概率论和数理统计的 Matlab 实现

1　概　　述

自然界和社会上会发生各种各样的现象,其中有的现象在一定条件下是一定要发生的,有的则表现出一定的随机性，但总体上又有一定的规律可循。一般称前者为确定性事件，后者为不确定性事件(或称随机事件)。概率论和数理统计就是研究和揭示不确定事件统计规律性的一门数学学科。

作为一门实用性很强的数学分支，概率论和数理统计的理论和方法已经广泛应用于管理、经济、心理、教育、体育、医学、生物、化学、机械、水文、地质、林业、气象、工业生产、建筑、通讯、自动控制等几乎所有社会和科学技术领域。

Matlab6.0 的统计工具箱相对于前面一些版本，改进较大。目前已经可以与 SPSS、SAS 等软件的统计功能相媲美。具体而言，它包括下面几个方面的内容：

- 概率分布　给出了常见的 20 种概率分布类型的概率密度函数、累加分布函数（分布函数)、逆累加分布函数、参数估计函数、随机数生成函数和统计量计算函数。
- 参数估计　提供了多种分布类型分布参数及其置信区间的估计方法。
- 样本描述　提供了描述中心趋势和离中趋势的统计量函数，缺失数据条件下的样本描述方法以及其它一些统计量计算函数。
- 方差分析　包括单因子方差分析、双因子方差分析和多因子方差分析。
- 多元方差分析　包括单因素多元方差分析、分组聚类和多元比较等。
- 回归分析　包括多元线性回归（包括逐步回归)、岭回归、一般线性模型拟合、多项式拟合、稳健回归、响应面分析（包括二维响应面分析和多维响应面分析)、非线性回归。
- 假设检验　包括单样本 t 检验、双样本 t 检验和 z 检验。
- 分布的检验　包括 Jarque-Bera 正态性检验、Kolmogorov-smirnov 单样本检验、Kolmogorov-smirnov 双样本检验和 Lilliefors 正态性检验。
- 非参数检验　包括 friedman 检验、Kruskalwallis 检验、秩和检验、符号秩检验和符号检验。
- 判别分析
- 聚类分析
- 因子分析
- 统计过程控制　提供了常用的过程管理图和过程性能图。
- 试验设计　包括完全析因设计、不完全析因设计和 D-优化设计。
- 统计图　包括箱形图、经验累加分布函数图、误差条图、函数交互等值线图、交互画线、交互点标注、散点矩阵图、散点图、添加最小二乘拟合线、正态概率图、帕累托图、q-q 图、回归个案次序图、参考多项式曲线、添加参考线、交互插值等值线图和威布尔图。

2 概 率 分 布

试验得到的数据通常呈现一定的规律性，引入随机变量以后，可以将随机数据表达为随机变量的函数。常见的随机变量有离散型随机变量和连续型随机变量两种。

- 当变量全部可以取到的值是有限个或可列无限多个时，称为离散型随机变量。
- 如果对于随机变量 x 的分布函数 $F(x)$，存在非负函数 $f(x)$，使得对于任意实数 x 有

$$F(x)=\int_{-\infty}^{x} f(t)dt, \tag{1}$$

  则称 X 为连续型随机变量。

对应于离散型随机变量和连续型随机变量，有离散型概率分布函数和连续型概率分布函数。

2.1 概率密度函数

2.1.1 基本数学原理

对于离散型概率分布和连续型概率分布，二者的概率密度函数定义有所不同。上面（1）式中，函数 $f(x)$ 称为 X 的概率密度函数。该函数具有以下性质：

- $f(x)\geq 0.$
- $\int_{-\infty}^{\infty} f(x)dx=1.$
- $P\{x_1<X\leq x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2} f(x)dx,(x_1\leq x_2)$
- 若 f(x) 在点 x 处连续，则有 $F^{'}(x)=f(x)$。

对于离散型概率分布，则不称其为概率密度函数，而叫做概率分布或分布律。设离散型随机变量 x 所有可能取的值为 $x_k(k=1, 2, ...)$，x 取各个可能值的概率，即事件 $\{X=x_k\}$ 的概率为：

$$P\{X=x_k\}=p_k, \qquad k=1, 2, ...$$

p_k 即称为分布律。它有下面两个性质：

- $p_k\geq 0, k=1,2,...;$
- $\sum_{k=1}^{\infty} p_k=1.$

2.1.2 相关函数介绍

对于连续型概率分布函数，表 2 给出了对应函数的概率密度函数及其数学意义和调用

格式。下面选择正态分布概率密度函数和指数分布概率密度函数进行重点介绍。

normpdf 函数

功能：计算正态概率密度函数。

语法：Y = normpdf(X,MU,SIGMA)

描述：normpdf(X,MU,SIGMA) 计算参数为 MU 和 SIGMA 的数据 X 的正态概率密度函数。参数 SIGMA 必须为正。

正态概率密度函数的计算公式为：

$$y = f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

似然函数为概率密度函数，它被视为参数的函数。最大似然估计量（MLEs）是使 x 处的似然函数最小化时的值。

若 x 服从标准正态分布，则 x + μ 也服从均值为 μ，标准差为 σ 的正态分布。相反地，若 y 服从参数为 μ 和 σ 的正态分布，则 x = (y -μ)/σ 服从标准正态分布。

举例：

```
mu = [0:0.1:2];
[y i] = max(normpdf(1.5,mu,1));
MLE = mu(i)
MLE =
    1.5000
```

exppdf 函数

功能：计算指数概率密度函数。

语法：Y = exppdf(X,MU)

描述：exppdf(X,MU) 计算 X 处的参数为 MU 的指数概率密度函数。MU 参数必须为正。

指数概率密度函数为：

$$y = f(x|\mu) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$$

指数概率密度函数等价于第一个参数（a）等于 1 时的伽玛概率密度函数。

举例：

```
y = exppdf(5,1:5)
y =
    0.0067    0.0410    0.0630    0.0716    0.0736
y = exppdf(1:5,1:5)
y =
    0.3679    0.1839    0.1226    0.0920    0.0736
```

表 2 常见分布的概率密度

函数名	对应的分布	数 学 意 义	调 用 格 式
betapdf	贝塔分布	$y=f(x\|a,b)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}I_{(0,1)}(x)$ （$0<x<1$）	Y = betapdf(X,A,B)
binopdf	二项分布	$y=f(x\|n,p)=\binom{n}{x}p^{x}q^{(1-x)}I_{(0,1,\ldots,n)}(x)$	Y=binopdf(X,N,P)
chi2pdf	卡方分布	$y=f(x\|v)=\frac{x^{(v-2)/2}e^{-x/2}}{2^{v/2}\Gamma(v/2)}$	Y=chi2pdf(X,V)
exppdf	指数分布	$y=f(x\|\mu)=\frac{1}{\mu}e^{-\frac{x}{\mu}}$	Y=exppdf(X,MU)
fpdf	F 分布	$y=f(x\|v_1,v_2)=\frac{\Gamma\left[\frac{v_1+v_2}{2}\right]}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}}\frac{x^{\frac{v_1-2}{2}}}{\left[1+\left(\frac{v_1}{v_2}\right)x\right]^{\frac{v_1+v_2}{2}}}$	Y=fpdf(X,V1,V2)
gampdf	伽玛分布	$y=f(x\|a,b)=\frac{1}{b^{a}\Gamma(a)}x^{a-1}e^{-\frac{x}{b}}$	Y=gampdf(X,A,B)
geopdf	几何分布	$y=f(x\|p)=pq^{x}I_{(0,1,K)}(x)$ 其中，$q=1-p$	Y=geopdf(X,P)
hygepdf	超几何分布	$y=f(x\|M,K,n)=\frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$	Y=hygepdf(X,M,K,N)
normpdf	正态（高斯）分布	$y=f(x\|\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	Y=normpdf(X,MU,SIGMA)
lognpdf	对数正态分布	$y=f(x\|\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}$	Y=lognpdf(X,MU,SIGMA)
nbinpdf	负二项分布	$y=f(x\|r,p)=\binom{r+x+1}{x}p^{x}q^{x}I_{(0,1,\ldots)}(x)$ 其中，$q=1-p$	Y=nbinpdf(X,R,P)
ncfpdf	非中心F分布	假设随机变量 $\chi_1^2(\mu)$ 服从自由度为 f_1、非中心参数为 μ 的非中心卡方分布，χ_2^2 服从自由度为 f_2 的卡方分布，且 $\chi_1^2(\mu)$ 和 χ_2^2 相互独立，则随机变量 $F=\frac{\chi_1^2(\mu)/f_1}{\chi_2^2/f_2}$ 的分布称为自由度为（f_1，f_2）、非中心参数为 μ 的非中心 F 分布。	Y=ncfpdf(X,NU1,NU2,DELTA)
nctpdf	非中心t分布	如果 U 服从参数为 μ 和 1 的正态分布，$\chi_{(v)}^2$ 服从自由度为 v 的 χ^2 分布，并且 U 与 $\chi_{(v)}^2$ 相互独立，则称随机变量	Y=nctpdf(X,V,DELTA)

		$t(\mu)=\dfrac{U+\mu}{\sqrt{\chi^2_{(v)}/v}}$ 的分布为自由度为 v、非中心参数为 μ 的非中心 t 分布。	
ncx2pdf	非中心卡方分布	如果随机变量 Xi 服从参数为 μ_i（i=1, ..., v）和 σ^2 的正态分布，并且相互独立，则随机变量 $\chi^2(\mu)=(X_1^2+\cdots+X_v^2)/\sigma^2$ 所服从的分布称为自由度为 v、非中心参数为 $\mu^2=(\mu_1^2+\cdots+\mu_v^2)/\sigma^2$ 的非中心 χ^2 分布。	Y=ncx2pdf(X, V, DELTA)
poisspdf	泊松分布	$y=f(x\|\lambda)=\dfrac{\lambda^x}{x!}e^{-\lambda}I_{(0,1,K)}(x)$	Y=poisspdf(X, LAMBDA)
raylpdf	雷利分布	$y=f(x\|b)=\dfrac{x}{b^2}e^{\left(\frac{-x^2}{2b^2}\right)}$	Y=raylpdf(X, B)
tpdf	学生氏 t 分布	$y=f(x\|v)=\dfrac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\dfrac{1}{\sqrt{v\pi}}\dfrac{1}{\left(1+\frac{x^2}{v}\right)^{\frac{v+1}{2}}}$	Y=tpdf(X, V)
unidpdf	离散均匀分布	$y=f(x\|N)=\dfrac{1}{N}I_{(1,\ldots,N)}(x)$	Y=unidpdf(X, N)
unifpdf	连续均匀分布	$y=f(x\|a,b)=\dfrac{1}{b-a}I_{[a,b]}(x)$	Y=unifpdf(X, A, B)
weibpdf	威布尔分布	$y=f(x\|a,b)=abx^{b-1}e^{-ax^b}I_{(0,\infty)}(x)$	Y=weibpdf(X, A, B)

2.2 累加分布函数

2.2.1 基本数学原理

对于连续型随机变量，其分布函数的定义为：若 X 为随机变量，x 为任意实数，则函数

$$F(x)=p\{X\le x\}$$

称为 X 的分布函数。如果知道 X 的分布函数，就知道了 X 落在任一区间（x_1, x_2］上的概率。

分布函数 $F(x)$ 具有以下一些性质：

- $F(x)$ 是不减函数；
- $0\le F(x)\le 1,$ 且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,$$

$$F(\infty)=\lim_{x\to\infty}F(x)=1,$$

- $F(x+0)=F(x)$，即 $F(x)$ 是右连续的。

2.2.2 相关函数介绍

normcdf 函数

功能：计算累加正态分布函数。

语法：P = normcdf(X,MU,SIGMA)

描述：normcdf(X,MU,SIGMA) 计算服从参数为 MU 和 SIGMA 的正态分布数据 X 的累加分布函数。参数 SIGMA 必须为正。

累加正态分布函数为：

$$p=F(x|\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-(t-\mu)^2}{2\sigma^2}}dt$$

结果 p 为取自参数为 μ 和σ的正态分布的单个观测量落在区间(-∞ x]中的概率。

举例：

下面的例子求取自标准正态分布的一个观测量落在区间[-1 1]中的概率。

```
p = normcdf([-1 1]);
p(2) - p(1)
ans =
    0.6827
```

更一般地，若观测量取自参数为σ和 μ 的正态分布，则它落在该区间中的概率为 68%。

expcdf 函数

功能：计算累加指数分布函数。

语法：P = expcdf(X,MU)

描述：expcdf(X,MU) 计算参数为 MU 的数据 X 的累加指数分布函数。指数 MU 必须为正。

累加指数分布函数的计算公式为：

$$p=F(x|\mu)=\int_0^x\frac{1}{\mu}e^{-\frac{t}{\mu}}dt=1-e^{-\frac{x}{\mu}}$$

结果 p 为源于指数分布的单个观测量落在区间[0 x]中的概率。

举例：

指数为 μ 的指数分布数据的中值等于 μ*log(2)，下例进行演示。

```
mu = 10:10:60;
p = expcdf(log(2)*mu,mu)
p =
```

```
    0.5000    0.5000    0.5000    0.5000    0.5000    0.5000
```

下面计算指数分布随机变量小于或等于均值 μ 的概率。

```
mu = 1:6;
x = mu;
p = expcdf(x,mu)
p =
    0.6321    0.6321    0.6321    0.6321    0.6321    0.6321
```

表 3 中为常见分布的累加函数及其调用格式。

表 3 常见分布的累加函数

函数名	累加函数对应的分布	数 学 意 义	调 用 格 式
betacdf	贝塔分布	$p=F(x\|a,b)=\dfrac{1}{B(a,b)\int_0^x t^{a-1}(1-t)^{b-1}dt}$	P=betacdf(X,A,B)
binocdf	二项分布	$y=F(x\|n,p)=\sum_{i=0}^{x}\binom{n}{i}p^i q^{(1-i)}I_{(0,1,\ldots,n)}(i)$	Y=binocdf(X,N,P)
chi2cdf	卡方分布	$p=F(x\|v)=\int_0^x \dfrac{t^{(v-2)/2}e^{-t/2}}{2^{v/2}\Gamma(v/2)}dt$	P=chi2cdf(X,V)
expcdf	指数分布	$p=F(x\|\mu)=\int_0^x \dfrac{1}{\mu}e^{-\frac{x}{\mu}}dt=1-e^{-\frac{x}{\mu}}$	P=expcdf(X,MU)
fcdf	F 分布	$F(x\|v_1,v_2)=\int \dfrac{\Gamma\left[\frac{v_1+v_2}{2}\right]}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}\left(\dfrac{v_1}{v_2}\right)^{\frac{v_1}{2}}\dfrac{t^{\frac{v_1-2}{2}}}{\left[1+\left(\frac{v_1}{v_2}\right)\right]^{\frac{v_1+v_2}{2}}}dt$	P=fcdf(X,V1,V2)
gamcdf	伽玛分布	$p=F(x\|a,b)=\dfrac{1}{b^a\Gamma(a)}\int_0^x t^{a-1}e^{-\frac{1}{b}}dt$	P=gamcdf(X,A,B)
geocdf	几何分布	$y=F(x\|p)=\sum_{i=0}^{floor(x)}pq^i, q=1-p$	Y=geocdf(X,P)
hygecdf	超几何分布	$p=F(x\|M,K,N)=\sum_{i=0}^{x}\dfrac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}}$	P=hygecdf(X,M,K,N)
logncdf	对数正态分布	$p=f(x\|\mu,\sigma)=\dfrac{1}{x\sigma\sqrt{2\pi}}\int_0^x \dfrac{e^{\frac{-(\ln t-\mu)^2}{2\sigma^2}}}{t}dt$	P=logncdf(X,MU,SIGMA)
nbincdf	负二项分布	$y=F(x\|r,p)=\sum_{i=0}^{x}\binom{r+i+1}{i}p^r q^i I_{(0,1,\ldots)}(i)$	Y=nbincdf(X,R,P)
ncfcdf	非中心 F 分布	$F(x\|v_1,v_2,\delta)=\sum\left(\dfrac{\left(\frac{1}{2}\delta\right)^j}{j!}e^{-\frac{\delta}{2}}\right)I\left(\dfrac{\frac{v_1x}{v_2+v_1x}}{v_1/2+j},v_2/2\right)$	P=ncfcdf(X,NU1,NU2,DELTA)

nctcdf	非中心 t 分布	$\Pr((-t)<x<t\mid v,\delta)=\sum\left(\frac{\left(\frac{1}{2}\delta\right)^j}{j!}e^{-\frac{\delta^2}{2}}\right)\left(\frac{\left(\frac{x^2}{v+x^2}\right)}{v/2+j},v/2\right)$	P=nctcdf(X, NU, DELTA)
ncx2cdf	非中心卡方分布	$F(x\mid v,\delta)=\sum\left(\frac{\left(\frac{1}{2}\delta\right)^j}{j!}e^{-\frac{\delta}{2}}\right)\Pr[\chi^2_{v+2j}\le x]$	P=ncx2cdf(X, V, DELTA)
normcdf	正态（高斯）分布	$p=f(x\mid\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-(t-\mu)^2}{2\sigma^2}}dt$	P=normcdf(X, MU, SIGMA)
poisscdf	泊松分布	$p=F(x\mid\lambda)=e^{-\lambda}\sum_{i=0}^{floor(x)}\frac{\lambda^i}{i!}$	P=poisscdf(X, LAMBDA)
raylcdf	雷利分布	$y=F(x\mid b)=\int_0^x\frac{1}{b^2}e^{\left(\frac{-t^2}{2b^2}\right)}dt$	P=raylcdf(X, B)
tcdf	学生氏 t 分布	$p=F(x\mid v)=\int\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\frac{1}{\sqrt{v\pi}}\frac{1}{\left(1+\frac{t^2}{v}\right)^{\frac{v+1}{2}}}dt$	P=tcdf(X, V)
unidcdf	离散均匀分布	$p=F(x\mid N)=\frac{floor(x)}{N}I_{(1,\ldots,N)}(x)$	P=unidcdf(X, N)
unifcdf	连续均匀分布	$p=F(x\mid a,b)=\frac{x-a}{b-a}I_{[a,b]}(x)$	P=unifcdf(X, A, B)
weibcdf	威布尔分布	$p=F(x\mid a,b)=\int_0^x abt^{b-1}e^{-at^b}dt=1-e^{-ax^b}I_{(0,\infty)}(x)$	P=weibcdf(X, A, B)

2.3 参数估计

2.3.1 基本数学原理

参数估计的内容包括点估计和区间估计。

点估计是用单个数值作为参数的估计，常用的方法有矩法和极大似然法等。

- 矩法　某些情况下，待估参数往往是总体原点矩或原点矩的函数，此时可以用取自该总体的样本的原点矩或样本原点矩的函数值作为待估参数的估计，这种方法称为矩法。如，样本均值总是总体均值的矩估计量，样本方差总是总体方差的矩估计量，样本标准差总是总体标准差的矩估计量。
- 极大似然法　极大似然法是在待估参数的可能取值范围内进行挑选，使似然函数值（即样本取固定观察值或样本取值落在固定观察值邻域内的概率）最大的那个数值即为极大似然估计量。由于极大似然估计量得到的估计量通常不仅仅满足无偏性、有效性等基本条件，还能保证其为充分统计量。所以，在点估计和区间估计中，一般推荐使用极大似然法。

区间估计不仅仅给出参数的近似取值，还给出了该取值的误差范围。求参数的区间估计，首先要求出该参数的点估计，然后构造一个含有该参数的随机变量，并根据一定的置信水平求该估计值的误差范围。

2.3.2 相关函数介绍

normfit 函数

功能：对正态分布数据进行参数估计，求参数的置信区间。

语法：[muhat,sigmahat,muci,sigmaci] = normfit(X)

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

描述：[muhat,sigmahat,muci,sigmaci] = normfit(X) 对于给定的服从正态分布的数据矩阵 X，返回参数 μ 和 σ 的估计值 muhat 和 sigmahat。muci 和 sigmaci 为 95%置信区间。muci 和 sigmaci 向量分别有两行，其列数与数据矩阵 X 的列数相同。上下两行的数据分别为置信区间的下限和上限。

[muhat,sigmahat,muci,sigmaci] = normfit(X,alpha) 进行参数估计并计算 100（1-alpha）置信区间。如 alpha = 0.01 时，给出 99%置信区间。

举例：

本例中，数据为两列随机正态矩阵。两列都有 μ = 10 和 σ = 2。

```
r = normrnd(10,2,100,2);
[mu,sigma,muci,sigmaci] = normfit(r)
mu =
   10.1455   10.0527
sigma =
    1.9072    2.1256
muci =
    9.7652    9.6288
   10.5258   10.4766
sigmaci =
    1.6745    1.8663
    2.2155    2.4693
```

参见：

betafit, binofit, expfit, gamfit, poissfit, unifit, weibfit

betafit 函数

功能：对服从贝塔分布的数据进行参数估计，并计算置信区间。

语法：phat = betafit(x)

[phat,pci] = betafit(x,alpha)

描述：betafit 函数计算服从贝塔分布的数据 x 的参数的最大似然估计，有两个输出变量。该函数也可以以 2×2 矩阵的形式返回参数的置信区间，矩阵的第一列包括 A 参数的下界和上界，第二列包括 B 参数的下界和上界。

alpha 参数为输入变量的可选项，它控制置信区间的宽度。在缺省情况下，alpha 等于 0.05，对应于 95%置信区间。

举例：

本例首先用 betarnd 函数生成 100 个服从贝塔分布的数据。假设参数真值分别为 4 和 3，现利用 betafit 函数进行参数估计，并计算参数的置信区间。

```
r = betarnd(4,3,100,1);
[p,ci] = betafit(r,0.01)
p =
    3.9010     2.6193
ci =
    2.5244     1.7488
    5.2777     3.4899
```

参见：

betalike, mle

betalike 函数

功能：计算负贝塔对数似然函数。

语法：logL = betalike(params,data)

[logL,info] = betalike(params,data)

描述：logL = betalike(params,data) 对于给定的数据 data，返回两个 beta 参数的贝塔对数似然负函数。logL 的长度为 data 的长度。

[logL,info] = betalike(params,data) 该形式还返回费歇尔信息矩阵 info。费歇尔信息矩阵的对角线上为各参数的渐进方差。

Betalike 函数是一个工具函数，用于求取贝塔函数的最大似然估计。要求数据样本满足相互独立的似然假设。由于 betalike 函数返回负的伽玛对数似然函数，所以使用 fmins 函数最小化 betalike 的效果与最大化 likelihood 函数的相同。

举例：

继续使用 betafit 函数的例子。

```
r = betarnd(4,3,100,1);
[logl,info] = betalike([3.9010 2.6193],r)
logl =
  -33.0514
info =
    0.2856     0.1528
    0.1528     0.1142
```

参见：

betafit, fmins, gamlike, mle, weiblike

mle 函数

功能： 进行最大似然估计。

语法： phat = mle('dist',data)

[phat,pci] = mle('dist',data)

[phat,pci] = mle('dist',data,alpha)

[phat,pci] = mle('dist',data,alpha,p1)

描述： phat = mle('dist',data) 用 data 向量中的样本返回‘dist’指定的分布的最大似然估计（MLEs）。

[phat,pci] = mle('dist',data) 返回最大似然估计和 95%置信区间。

[phat,pci] = mle('dist',data,alpha) 返回指定分布的最大似然估计值和 100（1-alpha）置信区间。

[phat,pci] = mle('dist',data,alpha,p1) 该形式仅用于二项分布，其中 p1 为试验次数。

举例：

```
rv = binornd(20,0.75)
rv =
    16
[p,pci] = mle('binomial',rv,0.05,20)
p =
    0.8000
pci =
    0.5634
    0.9427
```

参见：

betafit, binofit, expfit, gamfit, normfit, poissfit, weibfit

表 4 中为常见分布的参数估计函数及其调用格式。

表 4 常见分布的参数估计函数及其调用格式

函 数 名	参数估计对应的分布	调 用 格 式
betafit	贝塔分布	phat = betafit(x) [phat,pci] = betafit(x,alpha)
betalike	贝塔对数似然函数	logL=betalike(params,data) [logL,info]=betalike(params,data)
binofit	二项分布	phat=binofit(x,n) [phat,pci]=binofit(x,n) [phat,pci]=binofit(x,n,alpha)
expfit	指数分布	muhat=expfit(x) [muhat,muci]=expfit(x) [muhat,muci]=expfit(x,alpha)
gamfit	伽玛分布	phat=gamfit(x) [phat,pci]=gmfit(x) [phat,pci]=gamfit(x,alpha)

gamlike	伽玛似然函数	logL=gamlike(params,data) [logL,info]=gamlike(params,data)
mle	极大似然估计	phat=mle('dist',data) [phat,pci]=mle('dist',data) [phat,pci]=mle('dist',data,alpha) [phat,pci]=mle('dist',data,alpha,p1)
normlike	正态对数似然函数	L=normlike(params,data)
normfit	正态分布	[muhat,sigmahat,muci,sigmaci]=normfit(X) [muhat,sigmahat,muci,sigmaci]=normfit(X,alpha)
poissfit	泊松分布	lambdahat=poissfit(X) [lambdahat,lambdaci]=poissfit(X) [lambdahat,lambdaci]=poissfit(X,alpha)
unifit	均匀分布	[ahat,bhat]=unifit(X) [ahat,bhat,ACI,BCI]=unifit(X) [ahat,bhat,ACI,BCI]=unifit(X,alpha)
weibfit	威布尔分布	phat=weibfit(x) [phat,pci]=weibfit(x) [phat,pci]=weibfit(x,alpha)
weiblike	威布尔对数似然函数	logL=weiblike(params,data) [logL,info]=weiblike(params,data)

2.4 逆累加分布函数

2.4.1 基本数学原理

逆累加分布函数是累加分布函数的逆函数。利用逆累加分布函数，可以求得满足给定概率时随机变量对应的置信区间的最小值和最大值。

2.4.2 相关函数介绍

norminv 函数

功能：计算正态累加分布函数的逆函数。

语法：X = norminv(P,MU,SIGMA)

描述：norminv(P,MU,SIGMA) 计算 P 处参数为 MU 和 SIGMA 的正态累加函数的逆函数。SIGMA 必须为正，P 值必须属于[0 1]区间。

下面用正态累加分布函数的形式来定义正态逆函数。

其中

$$p = F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}}dt \}$$

结果 x 为上面参数为 μ 和 σ 的整型方程的解，其中提供了概率 p。

举例：

求包含标准正态分布数据 95%的值的区间。

```
x = norminv([0.025 0.975],0,1)
x =
```

    -1.9600   1.9600

注意，x 区间不是符合要求的唯一区间，但是是最短的。

xl = norminv([0.01 0.96],0,1)

xl =

    -2.3263   1.7507

区间 xl 也包含 95%的值，但它比 x 长。

expinv 函数

功能：计算指数累加分布函数的逆函数。

语法：X = expinv(P,MU)

描述：

expinv(P,MU) 计算 P 处参数为 MU 的指数累加分布函数的逆函数。MU 必须为正，P 值必须界于 0 和 1 之间。

指数累加分布函数的逆函数为:

$$x = F(p|\mu) = -\mu\ln(1-p)$$

结果 x 表示取自参数为 μ 的指数分布的观测值落在[0 x]区间内的概率为 p 时对应的值。

举例：

假设电灯的使用寿命服从mu等于700小时的指数分布,求电灯使用寿命的中值。

expinv(0.50,700)

ans =

  485.2030

所以，假设你买了一箱使用寿命为"700 小时"的电灯，若 700 小时为电灯的平均寿命，则其中一半的使用寿命不会超过 500 小时。

表 5 为常见分布的累加分布逆函数及其调用格式。

表 5 常见分布的累加分布逆函数

函数名	累加分布函数逆函数对应的分布	调 用 格 式
betainv	贝塔分布	X=betainv(P,A,B)
binoinv	二项分布	X=binoinv(Y,N,P)
chi2inv	卡方分布	X=chi2inv(P,V)
expinv	指数分布	X=expinv(P,MU)
finv	F 分布	X=finv(P,V1,V2)
gaminv	伽玛分布	X=gaminv(P,A,B)
geoinv	几何分布	X=geoinv(Y,P)
hygeinv	超几何分布	X=hygeinv(P,M,K,N)
loginv	对数正态分布	X=logncdf(X,MU,SIGMA)
nbininv	负二项分布	Y=nbincdf(X,R,P)
ncfinv	非中心 F 分布	P=ncfcdf(X,NU1,NU2,DELTA)
nctinv	非中心 t 分布	P=nctcdf(X,NU,DELTA)

ncx2inv	非中心卡方分布	P=ncx2cdf(X,V,DELTA)
icdf		
norminv	正态（高斯）分布	X=norminv(P,MU,SIGMA)
poissinv	泊松分布	X=poissinv(P,LAMBDA)
raylinv	雷利分布	X=raylinv(P,B)
tinv	学生氏 t 分布	X=tinv(P,V)
unidinv	离散均匀分布	X=unidinv(P,N)
unifinv	连续均匀分布	X=unifcdf(X,A,B)
weibinv	威布尔分布	X=weibcdf(X,A,B)

2.5 随机数的生成

2.5.1 随机数生成的基本原理

生成给定分布的随机数，需要首先生成服从均匀分布的随机数。常用的生成均匀分布随机数的方法是同余法，其递推公式为：

$$x_i=(ax_{i-1}+c)\bmod m$$

给定初值时，可以迭代出均匀随机数 $x_1,x_2,\ldots,x_n$，将它们进行正规化（则随机数界于0 和 1 之间）或极差标准化（则随机数界于-1 和 1 之间），可以得到均匀分布的随机数。

获得均匀分布的随机数以后，可以用多种方法构造基于该随机数的随机变量，常用的方法是反函数法，即利用随机变量 x 的分布函数 $F(x)$ 的反函数 $F^{-1}(x)$ 来推求随机变量。基本算法是：

- 产生均匀分布随机数 r；
- 令 $x=F^{-1}(x)$，然后返回。

下面结合正态分布随机变量的生成进行具体介绍：

正态分布密度函数为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\qquad(-\infty<x<+\infty)$$

其中，μ 为期望；σ^2 为方差，分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$F(x)$ 为非可积函数，由中心极限定理，有

$$\frac{x-\mu}{\sigma}=\sqrt{\frac{12}{n}\left(\sum_{i=1}^{n}r_i-\frac{n}{2}\right)}$$

当 n=12 时，可达到较好精度，故

$$\boldsymbol{x}=\left(\sum_{i=1}^{12}\boldsymbol{r}_i-6\right)\boldsymbol{\sigma}+\boldsymbol{\mu}$$

则 $\boldsymbol{x}$ 就是基于均匀分布随机数 $\boldsymbol{r}_i$ 的服从正态分布的随机数。

2.5.2 相关函数介绍

normrnd 函数

功能：生成服从正态分布的随机数。

语法：

```
R = normrnd(MU,SIGMA)
R = normrnd(MU,SIGMA,m)
R = normrnd(MU,SIGMA,m,n)
```

描述：

R = normrnd(MU,SIGMA) 生成均值为 MU，标准差 SIGMA 的正态分布随机数。

R = normrnd(MU,SIGMA,m) 生成均值为 MU，标准差 SIGMA 的正态分布随机数。m 为 1×2 的向量，包含 R 的行和列的维数。

R = normrnd(MU,SIGMA,m,n) 生成均值为 MU，标准差 SIGMA 的正态分布随机数。m 和 n 分别为行和列的维数。

举例：

```
n1 = normrnd(1:6,1./(1:6))
n1 =
    2.1650    2.3134    3.0250    4.0879    4.8607    6.2827
n2 = normrnd(0,1,[1 5])
n2 =
    0.0591    1.7971    0.2641    0.8717   -1.4462
n3 = normrnd([1 2 3;4 5 6],0.1,2,3)
n3 =
    0.9299    1.9361    2.9640
    4.1246    5.0577    5.9864
```

random 函数

功能：生成指定分布的随机数。

语法：y = random('name',A1,A2,A3,m,n)

描述：

random 为一工具函数，通过将分布名称指定为参数，利用它可以生成服从各种分布的随机数。

y = random('name',A1,A2,A3,m,n) 返回一个随机数矩阵。'name' 为一字符串，包含分布名；A1、A2 和 A3 为分布参数矩阵。有的分布不需要这么多的参数。最后两个参数分别表示 x 矩阵和 y 矩阵的大小。如果分布参数为矩阵，则这些

参数是可选的，但它们必须与其它矩阵变量的大小相匹配。

举例：

```
rn = random('Normal',0,1,2,4)
rn =
    1.1650    0.0751   -0.6965    0.0591
    0.6268    0.3516    1.6961    1.7971
rp = random('Poisson',1:6,1,6)
rp =
     0     0     1     2     5     7
```

表 6 中为常见分布随机数生成函数的调用格式。

表 6 常见分布随机数生成函数的调用格式

函　数	分　布	调　用　格　式		
		格 式 一	格 式 二	格 式 三
betarnd	贝塔分布	R=betarnd(A,B)	R=betarnd(A,B,m)	R=betarnd(A,B,m,n)
binornd	二项分布	R=binornd(N,P)	R=binornd(N,P,mm)	R=binornd(N,P,mm,nn)
chi2rnd	卡方分布	R=chi2rnd(V)	R=chi2rnd(V,m)	R=chi2rnd(V,m,n)
exprnd	指数分布	R=exprnd(MU)	R=exprnd(MU,m)	R=exprnd(MU,m,n)
frnd	F 分布	R=frnd(V1,V2)	R=frnd(V1,V2,m)	R=frnd(V1,V2,m,n)
gamrnd	伽玛分布	R=gamrnd(A,B)	R=gamrnd(A,B,m)	R=gamrnd(A,B,m,n)
geornd	几何分布	R=geornd(P)	R=geornd(P,m)	R=geornd(P,m,n)
hygernd	超几何分布	R=hygernd(M,K,N)	R=hygernd(M,K,N,mm)	R=hygernd(M,K,N,mm,nn)
lognrnd	对数正态分布	R=lognrnd(MU,SIGMA)	R=lognrnd(MU,SIGMA,m)	R=lognrnd(MU,SIGMA,m,n)
nbinrnd	负二项分布	R=nbinrnd(R,P)	R=nbinrnd(R,P,m)	R=nbinrnd(R,P,m,n)
ncfrnd	非中心 F 分布	R=ncfrnd(NU1,NU2,DELTA)	R=ncfrnd(NU1,NU2,DELTA,m)	R=ncfrnd(NU1,NU2,DELTA,m,n)
nctrnd	非中心 t 分布	R=nctrnd(V,DELTA)	R=nctrnd(V,DELTA,m)	R=nctrnd(V,DELTA,m,n)
ncx2rnd	非中心卡方分布	R=ncx2rnd(V,DELTA)	R=ncx2rnd(V,DELTA,m)	R=ncx2rnd(V,DELTA,m,n)
normrnd	正态（高斯）分布	R=normrnd(MU,SIGMA)	R=normrnd(MU,SIGMA,m)	R=normrnd(MU,SIGMA,m,n)
poissrnd	泊松分布	R=poissrnd(LAMBDA)	R=poissrnd(LAMBDA,m)	R=poissrnd(LAMBDA,m,n)
raylrnd	瑞利分布	R=raylrnd(B)	R=raylrnd(B,m)	R=raylrnd(B,m,n)
trnd	学生氏 t 分布	R=trnd(V)	R=trnd(V,m)	R=trnd(V,m,n)
unidrnd	离散均匀分布	R=unidrnd(N)	R=unidrnd(N,mm)	R=unidrnd(N,mm,nn)
unifrnd	连续均匀分布	R=unifrnd(N)	R=unifrnd(N,mm)	R=unifrnd(N,mm,nn)
weibrnd	威布尔分布	R=weibrnd(A,B)	R=weibrnd(A,B,m)	R=weibrnd(A,B,m,n)

2.6 分布函数的统计量估计

normstat 函数

功能：求正态分布的均值和方差。

语法：[M,V] = normstat(MU,SIGMA)

描述：

对于正态分布：

$$y = f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

均值为 μ，方差为 σ^2。

举例：

```
n = 1:5;
[m,v] = normstat(n'*n,n'*n)
[m,v] = normstat(n'*n,n'*n)
m =
     1     2     3     4     5
     2     4     6     8    10
     3     6     9    12    15
     4     8    12    16    20
     5    10    15    20    25
v =
     1     4     9    16    25
     4    16    36    64   100
     9    36    81   144   225
    16    64   144   256   400
    25   100   225   400   625
```

expstat 函数

功能：计算指数分布的均值和方差。

语法：[M,V] = expstat(MU)

描述：

对于指数分布函数：

$$y = f(x|\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$$

均值为 μ，方差为 μ^2。

举例：

```
[m,v] = expstat([1 10 100 1000])
```

```
m =
           1          10         100        1000
v =
           1         100       10000     1000000
```

表 7 中为常见分布的统计量估计函数及其调用格式。

表 7 常见分布的统计量估计

函数名	分 布	参 数 意 义		调 用 格 式
		均 值	方 差	
betastat	贝塔分布	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$	[M, V]=betastat(A, B)
binostat	二项分布	np	npq	[M, V]=binostat(N, P)
chi2stat	卡方分布	n	$2n$	[M, V]=chi2stat(NU)
expstat	指数分布	μ	μ^2	[M, V]=expstat(MU)
fstat	F 分布	$\frac{v_2}{v_2-2}, v_2>2$	$\frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)}, v_2>4$	[M, V]=fstat(V1, V2)
gamstat	伽玛分布	ab	ab^2	[M, V]=gamstat(A, B)
geostat	几何分布	q/p	q/p^2	[M, V]=geostat(P)
hygestat	超几何分布	$N\frac{K}{M}$	$N\frac{K}{M}\frac{M-K}{M}\frac{M-N}{M-1}$	[M, V]=hygestat(M, K, N)
lognstat	对数正态分布	$e^{\left(\mu+\frac{\sigma^2}{2}\right)}$	$e^{\left(2\mu+2\sigma^2\right)}-e^{\left(2\mu+\sigma^2\right)}$	[M, V]=lognstat(MU, SIGMA)
nbinstat	负二项分布	$\frac{rq}{p}$	$\frac{rq}{p^2}$	[M, V]=nbinstat(R, P)
ncfstat	非中心 F 分布	$\frac{v_2(\delta+v_1)}{v_1(v_2-2)}, v_2>2$	$2\left(\frac{v_2}{v_1}\right)^2\left[\frac{(\delta+v_1)^2+(2\delta+v_1)(v_2-2)}{(v_2-2)^2(v_2-4)}\right]$, $v_2>4$	[M, V]=ncfstat(NU1, NU2, DELTA)
nctstat	非中心 t 分布	$\frac{\delta(v/2)^{1/2}\Gamma((v-1)/2)}{\Gamma(v/2)}$	$\frac{v}{(v-2)}(1+\delta^2)-\frac{v}{2}\delta^2\left[\frac{\Gamma((v-1)/2)}{\Gamma(v/2)}\right]^2$	[M, V]=nctstat(NU, DELTA)
ncx2stat	非中心卡方分布	$v+\delta$	$2(v+2\delta)$	[M, V]=ncx2stat(NU, DELTA)
normstat	正态（高斯）分布	μ	σ^2	[M, V]=normstat(MU, SIGMA)
poisstat	泊松分布	λ	λ	[M, V]=poisstat(LAMBDA)
raylstat	瑞利分布	$b\left(\frac{\pi}{2}\right)^{\frac{1}{2}}$	$\frac{2-\pi}{2}b^2$	[M, V]=raylstat(B)
tstat	学生氏 t 分布	若 v>1，则均值为 0；若 v=1，则均值不存在	$\frac{v}{v-2}, v>2$	[M, V]=tstat(NU)
unidstat	离散均匀分布	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	[M, V]=unidstat(N)
unifstat	连续均匀分布	(a+b)/2	(b-a)/12	[M, V]=unidstat(A, B)
weibstat	威布尔分布	$a^{-\frac{1}{b}}\Gamma(1-b^{-1})$	$a^{-\frac{2}{b}}\left[\Gamma(1+2b^{-1})-\Gamma^2(1+b^{-1})\right]$	[M, V]=weibstat(A, B)