概率论与数理统计

范正平
fanzhp@mail.sysu.edu.cn

中山大学 数据科学与计算机学院

Chapter 7
参数估计

参数估计

参数估计问题是利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数.

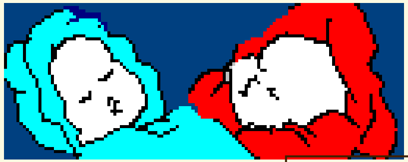
例如：在某炸药制造厂，一天中发生着火现象的次数 X 是一个随机变量，假设它服从以 $\lambda > 0$ 为参数的泊松分布，参数 λ 为未知，设有以下的样本值，试估计参数 λ.

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

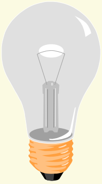
在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数.

参数估计(Cont.)

估计新生儿的体重

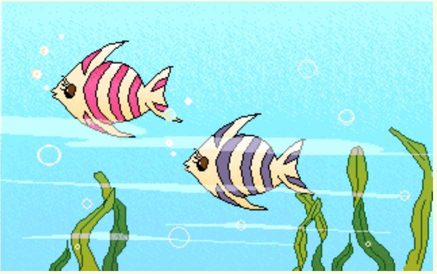


估计废品率

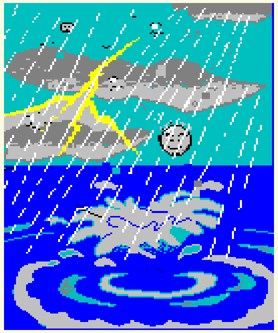


均事先假定总体分布形式已知，只有分布中的参数未知.

估计湖中鱼数



估计降雨量



…

…

参数估计问题的定义

设有一个统计总体，总体的分布函数为

F(x,θ)，其中θ为未知参数 (θ 可以是向量) .

现从该总体抽样，得样本

$$X_1, X_2, \ldots, X_n$$

要依据该样本对参数 θ 作出估计, 或估计 θ

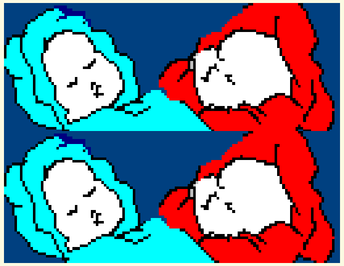
的某个已知函数 $g(\theta)$.

这类问题称为参数估计.

参数估计

例 已知某地区新生婴儿的体重 $X \sim N(\mu,\sigma^2)$,

$(\mu,\sigma$ 未知$)$



…

随机抽查100个婴儿，得100个体重数据

10,7,6,6.5,5,5.2, …

而全部信息就由这100个数组成.

据此,我们应如何估计 μ 和 σ 呢？

参数估计(Cont.)

为估计 μ:

我们需要构造出适当的样本的函数 $T(X_1,X_2,\ldots X_n)$ ，每当有了样本，就代入该函数中算出一个值，用来作为 μ 的估计值 .

$T(X_1,X_2,\ldots X_n)$ 称为参数 μ 的估计量,

把样本值代入$T(X_1,X_2,\ldots X_n)$ 中，得到 μ 的一个估计值 .

问题是:

使用什么样的统计量去估计 μ？

可以用样本均值;

也可以用样本矩;

还可以用别的统计量 .

我们知道, 若 $X \sim N(\mu,\sigma^2)$, 则 $E(X)=\mu$.

由大数定律,

样本体重的平均值

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$$

自然想到把样本体重的平均值作为总体平均体重的一个估计.

用样本体重的均值 $\overline{X}$ 估计 μ.

类似地，用样本体重的方差 S^2 估计 σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

寻求参数估计量的方法

1. 极大似然法

2. 矩估计法

……

最大似然估计法

似然函数的定义

1) 设总体 X 属离散型

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 $\prod_{i=1}^{n}p(x_i;\theta)$.

最大似然估计法(Cont.)

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

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

$L(\theta)$称为样本的似然函数.

最大似然估计法(Cont.)

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

(其中Θ是θ可能的取值范围)

这样得到的$\hat{\theta}$与样本值$x_1, x_2, \cdots, x_n$有关,记为

$\boxed{\hat{\theta}(x_1, x_2, \cdots, x_n)}$, 参数$\theta$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

最大似然估计法(Cont.)

似然函数的定义

(2) 设总体 X 属连续型

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

最大似然估计法(Cont.)

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

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

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

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

求最大似然估计量的步骤(Cont.):

(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 θ_i $(i=1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

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

$$X_i \sim \begin{Bmatrix} 0 & 1 \\ 1-p & p \end{Bmatrix}$$

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

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

例 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

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

$$= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2]$$

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

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

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

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

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

例 设总体 X 在 $[0,\theta]$ 上服从均匀分布,其中 θ 未知, $x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值,求 θ 的最大似然估计值.

解 X 的概率密度为 $f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta,\\ 0, & 其他.\end{cases}$

似然函数为

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0<x_1,x_2\cdots,x_n\le\theta,\\ 0, & 其他\end{cases}$$

记 $x_{(n)}=\max(x_1,x_2,\cdots,x_n)$,

由于 $x_1,x_2,\cdots,x_n\le\theta$ 等价于 $x_{(h)}\le\theta$,

用求导方法无法最终确定

只能用最大似然原则来求

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^{n}}, & \theta \ge x_{(n)},\\ 0, & \text{其他}.\end{cases}$$

即知当$\theta < x_{(n)}$时$L(\theta)=0$;

而当$\theta \ge x_{(n)}$时$L(\theta)$随θ的增大而减少.

故$L(\theta)$在$x_{(n)}$取得最大值，即得θ的最大似然估计值为

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

矩估计法

1. 设 X 是随机变量,若$E(X^k)$, $k=1,2,\cdots$ 存在,称它为 X 的 k **阶原点矩**,简称 k **阶矩**.

2. k 阶样本（原点)矩

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

矩估计法的核心思想

由辛钦大数定理定理，

若总体 X 的数学期望 $E(X)=\mu$ 存在, 则有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} E(X)=\mu$$

$$\Downarrow$$

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X^k)=\mu_k \ (k=1,2,\cdots)$$

$$g(A_1,A_2,\cdots,A_k) \xrightarrow{P} g(\mu_1,\mu_2,\cdots,\mu_k)$$

其中 g 为连续函数.

矩估计法的核心思想(Cont.)

设 X 为连续型随机变量,其概率密度为

$f(x;\theta_1,\theta_2,\cdots,\theta_k)$, 或 X 为离散型随机变量,

其分布律为 $P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_k)$,

其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数,

若 $X_1,X_2,\cdots,X_n$ 为来自 X 的样本,

假设总体 X 的前 k 阶矩存在,

且均为 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数, 即

矩估计法的核心思想(Cont.)

$$\mu_l = E(X^l) = \int_{-\infty}^{+\infty} x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)\mathrm{d}x$$ **(X为连续型)**

或 $\mu_l = E(X^l) = \sum_{x\in R_X} x^l p(x;\theta_1,\theta_2,\cdots,\theta_k)$, **($X$为离散型)**

因为样本矩 $A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l$ 依概率收敛于相应的总体矩 μ_l $(l=1, 2,\cdots, k)$,

样本矩的连续函数依概率收敛于相应的总体矩的连续函数.

矩估计法的具体做法

矩估计法的具体做法：令 $\mu_l = A_l, \quad l = 1, 2, \cdots, k.$

这是一个包含 k 个未知参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的方程组,

解出其中 $\theta_1, \theta_2, \cdots, \theta_k$.

用方程组的解 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ 分别作为 $\theta_1, \theta_2, \cdots, \theta_k$ 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的概率密度为

$$f(x;\mu,\theta)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x>\mu,\\ 0, & \text{其他},\end{cases}$$

其中 $\mu,\theta(\theta>0)$ 为待估参数，设 $X_1,X_2,\cdots X_n$ 是来自X的一个样本，求 μ,θ 的矩估计量.

解　总体X 的一阶、二阶矩分别为

$$\mu_1 = E(X) = \int_{\mu}^{\infty} x \frac{1}{\theta} e^{-(x-\mu)/\theta} dx = \mu + \theta,$$

$$\mu_2 = E(X^2) = \int_{\mu}^{\infty} x^2 \frac{1}{\theta} e^{-(x-\mu)/\theta} dx = \mu^2 + 2\theta(\mu + \theta).$$

分别以一阶、二阶样本矩 A_1，A_2

代替上两式中的 μ_1, μ_2， 有

$$\begin{cases} A_1 = \mu + \theta, \\ A_2 = \mu^2 + 2\theta(\mu + \theta). \end{cases}$$

$$A_1=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$$

从中解得 θ,μ, 即得到 θ,μ 的矩估计量为

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

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

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 $\dfrac{\hat{\theta}}{2}=A_1=\overline{X}$,

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

例 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 $$\frac{a+b}{2} = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\frac{(a-b)^2}{12} + \frac{(a+b)^2}{4} = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

即 $\begin{cases} a+b=2A_1, \\ b-a=\sqrt{12(A_2-A_1^2)}. \end{cases}$

解方程组得到*a*, *b*的矩估计量分别为

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

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

例　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　$\mu_1 = E(X) = \mu,$

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

令 $\begin{cases} \mu = A_1, \\ \sigma^2 + \mu^2 = A_2. \end{cases}$

解方程组得到矩估计量分别为 $\hat{\mu} = A_1 = \overline{X},$

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

矩估计法的优缺点

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．

其主要原因在于建立矩法方程时，选取那些总体矩用相应样本矩代替带有一定的随意性．

估计量的评选标准

问题的提出

对于同一个参数，用不同的估计方法求出的估计量可能不相同，

问题

(1)对于同一个参数究竟采用哪一个估计量好?

(2)评价估计量的标准是什么?

下面介绍几个常用标准.

常用的几条标准是：

1. 无偏性
2. 有效性
3. 相合性

无偏性

估计量是随机变量，对于不同的样本值会得到不同的估计值．我们希望估计值在未知参数真值附近摆动，而它的期望值等于未知参数的真值. 这就导致无偏性这个标准．

设 $\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数 θ 的估计量，若

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

则称 $\hat{\theta}$ 为 θ 的无偏估计．

例 设总体X的k阶矩$\mu_k = E(X^k)$存在

$(X_1, X_2, \cdots, X_n)$是总体X的样本，

证明：不论X服从什么分布，$A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$

是μ_k的无偏估计量.

证 因为$X_1, X_2, \cdots, X_n$与X同分布，

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

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

$$= \frac{1}{n} \cdot n \cdot \mu_k = \mu_k$$

例 对于均值μ, 方差$\sigma^2>0$的总体, 若

μ, σ^2 均为未知, 则σ^2的估计量$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

是有偏的(即不是无偏估计).

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

因为 $E(A_2)=\mu_2=D(X)+[E(X)]^2=\sigma^2+\mu^2,$

又因为 $E(\overline{X}^2)=D(\overline{X})+[E(\overline{X})]^2=\frac{\sigma^2}{n}+\mu^2,$

所以 $E(\hat{\sigma}^2)=E(A_2-\overline{X}^2)=E(A_2)-E(\overline{X}^2)$

$$=\frac{n-1}{n}\sigma^2 \neq \sigma^2,$$ 所以 $\hat{\sigma}^2$ 是有偏的.

若以 $\dfrac{n}{n-1}$ 乘 $\hat{\sigma}^2$, 所得到的估计量就是无偏的.

(这种方法称为无偏化).

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

因为 $\dfrac{n}{n-1}\hat{\sigma}^2=S^2 \quad =\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,$

即 S^2 是 σ^2 的无偏估计，故通常取 S^2 作 σ^2 的估计量.

例　设总体 $X \sim N(\mu,\sigma^2)$,

$(X_1, X_2, \cdots, X_n)$ 为 X 的一个样本

求常数 k, 使 $k\sum_{i=1}^{n}|X_i - \overline{X}|$ 为 σ 的无偏估计量

解 $E\left(k\sum_{i=1}^{n}|X_i - \overline{X}|\right) = k\left(\sum_{i=1}^{n}E|X_i - \overline{X}|\right)$

注意到 $X_i - \overline{X}$ 是 $X_1, X_2, \ldots, X_n$ 的线性函数,

$$X_i - \overline{X} = \frac{1}{n}\left(-X_1 - X_2 \cdots + (n-1)X_i - \cdots - X_n\right)$$

$$E(X_i - \overline{X}) = 0, \quad D(X_i - \overline{X}) = \frac{n-1}{n}\sigma^2$$

$$X_i - \overline{X} \sim N\left(0, \frac{n-1}{n}\sigma^2\right)$$

$$E(|X_i - \overline{X}|) = \int_{-\infty}^{+\infty} |z| \frac{1}{\sqrt{2\pi}\sqrt{\frac{n-1}{n}}\sigma} e^{-\frac{z^2}{2\frac{n-1}{n}\sigma^2}} dz$$

$$= 2\int_0^{+\infty} z \frac{1}{\sqrt{2\pi}\sqrt{\frac{n-1}{n}}\sigma} e^{-\frac{z^2}{2\frac{n-1}{n}\sigma^2}} dz$$

$$= \frac{2}{\sqrt{2\pi}}\sqrt{\frac{n-1}{n}}\sigma$$

故

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

$$=kn\frac{2}{\sqrt{2\pi}}\sqrt{\frac{n-1}{n}}\sigma\overset{\text{令}}{=}\sigma$$

$$\longrightarrow\quad k=\sqrt{\frac{\pi}{2n(n-1)}}$$

例

设总体 X 服从参数为 θ 的指数分布, 概率密度

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

其中参数 $\theta>0$, 又设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 试证 $\overline{X}$ 和 $nZ=n[\min(X_1,X_2,\cdots,X_n)]$ 都是 θ 的无偏估计.

证 $E(X)=\theta,\quad E(\overline{X})=\theta$

所以 $\overline{X}$ 是参数 θ 的无偏估计量．而

$Z=\min(X_1,\dots,X_n)$ 具有概率密度

$$f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-nx/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

$$F_N(z)=1-[1-F(z)]^n$$

故知 $E(Z)=\dfrac{\theta}{n},\quad E(nZ)=\theta$

即 nZ 也是参数 θ 的无偏估计量．

由以上例子可知,一个参数可以有不同的无偏估计量.

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 $\hat{\theta}_2$ 好.

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.

这就引进了有效性这一概念 .

有效性

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$与$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$

都是θ的无偏估计量，若有 $D(\hat{\theta}_1)\le D(\hat{\theta}_2)$,

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

例　设总体 X 服从参数为 θ 的指数分布, 概率密度

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

其中参数 $\theta>0$, 又设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, $\overline{X}$ 和 $nZ=n[\min(X_1, X_2, \cdots, X_n)]$ 都是 θ 的无偏估计.

试证当 $n>1$ 时, θ 的无偏估计量 $\overline{X}$ 较 nZ 有效.

证 $D(X)=\theta^2,$

故有 $D(\overline{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\theta^2}{n}$

而 $D(Z)=\frac{\theta^2}{n^2},$ 故有 $D(nZ)=\theta^2.$

当 $n>1$ 时， $D(nZ)>D(\overline{X}),$ 故 $\overline{X}$ 较 nZ 有效.

相合性

若$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为参数$\theta$的估计量，若对于任意$\theta\in\Theta$，当$n\to\infty$时，$\hat{\theta}(X_1,X_2,\cdots,X_n)$依概率收敛于$\theta$，则称$\hat{\theta}$为$\theta$的相合估计量.

θ 为 θ 的**相合估计量**

$\Leftrightarrow$ 对于任意 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|\theta-\theta|<\varepsilon\}=1,\ \theta\in\Theta$$

相合性是对估计量的一个基本要求,
不具备相合性的估计量是不予以考虑的.

由辛钦定理

若总体 X 的数学期望 $E(X)=\mu$ 存在，则有

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X^k)=\mu_k \ (k=1,2,\cdots)$$

故

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$$ 为 $E(X^k)=\mu_k \ (k=1,2,\cdots)$ 的相合估计量.

关于相合性的两个常用结论

1. 样本 k 阶矩是总体 k 阶矩的相合性估计量. 由大数定律证明

2. 设 $\hat{\theta}$ 是 θ 的无偏估计量, 且 $\lim\limits_{n\to\infty} D(\hat{\theta}) = 0$, 则 $\hat{\theta}$ 是 θ 的相合估计量. 用切比雪夫不等式证明

矩法得到的估计量一般为相合估计量

在一定条件下, 极大似然估计具有相合性

例 $X \sim f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0, \\ 0 & x \le 0 \end{cases}$ $\theta > 0$ 为常数

则 $\overline{X}$ 是θ的无偏、有效、相合估计量.

证 已证明 $\overline{X}$是θ的无偏、有效估计量.

$$\lim_{n\to\infty} D(\overline{X}) = \lim_{n\to\infty} \frac{\theta^2}{n} = 0$$

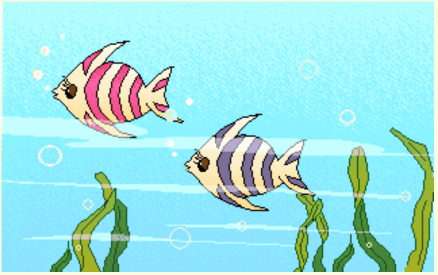
所以 $\overline{X}$ 是 θ 的相合估计量, 证毕.

区间估计

前面，我们讨论了参数点估计. 它是用样本算得的一个值去估计未知参数. 但是，点估计值仅仅未知参数的一个近似值，它没有反映出这个近似值的误差范围，使用起来把握不大. 区间估计正好弥补了点估计的这个缺陷 .

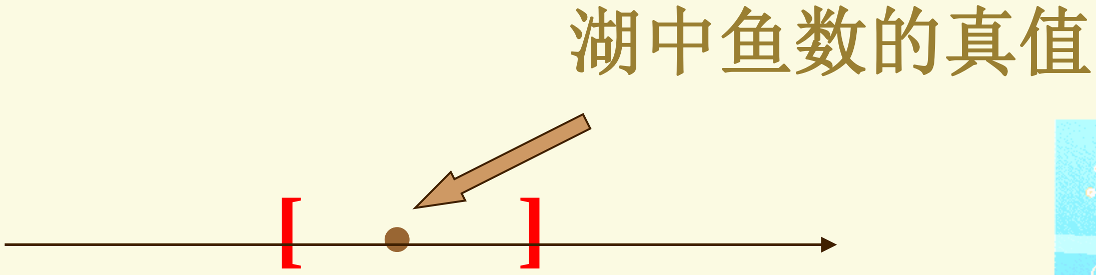
譬如，在估计湖中鱼数的问题中，若我们根据一个实际样本，得到鱼数 N 的极大似然估计为1000条.

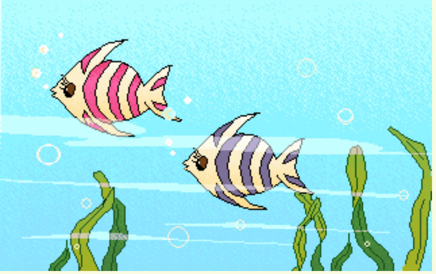
实际上，N的真值可能大于1000条，也可能小于1000条.



若我们能给出一个区间，在此区间内我们合理地相信 N 的真值位于其中. 这样对鱼数的估计就有把握多了.

也就是说，我们希望确定一个区间，使我们能以比较高的可靠程度相信它包含真参数值.





这里所说的“可靠程度”是用概率来度量的，称为置信度或置信水平.

习惯上把置信水平记作 $1-\alpha$ ，这里α是一个很小的正数.

置信水平的大小是根据实际需要选定的.

例如，通常可取置信水平 $1-\alpha$ =0. 95或0. 9等.

根据一个实际样本，由给定的置信水平，我们求出一个尽可能小的区间 $(\underline{\theta},\overline{\theta})$ 使

$$P\{\underline{\theta}<\theta<\overline{\theta}\}=1-\alpha$$

称区间 $(\underline{\theta},\overline{\theta})$ 为 θ 的置信水平为 $1-\alpha$ 的置信区间.

置信区间定义

设 θ 是一个待估参数，给定 $\alpha > 0$, 若由样本

$X_1, X_2, \ldots X_n$确定的两个统计量

$$\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n)$$

$$\overline{\theta} = \overline{\theta}(X_1, X_2, \cdots, X_n)$$

$$(\underline{\theta} < \overline{\theta})$$

满足

$$P\{\underline{\theta} < \theta < \overline{\theta}\} = 1 - \alpha$$

则称区间 $(\underline{\theta}, \overline{\theta})$ 是 θ 的置信水平（置信度）为$1-\alpha$

的置信区间.

$\underline{\theta}$ 和 $\overline{\theta}$ 分别称为置信下限和置信上限.

例 设$X_1,\ldots X_n$是取自 $N(\mu,\sigma^2)$的样本，σ^2已知，求参数μ的置信度为$1-\alpha$的置信区间.

解 μ 的无偏估计为 $\overline{X}$

取 $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

对给定的置信水平 $1-\alpha$,

查正态分布表得 $z_{\alpha/2}$, 使

$$P\{|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}|\le z_{\alpha/2}\}=1-\alpha$$

从中解得

$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}=1-\alpha$$

$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}$$

$$=1-\alpha$$

于是所求μ 的 置信区间为

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

从解题的过程，我们归纳出求置信区间的一般步骤如下:

1. 明确问题，是求什么参数的置信区间？置信水平$1-\alpha$是多少?

2. 寻找参数θ的一个良好的点估计

$$T(X_1,X_2,\dots X_n)$$

3. 寻找一个待估参数 θ和估计量 T 的函数 $W(T,\theta)=W(X_1,X_2,\dots X_n,\ \theta)$ 且其分布为已知.

4. 对于给定的置信水平$1-\alpha$，根据W($X_1,X_2,\ldots X_n$, θ)的分布，确定常数*a*, *b*，使得

$$P(a < W(X_1,X_2,\ldots X_n, \theta) < b) = 1-\alpha$$

5. 对$a < W(X_1,X_2,\ldots X_n, \theta) < b$作等价变形,得到如下形式:

$$\underline{\theta} < \theta < \overline{\theta}$$

即

$$P\{\underline{\theta} < \theta < \overline{\theta}\} = 1-\alpha$$

于是$(\underline{\theta},\overline{\theta})$就是$\theta$的100($1-\alpha$)%的置信区间.

可见，确定区间估计很关键的是要寻找一个

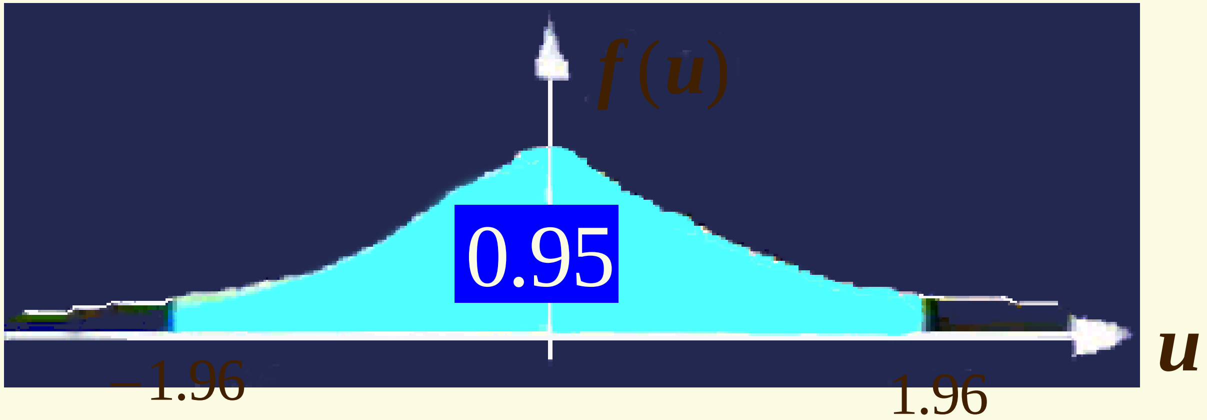
待估参数 θ 和估计量T 的函数 $W(X_1, X_2, \cdots X_n, \theta)$且$W(X_1, X_2, \cdots X_n, \theta)$ 的分布为已知，不依赖于任何未知参数 .

而这与总体分布有关，所以，总体分布的形式是否已知，是怎样的类型，至关重要.

需要指出的是，给定样本，给定置信水平 ，置信区间也不是唯一的.

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

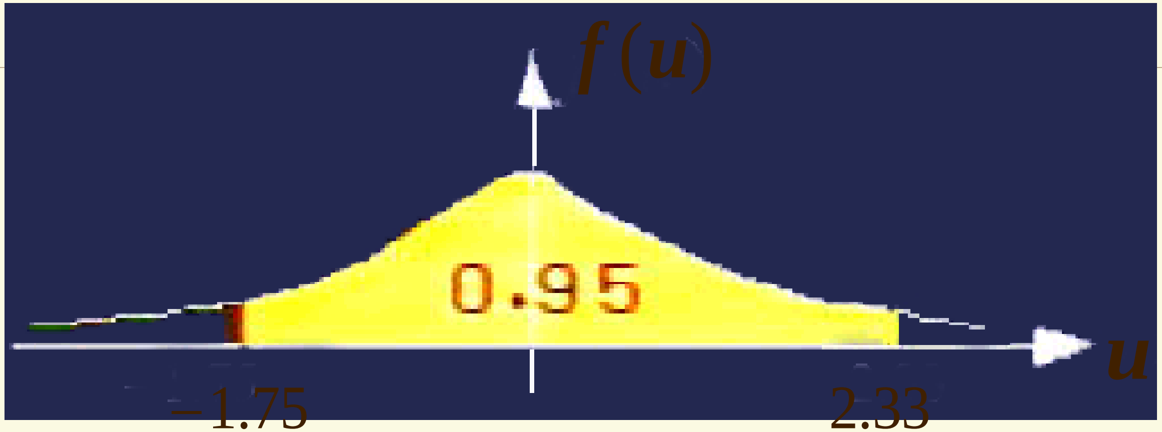
例如，由 $P(-1.96\le Z\le 1.96)=0.95$



我们得到均值 μ 的置信水平为 $1-\alpha=0.95$ 的置信区间为

$$\left[\bar{X}-1.96\sigma/\sqrt{n},\ \bar{X}+1.96\sigma/\sqrt{n}\right]$$

由 $P(-1.75 \le Z \le 2.33) = 0.95$



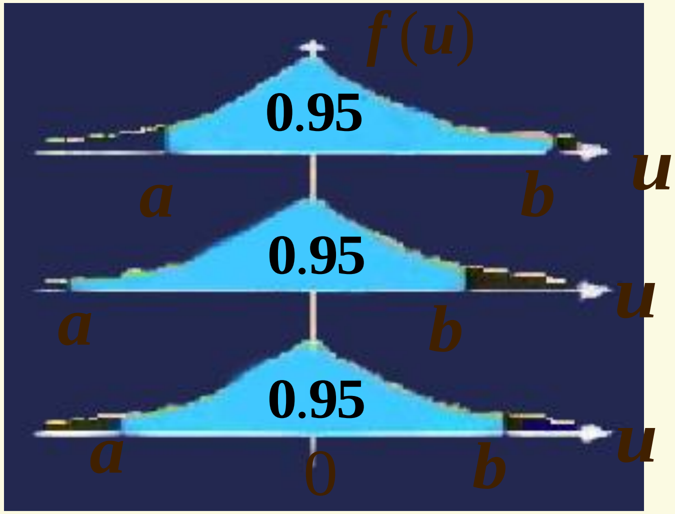
我们得到均值 μ 的置信水平为 $1-\alpha = 0.95$ 的置信区间为

$$[\bar{X} - 1.75\sigma/\sqrt{n}, \bar{X} + 2.33\sigma/\sqrt{n}]$$

这个区间比前面一个要长一些.

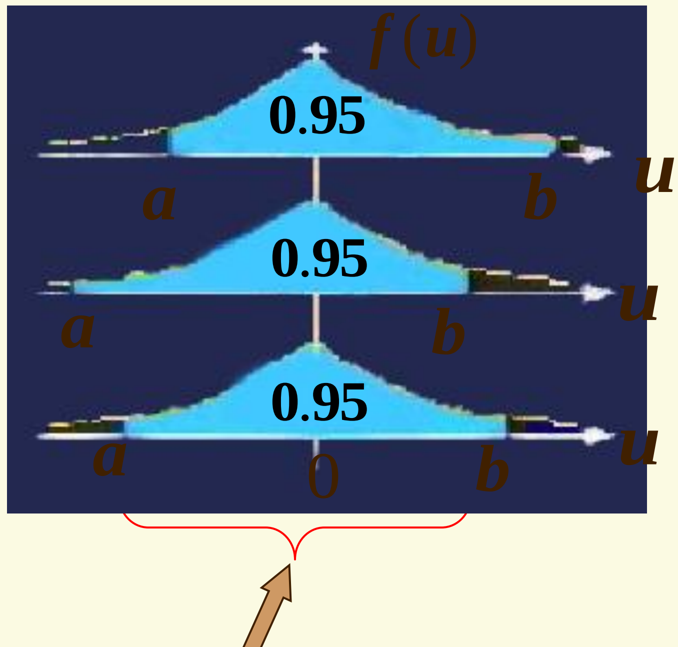
类似地，我们可得到若干个不同的置信区间.

任意两个数*a*和*b*，只要它们的纵标包含$f(u)$下95%的面积，就确定一个95%的置信区间.



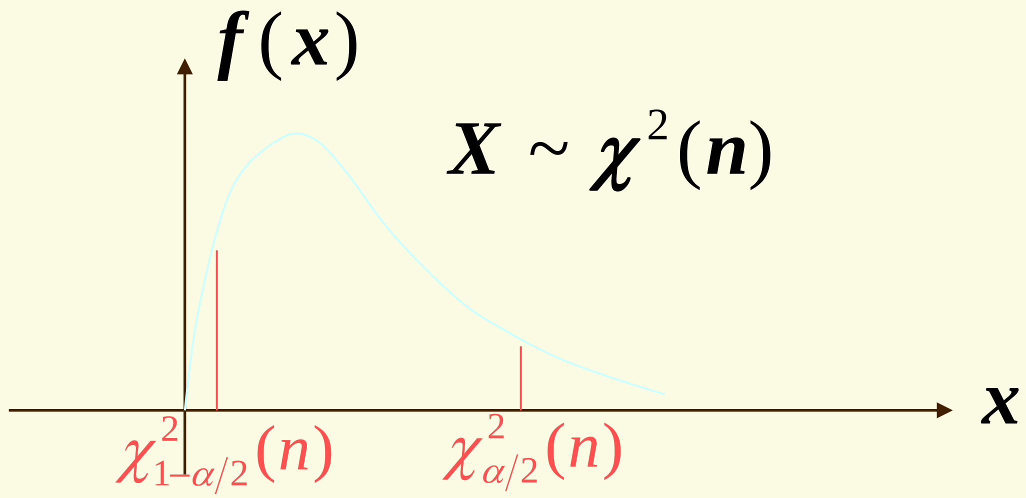
我们总是希望置信区间尽可能短.

在概率密度为单峰且对称的情形，当$a=-b$时求得的置信区间的长度为最短.



$a=-b$

即使在概率密度不对称的情形，如 χ^2分布，*F*分布，习惯上仍取对称的分位点来计算未知参数的置信区间.



正态总体均值与方差的区间估计

一、单个总体 $N(\mu,\sigma^2)$ 的情况

$X \sim N(\mu,\sigma^2)$,并设 $X_1,\dots,X_n$ 为来自总体的样本，$\bar{X},S^2$ 分别为样本均值和样本方差.

1. 均值 μ 的置信区间

1° σ^2 为已知

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

可得到 μ 的置信水平为 $1-\alpha$ 的置信区间为

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

2° σ^2 为未知

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

此分布不依赖于任何未知参数

由 $$P\left\{\left|\frac{\bar{X}-\mu}{S/\sqrt{n}}\right| < t_{\alpha/2}(n-1)\right\} = 1-\alpha$$

可得到 μ 的置信水平为 $1-\alpha$ 的置信区间为

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

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

例 有一大批糖果.现从中随机地取 16 袋，称得重量(以克计)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地服从正态分布,试求总体均值 μ 的置信水平0.95为的置信区间.

解 这里 $1-\alpha=0.95, \alpha/2=0.025, n-1=15,$

$t_{0.025}(15)=2.1315.$

$$\bar{x}=\frac{1}{16}\sum_{i=1}^{16}x_i=503.75,$$

$$s=\sqrt{\frac{1}{15}\sum_{i=1}^{16}(x_i-\bar{x})^2}=6.2022.$$

于是得到μ的置信水平为 0.95 的置信区间为

$$\left(\bar{x} \pm \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$$

即 $(500.4, 507.1)$

2. 方差σ^2的置信区间，当μ未知时

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

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

可得到σ^2的置信水平为$1-\alpha$的置信区间为

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

由

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

可得到标准差 σ 的置信水平为 $1-\alpha$ 的置信区间为

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

方差σ^2的置信区间　当μ已知时

当μ已知时，可利用下式求得置信区间

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

例 有一大批糖果.现从中随机地取 16 袋，称得重量(以克计)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地服从正态分布,试求总体标准差 σ 的置信水平0.95为的置信区间.

解 这里 $\alpha/2 = 0.025, 1-\alpha/2 = 0.975, n-1 = 15,$

$$\chi^2_{0.025}(15) = 27.488, \quad \chi^2_{0.975}(15) = 6.262.$$

$$s = \sqrt{\frac{1}{15}\sum_{i=1}^{16}(x_i - \overline{x})^2} = 6.2022.$$

于是得到 σ 的置信水平为 0.95 的置信区间为

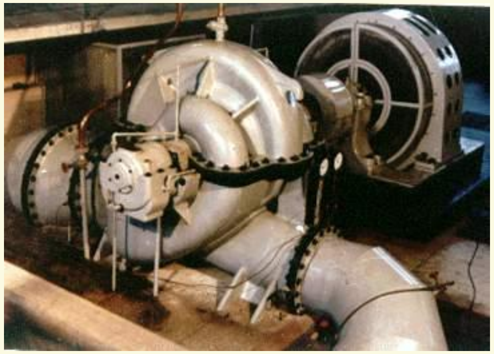
$$\left(\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{\alpha/2}(n-1)}}, \frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}}\right)$$

即 $$(4.58, 9.60).$$

单侧置信区间

之前讨论的置信区间中置信限都是双侧的，但对于有些实际问题，人们关心的只是参数在一个方向的界限.

例如对于设备、元件的使用寿命来说，平均寿命过长没什么问题，过短就有问题了.



这时，可将置信上限取为$+\infty$，而只着眼于置信下限，这样求得的置信区间叫单侧置信区间.

于是引入单侧置信区间和置信限的定义：

定义 设 θ 是一个待估参数，给定 $\alpha>0$,
若由样本$X_1,X_2,\ldots X_n$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

$$P\{\theta\geq\underline{\theta}\}=1-\alpha$$

则称区间$[\underline{\theta},+\infty)$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.

若由样本$X_1, X_2, \ldots X_n$确定的统计量

$$\bar{\theta} = \bar{\theta}(X_1, X_2, \cdots, X_n)$$

对于任意 $\theta \in \Theta$，满足

$$P\{\theta \le \bar{\theta}\} = 1 - \alpha$$

则称区间$(-\infty, \bar{\theta}]$是θ的置信水平为$1-\alpha$的单侧置信区间. $\bar{\theta}$ 称为 θ的置信水平为 $1-\alpha$ 的单侧置信上限.

例 从一批灯泡中随机抽取5只作寿命试验，测得寿命X（单位：小时）如下：

1050，1100，1120，1250，1280

设灯泡寿命服从正态分布. 求灯泡寿命均值μ的置信水平为0.95的单侧置信下限.

解 μ的点估计取为样本均值 $\overline{X}$，

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

方差 σ^2未知

对给定的置信水平$1-\alpha$, 确定分位点$t_\alpha(n-1)$

使
$$P\{\frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_\alpha(n-1)\} = 1-\alpha$$

即
$$P\{\mu \ge \bar{X} - t_\alpha(n-1)\frac{S}{\sqrt{n}}\} = 1-\alpha$$

于是得到μ的置信水平为$1-\alpha$的单侧置信区间为

$$[\bar{X} - t_\alpha(n-1)\frac{S}{\sqrt{n}}, \infty]$$

即 μ 的置信水平为$1-\alpha$的单侧置信下限为

$$\overline{X}-t_{\alpha}(n-1)\frac{S}{\sqrt{n}}$$

将样本值代入得

μ 的置信水平为0.95的单侧置信下限是

1065小时

作业

Exes. 2, 3 , 4, 10, 14, 16, 17, 21, 25