MATLAB统计分析与应用

——概率分布与随机数

主讲人：谢中华

ssstudy.com

科学软件学习网

主要内容

- 概率密度、分布和逆概率分布函数值的计算
- 生成一元分布随机数
- 生成多元分布随机数
- 蒙特卡罗模拟及其应用

第一节 概率分布计算

MATLAB统计工具箱中有这样一系列函数，函数名以pdf三个字符结尾的函数用来计算常见连续分布的密度函数值或离散分布的概率函数值，函数名以cdf三个字符结尾的函数用来计算常见分布的分布函数值，函数名以inv三个字符结尾的函数用来计算常见分布的逆概率分布函数值

【例7.1-1】求均值为1.2345，标准差（方差的算术平方根）为6的正态分布在 处的密度函数值与分布函数值。

```
>> x = 0:10;    %产生一个向量
>> Y = normpdf(x, 1.2345, 6)    %求密度函数值
>> P = normcdf(x, 1.2345, 6)    %求分布函数值
```

【例 7.1-2】求标准正态分布、t 分布、χ^2 分布和 F 分布的上侧分位数。

(1) 标准正态分布的上侧 0.05 分位数 $u_{0.05}$；

(2) 自由度为 50 的 t 分布的上侧 0.05 分位数 $t_{0.05}(50)$；

(3) 自由度为 8 的 χ^2 分布的上侧 0.025 分位数 $\chi^2_{0.025}(8)$；

(4) 第一自由度为 7，第二自由度为 13 的 F 分布的上侧 0.01 分位数 $F_{0.01}(7, 13)$；

(5) 第一自由度为 13，第二自由度为 7 的 F 分布的上侧 0.99 分位数 $F_{0.99}(13, 7)$.

```
>> u = norminv(1-0.05, 0, 1)

>> t = tinv(1-0.05, 50)

>> chi2 = chi2inv(1-0.025, 8)

>> f1 = finv(1-0.01, 7, 13)

>> f2 = finv(1-0.99, 13, 7)
```

第二节　生成一元分布随机数

一、均匀分布和标准正态分布随机数

1. rand函数

调用格式：

```
Y = rand
Y = rand(n)
Y = rand(m,n)
Y = rand([m n])
Y = rand(m,n,p, ...)
Y = rand([m n p ...])
Y = rand(size(A))
```

2. randn函数

调用格式：

与rand函数类似

【例7.2-1】调用rand函数生成10×10的随机数矩阵，并将矩阵按列拉长，然后调用hist函数画出频数直方图。

```
>> x = rand(10)
>> y = x(:);
>> hist(y)
>> xlabel('[0,1]上均匀分布随机数');
>> ylabel('频数');
```

【例7.2-1续】设置随机数生成器的算法为

Mersenne Twister算法，生成均匀分布随机数矩阵

```
>> rand('twister',1);
>> x1 = rand(2,6)
```

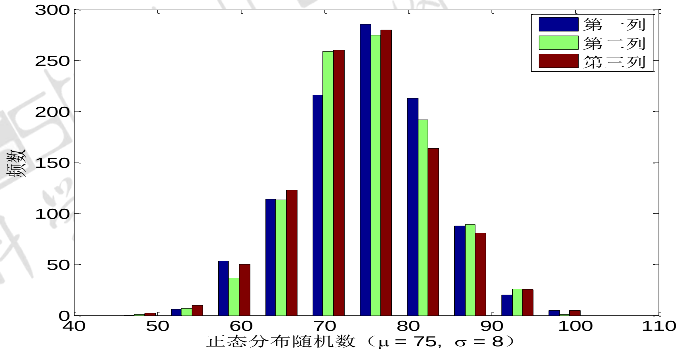
二、常见一元分布随机数

MATLAB统计工具箱中函数名以rnd三个字符结尾的函数用来生成常见分布的随机数。例如：

betarnd	Beta分布
exprnd	指数分布
gamrnd	Gamma分布
lognrnd	对数正态分布
normrnd	正态分布
poissrnd	泊松分布
randsample	从有限总体中随机抽样
random	指定分布

【例7.2-2】调用normrnd函数生成1000×3的正态分布随机数矩阵，其中均值 为75，标准差为8 ，并作出各列的频数直方图

```
>> x = normrnd(75, 8, 1000, 3);
>> hist(x)
>> xlabel('正态分布随机数（\mu = 75, \sigma = 8）');
>> ylabel('频数');
>> legend('第一列', '第二列', '第三列')
```

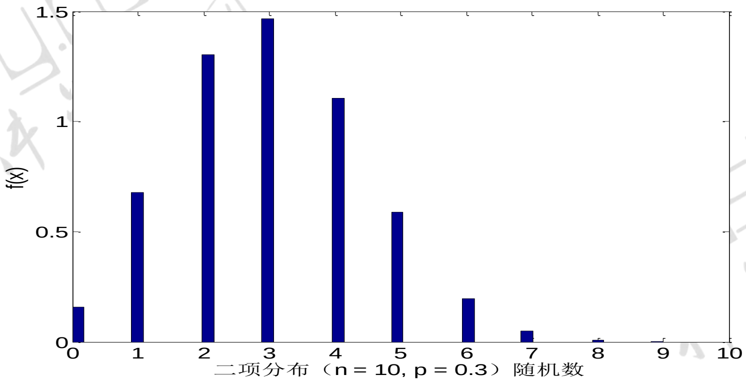


【例7.2-3】调用normrnd函数生成1000×3的正态分布随机数矩阵，其中第各列均值分别为0，15，40，标准差分别为1，2，3，并作出各列的频数直方图

```
>> x = normrnd(repmat([0 15 40], 1000, 1), ...
repmat([1 2 3], 1000, 1), 1000, 3);
>> hist(x, 50)
>> xlabel('正态分布随机数');
>> ylabel('频数');
>> legend('\mu = 0,  \sigma = 1','\mu = 15,
\sigma = 2',...
'\mu = 40,  \sigma = 3')
```

【例7.2-4】调用random函数生成10000×1的二项分布随机数向量，然后作出频率直方图。其中二项分布的参数为 n=10，p=0.3

```
>> x = random('bino', 10, 0.3, 10000, 1);
>> [fp, xp] = ecdf(x);
>> ecdfhist(fp, xp, 50);
>> xlabel('二项分布（n = 10, p = 0.3）随机数');
>> ylabel('f(x)');
```



【例7.2-5】调用random函数生成10000×1的卡方分布随机数向量，然后作出频率直方图，并与自由度为10的卡方分布的密度函数曲线作比较。其中卡方分布的参数（自由度）为 10。

```
>> x = random('chi2', 10, 10000, 1);
>> [fp, xp] = ecdf(x);
>> ecdfhist(fp, xp, 50);
>> hold on
>> t = linspace(0, max(x), 100);
>> y = chi2pdf(t, 10);
>> plot(t, y, 'r', 'linewidth', 3)
>> xlabel('x  ( \chi^2(10) )');
>> ylabel('f(x)');
>> legend('频率直方图', '密度函数曲线')
```

三、指定离散分布（一元）随机数

1. 一元离散分布的分布律

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_n \\ \hline p & p_1 & p_2 & \cdots & p_n \end{array}.$$

2. 生成离散分布随机数的MATLAB函数

MATLAB统计工具箱中的randsample函数

MATLAB通讯系统工具箱中的randsrc函数

【例 7.2-6】 设离散总体 X 的分布列为

X	-2	-1	0	1	2
p	0.05	0.2	0.5	0.2	0.05

分别调用 **randsample** 和 **randsrc** 函数生成 **100** 个服从该分布的随机数。

```
>> xvalue = [-2 -1 0 1 2];
>> xp = [0.05 0.2 0.5 0.2 0.05];
>> x = randsample(xvalue, 100, true, xp);
>> reshape(x,[10 10])
>> y = randsrc(10,10,[xvalue;xp])
```

【例 7.2-7】 设离散总体 X 的分布列为

X	A	B	C	D	E
p	0.3	0.2	0.25	0.2	0.05

调用 **randsample** 函数生成 **100** 个服从该分布的随机字母。

```
>> xvalue = ['ABCDE'];
>> xp = [0.3 0.2 0.25 0.2 0.05];
>> x = randsample(xvalue, 100, true, xp);
```

【例7.2-8】调用randi函数生成10×10的随机整数矩阵（取值范围为[0,10]），并调用tabulate函数统计各数字出现的频数和频率。

```
>> x = randi(10,[10,10])
>> tabulate(x(:))
```

四、指定连续分布（一元）随机数

1. 一元连续分布的密度函数

$$f(x), \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$

2. 生成连续分布随机数的MATLAB函数

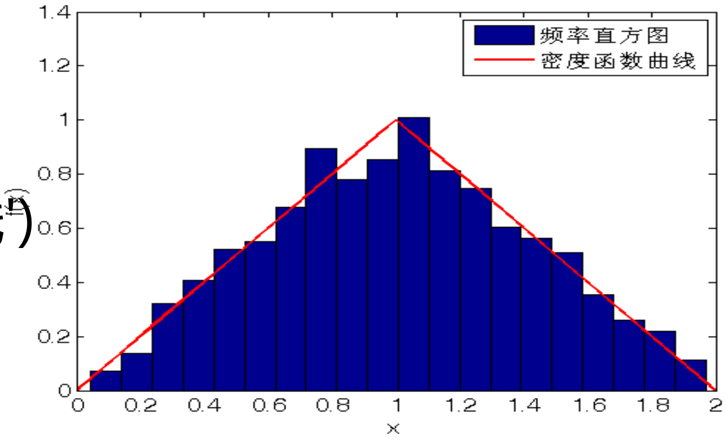
MATLAB统计工具箱中的slicesample函数

【例7.2-9】总体X服从三角分布，其概率密度函数如下：

$$f(x)=\begin{cases} x, & 0\le x<1 \\ 2-x, & 1\le x<2 \\ 0, & \text{其他} \end{cases}$$

下面调用slicesample函数生成1000个服从该分布的随机数，作出频率直方图，并与真实的密度函数曲线作比较。

```
>> pdffun = @(x)x*(x>=0 & x<1)+(2-x)*(x>=1 & x<2);
>> x = slicesample(1.5,1000,'pdf',pdffun);
>> [fp,xp] = ecdf(x);
>> ecdfhist(fp, xp, 20);
>> hold on
>> fplot(pdffun, [0 2], 'r')
>> xlabel('x');
>> ylabel('f(x)');
>> legend('频率直方图', '密度函数曲线')
```



五、一元混合分布随机数

1. 混合分布的密度函数

$$f(x)=\sum_{i=1}^{n}p_i f_i(x),\qquad \sum_{i=1}^{n}p_i=1$$

2. 生成混合分布随机数的MATLAB函数

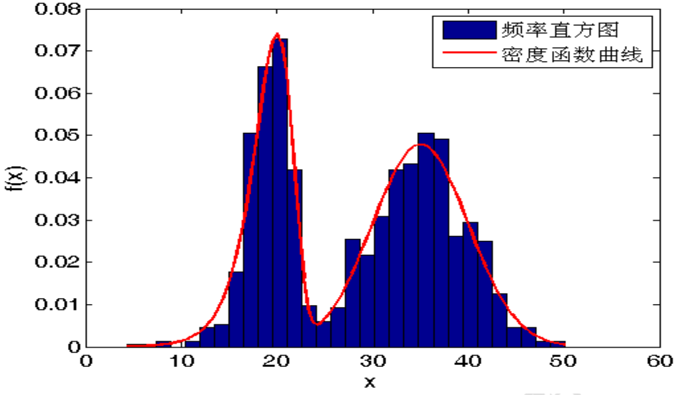
没有统一函数

【例 7.2-10】设随机变量 X 服从由正态分布和 I 型极小值分布（即 Gumbel 分布）混合而成的混合分布，其中正态分布的参数为 $\mu=35,\ \sigma=5$，Gumbel 分布的参数为 $\mu=20,\ \sigma=2$，两种分布的比例分别为 0.6 和 0.4。随机变量 X 的密度函数为：

$$f(x)=\frac{0.6}{5\sqrt{2\pi}}\exp\left(\frac{-(x-35)^2}{50}\right)+\frac{0.4}{2}\exp\left(\frac{x-20}{2}\right)\exp\left(-\exp\left(\frac{x-20}{2}\right)\right)$$

试生成 1000×1 的服从该混合分布的随机数向量，然后作出频率直方图，并与真实密度曲线作比较.

```
rand('seed',1);
randn('seed',1);
x = normrnd(35,5,1000,1);
y = evrnd(20,2,1000,1);
z = randsrc(1000,1,[1,2;0.6,0.4]);
data = x.*(z==1) + y.*(z==2);
pdffun = @(t,mu1,sig1,mu2,sig2)...
0.6*normpdf(t,mu1,sig1)+0.4*evpdf(t,mu2,sig2);
xd = linspace(min(data),max(data),100);
yd = pdffun(xd,35,5,20,2);
[fi,xi] = ecdf(data);
ecdfhist(fi,xi,30);
hold on;
plot(xd,yd,'r','linewidth',2);
xlabel('x');
ylabel('f(x)');
legend('频率直方图', '密度函数曲线');
```



第三节 生成多元分布随机数

生成多元分布随机数的MATLAB函数

函数名	分　　布
iwishrnd	逆 Wishart 分布
mnrnd	多项分布
mvnrnd	多元正态分布
mvtrnd	多元 t 分布
wishrnd	Wishart 分布

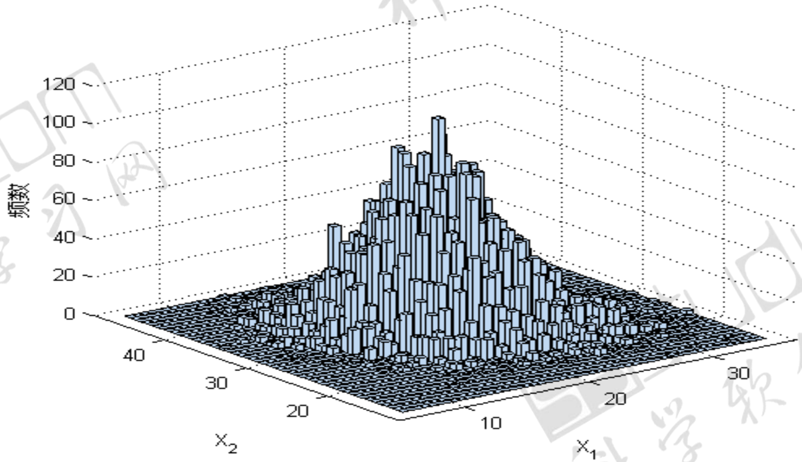
【例 7.3-1】若随机向量 $X=(X_1, X_2, \cdots, X_m)'$ 的分布列为

$$P(X_1=k_1, X_2=k_2, \cdots, X_m=k_m)=\frac{n!}{k_1!k_2!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m}.$$

其中 $0<p_i<1,\ i=1,2,\cdots,m,\ k_1+k_2+\cdots+k_m=n,\ p_1+p_2+\cdots+p_m=1$。

则称随机向量 $\boldsymbol{X}$ 服从参数为 n 和 $p=\left(p_1, p_2, \cdots, p_m\right)$ 的多项分布。调用 **mnrnd** 函数生成 **10000** 组 **3** 项分布随机数，其中 **3** 项分布的参数为 $n=100$， $p=(0.2, 0.3, 0.5)$。并利用 **hist3** 函数作出前两列的频数直方图。

```
>> n = 100;
>> p = [0.2  0.3  0.5];
>> r = mnrnd(n, p, 10000);
>> hist3(r(:,1:2),[50,50])
>> xlabel('X_1')
>> ylabel('X_2')
>> zlabel('频数')
```



【例 7.3-2】若随机向量 $\mathbf{X}=(X_1, X_2, \cdots, X_m)'$ 的密度函数为

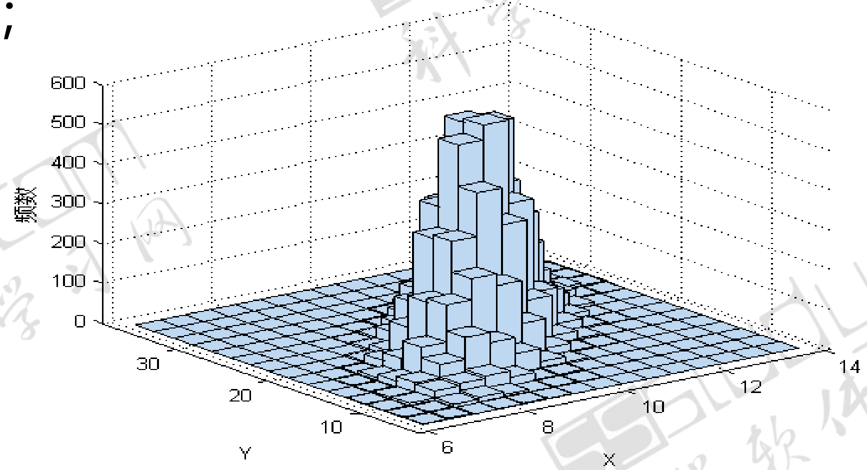
$$f(\mathbf{x})=(2\pi)^{-m/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),\quad \mathbf{x}\in\mathbb{R}^m$$

其中 $\mathbf{x}=(x_1, x_2, \cdots, x_m)'$，$\boldsymbol{\mu}=(\mu_1, \mu_2, \cdots, \mu_m)'$，$\boldsymbol{\Sigma}$ 为 m 阶正定矩阵。则称随机向量x服从参数为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的非退化 **m 元正态分布**，记为 $\mathbf{X}\sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$。$\boldsymbol{\mu}$ 为均值向量，$\boldsymbol{\Sigma}$ 为协方差矩阵。

利用 mvnrnd 函数生成 10000 组 2 元正态分布随机数，并利用 hist3 函数作出频数直方图。其中分布的参数为

$$\boldsymbol{\mu}=\begin{pmatrix}10\\20\end{pmatrix},\qquad \boldsymbol{\Sigma}=\begin{pmatrix}1 & 3\\3 & 16\end{pmatrix}。$$

```
>> mu = [10  20];
>> sigma = [1  3; 3  16];
>> xy = mvnrnd(mu, sigma, 10000);
>> hist3(xy, [15, 15]);
>> xlabel('X')
>> ylabel('Y')
>> zlabel('频数')
```



第四节 蒙特卡罗模拟及其应用

蒙特卡洛(Monte Carlo)方法，或称计算机随机模拟方法，是一种基于“随机数”的计算方法。这一方法源于美国在第二次世界大战期间研制原子弹的“曼哈顿计划”。该计划的主持人之一，数学家冯•诺伊曼用摩纳哥的驰名世界的赌城Monte Carlo来命名这种方法，因此称之为Monte Carlo方法。

一、用蒙特卡洛方法求圆周率

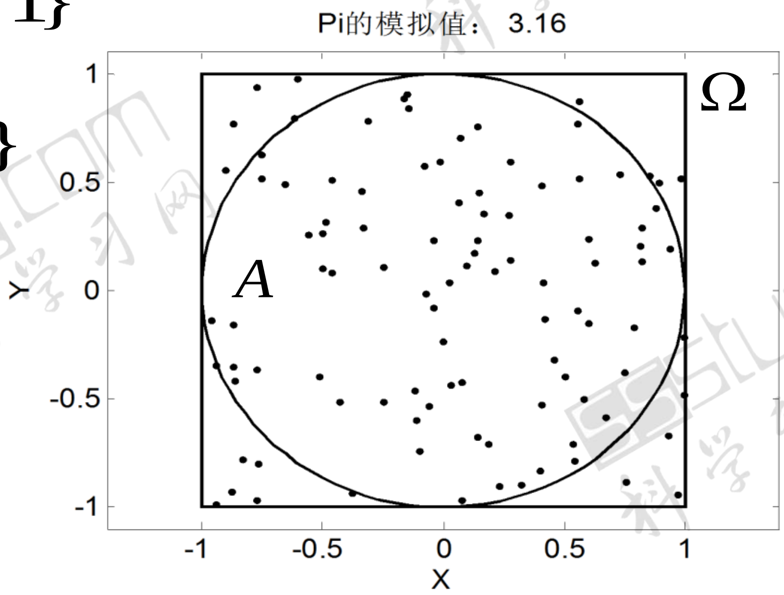
【例7.4-1】用随机投点法求圆周率 。

1. 随机投点法原理图解

$$\Omega = \{-1 \le x, y \le 1\}$$

$$A = \{x^2 + y^2 \le 1\}$$

$$\pi = 4P(A)$$



2. 程序实现

程序略…

二、用蒙特卡洛方法求积分

$$I = \int_{D^n} f(\mathbf{x})dx$$

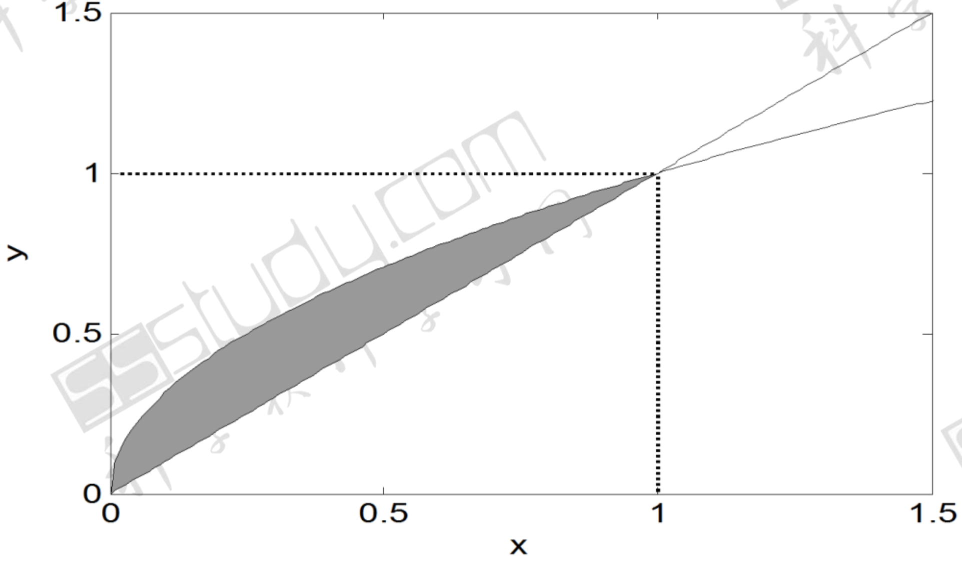
1. 原理

选取一个包含积分区域 D^n 的超立方体区域 C^n（其体积记为 M_C），在 C^n 内随机投入 N 个均匀分布的点 $\mathbf{x}_i$，则当 N 足够大时，

$$I = \int_{D^n} f(\mathbf{x})d\mathbf{x} \approx \frac{M_C}{N} \sum_{\mathbf{x}_i \in D^n} f(\mathbf{x}_i)$$

2. 定积分

【例 7.4-2】如图所示，求曲线 $y=\sqrt{x}$ 与直线 $y=x$ 所围成的阴影区域的面积。



```
function Sm = quad1mont2(n)
%   Sm = quad1mont2(n),求曲线 y = sqrt(x)与直线 y = x 所围成的阴影区域%   的面积的蒙特卡洛模拟值Sm.输入参数n是随机投点的个数，可以是正%   整数标量或向量.

fun = @(x)sqrt(x)-x;          % 定义被积函数
% 计算阴影区域的面积的蒙特卡洛模拟值
for i = 1:length(n)
    x = rand(n(i),1);         % 随机投点
    Sm(i) = mean(fun(x));     % 积分的模拟值
end
```

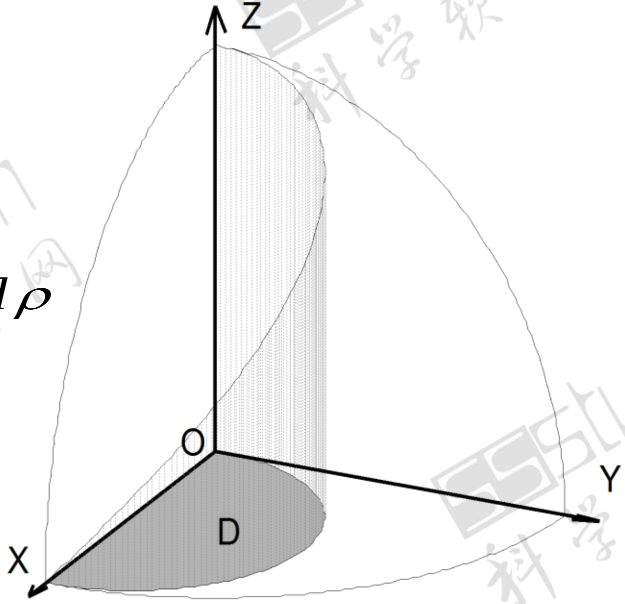
3. 二重积分

【例 7.4-3】求球体 $x^2+y^2+z^2\le 4$ 被圆柱面 $x^2+y^2=2x$ 所截得的（含在圆柱面内的部分）立体的体积。

$$V=4\iint_D \sqrt{4-\rho^2}\,\rho d\rho d\theta$$

$$=4\int_0^{\pi/2}d\theta\int_0^{2\cos\theta}\sqrt{4-\rho^2}\,\rho d\rho$$

$$=9.6440$$



```
function Vm = quad2mont2(n)
%   Vm = quad2mont(n),求球面 x^2+y^2+z^2 = 4 被圆柱面
x^2+y^2 = 2*x %   所截得的（含在圆柱面内的部分）立体的体积
的蒙特卡洛模拟值Vm. %   输入参数n是随机投点的个数，可以是
正整数标量或向量.
fun = @(x,y)sqrt(4-x.^2-y.^2);    % 定义被积函数
% 求体积的蒙特卡洛模拟值
for i = 1:length(n)
    x = 2*rand(n(i),1);
    y = 2*rand(n(i),1)-1;
    id = (x-1).^2 + y.^2 <= 1;
    Vm(i) = 8*sum(fun(x(id),y(id)))/n(i);
end
```

4. 多重积分

【例7.4-4】计算3重积分

$$\int_1^2\left(\int_x^{2x}\left(\int_{xy}^{2xy} xyz\ dz\right)dy\right)dx$$

```
function [V0,Vm] = quad3mont(n)
%   [V0,Vm] = quad3mont(n), 蒙特卡洛方法计算3重积分，返回理论值V0
%   和模拟值Vm. 输入参数n是随机投点的个数，可以是正整数标量或向量.

fun = @(x,y,z)x.*y.*z;
ymin = @(x)x; ymax = @(x)2*x; zmin = @(x,y)x.*y; zmax =
@(x,y)2*x.*y;
V0 = integral3(fun,1,2,ymin,ymax,zmin,zmax);
fun = @(x)prod(x);
for i = 1:length(n)
    x = unifrnd(1,2,1,n(i));
    y = unifrnd(1,4,1,n(i));
    z = unifrnd(1,16,1,n(i));
    X = [x;y;z];
    id = (y>=x)&(y<=2*x)&(z>=x.*y)&(z<=2*x.*y);
    Vm(i) = (4-1)*(16-1)*sum(fun(X(:,id)))/n(i);
end
```

Thank You