第六章　参数估计

总体是由总体分布来刻画的.

总体分布类型的判断——在实际问题中, 我们根据问题本身的专业知识或以往的经验或适当的统计方法, 有时可以判断总体分布的类型.

总体分布的未知参数的估计——总体分布的参数往往是未知的,需要通过样本来估计. 通过样本来估计总体的参数, 称为参数估计, 它是统计推断的一种重要形式.

本章讨论:

1. 参数估计的常用方法.
2. 估计的优良性准则.
3. 若干重要总体的参数估计问题.

6.1　点估计

设总体分布的形式已知，但它的一个或多个参数未知，借助于从这个总体中抽取的一些样本来估计这些未知参数或者其函数的值，这种问题称为参数估计问题。

例如假设设总体分布$F_\theta(x)$的形式已知，θ为待估参数，$X_1, \cdots, X_n$为从此总体中抽取的一个样本，而$x_1, \cdots, x_n$为样本的观察值. 为此，构造适当的统计量$\hat{\theta}(X_1, \cdots, X_n)$(称其为$\theta$的估计量, Estimator)，在有了样本的观察值后，带入到$\hat{\theta}(X_1, \cdots, X_n)$得到$\theta$的估计值(Estimate) $\hat{\theta}(x_1, \cdots, x_n)$。

常见的参数估计方法有:

(1) 矩估计法

(2) 极大似然法

(3) 贝叶斯方法

这里我们主要介绍前面两种方法.

6.1.1　矩估计方法

矩是基于一种简单的“替换”思想建立起来的一种估计方法. 其基本思想是用样本矩估计总体矩. 由大数律，如果未知参数和总体的某个(些)矩有关系，我们很自然的构造未知参数的估计。

同以前的记法：

$$\text{样本}k\text{阶矩:}\quad a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

$$\text{总体}k\text{阶矩:}\quad \alpha_k = EX^k \qquad \mu_k = E(X - EX)^2$$

因此在k阶矩存在的情况下，有

$$a_k \stackrel{a.s}{\to} \alpha_k, \quad m_k \stackrel{a.s}{\to} \mu_k$$

从而我们可以使用a_k, m_k分别估计α_k, μ_k。设总体F包含k个未知参数$\theta_1, \cdots, \theta_k$：$F(x;\theta_1,\cdots,\theta_k)$，若方程组

$$\begin{cases} \alpha_1 = f_1(\theta_1, \cdots, \theta_k) \\ \qquad\vdots \\ \alpha_k = f_k(\theta_1, \cdots, \theta_k) \end{cases}$$

可以反解得到

$$\begin{cases} \theta_1 = g_1(\alpha_1, \cdots, \alpha_k) \\ \qquad\vdots \\ \theta_k = g_k(\alpha_1, \cdots, \alpha_k) \end{cases}$$

由大数律，我们可以得到参数$\theta_1, \cdots, \theta_k$的一个估计:

$$\begin{cases} \hat{\theta}_1 = g_1(a_1, \cdots, a_k) \\ \qquad\vdots \\ \hat{\theta}_k = g_k(a_1, \cdots, a_k) \end{cases}$$

这里我们用的都是原点矩α_k，当然也可以使用中心矩μ_k，或者两个都使用。在这种情况下，只需要把相应的总体矩换成样本矩。我们称这种估计方法为矩估计法，得到的估计量称为矩估计量。**矩估计方法应用的原则是：能用低阶矩处理的就不用高阶矩。**

矩估计法的优点是简单易行,并不需要事先知道总体是什么分布. 缺点是，当总体类型已知时，没有充分利用分布提供的信息. 一般场合下, 矩估计量不具有唯一性.

例6.1. 设$X_1, \cdots, X_n$为从总体$X \sim B(n,p)$中抽取的样本，求参数p的矩估计量。

解: 由于$EX = np$，因此p的一个矩估计量为

$$\hat{p} = \bar{X}.$$

例6.2. 设$X_1, \cdots, X_n$为从总体$X \sim N(a, \sigma^2)$中抽取的样本，求参数a, σ^2的矩估计量。

解: 由于

$$EX = a, \quad D(X) = \sigma^2$$

所以a, σ^2的一个矩估计量为

$$\hat{a} = \bar{X}, \quad \hat{\sigma}^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

我们知道$ES^2 = \sigma^2$，因此，σ^2的另一个矩估计量为$\hat{\sigma}^2 = S^2$. □

6.1.2 极大似然估计方法(MLE)

这种方法是基于如下的看法:

定义 6.1.1. *设样本X(不一定是简单样本)有概率函数$f(x,\theta)$，这里参数$\theta \in \Theta$，而当固定x时把$f(x,\theta)$看成为θ的函数，称为似然函数。*

当固定参数θ时，$f(x,\theta)$可以看成是得到样本观察值x的可能性，这样，当把参数θ看成变动时，也就得到了"在不同的θ值下能观察到x的可能性大小"；由于我们已经观察到了x，所以我们要寻求在哪一个θ的值下，使得能观察到x的可能性最大。这个θ的值我们称为极大似然估计值。即

定义 6.1.2. *设$X_1, \cdots, X_n$为从具有概率函数f的总体中抽取的样本，$x = (x_1, \cdots, x_n)$为样本的观察值。若$\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$为一个统计量，满足*

$$f(x, \hat{\theta}) = \sup_{\theta \in \Theta} f(x, \theta)$$

则称$\hat{\theta}$为参数θ的极大似然估计量(MLE)。若待估参数为θ的函数$g(\theta)$，则$g(\theta)$ 的极大似然估计量为$g(\hat{\theta})$。

求极大似然估计量相当于求似然函数的极大值。我们称

$$L(\theta) = f(x_1, \cdots, x_n; \theta)$$

为似然函数。在简单样本的情况下，

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

而把似然函数的对数称为对数似然函数:(在一些情况下，处理对数似然函数更方便)

$$l(\theta) = logL(\theta)$$

当似然函数为非单调函数时，我们可以求其聚点:

$$\frac{dl(\theta)}{d\theta} = 0 \quad (\text{或者}\frac{dL(\theta)}{d\theta} = 0)$$

然后判断此聚点是否是最大值点。简单总结为

求极大似然估计(MLE)的一般步骤是：
(1) 由总体分布导出样本的联合概率函数(或联合密度);
(2) 把样本联合概率函数(或联合密度)中自变量看成已知常数, 而把参数看作自变量, 得到似然函数$L(\theta)$;
(3) 求似然函数$L(\theta)$的最大值点(常常转化为求$lnL(\theta)$最大值点)，即得MLE;
(4) 在最大值点的表达式中, 用样本值代入就得参数的极大似然估计值.

例6.3. 设$X_1, \cdots, X_n$为从总体$X \sim N(a, \sigma^2)$中抽取的样本，求参数a, σ^2的极大似然估计量。

解: 易得对数似然函数为

$$l(a, \sigma^2) = c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 - \frac{n}{2}log(\sigma^2)$$

令

$$\begin{cases} \frac{\partial l(a,\sigma^2)}{\partial a} = 0 \\ \frac{\partial l(a,\sigma^2)}{\partial \sigma^2} = 0 \end{cases}$$

得到

$$\begin{cases} a = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - a)^2 \end{cases}$$

容易验证此聚点是唯一的最大值点，因此得到a, σ^2的极大似然估计量:

$$\begin{cases} \hat{a} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2. \end{cases}$$

□

例6.4. 设$X_1, \cdots, X_n$为从具有如下形式密度的总体中抽取的样本*:*

$$f(x; a, b) = \begin{cases} \frac{1}{b}exp\{-\frac{x-a}{b}\} & , x > a \\ 0 & x \le a \end{cases}$$

求参数a, b的极大似然估计量.

解: 易得似然函数为

$$L(a, b) = \prod_{i=1}^{n} f(x_1; a, b) = \frac{1}{b^n}exp\{-\frac{1}{b}\sum_{i=1}^{n}(x_i - a)\}I_{x_{(1)}>a}$$

在固定b时，显然似然函数为a的单调增函数，因此$L(a)$的聚点为$\hat{a}=x_{(1)}$。再令$\frac{\partial L(a,b)}{\partial b}=0$，得到$b=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{(1)})$，容易验证此解是最大值点。从而得到$a,b$的极大似然估计量:

$$\begin{cases}\hat{a}=X_{(1)}\\ \hat{b}=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-X_{(1)}).\end{cases}$$

□

例6.5. 设$X_1,\cdots,X_n$为从如下分布中抽取的简单样本，求θ的MLE.

$$f(x)=\frac{1}{x!(2-x)!}[\theta^x(1-\theta)^{2-x}+\theta^{2-x}(1-\theta)^x],\quad x=0,1,2;\quad \theta\in(0,\frac{1}{2})$$

解: 由题设知$f(x)$为离散型，其分布律为

X	0	1	2
P	$\frac{1}{2}[(1-\theta)^2+\theta^2]$	$2\theta(1-\theta)$	$\frac{1}{2}[(1-\theta)^2+\theta^2]$

若直接从此分布出发，则不能得到θ的MLE的显式表达。为此，我们重新参数化，记$\eta=2\theta(1-\theta)$. 则由题设知$\eta>1/2$。则

X	0	1	2
P	$\frac{1}{2}(1-\eta)$	η	$\frac{1}{2}(1-\eta)$

再记$n_i=\#\{X_1,\cdots,X_n$中等于i的个数$\}$, $\quad i=0,1,2$, 则得到似然函数为

$$L(\eta)=(\frac{1}{2}(1-\eta))^{n_0}\eta^{n_1}(\frac{1}{2}(1-\eta))^{n_2}=(\frac{1}{2}(1-\eta))^{n-n_1}\eta^{n_1}$$

求解并注意η的下界即得到η的MLE为

$$\hat{\eta}=\min\{\frac{n_1}{n},\frac{1}{2}\}$$

再由$\theta=\frac{1-\sqrt{1-2\eta}}{2}$得到$\theta$的MLE为

$$\hat{\theta}=\frac{1-\sqrt{1-2\hat{\eta}}}{2}$$

□

6.1.3 估计量的评选标准

我们看到对同一个参数，有多个不同的估计量，因此，评选不同估计量的优劣性是需要考虑的。

1. 无偏性

设$\hat{g}(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的一个估计量，若

$$E\hat{g}(X_1,\cdots,X_n)=g(\theta)$$

则称$\hat{g}(X_1,\cdots,X_n)$为$g(\theta)$的无偏估计量(Unbiased Estimator)。无偏性是对一个估计量的最基本的要求。无偏性能够消除系统误差，因此在有多个估计量可供选择时，我们优先考虑无偏估计量。

2. 有效性(Efficiency)

设$\hat{g}_1(X_1,\cdots,X_n)$和$\hat{g}_2(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的两个不同的无偏估计量，若对任意的$\theta\in\Theta$,有

$$Var(\hat{g}_1(X_1,\cdots,X_n))\le Var(\hat{g}_2(X_1,\cdots,X_n))$$

而且至少对某个$\theta_0\in\Theta$使得严格不等式成立。则称$\hat{g}_1$较$\hat{g}_2$有效。

3. 相合性和渐近正态性

定义 6.1.3. *设总体分布依赖于参数$\theta_1,\cdots,\theta_k,g(\theta_1,\cdots,\theta_k)$是待估参数函数。设$X_1,\cdots,X_n$为自该总体中抽取的样本，$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个估计量，如果对任意的$\epsilon>0$和$\theta_1,\cdots,\theta_k$的一切可能值都有*

$$\lim_{n\to\infty}P_{\theta_1,\cdots,\theta_k}(|T(X_1,\cdots,X_n)-g(\theta_1,\cdots,\theta_k)|\ge\epsilon)=0$$

我们则称$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个(弱)相合估计量(Consistent Estimator)。

相合性是对一个估计量的最基本的要求，如果一个估计量没有相合性，那么无论样本大小多大，我们也不能把未知参数估计到任意预定的精度。这种估计量显然是不可取的。

矩估计量是满足相合性的，极大似然估计量在很一般的条件下也是满足相合性的。

估计量是样本$X_1,\cdots,X_n$的函数，其确切的分布一般不是容易得到。但是，许多形式很复杂的统计量(未必是和)，当n很大时，其分布都渐近于正态分布，这个性质称为统计量的“渐近正态性”。

无偏性和有效性都是对固定的样本大小n而言的，这种性质称为估计量的“小样本性质”，而相合性和渐近正态性都是考虑在样本大小趋于无穷时的性质，这种性质称为“大样本性质”。

例6.6. 设从总体

X	0	1	2	3
P	$\theta/2$	θ	$3\theta/2$	$1-3\theta$

抽取的一个简单样本$X_1,\cdots,X_{10}$的观察值为$(0,3,1,1,0,2,0,0,3,0)$，

(1) 求θ的矩估计量$\hat{\theta}_M$和极大似然估计量$\hat{\theta}_L$，并求出估计值。

(2) 上述估计量是否为无偏的？若不是，请作修正.

(3) 比较修正后的两个估计量，指出那个更有效.

解：略.

6.1.4 最小方差无偏估计(MVUE)*

由有效性的定义，我们自然会问在一起可能的无偏估计里，能否找到具有最小方差的无偏估计量？如果存在这样的估计量，我们称其为最小方差无偏估计量，即

定义 6.1.4. *设$\hat{g}(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的一个无偏估计量，若对$g(\theta)$的任一无偏估计量$\hat{f}(X_1,\cdots,X_n)$, 都有*

$$Var(\hat{g}(X_1,\cdots,X_n)) \le Var(\hat{f}(X_1,\cdots,X_n)) \quad \forall\,\theta\in\Theta$$

则称$\hat{g}$为$g(\theta)$的最小方差无偏估计(MVUE)。

这里我们介绍一种求MVUE的方法:

Cramer-Rao不等式法

设样本有概率函数$f(x,\theta)$，为确定计，设$f(x,\theta)$为pdf(离散的情况类似)。参数θ为一维的，在$\Theta=(a,b)$(a,b可为无穷)上取值，$g(\theta)$为待估函数。设$\hat{g}(X)$为$g(\theta)$的一个无偏估计量，则有(假定以下推导所需的条件都满足)

$$E\hat{g}(X)=\int\hat{g}(x)f(x,\theta)dx=g(\theta)\quad\forall\,\theta\in\Theta$$

两边求导数，得到

$$\int\hat{g}(x)\frac{\partial f(x,\theta)}{\partial\theta}dx=g'(\theta)$$

注意到$\int f(x,\theta)dx=1$，对θ求导得到

$$\int\frac{\partial f(x,\theta)}{\partial\theta}dx=0$$

所以有

$$\int[\hat{g}(x)-g(\theta)]\frac{\partial f(x,\theta)}{\partial\theta}dx=g'(\theta)$$

$\Longleftrightarrow$

$$\int[\hat{g}(x)-g(\theta)]\sqrt{f(x,\theta)}\Big[\frac{1}{\sqrt{f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial\theta}\Big]dx=g'(\theta)$$

由Cauchy-Schwarz不等式得到

$$\begin{aligned}[g'(\theta)]^2 &\le \int[\hat{g}(x)-g(\theta)]^2f(x,\theta)dx\cdot\int\Big[\frac{1}{f(x,\theta)}\frac{\partial f(x,\theta)}{\partial\theta}\Big]^2f(x,\theta)dx\\ &= Var(\hat{g}(X))\cdot E[\frac{\partial logf(X,\theta)}{\partial\theta}]^2\end{aligned}$$

即

$$Var(\hat{g}(X)) \geq [g'(\theta)]^2 \Big/ E[\frac{\partial log f(X,\theta)}{\partial\theta}]^2$$

此即为$Cramer-Rao$不等式。

特别，

●当$g(\theta)=\theta$时，

$$Var(\hat{g}(X)) \geq 1 \Big/ E[\frac{\partial log f(X,\theta)}{\partial\theta}]^2$$

●当样本为简单样本时，$X_1,\cdots,X_n \sim f_\theta(x)$，则$f(x,\theta)=\prod_{i=1}^{n} f_\theta(x_i)$，容易得到

$$E[\frac{\partial log f(X,\theta)}{\partial\theta}]^2 = nE[\frac{\partial log f_\theta(X_1)}{\partial\theta}]^2$$

于是

$$Var(\hat{g}(X)) \geq [g'(\theta)]^2 \Big/ nI(\theta)$$

其中$I(\theta)=E[\frac{\partial log f_\theta(X_1)}{\partial\theta}]^2$.

在以上的推导中，需要满足很多条件，总计如下：

定理 6.1.1. *设$X_1,\cdots,X_n$为简单样本，总体有概率函数$f_\theta(x)$，参数$\theta\in\Theta=(a,b)$(a,b可为无穷). $g(\theta)$为(a,b)上的可微函数。设存在函数$G(t,\theta)$，使得*

1. $EG^2(X_1,\theta)<\infty, \quad \forall\,\theta\in\Theta$;

2. *对任意$\theta\in\Theta$，存在$\epsilon_\theta>0$，使得当$|\psi-\theta|<\epsilon_\theta$时，有*

$$|\frac{\partial log f_\psi(t)}{\partial\psi}| \leq G(t,\theta)$$

则当$\hat{g}(X)$为$g(\theta)$的一个无偏估计时，有

$$Var(\hat{g}(X)) \geq [g'(\theta)]^2 \Big/ nI(\theta)$$

利用C-R不等式求MVUE的方法: 首先由直观或者其他途径找一个可能是最好的无偏估计，然后计算其方差，看是否达到了C-R不等式的下界，若达到了，就是MVUE。同时，还要仔细验证不等式推导中的所有条件都要满足。

4. 估计的效率

若$\hat{g}$为$g(\theta)$的一个无偏估计，其方差为$Var(\hat{g})$，则称

$$e_{\hat{g}}(\theta)=\frac{[g'(\theta)]^2}{nI(\theta)} \Big/ Var(\hat{g})$$

为无偏估计$\hat{g}$的效率。一般有$e_{\hat{g}}(\theta) \leq 1$。当$e_{\hat{g}}(\theta) = 1$时，称$\hat{g}$为有效估计。

若$\hat{g}$为$g(\theta)$的一个相合渐近正态估计，有渐近方差$\sigma^2(\theta)$，则称

$$ae_{\hat{g}}(\theta) = \frac{[g'(\theta)]^2}{I(\theta)} \Big/ \sigma^2(\theta)$$

为$\hat{g}$(在θ处)的渐近效率。极大似然估计的渐近效率为1，而矩估计除了几个常见的例子外，渐近效率一般都远抵于1。通常人们所说的矩估计不如似然估计，大抵上就是指这个。

例6.7. 设$X_1, \cdots, X_n$为从总体$N(\theta, 1)$里抽取的简单样本，则$\bar{X}$为θ的$MVUE$。

解: 因为

$$I(\theta) = E[\frac{\partial log f_\theta(X_1)}{\partial\theta}]^2 = 1$$

所以由C-R不等式知θ的任一无偏估计的方差都不小于1/n，而$Var(\bar{X}) = 1/n$，因此$\bar{X}$为θ的一个MVUE。□

还有其他一些求MVUE的方法，详细地可以参考陈希孺的《数理统计教程》。

6.2 区间估计

6.2.1 置信区间

区间估计是用一个区间去估计未知的参数。其好处是把可能的误差用明显的形式表达出来。不难看出，这里要满足两个条件：

- 估计的可靠性，即θ要以很大的概率落在区间$[\underline{\theta}, \bar{\theta}]$里，i.e.，

$$P_\theta(\underline{\theta} \leq \theta \leq \bar{\theta}) = 1 - \alpha$$

- 估计的精度要尽可能高，即要求区间$[\underline{\theta}, \bar{\theta}]$要尽可能的短。

但这两个要求是相互矛盾的，因此区间估计的原则是在已有的样本资源限制下，找出更好的估计方法以尽量提高可靠性和精度。Neyman 提出了广泛接受的准则：**先保证可靠性，在此前提下尽可能提高精度**。为此，引入如下定义:

定义 6.2.5. 设总体分布$F(x, \theta)$含有一个或多个未知的参数θ，$\theta \in \Theta$，对给定的值$\alpha, (0 < \alpha < a)$，若由样本$X_1, \cdots, X_n$确定的两个统计量$\bar{\theta} = \bar{\theta}(X_1, \cdots, X_n)$ 和$\underline{\theta} = \underline{\theta}(X_1, \cdots, X_n)$，满足

$$P_\theta(\underline{\theta} \leq \theta \leq \bar{\theta}) =^{[注1]} 1 - \alpha \qquad \forall\, \theta \in \Theta$$

[注1]有时候，不能证明对一切θ等式成立，但知道不会小于$1 - \alpha$. 此时$1 - \alpha$称为置信水平(Confidence level)。这两个术语并不严格区分。

称$1-\alpha$为置信系数*(Confidence coefficient)*，而称$[\underline{\theta},\bar{\theta}]$为$\theta$的置信水平为$1-\alpha$的置信区间*(Confidence Interval)*。

区间估计就是在给定的置信水平之下，去寻找有优良精度的区间。

一般，我们首先寻求参数θ的一个估计(多数是基于其充分统计量构造的)，然后基于此估计量构造参数θ的置信区间，介绍如下:

1. 枢轴变量法 设待估参数为$g(\theta)$，

1. 找一个与待估参数$g(\theta)$有关的统计量T，一般是其一个良好的点估计*(*多数是基于充分统计量构造或者是通过MLE构造*)*；

2. 设法找出T与$g(\theta)$的某一函数$S(T,g(\theta))$的分布，其分布F要与参数θ无关*(*S即为枢轴变量*);*

3. 对任何常数$a<b$，不等式$a\le S(T,g(\theta))\le b$要能表示成等价的形式$A\le g(\theta)\le B$，其中A,B只与T,a,b有关而与参数无关；

4. 取分布F的上$\alpha/2$分位数$\omega_{\alpha/2}$和上$(1-\alpha/2)$分位数$\omega_{1-\alpha/2}$，有$F(\omega_{\alpha/2})-F(\omega_{1-\alpha/2})=1-\alpha$. 因此

$$P(\omega_{1-\alpha/2}\le S(T,g(\theta))\le\omega_{\alpha/2})=1-\alpha$$

由3我们就可以得到所求的置信区间.

例6.8. 设$X_1,\cdots,X_n$为从正态总体$N(\mu,\sigma^2)$中抽取得样本，求参数μ,σ^2的$1-\alpha$置信区间。

解：由于μ,σ^2的估计$\bar{X},S^2$而且满足[注2]

$$T_1=\sqrt{n}(\bar{X}-\mu)/S\sim t_{n-1}$$
$$T_2=(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$$

所以T_1,T_2就是我们所要寻求的枢轴变量，从而易的参数μ,σ^2的$1-\alpha$置信区间分别为

$$\mu:\qquad\left[\bar{X}-\frac{1}{\sqrt{n}}St_{\alpha/2}(n-1),\bar{X}+\frac{1}{\sqrt{n}}St_{\alpha/2}(n-1)\right]^{[注3]},$$

$$\sigma^2:\qquad\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]^{[注4]}.$$

[注2]参见定理**??**

[注3]由于t分布对称，因此不难证明此区间就是最短的区间。

[注4]由于χ^2分布不对称，因此此区间只是习惯的一个取法。另外，当μ已知时，需要修改样本方差S^2为

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

此时，$nS^2/\sigma^2\sim\chi^2_n$。

□

例6.9. 设$X_1, \cdots, X_n$为从正态总体$N(\mu_1, \sigma_1^2)$中抽取得样本，$Y_1, \cdots, Y_m$为从正态总体$N(\mu_2, \sigma_2^2)$中抽取得样本，两组样本相互独立。求参数$\mu_1 - \mu_2, \sigma_1^2/\sigma_2^2$的$1-\alpha$置信区间。

解：方法完全类似于前面的例子，主要利用定理**??**。此处略. □

例6.10. $X_1, \cdots, X_n$为从均匀总体$U(0, \theta)$中抽取得样本，求参数θ的$1-\alpha$置信区间。

解: 由于参数θ的充分统计量为$X_{(n)} = \max\limits_{1 \le i \le n} X_i$而且其概率密度为

$$f(t) = \frac{n}{\theta^n} t^{n-1} I_{0<t<\theta}$$

由此立得$X_{(n)}/\theta$有概率密度

$$p(t) = nt^{n-1} I_{0<t<1}$$

与参数θ无关。取实数$0 < a < b < 1$，使得

$$P(a \le \frac{X_{(n)}}{\theta} \le b) = 1 - \alpha$$

$\Longrightarrow$

$$b^n - a^n = 1 - \alpha$$

从而在尽可能提高精度的提前下，可以使用数值解法就出最短的区间$[a_0, b_0]$，进而得到参数θ的$1-\alpha$置信区间：

$$\left[\frac{X_{(n)}}{b_0}, \frac{X_{(n)}}{a_0}\right].$$

□

2. 大样本法

大样本法就是利用极限分布，主要是中心极限定理，以建立枢轴变量。通过以下例子说明:

例6.11. 某事件A在每次实验中发生的概率都是p，作n次独立的实验，以Y_n记A发生的次数。求p的$1-\alpha$置信区间。

解：设n比较大，则由中心极限定理知，$(Y_n - np)/\sqrt{npq} \sim AN(0,1)$，从而$(Y_n - np)/\sqrt{npq}$可以作为枢轴变量。由

$$P(-u_{\alpha/2} \le (Y_n - np)/\sqrt{npq} \le u_{\alpha/2}) \approx 1 - \alpha \qquad (*)$$

可以等价表示成

$$P(A \le p \le B) \approx 1 - \alpha$$

其中A, B为方程

$$(Y_n - np)/\sqrt{npq} = u_{\alpha/2}$$

的解, 即

$$A, B = \frac{n}{n + u_{\alpha/2}^2}\left[\hat{p} + \frac{u_{\alpha/2}^2}{2n} \pm u_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{u_{\alpha/2}^2}{4n^2}}\ \right]$$

A取负号，B取正号，$\hat{p} = Y_n/n$。□

由于(*)式只是近似成立，故区间估计也只是近似成立，当n较大时才相去不远。详细的说明参见课本p203。

这种方法得到的置信区间称为“Score”/Approximate Interval, 我们还可以视方差是”已知”的，最后再将其估计，故得到Wald Confidence Interval：

$$\hat{p} \pm u_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

另外还有Adjusted Wald Confidence Interval, Exact/Clopper-Pearson Interval 等[注5].

6.2.2 置信界

在实际中，有时我们只对参数θ的一端的界限感兴趣。

定义 6.2.6. *设总体分布$F(x, \theta)$含有一个未知的参数θ，$\theta \in \Theta$，对给定的值$\alpha, (0 < \alpha < a)$，若由样本$X_1, \cdots, X_n$确定的两个统计量$\bar{\theta} = \bar{\theta}(X_1, \cdots, X_n)$ 和$\underline{\theta} = \underline{\theta}(X_1, \cdots, X_n)$，*

1.若

$$P_\theta(\theta \le \bar{\theta}) \geqq 1 - \alpha \qquad \forall\, \theta \in \Theta$$

则称$\bar{\theta}$为θ的一个置信系数为$1-\alpha$的置信上界(upper confidence bound)。

2.若

$$P_\theta(\theta \ge \underline{\theta}) \geqq 1 - \alpha \qquad \forall\, \theta \in \Theta$$

则称$\underline{\theta}$为θ的一个置信水平为$1-\alpha$的置信下界(lower confidence bound)。

而$(-\infty, \bar{\theta}]$ 和$[\underline{\theta}, +\infty)$都称为是单边的置信区间(One-Sided Confidence Interval).

寻求置信上、下界的方法和寻求置信区间的方法完全类似。

例6.12. 设$X_1, \cdots, X_n$为从期望为λ^{-1}的指数总体中抽取的样本，求参数λ的$1-\alpha$置信上、下界。

[注5]Jeff Sauro, James R. Lewis, 2005, Estimating completion rates from small samples using Binomial confidence intervals: comparisons and recommendations, proceedings of the human factors and ergonomics society 49th annual meeting.

解：由于$\sum\limits_{i=1}^{n} X_i$为参数$\lambda$的充分统计量，而且

$$2\lambda\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

因此易得λ的$1-\alpha$置信上、下界分布为

$$\bar{\lambda} = \frac{\chi^2_{\alpha}(2n)}{2\sum\limits_{i=1}^{n} X_i}, \qquad \underline{\lambda} = \frac{\chi^2_{1-\alpha}(2n)}{2\sum\limits_{i=1}^{n} X_i}.$$

□

例6.13. 设$Y_1,\cdots,Y_n\ i.i.d \sim B(1,p)$，$n$已知且比较大。求参数$p$的$1-\alpha$置信下界。

解：由中心极限定理及大数律知

$$\frac{Y-np}{\sqrt{npq}} \sim AN(0,1), \quad S \xrightarrow{a.s.} \sqrt{pq}$$

其中$Y=\sum\limits_{i=1}^{n} Y_i$，$S$为样本标准差，简单计算得到$S^2=\frac{Y(n-Y)}{n(n-1)}$。易知有

$$\frac{Y-np}{\sqrt{n}S} \sim AN(0,1)$$

因此有

$$P(\frac{Y-np}{\sqrt{n}S} \le u_{\alpha}) \approx 1-\alpha$$

从而得到p的渐近$1-\alpha$置信下界

$$\underline{p} = \frac{1}{n}[Y-\sqrt{n}Su_{\alpha}].$$

□

6.2.3 确定样本大小

置信区间越窄就越好，为什么呢？作为一个一般的原则，我们已经知道更多的测量可以得到更精确的推断。有时候，对精度是有要求的，甚至于是在测量之前就提出此要求，因此相应的样本大小就要事先确定下来。我们以如下的例子说明如何确定样本大小，一般的方法类似。

例6.14. 假设某种成分的含量服从正态分布$N(\mu,\sigma^2)$，σ^2已知。要求平均含量μ的$(1-\alpha)\%$置信区间的长度不能长于ω。试确定测量样本大小。

解: 由于σ^2已知，我们已经知道可以根据$\bar{X} \sim N(\mu, \sigma^2/n)$来构造$\mu$的95% 置信区间。因此易知区间长度为$2u_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. 从而由

$$2u_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \omega$$

得到

$$n \geq \left(\frac{2u_{\alpha/2}\sigma}{\omega}\right)^2.$$

比如当$\sigma = 0.1, \omega = 0.05, \alpha = 0.05$, 可以得到$n \geq \left(\frac{2*1.96*0.1}{0.05}\right)^2 = 61.4656$. 即为达到要求至少需要测量62次。

6.3 总结

在参数点估计理论里, 矩估计是一种直接的估计方法, 但由于其效率一般比较低，而在实际中用的不是很多。相对地，由于极大似然估计的渐近效率为1，且一般具有良好的性质，而在实际中得到广泛的应用。

无偏性是对一个估计量的基本要求。因此，在实际中，总是优先考虑无偏估计量。$Cramer-Rao$尽管是用$Cramer$和Rao两个著名统计学家的名字命名，但是却是最早由$Fisher$在1922年提出的。实际上，参数估计的大部分方法和估计量的优良性准则，都是$Fisher$建立的。

区间估计一般是通过参数的极大似然估计量来构造。求出极大似然估计量后和参数作某种变换后分布于参数无关，则得到了枢轴变量。然后再构造置信区间。对离散型总体，一般MLE和参数作某种变换没有确切的和参数无关的分布，而当样本容量很大时，可以利用大样本方法进行构造。

参考文献

[1] 陈希孺, 概率论与数理统计, 合肥: 中国科学技术大学出版社, 1995.

[2] 陈希孺, 倪国熙, 数理统计学教程, 上海: 上海科学技术出版社, 1984.

[3] Lehmann, E. L., Theory of point estimation. New York: John Wiley & Sons, 1983.