ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1970 год Выпуск 42

УДК 518.5 +519.2

К ОЦЕНКЕ ДАТЧИКА ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

М.В. Антипов

Для получения последовательности случайных чисел при вычислениях на ЭВМ часто пользуются датчиком псевдослучайных чисел. Датчик представляет набор операций, реализующий рекуррентный алгоритм получения очередного числа. Нас будет интересовать датчик

$$r_{n+1} \equiv k r_n \pmod{2^p}; \quad r, k, p - \text{целые}, \tag{1}$$

довольно хорошо проверенный и давший неплохие результаты [1–4]. Псевдопоследовательность, полученная этим датчиком, имеет максимальный период, равный 2^{p-2} при нечетном r_0 и при $k \equiv 3 \pmod 8$; $k \equiv 5 \pmod 8$ [1].

Рассмотрим полные периоды датчика при различных k. Все псевдослучайные числа нечетны и, следовательно, равны 1; 3; 5; 7 $(mod\ 8)$.

Составим табличку для двух видов k датчика (1):

$r_n \pmod 8$	1	3	5	7
$r_{n+1} \equiv 3 r_n \pmod 8$	3	1	7	5
$r_{n+1} \equiv 5 r_n \pmod 8$	5	7	1	3

Отсюда находим четыре вида псевдопоследовательностей:

$$
\begin{aligned}
&k \equiv 3 \pmod 8, && \tau_n \equiv 1,3 \pmod 8,\\
& && R_1 = \{1, 3, 9, \ldots, 2^p-7, 2^p-5\};\\
&k \equiv 3 \pmod 8, && \tau_n \equiv 5,7 \pmod 8,\\
& && R_2 = \{5, 7, 13, \ldots, 2^p-3, 2^p-1\};\\
&k \equiv 5 \pmod 8, && \tau_n \equiv 1,5 \pmod 8,\\
& && R_3 = \{1, 5, 9, \ldots 2^p-7, 2^p-3\};\\
&k \equiv 5 \pmod 8, && \tau_n \equiv 3,7 \pmod 8,\\
& && R_4 = \{3, 7, 11, \ldots, 2^p-5, 2^p-1\}.
\end{aligned}
\qquad (2)
$$

R_i (i = I,...,4) - множество, полученное упорядочением всех чисел полного периода псевдопоследовательности. Зная τ_n и k, легко определить, какой из четырех последовательностей принадлежит наша. Найдем среднее полного периода для каждой последовательности:

$$\tau_{cp} = \frac{1}{2^{p-2}} \sum_{i=0}^{2^{p-2}-1} \tau_i .$$

I. $\tau_{cp}(R_1) = 2^{p-1}-2$, 3. $\tau_{cp}(R_3) = 2^{p-1}-1$,

2. $\tau_{cp}(R_2) = 2^{p-1}+2$, 4. $\tau_{cp}(R_4) = 2^{p-1}+1$

или после нормировки, т.е. деления на 2^p , соответственно получим:

I. $\tau_{cp}(R_1) = 1/2 - 1/2^{p-1}$, 3. $\tau_{cp}(R_3) = 1/2 - 1/2^p$,

2. $\tau_{cp}(R_2) = 1/2 + 1/2^{p-1}$, 4. $\tau_{cp}(R_4) = 1/2 + 1/2^p$.

Отсюда видно, что псевдопоследовательности с $k \equiv 5 \pmod 8$ более предпочтительны, чем с $k \equiv 3 \pmod 8$,так как их средние ближе к I/2. Однако для современных вычислительных машин p достаточно велико и различием между множителями k можно пренебречь.

Из (I) очевидно,что $\tau_{n+s} \equiv k^s \tau_n \pmod{2^p}$, а из (2),что $\min_i \{R_1 \cap R_3\} = 1$ и $\min_i \{R_2 \cap R_4\} = 7$. Тогда $\tau_n \equiv k^n \pmod{2^p}$ определяет последовательность R_1 (если $k \equiv 3 \pmod 8$) и R_3 (если $k \equiv 5 \pmod 8$)), а $\tau_n \equiv 7k^n \pmod{2^p}$определяет последовательности R_2 и R_4.

Если в (I)заменить τ_n на дополнение τ_n ,т.е. на $2^p - \tau_n$,то это будет означать переход к другой последовательности. В

самом деле, если $\tau_n \in R_1$, то $2^p - \tau_n \in R_2$, а при $\tau_n \in R_3$ $2^p - \tau_n \in R_4$.

Так как $k(2^p - \tau_n) \pmod{2^p} \equiv (2^p - k)\tau_n \pmod{2^p} \equiv (2^p - k\tau_n) \pmod{2^p}$ и $\tau_{n+1} \equiv (2^p - k)(2^p - \tau_n) \pmod{2^p} \equiv k\tau_n \pmod{2^p}$, то отсюда следует, что k и $2^p - k$ равноценны в смысле статистических качеств полученных псевдопоследовательностей, ибо дополнение $2^p - \tau_n$ – это изменение всех двоичных разрядов числа τ_n на обратный разряд, кроме младшего: он всегда равен единице.

Пусть в датчике (I):

$$\tau_n = \varepsilon_1\, \varepsilon_2\, \varepsilon_3 \cdots \varepsilon_{p-1}\, \varepsilon_p; \qquad \text{где} \quad \varepsilon_p = \eta_p = \xi_p = 1 ;$$

$$k = \eta_1\, \eta_2\, \eta_3 \cdots \eta_{p-1}\, \eta_p; \qquad \varepsilon_i, \eta_i, \xi_i = \{0,1\} (1 \le i \le p-1);$$

$$\tau_{n+1} = \xi_1\, \xi_2\, \xi_3 \cdots \xi_{p-1}\, \xi_p .$$

Так запишется двоичное представление чисел τ_n, k , τ_{n+1} . Так как $k \equiv 3; 5 \pmod 8$,то $\eta_{p-2} + \eta_{p-1} \equiv 1 \pmod 2$. Найдем результат произведения, т.е. найдем значения разрядов нового псевдослучайного числа.

$$
\begin{aligned}
&\xi_{p-1} \equiv (\xi_{p-1} + \eta_{p-1}) \pmod 2 ;\\
&\xi_{p-2} \equiv [(2\varepsilon_{p-2} + 2\eta_{p-1}\varepsilon_{p-1} + 2\eta_{p-2} + \varepsilon_{p-1} + \eta_{p-1})/2] \pmod 2 ;\\
&\xi_{p-3} \equiv [(4\varepsilon_{p-3} + 4\eta_{p-1}\varepsilon_{p-2} + 4\eta_{p-2}\varepsilon_{p-1} + 4\eta_{p-3} + 2\varepsilon_{p-2} + 2\eta_{p-1}\varepsilon_{p-1} +\\
&\quad + 2\eta_{p-2} + \varepsilon_{p-1} + \eta_{p-1})/2^2] \pmod 2 ;\\
&\xi_1 \equiv [(2^{p-2}\varepsilon_1 + 2^{p-2}\eta_{p-1}\varepsilon_2 + 2^{p-2}\eta_{p-2}\varepsilon_3 + \cdots + 2^{p-2}\eta_2\varepsilon_{p-1} + 2^{p-2}\eta_1 +\\
&\quad + 2^{p-3}\varepsilon_2 + 2^{p-3}\eta_{p-1}\varepsilon_3 + 2^{p-3}\eta_{p-2}\varepsilon_4 + \cdots + 2^{p-3}\eta_3\varepsilon_{p-1} + 2^{p-3}\eta_2 +\\
&\quad + 2^{p-4}\varepsilon_3 + 2^{p-4}\eta_{p-1}\varepsilon_4 + 2^{p-4}\eta_{p-2}\varepsilon_5 + \cdots + 2^{p-4}\eta_4\varepsilon_{p-1} + 2^{p-4}\eta_3 +\\
&\quad + 2^2\varepsilon_{p-3} + 2^2\eta_{p-1}\varepsilon_{p-2} + 2^2\eta_{p-2}\varepsilon_{p-1} + 2^2\eta_{p-3} + 2\varepsilon_{p-2} + 2\eta_{p-1}\varepsilon_{p-1} +\\
&\quad + 2\eta_{p-2} + \varepsilon_{p-1} + \eta_{p-1})/2^{p-2}] \pmod 2 .
\end{aligned}
\qquad (3)
$$

Дальнейшие рассуждения будут касаться только значения ξ_1 – старшего и, естественно, наиболее важного разряда нового псевдослучайного числа z_{n+1}. Легко заметить, что подмодульное выражение для разряда ξ_i входит в подмодульное выражение разрядов $\xi_{i-1}, \xi_{i-2}, \dots, \xi_1$ (более старших) линейной независимой частью. Поэтому если будут замечены корреляционные связи между старшими разрядами псевдослучайных чисел, то между младшими разрядами эти связи могут быть выражены только в более сильной степени. Кроме того, ξ_i будет старшим разрядом для датчика $z_{n+1} \equiv k z_n \pmod{2^{\rho+1-i}}$, и, заменяя ρ на $\rho+1-i$ в (3), получим интересующее нас ξ_i по формуле для ξ_1.

Сгруппируем подмодульное выражение для ξ_1 из формул (3):

$$\xi_1 \equiv \Big[\Big\{2^{\rho-2}\varepsilon_1 + (2^{\rho-2}\eta_{\rho-1} + 2^{\rho-3})\varepsilon_2 + (2^{\rho-2}\eta_{\rho-2} + 2^{\rho-3}\eta_{\rho-1} + 2^{\rho-4})\varepsilon_3 + \dots$$
$$+ (2^{\rho-2}\eta_2 + 2^{\rho-3}\eta_3 + 2^{\rho-4}\eta_4 + \dots + 2^2\eta_{\rho-2} + 2\eta_{\rho-1} + 1)\varepsilon_{\rho-1} +$$
$$+ 2^{\rho-2}\eta_1 + 2^{\rho-3}\eta_2 + 2^{\rho-4}\eta_3 + \dots + 2^2\eta_{\rho-3} + 2\eta_{\rho-2} + \eta_{\rho-1}\Big\} \Big/ 2^{\rho-2}\Big] \pmod 2. \qquad (4)$$

Здесь и дальше $[x]$ означает взятие целой части числа x.

Пусть все $\eta_i = 0 \ (1 \le i \le \rho-1)$, кроме $i = \rho - s$. Тогда

$$\xi_1 \equiv \Big[\varepsilon_1 + \frac{\varepsilon_2}{2} + \frac{\varepsilon_3}{4} + \dots + \frac{\varepsilon_{\rho-1}}{2^{\rho-2}} + \varepsilon_{s+1} + \frac{\varepsilon_{s+2}}{2} + \dots + \frac{\varepsilon_{\rho-1}}{2^{\rho-2-s}} + \frac{1}{2^{\rho-1-s}}\Big] \pmod 2.$$

В данном примере $\eta_{\rho-s} = 1$. ξ_1 определяется в основном значениями ε_1 и ε_{s+1}; ε_2 и ε_{s+2} оказывают влияние тогда, когда $\varepsilon_2 \cdot \varepsilon_{s+2} = 1$ или когда $\varepsilon_2 + \varepsilon_{s+2} = 1$, но $\varepsilon_3 \cdot \varepsilon_{s+3} = 1$ и т.д. Налицо прямая связь между ξ_1 и $(\varepsilon_1 + \varepsilon_{s+1}) \pmod 2$ и слабая (в том смысле, что она мало влияет на результат) зависимость ξ_1 от остальных разрядов.

С другой стороны, пусть все $\eta_i = 1$, кроме $i = \rho - s$.

$$\xi_1 \equiv \Big[2\varepsilon_1 + 2\varepsilon_2 + \dots + 2\varepsilon_{\rho-1} + 2 - \varepsilon_1 - \frac{\varepsilon_2}{2} - \dots$$
$$\dots - \frac{\varepsilon_{\rho-1}}{2^{\rho-2}} - \frac{1}{2^{\rho-2}} - \varepsilon_{s+1} - \frac{\varepsilon_{s+2}}{2} - \dots\Big] \pmod 2.$$

Здесь $\eta_{\rho-s} = 0$. Как и в предыдущем примере, наблюдается прямая связь между ξ_1 и $(\varepsilon_1 + \varepsilon_{s+1}) \pmod 2$ и слабая зависимость ξ_1 от остальных разрядов z_n. Впрочем, этот результат можно было ожидать в связи с замечанием об эквивалентности k и $2^\rho - k$. Так как плохо, если в двоичном разложении много нулей или много единиц, то очевидно, что число нулей и единиц в двоичном разложении k должно быть близким друг другу. Идеальным был бы тот недостижимый случай, когда все $\eta_i = \frac{1}{2}, (1 \le i \le \rho-1)$. Тогда

$$\xi_1 \equiv \Big[\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{\rho-1} + 1 - \frac{1}{2^{\rho-1}}\Big] \pmod 2 \equiv \Big(\sum_{i=1}^{\rho-1} \varepsilon_i\Big) \pmod 2.$$

Здесь ξ_1 зависит в одинаковой степени от всех разрядов z_n. Достигнута максимальная "случайность" ξ_1.

Казалось бы, просто чередуя нули и единицы в двоичном разложении k, получим неплохие результаты, однако это не так. Пусть (для простоты) ρ – четное; $\eta_{\rho-1} = 0$, $\eta_{\rho-2} = 1$, $\eta_{\rho-3} = 0, \dots, \eta_2 = 1$, $\eta_1 = 0$. Тогда

$$\xi_1 \equiv \Big[\Big\{\frac{2^\rho - 2^{\rho-2}}{3}\varepsilon_1 + 2\,\frac{2^{\rho-2} - 2^{\rho-4}}{3}\varepsilon_2 + \frac{2^\rho - 2^{\rho-4}}{3}\varepsilon_3 + 2\,\frac{2^{\rho-2} - 2^{\rho-6}}{3}\varepsilon_4 + \dots$$
$$\dots + \frac{2^\rho - 1}{3}\varepsilon_{\rho-1} + 2\,\frac{2^{\rho-2} - 1}{3}\Big\} \Big/ 2^{\rho-2}\Big] \pmod 2 \equiv \Big[\frac{1}{3}\Big\{4\sum_{i=0}^{\rho/2-1}\varepsilon_{2i+1} +$$
$$+ 2\sum_{i=1}^{\rho/2-1}\varepsilon_{2i} - \sum_{i=1}^{\rho-1}\varepsilon_i/2^{i-1} + 2 - 1/2^{\rho-3}\Big\}\Big] \pmod 2; \qquad (5)$$

$1 < \sum_{i=1}^{\rho-1}\varepsilon_i/2^{i-1} + 1/2^{\rho-3} < 2$, если $\varepsilon_1 = 1$, и $0 < \sum_{i=1}^{\rho-1}\varepsilon_i/2^{i-1} + 1/2^{\rho-3} < 1$, если $\varepsilon_1 = 0$ для всех наборов ε_i, для которых $\varepsilon_2 \cdot \varepsilon_3 \cdot \dots \cdot \varepsilon_{\rho-4} \cdot \varepsilon_{\rho-3} = 0$.

Предполагая наборы ε_i случайными, находим вероятность появления набора иного вида $1/2^{\rho-6}$. При больших ρ такими наборами можно пренебречь.

Обозначим $\sum_{i=0}^{\rho/2-1}\varepsilon_{2i+1} = N_1$, $\sum_{i=1}^{\rho/2-1}\varepsilon_{2i} = N_2$.

I. $\varepsilon_1 = 1$. Тогда выражение (5) можно записать так: $\xi_1 \equiv \Big[\frac{1}{3}\Big\{4(N_1 - 1) + 2N_2 + 4\Big\}\Big] \pmod 2$. Составим небольшую табличку значений ξ_1 в зависимости от $N_1 - 1$ и N_2 (в предположении о случайности и равновероятности $\varepsilon_i = 0$ и $\varepsilon_i = 1$).

2. $\varepsilon_1 = 0$. Тогда $\xi_1 \equiv \left[\frac{1}{3}\left\{4N_1 + 2N_2 + 1\right\}\right] (mod 2)$.

Составим табличку и для этого случая.

N_1-1 \ N_2	0	I	2
0	I	0	0
I	0	I	0
2	0	0	I

N_1 \ N_2	0	I	2
0	0	I	I
I	I	0	I
2	I	I	0

При $\varepsilon_1 = 1$ вероятность $P\{\xi_1 = 0\} = \frac{2}{3}$, а при $\varepsilon_1 = 0$ $P\{\xi_1 = 1\} = \frac{2}{3}$. Датчик с таким k должен давать слишком много коротких серий. Проверка на ЭВМ полностью подтвердила это заключение. Слишком правильное чередование нулей и единиц в k отрицательно влияет на статистические качества псевдопоследовательности.

Пусть двоичное разложение k имеет s единиц ($s < p/2$).Из формулы (4) видно, что ξ_1 зависит от s разрядов ε_i , s других разрядов ε_j с коэффициентом $\frac{1}{2}$, s третьих разрядов ε_ℓ с коэффициентом $1/4$ и т.д. Допуская, что $\frac{1}{2}(\varepsilon_i + \varepsilon_j)$ приравниваем к одному разряду, получаем зависимость ξ_1 от $2s$ разрядов ε_i . Датчик с таким k должен давать вероятностное смещение порядка $1/2^{2s}$. Однако, как показывает нижеприведенный пример, даже при сравнительно небольших s обнаружить аномали весьма трудно. Разумеется, этот вывод справедлив и для s нулей ($s < p/2$) в двоичном разложении k .

В качестве примера рассмотрим довольно хорошо проверенный датчик [5] с $k = 2^{18} + 11$. В двоичном коде k запишется так:

000 000 000 000 000 001 000 000 000 000 001 011.

$p = 36$. Подставляя в (4), получим:

$$\xi_1 = \left[\varepsilon_1 + \frac{3}{2}\varepsilon_2 + \frac{3}{4}\varepsilon_3 + \frac{11}{8}\varepsilon_4 + \frac{11}{16}\varepsilon_5 + \cdots + \frac{11}{2^{18}}\varepsilon_{19} + \left(\frac{1}{2} + \frac{11}{2^{19}}\right)\varepsilon_{20} + \cdots + \left(\frac{1}{2^{18}} + \frac{11}{2^{36}}\right)\varepsilon_{35} + \frac{1}{2^{37}} + \frac{5}{2^{36}}\right] (mod 2).$$

Здесь ξ_1 определяется в основном значениями ε_1 , ε_2 , ε_3 , ε_4 , ε_{19} , ε_{20} , т.е. 7 – 8 разрядами из 36. Избыток нулей в разложении k привел к тому, что на старший разряд нового числа не оказывают почти никакого влияния $4/5$ всех значащих разрядов. τ_n . Псевдопоследовательность с таким k испытывает вероятностное смещение порядка $1/2^5$, однако тестовая проверка датчика не выявила этого отклонения [5].

Дополнительно была предпринята ещё одна тестовая проверка: получено τ_n .

Ищется такое $s \geq 1$, чтобы $|\tau_n - \tau_{n+\ell}| > \frac{1}{2}$ для $\ell < s$, но $|\tau_n - \tau_{n+s}| < \frac{1}{2}$, т.е. ищутся серии чисел, отстоящих от τ_n более чем на $\frac{1}{2}$. Набирается статистика по числу подобных серий.Вероятность серии длины s равна

$$P(s) = 2\int_0^{1/2} (1/2 - x)^{s-1}(1/2 + x)dx = \frac{s+2}{s(s+1)2^s}; \quad \sum_{s=1}^{\infty} \frac{s+2}{s(s+1)2^s} = 1.$$

Длина выборки $2^{17} \approx 130000$. Проверялись 4 датчика.

1. $k = 2^{18} + 11$.
2. $k = 2525\ 2525\ 2525_8$.
3. $k = 4301\ 7417\ 0715_8$.
4. $k = 7700\ 3701\ 6145_8$.

Проверка показала, что датчики 3 и 4 хорошо согласуются с гипотезой о случайности процесса. Датчик 2 с чередованием нулей и единиц дает очень плохой результат, а датчик 1 удовлетворяет критерию случайности с 5%-ным уровнем значимости (находился χ^2), но отвергается с 10%-ным уровнем.

Как и прежде, аномалию достаточно четко обнаружить не удалось, т.к. величина $\frac{1}{2^{2s}} = \frac{1}{256}$ достаточно мала (см.выше).

Для того, чтобы избежать периодического чередования нулей и единиц в двоичном разложении k , предлагается находить такие суммы $S(m)$:

$$S(m) = \sum_{i=1}^{p} \left\{\eta_i + \eta_{(i+m)(mod\,p)}\right\} (mod 2),$$

где $m = 1, 2, \ldots, \left[p/2\right]$.

Если k выбрано с числом нулей и единиц, близким друг другу, и при каждом m величина $S(m)$ далека как от 0, так и от p, то такое k можно рекомендовать для счета. Представляется затруднительным дать точные границы для $S(m)$. Предлагается пользоваться таким правилом: $|S(m)-[p/2]| \leq [p/4]$ для каждого m.

Тема предложена Б.В. Чириковым.

ЛИТЕРАТУРА

1. E. BOFINGER, V.J. BOFINGER. On a Periodic Property of Pseudo-Random Sequences. - Journ. Assoc.Comp.Mach.,5(1958) p.261-265.

2. M. GREENBERGER. Notes on a New Pseudo-Random Number Generator. ibid. 8 (1961), p.163-167.

3. D. MacLAREN, G. MARSAGLIA. Uniform Random Number Generators ibid. 12, N 1 (1965), p.83-89.

4. V.D. BARNETT. The Behaviour of Pseudo-Random Sequences Generated on Computers by the Multiplicative Congruen - tial Method. - Math. Comp. 16 (1962), p.63-69.

5. М.В. АНТИПОВ, Ф.М. ИЗРАЙЛЕВ, Б.В. ЧИРИКОВ. Статистическая проверка датчика псевдослучайных чисел. - Вычислительные системы, Новосибирск, "Наука" СО, 1968, вып.30, стр. 77-85.

Поступила в редакцию
17 июля 1970 г.