第一章 参数估计

§1.0　统计量简介

统计的核心任务是把样本中所包含的总体的信息提取出来，面对分布族$\{P_\theta, \theta \in \Theta\}$，其本质就是对参数$\theta$的推断和分析。

定义：设$X_1, X_2, \dots, X_n$为来自总体X简单样本，若样本的函数$T(X_1, X_2, \dots, X_n)$中不包含任何未知参数，则称此函数为统计量。

例如：设$X_1, X_2, \dots, X_n$是来自正态总体$N(\mu, \sigma_0^2)$简单样本，其中μ是未知参数，σ_0^2已知，则$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，$\frac{(n-1)S^2}{\sigma_0^2}$都是统计量，而

$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2$不是统计量，因为它含有未知参数$\mu$

常用统计量:

样本均值（Sample Mean）：

$$\boxed{\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\,.}$$

样本方差（Sample Variance）：

$$\boxed{S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\,.}$$

样本标准差（Sample Standard Deviation）：

$$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}.$$

样本矩（Sample Moment）

k阶原点矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,k=1,2\cdots.$$

k阶中心矩

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k,k=1,2\cdots.$$

由大数定律可知，上述各种样本原点矩、样本中心矩、样本方差等都依概率收敛到对应的总体原点矩、总体中心矩、总体方差。

顺序统计量

$X_1,X_2\ldots X_n$观察值$x_1,x_2,\ldots,x_n$按从小到大递增顺序排序，记为$x_{(1)},x_{(2)}\ldots,x_{(n)}$,它满足$x_{(1)}\le x_{(2)}\le\ldots\le x_{(n)}$。定义排在$k$位置的数$x_{(k)}(1\le k\le n)$为随机变量$X_{(k)}$的观察值，显然$X_{(1)}\le X_{(2)}\le\ldots\le X_{(n)}$。由于$X_{(k)}$是样本$X_1,X_2\ldots X_n$的函数，且不含任何未知参数，因而$X_{(k)}$是统计量，称$X_{(1)},X_{(2)},\ldots,X_{(n)}$为顺序统计量。

特别地，$X_{(1)}=\min\{X_1,X_2\ldots X_n\}$ 表示样本 $X_1,X_2,\ldots,X_n$ 最小的，其观察值是样本观察值中最小。而 $X_{(n)}=max\{X_1,X_2\ldots X_n\}$ 表示样本 $X_1,X_2,\ldots,X_n$ 最大，观察值是样本观察值中最大。

如果总体 X 的分布函数为 $F(x)$，则容易求得

$X_{(1)}$ 的分布函数为 $F_{X_{(1)}}(t)=1-[1-F(t)]^n$，

$X_{(n)}$ 的分布函数为 $F_{X_{(n)}}(t)=F^n(t)$。

基于顺序统计量，可以构造一些常用的统计量，如样本极差，样本分位数等。

样本极差定义为$R=X_{(n)}-X_{(1)}$，它是一个描述数据分散程度的常用统计量。与样本方差S^2和样本二阶中心矩$\hat{\sigma}^2$相比，它仅仅使用了样本的最小值和最大值，计算简便，这也决定了它只能用来粗略描述数据分散程度，没有S^2和$\hat{\sigma}^2$精细准确。

样本中位数的定义如下：

$$m_{0.5}=\begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n\text{是奇数} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)}+X_{\left(\frac{n}{2}+1\right)}\right) & n\text{是偶数} \end{cases}$$

它表示把样本观察值从小到大进行排序时处在中间的数，反映了样本的平均值。

更一般的是样本的$p(0 < p < 1)$分位数m_p, 其定义为

$$m_p = \begin{cases} X_{([np+1])}, & np\text{不是整数} \\ \dfrac{1}{2}\left(X_{(np)} + X_{(np+1)}\right) & np\text{是整数} \end{cases}$$

由$X_{(1)}$、$X_{(n)}$、$m_{0.25}$、$m_{0.5}$、$m_{0.75}$五个数字组成五数概括，反映数据的大致分布

例：设$X_1, X_2 \dots X_n$是来自总体X的简单样本，且总体方差$Var(X)$存在，试证：

（1）$E(\overline{X}) = E(X)$;（2）$Var(\overline{X}) = \frac{1}{n} Var(X)$;

（3）$E(S^2) = Var(X)$.

证明：（1） $E(\bar{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=E(X)$

（2） $Var(\bar{X})=Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)$

$$=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{1}{n}Var(X)$$

（3） $E(S^2)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$

$$=E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\right]$$

$$=\frac{1}{\mathrm{n}-1}\left[\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\bar{X}^2\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{\mathbf{n}-1}\left[nE(X^2) - nE(\bar{X}^2)\right] \\
&= \frac{n}{n-1}\left[E(X^2) - E(\bar{X}^2)\right] \\
&= \frac{n}{n-1}\{Var(X) + [E(X)]^2 - Var(\bar{X}) - [E(\bar{X})]^2\} \\
&= \frac{n}{n-1}[Var(X) - Var(\bar{X})] \\
&= \frac{n}{n-1}\left[Var(X) - \frac{1}{n}Var(X)\right] \\
&= Var(X)
\end{aligned}
$$

统计量既然是对样本的加工或压缩，在这个过程中可能有损失有关参数的一部分信息，现在问题是在这个过程中是否存在某些统计量，既起到压缩作用，又不损失参数的信息，这样的统计量称为充分统计量。

例：设样本的观察值为$x_1, x_2, \ldots, x_n$,则样本的联合分布函数为

$P(\mathrm{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n}) = \theta^s(1-\theta)^{n-s}$,其中$x_i = 0$或$1$，$s = \sum_{i=1}^{n} x_i$

给定s的条件下，样本的条件分布为

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | s) = \frac{1}{\binom{n}{s}}$$

定义：　设总体分布族为$\{P_\theta, \theta \in \Theta\}$, $T(X)$是统计量。如果在给定$T(X) = t$的条件下，X的条件分布与参数θ无关，则称统计量$T(X)$是参数θ的充分统计量(**Sufficient Statistics**)

一般情况下，利用条件分布证明统计量的充分性是比较困难的。但存在证明充分性的一个充分必要准则，这是下面的因式分解定理(**Factorization theorem**)。

定理：　设总体分布族为$\{P_\theta, \theta \in \Theta\}$,$T(X)$是充分的，当且仅当存在一个定义在$I \times \theta$上的函数$g(t,\theta)$及定义在$R^n$上的函数$h(x)$使得

$$p(x,\theta) = g(T(x),\theta)h(x)$$

对所有的$x \in R^n$都成立，其中I是$T(x)$的值域，$p(x,\theta)$是样本的联合概率密度函数或分布率。

例： 设$x_1, x_2, \dots, x_n$来自Poission分布总体$P(x;\lambda)$ 简单样本，求参数λ的充分统计量。

解 样本$x_1, x_2, \dots, x_n$的联合分布列为

$$p(x_1, x_2, \dots, x_n; \lambda) = \frac{e^{-n\lambda}}{x_1!\, x_2! \dots x_n!} \lambda^{\sum_{i=1}^n x_i}$$

取$T(x) = \sum_{i=1}^n x_i$, $g(t, \lambda) = e^{-n\lambda}\lambda^t$, $h(x) = \frac{1}{x_1! x_2! \dots x_n!}$,
则有

$$p(x;\lambda) = g(T(x), \lambda)h(x)$$

可知，$T(x) = \sum_{i=1}^n x_i$为参数λ的充分统计量

例：设$x_1, x_2, \cdots x_n$是来自正态总体$N(\mu, \sigma^2)$的简单样本，其中μ和σ^2都是未知的，令参数$\theta = (\mu, \sigma^2)$,试证明

$T(x) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$及

$S(x) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} (x_i - \overline{x})^2\right)$都是$\theta$的充分统计量。

证明： 样本联合密度函数为$p(x_1, x_2, \cdots, x_n; \theta)$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right)\right\}$$

若取$\boldsymbol{T(x)=\left(T_1(x),T_2(x)\right)=\left(\sum_{i=1}^n x_i,\sum_{i=1}^n x_i^2\right)}$,
$\boldsymbol{h(x)\equiv 1}$及

$$\boldsymbol{g(t,\theta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(t_2-2\mu t_1+n\mu^2\right)\right\}}$$

其中$\boldsymbol{t=(t_1,t_2)}$，则有

$$\boldsymbol{p(x_1,x_2,\cdots,x_n;\theta)=g(t,\theta)h(x)}$$

由因式分解定理可知，

$\boldsymbol{T(x)=\left(\sum_{i=1}^n x_i,\sum_{i=1}^n x_i^2\right)}$是$\boldsymbol{\theta}$的充分统计量。

又因为样本的联合密度函数可表示为

$$p(x_1,x_2,\cdots,x_n;\theta)$$
$$=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2\right]\right\}$$

若取$S(x)=\big(S_1(x),S_2(x)\big)=\left(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}(x_i-\overline{x})^2\right)$,

$h(x)\equiv 1$及

$$g(s,\theta)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\left[s_2+n\left(\frac{s_1}{n}-\mu\right)^2\right]\right\}$$

其中$\boldsymbol{s}=(\boldsymbol{s}_1,\boldsymbol{s}_2)$，则有

$$\boldsymbol{p}(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n;\boldsymbol{\theta})=\boldsymbol{g}(\boldsymbol{s},\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{x})$$

由因式分解定理可知，$\boldsymbol{S}(\boldsymbol{x})=\left(\sum_{i=1}^{n}\boldsymbol{x}_i,\sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^2\right)$也是$\boldsymbol{\theta}$的充分统计量。

例：设 $x_1, x_2, \cdots, x_n$ 是来自均匀分布总体 $U[a,b]$ 的简单样本，其中 a 和 b 都是未知的，且 $a < b$，令参数 $\theta = (a,b)$，求 θ 的充分统计量。

解：　样本的联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

$$= \frac{1}{(b-a)^n} I_{\{a \leqslant x_{(1)} \leqslant x_{(n)} \leqslant b\}}(x_1, x_2, \cdots, x_n)$$

$$= \frac{1}{(b-a)^n} I_{\{a \leqslant x_{(1)} \leqslant x_{(n)} \leqslant b\}}\left(x_{(1)}, x_{(n)}\right)$$

其中

$$I_A(u)=\begin{cases}1, & u\in A\\ 0, & u\notin A\end{cases}$$

称其为示性函数。

令$T(x)=\left(x_{(1)},x_{(n)}\right)$,

$$g(t,\theta)=\frac{1}{(b-a)^n}I_{\langle a\leqslant t_1\leqslant t_2\leqslant b|}(t_1,t_2),\ h(x)\equiv 1$$

其中$t=(t_1,t_2)$,则联合密度函数有分解式

$$p(x_1,x_2,\cdots,x_n;\theta)=g(t,\theta)h(x)$$

由因式分解定理可知，$T(x)=\left(x_{(1)},x_{(n)}\right)$ 是 θ 的充分统计量。

本例说明，一方面，参数分布族$\{P_{\theta}: \theta \in \Theta\}$中参数$\theta$的充分统计量$T$不唯一。一般情况下，若$T(x)$是充分统计量，$g(t)$是一一对应的实函数（可以是向量函数），则$g(T(x))$也是充分统计量。

另一方面，当参数维数和充分统计量维数相等的时候，应将充分统计量看成一个有机整体，不能拆分，认为充分统计量的分量是相应参数向量的充分统计量是错误的。

值得指出，对一个分布族$\{P_{\theta}: \theta \in \Theta\}$而言，充分统计量往往不唯一，都具有保留分布信息的特点，但在对样本进行加工处理剔除与参数无关的信息能力，或者说压缩数据能力方面各不相同。

例如，总体分布族是两点分布族$\{B(1,\theta): 0 < \theta < 1\}$，统计量$(x_1, x_2, \dots, x_n)$、$(\sum_{i=1}^{n-1} x_i, x_n)$、$\sum_{i=1}^{n} x_i$都是充分统计量。

上述统计量都保留了与参数$\boldsymbol{\theta}$有关的信息“1”的个数，但在剔除无关信息“1”的 位置方面，

$(x_1, x_2, \ldots, x_n)$最差，$(\sum_{i=1}^{n-1} x_i, x_n)$次之，而$\sum_{i=1}^{n} x_i$做的最好，它除了保留所有与$\boldsymbol{\theta}$有关的信息外，还剔除了所有与$\boldsymbol{\theta}$无关的信息，或者从压缩数据能力方面看，$\sum_{i=1}^{n} x_i$最强，再没有比$\sum_{i=1}^{n} x_i$更强的充分统计量了。这种统计量称为**最小充分统计量**。

§1.1　频率估计

参数估计： 所谓参数估计，就是对来自总体的样本$x_1, x_2, \cdots, x_n$构造合适的统计量$T(x_1, x_2, \cdots, x_n)$作为参数$q(\theta)$的估计，记为$\hat{q}(x_1, x_2, \cdots, x_n) = T(x_1, x_2, \cdots, x_n)$，也简称为$\hat{q}(x) = T(x)$，其中$x = (x_1, x_2, \cdots, x_n)$。

考虑n次独立重复实验，每次实验有m种可能的结果$D_1, D_2, \cdots, D_m$，每个结果D_i发生的概率$P(D_i) = p_i$是未知的，且$\sum_{i=1}^{m} p_i = 1$。用n_i表示n次独立重复试验结果D_i发生的次数，则$(n_1, n_2, \cdots, n_m)$的精确分布称为多项分布，其分布列为

$$p(n_1, n_2, \cdots, n_m) = \frac{n!}{n_1!\, n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

其中$\sum_{i=1}^{m} \boldsymbol{n_i} = \boldsymbol{n}$。

多项分布是二项分布推广。概率$\boldsymbol{p_i}$最直观估计是$\frac{\boldsymbol{n_i}}{\boldsymbol{n}}$，即$\widehat{\boldsymbol{p}}_i=\frac{\boldsymbol{n_i}}{\boldsymbol{n}}$，称$\widehat{\boldsymbol{p}}_i$为$\boldsymbol{p_i(i=1,2,\cdots,m)}$的频率替换估计 (***Frequency Substitution Estimate***)。

一般的，考虑$\boldsymbol{p_i}$的连续函数$\boldsymbol{q(p_1,p_2,\cdots,p_m)}$的估计问题。由频率替换原理，最自然的方法就是用样本频率$\frac{\boldsymbol{n_1}}{\boldsymbol{n}},\frac{\boldsymbol{n_2}}{\boldsymbol{n}},\cdots,\frac{\boldsymbol{n_m}}{\boldsymbol{n}}$来替换未知概率$\boldsymbol{p_1,p_2,\cdots,p_m}$，使用$\boldsymbol{q\left(\frac{n_1}{n},\frac{n_2}{n},\cdots,\frac{n_m}{n}\right)}$来估计$\boldsymbol{q(p_1,p_2,\cdots,p_m)}$。

例1.1.1： 设总体X服从正态分布$N(\mu,\sigma_0^2)$，其中σ_0^2已知，$x_1,x_2,\cdots,x_n$是简单样本。由于某种原因，原始数据$x_1,x_2,\cdots,x_n$已不存在，仅知道n个观察值中有m个数不小于1，求μ的频率替换估计。

解　由已知条件，需求概率$p=P\{X\geq 1\}$。因为

$$p=P\{X\geq 1\}=P\left\{\frac{X-\mu}{\sigma_0}\geq\frac{1-\mu}{\sigma_0}\right\}=1-\Phi\left(\frac{1-\mu}{\sigma_0}\right)$$

其中$\Phi(x)$是标准正态分布的分布函数，所以

$$\mu=1-\sigma_0\Phi^{-1}(1-p)$$

用频率$\frac{m}{n}$替换相应概率p，可得μ的频率替换估计为

$$\mu=1-\sigma_0\Phi^{-1}\left(1-\frac{m}{n}\right)$$

例1.1.2(Hardy-Weinberg模型)： 考虑具有两个等位基因的单个基因的遗传平衡问题。如果三种不同的基因型是可辨识的，分别用l_1, l_2, l_3表示，并假设个体三种基因型发生的概率分别为

$$p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2$$

其中$\theta(0 < \theta < 1)$表示等位基因之一发生的概率。独立观测n个个体，三种基因型出现的次数分别为n_1, n_2, n_3，试求参数θ的频率替换估计。

解： n个个体中，三种基因型发生的频率分别为

$$\hat{p}_1 = \frac{n_1}{n}, \hat{p}_2 = \frac{n_2}{n}, \hat{p}_3 = \frac{n_3}{n}$$

而θ有三种表达式，分别为

$$\theta = \sqrt{p_1},\ \theta = 1 - \sqrt{p_3},\ \theta = p_1 + \frac{p_2}{2}$$

由频率替换原理可得参数θ的三个不同的估计分别为

$$\hat{\theta}_1 = \sqrt{\frac{n_1}{n}}, \hat{\theta}_2 = 1 - \sqrt{\frac{n_3}{n}}, \hat{\theta}_3 = \frac{n_1}{n} + \frac{n_2}{2n}$$

§1.2　矩估计

矩估计法的主要思想是基于替换原理。设总体分布族为$\{P_\theta:\theta\in\Theta\},\theta=(\theta_1,\theta_2,\cdots,\theta_s)$，其$r(r\geq s)$阶原点矩$E_\theta(X^r)$存在，令

$$\mu_k=E_\theta\left(X^k\right)=g_k(\theta_1,\theta_2,\cdots,\theta_s),\ k=1,2,\cdots,r$$

选择其中$s(s\leq r)$个方程，不妨设是前$s(s\leq r)$个，若通过反解，可将θ_j表示成μ_i函数，即

$$\begin{cases} \boldsymbol{\theta_1} = h_1(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \cdots, \boldsymbol{\mu_s}) \\ \boldsymbol{\theta_2} = h_2(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \cdots, \boldsymbol{\mu_s}) \\ \qquad \cdots\cdots \\ \boldsymbol{\theta_s} = h_s(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \cdots, \boldsymbol{\mu_s}) \end{cases}$$

相应的样本前$\boldsymbol{r}$个原点矩为

$$\boldsymbol{A_k} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{X_i^k},\ \boldsymbol{k} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{r}$$

用样本矩$\boldsymbol{A_k}$替换相应总体矩$\boldsymbol{\mu_k}$，可得参数$\boldsymbol{\theta}$的估计为

$$
\begin{cases}
\hat{\boldsymbol{\theta}}_1 = h_1(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_s) = h_1\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i, \dfrac{1}{n}\sum_{i=1}^{n} x_i^2, \cdots, \dfrac{1}{n}\sum_{i=1}^{n} x_i^s\right) \\
\hat{\boldsymbol{\theta}}_2 = h_2(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_s) = h_2\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i, \dfrac{1}{n}\sum_{i=1}^{n} x_i^2, \cdots, \dfrac{1}{n}\sum_{i=1}^{n} x_i^s\right) \\
\cdots\cdots \\
\hat{\boldsymbol{\theta}}_s = h_s(\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_s) = h_s\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i, \dfrac{1}{n}\sum_{i=1}^{n} x_i^2, \cdots, \dfrac{1}{n}\sum_{i=1}^{n} x_i^s\right)
\end{cases}
$$

这种方法称为矩估计法，所得的估计量称为矩估计量(***Moment Estimate***)。

例1.2.1： 设总体X的二阶矩$\mu_2=E\left(X^2\right)$存在，$x_1,x_2,\cdots,x_n$是简单样本，试求总体均值$\mu=E(X)$和方差$\sigma^2=\mathrm{Var}(X)$的矩估计。

解　因为$\mu_1=E(X)=\mu$，$\mu_2=E(X^2)=\mathrm{Var}(X)+[E(X)]^2=\sigma^2+\mu^2$，解方程组

$$\begin{cases}\mu_1=\mu\\ \mu_2=\sigma^2+\mu^2\end{cases}$$

可得，$\mu=\mu_1,\sigma^2=\mu_2-{\mu_1}^2$。

用A_1和A_2分别替换μ_1和μ_2，可得总体均值μ和方差σ^2的矩估计分别为$\hat{\mu}=\bar{x}$及$\sigma^2=\frac{1}{n}\sum_{i=1}^n x_i^2-\bar{x}^2=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2$。

例1.2.2： 设总体X在$[a,b]$上服从均匀分布，其概率密度函数为

$$p(x;a,b)=\begin{cases}\dfrac{1}{b-a}, x\in[a,b]\\ 0, x\notin[a,b]\end{cases}$$

其中$b>a$，$x_1,x_2,\cdots,x_n$是简单样本，试求a和b的矩估计。

解　由于$\mu_1=E(X)=\frac{1}{2}(a+b), v_2=\mathrm{Var}(X)=\frac{1}{12}(b-a)^2$，解方程组

$$\begin{cases}\mu_1=\dfrac{1}{2}(a+b)\\ v_2=\dfrac{1}{12}(b-a)^2\end{cases}$$

可得$a=\mu_1-\sqrt{3v_2}$和$b=\mu_1+\sqrt{3v_2}$。

因此，用样本均值$\overline{x}$替换总体均值μ_1，样本二阶中心矩B_2替换总体二阶中心矩v_2，可得参数a和b的矩估计分别为$\hat{a}=\overline{x}-\sqrt{3}\hat{\sigma},\hat{b}=\overline{x}+\sqrt{3}\hat{\sigma}$，其中$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$。

注意：

1.总体存在适当阶的矩。

反例，考虑Cauchy分布，其密度函数为

$$f(x,\theta)=\frac{1}{\pi\left(1+(x-\theta)^2\right)},-\infty<x<+\infty$$

其各阶矩均不存在。

2.对相同的参数$q(\theta)$，存在多个矩估计。

例如，考虑总体是参数为λ的Poisson分布，即使λ总体的均值，又是总体的方差。

§1.3　极大似然估计

极大似然法是另一种常用的估计法，先举例说明其思想。

例1.3.1 已知甲、乙两射手命中靶心的概率为0.9和0.4，观察一张靶纸知10枪六中，且这张靶纸肯定是射手甲乙其中一人所射，问究竟是谁所射?

根据题意，可建立统计模型$\{P_\theta, \theta \in \Theta\}$，其中$\Theta = \{0.9, 0.4\}$，当$\theta$取定时，$P_\theta$具有两点分布形式。设$x_1, x_2, \cdots, x_{10}$是样本值，则射中的概率为

$$L(\theta) = \theta^{\sum_{i=1}^{10} x_i}(1-\theta)^{10-\sum_{i=1}^{10} x_i}, \theta \in \Theta$$

通过计算可知$L(0.9) \approx 0.00005, L(0.4) \approx 0.0005$，比较这两个值，可知射中的概率$L(\theta)$在$\theta = 0.4$时最大，因此更愿接受这张靶纸是乙所射的，即用0.4作为θ的估计，$\hat{\theta} = 0.4$。

从这个例子可以看出，极大似然法是基于这样一个统计原理：在一次试验中，某一事件已经发生，则必认为发生该事件的概率最大（小概率事件原理）。

设$x_1, x_2, \cdots, x_n$来自总体的样本，总体的分布族为$p(x, \theta), \theta \in \Theta$，则$x_1, x_2, \cdots, x_n$的联合分布为

$$L(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p(x_i, \theta)$$

其中$\boldsymbol{L}(\boldsymbol{\theta};\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$称之为似然函数（***Likelihood Function***），简记为$\boldsymbol{L}(\boldsymbol{\theta})$，它是定义在参数空间$\boldsymbol{\Theta}$上函数

如果存在$\widehat{\boldsymbol{\theta}}\in\boldsymbol{\Theta}$，使$\boldsymbol{L}(\boldsymbol{\theta})$达到最大，即

$$\boldsymbol{L}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{x});\boldsymbol{x}\right)=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\{\boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{x})\}=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{\prod_{i=1}^{n}\boldsymbol{p}(\boldsymbol{x}_i,\boldsymbol{\theta})\right\}$$

则称为参数$\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$的极大似然估计（***Maximum Likelihood Estimate***），简称为**MLE**。相应的，称$\boldsymbol{q}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{x})\right)$为$\boldsymbol{q}(\boldsymbol{\theta})$的**MLE**。

注意：

当总体X为离散随机变量时，$p(x,\theta)$为分布率（分布列）；而当总体为连续随机变量时，$p(x,\theta)$为概率密度函数。

从以上分析可知，求θ的MLE$\widehat{\theta}(x)$，等价于求$L(\theta,x)$在$\theta\in\Theta$上的最大值。

求极大似然估计的具体过程可归如下：

1. 当参数空间仅含有限个元时，可以用穷举法求使$\boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{x})$在$\boldsymbol{\Theta}$上取得最大的$\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$。

2. 当参数空间$\boldsymbol{\Theta}$包含$\boldsymbol{m}$维欧氏空间的一个域时，可以先对$\boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{x})$取对数，得对数似然函数$\ln\boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{x})$，令偏导数等于零，即

$$\frac{\partial \ln \boldsymbol{L}(\boldsymbol{\theta},\boldsymbol{x})}{\partial \boldsymbol{\theta}_{\mathrm{i}}}=\mathbf{0},\ \boldsymbol{i}=\mathbf{1},\mathbf{2},\cdots,\boldsymbol{n}$$

这个方程称为似然方程（**Likelihood Function**），求解似然方程可获得参数的**MLE**$\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$。

例1.3.2： 考虑n次独立重复试验，每次试验有m种可能结果$v_1, v_2, \cdots, v_m$，每个结果发生概率p_i是未知的，且$\sum_{i=1}^{m} p_i = 1$。 $x_1, x_2, \cdots, x_n$是进行n次试验得到的简单样本。求p_i的极大似然估计。

解　若用$n_i (i = 1, 2, \cdots, m)$表示样本$x_1, x_2, \cdots, x_n$中取值$v_i$的次数，则$\sum_{i=1}^{m} n_i = n$ 。此时，似然函数为

$$L(p_1, p_2, \cdots, p_m) = p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

注意限制$\sum_{i=1}^{m} n_i = n$和$\sum_{i=1}^{m} p_i = 1$。对数似然函数为

$$\ln L(p_1, p_2, \cdots, p_m) = \sum_{i=1}^{m-1} n_i \ln p_i + n_m \ln\left(1 - \sum_{i=1}^{m-1} p_i\right)$$

求偏导，可得似然方程组

$$\frac{\partial}{\partial p_i} \ln L(p_1, p_2, \cdots, p_m) = \frac{n_i}{p_i} - \frac{n_m}{1 - \sum_{i=1}^{m-1} p_i} = \frac{n_i}{p_i} - \frac{n_m}{p_m} = 0$$

$$i = 1, 2, \cdots, m-1$$

求解可得参数的极大似然估计为$\hat{p}_i = \frac{n_i}{n}, i = 1, 2, \cdots, m$

例1.3.3：考虑Hardy-Weinberg模型，$x_1, x_2, \cdots, x_n$是简单样本，记三种不同基因型发生的次数分别为n_1, n_2, n_3，$n = n_1 + n_2 + n_3$，试求等位基因之一发生的概率θ的极大似然估计。

解　似然函数为

$$L(\theta) = p_1^{n_1} p_2^{n_2} p_3^{n_3} = 2^{n_2} \theta^{2n_1+n_2} (1-\theta)^{n_2+2n_3}$$

对数似然函数为

$$\ln L(\theta) = n_2 \ln 2 + (2n_1 + n_2) \ln \theta + (n_2 + 2n_3) \ln(1-\theta)$$

求偏导，有

$$\frac{\partial \ln L(\theta)}{\partial \theta}=\frac{2n_1+n_2}{\theta}-\frac{n_2+2n_3}{1-\theta}$$

令$\frac{\partial \ln L(\theta)}{\partial \theta}=0$，可得的极大似然估计为

$$\hat{\theta}=\frac{n_1}{n}+\frac{n_2}{2n}$$

例1.3.4: 设总体X为正态分布族$\left\{N(\mu,\sigma^2):(\mu,\sigma^2)\in\Theta\right\}$，$x_1,x_2,\cdots,x_n$是简单样本，在参数空间$\Theta=\left\{(\mu,\sigma^2):\mu\in\mathrm{R},\sigma^2>0\right\}$和$\Theta_0=\left\{(\mu,\sigma^2):\mu\geq 0,\sigma^2>0\right\}$，分别求参数$\mu$和$\sigma^2$的极大似然估计。

解　首先考虑参数空间$\Theta=\left\{(\mu,\sigma^2):\mu\in\mathrm{R},\sigma^2>0\right\}$。

由于总体的密度函数为

$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\{N(\mu,\sigma^2):(\mu,\sigma^2)\in\Theta\}$，$x_1,x_2,\cdots,x_n$是简单样本，在所以，对数似然函数为

$$\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

分别对μ及σ^2求偏导数，可得似然方程组

$$\left.\begin{aligned}\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\\ -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0\end{aligned}\right\}$$

解方程组可得

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

由于$\ln L(\mu,\sigma^2)$在参数空间Θ上偏导数存在，且有唯一的驻点$(\hat{\mu},\hat{\sigma}^2)\in\Theta$，因此$\hat{\mu}$和$\hat{\sigma}^2$分别是$\mu$和$\sigma^2$的极大似然估计。

其次，考虑参数空间$\Theta_0=\{(\mu,\sigma^2):\mu\geq 0,\sigma^2>0\}$。若方程组的解中的第一个分量$\bar{x}\geq 0$，则驻点$(\bar{x},\hat{\sigma}^2)\in\Theta_0$，且$\ln L(\bar{x},\hat{\sigma}^2)=\sup\limits_{\theta\in\Theta_0}\{\ln L(\theta)\}$，所以$\hat{\mu}_0=\bar{x}$和$\hat{\sigma}_0^2=\hat{\sigma}^2$分别是参数$\mu$和$\sigma^2$的极大似然估计；若$\bar{x}<0$，则$(\bar{x},\hat{\sigma}^2)\notin\Theta_0$，因而$\bar{x}$和$\hat{\sigma}^2$就不再是参数$\mu$和$\sigma^2$的极大似然估计。由于$\bar{x}<0$，且对任意的$(\mu,\sigma^2)\in\Theta_0$，有

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\frac{n}{\sigma^2}(\bar{x}-\mu)<0$$

所以，$\ln L(\mu,\sigma^2)$是$\mu(\mu\geq 0)$的单调减函数，因而对数似然函数$\ln L(\mu,\sigma^2)$在参数空间Θ_0的点$\left(0,\frac{1}{n}\sum_{i=1}^{n}x_i\right)$处取得最大，由此可得$\hat{\mu}_0=0$和$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}x_i{}^2$分别是$\mu$和$\sigma^2$的极大似然估计。综上所述，在参数空间$\Theta_0$上，统计量$\hat{\mu}_0=\max(\bar{x},0)$和$\hat{\sigma}_0^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_0)^2$分别是参数$\mu$和$\sigma^2$的极大似然估计。

例1.3.5： 设$x_1, x_2, \cdots, x_n$是来自均匀分布总体$U[a,b]$的简单样本，试求参数a和b的极大似然估计。

解　似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, a \le x_{(1)} \le x_{(n)} \le b \\ 0, \text{其他}\end{cases}$$

参数空间$\Theta=\{(a,b): a<b\}$。从上式可知，$L(a,b)$只可能在范围$a \le x_{(1)} \le x_{(n)} \le b$内取得最大值，但两个偏导数:

$$\frac{\partial \ln L(a,b)}{\partial a} = \frac{n}{b-a}, \frac{\partial \ln L(a,b)}{\partial b} = -\frac{n}{b-a}$$

均不可能为零，导致似然方程组无法用求导法求参数a和b的极大似然估计，而直接使用定义。由于样本$x_1, x_2, \cdots, x_n$给定时，其最小$x_{(1)}$和最大$x_{(n)}$亦给定，而参数满足$a \le x_{(1)}$及$b \ge x_{(n)}$。

因此当$a = x_{(1)}, b = x_{(n)}$时，$(b-a)^n$取最小值$\left(x_{(n)} - x_{(1)}\right)^n$，

相应的，$L(a,b)$取得最大值$\frac{1}{\left(x_{(n)} - x_{(1)}\right)^n}$。所以，$a$和$b$的极大似然估计分别为$\hat{a} = x_{(1)}$和$\hat{b} = x_{(n)}$。

例1.3.6： 总体X的密度函数为

$$f(x)=\begin{cases}2e^{-2(x-\theta)}, x>\theta\\ 0, x\leq 0\end{cases}$$

其中$\theta>0$是未知参数。$x_1,x_2,\cdots,x_n$是来自总体的简单样本。求θ的极大似然估计。

解 似然函数为

$$L(\theta;x_1,x_2,\cdots,x_n)=\begin{cases}2^n e^{-\sum_{i=1}^{n}2(x_i-\theta)}, x_{(1)}>\theta\\ 0, \text{其他}\end{cases}$$

样本$x_1,x_2,\cdots,x_n$给定时，由于$L(\theta;x_1,x_2,\cdots,x_n)$是单调增函数，所以在$\hat{\theta}=x_{(1)}$处取得上确界。因此的$\theta$极大似然估计为$x_{(1)}$。

例1.3.7： 设$x_1,x_2,\cdots,x_n$是来自正态总体$N(\mu,\sigma^2)$的简单样本，其中μ和σ^2都是未知参数，对给定的正常数a，存在c，使得$P\{X\le c\}=aP\{x\ge c\}$成立，求c的极大似然估计。

解　由条件$P\{X\le c\}=aP\{x\ge c\}$可得

$P\{X\le c\}=\frac{a}{a+1}$，即　$P\left\{\frac{X-\mu}{\sqrt{\sigma^2}}\le\frac{c-\mu}{\sqrt{\sigma^2}}\right\}=\frac{a}{a+1}$

由于$X\sim N(\mu,\sigma^2)$，所以$\frac{X-\mu}{\sqrt{\sigma^2}}\sim N(0,1)$。因而$\frac{c-\mu}{\sqrt{\sigma^2}}$是标准正态分布的$\frac{a}{a+1}$分位数，即$\frac{c-\mu}{\sqrt{\sigma^2}}=z_{\frac{a}{a+1}}$，求得$c=\mu+z_{\frac{a}{a+1}}\sqrt{\sigma^2}$。

又由例1.3.4可知，μ和σ^2的极大似然估计分别为

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

因而的极大似然估计为$\hat{c}=\bar{x}+z_{\frac{a}{a+1}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

若参数$\boldsymbol{\theta}$有一个充分统计量$\boldsymbol{T(x)}$，且极大似然估计$\widehat{\boldsymbol{\theta}}$存在，则由因子分解定理得似然函数$\boldsymbol{L(\theta; x) = g(T(x),\theta)h(x)}$，所以$\widehat{\boldsymbol{\theta}}$是充分统计量$\boldsymbol{T(x)}$函数，这是极大似然估计的优点。而矩估计却不具有这样的性质，例如，均匀分布总体$\boldsymbol{U[a, b]}$中参数$\boldsymbol{a}$和$\boldsymbol{b}$的矩估计$\widehat{\boldsymbol{a}}$和$\widehat{\boldsymbol{b}}$并不是充分统计量$\boldsymbol{T(x) = \left(x_{(1)}, x_{(n)}\right)}$的函数。

若$\boldsymbol{\theta}$频率估计、矩估计、极大似然估计分别为$\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \widehat{\boldsymbol{\theta}}_3$，则$\boldsymbol{f(\theta)}$的频率估计、矩估计、极大似然估计分别为$\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}_1\right)$, $\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}_2\right)$, $\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}_3\right)$。

§1.4　估计优良性

假设用$T(x)$作为参数$q(\theta)$的估计量，评价估计优劣的一个自然准则可定义如下：

$$MSE_{\theta}(T)=R(\theta,T)=E\big(T(x)-q(\theta)\big)^2$$

称上式为均方误差(*Mean Squared Error*)，简记为MSE。

如果$R(\theta,T)<+\infty$，则MSE由T的均值和方差确定，即

$$R(\theta,T)=Var_{\theta}\big(T(x)\big)+b^{2}(\theta,T)$$

其中$b(\theta,T)=E_{\theta}\big(T(x)-q(\theta)\big)$

称$b(\theta,T)$为用$T(x)$估计$q(\theta)$产生的偏差(bias)。

例1.4.1： 设$x_1, x_2, \cdots, x_n$是来自正态总体$N(\mu, \sigma^2)$的简单样本，μ和σ^2都是未知，试求总体均值μ和方差σ^2的极大似然估计$\overline{x}$和$\hat{\sigma}^2$的均方误差。

解　令$\theta = (\mu, \sigma^2), \Theta = \left\{(\mu, \sigma^2): \mu \in R, \sigma^2 > 0\right\}$。由于

$$\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$$

所以

$$b(\mu,\overline{x})=0,Var_\theta(\overline{x})=\frac{\sigma^2}{n}$$

$$b(\sigma^2,\hat{\sigma}^2)=-\frac{\sigma^2}{n},Var_\theta\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right)=2(n-1)$$

从而

$$MSE_\theta(\overline{x})=Var(\overline{x})+b^2(\mu,\overline{x})=\frac{\sigma^2}{n}$$

$$MSE_\theta(\hat{\sigma}^2)=Var_\theta(\hat{\sigma}^2)+b(\sigma^2,\hat{\sigma}^2)=\frac{(2n-1)\sigma^4}{n^2}$$

设$S(x)$和$T(x)$是参数$q(\theta)$的两个估计，如果对所有$\theta \in \Theta$不等式

$$R(\theta, T) \leq R(\theta, S)$$

成立，且对某些θ严格不等式成立，则称T比S好，也说S是非容许的(Inadmissible)。

例1.4.2： 设$x_1, x_2, \cdots, x_n$是来自正态总体$N(\mu, \sigma^2)$的简单样本，μ和σ^2未知。在均方误差准则下，试判断总体方差σ^2的两个估计样本方差S^2和样本二阶中心矩$\hat{\sigma}^2$的优劣。

解 令$\Theta = \{(\mu, \sigma^2): \mu \in R, \sigma^2 > 0\}$。由于

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

所以

$$E_\theta\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1, Var_\theta\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$$

从而

$$MSE_\theta(S^2)=Var_\theta(S^2)+b^2\left(\sigma^2,S^2\right)=\frac{2\sigma^4}{n-1}$$

再由例1.4.1可知，$MSE_\theta(\hat{\sigma}^2)=\frac{(2n-1)\sigma^4}{n^2}$。所以，当$n>1$时，对所有$\theta\in\Theta$都有

$$MSE_\theta(\hat{\sigma}^2)<MSE_\theta(S^2)$$

因此，在均方误差标准下要比优。

从均方误差可知，我们自然希望估计的MSE越小越好。

用$\boldsymbol{G_q}$表示$\boldsymbol{q(\theta)}$所有可能估计组成的类，如果在$\boldsymbol{G_q}$中存在一个元$\boldsymbol{T^*}$使得对任一$\boldsymbol{T \in G_q}$，有

$$\boldsymbol{R(\theta, T^*) \le R(\theta, T)}$$

对所有的$\boldsymbol{\theta \in \Theta}$成立，则$\boldsymbol{T^*(x)}$应是$\boldsymbol{q(\theta)}$的最好估计。

遗憾的是，这样的估计T^*并不存在，因为倘若这样的估计$T^*(x)$存在，那么对任一$\theta_0 \in \Theta$，令$S(x) = q(\theta_0)$，则$R(\theta_0, S) = 0$，这样

$$R(\theta_0, T^*) = E_{\theta_0}\left(T^*(x) - q(\theta_0)\right)^2 \le R(\theta_0, S) = 0$$

即$T^*(x) = q(\theta_0)$。由θ_0的任意性，因此这样$T^*(x)$不存在

平凡估计
(*Trivial Estimate*)

由此可见，均方误差一致达到最小的最优估计并不存在，那么应如何评判和寻找优良的估计呢？方法之一是对估计提出一些合理性的要求，将那些诸如不合理的平凡估计排除在外，然后在满足合理性要求的估计类中寻找优良的估计。无偏性便是一种常用的合理性要求。

定义1.4.1 设统计模型为$\{P_\theta, \theta \in \Theta\}$，$q(\theta)$未知参数，$X_1, X_2, \cdots, X_n$是来自总体样本，$T$是一个统计量，如果对所有的$\theta \in \Theta$有

$$E\big(T(X)\big) = q(\theta)$$

成立，即$b(\theta, T) = 0$，则称$T(X)$是参数$q(\theta)$的无偏估计量（*Unbiased Estimate*）。

例1.4.3： 设$x_1, x_2, \cdots, x_n$是来自正态总体$N(\mu, \sigma^2)$的简单样本，求μ和σ^2的无偏估计。

解　由于

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$$

所以$E_\theta(\bar{x}) = \mu, E_\theta(\hat{\sigma}^2) = \frac{(n-1)\sigma^2}{n}$。由此可见，样本均值$\bar{x}$是总体均值$\mu$的无偏估计，样本二阶中心矩$\hat{\sigma}^2$是总体方差$\sigma^2$的有偏估计。

由于$E_\theta\left(\frac{n\hat{\sigma}^2}{n-1}\right)=\sigma^2$，所以$\sigma^2$的无偏估计为

$$\frac{n\hat{\sigma}^2}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=S^2$$

即样本方差S^2是总体方差σ^2的无偏估计。

注意:

（1）无偏估计可能不存在。

如果参数$q(\theta)$的无偏估计存在，则称$q(\theta)$是可估的。

若无特别声明，均认为$q(\theta)$是可估参数。

（2）对可估参数$q(\theta)$，无偏估计一般是不唯一的。

（3）无偏估计不一定是最好的估计，即它可能是非容许的。

（4）在函数变换下，无偏性可能消失，即对$\boldsymbol{\theta}$而言，

$\hat{\boldsymbol{\theta}}$是无偏的，但$q(\hat{\boldsymbol{\theta}})$可能是$q(\boldsymbol{\theta})$的有偏估计。

设$q(\boldsymbol{\theta})$是可估参数，令

$$U_q = \{T(x): E_\theta(T(x)) = q(\theta), Var_\theta(T(x)) < \infty, \theta \in \Theta\}$$

即U_q表示参数$q(\theta)$的所有无偏估计组成的类。

§1.5　一致最小方差无偏估计

设统计模型为$\{P_\theta, \theta \in \Theta\}$，$q(\theta)$是可估参数，$U_q$是$q(\theta)$的无偏估计类，$q(\theta)$的一致最小方差无偏估计定义如下。

定义1.5.1 如果存在无偏估计$T(x) \in U_q$，使得对任一的$S(x) \in U_q$，有

$$Var_\theta\big(T(x)\big) \le Var_\theta\big(S(x)\big)$$

对所有的$\theta \in \Theta$都成立

则称$T(x)$为$q(\theta)$的一致最小方差无偏估计，简称为**UMVUE**。

(*Uniformly Minimum Variance Unbiased Estimate*)

定理**1.5.1**（存在性）设参数$q(\theta)$是可估的，则$T(x) \in U_q$是$q(\theta)$一致最小方差无偏估计的充分必要条件是对任意的$T_0(x) \in U_0$，等式

$$E_0\left(T_0(x)T(x)\right) = 0$$

对所有的$\theta \in \Theta$都成立

推论　设$T_1(x)$和$T_2(x)$分别是可估函数$q_1(\theta)$和$q_2(\theta)$的一致最小方差无偏估计，则对任一常数a和b，$aT_1(x)+bT_2(x)$是$aq_1(\theta)+bq_2(\theta)$的一致最小方差无偏估计。

定理1.5.2（唯一性）设参数$q(\theta)$是可估的，且$T(x)$和$S(x)$都是$q(\theta)$的一致最小方差无偏估计，则对所有的$\theta\in\Theta$，有　$P_\theta\{T(x)=S(x)\}=1$

即在概率1下一致最小方差无偏估计是唯一的。

定理1.5.3 设统计模型为$\{P_\theta, \theta \in \Theta\}$，$S(x)$是其充分统计量，$\varphi(x)$是$q(\theta)$的无偏估计量，令

$$T(x) = E\big(\varphi(x)|S(x)\big)$$

则$T(x)$也是$q(\theta)$的无偏估计量，且对所有的$\theta \in \Theta$，有

$$Var_\theta\big(T(x)\big) \le Var_\theta\big(\varphi(x)\big)$$

证明：由条件期望的性质，得 迭代期望定律：E（E（X | Y））= E（X）

$$E(T(x)) = E\left(E(\varphi(x)|S(x))\right) = E(\varphi(x)) = q(\theta)$$

所以$T(x)$也是$q(x)$的无偏估计。

$$Var_\theta(\varphi(x)) = E_\theta(\varphi(x) - q(\theta))^2$$

$$= E_\theta(\varphi(x) - T(x) + T(x) - q(\theta))^2$$

$$= E_\theta(\varphi(x) - T(x))^2 + E_\theta(T(x) - q(\theta))^2$$

$$+ 2E_\theta\left((\varphi(x) - T(x))(T(x) - q(\theta))\right)$$

但$E_\theta\Big(\big(\varphi(x)-T(x)\big)\big(T(x)-q(\theta)\big)\Big)$

$$= E_\theta\{E_\theta[(\varphi(x)-T(x))(T(x)-q(\theta))|S(x)]\}$$

$$= E_\theta\{(T(x)-q(\theta))E_\theta[(\varphi(x)-T(x))|S(x)]\}$$

$$= 0$$

这是因为

$$E_\theta[(\varphi(x)-T(x))|S(x)]$$

$$= E_\theta[\varphi(x)|S(x)] - E_\theta[T(x)|S(x)] = T(x)-T(x)$$

$$= 0$$

这样$Var_\theta(\varphi(x)) = E_\theta(\varphi(x)-q(\theta))^2 = E_\theta(\varphi(x)-T(x))^2 + E_\theta(T(x)-q(\theta))^2 \geq E_\theta(T(x)-q(\theta))^2 = Var_\theta(T(x))$

由此定理可知，利用充分统计量可以降低无偏估计量的方差，因此，为了寻找UMVUE，可以通过取充分统计量的条件期望（它是充分统计量的函数且是无偏的）来缩小无偏估计类。

若令$U_q^T=\{E(\varphi(x)|T(x)):all\ \varphi(x)\in U_q\}\subset U_q$，则**UMVUE**若存在应属于$U_q^T$。而$U_q^T$实际上是由$U_q$中充分统计量的所有函数组成的类，这是因为若$h(T(x))\in U_q$，则由

$$E\left(h(T(x))|T(x)\right)=h(T(x))$$

可知$h(T(x))\in U_q^T$。这样可以在充分统计量的函数类U_q^T中寻找**UMVUE**。但可能不唯一。为了在概率意义下确定唯一性，还需对统计量提出合理要求，这便是统计量的完全性。

定义1.5.2 设$g(t)$是定义在统计量$T(x)$的值域上的任一实值函数，如果对所有的$\theta \in \Theta$，$E_\theta\big(g(T)\big) = 0$成立时，必成立$g(T) \equiv 0$，则称统计量$T(x)$是完全的($Complete$)。

例1.5.1 设$x_1, x_2, \cdots, x_n$是来自两点分布$b(1, \theta)$的样本$(0 < \theta < 1)$，证明$\bar{x}$是完全统计量。

证明　因为$\bar{x}$服从类似$b(n, \theta)$分布，所以

$$E_\theta\big(g(\bar{x})\big) = \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k}\theta^k(1-\theta)^{n-k}$$

$$= (1-\theta)^n \sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k}\left(\frac{\theta}{1-\theta}\right)^k$$

令$E_\theta\big(g(\bar{x})\big) = 0$，有

$$\sum_{k=0}^{n} g\left(\frac{k}{n}\right)\binom{n}{k}\left(\frac{\theta}{1-\theta}\right)^{k}=0$$

因为上式的左边是$\frac{\theta}{1-\theta}$的多项式，因此对所有的

$\theta\big(\theta \in (0,1)\big)$

即$g\left(\frac{k}{n}\right)=0, k=0,1,\cdots,n$

故对分布族$b(1,\theta)$而言，$\overline{x}$是完全统计量。

定理1.5.4 设$\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$是来自总体$\{P_\theta,\theta\in\Theta\}$的一个样本，其密度函数可表示为

$$p(x,\theta)=c(w)h(x)\exp\left\{\sum_{k=1}^{m}w_kT_k(x)\right\}$$

其中$w=w(\theta)=\big(w_1(\theta),\cdots,w_m(\theta)\big)\in\Omega\subset R^m$。

$\big(T_1(x),\cdots,T_m(x)\big)=\left(\sum_{i=1}^{n}T_1(x_i),\cdots,\sum_{i=1}^{n}T_m(x_i)\right)$,

如果Ω有内点，则统计量$\big(T_1(x),\cdots,T_m(x)\big)$是完全充分的。

例1.5.2： 设$x_1, x_2, \cdots, x_n$是来自两点分布$B(1, \theta)$总体的简单样本，其中$0 < \theta < 1$，求θ的完全充分统计量。

解　两点分布$B(1, \theta)$分布列为$p(x; \theta) = \theta^x (1-\theta)^{1-x}$，其中$x = 0, 1$。样本的联合分布列可分解为

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= (1-\theta)^n \exp\left\{\left(\ln\frac{\theta}{1-\theta}\right)\sum_{i=1}^{n} x_i\right\}$$

取$\boldsymbol{w(\theta)} = \ln\frac{\boldsymbol{\theta}}{\boldsymbol{1-\theta}}, \boldsymbol{c(\theta)} = (\boldsymbol{1}-\boldsymbol{\theta})^{\boldsymbol{n}}$,

$\boldsymbol{h(x_1, x_2, \cdots, x_n) \equiv 1}$，　$\boldsymbol{T(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} x_i}$。

由于$\boldsymbol{w(\theta)}$的值域$\boldsymbol{\Lambda} = (-\infty, +\infty)$中包含内点，所以由定理1.5.4可知$\boldsymbol{T(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} x_i}$是完全统计量。

例1.5.3　设总体X服从对数正态分布$LN(\mu,\sigma^2)$，$x_1,x_2,\cdots,x_n$是简单样本，求完全统计量。

解　对数分布密度函数为

$$p(x;\theta)=\frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\mu^2}{2\sigma^2}}\exp\left\{\frac{\mu}{\sigma^2}\ln x-\frac{1}{2\sigma^2}(\ln x)^2\right\}\cdot\frac{1}{x}$$

其中$\theta=(\mu,\sigma^2)\in\Theta=(0,+\infty)\times(0,+\infty),x>0$

因此样本的联合密度为

$$\begin{aligned}&p(x_1,x_2,\cdots,x_n;\theta)\\&=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{n\mu^2}{2\sigma^2}}\cdot\frac{1}{\prod_{i=1}^n x_i}\cdot\exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^n \ln x_i-\frac{1}{2\sigma^2}\sum_{i=1}^n(\ln x_i)^2\right\}\end{aligned}$$

这样 $\left(T_1(x),T_2(x)\right)=\left(\sum_{i=1}^n \ln x_i,\sum_{i=1}^n(\ln x_i)^2\right)$

$$w=w(\theta)=\left(\frac{\mu}{\sigma^2},-\frac{1}{2\sigma^2}\right)\in\Omega=(0,+\infty)\times(-\infty,0)$$

$$w:(0,+\infty)\times(0,+\infty)=\Theta\to\Omega$$

$$=(0,+\infty)\times(-\infty,0)$$

由于二维区域$\boldsymbol{\Omega}$有内点，所以$\left(\boldsymbol{T_1(x)}, \boldsymbol{T_2(x)}\right) = \left(\sum_{i=1}^{n} \ln \boldsymbol{x_i}, \sum_{i=1}^{n} (\ln \boldsymbol{x_i})^2\right)$是完全充分统计量。

有关应用定理1.5.4的说明：

从这个例子可以看出，实际上只需把总体的密度函数分解成$\boldsymbol{p(x,\theta)} = \boldsymbol{c(w)h(x)}\exp\left\{\sum_{k=1}^{m} \boldsymbol{w_k T_k(x)}\right\}$，

$\boldsymbol{T_k(x)}=\sum_{i=1}^{n} \boldsymbol{T_k(x_i)}$的形式可直接由指数函数的性质和样本的独立性获得。

定理1.5.5（Lehmann-Scheffe）

设$\boldsymbol{S}(\boldsymbol{x})$是完全充分统计量，$\boldsymbol{\varphi}(\boldsymbol{x})$是$\boldsymbol{q}(\boldsymbol{\theta})$无偏估计，则$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{E}_{\boldsymbol{\theta}}\left(\boldsymbol{\varphi}(\boldsymbol{x})|\boldsymbol{S}(\boldsymbol{x})\right)$是$\boldsymbol{q}(\boldsymbol{\theta})$的UMVUE，进一步如果对所有$\boldsymbol{\theta} \in \boldsymbol{\Theta}$，$\boldsymbol{Var}_{\boldsymbol{\theta}}\left(\boldsymbol{T}(\boldsymbol{x})\right) < \infty$，则$\boldsymbol{T}(\boldsymbol{x})$是$\boldsymbol{q}(\boldsymbol{\theta})$唯一的UMVUE。

证明 由定理1.5.3可知$T(x)$是$q(\theta)$的无偏估计，且对一切$\theta\in\Omega$，有

$$Var_{\theta}\big(T(x)\big)\le Var_{\theta}\big(\varphi(x)\big)$$

即$T(x)\in U_q^S$，因此只要能证明U_q^S中元在概率1下都是相同的即可。

设$T_1=h_1(S)$和$T_2=h_2(S)$是中任意两个元，则对一切$\theta\in\Omega$，有

$$E_{\theta}\big(h_1(S)-h_2(S)\big)=E_{\theta}\big(h_1(S)\big)-E_{\theta}\big(h_2(S)\big)=0$$

由于$\boldsymbol{S}$是完全统计量，根据定义一切$\boldsymbol{\theta} \in \boldsymbol{\Omega}$，有$\boldsymbol{P_\theta\{h_1(S) = h_2(S)\} = 1}$，这实际上说明$\boldsymbol{T(x) = E_\theta\big(\varphi(x)|S(x)\big)}$不依赖于无偏估计$\boldsymbol{\varphi(x) \in U_q}$。所以在概率1下，$\boldsymbol{T(x)}$是$\boldsymbol{q(\theta)}$唯一的一致最小方差无偏估计。

注：Lehmann-Scheffe定理实际上给出了两种寻找UMVUE的方法，但必须知道完全充分统计量$T(x)$

（1）若$h\big(T(x)\big)$是$q(\theta)$无偏统计量，则$h\big(T(x)\big)$也是$q(\theta)$的UMVUE。即寻找完全充分统计量的函数使之成为$q(\theta)$的无偏估计。

（2）若能获得$q(\theta)$的一个无偏估计量$\varphi(x)$，则$E\big(\varphi(x)|S(x)\big)$就是$q(\theta)$的UMVUE。

例1.5.4： 设$x_1,x_2,\cdots,x_n$是来自Poisson总体$P(\lambda)$的简单样本，λ是正的未知参数。试求$q(\lambda)=e^{-\lambda}$的一致最小方差无偏估计。

解　由于总体X的分布列可分解为

$$p(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}=\frac{1}{x!}\exp\{(\ln\lambda)x\},x=0,1,2,\cdots$$

从而样本$x_1,x_2,\cdots,x_n$的联合分布列有分解式

$$p(x_1,x_2,\cdots,x_n;\lambda)=e^{-n\lambda}\prod_{i=1}^{n}\frac{1}{x_i!}\exp\left\{(\ln\lambda)\sum_{i=1}^{n}x_i\right\}$$

取 $S(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}x_i$, $w(\lambda)=\ln\lambda$, $c(\lambda)=e^{-n\lambda}$, $h(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}\frac{1}{x_i!}$。由于$w(\lambda)=\ln\lambda$的值域为$\Lambda=(-\infty,+\infty)$，且$\Lambda$中有内点，则由定理1.5.4可知，统计量$S(x_1,x_2,\cdots,x_n)=\sum_{k=1}^{n}x_k$是完全充分的。

要直接寻找完全充分统计量$S(x_1,x_2,\cdots,x_n)$的某个函数$h\big(S(x_1,x_2,\cdots,x_n)\big)$，使之成为$q(\lambda)=e^{-\lambda}$的无偏估计是困难的，相对来说容易获得$q(\lambda)$的一个无偏估计。

为此，令$\varphi(x_1,x_2,\cdots,x_n)=I_{\{x_1=0\}}(x_1,x_2,\cdots,x_n)$，则

$$E_\lambda\big(\varphi(x_1,x_2,\cdots,x_n)\big)=P_\lambda\{x_1=0\}=e^{-\lambda}=q(\lambda)$$

因此$\varphi(x_1,x_2,\cdots,x_n)$是$q(\lambda)$的无偏估计。由定理1.5.5知

$$T(x_1,x_2,\cdots,x_n)=E_\lambda\big(\varphi(x_1,x_2,\cdots,x_n)|S(x_1,x_2,\cdots,x_n)\big)$$

就是$q(\lambda)$的一致最小方差无偏估计。

下面求条件期望，由于

$$E_\lambda(\varphi(x_1,x_2,\cdots,x_n)|S(x_1,x_2,\cdots,x_n)=s)$$

$$=P\left\{x_1=0|\sum_{i=1}^{n}x_i=s\right\}=\frac{P\{x_1=0\}P\{\sum_{k=2}^{n}x_k=s\}}{P\{\sum_{k=1}^{n}x_k=s\}}$$

由独立Poisson随机变量之和仍服从Poisson分布，有

$$\sum_{k=2}^{n}x_k\sim P\big((n-1)\lambda\big),\sum_{k=1}^{n}x_k\sim P(n\lambda)$$

代入可得

$$E_\lambda\left(\varphi(x_1,x_2,\cdots,x_n)\middle|\sum_{i=1}^{n}x_i=s\right)=\left(\frac{n-1}{n}\right)^s$$

因此有

$$E_\lambda\left(\varphi(x_1,x_2,\cdots,x_n)\middle|\sum_{i=1}^{n}x_i\right)=\left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n}x_i}$$

因而统计量$T(x_1,x_2,\cdots,x_n)=\left(\frac{n-1}{n}\right)^{n\bar{x}}$是$q(\lambda)=e^{-\lambda}$的一致最小方差无偏估计

例1.5.5 设总体X服从正态分布$N(\mu,\sigma^2)$，$\theta=(\mu,\sigma^2)$未知，$x_1,x_2,\cdots,x_n$是来自总体的样本，求参数μ和σ^2的UMVUE。

解　首先求完全充分统计量。由于

$$p(x;\theta)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\mu^2}{2\sigma^2}}\exp\left\{\frac{\mu}{\sigma^2}x-\frac{1}{2\sigma^2}x^2\right\}$$

由于$w=\left(\frac{\mu}{\sigma^2},-\frac{1}{2\sigma^2}\right)$的值域包含内点，所以由定理4.2可知完全充分统计量为

$$T(x)=\left(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2\right)$$

而我们已经知道$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$是$\mu$的无偏估计，且是完全充分统计量$T(x)$的函数，故当$\sigma^2$未知时，$\mu$的UMVUE为$\overline{x}$。

（注：无论σ^2是已知或未知，$\overline{x}$都是μ的UMVUE）

又$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$是$\sigma^2$的无偏估计，且是完全充分统计量$T(x)$的函数，故当$\mu$未知时，$\sigma^2$的UMVUE为样本方差$S^2$。

（注，当μ已知时，S^2不是σ^2的UMVUE）

例1.5.6 设总体X在$[0,\theta]$上服从均匀分布，其中是θ是未知参数，$x_1,x_2,\cdots,x_n$是来自总体的样本，试求参数θ的UMVUE。

解　由于

$$p(x_1,x_2,\cdots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n}, 0\le x_{(1)}\le x_{(n)}\le\theta\\ 0, otherwise\end{cases}$$

由因子分解定理可知$x_{(n)}=\max\{x_1,x_2,\cdots,x_n\}$是充分统计量。下证它也是完全的。

由$P\{x_{(n)} \le t\} = (P\{x_1 \le t\})^n$可知$x_{(n)}$的密度函数为

$$p(t;\theta) = \begin{cases} n\theta^{-n}t^{n-1}, 0 \le t \le \theta \\ 0, otherwise \end{cases}$$

对任何函数$g(t)$及$\theta > 0$，由

$$E_\theta\left(g\left(x_{(n)}\right)\right) = n\theta^{-n}\int_0^\theta g(t)t^{n-1}\mathrm{dt} = 0$$

可得对所有的$\theta>0$，有$\int_0^\theta g(t)t^{n-1}\mathrm{dt}=0$，这个只有在$g(t)\equiv 0$时才能成立，因而$x_{(n)}$也是完全的。

又因为

$$E_\theta\left(x_{(n)}\right)=\frac{n}{\theta^n}\int_0^\theta t^n\mathrm{dt}=\frac{n\theta}{n+1}$$

所以θ的无偏估计为

$$\widehat{\theta}=\frac{n+1}{n}x_{(n)}$$

是完全充分统计量$x_{(n)}$的函数，故它就是θ的UMVUE。

§1.6　信息不等式

在上一节，我们知道如果UMVUE存在，则它在无偏估计类中是最好的，且其方差不可能是零，因为参数$q(\theta)$的方差为零的平凡估计不是无偏估计。那么，现在的问题是：

对$q(\theta)$的无偏估计类U_q，在一定条件下

（1）既然无偏估计的方差不是零，则必然存在一个下界，这个下界到底是多少？

（2）若UMVUE存在，那么它的方差是否可以达到这个下界？

问题（1）已由**Cramer-Rao不等式（信息不等式）**揭示；问题（2）不一定成立，我们举例予以阐述。

设分布族为$\{P_\theta, \theta \in \Theta\}$，密度函数为$p(x,\theta)$，$\Theta$为直线上的一个开区间。满足下述条件的分布族$\{P_\theta, \theta \in \Theta\}$称为Cramer-Rao族。

（1）支撑$A=\{x:p(x,\theta)>0\}$与θ无关，且对任一$x\in A$，偏导数$\frac{\partial}{\partial\theta}ln\,p(x,\theta)$存在。

（2）若对所有$\theta\in\Theta$，$T(x)$是满足$E_\theta|T|<\infty$任一统计量，则对$T(x)p(x,\theta)$，积分和微分可交换次序，即

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x)p(x,\theta)dx_1\ldots dx_n$$
$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x)\frac{\partial}{\partial\theta}p(x,\theta)dx_1\ldots dx_n$$

当仅有（1）成立时，我们可以定义所谓的Fisher信息量（Fisher Information Number）

$$I(\theta)=E\left[\frac{\partial}{\partial\theta}\ln p(x,\theta)\right]^2 \quad (0\le I(\theta)\le\infty)$$

如果

$$\frac{d^2}{d\theta^2}\int_{-\infty}^{+\infty}p(x;\theta)dx=\int_{-\infty}^{+\infty}\frac{\partial^2 p(x;\theta)}{\partial\theta^2}dx$$

成立，可以证明 $I(\theta)=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln p(x;\theta)\right]$

例1.6.1 设总体分布是Poisson分布族，即

$$p(x,\theta)=\frac{\theta^x}{x!}e^{-\theta},x=0,1,\dots$$

则

$$\frac{\partial}{\partial\theta}\ln p(x,\theta)=\frac{x}{\theta}-1$$

$$\text{因而}I(\theta)=E\left(\frac{x}{\theta}-1\right)^2=Var\left(\frac{x}{\theta}\right)=\frac{1}{\theta}$$

如果$X_1,X_2,\dots,X_n$是来自总体的样本，可以证明

$$I(\theta)=nI_1(\theta)\text{，其中}I_1(\theta)=E\left(\frac{\partial}{\partial\theta}\ln p(X_1,\theta)\right)^2$$

统计量$T(x_1,x_2,\ldots,x_n)$是对样本$x_1,x_2,\ldots,x_n$进行加工或压缩，提取参数θ的信息过程，在这个过程中有可能损失的信息。若将统计量$T(x_1,x_2,\ldots,x_n)$的密度函数或分布列记为$p_T(t;\theta)$，则它所包含的的信息量为

$$I_T(\theta)=E_\theta\left[\frac{\partial}{\partial\theta}\ln p_T(t;\theta)\right]^2$$

因此，直观上应该有$I_T(\theta)\le I_n(\theta)$。

理论上，在一定的条件下可以证明这个不等式确实成立，而且等号当且仅当$T(x_1, x_2, \ldots, x_n)$是充分统计量时成立，说明加工处理中信息没有损失，这更进一步加深了对充分统计量的理解，也说明将$I(\theta)$作为信息多少的度量是合理的.

定理1.6.1（Information Inequality）

设$T(x)$是对$\theta \in \Theta$所有满足$Var_\theta(T(X)) <+\infty$的统计量，记$\varphi(\theta) = E_\theta(T(X))$。如果分布族是Cramer-Rao正则族，且$0 < I(\theta) <+\infty$，则对所有的$\theta \in \Theta$，$\varphi(\theta)$是可微的，且

$$Var_\theta(T(X)) \geq \frac{(\varphi^2(\theta))^2}{I(\theta)}$$

证明 由于对所有$\theta \in \Theta$，有

$$\varphi(\theta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} T(x) p(x,\theta) dx_1 \dots dx_n$$

等式两边求导可得

$$\varphi'(\theta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} T(x) \frac{\partial}{\partial\theta} p(x,\theta) dx_1 \dots dx_n$$

$$= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} T(x) \frac{\partial}{\partial\theta} (\ln p(x,\theta))\, p(x,\theta) dx_1 \dots dx_n$$

$$= E\left(T(x) \frac{\partial}{\partial\theta} \ln p(x,\theta)\right)$$

又因为对所有的$\theta \in \Theta$，有

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p(x,\theta) dx_1 \dots dx_n = 1$$

等式两边求导可得

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} p(x,\theta) dx_1 \dots dx_n = 0$$

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \frac{\partial \ln p(x,\theta)}{\partial \theta} p(x,\theta) dx_1 \dots dx_n = 0$$

这样就有

$$E\left(\frac{\partial \ln p(x,\theta)}{\partial \theta}\right) = 0$$

从而有

$$\begin{aligned}\varphi'(\theta) &= E\left(T(x)\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)\\ &= Cov\left(T(x),\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)\end{aligned}$$

有$|\varphi'(\theta)| = \left|Cov\left(T(x),\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)\right|$

$$\le \sqrt{Var(T(X))}\sqrt{Var\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)}$$

而

$$Var\left(\frac{\partial \ln p(x,\theta)}{\partial \theta}\right) = E\left(\frac{\partial \ln p(x,\theta)}{\partial \theta}\right)^2 = I(\theta)$$

所以有

$$|\varphi'(\theta)| \le \sqrt{Var(T(X))}\sqrt{I(\theta)}$$

即

$$Var_\theta(T(X)) \ge \frac{\left(\varphi'(\theta)\right)^2}{I(\theta)}$$

在信息不等式中，下界通过$\boldsymbol{\varphi(\theta)}$依赖于$\boldsymbol{q(\theta)}$，因它是$\boldsymbol{T(x)}$的数学期望，也就是说对不同统计量而言，下界是变化的，如果将此定理应用于参数$\boldsymbol{q(\theta)}$的无偏估计类$\boldsymbol{U_q}$，就有：

对参数$\boldsymbol{q(\theta)}$的任意无偏估计$\boldsymbol{T(x)} \in \boldsymbol{U_q}$，有

$$\boldsymbol{Var_\theta(T(X)) \geq \frac{\left(q'(\theta)\right)^2}{I(\theta)}}$$

特别地，当$q(\theta)=\theta$时，对任一$T(X)\in U_\theta$，有

$$Var_\theta(T(X))\geq\frac{1}{I(\theta)}$$

通常称量$\frac{1}{I(\theta)}$为Cramer-Rao下界。

注意：（1）在以上三个不等式中$I(\theta)=nI_1(\theta)$

通常将$I_1(\theta)$看成一次观察所能获得的关于参数θ的信息，即一个观测值X_1所含θ的信息，那么$I(\theta)$就表示样本$X_1, X_2, \ldots, X_n$所含的信息。

（2）在将定理应用于无偏估计类U_q时，一定要注意定理的条件是否满足。

例1.6.2 设$\boldsymbol{x_1, x_2, \ldots, x_n}$是来自总体的简单样本，其密度函数为

$$p(x;\theta)=\begin{cases} e^{-(x-\theta)}, & x>\theta \\ 0, & x\le\theta \end{cases}$$

其中参数$\theta>0$。证明$\hat{\theta}=x_{(1)}-\frac{1}{n}$是$\theta$的无偏估计，且$Var(\hat{\theta})<\frac{1}{nI(\theta)}$。

证明　首先证明$\hat{\theta}=x_{(1)}-\frac{1}{n}$是$\theta$的无偏估计。由样本的独立性和同分布性，有$x_{(1)}$的分布函数为

$$F_{x_{(1)}}(y;\theta)=P_\theta\{x_{(1)}\le y\}=1-P_\theta\{x_{(1)}>y\}$$
$$=1-[1-F(y;\theta)]^n$$

其中$F(y;\theta)$是总体X的分布函数。关于y求导，可得$x_{(1)}$的密度函数为

$$g(y;\theta)=\begin{cases} ne^{-n(y-\theta)}, & y>\theta \\ 0, & y\le\theta \end{cases}$$

因此，$x_{(1)}$的数学期望为

$$E_\theta\left(x_{(1)}\right) = n\int_0^{+\infty} y e^{-n(y-\theta)}\,dy$$

$$= \theta + \int_0^{+\infty} e^{-n(y-\theta)}dy = \theta + \frac{1}{n}$$

移项可得$E_\theta\left(x_{(1)} - \frac{1}{n}\right) = \theta$，从而$\hat{\theta} = x_{(1)} - \frac{1}{n}$

是θ的无偏估计。

下面计算$\widehat{\theta}=x_{(1)}-\frac{1}{n}$的方差。

$$Var(\widehat{\theta})=Var_{\theta}(x_{(1)})=E_{\theta}(x_{(1)}^{2})-[E(x_{(1)})]^{2}$$

$$=n\int_{0}^{+\infty}y^{2}e^{-n(y-\theta)}\,dy-\left(\theta+\frac{1}{n}\right)^{2}$$

$$=\theta^{2}+2\int_{0}^{+\infty}ye^{-n(y-\theta)}\,dy-\left(\theta+\frac{1}{n}\right)^{2}=\frac{1}{n^{2}}$$

由于

$$I(\theta)=E_\theta\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2=\int_0^{+\infty}e^{-(y-\theta)}\,dy=1$$

所以，C-R下界为$\frac{1}{nI(\theta)}=\frac{1}{n}$。从而，当样本容量$n>1$时，有

$$Var_\theta\left(x_{(1)}-\frac{1}{n}\right)<\frac{1}{nI(\theta)}$$

当一致最小方差无偏估计存在、不等式成立时一致最小方差无偏估计的方差未必能达到C-R下界，举例如下。

例1.6.3 设$X_1, X_2, \ldots, X_n$是来自正态总体$N(\mu, 1)$的一个简单样本。试求参数μ^2的UMVUE，并证明其方差大于信息不等式的下界。

解　由于

$$p(x, \mu) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{(x-\mu)^2}{2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{\mu^2}{2}\right\} exp\left\{-\frac{x^2}{2}\right\} exp\{\mu x\}$$

由定理1.5.3知完全充分统计量为$\sum_{i=1}^{n} X_i$，所以μ的

UMVUE为$\bar{X}$，且服从$N(\mu, \frac{1}{n})$，而由

$$\frac{1}{n} = Var(\bar{X}) = E(\bar{X}^2) - \left(E(\bar{X})\right)^2 = E(\bar{X}^2) - \mu^2$$

有$E\left(\bar{X}^2 - \frac{1}{n}\right) = \mu^2$，这样$\bar{X}^2 - \frac{1}{n}$是$\mu^2$无偏估计，

且是$\bar{X}$的函数，所以它是μ^2的UMVUE。为计算

UMVUE方差，令$Z = \sqrt{n}(\bar{X} - \mu)$

则Z服从标准正态分布$N(0,1)$。则

$$Var\left(\bar{X}^2-\frac{1}{n}\right)=Var(\bar{X}^2)$$

$$=\frac{1}{n^2}Var\left\{\left(Z+\sqrt{n}\mu\right)^2\right\}=\frac{2}{n^2}+\frac{4\mu^2}{n}$$

而

$$I_1(\mu)=E\left(\frac{\partial}{\partial\mu}\ln\frac{1}{\sqrt{2\pi}}exp\left\{-\frac{(x-\mu)^2}{2}\right\}\right)^2$$

$$=E\left(\frac{\partial}{\partial\mu}\left\{-\frac{(x-\mu)^2}{2}\right\}\right)^2=E(x-\mu)^2=1$$

所以

$$Var\left(\overline{X}^2-\frac{1}{n}\right)=\frac{2}{n^2}+\frac{4\mu^2}{n}>\frac{4\mu^2}{n}=\frac{((\mu^2)')^2}{nI_1(\mu)}$$

这说明μ^2的UMVUE的方差未达到信息不等式的下界。

如果参数$q(\theta)$的无偏估计的方差$\hat{q}(X)$取得信息不等式的下界，即

$$Var(\hat{q}(X))=\frac{(q'(\theta))^2}{I(\theta)}$$

则$\hat{q}(X)$必是参数$q(\theta)$的UMVUE。

例1.6.4 设$X_1, X_2, \ldots, X_n$来自正态总体$N(0, \sigma^2)$的一个简单样本。试求参数σ^2的UMVUE。

解 由于$I_1(\sigma^2) = -E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\ln p(X_1, \sigma^2)\right)$

$$= -E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\ln \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right)$$

$$= -E\left(\frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}\right) = \frac{1}{2(\sigma^2)^2}$$

从而$I_1(\sigma^2)=\frac{n}{2(\sigma^2)^2}$。由信息不等式知，对任一无偏估计$\hat{\sigma}^2(X)\in U_{\sigma^2}$，有

$$Var\left(\hat{\sigma}^2(X)\right)\geq\frac{\left(\left(\sigma^2\right)'\right)^2}{I(\theta)}=\frac{2\left(\sigma^2\right)^2}{n}$$

若取$\hat{\sigma}^2(X)=\frac{1}{n}\sum_{i=1}^{n}{X_i}^2$，由$\frac{{X_i}^2}{\sigma^2}$服从$\chi^2(1)$可知

$$\frac{n\hat{\sigma}^2(X)}{\sigma^2}=\sum_{i=1}^{n}\frac{{X_i}^2}{\sigma^2}\sim\chi^2(n)$$

所以

$$E\left(\frac{n\hat{\sigma}^2(X)}{\sigma^2}\right)=n,Var\left(\frac{n\hat{\sigma}^2(X)}{\sigma^2}\right)=2n$$

即

$$E(\hat{\sigma}^2(X)) = \sigma^2, Var(\hat{\sigma}^2(X)) = \frac{2(\sigma^2)^2}{n}$$

故

$$Var(\hat{\sigma}^2(X)) = \frac{2(\sigma^2)^2}{n} = \frac{\left((\sigma^2)'\right)^2}{I(\theta)}$$

从而$\hat{\sigma}^2(X) = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2$是参数$\sigma^2$的UMVUE。

定义1.6.1 设分布族$\{P_\theta, \theta \in \Theta\}$是C-R正则族，$q(\theta)$是可估参数，如果存在某无偏估计$\hat{q}(X) \in U_q$，其方差达到信息不等式的下界，即

$$Var(\hat{q}(X)) = \frac{(q'(\theta))^2}{I(\theta)}$$

则称$\hat{q}(X)$为$q(\theta)$的有效估计(Efficient Estimate)。

定义1.6.2 对参数$q(\theta)$的任一无偏估计$\hat{q}(X) \in U_q$，令

$$e(\hat{q}(X)) = \frac{(q'(\theta))^2}{I(\theta)} / Var(\hat{q}(X))$$

则称$e\big(\hat{q}(X)\big)$为估计$q(\theta)$的有效率（Efficiency）。

显然$0 \le e(\hat{q}(X)) \le 1$。因此，有效估计是有效率为1的无偏估计。

理论上可以证明只有在指数分布族

$$\left\{ p(x;\theta) = c(\theta)exp\left\{ \sum_{k=1}^{m} \omega_k(\theta) T_k(\theta) \right\} h(x), \theta \in \Theta \right\}$$

下，可估参数的有效估计才可能存在，且也并非任一可估参数都有有效估计，也就是说即使在指数分布族下，有效估计存在的情况少之又少。

定义1.6.3 设$\{T_n(X)\}$是参数$q(\theta)$的估计序列，如果对所有的$\theta \in \Theta$，都有$\lim\limits_{n\to\infty} E_\theta\big(T_n(X)\big) = q(\theta)$，则称$T_n(X)$为参数$q(\theta)$的渐进无偏估计（Asymptotic Unbiased Estimate）。

例如对正态总体$N(\mu, \sigma^2)$，我们知道$\hat{\sigma}_n^2$是总体方差σ^2的有偏估计，且

$$E(\hat{\sigma}_n^2) = \frac{n-1}{n}\sigma^2$$

这样有$\lim\limits_{n\to\infty} E(\hat{\sigma}_n^2) = \lim\limits_{n\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2$

故$\hat{\sigma}_n^2$是总体方差σ^2的渐进无偏估计。

定义1.6.4 设$q(\theta)$是可估参数，如果存在无偏估计序列$\hat{q}_n(X) \in U_q$，使得

$$\lim_{n\to\infty} e\left(\hat{q}(X)\right) = \lim_{n\to\infty}\frac{\left(q'(\theta)\right)^2}{I(\theta)}/Var(\hat{q}(X)) = 1$$

成立，则称$\hat{q}_n(X)$为$q(\theta)$的渐进有效估计（**Asymptotic Efficient Estimate**）。

例如$X_1, X_2, \ldots, X_n$是来自正态总体$N(0,\sigma^2)$的样本，由

例1.6.4知方差σ^2的有效估计是$\frac{1}{n}\sum_{i=1}^{n} {X_i}^2$。

由于

$$\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

所以

$$E\left(\frac{(n-1)S_n^2}{\sigma^2}\right) = n-1, Var\left(\frac{(n-1)S_n^2}{\sigma^2}\right) = 2(n-1)$$

即$E(S_n^2)=\sigma^2, Var(S_n^2)=\frac{2(\sigma^2)^2}{n-1}$

而Cramer-Rao下界为

$$\frac{1}{I(\sigma^2)}=\frac{2(\sigma^2)^2}{n}$$

这说明无偏估计S^2既不是σ^2的UMVUE，也不是有效估计。

但是

$$\lim_{n\to\infty} e(S_n^2) = \lim_{n\to\infty} \frac{\frac{1}{I(\sigma^2)}}{Var(S_n^2)}$$

故样本方差S_n^2是σ^2的渐进有效估计。需要说明的是当UMVUE的方差较大时，方差小的有偏估计也不失为一个好的估计。

§1.7　相合估计

当样本容量变大时，要求参数的估计量具有这种极限性质实际上是对估计量的基本要求，这就是下面要介绍估计量的相合性（Consistency）准则。

定义1.7.1 设$\hat{q}_n(X)=\hat{q}_n(X_1,\ldots,X_n)$是参数$q(\theta)$的任一估计序列，如果$\{\hat{q}_n\}$依概率收敛于参数真值$q(\theta)$，即对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|\hat{q}_n(X)-q(\theta)|>\varepsilon\}=0$$

则称$\hat{q}_n(X)$是$q(\theta)$的相合估计。(Consistent Estimate)

相合性只是反映了$n\to\infty$时估计量的性质，即大样本性质，当样本容量有限时是无意义的。一般情形下证明估计的相合性可使用定义或大数定律。

例1.7.1 设$X_1,\ldots,X_n$是来自$[0,\theta]$上均匀分布总体的一个简单样本，试证明θ的MLE估计$X_{(n)}$是其相合估计。

证明　可知$X_{(n)}$的密度函数为

$$p(t;\theta)=\begin{cases} n\theta^{-n}t^{n-1}, 0\le t\le\theta \\ 0, \qquad\qquad otherwise \end{cases}$$

且$E_\theta\left(X_{(n)}\right)=\dfrac{n\theta}{n+1},\lim\limits_{n\to\infty}E_\theta\left(X_{(n)}\right)=\lim\limits_{n\to\infty}\dfrac{n\theta}{n+1}=\theta$

所以$X_{(n)}$是θ渐进无偏估计。

又因为对$\forall\varepsilon>0$，有

$$P\{|X_{(n)}-\theta|>\varepsilon\}=P\{X_{(n)}\le\theta-\varepsilon\}$$
$$=n\theta^{-n}\int_0^{\theta-\varepsilon}t^{n-1}dt=\left(\frac{\theta-\varepsilon}{\theta}\right)^n$$

这样$\lim\limits_{n\to\infty}P\{|X_{(n)}-\theta|>\varepsilon\}=\lim\limits_{n\to\infty}\left(\frac{\theta-\varepsilon}{\theta}\right)^n=0$

故$X_{(n)}$是θ的相合估计。

下面的定理在证明估计的相合性时很有用。

定理1.7.1 如果$\hat{q}_n(X)$是参数$q(\theta)$相合估计，且函数$h(y)$在$y=q(\theta)$处连续，则$h(\hat{q}_n)$是$h(q(\theta))$的相合估计。

例1.7.2 在Hardy-Weinberg模型中同位基因之一发生频率$\boldsymbol{\theta}$的三个频率替换估计为

$$\hat{\theta}_1=\sqrt{\frac{n_1}{n}},\hat{\theta}_2=1-\sqrt{\frac{n_3}{n}},\hat{\theta}_3=\frac{n_1}{n}+\frac{n_2}{2n}$$

又因为相应的函数

$$\theta=\sqrt{p_1},\theta=1-\sqrt{p_3},\theta=p_1+\frac{p_2}{2}$$

都是连续函数，由大数定律知$\frac{n_i}{n}$是p_i的相合估计，

故由定理1.7.1知$\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3$都是$\theta$的相合估计

注：

（1）这里仅介绍（弱）相合性（依概率收敛），还有强相合性（依概率1收敛或几乎必然收敛）就不涉及。

（2）相合性本身不能说明估计达到某一可靠度时，要求样本容量至少为多少。

（3）对同一参数，满足相合性的估计也许有多个。

对于（3），当存在多个相合估计时，其优劣往往可通过比较其渐进分布的渐进方差的大小来进行，最常用的渐进分布为正态分布。

定义1.7.2 设$\hat{q}_n(X) = \hat{q}_n(X_1, \dots, X_n)$是$q(\theta)$估计序列，如果存在数列$\mu_n(\theta)$和$\sigma_n^2(\theta)$，对任意实数$x$，都有

$$\lim_{n\to\infty} P\left\{\frac{\hat{q}_n(X)-\mu_n(\theta)}{\sigma_n(\theta)} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} exp\left\{-\frac{u^2}{2}\right\} du$$

则称$\hat{q}_n$具有渐近正态性，简记为$\frac{\hat{q}_n(X)-\mu_n(\theta)}{\sigma_n(\theta)} \xrightarrow{L} N(0,1)$

渐进正态性并不能确定用

$$\boldsymbol{\Phi}((x-\mu_n(\theta))/\sigma_n(\theta))$$

近似$P\{\hat{q}_n(X)<x\}$达到某精度时的最小样本容量n

一般情形下，可取

$$\mu_n(\theta)=E\left(\hat{q}_n(X)\right),\sigma_n^2(\theta)=Var(\hat{q}_n(X))$$

（1）对频率替换估计，若$h(p)$有连续偏导数，则

$$\sqrt{n}\left(h\left(\frac{n_1}{n},\dots,\frac{n_k}{n}\right)-h(p_1,\dots,p_k)\right)\xrightarrow{L}N(0,\sigma_h^2)$$

其中$\sigma_h^2=\sum_{i=1}^k p_i\left(\frac{\partial h(p)}{\partial p_i}\right)^2-\left(\sum_{i=1}^k p_i\frac{\partial h(p)}{\partial p_i}\right)^2$

例1.7.3 在Hardy-Weinberg模型中，θ的三个频率替换估计

$$\hat{\theta}_1=\sqrt{\frac{n_1}{n}},\hat{\theta}_2=1-\sqrt{\frac{n_3}{n}},\hat{\theta}_3=\frac{n_1}{n}+\frac{n_2}{2n}$$

都是相合估计，究竟哪个最优呢？比较其渐进正态分布的方差。它们相应的函数为

$$h_1=\sqrt{p_1},h_2=1-\sqrt{p_3},h_3=p_1+\frac{p_2}{2}$$

从而有$\sqrt{n}(\hat{\theta}_i-h_i)\xrightarrow{L}N(0,\sigma_h^2)$

其中

$$\frac{\sigma_1^2}{n}=\frac{1}{4n}(1-\theta^2),\frac{\sigma_2^2}{n}=\frac{1}{4n}(1-(1-\theta)^2)$$

$$\frac{\sigma_3^2}{n}=\frac{1}{2n}\theta(1-\theta)$$

比较这三个渐进方差，可知$\frac{\sigma_3^2}{n}$最小，因此可以认为

$\hat{\theta}_3=\frac{n_1}{n}+\frac{n_2}{2n}$是$\theta$的较好估计，实际上$\hat{\theta}_3$是$\theta$的MLE。

在一定条件下，可证明矩估计，MLE也有渐进正态性

（2）设$h(A_1, A_2, \ldots, A_r)$是参数

$q(\theta) = h\big(\mu_1(\theta), \mu_2(\theta), \ldots, \mu_r(\theta)\big)$的矩估计，

若总体$2r$阶原点矩$\mu_{2r} = E_\theta(X^{2r})$有限，且函数

$h(y_1, y_2, \ldots, y_r)$具有连续偏导数，则

$$\sqrt{n}(h(A_1, A_2, \ldots, A_r) - h(\mu_1, \mu_2, \ldots, \mu_r)) \sim AN(0, \sigma_h^2)$$

其中

$$\sigma_h^2 = Var\left[\sum_{k=1}^{r} \frac{\partial}{\partial \mu_k} h(\mu_1, \mu_2, \dots, \mu_r) X^k\right]$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{r} \frac{\partial}{\partial \mu_i} h(\mu_1, \mu_2, \dots, \mu_r) \frac{\partial}{\partial \mu_j} h(\mu_1, \mu_2, \dots, \mu_r) Cov\left(X^i, X^j\right)$$

$$= \sum_{k=2}^{2r} b_k \mu_k - \left[\sum_{k=1}^{r} \mu_k \frac{\partial}{\partial \mu_k} h(\mu_1, \mu_2, \dots, \mu_r)\right]^2$$

$$b_k = \sum_{\substack{i+j=k \\ 1 \le i,j \le r}} \frac{\partial}{\partial \mu_i} h(\mu_1, \mu_2, \dots, \mu_r) \frac{\partial}{\partial \mu_j} h(\mu_1, \mu_2, \dots, \mu_r),$$

$$k = 2, 3, \dots, 2r$$

在很广的条件下，参数$\boldsymbol{q(\theta)}$的极大似然估计$\hat{\boldsymbol{q}}_{\boldsymbol{n}}$是渐进正态估计，其渐进方差$\dfrac{\boldsymbol{\sigma^2(\theta)}}{\boldsymbol{n}}=\dfrac{[\boldsymbol{q'(\theta)}]^2}{\boldsymbol{nI(\theta)}}$(C-R下界)，

或写成$\hat{\boldsymbol{q}}_{\boldsymbol{n}}\sim \boldsymbol{AN}\left(\boldsymbol{q(\theta)},\dfrac{[\boldsymbol{q'(\theta)}]^2}{\boldsymbol{nI(\theta)}}\right)$的形式。