

Report1

胡琦浩 PB21000235

2023 年 10 月 2 日

1 问题

用 Schrage 方法编写随机数子程序, 用指定间隔 (非连续 $l > 1$) 两个随机数作为点的坐标值绘出若干点的平面分布图。再用 $\langle x_k \rangle$ 测试均匀性 (取不同量级的 N 值, 讨论偏差与 N 的关系)、 $C(l)$ 测试其 2 维独立性 (总点数 $N > 10^7$)

2 关于代码一点说明

```
40 for n in range(10000001):          #产生N>10^7个随机数(通过改变range里面的值实现不同的N的大小)
41     data.append(test.GetElem(i))
42     i=test.function(i)              #迭代一下
43     if n<43:                        #获得第二题前43个I值
44         I.append(i)
```

图 1: main.py 中的源代码片段

如图所示, 由于我编写的程序中 N 的值需要手动修改。在测试 k 阶矩和卡方分布时, 我将 N 的值依次设为 $10^7 + 1, 10^6 + 1, 10^5 + 1$, 得到了此实验报告的数据。但最后我将原代码中的 N 值又改回了 $10^7 + 1$ 并生成了可执行文件, 由于时间的差异导致初始值的不同, 我的可执行文件中所展示 $N = 10^7 + 1$ 时的数据与此报告中的数据略有差异, 但并不影响此报告结论的得出。

此外, 由于最后程序设置的 $N = 10^7 + 1$ 比较大, 执行程序输出结果大概要 1 分钟左右。

3 方法

用 16807 产生器产生随机数

$$I_{n+1} = (aI_n + b) \bmod m \quad (1)$$

$$x_n = I_n / m \quad (2)$$

我们取 $a = 7^5 = 16807$, $b = 0$, $m = 2^{31} - 1 = 2147483647$

考虑到 aI_n 可能越界, 故采用 Schrage 方法:

$$I_{n+1} = \begin{cases} a(I_n \bmod m) - r[I_n/q] & I_{n+1} \geq 0 \\ a(I_n \bmod m) - r[I_n/q] + m & I_{n+1} < 0 \end{cases}$$

式中: $q = 127773$, $r = 2836$

在代码实现中, 利用递推算法依次得到 I_n 与 x_n

初始 I_0 由计算机中的时间获得:

$$I_0 = i_y + 70(i_m + 12(i_d + 31(i_h + 23(i_n + 59i_s)))) \quad (3)$$

4 实验结果

4.1 随机数的平面分布图

此图中取了 10000 个点作图, 间隔为 2. 由图可以看出无明显的分层现象, 故随机数之间无较强的相关性

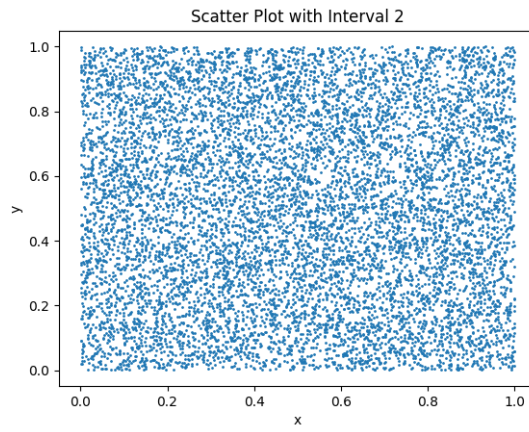


图 2: 随机数分布图

4.2 检验随机数均匀性

4.2.1 利用 k 阶矩检验

$$\langle x^k \rangle = \frac{1}{N} \sum_{i=1}^N x_i^k \Rightarrow \int_0^1 x^k p(x) dx = \frac{1}{k+1} \quad (4)$$

则:

$$\left| \langle x^k \rangle - \frac{1}{k+1} \right| \approx O\left(\frac{1}{\sqrt{N}}\right) \quad (5)$$

```
当K=3时, <x^k>=0.24986549839551006, expectation=0.25
当K=4时, <x^k>=0.19988591099026753, expectation=0.2
当K=5时, <x^k>=0.16656959800212923, expectation=0.16666666666666666
当K=6时, <x^k>=0.1427733220654046, expectation=0.14285714285714285
当K=7时, <x^k>=0.12492634060686453, expectation=0.125
当K=8时, <x^k>=0.11104528602338971, expectation=0.11111111111111111
当K=9时, <x^k>=0.09994029596835018, expectation=0.1
当K=10时, <x^k>=0.09085424935424406, expectation=0.09090909090909091
```

图 3: $N = 10^7 + 1$

```
当K=3时, <x^k>=0.24978240662129775, expectation=0.25
当K=4时, <x^k>=0.19978235999076951, expectation=0.2
当K=5时, <x^k>=0.16645438435819662, expectation=0.16666666666666666
当K=6时, <x^k>=0.14265231256370367, expectation=0.14285714285714285
当K=7时, <x^k>=0.124803266295954, expectation=0.125
当K=8时, <x^k>=0.11092247204871505, expectation=0.11111111111111111
当K=9时, <x^k>=0.09981915957447446, expectation=0.1
当K=10时, <x^k>=0.09073562255537919, expectation=0.09090909090909091
```

图 4: $N = 10^6 + 1$

不难看出, 随着 N 的值越来越大, $\langle x^k \rangle$ 的值与理论值越来越接近. 当然, 极其接近理论值的 $\langle x^k \rangle$ 表明由 Schrage 得到的随机数均匀性很好

4.2.2 利用卡方检验

将区间 $[0,1]$ 分为 K 个子区间, 统计随机数落在第 k 个子区间的实际频数 n_k , 它应当趋近于理论频数 $m_k = \frac{N}{K}, (k = 1, 2, \dots, K)$ 令统计量:

$$\chi^2 = \sum_{k=1}^K \frac{(n_k - m_k)^2}{m_k} \quad (6)$$

服从自由度为 $K - 1$ 的卡方分布

```
当K=3时, <x^k>=0.24917590128176625, expectation=0.25
当K=4时, <x^k>=0.1992557071351527, expectation=0.2
当K=5时, <x^k>=0.16598073401455707, expectation=0.16666666666666666
当K=6时, <x^k>=0.14221091400723213, expectation=0.14285714285714285
当K=7时, <x^k>=0.12438130652860102, expectation=0.125
当K=8时, <x^k>=0.11051236228265557, expectation=0.11111111111111111
当K=9时, <x^k>=0.09941639991690916, expectation=0.1
当K=10时, <x^k>=0.09033753486346885, expectation=0.09090909090909091
```

图 5: $N = 10^5 + 1$

```
在K=1时, 卡方值为0.0
在K=2时, 卡方值为4.940683605931639
在K=3时, 卡方值为2.593522140647786
在K=4时, 卡方值为5.671525332847467
在K=5时, 卡方值为4.208636979136302
在K=6时, 卡方值为6.8571850142815
在K=7时, 卡方值为7.142444885755512
在K=8时, 卡方值为7.623499137650087
在K=9时, 卡方值为5.540350445964956
在K=10时, 卡方值为6.660510233948976
```

图 6: $N = 10^7 + 1$

```
在K=1时, 卡方值为0.0
在K=2时, 卡方值为0.1592008407991592
在K=3时, 卡方值为0.03768796231203769
在K=4时, 卡方值为0.7947702052297948
在K=5时, 卡方值为1.7396922603077396
在K=6时, 卡方值为1.1211998788001214
在K=7时, 卡方值为2.2813497186502816
在K=8时, 卡方值为3.384547615452384
在K=9时, 卡方值为6.766009233990765
在K=10时, 卡方值为2.5319664680335316
```

图 7: $N = 10^6 + 1$

```
在K=1时, 卡方值为0.0
在K=2时, 卡方值为0.34224657753422466
在K=3时, 卡方值为3.1444685553144467
在K=4时, 卡方值为0.7609023909760904
在K=5时, 卡方值为2.4698153018469817
在K=6时, 卡方值为12.757002429975701
在K=7时, 卡方值为15.899841001589984
在K=8时, 卡方值为8.9910200897991
在K=9时, 卡方值为14.692853071469285
在K=10时, 卡方值为11.475975240247596
```

图 8: $N = 10^5 + 1$

在显著水平 $\alpha = 0.05$ 条件下, 由卡方分布表可得:

附表 3 χ^2 分布表

n	α						
	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.455	1.323	2.706	3.841	5.024	6.636	7.879
2	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	3.357	5.985	7.779	9.488	11.143	13.277	14.860
5	4.351	6.626	9.236	11.071	12.833	15.086	16.750
6	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	7.344	10.219	13.362	15.507	17.535	20.090	21.955
9	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	11.340	14.845	18.549	21.026	23.337	26.217	28.289
13	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	14.339	18.245	22.307	24.996	27.488	30.578	32.801
16	15.338	19.369	23.542	26.296	28.845	32.000	34.267
17	16.338	20.489	24.769	27.587	30.191	33.409	35.718
18	17.338	21.605	25.989	28.869	31.526	34.805	37.156
19	18.338	22.718	27.204	30.144	32.852	36.191	38.582
20	19.337	23.828	28.412	31.410	34.170	37.566	39.997
21	20.337	24.935	29.615	32.671	35.479	38.932	41.401
22	21.337	26.030	30.813	33.924	36.781	40.289	42.796
23	22.337	27.141	32.007	35.172	38.076	41.638	44.181
24	23.337	28.241	33.196	36.415	39.364	42.980	45.559
25	24.337	29.339	34.382	37.652	40.646	44.314	46.928
26	25.336	30.435	35.563	38.885	41.923	45.642	48.290
27	26.336	31.528	36.741	40.113	43.194	46.963	49.645
28	27.336	32.620	37.916	41.337	44.461	48.278	50.993
29	28.336	33.711	39.087	42.557	45.722	49.588	52.336
30	29.336	34.800	40.256	43.773	46.979	50.892	53.672
35	34.336	40.223	46.059	49.802	53.203	57.342	60.275
40	39.335	45.616	51.805	55.758	59.342	63.691	66.766
45	44.335	50.985	57.565	61.656	65.410	69.957	73.166

图 9: 卡方分布表

χ 均小于 χ_{α} , 故在置信度 $1 - \alpha = 0.95$ 下, 随机数的均匀性很好

4.3 二维独立性检验

用相邻两个随机数的自相关函数（或相关系数）来标识伪随机数序列的独立性情况，相关系数越小，独立性越好. 间距为 l 的自相关函数是:

$$C(l) = \frac{< x_n x_{n+l} > - < x_n >^2}{< x_n^2 > - < x_n >^2}$$

(7)

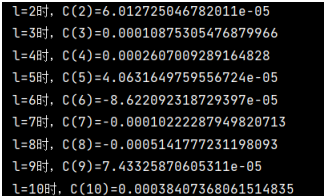


图 10: $N = 10^7 + 1$

可以看出: 不同 l 值所对应的 $C(l)$ 值均接近 0, 故数据独立性很高

5 总结

主要学习并掌握了由 Schrage 方法编写的 16807 随机数产生程序，以及各种对随机数相关性，均匀性，独立性的检验方法，收获颇多