概率论与数理统计

崔文泉 (wqcui@ust.edu.cn)

2021 Autumn

第二章 参数估计

目录

§ 2.1点估计

设总体分布函数的形式已知，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的参数的值，这种问题称为参数估计问题。

例题: 在某炸药制造厂，一天中发生着火现象的次数X是一个随机变量，假设它服从以$\lambda > 0$为参数的泊松分布，参数λ为未知。现有以下的样本值，试估计参数λ.

着火次数k	0	1	2	3	4	5	6	
发生k次着火的天数n_k	75	90	54	22	6	2	1	$\sum = 250$

§ 2.1点估计

设总体分布函数的形式已知，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的参数的值，这种问题称为参数估计问题。

例题: 在某炸药制造厂，一天中发生着火现象的次数X是一个随机变量，假设它服从以$\lambda > 0$为参数的泊松分布，参数λ为未知。现有以下的样本值，试估计参数λ.

着火次数k	0	1	2	3	4	5	6	
发生k次着火的天数n_k	75	90	54	22	6	2	1	$\sum = 250$

解：由于$X \sim \pi(\lambda)$故有$\lambda = E(X)$。我们自然想到用样本的均值老估计总体均值$E(X)$。现由已知数据计算得到

$$\bar{x} = \frac{\sum_{k=0}^{6} k n_k}{\sum_{k=0}^{6} n_k} = 1.22$$

即$E(X) = \lambda$的估计为1.22。

§2.1点估计

点估计问题的一般提法如下：设总体X的分布函数$F(x;\theta)$的形式为已知，θ是待估参数。$X_1,X_2,\cdots,X_n$是X的一个样本值。点估计问题就是要构造一个合适的统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$用它的观察值$\hat{\theta}(x_1,x_2,\cdots,x_n)$作为未知参数$\theta$的近似值。我们称$\hat{\theta}(X_1,X_2,\cdots,X_n)$为$\theta$的**估计量**，称$\hat{\theta}(x_1,x_2,\cdots,x_n)$为$\theta$的**估计值**。在不致混淆的情况下统称估计量和估计值为**估计**，并都简记为$\hat{\theta}$。由于估计量是样本的函数。因此对于不同的样本值，θ的估计值一般是不相同的。例如在例题中，我们用样本均值来估计总体均值。即有估计量与估计值：

$$\hat{\lambda}=\widehat{E(X)}=\frac{1}{n}\sum_{k=1}^{n}X_k, n=250.$$

$$\hat{\lambda}=\widehat{E(X)}=\frac{1}{n}\sum_{k=1}^{n}x_k=1.22.$$

下面介绍构造估计量的方法：**矩估计方法与极大似然估计**

§2.1.1矩估计方法

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：

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

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

因此在k阶矩存在的情况下，有

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

从而我们可以使用a_k, m_k分别估计α_k, μ_k。

§ 2.1.1矩估计方法

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

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

§ 2.1.1矩估计方法

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

$$\left\{\begin{array}{c}\hat{\theta}_1=g_1(a_1,\cdots,a_k)\\ \vdots \\ \hat{\theta}_k=g_k(a_1,\cdots,a_k)\end{array}\right.$$

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。

§ 2.1.1矩估计方法

例题2.1.1 ：设总体X在$[a, b]$上服从均匀分布，a, b未知。$X_1, X_2, \cdots, X_n$来自X的样本，试求a, b的估计量。

§2.1.1矩估计方法

例题2.1.1：设总体X在$[a,b]$上服从均匀分布，a,b未知。$X_1,X_2,\cdots,X_n$来自X的样本，试求a,b的估计量。

解: $\mu_1 = E(X) = (a+b)/2$
$\mu_2 = E(X^2) = D(X) + [E(X)]^2$

即

$$\begin{cases} a+b = 2\mu_1 \\ b-a = \sqrt{12(\mu_2-\mu_1^2)} \end{cases}$$

$a = \mu_1 - \sqrt{3(\mu_2-\mu_1^2)}, \quad b = \mu_1 + \sqrt{3(\mu_2-\mu_1^2)}$

分别以A_1,A_2代替μ_1,μ_2，得到a,b的矩估计量分别为(注意到$\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2$):

$$\hat{a} = A_1 - \sqrt{3(A_2-A_1^2)} = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^n (X_i-\bar{X})^2}$$

$$\hat{b} = A_1 + \sqrt{3(A_2-A_1^2)} = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^n (X_i-\bar{X})^2}$$

§ 2.1.1矩估计方法

例题2.1.2：设总体X的均值μ及方差σ^2都存在，且有$\sigma^2 > 0$.但μ, σ^2均为未知。又设$X_1, X_2, \cdots, X_n$是来自X的样本。试求μ, σ^2的矩估计量。

§ 2.1.1矩估计方法

例题2.1.2：设总体X的均值μ及方差σ^2都存在，且有$\sigma^2 > 0$.但μ, σ^2均为未知。又设$X_1, X_2, \cdots, X_n$是来自X的样本。试求μ, σ^2的矩估计量。

解: $\begin{cases} \mu_1 = E(X) = \mu \\ \mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2 \end{cases}$

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

分别以A_1, A_2代替μ_1, μ_2，得到μ和σ^2的矩估计量为

$$\hat{\mu} = A_1 = \bar{X}$$

$$\hat{\sigma^2} = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

所得结果表明，总体均值与方差的矩估计量的表达式不因不同的总体分布而异。

§2.1.1矩估计方法

例题2.1.3：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。

§ 2.1.1矩估计方法

例题2.1.3：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

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

§ 2.1.1矩估计方法

例题2.1.3：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

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

例题2.1.4：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。

§ 2.1.1矩估计方法

例题2.1.3：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

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

例题2.1.4：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$\hat{\sigma}^2=S^2$.

§ 2.1.2极大似然估计方法(MLE)

定义2.1.1：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。

定义2.1.2：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$g(\hat{\theta})$。

§ 2.1.2极大似然估计方法(MLE)

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。

§2.1.2极大似然估计方法(MLE)

例题2.1.5：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。

§ 2.1.2极大似然估计方法(MLE)

例题2.1.5：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

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

§ 2.1.2极大似然估计方法(MLE)

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:

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

§ 2.1.2极大似然估计方法(MLE)

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:

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

例题2.1.6: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:

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

求参数a, b的极大似然估计量.

§ 2.1.2极大似然估计方法(MLE)

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:

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

§2.1.2极大似然估计方法(MLE)

例题2.1.7:设$X \sim b(1,p)$.$X_1, X_2, \cdots, X_n$是来自X的一个样本，试求参数p的最大似然估计量。

§ 2.1.2极大似然估计方法(MLE)

例题2.1.7:设$X \sim b(1,p).X_1, X_2, \cdots, X_n$是来自$X$的一个样本，试求参数$p$的最大似然估计量。

解：设$x_1, x_2, \cdots, x_n$是相应于样本$X_1, X_2, \cdots, X_n$的一个样本值。X的分布律为

$$P(X=x) = p^x(1-p)^{1-x} \qquad x = 0, 1$$

故似然函数为

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

而

$$\ln L(p) = (\sum_{i=1}^{n} x_i) \ln p + (n - \sum_{i=1}^{n} x_i) \ln(1-p)$$

§ 2.1.2极大似然估计方法(MLE)

令

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0$$

解得p的最大似然估计值

$$\hat{p} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$$

p的最大似然估计量为

$$\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$$

我们看到这一估计量与矩估计量是相同的

§ 2.1.2极大似然估计方法(MLE)

例题2.1.8:设总体X在$[a,b]$上服从均匀分布，a,b未知，$x_1,x_2,\cdots,x_n$是一个样本值，试求a,b的最大似然估计值。

§ 2.1.2极大似然估计方法(MLE)

例题2.1.8:设总体X在$[a,b]$上服从均匀分布，a,b未知，$x_1,x_2,\cdots,x_n$是一个样本值，试求a,b的最大似然估计值。

解:

记$x_{(1)}=min(x_1,x_2,\cdots,x_n), x_{(n)}=max(x_1,x_2,\cdots,x_n)$.
X的概率密度是

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & others\end{cases}$$

似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n} & a\leqslant x_1,x_2,\cdots,x_n\leqslant b\\ 0 & others\end{cases}$$

§2.1.2极大似然估计方法(MLE)

由于$a \leqslant x_1, x_2, \cdots, x_n \leqslant b$,等价于$a \leqslant x_{(1)}, b \geqslant x_{(n)}$.似然函数可以写成

$$L(a,b) = \begin{cases} \dfrac{1}{(b-a)^n} & a \leqslant x_{(1)}, b \geqslant x_{(n)} \\ 0 & others \end{cases}$$

于是对于满足条件$a \leqslant x_{(1)}, b \geqslant x_{(n)}$的任意$a, b$有

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

即$L(a,b)$在$a = x_{(1)}, b = x_{(n)}$时取到最大值$(x_{(n)} - x_{(1)})^{-n}$。故a, b的最大似然估计值为

$$\hat{a} = x_{(1)} = min_{1 \leqslant i \leqslant n} x_i \quad \hat{b} = x_{(n)} = max_{1 \leqslant i \leqslant n} x_i$$

§ 2.1.3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。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。

§ 2.1.3估计量的评选标准

例题2.1.9:设总体X的k阶矩$\mu_k = E(X^k)(k \geqslant 1)$存在，又设$X_1, X_2, \cdots, X_n$是$X$的一个样本。证明不论总体服从什么分布，$k$阶样本矩$A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$是$k$阶总体矩$\mu_k$的无偏统计量。

§2.1.3估计量的评选标准

例题2.1.9:设总体X的k阶矩$\mu_k = E(X^k)(k \geqslant 1)$存在，又设$X_1, X_2, \cdots, X_n$是$X$的一个样本。证明不论总体服从什么分布，$k$阶样本矩$A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$是$k$阶总体矩$\mu_k$的无偏统计量。

证：$X_1, X_2, \cdots, X_n$与X同分布，故有

$$E(X_i^k) = E(X^k) = \mu_k \qquad i = 1, 2, \cdots, n.$$

即有

$$E(A_k) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \mu_k$$

§ 2.1.3估计量的评选标准

例题2.1.10:设总体X服从指数分布，其概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & others\end{cases}$$

其中参数$\theta>0$为未知，又设$X_1,X_2,\cdots,X_n$是X的一个样本，试证$\bar{X}$和$nZ=n(\min(X_1,X_2,\cdots,X_n))$都是$\theta$的无偏估计量。

§2.1.3估计量的评选标准

例题2.1.10:设总体X服从指数分布，其概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & others\end{cases}$$

其中参数$\theta>0$为未知，又设$X_1,X_2,\cdots,X_n$是X的一个样本，试证$\bar{X}$和$nZ=n(\min(X_1,X_2,\cdots,X_n))$ 都是θ的无偏估计量。

证：因为$E(\bar{X})=E(X)=\theta$，所以$\bar{X}$试θ的无偏估计量。
而$Z=\min(X_1,X_2,\cdots,X_n)$具有概率密度

$$f_{min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-n\frac{x}{\theta}} & x>0\\ 0 & others\end{cases}$$

§2.1.3估计量的评选标准

故知

$$E(Z)=\frac{\theta}{n}$$

$$E(nZ)=\theta$$

即nZ也是参数θ的无偏估计量。

由此可见一个未知参数可以有不同的无偏估计量。事实上，在本例中$X_1,X_2,\cdots,X_n$中的每一个都可作为θ的无偏估计量。

2. 有效性

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

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

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。

2. 有效性

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

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

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。

例题2.1.11:(续例2.1.10)试证当$n>1$时，θ的无偏估计量$\bar{X}$较θ的无偏估计量nZ有效。

2. 有效性

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。

例题2.1.11:(续例2.1.10)试证当$n>1$时，θ的无偏估计量$\bar{X}$较θ的无偏估计量nZ有效。

证：由于$D(X)=\theta^2$，故有$D(\bar{X})=\theta^2/n$.再者，由于

$$D(Z)=\theta^2/n^2$$

故有$D(nZ)=\theta^2$.
当$n>1$时$D(nZ)>D(\bar{X})$,故$\bar{X}$较nZ有效。

§2.1.3估计量的评选标准

3. 相合性和渐近正态性

定义2.1.3: 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

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

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

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。

§ 2.1.3估计量的评选标准

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，当n很大时，我们也可以由中心极限定理得到其极限分布，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”。

§ 2.1.4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2.1.4: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

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

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:

§ 2.1.4最小方差无偏估计(MVUE)

Cramer-Rao不等式： 在一定的条件下，对当$g(\theta)$的任一无偏估计$\hat{g}(X)$，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。

§ 2.1.4最小方差无偏估计(MVUE)

例题2.1.12: 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。

§ 2.1.4最小方差无偏估计(MVUE)

例题2.1.12: 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。

其它结果参加教材。

§ 2.2.1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.,

$$P_\theta(\underline{\theta} \le \theta \le \bar{\theta}) \geqq 1 - \alpha$$

- 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:

§ 2.2.1置信区间

定义2.2.5: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

$$P_\theta(\underline{\theta}\leq\theta\leq\bar{\theta})\geq^1 1-\alpha \qquad \forall\ \theta\in\Theta$$

称$1-\alpha$为置信水平，而称$[\underline{\theta},\bar{\theta}]$为$\theta$的置信水平为$1-\alpha$的置信区间。

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)，然后基此估计量构造参数θ的置信区间，介绍如下:

[1]这里表示等于或者大于等于，在一些场合(比如离散场合)，取等号时方程没有解，此时就取大于等于。

§2.2.1置信区间

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$a \leq S(T, g(\theta)) \leq b$要能表示成等价的形式$A \leq g(\theta) \leq B$，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

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

由3我们就可以得到所求的置信区间.

§ 2.2.1置信区间

例题2.2.1: 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。

§2.2.1置信区间

例题2.2.1: 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

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

§ 2.2.1置信区间

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

$$\mu : [\bar{X} - \frac{1}{\sqrt{n}} S t_{\alpha/2}(n-1), \bar{X} + \frac{1}{\sqrt{n}} S t_{\alpha/2}(n-1)],$$

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

§ 2.2.1置信区间

例题2.2.2: 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。

§2.2.1置信区间

例题2.2.2: 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1.4.3。此处略.

§2.2.1置信区间

例题2.2.2: 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1.4.3。此处略.

例题2.2.3: $X_1, \cdots, X_n$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。

§2.2.1置信区间

例题2.2.2: 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1.4.3。此处略.

例题2.2.3: $X_1,\cdots,X_n$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

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

§ 2.2.1置信区间

与参数θ无关。取实数$0 < a < b < 1$，使得

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

$\Longrightarrow$

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

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：

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

§ 2.2.1置信区间

例题2.2.4:有一大批糖果，现从中随机地取16袋，称得重量(以g计)如下：

506	508	499	503	504	510	497	512
514	505	493	496	506	502	509	496

设袋装糖果的重量近似服从正态分布，试求总体均值μ的置信水平为0.95的置信区间。

§ 2.2.1置信区间

例题2.2.4:有一大批糖果，现从中随机地取16袋，称得重量(以g计)如下：

506	508	499	503	504	510	497	512
514	505	493	496	506	502	509	496

设袋装糖果的重量近似服从正态分布，试求总体均值μ的置信水平为0.95的置信区间。

解：$1-\alpha=0.95, \alpha/2=0.025, n-1=15, t_{0.025}(15)=2.1315$
由给出的数据算的$\bar{x}=503.75, s=6.2022$
由例2.2.1可知均值μ的一个置信水平为0.95的置信区间为:

$$(503.75 \pm \frac{6.2022}{\sqrt{16}} \times 2.1315)=(500.4, 507.1)$$

就是说估计袋装糖果重量的均值在500.4g和507.1g之间，估计的可信程度为95%.若以区间内任一值作为μ的近似值，其误差不大于$\frac{6.2022}{\sqrt{16}} \times 2.1315 \times 2=6.61(g)$,这误差估计的可信程度为95%

§ 2.2.1置信区间

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。

§ 2.2.1置信区间

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。

例题2.2.5: 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。

§2.2.1置信区间

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。

例题2.2.5: 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。

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

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

可以等价表示成

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

§ 2.2.1置信区间

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。

§ 2.2.2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2.2.6: 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

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

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

2.若

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

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

寻求置信上、下界的方法和寻求置信区间的方法完全类似

§2.2.2置信界

例题2.2.6: 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。

§2.2.2置信界

例题2.2.6: 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。

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

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

§ 2.2.2置信界

例题2.2.7: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。

§2.2.2置信界

例题2.2.7: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$Y=\sum_{i=1}^{n} Y_i$，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

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

§ 2.2.2置信界

因此有

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

从而得到p的渐近$1-\alpha$置信下界

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

§ 2.3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