概率论与数理统计

第三十一讲

抽样分布定理

样本 $X_1, X_2, \ldots, X_n$ ⇒ 总体 $X \sim F(x)$

↑

$\boxed{\text{统计量 } g(X_1, X_2, \ldots, X_n)}$

问题

统计量 $g(X_1, X_2, \ldots, X_n)$ 服从什么分布？

本讲主要介绍在正态总体 $N(\mu, \sigma^2)$ 下，样本均值 $\bar{X}$、样本方差 S^2 及其函数的分布.

定理1 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，则有

1) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ ；

2) $\bar{X}$ 与 S^2 相互独立；

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

其中 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$．

定理2 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，则有

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

定理3 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $N(\mu_1, \sigma_1^2)$ 的样本；$Y_1, Y_2, \ldots, Y_m$ 是来自总体 $N(\mu_2, \sigma_2^2)$ 的样本，且两个样本相互独立，则有

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1).$$

其中 S_1^2，S_2^2 分别是两个样本的样本方差.

定理4 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $N(\mu_1, \sigma^2)$ 的样本；$Y_1, Y_2, \ldots, Y_m$ 是来自总体 $N(\mu_2, \sigma^2)$ 的样本，且两个样本相互独立，则有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t(n+m-2).$$

其中 $\bar{X},\bar{Y},S_1^2,\ S_2^2$ 分别是两个样本的样本均值和样本方差，且

$$S_\omega^2=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2},\quad S_\omega=\sqrt{S_\omega^2}.$$

证明 由定理1 知

$$\bar{X} \sim N(\mu_1, \frac{\sigma^2}{n}), \quad \bar{Y} \sim N(\mu_2, \frac{\sigma^2}{m})$$

$$\Rightarrow \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,\ 1)$$

另外又有

$$\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2(n-1), \quad \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(m-1)$$

由 χ^2 分布的可加性有

$$\frac{(n-1)S_1^2+(m-1)S_2^2}{\sigma^2}\sim\chi^2(n+m-2)$$

因为 $\bar{X},\bar{Y}$ 与 S_1^2,S_2^2 相互独立，由 t 分布定义有：

$$\frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\cancel{\sigma}\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}}{\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{\cancel{\sigma}^2}/(n+m-2)}}\sim t(n+m-2)$$

由 χ^2 分布的可加性有

$$\frac{(n-1)S_1^2+(m-1)S_2^2}{\sigma^2}\sim\chi^2(n+m-2)$$

因为 $\bar{X},\bar{Y}$ 与 S_1^2,S_2^2 相互独立，由 t 分布定义有：

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t(n+m-2).$$

例 从总体 $N(20，16)$ 抽取了样本容量为25的样本，求样本均值落在 18与22之间的概率.

解 由定理1有 $\bar{X} \sim N\left(20, \dfrac{16}{25}\right) \Longrightarrow \dfrac{\bar{X}-20}{4/5} \sim N(0,1)$

由此得到

$$P\{18<\bar{X}<22\} = P\left\{\frac{18-20}{4/5} < \frac{\bar{X}-20}{4/5} < \frac{22-20}{4/5}\right\}$$

$$= \Phi(2.5) - \Phi(-2.5)$$

$$= 2\Phi(2.5) - 1 = 0.9876$$

例 设 $X_1, X_2, \ldots, X_{10}$ 是来自总体 $N(0, 4)$ 的样本，试确定常数 C，使得

$$P\left\{\sum_{i=1}^{10} X_i^2 > C\right\} = 0.05.$$

例 设 $X_1, X_2, \ldots, X_n, X_{n+1}$ 是来自总体 $N(\mu, \sigma^2)$ 的样本，记

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

试证：

$$\sqrt{\frac{n}{n+1}}\,\frac{X_{n+1} - \bar{X}_n}{S_n} \sim t(n-1).$$

证明 因为 $X_{n+1} \sim N(\mu,\ \sigma^2),\ \bar{X}_n \sim N(\mu,\ \dfrac{\sigma^2}{n})$

$$\Longrightarrow X_{n+1} - \bar{X}_n \sim N(0,\ \frac{n+1}{n}\sigma^2)$$

$$\Longrightarrow \frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{(n+1)/n}} \sim N(0,\ 1)$$

又因为 $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$

且 $X_{n+1}-\bar{X}_n$ 与 S_n^2 相互独立，由 t 分布定义有：

$$\frac{\dfrac{X_{n+1}-\bar{X}_n}{\cancel{\sigma}\sqrt{(n+1)/n}}}{\sqrt{\dfrac{\cancel{(n-1)}S_n^2}{\cancel{\sigma^2}}\Big/\cancel{(n-1)}}}\sim t(n-1)$$

$$\Rightarrow\quad \sqrt{\frac{n}{n+1}}\,\frac{X_{n+1}-\bar{X}_n}{S_n}\sim t(n-1).$$

例 设 X_1，X_2，X_3，X_4 是来自总体 $N(0，1)$ 的样本，试确定常数C，使得

$$P\left\{ \frac{(X_1 - X_2)^2}{(X_1 - X_2)^2 + (X_3 - X_4)^2} > C \right\} = 0.95.$$