参数估计与区间估计

参数估计与区间估计

- 求取参数与区间估计的函数调用格式

$$[\mu,\sigma^2,\Delta\mu,\Delta\sigma^2]=\mathtt{normfit}(\boldsymbol{x},P_{\mathrm{ci}})$$

$$[\mu,\sigma^2,\Delta\mu,\Delta\sigma^2]=\mathtt{fittest}(\text{'}\mathtt{norm}\text{'},\boldsymbol{x},P_{\mathrm{ci}})$$

- 其中， $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 是实测一组数据
- μ 是该分布的均值，σ^2是该分布的方差
- $\Delta\mu$ 及 $\Delta\sigma^2$ 是置信区间
- P_{ci} 为用户指定的置信度

给定分布的均值与方差计算

- 函数 `norminv()` 可用于求出相关值，这样就可以得出所需的参数
  - Gamma分布的参数（a, λ）——`gamfit()` 函数
  - Rayleigh 分布的参数估计函数为`raylfit()`
  - 均匀分布的参数估计函数为`unifit()`
  - Poisson 分布的参数估计函数为`poissfit()`
- 可以调用 `fittest` 函数

例9-29 Gamma分布的参数估计

- 试用gamrnd()函数生成一组a=1.5,λ=3的伪随机数
  - 用参数估计的方法以不同的置信度进行估计
  - 选择置信度为90%，92%，95%，98%
  - 置信度，估计值，估计区间

```
>> p=gamrnd(1.5,3,30000,1);
   Pv=[0.9,0.92,0.95,0.98]; A=[];
   for i=1:length(Pv)
      [a,b]=gamfit(p,Pv(i));
      A=[A; Pv(i),a(1),b(:,1)', a(2),b(:,2)'];
   end, A
```

多元线性回归与区间估计

- 输出信号 y
  - n 路输入信号 $x_1, x_2, \cdots, x_n$
  - 线性组合

$$y = a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n$$

- 其中，$a_1, a_2, \cdots, a_n$ 为待定系数
- 线性回归
- 求解线性代数方程

得到的实测数据满足的关系式

➢ 实测的数据

$$
\begin{aligned}
y_1 &= x_{11}a_1 + x_{12}a_2 + \cdots + x_{1,n}a_n + \varepsilon_1 \\
y_2 &= x_{21}a_1 + x_{22}a_2 + \cdots + x_{2,n}a_n + \varepsilon_2 \\
&\vdots \\
y_m &= x_{m1}a_1 + x_{m2}a_2 + \cdots + x_{m,n}a_n + \varepsilon_m
\end{aligned}
$$

➢ 观测数据组数与未知待定参数个数不同

➢ 尽量使得总体误差最小

参数估计的最小二乘求解

- 目标函数选择为使得残差的平方和最小：

$$J = \min_{\boldsymbol{a}} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$$

- 系数向量 $\boldsymbol{a}$ 为 $\hat{\boldsymbol{a}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$
- 求最小二乘解的函数调用格式

$\boldsymbol{a}$=$\boldsymbol{X}$\\$\boldsymbol{y}$ $\quad\quad$ $\boldsymbol{a}$=inv($\boldsymbol{X}$'*$\boldsymbol{X}$)*$\boldsymbol{X}$'*$\boldsymbol{y}$

- 求解函数，1−α为用户指定的置信度

[$\hat{\boldsymbol{a}}$,$\boldsymbol{a}_{\mathrm{ci}}$]=regress($\boldsymbol{y}$,$\boldsymbol{X}$,$\alpha$)

例9-30 参数估计与区间估计

- 给定线性回归方程如下，生成120组随机输入值 x_i
- 计算输出向量 y，估计出系数 a_i

$$y = x_1 - 1.232x_2 + 2.23x_3 + 2x_4 + 4x_5 + 3.792x_6$$

- 用最小二乘计算公式：

```
>> a=[1 -1.232 2.23 2 4,3.792]';
   X=randn(120,6);y=X*a; a1=inv(X'*X)*X'*y
```

- 计算出 98% 的置信度的置信区间

```
>> [a,aint]=regress(y,X,0.02)
```

混叠噪声的信号参数估计

- 给输出样本叠加$N(0,0.5)$区间的正态分布噪声
- 再绘制参数估计的置信区间

```
>> yhat=y+sqrt(0.5)*randn(120,1);
[a,aint]=regress(yhat,X,0.02)
errorbar(1:6,a,aint(:,1)-a,aint(:,2)-a)
```

- 将噪声方差设为0.1

```
>> yhat=y+sqrt(0.1)*randn(120,1);
[a,aint]=regress(yhat,X,0.02);
errorbar(1:6,a,aint(:,1)-a,aint(:,2)-a)
```

非线性函数的最小二乘参数估计与区间估计

- 假设数据 $x_i, y_i, i = 1, 2, \cdots, N$ 满足原型函数

$$\hat{y}(x) = f(\boldsymbol{a}, x)$$

- 原函数严格写成

$$\hat{y}(x) = f(\boldsymbol{a}, x) + \varepsilon$$

- 引入目标函数：

$$I = \min_{\boldsymbol{a}} \sum_{i=1}^{N} [y_i - \hat{y}(x_i)]^2 = \min_{\boldsymbol{a}} \sum_{i=1}^{N} [y_i - f(\boldsymbol{a}, x_i)]^2$$

参数估计MATLAB求解

- 参数估计的函数调用格式
  - 最小二乘拟合

    $[a,r,J]=\texttt{nlinfit}(x,y,fun,a_0)$

  - 由置信度为 95% 的置信区间

    $c=\texttt{nlparci}(a,r,J)$

- 与函数lsqcurvefit()的功能相似

例9-31 参数与区间估计

- 原型函数 $y(x)=a_1\mathrm{e}^{-a_2x}+a_3\mathrm{e}^{-a_4x}\sin(a_5x)$
- 95% 置信度的置信区间
- 叠加均匀分布的噪声信号再进行参数与区间估计
- MATLAB求解语句

```
>> f=@(a,x)a(1)*exp(-a(2)*x)+...
       a(3)*exp(-a(4)*x).*sin(a(5)*x);
   x=0:0.1:10; y=f([0.12,0.213,0.54,0.17,1.23],x);
   [a,r,j]=nlinfit(x,y,f,[1;1;1;1;1]); a
   ci=nlparci(a,r,j)
```

叠加噪声后的估计

- 样本点数据 y_i
- 叠加上 $[0, 0.02]$ 区间均匀分布的噪声信号
- MATLAB求解语句：

```
>> y=f([0.12,0.213,0.54,0.17,1.23],x)...
        +0.02*rand(size(x));
   [a,r,j]=nlinfit(x,y,f,[1;1;1;1;1]),
   ci=nlparci(a,r,j)
   errorbar(1:5,a,ci(:,1)-a,ci(:,2)-a)
```