

УДК 518.5+519.2

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ДАТЧИКА ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

М.В.Антипов, Ф.М. Израйлев, Б.В.Чириков

Для решения задач методом статистических испытаний — методом Монте-Карло — необходимо иметь набор случайных чисел (чаще всего равномерно распределенных в $[0,1]$). Одним из наиболее распространенных и очень удобных для применения на ЭВМ методов получения таких чисел является использование некоторого простого преобразования (алгоритма), генерирующего последовательность так называемых псевдослучайных чисел. Последнее может быть легко трансформировано в псевдослучайный адрес для получения функции случайного аргумента, заданной в виде таблицы в памяти ЭВМ $[1]$.

По-видимому, из большого количества изобретенных алгоритмов наилучшим является мультипликативный $[2]$:

$$z_{n+1} = kz_n \pmod{2^P}; \quad z, k - \text{целые.} \quad (1)$$

Дело в том, что аналогичное преобразование для действительных чисел

$$x_{n+1} = \{kx_n\} \quad (2)$$

имеет положительную колмогоровскую энтропию $[3,4]$ и, следовательно, является наилучшей из известных в настоящее время имитаций случайного процесса, возможной для динамической системы. Поскольку, однако, теоремы эргодической теории справедливы с точностью до меры нуль, переход к целым числам (1) может привести к появлению аномалий. Хорошо известным примером таких аномалий является существование периода псевдослучайной последовательности. Возможны, однако, и более тонкие нарушения статистических свойств. Эти нарушения могли не обнаружиться при предыдущих исследованиях $[5-9]$ датчика (1). В настоящей рабо-

те приводятся результаты дополнительной проверки датчика (I), отличающейся от [5-9] большим числом методов, повышением статистической точности, а также увеличением объема последовательности.

Согласно [10] максимальный период (2^{p-2}) последовательности (I) достигается при

$$k \equiv 3; 5 \pmod{8}; z - \text{нечетное}, \quad (3)$$

а коэффициент парной корреляции соседних псевдослучайных чисел [II] будет

$$\rho \approx 1/k. \quad (4)$$

Поскольку целочисленное умножение по модулю 2^p на k и $(k-2^p)$ эквивалентно, то при $k \gg 2^{p-1}$ корреляции будут увеличиваться по сравнению с (4). Согласно [II] увеличение корреляций возможно даже при $k > 2^{(p/2)}$ в зависимости от конкретного значения k .

Для проверки качества псевдослучайной последовательности была использована система из пяти тестов:

1. Проверка равномерности. Интервал $[0,1]$ разбивался на 256 равных частей, и подсчитывалось количество попаданий псевдослучайного числа в каждый элементарный интервал длиной $1/256$. Величина χ^2 с 255 степенями свободы распределена приблизительно нормально, с параметрами $(255, 2 \times 255)$.^{х)} Поэтому и $(\chi^2 - 255)/510$ распределена приблизительно нормально, с параметрами $(0,1)$.

2. Проверка парной корреляции. Единичный квадрат разбивался на 256 частей (каждая сторона разбивалась на 16 частей). Первые четыре разряда ненормализованных псевдослучайных чисел z_n и z_{n+1} давали координаты одного из 256 элементарных квадратов. Дальнейшая обработка повторяет предыдущую.

3. Проверка комбинаций. Подсчитывалось количество единиц в первых 20 разрядах ненормализованного псевдослучайного числа. Так как известно гипотетическое распределение вероятностей $P\{s=i\} = c_i / 2^{20}$, где $0 \leq i \leq 20$ - число единиц, то легко получить распределение χ^2 с 20 степенями свободы.

4. Проверка серий. Подсчитывалось количество серий псевдослучайной последовательности вида $0 < z_n < 0,5$ и $0,5 < z_n < 1$ длины 1, 2, 3, ... Гипотетическая вероятность серии длины $i > 0$ того или другого вида равна $P\{s=i\} = 1/2^{i+2}$, а математическое ожидание общего количества серий $R = (N+2)/2 = 50001$, где N - длина псевдослучайной последовательности. Подсчитывалось общее ко-

личество серий и величина χ^2 с 30 степенями свободы (учитывалось количество серий первого вида длины от 1 до 16 и второго вида от 1 до 15).

5. Проверка интегральным методом. Рассматривались суммы вида:

$$I_s = \sum_{i=1}^N (-1)^{[i/s]} z_i,$$

где $[]$ - целая часть числа. Тогда при $s = 1, 2, \dots$ образуются знакопеременные суммы, у которых гипотетическое математическое ожидание равно нулю, а дисперсия - $N/12$. Величины $J_s = I_s / \sqrt{N/12}$ распределены асимптотически нормально, с параметрами $(0,1)$. Выбирая 10 интервалов величины J_s таким образом, чтобы гипотетическая вероятность попадания в любой из них была равна $1/10$, получаем распределение χ^2 с 9-ю степенями свободы. Нужно заметить, что при этой проверке стандартное N заменялось на $N_s = 2S[N/2S]$ для "зануления" математического ожидания J_s . Проверка интегральным методом означает, по существу, вычисление частотного спектра псевдослучайного процесса. Для случайного процесса спектр должен быть непрерывным.

Первоначальная проверка датчика (I) производилась на машине "М-20", где он имел вид:

065 $\langle z_0 \rangle$ $\langle k \rangle$ 0000 - произведение без нормализации и округления.

047 0000 0000 $\langle z \rangle$ - выдача младших разрядов произведения.

$\langle k \rangle$: 100 0000 0100 0013

$\langle z_0 \rangle$: 100 5633 4012 5643

На каждом шаге в ячейке $\langle z \rangle$ образуется ненормализованное псевдослучайное число.

На машине "Минск-22" датчик (I) реализуется следующими командами:

- 70 00 $\langle z_0 \rangle$ $\langle k \rangle$ - вывод младших разрядов произведения.

- 33 00 $\mathcal{E}$ $\mathcal{E}$ - условный переход по переполнению.

+ 12 00 0000 $\langle z_0 \rangle$ - занесение.

+ 72 00 $\langle c \rangle$ $\langle z \rangle$ - логическое умножение.

$\langle c \rangle$: - 7777 7777 7400

$\langle z_0 \rangle$: - 5633 4012 5643

$\langle k \rangle$: - 0000 0100 0013

х) Первый параметр - среднее значение, второй - дисперсия.

В ячейке $\langle z \rangle$ получается ненормализованное псевдослучайное число вида:

$$+ \varepsilon_1, \varepsilon_2, \dots, \sum_{s=1}^n 00000000.$$

В машине "Минск-22" при операции целочисленного умножения мантисса числа занимает как и в "М-20" 36 разрядов слова. Поэтому свойства датчика и, в частности, его период одинаковы на обеих машинах.

Иногда необходимо иметь хорошую в статистическом отношении последовательность, составленную из какого-либо двоичного разряда или группы разрядов псевдослучайного числа (например, адресной части [1]). Тогда период для j -го разряда равен 2^{P-1-j} . Действительно, как нетрудно убедиться, рассмотрение j -го разряда (или группы разрядов, начиная с j -го) означает, по существу, что вместо датчика $z_{i+1} \equiv \kappa z_i \pmod{2^P}$, реализуется датчик $z_{i+1} \equiv \kappa z_i \pmod{2^{P+1-j}}$ (разряды 1, 2, ..., $j-1$ не влияют на j -ый разряд очередного псевдослучайного числа; при $j = 1$ возвращаемся к исходному случаю). Для увеличения периода можно либо сдвигать полученное псевдослучайное число вправо, если нужна небольшая группа разрядов, либо применять возмущение κ [12]:

$$\begin{aligned} z_{i+1} &\equiv \kappa_j z_i \pmod{2^P}, \\ \kappa_{e+1} &\equiv \kappa_e + c \pmod{2^P}, \end{aligned} \quad (5)$$

где $c = 8$ - минимальная константа, для которой $\kappa_e \equiv 3 \pmod{8}$ для всех e . Согласно [12] период увеличивается при этом в $\sqrt{L}$ раз, где L - число шагов, через которое вводится возмущение.

Результаты тестовых проверок датчика (I) сведены в таблицу I.

Т а б л и ц а I

Тест	Равномерность	Парные корреляции	Серии $\chi^2_{R^{30}}$	Комбинации χ^2_{20}	Интегральный χ^2_9
Данные проверки	1,10	0,63	34,5 49996	22,2	6,23
Ожидаемый интервал	$0 \pm 1,96$	$0 \pm 1,96$	30 ± 15 50001 ± 310	$20 \pm 12,5$	$< 16,9$

В первой строке таблицы указаны полученные в результате проверки значения тестовых величин, описанных выше. Во второй строке указаны ожидаемые с вероятностью 95% интервалы тестовых величин для случайной последовательности.^{х)} Длина последовательности во всех испытаниях, кроме интегрального метода (см. выше), равна $N = 10^5$.

Результаты проверки интегральным методом при самых различных значениях S сведены в таблицу 2. Ожидаемый 95%-й интервал равен $0 \pm 1,96$.

Т а б л и ц а 2

s	J_s	S	J_s	S	J_s	S	J_s	S	J_s
1	1,479	31	0,611	801	0,166	1503	-0,086	3401	0,400
2	1,169	51	-0,339	901	0,687	1523	-0,026	3601	0,167
3	-0,667	77	-0,207	1000	-0,996	1555	-0,912	3801	0,913
5	0,634	201	-0,291	1001	-0,596	1655	-0,932	4001	0,805
6	-0,200	251	0,142	1101	-0,020	1755	0,621	4401	-0,560
11	-0,793	301	0,887	1201	1,647	2001	0,675	4801	-0,442
13	-1,163	401	0,319	1301	-0,765	2201	-0,175	5201	0,475
17	-0,797	501	0,835	1401	0,403	2401	0,865	5601	0,910
19	-0,567	601	0,844	1501	0,337	2801	0,566	7001	-0,115
21	0,340	701	1,158	1502	0,097	3201	0,772	10000	0,481

Как отмечалось выше, результаты всех этих испытаний полностью переносятся на машину "Минск-22". Тем не менее на ней была проведена проверка датчика на равномерность для большей длины последовательности $N = 2^{23} \approx 10^7$ с возмущением (5), но с меньшим числом ячеек - 128. Получена величина $\chi^2_{127} = 123,9$ при ожидаемом 95%-м интервале: 127 ± 16 .

Наконец, была предпринята дополнительная проверка датчика (I), использующая уникальные возможности машины БЭСМ-6. Были выбраны следующие параметры датчика (в восьмеричной записи в ячейке БЭСМ-6):

$$\begin{aligned} \langle \kappa \rangle &: 4013064256500425; \\ \langle z_0 \rangle &: 4013543660414035; \quad \frac{\kappa}{2^P} \approx \frac{11}{16}. \end{aligned} \quad (6)$$

^{х)} Интервалы указаны по нормальному распределению (кроме последнего столбца).

Точные значения параметров несущественны при выполнении условий (3). Даже очень "круглая" константа $\langle K \rangle$: 400000000200003 не ухудшает статистические свойства датчика. Однако заметим, что это, по-видимому, не всегда так [5,8]. Поэтому лучше выбирать "некруглые" параметры (6).

Использовались три метода проверки: равномерность (16384 ячейки); парные корреляции z_{n+1}, z_n (128 x 128 ячеек) и 14-кратные корреляции соседних чисел по одному двоичному разряду ($2 \times 2 \dots = 2^{14}$ ячеек).

Основные результаты приведены в табл. 3. Критерием случайности для всех трех методов служило отклонение от равномерного распределения во всем массиве из $2^{14}=16384$ ячеек. Характеристикой отклонения является отношение дисперсии (D) к среднему значению (M) количества псевдослучайных чисел в одной ячейке. Ожидаемое значение отношения для случайной последовательности равно (с доверительной вероятностью 95%):

$$D/M = 1.0 \pm 0,022. \quad (7)$$

В табл. 3 приведены также значения $\sqrt{D}/M$ - статистической точности проверки.

Т а б л и ц а 3

	N	10^3	10^4	10^5	10^6	10^7	10^8
Равномерность	$\sqrt{D}/M\%$	405	128	40	13	4,0	1,3
	D/M	1,003	1,003	1,003	1,006	1,000	0,977
	m	15415	8911	27	0	0	0
Парная корреляция	$\sqrt{D}/M\%$	407	128	40	13	4,0	1,3
	D/M	1,013	1,000	0,998	0,983	0,987	1,004
	m	15420	8905	38	0	0	0

В качестве дополнительного контроля статистических свойств был произведен подсчет числа пустых ячеек массива 512×1023 для парных корреляций. Массив является логическим, причем каждый элемент занимает один двоичный разряд [13]. Всего используется 16384 слова по 32 разряда в каждом. Все размеры являются степенью двойки, что резко упрощает программу. Результаты приведены в табл. 4, где m и $m_{теор}$ - полученное и ожидаемое число незаполненных ячеек в массиве.

Т а б л и ц а 4

	N	10^3	10^4	10^5	10^6	10^7
Парная корреляция	m	523288	514376	433175	77952	0
	$m_{теор}$	522700 ± 700	514300 ± 700	433600 ± 650	78000 ± 280	0

В табл. 5 представлены результаты проверки статистических свойств для 14-кратных корреляций 1-го и 14-го двоичных разрядов. Для увеличения периода в последнем случае применялось возмущение константы K (5).

Т а б л и ц а 5

	N	10^3	10^4	10^5	10^6	10^7	10^8
1 разряд	$\sqrt{D}/M\%$	412	127	41	13	4,0	1,3
	D/M	1,037	0,982	1,008	0,997	0,974	1,002
	m	15431	8848	31	0	0	0
14 разряд	$\sqrt{D}/M\%$	414	128	41	13	4,0	1,3
	D/M	1,045	1,006	1,013	0,994	0,989	1,008
	m	15431	8889	33	0	0	0

Наконец, для парных корреляций было построено вторичное распределение отклонений от среднего, что является более тонким методом проверки статистических свойств датчика [5]. Случайной величиной здесь является отклонение числа попаданий в ячейку двумерного массива от среднего значения, нормированное на корень из дисперсии. Распределение производилось в интервал $(-4,4)$, разделенный на 128 ячеек. График полученного распределения и сравнение с гауссовской кривой приведены на рисунке. Разброс точек вызывается двумя причинами: статистическим разбросом $\pm 5\%$, который хорошо согласуется с большинством точек на рисунке, и разбросом за счет целочисленности случайной величины. Минимальное изменение случайной величины составляет примерно 1/5 размера ячейки распределения, что может вызвать колебания $\pm 20\%$. Это объясняет выпадение нескольких точек (особенно одной). Небольшой выход последнего значения D/M в табл. 3 из 95%-го интервала (7) объясняется недостаточной длиной периода

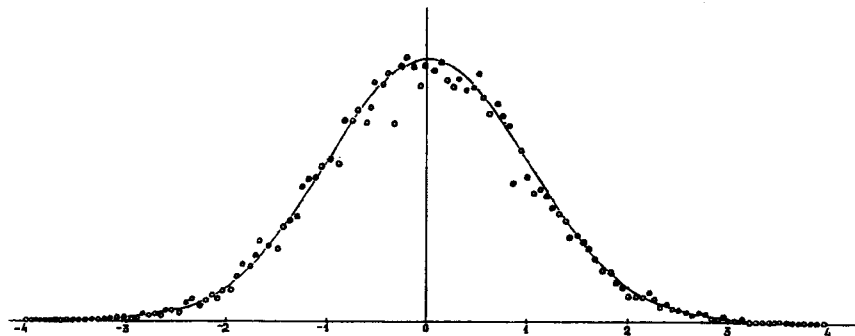


Рис.

для I4-го разряда, который еще существен при проверке на равномерность. Возмущение константы в этом случае не производилось; период I4-го разряда равен $2^{25} \approx 3 \cdot 10^7 < N = 10^8$.

Подводя итоги, можно сказать, что ни в одном из проведенных испытаний не было обнаружено отклонения свойств последовательности (I) от случайной.

Пользуемся случаем выразить искреннюю благодарность Ю.М. Волошину, Ю.Г. Косареву и А.И. Хисамутдинову за интересные дискуссии и полезные советы, а также группе обслуживания БЭСМ-6 и особенно В.П. Минаеву за большую помощь при проведении вычислений.

Л и т е р а т у р а

1. Ю.Г. Косарев. Примеры использования таблиц для сокращения времени счета - Данный сборник, стр. 46-54.
2. D.H. Lehmer. *Annals Comp.Lab.Harvard Univ.*, 26, 141, 1951.
3. В.А. Рохлин. *Изв. АН СССР, мат.*, 25, № 4, 499 (1961).
4. А.Г. Постников. Эргодические вопросы теории сравнений и теории диофантовых приближений. - Труды мат. ин-та им. Стеклова, XXXII, 1966 г.
5. J. Allard, A. Dobell, T. Hull. *Journ Assoc. Comp. Mach.*, 10, 131, 1963.
6. R. Kronmal, *ibid*, 11, 357, 1964.
7. A. Rotenberg, *ibid*, 7, 75, 1960.
8. T. Hull, A. Dobell, *ibid*, 11, 31, 1964.
9. D. MacLaren, G. Marsaglia, *ibid*, 12, 1965.
10. E. Bofinger, *ibid*, 5, 261, 1958.

11. M. Greenberger, *ibid*, 8, 163, 1961.

12. И.М. Соболев. Теория вероятностей и ее применение, 9, 367, 1964.

13. Ю.М. Волошин, А.П. Ершов, Г.И. Кожухин. Входной язык систем автоматизации программирования. Изд. СО АН СССР, 1964.

14. А.И. Голенко. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах, 1965.

Институт ядерной физики
СО АН СССР

Поступила в редакцию
15.01.1967 г.