第六章　参数估计

§ 6.1　点估计

一　　点估计概念

二　　矩估计法

三　　极大似然估计法

实例 制衣厂为了合理的确定服装各种尺码的生产比例，需要调查人们身长的分布。现从男性成人人群中随机选取100人,得到他们的身长数据为:....

若已知X服从正态分布$N(\mu,\sigma2)$,试估计参数的$\mu,\sigma2$值

已知“总体”的分布类型,对分布中的未知参数进行统计推断.

一、点估计问题

设总体 ***X*** 的分布函数形式已知, 但它的一个或多个参数为未知, 借助于总体 ***X*** 的一个样本来估计总体未知参数的值的问题称为点估计问题.

总体参数, 指总体分布$F(x;\theta)$的数学表达式中所含的参数θ

例如, 正态分布$N(\mu,\sigma^2)$的参数为μ,σ^2; 泊松分布$P(\lambda)$的参数为λ等等。

二　参数的点估计

定义1.1 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，θ 为总体分布 $F(x;\theta)$ 中的未知参数，构造一个统计量 $T = T(X_1, X_2, \cdots, X_n)$ 作为 θ 的估计，则称 $T = T(X_1, X_2, \cdots, X_n)$ 为 θ 的估计量；

若样本 $X_1, X_2, \cdots, X_n$ 的一个观察值为 $x_1, x_2, \cdots, x_n$，则称 $T = T(x_1, x_2, \cdots, x_n)$ 为

θ 的估计值，统称为参数 θ 的点估计，记作 $\hat{\theta}$.

即 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$为 θ 的点估计量

$\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$为 θ 的点估计值

注1. 点估计实际上是指用统计量的值去估计未知参数的值,又指用来估计未知参数的统计量,例如,用样本均值估计总体数学期望,用样本方差估计总体方差,用频率估计概率。

注2　若总体分布 $F(x;\theta_1,\theta_2,\cdots,\theta_r)$中含有 r个不同的未知参数，则须由样本 $X_1,X_2,\cdots,X_n$建立r个统计量 $T_i(X_1,X_2,\cdots,X_n),(1\le i\le r)$来作为相应参数 $\theta_i(1\le i\le r)$的点估计。

例如正总体 $N(u,\sigma^2)$有两个未知参数 μ及σ^2

而$E(X)=\mu, D(X)=\sigma^2$,可分别用

样本均值 $T_1=\overline{X}$与样本方差 $T_2=S^2$加以估计．

矩有原点矩与中心矩:

原点矩:

指随机变量 ξ k次幂的数学期望 $E\xi^k$

$E\xi$ 1阶原点矩　　$E\xi^2$ 2阶原点矩

设X是总体，$X_1, X_2, \ldots X_n$ 是X的一个样本.

$E(X^k)$ 称为总体X的*k*阶原点矩;

$\frac{1}{n}\sum_{i=1}^{n} X_i^k$ 称为样本的*k*阶原点矩.

中心矩：$E(\xi-\mu)^k$

是指随机变量ξ的离差 $\xi-\mu$ 的k次幂的数学期望

$E(\xi-\mu)^2$ 2阶中心矩.

设X是总体，$X_1, X_2, \ldots X_n$是X的一个样本.

$E(X-\mu)^k$ 称为总体X的k阶中心矩;

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 称为样本的k阶中心矩;

二 矩估计法

定义1.2 设总体 X 的分布函数 $F(x;\theta_1,\theta_2,\cdots,\theta_r)$ 中有 r 个未知参数 $\theta_1,\theta_2,\cdots,\theta_r$，假定总体 X 的 k 阶原点矩 $E(X^k)(1\le k\le r)$ 存在，一般来说，$E(X^k)$ 依赖于未知参数 $\theta_1,\dots,\theta_r$，记作

$$E(X^k)=V_k(\theta_1,\theta_2,\cdots,\theta_r)\qquad(1\le k\le r)\qquad(1.1)$$

令它等于 k 阶样本原点矩

$$\overline{X^k}=\frac{1}{n}\sum_{i=1}^{n}X_i^k\qquad(1.2)$$

即令

$$E(X^k)=V_k(\theta_1,\theta_2,\cdots,\theta_r)=\frac{1}{n}\sum_{i=1}^{n}X_i^k \qquad (1\le k\le r)$$

由上面的方程组解出r个值

$$\hat{\theta}_i=\hat{\theta}_i(X_1,X_2,\cdots,X_n) \qquad (1\le k\le r)$$

分别取 $\hat{\theta}_i$ 作为θ_i的估计量, 这种求估计量的方法称之为矩估计法, 由此得到的估计量称为矩估计量。若有一样本值$x_1, x_2,\ldots,x_n$, 则称

$$\hat{\theta}_i=\hat{\theta}_i(x_1,x_2,\cdots,x_n)$$

为矩估计值。

矩估计法:

其基本思想是替换原理, 用样本*k*阶矩作为总体*k*阶矩的估计量, 建立含有待估参数的方程或方程组, 从而解出待估参数

其特点是不需要假定总体分布有明确的分布类型。

矩估计法的具体步骤:

(1).求出$\mu_k = E(X^k) = V_k(\theta_1,\theta_2,\cdots,\theta_r) \quad k = 1,2,\cdots$

(2).令 $\mu_k = \overline{X^k}$, 其中$\overline{X^k} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k; k = 1,2,\cdots,r$

这是一个包含 r 个未知参数$\theta_1,\theta_2,\cdots,\theta_r$的方程组.

(3).解出其中 $\theta_1,\theta_2,\cdots,\theta_r$, 用$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_r$表示.

(4).用方程组的解 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_r$分别作为 $\theta_1,\theta_2,\cdots,\theta_r$的估计量,这个估计量称为矩估计 量.

矩估计量的观察值称为矩估计值.

例 设总体 X 在 $[0,\theta]$ 上服从均匀分布，其中 θ $(\theta > 0)$ 未知，$(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，求 θ 的估计量.

解 因为 $\mu_1 = E(X) = \dfrac{\theta}{2}$,

根据矩估计法, 令 $\dfrac{\hat{\theta}}{2} = \overline{X}$,

所以 $\hat{\theta} = 2\overline{X}$ 为所求 θ 的估计量.

例2.2 设总体 X 在 $[a,b]$ 上服从均匀分布，其中 a, b 未知，$(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，求 a, b 的估计量.

解

$$\mu_1 = E(X) = \frac{a+b}{2},$$

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \frac{(a-b)^2}{12} + \frac{(a+b)^2}{4},$$

令 $$\frac{a+b}{2} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\frac{(a-b)^2}{12} + \frac{(a+b)^2}{4} = \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

即 $$\begin{cases} a+b=2\overline{X}, \\ b-a=\sqrt{12(\overline{X^2}-\overline{X}^2)}. \end{cases}$$

解方程组得到*a*, *b*的矩估计量分别为

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

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

例2.3 设总体 X 的均值 μ 和方差 σ^2 都存在，且有 $\sigma^2 > 0$，但 μ 和 σ^2 均为未知，又设 $X_1, X_2, \cdots, X_n$ 是一个样本，求 μ 和 σ^2 的矩估计量.

解

$$\mu_1 = E(X) = \mu,$$

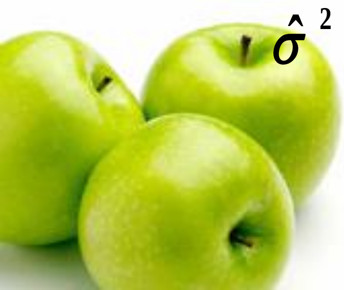
$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2,$$

$$\text{令}\begin{cases} \mu = \overline{X}, \\ \sigma^2 + \mu^2 = \overline{X^2}. \end{cases}$$

解方程组得到矩估计量分别为 $\hat{\mu} = \overline{X}$,

$$\hat{\sigma}^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 =: S_n^2$$

(二阶中心距)



上例表明：**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.**

例 $X \sim N(\mu, \sigma^2)$, μ, σ^2 未知，即得 μ, σ^2 的矩估计量

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

一般地,

用样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 作为总体 X 的均值的矩估计，

用样本二阶中心矩 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 作为总体

X 的方差的矩估计.

注. 1 定义中选用的是原点矩，也可以用中心矩，只要给定总体矩，采用相应的样本矩就可以。

2 若 $\hat{\theta}$为 θ的矩估计量， $g(\theta)$为 θ的连续函数，亦称 $g(\hat{\theta})$为$g(\theta)$的矩估计量。

例如， S_n^2为总体方差 σ^2的矩估计量，则 $S_n = \sqrt{S_n^2}$为标准差 σ的矩估计量。

3 矩估计的关键是计算总体矩，因此使用矩估计法其前提是总体矩必须存在。

例2.4 设$X_1,X_2,\ldots,X_n$是来自总体X的样本，当X的分布为

(1)正态分布$N(\mu,\sigma^2)$

(2)指数分布$E(\lambda)$

(3)均匀分布$U(a,b)$

(4)二项分布$B(n,p)$

(5)泊松分布$P(\lambda)$

试求其中未知参数的矩估计。

解:(1) 因为 $X \sim N(\mu,\sigma^2)$，$E(X)=\mu$，$D(X)=\sigma^2$

故有 $\hat{\mu}=\overline{X}$，$\hat{\sigma}^2=S_n^2$

(2) $X \sim E(\lambda)$，$E(X)=\dfrac{1}{\lambda}$

故 $\dfrac{1}{\lambda}=\overline{X}$，即 $\hat{\lambda}=\dfrac{1}{\overline{X}}$

(3) $X \sim U(a,b)$，$E(X)=\dfrac{a+b}{2}$，$D(X)=\dfrac{(b-a)^2}{12}$

故有 $\begin{cases} \dfrac{a+b}{2}=\overline{X} \\ \dfrac{(b-a)^2}{12}=S_n^2 \end{cases}$ 即 $\begin{cases} a+b=2\overline{X} \\ b-a=2\sqrt{3}S_n \end{cases}$

解出得：$\hat{a}=\overline{X}-\sqrt{3}S_n \quad \hat{b}=\overline{X}+\sqrt{3}S_n$

(4) $X \sim B(n,p)$, $E(X)=np$, $D(X)=np(1-p)$

故有 $\begin{cases} np = \overline{X} \\ np(1-p) = S_n^2 \end{cases}$ 即 $\begin{cases} \hat{p} = 1 - \dfrac{S_n^2}{\overline{X}} \\ \hat{n} = \dfrac{\overline{X}}{\hat{p}} = \dfrac{\overline{X}^2}{\overline{X} - S_n^2} \end{cases}$

(5) $X \sim P(\lambda)$, $E(X)=D(X)=\lambda$

故 $\hat{\lambda} = \overline{X}$ 或 $\hat{\lambda} = S_n^2$

注: 由此例可知, 矩估计量不唯一。

例2.5 设总体X的概率密度为

$$f(x;\theta)=\begin{cases}(\theta+1)x^{\theta} & 0<x<1 \\ 0 & 其它\end{cases} \qquad (\theta>-1)$$

$X_1,X_2,\ldots,X_n$是来自总体X的样本。0.1,0.2,0.9,0.8,0.7,0.7为一个样本观察值，试求θ的矩估计值。

解： $$E(X)=\int_{-\infty}^{+\infty}xf(x;\theta)dx=\int_0^1 x(\theta+1)x^{\theta}dx=(\theta+1)\int_0^1 x^{\theta+1}dx$$

$$=(\theta+1)\frac{1}{\theta+2}x^{\theta+2}\Big|_0^1=\frac{\theta+1}{\theta+2}$$

令 $\dfrac{\theta+1}{\theta+2}=\overline{X}$ 解之得 θ的矩估计 $\hat{\theta}=\dfrac{2\overline{X}-1}{1-\overline{X}}$

由样本值 0.1,0.2,0.9,0.8,0.7,0.7计算得 $\overline{x}=0.5667$

故 θ的矩估计值为 $\hat{\theta}=\dfrac{2\overline{x}-1}{1-\overline{x}}=0.3079$

例2.6 设总体X的概率密度为

$$f(x,\theta)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}} \qquad -\infty<x<+\infty,\ \theta>0$$

试求 θ的矩估计量 $\hat{\theta}$。

法1 解：虽然 $f(x;\theta)$中仅含一个未知参数 θ，但因

$$E(X)=\int_{-\infty}^{+\infty}x\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=0$$

不含 θ，不能由此解出，故需 继续求出总体二阶原点 矩：

$$E(X^2)=\int_{-\infty}^{+\infty}x^2\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\int_{0}^{+\infty}x^2\frac{1}{\theta}e^{-\frac{x}{\theta}}dx$$

$$=\theta^2+\theta^2=2\theta^2$$

故令 $2\theta^2=\overline{X^2}=\frac{1}{n}\sum_{i=1}^{n}X_i^2$，得 θ的矩估计量 $\hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}$

法2 考虑$|X|$的数学期望，

$$E(|X|)=\int_{-\infty}^{+\infty}|x|\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\int_{0}^{+\infty}x\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\theta$$

可得出 θ的另一矩估计量 $\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$

对同一参数，其矩估计量不唯一。

2. 最大似然估计法

(1) 设总体 X 属离散型

似然函数的定义

设分布律 $P\{X=k\}=p(x;\theta)$, θ 为待估参数，$\theta\in\Theta$，

(其中 Θ 是 θ 可能的取值范围)

$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，

则 $X_1,X_2,\cdots,X_n$ 的联合分布律为 $\prod_{i=1}^{n}p(x_i;\theta)$.

又设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值 .

则样本 $X_1, X_2, \cdots, X_n$ 取到观察值 $x_1, x_2, \cdots, x_n$ 的概率 ,

即事件 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ 发生的概率为

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

$L(\theta)$称为样本似然函数.

最大似然估计法

得到样本值 $x_1, x_2, \cdots, x_n$ 时，选取使似然函数 $L(\theta)$

取得最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值，

即 $L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max\limits_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta).$

参数 θ 的最大似然估计值.

参数 θ 的最大似然估计量.

(2) 设总体 X 属连续型

似然函数的定义

设概率密度为 $f(x;\theta), \theta$ 为待估参数，$\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，

则 $X_1, X_2, \cdots, X_n$ 的联合密度为 $\prod_{i=1}^{n} f(x_i;\theta)$.

又设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值.

则随机点 $(X_1, X_2, \cdots, X_n)$ 落在点 $(x_1, x_2, \cdots, x_n)$ 的邻域（边长分别为 $\mathrm{d}x_1, \mathrm{d}x_2, \cdots, \mathrm{d}x_n$ 的 n 维立方体）内的概率近似地为 $\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_i$，

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

$L(\theta)$ 称为样本的似然函数.

若 $L(x_1, x_2, \cdots, x_n;\hat{\theta}) = \max\limits_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n;\theta).$

$\boxed{\hat{\theta}(x_1, x_2, \cdots, x_n)}$ 参数 θ 的最大似然估计值，

$\boxed{\hat{\theta}(X_1, X_2, \cdots, X_n)}$ 参数 θ 的最大似然估计量.

1 极大似然估计概念

定义1.3 设总体 X 的分布函数 $F(x;\theta)$ 的形式已知，θ 为未知参数，Θ 为 θ 的可能取值范围，$x_1, x_2, \cdots, x_n$ 为 X 的一个样本值，

若存在 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$，使得似然函数

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$$

则称 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ 为 θ 的极大似然估计值，

$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$为 θ 的极大似然估计量，

统称为 θ 的极大似然估计。.

注1 :似然函数实际上是样本的密度函数或概率分布:

$$L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \quad (X\text{为连续型}) \qquad (1.4)$$

$$L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} P\{X_i = x_i; \theta\} \quad (X\text{为离散型}) \qquad (1.5)$$

注2　若总体分布中含有两个以上的未知参数 $\theta_1,\theta_2,\cdots,\theta_r$ 时，则 θ_i 的极大似然估计 $\hat{\theta}_i$ 满足

$$\boxed{L(x_1,\cdots,x_n;\hat{\theta}_1,\cdots,\hat{\theta}_r)=\max L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_r)}$$

2 求最大似然估计量的步骤:

(一) 写出似然函数

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

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

(二) 取对数

$$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\theta) \quad 或 \quad \ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta);$$

(三) 对 θ 求导 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$, 并令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta} = 0}$, 对数似然方程

解方程即得未知参数 θ 的最大似然估计值 $\hat{\theta}$.

最大似然估计法也适用于分布中含有多个未知参数的情况. 此时只需令

$$\boxed{\frac{\partial}{\partial \theta_i}\ln L = 0, \qquad i = 1,2,\cdots,k.}$$ 对数似然方程组

解出由 k 个方程组成的方程组, 即可得各未知参数 $\theta_i\ (i = 1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例2.6 设 $X \sim B(1, p)$, $X_1, X_2, \cdots, X_n$是来自 X 的一个样本，求 p的最大似然估计量.

解 设 $x_1, x_2, \cdots, x_n$为相应于样本 $X_1, X_2, \cdots, X_n$的一个样本值，

X的分布律为 $P\{X = x\} = p^x(1-p)^{1-x}$, $x = 0, 1$,

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

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

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

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

解得 p 的最大似然估计值 $$p=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

p 的最大似然估计量为 $$\hat{p}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}.$$

这一估计量与矩估计量是相同的.

例2.7 设 X 服从参数为 $\lambda\ (\lambda > 0)$ 的泊松分布，$X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本，求 λ 的最大似然估计量.

解 因为 X 的分布律为

$$P\{X = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, \quad (x = 0, 1, 2, \cdots, n)$$

所以 λ 的似然函数为

$$L(\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i!)},$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \sum_{i=1}^{n}(x_i!),$$

令 $$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda) = -n + \frac{\sum\limits_{i=1}^{n} x_i}{\lambda} = 0,$$

解得 λ 的最大似然估计值 $\lambda = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \overline{x},$

λ 的最大似然估计量为 $\hat{\lambda} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i = \overline{X}.$

这一估计量与矩估计量是相同的.

例2.8 设总体 $X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数，$x_1,x_2,\cdots,x_n$是来自 X 的一个样本值，求μ和σ^2的最大似然估计量.

解 X的概率密度为

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

X 的似然函数为

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

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

令 $$\begin{cases} \dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) = 0, \\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) = 0, \end{cases}$$

$$\begin{cases} \dfrac{1}{\sigma^2}\left[\displaystyle\sum_{i=1}^{n} x_i - n\mu\right] = 0, \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0, \end{cases}$$

由 $\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\mu\right] = 0$ 解得 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$

由 $-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$ 解得

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

故 μ 和 σ^2 的最大似然估计量分别为

$\hat{\mu} = \bar{X},\quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$ 它们与相应的矩估计量相同.

例2.9 设总体X的概率密度为

$$f(x)=\begin{cases}(\theta+1)x^{\theta} & 0<x<1\\ 0 & \text{其它}\end{cases} \qquad \theta>-1$$

0.1,0.2,0.9,0.8,0.7,0.7为一个样本观察值,试求θ的极大似然估计。

解 : $i)\ L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} (\theta + 1) x_i^{\theta} = (\theta + 1)^n (\prod_{i=1}^{n} x_i)^{\theta}$

$0 \le x_i \le 1$

$ii)\ \ln L(x_1, x_2, \cdots, x_n; \theta) = n \ln(\theta + 1) + \theta \sum_{i=1}^{n} \ln x_i$

$iii)$令 : $\frac{d}{d\theta}\left[\ln L(x_1, x_2, \cdots, x_n; \theta)\right] = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln x_i = 0$

$iv)$解之得: $\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln x_i} - 1$ 为 θ的极大似然估计值。

而 $\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln X_i} - 1$为 θ的极大似然估计量。

又由样本值计算得 $\sum_{i=1}^{n} \ln x_i = -4.9539$ 故得 $\hat{\theta} = \frac{-6}{-4.9539} - 1 = 0.2112$

例6.1.6　设一个试验有三种可能结果，其发生概率 $p_1 = \theta^2$， $p_2 = 2\theta(1-\theta)$， $p_3 = (1-\theta)^2$

现做了n次试验，观测到三种结果发生的次数分别为 n_1, n_2, n_3 $(n_1 + n_2 + n_3 = n)$，则似然函数为

$$L(\theta) = (\theta^2)^{n_1}[2\theta(1-\theta)]^{n_2}[(1-\theta)^2]^{n_3}$$

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

其对数似然函数为

$$\text{n}\, L(\theta) = (2n_1 + n_2)\ln\theta + (2n_3 + n_2)\ln(1-\theta) + n_2 \ln 2$$

将之关于θ求导，并令其为0得到似然方程

$$\frac{2n_1+n_2}{\theta}-\frac{2n_3+n_2}{1-\theta}=0$$

解之，得

$$\hat{\theta}=\frac{2n_1+n_2}{2(n_1+n_2+n_3)}=\frac{2n_1+n_2}{2n}$$

由于

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}=-\frac{2n_1+n_2}{\theta^2}-\frac{2n_3+n_2}{(1-\theta)^2}<0$$

所以$\hat{\theta}$ 是极大值点。

(2) 利用极大似然估计定义求估计法

i) 建立似然函数

$$L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_r) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \cdots, \theta_r)$$

ii) 由 $x_1, x_2, \cdots, x_n$ 确定顺序统计值

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

iii) 利用顺序统计值的函数 作出 $\theta_1,\theta_2,\cdots,\theta_r$的

相应估计 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$,使其满足条件

$$L(x_1,\cdots,x_n;\hat{\theta}_1,\cdots,\hat{\theta}_n) = \max_{(\theta_1,,\ldots,\theta_r)\in\Theta} L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_r)$$

则 $\hat{\theta}_i$ 即为 $\theta_i (1 \le i \le r)$的极大似然估计。

例2.10 设总体 X 在 $[a,b]$ 上服从均匀分布，其中 a，b 未知，$x_1, x_2, \cdots, x_n$ 是来自总体 X 的一个样本值，求 a, b 的最大似然估计量.

解 记 $x_{(1)} = \min(x_1, x_2, \cdots, x_n)$,

$$x_{(n)} = \max(x_1, x_2, \cdots, x_n),$$

X 的概率密度为

$$f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{其他}. \end{cases}$$

因为 $a \le x_1, x_2, \cdots, x_n \le b$ 等价于 $a \le x_{(1)}, x_{(n)} \le b$,

作为 a, b 的函数的似然函数为

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

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

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

即似然函数 $L(a,b)$ 在 $a=x_{(1)}$, $b=x_{(n)}$ 时

取到最大值 $(x_{(n)}-x_{(1)})^{-n}$,

a,b 的最大似然估计值

$$\hat{a}=x_{(1)}=\min_{1\le i\le n} x_i, \qquad \hat{b}=x_{(n)}=\max_{1\le i\le n} x_i,$$

a,b 的最大似然估计量

$$\hat{a}=\min_{1\le i\le n} X_i, \qquad \hat{b}=\max_{1\le i\le n} X_i.$$

例2.11 设$X_1,X_2,\ldots X_n$是取自总体X的一个样本

$$X \sim f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu\\ 0, & \text{其它}\end{cases}\qquad \theta,\mu\text{为未知参数}$$

其中$\theta>0$,求θ,μ的极大似然估计.

解： 似然函数为

$$L(x_1,x_2,\cdots x_n\theta,\mu)=\begin{cases}\prod\limits_{i=1}^{n}\dfrac{1}{\theta}e^{-(x_i-\mu)/\theta}, & x_i\ge\mu\\ 0, & \text{其它}\end{cases}$$

$i=1,2,\ldots,n$

$$\begin{cases} \frac{1}{\theta} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)} & x_i \geq \mu \end{cases}$$

对数… $i = 1,2,\ldots,n$

$$\ldots \frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)$$

用求导方法无法最终确定θ、μ,
用极大似然原则来求.

对 θ, μ 分别求偏导并令其为0,

$$\frac{\partial \ln L(\theta,\mu)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \qquad (1)$$

$$\frac{\partial \ln L(\theta,\mu)}{\partial \mu} = \frac{n}{\theta} > 0 \qquad (2)$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu$$

由于 $$L(\theta,\mu)=\begin{cases}\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}(x_i-\mu)}, & x_i\ge\mu\\ 0, & 其它\end{cases}$$

对 $\mu\le\min x_i, L(\theta,\mu)>0$，且是 μ 的增函数

μ 取其它值时，$L(\theta,\mu)=0$.

故使 $L(\theta,\mu)$ 达到最大的 μ，即 μ 的MLE，

是 $$\hat{\mu}=\min_{1\le i\le n} x_i$$

于是 $$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i-\hat{\mu}$$

例1.11　设总体X的密度函数为

$$f(x;\theta)=\begin{cases}1 & \theta-\dfrac{1}{2}\le x\le\theta+\dfrac{1}{2}\\ 0 & \text{其它}\end{cases}\qquad -\infty<x<+\infty$$

试求未知参数θ 的极大似然估计.

解：

$$i)L(x_1,x_2,\ldots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)=1\qquad \theta-\frac{1}{2}\le x\le\theta+\frac{1}{2}$$

$ii)\, x_1,x_2,\ldots,x$的顺序统计值为：$x_{(1)}\le x_{(2)}\le\ldots\le x_{(n)}$

$iii)\ L(x_1, x_2, \ldots, x_n;\theta)$与$\theta$无关而要：$\theta-\dfrac{1}{2}\le x_1, x_2, \ldots, x_n \le \theta+\dfrac{1}{2}$

故 $$\theta-\frac{1}{2}\le x_{(1)}\le x_{(2)}\le \ldots \le x_{(n)}\le \theta+\frac{1}{2}$$

故由极大似然估计定义,应有

$$\hat{\theta}-\frac{1}{2}=x_{(1)}=\min_{1\le i\le n}\{x_i\},\qquad \hat{\theta}+\frac{1}{2}=x_{(n)}=\max_{1\le i\le n}\{x_i\}$$

极大似然估计的不变性

设 $\hat{\theta}$ 是 θ 的极大似然估计值, $u(\theta)$ $(\theta \in \Theta)$ 是 θ 的函数, 且有单值反函数

$$\theta = \theta(u), \qquad u \in U$$

则 $\hat{u} = u(\hat{\theta})$ 是 $u(\theta)$ 的极大似然估计值.

如 在正态总体$N(\mu,\sigma^2)$中，σ^2的极大似然估计值为

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

$\sigma=\sqrt{\sigma^2}$ 是σ^2的单值函数，且具有单值反函数，故σ的极大似然估计值为

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

$\lg\sigma$的极大似然估计值为

$$\widehat{\lg\sigma}=\lg\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

三、小结

两种求点估计的方法: $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

在统计问题中往往先使用最大似然估计法，在最大似然估计法使用不方便时，再用矩估计法.

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

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

§6.2 点估计的评价标准

6.2.1 相合性

我们知道，点估计是一个统计量，因此它是一个随机变量，在样本量一定的条件下，我们不可能要求它完全等同于参数的真实取值。但如果我们有足够的观测值，根据格里纹科定理，随着样本量的不断增大，经验分布函数逼近真实分布函数，因此完全可以要求估计量随着样本量的不断增大而逼近参数真值，这就是相合性，严格定义如下。

定义6.2.1 设$\theta\in\Theta$为未知参数，$\hat{\theta}_n=\hat{\theta}_n(X_1,\cdots,X_n)$是$\theta$的一个估计量，$n$是样本容量，若对任何一个$\varepsilon>0$，有

$$\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|>\varepsilon)=0 \qquad (6.2.1)$$

则称$\hat{\theta}_n$为θ参数的相合估计。

相合性被认为是对估计的一个最基本要求，如果一个估计量，在样本量不断增大时，它都不能把被估参数估计到任意指定的精度，那么这个估计是很值得怀疑的。 通常，不满足相合性要求的估计一般不予考虑。证明估计的相合性一般可应用大数定律或直接由定义来证.

若把依赖于样本量*n*的估计量 $\hat{\theta}_n$看作一个随机变量序列, 相合性就是$\hat{\theta}_n$ 依概率收敛于θ, 所以证明估计的相合性可应用依概率收敛的性质及各种大数定律。

在判断估计的相合性时下述两个定理是很有用的。

定理6.2.1 设$\hat{\theta}_n=\hat{\theta}_n(x_1,\cdots,x_n)$是$\theta$的一个估计量，若

$$\lim_{n\to\infty}E(\hat{\theta}_n)=\theta,\quad \lim_{n\to\infty}Var(\hat{\theta}_n)=0$$

则$\hat{\theta}_n$是θ的相合估计，

定理6.2.2 若$\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk}$分别是$\theta_1,\ldots,\theta_k$的相合估计，$\eta=g(\theta_1,\ldots,\theta_k)$是$\theta_1,\ldots,\theta_k$的连续函数，则$\hat{\eta}_n=g(\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk})$是$\eta$的相合估计。

例6.2.2 设 $X_1, X_2, \ldots, X_n$ 是来自均匀总体$U(0, \theta)$的样本，证明θ的极大似然估计是相合估计。

证明：在例6.1.7中我们已经给出θ的极大似然估计是 $X_{(n)}$。由次序统计量的分布，我们知道 $X_{(n)}$ 的分布密度函数为 $p(y)=ny^{n-1}/n$，$y<\theta$, 故有

$$E\hat{\theta} = \int_0^{\theta} ny^n \mathrm{d}y / \theta^n = \frac{n}{n+1}\theta \to \theta$$

$$E\hat{\theta}^2 = \int_0^{\theta} ny^{n+1} \mathrm{d}y / \theta^n = \frac{n}{n+2}\theta^2$$

$$\operatorname{Var}(\hat{\theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 \to 0,$$

由定理6.2.1可知，$x_{(n)}$是θ的相合估计。

由大数定律及定理6.2.2，我们可以看到：矩估计一般都具有相合性。比如：

- 样本均值是总体均值的相合估计；
- 样本标准差是总体标准差的相合估计；
- 样本变异系数是总体变异系数的相合估计。

例8 试证：样本均值 $\overline{X}$ 是总体均值 μ 的相合估计量，样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 及样本的二阶中心矩 $B_2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 都是总体方差 σ^2 的相合估计量．

证明 由大数定律知，

$$\forall \varepsilon>0,\ \text{有}\ \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1,$$

所以 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 是 μ 的相合估计量．

又 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)$

$$=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=A_2-\overline{X}^2,$$

(A_2是样本二阶原点矩)

由大数定律知,

$$A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2\text{依概率收敛于 }E(X^2),$$

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\text{依概率收敛于 }E(X),$$

故 $B_2 = A_2 - \overline{X}^2$

依概率收敛于 $E(X^2) - [E(X)]^2 = \sigma^2$,

所以 B_2 是 σ^2 的相合估计量 .

又 $\lim\limits_{n\to\infty} \dfrac{n}{n-1} = 1$,

所以 $S^2 = \dfrac{n}{n-1} B_2$ 也是 σ^2 的相合估计量 .

二 无偏性

1 无偏性的定义

定义2.1 设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为未知参数 θ 的一个估计量，若

$$E(\hat{\theta})=\theta \qquad (2.1)$$

则称 $\hat{\theta}$为 θ的无偏估计量，否则称　为有偏估计量。

例3.1 设总体 X 的 k 阶矩 $\mu_k = E(X^k)\ (k \geq 1)$ 存在，又设 $X_1, X_2, \cdots, X_n$ 是 X 的一个样本，试证明不论总体服从什么分布，k 阶样本矩 $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 k 阶总体矩 μ_k 的无偏估计.

证 因为 $X_1, X_2, \cdots, X_n$ 与 X 同分布，

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

即 $E(\overline{X^k}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \mu_k.$

故 k 阶样本矩 $\overline{X^k}$ 是 k 阶总体矩 μ_k 的无偏估计 .

特别的:

不论总体 X 服从什么分布，只要它的数学期望存在,

$\overline{X}$ 总是总体 X 的数学期望 $\mu_1 = E(X)$ 的无偏估计量 .

例: 从总体X~N(μ,σ²)中取一样本($X_1,X_2,\dots,X_n$) EX=μ,DX=σ²,试证明样本均值 $\overline{X}$ 与样本方差 S^2 分别是 μ 及 σ^2 的无偏估计量.

证明: 设 $\hat{\mu}=\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 只需证明:

因为($X_1,X_2,\dots,X_n$)是总体X~N(μ,σ²)的样本.

$X_i \sim N\left(\mu,\sigma^2\right)$ 则有: $\overline{X} \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$

所以: $E(\hat{\mu})=E(\overline{X})=\mu$

故:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$ 是总体均值的无偏估计量.

下面证明样本方差S^2是总体方差的无偏估计量.

设 $$\hat{\sigma}^2=S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

只需证明: $E\hat{\sigma}^2=ES^2=\sigma^2$

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

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

$$= \frac{1}{n-1}(\sum_{i=1}^{n} EX_i^2 - nE\overline{X}^2)$$

$$DX = EX^2 - \left(EX\right)^2$$

$$= \frac{1}{n-1}(\sum_{i=1}^{n}\left[\underline{DX_i + \left(EX_i\right)^2}\right] - n\left[\underline{D\overline{X} + \left(E\overline{X}\right)^2}\right])$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\left[DX_i + (EX_i)^2\right] - n\left[D\overline{X} + (E\overline{X})^2\right]\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\left[\sigma^2 + \mu^2\right] - n\left[\frac{\sigma^2}{n} + \mu^2\right]\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - \left[\sigma^2 + n\mu^2\right]\right) = \sigma^2$$

故：样本方差S^2是总体方差的无偏估计量.

说明1:

$$S_n^{\ 2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$ 不是无偏估计量.

因为: $$ES_n^{\ 2} = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)$$

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

$$= \frac{n-1}{n}\sigma^2$$

例3.3 设总体 X 在 $[0,\theta]$上服从均匀分布，参数 $\theta>0$, $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证明 $2\overline{X}$ 和 $\dfrac{n}{n+1}\max(X_1,X_2,\cdots,X_n)$ 都是 θ 的无偏估计.

证 因为 $E(2\overline{X})=2E(\overline{X})=2E(X)=2\times\dfrac{\theta}{2}=\theta,$

所以 $2\overline{X}$ 是 θ 的无偏估计量.

因为 $X_{(n)}=\max(X_1,X_2,\cdots,X_n)$的概率密度为

$$f(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}, & 0\le x\le\theta,\\ 0, & \text{其他}\end{cases}$$

所以 $E(X_{(n)}) = \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n} \mathrm{d}x$

$$= \frac{n}{n+1}\theta,$$

故有 $E\left(\frac{n+1}{n} X_{(n)}\right) = \theta,$

故 $\frac{n}{n+1}\max(X_1, X_2, \cdots, X_n)$ 也是 θ 的无偏估计量 .

2 渐近无偏估计

定义3.2 若 $\hat{\theta}_n$ 为 θ 的有偏估计量，但有

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta \qquad (2.2)$$

则称 $\hat{\theta}_n$ 为 θ 的渐近无偏估计量。

例3.4 样本二阶中心矩 S_n^2 是 σ^2 的渐近无偏估计量。

注：如果$\hat{\theta}$是参数θ的一个估计，我们常用$g(\hat{\theta})$作为$g(\theta)$的估计.但当$\hat{\theta}$是θ的无偏估计时，$g(\hat{\theta})$未必是$g(\theta)$的无偏估计.

例如,样本标准差S不是总体标准差σ的无偏估计.

$\because E(S^2)=\sigma^2$　　$\therefore D(S)+E^2(S)=\sigma^2$

$\because D(S)\geq 0$　　$\therefore E^2(S)=\sigma^2-D(S)\leqslant\sigma^2$

$\therefore E(S)\leq\sigma$

即：一般说S不是σ的无偏估计

无偏估计量的函数未必是无偏估计量

设总体X的方差$D(X)$存在，且$D(X)>0$，$(X_1, X_2, \cdots, X_n)$为来自总体X的样本，试选择适当的常数C，使得

$$C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$$

为$D(X)$的无偏估计.

解 $\because$ $$E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=C\sum_{i=1}^{n-1}E(X_{i+1}-X_i)^2$$

$$=C\sum_{i=1}^{n-1}\{D(X_{i+1}-X_i)+[E(X_{i+1}-X_i)]^2\}$$

而$X_1, X_2, \cdots, X_n$相互独立，且与X同分布

$$\therefore \quad E(X_i) = E(X), \quad D(X_i) = D(X) \quad (i = 1, 2, \cdots, n)$$

$$D(X_{i+1} - X_i) = D(X_{i+1}) + D(X_i) = 2D(X)$$

$$E(X_{i+1} - X_i) = E(X_{i+1}) - E(X_i) = 0$$

$$\therefore \quad E[C\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2]$$

$$= C\sum_{i=1}^{n-1}\{D(X_{i+1} - X_i) + [E(X_{i+1} - X_i)]^2\}$$

$$= C\sum_{i=1}^{n-1} 2D(X) = C \cdot 2(n-1)D(X)$$

依题意，要求： $E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=D(X)$

即 $C\cdot 2(n-1)D(X)=D(X)$

$\because\ D(X)>0\quad \therefore\ C=\dfrac{1}{2(n-1)}.$

三、有效性

比较参数θ的两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，如果在样本容量n相同的情况下，$\hat{\theta}_1$的观察值在真值θ的附近较$\hat{\theta}_2$更密集，则认为$\hat{\theta}_1$较 $\hat{\theta}_2$ 有效.

由于方差是随机变量取值与其数学期望的偏离程度，所以无偏估计以方差小者为好.

1 有效方差

定义3.3 设 $\hat{\theta}_1$与 $\hat{\theta}_2$是 θ 的无偏估计量，若

$$\boxed{D(\hat{\theta}_1) < D(\hat{\theta}_2) \qquad (2.3)}$$

则称 $\hat{\theta}_1$较 $\hat{\theta}_2$有效。

例7 (续例4) 在例 4中已证明 $\hat{\theta}_1 = 2\overline{X}$ 和 $\hat{\theta}_2 = \dfrac{n+1}{n}\max\{X_1, X_2, \cdots, X_n\}$ 都是 θ 的无偏估计量，现证当 $n \geq 2$ 时，$\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效．

证明 由于 $D(\hat{\theta}_1) = 4D(\overline{X}) = \dfrac{4}{n}D(X) = \dfrac{\theta^2}{3n}$,

$$D(\hat{\theta}_2) = D\left(\frac{n+1}{n}X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 D\left(X_{(n)}\right),$$

又因为 $E(X_{(n)}) = \dfrac{n+1}{n}\theta$,

$$E(X_{(n)}^2)=\int_0^\theta \frac{n}{\theta^n}x^{n+1}\mathrm{d}x=\frac{n}{n+2}\theta^2,$$

$$D(X_{(n)})=E(X_{(n)}^2)-[E(X_{(n)})]^2$$

$$=\frac{n}{(n+1)^2(n+2)}\theta^2,$$

故 $D(\hat{\theta}_2)=\dfrac{1}{n(n+2)}\theta^2,$

又 $n\geq 2$，所以 $D(\hat{\theta}_2)<D(\hat{\theta}_1)$， $\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效 .

例3 设总体 X 的均值和方差均存在，$X_1, \cdots, X_n$ 是总体 X 的样本，$C_1, C_2, \cdots, C_n$ 为不全相同且满足 $\sum_{i=1}^{n} C_i = 1$ 的任一组常数，

证明：（1）样本的线性函数 $\sum_{i=1}^{n} C_i X_i$ 是总体均值 μ 的无偏估计量；

（2）总体均值的无偏估计量 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 较 $\sum_{i=1}^{n} C_i X_i$ 有效.

证（1） $\because E(\sum_{i=1}^{n} C_i X_i) = \sum_{i=1}^{n} C_i \cdot EX_i = \mu \sum_{i=1}^{n} C_i = \mu$

$\therefore \sum_{i=1}^{n} C_i X_i$ 是 μ 的无偏估计量；

（2） 由柯西—许瓦兹不等式 知

$$1 = (\sum_{i=1}^{n} C_i)^2 < \sum_{i=1}^{n} 1^2 \cdot \sum_{i=1}^{n} C_i^2 = n\sum_{i=1}^{n} C_i^2 \quad \Rightarrow \quad \sum_{i=1}^{n} C_i^2 > \frac{1}{n}$$

$$\therefore D(\sum_{i=1}^{n} C_i X_i) = \sum_{i=1}^{n} C_i^2 \cdot DX_i = \sigma^2 \sum_{i=1}^{n} C_i^2 > \frac{\sigma^2}{n} = D\overline{X}.$$

这表明，在 μ 的所有线性无偏估计量中，样本均值 $\overline{X}$ 是最有效的.

例4 设$X_1,X_2,\ldots,X_n$和$Y_1,Y_2,\ldots,Y_m$是两组随机样本,分别取自总体$X\sim N(\mu,1)$和$Y\sim N(\mu,2^2)$,

$$T=a\sum_{i=1}^{n}X_i+b\sum_{j=1}^{m}Y_j.$$

(1)当a,b满足什么关系时,T是μ的无偏估计?

(2)当a,b分别取何值时,T最有效?

解:(1) $E(T)=a\cdot n\mu+b\cdot m\mu\;=(na+mb)\mu$

$\Rightarrow$当$na+mb=1$时, $E(T)=\mu$

此时,T是μ的无偏估计

$$(2)\ D(T)=a^2\cdot n+b^2\cdot 4m = na^2 + 4m\left(\frac{1-na}{m}\right)^2$$

$$= na^2 + \frac{4(1-na)^2}{m}$$

$$令\ \frac{dD}{da} = 0 \Rightarrow 2na - \frac{8n(1-na)}{m} = 0$$

$$\Rightarrow (4n+m)a=4 \Rightarrow a = \frac{4}{4n+m}$$

$D''(a)>0 \Rightarrow$此时$D(T)$最小,即T最有效

$$\therefore a = \frac{4}{4n+m},\quad b = \frac{1}{4n+m}$$

6.2.4　均方误差

在均方误差的标准下，无偏估计不一定比有偏估计更优。

评价一个点估计的好坏一般可以用：点估计值$\hat{\theta}$与参数真值θ的距离平方的期望，这就是下式给出的均方误差

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

均方误差是评价点估计的最一般的标准。我们希望估计的均方误差越小越好。

注意到 $MSE(\hat{\theta})=\mathrm{Var}(\hat{\theta})+(E\hat{\theta}-\theta)^2$ ，因此

(1) 若 $\hat{\theta}$ 是θ的无偏估计，$MSE(\hat{\theta})=\mathrm{Var}(\hat{\theta})$ ，这说明用方差考察无偏估计有效性是合理的。

(2) 当$\hat{\theta}$ 不是θ的无偏估计时，就要看其均方误差$MSE(\hat{\theta})$。

下面的例子说明：在均方误差的含义下有些有偏估计优于无偏估计。

例6.2.8 对均匀总体$U(0,\theta)$，由θ的极大似然估计得到的无偏估计是 $\hat{\theta}=\dfrac{(n+1)}{n}X_{(n)}$ ，它的均方误差

$$MSE(\hat{\theta})=\mathrm{Var}(\hat{\theta})=\frac{\theta^2}{n(n+2)}$$

现我们考虑 θ 的形如 $\hat{\theta}_\alpha=\alpha\cdot x_{(n)}$ 的估计，其均方差为

$$MSE(\hat{\theta}_\alpha)=\alpha^2\frac{n}{(n+1)^2(n+2)}\theta^2+\left(\frac{n\cdot\alpha}{n+1}-1\right)^2\theta^2$$

用求导的方法不难求出当 $\alpha_0=(n+2)/(n+1)$ 时上述均方误差达到最小，且其均方误差

$$MSE(\hat{\theta}_0)=\frac{\theta^2}{(n+1)^2}<\frac{\theta^2}{n(n+2)}=MSE(\hat{\theta})$$

所以 在均方误差的标准下，有偏估计 $\hat{\theta}_0$ 优于无偏估计 $\hat{\theta}$ 。

6.4.2　最小方差无偏估计

<u>定义6.4.2</u>　对参数估计问题，设 $\hat{\theta}$ 是θ的一个无偏估计，如果对另外任意一个θ的无偏估计$\tilde{\theta}$，在参数空间Θ上都有

$$\mathrm{Var}_{\theta}(\hat{\theta}) \le \mathrm{Var}_{\theta}(\tilde{\theta})$$

则称 $\hat{\theta}$ 是θ的一致最小方差无偏估计简记为 **UMVUE**。

1 最小方差无偏估计的判别法

<u>定理6.4.1</u> 设 X=($X_1, X_2, \ldots, X_n$) 是来自某总体的一个样本，$\hat{\theta} = \hat{\theta}(X)$是$\theta$的一个无偏估计，$\mathrm{Var}(\hat{\theta}) < \infty$. 如果对任意一个满足$E(\varphi(X))=0$的$\varphi(X)$，都有

$$\mathrm{Cov}_{\theta}(\hat{\theta}, \varphi) = 0, \qquad \forall \theta \in \Theta$$

则 $\hat{\theta}$是θ的UMVUE。

注

1　此定理是最小方差无偏估计的判别法，但无法寻求最小方差无偏估计的存在性.

2　由于$\varphi(X)$的任意性，因而很难利用定理判别.

由下例可以看出，利用判别定理进行判别，非常复杂，况且也无法利用此定理去寻求UMVUE.

例6.4.2　设 $X_1, X_2, \ldots, X_n$ 是来自指数分布 $Exp(1/\theta)$ 的样本，则 $T = X_1 + \ldots + X_n$ 是 θ 的充分统计量，$\bar{X} = T/n$ 是 θ 的无偏估计。设 $\varphi = \varphi(X_1, X_2, \ldots, X_n)$ 是0的任一无偏估计，则

$$\int_0^{\infty} \cdots \int_0^{\infty} \varphi(x_1, \cdots, x_n) \cdot e^{-(x_i + \cdots + x_n)/\theta} \mathrm{d}x_1 \cdots \mathrm{d}x_n = 0$$

两端对 θ 求导得

$$\int_0^{\infty} \cdots \int_0^{\infty} \frac{n\bar{x}}{\theta^2} \varphi(x_1, \cdots, x_n) \cdot e^{-(x_i + \cdots + x_n)/\theta} \mathrm{d}x_1 \cdots \mathrm{d}x_n = 0$$

这说明 $E(\bar{x} \cdot \varphi) = 0$，从而 $\mathrm{Cov}(\bar{x}, \varphi) = E(\bar{x} \cdot \varphi) - E(\bar{x}) \cdot E(\varphi) = 0$，由定理6.4.1，它是 θ 的UMVUE。

例 设$X=(X_1,X_2,\cdots,X_n)^T$是来自总体(μ,σ^2)的一个样本，已知$\overline{X}$和S_n^{*2}是μ和σ^2的无偏估计，证明$\overline{X}$和S_n^{*2}分别是μ和σ^2的$MVUE$.

证 设$L(X)$满足$EL(X)=0$,则

$$\int\cdots\int L\cdot\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathrm{d}x=0$$

两边关于μ求导，则

$$\int\cdots\int L\cdot\sum_{i=1}^{n}x_i\cdot\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathrm{d}x=0$$

因而 $E(L(X)\overline{X})=0$

故$\overline{X}$是μ的$MVUE$.

对此式$\int\cdots\int L\cdot\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathbf{d}x=0$关于$\mu$

求二阶导数，则

$$\int\cdots\int L\cdot(\sum_{i=1}^{n}x_i\cdot)^2\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathbf{d}x=0$$

对此式$\int\cdots\int L\cdot\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathbf{d}x=0$关于$\sigma^2$

求导数，则

$$\int\cdots\int L\cdot\sum_{i=1}^{n}(x_i-\mu)^2\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathbf{d}x=0$$

又由于$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i-\mu)^2-n(\bar{x}-\mu)^2$,可得

$$\int\cdots\int L\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}\mathrm{d}x=0$$

因而 $E(L(X)S_n^{*2})=0$

故S_n^{*2}是σ^2的$MVUE$.

6.4.2 充分性原则

以下定理说明：充分统计量是一个有力工具

定理6.4.2　设总体概率函数是 $f(x, \theta)$，　$X_1, X_2, \ldots, X_n$ 是其样本，$T=T(X_1, X_2, \ldots, X_n)$是$\theta$的充分统计量，则

对θ的任一无偏估计$\hat{\theta}=\hat{\theta}(x_1, \cdots, x_n)$，令$\tilde{\theta}=E(\hat{\theta} \mid T)$，

$\tilde{\theta}$ 也是θ的无偏估 $\mathrm{Var}(\tilde{\theta}) \le \mathrm{Var}(\hat{\theta})$

定理6.4.2说明：如果无偏估计不是充分统计量的函数，则将之对充分统计量求条件期望可以得到一个新的无偏估计，该估计的方差比原来的估计的方差要小，从而降低了无偏估计的方差。换言之，考虑θ的估计问题只需要在基于充分统计量的函数中进行即可，该说法对所有的统计推断问题都是正确的，这便是所谓的充分性原则。

注 最小方差无偏估计计算方法

1、构造一个充分完备统计量$T(X_1,\cdots,X_n)$和一个θ的无偏估计$\hat{\theta}$.

2、计算数学期望$E(\hat{\theta}|T)$,即得θ的一个$MVUE$.

例6.3.1　设 $x_1, x_2, \ldots, x_n$ 是来自$b(1, p)$的样本，则 $T = n\bar{x}$ 是p 的充分统计量。为估计$\theta=p^2$，可令

$$\hat{\theta}_1 = \begin{cases} 1, & x_1 = 1, x_2 = 1 \\ 0, & \text{其它} \end{cases}$$

由于 $E(\hat{\theta}_1) = P(x_1 = 1, x_2 = 1) = p \cdot p = \theta$ ，所以 $\hat{\theta}_1$ 是θ的无偏估计。这个只使用了两个观测值的估计并不好. 下面我们用**Rao−Blackwell**定理对之加以改进：求 $\hat{\theta}_1$ 关于充分统计量 $T = \sum_{i=1}^{n} x_i$ 的条件期望，得

$$\overset{\frown}{\theta} = E(\overset{\frown}{\theta}_1 \mid T = t) = \binom{n-2}{t-2} \Big/ \binom{n}{t} = \frac{t(t-1)}{n(n-1)}$$

6.4.4 Cramer-Rao不等式

上一节介绍了最小方差无偏估计以及相应的寻求方法。自然会引入另一个问题：最小方差无偏估计是否可以任意的小？是否有下界？事实上， Rao-Cramer不等式可以回答此问题。

定义6.4.3　设总体的概率函数 $P(x,\theta)$, $\theta\in\Theta$满足下列条件：

(1) 参数空间Θ是直线上的一个开区间；

(2) 支撑 $S=\{x: P(x,\theta)>0\}$与θ无关；

(3) 导数 $\frac{\partial}{\partial\theta}p(x;\theta)$ 对一切$\theta\in\Theta$都存在；

(4) 对$P(x,\theta)$，积分与微分运算可交换次序；

(5) 期望 $E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2$ 存在；

则称 $I(\theta)=E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2$

为总体分布的费希尔(Fisher) 信息量。

费希尔信息量是数理统计学中一个基本概念，很多的统计结果都与费希尔信息量有关。如极大似然估计的渐近方差，无偏估计的方差的下界等都与费希尔信息量$I(\theta)$有关。

例6.3.3　设总体为泊松分布$P(\lambda)$分布，

则

$$\ln p(x;\lambda) = x\ln\lambda - \lambda - \ln(x!)$$

$$\frac{\partial}{\partial\lambda}\ln p(x;\lambda) = \frac{x}{\lambda} - 1$$

于是 $I(\lambda) = E\left(\frac{X-\lambda}{\lambda}\right)^2 = \frac{1}{\lambda}$

例6.3.4　设总体为指数分布，其密度函数为

$$p(x;\theta)=\frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\},\quad x>0,\ \theta>0$$

可以验证定义6.3.2的条件满足，且

$$\frac{\partial}{\partial\theta}\ln p(x;\theta)=\frac{1}{\theta}-\frac{x}{\theta^{2}}=-\frac{x-\theta}{\theta^{2}}$$

于是

$$I(\theta)=E\left(\frac{x-\theta}{\theta^{2}}\right)^{2}=\frac{\mathrm{Var}(x)}{\theta^{4}}=\frac{1}{\theta^{2}}$$

定理6.3.4（Cramer−Rao不等式）

设定义6.3.2的条件满足，$X_1, X_2, \ldots, X_n$ 是来自该总体的样本，$T=T(X_1, X_2, \ldots, X_n)$是$g(\theta)$的任一个无偏估计，$g'(\theta)=\dfrac{\partial g(\theta)}{\partial\theta}$ 存在，且对$\theta\in\Theta$中一切θ，微分可在积分号下进行，则有

$$\mathrm{Var}(T)\geq\frac{[g'(\theta)]^2}{nI(\theta)}$$

- 上式称为克拉美−罗（C-R）不等式；
- $[g'(\vartheta)]^2/(nI(\theta))$称为$g(\theta)$的无偏估计的方差的C−R下界，简称$g(\theta)$的C−R下界。
- 特别，对θ的无偏估计$\hat{\theta}$，有 $\mathrm{Var}(\hat{\theta}) \geq (nI(\theta))^{-1}$
- 如果等号成立，则称 $T=T(x_1, \ldots, x_n)$ 是$g(\theta)$的有效估计，有效估计一定是UMVUE。

例6.4.6　设总体分布列为$p(x,\theta)=\theta^{x}(1-\theta)^{1-x}$，$x=0,1$，它满足定义6.4.3的所有条件，可以算得该分布的费希尔信息量$I(\theta)=\dfrac{1}{\theta(1-\theta)}$，若 $x_1, x_2, \ldots, x_n$ 是该总体的样本，则θ的C−R下界为$(nI(\theta))^{-1}=\theta(1-\theta)/n$。因为$\bar{x}$是$\theta$的无偏估计，且其方差等于$\theta(1-\theta)/n$，达到C−R 下界，所以$\bar{x}$是$\theta$的有效估计，它也是$\theta$的UMVUE。

例6.4.7　设总体为指数分布$Exp(1/\theta)$，它满足定义6.3.2的所有条件，例6.3.4中已经算出该分布的费希尔信息量为$I(\theta)=\theta^{-2}$，若$x_1, x_2, \ldots, x_n$是样本，则θ的C−R下界为$(nI(\theta))^{-1}=\theta^2/n$。而 $\bar{x}$ 是θ的无偏估计，且其方差等于θ^2/n，达到了C−R下界，所以，$\bar{x}$是θ的有效估计，它也是θ的UMVUE。

能达到C−R下界的无偏估计不多：

例6.4.8　设总体为$N(0,\sigma^2)$，满足定义6.3.2的条件，且费希尔信息量为$I(\sigma^2)=\dfrac{1}{2\sigma^4}$，

令 $\sigma=g(\sigma^2)=\sqrt{\sigma^2}$，

则σ 的C−R下界为$\dfrac{[g'(\sigma^2)]^2}{nI(\sigma^2)}=\dfrac{\sigma^2}{2n}$，

而σ的UMVUE为 $\hat{\sigma}=\sqrt{\dfrac{n}{2}}\cdot\dfrac{\Gamma(n/2)}{\Gamma((n+1)/2)}\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}x_i^2}$

其方差大于C−R下界。这表明所有σ的无偏估计的方差都大于其C−R下界。

费希尔信息量的主要作用体现在极大似然估计。

<u>定理6.3.5</u>　设总体*X*有密度函数 $p(x;\theta)$，$\theta\in\Theta$，Θ为非退化区间，假定

(1) 对任意的*x*，偏导数 $\dfrac{\partial \ln p}{\partial\theta}$，$\dfrac{\partial^2 \ln p}{\partial\theta^2}$ 和 $\dfrac{\partial^3 \ln p}{\partial\theta^3}$ 对所有$\theta\in\Theta$都存在；

(2) $\forall\theta\in\Theta$,

有 $\left|\dfrac{\partial p}{\partial\theta}\right|< F_1(x)$, $\left|\dfrac{\partial^2 p}{\partial\theta^2}\right|< F_2(x)$, $\left|\dfrac{\partial^3 \ln p}{\partial\theta^3}\right|< F_3(x)$,

其中函数$F_1(x)$, $F_2(x)$, $F_3(x)$可积.

(3) $\forall\theta\in\Theta$, $0 < I(\theta) \equiv \int_{-\infty}^{\infty}\left(\frac{\partial \ln p}{\partial\theta}\right)^2 p(x;\theta)\mathrm{d}x < \infty$

若 $x_1, x_2, \ldots, x_n$ 是来自该总体的样本，则存在未知参数θ的极大似然估计 $\hat{\theta}_n = \hat{\theta}_n(x_1,\cdots,x_n)$ ，且 $\hat{\theta}_n$ 具有相合性和渐近正态性：

$$\hat{\theta}_n \dot{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

定理 设总体X的分布函数为$F(x,\theta),\theta\in\Theta$是未知参数，$X=(X_1,X_2,\cdots,X_n)^T$是来自总体$X$的一个样本，如果$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是$\theta$的无偏估计量，则$\hat{\theta}$是$\theta$的有效估计的充分必要条件为：

1、$\hat{\theta}$是θ的充分估计量；

2、$$\frac{\partial \ln L(x,\theta)}{\partial\theta}=C(\theta)[\hat{\theta}(x_1,x_2,\cdots,x_n)-\theta]$$

其中$L(x,\theta)$是样本的联合分布密度，$C(\theta)$仅依赖参数θ.

证明从略。

例 设 X 服从两点分布 $B(1,p)$, X_1, $X_2,\cdots,X_n$ 是来自 X 的一个样本,证明 p 的最大似然估计量是有效估计.

解　因为 X 的分布律为

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

所以 p 的似然函数为

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

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

$$\frac{\mathrm{d}}{\mathrm{dp}}\ln L(p)=\frac{\sum\limits_{i=1}^{n} x_i}{p}-\frac{n-\sum\limits_{i=1}^{n}(x_i)}{1-p}=\frac{n}{p(1-p)}\{\bar{x}-p\}=0$$

解得 p 的最大似然估计值 $\hat{p}=\frac{1}{n}\sum\limits_{i=1}^{n} x_i=\bar{x},$

又因为 $\frac{\partial \ln L(x,p)}{\partial p}=C(p)[\hat{p}(x_1,x_2,\cdots,x_n)-p],$ 其中

$C(p)=\frac{n}{p(1-p)},$ 同时 $\bar{X}$ 是p的充分统计量，因而由定理3.11可知，$\bar{X}$ 是p的有效估计.

§6.4 贝叶斯估计

6.4.1 统计推断的基础

- 经典学派的观点：统计推断是根据样本信息对总体分布或总体的特征数进行推断，这里用到两种信息：总体信息和样本信息；
- 贝叶斯学派的观点：除了上述两种信息以外，统计推断还应该使用第三种信息：先验信息。

（1）总体信息:总体分布提供的信息。

（2）样本信息:抽取样本所得观测值提供的信息。

（3）先验信息:人们在试验之前对要做的问题在经验上和资料上总是有所了解的，这些信息对统计推断是有益的。先验信息即是抽样（试验）之前有关统计问题的一些信息。一般说来，先验信息来源于经验和历史资料。先验信息在日常生活和工作中是很重要的。

基于上述三种信息进行统计推断的统计学称为贝叶斯统计学。它与经典统计学的差别就在于是否利用先验信息。贝叶斯统计在重视使用总体信息和样本信息的同时，还注意先验信息的收集、挖掘和加工，使它数量化，形成先验分布，参加到统计推断中来，以提高统计推断的质量。忽视先验信息的利用，有时是一种浪费，有时还会导出不合理的结论。

贝叶斯学派的基本观点：任一未知量θ都可看作随机变量，可用一个概率分布去描述，这个分布称为先验分布；在获得样本之后，总体分布、样本与先验分布通过贝叶斯公式结合起来得到一个关于未知量θ新的分布——<u>后验分布</u>；任何关于θ的统计推断都应该基于θ的后验分布进行。

6.4.2 贝叶斯公式的密度函数形式

- 总体X依赖于参数θ的概率函数在贝叶斯统计中记为$f(X \mid \theta)$，它表示在随机变量 θ取某个给定值时总体的条件概率函数；
- 根据参数θ的先验信息可确定先验分布$\pi(\theta)$；
- 从贝叶斯观点看，样本 $X_1, X_2, \ldots, X_n$的产生分两步进行：首先从先验分布$\pi(\theta)$产生一个样本θ_0，然后从$f(X \mid \theta_0)$中产生一组样本。这时样本的联合条件概率函数为 $p(X_1, \cdots, X_n \mid \theta_0) = \prod_{i=1}^{n} f(X_i \mid \theta_0)$ ，
- 这个分布综合了总体信息和样本信息；

- θ_0 是未知的，它是按先验分布$\pi(\theta)$产生的。为把先验信息综合进去，不能只考虑θ_0，对θ的其它值发生的可能性也要加以考虑，故要用$\pi(\theta)$进行综合。这样一来，样本$X_1,\ldots,X_n$和参数θ的联合分布为：

$$h(X_1, X_2, \ldots, X_n, \theta) = f(X_1, X_2, \ldots, X_n | \theta)\pi(\theta),$$

这个联合分布把总体信息、样本信息和先验信息三种可用信息都综合进去了；

- 在没有样本信息时，人们只能依据先验分布对θ作出推断。在有了样本观察值 x_1, x_2, …, x_n之后，则应依据 $h(x_1, x_2, \dots, x_n, \theta)$对$\theta$作出推断。由于

$$h(x_1,x_2,\dots,x_n,\theta)=\pi(\theta|x_1,x_2,\dots,x_n)m(x_1,x_2,\dots,x_n),$$

其中 $m(x_1,\cdots,x_n)=\int_\Theta h(x_1,\cdots,x_n,\theta)d\theta=\int_\Theta p(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta$

是x_1, x_2, …, x_n的边际概率函数，它与θ无关，不含θ的任何信息。因此能用来对θ作出推断的仅是条件分布$\pi(\theta|x_1, x_2, \dots, x_n)$，它的计算公式是

$$\pi(\theta|x_1,\cdots,x_n)=\frac{h(x_1,\cdots,x_n,\theta)}{m(x_1,\cdots,x_n)}=\frac{p(x_1,\cdots,x_n|\theta)\pi(\theta)}{\int_\Theta p(x_1,\cdots,x_n|\theta)\pi(\theta)d\theta}$$

这个条件分布称为θ的后验分布，它集中了总体、样本和先验中有关θ的一切信息。

后验分布$\pi(\theta | x_1, x_2, \ldots, x_n)$的计算公式就是用密度函数表示的贝叶斯公式。它是用总体和样本对先验分布$\pi(\theta)$作调整的结果，贝叶斯统计的一切推断都基于后验分布进行。

6.4.3 贝叶斯估计

基于后验分布$\pi(\theta|x_1, x_2, \ldots, x_n)$对$\theta$所作的贝叶斯估计有多种，常用有如下三种：

- 使用后验分布的密度函数最大值作为θ的点估计，称为最大后验估计；
- 使用后验分布的中位数作为θ的点估计，称为后验中位数估计；
- 使用后验分布的均值作为θ的点估计，称为后验期望估计。

用得最多的是后验期望估计，它一般也简称为贝叶斯估计，记为$\hat{\theta}_B$。

例6.4.2　设某事件A在一次试验中发生的概率为θ，为估计θ，对试验进行了*n*次独立观测，其中事件*A*发生了*X*次，显然 $X|\theta \sim b(n,\theta)$，即

$$P(X=x|\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\qquad x=0,1,\cdots,n$$

假若我们在试验前对事件*A*没有什么了解，从而对其发生的概率θ也没有任何信息。在这种场合，贝叶斯本人建议采用“同等无知”的原则使用区间（0,1）上的均匀分布*U*(0,1)作为θ的先验分布，因为它取（0,1）上的每一点的机会均等。贝叶斯的这个建议被后人称为贝叶斯假设。

由此即可利用贝叶斯公式求出θ的后验分布。具体如下：

先写出*X*和θ的联合分布

$$h(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\qquad x=0,1,\cdots,n,\quad 0<\theta<1$$

然后求*X*的边际分布

$$m(x)=\int_0^1\binom{n}{x}\theta^{x}(1-\theta)^{n-x}d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

最后求出θ的后验分布

$$\pi(\theta\mid x)=\frac{h(x,\theta)}{m(x)}=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\qquad 0<\theta<1$$

最后的结果说明$\theta|X \sim Be(x+1,n-x+1)$，其后验期望估计为

$$\hat{\theta}_B=E(\theta\mid x)=\frac{x+1}{n+2}\qquad(6.4.4)$$

伽玛函数表达式：

$$\Gamma(x)=\int_0^{+\infty} e^{-t}t^{(x-1)}dt=(x-1)\Gamma(x-1)$$

$$\Gamma(1)=1,\Gamma(n+1)=n!$$

U(0,1)均匀分布是贝塔分布的一个特例Be（1，1）

某些场合，贝叶斯估计要比极大似然估计更合理一点。比如:“抽检3个全是合格品”与“抽检10个全是合格品”，后者的质量比前者更信得过。这种差别在不合格品率的极大似然估计中反映不出来（两者都为0），而用贝叶斯估计两者分别是0.2 和 0.83。

由此可以看到，在这些极端情况下，贝叶斯估计比极大似然估计更符合人们的理念。

例6.4.3　设$x_1, x_2, \ldots, x_n$是来自正态分布

$N(\mu, \sigma_0^2)$的一个样本，其中σ_0^2已知，μ未知，假设μ的先验分布亦为正态分布$N(\theta, \tau^2)$，其中先验均值θ和先验方差τ^2均已知，试求μ的贝叶斯估计。

解：样本x的分布和μ的先验分布分别为

$$p(x \mid \mu) = (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

$$\pi(\mu) = (2\pi\tau^2)^{-1/2} \exp\left\{-\frac{1}{2\tau^2}(\mu - \theta)^2\right\}$$

由此可以写出x与μ的联合分布

$$h(X,\mu)=k_1\cdot\exp\left\{-\frac{1}{2}\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^{n}x_i^2}{\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{\tau^2}\right]\right\}$$

其中 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ ， $k_1=(2\pi)^{-(n+1)/2}\tau^{-1}\sigma_0^{-n}$ 。若记

$$A=\frac{n}{\sigma_0^2}+\frac{1}{\tau^2},\quad B=\frac{n\bar{x}}{\sigma_0^2}+\frac{\theta}{\tau^2},\quad C=\frac{\sum_{i=1}^{n}x_i^2}{\sigma_0^2}+\frac{\theta^2}{\tau^2}$$

则有

$$h(X,\mu)=k_1\exp\{-\frac{1}{2}[A\mu^2-2B\mu+C]\}$$

$$=k_1\exp\{-\frac{(\mu-B/A)^2}{2/A}-\frac{1}{2}(C-B^2/A)\}$$

注意到*A,B,C*均与μ无关，由此容易算得样本的边际密度函数

$$m(x)=\int_{-\infty}^{\infty}h(x,\mu)d\mu=k_1\exp\left\{-\frac{1}{2}(C-B^2/A)\right\}(2\pi/A)^{1/2}$$

应用贝叶斯公式即可得到后验分布

$$\pi(\mu\,|\,x)=\frac{h(x,\mu)}{m(x)}=(2\pi/A)^{1/2}\exp\left\{-\frac{1}{2/A}(\mu-B/A)^2\right\}$$

这说明在样本给定后，μ的后验分布为

***N*(*B*/*A*,1/*A*)**，即 $\mu\,|\,x\sim N\left(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{-2}+\tau^{-2}}\right)$

后验均值即为其贝叶斯估计：

$$\hat{\mu} = \frac{n/\sigma_0^2}{n/\sigma_0^2 + 1/\tau^2}\bar{x} + \frac{1/\tau^2}{n/\sigma_0^2 + 1/\tau^2}\theta$$

它是样本均值 $\bar{x}$ 与先验均值θ的加权平均。

6.4.4 共轭先验分布

若后验分布$\pi(\theta|x)$与$\pi(\theta)$属于同一个分布族，则称该分布族是θ的共轭先验分布(族)。

- 后验分布和先验分布是同一个类型
- 优点：易于解释、继续试验
- 已知：$p(x|\theta)$，选 $\pi(\theta)$
- 使得 $h(\theta|r) \propto p(x|\theta) * \pi(\theta)$ 与先验分布同类型
- 若p(x|θ)服从正态分布，选正态分布
- 若p(x|θ)服从两点分布，选Beta分布
- 若p(x|θ)服从指数分布，选逆Gamma分布

§6.5　区间估计

点估计不能反映估计的精度, 故而本节引入了区间估计.

定义6.1 设 $X_1, X_2, \cdots, X_n$ 是来自分布函数为 $F(x;\theta)$ 的总体 X 的一个样本， θ 为未知参数，对于给定值 α ，$0<\alpha<1$，若存在两个统计量

$$\hat{\theta}_L = \hat{\theta}_L(X_1, X_2, \cdots X_n), \quad \hat{\theta}_U = \hat{\theta}_U(X_1, X_2, \cdots X_n),$$

使得

$$\boxed{P\{\hat{\theta}_L \le \theta \le \hat{\theta}_U\} = 1-\alpha \qquad (3.1)}$$

则称随机区间$[\hat{\theta}_L,\hat{\theta}_U]$为$\theta$的置信水平为$1-\alpha$的置信区间，或简称$[\hat{\theta}_L,\hat{\theta}_U]$是$\theta$的$1-\alpha$置信区间.

$\hat{\theta}_L$和$\hat{\theta}_U$分别称为θ的（双侧）置信下限和置信上限.

注：置信水平为 $1-\alpha$ 的置信区间的含义是：

利用样本值所构造的一个随机区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 能盖住未知参数 θ 的概率为 $1-\alpha$。

因为这个随机区间会随样本观察值的不同而不同，它有时盖住了参数 θ，有时没有盖住 θ，

因此用这种方法作区间估计时，100 次中大约有 $100(1-\alpha)$ 个区间能盖住未知参数 θ，100α 个左右区间不能盖住 θ。

例6.5.1 设$x_1, x_2, \ldots, x_{10}$是来自$N(\mu,\sigma^2)$的样本，则μ的置信水平为1−α的置信区间为

$$\left[\bar{x} - t_{1-\alpha/2}(9)\, s\big/\sqrt{10}\,,\quad \bar{x} + t_{1-\alpha/2}(9)\, s\big/\sqrt{10}\right]$$

其中，$\bar{x}$ ，s 分别为样本均值和样本标准差。这个置信区间的由来将在6.5.3节中说明，这里用它来说明置信区间的含义。

若取α=0.10，则$t_{0..95}(9)$=1.8331，上式化为

$$\left[\bar{x} - 0.5797\, s,\quad \bar{x} + 0.5797\, s\right]$$

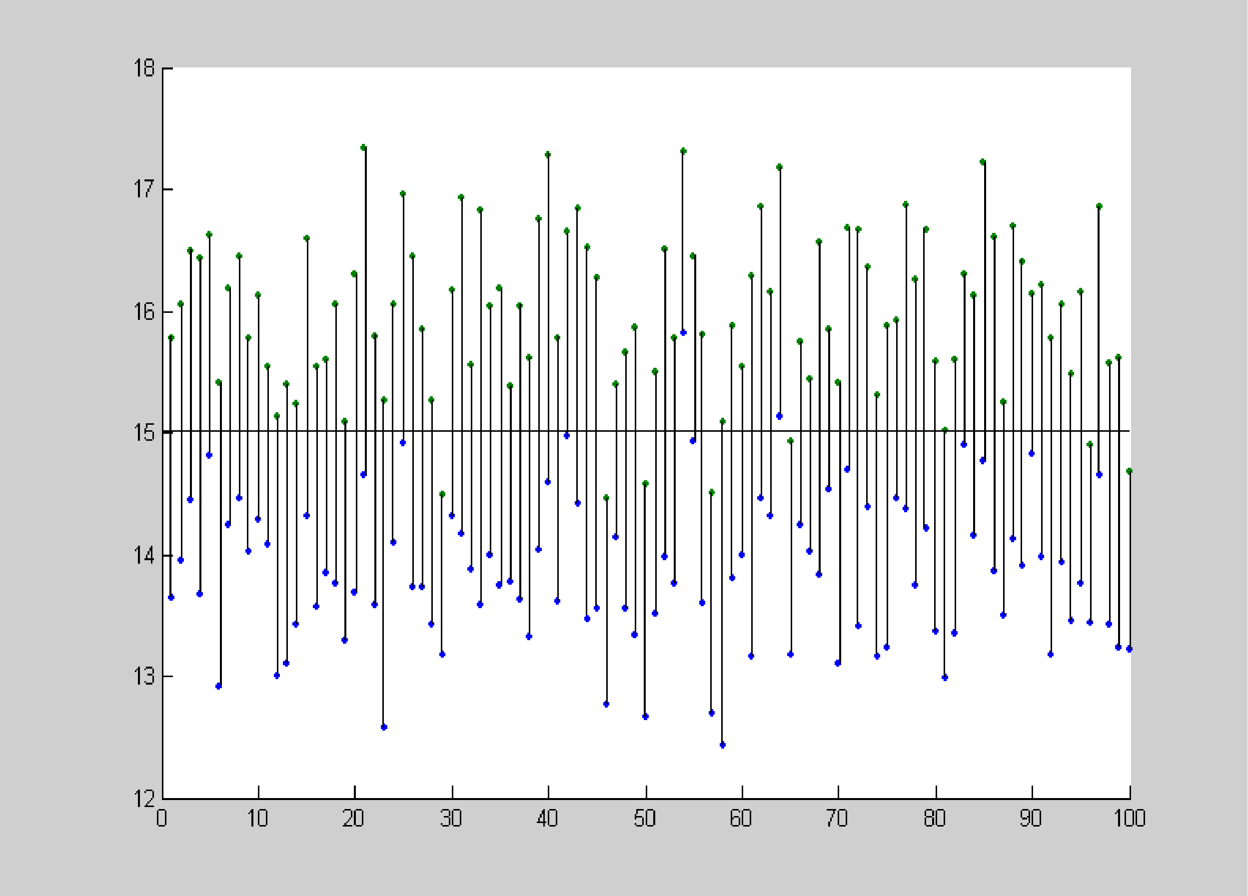
现假定μ=15,σ^2=4，则我们可以用随机模拟方法由N(15,4)产生一个容量为10的样本，如下即是这样一个样本：14.85 13.01 13.50 14.93 16.97 13.80 17.9533 13.37 16.29 12.38

由该样本可以算得 $\bar{x} = 14.7053, s = 1.8438$

从而得到μ的一个区间估计为

$$[14.7053 - 0.5797 \times 1.8438,\quad 14.7053 + 0.5797 \times 1.8438] = [13.6427,\quad 15.7679]$$

该区间包含μ的真值--15。现重复这样的方法100次，可以得到100个样本，也就得到100个区 间，我们将这100个区间画在图6.5.1上。



由图6.5.1可以看出，这100个区间中有91个包含参数真值15，另外9个不包含参数真值。

图6.5.1 μ 的置信水平为0.90的置信区间

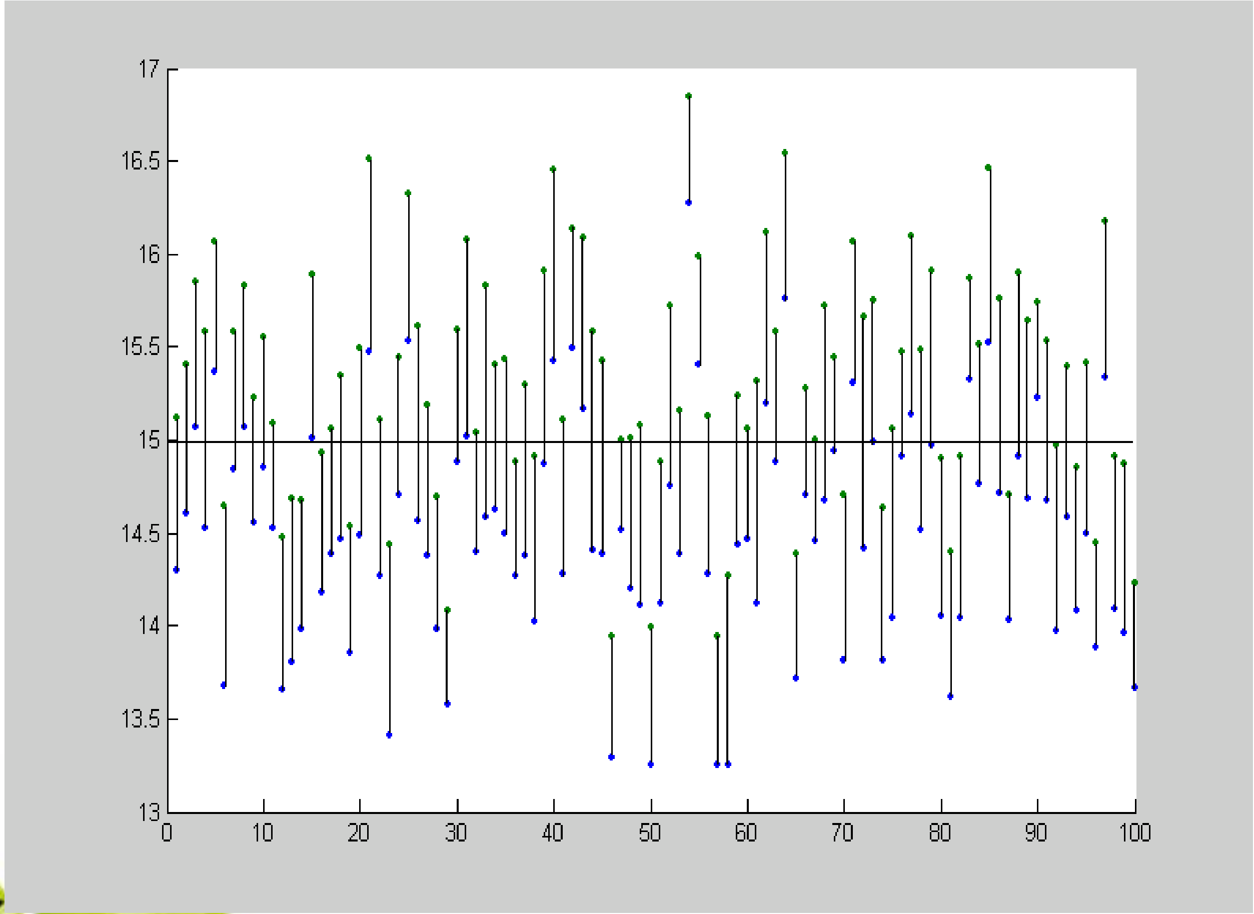


图6.5.2　μ 的置信水平为0.50的置信区间

取 α=0.50，我们也可以给出100个这样的区间，见图6.5.2。可以看出，这100个区间中有50个包含参数真值15，另外50个不包含参数真值。

例6.1 设总体 $X \sim N(\mu,\sigma^2)$，σ^2为已知，μ为未知，设$X_1, X_2, \cdots, X_n$是来自 X的样本，试求 μ的置信度为$1-\alpha$的置信区间。

解：(1)在 σ^2已知条件下，求置信度 为$1-\alpha$的未知参数 μ的置信区间，

（2）μ的点估计为 $\overline{X}$，故构造含 μ的样本函数为

$$\boxed{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}$$

(3)$\sqrt{n}\dfrac{\overline{X}-\mu}{\sigma}$落在任一区间 $[a,b]$内取值的概率都能得到，即

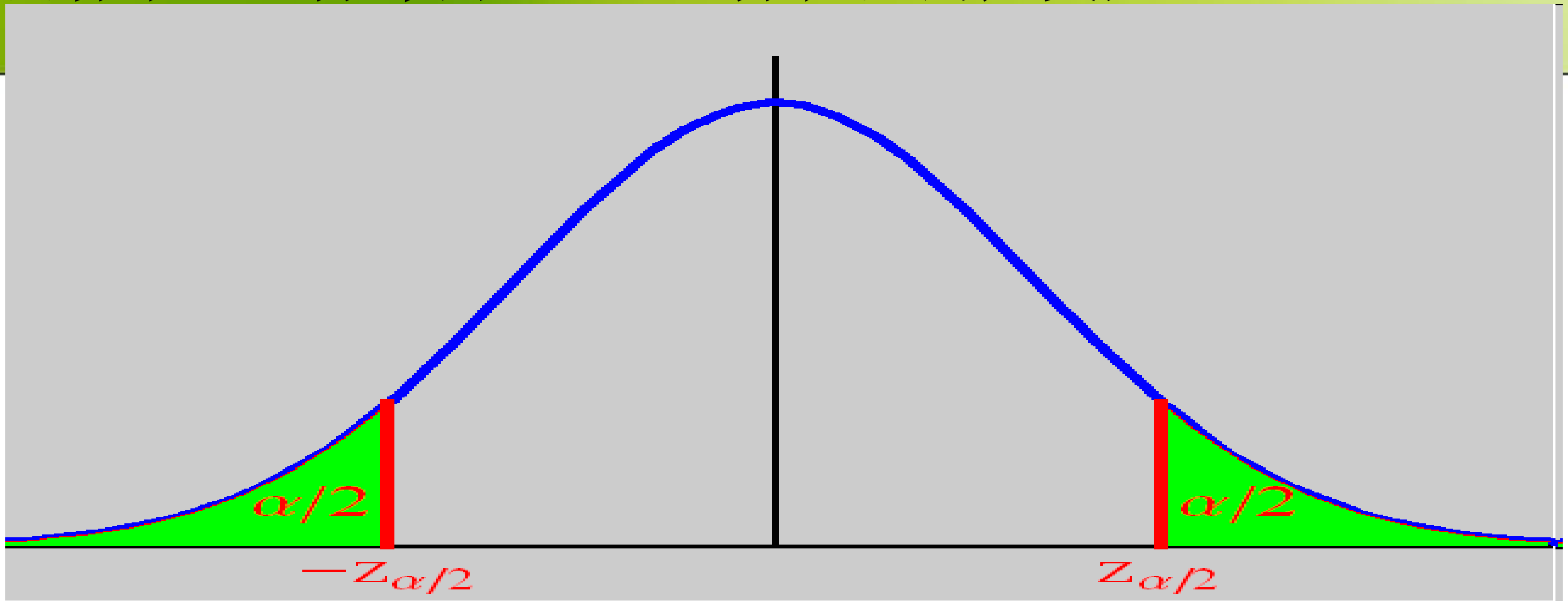
$$P\left\{a \le \sqrt{n}\frac{\overline{X}-\mu}{\sigma} \le b\right\} = \Phi(b) - \Phi(a).$$

若令 $P\left\{a \le \sqrt{n}\dfrac{\overline{X}-\mu}{\sigma} \le b\right\} = 1-\alpha$.

区间端点 $[a,b]$有无穷选择。若要使区间长度最短，只有取关于原点对称的区间。

即 $a = -z_{\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}, b = z_{\frac{\alpha}{2}}$.

由标准正态分布的上 α 分位点的定义知



即 $P\left\{\overline{X}-\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\}=1-\alpha,$

由 $\Phi(z_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$，查表得 $z_{\alpha/2}$ 的值；

于是得 μ 的一个置信水平为 $1-\alpha$ 的置信区间

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right).$$

这样的置信区间常写成 $\left(\overline{X} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right).$

其置信区间的长度为 $2 \times \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}.$

注意：置信水平为$1-\alpha$的置信区间是不唯一的

如果在本例中取 $n=16,\ \sigma=1,\ \alpha=0.05$,

查表可得 $z_{\alpha/2}=z_{0.025}=1.96$,

得一个置信水平为 0.95 的置信区间 $\left(\overline{X}\pm\dfrac{1}{\sqrt{16}}\times 1.96\right)$.

由一个样本值算得样本均值的观察值 $\overline{x}=5.20$,

则置信区间为(5.20 ± 0.49)，即 $(4.71,\ 5.69)$.

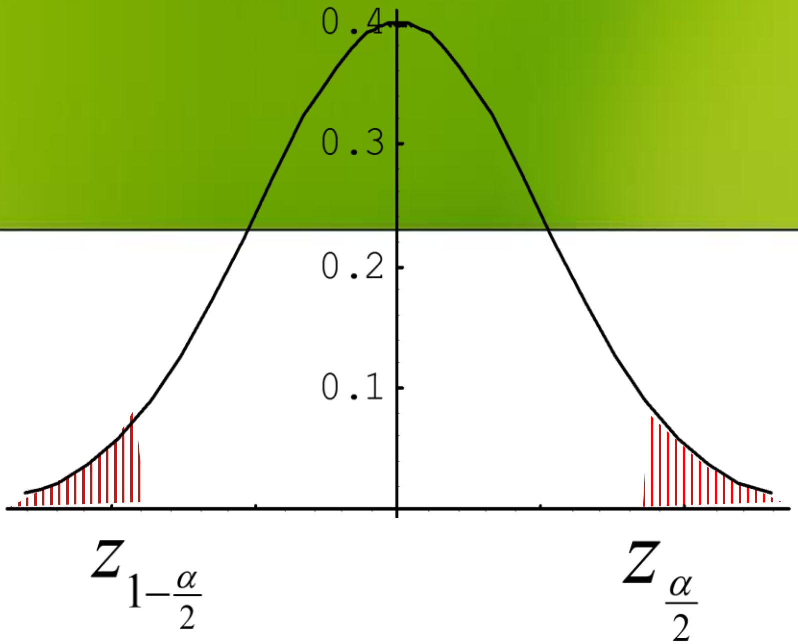
就是说估计总体的均值在4.71与5.69之间, 这个估计的可信程度为95%.

若依此区间内任一值作 为 μ 的近似值 ，

其误差不大于 $$\left(\frac{1}{\sqrt{16}}\times 1.96\times 2\right)=0.98$$

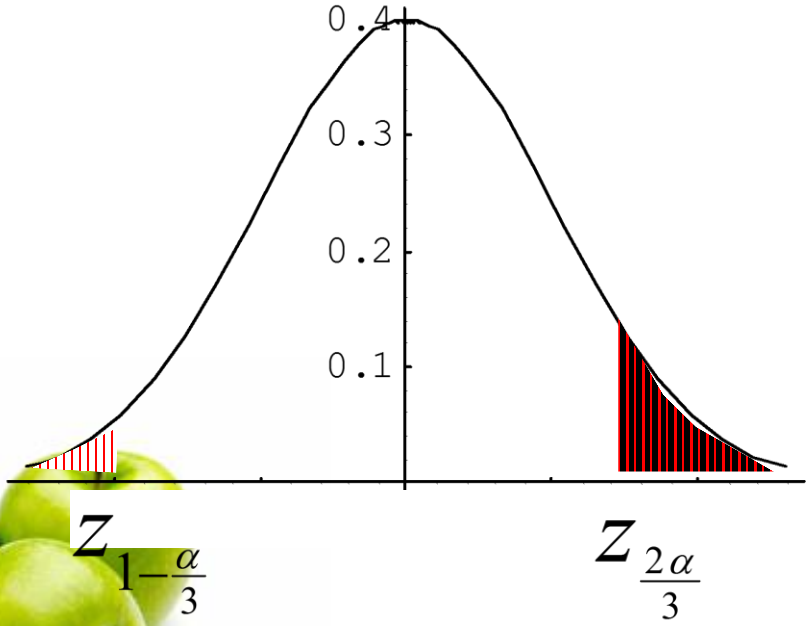
这个误差的可信度为95%.

取 $\alpha = 0.05$



$$z_{\frac{\alpha}{2}} - z_{1-\frac{\alpha}{2}} = 1.96 - (-1.96)$$

$$= 3.92$$



$$z_{\frac{2\alpha}{3}} - z_{1-\frac{\alpha}{3}} = 1.84 - (-2.13)$$

$$= 3.97$$

说明: 对于概率密度的图形是单峰且关于纵坐标轴对称的情况, 易证取a和b关于原点对称时,能使置信区间长度最小.

6.5.2　枢轴量法

构造未知参数θ的置信区间的最常用的方法是枢轴量法，其步骤可以概括为如下三步：

1. 设法构造一个样本和θ的函数 $G=G(x_1, x_2, \ldots, x_n, \theta)$ 使得G的分布不依赖于未知参数。一般称具有这种性质的G为枢轴量。
2. 适当地选择两个常数c,d，使对给定的α $(0<\alpha<1)$ 有
$$P(c \le G \le d)=1-\alpha$$
3. 假如能将$c \leqslant G \leqslant d$ 进行不等式等价变形化为
$$\hat{\theta}_L \le \theta \le \hat{\theta}_U$$
则$[\hat{\theta}_L, \hat{\theta}_U]$是$\theta$的$1-\alpha$ 同等置信区间。

关于置信区间的构造有两点说明：

- 满足置信度要求的c与d通常不唯一。若有可能，应选平均长度$E(\hat{\theta}_U-\hat{\theta}_L)$达到最短的$c$与$d$，这在$G$的分布为对称分布场合通常容易实现。
- 实际中，选平均长度$E(\hat{\theta}_U-\hat{\theta}_L)$尽可能短的$c$与$d$，这往往很难实现，因此，常这样选择 c与d，使得两个尾部概率各为$\alpha/2$，即$P(G<c)=P(G>d)=\alpha/2$，这样的置信区间称为等尾置信区间。这是在G的分布为偏态分布场合常采用的方法。

6.6.3 单个正态总体参数的区间估计

设给定置信水平为 $1-\alpha$，并设 $X_1, X_2, \cdots, X_n$ 为总体 $N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 分别是样本均值和样本方差．

1. 均值 μ 的置信区间

(1) σ^2 为已知，由上节例6可知:

μ 的一个置信水平为 $1-\alpha$ 的置信区间 $\left(\overline{X} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$.

例6.2 包糖机某日开工包了12包糖，称得质量(单位：克)分别为506, 500, 495, 488, 504, 486, 505, 513, 521, 520, 512, 485. 假设重量服从正态分布，且标准差为 $\sigma = 10$，试求糖包的平均质量 μ 的 $1-\alpha$ 置信区间 (分别取 $\alpha = 0.10$ 和 $\alpha = 0.05$).

解 $\sigma = 10$, $n = 12$,

计算得 $\bar{x} = 502.92$,

(1) 当 $\alpha = 0.10$ 时, $1 - \dfrac{\alpha}{2} = 0.95$,

查表得 $z_{\alpha/2} = z_{0.05} = 1.645$,

$$\bar{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}=502.92-\frac{10}{\sqrt{12}}\times 1.645=498.17,$$

$$\bar{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}=502.92+\frac{10}{\sqrt{12}}\times 1.645=507.67,$$

即 μ 的置信度为 90% 的置信区间为

(498.17，507.67).

(2) 当 $\alpha=0.05$ 时， $1-\frac{\alpha}{2}=0.975$,

查表得

$$z_{\alpha/2} = z_{0.025} = 1.96,$$

同理可得 μ 的置信度为 95% 的置信区间为

(497.26, 508.58).

从此例可以看出,

当置信度 $1-\alpha$ 较大时, 置信区间也较大;

当置信度 $1-\alpha$ 较小时, 置信区间也较小.

(2) σ^2为未知,

μ的置信度为$1-\alpha$的置信区间$\left(\overline{X} \pm \dfrac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$.

推导过程如下:

由于区间 $\left(\overline{X} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$中含有未知参数 σ,不能

直接使用此区间 ,

但因为 S^2 是 σ^2 的无偏估计 ,可用 $S = \sqrt{S^2}$ 替换 σ,

又根据第四章定理 4.5.1(4) 知 $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$,

则 $$P\left\{-t_{\alpha/2}(n-1) < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

即 $$P\left\{\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1) < \mu < \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

于是得μ的置信度为$1-\alpha$的置信区间

$$\left(\overline{X}\pm\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right).$$

例6.3 （续例6.2）如果只假设糖包的重量服从正态分布

$N(\mu,\sigma^2)$，试求糖包重量 μ 的 95% 的置信区间.

解 此时 σ 未知， $n = 12$，

$\alpha = 0.05$， $\bar{x} = 502.92$， $s = 12.35$，

查 $t(n-1)$ 分布表可知： $t_{0.025}(11) = 2.201$，

于是 $\dfrac{s}{\sqrt{n}} t_{\alpha/2}(n-1) = \dfrac{12.35}{\sqrt{12}} \times 2.201 = 7.85$，

得 μ 的置信度为 95% 的置信区间 $(495.07, 510.77)$.

2. 方差 σ^2 的置信区间

方差 σ^2 的置信度为 $1-\alpha$ 的置信区间

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

推导过程如下:

因为 S^2 是 σ^2 的无偏估计 ，

根据第四章定理4.5.1知 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$

则 $$P\left\{\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

即 $$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right\} = 1-\alpha,$$

于是得方差 σ^2 的置信度为 $1-\alpha$ 的置信区间

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

进一步可得:

标准差σ的一个置信度为$1-\alpha$的置信区间

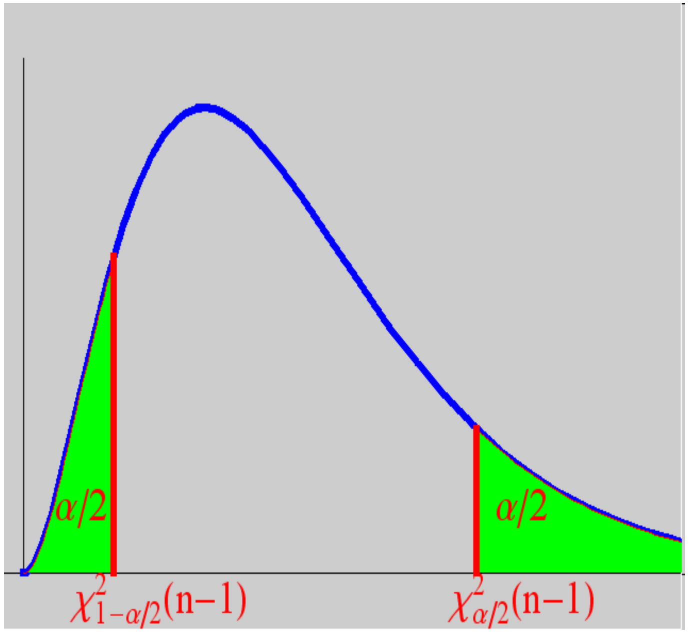
$$\left(\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{\alpha/2}(n-1)}},\quad \frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}}\right).$$

注意: 在密度函数不对称时,

如 χ^2 分布和 F 分布，

习惯上仍取对称的分位点来

确定置信区间(如图).



例6.4 (续例6. 1) 求例 1中总体方差 σ^2 和标准差 σ 的置信度为 0.95 的置信区间 .

解 $\dfrac{\alpha}{2}=0.025,\qquad 1-\dfrac{\alpha}{2}=0.975,\qquad n-1=11,$

查 $\chi^2(n-1)$ 分布表可知 :

$$\chi^2_{0.025}(11)=21.920,\qquad \chi^2_{0.975}(11)=3.816,$$

方差 σ^2 的置信区间 $(78.97,\ 453.64)$;

标准差 σ 的置信区间 $(8.87,\ 21.30)$.

表3.1单个正态总体参数的区间估计一览表

被估计参数	条件	估计统计量	服从分布	置信度为1-α的置信区间	α的分位数
μ	σ已知	$U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	N(0,1)	$(\overline{X}\pm\dfrac{\sigma}{\sqrt{n}}u_{\alpha/2})$	$\Phi(u_{\alpha/2})=1-\alpha/2$
μ	σ未知	$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	t(n-1)	$(\overline{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1))$	$P\{T>t_{\alpha/2}(n-1)=\alpha/2$
σ^2	μ已知	$\chi^2=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$	$\chi^2(n)$	$(\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)},\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)})$	$P\{\chi^2>\chi^2_{\alpha/2}(n)\}=\alpha/2$ $P\{\chi^2>\chi^2_{1-\alpha/2}(n)\}=\alpha/2$
σ^2	μ未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$	$(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)})$	$P\{\chi^2>\chi^2_{\alpha/2}(n-1)\}=\alpha/2$ $P\{\chi^2>\chi^2_{1-\alpha/2}(n-1)\}=\alpha/2$
σ	μ未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$	$(\sqrt{\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}},\sqrt{\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}})$	$P\{\chi^2>\chi^2_{\alpha/2}(n-1)\}=\alpha/2$ $P\{\chi^2>\chi^2_{1-\alpha/2}(n-1)\}=\alpha/2$

例3.3 从一台机床加工的轴承中，随机地抽取 200 件，测得其椭圆度，得样本观察值 $\overline{X}=0.081$ 毫米，并由累积资料知椭圆度 X 服从 $N(\mu,0.025^2)$，试在置信度 0.95 下，求 μ 的置信区间的相应于样本观察值的一个现实区间。

解：由一览表中查得，σ^2 已知条件下，μ 置信度为 $1-\alpha=0.95$ 的置信区间为

$$(\overline{X}\pm\frac{\sigma}{\sqrt{n}}z_{\alpha/2})$$

其中 $z_{\alpha/2}=z_{\frac{0.05}{2}}=1.96$，且已知 $\sigma=0.025,n=200,\overline{x}=0.081$

故所求现实区间为

$$(0.081\pm\frac{0.025}{\sqrt{200}}\times1.96)=(0.081\pm0.035)=(0.0775,0.0845)$$

例3.4 已知某种白帜灯泡的寿命 $X \sim N(\mu,\sigma^2)$, 在一批该种灯泡中随机地抽取 10 只, 测得其寿命 (以小时计算) 于下：

1067 919 1196 785 1126 936 918 1156 920 948

试求未知参数 μ,σ^2 及 σ 的知心度为 0.95的置信区间。

解：(1)查一览表中查得 $,\sigma^2$ 未知条件下，μ 的置信度为 $1-\alpha=0.95$ 的置信区间为：

$$\left(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$$

查 t 分布表得 $t_{\alpha/2}(n-1)=t_{0.05/2}(10-1)=2.2622$

由数据计算得 $\overline{x}=997.1, s=131.5476$

故得 μ 的置信度为 0.95的置信区间为

$$\left(997.1 \pm \frac{131.5476}{\sqrt{10}} \times 2.2622\right)=(997.1 \pm 94.1)=(903, 1091.2)$$

(2)由一览表查得：μ未知条件下，σ^2的置信度为 $1-\alpha=0.95$的置信区间为

$$\left(\frac{(n-1)S^2}{x^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{x^2_{1-\alpha/2}(n-1)}\right)$$

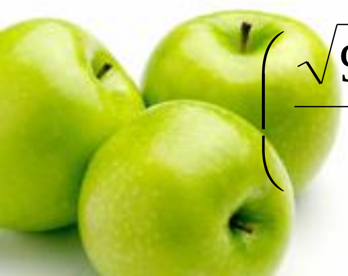
查x^2分布表得 $x^2_{\alpha/2}(n-1)=x^2_{0.05/2}(9)=2.7$

故得 σ^2置信区间为 0.95的置信区间为

$$\left(\frac{9\times 131.5476^2}{19.023},\frac{9\times 131.5476^2}{2.7}\right)=(8187.1,57682.6)$$

σ的置信度为 0.95的置信区间为

$$\left(\frac{\sqrt{9}\times 131.5476}{\sqrt{19.023}},\frac{\sqrt{9}\times 131.5476}{\sqrt{2.7}}\right)=(90.5,240.2)$$



设单个总体 $X \sim N(\mu,\sigma^2)$,

1. 单个总体均值 μ 的置信区间

$$\begin{cases} (1)\ \sigma^2 \text{为已知}, & \left(\overline{X} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right). \\ (2)\ \sigma^2 \text{为未知}, & \left(\overline{X} \pm \dfrac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right). \end{cases}$$

2. 单个总体方差 σ^2 的置信区间

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

二、两个总体$N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$的情况

设给定置信度为 $1-\alpha$，并设 $X_1, X_2, \cdots, X_n$ 为第一个总体 $N(\mu_1,\sigma_1^2)$的样本，$Y_1, Y_2, \cdots, Y_n$ 为第二个总体 $N(\mu_2,\sigma_2^2)$的样本，$\overline{X}, \overline{Y}$分别是第一、二个总体的样本均值，$S_1^2, S_2^2$分别是第一、二个总体的样本方差.

讨论两个整体总体均值差和方差比的估计问题.

1. 两个总体均值差$\mu_1-\mu_2$的置信区间

(1) σ_1^2和σ_2^2均为已知

$\mu_1-\mu_2$的一个置信度为$1-\alpha$的置信区间

$$\left(\overline{X}-\overline{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right).$$

推导过程如下:

因为 $\overline{X},\overline{Y}$ 分别是 μ_1,μ_2 的无偏估计，

所以 $\overline{X}-\overline{Y}$ 是 $\mu_1-\mu_2$ 的无偏估计，

由 $\overline{X}, \overline{Y}$ 的独立性及

$$\overline{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \qquad \overline{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right),$$

可知

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right),$$

或

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1),$$

于是得$\mu_1-\mu_2$的一个置信度为$1-\alpha$的置信区间

$$\left(\overline{X}-\overline{Y}\pm z_{\alpha/2}\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}}\right).$$

(2) $\sigma_1^{\ 2}$和$\sigma_2^{\ 2}$均为未知 ，

只要 n_1和n_2都很大 (实用上 >50即可), 则有

$\mu_1-\mu_2$的一个置信度为$1-\alpha$的近似置信区间

$$\left(\overline{X}-\overline{Y}\pm z_{\alpha/2}\sqrt{\frac{S_1^{\ 2}}{n_1}+\frac{S_2^{\ 2}}{n_2}}\right).$$

(3) $\sigma_1^2 = \sigma_2^2 = \sigma^2$, 但 σ^2 为未知，

$\mu_1 - \mu_2$的一个置信度为$1-\alpha$的置信区间

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

其中 $S_w^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, $\quad S_w = \sqrt{S_w^2}$.

例7 为比较Ⅰ, Ⅱ两种型号步枪子弹的枪口速度, 随机地取Ⅰ型子弹10发, 得到枪口速度的平均值为 $\bar{x}_1 = 500\,(\mathrm{m/s})$,标准差 $s_1 = 1.10\,(\mathrm{m/s})$, 随机地取Ⅱ型子弹20发, 得枪口速度平均值为 $\bar{x}_2 = 496\,(\mathrm{m/s})$, 标准差 $s_2 = 1.20\,(\mathrm{m/s})$, 假设两总体都可认为近似地服从正态分布,且由生产过程可认为它们的方差相等, 求两总体均值差 $\mu_1-\mu_2$的置信度为0.95的置信区间.

解 由题意, 两总体样本独立且方差相等(但未知),

$\frac{\alpha}{2} = 0.025,\quad n_1 = 10,\quad n_2 = 20,\quad n_1 + n_2 - 2 = 28,$

查 $t(n-1)$ 分布表可知：$t_{0.025}(28) = 2.0484$，

$$s_w^{\ 2} = \frac{9\times 1.10^2 + 19\times 1.20^2}{28},\quad s_w = \sqrt{S_w^{\ 2}} = 1.1688,$$

于是得$\mu_1-\mu_2$的一个置信度为0.95的置信区间

$$\left(\overline{x}_1 - \overline{x}_2 \pm S_w \times t_{0.025}(28)\sqrt{\frac{1}{10}+\frac{1}{20}}\right) = (4 \pm 0.93),$$

即所求置信区间为 $(3.07,\ 4.93)$.

例8 为提高某一化学生产过程的得率, 试图采用一种新的催化剂, 为慎重起见, 在试验工厂先进行试验. 设采用原来的催化剂进行了$n_1 = 8$ 次试验, 得到得率的平均值$\bar{x}_1 = 91.73$. 样本方差 $s_1^2 = 3.89$, 又采用新的催化剂进行了$n_2 = 8$ 次试验, 得到得率的平均值$\bar{x}_2 = 93.75$, 样本方差 $s_2^2 = 4.02$,假设两总体都可认为近似地服从正态分布, 且方差相等, 求两总体均值差 $\mu_1 - \mu_2$的置信水平为0.95的置信区间.

解 由题意, 两总体样本独立且方差相等(但未知),

且 $s_w{}^2 = \dfrac{(n_1-1)S_1{}^2 + (n_2-1)S_2{}^2}{n_1+n_2-2} = 3.96,$

于是得$\mu_1 - \mu_2$的一个置信水平为0.95的置信区间

$$\left(\bar{x}_1 - \bar{x}_2 \pm s_w \times t_{0.025}(14)\sqrt{\frac{1}{8}+\frac{1}{8}}\right) = (-2.02 \pm 2.13),$$

即所求置信区间为 $(-4.15，0.11).$

2.两个总体方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的置信区间

仅讨论总体均值 μ_1, μ_2 为未知的情况.

$\dfrac{\sigma_1^2}{\sigma_2^2}$ 的一个置信度为 $1-\alpha$ 的置信区间

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right).$$

推导过程如下:

由于 $\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$, $\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$,

且由假设知 $\dfrac{(n_1-1)S_1^2}{\sigma_1^2}$ 与 $\dfrac{(n_2-1)S_2^2}{\sigma_2^2}$ 相互独立，

根据*F*分布的定义, 知 $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1),$

即 $$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)} \sim F(n_1-1, n_2-1),$$

$$P\left\{F_{1-\alpha/2}(n_1-1,n_2-1)<\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}<F_{\alpha/2}(n_1-1,n_2-1)\right\}$$

$$=1-\alpha,$$

$$P\left\{\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right\}$$

$$=1-\alpha,$$

于是得 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的一个置信度为 $1-\alpha$ 的置信区间

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\ \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right).$$

例9 研究由机器 *A* 和机器 *B* 生产的钢管内径, 随机抽取机器 *A* 生产的管子 18 只, 测得样本方差为 $s_1^2 = 0.34(\mathrm{mm}^2)$; 抽取机器*B*生产的管子 13 只, 测得样本方差为 $s_2^2 = 0.29(\mathrm{mm}^2)$. 设两样本相互独立,且设由机器 *A* 和机器 *B* 生产的钢管内径分别服从正态分布 $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$, $\mu_i,\sigma_i^2(i=1,2)$ 均未知, 求方差比 σ_1^2/σ_2^2 的置信度为0.90的置信区间.

解 $n_1 = 18$, $n_2 = 13$, $\alpha = 0.10$,

$$s_1^2 = 0.34(\mathrm{mm}^2), \qquad s_2^2 = 0.29(\mathrm{mm}^2),$$

$$F_{\alpha/2}(n_1-1, n_2-1) = F_{0.05}(17,12) = 2.59,$$

$$F_{1-\alpha/2}(17,12) = F_{0.95}(17,12) = \frac{1}{F_{0.05}(12,17)} = \frac{1}{2.38},$$

于是得 $\dfrac{\sigma_1^{\ 2}}{\sigma_2^{\ 2}}$ 的一个置信度为0.90的置信区间

$$\left(\frac{0.34}{0.29}\times\frac{1}{2.59},\ \frac{0.34}{0.29}\times 2.38\right) = (0.45,\ 2.79).$$

例10 甲、乙两台机床加工同一种零件,在机床甲加工的零件中抽取9个样品,在机床乙加工的零件中抽取6个样品,并分别测得它们的长度(单位:mm),由所给数据算得 $s_1^2=0.245$, $s_2^2=0.357$,在置信度0.98下,试求这两台机床加工精度之比 σ_1/σ_2 的置信区间. 假定测量值都服从正态分布,方差分别为 σ_1^2, σ_2^2.

解 $n_1=9,\quad n_2=6,\quad \alpha=0.02,$

$$F_{1-\alpha/2}(n_1-1, n_2-1)=F_{0.99}(8,5)=10.3,$$

$$F_{\alpha/2}(8,5)=F_{0.01}(8,5)=\frac{1}{F_{0.99}(5,8)}=\frac{1}{6.63},$$

于是得 $\dfrac{\sigma_1}{\sigma_2}$ 的一个置信度为0.98的置信区间

$$\left(\sqrt{\frac{S_1^{\ 2}}{S_2^{\ 2}}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)}},\ \sqrt{\frac{S_1^{\ 2}}{S_2^{\ 2}}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}}\right)$$

$$=\left(\sqrt{\frac{0.245}{0.357\times 10.3}},\ \sqrt{\frac{0.245\times 6.63}{0.357}}\right)=(0.258\ ,\ \ 2.133\ ).$$

第六节 (0－1)分布参数的区间估计

一、置信区间公式

二、典型例题

在样本容量充分大时，可以用渐近分布来构造近似的置信区间。一个典型的例子是关于比例p 的置信区间。

一、置信区间公式

设有一容量 $n > 50$ 的大样本，它来自 $(0-1)$ 分布的总体 X，X 的分布律为 $f(x;p) = p^x(1-p)^{1-x}$，$x = 0, 1$，其中 p 为未知参数，则 p 的置信度为 $1-\alpha$ 的置信区间是

$$\left(\frac{-b-\sqrt{b^2-4ac}}{2a}, \frac{-b+\sqrt{b^2-4ac}}{2a}\right),$$

其中 $a = n + z_{\alpha/2}^2$，$b = -(2n\overline{X} + z_{\alpha/2}^2)$，$c = n\overline{X}^2$.

推导过程如下:

因为(0–1)分布的均值和方差分别为

$$\mu = p, \quad \sigma^2 = p(1-p),$$

设 $X_1, X_2, \cdots, X_n$ 是一个样本 ，因为容量n较大,

由**中心极限定理**知 $\dfrac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} = \dfrac{n\overline{X} - np}{\sqrt{np(1-p)}}$

近似地服从 $N(0,1)$ 分布，

$$P\left\{-z_{\alpha/2} < \frac{n\overline{X} - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right\} \approx 1-\alpha,$$

不等式 $-z_{\alpha/2} < \dfrac{n\overline{X} - np}{\sqrt{np(1-p)}} < z_{\alpha/2}$

等价于 $(n+z_{\alpha/2}^2)p^2-(2n\overline{X}+z_{\alpha/2}^2)p+n\overline{X}^2<0,$

令 $p_1=\dfrac{-b-\sqrt{b^2-4ac}}{2a}$， $p_2=\dfrac{-b+\sqrt{b^2-4ac}}{2a}$，

其中 $a=n+z_{\alpha/2}^2$， $b=-(2n\overline{X}+z_{\alpha/2}^2)$， $c=n\overline{X}^2.$

则 p 的近似置信水平为 $1-\alpha$ 的置信区间是

$(p_1，p_2).$

二、典型例题

例1 设从一大批产品的100个样品中, 得一级品60个, 求这批产品的一级品率 p 的置信水平为0.95的置信区间.

解 一级品率 p 是(0-1)分布的参数,

$$n = 100, \quad \bar{x} = \frac{60}{100} = 0.6,$$

$$1 - \alpha = 0.95, \quad z_{\alpha/2} = z_{0.025} = 1.96,$$

则 $a = n + z_{\alpha/2}^2 = 103.84,$

$$b = -(2n\overline{X} + z_{\alpha/2}^2) = -(2n\overline{x} + z_{\alpha/2}^2) = -123.84,$$

$$c = n\overline{X}^2 = n\overline{x}^2 = 36,$$

于是 $$p_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} = 0.50,$$

$$p_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} = 0.69,$$

p 的置信水平为0.95的置信区间为 $(0.50, 0.69)$.

例2 设从一大批产品的120个样品中，得次品9个，求这批产品的次品率 p 的置信水平为0.90的置信区间.

解 $n = 120$， $\bar{x} = \dfrac{9}{100} = 0.09$， $1 - \alpha = 0.90$，

则 $a = n + z_{\alpha/2}^{2} = 122.71$，

$$b = -(2n\overline{X} + z_{\alpha/2}^{2}) = -(2n\bar{x} + z_{\alpha/2}^{2}) = -24.31,$$

$$c = -n\overline{X}^{2} = -n\bar{x}^{2} = 0.972,$$

于是 $$p_1 = \frac{-b-\sqrt{b^2-4ac}}{2a} = 0.056,$$

$$p_2 = \frac{-b+\sqrt{b^2-4ac}}{2a} = 0.143,$$

p 的置信水平为0.90的置信区间为 $(0.056, 0.143)$.

通常，未知参数p的置信度为$1-\alpha$的置信区间近似为

$$\overline{X} \pm z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})/n}$$

第七节　单侧置信区间

一、问题的引入

二、基本概念

三、典型例题

四、小结

一、问题的引入

在以上各节的讨论中，对于未知参数 θ，我们给出两个统计量 $\underline{\theta}, \overline{\theta}$，得到 θ 的双侧置信区间 $(\underline{\theta}, \overline{\theta})$.

但在某些实际问题中，例如，对于设备、元件的寿命来说，平均寿命长是我们希望的，我们关心的是平均寿命 θ 的“下限”；与之相反，在考虑产品的废品率 p 时，我们常关心参数 p 的“上限”，这就引出了单侧置信区间的概念.

二、基本概念

1. 单侧置信区间的定义

对于给定值 $\alpha\ (0<\alpha<1)$，若由样本 $X_1, X_2, \cdots, X_n$ 确定的统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n)$，对于任意 $\theta \in \Theta$ 满足

$$P\{\theta > \underline{\theta}\} \geq 1-\alpha,$$

则称随机区间 $(\underline{\theta}, +\infty)$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间，$\underline{\theta}$ 称为 θ 的置信水平为 $1-\alpha$ 的单侧置信下限．

又如果统计量 $\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$, 对于任意 $\theta\in\Theta$满足

$$P\{\theta<\overline{\theta}\}\geq 1-\alpha,$$

则称随机区间 $(-\infty,\overline{\theta})$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间, $\overline{\theta}$ 称为 θ 的置信水平为 $1-\alpha$的单侧置信上限.

2. 正态总体均值与方差的单侧置信区间

设正态总体 X 的均值是 μ, 方差是 σ^2 (均为未知),

$X_1, X_2, \cdots, X_n$ 是一个样本, 由 $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$,

有 $P\left\{\dfrac{\overline{X}-\mu}{S/\sqrt{n}} < t_\alpha(n-1)\right\} = 1-\alpha$,

即 $P\left\{\mu > \overline{X} - \dfrac{S}{\sqrt{n}} t_\alpha(n-1)\right\} = 1-\alpha$,

于是得 μ 的一个置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha}(n-1),+\infty\right),$$

μ 的置信水平为 $1-\alpha$ 的置信下限 $\underline{\mu}=\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha}(n-1).$

又根据 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$

有 $P\left\{\dfrac{(n-1)S^2}{\sigma^2}>\chi^2_{1-\alpha}(n-1)\right\}=1-\alpha,$

即 $P\left\{\sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right\} = 1-\alpha,$

于是得 σ^2 的一个置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(0,\ \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right),$$

σ^2 的置信水平为 $1-\alpha$ 的单侧置信上限

$$\overline{\sigma^2} = \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}.$$

三、典型例题

例1 设从一批灯泡中, 随机地取5只作寿命试验, 测得寿命(以小时计)为 1050, 1100, 1120, 1250, 1280, 设灯泡寿命服从正态分布, 求灯泡寿命平均值的置信水平为 0.95 的单侧置信下限.

解 $1-\alpha = 0.95$, $n = 5$, $\bar{x} = 1160$, $s^2 = 9950$,

$$t_\alpha(n-1) = t_{0.05}(4) = 2.1318,$$

μ 的置信水平为 0.95 的置信下限

$$\underline{\mu} = \bar{x} - \frac{s}{\sqrt{n}} t_\alpha(n-1) = 1065.$$

例2 设总体 X 在$[0,\theta]$上服从均匀分布，其中 θ $(\theta > 0)$ 未知，$(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，给定 α，求 θ 的置信水平为 $1-\alpha$ 的置信下限和置信上限.

解 令 $X_h = \max\{X_1, X_2, \cdots, X_n\}$,

对于给定的 α，找 $0 < \underline{\theta} \le 1$，使 $P\left\{\theta > \dfrac{X_h}{\underline{\theta}}\right\} = 1-\alpha$,

即 $1-\alpha = \int_0^{\underline{\theta}} nz^{n-1}\mathrm{d}z = \underline{\theta}^n$，于是 $\underline{\theta} = \sqrt[n]{1-\alpha}$,

所以 $P\left\{\dfrac{X_h}{\sqrt[n]{1-\alpha}} < \theta\right\} = 1-\alpha$,

θ 的置信水平为 $1-\alpha$ 的置信下限 $\underline{\theta} = \dfrac{X_h}{\sqrt[n]{1-\alpha}}$.

对于给定的 α, 找 $0 < \overline{\theta} < 1$, 使 $P\left\{\theta < \dfrac{X_h}{\overline{\theta}}\right\} = 1-\alpha$,

即 $1-\alpha = \int_{\overline{\theta}}^{1} nz^{n-1}\mathrm{d}z = 1-\overline{\theta}^{\,n}$, 于是 $\overline{\theta} = \sqrt[n]{\alpha}$,

所以 $P\left\{\theta < \dfrac{X_h}{\sqrt[n]{\alpha}}\right\} = 1-\alpha$,

θ 的置信水平为 $1-\alpha$ 的置信上限 $\overline{\theta} = \dfrac{X_h}{\sqrt[n]{\alpha}}$.

四、小结

正态总体均值 μ 的置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(-\infty, \boxed{\bar{X}+\frac{S}{\sqrt{n}}t_\alpha(n-1)}\right), \quad \left(\boxed{\bar{X}-\frac{S}{\sqrt{n}}t_\alpha(n-1)}, +\infty\right),$$

单侧置信上限 $\bar{\mu}$　　单侧置信下限 $\underline{\mu}$

正态总体方差 σ^2 的置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(0, \boxed{\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}}\right).$$

单侧置信上限 $\overline{\sigma^2}$

第四节　样本容量的确定

㈠确定样本容量的意义

㈡估计均值时的样本容量

㈢估计比率时的样本容量

一、确定样本容量的意义

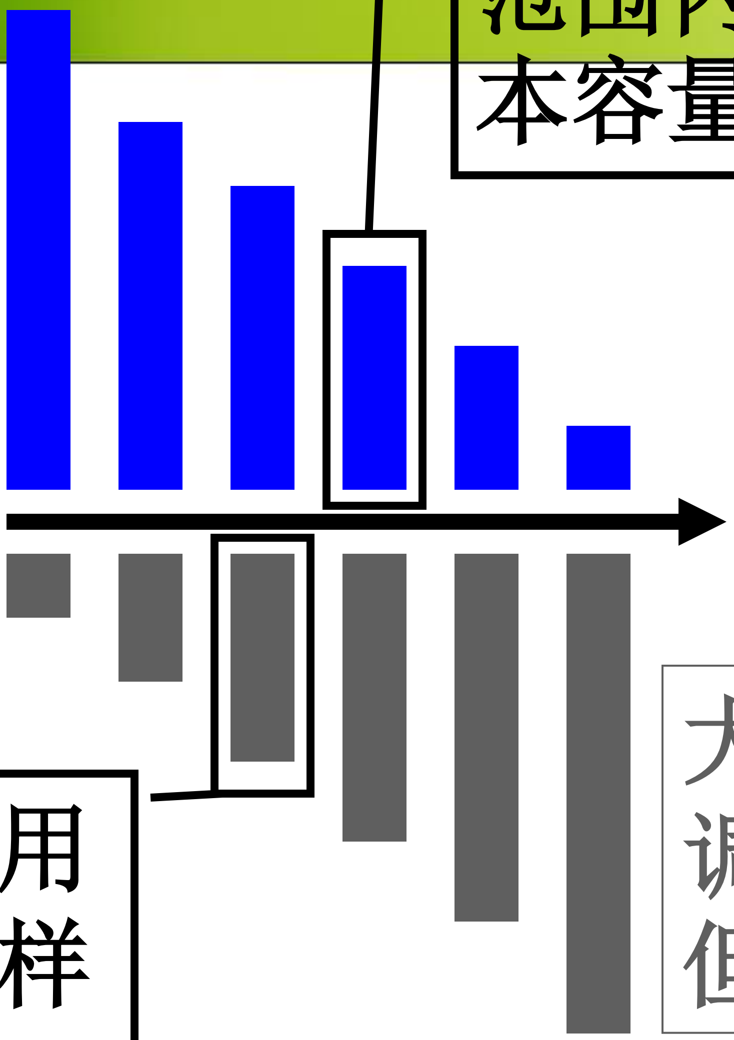
小样本容量
节省费用但
调查误差大

找出在规定误差范围内的最小样本容量

调查误差

样本容量

调查费用



找出在限定费用范围内的最大样本容量

大样本容量
调查精度高
但费用较大

二、估计均值时的样本容量

1、已知σ ²时，μ 的置信区间

$$[\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}]$$

区间长度：$L=2Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

在给定置信度1-α和区间长度可容许的上限值l都已经给定的前提下，要使$L\le l$，等价于

$$L=2Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le l\rightarrow n\ge(\frac{2\sigma}{l}Z_{\alpha/2})^2$$

例　在交通工程中需要测定车速（单位 km/h),由以往的经验知道，测量值为X， $X \sim N(\mu,\sigma^2)$ $\sigma^2 = 3.58^2$

1、至少作多少次观测，才能以0.99的可靠性保证平均测量值的误差在 ± 1 之间。

2、现在作了150次观测，试问平均测量值的误差在 ± 1 之间的概率有多大?

解　由题意知 $X \sim N(\mu, 3.58^2)$ 用平均测量值 $\overline{X}$ 来估计 μ

其误差 $\left|\overline{X} - \mu\right|$ 由题意要求 $P\{\left|\overline{X} - \mu\right| < 1\} \geq 0.99$ 　$\alpha = 0.01$

由置信区间的概念，所求 μ 的0.99的 置信区间为

$[\overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{0.025}]$ 　即　$P\{\left|\overline{X} - \mu\right| < \frac{\sigma}{\sqrt{n}} Z_{0.005}\} \geq 0.99$

$$[\overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{0.025}]$$ 即 $$P\{|\overline{X} - \mu| < \frac{\sigma}{\sqrt{n}} Z_{0.005}\} \geq 0.99$$

令 $$\frac{\sigma}{\sqrt{n}} Z_{0.005} = 1 \quad \Rightarrow \quad n = 3.58^2 \cdot (2.576)^2 = 86.047$$

则钢索所能承受的平均张力为 6650.9 kg/cm²

$$\therefore \quad n = 86 \qquad z_{0.005} = 2.57$$

至少要作86次观测，才能以0. 99的可靠性保持平均测量误差在 ± 1 之间。

2\. $n = 150 \qquad \sigma = 3.58$

$$P\{|\overline{X} - \mu| < 1\} = P\{\left|\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}\right| < \frac{\sqrt{n}}{\sigma}\} = 2\Phi(\frac{\sqrt{n}}{\sigma}) - 1$$

$$= 2\Phi(3.421) - 1 = 2 \cdot 0.9997 - 1 = 0.9994$$

2、未知 σ^2时，μ的置信区间

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\ \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right]$$

区间长度：$L=2t_{\alpha/2}(n-1)\dfrac{s}{\sqrt{n}}$

在给定置信度1-α和区间长度可容许的上限值l都已经给定的前提下，要使$L\le l$，等价于

$$L=2t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\le l$$

n的下界依赖于$s=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$和$t_{\alpha/2}(n-1)$

------------Fisher问题

三、估计比率时的样本容量

1、设总体X服从（0-1）分布

$$P(X=k)=\theta^{k}(1-\theta)^{1-k} \qquad k=0,1$$

通常，未知参数θ的置信度为$1-\alpha$的置信区间近似为

$$\overline{X} \pm z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})/n}$$

在给定置信度1-α和区间长度可容许的上限值l都已经给定的前提下，要使$L \le l$，等价于

$$L=2Z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \le l \qquad \text{因此}, n \ge 4\overline{x}(1-\overline{x})\left(\frac{Z_{\alpha/2}}{l}\right)^{2}$$