объединенный
институт
ядерных
исследований
дубна

P11-80-520

Г.А.Ососков

ПРОГРАММНЫЙ ГЕНЕРАТОР ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ ДЛЯ МИКРО-ЭВМ

Направлено на 6 Международный симпозиум
по мини- и микро-ЭВМ и их приложениям
/Будапешт, 9-12 сентября, 1980 г./

1980

1. ВВЕДЕНИЕ

Широкое распространение методов статистического моделирования /Монте-Карло/ стимулировало целый ряд исследований по выработке последовательностей случайных чисел на ЭВМ. Во всех современных руководствах по методам Монте-Карло подчеркиваются несомненные преимущества методов программной генерации последовательностей чисел, которые в силу рекуррентной природы их получения должны быть зависимы, но при тщательно выбранном алгоритме генерации могут вести себя как случайные с точки зрения самых строгих статистических критериев. Поэтому такие числа называют псевдослучайными. Для ЭВМ с длиной слова более 32 разрядов наиболее употребительными являются программные генераторы случайных чисел /ГСЧ/, основанные на известном методе вычетов П.Лемера /1/. Последовательность целых чисел $\{X_n\}$, образуемая рекуррентно с помощью линейного сравнению по модулю 2^p

$$X_{n+1} \equiv M \cdot X_n + C (\mathrm{mod}\, 2^p), \quad n = 1, 2, \ldots \qquad /1/$$

порождает последовательность чисел $U_n = X_n \cdot 2^{-p}$, которая при правильном выборе параметров M, C и X_0 будет равномерно распределена в интервале /0,1/. При $C \neq 0$ ГСЧ называется смешанным конгруентным для отличия от мультипликативного ГСЧ, у которого $C = 0$.

Априори можно предположить, что не при всех значениях M, C и X_0 в последовательности /1/ встретятся все 2^p возможных чисел от 0 до $2^p - 1$. Может оказаться, что после некоторой неповторяющейся группы чисел $X_0, X_1, \ldots X_L$ мы получим, что X_{L+1} повторяет какое-то число $X_k (k < L)$, так что эта группа из $T = L - k$ чисел начнет потом повторяться в силу /1/. В этом случае начальный отрезок числовой последовательности называется отрезком апериодичности, а T - периодом. Так будет, например, при четном M, когда независимо от выбора C и X_0 последовательность будет иметь отрезок апериодичности, не превосходящий p, и период $T = 1$. Поэтому всюду в дальнейшем будет считаться, что M - нечетно, т.е. имеет вид $M = 2^k \cdot m \pm 1$ $(k \geq 2)$. Таким же естественным ограничением будет нечетность C, поскольку в противном случае мы фактически сократим нашу последовательность вдвое, так как в зависимости от четности X_0 будем получать в /1/ либо только четные, либо только нечетные числа.

При этих ограничениях можно показать[2], что в зависимости от вида M период T будет определяться соотношением

$$T = \begin{cases} 2^p, & M = 2^k \cdot m + 1 \\ 2^{p-k+1}, & M = 2^k \cdot m - 1 \end{cases} \qquad k \geq 2. \qquad /2/$$

В случае мультипликативного ГСЧ (C=0) имеем

$$T = 2^{p-k-\ell}, \qquad /3/$$

где ℓ определяется представлением стартового значения $X_0 = 2^\ell \cdot B$ / B - нечетное/.

Таким образом, при нечетном C первое из двух возможных представлений $M = 2^k \cdot m + 1$ обеспечивает полный период последовательности $T = 2^p$ независимо от X_0, а в случае мультипликативного ГСЧ нечетность X_0 обеспечивает при $k = 2$ и нечетном m вчетверо меньший максимальный период $T = 2^{p-2}$. ГСЧ с полным или максимальным периодом вовсе не обязательно должны иметь хорошие свойства. Для этого следует выбрать подходящие значения параметров m и C. Широкий диапазон рекомендаций по выбору параметров можно найти, например, в книгах[3,4], где для ЭВМ с большой длиной слова /35 разрядов и выше/ рекомендуются главным образом мультипликативные ГСЧ с множителями M вида 5^{15}, 5^{17} и т.п.

2. ТЕСТЫ ДЛЯ ПРОВЕРКИ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

Эти рекомендации основаны на различных тестах для проверки качества ГСЧ, которые, следуя Кнуту[3], можно разбить на два класса: эмпирические и теоретические тесты.

Тесты первого класса основаны на различных статистических критериях и прилагаются к последовательности чисел, рассматриваемой как случайная выборка, безотносительно от способа ее получения. Такие тесты удобны тем, что они пригодны для проверки любых ГСЧ не обязательно вида /1/. Достаточно полный набор эмпирических тестов можно найти, например, в[3].

Теоретические тесты второго класса не требуют рассмотрения выборки, а основаны на теоретико-числовом исследовании рекуррентного соотношения, порождающего псевдослучайную последовательность, для получения аналитических выводов, относительно ее статистических свойств. Например, для таких важных статистических критериев проверки на случайность, как длина периода /см. /2-3/ выше/, число серий нулей и единиц в выбранном разряде двоичного представления чисел проверяемой последовательности, коэффициент корреляции между этими числами, в[2] выведены аналитические оценки, сделанные по всему периоду изменения чисел.

Наиболее полным из теоретических тестов, применимых, однако, только к ГСЧ с полным периодом, считаются спектральный тест, подробно описанный в книге[3], и менее известный тест, проверяющий решетчатость[5], основанные на исследовании решетчатой структуры распределения точек с псевдослучайными координатами в n-мерном пространстве.

3. ПСЕВДОСЛУЧАЙНЫЕ ЧИСЛА ДЛЯ МИНИ- И МИКРО-ЭВМ

Широкое распространение мини- и микро-ЭВМ для целей управления и контроля аппаратуры показало, что проблема программной генерации псевдослучайных чисел осталась по-прежнему актуальной, хотя главный акцент их использования переносится со сложных задач многомерного моделирования на задачи тестовой проверки аппаратуры и каналов связи /см., например,[6] /.

Непосредственная реализация на малых ЭВМ, рекомендованных теорией линейных конгруентных методов, может встретить серьезные затруднения. Если для 16-разрядной ЭВМ можно использовать, например, ГСЧ /1/ с $M = 44373 \equiv 5^{15} (\mathrm{mod}\, 2^{16})$, что при нечетном C гарантирует сравнительно малый период $T = 65536$, то для ЭВМ с длиной слова 12 и менее бит для обеспечения приемлемого периода ГСЧ придется использовать арифметику с удвоенной /или даже утроенной/ точностью. Кроме того, отсутствие у многих мини-ЭВМ аппаратного блока расширенной арифметики делает неэффективным алгоритмы, основанные на умножении на очень большие множители.

В этой связи возник интерес к другим способам реализации линейных конгруентных ГСЧ типа /1/. Выбор множителя в /1/ вида $M = 2^k + 1$ позволяет реализовать умножение на M по модулю 2^p с помощью двух простых операций: сдвига числа X_n влево на k разрядов /с потерей старших разрядов/ с последующим сложением X_n с результатом сдвига.

Вышеупомянутые аналитические оценки для числа серий R и коэффициентов корреляции ρ псевдослучайной последовательности с полным периодом 2^p показывают, что для ГСЧ с M вида $2^k + 1$ оптимальное с точки зрения малости среднеквадратичных ошибок значение константы сдвига k достигается при $k = p/2$. Заметим, что сдвиг на половину разрядной сетки является также достаточно медленной операцией, однако если увеличить вдвое разрядность случайных чисел /что и требовалось для получения достаточно большого периода ГСЧ/, то сдвиг теперь уже на $k = p$ разрядов можно осуществить как пересылку числа из одной ячейки ЭВМ, содержащей правую половину случайного числа, в другую ячейку, где находятся старшие разряды X_n.

Программы для мини-ЭВМ с 12 и 16-разрядными словами, реализующие этот новый ГСЧ /назовем его ГСЧМ/, приводятся в /7/ вместе с результатами статистических тестов, показавшими вполне удовлетворительные свойства одномерных случайных чисел и распределений их пар на плоскости. Реализация ГСЧМ на микро-ЭВМ с однобайтовыми словами путем размещения X_n в четырех байтах не представляет затруднений.

Тем не менее, в полном соответствии с предостережением Д.Кнута об опасности применения множителей M вида 2^k+1 попытки использования ГСЧМ для генерации точек в трехмерном пространстве сразу показали на наличие их неравномерности в малых объемах. Например,обнаружилось, что никакие две последовательные точки из группы в 10 тыс. не попадают вместе в сферу радиуса 10^{-1}, хотя число таких случаев с вероятностью 99% должно было превысить 20.

Причиной этого является решетчатая структура распределения псевдослучайных векторов (U_n, U_{n+1}, U_{n+2}) в единичном кубе, лежащих в параллельных плоскостях /3/, п.3.3.4,упр.27/.В табл.1 приведены данные применения теста на решетчатость, в котором вычисляется максимальное отношение сторон ячеек решетки, образованной этими гиперплоскостями /чем больше отличается это отношение от 1, тем сильнее уклоняется распределение псевдослучайных чисел от равномерного/. Из данных табл.1 можно сделать вывод об отсутствии универсальной константы сдвига k, пригодной для конструирования ГСЧМ, "работающих" при любой размерности.

Таблица 1

Максимальное отношение сторон ячеек пространственной решетки для ГСЧМ с модулем 2^{2p}

p	k	$M=2^k+1$	Размерность пространства 2	3	4
12	6	65	3970,1	61,1	1,1
	8	257	254,0	1,0	40,6
	12	4097	1,0	1182,4	1121,7
16	8	257	65026,0	253,0	1,0
	11	2049	1023,0	1,4	228,0
	16	65537	1,0	18918,6	17947,8

4. УЛУЧШЕНИЕ КАЧЕСТВА ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

Если подобное затруднение с генерацией псевдослучайных чисел на малой ЭВМ для использования их в качестве координат точек в многомерных пространствах не может быть преодолено на пути использования расширенной арифметики с удвоенной точностью, то можно обратиться к другим способам получения псевдослучайных чисел, а также методам улучшения качества для имеющихся ГСЧ. Подобные методы либо не используют линейный конгруентный генератор, либо направлены на разрушение решетчатой структуры псевдослучайной последовательности. Поэтому теоретические тесты для их проверки неприменимы, и следует использовать статистические способы проверки.

В дополнение к известным методам проверки /3/ нами использовались специальные тесты для проверки на равномерность в малых объемах. С этой целью вычислялись 9 значений χ^2 по частным гистограммам, разбивающим на 1000 ячеек области, расположенные в углах и центре единичного куба.

Второй тест - число совпадений двух подряд идущих трехмерных векторов с точностью до 1-2 десятичных знаков.

Наиболее простой способ улучшения качества псевдослучайной последовательности - подвергнуть достаточно большие ее отрезки перемешиванию с помощью таблицы на 64÷128 ячеек, задающей некоторый фиксированный порядок выдачи чисел последовательности.

Такое блочное перемешивание, выполненное с помощью табл.2, привело к значительному улучшению качества псевдослучайной последовательности, выдаваемой ГСЧМ с P=16.Результаты применения тестов на равномерности в 9 областях и на случайные совпадения приведены в табл.3.

Если на каждом шаге менять таблицу для перемешивания с помощью второго независимого ГСЧ /метод был предложен Макләреном и Марсальей /8/ /,то мы получим еще более равномерную последовательность. Метод успешно использовался при создании ГСЧ для ЭВМ БЭСМ-6 /7/.Однако в случае микро-ЭВМ программная реализация метода сдвоенных генераторов получается слишком громоздкой.

Гораздо более экономичный генератор для наиболее известных микро-ЭВМ типа ИНТЕЛ-8080 был разработан автором с помощью дальнейшего развития идеи об использовании представления случайных чисел в виде нескольких слов ЭВМ. Соответствующий рекурсивный алгоритм генератора /названного ГСЧИ/ основан на "перемешивании" четвертей предыдущего случайного числа для получения последующего. Детальное описание алгоритма дано в Приложении в виде подпрограммы-функции на языке ФОРТРАН, а также на автокоде ИНТЕЛ-8080. Следует отметить специальную обработку переполнений при сложении: единицы переноса между байтами суммируются по модулю 2.

Таблица 2

Порядок извлечения чисел, выдаваемых ГСЧМ, из таблицы на 64 ячейки при блочном перемешивании

8,55,46,47,42, 2, 5,62,28,25,39,58, 5,37,59,37, 8,5I,38,I8,29,I5,40, 52, 9,34,4I,48,45,33,48,36,20,32,57,52, 3,58,55,45,33,I3,I2,20,55, 8,I5,53,62,35,37,47,24,42,33,57,25,49,57, 9, 5,49,59

Таблица 3

Значения тестов для ГСЧМ после перемешивания с помощью табл.2. 1°. χ^2_{1000} , вычисленные для 9 подобластей единичного куба. Размер ячеек гистограмм 0,05x0,05x0,05 . 5% - критическое значение составляет 1145. 2°. Число совпадений триплетов с точностью 0,1, Значения 5% доверительного интервала даны в скобках

		Число испытаний /в тыс./		
		20	60	100
1°.	1	941	1097	1092
	2	987	1021	1115
	3	1003	1046	1124
	4	909	992	1052
	5	1040	1091	1098
	6	997	1004	1006
	7	1033	1047	1134
	8	1019	953	1084
	9	951	1042	1116
2°.		84/55,110/	252/203,298/	429/356,479/

В табл.4 приведены результаты проверки ГСЧИ по тестам на равномерность в малых объемах, а также по одному из наиболее сильных статистических критериев - тесту на монотонность подпоследовательностей нулей и единиц /см. /3/, разд.3.3.2/. Оценка периода ГСЧИ с помощью ЭВМ показала, что период превышает $2 \cdot 10^6$.

Таблица 4

Значения тестов для ГСЧИ. 1°. χ^2_{1000}, вычисленные для 9 подобластей единичного куба. Размер ячеек гистограмм тот же, что и в табл.3. 2°- число совпадений триплетов с точностью 0,1. 3°. χ^2_{10} для теста на монотонность. 5% - критическое значение = 18,3

		Число испытаний /в тыс./		
		20	60	100
1°	1	1009	965	1044
	2	1003	976	981
	3	958	923	999
	4	1051	1047	1032
	5	989	925	1015
	6	946	918	1025
	7	1105	992	989
	8	972	953	949
	9	965	976	1028
2°.		70/55,110/	240/203,298/	400/356,479/
3°.		5,37	5,50	6,46

5. ЗАКЛЮЧЕНИЕ

Из двух генераторов, предложенных выше, ГСЧИ более ориентирован на микро-ЭВМ типа ИНТЕЛ-8080.

Блочное перемешивание с помощью табл.2 чисел, выдаваемых ГСЧМ, может применяться для 12-16 разрядных мини-ЭВМ. Для простых вычислений, использующих только одномерные и двумерные случайные последовательности, ГСЧМ вполне применим и без всякого перемешивания.

Автор признателен доктору А.Аткинсону за полезные рекомендации и Х.Лайху за помощь в программировании на микро-ЭВМ.

ПРИЛОЖЕНИЕ

1. Фортранный вариант ГСЧИ

k - фиктивный параметр. Вызывающая программа должна содержать операторы, задающие начальные значения случайных чисел:

```
COMMON/IJ/ II(5),JJ(4)
DATA(II=205B,54B,321B,234B,205B),(JJ=273B,13B,311B,115B)
```

```
      FUNCTION RNGI(K)
      COMMON/IJ/ II(5),JJ(4)
      M=256
      MS=255
      DO 3L=1,4
      II(L)=II(L)+JJ(L)+II(L+1)
      II(5)=II(1)
      IF(II(L).AND.M.EQ.M) II(L+1)=II(L+1)+1
    3 CONTINUE
      I=SHIFT(II(4),8).OR.II(3)
      RNGI=I/FLOAT(M* M)
      RETURN
      END
```

2. Подпрограмма, реализующая ГСЧИ на автокоде ИНТЕЛ-8080

В регистрах B и C находится двухбайтовый параметр подпрограммы, который является адресом первого слова массива X, состоящего из 5 байтов. В качестве текущего случайного числа используются третий и четвертый байты массива X. Перед первым обращением к подпрограмме в массивы X и Y должны быть засланы их начальные значения /ими могут быть, например, те же числа, что подлежат засылке в массивы II и JJ в п.1/.

```
RNG: LXI   H,AX+1H
     MOV   M,B
     DCX   H
     MOV   M,C
     LXI   H,CAR
     MVI   M,ØH
M1:  LXI   H,I
     MVI   M,ØH
Q3:  MVI   A,2H
     LXI   H,I
     CMP   M
     JC    Q4
     LDA   CAR
     ANI   1H
     LHLD  I
     MVI   H,Ø
     LXI   B,Y
     DAD   B
     ADD   M
     LHLD  I
     MVI   H,Ø
     XCHG
     LHLD  AX
     DAD   D
     ADD   M
     MOV   M,A
     LXI   H,CAR
     MVI   M,ØH
     SBB   A
     CPI   ØFFH
     JNZ   Q1
     LXI   H,CAR
     INR   M
Q1:  LHLD  I
     MVI   H,Ø
     XCHG
     LHLD  AX
     DAD   D
     PUSH  H      ;1
     LHLD  I
     MVI   H,Ø
     LXI   B,X+1H
     DAD   B
     XCHG
     LHLD  AX
     DAD   D
     MOV   A,M
     POP   H      ;1
     ADD   M
     MOV   M,A
     SBB   A
     CPI   ØFFH
     JNZ   Q2
     LXI   H,CAR
     INR   M
Q2:  LXI   H,I
     INR   M
     JNZ   Q3
Q4:  LDA   CAR
     ANI   1H
     LXI   H,Y+3H
     ADD   M
     LXI   B,3H
     LHLD  AX
     DAD   B
     ADD   M
     PUSH  H      ;1
     LHLD  AX
     ADD   M
     POP   H      ;1
     MOV   M,A
     RET
```


ЛИТЕРАТУРА

1. Lehmer P.H. Mathematical Methods in Large-Scale Computing Units. Ann.Comp.Lab. Harvard University, 1951, p.26.
2. Акишин П.Г., Ососков Г.А. ОИЯИ, Р5-8411, Дубна, 1974.
3. Кнут Д. Искусство программирования для ЭВМ, т.2. "Мир", М., 1977.
4. Михайлов Г.А. Некоторые вопросы теории методов Монте-Карло. "Наука", Новосибирск, 1974.
5. Marsaglia G. The Structure of Linear Congruential Secuences. In: Applications of Number Theory to Numerical Analysis. Ed. by S.K.Zaremba. Acad.Press, N.-Y., 1972.
6. Соучек Б. Мини-ЭВМ в системах обработки информации. "Мир", М., 1976.
7. Ососков Г.А. Программные генераторы псевдослучайных чисел для малоразрядных ЭВМ. В кн.: Совм. научн.сб. ОИЯИ-ЦИФИ, вып.2, 1977, с.12.
8. Maclaren M.D., Marsaglia G. Uniform Random Number Generators. J.ACM, 1965, vol.12, No.11.

Рукопись поступила в издательский отдел
25 августа 1980 года.