§7.3 抽样分布定理

抽样分布：设$X_1, X_2, \cdots, X_n$是取自总体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$是统计量，称此统计量所服从的分布为**抽样分布**。

一、一个正态总体的抽样分布

设$X_1, X_2, \cdots, X_n$是取自总体 X 的一个样本，其样本均值为$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，样本方差为 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$。我们就先来看看这两个统计量的分布情况。

定理：设 $X_1, X_2, \cdots, X_n$ 是取自总体 X 的一个样本

且 $E(X)=u,\ D(X)=\sigma^2$ 则 $E(\overline{X})=u,\ D(\overline{X})=\sigma^2/n$。

证明：因 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，故

$$E(X_i)=E(X)=u,\ D(X_i)=D(X)=\sigma^2 (i=1,2,\cdots,n)$$

且 $X_1, X_2, \cdots, X_n$ 相互独立。于是有 $E(\overline{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)$

$$=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=u,\ D(\overline{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}.$$

该定理说明，样本容量越大，方差越小，从而样本均值越来越接近总体的平均值。

定理7.3：设 $X_1, X_2, \cdots, X_n$ 是取自总体 $X \sim N(u,\sigma^2)$ 的一个样本，则

(1) $\overline{X} \sim N(u, \frac{\sigma^2}{n});$ (2) $\frac{\overline{X}-u}{\sigma/\sqrt{n}} \sim N(0,1);$

（3）若记 $Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - u)^2$, 则 $Y \sim \chi^2(n)$。

证明：(1) 因 $X_1, X_2, \cdots, X_n$ 相互独立且都服从正态分布 $N(u,\sigma^2)$, 故 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 也服从正态分布，故有

$$\overline{X} \sim N(u, \frac{\sigma^2}{n})$$

$$E(\overline{X}) = u, \quad D(\overline{X}) = \frac{\sigma^2}{n};$$

(2) 将 X 标准化，有 $\dfrac{\overline{X}-u}{\sigma/\sqrt{n}} \sim N(0,1).$

（3）由条件知$X_1, X_2, \cdots, X_n$相互独立且

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

故 $\dfrac{X_i-u}{\sigma} \sim N(0,1),\ i=1,2,\cdots,n$ ，从而有

$\dfrac{X_1-u}{\sigma}, \dfrac{X_2-u}{\sigma}, \cdots, \dfrac{X_n-u}{\sigma}$ 相互独立，再由 χ^2 的

定义知:

$$Y=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-u)^2=\sum_{i=1}^{n}\left(\frac{X_i-u}{\sigma}\right)^2 \sim \chi^2(n).$$

例：已知某地区有110kw 的电网，在一般情况下电压的 值（单位：千瓦） $V \sim N(110, 5.5^2)$若在某天内随机进行了16次电压值测试。试问其样本均值 $\overline{V}$ 与110kw 的偏差小于 4kw 的概率是多少?

分析：问题的关键在于要知道正态总体样本均值 $\overline{V}$ 的分布。

解：由题意知，总体 $V \sim N(110, 5.5^2)$，样本容量为 n =16，由以上定理有 $\overline{V} \sim N(110, \frac{5.5^2}{16}) = N(110, 1.375^2)$。于是，所求概率为

$$P(|\overline{V} - 110| < 4) = P(|\frac{\overline{V} - 110}{1.375}| < \frac{4}{1.375} = 2.91) = 2\Phi(2.91) - 1 = 0.9964.$$

注：由上可见，不管总体方差如何，但只要样本容量足够大，则样本均值与总体均值之间大于 4kv 的概率是很小的 。由此可以想见，当容量充分大时，就可以认为样本均值就是总体均值。

在实际问题中，一个正态总体 $X \sim N(u,\sigma^2)$ 的参数 u,σ^2往往是未知的，这时，我们就可以用样本均值 $\overline{X}$ 来估计总体均值 u ，用样本方差 S^2或二阶中心矩 B_2 来估计总体方差 σ^2。

定理7.4：设$X_1,X_2,\cdots,X_n$是正态总体$X\sim N(u,\sigma^2)$的样本，则

(1)　$\overline{X}$与S^2相互独立；

(2)　$$\frac{(n-1)S^2}{\sigma^2}=\frac{nB_2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\sim\chi^2(n-1)。$$

证明不作要求。

定理7.5：设$X_1,X_2,\cdots,X_n$是正态总体$X\sim N(u,\sigma^2)$的样本，则

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

由样本均值的抽样分布知：

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), \text{即} U = \frac{\bar{X}-\mu}{\sigma / \sqrt{n}} \sim N(0,1).$$

由方差的抽样分布知：

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

由于 $\bar{X}$ 与 S^2 独立，则U与V也独立。

由t分布的定义知

$$T = \frac{U}{\sqrt{V / n-1}} = \frac{\bar{X}-\mu}{S / \sqrt{n}} \sim t(n-1),$$

注：需要说明的是，上述定理 2、3、4 仅适用于正态总体，而非正态总体的抽样问题要复杂得多。但中心极限定理告诉我们，只要样本容量足够大，抽样问题是可以转化为近似正态分布问题来加以解决，即

定理7.6：设$X_1, X_2, \cdots, X_n$是总体X的样本，若E(X)=u

$D(X)=\sigma^2$，则当n充分大时，近似地有

$$(1)\quad \overline{X} \sim N\left(u, \frac{\sigma^2}{n}\right); \qquad (2)\quad \frac{\overline{X}-u}{S/\sqrt{n}} \sim N(0,1).$$

（注：由中心极限定理可证）。

二、两个正态总体的抽样分布

设X，Y是两个正态总体，$X_1, X_2, \cdots, X_{n_1}$是来自 X 的样本，$Y_1, Y_2, \cdots, Y_{n_2}$ 是来自Y 的样本，记

$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i,$ 为X 的样本均值；

$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1} (X_i - \bar{X})^2$ 为X 的样本方差；

$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i,$ 为Y 的样本均值；

$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ 为Y 的样本方差。

定理7.7-9设两总体$X : N(u_1,\sigma_1{}^2)$，$Y : N(u_2,\sigma_2{}^2)$相互独立，则

(1) $\bar{X}-\bar{Y} : N(\mu_1-\mu_2,\sigma_1^2/n_1+\sigma_2^2/n_2)$，将其标准化，有$U=\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} : N(0,1)$；

(2) 若$\sigma_1=\sigma_2=\sigma$，则

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{1/n_1+1/n_2}} : t(n_1+n_2-2);$$

$$\frac{S_1{}^2}{S_2{}^2} : F(n_1-1,n_2-1).$$ 其中 $S_W=\sqrt{\dfrac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{n_1+n_2-2}}$.

证明：（1）由题意知$\bar{X} \sim N(u_1, \frac{\sigma_1^2}{n_1})$，$\bar{Y} \sim N(u_2, \frac{\sigma_2^2}{n_2})$

且$\overline{X}$与$\overline{Y}$相互独立。 由定理 3.5.5 知它们的线性组合

$$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$ 将其标准化，有

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

（2） 由（1）知，当$\sigma_1=\sigma_2=\sigma$时，有

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

又由方差的抽样分布知

$$\frac{(n_1-1)S_1^2}{\sigma_1^{\ 2}} \sim \chi^2(n_1-1),\quad \frac{(n_2-1)S_2^2}{\sigma_2^{\ 2}} \sim \chi^2(n_2-1)$$

且二者相互独立。由分布的可加性有

$$V=\frac{(n_1-1)S_1^2}{\sigma_1^{\ 2}}+\frac{(n_2-1)S_2^2}{\sigma_2^{\ 2}}=\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$$

且U与V相互独立。由t 分布的定义知

$$\frac{U}{\sqrt{V/(n_1+n_2-2)}}=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} : \ t(n_1+n_2-2).$$

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

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

且相互独立，由F分布的定义：

$$\left(\frac{(n_1-1)S_1^2}{\sigma_1{}^2}\Big/ n_1-1\right)\Big/\left(\frac{(n_2-1)S_2^2}{\sigma_2{}^2}\Big/ n_2-1\right) = \frac{S_1^2\big/\sigma_1{}^2}{S_2^2\big/\sigma_2{}^2} \sim F(n_1-1, n_2-1)$$

例 设总体X~N(0,0.64), $X_1, X_2, \cdots, X_n$ 是来自总体X

样本，记 $Y_n = \sum_{i=1}^{n} X_i^2$, 若要使得

$$P(Y_n \le 25) \ge 0.9,$$

问n至少要去多大?

解： $\because X \sim N(0, 0.8^2), \quad \therefore \dfrac{X}{0.8} \sim N(0,1).$

由于 $X_1, X_2, \cdots, X_n$ **是来自总体** X **的样本，从而**

$\dfrac{X_i}{0.8} \sim N(0,1), \quad i = 1, 2, \cdots, n$ **且它们相互独立。由** χ^2 **分布的定义知** $\dfrac{Y_n}{0.8^2} = \sum\limits_{i=1}^{n} \left(\dfrac{X_i}{0.8} \right)^2 \sim \chi^2(n)$**，于是有**

$$0.9 \le P(Y_n \le 25) = P\left(\frac{Y_n}{0.8^2} \le \frac{25}{0.8^2}\right) = P\left(\frac{Y_n}{0.8^2} \le 39.0625\right),$$

$$\chi^2_{0.9}(n) \le 39.0625, \quad n \le 28.$$

例 设 $X_1, X_2, \cdots, X_9$ 是来自正太总体 $X \sim N(\mu, \sigma^2)$ 的样本，令

$$Y_1 = \frac{1}{6}\sum_{i=1}^{6} X_i, Y_2 = \frac{1}{3}\sum_{i=7}^{9} X_i,$$

$$S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2, Z = \frac{Y_1 - Y_2}{S}.$$

求： a使得$P(Z \le a) = 0.90.$

解：显然$Y_1 \sim N(\mu, \frac{\sigma^2}{6})$，$Y_2 \sim N(\mu, \frac{\sigma^2}{3})$且 Y_1与 Y_2相互独立，

于是有 $Y_1 - Y_2 \sim N(0, \frac{\sigma^2}{6} + \frac{\sigma^2}{3}) = N(0, \frac{\sigma^2}{2})$

所以 $$\frac{Y_1 - Y_2}{\sqrt{\sigma^2/2}} = \frac{\sqrt{2}(Y_1 - Y_2)}{\sigma} \sim N(0,1)$$

因Y_2, S^2分别为 样本容量为3的样本 X_7, X_8, X_9

的均值与方差，故 Y_2，S^2 相互独立且

$\frac{2S^2}{\sigma^2} \sim \chi^2(2)$ 从而 $\frac{\sqrt{2}(Y_1 - Y_2)}{\sigma}$，$\frac{2S^2}{\sigma^2}$ 相互独立。

由 t 分布的定义有 $$\frac{\sqrt{2}(Y_1 - Y_2)/\sigma}{\sqrt{\left(2S^2/\sigma^2\right)/2}} = \frac{\sqrt{2}(Y_1 - Y_2)}{S} \sim t(2).$$

例 设总体 $X \sim N(\mu, \sigma^2), X_1, X_2, \cdots, X_n$ 是X的样本，求

(1) $$P\left(0.26\sigma^2 \le \frac{1}{10}\sum_{i=1}^{10}(X_i-\mu)^2 \le 2.3\sigma^2\right),$$

(2) $$P\left(0.26\sigma^2 \le \frac{1}{10}\sum_{i=1}^{10}(X_i-\bar{X})^2 \le 2.3\sigma^2\right),$$ 其中

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i.$$

解：因 $X \sim N(\mu,\sigma^2)$,故 $\dfrac{X_i-\mu}{\sigma} \sim N(0,1)$.

于是 $$\sum_{i=1}^{10}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(10).$$

$$P(0.26\sigma^2 \le \frac{1}{10}\sum_{i=1}^{10}(X_i-\mu)^2 \le 2.3\sigma^2)$$

$$= P(2.6 \le \sum_{i=1}^{10}\left(\frac{X_i-\mu}{\sigma}\right)^2 \le 23)$$

$$= P(2.6 \le \chi^2(10) \le 23) = 0.98.$$

由样本方差的抽样分布有

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

$$P(0.26\sigma^2\le\frac{1}{10}\sum_{i=1}^{10}(X_i-\bar{X})^2\le 2.3\sigma^2)$$

$$=P(2.6\le\sum_{i=1}^{10}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2\le 23)$$

$$=P(2.6\le\chi^2(10-1)\le 23)=0.97.$$

例题、设 $\overline{X}$ 与 S^2 分别为来自正态总体 $N(u,\sigma^2)$ 的样本均值和方差，则下列说法正确的是________。

$A.\left(\overline{X}\right)^2 = S^2$ $B.\dfrac{\left(\overline{X}\right)^2}{S^2} \sim F(1,n-1)$

$C.\left(\overline{X}\right)^2$ 与 S^2 相关 $D.\left(\overline{X}\right)^2$ 与 S^2 不相关

例题、设总体 $X \sim N(1,3)$，X_1, X_2, X_3 为来自 X 的样本，记 $\overline{X} = \dfrac{1}{3}\left(X_1 + X_2 + X_3\right)$，则由抽样分布定理，概率 $P(0 < \overline{X} < 2) =$ ____________。

$A.2-2\Phi(1)$ $B.2\Phi(1)-1$ $C.\Phi(1)$ $D.1-\Phi(1)$