Математические методы исследования

УДК 519.28

ИССЛЕДОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

© И. А. Кожевникова[1]

Статья поступила 29 апреля 2010 г.

Предложен метод исследования последовательностей псевдослучайных чисел на основании максимального отклонения оценки спектральной плотности, полученной осреднением по сдвигу во времени. При отсутствии временны́х зависимостей внутри таких последовательностей оценки их спектральной плотности должны быть близки к постоянной — спектральной плотности последовательности независимых одинаково распределенных случайных величин. Методом математического моделирования проведено сравнение глобальной и локальной проверок для датчика псевдослучайных чисел URAND. Проведенные расчеты показывают, что гипотеза независимости последовательности не отвергается на разумном уровне значимости либо для последовательностей, полученных с помощью прореживания, либо с помощью М-алгоритма.

Ключевые слова: периодограмма; оценка спектральной плотности с временны́м сдвигом; последовательность псевдослучайных чисел; белый шум 4-го порядка.

Рассмотрим метод исследования близости последовательностей псевдослучайных чисел и независимых одинаково распределенных случайных величин $\varepsilon(t)$ на основании оценки спектральной плотности. Известно, что спектральная плотность последовательности независимых одинаково распределенных случайных величин является постоянной на отрезке $[-\pi, \pi]$. Датчики псевдослучайных чисел находят широкое применение в прикладных исследованиях, проводимых методами математического моделирования.

Основная задача заключается в том, чтобы на выбранном уровне значимости проверить гипотезу о том, является ли последовательность псевдослучайных чисел последовательностью независимых и одинаково распределенных случайных величин.

Доверительные границы для диапазона частот $0 \le \lambda \le \pi$ спектральной плотности по выборкам больших длин N строятся с использованием оценки с временны́м сдвигом [1]. Для этого выборка длиной N разбивается на T непересекающихся отрезков длиной n. Оценка спектральной плотности с временны́м сдвигом имеет вид

$$\bar{f}_N(\lambda) = \frac{1}{T}\sum_{k=1}^{T} I_{n,k}(\lambda), \qquad (1)$$

где

$$I_{n,k}(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \varepsilon(k,t) e^{i\lambda t}\right|^2,$$

$$|\lambda| \le \pi, \ \varepsilon(k, t) = \varepsilon[t + (k-1)n] \ — \qquad (2)$$

периодограмма, вычисленная на k-м отрезке; $\varepsilon(t)$ — исследуемая последовательность псевдослучайных чисел; λ — частота в радианах. Предположим, что длина отрезка n постоянна, но большая, а число отрезков $T \to \infty$. При решении других задач, например оценивания спектральной плотности по выборке небольшой длины, допускается пересечение отрезков [2]. Нижняя граница длины отрезка разбиения будет определена методом математического моделирования.

В работе [3] только для мультипликативного датчика псевдослучайных чисел RANDU [4] на основании теоретически вычисленной периодограммы для алгоритма датчика и формулы (1) была проверена гипотеза независимости H_0. Поскольку оценка (1) имеет асимптотически нормальное распределение [5], то на основании теории стационарных гауссовских процессов с дискретным временем [6] были вычислены верхняя и нижняя границы множества на уровне значимости 0,05. Гипотеза H_0 на основании вычислений [3] была отвергнута.

Основным результатом этой работы является теорема 4, в которой была получена формула предельного распределения экстремальных значений при условии, что исходный процесс имеет гауссовское распределение только асимптотически.

Покажем основные свойства периодограммы (2). Приведенная ниже теорема 1 [7] устанавливает ее асимптотическую несмещенность.

Теорема 1. Пусть $X(t)$, $t = 0, \pm1, \pm2, \ldots$, — действительный временно́й ряд со спектральной плотностью $f(\lambda)$, математическим ожиданием $EX(t) = c_X$ и ковариационной функцией $\operatorname{cov}[X(t), X(t+u)] = c_{XX}(u)$. Предположим, что ковариационная функция удовлетворяет условию

$$\sum_{u=-\infty}^{\infty} |u c_{XX}(u)| < \infty,$$

[1] Московский государственный университет им. М. В. Ломоносова, Москва, Россия; e-mail: irina_kozhev@mail.ru

тогда для любого отрезка с номером k, $k = 1, 2, ..., T$, получим

$$EI_{n,k}(\lambda) = f(\lambda) + \frac{1}{2\pi n}\left(\frac{\sin\frac{n\lambda}{2}}{\sin\frac{\lambda}{2}}\right)^2 c_X^2 + O\left(\frac{1}{n}\right).$$

Заметим, что при $\lambda = 2\pi s/n$, s — целое число, $s \neq jn$, $j = 0, \pm 1, \pm 2, ...$, второе слагаемое в правой части последнего выражения обращается в нуль и периодограмма, вычисленная только в точках $\lambda = 2\pi s/n$, является асимптотически несмещенной оценкой спектральной плотности. Применение любого спектрального окна осреднения по частотам лишает ее этого свойства [7]. Ковариационную структуру периодограммы в точках $\lambda = 2\pi s/n$ описывает теорема 2.

Теорема 2. Пусть $X(t)$, $t = 0, \pm 1, \pm 2, ...,$ — действительный временнóй ряд со спектральной плотностью $f(\lambda)$, причем $r \pm s \neq jn$, $j = 0, \pm 1, \pm 2, ...$; $\mu = 2\pi r/n$; $\lambda = 2\pi s/n$. Тогда

$$DI_{n,k}(\lambda) = f^2(\lambda) + O(1/n), \tag{3}$$

$$\mathrm{cov}[I_{n,k}(\mu), I_{n,k}(\lambda)] = O(1/n),$$

$$0 < \lambda < \pi,\ 0 < \mu < \pi,\ \lambda \neq \mu. \tag{4}$$

Эта теорема устанавливает два важных для статистической практики факта: во-первых, дисперсия периодограммы будет стремиться к постоянному уровню $f^2(\lambda)$, какой бы большой ни была длина выборки n, во-вторых, периодограммы, вычисленные в различных точках, асимптотически не коррелированы. Из теоремы 2 также следует, что ординаты статистик асимптотически не коррелированы. Используя выражения (3) и (4), получим

$$\sigma^2 = D\bar{f}_N(\lambda) = \frac{f^2(\lambda)}{T},\ \lambda \neq 0,\ \lambda \neq \pi. \tag{5}$$

Из этой формулы следует, что доверительный интервал оценки спектральной плотности, построенный на основании оценки (1) при фиксированном значении λ, будет в $\sqrt{T}$ раз уже доверительного интервала, построенного на основании периодограммы (2).

Введем некоторые понятия и определения, необходимые для изложения дальнейшего материала.

Определение 1. Семиинвариантом $S_k(t_1 \cdot t_2 \cdot ... \cdot t_k)$ порядка k называется коэффициент при слагаемом $(i)^k$ $t_1 \cdot t_2 \cdot ... \cdot t_k$ в разложении логарифма характеристической функции $\ln\left[E\exp\left(i\sum_{j=1}^{m}\xi_j t_j\right)\right]$ случайных величин $\xi_1, \xi_2, ..., \xi_m$ в ряд Тейлора в окрестности начала координат.

При доказательствах теорем и лемм потребуются семиинварианты до четвертого порядка включительно.

Определение 2. Последовательность некоррелированных и одинаково распределенных случайных величин $\varepsilon(t)$, $t = 0, 1, ..., E\varepsilon(t) = 0$, $D\varepsilon(t) = 1$, причем $E(\varepsilon(t))^4$ равномерно ограничены, называется белым шумом четвертого порядка, если все смешанные семиинварианты до четвертого порядка равны нулю.

В частности, некоррелированная стационарная в широком смысле последовательность является белым шумом второго порядка со спектральной плотностью $f(\lambda) = 1/2\pi$.

Используя выражение периодограммы через оценку ковариационной функции, для каждого фиксированного λ, $0 \leq |\lambda| \leq \pi$ представим $2\pi\bar{f}_N(\lambda) - 1$ в виде суммы некоррелированных одинаково распределенных случайных величин и слагаемого, которое стремится к 0 при $N \to \infty$:

$$2\pi\bar{f}_N(\lambda) - 1 = \frac{1}{T}\sum_{k=1}^{T} Z_{n,k}(\lambda) + $$

$$+ \frac{1}{Tn}\sum_{k=1}^{T}\sum_{t=0}^{n-1}[\varepsilon^2(t + n(k-1)) - 1], \tag{6}$$

$$Z_{n,k}(\lambda) = \frac{2}{n}\sum_{\tau=1}^{n-1}\sum_{t=0}^{n-1-\tau}\varepsilon(k,t)\varepsilon(k,t+\tau)e^{i\lambda\tau}, \tag{7}$$

где $\varepsilon(k, t)$ определено в формуле (2). Обозначим

$$U_N(\lambda) = \frac{1}{T}\sum_{k=1}^{T} Z_{n,k}(\lambda),$$

$$r_N = \frac{1}{Tn}\sum_{k=1}^{T}\sum_{t=0}^{n-1}[\varepsilon^2(t + n(k-1)) - 1]. \tag{8}$$

Из формулы (7) следует, что случайные величины $Z_{n,k}(\lambda)$ имеют одно и то же распределение при различных значениях $k = 1, ..., T$. Ниже приведены леммы, устанавливающие важные свойства случайных величин $Z_{n,k}(\lambda)$, а также то, что при разных значениях k случайные величины $Z_{n,k}(\lambda)$, $0 \leq |\lambda| \leq \pi$, асимптотически не коррелированы.

Лемма 1. Если последовательность $\varepsilon(t)$ — белый шум четвертого порядка, то при каждом фиксированном λ моменты случайных величин $Z_{n,k}(\lambda)$, определенных формулой (7), выражаются следующим образом:

$$EZ_{n,k}(\lambda) = 0,\ 0 \leq |\lambda| \leq \pi, \tag{9}$$

$$DZ_{n,k}(\lambda) = 1 - \frac{4}{n} + \frac{\sin^2 n\lambda}{n^2 \sin^2 \lambda},\ 0 \leq |\lambda| \leq \pi, \tag{10}$$

$$E\{Z_{n,k}(\lambda)Z_{n,s}(\lambda)\} = 0,$$

$$0 \leq |\lambda| \leq \pi,\ k \neq s,\ k, s = 1, ..., T. \tag{11}$$

Если $\frac{\ln n}{n} \le |\lambda_1| \le \pi$, $i = 1, 2$, то при тех же условиях выполняются следующие неравенства:

$$\begin{cases} |E(Z_{n,k}(\lambda_1)Z_{n,k}(\lambda_2))| < 1, \frac{1}{n} \le |\lambda_1 - \lambda_2| \le \frac{\ln n}{n}, \\ |E(Z_{n,k}(\lambda_1)Z_{n,k}(\lambda_2))| \le \frac{1}{\ln^2 n}, \frac{\ln n}{n} \le |\lambda_1 - \lambda_2| \le \pi. \end{cases} \quad (12)$$

Лемма доказывается вычислением моментов и устанавливает, что вторые моменты выборочных функций $Z_{n,k}(\lambda)$ равномерно ограничены. Это следует из формулы (10) и неравенства $\frac{\sin^2 n\lambda}{n^2 \sin^2 \lambda} \le 1$, $0 \le |\lambda| \le \pi$.

Лемма 2. Если последовательность $\varepsilon(t)$ — белый шум четвертого порядка, то четвертый семиинвариант выборочных функций $Z_{n,k}(\lambda)$ имеет следующий порядок:

$$S_4\{Z_{n,k}(\lambda)\} = O\left(\frac{\ln^2 n}{n}\right), \ 0 \le |\lambda| \le \pi.$$

Из лемм 1 и 2 следует, что при различных значениях k выборочная функция $U_N(\lambda)$, определенная формулой (8), является суммой асимптотически нормальных, независимых и одинаково распределенных случайных величин $Z_{n,k}(\lambda)$.

Поскольку $EZ_{n,k}(\lambda) = 0$, $0 \le |\lambda| \le \pi$, что установлено в формуле (9), то $EU_N(\lambda) = 0$, $0 \le |\lambda| \le \pi$. Используя выражение (10), получим

$$DU_N(\lambda) = \frac{1}{T^2}\sum_{k,p=1}^{T} E[Z_{n,k}(\lambda)Z_{n,p}(\lambda)] = Q\left(\frac{1}{T}\right),$$

$$0 \le |\lambda| \le \pi, \quad (13)$$

так как при $p \ne k$ величины $Z_{n,k}(\lambda)$ и $Z_{n,p}(\lambda)$ не коррелированы. Кроме того, если $\frac{\ln n}{n} \le \lambda_i \le \pi$, $i = 1, 2$, то из формулы (12) следует

$$E[U_N(\lambda_1)U_N(\lambda_2)] = \begin{cases} O\left(\frac{1}{T}\right), \frac{1}{n} \le |\lambda_1 - \lambda_2| < \frac{\ln n}{n}, \\ O\left(\frac{1}{Tn}\right), \frac{\ln n}{n} \le |\lambda_1 - \lambda_2| \le \pi. \end{cases}$$

Поскольку случайные величины $Z_{n,k}(\lambda)$, $k = 1, 2, ..., T$, одинаково распределены и имеют конечную дисперсию, то при $T \to \infty$ и, следовательно, $N \to \infty$ ($n = \mathrm{const}$) случайная величина $U_N(\lambda)$, определенная в формуле (8), будет иметь асимптотически нормальное распределение с нулевым математическим ожиданием и дисперсией, заданной в формуле (13) [8, 9].

Порядок сходимости к нулю по вероятности остаточного слагаемого r_N в формуле (8) на основании закона больших чисел устанавливает следующая теорема.

Теорема 3. Если последовательность $\varepsilon(t)$ — белый шум четвертого порядка, то для любого $\varepsilon > 0$ при $N \to \infty$ имеем

$$P\left\{\sup_{0\le|\lambda|\le\pi} |2\pi\bar{f}_N(\lambda) - 1 - U_N(\lambda)| > \varepsilon\right\} = O\left(\frac{1}{\sqrt{N}}\right), \quad (14)$$

где выборочные функции $\bar{f}_N(\lambda)$ и $U_N(\lambda)$ определены в формулах (1) и (8) соответственно.

Доказательство. Из выражения (6) следует, что

$$\sup_{0\le|\lambda|\le\pi} |2\pi\bar{f}_N(\lambda) - 1 - U_N(\lambda)| = \frac{1}{Tn}\sum_{k=1}^{T}\sum_{t=0}^{n-1}[\varepsilon^2(t + n(k-1)) - 1]$$

и принимает лишь неотрицательные значения. Найдем дисперсию случайной величины, определенную в правой части последней формулы. Поскольку

$$E\left\{\frac{1}{Tn}\sum_{k=1}^{T}\sum_{t=0}^{n-1}[\varepsilon^2(t + n(k-1)) - 1]\right\} = 0,$$

то

$$\frac{1}{T^2n^2}\sum_{k,p=1}^{T}\sum_{t,s=0}^{n-1} E[\varepsilon(t + (k-1)n)\varepsilon(t + (k-1)n)\times$$

$$\times\varepsilon(t + (p-1)n)\varepsilon(t + (p-1)n)] - 1 =$$

$$= \frac{1}{T^2n^2}\sum_{k,p=1}^{T}\sum_{t,s=0}^{n-1} S_4[\varepsilon(t + (k-1)n)\varepsilon(t + (k-1)n)\times$$

$$\times\varepsilon(t + (p-1)n)\varepsilon(t + (p-1)n)] +$$

$$+\frac{1}{T^2n^2}\sum_{k,p=1}^{T}\sum_{t,s=0}^{n-1}\Big\{E\varepsilon^2(t + (k-1)n)E\varepsilon^2(t + (p-1)n) +$$

$$+2E[\varepsilon(t + (k-1)n)\varepsilon(t + (p-1)n)]^2\Big\} - 1, \quad (15)$$

что следует из формул связи моментов и семиинвариантов [10] для белого шума четвертого порядка. По условию теоремы все смешанные семиинварианты последовательности $\varepsilon(k, t)$ равны 0, а несмешанные семиинварианты до второго порядка включительно постоянны, поэтому

$$D\left(\sup_{0\le|\lambda|\le\pi} |2\pi\bar{f}_N(\lambda) - 1 - U_N(\lambda)|\right) = O\left(\frac{1}{N}\right). \quad (16)$$

Выражение (14) следует из формулы (16) и неравенства Чебышёва. Теорема доказана.

Из формулы (16) можно заключить, что порядок сходимости по вероятности случайной величины r_N к нулю при $N \to \infty$ равен $O\left(\frac{1}{\sqrt{N}}\right)$.

Если случайные величины являются белым шумом четвертого порядка, то доверительные области для диапазона частот определяются на основании асимптотического распределения случайной величины

$$\sup_{0\le\lambda\le\pi}\frac{\bar f_N(\lambda)-f(\lambda)}{\sqrt{D\bar f_N(\lambda)}}=\sup_{0\le\lambda\le\pi}\sqrt{T}(2\pi\bar f_N(\lambda)-1),$$

где $\bar f_N(\lambda)$ и $D\bar f_N(\lambda)$ выражены формулами (1) и (5) соответственно: $f(\lambda)$ — спектральная плотность.

Теорема 4 приведена ниже без доказательства и позволяет проверить гипотезу H_0 на выбранном уровне значимости и таким образом оценить качество последовательности псевдослучайных чисел и ее отклонение от последовательности независимых одинаково распределенных случайных величин на основании отклонения оценки спектральной плотности (1) от постоянного значения. Глобальная проверка гипотезы независимости потребовала доказательства теоремы о том, что если последовательность случайных величин является асимптотически нормальной, то распределение ее максимумов подчиняется распределению $F(x)=\exp[-\exp(-x)]$, $-\infty<x<\infty$.

Условие 1. Пусть выборка длиной N разбита на T непересекающихся отрезков фиксированной длины n, длина отрезка n и число отрезков T являются целочисленными функциями длины N, причем $T\to\infty$ при $N\to\infty$ и при этом

$$\frac{n}{N}\to 0,\quad \frac{\ln T}{n}\to\infty.$$

Теорема 4. Пусть $\varepsilon(t)$, $t=0, 1, \ldots$, $E\varepsilon(t)=0$, $D\varepsilon(t)=1$, — белый шум четвертого порядка, тогда для любого $-\infty<x<\infty$ имеет место равенство

$$\lim_{N\to\infty}P\left\{\frac{\sup\limits_{0\le\lambda\le\pi}\sqrt{T}\left|2\pi\bar f_N(\lambda)-1\right|-b_n}{a_n}\le x\right\}=\exp(-e^{-x}),\quad(17)$$

где

$$a_n=\frac{1}{\sqrt{2\ln\left(\frac{n}{2}\right)}};\quad b_n=\sqrt{2\ln\left(\frac{n}{2}\right)}-\frac{\ln\ln\left(\frac{n}{2}\right)+\ln(2\pi)}{\sqrt{2\ln\left(\frac{n}{2}\right)}};\quad(18)$$

$\bar f_N(\lambda)$ выражена формулой (1); N — длина выборки; T (число непересекающихся отрезков, на которые разбита выборка) и n (длина каждого отрезка) — целочисленные функции длины выборки, удовлетворяющие условию 1.

Значения a_n и b_n, приведенные в формулах (18), определяются точно так же, как и в случае распределения максимального элемента последовательности одинаково нормально распределенных случайных величин. Теорема 4 может быть доказана только при условии, что длина отрезка разбиения n такова, что $O(\ln n)\to\infty$, т.е. n достаточно велико.

Этот результат носит теоретический характер и особенно интересен в тех случаях, когда оценка (1) для последовательности псевдослучайных чисел вычисляется теоретически на основании алгоритма датчика.

Доказательство этой теоремы основано на теореме М. Вудруфа [11] о больших отклонениях многомерных случайных величин.

Теорема 5. Если $\varepsilon(t)$, $t=0, 1, \ldots$, $E\varepsilon(t)=0$, $D\varepsilon(t)=1$, — белый шум четвертого порядка, то для всех значений λ_i, $i=1, 2$, $\ln n/n\le|\lambda_i|\le\pi$, существуют $\delta>0$ и N_0 такие, что при всех $N>N_0$ справедливо следующее неравенство:

$$\sup_{\frac{1}{n}\le|\lambda_1-\lambda_2|\le\pi}\frac{P(|U_N(\lambda_1)|\ge z_N, |U_N(\lambda_2)|\ge z_N)}{1-\Phi(z_N)}\le$$

$$\le B\exp(-\delta z_N),\quad(19)$$

где случайная величина U_N определена в формуле (8); $\Phi(x)$ — стандартная функция нормального распределения; B — некоторая постоянная величина; δ находится из соотношения

$$0<\delta<\frac{1}{4}|1-|E(Z_{n,k}(\lambda_1)Z_{n,k}(\lambda_2))||,\ k=1, \ldots, T,$$

где $Z_{n,k}$ задано в формуле (7).

Подобная теорема была доказана в работе [12] в случае независимых одинаково распределенных случайных величин для классических периодограммных оценок спектральной плотности:

$$f_N(\lambda)=\int_{-\pi}^{\pi}W_N(\lambda-x)I_N(x)\mathrm{d}x,\quad(20)$$

где $W_N(x)$ — некоторая весовая функция; $I_N(x)$ — периодограмма, определенная в формуле (2). В работе [13] задача о распределении максимального уклонения рассматривается для гауссовской стационарной последовательности для оценок вида

$$f_N(\lambda)=\frac{1}{2\pi N}\sum_{s,t=1}^{N}g\left(\frac{s-t}{N}\right)X(s)X(t)e^{i\lambda(t-s)}.\quad(21)$$

Оценки, рассмотренные в работах [12] и [13], используют сглаживающие весовые функции, приводящие к уменьшению дисперсии оценки спектральной плотности, но одновременно увеличивающие ее смещение. Поэтому для многих практических исследований оценки (20) и (21) значительно хуже оценки (1), что было подробно разобрано в работе [5].

Рассмотрим результаты дискуссии «Проблема качества датчиков псевдослучайных чисел», проведенной в журнале «Заводская лаборатория» [14, 16, 17, 21 – 23].

В публикации [14] на основании оценки спектральной плотности с временны́м сдвигом (1), определенной в статье [3], приведена методика выявления временны́х зависимостей в последовательности псевдослучайных чисел, вырабатываемых датчиком URAND [15]. Результатом применения указанного метода является получение критической длины последовательности N_α для заданного уровня значимости α. Это означает, что при $N > N_\alpha$ значения последовательности будут статистически связаны между собой. Гипотеза H_0 проверяется на основании нормального распределения статистики (1) при большом значении числа отрезков T, на которые разбита выборка. В этом состоит отличие методик работ [14] и [3]. Достоинством работы [14] является то, что в исследование включены прореженные последовательности.

Автор работы [16] справедливо считает, что в статье [3] множество, на котором выполняется гипотеза H_0, построенное для диапазона частот, шире соответствующего множества для конкретной частоты. Кроме того, он [16] рекомендует использовать датчики, основанные на М-алгоритме [17].

В комментарии [18, I] кратко повторен результат работы [3]: гипотеза H_0 проверена только для мультипликативного датчика псевдослучайных чисел RANDU. Впоследствии подобный результат был получен для датчика псевдослучайных чисел URAND [15] на уровне значимости 0,15. Вычисления проводились на ЭВМ БЭСМ-6, гипотеза H_0 проверялась на основании предельного распределения экстремальных значений стационарных гауссовских процессов с дискретным временем [6].

В комментарии [18, II] сформулированы важные понятия локальной и глобальной проверок последовательностей псевдослучайных чисел, которые не обсуждались в статье [3]. Под локальной проверкой понимается выявление отклонения оценки спектральной плотности от спектральной плотности последовательности независимых одинаково распределенных случайных величин для одной конкретной частоты или для фиксированного ряда частот. Глобальная проверка должна быть выполнена для всех частот, например, с помощью распределения статистики

$$\eta = \sup_{-\pi \le \lambda \le \pi} |\bar{f}_N(\lambda) - 0{,}01326|, \qquad (22)$$

где $\bar{f}_N(\lambda)$ — оценка спектральной плотности; 0,01326 — ненормированное значение спектральной плотности независимых одинаково распределенных случайных величин, имеющих равномерное распределение на [0; 1]. Отмечено, что локальная проверка является более жесткой и этим отличается от глобальной. Для глобальной проверки следует использовать методы, основанные на распределении статистики (22) при $N \to \infty$. Псевдослучайные последовательности URAND проверялись более жестко в работе [14], чем в [3]. Кроме того, в статье [3] с чисто теоретических позиций, но без числовых проверок обосновано [18, II], что из подхода А. Н. Колмогорова к случайности следует, что алгоритмические датчики псевдослучайных чисел имеют неустранимые недостатки и не всегда генерируют достаточно хорошую последовательность.

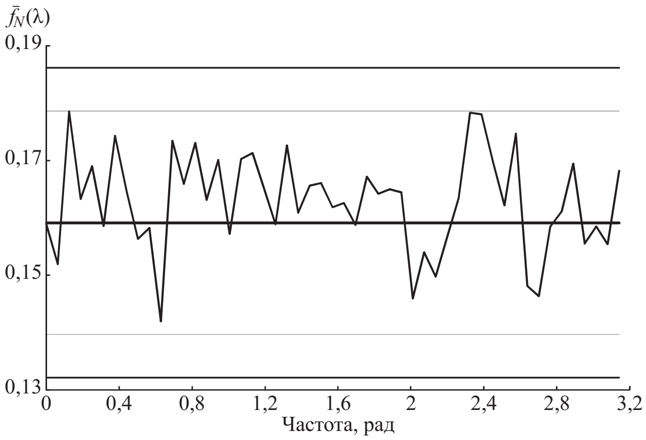
В комментарии [18, III] отмечено, что вопросы обоснования применимости метода математического моделирования в ситуациях, когда вместо идеального датчика используется процедура, поставляющая последовательности псевдослучайных чисел, являются сложными и далеко не решенными.

В комментарии [19] предлагается рассмотреть другой более употребительный класс методики проверки гипотезы H_0, суть которой заключается в следующем. Рассматриваются последовательности $\mathbf{z}_1, \mathbf{z}_2, \ldots$ векторов из k-мерного единичного куба $[0; 1]^k$, полученные из последовательности псевдослучайных чисел. Куб $[0; 1]^k$ разбивается на s непересекающихся частей: $A_1, \ldots, A_s$. Наблюдаемые частоты попадания в каждое из этих множеств сравниваются с теоретически ожидаемыми частотами, например, с помощью критериев χ_2, Колмогорова и т.д. Отмечены недостатки метода, включающие, в частности, вопрос, каким образом надо разбивать куб $[0; 1]^k$ на множества. Приводятся схема проверки и перечень проверенных датчиков: RANDU (фирма IBM) [4], URAND [15], датчики, основанные на М-алгоритме [17], и т.д. В заключение перечислены результаты проверки с помощью ортогональной системы функций.

Датчик RANDU [4] забракован на выборках в $10^4 – 10^5$. Среди остальных датчиков, за исключением RANDU [14], не основанных на М-алгоритме [17], наилучшим оказался URAND [15]. Датчики, основанные на М-алгоритме [17], прошли все проверки рассматриваемого метода.

В комментарии [21] разбирается статья [20], которая содержит основные сведения по научному обоснованию метода статистических испытаний. Ее автор считает, что однозначных рекомендаций по использованию конкретного датчика дать нельзя. Такая же точка зрения высказывается в работе [17]. Кроме того, в статье [20] рассматриваются псевдослучайные последовательности только фиксированной длины, что не всегда соответствует проводимым исследованиям. В заключение дискуссии [21] высказано пожелание подробно рассмотреть достоинства и недостатки различных типов датчиков.

Числовые результаты. Пусть последовательность случайных величин $\varepsilon(t)$, $t = 0, 1, \ldots$, $E\varepsilon(t) = 0$, $D\varepsilon(t) = 1$, имеет равномерное распределение. Спектральная плотность такой последовательности равномерно рас-



Оценка спектральной плотности с временны́м сдвигом $n = 100$, $T = 256$ (тонкая линия), полученная для последовательности псевдослучайных чисел на основании М-алгоритма: жирная линия — нормированная спектральная плотность последовательности независимых случайных величин; пунктирная и полужирная линии — границы множеств, на которых на уровне значимости 0,05 принимается гипотеза независимости при локальной и глобальной проверках

пределенных случайных величин постоянна на отрезке $[-\pi, \pi]$ и определяется следующим образом:

$$f(\lambda) = \frac{1}{2\pi}\sum_{t=-\infty}^{\infty} C(t)e^{-i\lambda t} = \frac{C(0)}{2\pi},\ 0 \le |\lambda| \le \pi,$$

где $C(t)$ — ковариационная функция, в данном случае равномерно распределенных случайных величин $C(0) = 1$, поэтому $f(\lambda) = 0{,}159156$. Для исследования использовался датчик псевдослучайных чисел URAND [15]. На основании асимптотического распределения (17) проверим гипотезу H_0 о том, является ли произвольная, но фиксированная последовательность псевдослучайных чисел последовательностью независимых одинаково распределенных случайных величин. Поскольку

$$\sup_{0\le\lambda\le\pi}\left|2\pi\bar{f}_N(\lambda) - 1\right| =$$

$$= \max\left\{\sup_{0\le\lambda\le\pi}(2\pi\bar{f}_N(\lambda) - 1) - \inf_{0\le\lambda\le\pi}(2\pi\bar{f}_N(\lambda) - 1)\right\},$$

то, выбирая уровень значимости $0 < \alpha < 1$, проверим гипотезу H_0. Из уравнения $1 - \alpha = \exp[-\exp(-x)]$ найдем $x = -\ln[-\ln(1 - \alpha)]$. Величина

$$f(\lambda) + \frac{(b_N + a_N x)f(\lambda)}{\sqrt{T}},\ 0 \le \lambda \le \pi,$$

задает границу левостороннего критерия, на котором принимается гипотеза H_0 на заданном уровне значимости α; a_N и b_N выражены формулами (18). Этот метод задает глобальную проверку [18, II]. Построим условно нижнюю границу, симметричную верхней относительно теоретического значения спектральной плотности $f(\lambda)$. Это позволяет увидеть, как расположены наименьшие значения относительно нижней границы.

Для определения подходящего уровня значимости для проверки гипотезы H_0 на основании распределения (17) используем выражение

$$\alpha = 1 - \exp\left\{-\exp\left[-\frac{1}{a_N}\left(\frac{(\bar{f}_N(\lambda_0) - f(\lambda))\sqrt{T}}{f(\lambda)} - b_N\right)\right]\right\},$$

где $\lambda_0 = \max\left(\sup_{0\le\lambda\le\pi}\bar{f}_N(\lambda),\ \inf_{0\le\lambda\le\pi}\bar{f}_N(\lambda)\right)$.

Следуя работе [14], проведем локальную проверку псевдослучайных последовательностей, вырабатываемых URAND [15]. Отличие псевдослучайных последовательностей состоит в том, что для ЭВМ БЭСМ-6 максимальное значение целого числа $m = 2^{40}$, а для Windows XP — $m = 2^{31}$. Прореживание последовательности проводилось по следующему алгоритму [14]: из псевдослучайной последовательности были взяты элементы с номерами $1, \tau + 1, 2\tau + 1, \ldots$, а промежуточные — пропущены, и так до тех пор, пока число элементов последовательности не достигнет значения n. Далее вычислялась оценка $\bar{f}_N(\lambda)$ по формуле (1), $|\lambda| \le \pi$, имеющая в каждой точке асимптотически нормальное распределение, и выбиралось максимальное значение оценки $\bar{f}_N(\lambda_0)$, которое достигалось в точке λ_0. Аналогично [14] была проверена гипотеза независимости H_0 последовательности на заданном уровне α значимости на основании

$$f(\lambda_0) \in \left[f(\lambda) - u_{1-\alpha/2}\frac{f(\lambda)}{\sqrt{T}},\ f(\lambda) + u_{1-\alpha/2}\frac{f(\lambda)}{\sqrt{T}}\right],$$

где $f(\lambda)$ — нормированное значение спектральной плотности для последовательности независимых одинаково равномерно распределенных на $[0; 1]$ случайных величин, равное 0,159156; $u_{1-\alpha/2}$ — квантиль стандартного нормального распределения.

Глобальную и локальную проверки гипотезы независимости иллюстрирует рисунок. Результаты этих проверок для различных последовательностей псевдослучайных чисел URAND приведены в таблице. Для вычислений было взято число отрезков $T = 256$, начальное число последовательности равно 1, уровень значимости $\alpha = 0{,}05$. Одна последовательность моделировалась с помощью М-алгоритма [17].

Из таблицы следует, что при прореживании последовательности псевдослучайных чисел гипотеза H_0 может быть выполнена по крайней мере для выборок, состоящих из 256 непересекающихся отрезков при $n \le 7000$. Из 16 выборок несовпадение результатов глобальной и локальной проверок гипотезы независимости обнаружено для шести выборок. Вычисление границ критического множества гипотезы независимости, а также определение соответствующих уровней значимости для распределения экстремальных значений и нормального распределения в рассматриваемых случаях подчеркивают строгость локальной проверки по сравнению с глобальной.

Прореживания выборок с отрезками $n = 500 - 6000$, $T = 256$, $\tau = 128$ приводит к тому, что такие последовательности псевдослучайных чисел удовлетворяют гипотезе независимости на уровне значимости 0,05. При отсутствии прореживания для соответствующих выборок гипотеза о независимости на уровне значимости 0,05 не выполняется. Применение М-алгоритма при построении выборки приводит к тому, что при глобальной и локальной проверках гипотеза независимости не отвергается, если длина отрезка разбиения $n = 100$.

В таблице для сравнения глобальной и локальной проверок приведены уровни значимости, построенные на основании значения модуля максимального отклонения оценки (1) от ее нормированного теоретического значения. Эти результаты показывают, на каком конкретном уровне выполняется гипотеза независимости.

Таким образом, результаты расчетов показали, что гипотеза о независимости не отвергается на разумном уровне значимости либо для последовательностей, полученных с помощью прореживания, либо с помощью М-алгоритма. Кроме того, было установлено, что локальная проверка, формирующая множество, на котором не отвергается гипотеза о независимости последовательности, более строгая, чем глобальная, основанная на построении критического множества для диапазона частот.

Результаты глобальной и локальной проверок для различных последовательностей псевдослучайных чисел

Длина отрезка n	Величина прореживания τ	Значение частоты λ_0, при которой оценка достигает максимального отклонения от теоретической спектральной плотности	Экстремальные значения оценки спектральной плотности с временным сдвигом		Границы множества для проверки гипотезы независимости на уровне значимости 0,05		Уровни значимости, вычисленные на основании модуля максимального отклонения оценки от теоретической спектральной плотности		Границы множества для проверки гипотезы независимости
			минимум	максимум	распределение экстремальных значений	нормальное распределение	распределение экстремальных значений	нормальное распределение	
					URAND				
100	*0*	*1,13097*	*0,140620*	*0,181027*			0,19500	0,02780	(0,137283; 0,181027)
100	16	1,38230	0,139719	0,189453			0,02008	0,00270	(0,128857; 0,189453)
100	*32*	*1,94779*	*0,133854*	*0,185837*	(0,132155; 0,186155)		0,054541	0,00736	(0,132473; 0,185837)
100	64	0,69115	0,127142	0,180681			0,012448	0,00138	(0,127142; 0,191168)
100	*128*	*1,63363*	*0,145760*	*0,178662*			0,344121	0,05003	(0,139648; 0,178662)
500	*0*	*2,22745*	*0,136692*	*0,191061*	(0,127825; 0,190485)		0,041434;	0,00138	(0,127249; 0,191061)
500	**128**	**0,46496**	**0,142510**	**0,178443**			>0,35	0,03238	(0,139867; 0,178443)
1000	0	0,18850	0,141212	0,192757	(0,126046; 0,192264)	(0,139658; 0,178651)	0,042157	0,00068	(0,125553; 0,192757)
1000	**128**	**1,22522**	**0,144586**	**0,175829**			>0,35	0,09296	(0,142481; 0,175829)
5000	0	1,63614	0,132222	0,195121	(0,122132; 0,196178)		0,075121	0,00032	(0,123189; 0,195121)
5000	**128**	**2,72439**	**0,143608**	**0,174349**			>0,35	0,12602	(0,143961; 0,174863)
6000	*0*	*0,18326*	*0,125533*	*0,192777*	(0,121707; 0,196603)		0,0212642	0,000064	(0,125533; 0,192777)
6000	**128**	**1,53833**	**0,141664**	**0,175864**			>0,35	0,09296	(0,142446; 0,175864)
7000	0	2,25656	0,126267	0,200711	(0,121350; 0,196960)		0,011118	0,000064	(0,117599; 0,200711)
7000	*128*	*2,27631*	*0,129785*	*0,187112*			>0,35	0,00496	(0,131197; 0,187112)
					М-алгоритм [17]: RANDOM(*n*), URAND				
100	—	**0,125664**	**0,141966**	**0,178584**	*(0,132155; 0,186155)*	(0,139658; 0,178651)	0,3500227	0,05118	(0,139726; 0,178584)

Примечание. Жирным шрифтом показаны те выборки, для которых гипотеза H_0 выполняется для обоих распределений; курсивом — выборки, для которых гипотеза H_0 выполняется только для распределения экстремальных значений; обычным шрифтом — выборки, для которых гипотеза H_0 не выполняется для обоих распределений.

ЛИТЕРАТУРА

1. **Отнес Р., Эноксон Л.** Прикладной анализ временны́х рядов. — М.: Мир, 1982. — 428 с.
2. **Журбенко И. Г., Кожевникова И. А.** О сравнительных характеристиках статистик спектральной плотности стационарных случайных процессов / Проблемы передачи информации. 1982. Т. XVIII. Вып. 1. С. 64 – 77.
3. **Журбенко И. Г., Кожевникова И. А., Смирнова О. С.** О построении и исследовании псевдослучайных последовательностей различными методами / Заводская лаборатория. 1985. Т. 51. № 5. С. 47 – 51.
4. Сборник научных программ на Фортране / Под ред. С. Я. Виленкина. Ч. I. — М.: Статистика. 1974. — 374 с.
5. **Журбенко И. Г.** Анализ стационарных и однородных случайных систем. — М.: Изд-во МГУ, 1987. — 238 с.
6. **Крамер Г., Лидбеттер М.** Стационарные случайные процессы. — М.: Мир, 1969. — 398 с.
7. **Бриллинджер Д.** Временны́е ряды. Обработка данных и теория. — М.: Мир, 1980. — 536 с.
8. **Ибрагимов И. А., Линник Ю. В.** Независимые и стационарно связанные величины. — М.: Наука, 1965. — 524 с.
9. **Лидбеттер М., Линдгрен Г., Ротсен Х.** Экстремумы случайных последовательностей и процессов. — М.: Мир, 1989. — 524 с.
10. **Ширяев А. Н.** Вероятность-1. — М.: Изд-во МЦНМО, 2004. — 519 с.
11. **Woodroofe M. B.** On the maximum deviation of sample spectral density / Ann. Math. Statist. 1967. Vol. 38. N 2. P. 475 – 481.
12. **Рудзкис Р.** Об общей оценке спектральной плотности / Литовский математический сборник. 1985. Т. 25. № 3. С. 163 – 174.
13. **Рудзкис Р.** О распределении максимального уклонения оценки спектральной плотности стационарной последовательности / Литовский математический сборник. 1985. Т. 25. № 4. С. 118 – 130.
14. **Рыданова Г. В.** Методика изучения временны́х зависимостей в последовательности псевдослучайных чисел / Заводская лаборатория. 1986. Т. 52. № 1. С. 56 – 58.
15. **Форсайт Дж., Малькольм М., Моулер К.** Машинные методы математических вычислений. — М.: Мир, 1980. — 279 с.
16. **Алексеев В. Г.** Об одном методе изучения последовательности псевдослучайных чисел / Заводская лаборатория. 1990. Т. 56. № 3. С. 84 – 86.
17. **Кнут Д.** Искусство программирования для ЭВМ. Получисленные алгоритмы. — М.: Мир, 1977. — 724 с.
18. **Журбенко И. Г., Кожевникова И. А., Орлов А. И., Благовещенский Ю. Н.** Комментарий к статье В. Г. Алексеева «Об одном методе изучения последовательности псевдослучайных чисел» / Заводская лаборатория. 1990. Т. 56. № 3. С. 68; С. 86 – 87; С. 88 – 89.
19. **Тюрин Ю. Н., Фигурнов В. Э.** О проверке датчиков псевдослучайных чисел / Заводская лаборатория. 1990. Т. 56. № 3. С. 89 – 92.
20. **Ермаков С. М.** О датчиках псевдослучайных чисел / Заводская лаборатория. 1993. Т. 59. № 7. С. 48 – 50.
21. **Орлов А. И.** Комментарий к статье Ермакова С. М. «О датчиках псевдослучайных чисел» / Заводская лаборатория. 1993. Т. 59. № 7. С. 51.