

概率统计

本章介绍 MATLAB 在概率统计中的若干命令和使用格式，这些命令存放于 Statistics Toolbox 中。这部分函数的主要形式为：

- 通用函数：函数名('name',参数表)
- 专用函数：Name 函数名（参数表）

常见分布函数名

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

第 1 节 随机数的产生

一、通用函数

命令 求指定分布的随机数

函数 random

格式 $y = \text{random}(\text{'name'}, A1, A2, A3, m, n)$ %name 的取值见上表；A1, A2, A3 为分布的参数；m, n 指定随机数的行和列

例：产生 12（3 行 4 列）个均值为 2，标准差为 0.3 的正态分布随机数

```
>> y=random('norm',2,0.3,3,4)
```

y =

```
2.3567    2.0524    1.8235    2.0342
1.9887    1.9440    2.6550    2.3200
2.0982    2.2177    1.9591    2.0178
```

二、常用的专用函数

1. 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 $R = \text{normrnd}(\text{MU}, \text{SIGMA})$ %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。

$R = \text{normrnd}(\text{MU}, \text{SIGMA}, m)$ %m 指定随机数的个数, 与 R 同维数。

$R = \text{normrnd}(\text{MU}, \text{SIGMA}, m, n)$ %m,n 分别表示 R 的行数和列数

例

>> $R = \text{normrnd}(10, 0.5, [2, 3])$ %mu 为 10, sigma 为 0.5 的 2 行 3 列个正态随机数

R =

```
9.7837    10.0627    9.4268
9.1672    10.1438    10.5955
```

2. 常见分布的随机数产生函数

常见分布的随机数的使用格式与上面相同

随机数产生专用函数表

函数名	调用形式	注 释
Unifrnd	$\text{unifrnd}(A, B, m, n)$	[A,B]上均匀分布(连续) 随机数
Unidrnd	$\text{unidrnd}(N, m, n)$	均匀分布(离散) 随机数
Exprnd	$\text{exprnd}(\text{Lambda}, m, n)$	参数为 Lambda 的指数分布随机数
Normrnd	$\text{normrnd}(\text{MU}, \text{SIGMA}, m, n)$	参数为 MU, SIGMA 的正态分布随机数
chi2rnd	$\text{chi2rnd}(N, m, n)$	自由度为 N 的卡方分布随机数
Trnd	$\text{trnd}(N, m, n)$	自由度为 N 的 t 分布随机数
Frnd	$\text{frnd}(N_1, N_2, m, n)$	第一自由度为 N_1 , 第二自由度为 N_2 的 F 分布随机数
Gamrnd	$\text{gamrnd}(A, B, m, n)$	参数为 A, B 的 γ 分布随机数
Betarnd	$\text{betarnd}(A, B, m, n)$	参数为 A, B 的 β 分布随机数
Lognrnd	$\text{lognrnd}(\text{MU}, \text{SIGMA}, m, n)$	参数为 MU, SIGMA 的对数正态分布随机数
Nbinrnd	$\text{nbinrnd}(R, P, m, n)$	参数为 R, P 的负二项式分布随机数
Ncfrnd	$\text{ncfrnd}(N_1, N_2, \text{delta}, m, n)$	参数为 N_1, N_2, delta 的非中心 F 分布随机数
Nctrnd	$\text{nctrnd}(N, \text{delta}, m, n)$	参数为 N, delta 的非中心 t 分布随机数
Ncx2rnd	$\text{ncx2rnd}(N, \text{delta}, m, n)$	参数为 N, delta 的非中心卡方分布随机数
Raylrnd	$\text{raylrnd}(B, m, n)$	参数为 B 的瑞利分布随机数
Weibrnd	$\text{weibrnd}(A, B, m, n)$	参数为 A, B 的韦伯分布随机数
Binornd	$\text{binornd}(N, p, m, n)$	参数为 N, p 的二项分布随机数
Geornd	$\text{geornd}(p, m, n)$	参数为 p 的几何分布随机数
Hygernd	$\text{hygernd}(M, K, N, m, n)$	参数为 M, K, N 的超几何分布随机数
Poissrnd	$\text{poissrnd}(\text{Lambda}, m, n)$	参数为 Lambda 的泊松分布随机数

第 2 节 随机变量的概率密度函数

一、通用函数

命令 通用函数计算概率密度函数值

函数 pdf

格式 $Y = \text{pdf}(\text{name}, K, A)$

$Y = \text{pdf}(\text{name}, K, A, B)$

$Y = \text{pdf}(\text{name}, K, A, B, C)$

说明 返回在 $X=K$ 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, 其取值如表 4-2。

例如二项分布: 设一次试验, 事件 A 发生的概率为 p, 那么, 在 n 次独立重复试验中,

事件 A 恰好发生 K 次的概率 P_K 为: $P_K=P\{X=K\}= \text{pdf}(\text{'bino'},K,n,p)$

例: 计算正态分布 $N(0, 1)$ 的随机变量 X 在点 0.6578 的密度函数值。

解: `>> pdf('norm',0.6578,0,1)`

```
ans =  
0.3213
```

二、常见的专用函数

1. 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 A 发生的概率; K —事件 A 发生 K 次; n —试验总次数

2. 泊松分布的概率值

函数 `poisspdf`

格式 `poisspdf(k, Lambda)` %等同于 `pdf('poiss',K,Lamda)`

3. 正态分布的概率值

函数 `normpdf(K,mu,sigma)` %计算参数为 $\mu = \text{mu}$, $\sigma = \text{sigma}$ 的正态分布密度函数在 K 处的值

4. 计算概率密度函数的专用函数

专用函数计算概率密度函数表

函数名	调用形式	注 释
Unifpdf	<code>unifpdf(x, a, b)</code>	[a,b]上均匀分布(连续)概率密度在 $X=x$ 处的函数值
unidpdf	<code>unidpdf(x,n)</code>	均匀分布(离散)概率密度函数值
Exppdf	<code>exppdf(x, Lambda)</code>	参数为 Lambda 的指数分布概率密度函数值
normpdf	<code>normpdf(x, mu, sigma)</code>	参数为 mu , sigma 的正态分布概率密度函数值
chi2pdf	<code>chi2pdf(x, n)</code>	自由度为 n 的卡方分布概率密度函数值
Tpdf	<code>tpdf(x, n)</code>	自由度为 n 的 t 分布概率密度函数值
Fpdf	<code>fpdf(x, n1, n2)</code>	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	<code>gampdf(x, a, b)</code>	参数为 a, b 的 γ 分布概率密度函数值
betapdf	<code>betapdf(x, a, b)</code>	参数为 a, b 的 β 分布概率密度函数值
lognpdf	<code>lognpdf(x, mu, sigma)</code>	参数为 mu , sigma 的对数正态分布概率密度函数值
nbinpdl	<code>nbinpdl(x, R, P)</code>	参数为 R, P 的负二项式分布概率密度函数值
Ncfpdf	<code>ncfpdl(x, n1, n2, delta)</code>	参数为 n_1, n_2, delta 的非中心 F 分布概率密度函数值
Nctpdf	<code>nctpdf(x, n, delta)</code>	参数为 n, delta 的非中心 t 分布概率密度函数值
ncx2pdf	<code>ncx2pdf(x, n, delta)</code>	参数为 n, delta 的非中心卡方分布概率密度函数值
raylpdl	<code>raylpdl(x, b)</code>	参数为 b 的瑞利分布概率密度函数值
weibpdf	<code>weibpdf(x, a, b)</code>	参数为 a, b 的韦伯分布概率密度函数值
binopdl	<code>binopdl(x,n,p)</code>	参数为 n, p 的二项分布的概率密度函数值
geopdl	<code>geopdl(x,p)</code>	参数为 p 的几何分布的概率密度函数值
hygepdf	<code>hygepdf(x,M,K,N)</code>	参数为 M, K, N 的超几何分布的概率密度函数值
poisspdf	<code>poisspdf(x,Lambda)</code>	参数为 Lambda 的泊松分布的概率密度函数值

第 3 节 随机变量的分布函数

一、通用函数

命令 通用函数 `cdf` 用来计算随机变量 $X \leq K$ 的概率之和(累积概率值)

函数 `cdf`

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 $X \leq K$ 的概率之和的累积概率值

例 求标准正态分布随机变量 X 落在区间 $(-\infty, 0.4)$ 内的概率（该值就是概率统计教材中的附表：标准正态数值表）。

解：

```
>> cdf('norm', 0.4, 0, 1)
ans =
    0.6554
```

二、专用函数

1. 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。

2. 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

例 设 $X \sim N(3, 2^2)$

(1) 求 $P\{2 < X < 5\}$, $P\{-4 < X < 10\}$, $P\{|X| > 2\}$, $P\{X > 3\}$

(2) 确定 c, 使得 $P\{X > c\} = P\{X < c\}$

解 (1) $p1 = P\{2 < X < 5\}$
 $p2 = P\{-4 < X < 10\}$
 $p3 = P\{|X| > 2\} = 1 - P\{|X| \leq 2\}$
 $p4 = P\{X > 3\} = 1 - P\{X \leq 3\}$

则有：

```
>> p1 = normcdf(5, 3, 2) - normcdf(2, 3, 2)
p1 =
    0.5328
>> p2 = normcdf(10, 3, 2) - normcdf(-4, 3, 2)
p2 =
    0.9995
>> p3 = 1 - normcdf(2, 3, 2) - normcdf(-2, 3, 2)
p3 =
    0.6853
>> p4 = 1 - normcdf(3, 3, 2)
p4 =
    0.5000
```

3. 专用函数

专用函数的概率分布函数

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 $F(x)=P\{X\leq x\}$
unidedcdf	unidedcdf(x,n)	均匀分布(离散)累积分布函数值 $F(x)=P\{X\leq x\}$
expcdf	expcdf(x, Lambda)	参数为 Lambda 的指数分布累积分布函数值 $F(x)=P\{X\leq x\}$
normcdf	normcdf(x, mu, sigma)	参数为 mu, sigma 的正态分布累积分布函数值 $F(x)=P\{X\leq x\}$
chi2cdf	chi2cdf(x, n)	自由度为 n 的卡方分布累积分布函数值 $F(x)=P\{X\leq x\}$
tcdf	tcdf(x, n)	自由度为 n 的 t 分布累积分布函数值 $F(x)=P\{X\leq x\}$
fcdf	fcdf(x, n1, n2)	第一自由度为 n1, 第二自由度为 n2 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 $F(x)=P\{X\leq x\}$
betacdf	betacdf(x, a, b)	参数为 a, b 的 β 分布累积分布函数值 $F(x)=P\{X\leq x\}$
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 的负二项式分布累积分布函数值 $F(x)=P\{X\leq x\}$
ncfcdf	ncfcdf(x, n1, n2, delta)	参数为 n1, n2, 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 $F(x)=P\{X\leq x\}$
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 $F(x)=P\{X\leq x\}$
weibcdf	weibcdf(x, a, b)	参数为 a, b 的韦伯分布累积分布函数值 $F(x)=P\{X\leq x\}$
binocdf	binocdf(x, n, p)	参数为 n, p 的二项分布的累积分布函数值 $F(x)=P\{X\leq x\}$
geocdf	geocdf(x, p)	参数为 p 的几何分布的累积分布函数值 $F(x)=P\{X\leq x\}$
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 $F(x)=P\{X\leq x\}$

说明 累积概率函数就是分布函数 $F(x)=P\{X\leq x\}$ 在 x 处的值。

第 4 节 随机变量的逆累积分布函数

MATLAB 中的逆累积分布函数是已知 $F(x) = P\{X \leq x\}$, 求 x。

一、通用函数

命令 **icdf** 计算逆累积分布函数

格式 **icdf('name', P, a1, a2, a3)**

说明 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

如果 $P = \text{cdf}('name', x, a_1, a_2, a_3)$, 则 $x = \text{icdf}('name', P, a_1, a_2, a_3)$

例 在标准正态分布表中, 若已知 $\Phi(x) = 0.975$, 求 x

解: `>> x=icdf('norm',0.975,0,1)`

x =

1.9600

二、专用函数

1.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: $p=P\{X\leq x\}$ 。

例 设 $X \sim N(3, 2^2)$, 确定 c 使得 $P\{X > c\} = P\{X < c\}$ 。

解: 由 $P\{X > c\} = P\{X < c\}$ 得, $P\{X > c\} = P\{X < c\} = 0.5$, 所以

`>>X=norminv(0.5,3,2)`

X=

3

2.常用专用函数

常用临界值函数表

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

例 4-28 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 $N(175, 36)$, 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 $P\{X > h\} \leq 0.01$ 的 h , 即 $P\{X < h\} \geq 0.99$, 所以

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

```
h =
```

```
188.9581
```

第 5 节 随机变量的数字特征

一、均值

1. 利用 mean 求算术平均值

格式 $\text{mean}(X)$ % X 为向量, 返回 X 中各元素的平均值

$\text{mean}(A)$ % A 为矩阵, 返回 A 中各列元素的平均值构成的向量

$\text{mean}(A, \text{dim})$ %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 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, 即样本均值。

例

```
>> A=[1 3 4 5;2 3 4 6;1 3 1 5]
```

```
A =
```

```
1     3     4     5
```

```
2     3     4     6
```

```
1     3     1     5
```

```
>> mean(A)
```

```
ans =
```

```
1.3333    3.0000    3.0000    5.3333
```

2. 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。

例

```
>> A=[1 2 3;nan 5 2;3 7 nan]
A =
     1     2     3
    NaN     5     2
     3     7    NaN
>> nanmean(A)
ans =
    2.0000    4.6667    2.5000
```

二、数据比较

1. 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。

例

```
>> A=[1 2 3;4 5 2;3 7 0]
A =
     1     2     3
     4     5     2
     3     7     0
>> sort(A)
ans =
     1     2     0
     3     5     2
     4     7     3
>> [Y,I]=sort(A)
Y =
     1     2     0
     3     5     2
     4     7     3
I =
     1     1     3
     3     2     2
     2     3     1
```

2. 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。

例

```
>> A=[1 2 3;4 5 2;3 7 0]
```

```
A =
```

```
     1     2     3
     4     5     2
     3     7     0
```

```
>> sortrows(A)
```

```
ans =
```

```
     1     2     3
     3     7     0
     4     5     2
```

3. 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。

例 4-38

```
>> A=[1 2 3;4 5 2;3 7 0]
```

```
A =
```

```
     1     2     3
     4     5     2
     3     7     0
```

```
>> Y=range(A)
```

```
Y =
```

```
     3     5     3
```

三、数学期望

1. 计算样本均值

函数 **mean**

格式 用法与前面一样

例: 随机抽取 6 个滚珠测得直径如下: (直径: mm)

```
14.70 15.21 14.90 14.91 15.32 15.32
```

试求样本平均值

解: `>>X=[14.70 15.21 14.90 14.91 15.32 15.32];`

```
>>mean(X)     %计算样本均值
```

则结果如下:

```
ans =
```

```
15.0600
```

2. 由分布律计算均值

利用 **sum** 函数计算: 格式 **sum(A)**

例: 设随机变量 X 的分布律为:

X	-2	-1	0	1	2
P	0.3	0.1	0.2	0.1	0.3

求 $E(X)$ $E(X^2-1)$

解：在 Matlab 编辑器中建立 M 文件如下：

```
X=[-2 -1 0 1 2];
p=[0.3 0.1 0.2 0.1 0.3];
EX=sum(X.*p)
Y=X.^2-1
EY=sum(Y.*p)
```

运行后结果如下：

```
EX =
    0
Y =
    3    0   -1    0    3
EY =
   1.6000
```

四、方差

1. 求样本方差

函数 var

格式 $D=\text{var}(X)$ % $\text{var}(X)=s^2=\frac{1}{n-1}\sum_{i=1}^n(x_i-\bar{X})^2$,若 X 为向量, 则返回向量的样本方差。

$D=\text{var}(A)$ % A 为矩阵, 则 D 为 A 的列向量的样本方差构成的行向量。

$D=\text{var}(X, 1)$ %返回向量(矩阵) X 的简单方差(即置前因子为 $\frac{1}{n}$ 的方差)

$D=\text{var}(X, w)$ %返回向量(矩阵) X 的以 w 为权重的方差

2. 求标准差

函数 std

格式 $\text{std}(X)$ %返回向量(矩阵) X 的样本标准差(置前因子为 $\frac{1}{n-1}$) 即:

$$\text{std} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}$$

$\text{std}(X, 1)$ %返回向量(矩阵) X 的标准差(置前因子为 $\frac{1}{n}$)

$\text{std}(X, 0)$ %与 $\text{std}(X)$ 相同

$\text{std}(X, \text{flag}, \text{dim})$ %返回向量(矩阵)中维数为 dim 的标准差值, 其中 $\text{flag}=0$

时, 置前因子为 $\frac{1}{n-1}$; 否则置前因子为 $\frac{1}{n}$ 。

例 4-41 求下列样本的样本方差和样本标准差, 方差和标准差

14.70 15.21 14.90 15.32 15.32

解:

```
>>X=[14.7 15.21 14.9 14.91 15.32 15.32];
>>DX=var(X,1) %方差
```

```

DX =
    0.0559
>>sigma=std(X,1)    %标准差
sigma =
    0.2364
>>DX1=var(X)        %样本方差
DX1 =
    0.0671
>>sigma1=std(X)      %样本标准差
sigma1 =
    0.2590

```

五、常见分布的期望和方差专用函数（stat 结尾）

1. 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[A,B]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。

例

```

>>a = 1:6; b = 2.*a;
>>[M,V] = unifstat(a,b)
M =
    1.5000    3.0000    4.5000    6.0000    7.5000    9.0000
V =
    0.0833    0.3333    0.7500    1.3333    2.0833    3.0000

```

2. 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=SIGMA²。

例

```

>> n=1:4;
>> [M,V]=normstat(n'*n,n'*n)
M =
     1     2     3     4
     2     4     6     8
     3     6     9    12
     4     8    12    16
V =
     1     4     9    16
     4    16    36    64
     9    36    81   144
    16    64   144   256

```

3. 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。

例

```
>>n = logspace(1,5,5)
n =
    10    100   1000  10000 100000

>>[M,V] = binostat(n,1./n)
M =
    1     1     1     1     1
V =
    0.9000    0.9900    0.9990    0.9999    1.0000

>>[m,v] = binostat(n,1/2)
m =
     5     50    500   5000  50000
v =
1.0e+04 *
    0.0003    0.0025    0.0250    0.2500    2.5000
```

4. 常见分布的期望和方差函数

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

第 6 节 参数估计

一、通用函数

函数 mle

格式 `phat=mle('dist', X)` %返回用 **dist** 指定分布的最大似然估计值
`[phat, pci]=mle('dist', X)` %置信度为 95%
`[phat, pci]=mle('dist', X, alpha)` %置信度由 **alpha** 确定
`[phat, pci]=mle('dist', X, alpha, pl)` %仅用于二项分布, **pl** 为试验次数。
说明 **dist** 为分布函数名, 如: **beta**(β 分布)、**binom**(二项分布)等, **X** 为数据样本,
alpha 为显著水平 α , $(1-\alpha)\times 100\%$ 为置信度。**phat** 为估计值, **pci** 为置信区间

例

```
>> X=binornd(20,0.75) %产生二项分布的随机数
X =
    16
>> [p,pci]=mle('binom',X,0.05,20) %求概率的估计值和置信区间, 置信度为 95%
p =
    0.8000
pci =
    0.5634
    0.9427
```

二、专用函数

1. 正态分布的参数估计

函数 `normfit`

格式 `[muhat,sigmahat,muci,sigmaci] = normfit(X)`

`[muhat,sigmahat,muci,sigmaci] = normfit(X,alpha)`

说明 **muhat,sigmahat** 分别为正态分布的参数 μ 和 σ 的估计值, **muci,sigmaci** 分别为置信区间, 其置信度为 $(1-\alpha)\times 100\%$; **alpha** 给出显著水平 α , 缺省时默认为 0.05, 即置信度为 95%。

例 分别使用金球和铂球测定引力常数

(1) 用金球测定观察值为: 6.683 6.681 6.676 6.678 6.679 6.672

(2) 用铂球测定观察值为: 6.661 6.661 6.667 6.667 6.664

设测定值总体为 $N(\mu, \sigma^2)$, μ 和 σ 为未知。对(1)、(2)两种情况分别求 μ 和 σ 的置信度为 0.9 的置信区间。

解: 建立 M 文件: **LX0833.m**

```
X=[6.683 6.681 6.676 6.678 6.679 6.672];
Y=[6.661 6.661 6.667 6.667 6.664];
[mu,sigma,muci,sigmaci]=normfit(X,0.1) %金球测定的估计
[MU,SIGMA,MUCI,SIGMACI]=normfit(Y,0.1) %铂球测定的估计
```

运行后结果显示如下:

```
mu =
    6.6782
sigma =
    0.0039
muci =
    6.6750
    6.6813
```

```

sigmaci =
    0.0026
    0.0081

MU =
    6.6640

SIGMA =
    0.0030

MUCI =
    6.6611
    6.6669

SIGMACI =
    0.0019
    0.0071

```

由上可知，金球测定的 μ 估计值为 6.6782，置信区间为[6.6750， 6.6813]；

σ 的估计值为 0.0039，置信区间为[0.0026， 0.0081]。

泊球测定的 μ 估计值为 6.6640，置信区间为[6.6611， 6.6669]；

σ 的估计值为 0.0030，置信区间为[0.0019， 0.0071]。

2.常见参数估计专用函数

参数估计函数

函数名	调用形式	函数说明
Binofit	PHAT= binofit(X, N) [PHAT, PCI] = binofit(X,N) [PHAT, PCI]= binofit (X, N, ALPHA)	二项分布的概率的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
Poissfit	Lambdahat=poissfit(X) [Lambdahat, Lambdaci] = poissfit(X) [Lambdahat, Lambdaci]= poissfit (X, ALPHA)	泊松分布的参数最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的 λ 参数和置信区间
normfit	[muhat,sigmahat,muci,sigmaci] = normfit(X) [muhat,sigmahat,muci,sigmaci] = normfit(X, ALPHA)	正态分布的最大似然估计，置信度为 95% 返回水平 α 的期望、方差值和置信区间
betafit	PHAT=betafit (X) [PHAT, PCI]= betafit (X, ALPHA)	返回 β 分布参数 a 和 b 的最大似然估计 返回最大似然估计值和水平 α 的置信区间
unifit	[ahat,bhat] = unifit(X) [ahat,bhat,ACI,BCI] = unifit(X) [ahat,bhat,ACI,BCI]=unifit(X, ALPHA)	均匀分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
expfit	muhat =expfit(X) [muhat,muci] = expfit(X) [muhat,muci] = expfit(X,alpha)	指数分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计和置信区间
gamfit	phat =gamfit(X) [phat,pci] = gamfit(X) [phat,pci] = gamfit(X,alpha)	γ 分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回最大似然估计值和水平 α 的置信区间
weibfit	phat = weibfit(X) [phat,pci] = weibfit(X) [phat,pci] = weibfit(X,alpha)	韦伯分布参数的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的参数估计及其区间估计
Mle	phat = mle('dist',data) [phat,pci] = mle('dist',data) [phat,pci] = mle('dist',data,alpha) [phat,pci] = mle('dist',data,alpha,p1)	分布函数名为 dist 的最大似然估计 置信度为 95%的参数估计和置信区间 返回水平 α 的最大似然估计值和置信区间 仅用于二项分布，p1 为试验总次数

说明 各函数返回已给数据向量 X 的参数最大似然估计值和置信度为 $(1-\alpha) \times 100\%$ 的置信区间。 α 的默认值为 0.05，即置信度为 95%。

第 7 节 假设检验

一、 σ^2 已知，单个正态总体的均值 μ 的假设检验 (U 检验法)

函数 ztest

格式 $h = \text{ztest}(x, m, \text{sigma})$ % x 为正态总体的样本, m 为均值 μ_0 , sigma 为标准差, 显著性水平为 0.05(默认值)

$h = \text{ztest}(x, m, \text{sigma}, \text{alpha})$ %显著性水平为 alpha

$[h, \text{sig}, \text{ci}, \text{zval}] = \text{ztest}(x, m, \text{sigma}, \text{alpha}, \text{tail})$ % sig 为观察值的概率, 当 sig 为小概率时则对原假设提出质疑, ci 为真正均值 μ 的 $1-\text{alpha}$ 置信区间, zval 为统计量的值。

说明 若 $h=0$, 表示在显著性水平 alpha 下, 不能拒绝原假设;

若 $h=1$, 表示在显著性水平 alpha 下, 可以拒绝原假设。

原假设: $H_0: \mu = \mu_0 = m$,

若 $\text{tail}=0$, 表示备择假设: $H_1: \mu \neq \mu_0 = m$ (默认, 双边检验);

$\text{tail}=1$, 表示备择假设: $H_1: \mu > \mu_0 = m$ (单边检验);

$\text{tail}=-1$, 表示备择假设: $H_1: \mu < \mu_0 = m$ (单边检验)。

例 某车间用一台包装机包装葡萄糖, 包得的袋装糖重是一个随机变量, 它服从正态分布。当机器正常时, 其均值为 0.5 公斤, 标准差为 0.015。某日开工后检验包装机是否正常, 随机地抽取所包装的糖 9 袋, 称得净重为 (公斤)

0.497, 0.506, 0.518, 0.524, 0.498, 0.511, 0.52, 0.515, 0.512

问机器是否正常?

解: 总体 μ 和 σ 已知, 该问题是当 σ^2 为已知时, 在水平 $\alpha = 0.05$ 下, 根据样本值判断 $\mu = 0.5$ 还是 $\mu \neq 0.5$ 。为此提出假设:

原假设: $H_0: \mu = \mu_0 = 0.5$

备择假设: $H_1: \mu \neq 0.5$

$\gg X = [0.497, 0.506, 0.518, 0.524, 0.498, 0.511, 0.52, 0.515, 0.512];$

$\gg [h, \text{sig}, \text{ci}, \text{zval}] = \text{ztest}(X, 0.5, 0.015, 0.05, 0)$

结果显示为

$h =$

1

$\text{sig} =$

0.0248 %样本观察值的概率

$\text{ci} =$

0.5014 0.5210 %置信区间, 均值 0.5 在此区间之外

$\text{zval} =$

2.2444 %统计量的值

结果表明: $h=1$, 说明在水平 $\alpha = 0.05$ 下, 可拒绝原假设, 即认为包装机工作不正常。

二、 σ^2 未知, 单个正态总体的均值 μ 的假设检验 (t 检验法)

函数 ttest

格式 $h = \text{ttest}(x, m)$ % x 为正态总体的样本, m 为均值 μ_0 , 显著性水平为 0.05

$h = \text{ttest}(x, m, \text{alpha})$ % alpha 为给定显著性水平

`[h,sig,ci] = ttest(x,m,alpha,tail)` %sig 为观察值的概率, 当 sig 为小概率时则对原假设提出质疑, ci 为真正均值 μ 的 $1-\alpha$ 置信区间。

说明 若 $h=0$, 表示在显著性水平 α 下, 不能拒绝原假设;
若 $h=1$, 表示在显著性水平 α 下, 可以拒绝原假设。

原假设: $H_0: \mu = \mu_0 = m$,

若 $\text{tail}=0$, 表示备择假设: $H_1: \mu \neq \mu_0 = m$ (默认, 双边检验);

$\text{tail}=1$, 表示备择假设: $H_1: \mu > \mu_0 = m$ (单边检验);

$\text{tail}=-1$, 表示备择假设: $H_1: \mu < \mu_0 = m$ (单边检验)。

例 某种电子元件的寿命 X (以小时计) 服从正态分布, μ 、 σ^2 均未知。现测得 16 只元件的寿命如下

159 280 101 212 224 379 179 264 222 362 168 250
149 260 485 170

问是否有理由认为元件的平均寿命大于 225 (小时) ?

解: 未知 σ^2 , 在水平 $\alpha = 0.05$ 下检验假设: $H_0: \mu < \mu_0 = 225$, $H_1: \mu > 225$

```
>> X=[159 280 101 212 224 379 179 264 222 362 168 250 149 260 485 170];
```

```
>> [h,sig,ci]=ttest(X,225,0.05,1)
```

结果显示为:

$h =$

0

$\text{sig} =$

0.2570

$\text{ci} =$

198.2321 Inf %均值 225 在该置信区间内

结果表明: $H=0$ 表示在水平 $\alpha = 0.05$ 下应该接受原假设 H_0 , 即认为元件的平均寿命不大于 225 小时。

三、两个正态总体均值差的检验 (t 检验)

两个正态总体方差未知但等方差时, 比较两正态总体样本均值的假设检验

函数 `ttest2`

格式 `[h,sig,ci]=ttest2(X,Y)` %X, Y 为两个正态总体的样本, 显著性水平为 0.05

`[h,sig,ci]=ttest2(X,Y,alpha)` %alpha 为显著性水平

`[h,sig,ci]=ttest2(X,Y,alpha,tail)` %sig 为当原假设为真时得到观察值的概率, 当 sig 为小概率时则对原假设提出质疑, ci 为真正均值 μ 的 $1-\alpha$ 置信区间。

说明 若 $h=0$, 表示在显著性水平 α 下, 不能拒绝原假设;

若 $h=1$, 表示在显著性水平 α 下, 可以拒绝原假设。

原假设: $H_0: \mu_1 = \mu_2$, (μ_1 为 X 为期望值, μ_2 为 Y 的期望值)

若 $\text{tail}=0$, 表示备择假设: $H_1: \mu_1 \neq \mu_2$ (默认, 双边检验);

$\text{tail}=1$, 表示备择假设: $H_1: \mu_1 > \mu_2$ (单边检验);

$\text{tail}=-1$, 表示备择假设: $H_1: \mu_1 < \mu_2$ (单边检验)。

例 在平炉上进行一项试验以确定改变操作方法的建议是否会增加钢的产率, 试验是在同一只平炉上进行的。每炼一炉钢时除操作方法外, 其他条件都尽可能做到相同。先用标准方法炼一炉, 然后用建议的新方法炼一炉, 以后交替进行, 各炼 10 炉, 其产率分别为

(1) 标准方法: 78.1 72.4 76.2 74.3 77.4 78.4 76.0 75.5 76.7 77.3

(2) 新方法: 79.1 81.0 77.3 79.1 80.0 79.1 79.1 77.3 80.2 82.1

设这两个样本相互独立, 且分别来自正态总体 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$, μ_1 、 μ_2 、 σ^2 均未知。问建议的新操作方法能否提高产率? (取 $\alpha=0.05$)

解: 两个总体方差不变时, 在水平 $\alpha=0.05$ 下检验假设: $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 < \mu_2$

```
>> X=[78.1 72.4 76.2 74.3 77.4 78.4 76.0 75.5 76.7 77.3];
```

```
>> Y=[79.1 81.0 77.3 79.1 80.0 79.1 79.1 77.3 80.2 82.1];
```

```
>> [h,sig,ci]=ttest2(X,Y,0.05,-1)
```

结果显示为:

```
h =
```

```
1
```

```
sig =
```

```
2.1759e-004 %说明两个总体均值相等的概率很小
```

```
ci =
```

```
-Inf -1.9083
```

结果表明: $H=1$ 表示在水平 $\alpha=0.05$ 下, 应该拒绝原假设, 即认为建议的新操作方法提高了产率, 因此, 比原方法好。