06：球面上均匀分布点的随机抽样

许传奇 PB16021546

1 题目

在球坐标系(ρ,θ,ϕ)下产生球面上均匀分布的随机坐标点，给出其直接抽样方法。

2 原理与算法

2.1 原理

2.1.1 连续型变量的直接抽样法

设连续型变量x在区间$[a,b]$中取值，对应的概率密度函数为$p(x)$，则其对应的累计函数$\xi(x)$为：

$$\xi(x)=\int_a^x p(\epsilon)d\epsilon \tag{1}$$

累计函数$\xi(x)$单调不减且满足：

$$\xi(a)=0,\ \xi(b)=1 \tag{2}$$

因此，可以得到其反函数$x(\xi)=\xi^{-1}(\xi)$。

基于以上推导，直接抽样法的步骤为：

1. 在$[0,1]$上随机抽样，得到ξ的均匀分布;
2. 根据对应关系$x=x(\xi)$，计算出与ξ对应的x的值；
3. 不断重复以上步骤，得到的一系列的x就是区间$[a,b]$上以$p(x)$为密度函数的直接抽样。

2.1.2 球面上均匀分布点的直接抽样

球面上点对应的分布为：

$$\oiint_{sphere} p(x,y,z)dS=1 \tag{3}$$

其中$p(x,y,z)$为概率密度函数。

若为球面上的均匀分布，则由上式可得：

$$p(x,y,z)=\frac{1}{4\pi(x^2+y^2+z^2)} \tag{4}$$

将平面直角坐标系(x,y,z)变换变换到球坐标系(ρ,θ,ϕ)，变换公式如下：

$$\begin{cases} x = \rho sin\theta cos\phi \\ y = \rho sin\theta sin\phi \\ z = \rho cos\theta \end{cases} \tag{5}$$

则相应的概率分布函数变换为：

$$p(\rho,\theta,\phi) = p(x,y,z)\cdot|\frac{\partial(x,y,z)}{\partial(\rho,\theta,\phi)}| = \frac{sin\theta}{4\pi} \tag{6}$$

ρ仅能取球的半径R这一个值。因此，当球面确定时，概率分布函数仅有θ和ϕ两个变量，即有：

$$p(\theta,\phi) = \frac{sin\theta}{4\pi} \tag{7}$$

或者，由于投影关系$dS = dxdy/cos\theta$，也可以得到$p(\theta,\phi)$的表达式：

$$p(\theta,\phi)d\theta d\phi = p(x,y,z)\frac{dxdy}{cos\theta} = p(x,y,z)\cdot|\frac{\partial(x,y)}{\partial(\theta,\phi)}|\cdot\frac{d\theta d\phi}{cos\theta} = \frac{sin\theta}{4\pi}d\theta d\phi \tag{8}$$

对应的边缘分布为：

$$\begin{cases} p(\theta) = \int_0^{2\pi}\frac{sin\theta}{4\pi}d\phi = \frac{sin\theta}{2} \\ p(\phi) = \int_0^{\pi}\frac{sin\theta}{4\pi}d\theta = \frac{1}{2\pi} \end{cases} \tag{9}$$

θ、ϕ对应的累计函数为：

$$\begin{cases} \xi_1 = \int_0^{\theta}p(\theta)d\theta = \int_0^{\theta}\frac{sin\theta}{2}d\theta = \frac{1-cos\theta}{2} \\ \xi_2 = \int_0^{\phi}p(\phi)d\phi = \int_0^{\phi}\frac{1}{2\pi}d\phi = \frac{\phi}{2\pi} \end{cases} \tag{10}$$

对应的反变换为：

$$\begin{cases} cos\theta = 1-2\xi_1 \\ \phi = 2\pi\xi_2 \end{cases} \tag{11}$$

因此，在$[0,1]$上分别独立地生成均匀分布的随机数ξ_1、ξ_2，再利用(11)式，即可得到球面上均匀分布的点的$cos\theta$与ϕ的抽样。再将其带入式(5)中，即可计算出对应点的坐标。

2.2 算法

用之前题目中写到的16807随机数产生器分别产生两组随机数，产生器的种子值由随机函数随机生成,得到在$[0,1]$上分别独立地生成均匀分布的随机数ξ_1、ξ_2。

不失一般性，可以设$\rho = 1$。根据(5)式与(11)式，计算出对应点的x、y、z坐标，随后输出并画图。

3 源文件使用说明

编译并运行“06DirectSampling.cpp”，将弹出命令行，要求输入随机数的个数num。
输入随机数的个数后，程序运行并将数据输出到文件“num=输入的num.txt”中。
编译并运行“plot.py”即可绘制出三维散点图或XY平面投影图。

4 计算结果及具体分析

先生成点的个数为1000的情况，如下图所示：

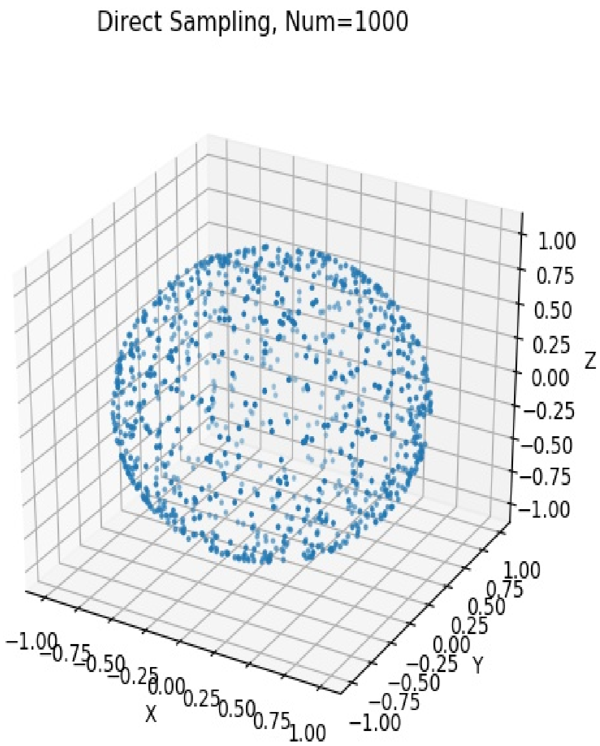

图 1: Num=1000时的点分布

再把点的个数设为10000，如下图所示：

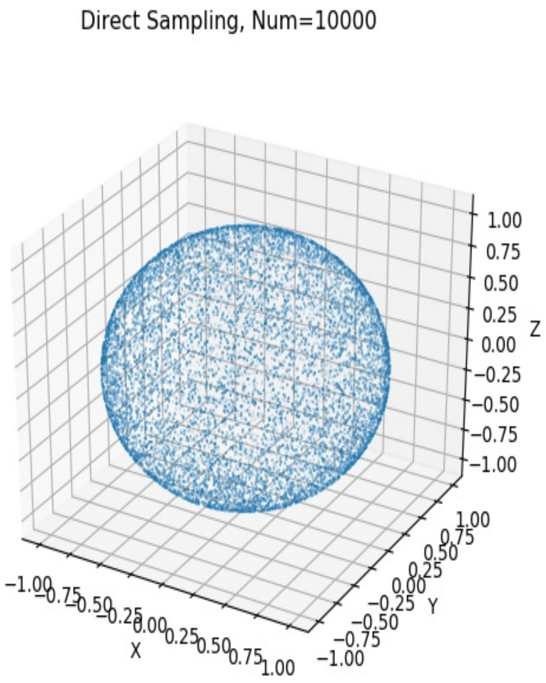

图 2: Num=10000时的点分布

随后换一个观察方向，得到下图：

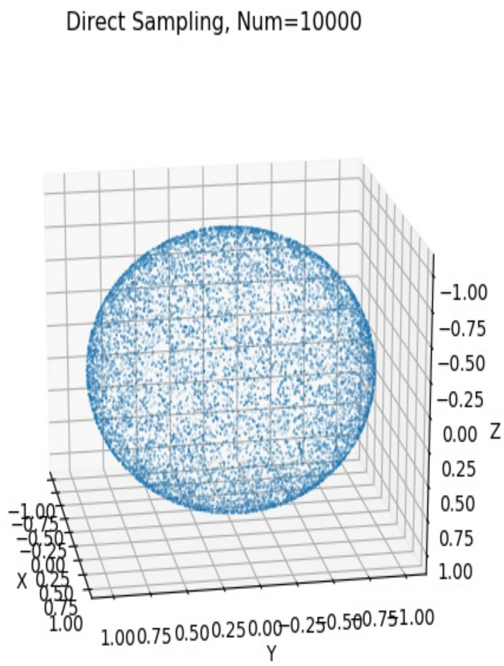

图 3: Num=10000时的点分布(另一观察方向)

生成该情况下在XY平面上的投影图：

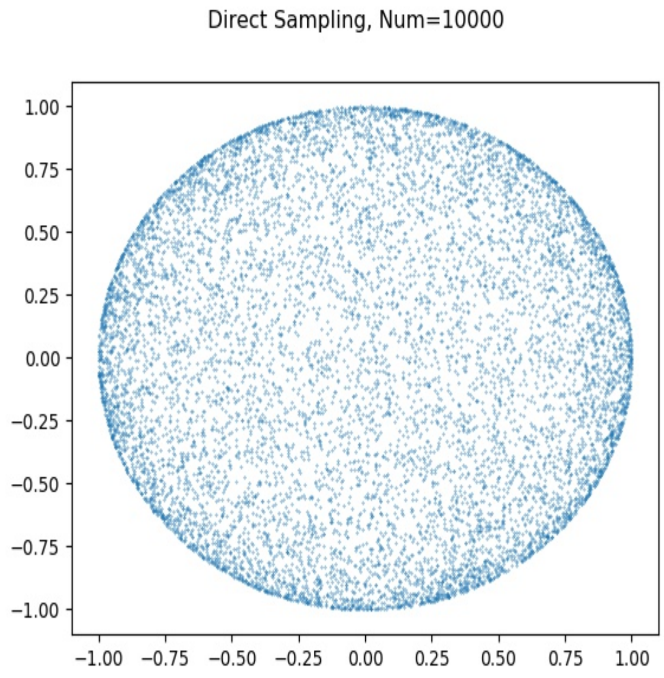

图 4: Num=10000时的点在XY平面上的投影

生成点的个数为100000时的点在XY平面上的投影图：

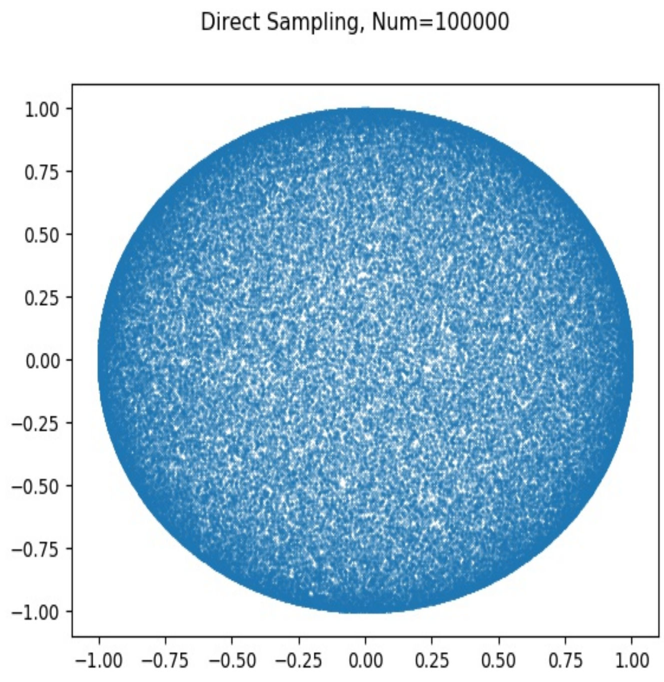

图 5: Num=100000时的点在XY平面上的投影

5 讨论

5.1 原理的讨论

本题原理较简单，仅需按照原理部分的推导，根据公式，由均匀分布的两个随机变量计算对应点的x、y、z坐标即可。

5.2 算法的改进

本题随机抽样用到之前所写的16807随机数产生器，且起始点由随机函数随机生成。

本次实验中，两个随机变量的抽样没使用固定的起始值，也没有采用一个起始值生成数组并将其均分再对应到两个抽样。这样做的目的是为了减小偶然误差。其次如果平分的话，可能会由于16807随机数产生器产生的n个连续的点分布在(n-1)维超平面上的原因导致点的分布有关联性。

我们运用直接抽样法抽取平面上均匀分布的点，在抽样和计算x、y、z坐标时，我们需要计算三角函数。但是，三角函数的计算耗时很多，一般不希望采用这样的抽样方式。因此我们可以通过其他的方法来提高效率，对于三维球面的情形，我们采用的就是Marsaglia抽样方法，即第七题所用到的抽样方法。

5.3 结果的讨论

在点数较少的时候，如图1所示，看得并不是很清楚，但也可以看出其均匀性较好。

在点数较多的时候，如图2所示，可以看出其确实分布比较均匀。通过旋转观察的方向，如图3所示，可以看出其图样并没有变化很大，也从侧面证明了分布的均匀性。

若点是在球面上均匀分布的，求得球面上均匀分布的点在XY平面上投影的几率分布函数为：

$$p(x,y) = \frac{1}{2\pi\sqrt{1-(x^2+y^2)}},\ x^2+y^2 \leq 1 \tag{12}$$

可知，在XY平面上的几率分布函数从圆心到圆周逐渐增大，因此点的分布从圆心到圆周逐渐密集，如图4与图5所示。图4点数较少，可能不明显，图5点数更多，所以这种关系更加明显。这也从侧面证明了分布的均匀性。