8.3　区间估计

基本概念

用点估计来估计总体参数, 即使是无偏有效的估计量, 也会由于样本的随机性, 从一个样本算得估计量的值不一定恰是所要估计的参数真值. 而且, 即使真正相等, 由于参数值本身是未知的, 也无从肯定这种相等. 到底二者相差多少呢? 这个问题换一种提法就是, 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. 这就是参数的区间估计问题.

区间估计的具体做法是,找两个统计量

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

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

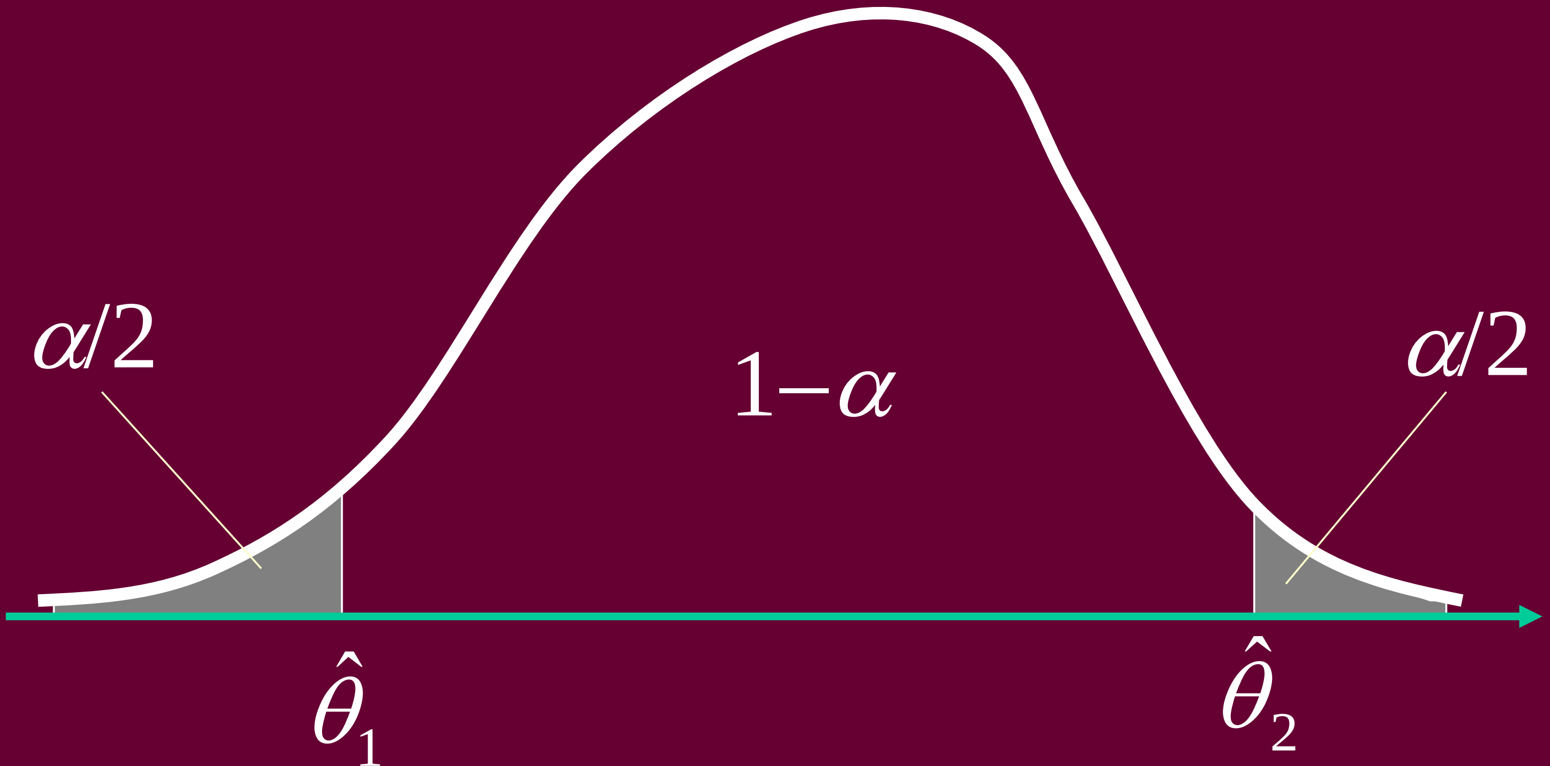
区间$(\hat{\theta}_1,\hat{\theta}_2)$称为置信区间,

$\hat{\theta}_2$和$\hat{\theta}_1$分别称为置信区间的上,下限

$1-\alpha$为置信系数,α称作检验水平,

通常$\alpha = 5\%$或1%

区间估计示意图



$1-\alpha$为<u>置信系数</u>, <u>置信概率</u>或<u>置信度</u>

α为<u>检验水平</u>

总体期望值$E\xi$的区间估计

第一种情形: 方差已知, 对$E\xi$进行区间估计

2. 正态总体

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

则$\overline{X} \sim N(\mu,\sigma^2/n)$,

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

查表可求得u_α使得

$$P(|U|<u_\alpha)=1-\alpha=2\Phi_0(u_\alpha)-1$$

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

则 $$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|<u_\alpha\right)=P\left(\left|\frac{\mu-\bar{X}}{\sigma/\sqrt{n}}\right|<u_\alpha\right)=1-\alpha$$

即 $$P(-u_\alpha<\frac{\mu-\bar{X}}{\sigma/\sqrt{n}}<u_\alpha)=1-\alpha$$

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

因此μ的置信度为$1-\alpha$的置信区间是

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

其中u_α由公式$\Phi_0(u_\alpha)=1-\alpha/2$,查表获得

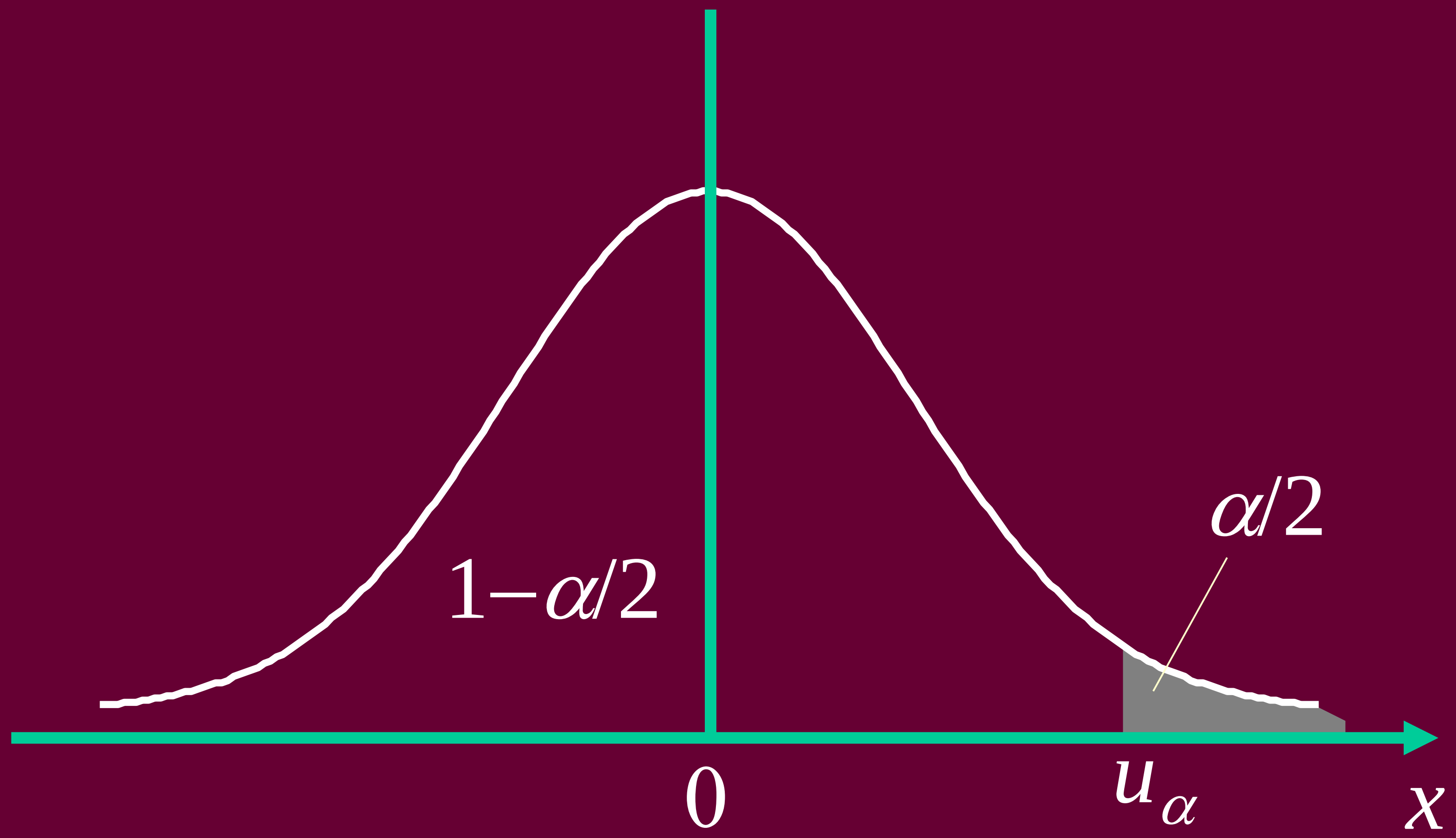
例如, 当α=0.05时, u_α=1.96, 有

$$\bar{x}-\frac{1.96\sigma}{\sqrt{n}}<\mu<\bar{x}+\frac{1.96\sigma}{\sqrt{n}}$$

当$\alpha=0.01$时,$u_\alpha=2.58$,有

$$\bar{x}-\frac{2.58\sigma}{\sqrt{n}}<\mu<\bar{x}+\frac{2.58\sigma}{\sqrt{n}}$$

查表示意图



例2 某灯泡厂某天生产了一大批灯泡, 假设灯泡的寿命ξ服从正态分布, $\xi \sim N(\mu,8)$, 从中抽取了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,试找出平均寿命区间(α=0.05).

解 因为α=0.05, 所以u_α=1.96,而n=10, s=2.8284

算出$\bar{x}=1147$,则

$$P(1147-\frac{2.8284\times 1.96}{\sqrt{10}}<\mu<1147-\frac{2.8284\times 1.96}{\sqrt{10}})$$

$$=0.95$$

即μ的置信区间为(1145.25, 1148.75)

例3 已知某炼铁厂的铁水含碳量在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μ, 已知α=5%, 所以u_α=1.96, μ的置信系数为95%的置信区间是

$$4.484-\frac{0.108}{\sqrt{9}}\times 1.96<\mu<4.484+\frac{0.108}{\sqrt{9}}\times 1.96$$

即 $4.413<\mu<4.555$

第二种情形:

方差$D\xi$未知, 对$E\xi$的区间估计

如果方差$D\xi$未知, 大样本下可用S^2来作为$D\xi$估计值, 仍可按正态分布的办法来进行区间估计

4. 方差未知的正态总体, 小样本下$E\xi$的区间估计

设样本$(X_1,\ldots,X_n)$来自正态总体N(μ,σ^2), 由于σ^2未知, 则令

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

对于给定的α和$n-1$查t分布临界值表确定 t_α使

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

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

即

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

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

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

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

即置信区间为$\left(\bar{X}-\frac{S}{\sqrt{n}}t_\alpha,\bar{X}+\frac{S}{\sqrt{n}}t_\alpha\right)$

例4 假定初生婴儿(男孩)的体重服从正态分布, 随机抽取12名婴儿, 测体重为3100, 2520, 3000, 3000, 3600, 3160, 3560, 3320, 2880, 2600, 3400, 2540. 试以95%的置信系数估计新生男婴儿的平均体重(单位:克).

解 设新生男婴儿体重为ξ克, 由于ξ服从正态分布, 方差σ^2未知, 因此要查t分布表.

对α=0.05, 因样本数n=12, 则查自由度为11的t分布表, 得$t_\alpha(12-1)=2.201$

再计算

$$\bar{x}=\frac{1}{12}(3100+\cdots+2540)\approx 3057$$

$$s=\sqrt{\frac{1}{11}\sum_{i=1}^{n}(x_i-3057)^2}\approx 375.3$$

因此，μ的置信度为95%的置信区间是

$$3057-\frac{375.3}{\sqrt{12}}\times 2.201<\mu<3057+\frac{375.3}{\sqrt{12}}\times 2.201$$

即 $2818<\mu<3295$

(二) 小样本下正态总体方差σ^2的估计

设样本$(X_1,\ldots,X_n)$来自正态总体$N(\mu,\sigma2)$. 则$\chi^2=(n-1)S^2/\sigma^2$服从具有$n-1$个自由度的χ^2分布, 对于给定的α, 查表可以确定a及b, 使得

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

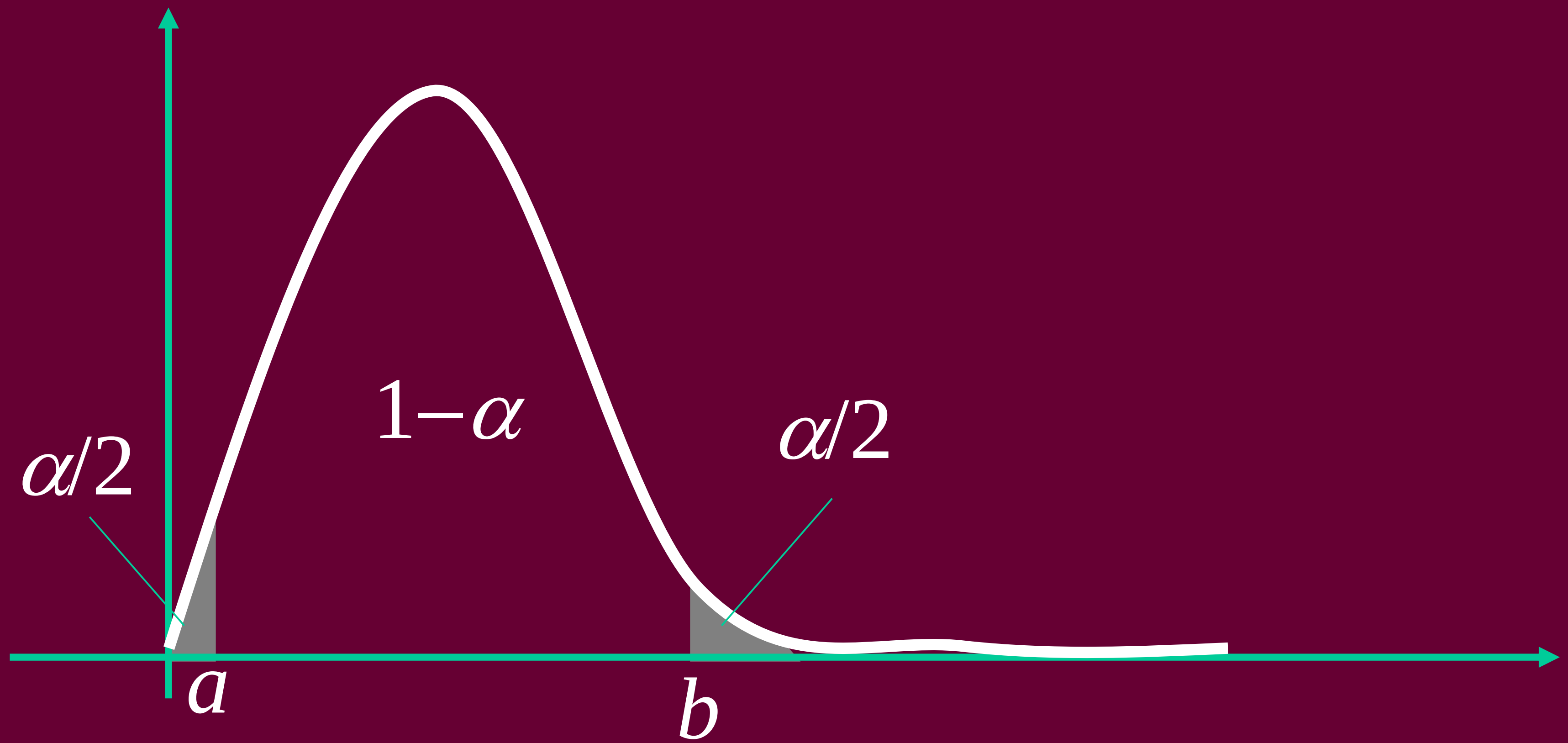
$$=1-\alpha$$

而$a<\dfrac{(n-1)S^2}{\sigma^2}<b \Leftrightarrow \dfrac{1}{a}>\dfrac{\sigma^2}{(n-1)S^2}>\dfrac{1}{b}$

$$\Leftrightarrow \frac{(n-1)S^2}{b}<\sigma^2<\frac{(n-1)S^2}{a}$$

在确定a, b时，一般是取

$$P(\chi^2 \le a) = P(\chi^2 \ge b) = \frac{\alpha}{2}$$



例5 假定初生婴儿(男孩)的体重服从正态分布, 随机抽取12名婴儿, 测体重为3100, 2520, 3000, 3000, 3600, 3160, 3560, 3320, 2880, 2600, 3400, 2540. 对婴儿体重的方差进行区间估计 (α=0.05)

解 计算可得$(n-1)s^2 \approx 1549467$,

$\alpha=0.05$, $n-1=11$, 查附表五, 得$a=3.82$, $b=21.9$, a,b满足

$$P(\chi^2 \geq a) = 1 - \frac{\alpha}{2} = 0.975$$

$$P(\chi^2 \geq b) = \frac{\alpha}{2} = 0.025$$

因此得σ^2的置信区间为

$$\frac{1549467}{21.9} < \sigma^2 < \frac{1549467}{3.82}$$

即 $70752 < \sigma^2 < 405620$