计算物理 A——Homework 1

何金铭 PB21020660

1 题目描述

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 $<x^k>$ 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）。

1.1 大致思路

1. 先用 Lehmer 线性同余法产生一组随机数，但由于计算过程中的数值可能会超出 32 位计算机的最大整数位，所以还采用了 Schrage 法进行取模运算，最终得到一组随机数。

2. 再将数据以一定的规则和间隔（l=2,3...）画至 o-xy 平面内，利用一些统计方法来检验其点列的均匀性与独立性。

2 算法过程

2.1 Lehmer 线性同余法

利用以下方程来产生随机数：

$$\begin{cases} I_{n+1} = (aI_n + b) \bmod m \\ x_n = \frac{I_n}{m} \end{cases} \tag{1}$$

当选取某些参数 a, b, m 时可以产生一些随机性、均匀性良好的随机数。比如 16807 产生器（$a = 16807, b = 0, m = 2147483647$）。以下的操作均取 16807 随机数产生器。

2.2 Schrage 法

由于在 $I_{n+1} = (aI_n + b) \bmod m$ 的操作中 aI_n 可能会超过计算机的最大整数 2^31，而报错。所以使用了 $Schrage$ 法。

设 $m = aq + r, 0 < z < m - 1$

$$\begin{aligned} az \bmod m &= \left\{\frac{z}{q}(aq+r) - \frac{rz}{q}\right\} \bmod m \\ &= \left\{a(z \bmod q) - r\left[\frac{z}{q}\right]\right\} \\ &= \begin{cases} \left\{a(z \bmod q) - r\left[\frac{z}{q}\right]\right\} & if \geq 0 \\ \left\{a(z \bmod q) - r\left[\frac{z}{q}\right]\right\} + m & otherwize \end{cases} \end{aligned} \tag{2}$$

3 程序说明

3.1 主要程序

schrage.c 16807 随机数生成器 ($Schrage$ 法) 的代码

draw.py 生成间隔为 l 的二维随机数点的坐标 (x_i, x_{i+l}), 并且作图

x_k.py $\langle x^k \rangle$ 检验与 $C(l)$ 检验的代码

3.2 程序结果

schrage.exe 16807 PRNG

16807.csv 一个种子数为"23942907" 的样本随机数文件

./out 文件夹路径, 里面存放了二维随机数点列的坐标, 其中文件名中的数字代表了随机点数的个数

./pic 文件夹路径, 里面存放了二维随机数点列的散点图, 其中文件名中的数字代表了随机点数的个数

3.3 其他说明

1. 数据都写于 CSV 文件中
2. 其中 Python 程序用到的库有:

- matplotlib.pyplot：用于作图
- numpy：用于数据处理
- csv：用于读写 CSV 文件

4 结果分析

4.1 随机数作图结果

4.1.1 随机数坐标

随机数坐标存于'./out' 中, 其中文件名中的数字代表此文件中随机数的个数, 比如'./out/100.csv' 中有 100 个随机点的坐标

4.1.2 随机数图像

以下为将 (x_i, x_{i+2}) 作为 x-y 上的点的结果

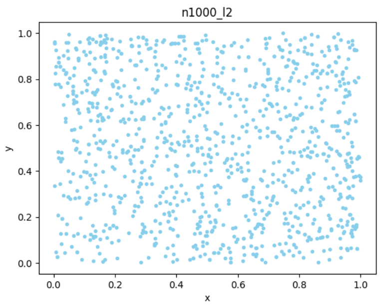

图 1: 1000 个随机数点

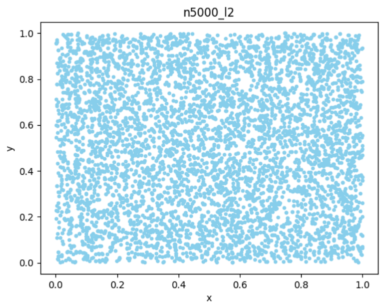

图 2: 5000 个随机数点

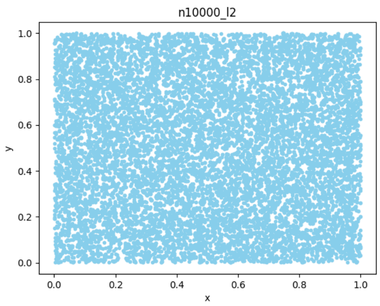

图 3: 10000 个随机数点

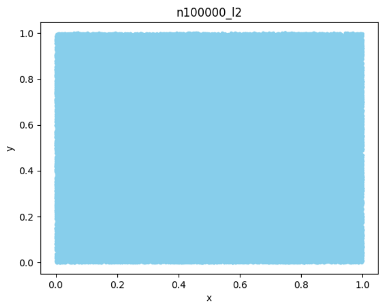

图 4: 100000 个随机数点

可以发现，当点数 $N = 5 \cdot 10^3, N = 10^4$ 时，随机数中的空白部分均匀的分布，体现了比较好的随机性。

4.2 随机数均匀性检验

主要利用 $\langle x^k \rangle$ 来检验其均匀性。

比较：

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

可得, 当满足：

$$\left| \langle x^k \rangle - \frac{1}{1+k} \right| \approx O(\frac{1}{\sqrt{N}}) \tag{3}$$

可认为生成的平均数为均匀分布

若将其值化在图像当中，结果将会更加明显（记 $\Delta = \left| \langle x^k \rangle - \frac{1}{1+k} \right| / \frac{1}{\sqrt{N}}$）：

$\left\vert\langle x^k\rangle - \frac{1}{1+k}\right\vert / \frac{1}{\sqrt{N}}$	$k=2$	$k=3$	$k=4$	$k=5$
$N=10^2$	0.103	0.169	0.209	0.232
$N=10^3$	0.530	0.548	0.542	0.524
$N=10^4$	0.546	0.547	0.518	0.482
$N=10^5$	0.244	0.167	0.098	0.043
$N=10^6$	0.267	0.219	0.187	0.166
$N=10^7$	0.254	0.211	0.179	0.156

表 1: 不同量级 N 与不同 k 时 $\left\vert\langle x^k\rangle - \frac{1}{1+k}\right\vert / \frac{1}{\sqrt{N}}$ 的值

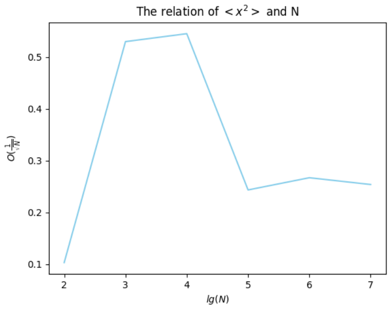

图 5: k=2 时,Δ 随 N 的变化

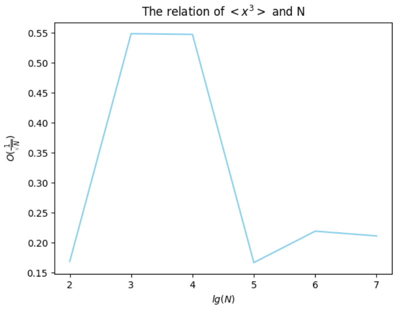

图 6: k=2 时,Δ 随 N 的变化

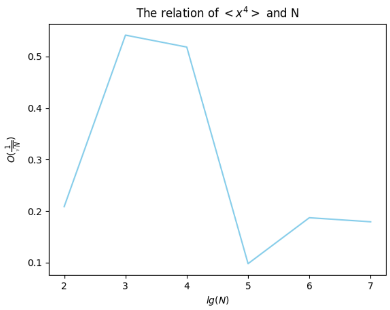

图 7: k=2 时,Δ 随 N 的变化

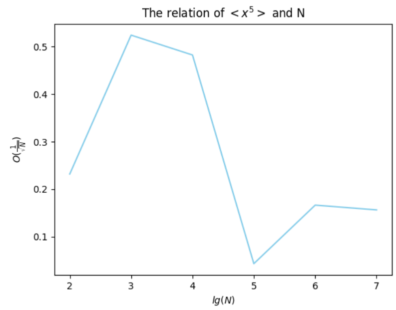

图 8: k=2 时,Δ 随 N 的变化

可以发现到 N 的数量级到达 10^6 时,$\left\vert\langle x^k\rangle - \frac{1}{1+k}\right\vert / \frac{1}{\sqrt{N}}$ 的值趋于稳定,大约为 0.2。发现 $\left\vert\langle x^k\rangle - \frac{1}{1+k}\right\vert$ 大约比 $\frac{1}{\sqrt{N}}$ 小一个量级，在 N 足够大的时候，可以认为随机数是均匀分布的。

值得一提的是，图像中的曲线波动是由于中心极限定理的作用。

4.3 随机数独立性检验

4.3.1 C(l) 自相关系数检验

自相关系数为:

$$C(l) = \frac{1}{N}\sum_{i=1}^{N} x_i x_{i+1} - \langle x\rangle^2 \Rightarrow \int_0^1\int_0^1 xyp(x,y)\,dx\,dy - \frac{1}{4} = 0 \tag{4}$$

当满足 $|C(l)| \approx O(\frac{1}{\sqrt{N}})$ 时，可认为随机数满足随机分布条件。

发现, 当 $N > 10^7$ 时,$C(l)$ 随着 N 的增大而逐渐变小, 大约比 $\frac{1}{\sqrt{N}}$ 小了 2 个量级左右.

	$N=1.2\cdot10^7$	$N=1.4\cdot10^7$	$N=1.6\cdot10^7$	$N=1.8\cdot10^7$	$N=2\cdot10^7$
$C(l)/\frac{1}{\sqrt{N}}$	0.0337	0.0013	0.021	0.024	0.005

表 2: 不同量级 N 与不同 k 时 $\left|\langle x^k\rangle-\frac{1}{1+k}\right|/\frac{1}{\sqrt{N}}$ 的值

5 总结

1. Schrage 法可以有效避免整数溢出的情况
2. 16807 随机数产生器的二维分布图具有良好的均匀性
3. $\langle x^k\rangle$ 的结果为偏差 Δ 约为 $\frac{1}{\sqrt{N}}$ 的 0.2 倍
4. $C(l)$ 检验的结果为 $C(l)$ 约为 $\frac{1}{\sqrt{N}}$ 的 0.01 倍