计算物理 A——Homework 6

何金铭 PB21020660

1 题目描述

对两个函数线型（Gauss 分布和类 Lorentz 型分布），设其一为 $p(x)$，另一为 $F(x)$，其中常数 $a \neq b \neq 1$, 用舍选法对 $p(x)$ 抽样。将计算得到的归一化频数分布直方图与理论曲线 $p(x)$ 进行比较，讨论差异，讨论抽样效率。

$$Gaussian :\sim exp{-ax^2}; \quad Lorentzian\ like :\sim \frac{1}{1+bx^4}$$

1.1 大致思路

1. 先观察曲线的大致分布，并且寻找一对合适的参数 (a, b)
2. 再利用第二类舍选法进行抽样，得到归一化频数直方图，再将其与原曲线对比分析
3. 分析抽样的效率

2 理论分析

2.1 Guass 分布为 $p(x)$

在满足归一化条件的情况下，不妨设

$$p(x) = \sqrt{\frac{a}{\pi}} e^{-ax^2},\ x \in \mathbb{R} \tag{1}$$

$$F(x) = k_2 \frac{1}{1+bx^4},\ x \in \mathbb{R} \tag{2}$$

要求：

$$p(x) \leq F(x),\ \forall x \in \mathbb{R},\ always\ right \tag{3}$$

并且要求 $F(x)$ 的图像尽可能的接近 $p(x)$，这对 $p(x)$ 和 $F(x)$ 中的参数提出了要求。

2.2 Lorentz 分布为 $p(x)$

需先求归一化的 $p(x)$ 但由于 $p(x)$ 的原函数解析式十分难求，可以采取先数值积分，之后再乘一个系数即可。

$$p(x) = k_2 \frac{1}{1+bx^4} \tag{4}$$

$$F(x) = k_1 e^{-ax^2} \tag{5}$$

其中 $k_2 = \int_{-\infty}^{\infty} \frac{1}{1+bx^4}\,dx$

3 算法过程

3.1 算法理论 1——Guass 分布为 $p(x)$

算法即为第二类舍选法，原理十分简单，在这里不再赘述

1. 先生成随机数 $\xi,\eta \in [0,1]$
2. 利用 $\xi = \int_{-\infty}^{\xi_1} F(x)\,dx / \int_{-\infty}^{\infty} F(x)\,dx$ 解得 ξ_1
3. 若满足 $\eta \cdot F(\xi_1) \le p(\xi_1)$，则取 ξ_1 为满足条件的样本点；否则重新进行 1 操作

其中值得一提的是 ξ_1 的求解：

$$\xi = \frac{\int_{-\infty}^{\xi_1} F(x)\,dx}{\int_{-\infty}^{\infty} F(x)\,dx} \tag{6}$$

其中 $\int_{-\infty}^{\infty} F(x)\,dx$ 代表了 $F(x)$ 下围成的面积，$\int_{-\infty}^{\xi_1} F(x)\,dx$ 代表了 $F(x)$ 则为概率积累函数于 $x=\xi_1$ 处的取值。

3.1.1 $F(x)$ 的近似

由于 $\int F(x)\,dx = \int k_2 \frac{1}{1+bx^4}\,dx$ 的解析式不易给出，以下给出近似计算方法。

$$\alpha = \frac{F(x_0)}{F(0)} = \frac{1}{1+bx_0^4} \tag{7}$$

当函数值变为原来的 α 倍时有，$x = x_0$

$$x_0 = \sqrt[4]{(\frac{1}{\alpha}-1)\frac{1}{b}} \tag{8}$$

可以确定一个置信区间，认为在区间 $[-x_0, x_0]$ 之间的函数围成的面积即为函数值。这里也可以定量讨论，在置信区间以外的函数值可以通过缩放给出范围：

$$\int_{x_0}^{\infty} \frac{k_2}{1+bx^4}\,dx \le \int_{x_0}^{\infty} \frac{k_2}{1+bx^2}\,dx = \frac{k_2}{\sqrt{b}} \arctan \sqrt{b}x \bigg|_{x_0}^{\infty} \tag{9}$$

则可得左右置信区间以外的函数积分值不会超过 $2\frac{k_2}{\sqrt{b}}(\frac{\pi}{2} - \arctan\sqrt{b}x_0)$

3.1.2 $p(x)$ 的近似

同时也需要对 $p(x)$ 进行近似。其置信区间也应该与上面抽样函数一致。

3.1.3 一个尝试的例子

不妨先取 $a=2, b=3, k_2=\sqrt{\frac{a}{\pi}}, x_0=5$

$$\int_{-\infty}^{\infty} F(x)\,dx = \int_{-x_0}^{x_0} F(x)\,dx = 2x_0 \langle F(x)\rangle \, x\in[-x_0,x_0] \tag{10}$$

利用顺序查找方法求解 ξ_1

对于每一个随机数 ξ，数值计算 $(\xi_1 - x_0)\langle F(x)\rangle, x\in[-x_0,\xi_1]$, 若 $|\xi - (\xi_1-x_0)\langle F(x)\rangle| < \epsilon$，$\epsilon$ 为可以接受的误差，则可解得 ξ_1

3.2 算法理论 2——Lorentz 分布为 $p(x)$

1. 先用 Box-Muller 法求出满足 $F(x)$ 的 x 的随机变量分布
2. 利用积分求出 $p(x)$ 的系数
3. 进行抽样

不妨设 $k_1 = 1.2, k_2 = 1, a = 0.5, b = 2$，则有

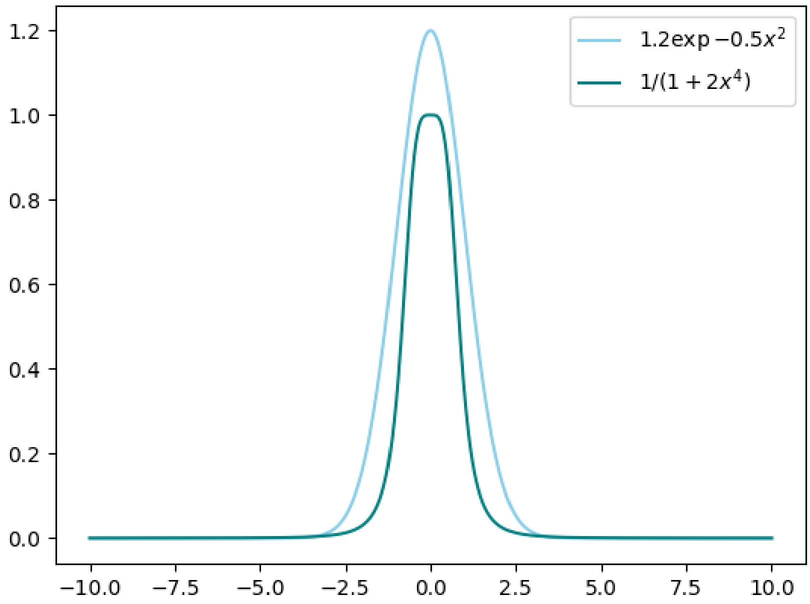

图 1: 函数图

若采用 3σ 原则，则有积分：

$$I = \int_{-3}^{3} \frac{1}{1+2x^4}\,dx = 1.85569 \tag{11}$$

所以有：

$$p(x) = \frac{1}{I}\frac{1}{1+2x^4} \tag{12}$$

于是有算法：

1. 生成随机数 $u, v, \eta \in [0,1]$
2. 有 $x = \sqrt{-2\log u}\cos 2\pi v$
3. 若 $\eta F(x) < p(x)$，则 x 为一个样本点；否则重新操作 1.

4 程序说明

4.1 主要程序

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

try.py 一个有些失败的 Guass 分布抽样（抽样速度过慢，可以忽略这个）

lorentz.py 一个参数为 $k_1 = 1.2, k_2 = 1, a = 0.5, b = 2$Lorentz 分布的抽样

4.2 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

lorentz.csv Lorentz 分布的抽样结果

./pic 文件夹路径, 里面存放了归一化频率直方图与理论曲线的对比图

4.3 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

5 结果分析

5.1 Guass 分布为 $p(x)$

由于前面的程序中顺序查找较慢，故这里不分析，分析另一种情况。
并且 Guass 分布已经有很多种方便的方法。

5.2 Lorentz 分布为 $p(x)$

在 $k_1 = 1.2, k_2 = 1, a = 0.5, b = 2$ 时分析结果

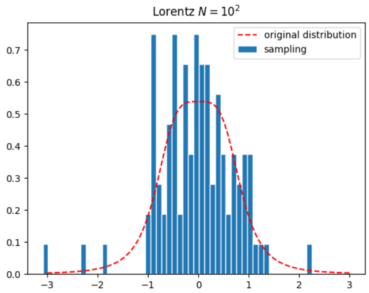

图 2: $N = 10^2$

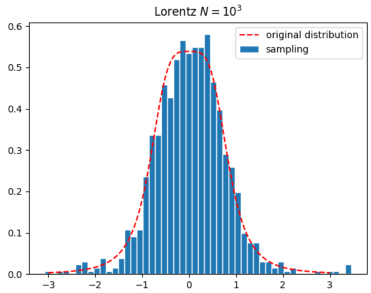

图 3: $N = 10^3$

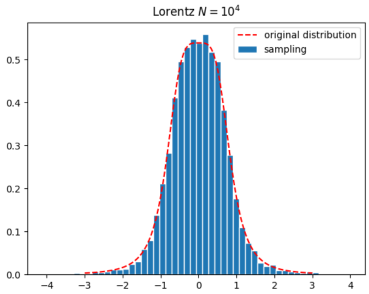

图 4: $N = 10^4$

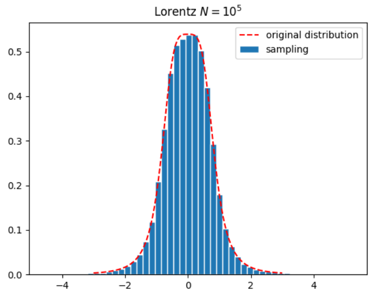

图 5: $N = 10^5$

发现在 N 很小的时候直方图与原曲线不重合，随着 N 的增大时，直方图和原曲线轨迹逐渐重合，当 $N > 10^5$ 时，肉眼已经区分不了了。

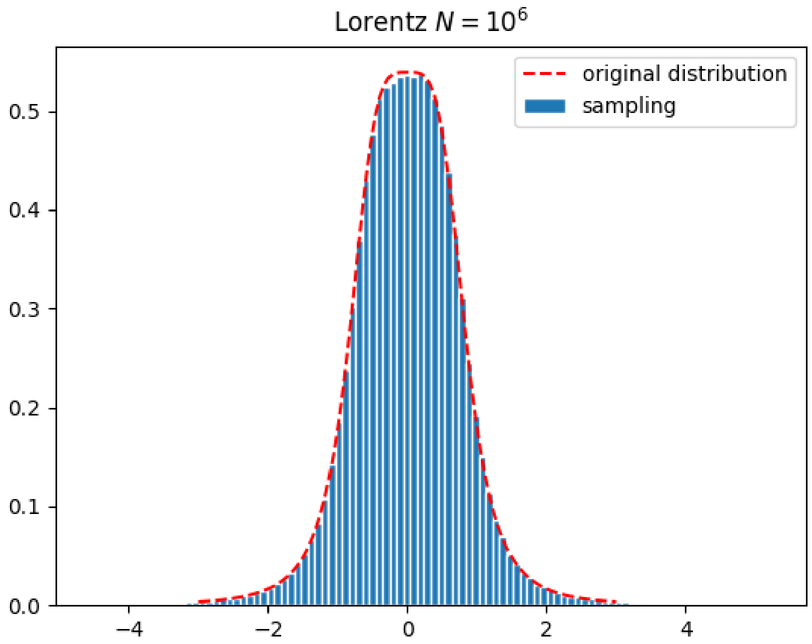

图 6: $N = 10^6$

以下讨论舍选法的效率问题。

N	10^2	10^3	10^4	10^5	10^6
effect	0.3861	0.3429	0.3351	0.3346	0.3340

表 1: 效率随抽样次数 N 的变化

发现在 $k_1 = 1.2, k_2 = 1, a = 0.5, b = 2$ 的条件下，对 $Lozentz$ 分布的抽样效率为 0.33 且发现效率随着 N 的增大逐渐趋于一个稳定值。

6 总结

1. 只分析了一种情况，另一种情况需要改进算法来提高运算速度。
2. 对于任意一个抽样，其归一化频数直方图与理论值的差距随着 N 的增大而减小。
3. 对于参数为 $k_1 = 1.2, k_2 = 1, a = 0.5, b = 2$ 的 Lorentz 分布的效率为 0.334，说明参数的选取有待改进。