粒子物理与核物理实验中的数据分析

陈少敏
清华大学

第七讲：最大似然法(I)

本讲要点

- 似然函数，最大似然估计量
- 指数与高斯概率密度函数的参数确定举
- 最大似然估计量的方差
  - 解析法
  - 蒙特卡罗法
  - RCF边界法
  - 图解法
- 不等精度观测结果的并合

参数估计量的好坏标准

符合程度(一致性)

$$\lim_{n\to\infty}\hat{\theta}=\theta,$$

$$\lim_{n\to\infty}P(|\hat{\theta}-\theta|>\varepsilon)=0,\ \text{对任何}\varepsilon>0\text{都成立。}$$

偏置大小(无偏性)

$$b=E[\hat{\theta}]-\theta=0$$

方差大小(有效性)

对任何估计量$\hat{\theta}'$，都有$\lim_{n\to\infty}\frac{V[\hat{\theta}_n]}{V[\hat{\theta}_n']}\le 1$，则$\hat{\theta}$渐进效佳估计量。

物理研究中如何寻找未知参数？

最大似然法是用来寻求未知参数适当估计值的一种方法

参数估计与概率大小的关系

考虑有数据样本 $\vec{x}=(x_1,\ldots,x_n)$，这里 x 服从pdf分布 $f(x;\theta)$。

目标：估计 θ。或者更为一般地，估计 $\vec{\theta}=(\theta_1,\ldots,\theta_m)$

如果 $f(x;\theta)$ 为真，则有

$$P(\text{对所有在}[x_i,x_i+dx_i]\text{观察到的}x_i)=\prod_{i=1}^{n} f(x_i,\theta)dx_i$$

如果假设(包括 θ 的取值)为真

➡ 可以预料会使观测结果具有高的概率。

如果假设的 θ 取值远离真值

➡ 会使观测结果具有低的概率。

似然函数

根据参数好坏与概率大小的关系，可以认为真实的θ应使得下式定义的似然函数

$$L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$$

有大的数值。

注意：虽然$L(\theta)=f_{sample}(\vec{x};\theta)$，但是$L(\theta)$只是$\theta$的函数。这是因为在实验完成以后，$\vec{x}$就可以被当做常数。

在经典统计理论里，$L(\theta)$并不是θ的概率密度函数。

θ不是一个随机变量，但$\hat{\theta}$却是。

在贝叶斯统计理论里，把$L(\theta)=L(\vec{x}\,|\,\theta)$看作给定$\theta$情况下，$\vec{x}$的概率密度函数，然后利用贝叶斯定理得到验后概率密度函数$p(\theta\,|\,\vec{x})$。

最大似然估计量

定义最大似然估计量$\hat{\theta}$为使得 $L(\theta)$ 最大的θ值。通过解下列方程

$$\frac{\partial L(\theta)}{\partial \theta_i} = 0 \qquad i = 1,\ldots,m$$

通常可以找到对于m 个参数的解$\hat{\theta}_1,\ldots,\hat{\theta}_m$。

有时候 $L(\theta)$ 可以有好几个极大值 → 取最大值

注意，1）该方法利用了所有信息，与如何划分数据分布区间无关；
2）定义的最大似然估计量并不保证它们总是最优的。

→ 需要对诸如无偏性，有效性等问题进行研究

多数情况下对于足够大样本，最大似然法的确能给出了期待的好结果。

即使是小样本的情况，虽然并不总是达到最优，但它通常仍然能给出最好的实用解。

最大似然估计量的唯一性

考虑 θ 的最大似然估计值是下列方程的解

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0$$

如果选用另一个等价参数 $h(\theta)$, 则 h 的最大似然估计值是下列方程的解

$$\frac{\partial \log L(\theta)}{\partial h} = 0$$

而对于

$$\frac{\partial \log L(\theta)}{\partial h} = \frac{\partial \log L(\theta)}{\partial \theta}\frac{\partial \theta}{\partial h}$$

只要 $\frac{\partial h}{\partial \theta} \neq 0$, 就有

$$\left.\frac{\partial \log L(\theta)}{\partial h(\theta)}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial \log L(\theta)}{\partial \theta}\frac{\partial \theta}{\partial h}\right|_{\theta=\hat{\theta}} = 0 \Rightarrow \hat{h} = h(\hat{\theta})$$

因此，***h*** 的最大似然估计值与参数选取无关，具有唯一性。

最大似然估计量的渐进性

如果 $\vec{x}=(x_1,...,x_n)$ 是分布 $f(x;\theta)$ 的随机样本，$\hat{\theta}$ 是参数 θ 的最大似然估计。则当样本容量 $n\to\infty$ 时，$\hat{\theta}$ 的分布趋近于一个正态分布，即

$$f(\hat{\theta};\theta)=N(\hat{\theta};\theta,V[\theta])$$

其中方差

$$V[\theta]=-\left(\overline{\frac{\partial^2 \log L(\theta)}{\partial^2\theta}}\right)^{-1}=-\frac{1}{n}\left(\overline{\frac{\partial^2 \log f(x;\theta)}{\partial^2\theta}}\right)^{-1}$$

在推断大样本的最大似然估计的误差时，可以利用测量误差理论中最常见的正态分布进行推断。

注意：样本容量多大，才能近似利用极限正态分布，才可以看作最有效的估计，这将依赖于观测量的概率密度函数的具体形式。但对于指数型分布会有一些最优性质。

例子:指数概率密度函数参数

考虑指数概率密度函数

$$f(t;\tau)=\frac{1}{\tau}e^{-t/\tau}$$

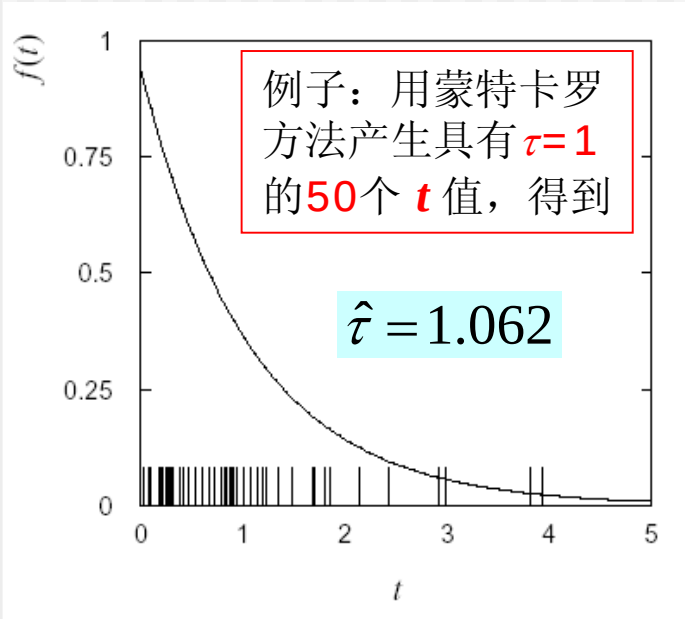
并假设有一数据样本 $t_1,\ldots,t_n$。通常为了方便起见，可采用对数形式(对同样的参数值，该定义并不会改变最大值的位置)。

$$\log L(\tau)=\sum_{i=1}^{n}\log f(t_i;\tau)=\sum_{i=1}^{n}\left(\log\frac{1}{\tau}-\frac{t_i}{\tau}\right)$$

令 $\frac{\partial \log L}{\partial \tau}=0$，并求解 τ，

$$\Rightarrow \hat{\tau}=\frac{1}{n}\sum_{i=1}^{n}t_i$$

是平均寿命的最大似然估计



最大似然估计的偏向性问题

对样本求平均

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

所得到平均值是 τ 的一个无偏估计量吗？

原则上可以通过找出概率密度函数(例如采用蒙特卡罗方法)

$$g(\hat{\tau};\tau)$$

并计算出偏置的大小

$$b = E[\hat{\tau}] - \tau$$

来检查估计量是否是无偏的

但是...

最大似然估计是无偏的

一种较简单的方法是计算 $E[\hat{\tau}]$

$$E[\hat{\tau}(t_1,\ldots,t_n)] = \int\ldots\int \hat{\tau}(\vec{t})\, f_{\text{joint}}(\vec{t};\tau)\, dt_1 \ldots dt_n$$

$$= \int\ldots\int \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)\frac{1}{\tau}e^{-t_1/\tau}\ldots\frac{1}{\tau}e^{-t_n/\tau}\, dt_1\ldots dt_n$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\int t_i \frac{1}{\tau}e^{-t_i/\tau}dt_i \prod_{j\neq i}\int \frac{1}{\tau}e^{-t_j/\tau}dt_j\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\tau$$

$$= \tau$$

因此，$\hat{\tau}$ 是 τ 的一个无偏估计量。

更简单的方法是证明样本的平均值 $\bar{t}$ 是 $E[t]$ 的一个无偏估计量，而对于指数型概率密度函数，$E[t]=\tau$。

最大似然估计量的函数

假设指数概率密度函数可写为

$$f(t;\lambda)=\lambda e^{-\lambda t}$$

这里 λ=1/τ，是衰变常数或寿命的倒数。

在测量值是时间 ***t*** 的情况下，如何找出 λ 的最大似然估计量？

根据最大似然估计量的唯一性，当 λ 是 ***t*** 的等价参数时，

使 $L_\lambda(\lambda)$ 达到最大值的是$\lambda(\hat{\tau})$，变量$\hat{\tau}$也同时使 $L_\tau(\tau)$ 达到最大。

→ 函数 $\lambda(\theta)$ 的最大似然估计量是 $\hat{\lambda}=\lambda(\hat{\tau})$

最大似然估计的函数(续)

所以，对于衰变常数，有

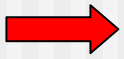
$$\hat{\lambda} = \frac{1}{\hat{\tau}} = \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right)^{-1}$$

那么$\hat{\lambda}$ 是λ 的一个无偏估计量吗？

一般而言，一个无偏估计量的非线性函数对参数的函数是有偏向性的。

对于$\hat{\lambda}$，可以证明

$$E[\hat{\lambda}] = \lambda \frac{n}{n-1}$$



$\hat{\lambda}$有偏置。但当$n \to \infty$时，该偏置将趋于零。

高斯概率密度函数中的参数

考虑一个样本服从高斯概率密度函数，其参数 μ，σ^2 未知。其对数似然函数为

$$\log L(\mu,\sigma^2)=\sum_{i=1}^{n}\log f(x_i;\mu,\sigma^2)=\sum_{i=1}^{n}\left(\log\frac{1}{\sqrt{2\pi}}+\frac{1}{2}\log\frac{1}{\sigma^2}-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

对 μ，σ^2 偏微分后的函数取零，解方程得到

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i,\qquad \widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

取“帽”号表示方程的解是参数的估计值。

前面已证明 $\hat{\mu}$ 是 μ 的无偏估计量。对于$\widehat{\sigma^2}$，我们有

$$E[\overline{\sigma^2}]=\frac{n-1}{n}\sigma^2$$

因此，σ^2 的最大似然估计量有偏向性。这种偏向性随 n 趋于无穷大时而消失。但是，

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

对任何概率密度函数的方差估计都是无偏的。

估计量的方差：数值方法

指数分布平均值的估计量为：

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

对应的概率密度函数 $g(\hat{\tau};\tau,n)$ 分布的宽度可以估计

$$V[\hat{\tau}] = E[\hat{\tau}^2] - (E[\hat{\tau}])^2$$

$$= \int\cdots\int \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)^2 \frac{1}{\tau}e^{-t_1/\tau}\cdots\frac{1}{\tau}e^{-t_n/\tau}dt_1\ldots dt_n$$

$$-\left(\int\cdots\int \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)\frac{1}{\tau}e^{-t_1/\tau}\cdots\frac{1}{\tau}e^{-t_n/\tau}dt_1\ldots dt_n\right)^2$$

$$= \frac{\tau^2}{n} \Rightarrow$$

$\hat{\tau}$ 的方差比 t 的方差小 n 倍。

估计量的方差：数值方法(续)

注意：对于未知真值 τ 的 $V[\hat{\tau}], \sigma_{\hat{\tau}}$ 函数，其估计可以采用

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{n}}$$

通常情况下，“统计误差”由上式给出。例如，

$$\hat{\tau} \pm \hat{\sigma}_{\hat{\tau}} = 1.062 \pm 0.150$$

这就意味着

最大似然法对 τ 的估计为1.062；
最大似然法对 $g(\hat{\tau};\tau,n)$ 的 σ 估计为 0.150。

如果 $g(\hat{\tau};\tau,n)$ 是高斯函数，则 $[\hat{\tau}-\hat{\sigma}_{\hat{\tau}}, \hat{\tau}+\hat{\sigma}_{\hat{\tau}}]$ 为所谓的

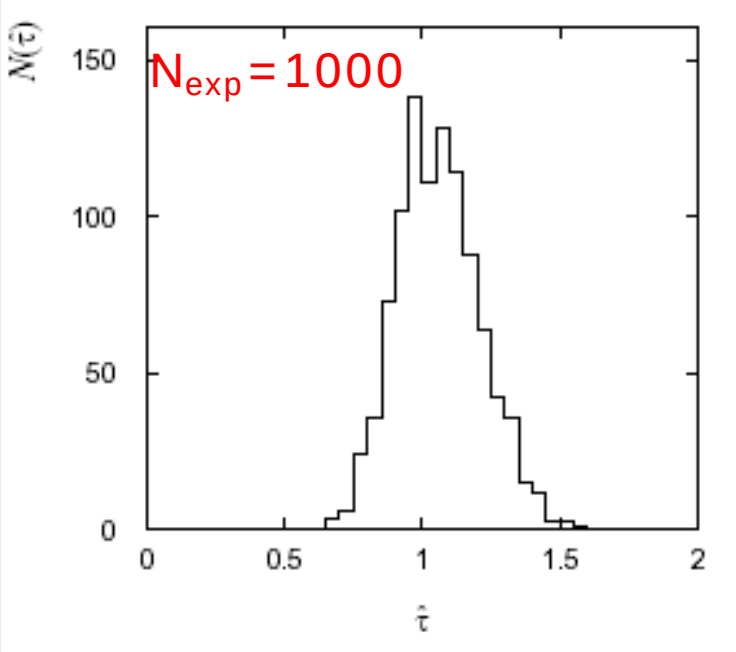
“68%的置信区间”

估计量的方差：蒙特卡罗方法

通常情况下，$\hat{\theta}$ 的具体形式 $g(\hat{\theta};\theta,n)$ 并不知道。对于此类情况，

可采用蒙特卡罗方法得到 $g(\hat{\theta};\theta,n)$

例如，对指数的pdf，我们有 $\hat{\tau}$=1.062。在蒙特卡罗中，把其作为 τ 的真值。产生 $n=50$ 的样本，并重复1000次实验。计算每次实验的 $\hat{\tau}$,并填入直方图。



蒙特卡罗实验可给出标准偏差

$$\hat{\sigma}_{\hat{\tau}}=\left[\frac{1}{N_{\exp}-1}\sum_{i=1}^{N_{\exp}}(\hat{\tau}_i-\bar{\hat{\tau}})^2\right]^{1/2}=0.151$$

类似于前面估计的 $\hat{\tau}/\sqrt{n}=0.150$。

注意：$g(\hat{\tau};\tau,n)$近似服从高斯分布。根据中心极限定理，可以断定在大样本极限下它确实是高斯分布。

RCF边界问题(信息不等式)

任何估计量(不仅仅是最大似然法)的方差下界为

$$V[\hat{\theta}] \geq \left(1+\frac{\partial b}{\partial \theta}\right)^2 \Big/ E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right]$$

(b为偏置)

这就是所谓的 Rao-Cramér-Frechet 不等式(信息不等式)。

如果等式满足，就可以说 $\hat{\theta}$ 是有效的。

最大似然估计量对大的样本统计量 ***n*** 几乎总是有效的。

通常假设上述结论为真，利用RCF边界估计 $V[\hat{\theta}]$

RCF边界问题(续一)

例如，对于前面的指数概率密度函数的例子，我们可以得到

$$\frac{\partial^2 \log L}{\partial \tau^2} = \frac{n}{\tau^2}\left(1 - \frac{2}{\tau}\frac{1}{n}\sum_{i=1}^{n} t_i\right) = \frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)$$

已知 $\boldsymbol{b} = \mathbf{0}$，所以

$$V[\hat{\tau}] \geq \frac{1}{E\left[-\frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)\right]} = \frac{1}{-\frac{n}{\tau^2}\left(1 - \frac{2E[\hat{\tau}]}{\tau}\right)} = \frac{\tau^2}{n}$$

=真的方差

→ 最大似然法的估计量 $\hat{\tau}$ 对任何样本统计量大小 n 都是有效的。

对于 $\vec{\theta}=(\theta_1,\ldots,\theta_m)$具有有效估计量以及零偏置的情况，

$$(V^{-1})_{ij} = E\left[-\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right] = -n\int f(x;\vec{\theta})\frac{\partial^2 \log f(x;\vec{\theta})}{\partial\theta_i \partial\theta_j}dx$$

→ 有效估计量的方差正比于1/n

RCF边界问题(续二)

在RCF边界里，$\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}$的期待值是参数真值的函数。可通过下式估计

$$\overline{(V^{-1})}_{ij} = -\left.\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right|_{\hat{\theta}=\vec{\theta}}$$

对于只有一个参数的情况，可以得到

$$\widehat{\sigma^2}_{\hat{\theta}} = \left.\left(-1/\frac{\partial^2 \log L}{\partial\theta^2}\right)\right|_{\theta=\hat{\theta}}$$

通常求 **log*L*** 的最大值是通过数值计算来完成，二阶导数的矩阵(也就是Hessian矩阵)是通过有限差值来估计。

调用 CERN 的 MINUIT 软件包中的 HESSE 程序

估计量的方差：图解法

考虑单参数 θ 情况下，将 $\log L(\theta)$ 在 $\hat{\theta}$ 附近展开，

$$\log L(\theta) = \log L(\hat{\theta}) + \left[\frac{\partial \log L}{\partial \theta}\right]_{\theta=\bar{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!}\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \ldots$$

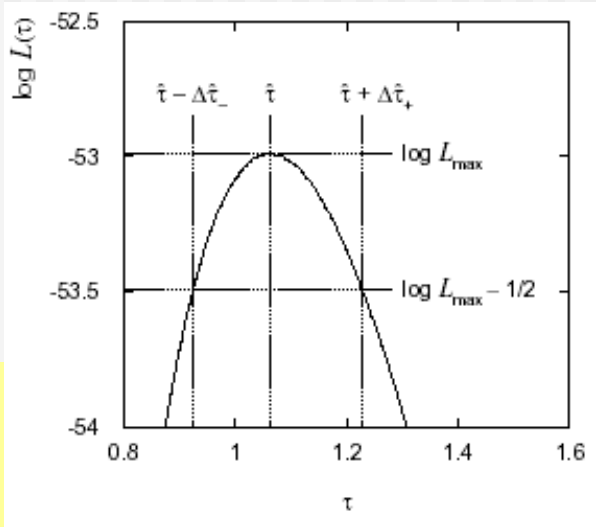
令 $\log L(\hat{\theta}) = \log L_{\max}$，并且上式第二项为零，因此有

$$\log L(\theta) = \log L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\sigma^2_{\hat{\theta}}}$$

也就是

$$\log L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \log L_{\max} - \frac{1}{2}$$

为了得到$\hat{\sigma}_{\hat{\theta}}$，可以让$\theta$偏离$\hat{\theta}$，使得$\log L$值减掉一个1/2数值。



指数函数例子

$\hat{\tau} = 1.062$

$\Delta\hat{\tau}_- = 0.137$

$\Delta\hat{\tau}_+ = 0.165$

$\hat{\sigma}_{\hat{\tau}} \approx \Delta\hat{\tau}_-$

$\approx \Delta\hat{\tau}_+ = 0.15$

不等精度观测结果的并合

如果 $(x_1,\ldots,x_n)$ 是对同一固定量 μ 的 n 个不等精度测量值，对应的标准误差是 $(\sigma_1,\ldots,\sigma_n)$，其中任一测量值 x_i 的分布是方差为 σ^2_i 的正态分布 $N(x_i,\mu,\sigma^2_i)$，而且各观测值相互独立，方差已知。则不等精度观测样本的似然函数为

$$L(\vec{x};\mu)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right]$$

因此，取对数并解似然方程

$$\frac{\partial \log L}{\partial \mu}=\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma_i^2}=0 \Rightarrow \hat{\mu}=\frac{\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}=\frac{\sum_{i=1}^{n}\omega_i x_i}{\sum_{i=1}^{n}\omega_i}$$

$$\widehat{\sigma^2_{\hat{\mu}}}=\frac{1}{\left(\sum_{i=1}^{n}\omega_i\right)^2}\sum_{i=1}^{n}\omega_i^2\sigma_i^2=\frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}$$

权重因子 $\omega_i=\frac{1}{\sigma_i^2}$

例子：求μ 寿命的世界平均值

世界上一共有五个实验精确测量了μ 的平均寿命(10^{-9} s)

τ_i	权重因子	$\tau_i-\hat{\tau}$	$\chi_i^2=\omega_i(\tau_i-\hat{\tau})^2$
2197.078 ± 0.073	187.65	0.048	0.432
2197.025 ± 0.155	41.62	–0.005	0.001
2196.95 ± 0.06	277.78	–0.08	1.778
2197.11 ± 0.08	156.25	0.08	1.000
2197.3 ± 0.3	11.11	0.27	0.810

$$\hat{\tau}=\sum_{i=1}^{5}\omega_i\tau_i\Big/\sum_{i=1}^{5}\omega_i=2197.03\times10^{-9}\text{秒},\quad \widehat{\sigma_{\hat{\tau}}}=\left(\sum_{i=1}^{5}\omega_i\right)^{-\frac{1}{2}}=0.04\times10^{-9}\text{秒}$$

$$\sum_{i=1}^{5}\chi_i^2\Big/(5-1)=1.005$$

与期待值相符。

小结

1. 似然函数，最大似然估计量

$L(\theta)$是已得到数据的联合概率密度函数，最大似然法在 L 处大处估计 $\hat{\theta}$ 。

2. 指数与高斯概率密度函数的参数确定：

所有问题均有解析解。$\widehat{\sigma^2}$有偏置(当 $n \to \infty$时，偏置将趋于零)。

3. 最大似然估计量的方差

a) 解析法：可能的时候，最好采用；

b) 蒙特卡罗法：有用，但可能费时；

c) RCF边界法：仅为不等式，但在样本足够大时，与最大似然趋于相等；

d) 图解法：让 θ 偏离 $\hat{\theta}$ 使得 $\log L$ 值刚好减掉一个 1/2数值。

4. 不等精度观测结果的并合

采用加权平均计算估计值与误差

习题

习题7.1 假设我们想通过质子与反质子弹性散射来研究反质子的极化，观测量为散射角 $x=\cos\phi$，已知对应的概率密度函数为 $f(x;\alpha)=\frac{1}{2}(1+\alpha x)$，其中 α 是反映反质子极化的参数。如果过去的实验已经测量出 α 在0.10 ± 0.02左右(即α估计值的相对误差为20%)，那么要想在统计上将相对误差减少到5%，总共需要多少个事例？

习题7.2 考虑一服从泊松分布的随机变量观测值 r，样本容量为n。请给出平均值 λ 的最大似然估计量。试证明该估计是无偏的，并找出它的方差，以及证明 λ估计值的方差等于最小方差的下限。