ИНСТИТУТ ЯДЕРНОЙ ФИЗИКИ СИБИРСКОГО ОТДЕЛЕНИЯ АН СССР

препринт 174

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

Статистическая проверка датчика псевдослучайных чисел

г.Новосибирск 1967

Институт ядерной физики Сибирского отделения АН СССР

Препринт

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

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ДАТЧИКА ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

Новосибирск
1967

A B S T R A C T

Results of various tests for a pseudorandom numbers generator of multiplicative type (1) are given with the sequence length up to 10^8, number of bins ≈ 16000 and statistical accuracy about 1%. The generator (1) seems to be the best one because corresponding transformation for real numbers (2) is the mixing with positive Kolmogorov'entropy; hence, it is the best imitation of random-like process which is possible for a dynamical system (for an algorithm). All test results obtained so far are consistent with randomness of sequence (1).

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 методов получения таких чисел, очень удобном для применения на ЭВМ, 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 /13/.

Из большого количества изобретенных алгоритмов наилучшим представляется мультипликативный /1/:

$$z_{n+1} \equiv K z_n \pmod{2^p} \qquad (1)$$

z, K – целые

Дело в том, что аналогичное преобразование для действительных чисел:

$$X_{n+1} = \{K X_n\} \qquad (2)$$

имеет положительную колмогоровскую энтропию /9.10/ и, следовательно, является наилучшей, известной в настоящее время, имитацией 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. Поэтому возникает необходимость проверки датчика (1). Такая проверка производилась уже в ряде работ /2-6/. В настоящей работе приводятся результаты дополнительной проверки датчика (1), отличающейся применением большого числа методов проверки, повышением её точности, а также увеличением объема последовательности.

Согласно /7/ максимальный период (2^{p-2}) последовательности (1) достигается при:

$K \equiv 3; 5 (\mathrm{mod}\ 8)$; z - нечетное (3)

а коэффициент парной корреляции соседних псевдослучайных чисел равен /8/:

$$\rho \approx 1/K \qquad (4)$$

Поскольку целочисленное умножение по модулю 2^p на K и $(K-2^p)$ эквивалентно, то при $K \geq 2^{p-1}$ корреляции будут увеличиваться по сравнению с (4). Согласно /8/ 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:

I. Проверка равномерности. Интервал $[0,1]$ разбивался на 256 равных частей и подсчитывалось количество попаданий псевдослучайного числа в каждый элементарный интервал длиной I/256. Величина χ^2 с 255 степенями свободы распределена приближенно нормально, с параметрами (255, 2 x 255)+). Тогда величина $(\chi^2 - 255)/510$ распределена приближенно нормально, с параметрами (0,I).

2. Проверка парной корреляции. Единичный квадрат разбивался на 256 частей (каждая сторона разбивалась на I6 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 $P\{s=i\} = C_{20}^{i}/2^{20}$, где $0 \leq i \leq 20$ - число единиц, то легко получить распределение χ^2 с 20-ю степенями свободы.

4. Проверка серий. Подсчитывалось количество серий псевдослучайной последовательности вида $0 < z_n < 0{,}5$ и $0{,}5 < z_n < 1$ длины I,2,3,... Гипотетическая вероятность серии длины $i > 0$ того или другого вида равна $P\{s=i\} = I/2^{i+2}$

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

а математическое ожидание общего количества серий $R = (N+2)/2 = 5000I$, где N - длина псевдослучайной последовательности. Подсчитывалось общее количество серий и величина χ^2 с 30-ю степенями свободы (учитывалось количество серий первого вида длины от I до I6 и второго вида от I до I5).

5. Проверка интегральным методом. Рассматривались суммы вида:

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

где $[\ ]$ - целая часть числа. Тогда при $s = I,2,\ldots$ образуются знакопеременные суммы, гипотетическое математическое ожидание которых равно нулю, а дисперсия - $N/I2$. Величины $J_s = I_s / \sqrt{N/12}$ распределены асимптотически нормально, с параметрами (0,I). Выбирая I0 интервалов величины J_s таким образом, чтобы гипотетическая вероятность попадания в любой из них была равна I/I0, 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  <z>  <K>  —   произведение без норм.и окр.
047   —    —   <z>  выдача мл.разрядов произв.
   <K>  : I00 0000 0I00 00I3
   <z0> : I00 5633 40I2 5643
```

На каждом шаге в ячейке <z> образуется ненормализованное псевдослучайное число.

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

ж	- 70 00	<τ_0>	<K>	вывод мл.разрядов произв.
	- 33 00	ж	ж	усл.переход по переполн.
	I2 00	-	<τ_0>	занесение
	72 00	<C>	<τ>	логическое умножение

<C> : - 77 77 7777 7400

<τ_0> : - 5633 40I2 5643

<K> : - 0000 0I00 00I3

Нормализованное псевдослучайное число вида + ξ_1 , ξ_2,... ξ_{28} 00000000 получается в ячейке <τ> .

Заметим, что хотя монтисса числа в машине Минск-22 содержит всего 28 разрядов по сравнению с 36 разрядами для М-20, операция целочисленного умножения позволяет использовать все 36 разрядов слова. Поэтому свойства датчика и, в частности, его период одинаковы на обоих машинах.

Иногда необходимо иметь хорошую в статистическом отношении последовательность, составленную из какого-либо двоичного разряда или группы разрядов псевдослучайного числа (например, адресной части /I3/). Тогда период для j - того разряда равен 2^{p-1-j} . Действительно, как нетрудно убедиться, рассмотрение j - того разряда (или группы разрядов, начиная с j - того) означает, по существу, что реализуется не датчик $\tau_{i+1} \equiv K\tau_i \pmod{2^p}$, а датчик $\tau_{i+1} \equiv K\tau_i \pmod{2^{p+1-j}}$ (разряды I,2,..., j - I не влияют на j -ый разряд очередного псевдослучайного числа; при j = I возвращаемся к исходному случаю). Для увеличения периода можно либо сдвигать полученное псевдослучайное число вправо, если нужна небольшая группа разрядов, либо применять возмущение K /I4/:

$$\tau_{i+1} \equiv K_j \tau_i \pmod{2^p} \qquad (5)$$

$$K_{\ell+1} \equiv K_\ell + C \pmod{2^p}$$

где C = 8 - минимальная константа для которой $K_\ell \equiv 3 \pmod 8$ для всех ℓ . Согласно /I4/ период увеличивается при этом в

$\sqrt{L}$ раз, где L - число шагов, через которое производится возмущение.

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

Таблица I

Тест	Равномерность	Парные корреляции	Серии χ^2_{30} R	Комбинации χ^2_{20}	Интегральн. χ^2_9
I	2	3	4	5	6
Данные проверки	I,I0	0,63	34.5 49996	22,2	6,23
Ожидаемый интеграл	0±I,96	0±I,96	30±I5 5000I±3I0	20±I2,5	I6,9

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+). Длина последовательности во всех испытаниях кроме интегрального метода (см.выше) равна $N = I0^5$.

Результаты проверки интегральным методом при самых различных значениях S сведены в таблицу II. Ожидаемый 95%-й интервал равен: 0 ± I,96.

Как отмечалось выше, результаты всех этих испытаний полностью переносятся на машину Минск-22. Тем не менее была проведена проверка датчика с возмущением (5) на машине Минск-22 на равномерность для бо́льшей длины последовательности $N = 2^{23} \approx I0^7$, но с меньшим числом ячеек - I28. Получена величина $\chi^2_{I27} = I23,9$ при ожидаемом 95% интервале; I27 ± I6.

+) Интервалы указаны по нормальному распределению (кроме последнего столбца).

Таблица П.

S	J_s	S	J_s	S	J_s	S	J_s	S	J_s
I	I,479	3I	0,6II	80I	0,I66	I503	-0,086	340I	0,400
2	I,I69	5I	-0,339	90I	0,687	I523	-0,026	360I	0,I67
3	-0,667	77	-0,207	I000	-0,996	I555	-0,9I2	380I	0,9I3
5	0,634	20I	-0,29I	I00I	-0,596	I655	-0,932	400I	0,805
6	-0,200	25I	0,I42	II0I	-0,020	I755	0,62I	440I	-0,560
II	-0,793	30I	0,887	I20I	I,647	200I	0,675	480I	-0,442
I3	-I,I63	40I	0,3I9	I30I	-0,765	220I	-0,I75	520I	0,475
I7	-0,797	50I	0,835	I40I	0,403	240I	0,865	560I	0,9I0
I9	-0,567	60I	0,844	I50I	0,337	280I	0,566	700I	-0,II5
2I	0,340	70I	I,I58	I502	0,097	320I	0,772	I0000	0,48I

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

$$\langle K\rangle : 40I3064256500425; \quad \frac{K}{2^P} \approx \frac{11}{16} \qquad (6)$$
$$\langle \tau_0\rangle : 40I3543660 4I4035;$$

Точные значения параметров несущественны при выполнении условий (3). Даже очень "круглая" константа $\langle K\rangle$: 4000000000200003 не ухудшает статистические свойства датчика. Заметим, однако, что это, по-видимому, не всегда так /2,5/. Поэтому лучше выбирать "некруглые" параметры (6).

Использовались три метода проверки: равномерность (I6384 ячейки); парные корреляции τ_{n+1} , τ_n (I28 x I28 ячеек) и I4-ти кратные корреляции соседних чисел по одному двоичному разряду ($2 \times 2 \ldots = 2^{I4}$ ячеек).

Основные результаты приведены в таблице Ш. Критерием случайности для всех трех методов служило отклонение от равномерного распределения во всем массиве из 2^{I4} = I6384 ячеек. Характеристикой отклонения является отношение дисперсии ($\mathcal{D}$) к среднему значению ($\mathcal{M}$) количества псевдослучайных чисел в одной ячейке. Ожидаемое значение отношения для случайной последовательности равно (с доверительной вероятностью 95%):

$$\mathcal{D}/\mathcal{M} = I.0 \pm 0,022 \qquad (7)$$

В таблице Ш приведены также значения $\sqrt{\mathcal{D}}/\mathcal{M}$ - статистической точности проверки.

В качестве дополнительного контроля статистических свойств был произведен подсчет числа пустых ячеек массива 5I2 x I023 для парных корреляций. Массив является логическим, причем каждый элемент занимает один двоичный разряд /I2/. Всего используется I6384 слова по 32 разряда в каждом. Все размеры являются степенью двойки, что резко упрощает программу. Результаты приведены в таблице IУ, где m и $m_{теор}$ - полученное и ожидаемое число незаполненных ячеек в массиве.

В таблице 5 представлены результаты проверки статистических свойств для I4-ти кратных корреляций I-го и I4-го двоичных разрядов. Для увеличения периода в последнем случае применялось возмущение константы K (5).

Наконец, для парных корреляций было построено вторичное распределение отклонений от среднего, что является более тонким методом проверки статистических свойств датчика /2/. 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 I28 ячеек. График полученного распределения и сравнение с гауссовской кривой приведены на рисунке. Разброс точек вызывается двумя причинами: статистический разброс ± 5%, который хорошо согласуется с большинством точек на рисунке и разброс за счет целочисленности случайной величины. Минимальное изменение случайной величины составляет примерно I/5 размера ячейки распределения, что может вызвать колебания ± 20%. Это объясняет выпадение нескольких точек (особенно одной).

Отметим также небольшой выход последнего значения $\mathcal{D}/\mathcal{M}$ в таблице Ш из 95% интервала (7). Это может объясняться недостаточной длиной периода для I4 разряда, который еще существенен при проверке на равномерность. Возмущение константы в этом случае не производилось; период I4 разряда равен $2^{25} \approx 3.I0^7 < N = I0^8$.

Подводя итоги, можно сказать, что ни в одном из проведен - ных испытаний не было обнаружено отклонения свойств последова - тельности (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 вычис-лений.

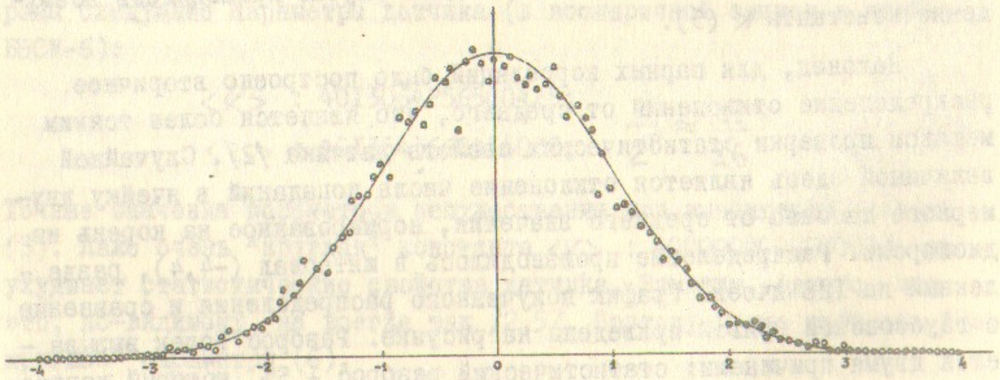

Таблица III

	N	10^3	10^4	10^5	10^6	10^7	$10^{\prime\prime 8}$
Равно-мерность	$\sqrt{D}/M$ %	405	I28	40	I3	4,0	I,3
	D/M	I,003	I,003	I,003	I,006	I,000	0,977
	m	I54I5	89II	27	0	0	0
Парная корре-ляция	$\sqrt{D}/M$ %	407	I28	40	I3	4,0	I,3
	D/M	I,0I3	I,000	0,998	0,983	0,987	I,004
	m	I5420	8905	38	0	0	0

Таблица IУ

	N	10^3	10^4	10^5	10^6	10^7
Парная корре-ляция	m	523288	5I4376	433I75	77952	0
	$m_{теор}$	522700 ±700	5I4300 ±700	433600 ±650	78000 ±280	0

Таблица У

	N	10^3	10^4	10^5	10^6	10^7	10^8
I разряд	$\sqrt{D}/M$ %	4I2	I27	4I	I3	4,0	I,3
	D/M	I,037	0,982	I,008	0,997	0,974	I,002
	m	I543I	8848	3I	0	0	0
I4 раз-ряд	$\sqrt{D}/M$ %	4I4	I28	4I	I3	4,0	I,3
	D/M	I,045	I,006	I,0I3	0,994	0,989	I,008
	m	I543I	8889	33	0	0	0

N - длина последовательности псевдослучайных чисел
m - число незаполненных ячеек
$m_{теор}$ - ожидаемое число незаполненных ячеек.

Литература

I. D.H.Lehmer, Annals Comp. Lab. Harvard University, 26, I4I, I95I

2. J.Allard, A.Dobell, T.Hull, Journ. Assoc. Comp. Mach., I0, I3I, I963.

3. H.Kronmal, ibid, 11, 357, 1964

4. A.Rotenberg, ibid, 7, 75, 1960

5. T.Hull, A.Dobell, ibid, 11, 31, 1964

6. D.MacLaren, G.Marsaglia, ibid, 12, 83, 1965

7. E.Bofinger, V.Bofinger, ibid, 5, 261, 1958

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

9. Рохлин В.А. Изв. АН СССР, мат., 25 № 4, 499 (I96I)

I0. Постников А.Г. Эргодические вопросы теории сравнений и теории диафантовых приближений. Труды мат. ин-та им.Стеклова, XXXП, I966.

II. Голенко А.И. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах, I965.

I2. Волошин Ю.М., Ершов А.П., Кожухин Г.И. Входной язык систем автоматизации программирования. Изд. СО АН СССР, I964.

I3. Косарев Ю.Г. Примеры использования таблиц для сокращения времени счета, сб. "Вычислительные системы", ИМ СО АН СССР,I968

I4. Соболь И.М., Теория вероятностей и её применения, 9, 367,I964.

Ответственный за выпуск Г.Б. ГЛАГОЛЕВ
Подписано к печати 21.XII-1967 г.
Усл. 0,5 печ. листа, тираж 250 экз.
Заказ № 174, Бесплатно.

Отпечатано на ротапринте в ИЯФ СО АН СССР