Министерство образования и науки Российской Федерации
Санкт-Петербургский Политехнический Университет Петра Великого
—
Институт компьютерных наук и кибербезопасности

ЛАБОРАТОРНАЯ РАБОТА № 1

«Изучение свойств линейно конгруэнтного генератора»

по дисциплине «Структуры данных»

Выполнила
студентка гр. 5131001/30003 Шевчук Ника

<подпись>

Преподаватель

Семьянов П.В

<подпись>

Санкт-Петербург
2024

1 ЦЕЛЬ РАБОТЫ

Разработать программу, которая реализует линейный конгруэнтный генератор псевдослучайных чисел.

2 ПОСТАНОВКА ЗАДАЧИ

1. Выбрать правильные параметры ЛКГ с длиной машинного слова 64 бита.
2. Проверить характеристики полученного ЛКГ.
3. Проверить, параметры ЛКГ критерием покера.
4. Проверить ЛКГ критерием хи-квадрат.

3 ТЕОРЕТИЧЕСКИЕ ИССЛЕДОВАНИЯ

3.1 Характеристики ЛКГ

Линейно конгруэнтный генератор – один из методов генерации псевдослучайных чисел. Применяется в простых случаях и не обладает криптографической стойкостью.

Суть метода заключается в вычислении последовательности случайных чисел X_n, полагая, что: $X_{n+1} = (aX_n + c) mod\ m$.

При выборе числа m нужно учитывать следующие условия:

1. Число m должно быть довольно большим, так как период не может иметь больше m элементов.
2. Значение числа m должно быть таким, чтобы $(aX_n + c) mod\ m$ вычислялось быстро.

Так как линейно конгруэнтный генератор определен числами m, a, c, то от выбора данных параметров будут зависеть его характеристики. Чтобы период ЛКГ был равен числу m, нужно, чтобы выполнялись следующие условия:

1. Числа c и m взаимно простые.
2. $b = a - 1$ кратно p для каждого простого p, являющегося делителем числа m.
3. b кратно 4, если m кратно 4.
4. x_0 - произвольное число.

5. a и c должны быть такими, что: $a\ mod\ 8 = 5$ и c – нечетное.

6. $0{,}01m < a < 0{,}99m$.

В зависимости от выбранных параметров можно рассматривать такие характеристики ЛКГ:

– Мощность – это зависимость коэффициентов a и m. По формуле $(a-1)^s \equiv 0(mod\ m)$ можем определить s - мощность. Если $s > 5$, то данный генератор имеет хорошую мощность

– Разброс. Если числа из промежутка $[a; b]$ появляются с одинаковой вероятностью, то разброс ЛКГ является хорошим.

– Период. Каждое число последовательности, созданной генератором псевдослучайных чисел, повторится через какой-то количество чисел. Это количество числе и называется периодом. Период не превышает коэффициент m. Чем больше эта характеристика, тем лучше выполняет свою задачу ЛКГ.

3.2 Покер-критерий

Классический покер-критерий рассматривает групп по пять последовательных целых чисел {Y5j, Y5j+1, Y5j+2, Y5j+3, Y5j+4} для 0<= j < n и проверяет, какие из следующих семи комбинаций соответствуют таким пятеркам чисел (порядок не имеет значения).

Все числа разные: abcde

Одна пара: aabcd

Две пары: aabbc

Три числа одного вида: aaabc

Полный набор: aaabb

Четыре числа одного вида: aaaab

Пять чисел одного вида: aaaaa

Хи -критерий основан на подсчете числа пятерок в каждой категории.

Уместно рассмотреть какую-нибудь упрощенную версию этого критерия, для которой можно использовать более простые программы. Хорошим компромиссом будет критерий, использующий более простой подсчет различных

значений в множестве пятерок. В этом случае можно выделить только пять категорий:

5 значений = все разные;

4 значения = одна пара;

3 значения = две пары или три числа одного вида; aabbc aaacd

2 значения = полный набор или четыре числа одного вида;

1 значение = пять чисел одного вида.

При такой схеме упрощаются подсчеты и критерий остается почти таким же хорошим.

В общем случае можно рассматривать n групп k последовательных чисел и подсчитывать число групп из k чисел с t различными числами. Затем применяется хи-критерий, в котором используются вероятности того, что в группе r различных чисел

$$p_r = \frac{d(d-1)\dots(d-r+1)}{d^k} \left\{ {k \atop r} \right\}.$$

(Числа Стирлинга определены в разделе 1.2.6 и могут быть подсчитаны по приведенным в нем формулам.) Так как вероятности pr, очень малы, когда r= 1 или 2, следует, вообще говоря, перед применением хи-критерия объединить несколько категорий, имеющих малые вероятности, в одну.

Чтобы получить формулу для pr, следует подсчитать, сколько d групп из k чисел, расположенных между 0 и d - 1, имеют точно r различных элементов, и разделить это число на d^k. Так как d(d - 1)… (d - r + 1) - это число упорядоченных наборов из r элементов множества, содержащего d элементов, достаточно показать, что)- это число способов разбиения множества из k элементов на точно r частей.

3.3 Критерий «Хи-квадрат»

Числа должны генерироваться с одинаковой вероятностью. В данном методе сравнивается теоретическая вероятность, с которой появляется

определенное число. В данной лабораторной работе этот критерий был реализован так: промежуток от 0 до максимального числа $(2^{64}-1)$ делится на 8 групп с равным количеством чисел. Тогда теоретическая вероятность того, что число попадет в одну из групп: $p_s = \frac{\text{Количество чисел в группе}}{\text{Количество чисел в последовательности}}$. Далее рассчитывается число V по формуле: $V = \sum_{s=1}^{k} \frac{(Y_s - np_s)^2}{np_s}$, где k - количество групп, Y_s – количество чисел, попавших в группу, n-количество чисел в последовательности.

Коэффициент V был получен для 8000 последовательностей, и посчитано среднее значение этого коэффициента. По таблице для метода хи-квадратов из книги Кнута «Искусство программирования, 2 том» в ряду с числом v (степень свободы, которая рассчитывается k-1) найти число, близкое к V. Посмотреть значение p. Если диапазон для V:

p < 1% или p > 99%, то данный генератор считается очень плохим,, 1% < p < 5% или 95% < p < 99%, то данный генератор считается подозрительным, 5% < p < 10% или 90% < p < 95%, то данный генератор считается хорошим, p > 10% или p < 90%, то данный генератор считается отличным, лучше использовать этот генератор.

```
Мощность-32
Хи_квадрат <очень хороший> - 4661
Хи_квадрат <хороший> - 339
Хи_квадрат <подозрительный> - 0
Хи_квадрат <очень плохой> - 0
```

	$p = 1\%$	$p = 5\%$	$p = 25\%$	$p = 50\%$	$p = 75\%$	$p = 95\%$	$p = 99\%$
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5754	1.386	2.773	5.991	9.210
$\nu = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 6$	0.8721	1.635	3.455	5.348	7.841	12.59	16.81
$\nu = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.72
$\nu = 12$	3.571	5.226	8.438	11.34	14.85	21.03	26.22
$\nu = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$\nu = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$\nu = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\nu = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$\nu > 30$	$\nu + \sqrt{2\nu}x_p + \frac{2}{3}x_p^2 - \frac{2}{3} + O(1/\sqrt{\nu})$						
$x_p =$	−2.33	−1.64	−.674	0.00	0.674	1.64	2.33

Рисунок 1 – Распределение Хи-квадрат

4.ОПИСАНИЕ РЕШЕНИЯ И ТЕСТИРОВАНИЕ ПРОГРАММЫ

3.4 Параметры ЛКГ

Учитывая условия из теоретического условия, были подобраны следующие параметры: $x_0 = \text{srand}(\text{time}(\text{NULL}))$, $m = 2^{64}$, $a = 7454901$, $c = 911321152141$.

3.5 Характеристики ЛКГ

3.5.1 Мощность

По формуле $(a-1)^s \equiv 0(mod\ m)$ можем рассчитать мощность ЛКГ. Это осуществляется благодаря циклу, если на i-итерации цикла число $(a-1)^i$ будет равно 0, значит число i– мощность, действие осуществляется за счет переполнения, поэтому операция взятия по модулю является необязательной.

3.5.2 Период

Период полученного ЛКГ вычислить точно не удалось, программа не посчитала значение после 8 часов работы, из этого можно сделать вывод, что период- число, близкое или равное значению 2^{64} .

3.5.3 Разброс

Количество чисел было разбито на 50 равных групп, путем взятия остатка от деления. После чего был подсчитан хи-квадрат для данного разбиения, его значение попадает в диапазон хорошего генератора.

Характеристики ЛКГ

Таблица 1 – Полученные практические значения характеристики ЛКГ

Характеристика	Теоретическое значение	Практическое значение
Мощность	s>5	s = 32
Период	2^{64}	2^{64}
Разброс	42.94 … 56	55,45

3.6 Хи-квадрат

Числа от 0 до максимального числа типа unsigned int (2^{64}) были разделены на 8 групп. С помощью циклов рассчитывается, сколько чисел попадает в группы. По формуле $V = \sum_{s=1}^{k} \frac{(Y_s - np_s)^2}{np_s}$, высчитывается значение V для каждой последовательности чисел (было взято 8000 последовательностей). Y_s – это количество чисел, которые попали в данную группу, np_s – количество чисел в последовательности, деленное на количество групп. Находится среднее значение V для 8000 последовательностей. np_s в моем случаем было равно 1000.

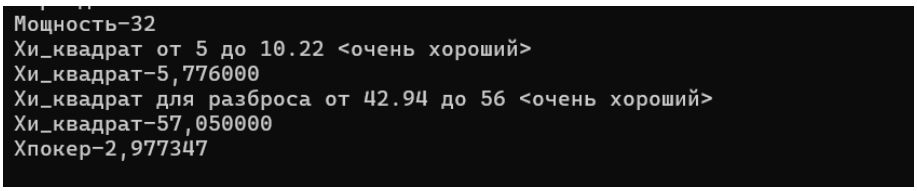


Рисунок 2 – Проверка Хи-квадрат

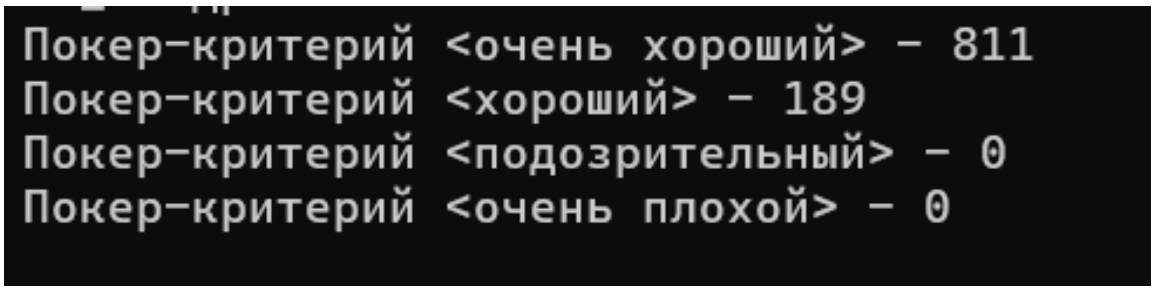
Рассчитывая среднее значение, смотрю, попадает ли оно в диапазон для данной степени свободы.

3.7 Покер-критерий

Значения Pr считается в отдельной функции pr, в которую сразу же передаются числа Стирлинга, взятые в таблице для числа 5.

$$p_r = \frac{d(d-1)\dots(d-r+1)}{d^k} \left\{ {k \atop r} \right\}.$$

Далее генерируются 10000 чисел, каждый раз добавляя в массив остатки от деления сгенерированного числа на 5, пузырьковой сортировкой числа распределяются по возрастанию, чтобы было легче определить, сколько из них различны. Каждый раз в соответственное место массива добавляется 1, а после, циклом вычисляется среднее значения покер-критерия.



4 ВЫВОДЫ

В результате выполнения лабораторной работы был изучен ЛКГ и подобраны оптимальные параметры для него. Данный генератор прошел проверку

на различные критерии, а полученные характеристики ЛКГ показывают, что генератор является хорошим и его можно использоваться для генерации псевдослучайных чисел.