§4 区间估计

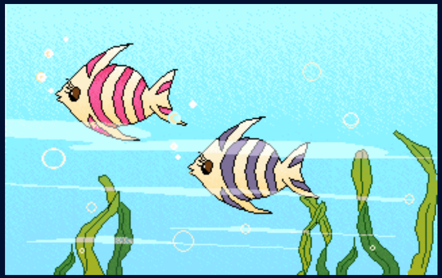
- 置信区间的定义
- 置信区间的意义
- 求置信区间的步骤
- 正态总体均值与方差区间估计

前面，我们讨论了参数点估计．它是用样本算得的一个值去估计未知参数．但是，点估计仅仅给出了未知参数的一个近似值，它没有反映出这种估计的精度．区间估计正好弥补了点估计的这个这个不足之处．

引例　在估计湖中鱼数的问题中，若我们根据一个实际样本，得到鱼数N的极大似然估计为1000条.

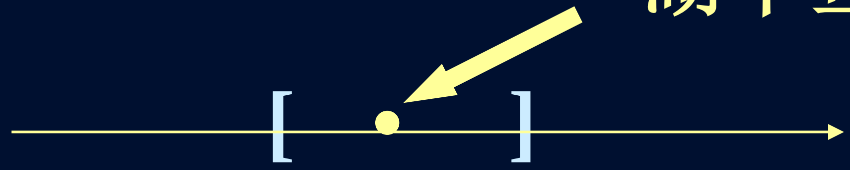
实际上，N的真值可能大于1000条，也可能小于1000条.

为此，我们希望确定一个区间来估计参数真值



a 使我们能以比较高的可靠程度相信它包含真参数值.

湖中鱼数的真值



这里所说的“可靠程度”是用概率来度量的

b 区间估计的精密度要高.

区间估计

设X_1，…，X_n为来自总体$X \sim F(x, \theta)$的一个样本，$\theta \in \Theta$为未知参数。构造以统计量

$$\hat{\theta}_L(X_1, \cdots, X_n), \ \hat{\theta}_U(X_1, \cdots, X_n)$$

为端点的区间 $[\hat{\theta}_L, \ \hat{\theta}_U]$ ，

每当有了样本观测值 $x_1, x_2, \ldots, x_n$，就代入该函数中算出一个区间，

$$[\hat{\theta}_L(x_1, \cdots, x_n), \ \hat{\theta}_U(x_1, \cdots, x_n)]$$

用来作为θ的区间估计

1. 要求 θ 以很大的可能被包含在区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 内，就是说，概率 $P\{\hat{\theta}_1 \le \theta \le \hat{\theta}_2\}$ 要尽可能大.

即要求估计尽量可靠.

2. 估计的精度要尽可能的高. 如要求区间长度 $\hat{\theta}_2 - \hat{\theta}_1$ 尽可能短，或能体现该要求的其它准则.

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.

置信区间的定义

设X_1，…，X_n为来自总体$X \sim F(x, \theta)$的一个样本，$\theta \in \Theta$为未知参数。若对于给定的α（$0 < \alpha < 1$），存在统计量

$$\hat{\theta}_L(X_1, \cdots, X_n),\ \hat{\theta}_U(X_1, \cdots, X_n)$$

使得对所有的$\theta \in \Theta$满足

$$P\{\hat{\theta}_L \le \theta \le \hat{\theta}_U\} = 1 - \alpha$$

$$P\{\hat{\theta}_L \leq \theta \leq \hat{\theta}_U\} = 1 - \alpha$$

则称随机区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 为参数θ 的置信度为1-α 的置信区间，

$\hat{\theta}_L, \hat{\theta}_U$分别称为置信度为1-$\alpha$的双侧置信区间的置信下限和上限。

置信度1-α 也称置信水平。

置信区间的意义

设$X \sim N(\mu, 1)$，$X_1, X_2, \ldots, X_n$ 是一组样本

未知参数 μ 的置信度为0.95的置信区间.

$$\left(\bar{X} - 1.96\sqrt{1/5}, \quad \bar{X} + 1.96\sqrt{1/5} \right)$$

即 $P\left(\bar{X} - 1.96\sqrt{1/5} \le \mu \le \bar{X} + 1.96\sqrt{1/5} \right) = 0.95$

若测得 一组样本值，算得 $\bar{x} = 1.86$

则得一区间 $(1.86 - 0.877, 1.86 + 0.877)$

它可能包含 μ 的真值，也可能不包含μ 的真值

反复抽样得到的区间中有95%包含 μ 的真值.

构造置信区间的方法

- 寻找一个样本的函数

$$g(X_1, X_2, \cdots, X_n, \theta)$$ — 称为枢轴量

它含有待估参数，不含其它未知参数，它的分布已知，(常由θ的点估计出发考虑).

例如 $$\overline{X} \sim N\left(\mu, \frac{1}{5}\right)$$

$$\Rightarrow \frac{\overline{X} - \mu}{\sqrt{1/5}} = g(X_1, X_2, \cdots, X_n, \mu) \sim N(0,1)$$

- 给定置信度 $1-\alpha$，定出两个常数 c , d ,使得

$$P(c < g(X_1, X_2, X_n, \theta) < d) = 1-\alpha$$

- 由$c < g(X_1, X_2, X_n, \theta) < d$ 解出

$$\hat{\theta}_L(X_1, X_2, \cdots, X_n) \quad \hat{\theta}_U(X_1, X_2, \cdots, X_n)$$

得置信区间$(\hat{\theta}_L, \hat{\theta}_U)$

一个正态总体$X \sim N(\mu, \sigma^2)$的情形
置信区间常用公式

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

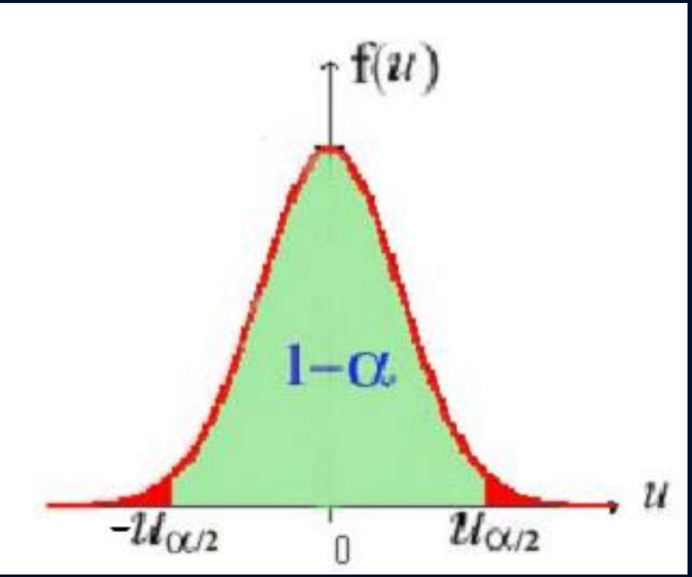
(1) 方差σ^2已知, μ的置信区间

解: 选 μ的点估计为$\overline{X}$

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

寻找一个待估参数和估计量的函数，要求其分布为已知.

对给定的置信水平$1-\alpha$,



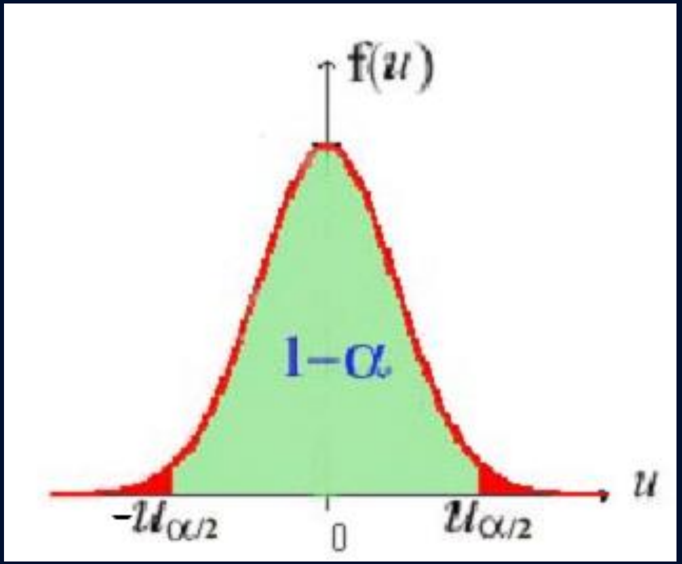
有 $$P\{|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}|\le z_{\alpha/2}\}=1-\alpha$$

从中解得

$$P\{\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\le\mu\le\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}=1-\alpha$$

$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}$$

$$=1-\alpha$$



于是所求μ的 置信区间为

$$[\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}]$$

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

(2) 方差σ^2未知, μ 的置信区间

解: 因方差未知，取

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

对给定的置信度$1-\alpha$，

有 $P\{|t|\le t_{\alpha/2}(n-1)\}=1-\alpha$

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

从中解得

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

$\theta=\mu+1$

$[\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\ \bar{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)]$ 即为均值 μ 的置信水平为 $1-\alpha$ 的区间估计.

思考：求 $\theta=\mu+1$ 的置信水平为 $1-\alpha$ 的区间估计

例1 某工厂生产一批滚珠，其直径 X 服从正态分布$N(\mu,\sigma^2)$, 现从某天的产品中随机抽取6件,测得直径为

15.1，14.8，15.2，14.9，14.6，15.1

(1) 若$\sigma^2=0.06$，求 μ 的置信度为95%的 置信区间;

(2) 若σ^2未知,求 μ 的置信度为95%的置信区间;

(1) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

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

由给定数据算得 $\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i = 14.95$

由公式 得 μ 的置信区间为

$$(14.95 - 1.96 \times 0.1,\quad 14.95 + 1.96 \times 0.1)$$

$$= (14.75,\quad 15.15)$$

(2)

$$\left(\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\ \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right)$$

由给定数据算得 $\bar{x} = 14.95$ 查表得 $t_{0.025}(5) = 2.5706$

$$s^2 = \frac{1}{5}\left(\sum_{i=1}^{6} x_i^2 - 6\bar{x}^2\right) = 0.051. \quad s = 0.226$$

μ 的置信区间为

$$\left(\bar{x}-\frac{s}{\sqrt{6}}t_{0.025}(5),\quad \bar{x}+\frac{s}{\sqrt{6}}t_{0.025}(5)\right)$$

$$=(14.71,\quad 15.187)$$

(3) 当 μ 未知时，方差 σ^2 的置信区间

取枢轴量 $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

对给定的置信度 $1-\alpha$，有

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

从中解得

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

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

于是 $\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$ 即为所求.

思考：求 σ 的置信水平为 $1-\alpha$ 的区间估计

例1续 某工厂生产一批滚珠，其直径 X 服从正态分布$N(\mu,\sigma^2)$, 现从某天的产品中随机抽取6件,测得直径为

15.1，14.8，15.2，14.9，14.6，15.1

(1) 若σ^2=0.06，求 μ 的置信度为95%的 置信区间;

(2) 若σ^2未知,求 μ 的置信度为95%的置信区间;

(3) 求方差σ^2的置信度为95%的置信区间.

(3) $s^2 = 0.051.$

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

查表得 $\chi^2_{0.025}(5) = 12.833$， $\chi^2_{0975}(5) = 0.831$

μ 的置信区间为

$$\left(\frac{5s^2}{\chi^2_{0.025}(5)}, \quad \frac{5s^2}{\chi^2_{0.975}(5)} \right) = (0.0199, \quad 0.3069)$$

正态总体参数的置信区间

总体个数	待估参数	条件	枢轴量	置信区间
一个	μ	σ 已知 $\sigma=\sigma_0$	$\sqrt{n}\left(\bar{X}_n-\mu\right)/\sigma_0 \sim N(0,1)$	$[\bar{X}_n-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}_n+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}]$
		σ 未知	$\sqrt{n}\left(\bar{X}_n-\mu\right)/S \sim t(n-1)$	$[\bar{X}_n-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\ \bar{X}_n+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)]$
	σ^2	μ 未知	$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$	$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$

上页 下页 返回

总体个数	待估参数	条件	枢轴量	置信区间
二个	$\mu_1-\mu_2$	σ_1,σ_2 已知	$\dfrac{\bar{X}_n-\bar{Y}_m-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}\sim N(0,1)$	$[\bar{X}_n-\bar{Y}_m-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}},\ \bar{X}_n-\bar{Y}_m+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}]$
		$\sigma_1=\sigma_2$ 未知	$\dfrac{\bar{X}_n-\bar{Y}_m-(\mu_1-\mu_2)}{S_W\sqrt{1/n+1/m}}\sim t(n+m-2)$	$[\bar{X}_n-\bar{Y}_m-t_{\alpha/2}(n+m-2)S_W\sqrt{\frac{1}{n}+\frac{1}{m}},\ \bar{X}_n-\bar{Y}_m+t_{\alpha/2}(n+m-2)S_W\sqrt{\frac{1}{n}+\frac{1}{m}}]$
	$\dfrac{\sigma_1^2}{\sigma_2^2}$	μ_1,μ_2 未知	$\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n-1,m-1)$	$[\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n-1,m-1)},\ \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n-1,m-1)}]$

作业2：13,14,17,19,20

上页 下页 返回