# 代码说明

## 第一题

$$Pdf_B = c \times x_i^{-4.5}$$

$$X_i \sim N(m, \sigma^2), \therefore Pdf_S = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-m)^2}{2\sigma^2}}$$

$$\because f_s = \frac{n_s}{n_s+n_b}, \therefore L_k = \prod_{i=1}^{n} (\frac{n_s}{n_s+n_b} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-m)^2}{2\sigma^2}} + \frac{n_b}{n_s+n_b} c x_i^{-4.5})$$

$$\ln L_k = \sum_{i=1}^{n} \ln (\frac{n_s}{n_s+n_b} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-m)^2}{2\sigma^2}} + \frac{n_b}{n_s+n_b} c x_i^{-4.5})$$

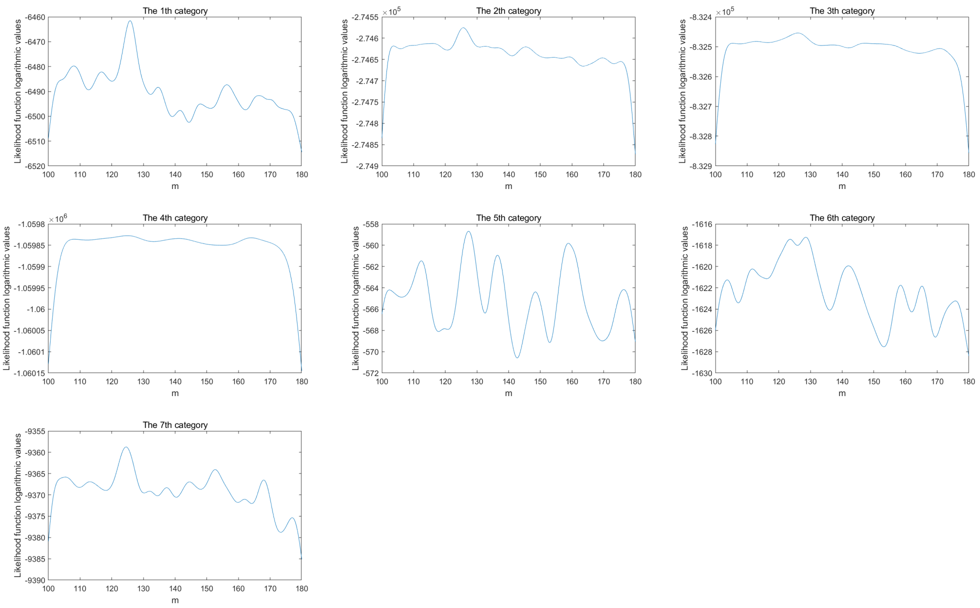
### 计算要求（续）

$$\begin{aligned} L &= \prod_{i=1}^{N} (\sum_{k=0}^{6} fs_k \times pdf_s + fb \times pdf_b) \\ &= \prod_{i=1}^{N} (fs \times pdf_s + fb \times pdf_b) \\ \ln L &= \sum_{i=1}^{N} \ln (fs \times pdf_s + fb \times pdf_b) \\ &= \sum_{i=1}^{N} \ln (fs \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-m)^2}{2\sigma^2}} + fb \times c x_i^{-4.5}) \end{aligned}$$

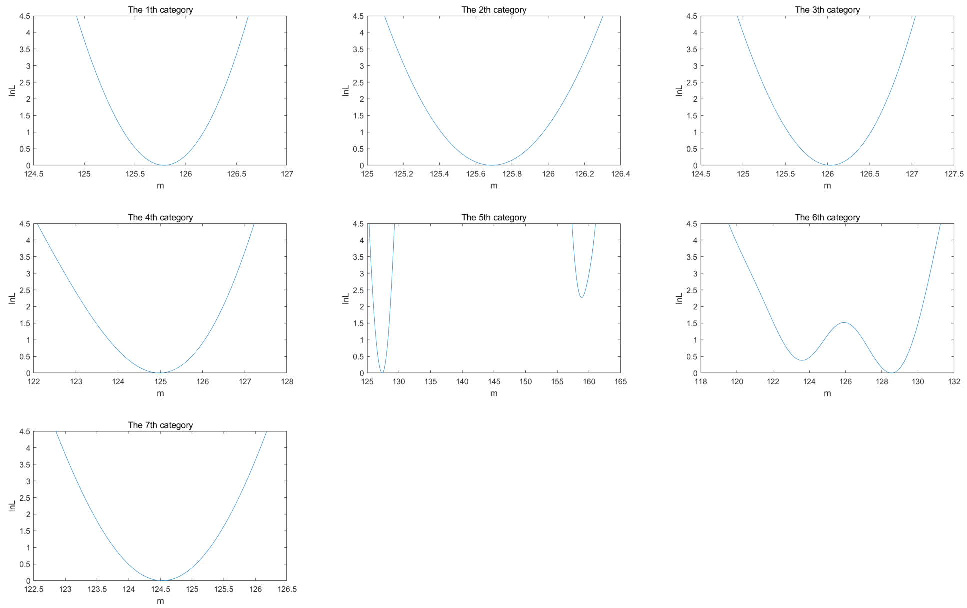
其中$\sigma = \sqrt{\frac{\sum_{i=0}^{7} ns_k \sigma_k^2}{ns}}$

### 计算结果

#### 0.01扫描作图——$\ln L$




## 0.01扫描作图——$2(\ln L(\hat{m})-\ln L)$



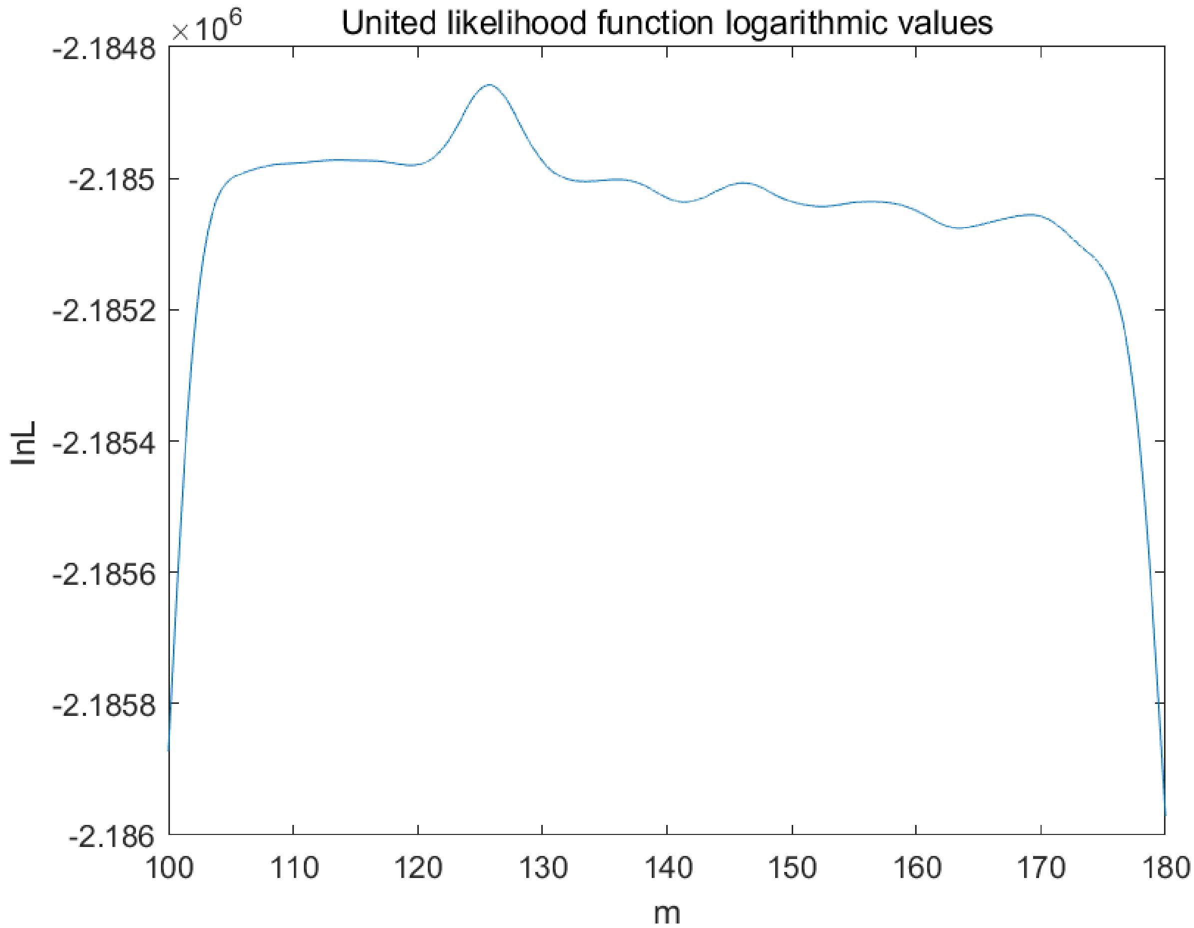

## 精细求解$\hat{m}_k$和$68.3\%CL$区间端点

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $\hat{m}_k$ | 125.7864 | 125.6887 | 126.0348 | 124.9765 | 127.3645 | 128.5342 | 124.5286 |

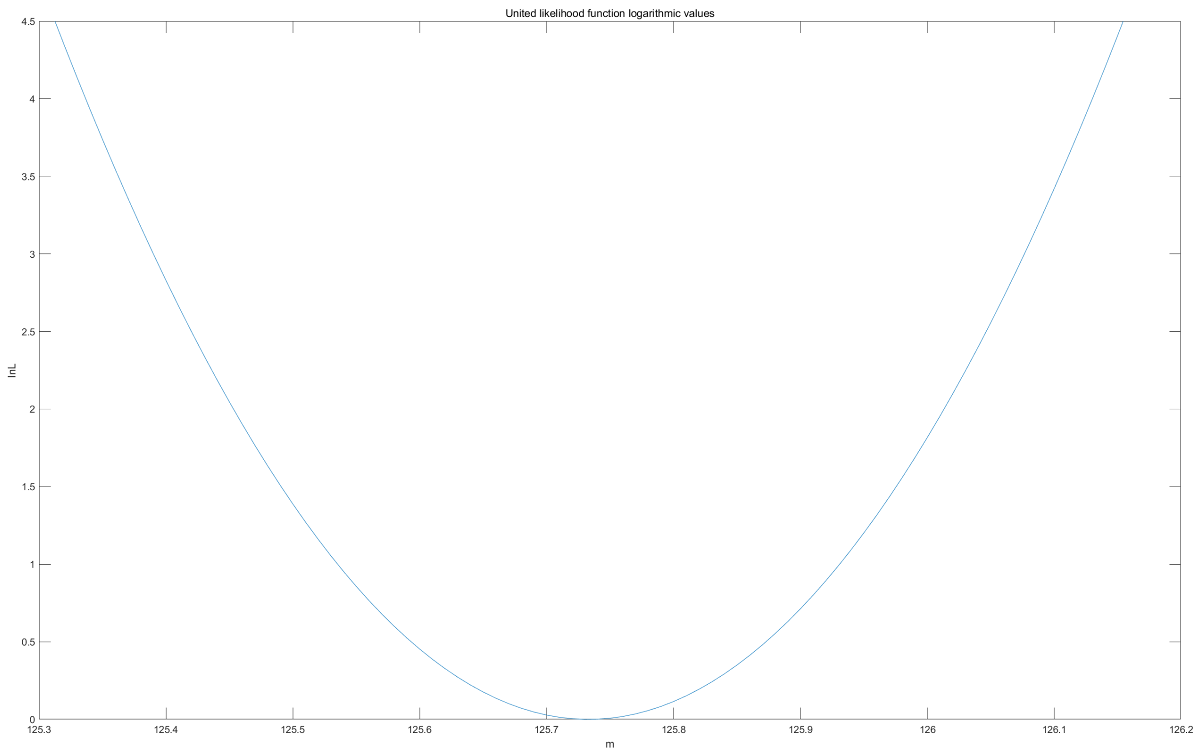
| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $m_{lo}$ | 125.3900 | 125.4088 | 125.5354 | 123.7780 | 126.4094 | 126.9747 | 123.7678 |
| $m_{up}$ | 126.1758 | 125.9733 | 126.5141 | 126.0513 | 128.2738 | 129.7347 | 125.2854 |

因此Category1,Category0,Category2精确度最高。

### 联合分布似然函数$\ln L$




### 联合分布似然函数$2(\ln L(\hat{m}) - \ln L)$

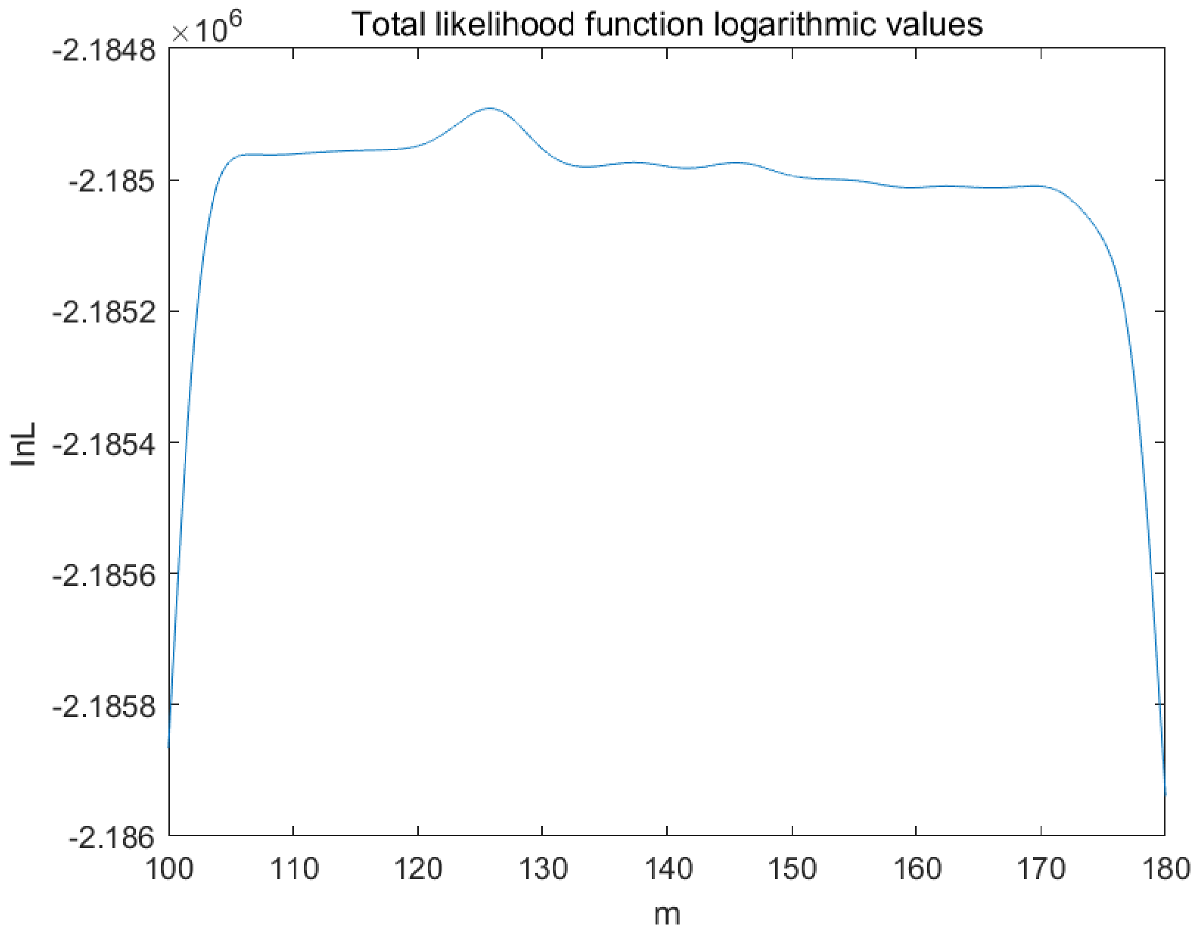


### 联合分布的参数值

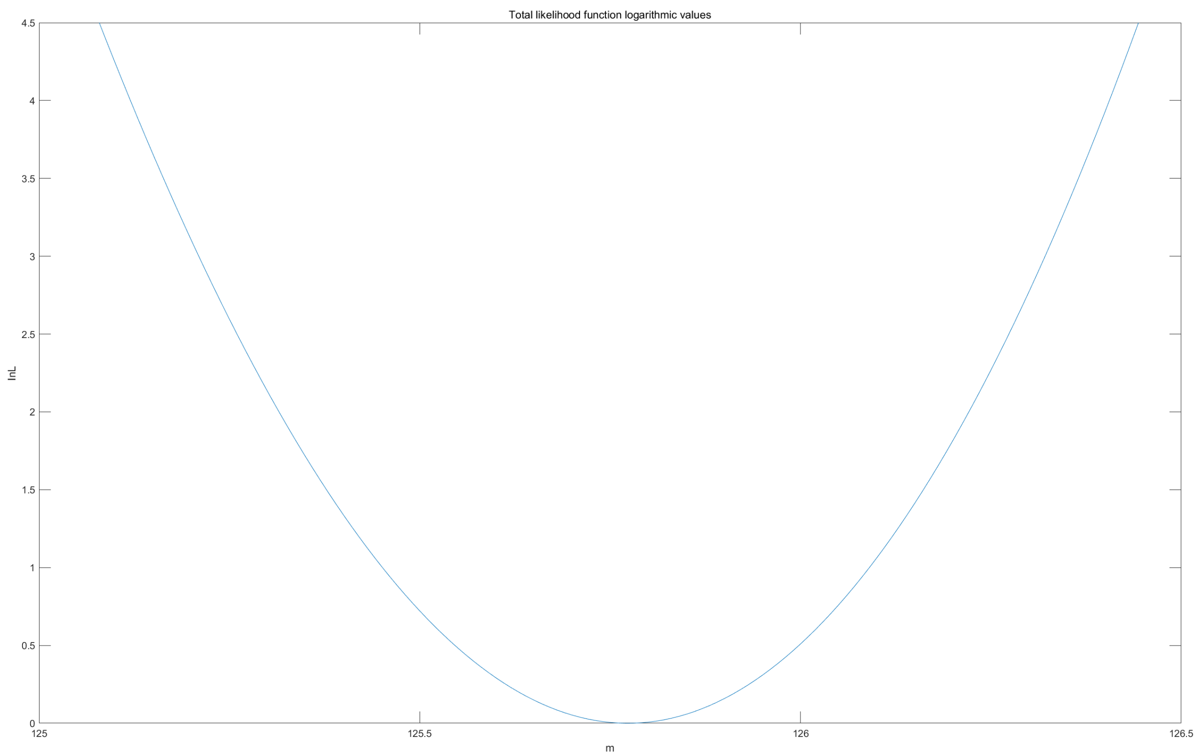
$$\hat{m}=125.73, 68.3\%CL\rightarrow[125.54,125.93]$$

由于实验得到的是$\hat{m}=125.78\pm0.21\rightarrow m\in[125.57,125.99]$宽度基本吻合，中心值略偏差。

### 混合分布似然函数$\ln L$




### 混合分布似然函数$2(\ln L(\hat{m})-\ln L)$



### 混合分布参数值

$$\hat{m}=125.77,68.3\%CL\rightarrow[125.45,126.09]$$

相比联合分布，精度下降了$64.1\%$。原因是相比联合分布，缺少了七组数据各自标准差等信息，对样本的描述变粗糙了。

## 代码

```
cat=cell(1,7);
for i=1:7
    str=['mgg_cms2020_cat',num2str(i-1),'.txt'];
    cat(i)={load(str)};
end
parameter=readmatrix('数据和模型.xlsx');
%%
syms c m
P=int(c*m^(-4.5),m,100,180);
c=double(solve(P==1,c));
%%
Sigma=parameter(:,3);
Ns=parameter(:,2);
Nb=parameter(:,4);
N=Ns+Nb;
M=100:.01:180;
f1=figure;
f1.Position=([0,0,2000,2000]);
for k=1:7
    sj=cat{k};
    sigma=Sigma(k);
    ns=Ns(k);
    nb=Nb(k);
    n=N(k);
    lnLk=0;
    for j=1:length(M)
        m=M(j);
        for i=1:length(sj)
            lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
        end
        siran(j)=lnLk;
        lnLk=0;
    end
    subplot(3,3,k)
    plot(M,siran)
    xlabel('m')
    ylabel('Likelihood function logarithmic values')
    Tit=['The ',num2str(k),'th category'];
    save([Tit,'.mat'],'siran')
    title(Tit)
    hatm=M(find(siran==max(siran)));
    Hatm(k)=hatm;
end
%%
f2=figure;
```

```
f2.Position=([0,0,2000,2000]);
for k=1:7
    sj=cat{k};
    sigma=Sigma(k);
    ns=Ns(k);
    nb=Nb(k);
    n=N(k);
    lnLk=0;
    m=Hatm(k);
    for i=1:length(sj)
        lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
    end
    pred=lnLk;
    Tit=['The',num2str(k),'th category'];
    load([Tit,'.mat'])
    noname=2.*(pred-siran);
    subplot(3,3,k)
    M=100:.01:180;
    plot(M,noname)
    xlabel('m')
    ylabel('lnL')
    Tit=['The ',num2str(k),'th category'];
    title(Tit);
    ylim([0,4.5]);
    jidian=find(M==m);
    if k~=6
        CLmin(k)=M(find(abs(noname(1:jidian)-1)==min(abs(noname(1:jidian)-1))));
        CLmax(k)=M(find(abs(noname(jidian:end)-1)==min(abs(noname(jidian:end)-1)))-1+jidian);
    elseif k==6
        left=find(M==126);right=find(M==128);
        CLmin(k)=M(find(abs(noname(left:right)-1)==min(abs(noname(left:right)-1)))-1+left);
        CLmax(k)=M(find(abs(noname(right:end)-1)==min(abs(noname(right:end)-1)))-1+right);
    end
end
%% 精细求解hatm
for k=1:7
    sj=cat{k};
    sigma=Sigma(k);
    ns=Ns(k);
    nb=Nb(k);
    n=N(k);
    M=Hatm(k)-0.01:0.0001:Hatm(k)+0.01;
    lnLk=0;
    for j=1:length(M)
        m=M(j);
        for i=1:length(sj)
```

```
            lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
        end
        siran(j)=lnLk;
        lnLk=0;
    end
    PreciseHatm(k)=M(find(siran==max(siran)));
    clear siran
end
%% 精细求解置信区端点
for k=1:7
    sj=cat{k};
    sigma=Sigma(k);
    ns=Ns(k);
    nb=Nb(k);
    n=N(k);
    M=CLmin(k)-0.01:0.0001:CLmin(k)+0.01;
    lnLk=0;
    for j=1:length(M)
        m=M(j);
        for i=1:length(sj)
            lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
        end
        siran(j)=lnLk;
        lnLk=0;
    end
    lnLk=0;
    m=PreciseHatm(k);
    for i=1:length(sj)
        lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
    end
    pred=lnLk;
    noname=2.*(pred-siran);
    PreciseCLmin(k)=M(find(abs(noname-1)==min(abs(noname-1))));
end
for k=1:7
    sj=cat{k};
    sigma=Sigma(k);
    ns=Ns(k);
    nb=Nb(k);
    n=N(k);
    M=CLmax(k)-0.01:0.0001:CLmax(k)+0.01;
    lnLk=0;
    for j=1:length(M)
        m=M(j);
        for i=1:length(sj)
            lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
        end
        siran(j)=lnLk;
        lnLk=0;
    end
    lnLk=0;
```

```
    m=PreciseHatm(k);
    for i=1:length(sj)
        lnLk=lnLk+log(ns/n*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+nb/n*c.*sj(i).^(-4.5));
    end
    pred=lnLk;
    noname=2.*(pred-siran);
    PreciseCLmax(k)=M(find(abs(noname-1)==min(abs(noname-1))));
end
%%
f3=figure;
f3.Position=([0,0,2000,2000]);
Unit=zeros(1,8001);
M=100:.01:180;
for k =1:7
    Tit=['The',num2str(k),'th category'];
    load([Tit,'.mat']);
    Unit=Unit+siran;
end
plot(M,Unit)
xlabel('m')
ylabel('lnL')
Tit='United likelihood function logarithmic values';
save([Tit,'.mat'],'Unit')
title(Tit)
unitedhatm=M(find(Unit==max(Unit)));
%%
f4=figure;
f4.Position=([0,0,2000,2000]);
pred=Unit(find(Unit==max(Unit)));
load([Tit,'.mat'])
noname=2.*(pred-Unit);
M=100:.01:180;
plot(M,noname)
xlabel('m')
ylabel('lnL')
title(Tit)
ylim([0,4.5])
jidian=find(Unit==max(Unit));
unitedCLmin=M(find(abs(noname(1:jidian)-1)==min(abs(noname(1:jidian)-1))));
unitedCLmax=M(find(abs(noname(jidian:end)-1)==min(abs(noname(jidian:end)-1)))-1+jidian);
%% Updated
sj=[];
for i=1:7
    sj=[sj,cat{i}'];
end
Sigma=parameter(:,3);
Ns=parameter(:,2);
Nb=parameter(:,4);
N=Ns+Nb;
sigma=(sum(Ns.*Sigma.^2)/sum(Ns))^(1/2);
fs=sum(Ns)/sum(N);
fb=sum(Nb)/sum(N);
M=100:.01:180;
```

```
f5=figure;
f5.Position=([0,0,2000,2000]);
lnLk=0;
for j=1:length(M)
    m=M(j);
    for i=1:length(sj)
        lnLk=lnLk+log(fs*1/(2*pi*sigma^2)^(1/2).*exp(-(sj(i)-m).^2./(2*sigma^2))+fb*c.*sj(i).^(-4.5));
    end
    siran(j)=lnLk;
    lnLk=0;
end
plot(M,siran)
xlabel('m')
ylabel('lnL')
Tit='Total likelihood function logarithmic values';
%save([Tit,'.mat'],'siran')
title(Tit)
Totalhatm=M(find(siran==max(siran)));
%%
f6=figure;
f6.Position=([0,0,2000,2000]);
pred=max(siran);
noname=2.*(pred-siran);
M=100:.01:180;
plot(M,noname)
xlabel('m')
ylabel('lnL')
title(Tit)
ylim([0,4.5])
jidian=find(siran==max(siran));
TotalCLmin=M(find(abs(noname(1:jidian)-1)==min(abs(noname(1:jidian)-1))));
TotalCLmax=M(find(abs(noname(jidian:end)-1)==min(abs(noname(jidian:end)-1)))-1+jidian);
```

# 第二题

在第$i$个实验中，产生$n_i$个数据，$n_i \sim P(s+b)$

贝叶斯参数区间估计：假设先验概率为$\pi(s)$

$$p(s|n) = \frac{P(n|s)\pi(s)}{\int P(n|s')\pi(s')ds'}$$

$\pi(s)$为阶跃函数。

$$\begin{aligned} 1-\alpha &= \frac{\int_0^{s_{up}} P(n|s)\pi(s)ds}{\int_0^{\infty} P(n|s')\pi(s')ds'} \\ &= \frac{\int_0^{s_{up}} \frac{(s+b)^n}{n!} e^{-(s+b)} ds}{\int_0^{\infty} \frac{(s+b)^n}{n!} e^{-(s+b)} ds'} \end{aligned}$$

$$def \quad \int_0^a x^n e^{-x} dx = \Gamma(n+1) F_{\chi^2}(2a, 2(n+1))$$

$$\therefore s_{up} = \frac{1}{2} F_{\chi^2}^{-1}(p, 2(n+1)) - b, p = 1-\alpha[1-F_{\chi^2}(2b, 2(n+1))]$$

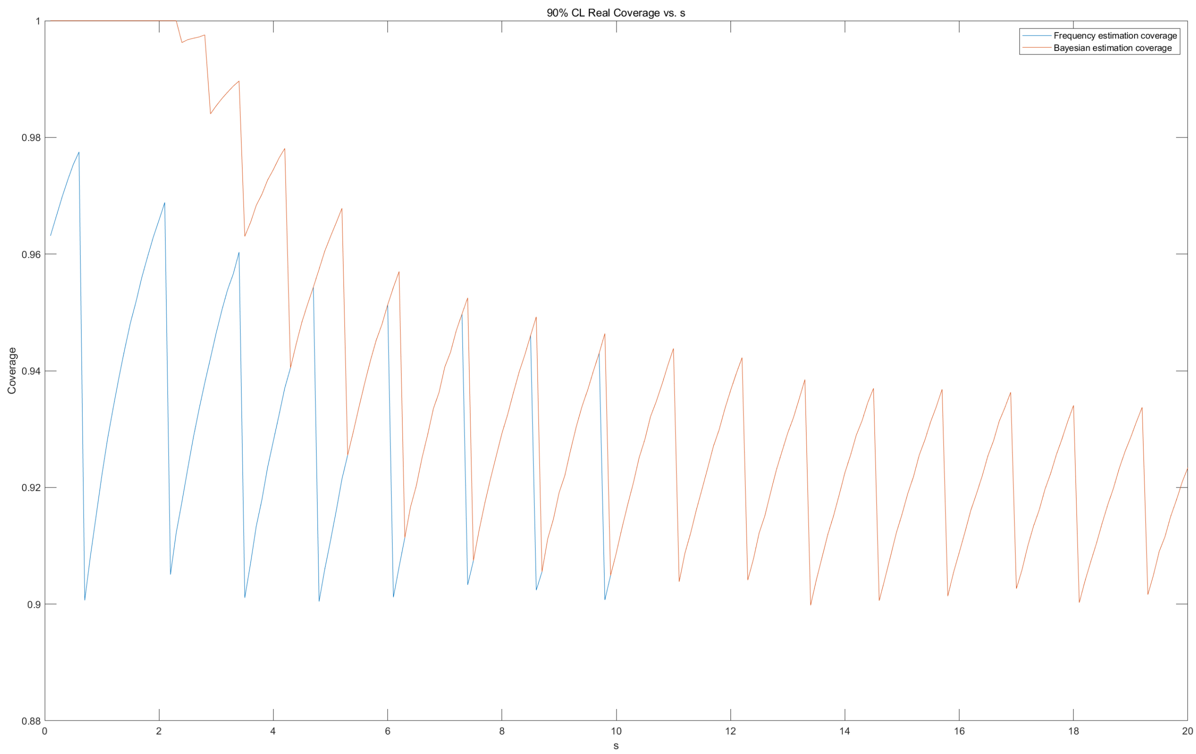
频率论估计：

$$\alpha=\sum_{k=0}^{n}\frac{(s_{up}+b)^k e^{(-s_{up}+b)}}{k!}$$

$$Theorem\ \sum_{k=0}^{n}\frac{\lambda^n e^{-\lambda}}{k!}=1-F_{\chi^2}(2\lambda,2(n+1))$$

$$\therefore s_{up}=\frac{1}{2}F_{\chi^2}^{-1}(1-\alpha,2(n+1))-b$$

计算结果：




两者均在$s=13.4$时覆盖率为$89.98\%$低于置信度，其余点均有足够的覆盖率。

## 曲线的主要特征

- 基本上始终高于0.9：90%置信区间被定义为至少提供 90% 的覆盖率，无论参数 $s$ 的真实值如何。而特别在离散分布中，由于样本离散，覆盖率一般都不会等于置信度。
- 曲线的波浪形：由图像可知，波浪的周期大致为1.原因应当是信号作为泊松过程产生，在整数点上有定义，因而会在整数周期上发生覆盖率的波动。
- 贝叶斯和频率论曲线的区别：贝叶斯曲线在$s=10$之前覆盖率优于频率论，在s很小时覆盖率达到了100%。前期拟合较好原因是s较小时信号的泊松分布不明显，先验概率相对符合。后期随着似然函数起主导作用，两种估计方法差异不明显。

## 代码

```
b=3.2;
s=6.8;
lambda=s+b;
N=poissrnd(lambda,[1,1000000]);
Bsig=0;
Fsig=0;
for i=1:length(N)
    n=N(i);
```

```
    p=1-0.1*(1-chi2cdf(2*b,2*(n+1)));
    BCLup=1/2*chi2inv(p,2*(n+1))-b;
    FCLup=1/2*chi2inv(0.9,2*(n+1))-b;
    if BCLup>=s
        Bsig=Bsig+1;
    end
    if FCLup>=s
        Fsig=Fsig+1;
    end
end
%%
S=0.1:0.1:20;
b=3.2;
for j=1:length(S)
    s=S(j);
    lambda=s+b;
    N=poissrnd(lambda,[1,1000000]);
    Bsig=0;
    Fsig=0;
    for i=1:length(N)
        n=N(i);
        p=1-0.1*(1-chi2cdf(2*b,2*(n+1)));
        BCLup=1/2*chi2inv(p,2*(n+1))-b;
        FCLup=1/2*chi2inv(0.9,2*(n+1))-b;
        if BCLup>=s
            Bsig=Bsig+1;
        end
        if FCLup>=s
            Fsig=Fsig+1;
        end
    end
    p1(j)=Bsig/1000000;
    p2(j)=Fsig/1000000;
end
save('Byes Cover Probablity.mat','p1')
save('Frequency Cover Probablity.mat','p2')
```