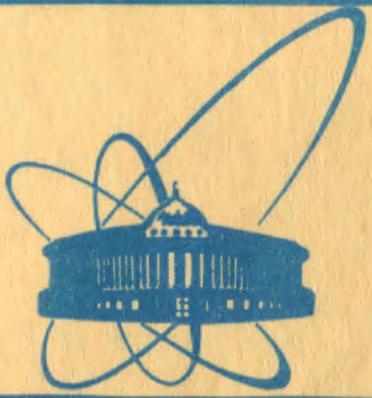

сообщения
объединенного
института
ядерных
исследований
дубна

10-81-536

В.Н.Аносов, Г.П.Лещенко

ГЕНЕРАТОР ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ С РАВНОМЕРНЫМ РАСПРЕДЕЛЕНИЕМ ДЛЯ МИНИ-ЭВМ

1981

1. ВВЕДЕНИЕ

В работе [1] описан генератор случайных чисел /ГСЧ/ с равномерным распределением на основе мультипликативного конгруэнтного метода для ЭВМ ЕС-1010, имеющий длину периода $2^{13}/10^4/$ при нечетных начальных числах, т.е. существенно меньше максимально возможной длины для ЭВМ с длиной слова в 16 разрядов $/2^{16}/$.

В работе [2] предложен алгоритм генерации псевдослучайных чисел для малоразрядных ЭВМ, в котором с целью достижения большого периода используется двойная длина слова ЭВМ, а для ускорения выработки случайных чисел множитель "a" берется в виде $a=2^p+1$. Однако в [3] отмечается, что использование множителей такого типа в ГСЧ приводит к ухудшению его характеристик. ГСЧ, использующий двойное слово и обычное умножение на "a", обладает следующими недостатками:

1/ из-за длины периода, равной $2^{32}/\sim 10^{10}/$, его невозможно будет проверить экспериментальными тестами на полной длине периода. Проверка сведется к теоретическим тестам, что снизит надежность оценки качества такого ГСЧ;

2/ длина периода 10^{10} не может быть использована полностью на малых ЭВМ из-за низкого быстродействия последних, тогда как пониженная скорость выработки случайных чисел такими ГСЧ, обусловленная операциями с двойной длиной слова, в ряде применений может оказаться существенным недостатком в случае использования ГСЧ при работе мини-ЭВМ на линии с установками в реальном масштабе времени.

В данной работе описан ГСЧ для мини-ЭВМ ЕС-1010, основанный на смешанном конгруэнтном методе, оперирующий с 16-разрядными числами, с экспериментально определенной длиной периода $2^{16}/\sim 6{,}5\cdot 10^4/$, проверенный экспериментальными тестами на случайность выдаваемых им чисел на полной длине периода. Время выработки одного числа составляет $\sim 3\cdot 10^{-5}$ с.

Текст программы ГСЧ /см. Приложение/, благодаря приведенным в нем комментариям, легко может быть переведен на язык ассемблер для любой другой мини-ЭВМ.

2. АЛГОРИТМ ГЕНЕРАЦИИ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

Для получения последовательности случайных чисел с равномерным распределением используется рекуррентное соотношение

$$X_{n+1} = (aX_n + c) \bmod m, \qquad /1/$$

где X_n - предыдущий член случайной последовательности; X_{n+1} - последующий член случайной последовательности; a, c, m - константы, выбор которых обсуждается ниже.

В качестве начального случайного числа можно брать любое целое число меньше 65536. Удобно за X_0 принимать либо текущую дату, либо текущий момент времени.

При выборе констант a, c, m исходят из следующих соображений /4/:

1/ период последовательности должен быть максимально возможным для ЭВМ данной разрядности;

2/ на полной длине периода не должны появляться одинаковые числа;

3/ остаток по модулю "m" должен получаться возможно более простым способом;

4/ мощность ГСЧ, определяемая как минимальное целое число "s", для которого $b^s = 0 (\bmod m)$, должна быть не менее 5-6, в противном случае последовательность нельзя считать случайной;

5/ коэффициент последовательной корреляции

$$C = \frac{n\Sigma(u_j v_j) - (\Sigma u_j)(\Sigma v_j)}{\sqrt{(n\Sigma u_j^2 - (\Sigma u_j)^2)(n\Sigma v_j^2 - (\Sigma v_j)^2)}} \qquad /2/$$

должен быть как можно ближе к нулевому значению.

На основе изложенных выше соображений константы a, c, m для ГСЧ ЕС-1010 выбирались следующим образом:

1/ поскольку период ГСЧ не может быть больше, чем "m", а также для удобства получения остатка по модулю "m" величина "m" принята равной 2^{16}.

2/ исходя из условия получения малой величины коэффициента последовательной корреляции выбираем "a" в соответствии с неравенствами

$$\sqrt{m} < a < m - \sqrt{m} \qquad /3/$$

и корректируем по соотношению $a \bmod 8 = 5$, выполнение которого гарантирует неповторение чисел на полной длине периода, а также высокую мощность ГСЧ; берем $a = 31413$;

3/ проверяем $b = a-1$ по двум условиям: а/ "b" кратно "p" для любого "p", являющегося делителем "m"; б/ "b" и "m" должны быть кратны "4";

4/ находим "c" = 0,2113248·m, выполнение которого совместно с условием п.2 гарантирует малую величину коэффициента последовательной корреляции;

5/ корректируем найденную величину "c" по условию, что "c" и "m" должны быть взаимно простыми числами для сохранения максимального периода ГСЧ, "c" = 6881;

6/ вычисляем мощность "s" = 5; согласно п.4 величина "s" удовлетворительная.

Подбор констант a, c, m по указанным выше рекомендациям априори гарантирует ГСЧ с хорошими характеристиками /5/. Тем не менее для надежной оценки качества ГСЧ, использующего найденные исходя из теоретических соображений константы, полезно также провести экспериментальную проверку с помощью системы тестов.

3. ТЕСТИРОВАНИЕ ГСЧ

Тестовая проверка ГСЧ на ЭВМ ЕС-1010 проводилась:

1/ по критерию χ^2,

2/ по критерию Колмогорова-Смирнова /КС-критерий/.

Оценка результатов,

а/ полученных по критерию χ^2, осуществлялась по КС-критерию;

б/ полученных по КС-критерию,- повторно по КС-критерию.

Остановимся подробнее на каждом способе.

Критерий χ^2

В пределах периода случайная последовательность чисел, выдаваемых ГСЧ, разбивалась на различные части по 1000, 5000, 10000 чисел в каждой. Числа каждой из перечисленных частей последовательности группировались по "k" категориям, после чего для каждой части последовательности вычислялась статистика V по формуле

$$V = \sum_{1 \le s \le k} \frac{(y_s - np_s)^2}{np_s}, \qquad /4/$$

где: p_s - вероятность попадания случайного числа в некоторую группу "s", где $1 \le s \le k$; y_s - количество чисел, которые действительно попали в группу "s" из данной части последовательности. По вычисленным значениям V из таблицы распределения χ^2 находятся вероятности их появления. Части последовательности, для которых эти вероятности превышают 99% или меньше 1%, трактуются, как недостаточно случайные. Если вероятности появле-

ния статистик попадают в диапазон 95-99% или 1-5%, то результаты считаются "подозрительными". При значениях вероятностей 90-95% или 5-10% результаты "слегка подозрительны". Считается, что часть последовательности удачно прошла испытание по критерию χ^2, если вероятности появления статистики, вычисляемой для нее, меньше 90% или больше 10%.

После оценки по таблице χ^2 статистик, вычисленных для всех частей последовательности на полном периоде, можно оценить качество ГСЧ на полном периоде, как удовлетворительное, если не менее 2/3 /т.е. ~ 68%/ вероятностей находятся в диапазоне 5÷95%. Результаты экспериментальной оценки ГСЧ ЕС-1010 по критерию χ^2 представлены в табл.1, где указано процентное отношение числа статистик, для которых вероятности находятся в диапазоне 5÷95%, к общему числу статистик на полном периоде.

Критерий Колмогорова-Смирнова /КС-критерий/

Поскольку в литературе есть данные о том, что в ряде случаев КС-критерий оказывается эффективнее при выяснении отклонений от случайности, чем критерий χ^2, ГСЧ ЕС-1010 был проверен и с помощью этого критерия.

Известно, что при использовании КС-критерия вычисляются следующие статистики:

$$K_n^+ = \sqrt{n} \cdot \max[F_n(x) - F(x)],$$

$$K_n^- = \sqrt{n} \cdot \max[F(x) - F_n(x)],$$

где K_n^+ - максимальное отклонение эмпирической функции распределения $F_n(x)$ от теоретической $F(x)$ для случаев $F_n(x) > F(x)$; K_n^- - максимальное отклонение $F_n(x)$ от $F(x)$ для случаев $F_n(x) < F(x)$; n - число испытаний, для которого вычислены статистики.

После вычисления статистик K_n^+ и K_n^- по таблице их распределения оцениваются вероятности появления таких статистик и далее по такому же принципу, как и для критерия χ^2, оценивается качество ГСЧ для всей последовательности случайных чисел и для отдельных ее интервалов.

Результаты испытания ГСЧ ЕС-1010 по КС-критерию приведены в табл.2, где показано процентное отношение числа испытаний, в которых вероятность появления статистик K_{1000}^+ и K_{1000}^- не выходила за пределы 5-95%, к общему числу испытаний.

Двойное применение критериев

Интегральную проверку ГСЧ по критериям χ^2 или КС-критериям на полной длине периода по изложенному выше "способу 2/3" можно дополнить следующим образом: поскольку статистики, вычисляемые для частей последовательности по критериям χ^2 и КС-критериям, сами являются случайными числами, то к ним можно повторно применить КС-критерий, что позволяет выявить глобальные отклонения от заданного закона распределения на полной длине периода ГСЧ. Результаты проверки ГСЧ ЕС-1010 с помощью двойного применения критериев приведены в табл.3, а также иллюстрированы рис.1,2.

Из табл.1-3 следует, что ГСЧ для ЭВМ ЕС-1010 вполне удовлетворительно прошел все тестовые проверки.

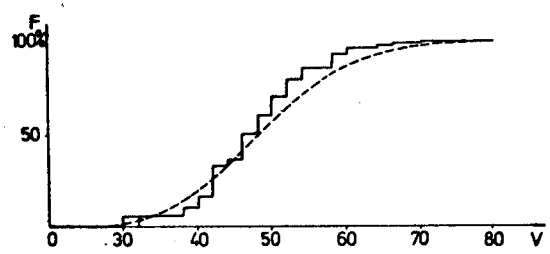

Рис.1

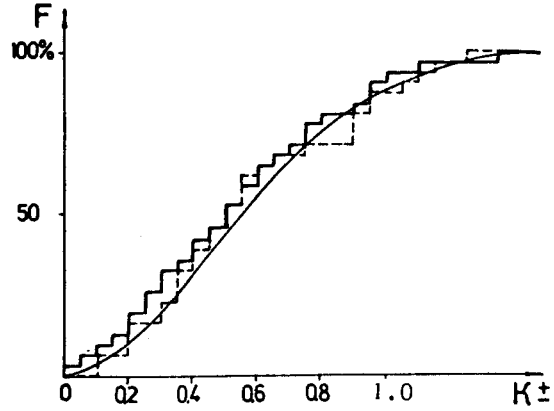

Рис.2

Таблица 1

k \ n	2000	5000	10000
10	93	74	82
20	97	96	82
50	93	96	86

Таблица 2

$K^{\pm}$ \ n	2000
K^+	90
K^-	87

Таблица 3

$\chi^2 \rightarrow$ КС	
$K^{\pm}$	P /%/
$K_{20}^+ = 0{,}670$	35
$K_{20}^- = 0{,}402$	60
КС→КС	
$K_{30}^+ = 0{,}657$	40
$K_{30}^- = 0{,}602$	45

```
 1                          LDSR   LDS
 2 0000                            RES    32
 3 0040 1AE1                C      DATA   6881
 4 0042 7AB5                A      DATA   31413
 5 0044 7FFF                W      DATA   32767
 6 0046                     X      RES    1
 7 0048 0000                       DATA   0
 8 004A 0000A               POINT  DATA   TARG
 9 004C   03                       DATA,1 3
10 004D   00                       DATA,1 0
11       004E              TARG   EQU    $
12 004E                    XO     RES    2
13 0052                    SLCH   RES    2
14 0056                    KLUCH  RES    2
15                                FIN
16                         RNDM   LPS    LDSR
17 0000 210C                      LDE    =12
18 0002 024A                      LDX    POINT
19 0004 0000                      CSV    M:MOVE
20 0006 6056                      LDA    @KLUCH
21 0008 C500A                     BAZ    M3
22 000A C700A                     BRU    M6
23 000C 604E               M3     LDA    @XO
24 000E 1146                      STA    X
25 0010 2001                      LDA    =1
26 0012 7156                      STA    @KLUCH
27 0014 0042               M6     LDA    A
28 0016 0C46                      MUL    X
29 0018 0540                      ADD    C
30 001A 1146                      STA    X
31 001C 2100                      LDE    =0
32 001E 220F                      LDX    =15
33 0020 0000                      CLS    FDVF
34 0022 7652                      DST    @SLCH
35 0024 F100                      RTS
```

ЛИТЕРАТУРА

1. Стандартные подпрограммы математической статистики на языке фортран, т.1, Будапешт, 1975.
2. Акишин П.Г., Ососков Г.А. ОИЯИ, Р5-8411, Дубна, 1974.
3. Кнут Д. Искусство программирования для ЭВМ. "Мир", М., 1977, т.2, с.39.
4. Там же, с.193.
5. Там же, с.67.

Рукопись поступила в издательский отдел
17 августа 1981 года.