

计算物理 A——Homework 7

何金铭 PB21020660

1 题目描述

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。

1.1 大致思路

虽然这里的 $p(x)$ 理论为一个离散的概率分布，然而又当间隔点足够小的时候，则可以视为一个近似的连续的概率密度函数。

1. 对 $p(x)$ 抽样可以选取直接抽样，也可以采取简单分布和第二类舍选法
2. 函数呈现多峰（且有一高峰）的形状，若采取舍选法，需要一个效率比较高的 $F(x)$
3. 再将直接抽样与舍选法的结果进行抽样效率的比较。

2 理论分析

首先画出实验数据来进行直观分析：

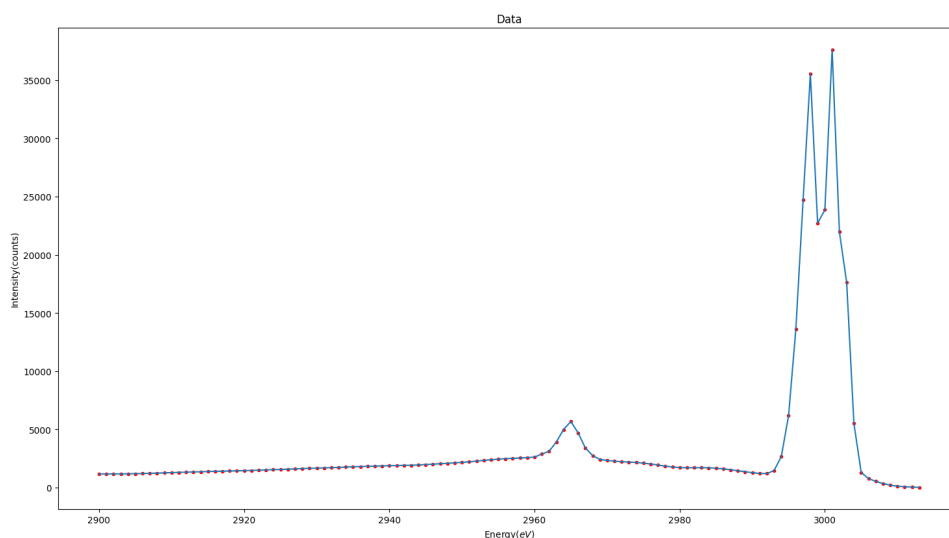
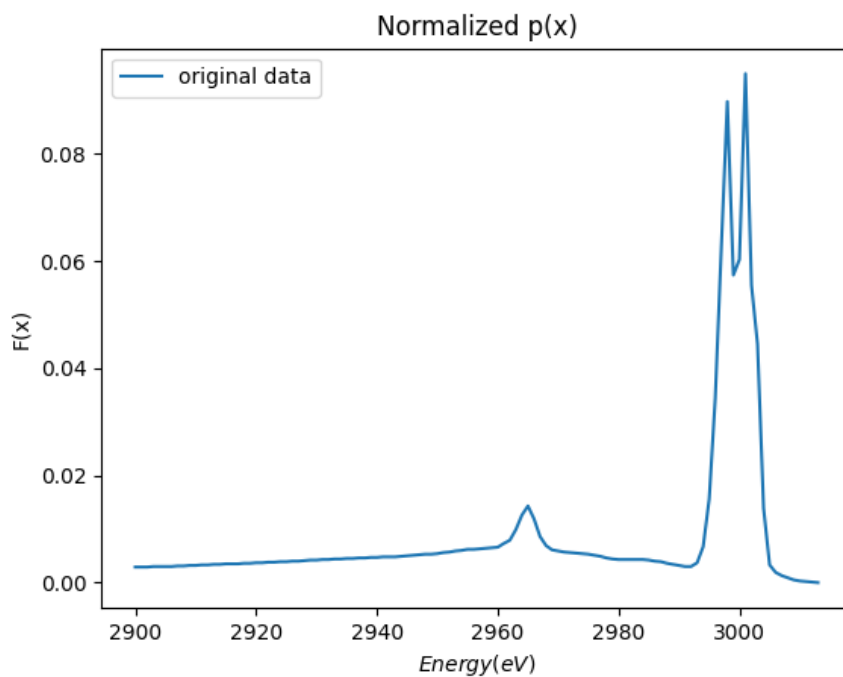


图 1: 实验数据

可以发现函数值在大部分区域中是平缓变化的，并且有一个小峰和一个大峰。

需要注意的是，这里做出的函数图像不是 $p(x)$ 的图像，需要经过归一化处理之后才能成为一个概率密度分布函数。

图 2: 归一化后的 $p(x)$

2.1 直接抽样法

原理十分简单，在前面几次作业中也有提及，详细可见算法过程。

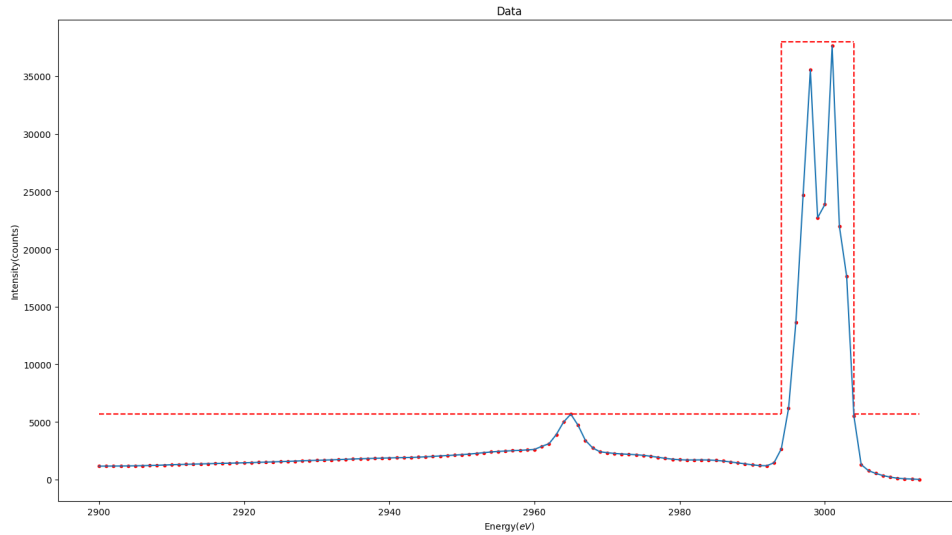
2.2 简单分别布法

简单分布是比较简单的，但是显然直接抽样的效率是很低的。其 $F(x) = 38000$

2.3 一个比较粗略的 $F(x)$

$$F(x) = \begin{cases} 5700 & x \in [2900, 2994] \\ 38000 & x \in [2994, 3004] \\ 5700 & x \in [3004, 3013] \end{cases} \quad (1)$$

以下为 $F(x)$ 的示意图。

图 3: $F(x)$

3 算法过程

3.1 直接抽样法

1. 生成随机数 $\xi \in [0, 1]$
2. 若 $\sum_{i=1}^{i=k-1} p(x_i) < \xi < \sum_{i=1}^{i=k} p(x_i)$, 则取 x_{k-1} 为样本点, 重复操作 1。

3.2 简单分布法 (给出算法但这里不演示)

1. 先生成随机数 $\xi, \eta \in [0, 1]$
2. 通过 $\xi = \frac{\xi_1 - 2900}{3013 - 2900}$ 来解得 ξ_1
3. 若 $\xi_1 \in [x_{i-1}, x_i], \eta F(\xi_1) \in [\min \{p(x_{i-1}), p(x_i)\}, \max \{p(x_{i-1}), p(x_i)\}]$ 则取 ξ_1 为满足条件的样本点; 否则重新进行 1 操作。

3.3 舍选法

1. 先生成随机数 $\xi, \eta \in [0, 1]$
2. 通过 $\xi = \int_{-\infty}^{\xi_1} F(x) dx / \int_{-\infty}^{\infty} F(x) dx$ 来解得 ξ_1
3. 若 $\xi_1 \in [x_{i-1}, x_i], \eta F(\xi_1) \in [\min \{p(x_{i-1}), p(x_i)\}, \max \{p(x_{i-1}), p(x_i)\}]$ 则取 ξ_1 为满足条件的样本点; 否则重新进行 1 操作。

其中第二个步骤中:

$$\xi_1 = \begin{cases} 169.667\xi + 2900 & \xi < 0.554028 \\ 25.45\xi + 2979.9 & 0.554028 < \xi < 0.946955 \\ 169.667\xi + 2843.333 & 0.946955 < \xi < 1 \end{cases} \quad (2)$$

4 程序说明

4.1 主要程序

schrage.c 16807 随机数生成器 (*Schrage* 法) 的代码

draw.py 由实验数据作图

sample.py 主函数, 其中有归一化函数, 直接抽样法函数, 舍选法函数等

data.csv 原始实验数据

4.2 程序结果

schrage.exe 16807 PRNG

16807.csv 一个种子数为"23942907" 的样本随机数文件

./pic 文件夹路径, 结果作图, 其中文件名中的数字代表了随机点数的次方数, "d" 代表直接抽样结果, "r" 代表舍选法结果。

unitlize.csv 原始实验数据归一化结果

direct.csv 直接抽样法抽样结果

trans.csv 舍选法抽样结果

4.3 其他说明

1. 数据都写于 CSV 文件中
2. 其中 Python 程序用到的库有:

- matplotlib.pyplot : 用于作图
- numpy : 用于数据处理
- csv : 用于读写 CSV 文件

5 结果分析

5.1 直接抽样结果

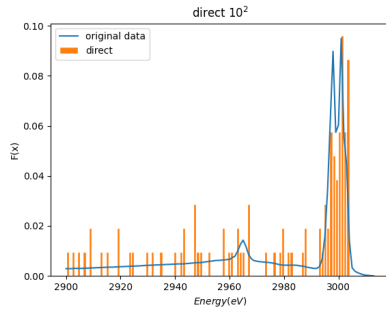


图 4: 直接抽样法 $N = 10^2$

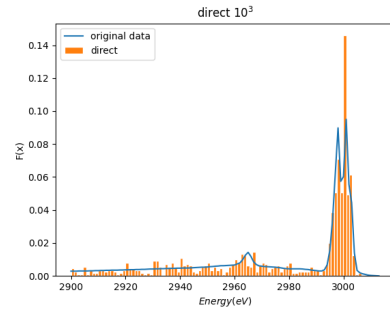


图 5: 直接抽样法 $N = 10^3$

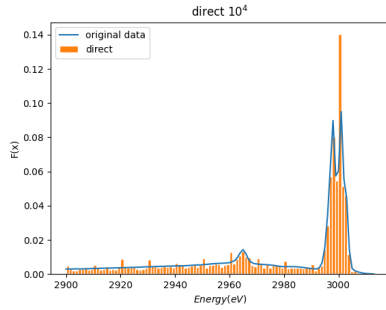


图 6: 直接抽样法 $N = 10^4$

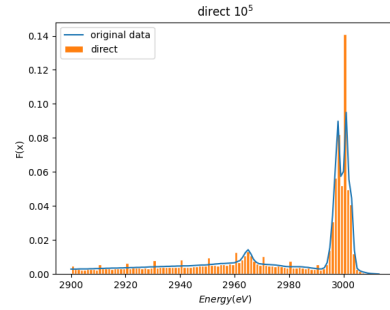


图 7: 直接抽样法 $N = 10^5$

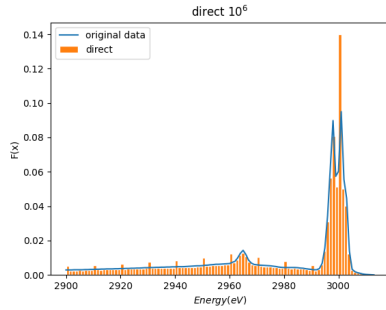


图 8: 直接抽样法 $N = 10^6$

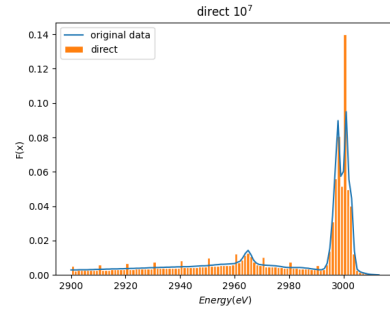
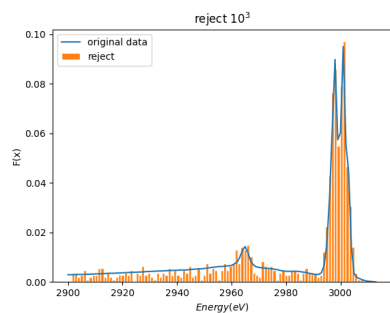
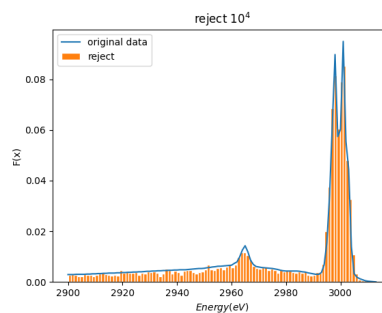
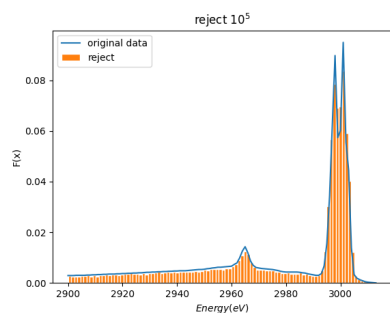
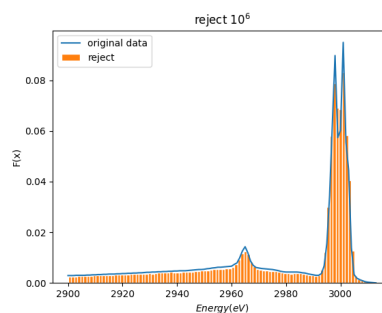


图 9: 直接抽样法 $N = 10^7$

发现直方图在 $N = 10^5$ 的时候趋于稳定状态。且发现直接抽样法会出现一些不符合原曲线的小凸峰。

5.1.1 第二类舍选法

图 10: 舍选法 $N = 10^3$ 图 11: 舍选法 $N = 10^4$ 图 12: 舍选法 $N = 10^5$ 图 13: 舍选法 $N = 10^6$

可以发现，当 $N = 10^5$ 时，直方图趋于稳定，且其图像较直接抽样法更加符合原曲线。

N	10^2	10^3	10^4	10^5	10^6
effect	0.461187	0.451104	0.458909	0.465309	0.464625

表 1: 舍选法误差随 N 的变化

6 总结

1. 从直接抽样法的结果图来看，其不如舍选法效率高，其 $N = 10^7$ 时的结果还不如舍选法 $N = 10^5$ 时的结果。
2. 发现舍选法的效率与其抽样次数无关
3. 此类舍选法的效率约为 0.4646，其值远高于简单分布