正态总体的抽样分布

下面给出几个有关正态总体统计量的分布定理

定理1 设（$X_1, X_2, \ldots, X_n$）是总体$\xi \sim N(\mu,\sigma^2)$的样本，$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$，

即有 $\frac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$

证： $\bar{X}$ 是$X_1, X_2, \ldots, X_n$的线性函数，而$X_1, X_2, \ldots, X_n$中的每一个X_i都服从正态分布而且相互独立，在概率论中，我们已经知道：相互独立的正态分布的线性组合仍为正态分布，因此，这里的$\bar{X}$服从正态分布。

由前面2.3节的定理2.1可知，$E\bar{X} = E\xi = \mu, D\bar{X} = \frac{D\xi}{n} = \frac{\sigma^2}{n}$，

所以有 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$，即有 $\frac{\bar{X}-\mu}{\sigma}\sqrt{n} = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

为了证明下面的定理，先给出一些定义

定义 设$X_1, X_2, \ldots, X_n$是n个随机变量. 称

$$X=\begin{bmatrix} X_1 \\ \ldots \\ X_n \end{bmatrix}$$ 为n维**随机向量**

$$EX=\begin{bmatrix} EX_1 \\ \ldots \\ EX_n \end{bmatrix}$$ 为X的**数学期望**

$$DX=E[(X-EX)(X-EX)^T]=E\left[\begin{bmatrix} X_1-EX_1 \\ \ldots \\ X_n-EX_n \end{bmatrix}\begin{bmatrix} X_1-EX_1 & \ldots & X_n-EX_n \end{bmatrix}\right]$$

$$=\begin{bmatrix} E(X_1-EX_1)^2 & \ldots & E[(X_n-EX_n)(X_1-EX_n)] \\ \ldots & \ldots. & \ldots \\ E[(X_n-EX_n)(X_1-EX_n)] & \ldots. & E(X_n-EX_n)^2 \end{bmatrix}$$

$$=\begin{bmatrix} DX_1 & \ldots & Cov(X_1,X_n) \\ \ldots & \ldots & \ldots \\ Cov(X_n,X_1) & \ldots & DX_n \end{bmatrix}$$ 为X的 **协方差矩阵**

定理2 设$X=\begin{bmatrix} X_1 \\ \cdots \\ X_n \end{bmatrix}$为$n$维随机向量，A=$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$为

$n\times n$常数矩阵，则有$E(AX)=AEX, D(AX)=A(DX)A^T$

(类似于$E(a\xi)=aE\xi, D(a\xi)=a^2D\xi$)

证：

因为$AX=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} X_1 \\ \cdots \\ X_n \end{bmatrix}=\begin{bmatrix} \sum_{j=1}^{n} a_{1j}X_j \\ \cdots \\ \sum_{j=1}^{n} a_{nj}X_j \end{bmatrix}$

所以$E(AX)=\begin{bmatrix} \sum_{j=1}^{n} a_{1j}EX_j \\ \cdots \\ \sum_{j=1}^{n} a_{nj}EX_j \end{bmatrix}=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} EX_1 \\ \cdots \\ EX_n \end{bmatrix}=AEX$

$$E(X^T A^T) = E[(AX)^T] = [E(AX)]^T = (AEX)^T$$

$$=(\mathrm{EX})^T A^T = E(X^T)A^T$$

$$D(AX) = E[(AX - AEX)(AX - AEX)^T]$$

$$= E[A(X - EX)(X - EX)^T A^T]$$

$$= AE[(X - EX)(X - EX)^T]A^T$$

$$= ADXA^T$$

定理 3 （Cochran定理）

说明： 设$X_i \sim N(0,1), i=1,2,\ldots,n$相互独立,$\sum_{i=1}^{n} X_i^2 = Q = Q_1 + Q_2 + \ldots + Q_k$.

其中，$Q_1, Q_2, \ldots, Q_k$都是$X_1, X_2, \ldots, X_n$的线性组合，已知Q_j的自由度为$f_j, j=1,2,\ldots,k$.

则$Q_j \sim \chi^2(f_j), j=1,2,\ldots,k, Q_1, Q_2, \ldots, Q_k$相互独立的充分必要条件是$f_1 + f_2 + \ldots + f_k = n$.

"Q_j"的自由度定义如下：如果Q_j是m项线性组合的平方和，而这m项线性组合又满足l个相互独立的线性关系式，Q_j的自由度就是$m-l$.

例如，$Q_j=(X_1-X_2)^2+(X_2-X_3)^2+(X_3-X_1)^2$,它是3项的平方和，但这2项又满足1个线性关系式:$(X_1-X_2)+(X_2-X_3)+(X_3-X_1)=0$,所以，$Q_j$的自由度是3-1=2.

又例如，$Q_j=(X_1+3X_2)^2+(X_1+3X_2)^2$.可以把$Q_j$看作是2项的平方和，但这2项又满足1个线性关系式：

$(X_1+3X_2)-(X_1+3X_2)=0$,所以，Q_j的自由度是2−1=1.

也可以把Q_j看作Q_j=2$(X_1+3X_2)^2$=$(\sqrt{2}\mathrm{X}_1+3\sqrt{2}\mathrm{X}_2)^2$

它是1项的平方和，满足0个线性关系式，所以，Q_j的自由度是1-0=0.

尽管看法不同，Q_j的自由度始终保持不变,

证明： 先证必要性

因为$Q_j \sim \chi^2(f_j), j=1,2,\ldots,k, Q_1,Q_2,\ldots,Q_k$相互独立，由$\chi^2$分布的可加性知$Q=Q_1+Q_2+\ldots+Q_k \sim \chi^2(f_1+f_2+\ldots+f_k)$

另一方面，因为$X_i \sim N(0,1), i=1,2,\ldots,n,$相互独立，由$\chi^2$分布的定义可知$Q=\sum_{i=1}^{n} X_i^2 \sim \chi^2(n).$

对比上面两式，可知$f_1+f_2+\ldots+f_k=n$

充分性的证明,要用到线性代数正交变换知识(略)

定理4　（Fisher定理）

设$(X_1,X_2,\dots,X_n)$是总体$\xi\sim N(\mu,\sigma^2)$的样本. $\bar{X}$ 是样本均值，

S^2是样本方差，S^{*2}是样本方差，则有

(1) $\bar{X}$ 与S^{*2}(或与S^2)相互独立；　　(2) $\dfrac{(n-1)S^{*2}}{\sigma^2}=\dfrac{nS^2}{\sigma^2}\sim\chi^2(n-1)$

证： 因为$(X_1,X_2,\dots,X_n)$是$\xi\sim N(\mu,\sigma^2)$的样本，所以，$X_i\sim N(\mu,\sigma^2)$,

$\dfrac{X_i-\mu}{\sigma}\sim N(0,1), i=1,2,\dots,n,$相互独立.

$$Q=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}+\bar{X}-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2+2\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)\left(\frac{X_i-\mu}{\sigma}\right)+\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2+0+n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2=Q_1+Q_2$$

$$\left(\text{其中}\sum_{i=1}^{n}\frac{X_i-\bar{X}}{\sigma}=\frac{\sum_{i=1}^{n}X_i-n\bar{X}}{\sigma}=0\right)$$

$Q_1=\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})^2$是$n$项的平方和，但这n项又满足1个线性关系式：

$\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})=\frac{\sum_{i=1}^{n}X_i-n\bar{X}}{\sigma}=0$.所以，$Q_1$的自由度$f_1=n-1$

$Q_2=n\sum_{i=1}^{n}(\frac{\bar{X}-\mu}{\sigma})^2=\sum_{i=1}^{n}(\frac{\bar{X}-\mu}{\sigma}\sqrt{n})^2$是1项的平方和，满足0个线性关系式

所以，Q_2的自由度$f_1=n-1$.因为$f_1+f_2=(n-1)+1=n$

$$Q_1=\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})}{\sigma^2}=\frac{(n-1)S^{*2}}{\sigma^2}=\frac{nS^2}{\sigma^2}\sim\chi^2(n-1)$$

$$Q_2=(\frac{\bar{X}-\mu}{\sigma}\sqrt{n})^2\sim\chi^2(n-1)$$

而且Q_1与Q_2相互独立，即$\bar{X}$与S^{*2}(或与S^2)相互独立

定理5 设$(X_1, X_2, \dots, X_n)$是总体$\xi \sim N(\mu, \sigma^2)$的样本. $\bar{X}$ 是样本均值，

S^2是样本方差，则有$\dfrac{nS^2 + n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2(n)$

证明： 因$\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$, 根据$\chi^2$分布的定义有

$$\frac{n(\bar{X}-\mu)^2}{\sigma^2} = \left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right)^2 \sim \chi^2(1)$$

$\dfrac{nS^2}{\sigma^2} \sim \chi^2(n-1)$, 而且 $\bar{X}$ 与S^2相互独立，即$\dfrac{n(\bar{X}-\mu)^2}{\sigma^2}$与$\dfrac{nS^2}{\sigma^2}$相互独立

$$\Rightarrow \quad \frac{nS^2 + n(\bar{X}-\mu)^2}{\sigma^2} = \frac{n(\bar{X}-\mu)^2}{\sigma^2} + \frac{nS^2}{\sigma^2} \sim \chi^2(n)$$

定理6 设 $(X_1,X_2,\cdots,X_n)$ 是总体 $\xi\sim N(\mu,\sigma^2)$ 的样本，$\bar{X}$ 是样本均值，S^* 是修正样本标准差，

则有 $\dfrac{\bar{X}-\mu}{S^*}\sqrt{n}\sim t(n-1)$ 。

分析 根据 t 分布的定义，$\dfrac{\boxed{N(0,1)}}{\sqrt{\boxed{\chi^2(n-1)}\Big/(n-1)}}\sim t(n-1)$ 。

证 由定理 2.5 可知 $\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$ 。

由定理 2.8 可知 $\dfrac{(n-1)S^{*2}}{\sigma^2}\sim\chi^2(n-1)$ ，而且 $\bar{X}$ 与 S^{*2} 相互独立，即 $\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n}$ 与 $\dfrac{(n-1)S^{*2}}{\sigma^2}$ 相互独立，所以，根据 t 分布的定义，有 $\dfrac{\bar{X}-\mu}{S^*}\sqrt{n}=\dfrac{\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\dfrac{(n-1)S^{*2}}{\sigma^2}\Big/(n-1)}}\sim t(n-1)$ 。

定理7 设 $(X_1,X_2,\cdots,X_m)$ 是总体 $\xi\sim N(\mu_1,\sigma_1^2)$ 的样本，$(Y_1,Y_2,\cdots,Y_n)$ 是总体 $\eta\sim N(\mu_2,\sigma_2^2)$ 的样本，两个样本相互独立，$\bar{X}$，$\bar{Y}$ 是 ξ，η 的样本均值，则有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1)\text{ 。}$$

证： 有定理1知 $\bar{X}\sim N(\mu_1,\frac{\sigma_1^2}{m})$，$\bar{Y}\sim N(\mu_2,\frac{\sigma_2^2}{n})$，它们相互独立（因为两个样本相互独立），所以 $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})$，即有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1)\text{ 。}$$

定理8 设 $X_1, X_2, \cdots, X_m$）是总体 $\xi \sim N(\mu_1, \sigma_1^2)$ 的样本，（$Y_1, Y_2, \cdots, Y_n$）是总体 $\eta \sim N(\mu_2, \sigma_2^2)$ 的样本，其中 $\sigma_1 = \sigma_2$ ，两个样本相互独立，$\overline{X}$，$\overline{Y}$ 是 ξ，η 的样本均值，S_x^{*2}，S_y^{*2} 是 ξ，η 的修正样本方差，则有

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$$

，其中，$S_w = \sqrt{\dfrac{(m-1)S_x^{*2}+(n-1)S_y^{*2}}{m+n-2}}$ 。

分析

根据 t 分布的定义，$\dfrac{\boxed{N(0,1)}}{\sqrt{\boxed{\chi^2(m+n-2)}\big/(m+n-2)}} \sim t(m+n-2)$ 。$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma^2}{m}+\dfrac{\sigma^2}{n}}} \sim N(0,1)$ 。

$$\frac{(m-1)S_x^{*2}}{\sigma^2} \sim \chi^2(m-1) \text{ ，} \frac{(n-1)S_y^{*2}}{\sigma^2} \sim \chi^2(n-1) \text{ ，} \frac{(m-1)S_x^{*2}}{\sigma^2}+\frac{(n-1)S_y^{*2}}{\sigma^2} \sim \chi^2(m+n-2) \text{ 。}$$

$$\sqrt{\left[\frac{(m-1)S_x^{*2}}{\sigma^2}+\frac{(n-1)S_y^{*2}}{\sigma^2}\right]\Big/(m+n-2)} = \sqrt{\frac{(m-1)S_x^{*2}+(n-1)S_y^{*2}}{m+n-2}} = S_w$$

定理9 设 $(X_1, X_2, \cdots, X_m)$ 是总体 $\xi \sim N(\mu_1, \sigma_1^2)$ 的样本，$(Y_1, Y_2, \cdots, Y_n)$ 是总体 $\eta \sim N(\mu_2, \sigma_2^2)$ 的样本，两个样本相互独立，$\bar{X}$，$\bar{Y}$ 是 ξ，η 的样本均值，S_x^2，S_y^2 是 ξ，η 的样本方差，则有

$$\frac{[S_x^2 + (\bar{X} - \mu_1)^2] \;/\sigma_1^2}{[S_y^2 + (\bar{Y} - \mu_2)^2] \;/\sigma_2^2} \sim F(m, n)$$ 。

分析 根据 F 分布的定义，有 $\dfrac{\boxed{\chi^2(m)}/m}{\boxed{\chi^2(n)}/n} \sim F(m,n)$ 。

证： 有定理5知 $\dfrac{mS_x^2 + m(\bar{X} - \mu_1)^2}{\sigma_1^2} \sim \chi^2(m)$，$\dfrac{nS_y^2 + n(\bar{Y} - \mu_2)^2}{\sigma_2^2} \sim \chi^2(n)$，它们相互独立（因为两个样本相互独立），所以，根据 F 分布的定义，有

$$\frac{[S_x^2 + (\bar{X} - \mu_1)^2] \;/\sigma_1^2}{[S_y^2 + (\bar{Y} - \mu_2)^2] \;/\sigma_2^2} = \frac{\dfrac{mS_x^2 + m(\bar{X} - \mu_1)^2}{\sigma_1^2}\Big/ m}{\dfrac{nS_y^2 + n(\bar{Y} - \mu_2)^2}{\sigma_2^2}\Big/ n} \sim F(m, n)$$ 。

定理10 设（$X_1, X_2, \cdots, X_m$）是总体 $\xi \sim N(\mu_1, \sigma_1^2)$ 的样本，（$Y_1, Y_2, \cdots, Y_n$）是总体 $\eta \sim N(\mu_2, \sigma_2^2)$ 的样本，两个样本相互独立，S_x^{*2}，S_y^{*2} 是 ξ，η 的修正样本方差，则有 $\dfrac{S_x^{*2}/\sigma_1^2}{S_y^{*2}/\sigma_2^2} \sim F(m-1, n-1)$ 。

分析 根据 F 分布的定义，有 $$\frac{\boxed{\chi^2(m-1)}\Big/(m-1)}{\boxed{\chi^2(n-1)}\Big/(n-1)} \sim F(m-1, n-1)\ 。$$

证： 有定理10知 $\dfrac{(m-1)S_x^{*2}}{\sigma^2} \sim \chi^2(m-1)$ ，$\dfrac{(n-1)S_y^{*2}}{\sigma^2} \sim \chi^2(n-1)$ ，它们相互独立（因为两个样本相互独立），所以，根据 F 分布的定义，有

$$\frac{S_x^{*2}/\sigma_1^2}{S_y^{*2}/\sigma_2^2} = \frac{\dfrac{(m-1)S_x^{*2}}{\sigma_1^2}\Big/(m-1)}{\dfrac{(n-1)S_y^{*2}}{\sigma_2^2}\Big/(n-1)} \sim F(m-1, n-1)\ 。$$

例 4 设随机变量 X 和 Y 都服从标准正态分布，则

（A） $X+Y$ 服从正态分布　　（B） X^2+Y^2 服从 χ^2 分布

（C） X^2 和 Y^2 都服从 χ^2 分布（D） X^2/Y^2 服从 F 分布

分析：如果加上 X，Y 相互独立的条件四个选项都对.

取 $Y=-X$ 可排除（A），（B），（D）。

答案：（C）