样本均值与方差性质证明

Tuesday, December 19, 2023 10:31 AM

对于任意分布的总体，设其有均值 μ，方差 σ^2，而 $X_1, X_2, \cdots X_n$ 是来自该总体的一个样本，而样本均值的样本方差为 $\overline{X}$、S^2.

从而：

(1) $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \times n\mu = \mu$

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow[\text{服从样本分布}]{\text{相互独立}} \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

(2).

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n} E\left((X_i - \overline{X})^2\right)$$

需要说明：$(X_i - \overline{X})^2$ 计算数学期望时，注意 $\overline{X}$ 的期望计算。

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - 2n E(X_i\overline{X}) + nE(\overline{X}^2)\right]$$

其中：$E(\overline{X}) = \mu$，　而：$D(\overline{X}) = E(\overline{X}^2) - E^2(\overline{X}) = \frac{\sigma^2}{n} \to$ 得：$E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$.

而：$E(X_i^2) = D(X_i) + \mu^2 = \sigma^2 + \mu^2$

显然 $E(X_i\overline{X})$ 不易求，有：$\sum_{i=1}^{n}(2X_i\overline{X} - \overline{X}^2) = 2\overline{X}\cdot n\overline{X} - n\overline{X}^2 = n\overline{X}^2$

故：$E = \frac{1}{n-1}\left[nE(X_i^2) - nE(\overline{X}^2)\right] = \frac{1}{n-1}\left[n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right] = \sigma^2$.

即：$E(S^2) = \sigma^2$. ☆

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

定理 ① 对于正态总体有：$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

② 对于正态总体 $N(\mu, \sigma^2)$ 的样本，设 $\overline{X}$、S^2 为均值和样本方差

则：$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，　$\overline{X}$ 与 S^2 相互独立.

有：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left((X_i - \overline{X})^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

故有：$\frac{1}{\sigma^2}(n-1)S^2 = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)$ 由于 $X_i \sim N(\mu, \sigma^2)$，

标准化：

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

故：对：$\frac{X_i - \overline{X}}{\sigma}$ 显 $E\left(\frac{X_i - \overline{X}}{\sigma}\right) = \frac{1}{\sigma}[\mu - \mu] = 0$

$$D\left(\frac{X_i - \overline{X}}{\sigma}\right) = \frac{1}{\sigma^2}\left[D(X_i) - D(\overline{X})\right]$$

即：$\frac{X_i - \overline{X}}{\sigma} \sim N\left(0, \frac{n-1}{n}\right)$

$$\frac{1}{\sigma^2}\left[\sigma^2 - \frac{1}{n}\sigma^2\right] = \frac{n-1}{n}$$

$D\left(\frac{X_i-\bar{X}}{\sigma}\right)=\frac{1}{\sigma^2}\left[D(X_i)-D(\bar{X})\right]$

$=\frac{1}{\sigma^2}\left[\sigma^2-\frac{1}{n}\sigma^2\right]=\frac{n-1}{n}$

即：$\frac{X_i-\bar{X}}{\sigma}\sim N\left(0,\frac{n-1}{n}\right)$，

故有：$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)$ 为n个方差 $\frac{n-1}{n}$ 正态分布相加、

正解：此时，凑：$\frac{X-\mu}{\sigma}\sim N(0,1)$，我们令 $Z_i=\frac{X_i-\mu}{\sigma}$，$i=1,2,3,\cdots\cdots n$.

则有：$\bar{Z}=\frac{1}{n}\sum_{i=1}^{n}Z_i=\frac{\bar{X}-\mu}{\sigma}$，显然：$E(\bar{Z})=0$，$D(\bar{Z})=\frac{1}{n}$.

则：

$$\frac{(n-1)S^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\longrightarrow=\frac{1}{\sigma^2}\left[(X_i-\mu)-(\bar{X}-\mu)\right]^2=\frac{1}{\sigma^2}\left[Z-\bar{Z}\right]^2$$

从而有：

$$\sum_{i=1}^{n}\left[Z_i-\bar{Z}\right]^2=\underbrace{\left[\sum_{i=1}^{n}Z_i^2-n\bar{Z}^2\right]}$$ ☆

取一n阶正交矩阵 $A=(a_{ij})_{n\times n}$，第一行元素均为 $\frac{1}{\sqrt{n}}$，并作正交变换

其中Y，Z为列向量

$Y=AZ$，从而有：

$Y_i=\sum_{j=1}^{n}a_{ij}Z_j$，$i=1,2,\cdots n$，显然有 $Z\sim N(0,1)$，

则 Y_1，Y_2，$\cdots$ 仍然为正态变量，且：

$E(Y_i)=E\left(\sum_{j=1}^{n}a_{ij}Z_j\right)=0$

此时，考虑 Z_{ij} 两两不相关，则有：

$Cov(Z_i,Z_j)=\delta_{ij}$，故：$Cov(Y_i,Y_j)=Cov\left(\sum_{k=1}^{n}a_{ik}Z_k,\sum_{k=1}^{n}a_{jk}Z_k\right)$

$Cov(Y_i,Y_j)=\underbrace{\sum_{k=1}^{n}\sum_{l=1}^{n}}a_{ik}a_{jl}\delta_{ij}$ $\Rightarrow$ 由正交矩阵性质：$\sum_{k=1}^{n}\sum_{l=1}^{n}a_{ik}a_{jl}=\begin{cases}1 & i=j\\0 & i\neq j\end{cases}$

$=\delta_{ij}$

从而：由 $Y_1=\sum_{j=1}^{n}a_{ij}Z_j=\sum_{j=1}^{n}\frac{1}{\sqrt{n}}Z_j=\frac{n}{\sqrt{n}}\bar{Z}=\sqrt{n}\bar{Z}$，

则有：

$\sum_{j=1}^{n}Y_i^2=Y^TY=(AZ)^T(AZ)=Z^TA^TAZ=Z^TZ=\sum_{i=1}^{n}Z_i^2$

故有：$\frac{(n-1)}{\sigma^2}S^2=\sum_{i=1}^{n}(Z_i-\bar{Z})^2=\underbrace{\sum_{i=1}^{n}Z_i^2}_{\text{代换}\sum_{i=1}^{n}Y_i^2}-n\bar{Z}^2=\sum_{i=1}^{n}Y_i^2-Y_1^2=\sum_{i=2}^{n}Y_i^2$

由于 Y_i 服从 $N(0,1)$，则 $\sum_{i=2}^{n}Y_i^2\sim\chi^2(n-1)$，得证.

另外：$\bar{X}=\sigma\bar{Z}+\mu$，$=\sigma\frac{Y_1}{\sqrt{n}}+\mu$，而 $S^2=\frac{\sigma^2}{n-1}\sum_{i=2}^{n}Y_i^2$.

另外：$\bar{X}=\sigma\bar{Z}+\mu,=\sigma\frac{Y_1}{\sqrt{n}}+\mu,$ 而 $S^2=\frac{\sigma^2}{n-1}\sum_{i=2}^{n}Y_i^2$。

故：$\bar{X}$ 仅依赖于 Y_1，S^2 仅依赖于 $Y_2,Y_3,\cdots Y_n$，即：$\bar{X}$ 与 S^2 相互独立。

③. 设 $X_1,X_2,\cdots X_n$ 为来自 $N(\mu,\sigma^2)$ 的样本，而 $\bar{X}$、S^2 分别为样本的均值与方差，则：

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

证：

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}=\frac{\bar{X}-\mu}{\sigma}\cdot\frac{\sigma\sqrt{n}}{S}=\frac{\frac{\bar{X}-\mu}{\sigma}}{\frac{S}{\sigma\sqrt{n}}}$$

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

则：$\underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{\text{取为}M}\sim N(0,1)$

在②中已证得：$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

则：

$$=\frac{M}{\underbrace{\sqrt{\frac{(n-1)S^2}{\sigma^2}}}_{\text{取为}N}\cdot\frac{1}{\sqrt{n-1}}}=\frac{M}{N/\sqrt{n-1}}$$

其中 $M\sim N(0,1)$，$N\sim\chi^2(n-1)$

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

④. 设有正态总体 X,Y，均值分别为 $\bar{X},\bar{Y}$，方差 S_1^2,S_2^2，

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

证：

(1) $\frac{S_1^2}{\sigma_1^2}\Big/\frac{S_2^2}{\sigma_2^2}$，有：$\frac{(n_1-1)S^2}{\sigma^2}\sim\chi^2(n_1-1)$，则有：

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

(2). 若有 $\sigma_1^2=\sigma_2^2=\sigma^2$ 则：

$$(\bar{X}-\bar{Y})-(\mu_1-\mu_2)=(\bar{X}-\mu_1)-(\bar{Y}-\mu_2)$$

$$\sim N(0,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

此时：有：$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)\xrightarrow{\text{条件}}=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$

此时：有：$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1) \xrightarrow{\text{条件}} = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$

又：$\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$，$\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$，由$\chi^2$分布可加性：

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

∴可构造：$\dfrac{(\bar{X}-\mu_1)-(\bar{Y}-\mu_2)}{\cancel{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot\sqrt{\frac{1}{n_1+n_2-2}\left[\frac{(n_1-1)S_1^2}{\cancel{\sigma^2}}+\frac{(n_2-1)S_2^2}{\cancel{\sigma^2}}\right]}}$

$$= \frac{(\bar{X}-\mu_1)-(\bar{Y}-\mu_2)}{\underbrace{\frac{1}{\sqrt{(n_1+n_2-2)}}\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}_{S_w}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$