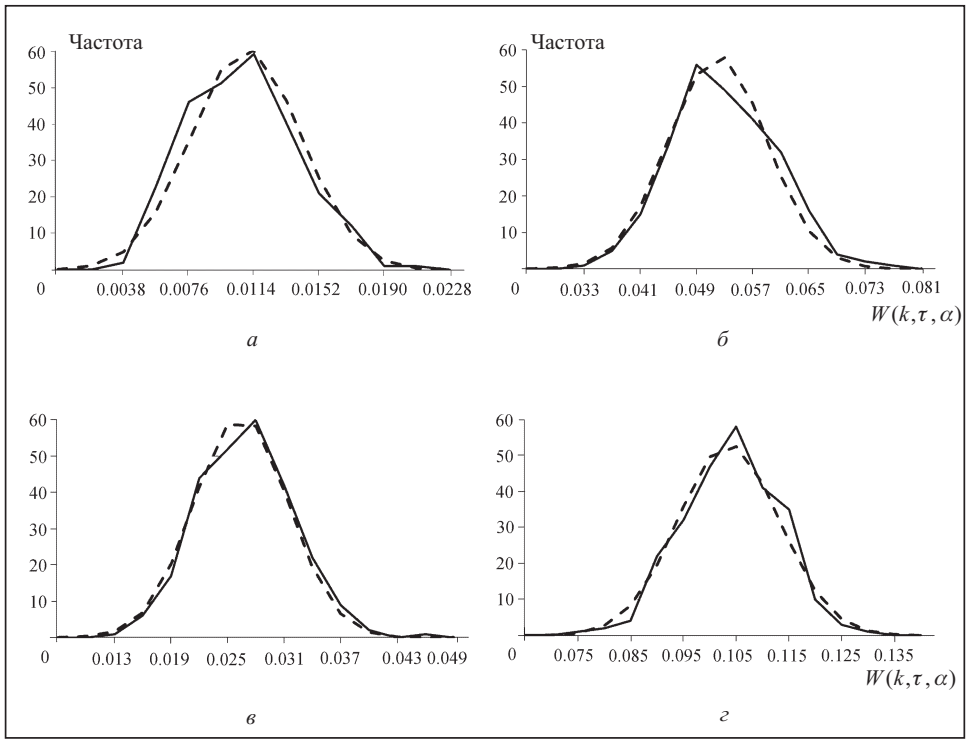

Рис. 2. Графики полигонов частот попадания статистики $\chi^2(k,\tau)$ в критическую область для $k = 1$ и $\alpha = 0.01$ (*а*); $k = 1$ и $\alpha = 0.05$ (*б*); $k = 2$ и $\alpha = 0.025$ (*в*); $k = 2$ и $\alpha = 0.1$ (*г*)

Таблица 1. Результаты проверки гипотезы о нормальности распределения относительной частоты попадания статистики $\chi_n^2(k,\tau)$ в критическую область

Уровень значимости α	$\alpha = 0.01$		$\alpha = 0.025$		$\alpha = 0.05$		$\alpha = 0.1$	
Длина k-граммы	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
p-value	0.445	0.495	0.676	0.798	0.287	0.808	0.391	0.451

Соответствие эмпирических распределений значений $W(k,\tau,\alpha)$ для атмосферного шума теоретическому предельному (нормальному) подтверждается результатами проверки по критерию χ^2 Пирсона [13], о чем свидетельствуют достигнутые уровни значимости (p-value), приведенные в табл. 1.

Проверка соответствия эмпирического распределения значений функции $W(k,\tau,\alpha)$ (при $\tau \geq 0$ и фиксированных α и k) нормальному закону с математическим ожиданием α и дисперсией $\sigma^2 = \alpha(1-\alpha)/N$ является одним из этапов разрабатываемого метода тестирования. Непрохождение данного этапа приводит к непрохождению всего теста. Отметим, что квантовый ГСЧ [20] и «Вихрь Мерсенна» MT19937 [10] успешно преодолевают данный этап.

В силу того, что для белого шума с.в. $W(k,\tau,\alpha)$ является нормально распределенной, для определения вероятности ее попадания в интервал $(W_{\min}(\alpha,\gamma), W_{\max}(\alpha,\gamma))$ используется выражение

$$\Phi((P_{\max}(\alpha,\gamma)-\alpha)/\sigma) - \Phi((P_{\min}(\alpha,\gamma)-\alpha)/\sigma),$$

где $\Phi(x)$ — функция Лапласа. Тогда нижний $W_{\min}(\alpha,\gamma)$ и верхний $W_{\max}(\alpha,\gamma)$ пределы интервала имеют вид:

$$W_{\min}(\alpha,\gamma) = \alpha - t_{(1+\gamma)/2}\sqrt{\alpha(1-\alpha)/N},\ W_{\max}(\alpha,\gamma) = \alpha + t_{(1+\gamma)/2}\sqrt{\alpha(1-\alpha)/N}, \quad (6)$$

где $t_{(1+\gamma)/2}$ — квантиль стандартного нормального распределения с уровнем $(1+\gamma)/2$.

ВЫЧИСЛЕНИЕ БАРЬЕРНОЙ ФУНКЦИИ

Рассмотрим статистические особенности при $\tau \in [1-L_{segm}, -1]$.

Определение 1. Барьерной функцией $\tau_{\min}(k,\alpha,\gamma)$ будем называть случайную функцию параметра α, равную минимальному значению $-\tau$, $\tau < 0$, при котором относительная частота $W(k,\tau,\alpha)$ попадания статистики $\chi_n^2(k,\tau)$ в критическую область $(\chi^2_{1-\alpha,m};\infty)$ превышает верхнюю границу $W_{\max}(\alpha,\gamma)$:

$$\tau_{\min}(k,\alpha,\gamma) = \min(-\tau > 0 : W(k,\tau,\alpha) > W_{\max}(\alpha,\gamma)). \quad (7)$$

Статистический смысл барьерной функции $\tau_{\min}(k,\alpha,\gamma)$ состоит в том, что при фиксированных значениях (k,α,γ) она выявляет «порог» $-\tau_{\min}$, выше которого $(\tau > -\tau_{\min})$ наблюдается равномерное распределение k-грамм знаков эмпирической корреляционной функции (1) исследуемой последовательности случайных чисел.

Для нормально распределенной с.в. $W(k,\tau,\alpha)$ справедливы следующие утверждения:

— при $\alpha' < \alpha'' < 0.5$: $\chi^2_{1-\alpha',m} > \chi^2_{1-\alpha'',m}$, поэтому $W(k,\tau,\alpha') \leq W(k,\tau,\alpha'')$;

— при $\alpha' < \alpha'' < 0.5$: $\alpha'(1-\alpha') < \alpha''(1-\alpha'')$, следовательно, $W_{\max}(\alpha',\gamma) < W_{\max}(\alpha'',\gamma)$, а $W_{\max}(\alpha',\gamma) - W_{\min}(\alpha',\gamma) < W_{\max}(\alpha'',\gamma) - W_{\min}(\alpha'',\gamma)$.

Иными словами, при увеличении уровня значимости α критическая область $(\chi^2_{1-\alpha,m}; \infty)$ расширяется. Следовательно, относительная частота попадания статистики $\chi^2_n(k,\tau)$ в критическую область не уменьшается (при $N \to \infty$ неравенство приобретает вид строгого). При увеличении уровня значимости расширяется также интервал $(W_{\min}(\alpha,\gamma), W_{\max}(\alpha,\gamma))$.

Замечание 1. Обе части неравенства $W(k,\tau,\alpha) > W_{\max}(\alpha,\gamma)$ из (7) монотонно возрастающие по α. В то же время $\tau_{\min}(k,\alpha,\gamma)$ может быть не монотонной по α.

КРИТЕРИЙ БАРЬЕРНОЙ ФУНКЦИИ И ЕГО ПРИМЕНЕНИЕ

Реализация барьерной функции. Исследуем поведение барьерной функции $\tau_{\min}(k,\alpha,\gamma)$ в зависимости от уровня значимости α. Для этого при фиксированном γ, $k \in [1, K]$ и $N = 1000$ вычислим последовательность значений $\tau_{\min}(k,\alpha,\gamma)$ для $\alpha \in [\alpha_1, \alpha_2]$ с шагом $\Delta\alpha = (\alpha_2 - \alpha_1)/N_\alpha$, где $(N_\alpha + 1)$ — количество значений α.

Пусть $\alpha \in [0.01, 0.1]$, $N_\alpha = \{20, 30, 45, 90\}$, $\gamma = 0.95$, а $K = 2$. Графики зависимостей значений функции $\tau_{\min}(k,\alpha,\gamma)$ от $\alpha \in [0.01, 0.1]$ для $N_\alpha = \{20, 45\}$ и $k = 1$ представлены на рис. 3, где видны существенные отличия для различных источников. Примем точки на графиках реализациями некоторых с.в. и вычислим для них выборочные дисперсии. Полученные результаты (табл. 2) показывают, что дисперсия $D(\tau_{\min}(k,\alpha,\gamma))$ минимальна для атмосферного шума и значительно меньшая, чем у других источников.

Таблица 2. Выборочные дисперсии $D(\tau_{\min}(k,\alpha,\gamma))$ в зависимости от N_a и k

Источник	$N_a = 20$		$N_a = 30$		$N_a = 45$		$N_a = 90$	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
Атмосферный шум	0.43	2.03	2.77	2.48	1.69	1.99	1.48	2.03
Квантовый ГСЧ	27.53	3.73	25.10	5.91	23.35	5.91	25.63	5.86
Генератор MT19937	36.16	9.23	26.12	9.21	30.42	9.42	30.06	9.13

Критерий барьерной функции. Для разработки критерия, характеризующего постоянство барьерной функции, используем статистическую методологию интерпретации зависимостей $\tau_{\min}(k,\alpha,\gamma)$. Применяемый метод основан на идеологии критерия χ^2 Пирсона и состоит в том, что для количественной оценки наблюдаемых статистических нерегулярностей полученные последовательности значений барьерной функции $\tau_{\min}(k,\alpha,\gamma)$ при $\alpha \in [\alpha_1, \alpha_2]$ с шагом $\Delta\alpha$ и фиксированными γ и k рассматриваются как частоты при реализации статистического эксперимента с $N_\alpha + 1$ возможными исходами. В этом случае графики (см. рис. 3) являются полигонами частот дискретных с.в.

Поскольку в качестве источника эталонного белого шума выбран источник атмосферного шума [19], а также в соответствии с видом соответствующих ему на рис. 3 полигонов частот, теоретическим законом распределения является равномерный закон.

Таким образом, критерий барьерной функции предусматривает следующую проверку. Эмпирическое распределение с относительными частотами $\varphi_{k,\gamma}(\alpha) = \tau_{\min}(k,\alpha,\gamma) / \sum_{\alpha=\alpha_1}^{\alpha_2} \tau_{\min}(k,\alpha,\gamma)$ сравнивается с теоретическим предельным (равномерным) по критерию χ^2 Пирсона при заданном уровне значимости β. Иссле-

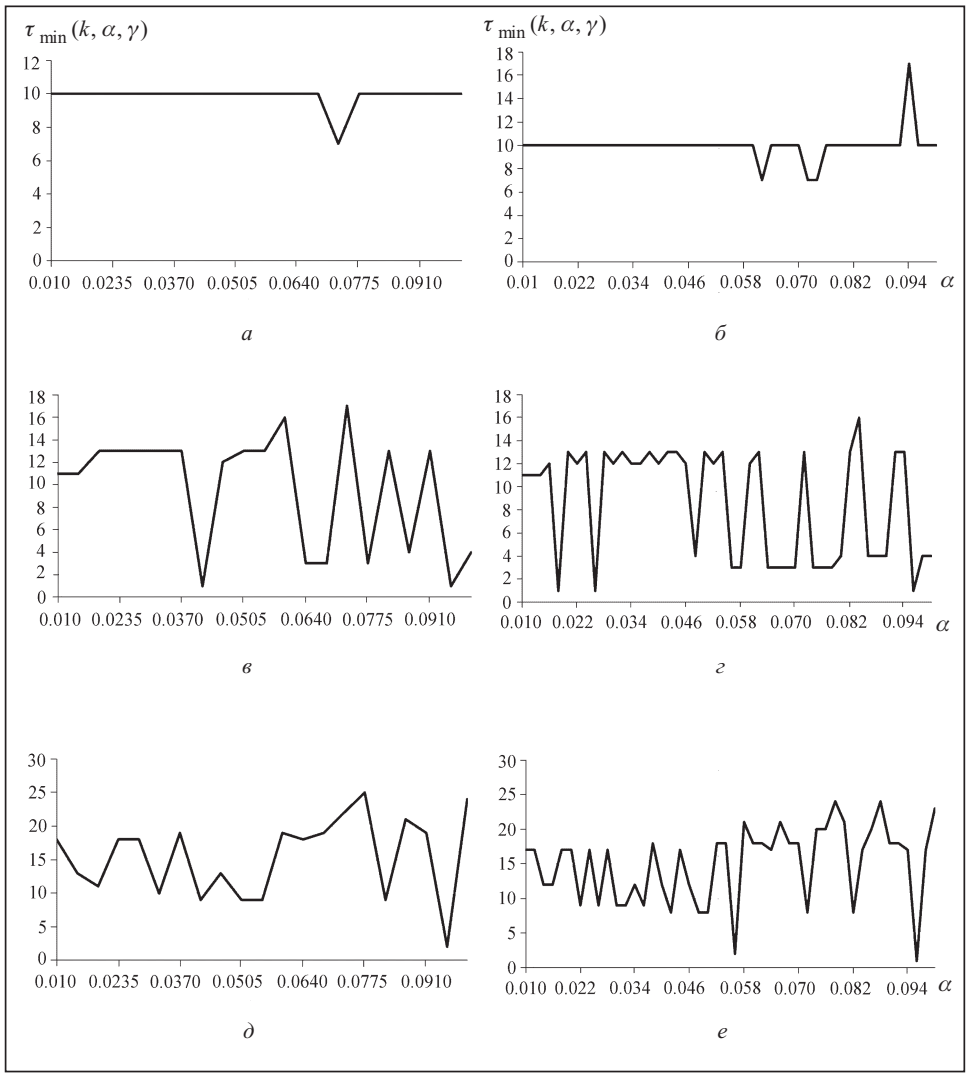

Рис. 3. Графики зависимостей $\tau_{\min}(k, \alpha, \gamma)$ от $\alpha \in [0.01, 0.1]$ при $k = 1$ для: атмосферного шума при $N_\alpha = 20$ (*а*) и $N_\alpha = 45$ (*б*); квантового ГСЧ при $N_\alpha = 20$ (*в*) и $N_\alpha = 45$ (*в*); генератора MT19937 при $N_\alpha = 20$ (*д*) и $N_\alpha = 45$ (*е*)

дуемый источник удовлетворяет критерию барьерной функции, если для всех $k \in [1, K]$ полученные статистики χ^2 не попадают в критическую область.

Применение критерия барьерной функции. Для исследуемых источников проверим соответствие полученных эмпирических распределений с относительными частотами $\varphi_{k,\gamma}(\alpha)$ равномерному закону по критерию χ^2 Пирсона. Достигнутые уровни значимости для $H_0: \{\varphi_{k,\gamma}(\alpha) = 1/(N_\alpha + 1)\}$ приведены в табл. 3.

Полученные результаты свидетельствуют о том, что равномерному закону соответствует распределение с абсолютными частотами $\tau_{\min}(k, \alpha, \gamma)$ только для последовательностей атмосферного шума. Квантовый ГСЧ и генератор MT19937 не удовлетворяют разработанному критерию. Непрохождение теста квантовым ГСЧ может являться следствием использования неоптимального преобразования квантовых явлений в последовательность случайных чисел.

Таблица 3. Результаты проверки гипотезы о равномерности эмпирического распределения с частотами $\tau_{\min}(k,\alpha,\gamma)$ при $a \in [0.01, 0.1]$ и $\gamma = 0.95$ в зависимости от N_a и k

Источник	Достигнутый уровень значимости (*p*-value)							
	$N_a = 20$		$N_a = 30$		$N_a = 45$		$N_a = 90$	
	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$	$k = 1$	$k = 2$
Атмосферный шум	1.0000	0.9977	0.9999	0.9989	1.0000	1.0000	1.0000	1.0000
Квантовый ГСЧ	0.0000	0.7538	0.0000	0.4915	0.0000	0.4446	0.0000	0.4744
Генератор MT19937	0.0004	0.0109	0.0164	0.0037	0.0000	0.0005	0.0000	0.0000

Заметим, что выбор значения N_α не оказывает существенного влияния на результаты эксперимента. Поэтому N_α можно выбирать как минимальное значение, при котором правомерно использование критерия χ^2 для проверки равномерности дискретного распределения: $\sum\limits_{\alpha} \tau_{\min}(k,\alpha,\gamma) > 100$ или $N_\alpha = \min(N_\alpha) : (N_\alpha + 1) \times \times \overline{\tau_{\min}}(k,\alpha,\gamma) > 100$, где $\overline{\tau_{\min}}(k,\alpha,\gamma) = \left(\sum\limits_{a} \tau_{\min}(k,\alpha,\gamma) \right) / (N_a + 1)$ — выборочное среднее.

МЕТОДИКА ПРИМЕНЕНИЯ КРИТЕРИЯ

Методика применения разработанного критерия включает следующие этапы:

— анализируются N последовательностей генератора ($N \geq 1000$);

— каждая последовательность длины L_{seq} ($L_{seq} = 256256$) разбивается на отрезки длины L_{segm} ($L_{segm} = M = 256$), расстояние между соседними отрезками $\tau \geq 1 - L_{segm}$, количество отрезков равно $i_{\max}(\tau) + 1$, $i_{\max}(\tau)$ вычисляется по (2);

— для каждой последовательности вычисляется множество оценок $\{r_i(\tau)\}$ (1) или $\{C_i(\tau)\}$ (3), $i \in [0, i_{\max}(\tau) - 1]$;

— множество $\{r_i(\tau)\}$ преобразуется в множество $\{z_i(\tau)\}$ (4);

— с помощью критерия χ^2 для последовательностей $\{z_i(\tau)\}$ при фиксированном τ проверяется гипотеза о равномерности распределения k-грамм для $k \in [1, K]$ ($K = 2$), в результате чего формируется множество значений статистики $\{\chi^2(k,\tau)\}$ для одной последовательности и $\{\chi^2_n(k,\tau)\}$ ($n \in [1, N]$) — для N последовательностей;

— для каждой пары (k,τ) вычисляется функция $W(k,\tau,\alpha)$ (5) при $\alpha \in [\alpha_1, \alpha_2]$ с шагом $\Delta\alpha = (\alpha_2 - \alpha_1)/N_\alpha$, $\alpha_1 = 0.01$, $\alpha_2 = 0.1$, $N_\alpha \geq 20$;

— с помощью критерия χ^2 проверяется соответствие эмпирического распределения значений функции $W(k,\tau,\alpha)$ для $\tau \geq 0$ (при фиксированных α и k) нормальному закону с математическим ожиданием α и дисперсией $\sigma^2 = \alpha(1-\alpha)/N$ (непрохождение данного этапа приводит к непрохождению всего теста);

— для каждого α определяется интервал $(W_{\min}(\alpha,\gamma), W_{\max}(\alpha,\gamma))$ (6);

— вычисляются значения барьерной функции $\tau_{\min}(k,\alpha,\gamma)$ (7);

— для вычисленных значений $\tau_{\min}(k,\alpha,\gamma)$ применяется критерий барьерной функции;

— тест пройден, если для всех $k \in [1, K]$ полученные статистики критерия барьерной функции не попадают в критическую область при заданном уровне значимости β.

ЗАКЛЮЧЕНИЕ

В результате проведенного исследования разработаны новые метод и критерий оценивания качества последовательностей равномерно распределенных случайных и псевдослучайных чисел, позволяющие выявлять в них не обнаруженные до настоящего времени статистические нерегулярности. Определено, что только выбранный в качестве эталона атмосферный шум [19] соответствует разработанному критерию, а квантовый ГСЧ [20] и генератор MT19937 [10] не удовлетворяют предъявляемым требованиям.

Полученные результаты имеют большое теоретическое и практическое значение при оценивании качества последовательностей случайных чисел и их можно использовать в специализированных прикладных пакетах тестирования технических средств формирования последовательностей случайных и псевдослучайных чисел.

СПИСОК ЛИТЕРАТУРЫ

1. Brown R.G. Introduction to random signal analysis and Kalman filtering. — New York: John Wiley and Sons, 1983. — 347 p.
2. Papoulis A. Probability, random variables, and stochastic processes. — 3rd ed. — New York: McGraw-Hill, 1991. — 666 p.
3. Иванов М.А., Чугунков И.В. Теория, применение и оценка качества генераторов псевдослучайных последовательностей. — М.: КУДИЦ-ОБРАЗ, 2003. — 240 с.
4. Knuth D.E. The art of computer programming: In 7 vol. Vol. 2: Seminumerical algorithms. — 3rd ed. — Reading (Massachusetts): Addison Wesley Longman Publishing Co., 1998. — 762 p.
5. Поточные шифры. Результаты зарубежной открытой криптологии. — http://www.ssl.stu.neva.ru/psw/crypto/potok/str_ciph.htm.
6. Robshaw M.J.B. Stream ciphers. Technical Report TR-701, V. 2.0. — RSA Lab., 1995. — 42 p.
7. Marsaglia G. DIEHARD battery of tests of randomness. — http://www.stat.fsu.edu/pub/diehard.
8. Rukhin A. et al. A statistical test suite for random and pseudorandom number generators for cryptographic applications. Spec. Pub. 800-22 rev. 1a. — Gaithersburg, MD, NIST, 2010. — 131 p.
9. L'ecuyer P., Simard R. TestU01: A C library for empirical testing of random number generators // ACM TOMS. — 2007. — **33**, N 4. — Article 22. — 40 p.
10. Matsumoto M., Nishimura T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator // ACM TOMACS. — 1998. — **8.** — P. 3–30.
11 Soto J. Statistical testing of random number generators // Proceedings of the 22nd National Information Systems Security Conference. — Gaithersburg, MD: NIST, 1999. — **10.** — 12 p.
12. Collins J.C. Testing, selection, and implementation of random number generators. Rep. ARL-TR-4498. — Aberdeen Proving Ground, MD: U.S. Army Research Laboratory, 2008. — 76 p.
13. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1969. — 512 с.
14. Kendall M.G. The advanced theory of statistics. — V. II. — London: C. Griffin&Co, 1946. — 521 p.
15. Dirac P.A.M. The principles of quantum mechanics. — 4th ed. — London: OUP, 1958. — 314 p.
16. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — 3-е изд. — М.: Наука, 1983. — 416 с.
17. Фауре Э.В., Щерба А.И., Лавданский А.А. Анализ корреляционных свойств последовательностей (псевдо) случайных чисел // Наука і техніка Повітряних Сил Збройних Сил України. — 2015. — № 1 (18) — С. 142–150.
18. Лавданский А.А., Фауре Э.В. Комбинационный метод формирования последовательности псевдослучайных чисел // Системний аналіз та інформаційні технології: Матеріали 16-ї Міжнародної науково-технічної конференції SAIT-2014, Київ, 26–30 травня 2014 р. ННК «ІПСА» НТУУ «КПІ». — К.: ННК «ІПСА» НТУУ «КПІ», 2014. — С. 403–404.
19. Haahr M. True random number service. — http://random.org/.
20. QRNG Service. — http://qrng.physik.hu-berlin.de/.

Поступила 22.09.2015