

REPORT 7

潘硕 PB24020526

2025 年 9 月 28 日

1 Question

文件 data.TXT 中存有一条光谱, 分别用直接抽样和舍选法对这个分布抽样. 比较原曲线和抽样结果, 讨论抽样效率.

2 Method

2.1 直接抽样法

data.TXT 中给出了光谱的离散分布数据 (counts - energy), 可用离散型变量分布进行抽样. $E \in [2900, 3013] \text{eV}$, $N \in [9, 37630]$, 考虑到 N 取值范围较大, 可取对数 $\log N$, 并将其归一化作为待抽样的分布 $p(E_i)$.

直接抽样法的具体步骤:

1. 在 $[0,1]$ 区间产生均匀分布的随机数 ξ_i ;
2. 若 ξ_i 满足:

$$\sum_{i=1}^{n-1} p(E_i) \leq \xi_i \leq \sum_{i=1}^n p(E_i) \quad (1)$$

则 ξ_i 取值为 E_n .

其中步骤二中, 插值查找的步骤封装于函数:

find_index.m

2.2 舍选抽样法

记 $\max N = M$, $E \in [2900, 3013] = [a, b]$, 并将 E 分成间隔均匀的 $(b - a + 1)$ 个点. 舍选抽样法的具体步骤为:

1. 在 $[0, 1] \times [0, 1]$ 区间产生均匀分布的随机样本点 (ξ_1, ξ_2) ;
2. 考虑到待抽样的 E 间隔相等, 可取 $\xi_x = [\xi_1 * (b - a + 1) + a]$ 作为由低到高分布第 ξ_x 个能量值, $\xi_y = M\xi_2$
3. 判断 $\xi_y \leq p(\xi_x)$, 则取得该点.

3 Experiment

取总样本点 $N = 10^7$, 绘制两次抽样结果直方图, 结果如下:

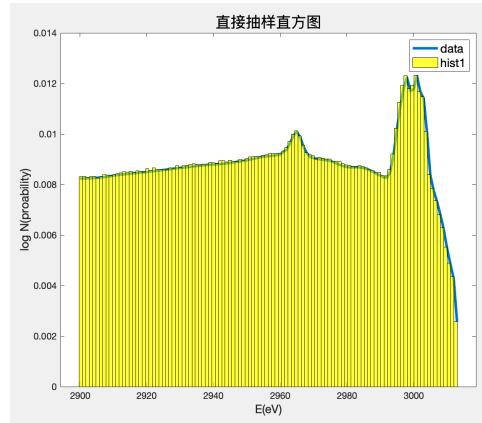


图 1: 直接抽样直方图

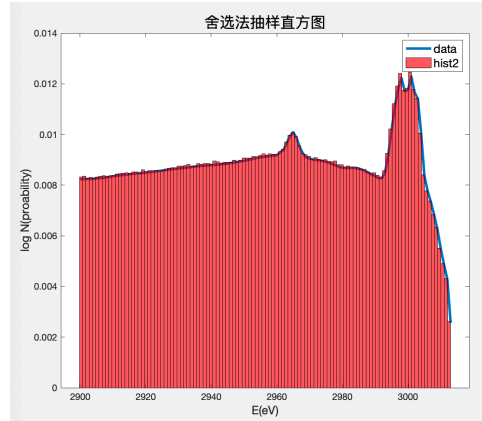


图 2: 舍选抽样直方图

两种抽样方法均与 data.TXT 原始数据拟合较好.

对两种方法的比较:

1. 抽样效率: $\eta_{\text{直方}} = 1 > \eta_{\text{舍选}} \approx 0.71$, 直方分布效率占优
2. 运行时间: $t_{\text{直方}} \approx 1.3s > t_{\text{舍选}} \approx 0.5s$, 直方分布运行速度稍慢.

4 Summary

本实验使用直接抽样法和舍选法, 对已知光谱的数据进行采样, 并得到了较好的采样结果. 实验进一步比较了两种采样结果. 直接抽样法的采样效率较高, 但是运行时间稍长. 可以通过优化插值查找函数来提高运行效率.