

计算物理作业七

于浩然 PB19020634 2021.10.20

1 作业题目

对一个实验谱数值曲线 $p(x)$ ，自设 $F(x)$ ，分别用直接抽样法和舍选法对 $p(x)$ 抽样. 比较原曲线和抽样得到的曲线以验证. 讨论其抽样效率.

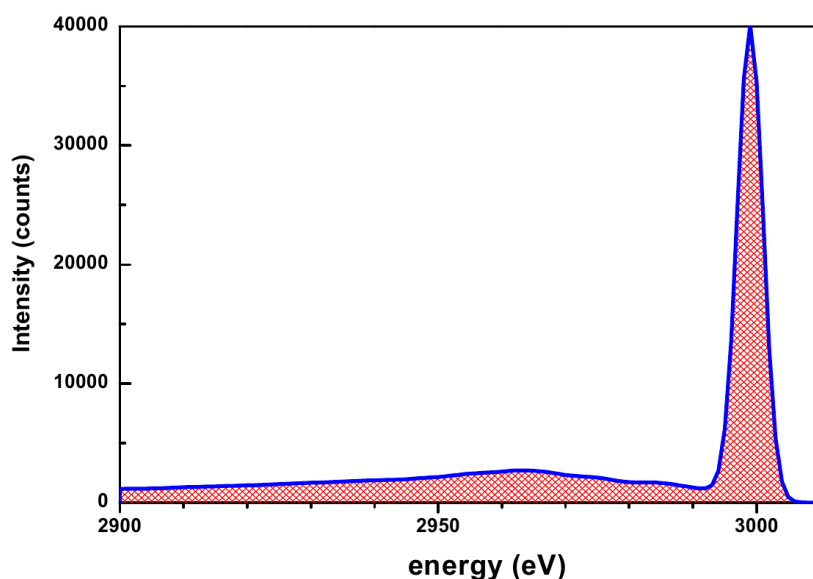


图 1: 实验谱数值曲线 $p(x)$ 示意图

2 算法简介

2.1 离散点的直接抽样法

设变量 x 是离散型的，取值为 x_1, x_2, \dots ，相应值出现的概率为 p_1, p_2, \dots . 如果从 $[0, 1]$ 区间中均匀抽样得到的随机数满足下式：

$$\sum_{i=1}^{n-1} p_i < \xi < \sum_{i=1}^n p_i \quad (1)$$

则物理量 x 取值为 x_n .

2.2 舍选抽样法

考虑到所给数据是一简单分布，为简单期间我们直接考虑用分段阶梯函数作为比较函数的舍选法. 根据 DATA.txt 文件所给数据，不妨取 $M_1 = 5700$ ， $M_2 = 38000$ ，比

较函数为如下形式：

$$F(x) = \begin{cases} M_1, & 2900 \leq x \leq 2990 \\ M_2, & 2990 \leq x \leq 3014 \end{cases} \quad (2)$$

根据舍选法的一般形式：

(1) 产生一对 $[0, 1]$ 上均匀分布的随机抽样值 (ξ_1, ξ_2) ，抽样表示式为：

$$\xi_1 = \int_{2900}^{\xi_x} F(x) dx / \int_{2900}^{3014} F(x) dx, \quad \xi_y = \xi_2 F(\xi_x) \quad (3)$$

即：

$$\xi_1 = \begin{cases} \frac{(\xi_x - 2900)M_1}{90M_1 + 24M_2}, & 2900 \leq \xi_x \leq 2990 \\ \frac{90M_1 + (\xi_x - 2990)M_2}{90M_1 + 24M_2}, & 2990 \leq \xi_x \leq 3014 \end{cases} \quad (4)$$

$$\xi_2 = \begin{cases} \xi_y / M_1, & 2900 \leq \xi_x \leq 2990 \\ \xi_y / M_2, & 2990 \leq \xi_x \leq 3013 \end{cases} \quad (5)$$

(2) 判断条件 $\xi_y \leq p(\xi_x)$ 是否成立：

- $2900 \leq \xi_x \leq 2990$

$$\xi_x = 2900 + (90 + 24M_2/M_1)\xi_1 \leq 2990 \Rightarrow 0 \leq \xi_1 \leq \frac{90}{90 + 24M_2/M_1} \quad (6)$$

判断条件为

$$M_1\xi_2 \leq p(2900 + (90 + 24M_2/M_1)\xi_1) \quad (7)$$

时，取 ξ_x 。

- $2990 \leq \xi_x \leq 3013$

$$\xi_x = 2990 + (90M_1/M_2 + 24)\xi_1 - 90M_1/M_2 \Rightarrow \frac{90}{90 + 24M_2/M_1} \leq \xi_x \leq 1 \quad (8)$$

判断条件为

$$M_2\xi_2 \leq p(2900 + (90M_1/M_2 + 24)\xi_1 - 90M_1/M_2) \quad (9)$$

时，取 ξ_x 。

若判断条件不成立，则舍。

3 编程实现

用 FORTRAN90 进行程序编写，分别展示如下。

3.1 离散点的直接抽样法

按照我们在 2.1 中所简述的方法，用一个子程序实现这一过程，将代码展示如下。由于所给实验谱数据为一个 2×114 的二维数组，在这里我们用一个长度为 228 的一维数组 `dat` 来存放这些数据，其中奇数项为 *energy*(eV) 值，偶数项为 *Intensity* 值。

```

1  SUBROUTINE DscSample(p) !离散(discrete)直接抽样
2      INTEGER(KIND=4) :: p, i, j
3      INTEGER(KIND=8) :: S = 0, dat(228) !奇数项dat(2i-1)为x值,偶
      数项dat(2i)为y值
4      REAL(KIND=8) :: sumprob(114) !几率求和值
5      REAL(KIND=8), DIMENSION(10**p) :: xi, x
6      OPEN (1, file='xi_1.dat') !从其中一个文件导入10^5个[0,1]上均
      匀分布的随机数
7      READ (1, *) xi
8      CLOSE (1)
9      OPEN (1, file='data.TXT') !读取实验谱数据
10     READ (1, *) dat
11     CLOSE (1)
12     DO i = 1, 114
13         S = S + dat(2 * i) !求粒子总数
14     END DO
15     sumprob(1) = real(dat(2)) / S !使用real()函数转换为实型使除
      法有意义
16     DO i = 2, 114
17         sumprob(i) = sumprob(i - 1) + real(dat(2 * i)) / S
18     END DO
19     DO i = 1, 10**p
20         DO j = 1, 114
21             IF(xi(i) < sumprob(j)) THEN
22                 x(i) = dat(2 * j - 1) !将x轴值作为抽出的数据
23                 EXIT
24             END IF
25         END DO
26     END DO
27     OPEN (1, file='dscsmp.dat') !将抽出数组x写入文件
28     WRITE (1, *) x
29     CLOSE (1)
30 END SUBROUTINE DscSample

```

另外为了将数据可视化与原曲线比较, 用 python 脚本绘制直方图, 代码如下:

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3 import math
4
5 plt.rcParams['savefig.dpi'] = 300
6 plt.rcParams['figure.dpi'] = 300
7
8 data = np.loadtxt('dscsmp.dat')
9 plt.xlabel('energy(eV)')
10 plt.ylabel('Intensity(counts)')
11 plt.hist(data, bins=114)
12 plt.savefig('fig_2.eps')

```

3.2 舍选抽样法

将舍选法所用数据 $M_1 = 5700$, $M_2 = 38000$ 代入参数, 可得:

$$\frac{90}{90 + 24M_2/M_1} = 0.360000 \equiv C \quad (10)$$

本 FORTRAN90 程序直接使用 Schrage 方法产生的一对 10^5 个随机数序列, 分别使用种子 461655342 和 96145235, 附于附件中.

由于 data.TXT 所给的数据为离散数据, 故我们在进行比较时使用 int() 函数将由均匀随机数序列生成的 [2900, 3014] 中的浮点型随机数 **转化为整型**, 再与 intensity 中的实验谱数据进行比较 (如 20 行、25 行). 在后面的计算结果中将讨论这种做法的可行性.

```

1 SUBROUTINE Sample(p) !舍选法抽样
2     INTEGER(KIND=4) :: dat(228), intensity(114), p , i
3     REAL(KIND=8) ,PARAMETER :: C = 0.36 !定义的常数C, 是xi_1的分
      界线
4     REAL(KIND=8) ,DIMENSION(10**p) :: xi_1, xi_2, xi_x, x
5     OPEN (1, file='data.TXT')
6     READ (1, *) dat
7     CLOSE (1)
8     DO i = 1, 114

```

```

9      intensity(i) = dat(2 * i) !将data中的y值写入数组，索引i
      对应energy为(2899+i)eV
10  END DO
11  OPEN (1, file='xi_1.dat') !将均匀分布的两个随机数序列写入数
      组xi_1, xi_2
12  READ (1, *) xi_1
13  CLOSE (1)
14  OPEN (1, file='xi_2.dat')
15  READ (1, *) xi_2
16  CLOSE (1)
17  DO i = 1, 10**p
18      IF (xi_1(i) < C) THEN
19          xi_x(i) = 2900 + (90 + 24 * 38000.0 / 5700) * xi_1(i)
20          IF (5700 * xi_2(i) <= intensity(int(xi_x(i)) - 2899)
21              ) THEN
22              x(i) = xi_x(i) !若上面的判断条件满足则取xi_x(i)
23              作为抽样结果x(i)
24          END IF
25      ELSE
26          xi_x(i) = 2990 + (90 * 5700.0 / 38000 + 24) * xi_1(i)
27          - 90 * 5700.0 / 38000
28          IF (38000 * xi_2(i) <= intensity(int(xi_x(i)) -
29              2899)) THEN
30              x(i) = xi_x(i)
31          END IF
32      END IF
33  END DO
34  OPEN (1, file='smp.dat')
35  DO i = 1, 10**p
36      IF(x(i) .ne. 0) THEN
37          WRITE (1,*) x(i) !将非零值(成功抽出的值)写入文件
38      END IF
39  END DO
40  CLOSE (1)
41  END SUBROUTINE Sample

```

用 python 绘制直方图，代码如下：

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 import math
4
5 plt.rcParams['savefig.dpi'] = 300
6 plt.rcParams['figure.dpi'] = 300
7
8 data = np.loadtxt('data.TXT')
9 plt.xlabel('energy(eV)')
10 plt.ylabel('Intensity(counts)')
11 plt.hist(ex, bins=114)
12 plt.savefig('fig_1.eps')
```

4 计算结果

4.1 直接抽样法

直方图显示如下:

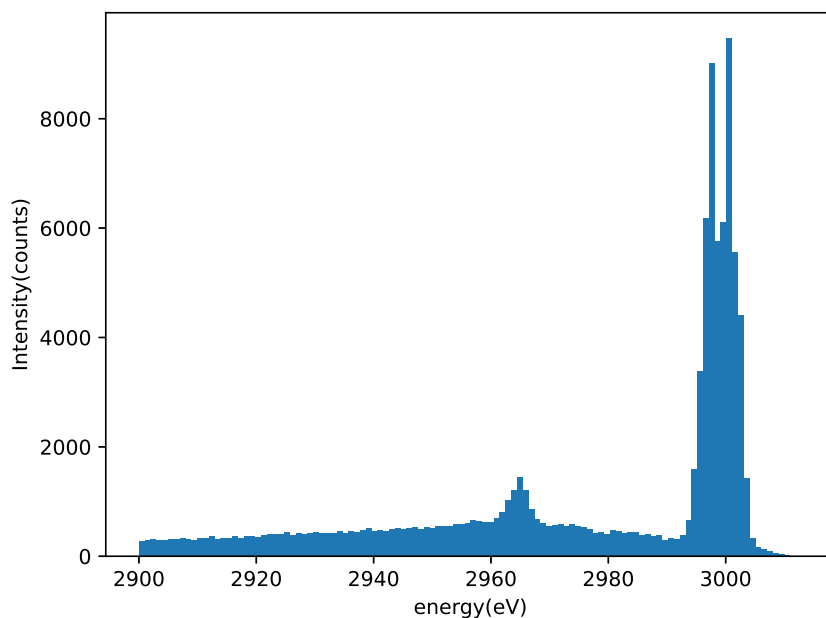


图 2: 直接抽样法结果直方图

图中纵轴为取值为特定整型值的抽样点的个数. 可见, 直接抽样法很好地显示了原曲线的特征, 对于这种取值点数较少的离散型抽样有很好的可行性.

4.2 舍选抽样法

将绘制的直方图显示如下：

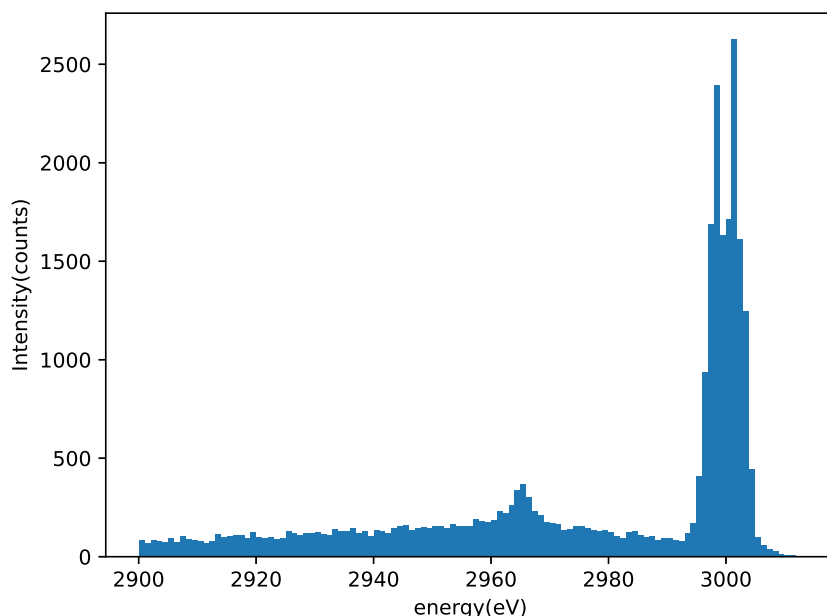


图 3: 舍选抽样法结果直方图

从此直方图可以明显看出实验谱数值曲线的轮廓，在 2665eV 的峰十分明显，在 3001eV 的峰也有明显的轮廓，但在峰中心的值偏少，然而在直接抽样法的结果中也出现了这一情况，表明这应是所给数据性质所导致的。舍选抽样法也得到了很好的结果。

从此数据实验的结果来看，我们对此离散数据谱的连续化是合理的。

考虑抽样效率，10000 个随机点结果抽出 27720 个点，可得抽样效率：

$$\frac{27720}{100000} = 0.2772 \quad (11)$$

5 结论

本题运用两种抽样法都得到了很好的模拟结果，我们加深了对舍选抽样的理解，并熟练了离散化数据直接抽样的步骤。