第三节　抽样分布

- 一、简介
- 二、基本定理

一、简介

统计量既然是依赖于样本的, 而后者又是随机变量, 故统计量也是随机变量, 因而就有一定的分布. 称这个分布为“抽样分布”． 也即抽样分布就是统计量的分布.

$$\text{抽样分布}\begin{cases}\text{精确抽样分布} & \text{(小样本问题中使用)}\\ \text{渐近分布} & \text{(大样本问题中使用)}\end{cases}$$

这一节, 我们来讨论正态总体的抽样分布.

二、基本定理

定理5.7 设随机变量序列$X_1, X_2, \cdots, X_n$ 相互独立,且

$$X_i \sim N(\mu_i, \sigma_i^2) \qquad (i = 1, 2, \cdots, n)$$

则它们的任一确定的线性函数

$$\sum_{i=1}^{n} C_i X \sim N\left(\sum_{i=1}^{n} C_i \mu_i, \sum_{i=1}^{n} C_i^2 \sigma_i^2\right).$$

其中$C_1, C_2, \cdots, C_n$为不全为零的常数.

证　由于$X_1, X_2, \cdots, X_n$独立且均为正态变量,

故他们的线性函数$\sum_{i=1}^{n} C_i X_i$仍为正态变量,又

$$E(\sum_{i=1}^{n} C_i X_i) = \sum_{i=1}^{n} C_i E(X_i) = \sum_{i=1}^{n} C_i \mu_i$$

$$D(\sum_{i=1}^{n} C_i X_i) = \sum_{i=1}^{n} C_i^2 D(X_i) = \sum_{i=1}^{n} C_i^2 \sigma_i^2$$

所以　$\sum_{i=1}^{n} C_i X_i \sim N(\sum_{i=1}^{n} C_i \mu_i, \sum_{i=1}^{n} C_i^2 \sigma_i^2)$

1. 样本来自单个正态总体

定理5.3 设样本$(X_1, X_2, \cdots, X_n)$是来自总体X，而

$$X \sim N(\mu, \sigma^2),$$

则 (1) 样本均值

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2 / n),$$

或

$$U = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0, 1). \longrightarrow$$

标准化样本均值

目的:

估计总体

数学期望μ.

样本关于$\overline{X}$的平均偏离程度

(2) $V = \dfrac{S_n^2}{\dfrac{\sigma^2}{n}} = \dfrac{nS_n^2}{\sigma^2} = \dfrac{(n-1)S_n^{*2}}{\sigma^2}$ 目的：估计σ^2.

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1)$$

$D(\overline{X}) = \dfrac{\sigma^2}{n}$：样本均值$\overline{X}$关于总体期望的偏离程度

其中S_n^2是样本方差.

(3) $\overline{X}$ 与 S_n^2 独立.

注 1° $V=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2(n-1)$,

自由度减少一个!

减少一个自由度的原因:

$\{\frac{X_i-\overline{X}}{\sigma}\}(i=1,2\cdots,n)$不相互独立.

事实上，它们受到一个条件的约束:

$$\boxed{\sum_{i=1}^{n} X_i = n\overline{X}}$$

$$\sum_{i=1}^{n}\frac{X_i-\overline{X}}{\sigma}=\frac{1}{\sigma}\left(\sum_{i=1}^{n} X_i - n\overline{X}\right)=\frac{1}{\sigma}\cdot 0=0.$$

2° 若X不服从正态分布，则由中心极限定理知，

当 $n >> 1$ (一般$n \geq 30$)时，

了解

$$U = \frac{\overline{X} - \mu}{\sigma} \cdot \sqrt{n} \overset{\text{近似}}{\sim} N(0, 1),$$

其中 $\mu = E(X)$, $\sigma^2 = D(X)$.

3° 在实际问题中，总体方差σ^2常常是未知的，

若将标准样本均值U中的σ用S_n^*代替，则有

推论1 设$(X_1, X_2, \cdots, X_n)$是总体$N(\mu,\sigma^2)$的样本，

$\bar{X}, S_n^{*2}$ 分别是样本均值和修正样本方差，则有

$$T = \frac{\bar{X}-\mu}{S_n^{*}/\sqrt{n}} = \frac{\bar{X}-\mu}{S_n/\sqrt{n-1}} \sim t(n-1).$$

证 $\because$ $U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), V = \dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1),$

且两者独立，由 t 分布的定义知

$$T = \frac{U}{\sqrt{\dfrac{V}{n-1}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \bigg/ \sqrt{\frac{(n-1)S_n^{*2}}{\sigma^2(n-1)}} \sim t(n-1).$$

$$S_n^{*2} = \frac{n}{n-1} S_n^2$$

例1 设$(X_1, X_2, \cdots, X_{10})$和$(Y_1, Y_2, \cdots, Y_{15})$是来自总体 $N(20,3)$的两个独立的样本，求

$$P\{|\bar{X}-\bar{Y}|>0.3\}.$$

解 $$\bar{X}=\frac{1}{10}\sum_{i=1}^{10}X_i \sim N(20, \frac{3}{10}),$$

$$\bar{Y}=\frac{1}{15}\sum_{i=1}^{15}Y_i \sim N(20, \frac{3}{15}),$$

$$\therefore \quad \bar{X}-\bar{Y} \sim N(0, \frac{3}{10}+\frac{3}{15}) = N(0, \frac{1}{2}),$$

故 $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{1}{2}}} \sim N(0,1)$

从而 $P\{|\overline{X}-\overline{Y}|>0.3\}=1-P\{|\overline{X}-\overline{Y}|\le 0.3\}$

$$=1-P\left\{\left|\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{2}}}\right|\le\frac{0.3}{\sqrt{\frac{1}{2}}}\right\}$$

$$=2[1-\Phi(0.3\sqrt{2})] \approx 2(1-0.6628)=0.6744.$$

2. 样本来自两个正态总体

定理5.4 设总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$

X与Y相互独立. 样本$(X_1, X_2, \cdots, X_{n_1})$

与$(Y_1, Y_2, \cdots, Y_{n_2})$分别来自总体$X$和$Y$，则

(1) $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2})$

或 $$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \sim N(0,1)$$

(2) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时,

(2) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时，

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2),$$

其中 $S_w^2=\dfrac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}$， $S_w=\sqrt{S_w^2}$.

S_1^{*2}和S_2^{*2}分别是来自两个总体样本的修正样本方差.

(3) $$F=\frac{S_1^{*2}/\sigma_1^2}{S_2^{*2}/\sigma_2^2}\sim F(n_1-1,n_2-1);$$

证 (1)、略

(2) 由定理5.3及定理5.4，知

$$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2})$$

$$\therefore \quad U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1),$$

由 $\frac{(n_1-1)S_1^{*2}}{\sigma^2} \sim \chi^2(n_1-1)$, $\frac{(n_2-1)S_2^{*2}}{\sigma^2} \sim \chi^2(n_2-1)$,

且它们相互独立, 故由 χ^2 分布的可加性知

$$V=\frac{(n_1-1)S_1^{*2}}{\sigma^2}+\frac{(n_2-1)S_2^{*2}}{\sigma^2}\sim\chi^2(n_1+n_2-2),$$

由于 U 与 V 相互独立,按 t 分布的定义

$$T=\frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

$$=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2).$$

(3) $\dfrac{(n_1-1)S_1^{*2}}{\sigma_1^2} \sim \chi^2(n_1-1)$, $\dfrac{(n_2-1)S_2^{*2}}{\sigma_2^2} \sim \chi^2(n_2-1)$,

由假设 S_1^{*2}, S_2^{*2} 独立, 则由 F 分布的定义知

$$\frac{(n_1-1)S_1^{*2}}{(n_1-1)\sigma_1^2} \Bigg/ \frac{(n_2-1)S_2^{*2}}{(n_2-1)\sigma_2^2} \sim F(n_1-1, n_2-1),$$

即 $F = \dfrac{S_1^{*2}/\sigma_1^2}{S_2^{*2}/\sigma_2^2} \sim F(n_1-1, n_2-1).$

例 设总体 $X \sim N(0,\sigma^2)$，$(X_1, X_2, \cdots, X_n, X_{n+1}, \cdots, X_{n+m})$ 是来自总体 X 的容量为 $n+m$ 的一个样本,试求统计量

$$T=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}}$$

的概率分布.

重点掌握

解 由于 $X_1, X_2, \cdots, X_{n+m}$ 独立且 $\dfrac{X_i}{\sigma} \sim N(0,1)$，则有

$$\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right) \sim N(0,n) \quad \text{或} \quad \frac{\sum_{i=1}^{n}(X_i/\sigma)}{\sqrt{n}} \sim N(0,1)$$

并且

$$\sum_{i=n+1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2(m)$$

又因为 $\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)/\sqrt{n}$ 与 $\sum_{i=n+1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2$ 相互独立，

再有t分布的定义得

$$\frac{\sum_{i=1}^{n}(X_i/\sigma)/\sqrt{n}}{\sqrt{\sum_{i=n+1}^{n+m}(X_i/\sigma)^2/m}} \sim t(m)$$

即

$$T=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}} \sim t(m)$$

备份题

例1 设$(X_1, X_2, \cdots, X_9)$是来自正态总体$N(0,1)$的样本，统计量

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2 + (X_7 + X_8 + X_9)^2$$

求常数C，使CY服从χ^2分布

解 因为 $X_1 + X_2 + X_3 \sim N(0,3)$,

$$X_4 + X_5 + X_6 \sim N(0,3),$$

$$X_7 + X_8 + X_9 \sim N(0,3),$$

故$\dfrac{1}{3}Y \sim \chi^2(3)$，从而$C = \dfrac{1}{3}$

例2 设$X_1,X_2,\cdots,X_n$是来自正态总体$N(\mu,\sigma^2)$的一个样本，$\overline{X}$和S_n^2是其样本均值和样本方差，又设$X_{n+1}\sim N(\mu,\sigma^2)$且与$X_1,X_2,\cdots,X_n$独立，试求统计量

$$F=\frac{(X_{n+1}-\overline{X})^2}{S_n^2}\times\frac{n-1}{n+1}$$

的概率分布

解：由于X_{n+1}与$X_1,X_2,\cdots,X_n$独立，则X_{n+1}与$\overline{X}$独立.且有

$$E(X_{n+1}-\overline{X})=E(X_{n+1})-E(\overline{X})=\mu-\mu=0$$

$$D(X_{n+1}-\overline{X})=D(X_{n+1})+D(\overline{X})=\sigma^2+\frac{\sigma^2}{n}=\frac{n+1}{n}\sigma^2$$

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

$$\frac{X_{n+1}-\overline{X}}{\sigma\sqrt{\frac{n+1}{n}}}\sim N(0,1),\quad \frac{(X_{n+1}-\overline{X})^2}{\sigma^2\frac{n+1}{n}}\sim\chi^2(1)$$

又 $\frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$，且$S_n^2$与$X_{n+1}-\overline{X}$独立，则有

$$\frac{\dfrac{(X_{n+1}-\bar{X})^2}{\sigma^2\dfrac{n+1}{n}}\Big/1}{\dfrac{nS_n^2}{\sigma^2}\Big/(n-1)} \sim F(1,n-1)$$

即 $$F=\frac{(X_{n+1}-\bar{X})^2}{S_n^2}\times\frac{n-1}{n+1}\sim F(1,n-1)$$