

Report7

PB20020480

王润泽

1 Question

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 Algorithm

2.1 直接抽样法

对于离散数据分布，首先将分布归一化处理。即对能量取值 $x(\text{eV})$

$$p(x_k) = \sum_{i=1}^k \frac{n_i}{\sum n_i} \quad x \in [2900, 3013]$$

那么对于 $[0, 1]$ 上均匀分布的 ξ ，根据 $p(x)$ 的累积函数，可以抽样出关于能量分布函数，即当满足

$$\sum_{i=1}^k p_i \leq \xi < \sum_{i=1}^{k+1} p_i$$

取 x 为 x_i

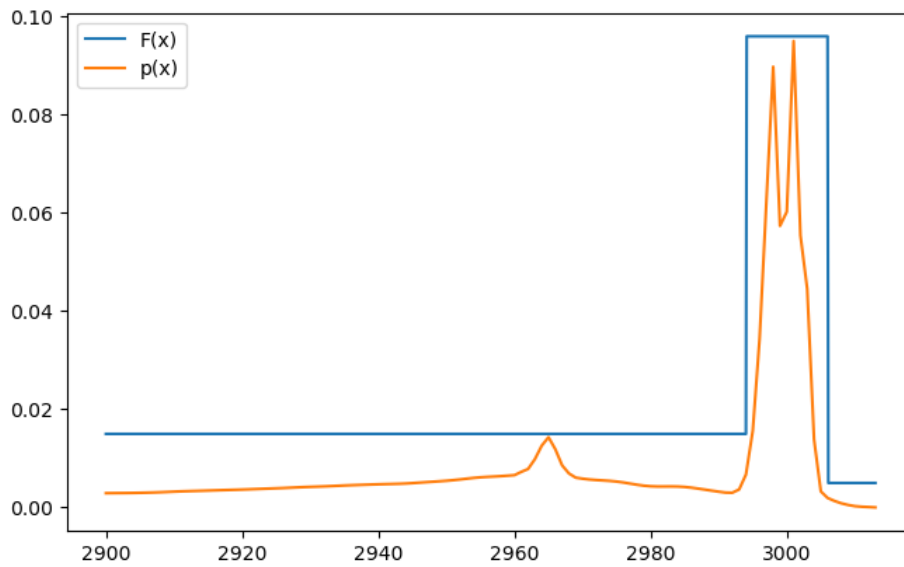
注：在代码中以 `MyDistribution` 类进行封装，`MyDistribution.ppf` 即输入 ξ 返回对应分位点 x

2.2 舍选法抽样

对于实验中数据没有直接的解析表达式，故采取舍选法选择抽样。

选取一个处处大于 $p(x)$ 的分段函数 $F(x)$

$$F(x) = \begin{cases} M_0 = 0.015, & 2900 \leq x \leq 2994 \\ M_1 = 0.096, & 2994 < x < 3006 \\ M_2 = 0.005, & 3006 \leq x \leq 3013 \\ 0, & otherwise \end{cases}$$



这样选择函数累积函数的反函数为:

$$F_{ppf}(p) = \begin{cases} 94p + 2900, & 0 \leq p \leq 0.5429 \\ 12p + 2994, & 0.5429 < p < 0.9865 \\ 7p + 3006, & 0.9865 \leq p \leq 1 \\ 0, & otherwise \end{cases}$$

这样算法抽样为:

1. 随机选择两个在[0, 1]之间均匀分布的随机抽样 ξ_1, ξ_y
2. 改变 ξ_1 的区间, 得到服从归一化的F(x)的抽样点 $\xi_x \in [2900, 3013]$

$$\xi_x = F_{ppf}(\xi_1)$$

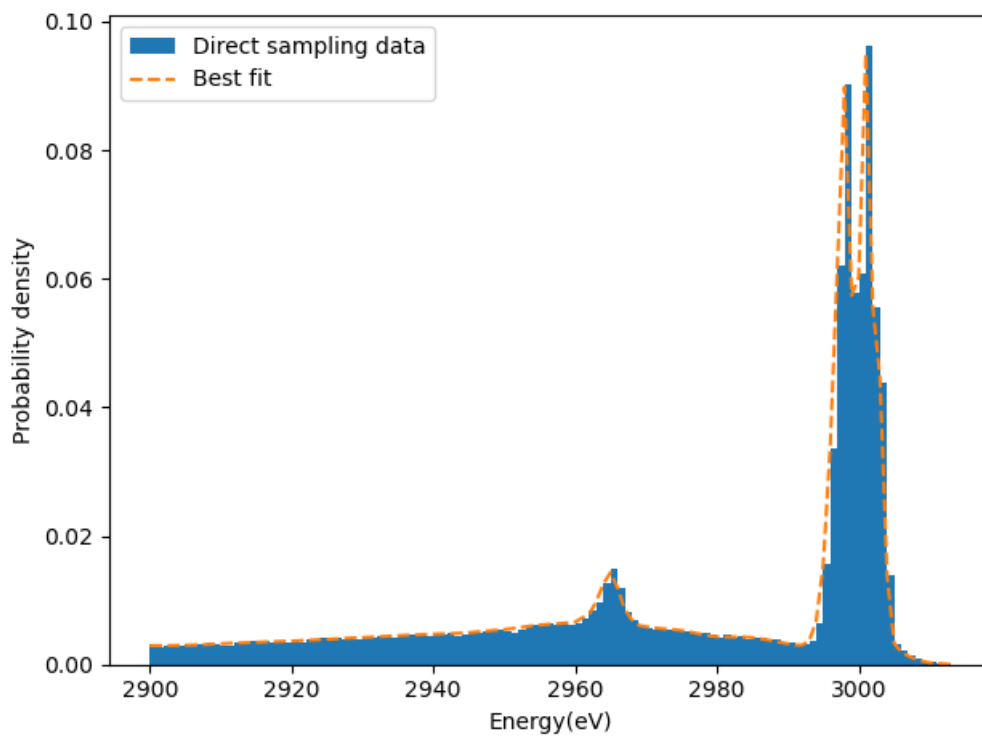
3. 判断下列条件是否成立

$$\begin{array}{ll} M_1 \xi_y \leq p(\xi_x) & \xi_x \in [2900, 2994] \\ M_2 \xi_y \leq p(\xi_x) & \xi_x \in (2994, 3006) \\ M_3 \xi_y \leq p(\xi_x) & \xi_x \in [3006, 3013] \end{array}$$

4. 否, 则舍去; 是, 则取 $x = \xi_x$

3.Experiment

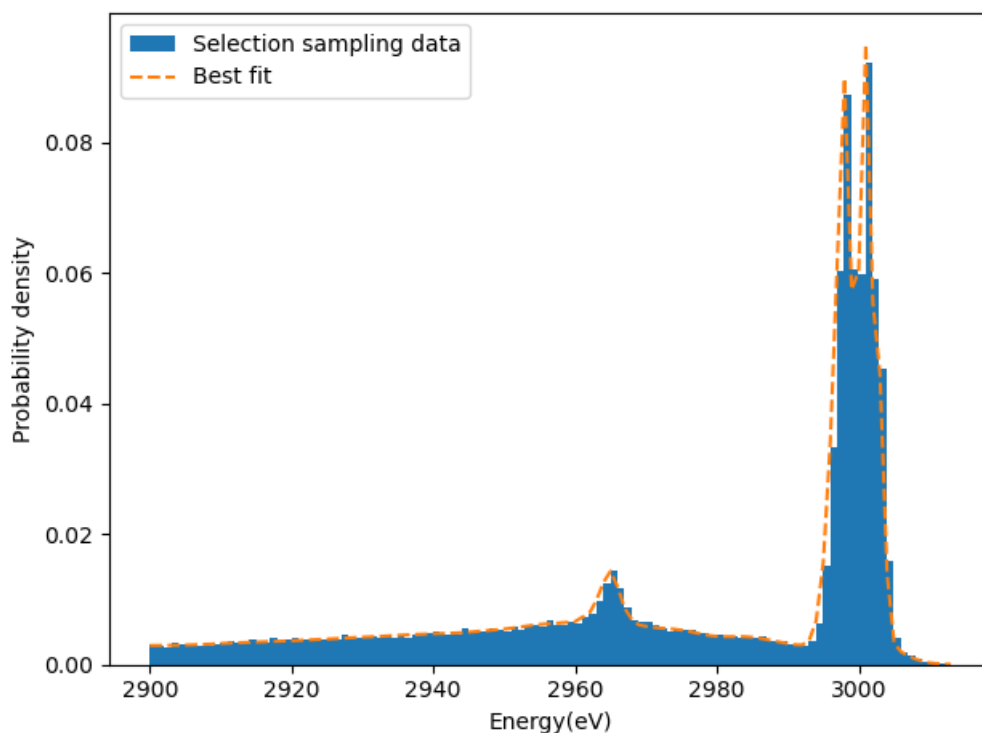
3.1 直接法抽样结果



由于是直接法抽样，所以抽样效率是 100%。

但缺点是，得到的随机变量是整数值，实际值往往应该是连续变量。

3.2 舍选法抽样结果



通过舍选法，得到连续变量的抽样效率只有有38.735%

Efficiency of Sampling:0.38735

从得到的直方图来看，仍然很吻合实验得到曲线，抽样效率低可能是由于选择函数选取的不够好。

4.Summary

4.1总结

本次实验采取了直接法与舍选法对实验所得分布进行抽样。

从结果来看得到的分布差异不大。但直接法得到是离散数据，舍选法得到的是连续变量。直接法抽样效率高，而舍选法抽样效率较低。

4.2展望

对于直接法，可以在将每个离散点插值得到一个连续函数，这样可以将离散抽样转化为连续变量抽样。

对于舍选法效率较低的问题，从图像中可以看到，主要有两个峰构成的实验分布，或许可以两个高斯分布的耦合来得到更加贴合实验分布的选择函数