第四、五节 区间估计

问题的引出

在参数的点估计中用样本构造一个估计量 $\hat{\theta}$ ，用 $\hat{\theta}$ 去估计 θ ，这仅仅是解决了一个求未知参数 θ 的一个 “近似值” 问题，而没有解决“近似值”的精确程度问题，即没有给出这个近似值的误差范围和估计的可信程度。

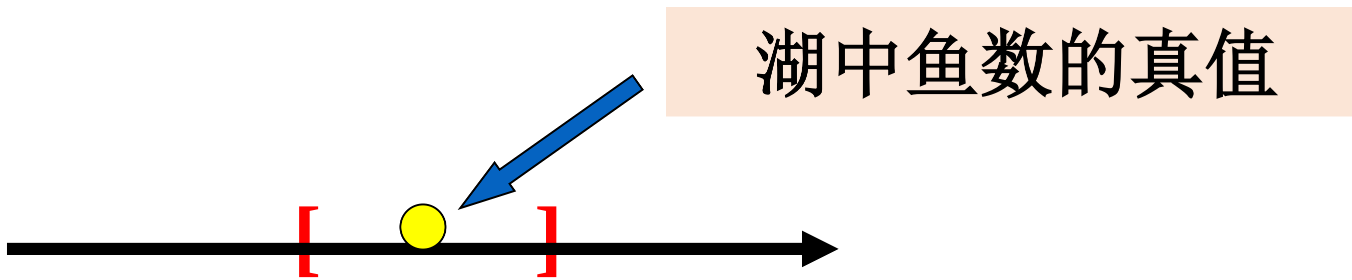
在参数的区间估计中则要用样本去给出未知参数 θ 的一个大致的范围，并使未知参数 θ 在其中有指定的概率。

具体: 若估计参数为 θ ，要考虑估计量 $\hat{\theta}$ 落在 $\underline{\theta} < \hat{\theta} < \overline{\theta}$ 的可能性有多大。

$$\text{即求 } P(\underline{\theta} < \hat{\theta} < \overline{\theta}) = ?$$

例如: 在估计湖中鱼数的问题中，若已知得到鱼数 N 的极大似然估计为1000条。而实际上 N 的真值可能大于1000条，也可能小于1000条。则在区间估计中就可以给出一个区间，在此区间内合理地相信 N 的真值位于其中（这样对鱼数的估计就更有把握）。

➢ 也就是说，所讨论的问题是希望确定一个区间，使得在该区间内能以比较高的可靠程度相信它包含未知参数的真值。



➢ 而这“可靠程度”是用概率来度量的，称为置信概率、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。称该区间为置信区间。

一. 置信区间

定义: 设总体 X 的分布函数 $F(x,\theta)$ 含有一个未知参数 θ，对于给定的值 $\alpha\ (0<\alpha<1)$，若由样本 $X_1, X_2 \cdots X_n$ 确定的两个统计量:

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

满足: $P(\underline{\theta} < \theta < \overline{\theta}) = 1-\alpha$ ，则称随机区间 $(\underline{\theta}, \overline{\theta})$ 是 θ 的置信度为 $1-\alpha$ 的置信区间,

$\underline{\theta}$和 $\overline{\theta}$ 为置信度为$1-\alpha$ 的双侧置信区间的置信下限和置信上限。$1-\alpha$ 称为置信水平。

注: ▲ 定义的含义是指:在反复抽样多次(各次得到的样本容量相等，均为n)，每次样本值确定一个区间 $(\underline{\theta}, \overline{\theta})$ ，每个这样的区间要么包含 θ 的真值，要么不包含 θ 的真值，按伯努利大数定理可知：在这么多的区间中包含θ真值的约占$100(1-\alpha)\%$不包含θ真值的仅占$100\alpha\%$

注:

▲对置信区间 $(\underline{\theta}, \bar{\theta})$ 有两个要求：一是要求 θ 以很大的可能被包含在该区间内，即要求：

$P\{\underline{\theta} < \theta < \bar{\theta}\}$ 尽可能大。二是要求估计的精度要尽可能的高，即要求区间 $\bar{\theta} - \underline{\theta}$ 尽可能短。

可靠度与精度是一对矛盾，一般是在保证可靠度的条件下尽可能提高精度。

一般地，寻求未知参数 θ 的置信区间的具体做法如下：

（1）寻求一个样本 $X_1, X_2, \cdots, X_n$ 和 θ 的函数 $W = W(X_1, X_2, \cdots, X_n; \theta)$，使得$W$的分布不依赖于 θ 以及其他未知参数，称具有这种性质的函数W为枢轴量。

（2）对于给定的置信水平 $1-\alpha$ ，定出两个常数 a, b使得 $P\{a < W(X_1, X_2, \cdots, X_n; \theta) < b\} = 1-\alpha$

若能从 $a < W(X_1, X_2, \cdots, X_n; \theta) < b$ 得到与之等价的θ 的不等式 $\underline{\theta} < \theta < \overline{\theta}$ ，其中

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

是两个统计量。那么 $(\underline{\theta}, \overline{\theta})$ 就是 θ 的一个置信水平为 $1-\alpha$ 的置信区间。

注：

枢轴量的构造，通常可以从 θ 的点估计着手考虑。

区间估计的方法步骤

1. **构建枢轴量$W(\theta,\hat{\theta}(X_1,X_2,\ldots,X_n))$：** 可从待估参数$\theta$的一个点估计量$\hat{\theta}$出发，构建关于$\theta$和$\hat{\theta}$的函数$W(\theta,\hat{\theta})$，且$W(\theta,\hat{\theta})\sim \mathrm{F(x)}$为已知不含未知参数的分布（此处综合考虑点估计量和第六章各抽样分布）。

2. **构造区间：** 在F(x)分布中，构造区间$P\left(a < W(\theta,\hat{\theta}) < b\right) = 1-\alpha$（置信水平）。

3. **求解区间：** 根据a< $W(\theta,\hat{\theta}) < b$，即得出取值范围$\underline{\theta} < \theta < \overline{\theta}$，即为其置信水平为$1-\alpha$的置信区间。

抽样分布回顾

设 $X_1, X_2 \cdots X_n$ 是总体$N(\mu,\sigma^2)$ 的一个样本，

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

定理 2. 设 $X_1, X_2, \ldots, X_n$ 是取自正态总体 $N(\mu,\sigma^2)$ 的样本，$\bar{X}$ 和 S^2 分别为样本均值和样本方差，

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

证明：$\because \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

抽样分布回顾

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

(2) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时，

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_W\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

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

讨论的问题： 给定置信度 $1-\alpha$ ，求置信区间。

讨论的对象： 正态随机变量情形的区间估计。

二. 正态总体均值的区间估计

1. 单个正态总体 $N(\mu,\sigma^2)$ 情形

问题：设 $X_1,\dots X_n$ 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: (1). 当方差 σ^2 已知的情形

选 μ 的点估计(无偏估计)为 $\bar{X}$

寻找未知参数的一个良好估计

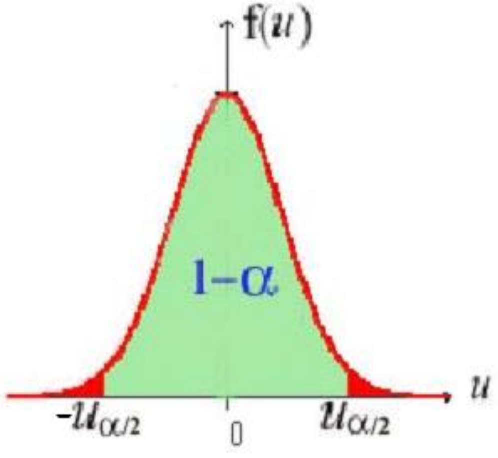
且枢轴量 $U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，而且

U 不依赖于任何未知参数。

- 现对于给定的置信水平 $1-\alpha$ (大概率)， 根据 U 的分布，确定一个区间，使得 U 取值于该区间的概率为 $1-\alpha$

- 故对于给定的置信水平，按照标准正态分布的分位点的定义有：

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



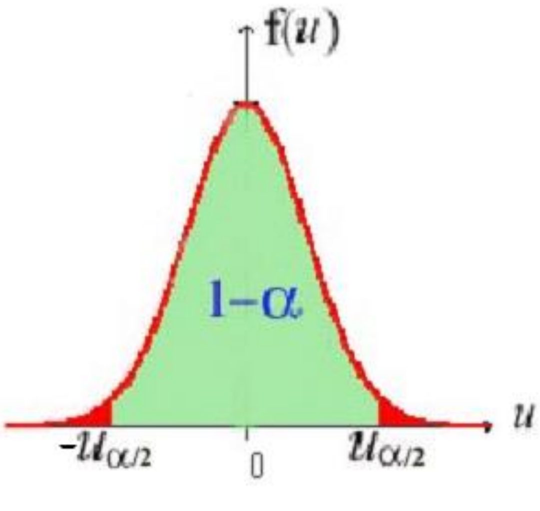
- 从中解得：

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

于是所求 μ 的置信度为$1-\alpha$ 置信区间为：

$$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\quad \bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right)$$

也可简记为： $\bar{X}\pm\frac{\sigma}{\sqrt{n}}u_{\alpha/2}$



例1. 某实验室测量铝的比重 16 次，得平均值 $\bar{X}=2.705$ ，设总体 $X\sim N(\mu, 0.029^2)$（高斯已证明测量误差是服从正态分布）

求: μ 的 95% 的置信区间.

例1. 某实验室测量铝的比重 16 次，得平均值

$\bar{X} = 2.705$ ，设总体 $X \sim N(\mu, 0.029^2)$

（高斯已证明测量误差是服从正态分布）

求: μ 的 95% 的置置信区间.

解: 由已知： $\because 1-\alpha = 95\% \quad \therefore \alpha = 5\%$,

查正态分布表得： $u_{\alpha/2} = u_{\frac{0.05}{2}} = u_{0.025}$

$$= u(1-0.025) = 1.96$$

得： $\dfrac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} = \dfrac{0.029}{\sqrt{16}} \times 1.96 = 0.014$

从而 μ 的 95% 的置信区间为：

$$(2.705 - 0.014,\quad 2.705 + 0.014) = (2.691,\ 2.719)$$

即用 $\overline{X} = 2.705$ 来估计 μ 值的可靠程度达到 95% 的区间范围是 **(2.691, 2.719)**

(2). 方差 σ^2 未知的情形

σ^2 未知，但考虑到样本方差是 σ^2 的无偏估计，

$\therefore$ 用 S^2 去代替 σ^2 得枢轴量：

(2). 方差σ^2未知的情形

σ^2 未知，但考虑到样本方差是 σ^2的无偏估计，

$\therefore$ 用 S^2 去代替 σ^2 得枢轴量：

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

它是不依赖于任何未知参数的。

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

从中解得：

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

于是所求 μ 的置信度为$1-\alpha$ 置信区间为：

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

例2. 确定某种溶液的化学浓度，现任取4个样品，测得样本均值为 $\bar{X} = 8.34\%$, $S = 0.03$

现溶液的化学浓度近似服从正态分布。

求: μ 的置信度为 95% 的置信区间

解:

例2. 确定某种溶液的化学浓度，现任取4个样品，测得样本均值为 $\bar{X} = 8.34\%$，$S = 0.03$

现溶液的化学浓度近似服从正态分布。

求: μ 的置信度为 95% 的置信区间

解: 由已知：$\because 1-\alpha = 95\%$ $\therefore \alpha = 5\%$

查 t 分布表得：$t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(3) = 3.1824$

得: $$\frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) = \frac{0.03}{\sqrt{4}} \times 3.1824 = 0.0478$$

从而 μ 的 95%的置信区间为:

$$(3.56\%，13.12\%)$$

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

问题: $X_1, X_2, \cdots X_n$ 是来自第一个总体 $N(\mu_1,\sigma_1^2)$ 的样本， $Y_1,Y_2\cdots Y_n$ 是来自第二个总体 $N(\mu_2,\sigma_2^2)$ 的样本, 它们相互独立，又设 $\bar{X}, \bar{Y}$ 分别是两个总体的样本均值; S_1^2, S_2^2 分别是两个总体的样本方差。给定置信度为 $1-\alpha$

求：两个总体均值差 $\mu_1-\mu_2$ 的置信区间。

解: (1). σ_1^2,σ_2^2 均为已知时

$\because$ $\bar{X},\bar{Y}$ 分别为 μ_1,μ_2 的无偏估计

$\therefore$ $\bar{X}-\bar{Y}$ 是 $\mu_1-\mu_2$ 的无偏估计

又由 $\bar{X},\bar{Y}$ 的独立性以及已知:

$$\bar{X}\sim N(\mu_1,\sigma_1^2/n_1),\quad \bar{Y}\sim N(\mu_2,\sigma_2^2/n_2),$$ 故有:

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

所以得枢轴量: $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$$

于是所求 $\mu_1-\mu_2$ 的置信度为 $1-\alpha$ 置信区间为:

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

(2). σ_1^2,σ_2^2 均为未知时

同单个总体在方差未知的情形下用 s^2 代替 σ^2 的构思相同，可以得到当 n_1,n_2 均很大时(一般大于50) $\mu_1-\mu_2$ 的一个置信度为 $1-\alpha$ 的近似置信区间:

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

(3). $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时

同单个总体方差未知的情形类似(又因$\sigma_1^2 = \sigma_2^2 = \sigma^2$)

得枢轴量: $$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_w \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

于是所求 $\mu_1 - \mu_2$ 的置信度为$1-\alpha$ 置信区间为:

$$\left(\bar{X} - \bar{Y} \pm t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) \cdot s_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

其中 $s_w^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$, $s_w = \sqrt{s_w^2}$

例3. 分别用金球和铂球测定引力常数(单位:$10^{-11}m^3kg^{-1}s^{-2}$)

设测定值总体为$N(\mu,\sigma^2)$, $\boxed{\mu, \sigma^2 \text{ 均为未知.}}$

(1) 用金球测定观察值为:

6.683, 6.681, 6.676, 6.678, 6.679, 6.672

(2) 用铂球测定观察值为:

6.661, 6.661, 6.667, 6.667, 6.664

1.分别就(1),(2)两种情况求μ的置信度为 0.9 的置信区间

2.若设用金球和用铂球测定时测定值总体的方差相等,

求两个测定值总体均值差的置信度为0.9的置信区间.

解: 1. 在(1)中：$\because \quad \bar{X}=\dfrac{40.069}{6}=6.678,$

$$S_1=\sqrt{\frac{1}{6-1}(0.0003)}=0.00387$$

又 $\because \quad 1-\alpha=0.9$ 　即：$\alpha=0.1$

$$\frac{S_1}{\sqrt{n}}t_{\alpha/2}(n-1)=\frac{0.00387}{\sqrt{6}}\cdot t_{\frac{0.1}{2}}(5)$$

$$=\frac{0.00387}{\sqrt{6}}\times 2.015=0.003$$

于是所求 μ 的置信度为 90% 置信区间为:

$$(\bar{X}-0.03,\ \bar{X}+0.003)=(6.675,\ 6.681)$$

用6.678作为引力常数的估计值的可靠程度为 90%的区间是(6.675，6.681)

在(2)中 $\because \overline{Y}=\dfrac{33.32}{5}=6.664$

$$S_2=\sqrt{\frac{1}{5-1}(0.000036)}=0.003$$

又$\because\ 1-\alpha=0.9,\qquad \therefore\ \alpha=0.1$

$$\frac{s_2}{\sqrt{n}}t_{\alpha/2}(n-1)=\frac{0.003}{\sqrt{5}}\cdot t_{\frac{0.1}{2}}(4)$$

$$=\frac{0.003}{2.236}\times 2.1318=0.0028$$

于是所求 μ 的置信度为 90% 置信区间为:

$$(\overline{Y}-0.028,\ \overline{Y}+0.0028)=(6.661,\ 6.667)$$

用6.664作为引力常数的估计值的可靠程度为 90%的区间是(6.661，6.667)

2. $\because$ 两个测定值总体的方差相等但又未知，

现要求 $\mu_1-\mu_2$ 的置信区间，可依公式计算:

$$\bar{X}-\bar{Y}=6.678-6.664=0.014$$

$$t_{\frac{0.1}{2}}(6+5-2)=t_{0.05}(9)=1.8331$$

$$s_w=\sqrt{\frac{(6-1)[\frac{1}{6-1}(0.0003)]+(5-1)[\frac{1}{5-1}(0.000036)]}{6+5-2}}$$

$$=\sqrt{0.0000373}=0.006$$

$$\sqrt{\frac{1}{6}+\frac{1}{5}}=\sqrt{0.36667}=0.6055$$

于是所求 $\mu_1-\mu_2$ 的置信度为 0.9 的置信区间为:

$$(\bar{X}-\bar{Y}\pm t_{\frac{\alpha}{2}}(n_1+n_2-2)\cdot s_w\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})$$

$$=(0.01,\ 0.02)$$

即用 0.014 作为 $\mu_1-\mu_2$ 的估计值的可靠程度达到 90% 的区间是(0.01, 0.02)

注: 一般，若 $\mu_1-\mu_2$ 的置信区间包含零，则可认为这两个总体的均值没有显著差别；若下限大于零，则可认为 μ_1 比 μ_2 大。

三. 正态总体方差的区间估计

1. 单个正态总体 $N(\mu,\sigma^2)$ 的情形

问题：设总体 $X \sim N(\mu,\sigma^2)$, μ, σ^2 未知。

$X_1, X_2, \cdots X_n$ 是总体 X 的一个样本，σ^2 是总体方差，给定置信度 $1-\alpha$

求:方差 σ^2 的置信区间。

三. 正态总体方差的区间估计

1. 单个正态总体 $N(\mu,\sigma^2)$ 的情形

问题: 设总体 $X \sim N(\mu,\sigma^2)$, μ, σ^2 未知。

$X_1, X_2, \cdots X_n$ 是总体 X 的一个样本，σ^2 是总体方差，给定置信度 $1-\alpha$

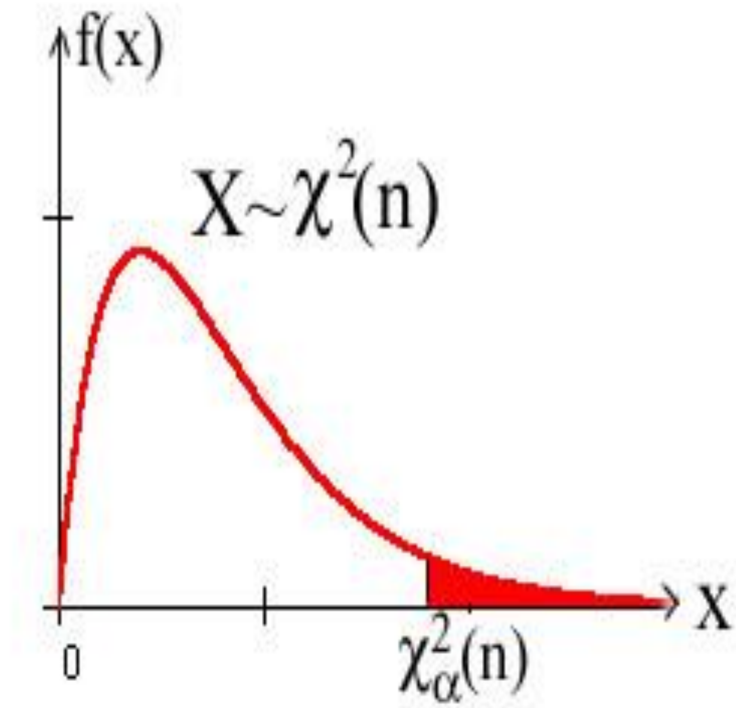
求: 方差 σ^2 的置信区间。

解: $\because$ S^2 是 σ^2 的无偏估计，且枢轴量:

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

是不依赖于任何未知参数的。

故对于给定的置信水平，按照 χ^2分布的上 α 分位点的定义有：



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

于是所求 σ^2 的置信度为 $1-\alpha$ 置信区间为:

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

标准差 σ 的一个置信度为 $1-\alpha$ 的置信区间:

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

例4. 求 例3 中的 (1)情况下，σ^2 的置信度为 0.9 的置信区间.

(1) 用金球测定观察值为:

6. 683, 6. 681, 6. 676,

6. 678, 6. 679, 6. 672

解: 在(1)中

$$(n-1)\cdot S^2=(6-1)\frac{1}{6-1}(0.0003)=0.0003$$

$$\chi^2_{\frac{0.1}{2}}(6-1)=\chi_{0.05}{}^2(5)=11.070$$

$$\chi^2_{1-\frac{0.1}{2}}(6-1)=\chi_{0.95}{}^2(5)=1.145$$

$\therefore$ σ^2 的置信度为0.9的置信区间为:

$$\left(\frac{0.0003}{11.070},\frac{0.0003}{1.145}\right)=(0.0000271,\ 0.000262)$$

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

问题：$X_1,X_2\cdots X_n$ 来自$N(\mu_1,\sigma_1^2)$的样本，$Y_1,Y_2\cdots Y_n$ 来自$N(\mu_2,\sigma_2^2)$的样本，它们相互独立。

S_1^2,S_2^2 ;μ_1,μ_2分别是两个总体的样本方差与样本均值，且样本均值均为未知，给定置信度$1-\alpha$

求：两个总体方差比σ_1^2/σ_2^2的置信区间.

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

问题: $X_1, X_2 \cdots X_n$ 来自 $N(\mu_1,\sigma_1^2)$ 的样本, $Y_1, Y_2 \cdots Y_n$ 来自 $N(\mu_2,\sigma_2^2)$ 的样本，它们相互独立。

S_1^2, S_2^2 ; μ_1, μ_2 分别是两个总体的样本方差与样本均值，且样本均值均为未知，给定置信度 $1-\alpha$

求：两个总体方差比 σ_1^2 / σ_2^2 的置信区间.

解: 解题思路同单个总体情况类似

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$$

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

且这两个统计量是相互独立的。

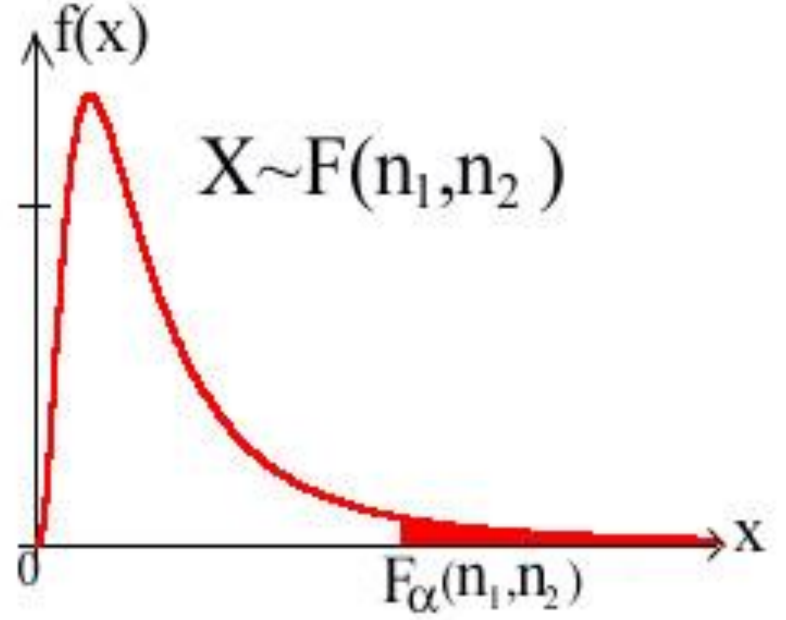
由 F 分布的定义得枢轴量:

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

是不依赖于任何未知参数的。

故对于给定的置信水平，按照 F 分布的上 α 分位点的定义有：

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



从中解得：

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

于是所求 σ_1^2 / σ_2^2 的置信度为 $1-\alpha$ 置信区间为：

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

例5. 在例3 中若测定值总体的方差分别为 σ_1^2,σ_2^2 ，求：这两个测定值总体方差比的置信度为 0.9 的置信区间.

解: $\because$ $$S_1^2 / S_2^2 = \frac{0.0003 / (6-1)}{0.000036 / (5-1)} = \frac{0.00006}{0.000009} = 6.667$$

$$\frac{1}{F_{\frac{0.1}{2}}(6-1,5-1)}=\frac{1}{F_{0.005}(5,4)}=\frac{1}{6.26}=0.159$$

$$\frac{1}{F_{1-\frac{0.1}{2}}(6-1,5-1)}=\frac{1}{F_{0.95}(5,4)}=\frac{1}{1\big/F_{0.05}(4,5)}$$

$$=F_{0.05}(4,5)=5.19$$

$\therefore$ 这两个测定值总体方差比 $\sigma_1^2\big/\sigma_2^2$ 的置信度为 0.9 的置信区间为:

$$(6.667\times 0.159,\quad 6.667\times 5.19)=(1.06,\quad 34.60)$$

注: ➢ 如果 $1 < a < \dfrac{\sigma_1^2}{\sigma_2^2}$ 即 $\sigma_1^2 > \sigma_2^2$

表明正态总体 $N(\mu_2, \sigma_2^2)$ 的波动较小

➢ 如果 $\dfrac{\sigma_1^2}{\sigma_2^2} < b < 1$ 即 $\sigma_1^2 < \sigma_2^2$

表明正态总体 $N(\mu_1, \sigma_1^2)$ 的波动较小

➢ 除了两种情况外，其它情况都不能判断出 $N(\mu_1, \sigma_1^2)$ 与 $N(\mu_2, \sigma_2^2)$ 中谁的波动性大。

➢ 但若 σ_1^2 / σ_2^2 的置信区间包含 1，则可认为 σ_1^2 与 σ_2^2 两者没有显著差别。

区间估计的方法步骤

1. **构建枢轴量$W(\theta,\hat{\theta}(X_1,X_2,\dots,X_n))$：** 可从待估参数$\theta$的一个点估计量$\hat{\theta}$出发，构建关于$\theta$和$\hat{\theta}$的函数$W(\theta,\hat{\theta})$，且$W(\theta,\hat{\theta})$~F(x)为已知不含未知参数的分布（此处综合考虑点估计量和第六章各抽样分布）。

2. **构造区间：** 在F(x)分布中，构造区间$P\left(a<W(\theta,\hat{\theta})<b\right)=1-\alpha$（置信水平）。

3. **求解区间：** 根据a< $W(\theta,\hat{\theta})<b$，即得出取值范围$\underline{\theta}<\theta<\overline{\theta}$，即为其置信水平为$1-\alpha$的置信区间。

第六节 （0-1）分布参数的区间估计

■ 这是一种特殊的离散型分布的区间估计，取自容量 $n > 50$ 的大样本。取大样本的目的是意在利用中心极限定理，使其近似服从标准正态分布。

问题: 设总体 X 服从参数为 p 的 (0-1)分布

$$X=\begin{cases}1 & \text{当事件}A\text{发生时}\\ 0 & \text{当事件}A\text{不发生时}\end{cases}$$

且 $P(X=1)=p$， $P(X=0)=1-p=q$

即 X 的分布律为:

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

其中 p 为未知参数

求： p 的置信水平为 $1-\alpha$ 的置信区间.

三. 棣莫弗---拉普拉斯定理

（De Moivere—laplace 中心极限定理）

定理3. 设随机变量 $\eta_1,\eta_2\cdots\eta_n\cdots$ 相互独立，η_n 服从参数为 n, p $(0<p<1)$ 的二项分布，则对任意 x 恒有:

$$\lim_{n\to\infty} P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

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

其中 p 为未知参数

求：p 的置信水平为 $1-\alpha$ 的置信区间.

解：由已知，

$$\because E(X)=\mu=p, \quad D(X)=\sigma^{2}=p(1-p)=pq$$

$X_1, X_2 \cdots X_n$ 是一个大样本，

所以由中心极限定理得：

$$\boxed{\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} = \frac{n\bar{X}-np}{\sqrt{np(1-p)}} \sim N(0,1)}$$

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

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

等价于:

$$(z_{\alpha/2}^2 + n)p^2 - (2n\bar{X} + z_{\alpha/2}^2)p + n\bar{X}^2 < 0$$

记: $p_1 = \dfrac{1}{2a}(-b - \sqrt{b^2 - 4ac})$

$$p_2 = \frac{1}{2a}(-b + \sqrt{b^2 - 4ac})$$

即为二次方程的两个根

其中：

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

∴ 得 p 的近似的置信度为 $1-\alpha$ 的置信区间为:

$$(p_1,\ p_2)$$

例6. 从一大批产品中任取100件产品进行检验，发现其中有 60 件是一级品。

试求：这批产品的一级品率 p 的置信度为 95% 的置信区间。

解: 由题意可知：一级品率 p 是(0-1)分布的参数

$$\because\ 1-\alpha=95\% \qquad \therefore\ \alpha=0.05$$

查表得： $z_{\frac{\alpha}{2}}=z_{0.025}=1.96$

又 $\because\ n=100$（大样本）， $\boxed{\bar{X}=\dfrac{60}{100}=0.6}$

经计算得：

$$a=n+z_{\frac{\alpha}{2}}^{2}=100+(1.96)^{2}=103.84$$

$$b=-(2n\bar{X}+z_{\frac{\alpha}{2}}^{2})$$

$$=-(2\times100\times0.6+(1.96)^{2})=-123.84$$

$$p_1 = -\frac{1}{2a}(b + \sqrt{b^2 - 4ac}) = 0.5020$$

$$p_2 = -\frac{1}{2a}(b - \sqrt{b^2 - 4ac}) = 0.6906$$

得一级品率 p 的置信度为 95% 的置信区间:

$$(0.5020,\ 0.6906)$$

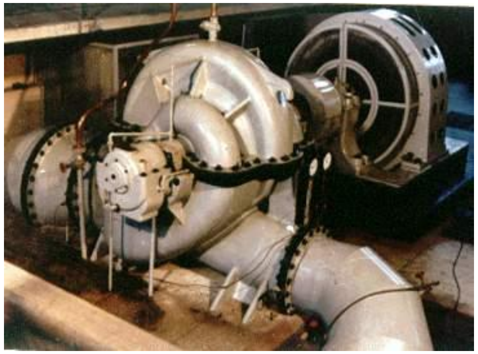
即用 $\bar{X} = 0.6$ 作为一级品率 p 的估计值的可靠程度达到 95% 的区间为 $(0.5020,\ 0.6906)$

第七节 单侧置信区间

问题的引出

前面介绍的置信区间中置信限都是双侧的，但在有些实际问题，人们所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称为单侧置信区间。

一. 单侧置信区间定义

定义: 给定 $\alpha\ (0<\alpha<1)$, 若由样本 $X_1, X_2 \cdots X_n$ 确定的 $\underline{\theta}=\underline{\theta}(X_1, X_2 \cdots X_n)$ (或$\overline{\theta}=\overline{\theta}(X_1, X_2, \cdots X_n)$)

满足: $P(\theta>\underline{\theta})=1-\alpha$ 或 $P(\theta<\overline{\theta})=1-\alpha$

则称随机区间: $(\underline{\theta}, +\infty)$ (或 $(-\infty, \overline{\theta})$) 是 θ 的置信度为 $1-\alpha$ 的单侧置信区间。

$\underline{\theta}$ 称为置信度为 $1-\alpha$ 单侧置信下限。

$\overline{\theta}$ 称为置信度为$1-\alpha$ 的单侧置信上限。

二. 单侧置信区间的求法

思路: 同双侧量区间的求法

不同处：在求单侧置信区间时不是查双侧 α 分位点，而是查单侧 α 分位点。

例7. 设有某部门对所属区域的职工家庭人均月收入进行调查，现抽取 20 个家庭，所得的月平均收入 $\bar{x} = 234.7$（元），$s^2 = 1590.85$

试以 95% 的置信度估计该区域职工家庭人均月收入的最低下限为多少？（单侧置信下限）

解: 用 μ 表示职工家庭人均月收入，X 表示该区域职工家庭月收入的总体。

现要根据所抽取的20 个家庭所得的月平均收入的数据，在方差未知的条件下求 $E(X)=\mu$ 的单侧置信下限。

由题设可知 μ 的置信度为 $1-\alpha$ 的单侧置信下限

$$P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}}<t_{\alpha}(n-1)\right)=1-\alpha \text{（即 } \frac{\overline{X}-\mu}{S\Big/\sqrt{n}}\sim t(n-1)\text{）}$$

即: $P(\mu > \overline{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1)) = 1-\alpha$

于是得到μ的一个置信水平为$1-\alpha$的单侧置信区间

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

即:
$$\underline{\mu} = \overline{X} - \frac{S}{\sqrt{n}} \cdot t_\alpha(n-1)$$

$\because \quad \overline{x} = 234.7$

$$t_\alpha(n-1) = t_{0.05}(20-1) = t_{0.05}(19) = 1.7291$$

所求的 μ 的单侧置信下限为:

$$\frac{s}{\sqrt{n}} = \sqrt{\frac{1590.85}{20}} = 8.92$$

$$\underline{\mu} = 234.7 - 8.92 \times 1.7291 = 234.7 - 15.4 = 219.3(元)$$

该区域职工家庭人均月收入的最低下限为**219.3 (元)**。

第七章作业（教材第五版）：

P176：1、2、3、5

P177： 8、10

P178： 12、13、15、16、17

P179： 21、22、23、25

注：作业不得抄袭；写上姓名、班级、学号和页码（如1/5）。第五章、第六章、第七章作业一起于12月1日（含）之前提交至教学云平台，标明题目属于第五章/第六章/第七章。