Report6

PB21000235 胡琦浩

一.问题

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a≠b≠1常数，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.

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

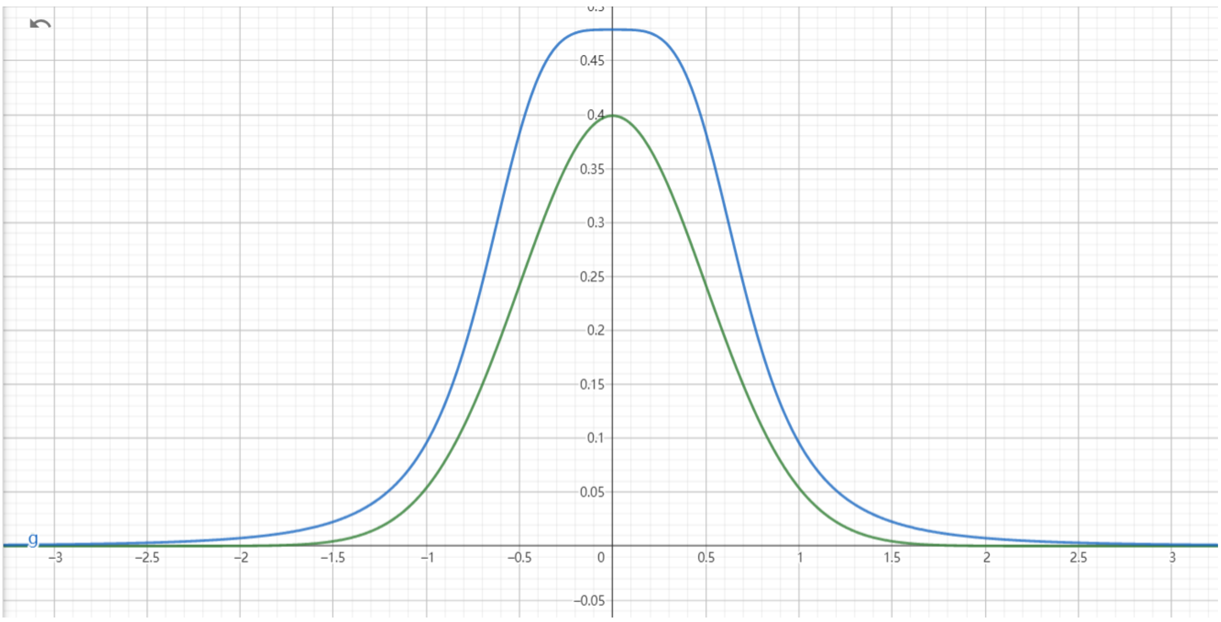
二.方法

2.1 数学推导

不妨令:a=2，b=4,记:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-2x^2} \quad F(x) = \frac{1}{\sqrt{2\pi}} \frac{1.2}{1+4x^4}$$

如图所示,蓝线代表F(x),绿线代表p(x).



在[-3,3]范围内,F(x)>p(x)恒成立。

由舍选法:

$$\xi_1 = \frac{\int_a^{\xi_x} F(x)\, dx}{\int_a^b F(x)\, dx}$$

$$\xi_2 = \frac{\xi_y}{F(\xi_x)}$$

对F(x)求不定积分可得:

$$G(x) = \int F(x)\, dx = \frac{1.2}{\sqrt{2\pi}} \int \frac{1}{1+4x^4}\, dx = \frac{1.2}{\sqrt{2\pi}} \times \frac{1}{8}(ln\frac{2x^2+2x+1}{2x^2-2x+1} + 2arctan(2x+1) - 2arctan(1-2x))$$

显然,较难求出ξ_x关于ξ_1的函数.

故采用数值解法:记:

$$f(\xi_x) = \xi_1 = \frac{G(\xi_x) - G(-3)}{G(3) - G(-3)}$$

显然f(ξ_x)为单调递增,定义域[-3,3],值域[0,1]的函数,可以使用二分法来得到ξ_x的数值解.

2.2 代码实现

程序中定义了p(x),F(x),G(x),f(x),以及f(x)的反函数f_1(x).

首先利用question1中的函数生成[0,1]的随机数储存到ξ_1与ξ_2中

然后利用函数f_1(x)根据ξ_1求出ξ_x

（贴出二分法实现过程）

```python
#利用二分法求f(x)反函数的值,即此函数为f(x)的反函数
def f_1(res):
    #初始范围
    a = -3
    b = 3
    c = (a+b)/2

    #精度为1e-6
    while math.fabs(f(c)-res) > 1e-6:
        if f(c) > res:
            b = c
        elif f(c) == res:
            return c
        else:
            a = c
        c = (a+b)/2
    return c
```

再利用舍选法:若ξ_y<p(ξ_x),则取x=ξ_x,否则重新选取

最后画出概率直方图,共分为101个区间

三.实验结果

```
总点数为N=1000,抽样点数为n=691,效率为0.691
总点数为N=5000,抽样点数为n=3332,效率为0.6664
总点数为N=10000,抽样点数为n=6606,效率为0.6606
总点数为N=50000,抽样点数为n=33388,效率为0.66776
总点数为N=100000,抽样点数为n=66732,效率为0.66732
```

理论值为:

$$\eta = \frac{\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-2x^2}\, dx}{\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} \frac{1.2}{1+4x^4}\, dx} = 0.66752$$

可以看出:随着总点数的增多,抽样效率越接近理论值。

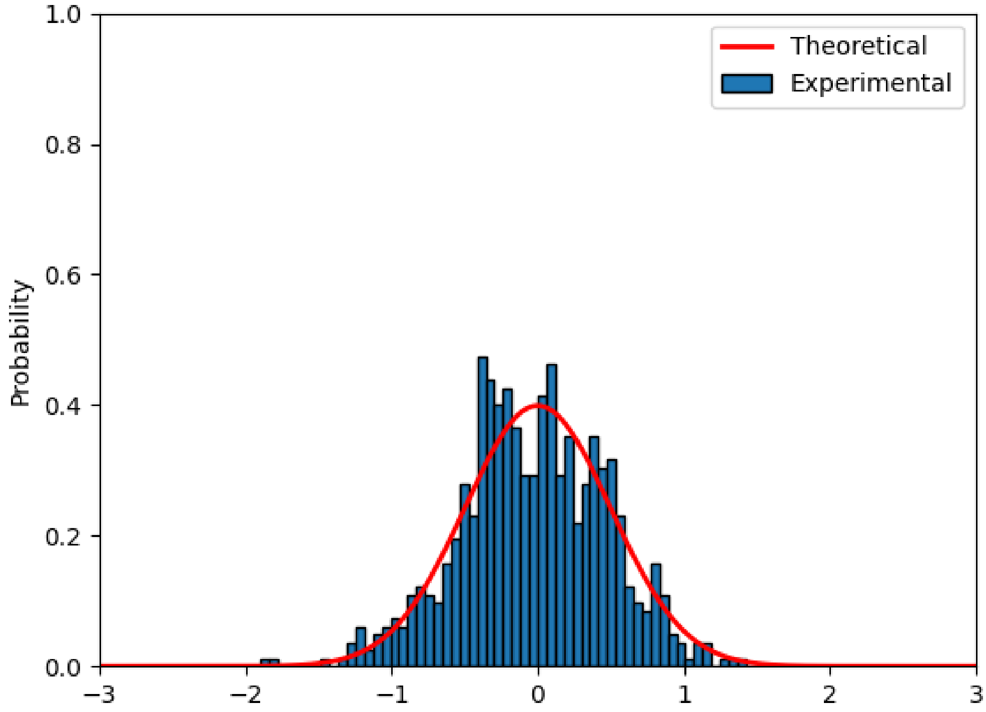


图1：N=1000

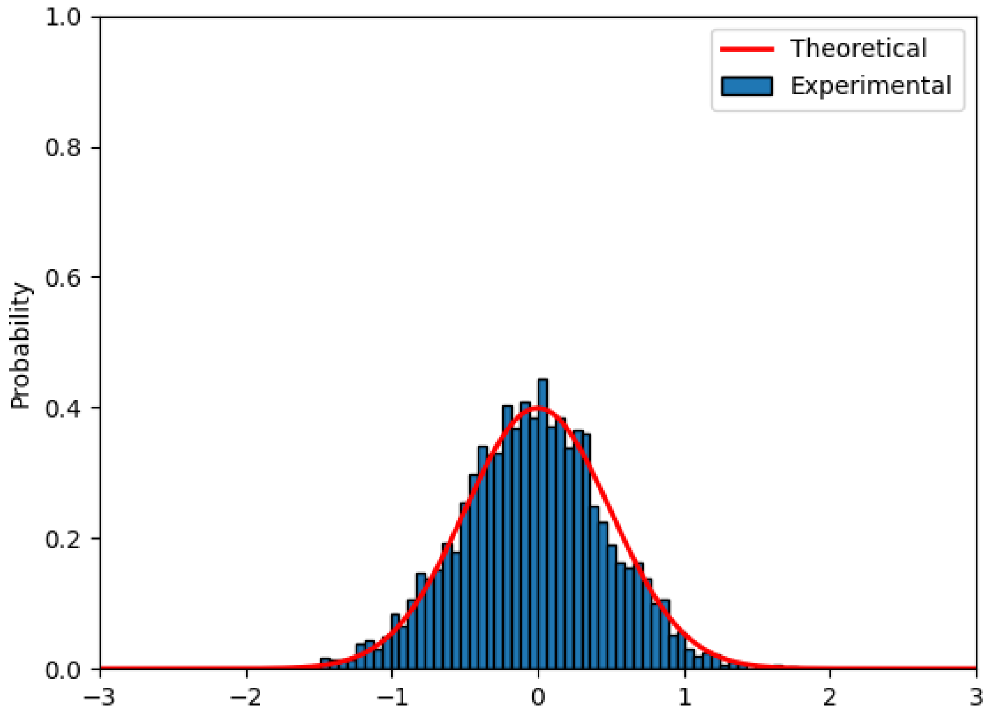


图2：N=5000

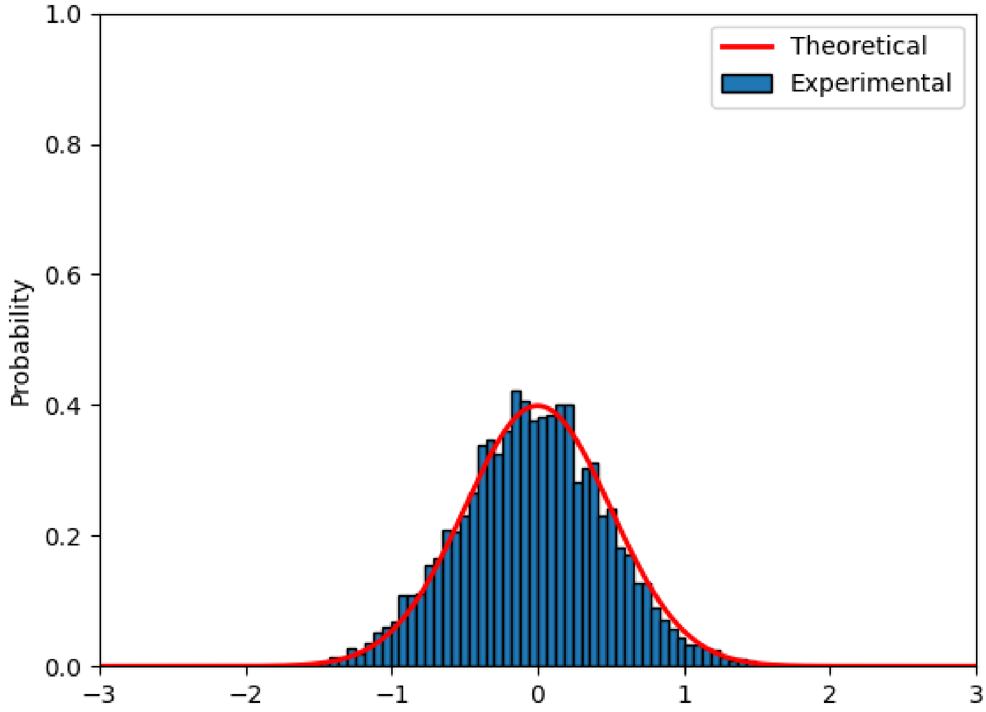


图3：N=10000

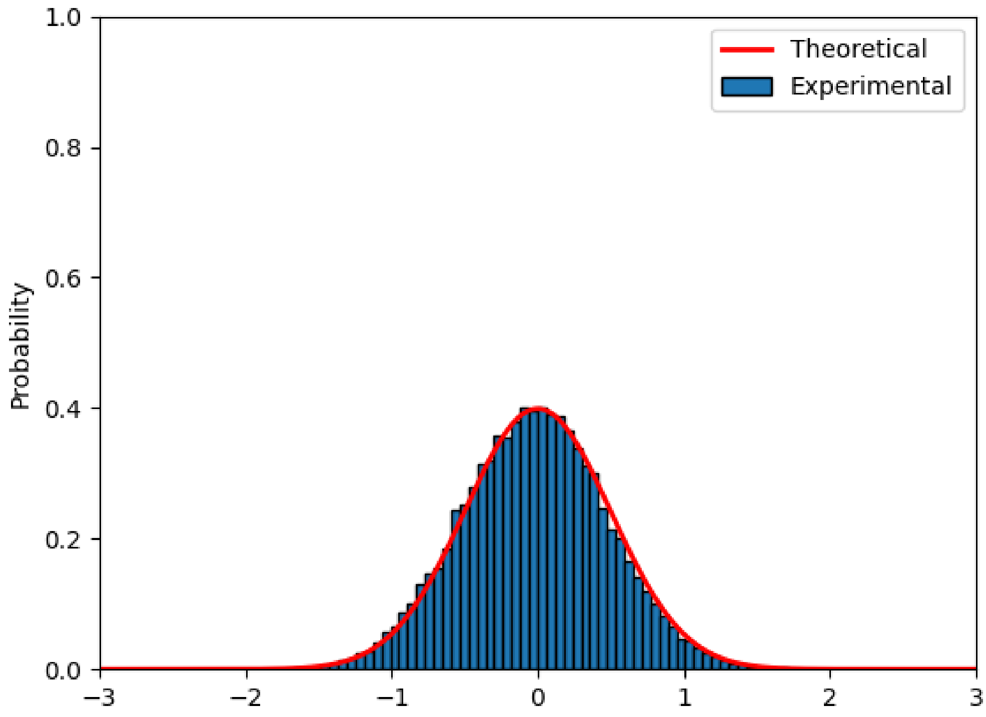


图4：N=50000

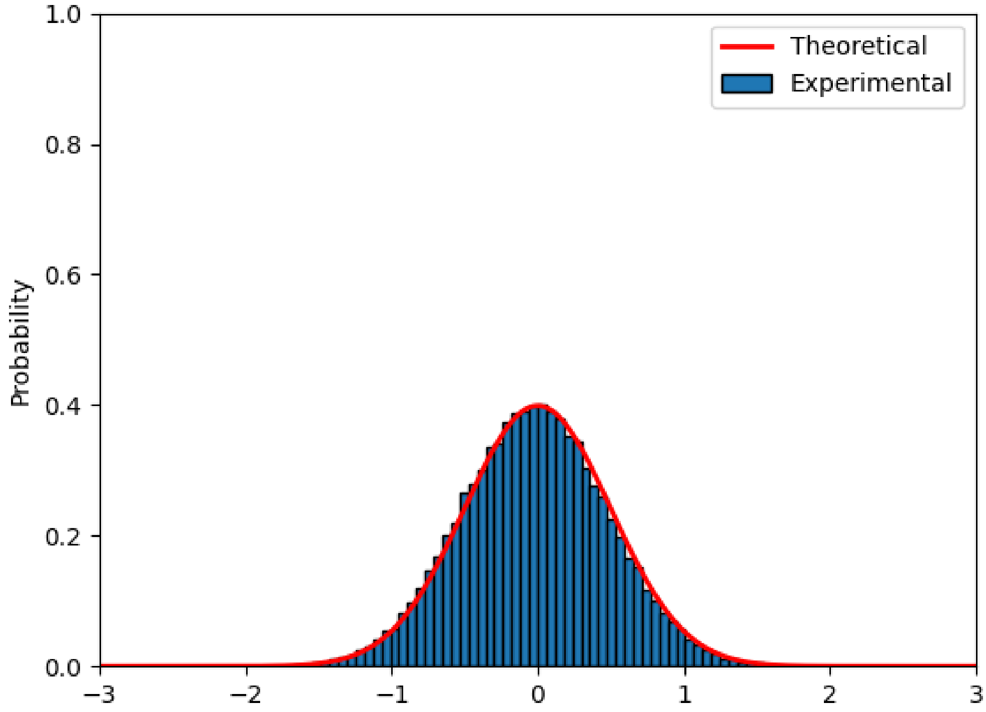


图5：N=100000

由图可以看出,随着总点数的增加，画得的概率直方图越吻合理论曲线

四.总结

本实验主要讨论了:当p(x)呈尖峰状时，采用一个函数F(x)包住f(x)来达到比用极值更加高效的舍选抽样法的目的。

当ξ_x关于ξ_1的函数较难求出时，没有依然用舍选法求解，而是采用数值求解更加简便

因此对舍选法了解更加深刻，需要根据实际情况选择不同的抽样法