07：Marsaglia抽样方法

许传奇 PB16021546

1 题目

对于球面上均匀分布的随机坐标点，给出它们在(x,y)平面上投影的几率分布函数。并由此验证Marsaglia抽样方法$x = 2u\sqrt{1-r^2}, y = 2v\sqrt{1-r^2}, z = 1-2r^2$确为球面上均匀分布的随机抽样。

2 原理与算法

2.1 原理

2.1.1 球面上均匀分布的点在XY平面上投影的几率分布函数

在第6题中，我们已经推导过球面上均匀分布的相关等式，如下：

球面上点对应的分布为：

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

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：

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

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：

$$\begin{cases} x = \rho sin\theta cos\phi \\ y = \rho sin\theta sin\phi \\ z = \rho cos\theta \end{cases} \tag{3}$$

为了求在XY平面上投影的几率分布函数，我们利用面积的投影关系：

$$dS = \frac{1}{|cos\theta|}dxdy = \frac{1}{\frac{|z|}{\sqrt{x^2+y^2+z^2}}}dxdy = \frac{\sqrt{x^2+y^2+z^2}}{|z|}dxdy \tag{4}$$

得到XY平面上的投影关系：

$$p'(x,y,z) = \frac{1}{4\pi|z|\sqrt{x^2+y^2+z^2}} \tag{5}$$

由于在半径为R的球面上，x、y、z有约束关系：

$$x^2 + y^2 + z^2 = R^2 \tag{6}$$

利用上式消去z，得到球面上均匀分布的点在XY平面上投影的几率分布函数p(x,y)：

$$p(x,y) = \frac{1}{2\pi R\sqrt{R^2 - (x^2+y^2)}},\ x^2+y^2 \le R^2 \tag{7}$$

此处需要注意的是，(5)式中是对全部球面的投影，由于式中分母上有z的绝对值，因此投影在XY平面上时，几率密度函数应当是两倍，对应于$(x, y, \pm z)$处的投影，从而保证了几率密度函数的归一性，即：

$$\iint_{circle} p(x,y)dxdy = \int_0^{2\pi} d\theta \int_0^R \frac{1}{2\pi R\sqrt{R^2 - r^2}} \cdot rdr = 1 \tag{8}$$

不失一般性，在本次作业中设置$R = 1$，此时的几率分布函数为：

$$p(x,y) = \frac{1}{2\pi\sqrt{1 - (x^2+y^2)}},\ x^2+y^2 \le 1 \tag{9}$$

2.1.2 Marsaglia抽样方法

在第六题中，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。

Marsaglia抽样方法的步骤为：

1. 随机抽取一对均匀分布的随机数(u,v)∈[-1,1]；
2. 计算$r^2 = u^2 + v^2$，如果$r^2 > 1$则重新抽样，直至$r^2 \le 1$；
3. 通过$x = 2u\sqrt{1-r^2}, y = 2v\sqrt{1-r^2}, z = 1 - 2r^2$计算点的x、y、z坐标，得到的一系列点即为三维球面上均匀分布的随机抽样。

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$[-1, 1]$上分别独立地生成使用者输入的随机数个数的均匀分布的随机数u、v。

当由于舍选导致生成的随机点个数小于需要的随机点个数时，程序重新生成随机数，并重复上述步骤，直到生成所需的个数的随机数为止。

再根据上述的Marsaglia抽样方法，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“07Marsaglia.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

4.1 Marsaglia方法生成球面上的均匀随机抽样

先生成点的个数为1000的情况，如下图所示：

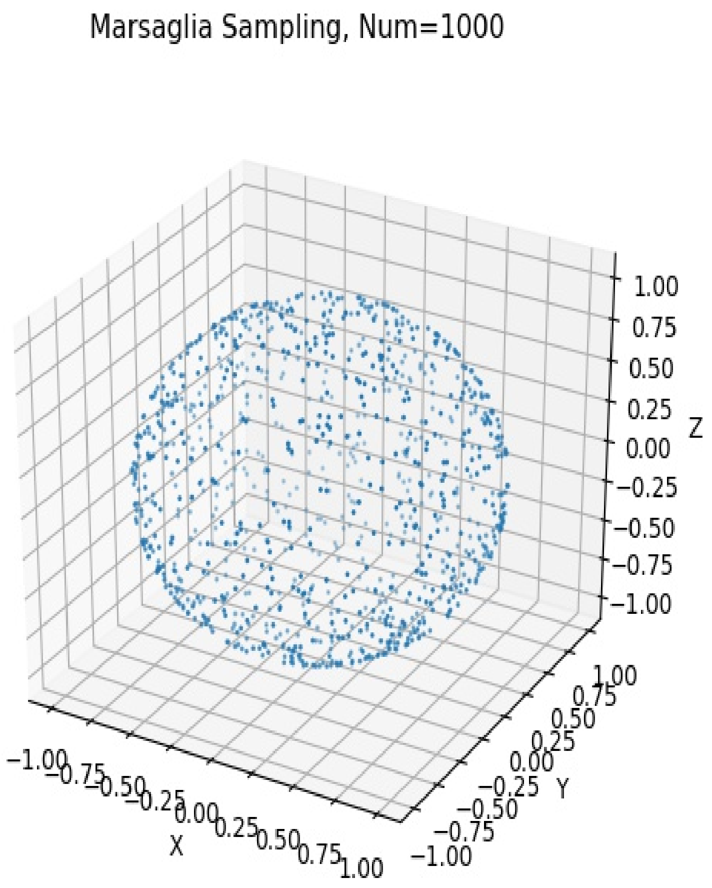

图 1: Num=1000时的点分布

再把点的个数设为10000，如下图所示：

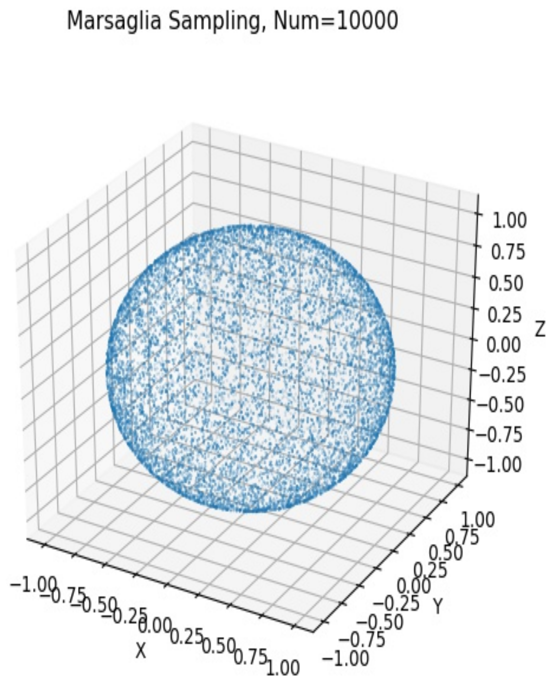

图 2: Num=10000时的点分布

随后换一个观察方向，得到下图：

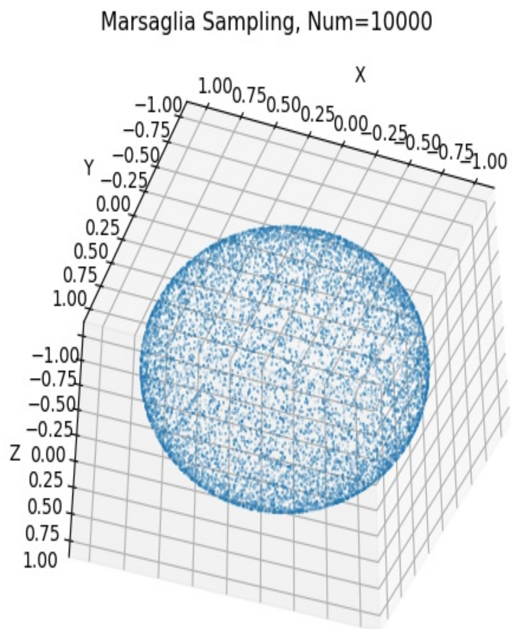

图 3: Num=10000时的点分布(另一观察方向)

4.2 Marsaglia方法在XY平面上的投影

生成该情况下在XY平面上的投影图：

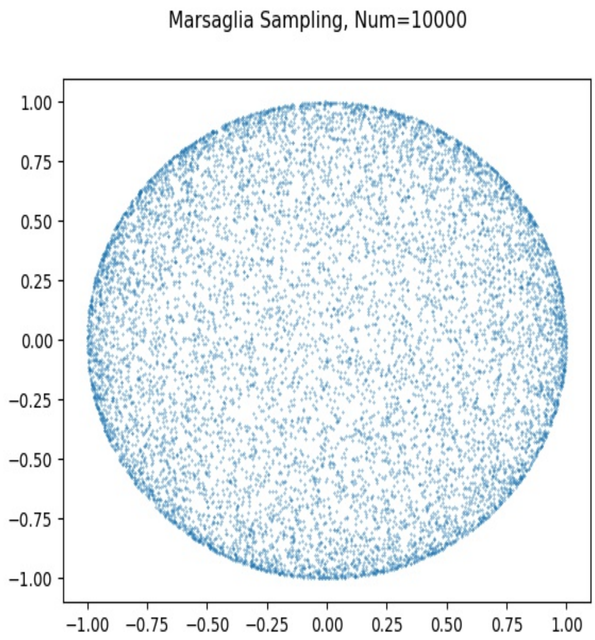

图 4: Num=10000时的点在XY平面上的投影

生成点的个数为100000时的点在XY平面上的投影图：

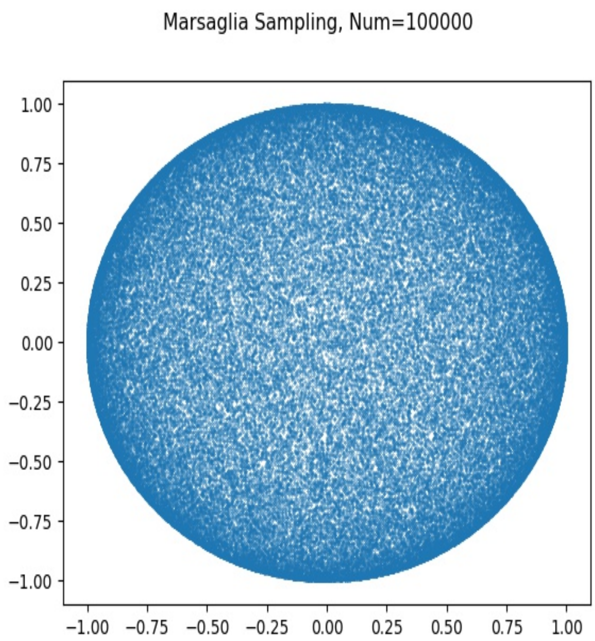

图 5: Num=100000时的点在XY平面上的投影

放大其边缘，如下图所示：

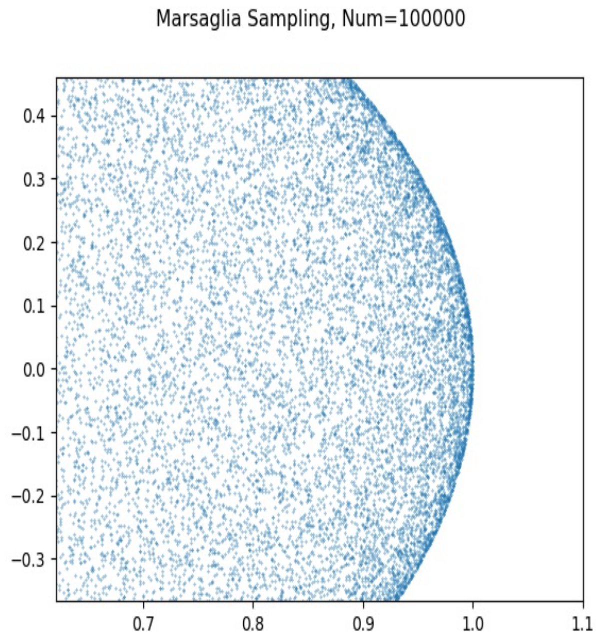

图 6: Num=100000时的点在XY平面上的投影的边缘部分

生成点的个数为1000000时的点在XY平面上的投影图：

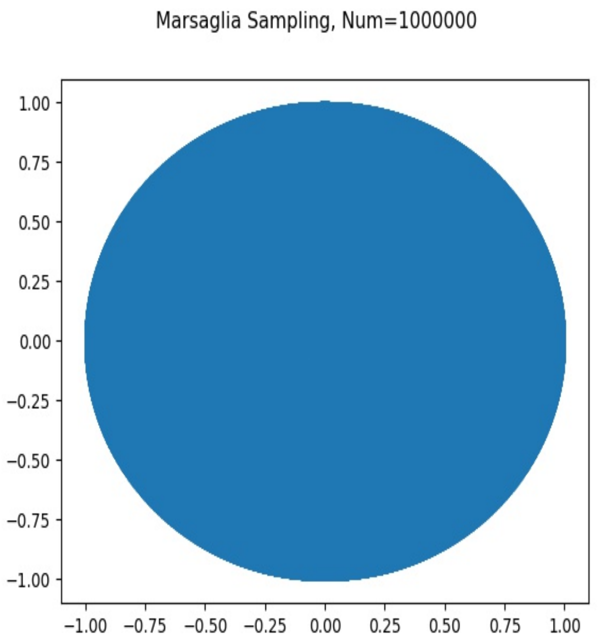

图 7: Num=1000000时的点在XY平面上的投影

放大其边缘，如下图所示：

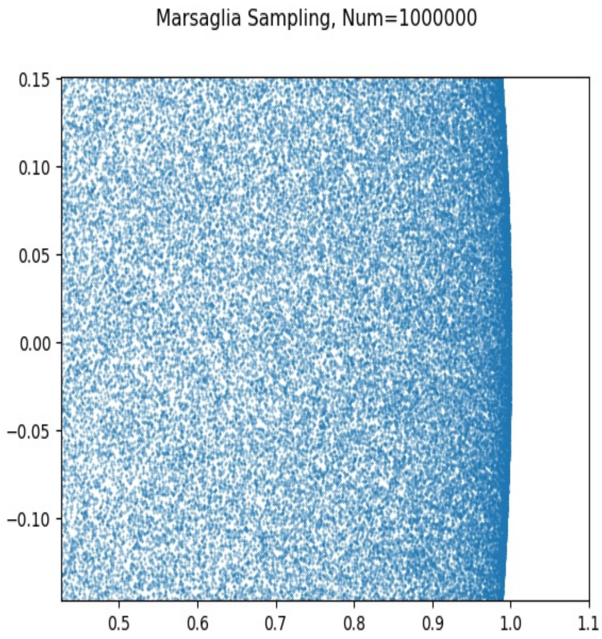

图 8: Num=1000000时的点在XY平面上的投影的边缘部分

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

由于本人的计算机知识并不是很丰富，下面的讨论可能会欠妥，源代码的编写也肯定会有很多需要改进的地方，希望以后学习中能够不断完善。

与第六题类似，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。

另外，在程序中需要注意一点，由于Marsaglia方法有舍选的步骤，因此可能会导致生成的随机数用完但还没达到所需的点数，此时需要再重新生成随机数，再重复Marsaglia方法，直至生成的随机点数达到所需的数目。

5.3 结果的讨论

1. 通过图片直观观察均匀性：

   与第六题类似，我们通过图片来直接观察Marsaglia方法是否是球面上均匀分布点的抽样。

   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。

2. 通过XY平面上的投影来观察均匀性：

   若点是在球面上均匀分布的，根据我们在原理部分求得的球面上均匀分布的点在XY平面上投影的几率分布函数：

$$p(x,y) = \frac{1}{2\pi\sqrt{1-(x^2+y^2)}},\ x^2+y^2 \le 1 \tag{10}$$

   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。通过图6和图8的局部放大图，能够更清晰地展现这种关系。这些都从侧面证明了分布的均匀性。

   下面推导Marsaglia方法抽样的点在XY上的投影确实是球面上均匀分布的点在XY平面上的投影。

   由于Marsaglia方法要求$u^2+v^2 \le 1$，所以u和v的区域为单位圆，u和v在此单位圆上均匀分布，因此其分布密度函数为：

$$p(u,v) = \frac{1}{\pi},\ u^2+v^2 \le 1 \tag{11}$$

   对于Marsaglia方法求得的x、y、z：

$$\begin{cases} x = & 2u\sqrt{1-r^2} \\ y = & 2v\sqrt{1-r^2} \\ z = & 1-2r^2 \end{cases} \tag{12}$$

   有如下关系：

$$x^2+y^2 = 4(u^2+v^2)[1-(u^2+v^2)] = 1-[1-2(u^2+v^2)]^2 \tag{13}$$

$$|\frac{\partial(x,y)}{\partial(u,v)}| = 4[1-2(u^2+v^2)] \tag{14}$$

   由式(13)和式(14)，可得：

$$p(x,y) = p(u,v)\cdot|\frac{\partial(x,y)}{\partial(u,v)}|^{-1} = \frac{1}{2\pi\sqrt{1-(x^2+y^2)}},\ x^2+y^2 \le 1 \tag{15}$$

   与之前一样，由于对式(13)开方会存在正负两种情况，因此需要将概率密度函数乘以2，对应于$u^2+v^2 > 1$和$u^2+v^2 < 1$的两个地方的投影，来满足密度函数的归一性。

   因为式(15)与式(10)完全一样，因此Marsaglia方法得到的在XY平面上的投影确实与球面上均匀分布的点在XY平面上的投影分布一样。

3. Marsaglia方法产生球面上均匀分布的点的一些不严格的证明方法：

   虽然我们证明了Marsaglia方法产生的点在XY平面上的投影与球面上均匀分布的点在XY平面上的投影分布一样，但并不能因此说明Marsaglia方法就产生了在球面上均匀分布的点。我想到了如下两种不严格地方法来证明Marsaglia方法确实产生的是球面上均匀分布的点：

   (a) 我们可以仿照上述步骤，求得Marsaglia方法抽样的点在YZ和XZ平面上分布的几率分布函数，可以发现这两个几率分布函数与球面上均匀分布的点在YZ和XZ平面上投影的几率分布函数一样。事实上，我们可以证得Marsaglia方法抽样的点在任意过球心的截面上投影的几率分布函数与球面上均匀分布的点在该面上投影的几率分布函数一样。这种任意性就证明了Marsaglia方法抽样确为球面上均匀分布的随机抽样。

   (b) 由于Marsaglia方法中$z = 1 - 2r^2 = 1 - 2(u^2 + v^2)$，我们可以将式(13)和式(14)写为如下形式：

$$x^2 + y^2 = 1 - z^2 \tag{16}$$

$$|\frac{\partial(x, y)}{\partial(u, v)}| = 4z \tag{17}$$

   因此，对应的XY平面上的投影密度函数为：

$$p'(x, y, z) = p(u, v) \cdot |\frac{\partial(x, y)}{\partial(u, v)}|^{-1} = \frac{1}{4\pi|z|} \tag{18}$$

   添加上额外变量z后，投影关系与我们在原理部分叙述的式(5)：$p'(x, y, z) = \frac{1}{4\pi|z|\sqrt{x^2+y^2+z^2}}$一样（两式中$\sqrt{x^2 + y^2 + z^2}$的差别是因为我们选取了$x^2 + y^2 + z^2 = 1$）。

   这也与之前的分析一致，z在[-1,1]上分布，球面上均匀分布的点在$|z|$越小处的分布概率密度越大，由式(18)可以分析出，几率分布函数在$|z|$越小的点越大，对应于XY平面上的投影边缘的点更加密集。

   因此，即便增加了了z坐标作为几率分布函数的新变量，对应的投影关系与球面上均匀分布的点在XY平面上的投影分布仍然一样，这也不严格地证明了Marsaglia方法确实产生了球面上的均匀分布的随机抽样。