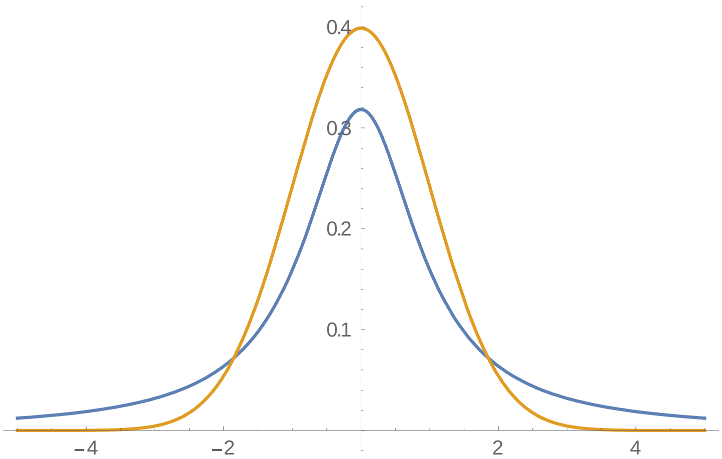
变换抽样与舍取法结合抽样

[算法及公式]：

1：抽样算法.

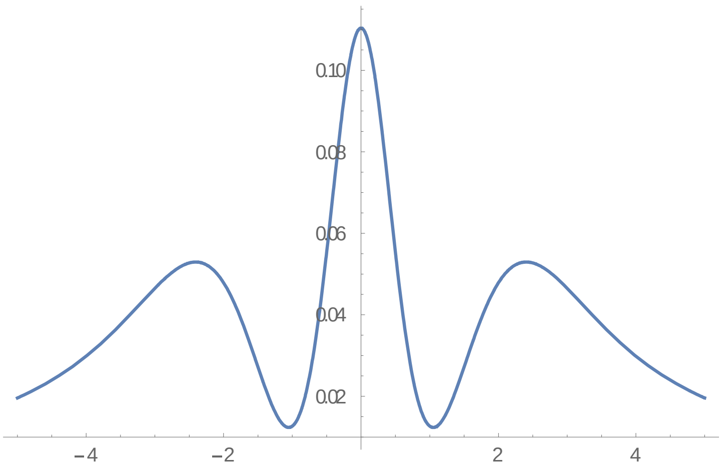
由于 Lorentz 分布的原函数是已知，故 Lorentz 分布作为 F[x]，Gauss 分布作为 p[x]。在区间[-5,5]上，做 p[x],F[x]的图像。



蓝色为 Gauss 分布。

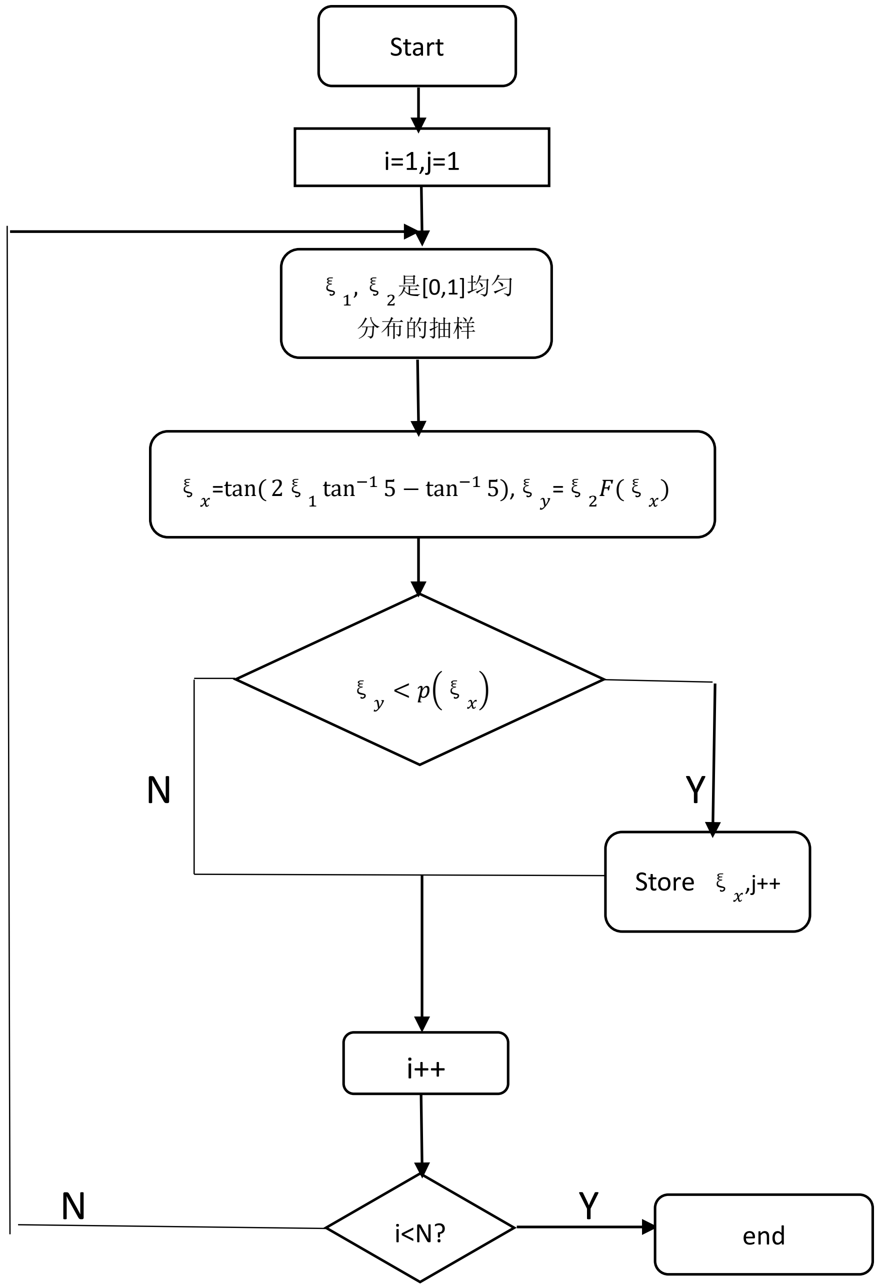
可见在区间[-5,5]上，F[x]并不恒大于 p[x]，故将 F[x]乘以一个放大因子 1.6，

即 F[x]=$\frac{1.6}{\pi}\frac{1}{1+x^2}$,p[x]=$\frac{1}{\sqrt{2\pi}}Exp(\frac{-x^2}{2})$再做 F[x]-p[x]的图像，



可见 F[x]-p[x]恒大于 0.

根据变换抽样与舍取法结合的抽样方法，取ξ$_1$，ξ$_2$是[0,1]均匀分布的抽样。



2：归一化频数直方图分布

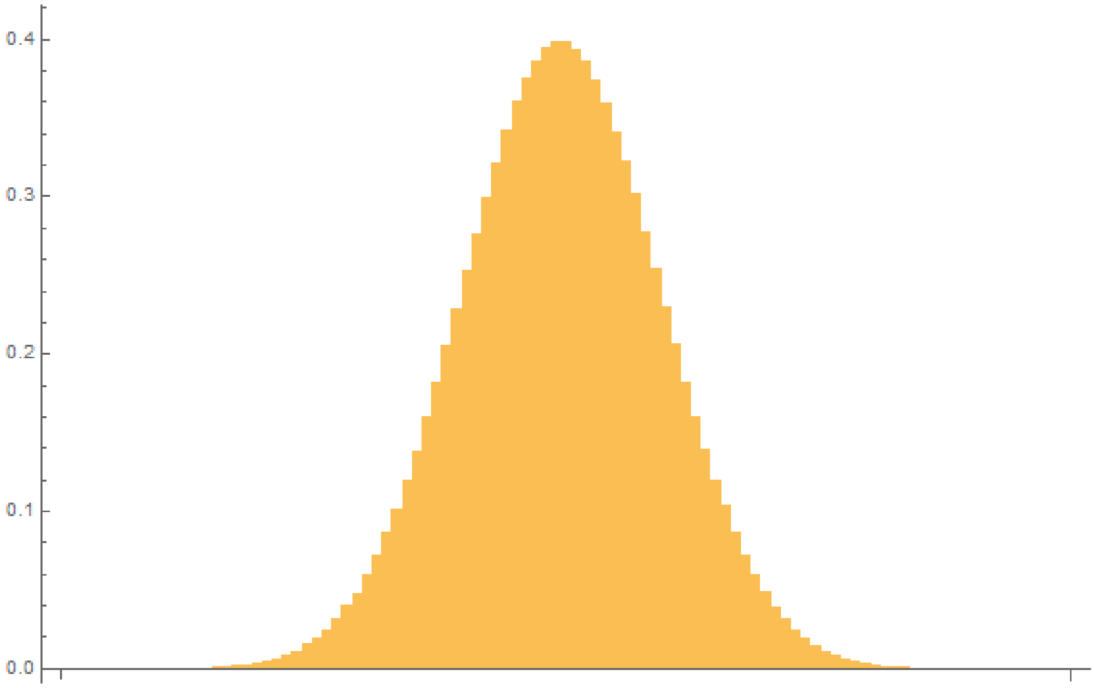
将区间[-5,5]分为 N 份，取长度为 N 的数组 st[N],初始化为 0.对ξ$_x$遍历，当ξ$_x$属于某个区间时，对应区间计数加 1.遍历完毕后，将 st 除以 N 再除以单个小区间长度，即得

到归一化概率密度分布.

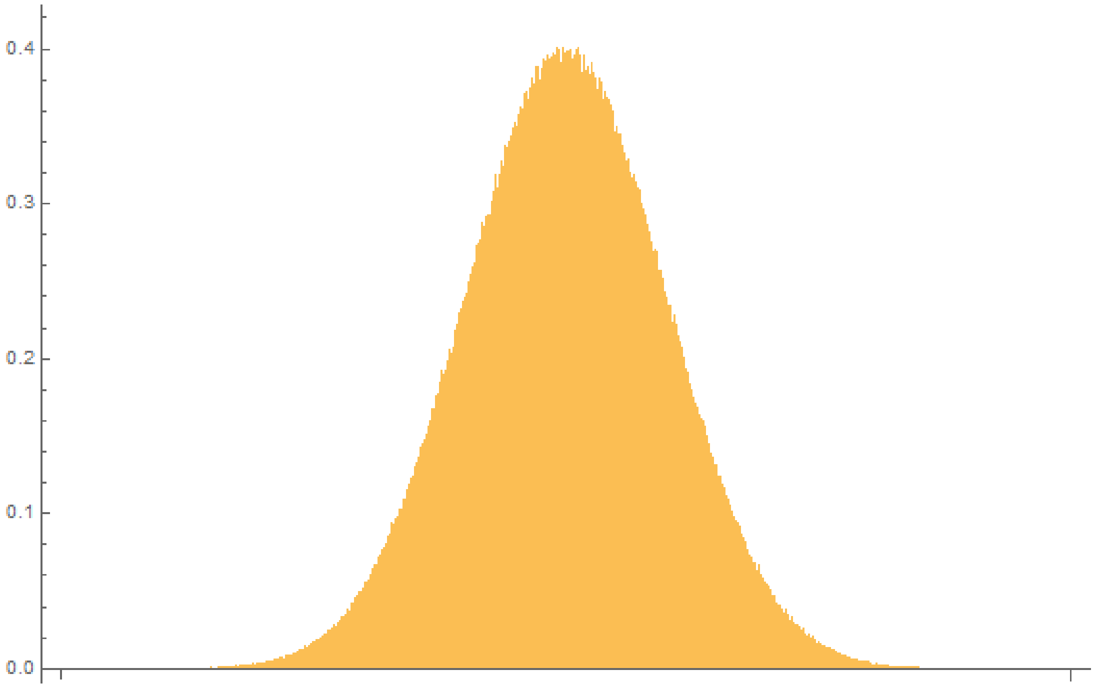
[结果与讨论]：

1： N 的选取

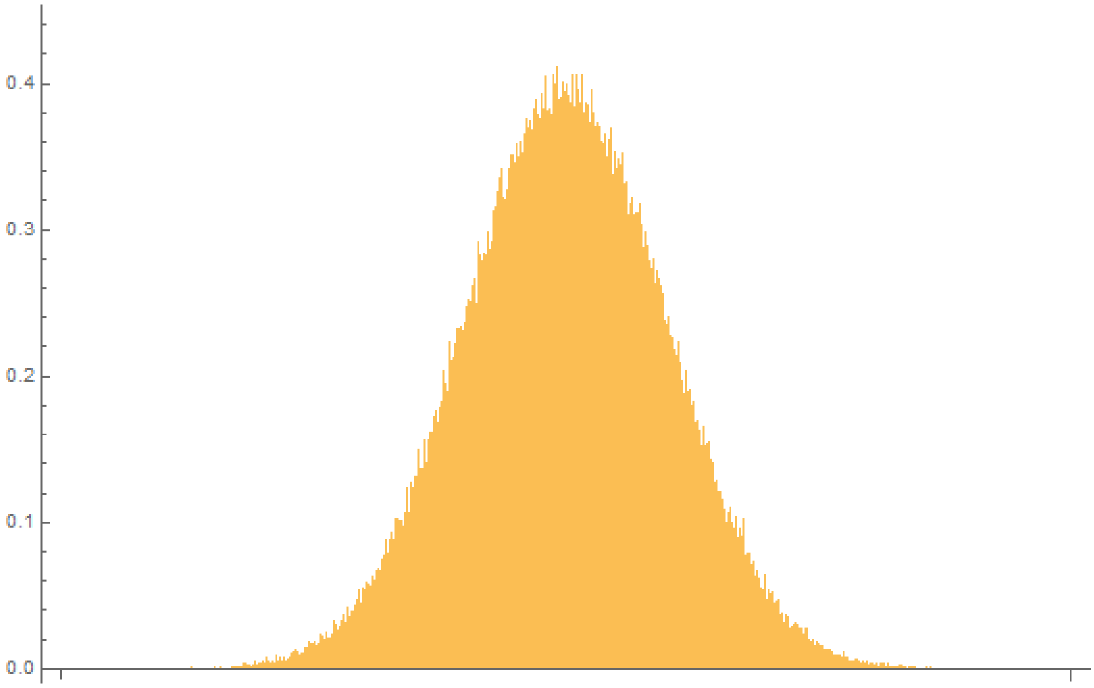
当区间数目 N 取不同值时，



N=100.



N=1000。



N=10000.

由上面三图可知，当 N（N=100）太小时，可见明显的条形分布，图形不光滑，当 N（N=10000）太大时，由于区间太小，即使同一位置附近相邻的两个区间的计数差别可能很大，体现在图形上为涨落很大。所以取 N=1000 较为合适。

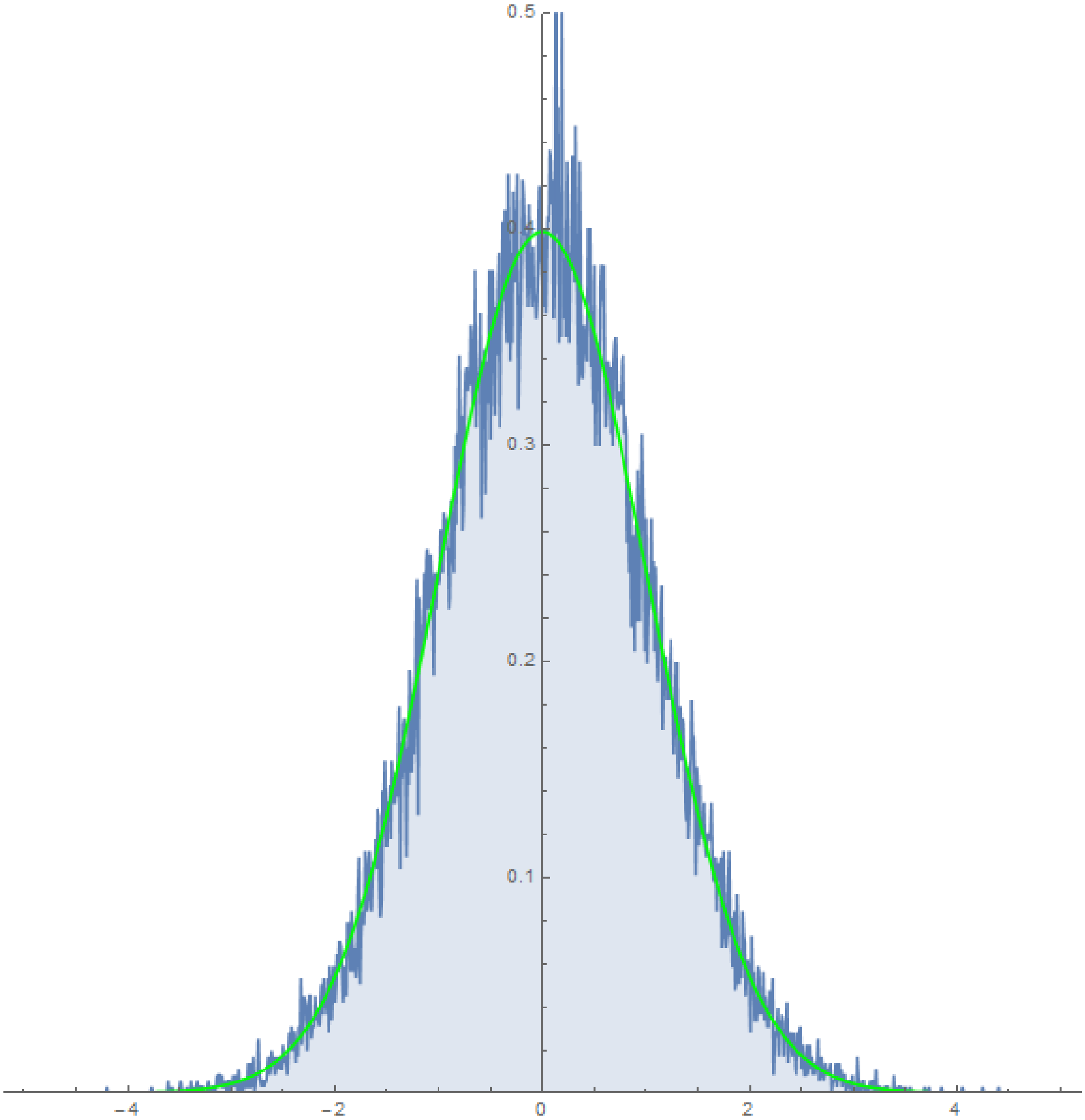
2：随机数数目的影响。

当 N 取定时，对不同的随机数数目 n，得到的归一化图形也不一样。

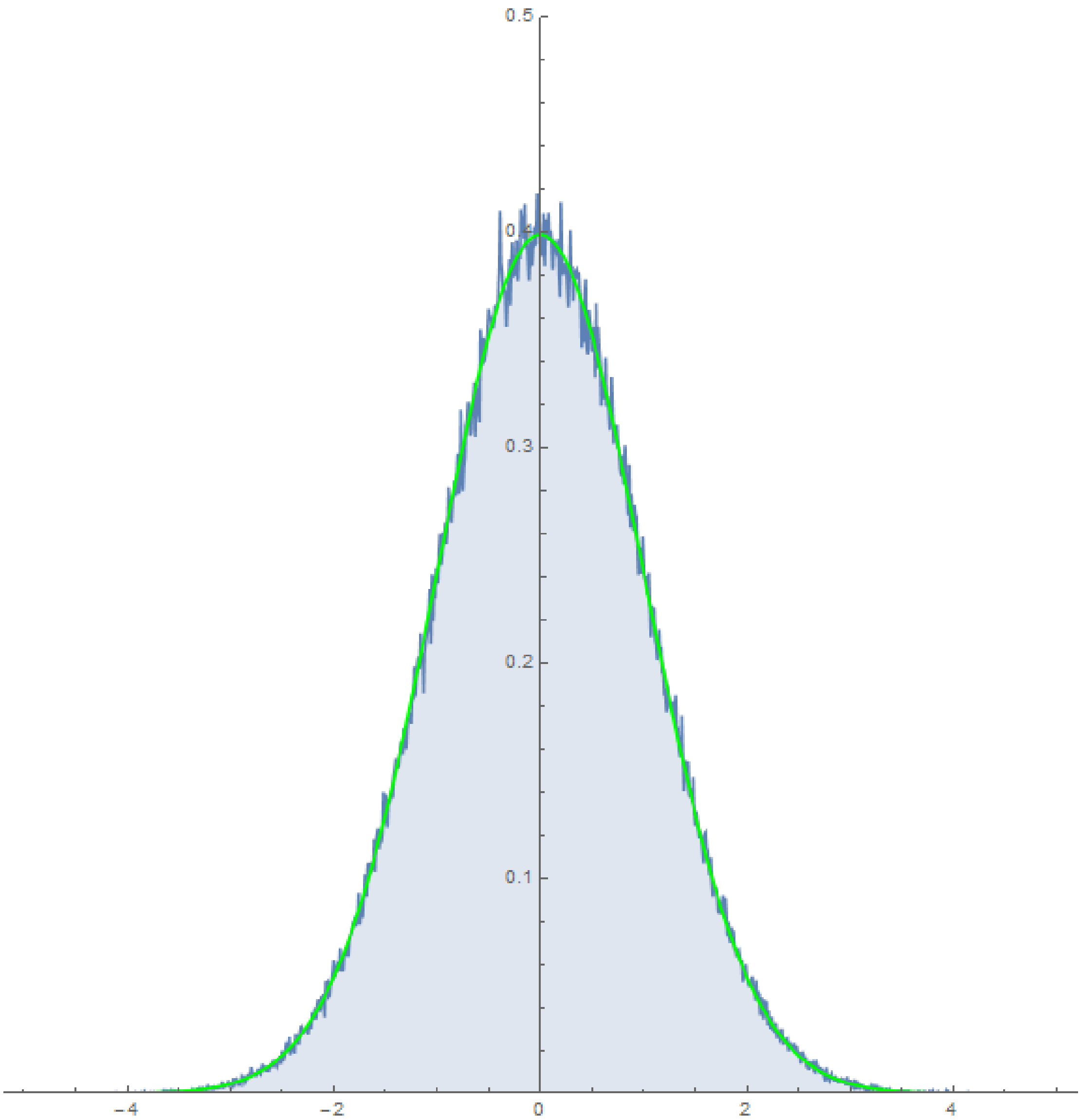
先进行理论分析，当 n 较小时，落入每个小区间的随机数较少，进行舍取法时的舍取概率可能偏离真实值较多，体现在图像上就是涨落较大。当 n 越大时，每个小区间的舍取概率越趋于真实概率，得到的曲线越光滑。

实际结果

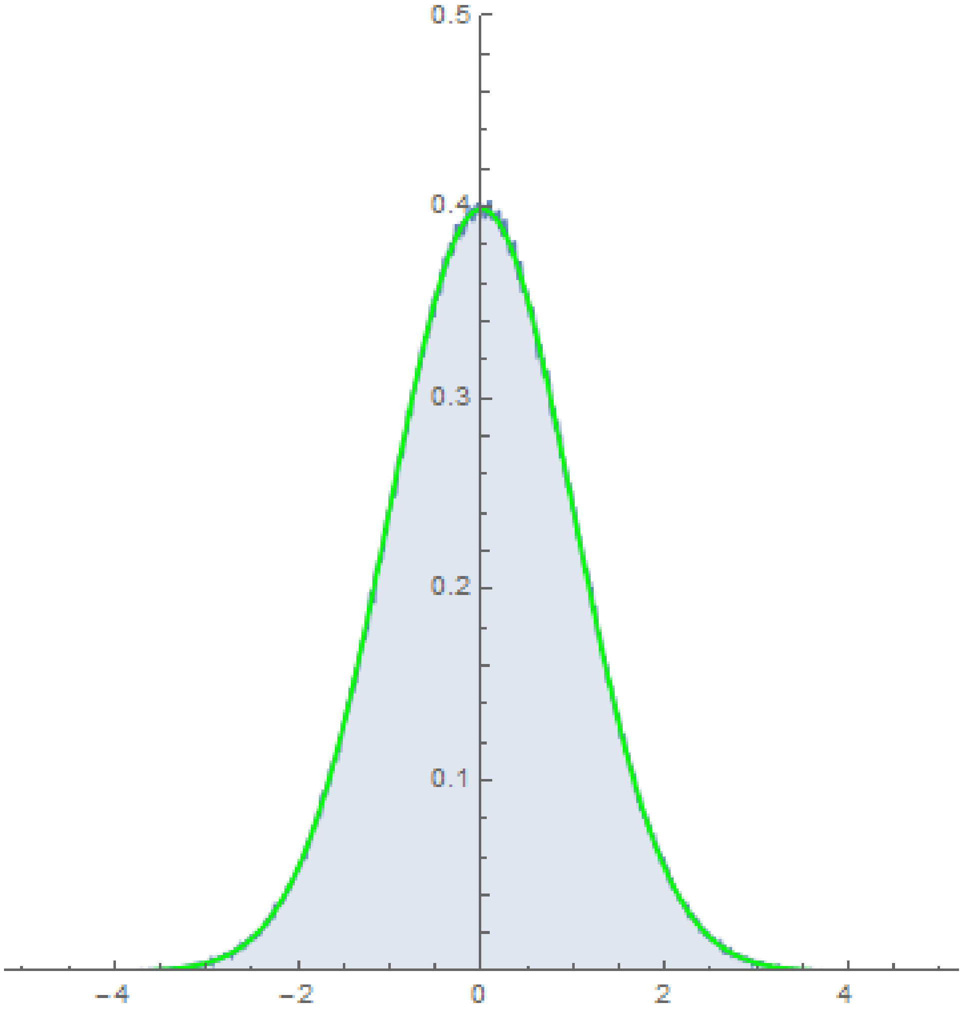
绿色曲线为 p(x)曲线



n=5w;



n=50w

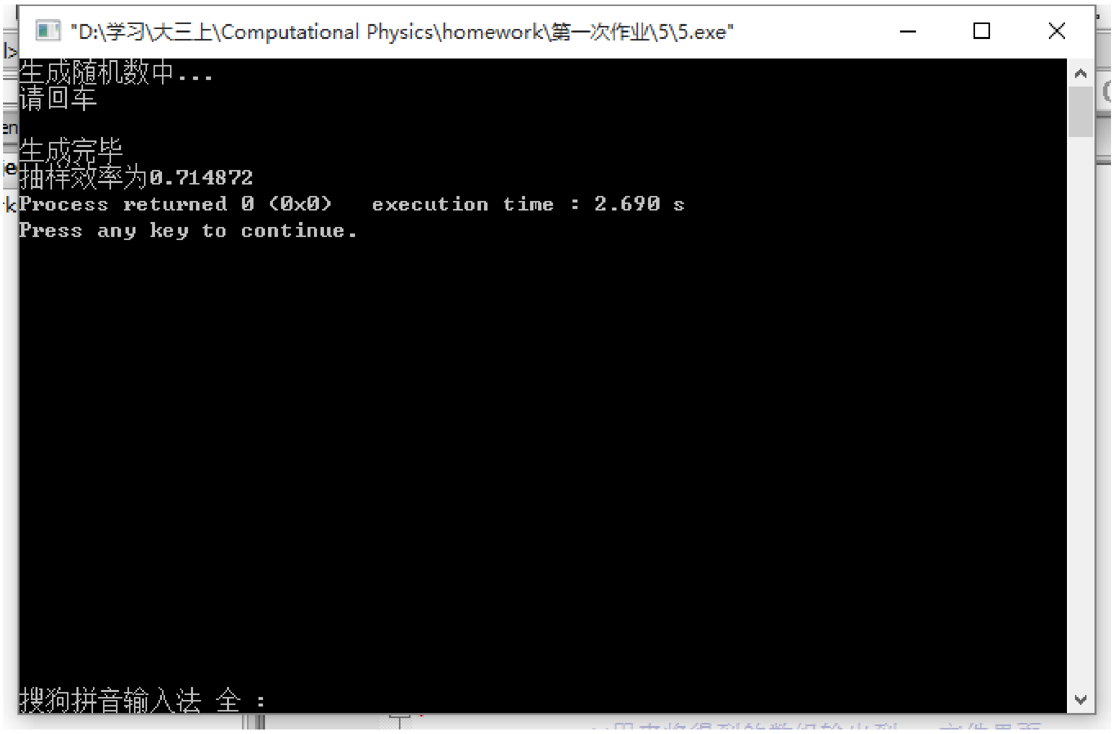


n=500w

可见当 n 取 500w 时，即每个小区间平均落入 5000 个点时，较为接近真实曲线。

3：抽样效率

N=1000，n=500w 时



抽样效率为 0.714872，

而理论抽样结果应为 p(x)的积分值与 F[x]的积分值之比。

由于 p[x]=$\frac{1}{\sqrt{2\pi}}Exp\left(\frac{-x^2}{2}\right)$,$\sigma=1$,5=5σ,所以在[-5,5]上 p[x]积分值非常接近 1，可取为 1，而 F[x]在[-5,5]上的积分值为 $\int_{-5}^{5}F[x]dx$=$\frac{1.6}{\pi}$ 2$\tan^{-1}$5=1.39893,所以理论抽样效率应为 $\frac{1}{1.39893}=0.7148297$，实际抽样效率与理论抽样效率非常接近。