第六节 (0-1)分布参数的区间估计

一、置信区间公式

二、典型例题

例1 设从一大批产品的100个样品中, 得一级品60个, 求这批产品的一级品率 p 的置信水平为0.95的置信区间.

- 非正态总体
- 0-1分布的区间估计

推导过程

因为(0–1)分布的均值和方差分别为

$$\mu = p, \quad \sigma^2 = p(1-p),$$

设 $X_1, X_2, \cdots, X_n$ 是一个样本, 因为容量n较大,

由中心极限定理知 $$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} = \frac{n\bar{X} - np}{\sqrt{np(1-p)}}$$

定理1（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n, \cdots$相互独立，服从同一分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2$ $(k=1,2,\cdots)$，则随机变量之和 $\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$对于任意x满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

近似地服

$P\{-$ … $\alpha,$

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

等价于 $(n + z_{\alpha/2}^2)p^2 - (2n\overline{X} + z_{\alpha/2}^2)p + n\overline{X}^2 < 0,$

记 $p_1 = \dfrac{-b - \sqrt{b^2 - 4ac}}{2a}, \quad p_2 = \dfrac{-b + \sqrt{b^2 - 4ac}}{2a},$

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

于是 p 的近似置信水平为 $1-\alpha$ 的置信区间是

$$(p_1, \ p_2).$$

一、置信区间公式

设有一容量$n > 50$的大样本，它来自$(0-1)$分布的总体X，X的分布律为 $f(x;p) = p^x(1-p)^{1-x}$，$x = 0, 1$，其中p为未知参数，现在来求p的置信水平为$1-\alpha$的置信区间.

$$\left(\frac{-b-\sqrt{b^2-4ac}}{2a}, \frac{-b+\sqrt{b^2-4ac}}{2a}\right),$$

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

二、典型例题

例1 设从一大批产品的100个样品中, 得一级品60个, 求这批产品的一级品率 p 的置信水平为0.95的置信区间.

解 一级品率p是(0-1)分布的参数,

$$n=100, \qquad \bar{x}=\frac{60}{100}=0.6,$$

$$1-\alpha=0.95, \qquad z_{\alpha/2}=z_{0.025}=1.96,$$

则

$$a=n+z_{\alpha/2}^2=103.84,$$

$$b=-(2n\overline{X}+z_{\alpha/2}^2)=-(2n\bar{x}+z_{\alpha/2}^2)=-123.84,$$

$$c=n\overline{X}^2=n\bar{x}^2=36,$$

于是

$$p_1 = \frac{-b-\sqrt{b^2-4ac}}{2a} = 0.50,$$

$$p_2 = \frac{-b+\sqrt{b^2-4ac}}{2a} = 0.69,$$

故p 的置信水平为0.95的置信区间为(0.50, 0.69).

例2 设从一大批产品的120个样品中，得次品9个，求这批产品的次品率 p 的置信水平为0.90的置信区间.

解 $n = 120, \quad \bar{x} = \dfrac{9}{100} = 0.09, \quad 1 - \alpha = 0.90,$

则 $a = n + z_{\alpha/2}^2 \ = 122.71,$

$$b = -(2n\overline{X} + z_{\alpha/2}^2) = -(2n\bar{x} + z_{\alpha/2}^2) = -24.31,$$

$$c = -n\overline{X}^2 = -n\bar{x}^2 \ = 0.972,$$

于是 $p_1 = \dfrac{-b - \sqrt{b^2 - 4ac}}{2a} = 0.056,$

$$p_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} = 0.143,$$

p 的置信水平为**0.90**的置信区间为 $(0.056, 0.143).$