计算物理第6题

PB18000039 徐祺云

一　作业题目

对两个函数线型($Gauss$分布和类$Lorentz$型分布)，设其一为$p(x)$，另一为 $F(x)$，用舍选法对$p(x)$抽样。将计算得到的归一化频数分布直方图与理论曲线$p(x)$进行比较，讨论差异。讨论抽样效率。

$$\text{Gaussian: } \sim \exp\left(-\frac{x^2}{2}\right); \quad \text{Lorentzian like: } \sim \frac{1}{1+x^4}$$

二　算法及主要公式

设$Guass$分布函数为$p(x)$，类$Lorentz$分布函数为$F(x)$，舍选法对$p(x)$抽样。我们知道，当曲线$p(x)$呈尖峰形状时，抽样效率很低，这时需要把变换抽样与舍选法结合起来，将简单分布舍选法里的$y = M$换为$y = F(x)$，要求是$F(x) > p(x), a \leqslant x \leqslant b$。

选定$a = -5, b = 5$，即在区间$[-5, 5]$上，设$p(x) = C_1 e^{-\frac{x^2}{2}}$，归一化常数 $C_1 = \left[\int_{-5}^{5} e^{\frac{-x^2}{2}} dx\right]^{-1} \approx 0.398943$，即

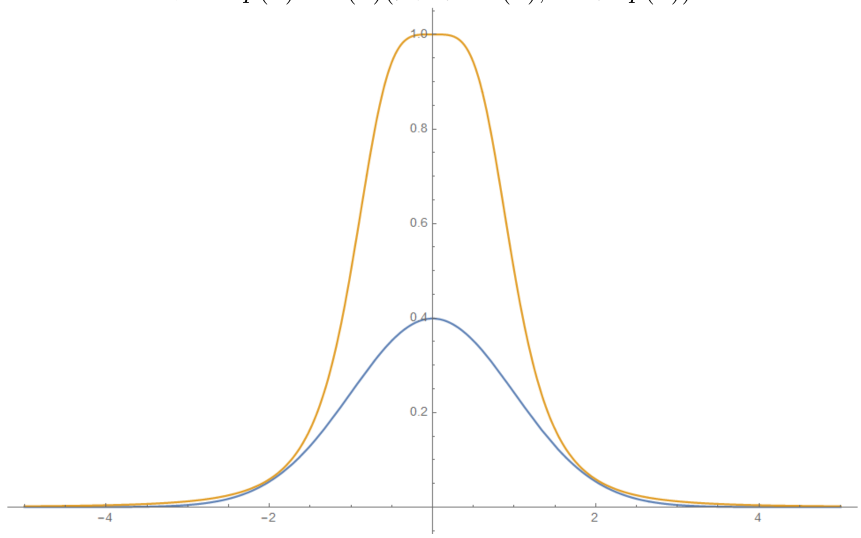
$$p(x) = 0.398943 e^{-\frac{x^2}{2}}$$

对$F(x) = \dfrac{C_2}{1+x^4}$，应取合适的$C_2$使得$F(x) > p(x), \forall x \in [-5, 5]$。不妨先取$C_2 = 1$，即

$$F(x) = \frac{1}{1+x^4}$$

下面验证$F(x)$满足条件。$Mathematica$绘制函数图像如下：

图 1: $p(x)$和$F(x)$(黄线是$F(x)$,蓝线是$p(x)$)



粗略可见$F(x) > p(x)$在区间上成立，但这样并不严谨。这里给出一个较为严谨的证明：

设 $G(x) = F(x) - p(x)$，则 $G'(x) = -\dfrac{4x^3}{(1+x^4)^2} - \dfrac{C_1 x^3}{2} e^{-\frac{x^2}{2}}$

$G'(x)$在$[-5, 5]$上有唯一零点$x_0 = 0$，且为极大值点；极小值点为边界点$x = \pm 5$：$G(x) \geqslant G(\pm 5) = 0.0016 > 0$

在比较函数$F(x)$的面积区内产生随机点(x_0, y_0)，由反函数法推出$\xi_x = \xi_x(\xi_1)$（此时ξ_x不是在$[a, b]$区间内均匀分布的，而是按权重$F(x)$的大小分布的）。

记$H(x) = \displaystyle\int_{-5}^{x} F(t)dt = \int_{-5}^{x} \frac{1}{1+t^4} dt$，解得

$$H(x) = \left.\frac{\left[2\arctan(1+\sqrt{2}t) - 2\arctan(1-\sqrt{2}t) + \ln(\frac{1+\sqrt{2}t+t^2}{1-\sqrt{2}t+t^2})\right]}{4\sqrt{2}}\right|_{-5}^{x}$$

$H(-5) = 0, H(5) = 1.10806 * 2 = 2.21612$

舍选法思路：

我们随机生成一组$\xi_1 \in [0, 1]$，$\xi_1 = \frac{H(x)}{H(5)} \Rightarrow$得到$x = H^{-1}(H(5)\xi_1)$，再随机生成一组$\xi_2 \in [0, 1]$，$y = \xi_2 F(x)$，判断这组$y$是否在$p(x)$面积区中，即，若$y < p(x)$成立，则取$x$点。

值得一提的是，由于$H(x)$解析式复杂，不易求反函数，这里采用精度为10^{-4}的二分法求解。

三　计算结果与分析

取$N = 10^6$，运行程序，过程如上算法分析所示，通过随机数$\xi_1, \xi_2 \in [0,1]$，最终得到一个数组$\{X_i\}$，导入$Origin$中，频数分布直方图如下：

图 2: $p(x,y)$

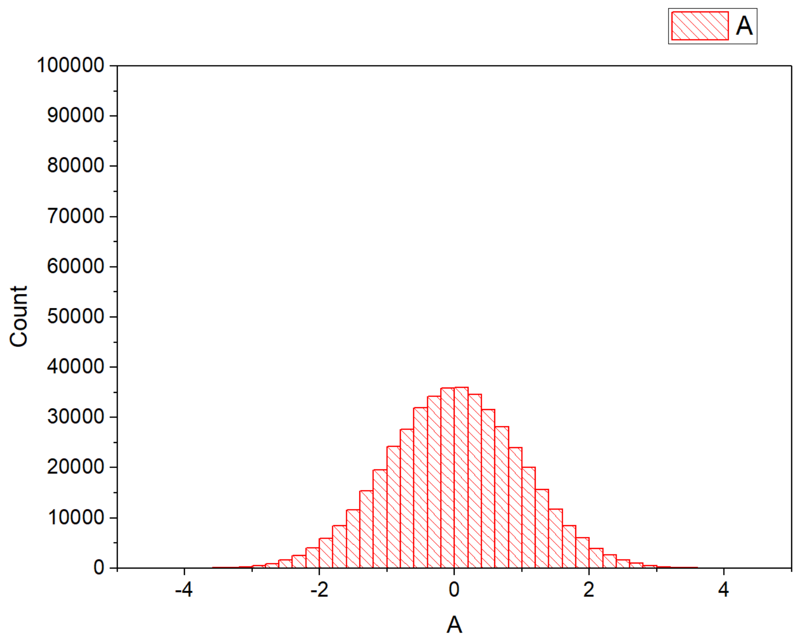


图2频数分布直方图和图1中的理论$p(x)$曲线对比，基本吻合，差异很小，说明舍选法的结果良好。

下面讨论抽样效率：

(1).运行的程序中，用num存储计数满足$y < p(x)$的点的个数，可以得到效率$num/N \approx 45.16\%$

理论上，在$F(x) = \dfrac{1}{1+x^4}$的情况下，该抽样的效率为：

$$\frac{\displaystyle\int_{-5}^{5} 0.398943 e^{\frac{-x^2}{2}} dx}{\displaystyle\int_{-5}^{5} \frac{1}{1+x^4} dx} \approx 45.12\%$$

理论值与实验值相近，由此可见，算法符合预期。

(2).另外，比较$F(x)$舍选法和直接舍选法的效率：

直接舍选法，选取$M = \max(p(x)) = 0.398943$，该抽样的效率为：

$$\frac{\displaystyle\int_{-5}^{5} 0.398943 e^{\frac{-x^2}{2}} dx}{\displaystyle\int_{-5}^{5} 0.398943 dx} \approx 25.07\%$$

由此可见，直接舍选法的效率较低，引入$F(x)$代替$y = M$后抽样的效率大大增加。这一点是很直观的：我们注意到抽样效率即为$p(x)$和$F(x)$(或$y = M$)的面积区之比。

四　结论

对于有尖峰形状的$p(x)$，直接舍选法的抽样效率很低，为了改良这一弊端，我们引入$F(x)$代替$y = M$，生成按权重$F(x)$分布于区间$[a, b]$中的抽样ξ_x，得到了频率直方图，与函数$y = p(x)$图像相吻合，最后计算得到的抽样效率与理论值一致，说明算法是良好的。另外，本体给出的C_2并非最高抽样效率的参数，可利用相关软件找到满足$F(x) > p(x), \forall x \in [a, b]$的最小的$C_2$，此时抽样效率最佳。