

## 第三节 区间估计

一、区间估计的基本概念

二、典型例题

三、小结



# 一、区间估计的基本概念

## 1. 置信区间的定义

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

$\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$  和  $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$  满足  $P\{\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)\} = 1 - \alpha$ , 则称随机区间  $(\underline{\theta}, \bar{\theta})$  是  $\theta$  的置信度为  $1 - \alpha$  的置信区间,  $\underline{\theta}$  和  $\bar{\theta}$  分别称为置信度为  $1 - \alpha$  的双侧置信区间的置信下限和置信上限,  $1 - \alpha$  为置信度.



## 关于定义的说明

被估计的参数  $\theta$  虽然未知,但它是一个常数,没有随机性,而区间  $(\underline{\theta}, \bar{\theta})$  是随机的.

因此定义中下表达式

$$P\{\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)\} = 1 - \alpha$$

的本质是:

随机区间  $(\underline{\theta}, \bar{\theta})$  以  $1 - \alpha$  的概率包含着参数  $\theta$  的真值,而不能说参数  $\theta$  以  $1 - \alpha$  的概率落入随机区间  $(\underline{\theta}, \bar{\theta})$ .





另外定义中的表达式

$$P\{\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)\} = 1 - \alpha$$

还可以描述为：

若反复抽样多次(各次得到的样本容量相等,都是 $n$ )

每个样本值确定一个区 间 $(\underline{\theta}, \bar{\theta})$ ,

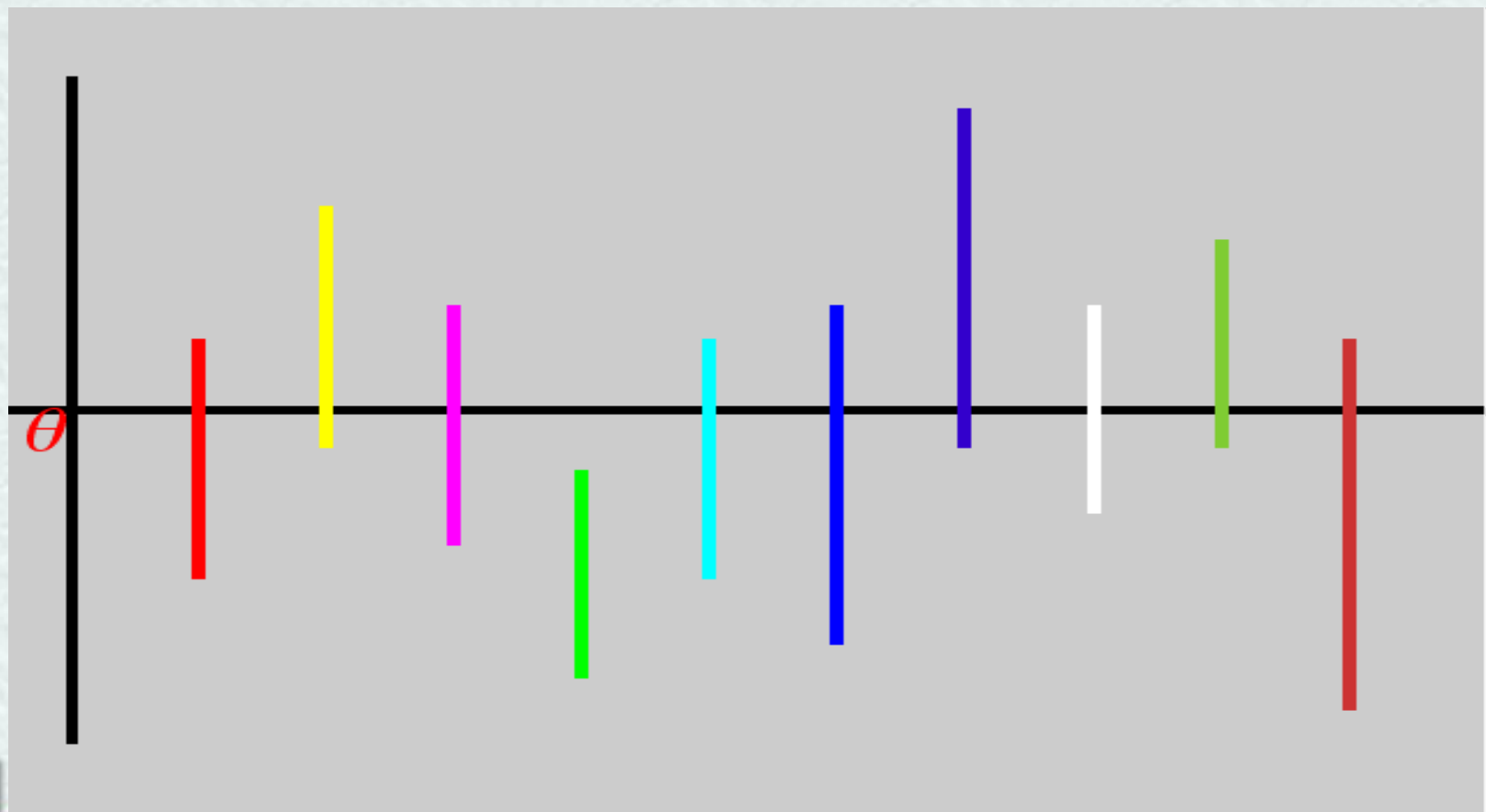
每个这样的区间或包含 $\theta$ 的真值或不包含 $\theta$ 的真值,

按**伯努利大数定理**,在这样多的区间中,

包含 $\theta$ 真值的约占 $100(1 - \alpha)\%$ ,不包含的约占 $100\alpha\%$ .



例如 若  $\alpha = 0.01$ , 反复抽样 1000 次,  
则得到的 1000 个区间中不包含  $\theta$  真值的约为 10 个.



## 2. 求置信区间的一般步骤 (共3步)

(1) 寻求一个样本  $X_1, X_2, \dots, X_n$  的函数：

$$Z = Z(X_1, X_2, \dots, X_n; \theta)$$

其中仅包含待估参数  $\theta$ , 并且  $Z$  的分布已知且不依赖于任何未知参数 (包括  $\theta$ ).

(2) 对于给定的置信度  $1 - \alpha$ , 定出两个常数  $a, b$ , 使  $P\{a < Z(X_1, X_2, \dots, X_n; \theta) < b\} = 1 - \alpha$ .





(3) 若能从  $a < Z(X_1, X_2, \dots, X_n; \theta) < b$  得到等价的  
 不等式  $\underline{\theta} < \theta < \bar{\theta}$ , 其中  $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ ,  
 $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$  都是统计量, 那么  $(\underline{\theta}, \bar{\theta})$  就  
 是  $\theta$  的一个置信度为  $1 - \alpha$  的置信区间.

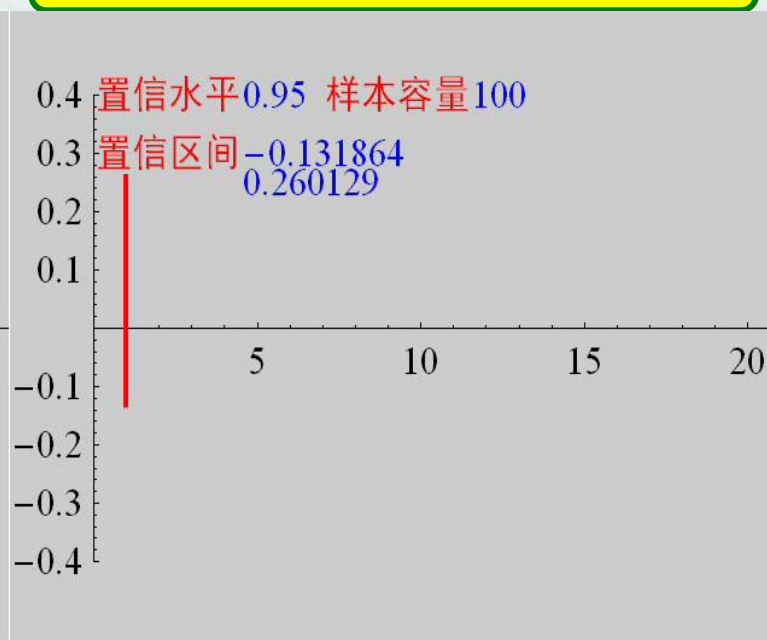
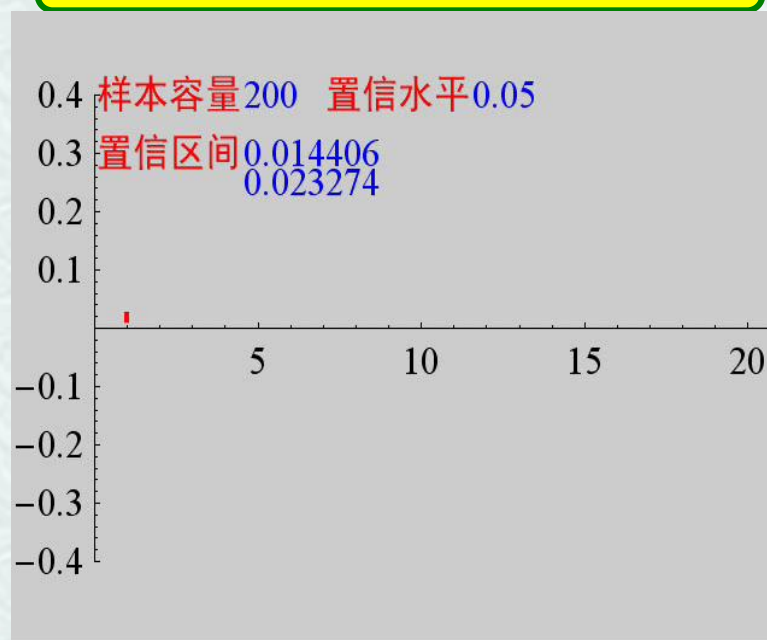


样本容量  $n$  固定, 置信水平  $1-\alpha$  增大, 置信区间长度增大, 可信程度增大, 区间估计精度降低.

置信水平  $1-\alpha$  固定, 样本容量  $n$  增大, 置信区间长度减小, 可信程度不变, 区间估计精度提高.

单击图形播放/暂停 ESC键退出

单击图形播放/暂停 ESC键退出





## 二、典型例题

例1 设总体  $X$  在  $[0, \theta]$  上服从均匀分布, 其中  $\theta$  ( $\theta > 0$ ) 未知,  $(X_1, X_2, \dots, X_n)$  是来自总体  $X$  的样本, 给定  $\alpha$ , 求  $\theta$  的置信水平为  $1 - \alpha$  的置信区间.

解 令  $X_h = \max\{X_1, X_2, \dots, X_n\}$ ,

由上节例4可知,  $\frac{n+1}{n} X_h$  是  $\theta$  的无偏估计,

因为  $X_h$  的概率密度为  $f(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 \leq x \leq \theta, \\ 0, & \text{其他.} \end{cases}$



考察包括待估参数  $\theta$  的随机变量  $Z = \frac{X_h}{\theta}$ ,

其概率密度为  $g(z) = \begin{cases} nz^{n-1}, & 0 \leq z \leq 1, \\ 0, & \text{其他.} \end{cases}$

对于给定的  $\alpha$ , 可定出两个常数  $a, b (0 < a < b \leq 1)$ ,

满足条件  $P\left\{a < \frac{X_h}{\theta} < b\right\} = 1 - \alpha$ ,

即  $1 - \alpha = \int_a^b nz^{n-1} dz = b^n - a^n$ ,

$\Rightarrow P\left\{\frac{X_h}{b} < \theta < \frac{X_h}{a}\right\} = 1 - \alpha$ ,  $\left(\frac{X_h}{b}, \frac{X_h}{a}\right)$  为置信区间.



例2 设  $X_1, X_2, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本, 其中  $\sigma^2$  为已知,  $\mu$  为未知, 求  $\mu$  的置信水平为  $1-\alpha$  的置信区间.

解 因为  $\bar{X}$  是  $\mu$  的无偏估计,

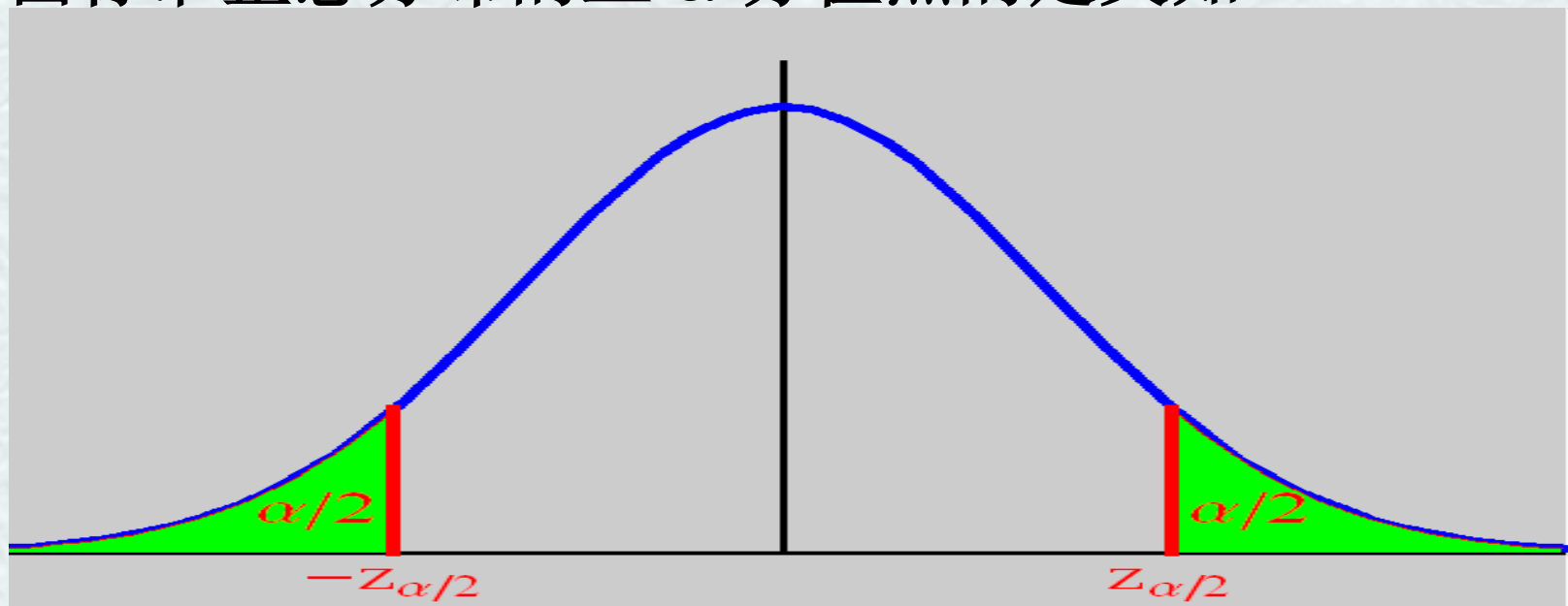
$$\text{且 } U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1),$$

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \text{ 是不依赖于任何未知参数的,}$$





由标准正态分布的上  $\alpha$  分位点的定义知



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

即 
$$P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right\} = 1 - \alpha,$$



于是得  $\mu$  的一个置信水平为  $1-\alpha$  的置信区间

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

这样的置信区间常写成  $\left( \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right).$

其置信区间的长度为  $2 \times \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$



**注意：置信水平为  $1-\alpha$  的置信区间是不唯一的。**

如果在例2中取  $n = 16$ ,  $\sigma = 1$ ,  $\alpha = 0.05$ ,

查表可得  $z_{\alpha/2} = z_{0.025} = 1.96$ ,

得一个置信水平为0.95的置信区间  $\left( \bar{X} \pm \frac{1}{\sqrt{16}} \times 1.96 \right)$ .

由一个样本值算得样本均值的观察值  $\bar{x} = 5.20$ ,

则置信区间为  $(5.20 \pm 0.49)$ , 即  $(4.71, 5.69)$ .





在例2中如果给定  $\alpha = 0.05$ ,

则又有  $P\left\{-z_{0.04} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_{0.01}\right\} = 0.95$ ,

即  $P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{0.01} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{0.04}\right\} = 0.95$ ,

故  $\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{0.01}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{0.04}\right)$  也是  $\mu$  的置信水平为0.95的置信区间.

其置信区间的长度为  $\frac{\sigma}{\sqrt{n}} (z_{0.04} + z_{0.01})$ .



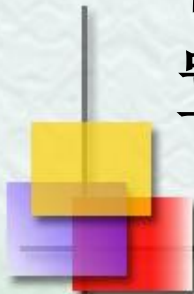
比较两个置信区间的长度

$$L_1 = 2 \times \frac{\sigma}{\sqrt{n}} z_{0.025} = 3.92 \times \frac{\sigma}{\sqrt{n}},$$

$$L_2 = \frac{\sigma}{\sqrt{n}} (z_{0.04} + z_{0.01}) = 4.08 \times \frac{\sigma}{\sqrt{n}},$$

显然  $L_1 < L_2$ . **置信区间短表示估计的精度高.**

**说明:** 对于概率密度的图形是单峰且关于纵坐标轴对称的情况, 易证取  $a$  和  $b$  关于原点对称时, 能使置信区间长度最小.



**例3** 设某工件的长度  $X$  服从正态分布  $N(\mu, 16)$ , 今抽9件测量其长度, 得数据如下(单位:mm):

142, 138, 150, 165, 156, 148, 132, 135, 160.

试求参数  $\mu$  的置信水平为 95% 的置信区间.

**解** 根据例2得  $\mu$  的置信度为  $1-\alpha$  的置信区间

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

由  $n = 9, \sigma = 4, \alpha = 0.05, z_{0.025} = 1.96, \bar{x} = 147.333$  知,

$\mu$  的置信度为 0.95 的置信区间为 (144.720, 149.946).





### 三、小结

点估计不能反映估计的精度, 故而本节引入了区间估计.

置信区间是一个随机区间  $(\underline{\theta}, \bar{\theta})$ , 它覆盖未知参数具有预先给定的概率(置信水平), 即对于任意的  $\theta \in \Theta$ , 有  $P\{\underline{\theta} < \theta < \bar{\theta}\} \geq 1 - \alpha$ .

求置信区间的一般步骤(分三步).

