Homework 2

孙锴

May 24, 2012

练习(2.36)**.** 本题我采用*C*语言生成随机数据并用*Excel*绘制直方图。由于*C*语言并没有可以直接生成正态分布随机数的方法，通过查阅若干资料，我采用*Box-Muller*法生成了正态分布随机数，并用2.5节的方法生成了n维球面上的随机点。所绘制的直方图如下：

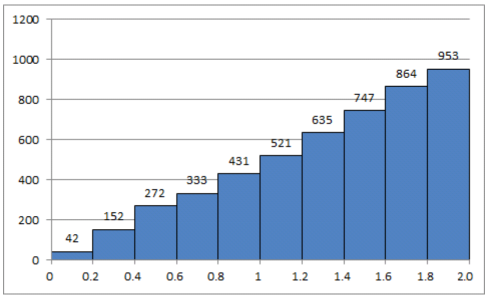

Figure 1: 3维球面随机两点距离分布

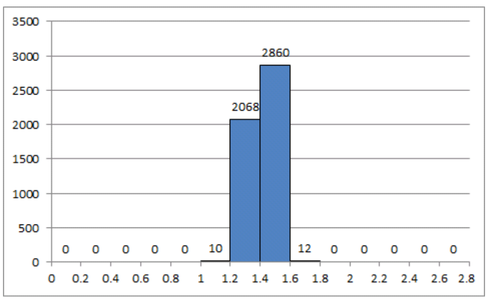

Figure 2: 100维球面随机两点距离分布

练习(2.37)**.** 原点及原点周围附近的将有非常大的概率密度，概率密度随着与原点距离的增加迅速下降，使得单位球面以外的区域的概率质量趋向

于*0*。
设空间维数为d，则要满足单位球面外的区域的概率质量为*0*，则必有$\int_{S^d} d\Omega \int_0^1 r^{d-1}p(r)dr = 1$
$\therefore \int_0^1 r^{d-1}e^{-\frac{r^2}{2\sigma^2}}dr = \frac{(2\pi)^{\frac{d}{2}}\sigma^d}{A(d)}$
$\therefore \int_0^1 r^{d-1}e^{-\frac{r^2}{2\sigma^2}}dr = \Gamma(\frac{d}{2})2^{\frac{d}{2}-1}\sigma^d$
用$t = \frac{r^2}{2\sigma^2}$换元，得到$\int_0^{\frac{1}{2\sigma^2}}(2t\sigma^2)^{\frac{d-1}{2}}e^{-t}2^{-\frac{1}{2}}\sigma t^{-\frac{1}{2}}dt = \Gamma(\frac{d}{2})2^{\frac{d}{2}-1}\sigma^d$
$\therefore \int_0^{\frac{1}{2\sigma^2}} t^{\frac{d}{2}-1}e^{-t}dt = \Gamma(\frac{d}{2})$
考虑等式左边，设$f(x) = x^{\frac{d}{2}-1}e^{-x} = x^a e^{-x}$，其中$a = \frac{d}{2}-1$，则$f'(x) = \frac{ax^{a-1}-x^a}{e^x}$，令$f'(x) = 0$，得$x = 0, a$，从而得到$f(x)$在$[0, a]$单调递增，在$[a, +\infty]$单调递减。不难看出$f(x)$在$[a, +\infty]$上减小速度非常快，在点$2a$处的值与在点$a$值的比$\frac{f(2a)}{f(a)} = (\frac{2}{e})^{\frac{d}{2}-1}$，更进一步，在点$ca$处*(*$c > 1$且为一个常数*)*，有$\frac{f(ca)}{f(a)} = (\frac{c}{e^{c-1}})^{(\frac{d}{2}-1)}$，因此在$d$极大且$\sigma$极小时，$\frac{f(ca)}{f(a)} \to 0$，从而$\int_0^{\frac{1}{2\sigma^2}} t^{\frac{d}{2}-1}e^{-t}dt = \int_0^{c(\frac{d}{2}-1)} t^{\frac{d}{2}-1}e^{-t}dt$，即要满足单位球面外的区域的概率质量趋向于*0*，只须令$\frac{1}{2\sigma^2} = ca$，得$\sigma = \frac{1}{\sqrt{c(d-2)}}$，从而得存在$\sigma$满足$\sigma = \Theta(\sqrt{\frac{1}{d}})$，又因为$c = 1$时不成立，所以显然有$\sigma = O(\sqrt{\frac{1}{d}})$，所以得到最终结论：
存在满足$\sigma = \Theta(\sqrt{\frac{1}{d}})$的满足本题条件的$\sigma$，且任意满足题目条件的$\sigma$都必须有$\sigma \le \sqrt{\frac{1}{d-2}}$

练习(2.50). *1.*许多实际问题可以用高维向量抽象表示、研究和解决，高维空间中的许多性质与三维空间不同。
*2.*计算*n*维球体积与表面积的方法，体积与表面积的关系。
*3.*当*n*趋向无穷时，球体积趋向于*0*且绝大多数点聚集于赤道表侧的窄圆环，表面积同样主要分布在赤道。
*4.*生成*n*维球面上的随机点的方法，高维高斯分布。
*5.*随机投影和*Johnson-Lindenstrauss*定理。