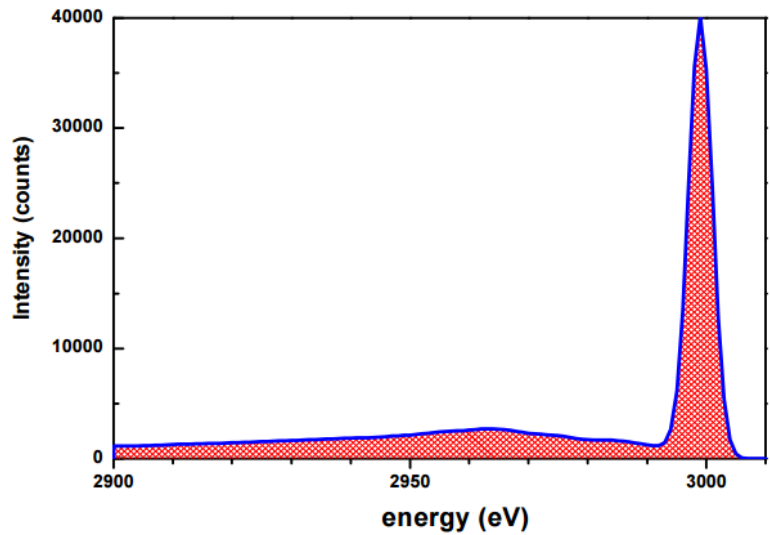


计算物理 作业报告10

PB14203209 张静宁 2017.11.07

第十题

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。



算法公式

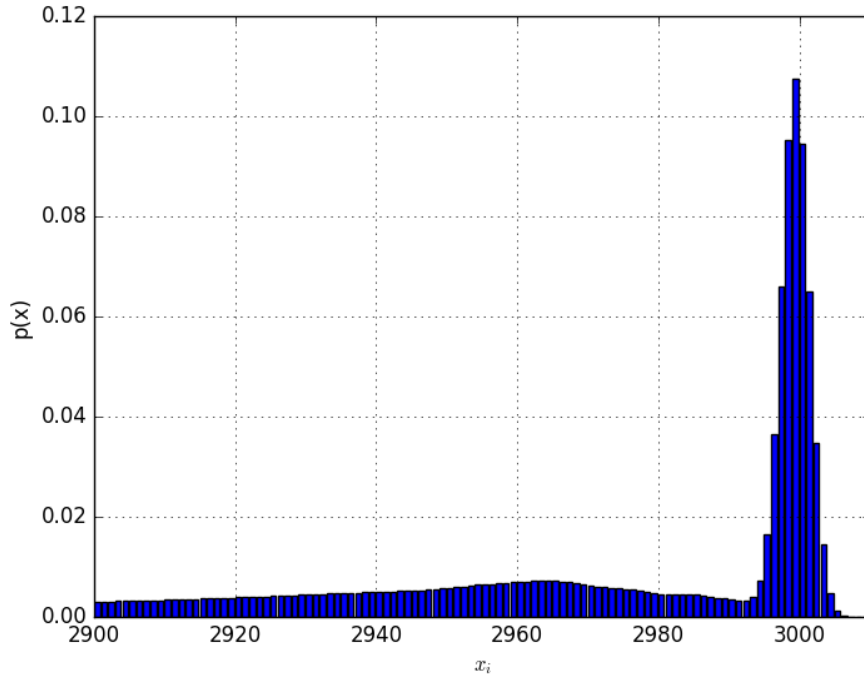
直接抽样

由于题目附加的data.TXT文件给出了111个实验数据点，由此可以得到离散型变量分布 X 。

横坐标为能量值，即为随机变量 $x_1 = 2900, x_2 = 2901, \dots, x_n = 3010, n = 111$,

纵坐标为该能量对应的频数 N_i ，把频数除以总的频数，可以得到每一个区间 $[x_i, x_{i+1}]$ 的概率，即 $p_i = \frac{N_i}{\sum_{i=1}^{n-1} N_i}$

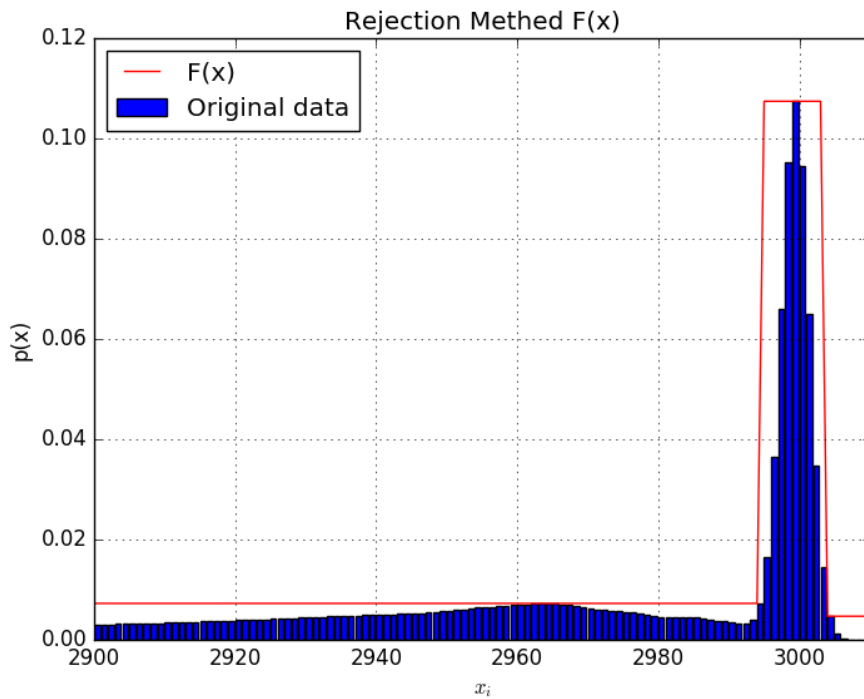
由于 $x_n = 3010$ 对应的频数为0，可以作为边界点剔除掉，即考虑300个区间里的频率。



故直接抽样方法为，从 $[0, 1]$ 区间中均匀抽样得到随机数 ξ ，如果 $\sum_{i=1}^{k-1} p_i < \xi \leq \sum_{i=1}^k p_i$ ，则能量 x 取值为 x_k 。

如果考虑连续型随机变量，则可以将相邻的数据点用线段连接， $p(x)$ 曲线即为所有折线，相当于分段线性插值。

舍选法



舍选法，采用分段阶梯函数 $F(x)$ 作为比较函数：

$$F(x) = \begin{cases} a = 0.00726337, & 2900 \leq x < 2995 \\ b = 0.10741919, & 2995 \leq x < 3004 \\ c = 0.00469857, & 3004 \leq x \leq 3010 \end{cases}$$

且 $F(x)$ 曲线下面积为 $S = \int_{2900}^{3010} F(x)dx = 1.68498$.

$$\xi_1 = \frac{\int_{2900}^{\xi_x} F(x)dx}{\int_{2900}^{3010} F(x)dx} = \begin{cases} a_1 * (x - 2900), & 2900 \leq x < 2995 \\ b_1 * (x - 2995) + m_1, & 2995 \leq x < 3004 \\ c_1 * (x - 3004) + m_2, & 3004 \leq x \leq 3010 \end{cases}$$

其中

$$a_1 = a/S = 0.00431066, b_1 = b/S = 0.063751, c_1 = c/S = 0.0027885, m_1 = 0.409512, m_2 = 0.983272.$$

由此可得 $\xi_x = \xi_x(\xi_1)$, $\xi_y = \xi_2 F(\xi_x)$

$$\xi_x(\xi_1) = \begin{cases} \xi_1/a_1 + 2900, & 0 \leq \xi_1 < m_1 \\ (\xi_1 - m_1)/b_1 + 2995, & m_1 \leq \xi_1 < m_2 \\ (\xi_1 - m_2)/c_1 + 3004, & m_2 \leq \xi_1 \leq 1 \end{cases}$$

其中 ξ_1, ξ_2 为 $[0, 1]$ 均匀随机抽样.

程序说明

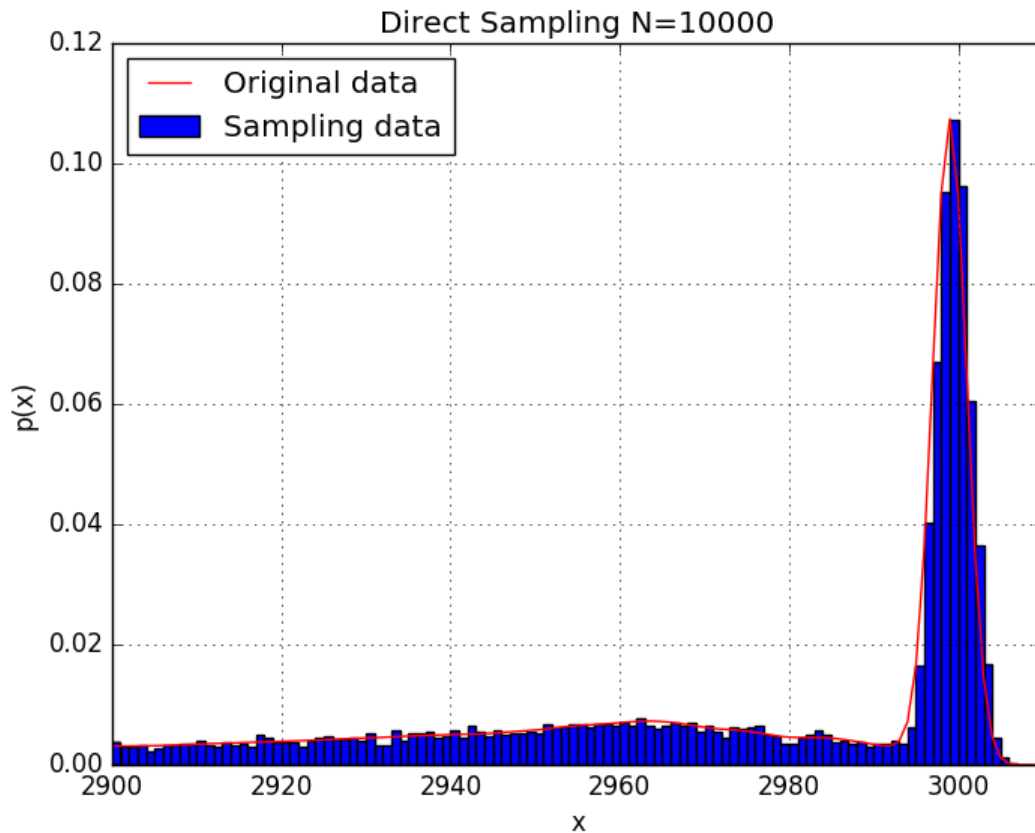
- direct_sample.c 读取数据文件，实现直接抽样，输出抽样值
- rejection_sample.c 读取数据文件，实现舍选法抽样，输出抽样值

计算结果

直接抽样

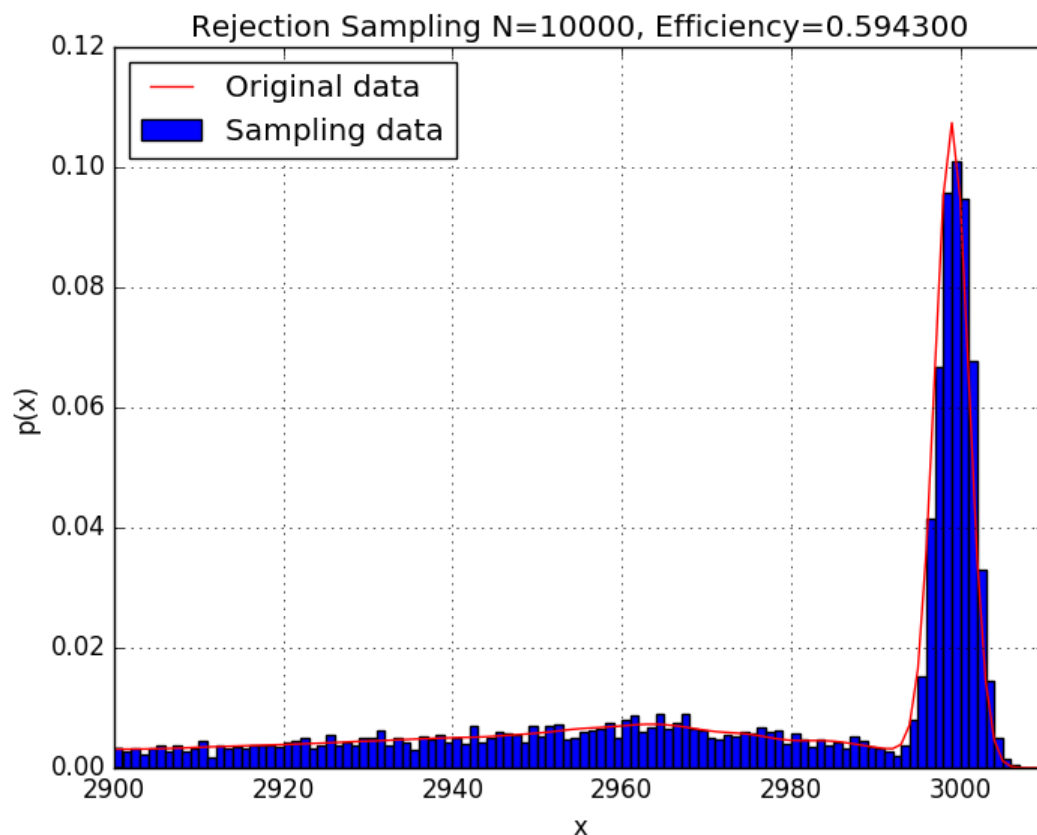
抽样 $N = 10^4$ 的点，直接抽样法得到的频率分布直方图和原曲线 $p(x)$ 较符合， $p(x)$ 为数据点连接而成的折线段.

由于是用 N_i 表示区间 $[x_i, x_{i+1}]$ 的频数，故直方图相比原曲线偏右移，直接抽样的效率为1.



舍选法抽样

抽样效率为 **0.5943**，比较高。



分析、总结

1. 直接抽样法、舍选法抽样得到的频率分布直方图和原曲线 $p(x)$ 都比较符合
2. 直接抽样的效率为1
3. 舍选法抽样的效率取决于比较函数的选取，阶梯分段函数的分得越细，抽样效率越高，但永远小于直接抽样效率。而且分段越系操作起来越麻烦，无论是表达式的推导还是写代码，都会增加工作量。
4. 因此，应该根据数据的类型、精度的要求来决定采用哪种采样方法，如本题给了111个有效数据点，由于 x 是等间隔选取的，故直接用阶梯函数来代替原函数 $p(x)$ ，同时比较函数也选取为阶梯函数。