# 第八章 参数估计

实际工作中碰到的随机变量往往是知道大致的分布类型，但不知道确切的分布。

需要根据样本来估计总体的参数。

这类问题称为参数估计。

通常有两种方法：

点估计：以样本的某一函数值作为总体中未知参数的估计值。

区间估计：依据样本把总体的参数确定在某一范围内。

## §1 估计量的优劣标准

设$\theta$为总体中要被估计的一个未知参数(比如期望)。

$\hat{\theta}$是$\theta$的估计值，它是样本$X_1, X_2, \ldots, X_n$的函数。

如样本平均值$\overline{X}$与样本方差$S^2$等。

希望估计量能代表真实参数。

三种常用的评价标准：

(一)一致估计

一般$\hat{\theta} \neq \theta$,但希望当$n \to \infty$时,$\hat{\theta}$与$\theta$越来越接近。

即样本容量增大时,$\hat{\theta}$依概率收敛于$\theta$

定义1 如果当$n\to\infty$时,$\hat{\theta}$依概率收敛于$\theta$,即任给$\varepsilon>0$, $\lim\limits_{n\to\infty}P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1$

则称$\hat{\theta}$为参数$\theta$的一致估计。

一致性只在样本容量较大时才起作用.

- 例1 若总体$\xi$服从$[0,\theta]$上的均匀分布, $X_1, X_2, \ldots, X_n$是一组样本。证明:$\hat{\theta}=2\overline{X}$是$\theta$的一致估计。

证: $EX_i=\dfrac{\theta}{2}$, $DX_i=\dfrac{\theta^2}{12}$

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

$$D\overline{X}=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{\theta^2}{12n}$$

利用切贝谢夫不等式，对任给$\varepsilon>0$

$$P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=P\left(\left|2\overline{X}-\theta\right|<\varepsilon\right)$$

$$=P\left(\left|\overline{X}-\frac{\theta}{2}\right|<\frac{\varepsilon}{2}\right)$$

$$\geq 1-\frac{\theta^2}{12n}\bigg/\left(\frac{\varepsilon}{2}\right)^2$$

$$=1-\frac{\theta^2}{3n\varepsilon^2}$$

即 $$1\geq P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)\geq 1-\frac{\theta^2}{3n\varepsilon^2}$$

故 $$\lim_{n\to\infty}P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1$$

即 $\hat{\theta}$是$\theta$的一致估计

## (二)无偏估计

如果有一系列抽样构成各个估计，

希望这些估计的期望值与参数的真实值相等。

即样本估计量在参数值的真实值周围摆动，没有系统误差。

- 定义2 如果$E\hat{\theta}=\theta$成立，则称估计$\hat{\theta}$为参数$\theta$的无偏估计。

如例1中的估计即为无偏估计

- 例2 从总体$\xi$中取一组样本$(X_1,\ldots,X_n)$，$E\xi=\mu$，$D\xi=\sigma^2$。试证样本平均值$\overline{X}$与样本方差$S^2$分别是$\mu$及$\sigma^2$的无偏估计。

证：$$E\overline{X}=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}EX_i=\mu$$

$$D\overline{X}=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{\sigma^2}{n}$$

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

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

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

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

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

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

注:无偏估计不唯一,如例3.

## (三)有效估计

无偏性只保证了$\hat{\theta}$的概率平均等于$\theta$，其取值可能与$\theta$相差很大。

要保证$\hat{\theta}$的取值集中于$\theta$附近，就要求$\hat{\theta}$的方差越小越好。

- 定义3 设$\hat{\theta}_1$与$\hat{\theta}_2$都是$\theta$的无偏估计量，若$D\hat{\theta}_1 < D\hat{\theta}_2$，则称$\hat{\theta}_1$是比$\hat{\theta}_2$有效的估计量。在$\theta$的一切无偏估计量中方差最小的估计量$\hat{\theta}$称为$\theta$的有效估计量。

例3 比较总体期望μ的两个无偏估计的有效性。

$$\overline{X} = \frac{1}{3}\sum_{i=1}^{3} X_i$$

$$X' = \sum_{i=1}^{3} a_i X_i \Big/ \sum_{i=1}^{3} a_i ,\left(\sum_{i=1}^{3} a_i \neq 0\right)$$

解：$E\overline{X} = \mu$

$$EX' = \sum_{i=1}^{3} a_i EX_i \Big/ \sum_{i=1}^{3} a_i = \mu$$

$$D\overline{X} = \sigma^2/3$$

$$DX' = \sum_{i=1}^{3} a_i^2 DX_i \Big/ \left(\sum_{i=1}^{3} a_i\right)^2 = \sigma^2 \sum_{i=1}^{3} a_i^2 \Big/ \left(\sum_{i=1}^{3} a_i\right)^2$$

利用 $a_i^2+a_j^2\ge 2a_ia_j$有

$$\left(\sum_{i=1}^{3}a_i\right)^2=\left(a_1+a_2+a_3\right)^2$$

$$=a_1^2+a_2^2+a_3^2+2a_1a_2+2a_1a_3+2a_2a_3$$

$$\le a_1^2+a_2^2+a_3^2+\left(a_1^2+a_2^2\right)+\left(a_1^2+a_3^2\right)+\left(a_2^2+a_3^2\right)$$

$$=3\left(a_1^2+a_2^2+a_3^2\right)=3\sum_{i=1}^{3}a_i^2$$

故 $$DX'=\sigma^2\sum_{i=1}^{3}a_i^2\Big/\left(\sum_{i=1}^{3}a_i\right)^2$$

$$\ge\sigma^2\sum_{i=1}^{3}a_i^2\Big/3\sum_{i=1}^{3}a_i^2=\frac{\sigma^2}{3}=D\overline{X}$$

故 $\overline{X}$与$X'$有效

## §2 点估计

### (一)矩法

利用样本的数字特征作为总体数字特征的估计。

用样本平均值$\overline{X}$估计总体的期望。

用样本方差$S^2$估计总体的方差

- 例1 某厂某天生产了一大批灯泡，从中抽取10个进行寿命试验，得数据如下(单位：小时)

1050 1100 1080 1120 1200

1250 1040 1130 1300 1200

问该天生产的灯泡平均寿命大约是多少？

解:$\overline{X}=1147$

灯泡的平均寿命约为1147小时

例2 用矩法估计事件发生的概率p

解：

| $\xi$ | 0 | 1 |
| --- | --- | --- |
| P | $1-p$ | $p$ |

$$E\xi = p$$

若$X_1,\ldots,X_n$为一组样本，则

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

- 例3 设总体$\xi$服从[a,b]上的均匀分布，用矩法估计a与b

解：设$X_1, \ldots, X_n$为一组样本

由于$E\xi = \frac{a+b}{2}$ $\quad D\xi = \frac{1}{12}(b-a)^2$

令 $\frac{a+b}{2} = \overline{X}$ $\quad \frac{1}{12}(b-a)^2 = S^2$

联立求解可得

$$\hat{a} = \overline{X} - \sqrt{3}S$$

$$\hat{b} = \overline{X} + \sqrt{3}S$$

- 例4 设总体$\xi \sim \varphi(x)=\begin{cases}(\alpha+1)x^{\alpha} & 0<x<1 \\ 0 & \text{其它}\end{cases}$

$\alpha$是未知参数，求$\alpha$的矩估计。

解：$E\xi=\int_0^1 x\cdot(\alpha+1)x^{\alpha}dx=\dfrac{\alpha+1}{\alpha+2}$

令 $\dfrac{\alpha+1}{\alpha+2}=\overline{X}$

解得 $\hat{\alpha}=\dfrac{2\overline{X}-1}{1-\overline{X}}$

矩估计的优点：直接、简便

缺点：未充分利用分布信息

设$(x_1,\ldots,x_n)$为总体$\xi$的一组样本观察值。

要选取总体分布中未知参数$\theta$的估计值$\hat{\theta}$,使得

$\hat{\theta}$作为参数时，上述样本出现的可能性最大。

这种方法称为最大似然法。

若$\xi$是离散型随机变量

$$P(\xi = x_i) = p(x_i, \theta)$$

则样本$x_1,\ldots,x_n$发生的概率为

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

其中$\theta$是未知参数，可以是一个值，也可以是向量。

若$\xi$是连续型随机变量，

$$\xi \sim \varphi(x,\theta)$$

则应将概率改为

$$L(x_1,\ldots,x_n,\theta)=\prod_{i=1}^{n}\varphi(x_i,\theta)$$

称L为样本的似然函数

若$\theta$是向量，则L是多元函数。

- 定义1 如果$L(x_1,\ldots,x_n,\theta)$在$\hat{\theta}$处达到最大值，则称$\hat{\theta}$是$\theta$的最大似然估计。

实际上,$\hat{\theta}$是似然函数L的最大值点。

由于$\ln L$与$L$同时达到最大值

求$\ln L$的最大值点往往更方便。

$\ln L$称为对数似然函数。

若$\theta$为向量，$\theta=(\theta_1,\ldots,\theta_m)$

解方程组
$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1} = 0 \\ \quad\ldots \\ \dfrac{\partial \ln L}{\partial \theta_m} = 0 \end{cases}$$

得到驻点$(\hat{\theta}_1,\ldots,\hat{\theta}_m)$,它常常就是最大值点。

- 例6 用最大似然法估计事件发生的概率p

解：

| $\xi$ | 0 | 1 |
| --- | --- | --- |
| P | $1-p$ | $p$ |

$$P(\xi = k) = p^k (1-p)^{1-k} \qquad k = 0,1$$

若$x_1, \dots, x_n$为一组样本观察值

$$L(p) = \prod_{i=1}^{n} P(\xi = x_i)$$

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

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

取对数得

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

求驻点：

$$\left[\ln L(p)\right]'=\left(\sum_{i=1}^{n}x_i\right)\frac{1}{p}-\left(n-\sum_{i=1}^{n}x_i\right)\frac{1}{1-p}=0$$

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

这是唯一的可能极值点，也是最大值点。

即由最大似然法，可以由频率估计概率。

例7　总体$\xi$服从几何分布$P(\xi=k)=p(1-p)^{k-1}$，$k=1,2,\ldots,$设$x_1,\ldots,x_n$为样本观察值，用最大似然法估计参数p

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

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

$$(\ln L)'=\frac{n}{p}-\left(\sum_{i=1}^{n}x_i-n\right)\frac{1}{1-p}=0$$

解得

$$\hat{p}=n\Big/\sum_{i=1}^{n}x_i=1\big/\overline{x}$$

例8 已知$\xi \sim \varphi(x,\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x>0 \\ 0 & 其它\end{cases}(\theta>0)$

若有一组样本

| | | | | | |
|---|---|---|---|---|---|
| 16 | 29 | 50 | 68 | 100 | 130 |
| 140 | 270 | 280 | 340 | 410 | 450 |
| 520 | 620 | 190 | 210 | 800 | 1100 |

用最大似然法估计$\theta$

解：用$x_1,\ldots,x_n$表示样本观察值

$$L=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i}{\theta}}=\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i}$$

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

$$(\ln L)' = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i \quad = 0$$

求解得到最大似然估计为

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

而 $\overline{x} = \frac{1}{18}\left(16 + 29 + \ldots + 800 + 1100\right) \approx 318$

故 $\hat{\theta} = 318$

例9 已知ξ服从正态分布$N(\mu,\sigma^2)$，$(x_1,...,x_n)$为ξ的一组样本观察值，用最大似然法估计μ，$\sigma^2$的值。

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

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

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

分别对μ与$\sigma^2$求偏导，得

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

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

求解可得

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

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

注意:$\hat{\sigma}^2$不是$\sigma^2$的无偏估计。

# §3 区间估计

点估计值未必等于真实值。

即使相等也无法判定。

根据估计量的分布，在一定的可靠程度下，指出被估计的总体参数所在的可能数值范围。

这类问题称为参数的区间估计。

找两个统计量$\hat{\theta}_1(X_1,\ldots,X_n)$与$\hat{\theta}_2(X_1,\ldots,X_n)$,使得

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$$

区间$(\hat{\theta}_1, \hat{\theta}_2)$称为置信区间。

$\hat{\theta}_1$及$\hat{\theta}_2$分别称为置信区间的上下限。

$1-\alpha$称为置信系数，也称置信概率或置信度。

$\alpha$是事先给定的一个小正数，它是指参数估计不准的概率。 一般设$\alpha=0.05$或$\alpha=0.01$

当样本不同时,$\hat{\theta}_1$与$\hat{\theta}_2$也会不同，而$\theta$是真实值。

$P(\hat{\theta}_1 < \theta < \hat{\theta}_2)$不是$\theta$落在区间$(\hat{\theta}_1, \hat{\theta}_2)$的概率。

而是随机区间$(\hat{\theta}_1, \hat{\theta}_2)$包含$\theta$的概率。

即平均100次抽样计算得到的100个区间中，约有$(1-\alpha)\times 100$个区间包含$\theta$

区间$(\hat{\theta}_1, \hat{\theta}_2)$不包含$\theta$的可能性不超过$\alpha$

即平均$1/\alpha$次估计中至多会有一次犯错误。

一般说来,$(\hat{\theta}_1, \hat{\theta}_2)$的区间长度越小(越精确)，则可靠程度$1-\alpha$也越小。

## (一)方差已知时，总体期望值Eξ的区间估计

### 1、总体分布未知

ξ为一般总体(非正态), Dξ已知

设$(X_1,\ldots,X_n)$为一组样本, $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

$$E\overline{X}=E\xi \qquad D\overline{X}=\frac{1}{n}D\xi$$

对$\overline{X}$应用切贝谢夫不等式

$$P\left(\left|\overline{X}-E\xi\right|<\varepsilon\right)\geq 1-\frac{D\overline{X}}{\varepsilon^2}$$

即 $$P\left(\left|\overline{X}-E\xi\right|<\varepsilon\right)\geq 1-\frac{D\xi}{n\varepsilon^2}$$

$\left|\overline{X}-E\xi\right|<\varepsilon$可改写为

$$\overline{X}-\varepsilon<E\xi<\overline{X}+\varepsilon$$

要使 $1-\dfrac{D\xi}{n\varepsilon^2}=1-\alpha$, 可取$\varepsilon=\sqrt{\dfrac{D\xi}{n\alpha}}$

于是

$$P\left(\overline{X}-\sqrt{\frac{D\xi}{n\alpha}}<E\xi<\overline{X}+\sqrt{\frac{D\xi}{n\alpha}}\right)\geq 1-\alpha$$

即$\left(\overline{X}-\sqrt{\dfrac{D\xi}{n\alpha}},\overline{X}+\sqrt{\dfrac{D\xi}{n\alpha}}\right)$是$E\xi$的置信度为$1-\alpha$的置信区间。

$\overline{X}\pm\sqrt{\dfrac{D\xi}{n\alpha}}$是置信区间的上、下限

置信区间的长度为$2\sqrt{\dfrac{D\xi}{n\alpha}}$

要提高精度，需要增大样本容量n

或增大α，即降低可靠程度1−α

- 例1 若已知某天某车间生产的灯泡寿命的方差是8，抽取10只灯泡，其平均寿命为1147小时，求出灯泡平均寿命的置信区间(α=0. 05)

解：$n=10 \quad \bar{x}=1147 \quad D\xi=8 \quad \alpha=0.05$

$$\sqrt{\frac{D\xi}{n\alpha}}=\sqrt{\frac{8}{10\times 0.05}}=4$$

故$E\xi$的置信区间为$(1147-4,1147+4)$

即(1143，1151)

## 2、正态总体

设$(X_1,\ldots,X_n)$是取自总体$N(\mu,\sigma^2)$的样本

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

查表确定$u_\alpha$,使

$$P\left(|U|<u_\alpha\right)=1-\alpha$$

即 $2\Phi_0(u_\alpha)-1=1-\alpha$

$$\Phi_0(u_\alpha)=1-\frac{\alpha}{2}$$

如$\alpha$＝0.05时，$u_\alpha$=1.96

$\alpha=0.01$时，$u_\alpha=2.58$

而 $|U|=\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right|<u_\alpha$可改写为

$$\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha}$$

即 $P\left(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha}\right)=1-\alpha$

故$\mu$的置信度为$1-\alpha$的置信区间为

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha},\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha}\right)$$

区间长度为$\frac{2\sigma}{\sqrt{n}}u_{\alpha}$

要提高精度，需要增大样本容量n

或减少$u_{\alpha}$,即降低$1-\frac{\alpha}{2}$,也就是降低$1-\alpha$

例2 若已知某天某车间生产的灯泡寿命服从正态分布$N(\mu,8)$，抽取10只灯泡，平均寿命为1147小时，求出灯泡平均寿命的置信区间（$\alpha$=0. 05）

解：$\alpha=0.05, u_{\alpha}=1.96$

$$n=10 \qquad \sigma=\sqrt{8} \qquad \bar{x}=1147$$

$$\text{置信区间为}\left(1147-\frac{\sqrt{8}}{\sqrt{10}}1.96, 1147+\frac{\sqrt{8}}{\sqrt{10}}1.96\right)$$

即(1145. 25，1148. 75)

此区间比例1中的区间短，更精确

这是因为它利用了分布的信息。

- 例3 已知某炼铁厂的铁水含碳量再正常生产情况下服从正态分布，其方差$\sigma^2$=$0.108^2$。现在测定了9炉铁水，其平均含碳量为4.484。按此资料计算该厂铁水平均含碳量的置信区间，并要求有95%的可靠性。

解：设该厂铁水平均含碳量为$\mu$

$$\alpha=0.05 \quad u_{\alpha}=1.96$$

$$\sigma=0.108 \quad n=9 \quad \bar{x}=4.484$$

置信度为95%的置信区间为

$$\left(4.484-\frac{0.108}{\sqrt{9}}1.96,4.484+\frac{0.108}{\sqrt{9}}1.96\right)$$

即 $(4.413,4.555)$

## 3、一般总体大样本

当样本容量很大时, 由中心极限定理,$\bar{X}$近似服从正态分布，可利用正态总体的区间估计。

- 例4 对某地家庭收入进行抽样检查，随机抽取100个家庭，其样本平均值为11900元，据现有资料，总体家庭收入的标准差是1500元。求置信度为95％的家庭收入均值的置信区间。

解：$n=100 \quad \bar{x}=11900 \quad \sigma=1500$

$\alpha=0.05 \qquad u_{\alpha}=1.96$

置信区间为$\left(11900-\frac{1500}{\sqrt{100}}1.96, 11900+\frac{1500}{\sqrt{100}}1.96\right)$

即 (11606,12194)

## (二)方差未知时, 总体期望值Eξ的区间估计

### 1、正态总体，小样本

由于σ未知，要用样本来估计

设样本$(X_1,\ldots,X_n)$取自正态总体$N(\mu,\sigma^2)$

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

查表确定$t_\alpha$,使得

$$P\left(|T|\geq t_\alpha\right)=\alpha$$

即 $$P\left(\left|\frac{(\overline{x}-\mu)}{S/\sqrt{n}}\right|<t_\alpha\right)=1-\alpha$$

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

即置信度为$1-\alpha$的置信区间为

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

区间长度为$\dfrac{2S}{\sqrt{n}}t_{\alpha}$

要提高精确度，需要增大样本容量n

或减少$t_{\alpha}$,即增加$\alpha$,降低可靠度$1-\alpha$

确定$t_{\alpha}$要查t分布的双侧临界值表, 样表如下：

$$P\left(\left|t(n)\right| > t_{\alpha}\right) = \alpha, \qquad n为自由度$$

| $n \backslash \alpha$ | 0.9 | 0.7 | 0.5 | 0.3 | 0.1 | 0.05 | 0.01 |
|---|---|---|---|---|---|---|---|
| 1 | 0.158 | 0.510 | 1.000 | 1.963 | 6.314 | 12.706 | 63.657 |
| 2 | 0.142 | 0.445 | 0.816 | 1.386 | 2.920 | 4.303 | 9.925 |
| 3 | 0.137 | 0.424 | 0.765 | 1.250 | 2.353 | 3.182 | 5.841 |
| 4 | 0.134 | 0.414 | 0.741 | 1.190 | 2.132 | 2.776 | 4.604 |
| 5 | 0.132 | 0.408 | 0.727 | 1.156 | 2.015 | 2.571 | 4.032 |
| 6 | 0.131 | 0.404 | 0.718 | 1.134 | 1.943 | 2.447 | 3.707 |
| 7 | 0.130 | 0.402 | 0.711 | 1.119 | 1.895 | 2.365 | 3.499 |
| 8 | 0.130 | 0.399 | 0.706 | 1.108 | 1.860 | 2.306 | 3.355 |
| 9 | 0.129 | 0.398 | 0.703 | 1.100 | 1.833 | 2.262 | 3.250 |
| 10 | 0.129 | 0.397 | 0.700 | 1.093 | 1.812 | 2.228 | 3.169 |
| 40 | 0.126 | 0.388 | 0.681 | 1.050 | 1.684 | 2.021 | 2.704 |
| $\infty$ | 0.126 | 0.385 | 0.674 | 1.036 | 1.645 | 1.960 | 2.576 |

- 例5 假定出生婴儿体重服从正态分布，随机抽取5名婴儿，测其体重为3100，2520，3000，3160，3560，试以95％的置信系数估计婴儿平均体重的区间。

解：$\bar{x} = 3068 \qquad s = \sqrt{137020} = 370.16$

n=5 　　　　　n-1=4

$\alpha = 0.05$ 查表可得 $t_\alpha = 2.776$

故婴儿平均体重的置信区间为：

$$\left(3068 - \frac{370.16}{\sqrt{5}} 2.776, 3068 + \frac{370.16}{\sqrt{5}} 2.776\right)$$

即 $(2608.46, 3527.54)$

## 2、一般总体、大样本

可用$S^2$代替方差,$\overline{X}$近似服从正态分布

按方差已知的正态总体区间估计计算，置信区间为

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

- 例6 对某地家庭收入进行抽样调查，随机抽取100个家庭，其样本平均值为11900元，样本标准差为2000元，求置信度为95％的家庭收入均值的置信区间。

解：$n=100 \quad \overline{x}=11900 \quad s=2000 \quad \alpha=0.05 \quad u_{\alpha}=1.96$

置信区间为$\left(11900-\frac{2000}{\sqrt{100}}1.96, 11900+\frac{2000}{\sqrt{100}}1.96\right)$

即 $(11508,12292)$

## (三)小样本下正态总体方差$\sigma^2$的区间估计

设$(X_1,\ldots,X_n)$来自正态总体$N(\mu,\sigma^2)$

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

查表确定a与b使得

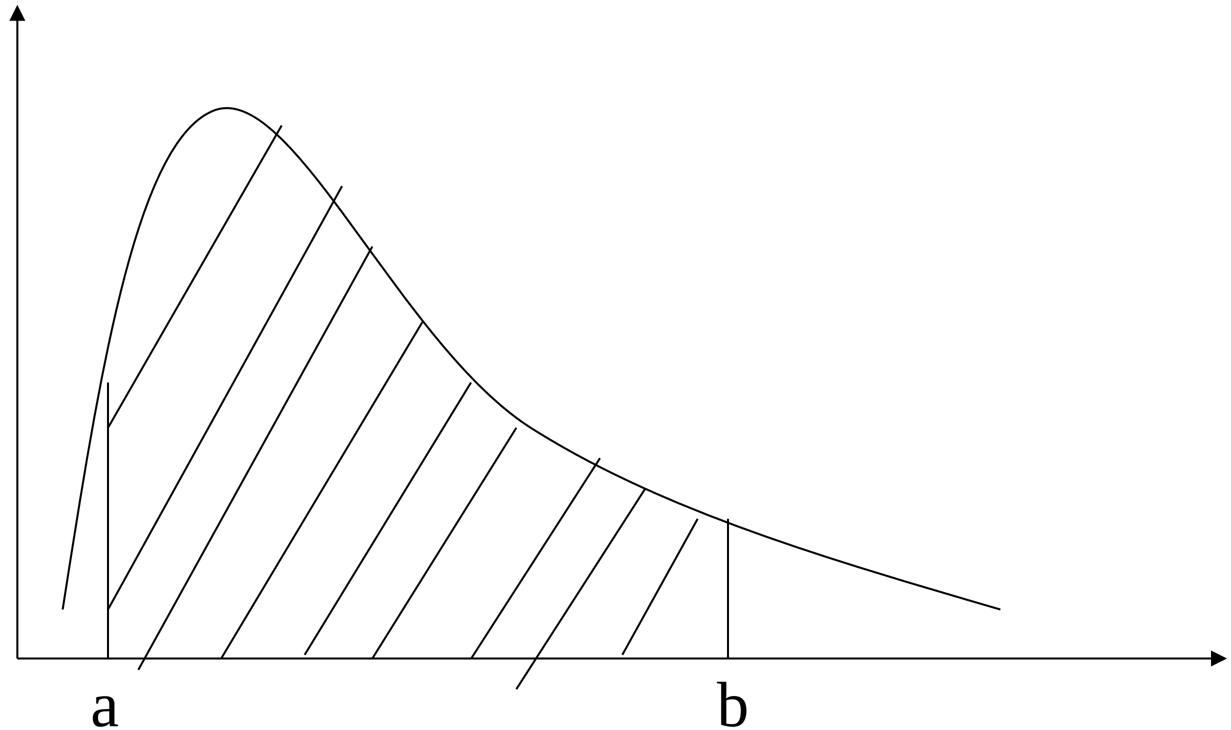
$$P\left(a<\chi^2<b\right)=P\left(a<\frac{(n-1)S^2}{\sigma^2}<b\right)=1-\alpha$$

即 $P\left(\frac{(n-1)S^2}{b}<\sigma^2<\frac{(n-1)S^2}{a}\right)=1-\alpha$

即置信度为$1-\alpha$的$\sigma^2$的置信区间为$\left(\frac{(n-1)S^2}{b},\frac{(n-1)S^2}{a}\right)$

$$P\left(\chi^2(n) \geq \chi^2_\alpha\right) = \alpha, n自由度$$

| n \ α | 0.99 | 0.975 | 0.95 | 0.05 | 0.025 | 0.01 |
|---|---|---|---|---|---|---|
| 1 | $0.0^3628$ | $0.0^3982$ | 0.00393 | 3.84 | 5.02 | 6.63 |
| 2 | 0.0201 | 0.0506 | 0.103 | 5.99 | 7.38 | 9.21 |
| 3 | 0.115 | 0.216 | 0.352 | 7.81 | 9.35 | 11.3 |
| 4 | 0.297 | 0.484 | 0.711 | 9.49 | 11.1 | 12.3 |
| 5 | 0.554 | 0.831 | 1.145 | 11.1 | 12.8 | 15.1 |
| 6 | 0.872 | 1.24 | 1.64 | 12.6 | 14.4 | 16.8 |
| 7 | 1.24 | 1.69 | 2.17 | 14.1 | 16.0 | 18.5 |
| 8 | 1.65 | 2.18 | 2.73 | 15.5 | 17.5 | 20.1 |
| 9 | 2.09 | 2.7 | 3.33 | 16.9 | 19.0 | 21.7 |
| 10 | 2.56 | 3.25 | 3.94 | 18.3 | 20.5 | 23.2 |
| 20 | 8.26 | 9.59 | 10.9 | 42.6 | 34.2 | 37.6 |
| 30 | 15.0 | 16.8 | 18.5 | 43.8 | 47.0 | 50.9 |




查表使$P\left(a<\chi^{2}<b\right)=1-\alpha$

即查$P(\chi^{2}\geq b)=\alpha/2$

$P(\chi^{2}\geq a)=1-\alpha/2$

例7 在例5中随机抽取5名新生婴儿, 其体重的样本方差为$s^2=137020$, 求新生婴儿体重的方差的区间估计($\alpha$=0. 05)

解: n=5 $s^2=137020$ $\alpha=0.05$

由$P(\chi^2\geq a)=1-\alpha/2=0.975$

$P(\chi^2\geq b)=\alpha/2=0.025$

查自由度为4的表得 a＝0.484，b＝11.1

故方差$\sigma^2$的区间估计为 $\left(\frac{4\times137020}{11.1},\frac{4\times137020}{0.484}\right)$

即 (49376.6,1132396.7)

标准差的区间估计为(222.2,1064.1)

例8　测量铅的比重，测量值服从正态分布$N(\mu,\sigma^2)$，测9次得到样本平均值为2.705，样本标准差为0.029，求铅的比重的均值的置信区间，测量方差的置信区间。($\alpha$=0.01)

解：$n=9$　　$\bar{x}=2.705$　　$s=0.029$

(1)方差未知，$\mu$的区间估计应为 $\left(\bar{x}-\frac{s}{\sqrt{n}}t_\alpha, x+\frac{s}{\sqrt{n}}t_\alpha\right)$

$n-1=8$　　$\alpha=0.01$　　$t_\alpha=3.355$

故$\mu$的置信区间为$\left(2.705-\frac{0.029}{\sqrt{9}}3.355, 2.705+\frac{0.029}{\sqrt{9}}3.355\right)$

即 $(2.6726, 2.7374)$

(2) $n-1=8,\ \alpha=0.01$

由 $P(\chi^2 \ge a) = 0.995$

$P(\chi^2 \ge b) = 0.005$

得 a=1.34,b=22.0

故$\sigma^2$的置信区间为$\left(\dfrac{8\times 0.029^2}{22}, \dfrac{8\times 0.029^2}{1.34}\right)$

即(0.0003058,0.005021)

标准差的区间为(0.0175,0.071)