

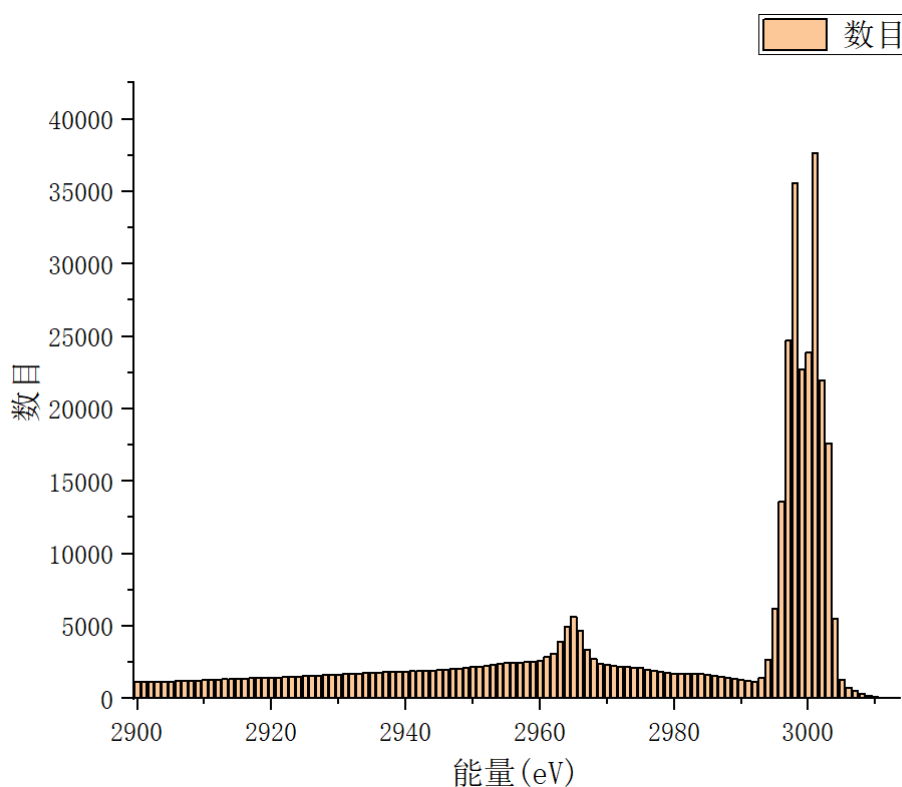
作业第七题报告

PB18020539黄韞飞

1.作业题目：对一个实验谱数值曲线 $p(x)$ ，自设 $F(x)$ ，分别用直接抽样和舍选法对 $p(x)$ 抽样。比较原曲线和抽样得到的曲线以验证，讨论抽样效率

2.公式与算法描述：

首先在ORIGIN中绘制出能量与计数的关系图，看到粒子的能量分布有一高一低两个峰。



$p(x)$ 的第一个峰在2965，峰值计数为5672。第二个峰在3000左右，峰值为37630。

直接抽样：分别生成随机数 $x \in [2900, 3013]$ 和 $y \in [0, 37630]$ ，判断点 (x, y) 是否在曲线 $p(x)$ 下方即可。

舍选法抽样：设分段函数 $F(x)$

$$F(x) = \begin{cases} 5672 & 2900 \leq x \leq 2990 \\ 37630 & 2990 < x \leq 3013 \\ 0 & \text{else.} \end{cases}$$

$F(x)$ 积分得到的分布函数为 $G(x) = \int_{2900}^x F(x)dx$

$$G(x) = \begin{cases} 5672(x - 2900) & 2900 \leq x \leq 2990 \\ 510480 + 37630(x - 2990) & 2990 < x \leq 3013 \end{cases}$$

设 ξ_1, ξ_2 是 $[0, 1]$ 上的均匀随机数

$$\xi_1 = \int_{2900}^{\xi_x} F(x)dx / \int_{2900}^{3013} F(x)dx, \xi_y = \xi_2 F(\xi_x)$$

如果 (ξ_x, ξ_y) 满足 $\xi_y < p(\xi_x)$ ，则取该点，否则舍去。

3.计算结果与讨论

取 $N = 10^6$ ，分别用直接抽样法和舍选法抽样，抽样效率：

Microsoft Visual Studio 调试控制台

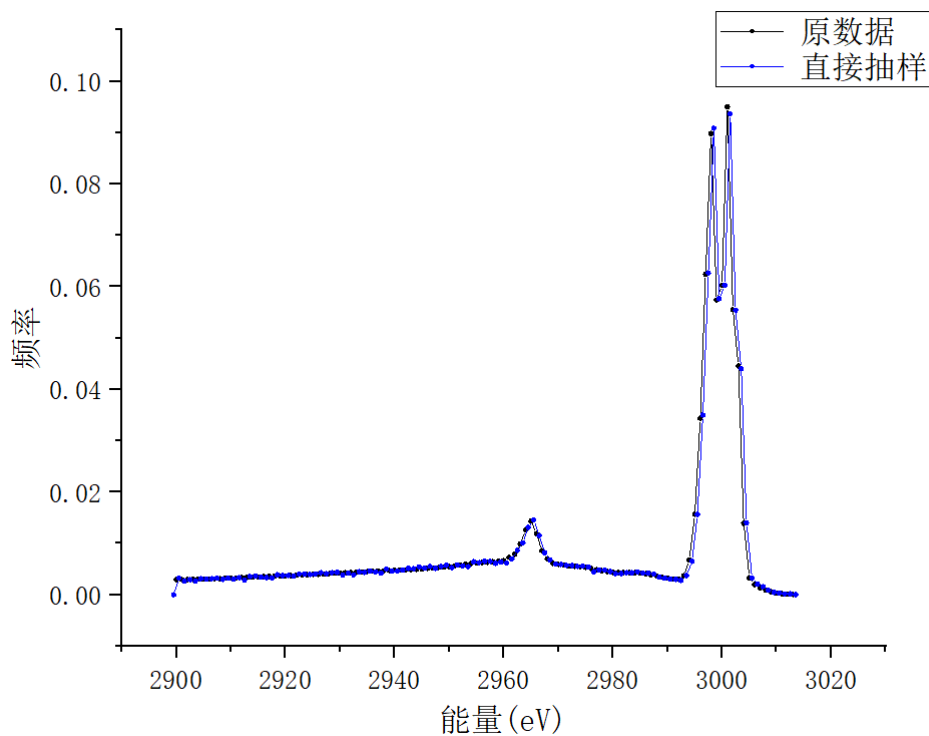
直接抽样：抽样点1000000个，有效点93024个，效率9.30%
舍选法：抽样点1000000个，有效点287185个，效率28.72%

由于直接抽样法在峰以外的地方舍去了绝大多数的点，所以抽样效率非常低，仅有9.30%。

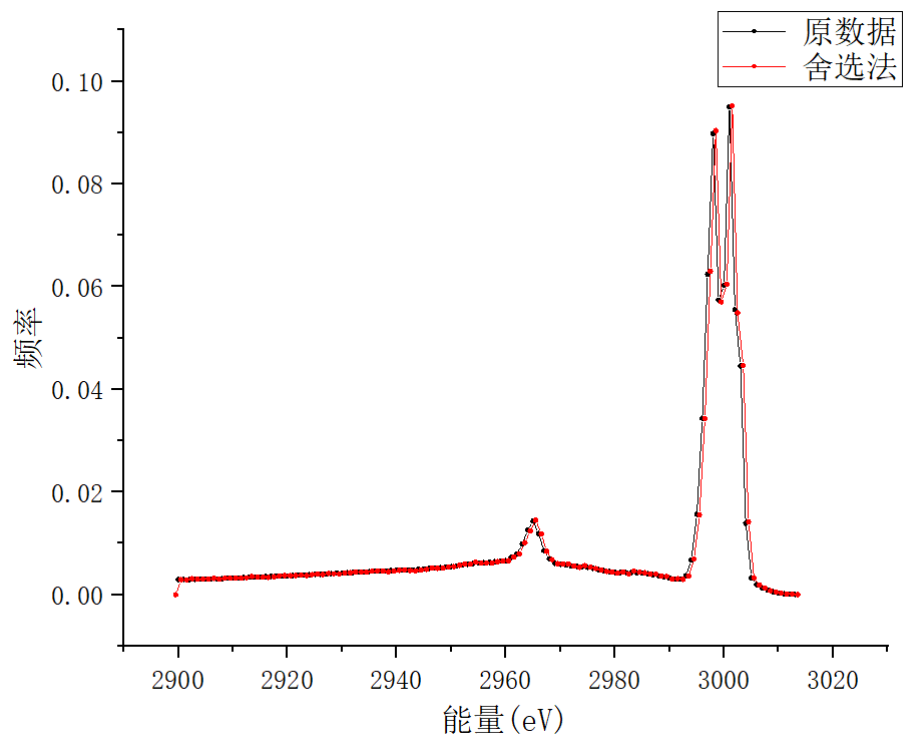
舍选法的抽样效率虽然高于直接抽样，但是由于在本题中所设的 $F(x)$ 是很简单的阶梯形函数，所以与 $p(x)$ 的符合不是特别好，导致抽样效率并不高。如果可以选取更复杂的 $F(x)$ ，比如把2900-3013区域分为4段或更多，虽然提高了抽样效率，但是也必然会提高求 $F(x)$ 的分布函数的复杂度。

将得到的数据导入ORIGIN，以1为步长统计得到的 ξ_x 的频率分布，并与原始数据data.TXT对比

直接抽样法：



舍选法：



可以看到，抽样结果与原数据的频率分布几乎完全相同。

4.结论

直接抽样法和舍选法都可以很好地得到数据 $p(x)$ 的抽样，舍选法的抽样效率更高，特别是对于这种尖峰状的实验数据，直接抽样法效率很低。