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

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

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

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} = \frac{n\overline{X} - np}{\sqrt{np(1-p)}}$$

近似地服从 $N(0,1)$ 分布，**于是有**

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

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

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

于是 p 的近似置信水平为 $1-\alpha$ 的置信区间是

$$(p_1,\ p_2).$$

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

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

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

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).

补充例题