第7章 概率和数理统计

- **概率**：又称或然率、机会率、机率（几率）或可能性，是概率论的基本概念。概率是对随机事件发生的可能性的度量，一般以一个在0到1之间的实数表示一个事件发生的可能性大小。越接近1，该事件更可能发生；越接近0，则该事件更不可能发生。人们常说某人有百分之多少的把握能通过这次考试，某件事发生的可能性是多少，这都是概率的实例。

- **数理统计**是伴随着概率论的发展而发展起来的一个数学分支，研究如何有效的收集、整理和分析受随机因素影响的数据，并对所考虑的问题作出推断或预测，为采取某种决策和行动提供依据或建议.


分析
LARGER
DATA
云计算
长时段跟踪
BIG DATA
直观
BIG
可视化数据仓库
Cloud
ADTIME
处理
大数据
实时性计算
Calculation
可视化

- 在MATLAB中，提供了专门的统计工具箱Staticstics，该工具箱有几百个专用求解概率和数理统计问题的函数。

7.1 随机数的产生

随机数是专门的随机试验的结果。

- 在统计学的不同技术中需要使用随机数：
  - 在从统计总体中抽取有代表性的样本的时候
  - 在将实验动物分配到不同的试验组的过程中
  - 在进行蒙特卡罗模拟法计算的时候
  - ……
- 产生随机数有多种不同的方法。这些方法被称为**随机数发生器**。随机数最重要的特性是：<u>它所产生的后面的那个数与前面的那个数毫无关系</u>。

7.1 随机数的产生

真正的随机数是使用物理现象产生的：比如掷钱币、骰子、转轮、使用电子元件的噪音、核裂变等等。这样的随机数发生器叫做物理性随机数发生器，它们的缺点是技术要求比较高。

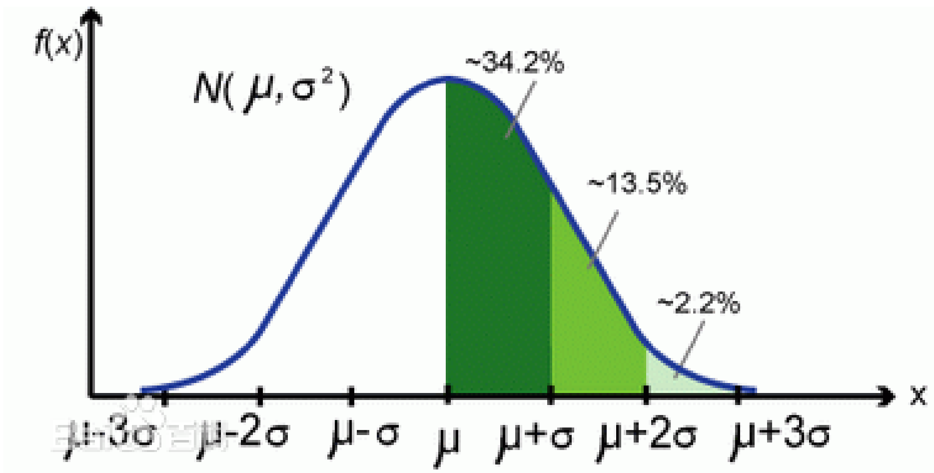
在实际应用中往往使用伪随机数就足够了。这些数列是“似乎”随机的数，实际上它们是通过一个固定的、可以重复的计算方法产生的。计算机或计算器产生的随机数有很长的周期性。它们不真正地随机，因为它们实际上是可以计算出来的，但是它们具有类似于随机数的统计特征。这样的发生器叫做伪随机数发生器。

在真正关键性的应用中，比如在密码学中，人们一般使用真正的随机数。

7.1 随机数的产生

- 二项分布的随机数据的产生
- 正态分布的随机数据的产生
- 常见分布的随机数产生
- 通用函数求各分布的随机数据

正态分布（Normal distribution）又名高斯分布（Gaussian distribution），是一个在数学、物理及工程等领域都非常重要的概率分布，在统计学的许多方面有着重大的影响力。若随机变量X服从一个数学期望为μ、方差为σ^2的高斯分布，记为N(μ，σ^2)。其概率密度函数为正态分布的期望值μ决定了其位置，其标准差σ决定了分布的幅度。因其曲线呈钟形，因此人们又经常称之为钟形曲线。我们通常所说的标准正态分布是$\mu = 0, \sigma = 1$的正态分布。



➢ 正态分布的随机数据的产生

- 命令 参数为μ、σ的正态分布的随机数据
- 函数 **normrnd**
- 格式
  - ➢ **R = normrnd(MU, SIGMA)**
    - 返回均值为**MU**，标准差为**SIGMA**的正态分布的随机数据，**R**可以是向量或矩阵。
  - ➢ **R = normrnd(MU, SIGMA, m)**
    - **m**指定随机数的个数，与**R**同维数。
  - ➢ **R = normrnd(MU, SIGMA, m, n)**
    - **m,n**分别表示**R**的行数和列数

```
clear all;
r=normrnd(0, 1)
R1=normrnd(0, 1, [3 5])   %产生一个3*5的矩阵
R2=normrnd(2, 4, [3 5])
```

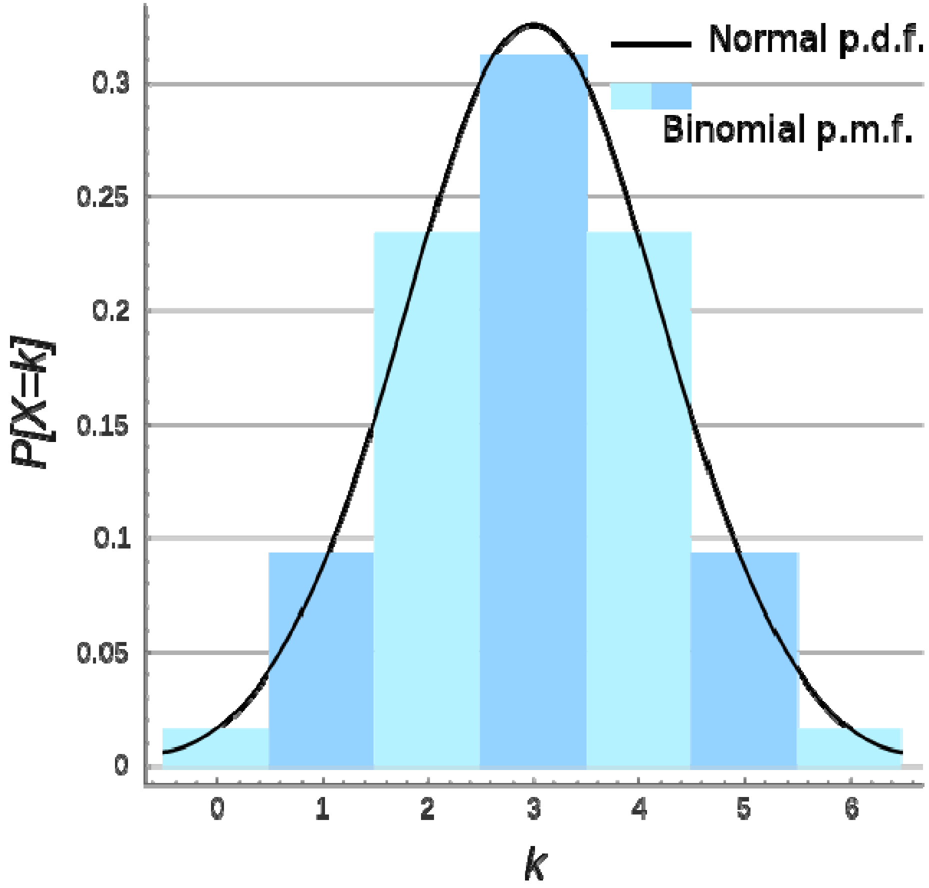
```
r =    0.5377

R1 =   1.8339    0.3188    0.3426   -1.3499   -0.0631
      -2.2588   -1.3077    3.5784    3.0349    0.7147
       0.8622   -0.4336    2.7694    0.7254   -0.2050


R2 =   1.5034    7.6688    4.8690    6.1388    3.1755
       7.9588    4.6860    8.5209    4.9075   -1.1491
       7.6361   -2.8299    3.9556    0.7862    5.5536
```

二项分布，即重复n次的伯努利试验，n个独立的是/非试验中成功的次数的离散概率分布，其中每次试验的成功概率为p。

$$p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}=b(k;n,p)$$
$$(k=0,1,\cdots,n),$$



n = 6、p = 0.5时的二项分布以及正态近似

➢ 二项分布的随机数据的产生

- 命令 参数为N，P的二项随机数据
- 函数 **binornd**
- 格式
  - ➢ **R = binornd(N, P)**
    - **N**、**P**为二项分布的两个参数，返回服从参数为**N**、**P**的二项分布的随机数。
  - ➢ **R = binornd(N, P, m)**
    - **m**指定随机数的个数。
  - ➢ **R = binornd(N, P, m, n)**
    - **m,n**分别表示**R**的行数和列数

```
clear all;

r=binornd(10, 0.5)

R=binornd(10, 0.5, 3, 4)   %产生一个3*4的矩阵

r =

    4


R =

    4     7     4     4
    6     4     6     5
    5     2     4     5
```

注意：每次运行结果是变化的！！！

➢常见分布的随机数产生

函数名	调用形式	注　　释
Unifrnd	unifrnd (A,B,m,n)	[A,B]上均匀分布(连续) 随机数
Unidrnd	unidrnd(N,m,n)	均匀分布（离散）随机数
Exprnd	exprnd(Lambda,m,n)	参数为 Lambda 的指数分布随机数
Normrnd	normrnd(MU,SIGMA,m,n)	参数为 MU，SIGMA 的正态分布随机数
chi2rnd	chi2rnd(N,m,n)	自由度为 N 的卡方分布随机数
Trnd	trnd(N,m,n)	自由度为 N 的 t 分布随机数
Frnd	frnd(N_1, N_2,m,n)	第一自由度为 N_1,第二自由度为 N_2 的 F 分布随机数
gamrnd	gamrnd(A, B,m,n)	参数为 A, B 的 γ 分布随机数
betarnd	betarnd(A, B,m,n)	参数为 A, B 的 β 分布随机数
lognrnd	lognrnd(MU, SIGMA,m,n)	参数为 MU, SIGMA 的对数正态分布随机数
nbinrnd	nbinrnd(R, P,m,n)	参数为 R，P 的负二项式分布随机数
ncfrnd	ncfrnd(N_1, N_2, delta,m,n)	参数为 N_1，N_2，delta 的非中心 F 分布随机数
nctrnd	nctrnd(N, delta,m,n)	参数为 N，delta 的非中心 t 分布随机数
ncx2rnd	ncx2rnd(N, delta,m,n)	参数为 N，delta 的非中心卡方分布随机数
raylrnd	raylrnd(B,m,n)	参数为 B 的瑞利分布随机数
weibrnd	weibrnd(A, B,m,n)	参数为 A, B 的韦伯分布随机数
binornd	binornd(N,P,m,n)	参数为 N, p 的二项分布随机数
geornd	geornd(P,m,n)	参数为 p 的几何分布随机数
hygernd	hygernd(M,K,N,m,n)	参数为 M，K，N 的超几何分布随机数
Poissrnd	poissrnd(Lambda,m,n)	参数为 Lambda 的泊松分布随机数

➢ 通用函数求各分布的随机数据

- 命令　求指定分布的随机数
- 函数　**random**
- 格式
  - ➢ **y = random('name', A1, A2, A3, m, n)**
    - **name**的取值见下页表；
    - **A1**，**A2**，**A3**为分布的参数；
    - **m**，**n**指定随机数的行和列

常见分布函数表

name 的取值			函数说明
'beta'	或	'Beta'	Beta 分布
'bino'	或	'Binomial'	二项分布
'chi2'	或	'Chisquare'	卡方分布
'exp'	或	'Exponential'	指数分布
'f'	或	'F'	F 分布
'gam'	或	'Gamma'	GAMMA 分布
'geo'	或	'Geometric'	几何分布
'hyge'	或	'Hypergeometric'	超几何分布
'logn'	或	'Lognormal'	对数正态分布
'nbin'	或	'Negative Binomial'	负二项式分布
'ncf'	或	'Noncentral F'	非中心 F 分布
'nct'	或	'Noncentral t'	非中心 t 分布
'ncx2'	或	'Noncentral Chi-square'	非中心卡方分布
'norm'	或	'Normal'	正态分布
'poiss'	或	'Poisson'	泊松分布
'rayl'	或	'Rayleigh'	瑞利分布
't'	或	'T'	T 分布
'unif'	或	'Uniform'	均匀分布
'unid'	或	'Discrete Uniform'	离散均匀分布
'weib'	或	'Weibull'	Weibull 分布

7.2 随机变量的概率密度计算

- 通用函数计算概率密度函数值
- 专用函数计算概率密度函数值
- 常见分布的密度函数作图

➢ 通用函数计算概率密度函数值

- 命令 通用函数计算概率密度函数值
- 函数 **pdf** (probability density function)
- 格式
  - ➢ **Y=pdf(name，x，A)**
  - ➢ **Y=pdf(name，x，A，B)**
  - ➢ **Y=pdf(name，x，A，B，C)**
- 说明
  - ➢ 返回在**X=x**处、参数为**A**、**B**、**C**的概率密度值，对于不同的分布，参数个数是不同；**name**为分布函数名，其取值如下表。

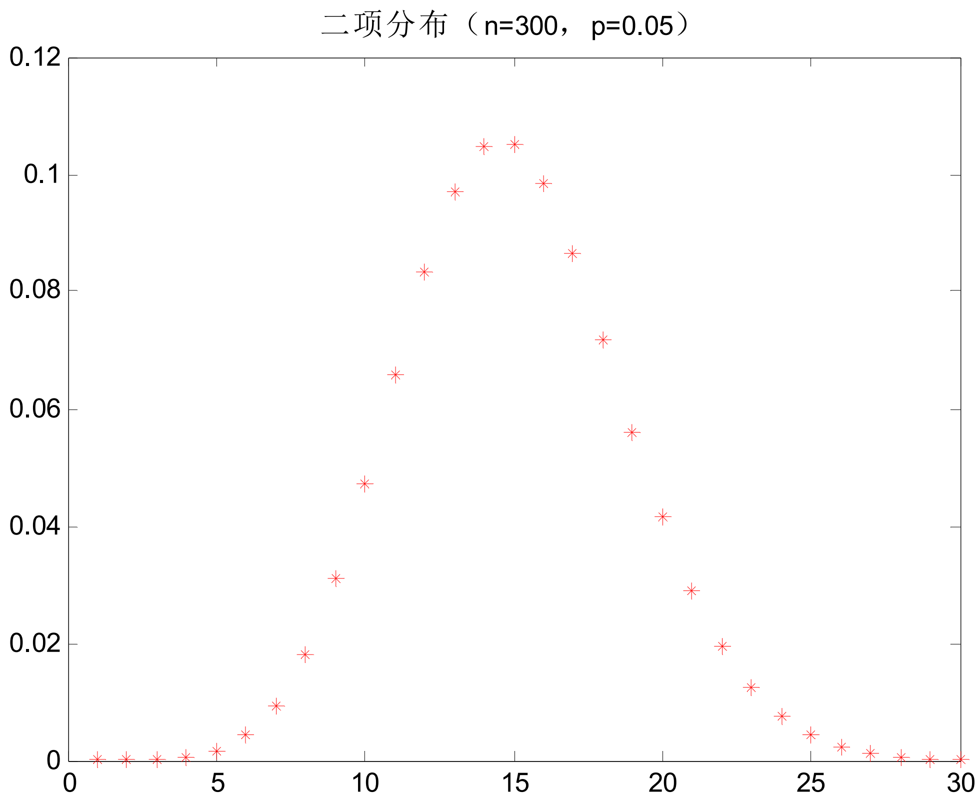
常见分布函数表

name 的取值			函数说明
'beta'	或	'Beta'	Beta 分布
'bino'	或	'Binomial'	二项分布
'chi2'	或	'Chisquare'	卡方分布
'exp'	或	'Exponential'	指数分布
'f'	或	'F'	F 分布
'gam'	或	'Gamma'	GAMMA 分布
'geo'	或	'Geometric'	几何分布
'hyge'	或	'Hypergeometric'	超几何分布
'logn'	或	'Lognormal'	对数正态分布
'nbin'	或	'Negative Binomial'	负二项式分布
'ncf'	或	'Noncentral F'	非中心 F 分布
'nct'	或	'Noncentral t'	非中心 t 分布
'ncx2'	或	'Noncentral Chi-square'	非中心卡方分布
'norm'	或	'Normal'	正态分布
'poiss'	或	'Poisson'	泊松分布
'rayl'	或	'Rayleigh'	瑞利分布
't'	或	'T'	T 分布
'unif'	或	'Uniform'	均匀分布
'unid'	或	'Discrete Uniform'	离散均匀分布
'weib'	或	'Weibull'	Weibull 分布

➢ 专用函数计算概率密度函数值

- 命令　二项分布的概率值
- 函数　**binopdf**
- 格式
  - ➢**binopdf (k, n, p)**
    - ⌖等同于**pdf(‘bino’, k, n, p)**，
    - ⌖**p —** 每次试验事件发生的概率；
    - ⌖**k—**事件发生**k**次；
    - ⌖**n—**试验总次数

```
clear all;
x=1:30;
y=binopdf(x, 300, 0.05);
figure;
plot(x, y, 'r*');
title('二项分布（n=300， p=0.05）');
```



事件A在每次试验中发生的概率是0.3, 计算在10次试验中A恰好发生6次的概率.

在命令窗口中输入:

```
p=binopdf(6, 10, 0.3)
```

结果显示:

```
p =      0.0368
```

表明: 参数是n=10, 概率是p=0.3的二项分布在X=6处的概率为0.0368.

事件A在每次试验中发生的概率是0.3, 求在4次试验中A发生次数的概率分布.

在命令窗口中输入:

```
p=binopdf(0:4, 4, 0.3)
```

%0: 4产生步长为 1 的等差数列 0, 1, 2, 3, 4.

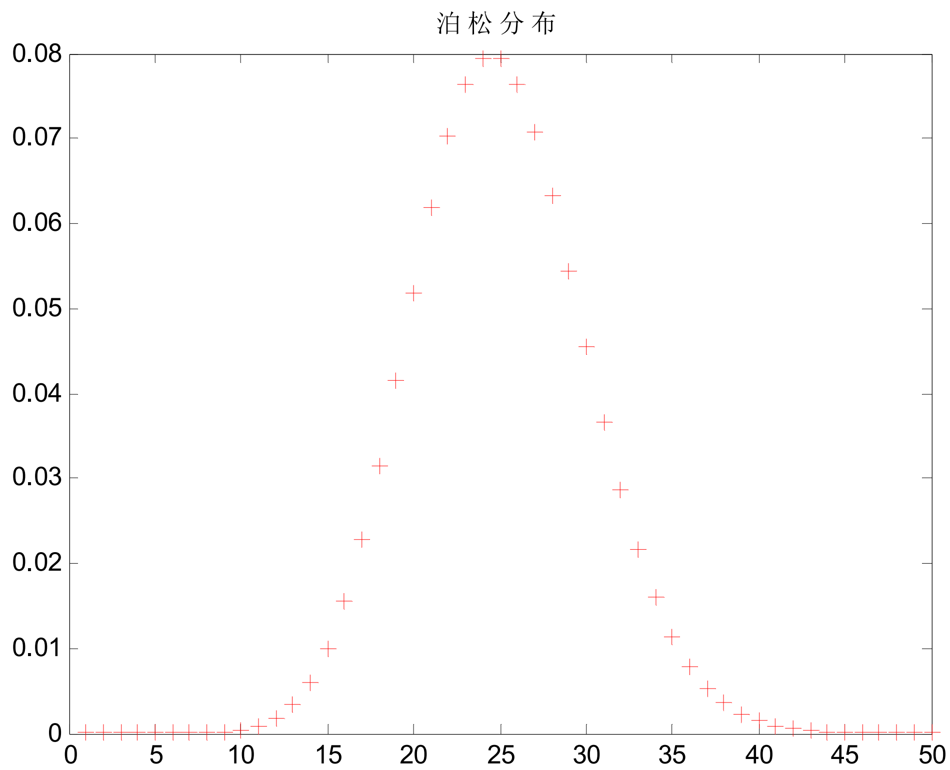
回车后显示:

p = 0.2401   0.4116   0.2646   0.0756   0.0081

计算的结果是: 参数是n=4, 概率是p=0.3的二项分布的分布律(当 x=0, 1, 2, 3, 4 时).

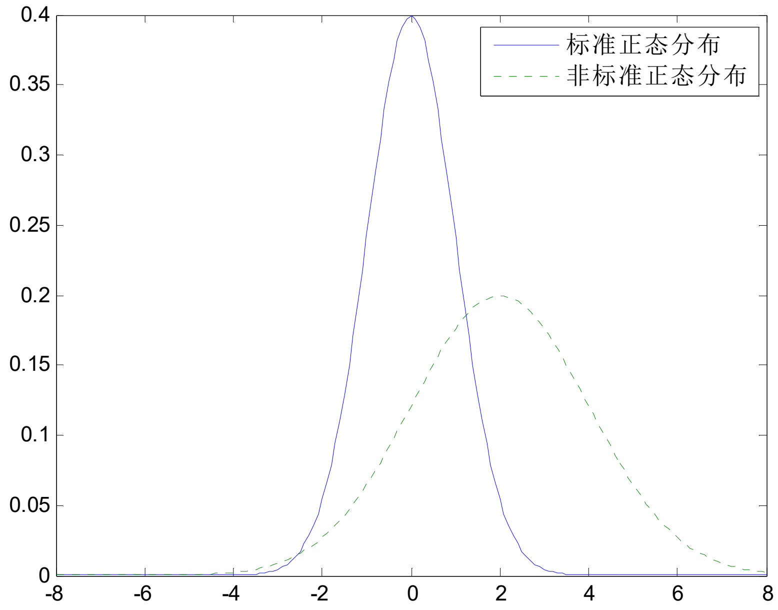
- 命令 泊松分布的概率值
- 函数 **poisspdf**
- 格式
  - **poisspdf(k, Lambda)**
  - 等同于**pdf('pois', k, Lambda)**

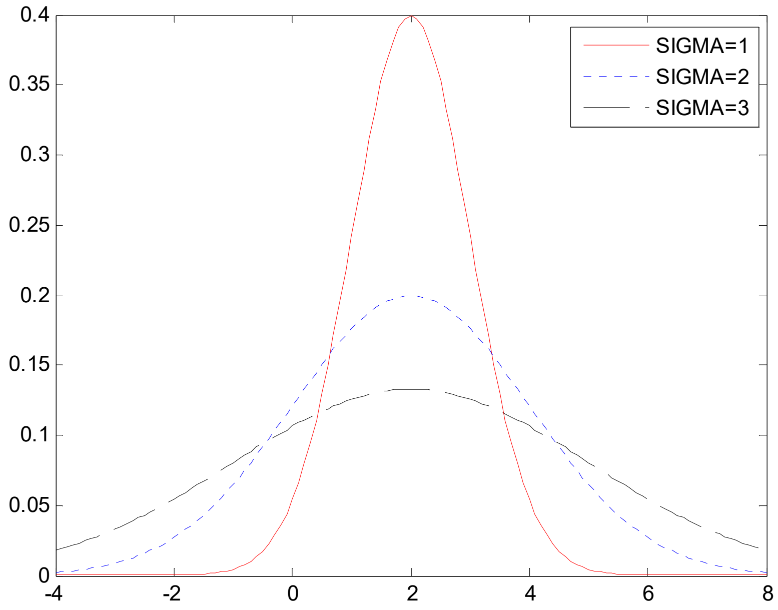
```
clear all;
x=1:50;
y=poisspdf(x, 25);       %泊松分布
figure;
plot(x,y,'r+');
title('泊松分布');
```

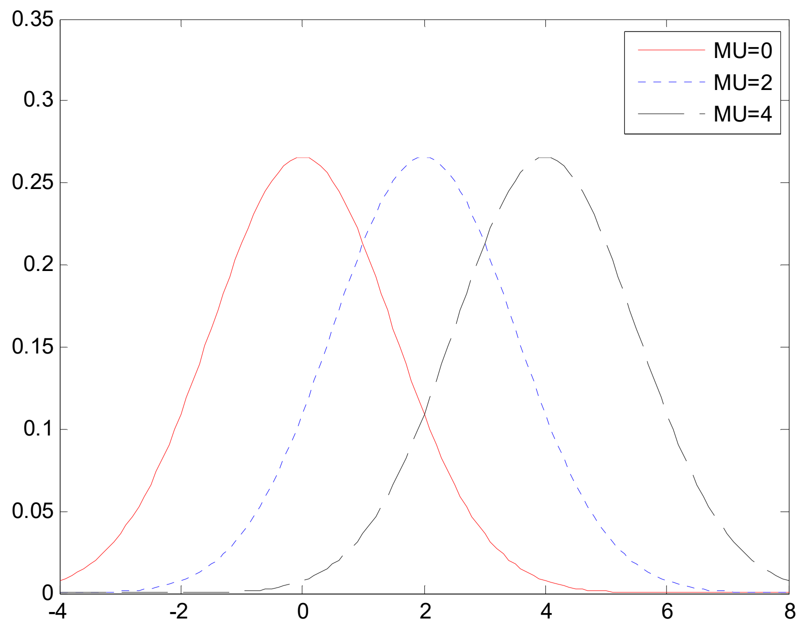


- 命令 正态分布的概率值
- 函数 **normpdf**
- 格式
  - **normpdf(x, mu, sigma)**
    - 计算参数为**μ=mu**，**σ=sigma**的正态分布密度函数在**x**处的值

```
clear all;
x=-8:0.1:8;
y1=normpdf(x, 0, 1);        %标准正态分布
y2=normpdf(x, 2, 2);        %非标准正态分布
figure;
plot(x, y1, x, y2, ':');
legend('标准正态分布','非标准正态分布');
x1=-4:0.1:8;
y3=normpdf(x1, 2, 1);       %SIGMA=1
y4=normpdf(x1, 2, 2);       %SIGMA=2
y5=normpdf(x1, 2, 3);       %SIGMA=3
figure;
plot(x1, y3, 'r-', x1, y4, 'b:', x1, y5, 'k--');
legend('SIGMA=1', 'SIGMA=2', 'SIGMA=3');
y6=normpdf(x1, 0, 1.5);             %MU=0
y7=normpdf(x1, 2, 1.5);             %MU=2
y8=normpdf(x1, 4, 1.5);             %MU=4
figure;
plot(x1, y6, 'r-', x1, y7, 'b:', x1, y8, 'k--');
legend('MU=0', 'MU=2', 'MU=4');
```


标准正态分布
非标准正态分布

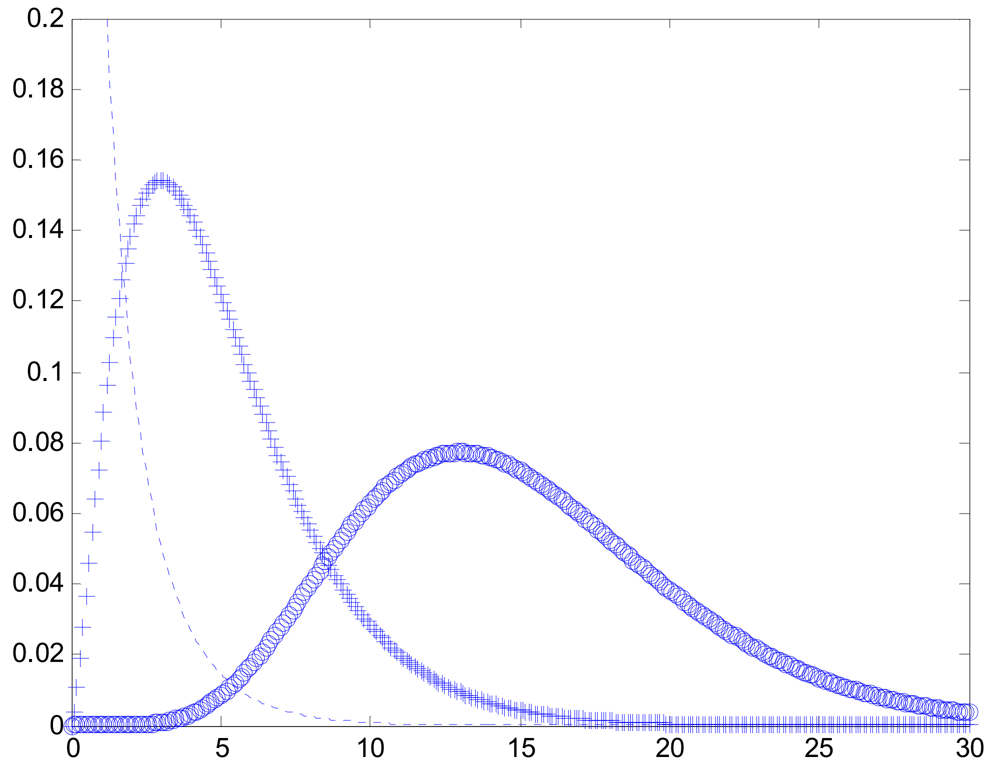

SIGMA=1
SIGMA=2
SIGMA=3


MU=0
MU=2
MU=4

专用函数计算概率密度函数表

函数名	调用形式	注　　　　释
Unifpdf	unifpdf (x, a, b)	[a,b]上均匀分布(连续)概率密度在 X=x 处的函数值
unidpdf	Unidpdf(x,n)	均匀分布（离散）概率密度函数值
Exppdf	exppdf(x, Lambda)	参数为 Lambda 的指数分布概率密度函数值
normpdf	normpdf(x, mu, sigma)	参数为 mu，sigma 的正态分布概率密度函数值
chi2pdf	chi2pdf(x, n)	自由度为 n 的卡方分布概率密度函数值
Tpdf	tpdf(x, n)	自由度为 n 的 t 分布概率密度函数值
Fpdf	fpdf(x, n_1, n_2)	第一自由度为 n_1,第二自由度为 n_2 的 F 分布概率密度函数值
gampdf	gampdf(x, a, b)	参数为 a, b 的 γ 分布概率密度函数值
betapdf	betapdf(x, a, b)	参数为 a, b 的 β 分布概率密度函数值
lognpdf	lognpdf(x, mu, sigma)	参数为 mu, sigma 的对数正态分布概率密度函数值
nbinpdf	nbinpdf(x, R, P)	参数为 R，P 的负二项式分布概率密度函数值
Ncfpdf	ncfpdf(x, n_1, n_2, delta)	参数为 n_1，n_2，delta 的非中心 F 分布概率密度函数值
Nctpdf	nctpdf(x, n, delta)	参数为 n，delta 的非中心 t 分布概率密度函数值
ncx2pdf	ncx2pdf(x, n, delta)	参数为 n，delta 的非中心卡方分布概率密度函数值
raylpdf	raylpdf(x, b)	参数为 b 的瑞利分布概率密度函数值
weibpdf	weibpdf(x, a, b)	参数为 a, b 的韦伯分布概率密度函数值
binopdf	binopdf(x,n,p)	参数为 n, p 的二项分布的概率密度函数值
geopdf	geopdf(x,p)	参数为 p 的几何分布的概率密度函数值
hygepdf	hygepdf(x,M,K,N)	参数为 M，K，N 的超几何分布的概率密度函数值
poisspdf	poisspdf(x,Lambda)	参数为 Lambda 的泊松分布的概率密度函数值

绘制卡方分布密度函数在自由度分别为1、5、15的图形




```
x=0:0.1:30;
y1=chi2pdf(x, 1);
plot(x, y1, ':')
hold on
y2=chi2pdf(x, 5);
plot(x, y2, '+')
y3=chi2pdf(x, 15);
plot(x, y3, 'o')
axis([0, 30, 0, 0.2])    %指定显示的图形区域
```

➢ 常见分布的密度函数作图

- 二项分布
- 卡方分布
- 非中心卡方分布
- 指数分布
- F分布
- 非中心F分布
- Γ分布
- 对数正态分布
- 负二项分布
- 正态分布
- 泊松分布
- 瑞利分布
- T分布
- 威布尔分布

课后练习

7.3 随机变量的累积概率值(分布函数值)

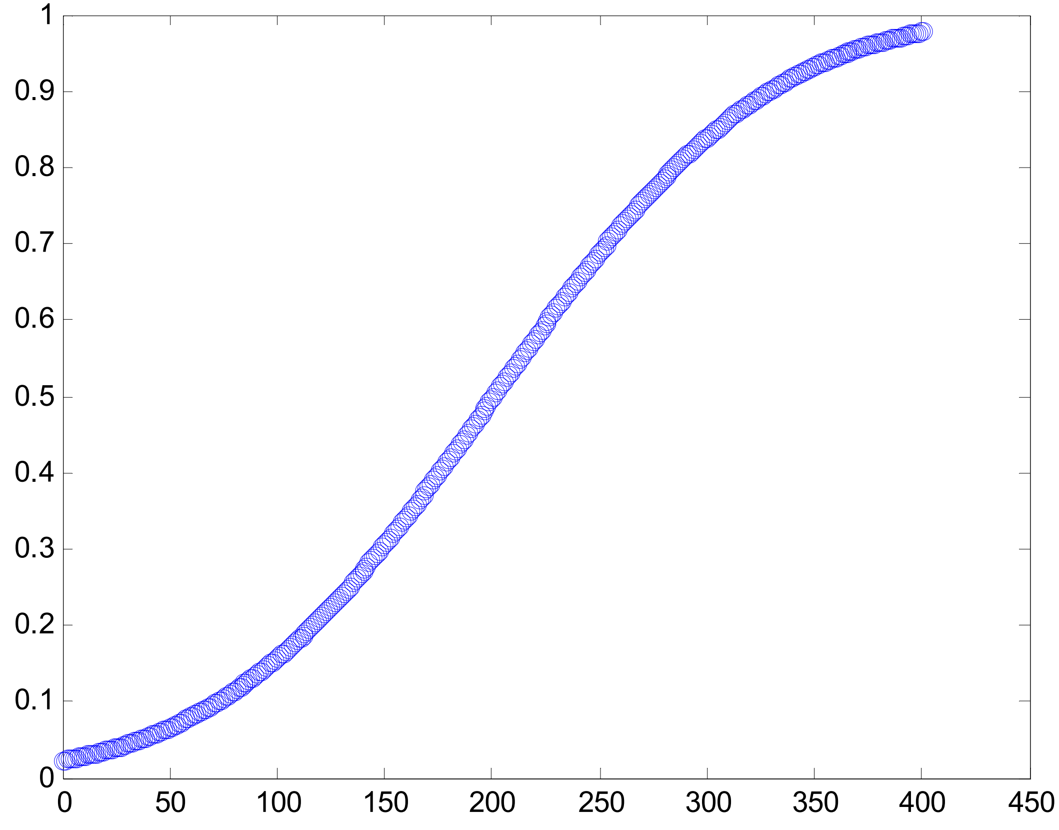
- 通用函数计算累积概率值
- 专用函数计算累积概率值（随机变量的概率之和）

➢ 通用函数计算累积概率值

- 命令 通用函数**cdf**用来计算随机变量的概率之和（累积概率值）
- 函数 **cdf (cumulative probability function)**
- 格式
  - ➢ **cdf('name', K, A)**
  - ➢ **cdf('name', K, A, B)**
  - ➢ **cdf('name', K, A, B, C)**
- 说明
  - ➢ 返回以**name**为分布、随机变量**X≤K**的概率之和的累积概率值，**name**的取值见“常见分布函数表”

```
p1 = cdf('Normal', -2:2, 0, 1)
p1 =
        0.0228  0.1587  0.5000  0.8413  0.9772

p2 = cdf('Normal', -2:0.01:2, 0,1) ;
plot(p2, 'o')
```



➢ 专用函数计算累积概率值（随机变量的概率之和）

- 命令 二项分布的累积概率值
- 函数 **binocdf**
- 格式 **binocdf (k, n, p)**
  - ➢ **n**为试验总次数
  - ➢ **p**为每次试验事件**A**发生的概率
  - ➢ **k**为**n**次试验中事件**A**发生的次数
  - ➢ 该命令返回**n**次试验中事件**A**恰好发生**k**次的概率。

- 命令 正态分布的累积概率值
- 函数 **normcdf**
- 格式 **normcdf(x, mu, sigma)**

➢返回**F(x)=** $\int_{-\infty}^{x} p(t)dt$ 的值，

➢**mu**、**sigma**为正态分布的两个参数

专用函数的累积概率值函数表

函数名	调用形式	注　　　　　　释
unifcdf	unifcdf (x, a, b)	[a,b]上均匀分布(连续)累积分布函数值 F(x)=P{X≤x}
unidcdf	unidcdf(x,n)	均匀分布（离散）累积分布函数值 F(x)=P{X≤x}
expcdf	expcdf(x, Lambda)	参数为 Lambda 的指数分布累积分布函数值 F(x)=P{X≤x}
normcdf	normcdf(x, mu, sigma)	参数为 mu，sigma 的正态分布累积分布函数值 F(x)=P{X≤x}
chi2cdf	chi2cdf(x, n)	自由度为 n 的卡方分布累积分布函数值 F(x)=P{X≤x}
tcdf	tcdf(x, n)	自由度为 n 的 t 分布累积分布函数值 F(x)=P{X≤x}
fcdf	fcdf(x, n_1, n_2)	第一自由度为 n_1,第二自由度为 n_2的 F 分布累积分布函数值
gamcdf	gamcdf(x, a, b)	参数为 a, b 的 γ 分布累积分布函数值 F(x)=P{X≤x}
betacdf	betacdf(x, a, b)	参数为 a, b 的 β 分布累积分布函数值 F(x)=P{X≤x}
logncdf	logncdf(x, mu, sigma)	参数为 mu, sigma 的对数正态分布累积分布函数值
nbincdf	nbincdf(x, R, P)	参数为 R，P 的负二项式分布概累积分布函数值 F(x)=P{X≤x}
ncfcdf	ncfcdf(x, n_1, n_2, delta)	参数为 n_1，n_2，delta 的非中心 F 分布累积分布函数值
nctcdf	nctcdf(x, n, delta)	参数为 n，delta 的非中心 t 分布累积分布函数值 F(x)=P{X≤x}
ncx2cdf	ncx2cdf(x, n, delta)	参数为 n，delta 的非中心卡方分布累积分布函数值
raylcdf	raylcdf(x, b)	参数为 b 的瑞利分布累积分布函数值 F(x)=P{X≤x}
weibcdf	weibcdf(x, a, b)	参数为 a, b 的韦伯分布累积分布函数值 F(x)=P{X≤x}
binocdf	binocdf(x,n,p)	参数为 n, p 的二项分布的累积分布函数值 F(x)=P{X≤x}
geocdf	geocdf(x,p)	参数为 p 的几何分布的累积分布函数值 F(x)=P{X≤x}
hygecdf	hygecdf(x,M,K,N)	参数为 M，K，N 的超几何分布的累积分布函数值
poisscdf	poisscdf(x,Lambda)	参数为 Lambda 的泊松分布的累积分布函数值 F(x)=P{X≤x}

- 某机床出次品的概率为0.01,求生产100件产品中：
  （1）恰有一件次品的概率；
  （2）至少有两件次品的概率。

可看作是100次独立重复试验，每次试验出次品的概率为0.01，恰有一件次品的概率：

p=binopdf(1, 100, 0.01)

显示结果为：

p=0.3697

（2）至少有两件次品的概率：

p=1-binocdf(1, 100, 0.01)

显示结果为：

p =0.2642

- 某公共汽车站从上午**7：00**起每**15**分钟来一班车。若某乘客在**7：00**到**7：30**间任何时刻到达此站是等可能的，试求他候车的时间不到**5**分钟的概率。
- 解：设乘客7点过X分钟到达此站，则X在[0，30]内服从均匀分布，当且仅当他在时间间隔（7：10，7：15）或（7：25，7：30）内到达车站时，候车时间不到5分钟。故其概率为：P1=P{10<X<15}+P{25<X<30}
- 程序：

```
p1=unifcdf(15, 0, 30)-unifcdf(10, 0, 30);
p2=unifcdf(30, 0, 30)-unifcdf(25, 0, 30);
p=p1+p2
```

- 则结果显示为：**p=0.3333**

pdf 和 cdf 的比较

设随机变量X服从参数是3的泊松分布，计算概率 P{X≤6}.

在命令窗口中输入:

```
p=poisscdf(6,3)
```

回车后显示:

p = 0.9665

结果表明：参数是 $\lambda=3$ 的泊松分布在 x=6 处的分布函数值 F(6)=P{X≤ 6}=0.9665 .

设随机变量 X服从参数是3的泊松分布，求概率 P{X=6}.

在命令窗口中输入:

```
p=poisspdf(6,3)
```

回车后显示:

p = 0.0504

结果表明：参数是 $\lambda=3$ 的泊松分布在x=6处的概率为0.0504.

7.4 随机变量的逆累积分布函数

已知 $F(x) = P\{X \le x\}$ ，求x。

- 通用函数计算逆累积分布函数值
- 专用函数-inv计算逆累积分布函数

➢ 通用函数计算逆累积分布函数值

- 命令 **icdf** 计算逆累积分布函数
- 格式

  ➢ **icdf('name', P, a_1, a_2, a_3)**

  ⌖ 返回分布为**name**，参数为a_1, a_2, a_3，累积概率值为**P**的临界值，这里**name**与前面表**1**相同。

  ⌖ 如果**P= cdf('name', x, a_1, a_2, a_3)**，则 **x = icdf('name', P, a_1, a_2, a_3)**

➢ 专用函数-inv计算逆累积分布函数

- 命令 正态分布逆累积分布函数
- 函数 **norminv**
- 格式
  - ➢**X=norminv(p, mu, sigma)**
    - **p**为累积概率值，
    - **mu**为均值，**sigma**为标准差，**X**为临界值
    - 满足：**p=P{X≤x}**。

常用临界值函数表

函数名	调用形式	注　　释
unifinv	x=unifinv (p, a, b)	均匀分布(连续)逆累积分布函数（P=P{X≤x}，求 x）
unidinv	x=unidinv (p,n)	均匀分布（离散）逆累积分布函数，x 为临界值
expinv	x=expinv (p, Lambda)	指数分布逆累积分布函数
norminv	x=Norminv(x,mu,sigma)	正态分布逆累积分布函数
chi2inv	x=chi2inv (x, n)	卡方分布逆累积分布函数
tinv	x=tinv (x, n)	t 分布累积分布函数
finv	x=finv (x, n_1, n_2)	F 分布逆累积分布函数
gaminv	x=gaminv (x, a, b)	γ 分布逆累积分布函数
betainv	x=betainv (x, a, b)	β 分布逆累积分布函数
logninv	x=logninv (x, mu, sigma)	对数正态分布逆累积分布函数
nbininv	x=nbininv (x, R, P)	负二项式分布逆累积分布函数
ncfinv	x=ncfinv (x, n_1, n_2, delta)	非中心 F 分布逆累积分布函数
nctinv	x=nctinv (x, n, delta)	非中心 t 分布逆累积分布函数
ncx2inv	x=ncx2inv (x, n, delta)	非中心卡方分布逆累积分布函数
raylinv	x=raylinv (x, b)	瑞利分布逆累积分布函数
weibinv	x=weibinv (x, a, b)	韦伯分布逆累积分布函数
binoinv	x=binoinv (x,n,p)	二项分布的逆累积分布函数
geoinv	x=geoinv (x,p)	几何分布的逆累积分布函数
hygeinv	x=hygeinv (x,M,K,N)	超几何分布的逆累积分布函数
poissinv	x=poissinv (x,Lambda)	泊松分布的逆累积分布函数

- 公共汽车门的高度是按成年男子与车门顶碰头的机会不超过**1%**设计的。设男子身高**X**（单位：**cm**）服从正态分布**N**（**175**，**6**），求车门的最低高度。
- 解：设h为车门高度，X为身高，求满足条件 $P\{X>h\}0.01$ 的h，即 $P\{X<h\}0.99$。
- 程序：

```
h=norminv(0.99, 175, 6)
```

- 结果：

```
h=
   188.9581
```

7.5 随机变量的数字特征

- 平均值、中值
- 数据比较
- 期望、方差、倾斜度
- 常见分布的期望和方差
- 协方差与相关系数

➢ 平均值、中值

- 命令 利用**mean**求算术平均值
- 格式
  - ➢ **mean(X)**
    - **X**为向量，返回**X**中各元素的平均值
  - ➢ **mean(A)**
    - **A**为矩阵，返回**A**中各列元素的平均值构成的向量
  - ➢ **mean(A, dim)**
    - 在给出的维数内的平均值
- 说明
  - ➢ **X**为向量时，算术平均值的数学含义是 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ ，即样本均值。

```
clear all;
A=[1 2 3 4; 2 0 1 4]
m1=mean(A)          %对列元素求算术平均值
m2=mean(A, 2)       %对行元素求算术平均值
```

```
A =    1    2    3    4
       2    0    1    4
m1 =   1.5000   1.0000   2.0000   4.0000
m2 =   2.5000
       1.7500
```

- 命令
  - ➢忽略**NaN**计算算术平均值
- 格式
  - ➢**nanmean(X)**
    - **X**为向量，返回**X**中除**NaN**外元素的算术平均值。
  - ➢**nanmean(A)**
    - **A**为矩阵，返回**A**中各列除**NaN**外元素的算术平均值向量。

```
clear all;
A=[1 2 nan 4; 2 nan 1 nan]
m1=mean(A)
m2=nanmean(A)          %对列元素求算术平均值
m3=nanmean(A, 2)         %对行元素求算术平均值
```

```
A =     1     2   NaN     4
        2   NaN     1   NaN
m1 =  1.5000     NaN     NaN     NaN
m2 =  1.5000  2.0000  1.0000  4.0000
m3 =  2.3333
      1.5000
```

- 命令 利用**median**计算中值（中位数）
- 格式
  - **median(X)**
    - **X**为向量，返回**X**中各元素的中位数。
  - **median(A)**
    - **A**为矩阵，返回**A**中各列元素的中位数构成的向量。
  - **median(A, dim)**
    - 求给出的维数内的中位数

把所有观察值高低排序后找出正中间的一个作为中位数。如果观察值有偶数个，通常取最中间的两个数值的平均数作为中位数。

```
clear all;
A1=[1 2 3 4]
m1=median(A1)          %向量的中位数
A2=[1 2 3; 2 3 4; 4 1 8]
m2=median(A2)          %列元素的中位数
m3=median(A2, 2)       %行元素的中位数
```

```
A1 =   1    2    3    4
m1 =   2.5000
A2 =   1    2    3
       2    3    4
       4    1    8
m2 =   2    2    4
m3 =   2
       3
       4
```

- 命令 忽略**NaN**计算中位数
- 格式
  - **nanmedian(X)**
    - **X**为向量，返回**X**中除**NaN**外元素的中位数。
  - **nanmedian(A)**
    - **A**为矩阵，返回**A**中各列除**NaN**外元素的中位数向量。

```
clear all;
A1=[1 2 nan 1 nan 3 4]
m1=nanmedian(A1)             %向量的中位数
A2=[1 2 nan; 2 nan 4; 4 1 8]
m2=median(A2)
m3=nanmedian(A2)             %列元素忽略NaN的中位数
m4=nanmedian(A2, 2)          %行元素忽略NaN的中位数
```

```
A1 =   1    2   NaN    1   NaN    3    4
m1 =   2
A2 =   1    2   NaN
       2   NaN    4
       4    1     8
m2 =   2   NaN   NaN
m3 =   2.0000    1.5000    6.0000
m4 =   1.5000
       3.0000
       4.0000
```

几何平均数（geometric mean）：
n个观察值连乘积的n次方根。

- 命令 利用**geomean**计算几何平均数
- 格式
  - **M=geomean(X)**
    - **X**为向量，返回**X**中各元素的几何平均数。
  - **M=geomean(A)**
    - **A**为矩阵，返回**A**中各列元素的几何平均数构成的向量。
- 说明
  - 几何平均数的数学含义是 $M=(\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$ ，

    其中：样本数据非负，主要用于对数正态分布。

```
clear all;
A1=[1 2 3 4]
m1=geomean(A1)      %向量的几何平均数
A2=[1 2 3; 2 3 4]
m2=geomean(A2)      %列元素的几何平均数
m3=geomean(A2, 2)   %行元素的几何平均数
```

```
A1 =   1     2     3     4
m1 =   2.2134
A2 =   1     2     3
       2     3     4
m2 =   1.4142    2.4495    3.4641
m3 =   1.8171
       2.8845
```

调和平均数（Harmonic Average）：又称倒数平均数，是总体各统计变量倒数的算术平均数的倒数。

- 命令 利用**harmmean**求调和平均值
- 格式
  - **M=harmmean(X)**
    - **X**为向量，返回**X**中各元素的调和平均值。
  - **M=harmmean(A)**
    - **A**为矩阵，返回**A**中各列元素的调和平均值构成的向量。
- 说明
  - 调和平均值的数学含义是 $M = \dfrac{n}{\sum\limits_{i=1}^{n} \dfrac{1}{x_i}}$，其中：样本数据非**0**，主要用于严重偏斜分布。

```
clear all;
A1=[1 2 3 4]
m1=harmmean(A1)     %向量的调和平均数
A2=[1 2 3; 2 3 4]
m2=harmmean(A2)     %列元素的调和平均数
m3=harmmean(A2, 2)   %行元素的调和平均数
```

```
A1 =   1     2     3     4
m1 =   1.9200
A2 =   1     2     3
       2     3     4
m2 =   1.3333    2.4000    3.4286
m3 =   1.6364
       2.7692
```

➢ 数据比较

- 命令　排序
- 格式
  - ➢ **Y=sort(X)**
    - **X**为向量，返回**X**按由小到大排序后的向量。
  - ➢ **Y=sort(A)**
    - **A**为矩阵，返回**A**的各列按由小到大排序后的矩阵。
  - ➢ **[Y,I]=sort(A)**
    - **Y**为排序的结果，**I**中元素表示**Y**中对应元素在**A**中位置。
  - ➢ **sort(A, dim)**
    - 在给定的维数**dim**内排序
- 说明
  - ➢ 若**X**为复数，则通过**|X|**排序。

```
clear all;
X=[1 3 4; 8 3 5; 2 7 4]
y1=sort(X)                    %按列由小到大排序
y2=sort(X, 2)                 %按行由小到大排序
y3=sort(X, 1, 'descend')      %按列由大到小排序
[Y1,I]=sort(X)
[Y2,I]=sort(X, 2)
```

```
X =     1   3   4        Y1 =    1   3   4
        8   3   5                2   3   4
        2   7   4                8   7   5
y1 =    1   3   4        I =     1   1   1
        2   3   4                3   2   3
        8   7   5                2   3   2
y2 =    1   3   4        Y2 =    1   3   4
        3   5   8                3   5   8
        2   4   7                2   4   7
y3 =    8   7   5        I =     1   2   3
        2   3   4                2   3   1
        1   3   4                1   3   2
```

- 命令 按行方式排序
- 函数 **sortrows**
- 格式
  - **Y=sortrows(A)**
    - **A**为矩阵，返回矩阵**Y**，**Y**按**A**的第**1**列由小到大，以行方式排序后生成的矩阵。
  - **Y=sortrows(A, col)**
    - 按指定列**col**由小到大进行排序
  - **[Y,I]=sortrows(A, col)**
    - **Y**为排序的结果，**I**表示**Y**中第**col**列元素在**A**中位置。
- 说明
  - 若**X**为复数，则通过**|X|**的大小排序。

- 命令 求最大值与最小值之差
- 函数 **range**
- 格式
  - **Y=range(X)**
    - **X**为向量，返回**X**中的最大值与最小值之差。
  - **Y=range(A)**
    - **A**为矩阵，返回**A**中各列元素的最大值与最小值之差。

```
clear all;
X1=[1 3 4 10 3 5]
y1=range(X1)                %向量
X2=[1 3 5; 2 4 6; 8 4 3]
y2=range(X2)                %矩阵
y3=range(X2, 2)

X1 =   1    3    4   10    3    5
y1 =   9
X2 =   1    3    5
       2    4    6
       8    4    3
y2 =   7    1    3
y3 =   4
       4
       5
```

- 命令 求最大值和最小值
- 函数**minmax( )**获取数据的最小值和最大值。

```
clear all;
X1=[2 5 4 12 3 15]
y1=minmax(X1)               %向量
X2=[1 3 5; 2 4 6; 8 4 3]
y2=minmax(X2)               %矩阵
```

```
X1 =   2    5    4    12    3    15
y1 =   2    15
X2 =   1    3    5
       2    4    6
       8    4    3
y2 =   1    5
       2    6
       3    8
```

➢ 期望

甲乙两个人赌博，他们两人获胜的机率相等，比赛规则是先胜三局者为赢家，赢家可以获得100法郎的奖励。比赛进行到第三局的时候，甲胜了两局，乙胜了一局，这时由于某些原因中止了比赛，那么如何分配这100法郎才比较公平？

甲获胜的概率为$1/2+(1/2)*(1/2)=3/4$，
乙获胜的概率为$(1/2)*(1/2)=1/4$。
因此，甲的期望值为$100*3/4=75$法郎，乙的期望值为25法郎。

在概率论和统计学中，一个离散性随机变量的期望值（或数学期望、或均值，亦简称期望）是试验中每次可能结果的概率乘以其结果的总和。换句话说，期望值是随机试验在同样的机会下重复多次的结果计算出的等同“期望”的平均值。

- 命令 计算样本均值
- 函数 **mean**

```
clear all;
X1=rand(1, 8)   %向量
y=mean(X1)
X2=rand(4, 7)   %矩阵
Y=mean(X2, 2)
```

```
X1 =    0.8147   0.9058   0.1270   0.9134   0.6324   0.0975   0.2785   0.5469
y =     0.5395
X2 =    0.9575   0.9572   0.4218   0.6557   0.6787   0.6555   0.2769
        0.9649   0.4854   0.9157   0.0357   0.7577   0.1712   0.0462
        0.1576   0.8003   0.7922   0.8491   0.7431   0.7060   0.0971
        0.9706   0.1419   0.9595   0.9340   0.3922   0.0318   0.8235
Y =     0.6576
        0.4824
        0.5922
        0.6076
```

➢ 方差

方差是各个数据分别与其和的平均数之差的平方的和的平均数，用字母D表示。在概率论和数理统计中，方差（Variance）用来度量随机变量和其数学期望（即均值）之间的偏离程度。在许多实际问题中，研究随机变量和均值之间的偏离程度有着重要意义。

标准差（Standard Deviation），中文环境中又常称均方差，是离均差平方和平均后的方根，用σ表示。标准差是方差的算术平方根。标准差能反映一个数据集的离散程度。平均数相同的，标准差未必相同。

- 命令 求样本方差
- 函数 **var**
- 格式

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

- D=var(X)
  - var(X)=s^2,若X为向量，则返回向量的样本方差。
- D=var(A)
  - A为矩阵，则D为A的列向量的样本方差构成的行向量。
- D=var(X, 1)
  - 返回向量（矩阵）X的简单方差（即置前因子为1/n的方差）
- D=var(X, w)
  - 返回向量（矩阵）X的以w为权重的方差

```
clear all;
A=rand(1, 8)        %向量
y=var(A)
B=rand(3, 5)        %矩阵
Y1=var(B)
Y2=var(B, 1)
W=[0.1 0.2 0.3]     %系数
Y3=var(B, W)
```

A =	0.6948	0.3171	0.9502	0.0344	0.4387	0.3816	0.7655	0.7952
y =	0.0926							
B =	0.1869	0.6463	0.2760	0.1626	0.9597			
	0.4898	0.7094	0.6797	0.1190	0.3404			
	0.4456	0.7547	0.6551	0.4984	0.5853			
Y1 =	0.0268	0.0030	0.0512	0.0431	0.0973			
Y2 =	0.0178	0.0020	0.0341	0.0287	0.0649			
W =	0.1000	0.2000	0.3000					
Y3 =	0.0110	0.0015	0.0211	0.0335	0.0430			

- 命令 求标准差
- 函数 **std**
- 格式

$$std = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} x_i - \overline{X}}$$

  - **std(X)**
    - 返回向量（矩阵）**X**的样本标准差（置前因子为**1/(n-1)**）：
  - **std(X,1)**
    - 返回向量（矩阵）**X**的标准差（置前因子为**1/n**）
  - **std(X, 0)**
    - 与**std (X)**相同
  - **std(X, flag, dim)**
    - 返回向量（矩阵）中维数为**dim**的标准差值，
    - 其中**flag=0**时，置前因子为**1/(n-1)**；否则置前因子为**1/n**。

```
clear all;
A=[2 3 4 1; 1 2 3 2;  1 2 3]
y1=std(A)
y2=std(A, 0)
y3=std(A, 1)
y4=std(A, 1, 2)
```

```
A =     2     3     4     1
        1     2     3     2
        8     1     2     3
y1 =    3.7859    1.0000    1.0000    1.0000
y2 =    3.7859    1.0000    1.0000    1.0000
y3 =    3.0912    0.8165    0.8165    0.8165
y4 =    1.1180
        0.7071
        2.6926
```

➢ 倾斜度

偏斜度是对统计数据分布偏斜方向及程度的度量。统计数据的频数分布有的是对称的，有的是不对称的，即呈现偏态。在偏态分布中，当偏斜度为正值时，分布正偏，即众数（在统计分布上具有明显集中趋势点的数值）位于算术平均数的左侧；当偏斜度为负值时，分布负偏，即众数位于算术平均数的右侧。

我们可以利用众数、中位数和算术平均数之间的关系判断分布是左偏态还是右偏态，但要度量分布偏斜的程度，就需要计算偏斜度。

- 命令 样本的偏斜度
- 函数 **skewness**
- 格式
  - **y = skewness(X)**
    - **X**为向量，返回**X**的元素的偏斜度；**X**为矩阵，返回**X**各列元素的偏斜度构成的行向量。
  - **y = skewness(X, flag)**
    - **flag=0**表示偏斜纠正，**flag=1**（默认）表示偏斜不纠正。
- 说明
  - 偏斜度样本数据关于均值不对称的一个测度，如果偏斜度为负，说明均值左边的数据比均值右边的数据更散；如果偏斜度为正，说明均值右边的数据比均值左边的数据更散，因而正态分布的偏斜度为 **0**；
  - 偏斜度的定义： $$y=\frac{E(x-\mu)^3}{\sigma^3}$$

    其中：**μ**为**x**的均值，**σ**为**x**的标准差，**E(.)**为期望值算子

```
clear all;
X=randn([4, 5])
y1=skewness(X)
y2=skewness(X, 0)
y3=skewness(X, 0, 2)
```

```
X =     -0.2256    0.5525   -1.4916   -0.6156   -0.7648
         1.1174    1.1006   -0.7423    0.7481   -1.4023
        -1.0891    1.5442   -1.0616   -0.1924   -1.4224
         0.0326    0.0859    2.3505    0.8886    0.4882
y1 =     0.2041   -0.0236    1.0484   -0.1452    0.8026
y2 =     0.3535   -0.0408    1.8159   -0.2515    1.3902
y3 =     0.2494
        -0.7180
         1.5482
         1.5669
```

➢ 常见分布的期望和方差

- 命令 均匀分布（连续）的期望和方差
- 函数 **unifstat**
- 格式
  - ➢ **[M,V] = unifstat(A, B)**
    - **A**、**B**为标量时，就是区间上均匀分布的期望和方差
    - **A**、**B**也可为向量或矩阵，则**M**、**V**也是向量或矩阵。

```
clear all;
a1=1;
b1=5;
[m1,v1]=unifstat(a1, b1)
a2=1:5;
b2=2.*a2;
[m2,v2]=unifstat(a2, b2)
```

```
m1 =   3
v1 =    1.3333
m2 =   1.5000    3.0000    4.5000    6.0000    7.5000
v2 =    0.0833    0.3333    0.7500    1.3333    2.0833
```

- 命令 正态分布的期望和方差
- 函数 **normstat**
- 格式
  - **[M,V] = normstat(MU, SIGMA)**
    - **MU**、**SIGMA**可为标量也可为向量或矩阵，则 **M=MU**，**V=SIGMA**^{2}。

```
clear all;
n1=2;
n2=3;
[m1, v1]=normstat(n1, n2)
n3=1:4;
[m2, v2]=normstat(n3'*n3, n3'*n3)
```

```
m1 =   2
v1 =   9
m2 =   1    2    3    4
       2    4    6    8
       3    6    9   12
       4    8   12   16
v2 =   1    4    9   16
       4   16   36   64
       9   36   81  144
      16   64  144  256
```

- 命令 二项分布的均值和方差
- 函数 **binostat**
- 格式
  - **[M,V] = binostat(N, P)**
    - **N**，**P**为二项分布的两个参数，可为标量也可为向量或矩阵。

```
clear all
n1=100
p1=0.3
[m1, v1]=binostat(n1, p1)
n2=logspace(1, 4, 4)
[m2, v2]=binostat(n2, 1./n2)
```

```
n1 =    100
p1 =    0.3000
m1 =              30
v1 =    21
n2 =    10        100       1000      10000
m2 =              1     1     1     1
v2 =    0.9000    0.9900    0.9990    0.9999
```

常见分布的均值和方差

函数名	调用形式	注　释
unifstat	[M,V]=unifstat (a, b)	均匀分布(连续)的期望和方差，M 为期望，V 为方差
unidstat	[M,V]=unidstat (n)	均匀分布（离散）的期望和方差
expstat	[M,V]=expstat (p, Lambda)	指数分布的期望和方差
normstat	[M,V]=normstat(mu,sigma)	正态分布的期望和方差
chi2stat	[M,V]=chi2stat (x, n)	卡方分布的期望和方差
tstat	[M,V]=tstat (n)	t 分布的期望和方差
fstat	[M,V]=fstat (n_1, n_2)	F 分布的期望和方差
gamstat	[M,V]=gamstat (a, b)	γ 分布的期望和方差
betastat	[M,V]=betastat (a, b)	β 分布的期望和方差
lognstat	[M,V]=lognstat (mu, sigma)	对数正态分布的期望和方差
nbinstat	[M,V]=nbinstat (R, P)	负二项式分布的期望和方差
ncfstat	[M,V]=ncfstat (n_1, n_2, delta)	非中心 F 分布的期望和方差
nctstat	[M,V]=nctstat (n, delta)	非中心 t 分布的期望和方差
ncx2stat	[M,V]=ncx2stat (n, delta)	非中心卡方分布的期望和方差
raylstat	[M,V]=raylstat (b)	瑞利分布的期望和方差
Weibstat	[M,V]=weibstat (a, b)	韦伯分布的期望和方差
Binostat	[M,V]=binostat (n,p)	二项分布的期望和方差
Geostat	[M,V]=geostat (p)	几何分布的期望和方差
hygestat	[M,V]=hygestat (M,K,N)	超几何分布的期望和方差
Poisstat	[M,V]=poisstat (Lambda)	泊松分布的期望和方差

➢ 协方差与相关系数

协方差用于衡量两个变量的总体误差。而方差是协方差的一种特殊情况，即当两个变量是相同的情况。

相关系数是用以反映变量之间相关关系密切程度的统计指标。相关系数是按积差方法计算，同样以两变量与各自平均值的离差为基础，通过两个离差相乘来反映两变量之间相关程度

- 命令 协方差
- 函数 **cov**
- 格式
  - **cov(X)**
    - 求向量**X**的协方差
  - **cov(A)**
    - 求矩阵**A**的协方差矩阵，该协方差矩阵的对角线元素是**A**的各列的方差，即：**var(A)=diag(cov(A))**。
  - **cov(X, Y)**
    - **X,Y**为等长列向量，等同于**cov([X Y])**。

```
clear all;
X1=rand(1, 5)        X1 =  0.4733  0.3517  0.8308  0.5853  0.5497
c1=cov(X1)           c1 =  0.0312
X2=rand(1, 5)        X2 =  0.9172  0.2858  0.7572  0.7537  0.3804
c2=cov(X2)           c2 =  0.0735
c3=cov(X1, X2)       c3 =  0.0312  0.0217
X3=rand(4, 4)              0.0217  0.0735
c4=cov(X3)           X3 =  0.5678  0.7792  0.4694  0.7943
c5=cov(X3(:, 1))           0.0759  0.9340  0.0119  0.3112
                           0.0540  0.1299  0.3371  0.5285
                           0.5308  0.5688  0.1622  0.1656
                     c4 =  0.0785  0.0272  0.0235  0.0128
                           0.0272  0.1219 -0.0261 -0.0042
                           0.0235 -0.0261  0.0400  0.0468
                           0.0128 -0.0042  0.0468  0.0749
                     c5 =  0.0785
```

- 命令 相关系数
- 函数 **corrcoef**
- 格式
  - **corrcoef(X, Y)**
    - 返回列向量**X,Y**的相关系数，等同于**corrcoef([X Y])**。
  - **corrcoef (A)**
    - 返回矩阵**A**的列向量的相关系数矩阵

```
clear all;
X=[1 2 3; 3 4 6; 7 4 2]
[R1,P1]=corrcoef(X)
x=[1 2 3 5 3 2];
y=[2 3 5 3 1 9];
R2=corrcoef(x, y)
```

```
X =     1     2     3
        3     4     6
        7     4     2
R1 =    1.0000    0.7559   -0.4193
        0.7559    1.0000    0.2774
       -0.4193    0.2774    1.0000
P1 =    1.0000    0.4544    0.7245
        0.4544    1.0000    0.8211
        0.7245    0.8211    1.0000
R2 =    1.0000   -0.1195
       -0.1195    1.0000
```

6 统计作图

- 正整数的频率表
- 经验累积分布函数图
- 最小二乘拟合直线
- 正态分布概率图
- 威布尔(Weibull)概率图
- 样本数据的盒图
- 给当前图形加一条参考线
- 在当前图形中加入一条多项式曲线
- 样本的概率图形
- 附加有正态密度曲线的直方图
- 在指定的界线之间画正态密度曲线

➢ 正整数的频率表

- 命令 正整数的频率表
- 函数 **tabulate**
- 格式
  - ➢ **table = tabulate(X)**
    - **X**为正整数构成的向量，返回**3**列：
      第**1**列中包含**X**的值**,**
      第**2**列为这些值的个数，
      第**3**列为这些值的频率。

```
clear all;
X=[2 3 4 1 2 4 6 5 8 3 2 1 6];
tabulate(X)
```

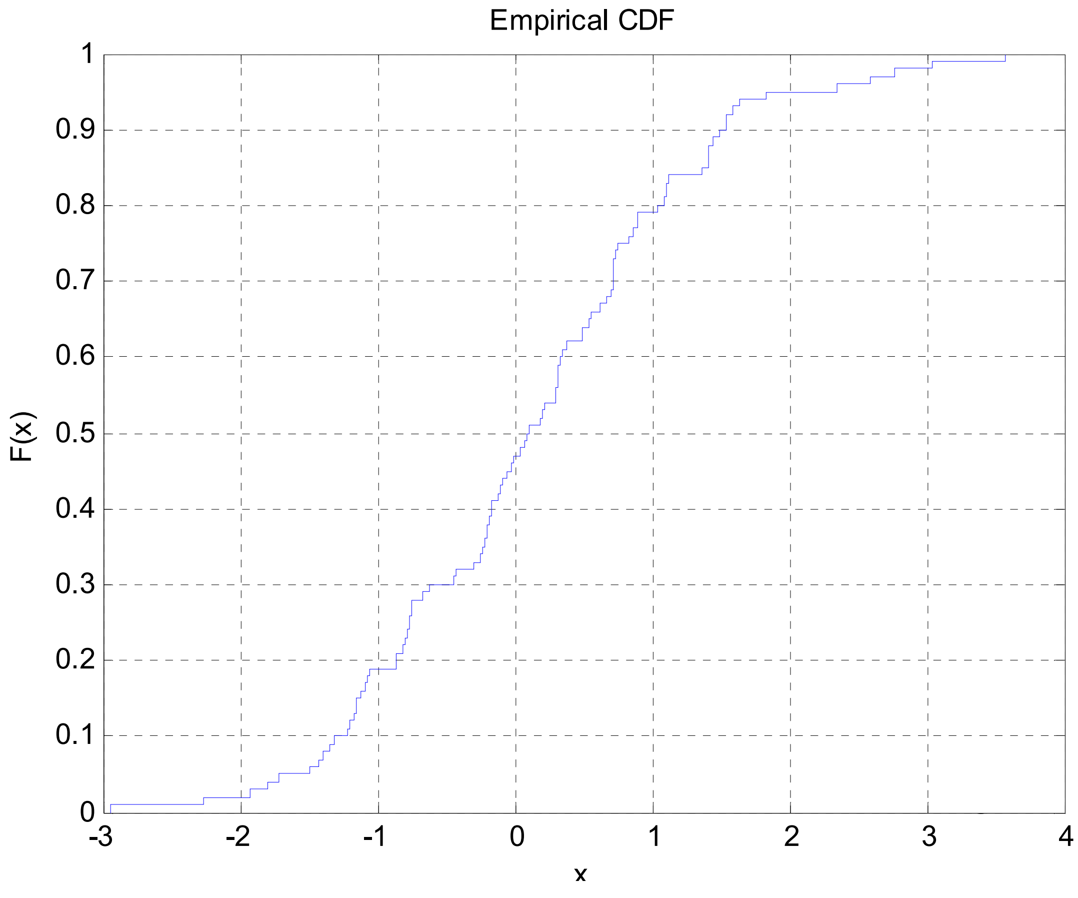
Value	Count	Percent
1	2	15.38%
2	3	23.08%
3	2	15.38%
4	2	15.38%
5	1	7.69%
6	2	15.38%
7	0	0.00%
8	1	7.69%

➢ 经验累积分布函数图

- 函数 **cdfplot**
- 格式
  - ➢ **cdfplot(X)**
    - 作样本**X**（向量）的累积分布函数图形
  - ➢ **h = cdfplot(X)**
    - **h**表示曲线的环柄
  - ➢ **[h,stats] = cdfplot(X)**
    - **stats**表示样本的一些特征

```
clear all;
X=normrnd(0, 1, 100, 1);
[h,stats]=cdfplot(X)
```

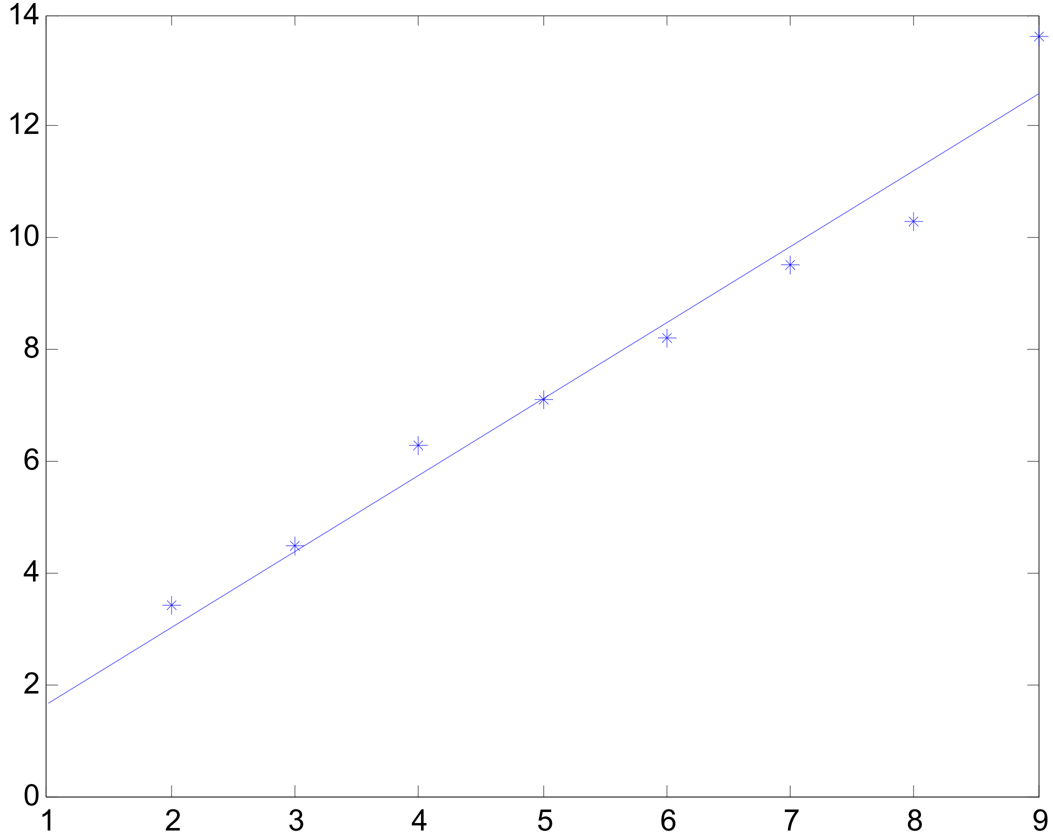
```
stats =
    min: -2.9443
    max: 3.5784
   mean: 0.1231
 median: 0.0954
    std: 1.1624
```



➢ 最小二乘拟合直线

- 函数 **lsline**
- 格式
  - ➢ **lsline**
    - 最小二乘拟合直线
  - ➢ **h = lsline**
    - **h**为直线的句柄

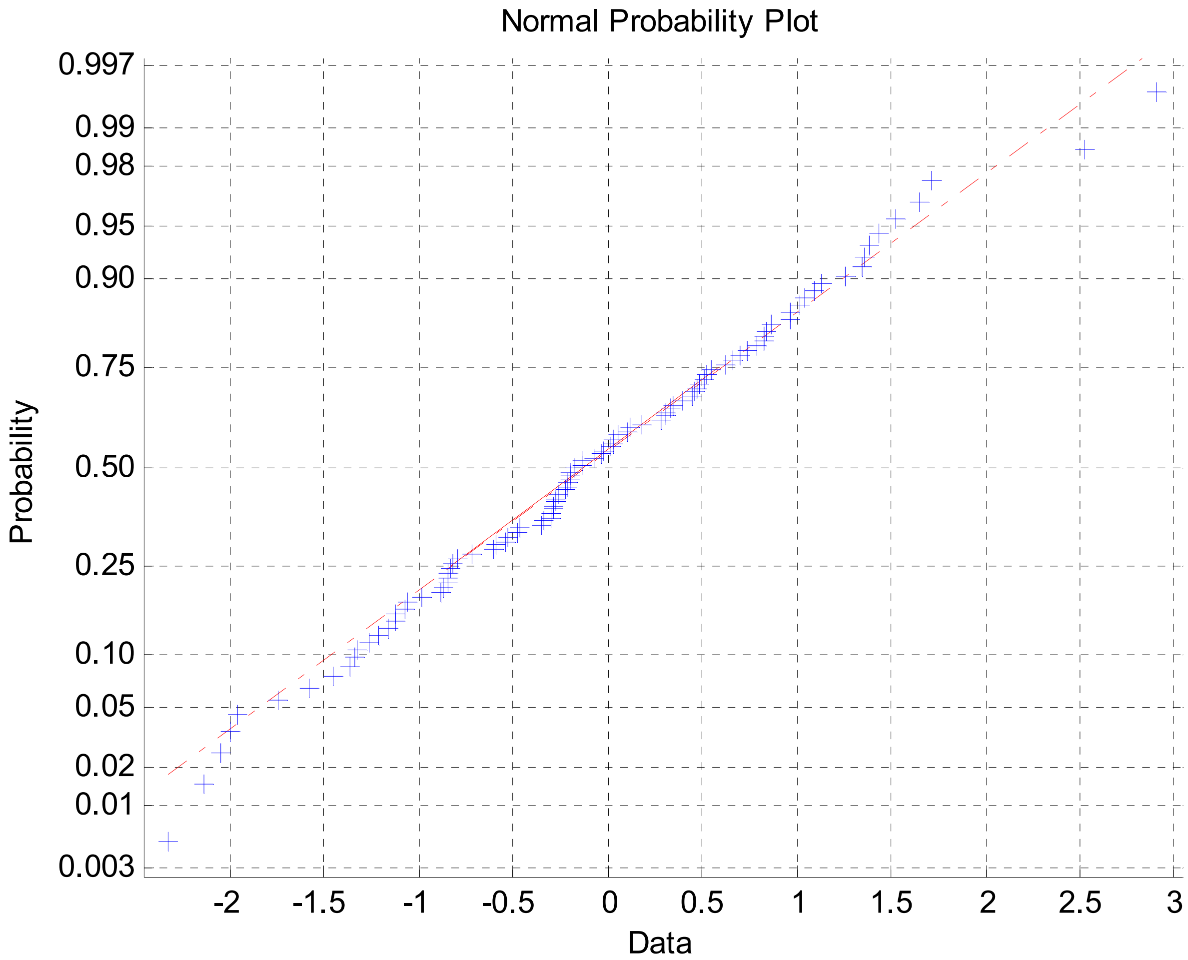
```
clear all;
X=[1.2 3.4 4.5 6.3 7.1 8.2 9.5 10.3 13.6];
figure;
plot(X,'b*');
h=lsline
```



➢ 正态分布概率图

- 函数 **normplot**
- 格式
  - ➢ **normplot(X)**
    - 若**X**为向量，则显示正态分布概率图形，
    - 若**X**为矩阵，则显示每一列的正态分布概率图形。
  - ➢ **h = normplot(X)**
  - ➢ 返回绘图直线的句柄
- 说明
  - ➢ 样本数据在图中用“**+**”显示；
  - ➢ 如果数据来自正态分布，则图形显示为直线，而其它分布可能在图中产生弯曲。

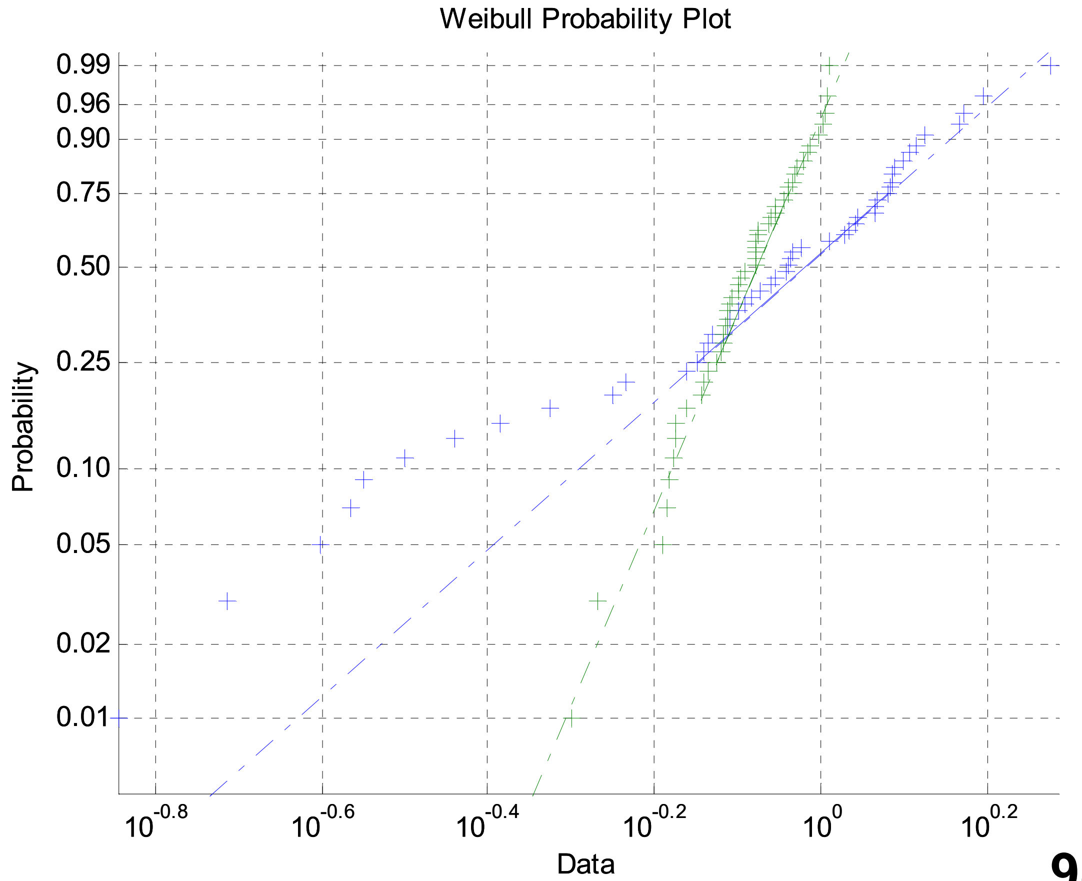
```
clear all;
X=normrnd(0, 1, 100, 1);
figure;
H=normplot(X)
```



➢ 威布尔(Weibull)概率图形

- 函数 **weibplot**
- 格式
  - ➢ **weibplot(X)**
    - 若**X**为向量，则显示威布尔**(Weibull)**概率图形，
    - 若**X**为矩阵，则显示每一列的威布尔概率图形。
  - ➢ **h = weibplot(X)**
    - 返回绘图直线的柄
- 说明
  - ➢ 绘制威布尔**(Weibull)**概率图形的目的是用图解法估计来自威布尔分布的数据**X**
  - ➢ 如果**X**是威布尔分布数据，其图形是直线的，否则图形中可能产生弯曲。

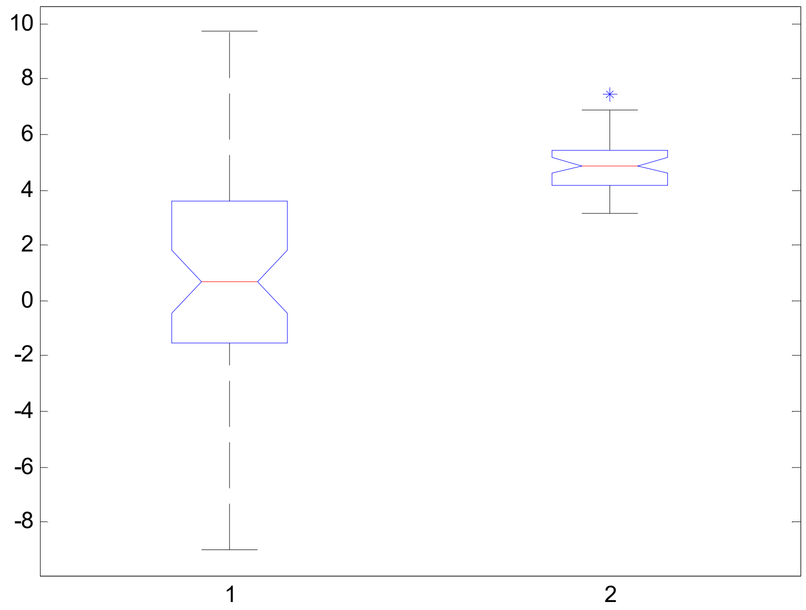
```
clear all;
a=weibrnd(1, 2, 50, 1);
b=weibrnd(3, 8, 50, 1);
X=[a b];
H=weibplot(X)
```

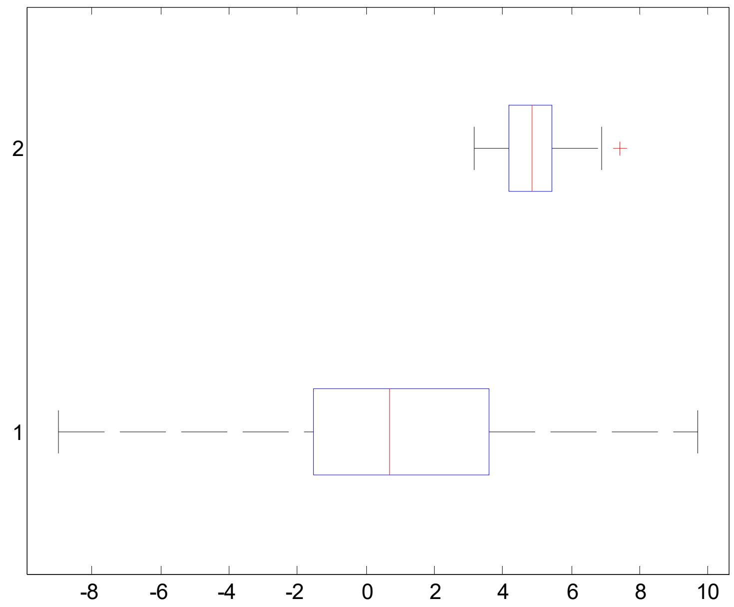


➢ 样本数据的盒图

- 函数 **boxplot**
- 格式
  - ➢ **boxplot(X)**
    - 产生矩阵**X**的每一列的盒图和“须”图，“须”是从盒的尾部延伸出来，并表示盒外数据长度的线，如果“须”的外面没有数据，则在“须”的底部有一个点。
  - ➢ **boxplot(X, notch)**
    - 当**notch=1**时，产生一凹盒图，**notch=0**时产生一矩箱图。
  - ➢ **boxplot(X, notch, 'sym')**
    - **sym**表示图形符号，默认值为“**+**”。
  - ➢ **boxplot(X, notch, 'sym', vert)**
    - 当**vert=0**时，生成水平盒图，**vert=1**时，生成竖直盒图（默认值**vert=1**）。
  - ➢ **boxplot(X, notch, 'sym', vert, whis)**
    - **whis**定义“须”图的长度，默认值为**1.5**，若**whis=0**则**boxplot**函数通过绘制**sym**符号图来显示盒外的所有数据值。

```
clear all;
a=normrnd(1, 4, 50, 1);
b=normrnd(5, 1, 50, 1);
x=[a b];
figure;
boxplot(x, 'notch', 'on', 'symbol', 'b*', 'orientation', 'vertical');
figure;
boxplot(x, 'notch', 'off', 'symbol', 'r+', 'orientation', 'horizontal')
```

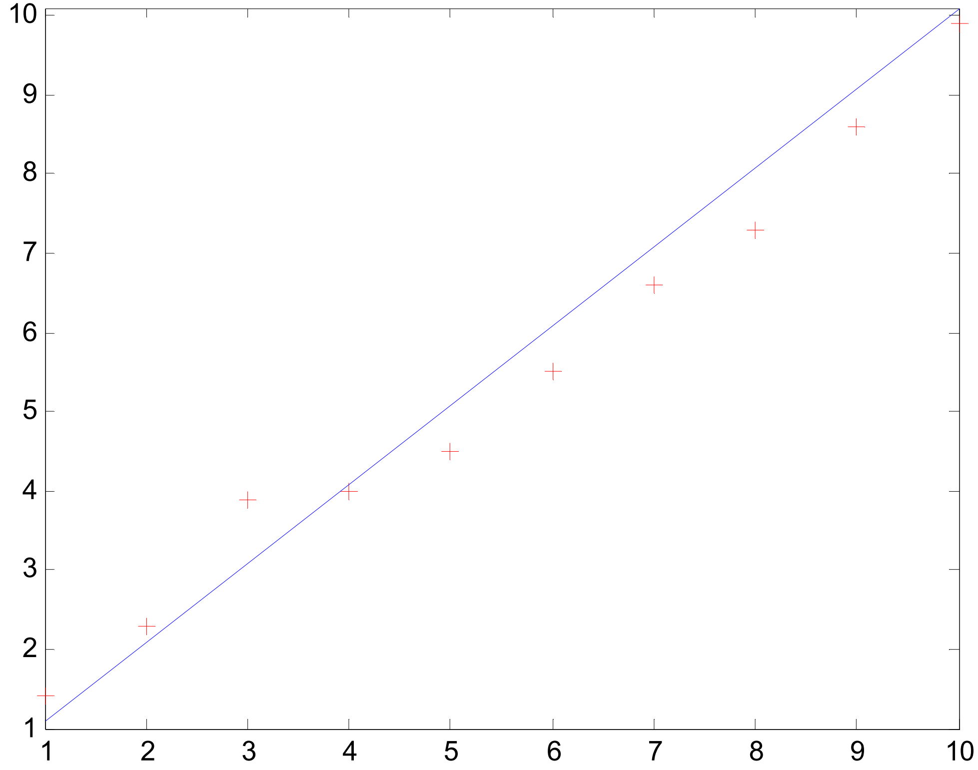




➢ 给当前图形加一条参考线

- 函数 **refline**
- 格式
  - ➢ **refline(slope, intercept)**
    - **slope**表示直线斜率，**intercept**表示截距
  - ➢ **refline(slope)**
    - **slope=[a b]**，图中加一条直线：**y=b+ax**。

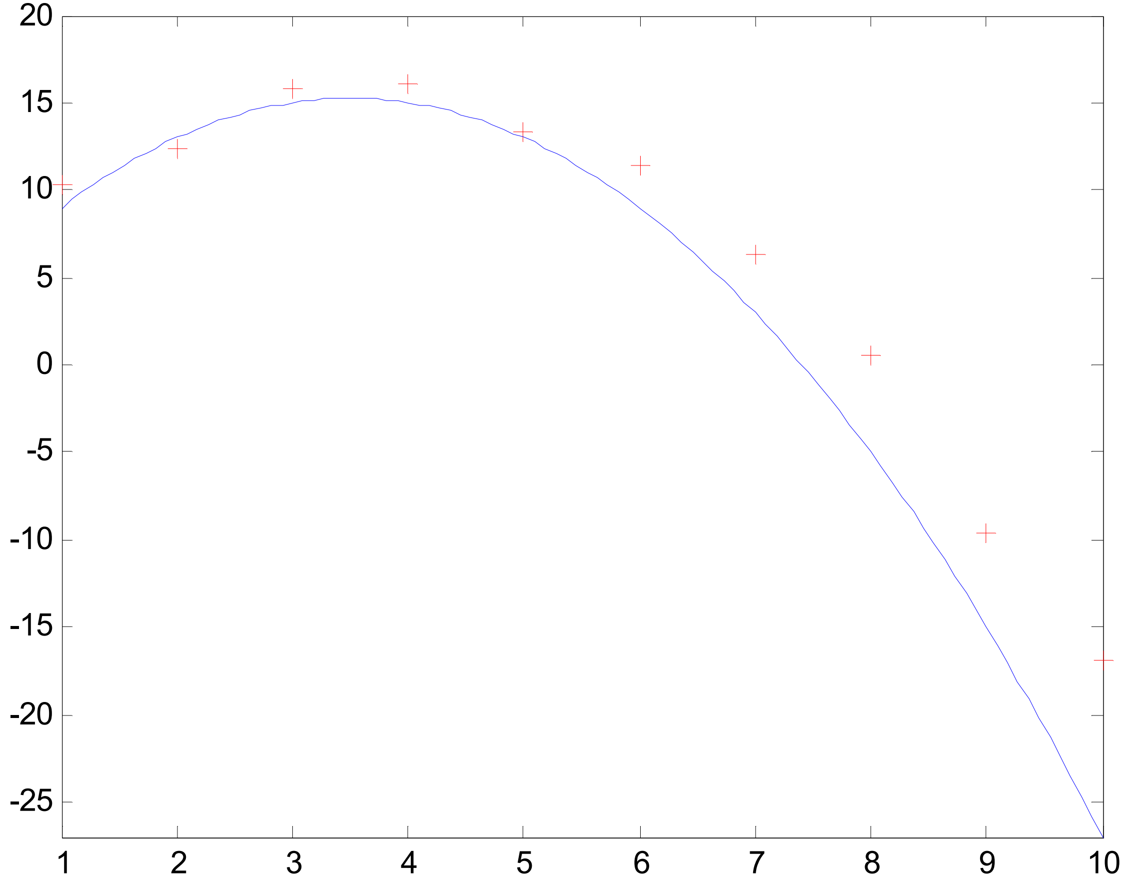
```
clear all;
X=[1.4 2.3 3.9 4.0 4.5 5.5 6.6 7.3 8.6 9.9];
figure;
plot(X, 'r+');
h=refline(1, 0.1)
```



➢ 在当前图形中加入一条多项式曲线

- 函数 **refcurve**
- 格式
  - ➢ **h = refcurve(p)**
    - 在图中加入一条多项式曲线，
    - **h**为曲线的环柄，
    - **p**为多项式系数向量，**p=[p1, p2, p3, …, pn]**，其中**p1**为最高幂项系数。

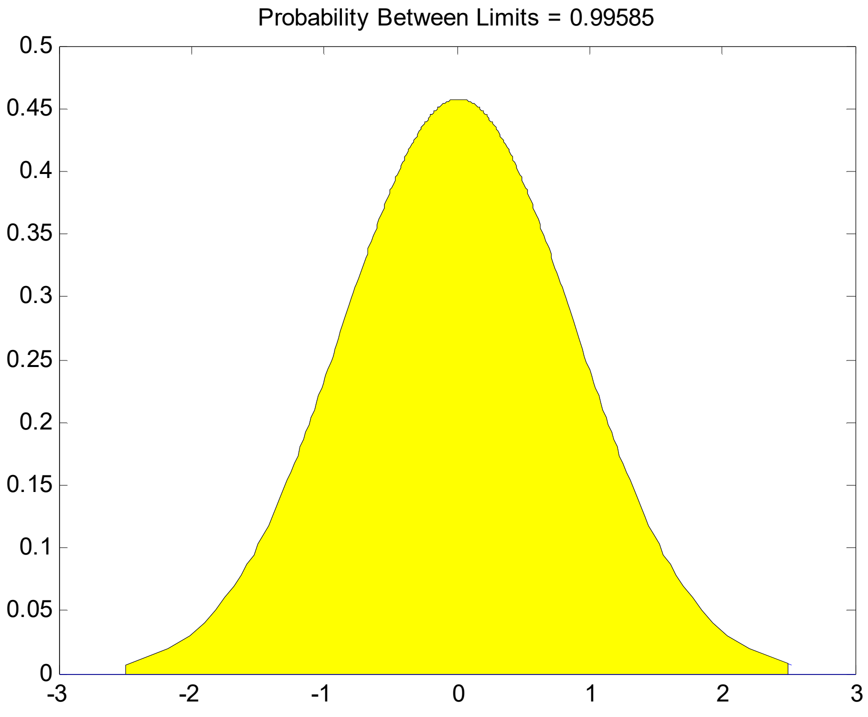
```
clear all;
X=[10.3 12.4 15.8 16.1 13.4 11.4 6.3 0.6 -9.6 -16.9];
figure;
plot(X, 'r+');
p=[-1 7 3];
h=refcurve(p)
```

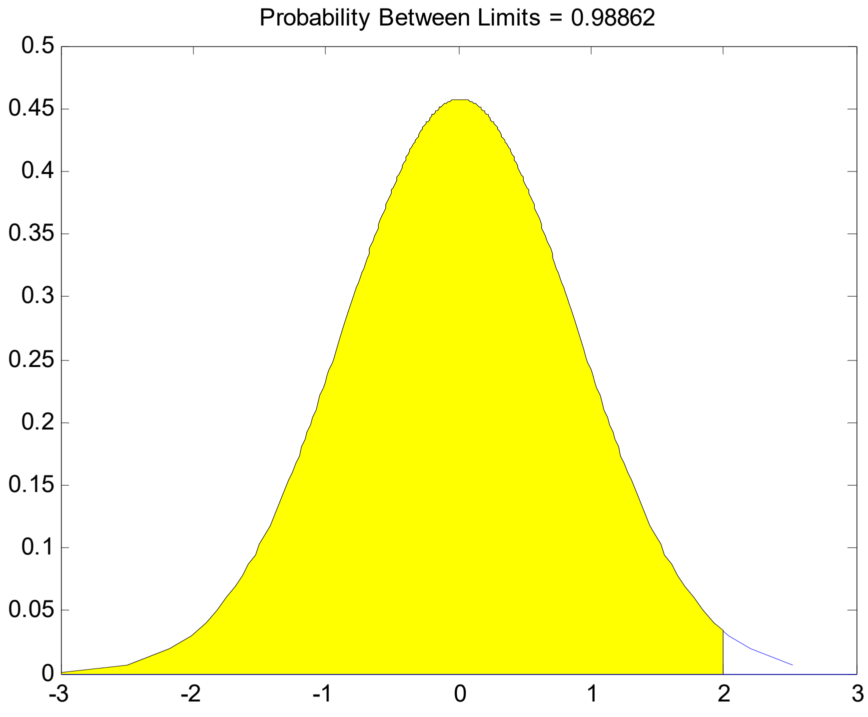


➢ 样本的概率图形

- 函数 **capaplot**
- 格式
  - ➢ **p = capaplot(data, specs)**
    - **data**为所给样本数据，
    - **specs**指定范围，
    - **p**表示在指定范围内的概率。
- 说明
  - ➢ 该函数返回来自于估计分布的随机变量落在指定范围内的概率

```
clear all;
data=normrnd(0, 1, 100, 1);
s1=[-2.5, +2.5];
s2=[-inf, 2];
figure;
[p, h]=capaplot(data, s1)
figure;
[p, h]=capaplot(data, s2)
```

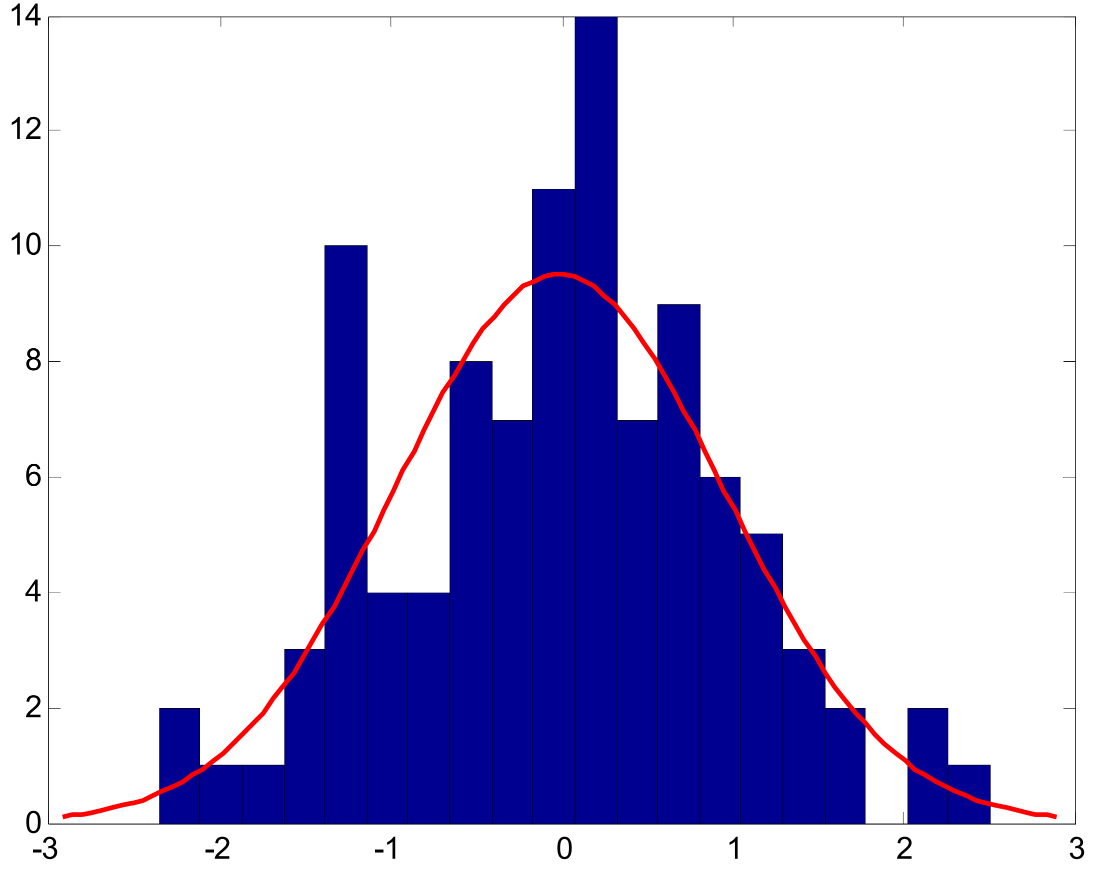




➢ 附加有正态密度曲线的直方图

- 函数 **histfit**
- 格式
  - ➢ **histfit(data)**
    - **data**为向量，返回直方图和正态曲线。
  - ➢ **histfit(data, nbins)**
    - **nbins**指定**bar**的个数，缺省时为**data**中数据个数的平方根。

```
clear all;
data=normrnd(0, 1, 100, 1);
figure;
h=histfit(data, 20)
```



为了研究400m赛跑后学生心率变化情况，体育老师统计了全班44名同学在赛跑后 1 分钟内的脉搏次数，结果如下：

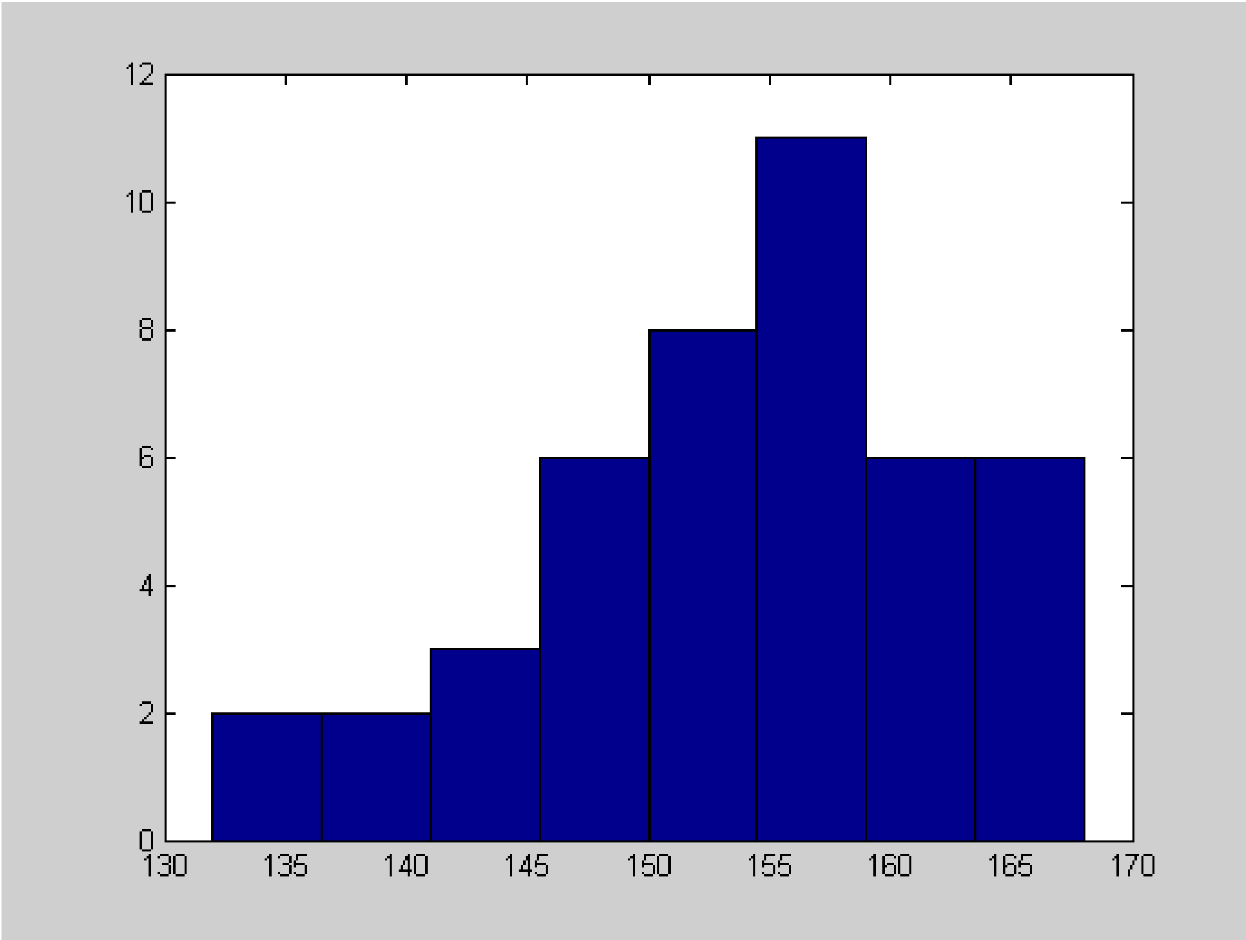
132	136	138	141	143	144	144	146
146	147	148	149	149	151	151	152
153	153	154	154	154	156	156	157
157	157	158	158	158	159	159	159
161	161	162	162	163	163	164	164
164	164	166	168				

按组距为 5 绘制频数分布直方图。

解：由给定数据可知，最小数据为132，最大数据为168，按组距为5，可取区间[130， 170]分为8等分，输入命令如下：

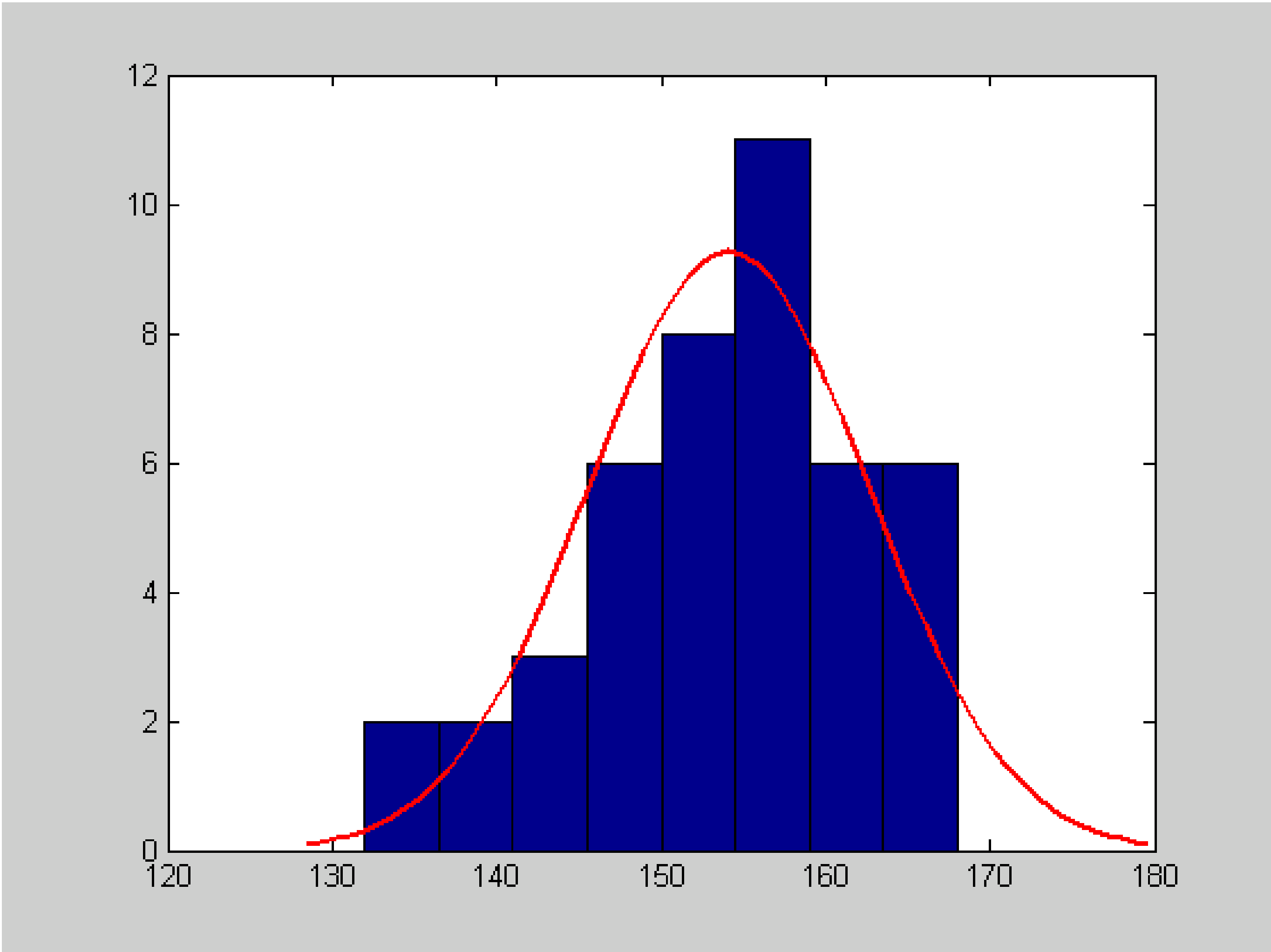
a=[132 136 138 141 143 144 144 146 146 147 148 149 149 151 151 152 153 153 154 154 154 156 156 157 157 157 158 158 158 159 159 159 161 161 162 162 163 163 164 164 164 164 166 168];

hist(a, 8)



histfit(a, 8)

可得附加有正态密度曲线的频数直方图



➢ 在指定的界线之间画正态密度曲线

- 函数 normspec
- 格式
  - ➢ p = normspec(specs, mu, sigma)
    - specs指定界线，
    - mu, sigma为正态分布的参数
    - p为样本落在上、下界之间的概率。

```
clear all;
s=[-inf  7];
mu=2;
sigma=4;
[p,h]=normspec(s, mu, sigma)
```