1968 г. ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ Выпуск 30

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]:

$$z_{n+1} \equiv \kappa z_n \pmod{2^p}; \qquad z, \kappa \text{ – целые.} \tag{1}$$

Дело в том, что аналогичное преобразование для действительных чисел

$$x_{n+1} = \{\kappa x_n\} \tag{2}$$

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 ($2^{\rho-2}$) последовательности (I) достигается при

$$\kappa \equiv 3;\ 5\ (mod 8);\ \tau \text{ - нечетное,} \qquad (3)$$

а коэффициент парной корреляции соседних псевдослучайных чисел [11] будет

$$\rho \approx 1/\kappa. \qquad (4)$$

Поскольку целочисленное умножение по модулю 2^{ρ} на κ и $(\kappa-2^{\rho})$ эквивалентно, то при $\kappa \gtrsim 2^{\rho-1}$ корреляции будут увеличиваться по сравнению с (4). Согласно [11] увеличение корреляций возможно даже при $\kappa > 2^{(\rho/2)}$ в зависимости от конкретного значения κ.

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 1/256. Величина x^2 с 255 степенями свободы распределена приближенно нормально, с параметрами (255,2x255).[x] Поэтому и $(x^2-255)/510$ распределена приближенно 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 τ_n и τ_{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 $0 \le i \le 20$ - число единиц, то легко получить распределение x^2 с 20 степенями свободы.

4. Проверка серий. Подсчитывалось количество серий псевдослучайной последовательности вида $0 < \tau_n < 0{,}5$ и $0{,}5 < \tau_n < 1$ длины 1,2,3,... Гипотетическая вероятность серии длины $l > 0$ 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-

---

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

личество серий и величина x^2 с 30 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:

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

где [ ]- целая часть числа. Тогда при s = I,2,... 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,I). Выбирая I0 интервалов величины J_s таким образом, чтобы гипотетическая вероятность попадания в любой из них была равна I/I0, получаем распределение X^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 ⟨τ0⟩ ⟨κ⟩ 0000  - произведение без нормализации и
                     округления.
047  0000  0000 ⟨τ⟩ - выдача младших разрядов произве-
                     дения.
⟨κ⟩: I00  0000  0I00 00I3
⟨τ0⟩: I00  5633  40I2 5643
```

На каждом шаге в ячейке ⟨τ⟩ образуется ненормализованное псевдослучайное число.

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

```
        - 70  00  ⟨τ0⟩  ⟨κ⟩ - вывод младших разрядов произве-
                              дения.
        - 33  00   æ    æ   - условный переход по переполне-
æ:                            нию.
        + I2  00  0000  ⟨τ0⟩ - занесение.
        + 72  00  ⟨c⟩   ⟨τ⟩ - логическое умножение.
⟨c⟩: - 7777  7777  7400
⟨τ0⟩:- 5633  40I2  5643
⟨κ⟩: - 0000  0I00  00I3
```

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

$$+\ \varepsilon_1\ ,\ \varepsilon_2\ ,\ \ldots\ ,\ \Sigma_{28}\ 00000000.$$

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 [I]). 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 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 κ [I2]:

$$\begin{aligned} z_{i+1} &\equiv \kappa_j z_j \pmod{2^p}, \\ \kappa_{i+1} &\equiv \kappa_i + C \pmod{2^p}, \end{aligned} \qquad (5)$$

где $C = 8$ - минимальная константа, для которой $\kappa_i \equiv 3 \pmod 8$ для всех i. Согласно [I2] 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.

Таблица I

Тест	Равномерность	Парные корреляции	Серии $\chi^2_{R\,50}$	Комбинации χ^2_{20}	Интегральный χ^2_9
Данные проверки	I,I0	0,63	34,5 49996	22,2	6,23
Ожидаемый интервал	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.[x] 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 ± 1.96.

Таблица 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 (1), использующая уникальные возможности машины БЭСМ-6. Были выбраны следующие параметры датчика (в восьмеричной записи в ячейке БЭСМ-6):

$$\langle K \rangle : 4013064256500425; \quad \frac{K}{2^P} \approx \frac{11}{16}. \qquad (6)$$
$$\langle \tau_0 \rangle : 4013543660414035;$$

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

Точные значения параметров несущественны при выполнении условий (3). Даже очень "круглая" константа $\langle K \rangle$: 4000000000200003 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 (I6384 ячейки); парные корреляции τ_{n+1}, τ_n (I28 x I28 ячеек) и I4 - кратные корреляции соседних чисел по одному двоичному разряду (2 x 2... = 2^{I4} ячеек).

Основные результаты приведены в табл. 3. Критерием случайности для всех трех методов служило отклонение от равномерного распределения во всем массиве из 2^{I4}=I6384 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%):

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

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

Таблица 3

	N	$I0^3$	$I0^4$	$I0^5$	$I0^6$	$I0^7$	$I0^8$
Равномерность	$\sqrt{D}/M$ %	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 корреляция	$\sqrt{D}/M$ %	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

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 [I3]. Всего используется I6384 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.

Таблица 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).

Таблица 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 Д/М в табл. 3 из 95%-го интервала (7) объясняется недостаточной длиной периода

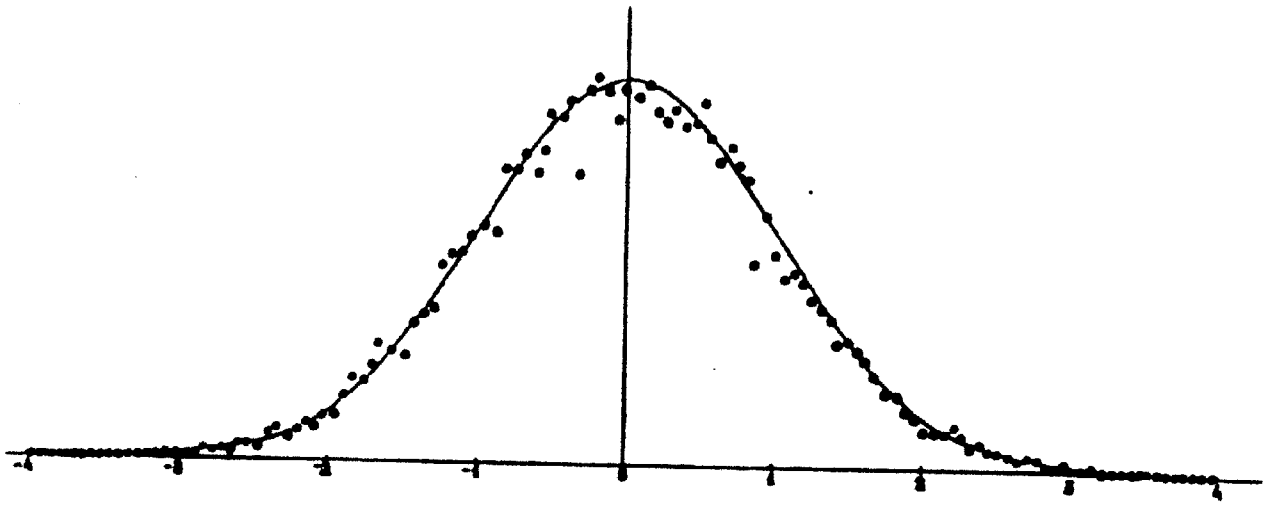

Рис.

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.

Литература

I. 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 (I96I).
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.

II. M.Greenberger, ibid,8 ,163,1961.
I2. И.М. Соболь. Теория вероятностей и ее применение, 2, 367, I964.
I3. Ю.М. Волошин, А.П. Ершов, Г.И. Кожухин. Входной язык систем автоматизации программирования. Изд. СО АН СССР, I964.
I4. А.И. Голенко. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах, I965.

Институт ядерной физики
СО АН СССР

Поступила в редакцию
I5.YI.I967 г.