6.2 数理统计中常见的抽样分布

北京化工大学数学系

苏贵福

统计量的概率分布称为抽样分布. 正态总体在数理统计中有着特别重要的地位. 本节将重点介绍几类来自正态总体的样本所构成的统计量的分布.

一. 样本均值的分布

定理1 设$X_1, X_2, \cdots, X_n$是来自总体$N(\mu, \sigma^2)$ 的一个样本. 则样本的任一确定的线性函数

$$U = \sum_{i=1}^{n} a_i X_i$$

也服从正态分布 $N\left(\mu \sum_{i=1}^{n} a_i, \sigma^2 \sum_{i=1}^{n} a_i^2\right)$.

证明 样本$X_1, X_2, \cdots, X_n$相互独立且与总体具有相同的分布,又因为独立正态分布的线性组合仍为正态分布(P57例7). 故只需求$E(U)$和$D(U)$.

根据期望和方差的概念，有

$$E(U)=E\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_iE(X_i)=\mu\sum_{i=1}^{n}a_i.$$

$$D(U)=D\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_i^2D(X_i)=\sigma^2\sum_{i=1}^{n}a_i^2.$$

因此U服从正态分布$N\left(\mu\sum_{i=1}^{n}a_i,\sigma^2\sum_{i=1}^{n}a_i^2\right)$.

特别地，当$a_1=a_2=\cdots=a_n=\frac{1}{n}$时，$U=\overline{X}$，进而

$$\overline{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right).$$

♠ 由上述讨论知$\overline{X}$与总体X的均值相同，但其方差是X方差的$\frac{1}{n}$，因而它与数学期望的μ的偏差更小.

二. χ^2分布

定义1 设随机变量$X_1, X_2, \cdots, X_n$相互独立且 都服从标准正态分布, 则统计量

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

所服从的分布称为自由度为n的$\underline{\underline{\chi^2\text{分布}}}$, 记作$\chi^2 \sim \chi^2(n)$.

定义2 当$\alpha > 0$时, 若积分$\int_0^{+\infty} x^{\alpha-1}e^{-x}dx$是收敛的, 则称函数

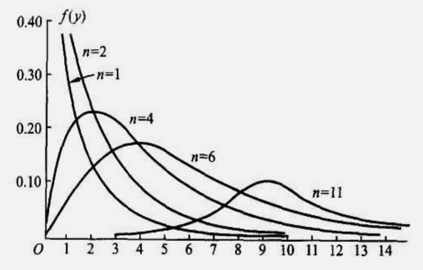
$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx$$

为Γ函数, 且有$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\Gamma(n+1) = n!$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$.

定理2 χ^2分布的密度函数为

$$\chi^2(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

下图刻画了$n = 1, 2, 4, 6$以及$n = 6$时χ^2分布的密度函数曲线:



性质1 $E(\chi^2(n)) = n,\ D(\chi^2(n)) = 2n.$

解 由于$X_i \sim N(0,1)$, 故$E(X_i^2) = D(X_i) + [E(X_i)]^2 = 1,\ i = 1, 2, \cdots, n.$
又因为

$$E(X_i^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3.$$

故$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 2.$

性质1 $E(\chi^2(n)) = n,\ D(\chi^2(n)) = 2n.$

解 由于$X_i \sim N(0,1)$, 故$E(X_i^2) = D(X_i) + [E(X_i)]^2 = 1,\ i = 1, 2, \cdots, n.$
又因为

$$E(X_i^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3.$$

故$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 2.$ 因此

$$E(\chi^2(n)) = E\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} E(X_i^2) = \sum_{i=1}^{n} D(X_i) = n.$$

$$D(\chi^2(n)) = D\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} D(X_i^2) = \sum_{i=1}^{n} 2 = 2n.$$

性质2 设$X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$, 且它们相互独立, 则

$$X_1 + X_2 \sim \chi^2(n_1 + n_2).$$

定理3 设$X_1, X_2, \cdots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本, 则样本方差S^2与样本均值$\overline{X}$相互独立, 且

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

三. t分布

定义3 设随机变量$X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且X与Y相互独立, 则称统计量

$$T = \frac{X}{\sqrt{Y/n}}$$

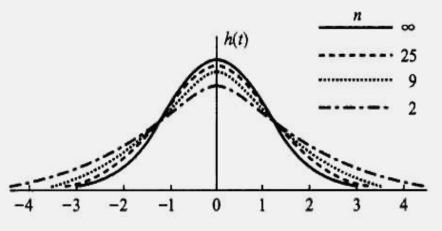
服从自由度为n的t分布, 记作$T \sim t(n)$.

t分布又称学生分布, 它的概率密度函数为

$$h(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < \infty.$$

易知t分布的密度函数关于$t=0$对称, 如下图所示.

$$\lim_{n\to\infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$



当$n \to \infty$时t分布趋于标准正态分布. 但对于较小的n而言t分布 与标准正态分布相差较大.

定理4 设$X_1, X_2, \cdots, X_n$是来自正态总体$N(\mu, \sigma^2)$的一个样本, 则

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \sim t(n-1).$$

证明 因为

$$X = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

且两者相互独立. 从而由t分布的定义有

$$\frac{X}{\sqrt{\frac{Y}{n-1}}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \sim t(n-1).$$

关于抽样分布的更多内容请参阅教材！