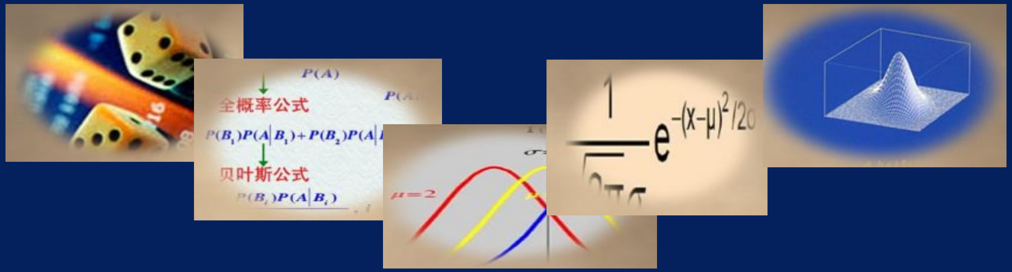
第七章　参数估计

7.4 区间估计（1）
基本概念



杨建芳

区间估计概念

假如 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是未知参数θ的一个点估计，那么一旦获得样本值$x_1, x_2, \cdots, x_n$，估计值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 就给出了一个确定的数，这个数给我们一个关于该参数的明确的数量概念．但是，我们必须注意到，点估计值只是θ 的一个近似值，它本身并没有反映这种近似值的精度，也就是说它并没有给出近似值的误差范围．

更进一步的，在数理统计学中仅仅知道误差范围 $(\hat{\theta}-\delta, \hat{\theta}+\delta)$ 也是不够的，由于样本的随机性，这个误差范围是一个随机区间，于是就连它是否包含θ的真值都成了疑问．因此，我们还必须建立一种统计推断的方法，希望通过它能确定这个区间包含θ真值的概率．

区间估计概念

为了弥补点估计的不足,本节讨论区间估计的概念. 区间估计是一种重要的统计推断方法, 它是由奈曼(J. Neyman)在1934年开始的一系列工作中引入的,这种思想从确立之日起就引起了众多统计学家的重视.

点估计是用一个点(即一个数)去估计未知参数;

区间估计就是用一个区间去估计未知参数.

区间估计概念

精度和可靠度

点估计：$\hat{\mu}=\bar{X},\hat{\sigma}^2=S^2$

误差不确定？

如何确定 a,b 的值？

考虑用区间 (a,b) 来估计未知参数

a,b关于样本的函数

其中 $a=a(X_1,X_2,\dots,X_n),b=b(X_1,X_2,\dots,X_n)$.

例：估计大学二年级学生的年龄

19岁（点估计）

（18,20）区间估计

（17,21）区间估计

(18,20)比（17,21）具有更好精度，而（17,21）比（18,20）具有更好的可靠度

区间估计概念

引例　正态总体 $X \sim N(\mu,\sigma^2)$，设 $X_1, X_2, \ldots, X_n$ 来自总体的样本。

$$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2$$

考虑用样本均值估计总体均值

$$|\bar{X} - \mu| < d$$

精度　可靠度

$$P\{\mu \in (\bar{X} - d, \bar{X} + d)\} = \underline{1-\alpha} \quad (d > 0)$$

$$P\{\sigma^2 \in (\frac{S^2}{B}, \frac{S^2}{A})\} = 1-\alpha \quad (0 < A < 1 < B)$$

区间估计概念

定义 设θ是总体的一个未知参数,给定α,$0<\alpha<1$,确定两个统计量

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

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

则称区间$(\hat{\theta}_1,\hat{\theta}_2)$为$\theta$的**置信水平（置信度）**$1-\alpha$的**置信区间**,

$\hat{\theta}_1,\hat{\theta}_2$分别称为置信下限和置信上限.

区间估计概念

置信区间的含义:固定样本容量n,然后进行多次抽样,每次抽样得到一个区间,由大数定律,当抽样的次数足够多时,包含θ的真值的区间大约占$100(1-\alpha)\%$. 即我们能以概率$(1-\alpha)$的可信程度保证,由样本值代入$(\hat{\theta}_1, \hat{\theta}_2)$中所得的区间包含$\theta$的真值.

注：置信区间是随机区间

区间估计概念

寻找未知参数θ的置信区间的基本步骤：

(1) 选取一个样本的函数 $W=W(X_1,X_2,\cdots$

(2) 对于给定的置信度$1-\alpha$，确定常数a,b，使得

$$P\{a<W(X_1,\cdots,X_n;\theta)<b\}=1-\alpha$$

(3) 利用不等式变形，求得未知参数θ的

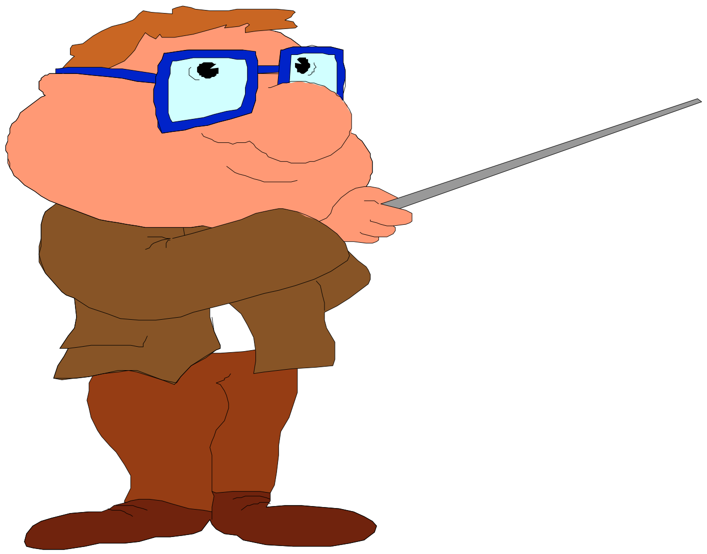
$$a<W(X_1,\cdots,X_n;\theta)<b \quad \Leftrightarrow \quad \hat{\theta}_1(X_1,\cdots,X_n)<\theta<\hat{\theta}_2(X_1,\cdots,X_n)$$

则区间$(\hat{\theta}_1,\hat{\theta}_2)$即为所求置信区间.

它只含待估参数θ，而不含其它未知参数，它的分布已知且不依赖于任何未知参数(当然也不依赖参数θ本身)；

一般利用$W=W(X_1,X_2,\cdots,X_n;\theta)$所服从分布的**分位点**来确定。

思考



1、如何理解置信区间是随机区间？

2、如何理解枢轴量的含义？枢轴量是不是统计量？

3、如何求置信区间？

第七章　参数估计

7.4 区间估计（2）

单正态总体均值的区间估计

杨建芳

区间估计基本步骤

寻找未知参数θ的置信区间的基本步骤:

(1) 选取一个样本的函数 $W=W(X_1,X_2,\cdots,X_n;\theta)$，（枢轴量）

(2) 对于给定的置信度$1-\alpha$，确定常数a,b，使得

$$P\{a<W(X_1,\cdots,X_n;\theta)<b\}=1-\alpha,$$

(3) 利用不等式变形, 求得未知参数θ的置信区间，若

$$a<W(X_1,\cdots,X_n;\theta)<b \quad \Leftrightarrow \quad \hat{\theta}_1(X_1,\cdots,X_n)<\theta<\hat{\theta}_2(X_1,\cdots,X_n)$$

则区间$(\hat{\theta}_1,\hat{\theta}_2)$即为所求置信区间.

（4）代值、计算.

单正态总体抽样分布定理

设总体 $X \sim N(\mu,\sigma^2)$，样本 $(X_1,X_2,\cdots,X_n)$，

(1) $\bar{X} \sim N(\mu,\dfrac{\sigma^2}{n})$，　$Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(2) $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

单正态总体均值的区间估计

单个正态总体$N(\mu,\sigma^2)$的情形

1. **σ^2已知时μ的区间估计**

有概率必有

随机变量

$$P\{\mu \in (\bar{X} - d, \bar{X} + d)\} = 1 - \alpha \quad (d > 0)$$

$$P\{|\bar{X} - \mu| < d\} = 1 - \alpha$$

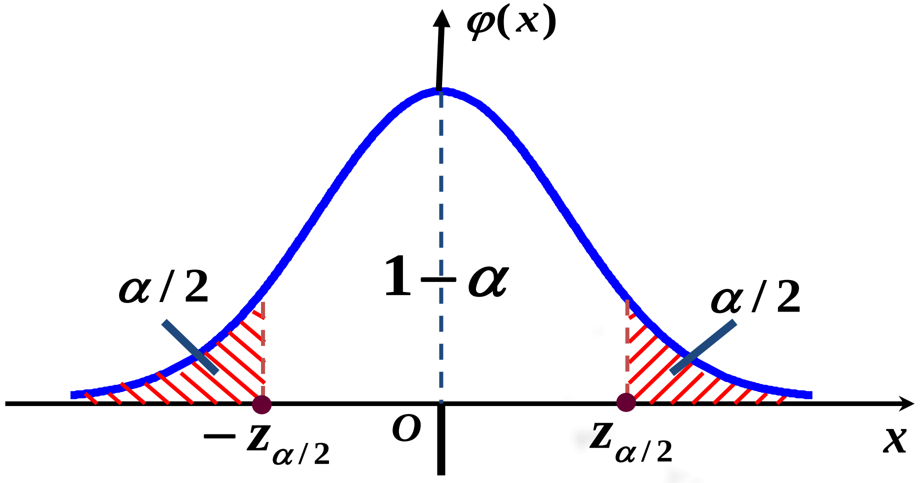
$$P\{|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| < \frac{d}{\sigma/\sqrt{n}}\} = 1 - \alpha$$

如何确定$\frac{d}{\sigma/\sqrt{n}}$?

单正态总体均值的区间估计

此时选取样本的函数为

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



对给定的置信水平 $1-\alpha$,

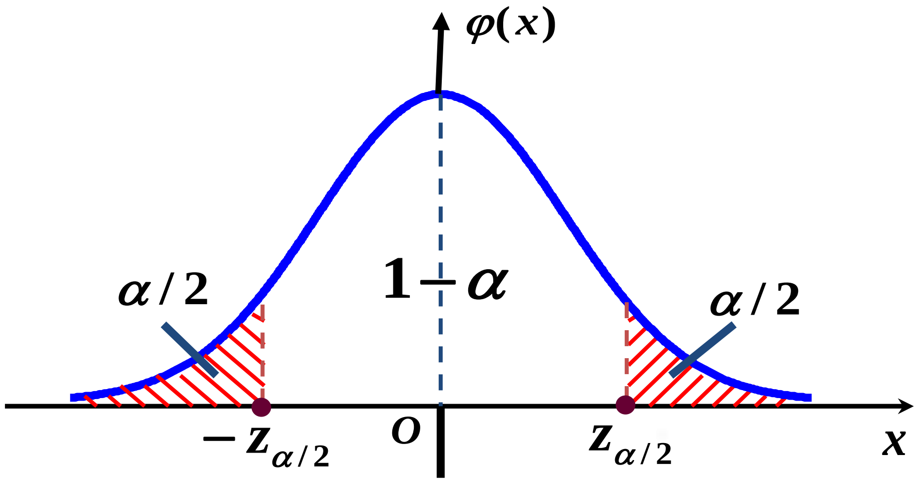
按标准正态分布的上 α 分位点的定义，得 $\dfrac{d}{\sigma/\sqrt{n}} = z_{\alpha/2}$

单正态总体均值的区间估计

基本步骤：

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

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



（3）得 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left|\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}\right| < z_{\alpha/2} \quad \Leftrightarrow \quad \bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

$\left(\bar{X} - \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{X} + \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$，简记为 $\left(\bar{X} \pm \dfrac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$.

（4）代值、计算.

单正态总体均值的区间估计

例1 某厂生产滚珠，直径 X 服从正态分布 $N(\mu,\sigma^2)$ 。为了估计 μ ，抽检 6 个滚珠，测得直径为 (mm)：14.70，15.21，14.90，14.91，15.32，15.32，

（1） 估计这批滚珠的直径均值 μ ；

（2） 若已知方差 $\sigma^2=0.05$，求 μ 的置信区间。$(\alpha=0.05)$

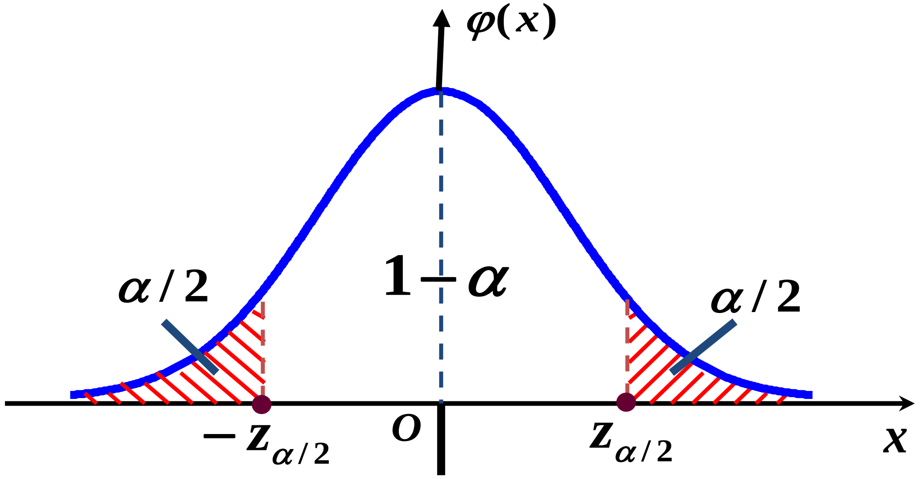
解 （1） $\overline{x}=15.06$，即为 μ 的点估计值。

单正态总体均值的区间估计

（2）置信区间

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

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



3° 置信区间为 $(\bar{X}\pm\dfrac{\sigma}{\sqrt{n}}z_{\alpha/2})$.

4° 代值计算 $\alpha=0.05$, $z_{\alpha/2}=1.96, n=6, \bar{x}=15.06, \sigma^2=0.05.$

所以 μ 的置信区间为$(15.06\pm1.96\sqrt{\dfrac{0.05}{6}})=(14.88\ ,\ 15.24)$.

单正态总体均值的区间估计

注 区间长度(即精度)为 $2z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$，希望它小一点，有两条途径：一是减小 $z_{\alpha/2}$，但这势必增大 α，从而降低区间估计的可靠性，因此一般用提高容量 $\boldsymbol{n}$ 的方法，在不降低可靠性的前提下，提高精度(缩小区间).

单正态总体均值的区间估计

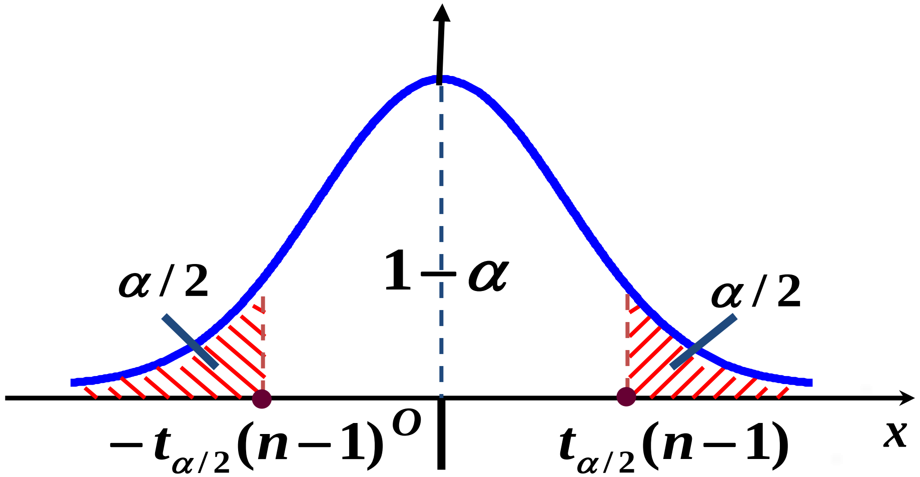
2. σ^2未知时μ的置信区间

设总体 $X \sim N(\mu, \sigma^2)$，

利用样本方差 S^2 代替 σ^2，

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

上一章已证得

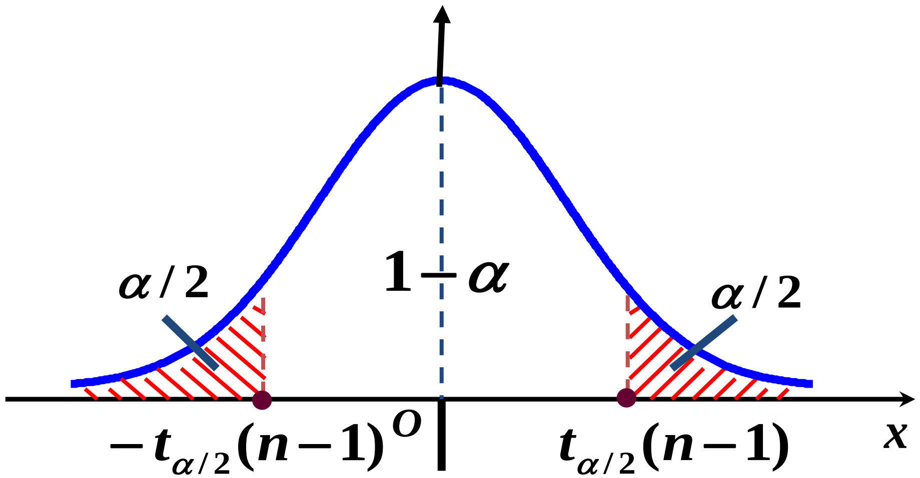


单正态总体均值的区间估计

基本步骤：

(1) $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，

(2) $P\{|\dfrac{\bar{X}-\mu}{S/\sqrt{n}}|\le t_{\alpha/2}(n-1)\}=1-\alpha$，



（3）得 μ 的置信水平为 $1-\alpha$ 的置信区间为

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

（4）代值、计算.

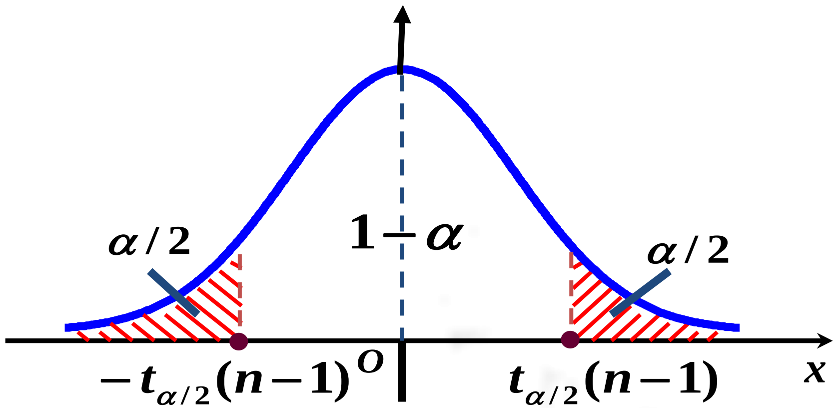
注 如果 n 较大，则用 $z_{\alpha/2}$ 代替 $t_{\alpha/2}(n-1)$。

单正态总体均值的区间估计

例2 为估计一物体的重量 μ，用天平秤了五次，得结果(g)：5.52，5.48，5.64，5.51，5.43。假定测量值是正态的，求 μ 的 95%的置信区间。

解： (1) $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，

(2) $P\{|\dfrac{\bar{X}-\mu}{S/\sqrt{n}}|\le t_{\alpha/2}(n-1)\}=1-\alpha$，



(3) 于是所求 μ 的置信区间为 $(\bar{X}\pm\dfrac{S}{\sqrt{n}}t_{\alpha/2}(n-1))$.

（4）代值、计算. $\alpha=0.05,n=5,\bar{x}=5.516,s=0.07765,t_{\alpha/2}(4)=2.7764$.

$$(\bar{X}\pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}})=(5.516\pm2.7764\times\frac{0.07765}{\sqrt{5}})=(5.481,5.550)$$

单正态总体均值的区间估计

设总体 $X \sim N(\mu,\sigma^2)$，样本 $(X_1, X_2, \cdots, X_n)$，

(1) $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$， $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(2) $T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

区间估计基本步骤

寻找未知参数θ的置信区间的基本步骤：

(1) 选取一个样本的函数 $W=W(X_1,X_2,\cdots,X_n;\theta)$，（枢轴量）

(2) 对于给定的置信度$1-\alpha$，确定常数a,b，使得

$$P\{a<W(X_1,\cdots,X_n;\theta)<b\}=1-\alpha,$$

(3) 利用不等式变形, 求得未知参数θ 的置信区间，若

$$a<W(X_1,\cdots,X_n;\theta)<b \iff \hat{\theta}_1(X_1,\cdots,X_n)<\theta<\hat{\theta}_2(X_1,\cdots,X_n)$$

则区间$(\hat{\theta}_1,\hat{\theta}_2)$ 即为所求置信区间.

（4）代值、计算.

求单正态总体均值的置信区间

1、如何选择枢轴量？枢轴量对应的分布？

2、怎么确定分位点？

3、自由度的取值？

第七章 参数估计

7.4 区间估计（3）

单正态总体方差的区间估计

杨建芳

区间估计的基本步骤

寻找未知参数θ的置信区间的基本步骤:

(1) 选取一个样本的函数 $W = W(X_1, X_2, \cdots, X_n; \theta)$，（枢轴量）

(2) 对于给定的置信度$1-\alpha$，确定常数a, b，使得

$$P\{a < W(X_1, \cdots, X_n; \theta) < b\} = 1 - \alpha ,$$

(3) 利用不等式变形, 求得未知参数θ 的置信区间，若

$$a < W(X_1, \cdots, X_n; \theta) < b \quad \Leftrightarrow \quad \hat{\theta}_1(X_1, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, \cdots, X_n)$$

则区间$(\hat{\theta}_1, \hat{\theta}_2)$ 即为所求置信区间.

（4）代值、计算.

单正态总体方差的区间估计

单个正态总体 $N(\mu,\sigma^2)$ 的情形

考虑构造关于 σ^2 的样本函数

1. μ 已知时 σ^2 的置信区间 （了解）

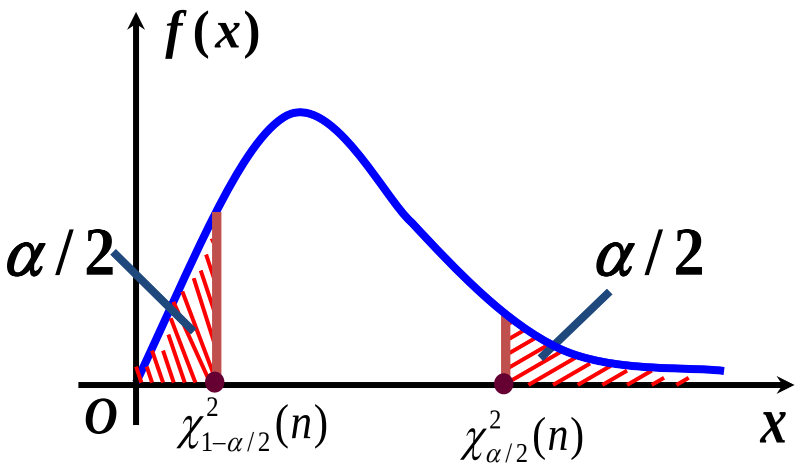
$$X_i \sim N(\mu,\sigma^2),\ \text{则}\ \frac{X_i-\mu}{\sigma} \sim N(0,1),$$

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

单正态总体方差的区间估计

基本步骤：

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



(2) $P\left\{\chi^2_{1-\alpha/2}(n) < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} < \chi^2_{\alpha/2}(n)\right\} = 1-\alpha$

（3）得σ^2的置信水平为$1-\alpha$的置信区间为$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}\right)$

（4）代值、计算.

单正态总体方差的区间估计

2. μ未知时σ^2的置信区间（更符合实际应用）

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

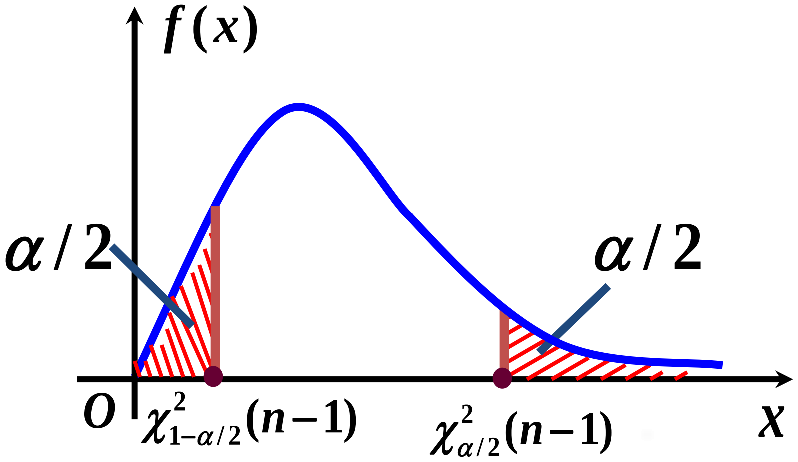
μ未知，用 $\bar{X}$ 来代替

$$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

单正态总体方差的区间估计

基本步骤：

(1) $\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$



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

（3）得 σ^2 的置信水平为 $1-\alpha$ 的置信区间为 $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

（4）代值、计算.

单正态总体方差的区间估计

例 某自动包装机包装洗衣粉,其重量服从正态分布,随机抽查12袋,测得重量分别为

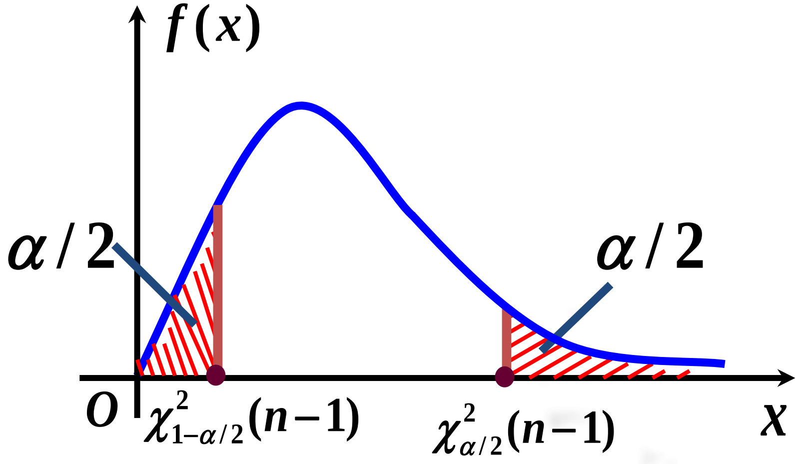
1001	1004	1003	997	999	1000
1004	1000	996	1002	998	999

如何估计该包装机所包装的洗衣粉重量的方差? $(\alpha = 0.05)$

单正态总体方差的区间估计

分析： μ 未知时 σ^2 的置信区间

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



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

（3）得 σ^2 的置信水平为 $1-\alpha$ 的置信区间为 $\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

单正态总体方差的区间估计

（4）代值、计算.

$$\chi^2_{0.975}(11)=3.816,\ \chi^2_{0.025}(11)=21.92,\ (n-1)s^2=76.25,$$

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)},\quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right)=\left(\frac{76.25}{21.92},\ \frac{76.25}{3.816}\right)=(3.48,\quad 19.98).$$

σ的0.95的置信区间为$(1.87,\ \ 4.47)$.

单正态总体方差的区间估计

设总体 $X \sim N(\mu,\sigma^2)$，样本$(X_1,X_2,\cdots,X_n)$，样本方差为S^2, 则

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

(2) μ未知时, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ （掌握）

区间估计的基本步骤

寻找未知参数θ的置信区间的基本步骤:

(1) 选取一个样本的函数 $W=W(X_1,X_2,\cdots,X_n;\theta)$, **（枢轴量）**

(2) 对于给定的置信度$1-\alpha$，确定常数a,b，使得

$$P\{a<W(X_1,\cdots,X_n;\theta)<b\}=1-\alpha\,,$$

(3) 利用不等式变形, 求得未知参数θ 的置信区间，若

$$a<W(X_1,\cdots,X_n;\theta)<b \quad\Leftrightarrow\quad \hat{\theta}_1(X_1,\cdots,X_n)<\theta<\hat{\theta}_2(X_1,\cdots,X_n)$$

则区间$(\hat{\theta}_1,\hat{\theta}_2)$ 即为所求置信区间.

（4）代值、计算.

单正态总体方差的区间估计

1、如何选择枢轴量？枢轴量的分布？

2、怎么确定分位点？

3、自由度的取值？

第七章　参数估计

7.4 区间估计（4）

双正态总体参数的区间估计

杨建芳

两个正态总体均值的区间估计

两个正态总体 $N(\mu_1,\sigma_1^2)$ **与** $N(\mu_2,\sigma_2^2)$ **的情形**

下面介绍两个正态总体**均值差**和**方差比**的区间估计：

设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$相互独立，

分别抽取样本$(X_1,X_2,\cdots,X_{n_1})$和$(Y_1,Y_2,\cdots,Y_{n_2})$

各自的样本均值和样本方差分别记为$\bar{X},\bar{Y}, S_1^2,S_2^2$.

基本思想： 用$\bar{X}-\bar{Y}$估计$\mu_1-\mu_2$, 用$\dfrac{S_1^2}{S_2^2}$估计$\dfrac{\sigma_1^2}{\sigma_2^2}$.

两个正态总体抽样分布定理

设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$ 相互独立， 分别抽取样本 $(X_1,X_2,\cdots,X_{n_1})$ 和 $(Y_1,Y_2,\cdots,Y_{n_2})$，各自的样本均值和样本方差分别记为 $\bar{X},\bar{Y},S_1^2,S_2^2$，则

$$(1)\quad Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$$

$$(2)\ T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中 $S_\omega^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

$$(3)\ F=\frac{S_1^2}{S_2^2}\Big/\frac{\sigma_1^2}{\sigma_2^2}\sim F(n_1-1,n_2-1).$$

两个正态总体均值的区间估计

1. σ_1^2 与 σ_2^2 已知时 $\mu_1-\mu_2$ 的置信区间

基本步骤：

标准化

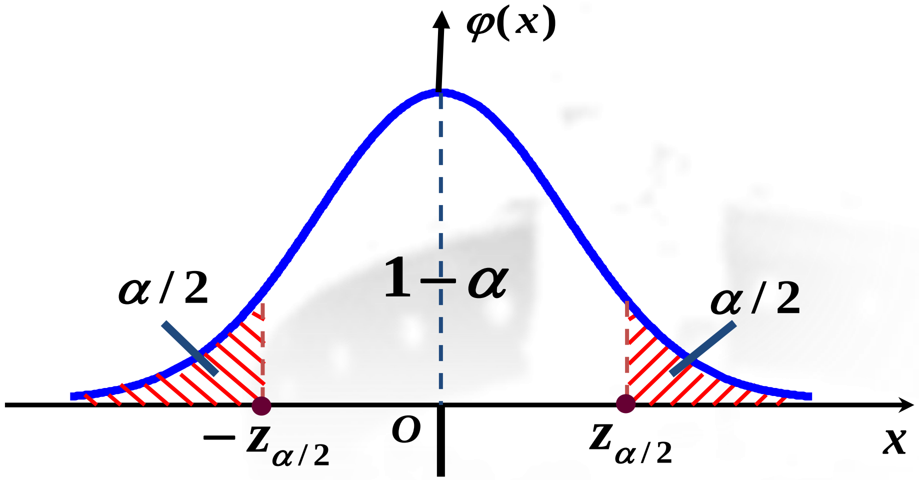
(1) $Z=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$ （枢轴量）

(2) $P\{|\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}|\le z_{\alpha/2}\}=1-\alpha$，

(3) 得 $\mu_1-\mu_2$ 的置信水平为 $1-\alpha$ 的置信区间

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

（4）代值、计算.



两个正态总体均值的区间估计

2. $\sigma_1^2=\sigma_2^2=\sigma^2$ **未知时** $\mu_1-\mu_2$ **的置信区间**

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Longrightarrow \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

总体方差未知，用联合样本方差替代

两个正态总体均值的区间估计

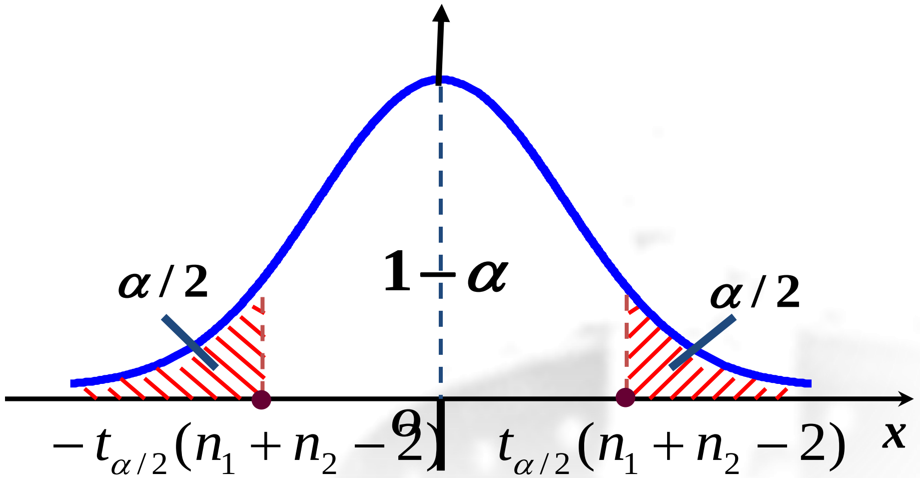
基本步骤：

(1) $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$ （枢轴量） $S_\omega^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

(2) $P\{|T|\le t_{\alpha/2}(n_1+n_2-2)\}=1-\alpha$，

(3) 得 $\mu_1-\mu_2$ 的置信水平为 $1-\alpha$ 的置信区间

$$\left(\bar{X}-\bar{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_\omega\sqrt{1/n_1+1/n_2}\right)$$



（4）代值、计算.

两个正态总体均值的区间估计

例1 某工厂一条生产灯泡的流水线, 在工艺改变前后分别抽检若干件产品的寿命, 得数据为

改变前：$n_1 = 6, \overline{x} = 1364, s_1^2 = 156$； 改变后：$n_2 = 9, \overline{y} = 1407, s_2^2 = 172$.

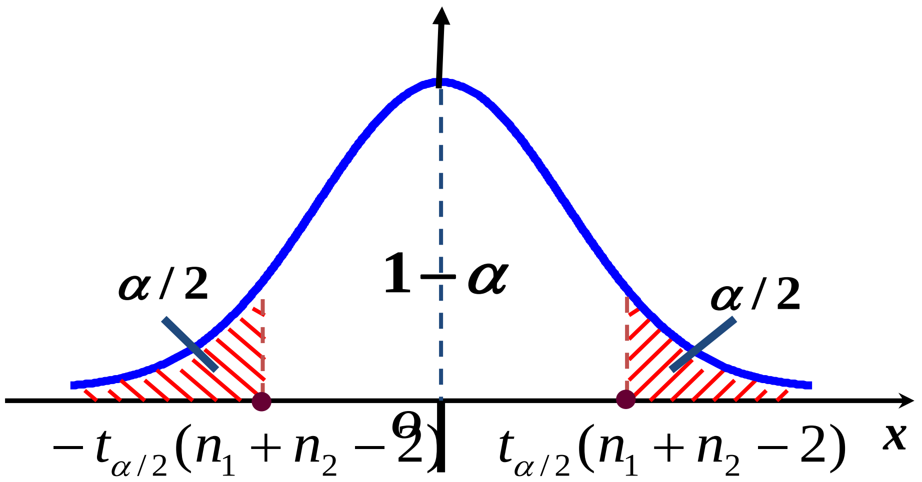
假定灯泡寿命服从正态分布, 且工艺改变前后方差不变, 试求工艺改变前后平均寿命之差 $\mu_2 - \mu_1$的95%的置信区间.

解 (1) $T = \dfrac{(\overline{Y} - \overline{X}) - (\mu_2 - \mu_1)}{S_\omega \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2)$

(2) $P\{|T| \le t_{\alpha/2}(n_1+n_2-2)\} = 1-\alpha$,

（3）得 $\mu_2-\mu_1$ 的置信水平为 $1-\alpha$ 的置信区间

$$\left(\bar{Y}-\bar{X} \pm t_{\alpha/2}(n_1+n_2-2)S_\omega\sqrt{1/n_1+1/n_2}\right)$$



（4）代值、计算.

$$s_\omega = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{5\times156+8\times172}{13}} = 12.88,\ \alpha=0.05,\ t_{0.025}(13)=2.16,$$

$$\bar{y}-\bar{x}=43,\ 2.16\times12.88\times\sqrt{1/6+1/9}=14.7,$$

所以 $\mu_2-\mu_1$ 的置信区间为

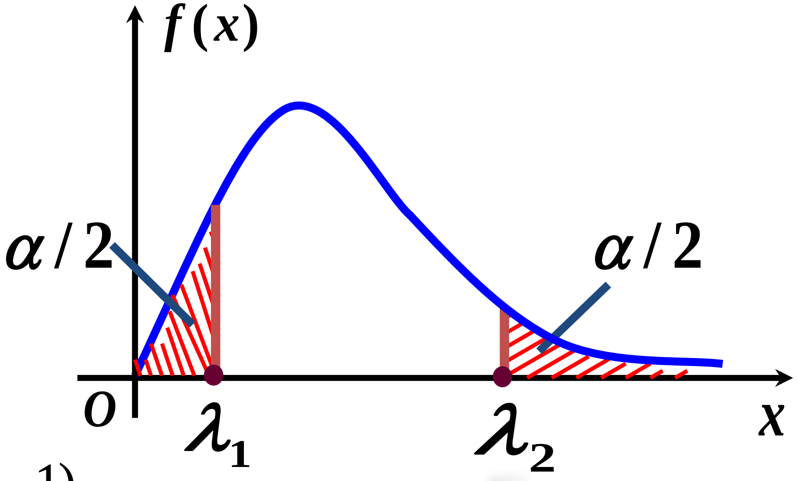
$$\left(\bar{y}-\bar{x}\pm t_{\alpha/2}(13)s_\omega\sqrt{1/n_1+1/n_2}\right)=(43-14.7,\ 43+14.7)=(28.3,\ 57.3).$$

两个正态总体方差的区间估计

3. 两个总体方差比 σ_1^2/σ_2^2 的置信区间

基本步骤：

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



(2) $P\{\lambda_1<F<\lambda_2\}=1-\alpha$

$$\lambda_1=F_{1-\alpha/2}(n_1-1,n_2-1)=\frac{1}{F_{\alpha/2}(n_2-1,n_1-1)},\lambda_2=F_{\alpha/2}(n_1-1,n_2-1),$$

(3) 所以 σ_1^2/σ_2^2 的置信水平为 $1-\alpha$ 的置信区间为

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

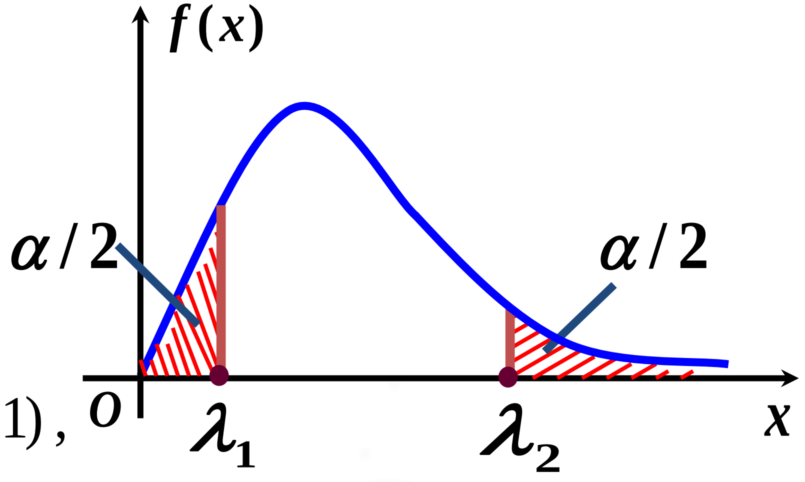
（4）代值、计算.

例2 两个相互独立的正态总体 $X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$，各取样本, $n_1 = 25$, $s_1^2 = 12.7$, $n_2 = 20$, $s_2^2 = 10.8$, 求 σ_1^2/σ_2^2 的 95%的置信区间.

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

(2) $P\{\lambda_1 < F < \lambda_2\} = 1-\alpha$

$\lambda_1 = F_{1-\alpha/2}(n_1-1,n_2-1) = \dfrac{1}{F_{\alpha/2}(n_2-1,n_1-1)}$, $\lambda_2 = F_{\alpha/2}(n_1-1,n_2-1)$,



(3) 所以 σ_1^2/σ_2^2 的置信水平为 $1-\alpha$ 的置信区间为

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

（4）代值、计算，得 $\left(\dfrac{1.176}{2.4523},\ 1.176\times 2.3273\right) = (0.347，2.737)$.

两个正态总体抽样分布定理

设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 相互独立，分别抽取样本 $(X_1, X_2, \cdots, X_{n_1})$ 和 $(Y_1, Y_2, \cdots, Y_{n_2})$，各自的样本均值和样本方差分别记为 $\bar{X}, \bar{Y}, S_1^2, S_2^2$，则

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

(2) $$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中 $S_\omega^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

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

区间估计的基本步骤

寻找未知参数θ的置信区间的基本步骤:

(1) 选取一个样本的函数 $W = W(X_1, X_2, \cdots, X_n; \theta)$，（枢轴量）

(2) 对于给定的置信度$1-\alpha$，确定常数a, b，使得

$$P\{a < W(X_1, \cdots, X_n; \theta) < b\} = 1 - \alpha,$$

(3) 利用不等式变形，求得未知参数θ 的置信区间，若

$$a < W(X_1, \cdots, X_n; \theta) < b \quad \Leftrightarrow \quad \hat{\theta}_1(X_1, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, \cdots, X_n)$$

则区间$(\hat{\theta}_1, \hat{\theta}_2)$ 即为所求置信区间.

（4）代值、计算.

双正态总体参数的区间估计

1、如何选择枢轴量？枢轴量的分布？

2、怎么确定分位点？

3、自由度的取值？

第七章 参数估计

7.7 单侧置信区间

杨建芳

正态总体单侧置信区间

单侧置信区间定义

设θ是总体的一个未知参数，给定$\alpha, 0<\alpha<1$,

$X_1, X_2, \ldots, X_n$来自总体的样本，确定统计量

$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$，使 $P(\theta > \hat{\theta}_1) = 1-\alpha$

$(\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$，使 $P(\theta < \hat{\theta}_2) = 1-\alpha)$

则称$\hat{\theta}_1$为参数θ的置信度为$1-\alpha$的单侧置信下限

（则称$\hat{\theta}_2$为参数θ的置信度为$1-\alpha$的单侧置信上限.)

正态总体单侧置信区间

$$N(\mu,\sigma^2)\quad \begin{cases} Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1) \\ t=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1) \\ \chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1) \end{cases}$$

注意到枢轴量都是待估参数的递减函数，因此求待估参数的下限，对于枢轴量而言便是上限，反之亦然

正态总体单侧置信区间

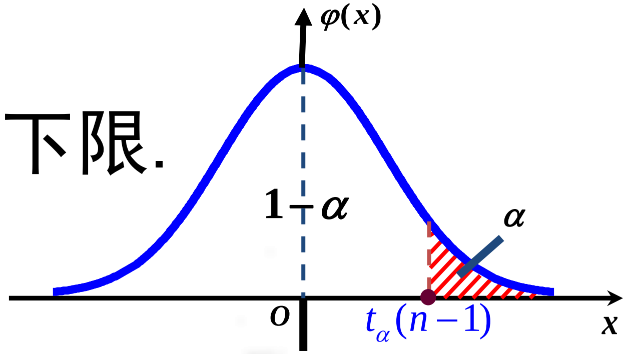
$$N(\mu_1,\sigma_1^2)\quad N(\mu_2,\sigma_2^2)\quad \begin{cases} Z=\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1) \\ t=\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_\omega\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2) \\ F=\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1) \end{cases}$$

注意到枢轴量都是待估参数的递减函数，因此求待估参数的下限，对于枢轴量而言便是上限，反之亦然

正态总体单侧置信区间

例1 已知某厂生产的滚珠的直径 X 可以认为是服从正态分布的. 今从某天生产的滚珠中随机抽取6个, 经计算测得直径样本均值 $\bar{x}=14.95\ \mathrm{mm}$, 样本标准差 $s=0.206\ \mathrm{mm}$ 。

试求：滚珠的平均直径的置信水平为0. 95的单侧置信下限.



解：（1）选择枢轴量 $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

（2）对于给定的置信水平$1-\alpha$, 有：$P\left\{\dfrac{\overline{X}-\mu}{S/\sqrt{n}} < t_\alpha(n-1)\right\}=1-\alpha$

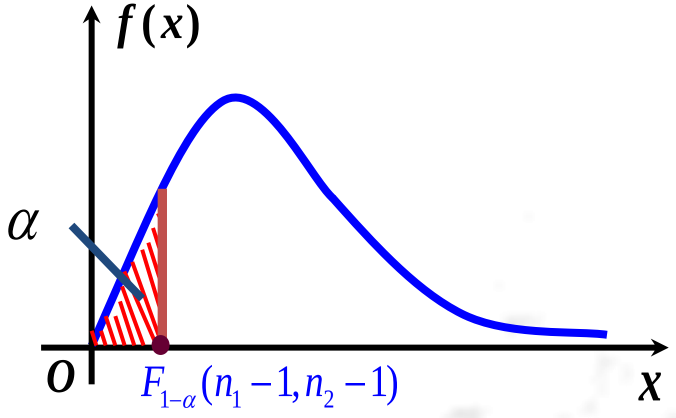
（3）因此求得μ的置信水平$1-\alpha$的单侧置信下限：$\underline{\mu}=\overline{X}-\dfrac{S}{\sqrt{n}}t_\alpha(n-1)$

(4) 代值、计算. $\underline{\mu}=\bar{x}-\dfrac{s}{\sqrt{n}}t_\alpha(n-1)=14.95-\dfrac{0.206}{\sqrt{6}}t_{0.05}(5)=14.95-\dfrac{0.206}{\sqrt{6}}\times 2.0150=14.781$

正态总体单侧置信区间

例 2 两个相互独立的正态总体 $X \sim N(\mu_1,\sigma^2)$，$Y \sim N(\mu_2,\sigma^2)$，各取样本，$n_1=20$，$s_1^2=12.7$，$n_2=25$，$s_2^2=10.8$，求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的95%的单侧置信上限.

解：(1) 选择枢轴量 $\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$



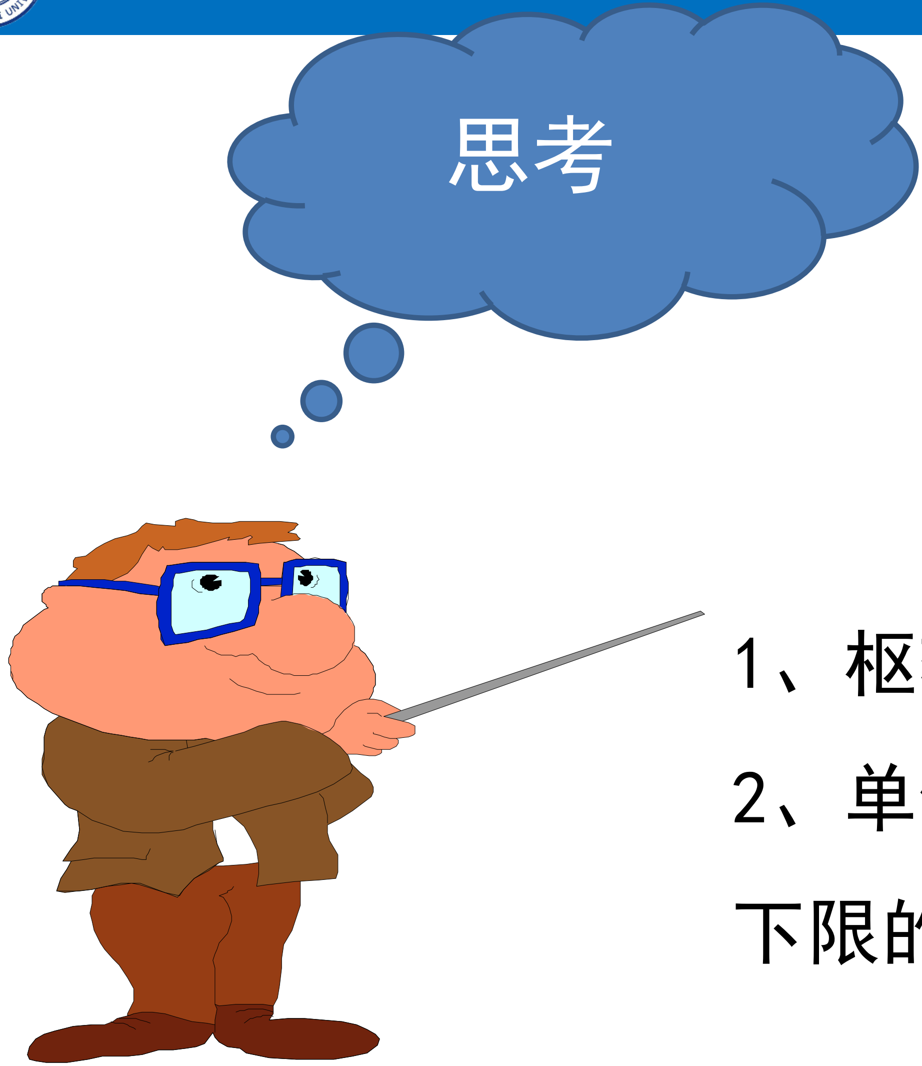
(2) 对于给定的置信水平$1-\alpha$，有：

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

(3) 因此求得σ_1^2/σ_2^2的置信水平$1-\alpha$的单侧置信上限：

$$\overline{\sigma_1^2/\sigma_2^2}=\frac{S_1^2/S_2^2}{F_{1-\alpha}(n_1-1,n_2-1)}=\frac{S_1^2}{S_2^2}F_\alpha(n_2-1,n_1-1)$$

(4) 代值、计算. $\overline{\sigma_1^2/\sigma_2^2}=\dfrac{s_1^2}{s_2^2}F_\alpha(n_2-1,n_1-1)=\dfrac{12.7}{10.8}F_{0.05}(24,19)=\dfrac{12.7}{10.8}\times 2.11\approx 2.48$



1、枢轴量与待估参数的关系？

2、单侧置信下限（上限）与双侧置信区间上下限的区别？

第七章 参数估计

小结

杨建芳

- 参数估计
  - 点估计法
    - 矩估计
    - 最大似然估计
  - 估计量评选标准
    - 无偏性
    - 有效性
    - 相合性
  - 区间估计
    - 双侧置信区间
    - 单侧置信区间

矩估计法(The Method of Moments)

基本原理： 总体矩=样本矩，令 $\mu_l = A_l$

基本步骤：

步骤1：求出总体原点矩 $u_l = E(X^l)$（含有待估未知参数）

步骤2：列方程或者方程组，令总 … 矩

$$u_l = A_l, l = 1,2,\dots,k$$

步骤3：解方程或者方程组，求出矩估计量

$$\hat{\theta}_l = \theta_l(A_1, A_2, \dots A_k), l = 1,2,\dots,k$$

注意：
选择原点矩；
几个未知参数选择几个矩；
从低阶到高阶选择矩。

k为未知参数的个数

只有一个未知参数时，
则令 $E(X) = X$

最大似然估计法（极大似然估计法）

基本原理：根据样本值来确定参数，使该

把样本联合概率分布律(或联合密度)中自变量看成已知常数, 而把参数 θ看作自变量

基本步骤：

（1） 由总体分布导出样本的联合分布律 （或

对数似然函数

似然函数 $L(\theta) = L(x_1, x_2, \ldots, x_n; \theta)$

(2) 求似然函数$L(\theta)$ 的最大值点(常常转化为求$\ln L(\theta)$的最大值点)，即θ的MLE：

一个未知参数：求导；

多个未知参数：求偏导.

$$令 \frac{\mathrm{d}\ln(L(\theta))}{\mathrm{d}\theta} = 0$$

(3) 解方程求出驻点，即为 θ的最大似然估计值 $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots x_n)$，

称 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots X_n)$ 为 θ 的最大似然估计量。

估计量的评选标准

一、无偏性 $E(\hat{\theta}) = \theta$

二、有效性 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$

三、一致性 $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta$

区间估计基本步骤

寻找未知参数θ的置信区间的基本步

(1) 选取一个样本的函数 $W = W(X$

(2) 对于给定的置信度$1-\alpha$，确定常

$$P\{a < W(X_1,\cdots,X_n;\theta) < b\} = 1-\alpha .$$

(3) 利用不等式变形，求得未知参数θ 的

$$a < W(X_1,\cdots,X_n;\theta) < b \quad \Leftrightarrow \quad \hat{\theta}_1(X_1,\cdots,X$$

则区间$(\hat{\theta}_1, \hat{\theta}_2)$ 即为所求置信区间.

（4）代值、计算.

它只含待估参数θ，而不含其它未知参数，它的分布已知且不依赖于任何未知参数（当然也不依赖参数θ本身）；

一般利用$W = W(X_1, X_2, \cdots, X_n; \theta)$

所服从分布的分位点来确定。

枢轴量的选取

$$N(\mu,\sigma^2)\quad \begin{cases} Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1) \\ T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1) \\ \chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim \chi^2(n-1) \end{cases}$$

枢轴量的选取

$$N(\mu_1,\sigma_1^2)$$
$$N(\mu_2,\sigma_2^2)$$

$$\begin{cases} Z = \dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1) \\ T = \dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_\omega\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2) \\ F = \dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1,n_2-1) \end{cases}$$

注意：

① 对于正态总体下枢轴量是关于待估参数的单调递减函数， 因此求待估参数的单侧置信下限，则意味着求枢轴量的上限，反之亦然.

② 双侧置信区间是去掉两端的小概率事件，左右两侧各取$\alpha/2$，因此对应分别为$\alpha/2$ 和$1-\alpha/2$分位点（正态分布和t 分布对应的是$\alpha/2$ 分位点与其相反数），对于单侧置信区间是去掉一端小概率事件，因此对应的是α 或者$1-\alpha$ 分位点.

$$z_{1-\alpha}=-z_{\alpha} \quad t_{1-\alpha}(n)=-t_{\alpha}(n) \quad F_{1-\alpha}(n_1,n_2)=\frac{1}{F_{\alpha}(n_2,n_1)}$$

③ 注意分布记号和分位点的区别

随机变量	分布	分位点（数）
X	$N(0,1)$	z_α
χ^2	$\chi^2(n)$	$\chi_\alpha^2(n)$
t	$t(n)$	$t_\alpha(n)$
F	$F(n_1,n_2)$	$F_\alpha(n_1,n_2)$

注意分位点
是一个实数

④ 注意各个分布的自由度.

谢谢大家！