第一章

统计学的基本概念(3)

§1.6　抽样分布

1.单正态总体下的抽样分布

2.次序统计量的分布

3.双正态总体下的抽样分布

统计量是样本的函数，故也是一个随机变量．统计量的分布即称为抽样分布.

1.单正态总体下的抽样分布

定理1.9　设$(X_1, X_2, \cdots, X_n)$是取自正态总体$N(\mu, \sigma^2)$的一个简单随机样本，则有

(1) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$，即$\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \sim N(0,1)$；

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

(3) $\bar{X}$与$S^2(S^{*2})$相互独立.

准备：①设 $Y_i=\dfrac{X_i-\mu}{\sigma}, i=1,2,\cdots,n$，则 $Y_1,\cdots,Y_n$ 独立同分布，且都服从标准正态分布；

②设 $Z_i=c_{i1}Y_1+c_{i2}Y_2+\cdots+c_{in}Y_n$，　$i=1,2,\cdots,n-1$，

而 $Z_n=\dfrac{1}{\sqrt{n}}Y_1+\dfrac{1}{\sqrt{n}}Y_2+\cdots+\dfrac{1}{\sqrt{n}}Y_n$ ，

且 $C\triangleq\begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \cdots & \vdots \\ c_{n-1,1} & \cdots & c_{n,1} \\ \dfrac{1}{\sqrt{n}} & \cdots & \dfrac{1}{\sqrt{n}} \end{pmatrix}$ 为正交阵，即 $C^{-1}=C^T$.

因此 $Z_1,\cdots,Z_n$ 也相互独立，且都服从标准正态分布.

③ 记 $Y=(Y_1,Y_2,\cdots,Y_n)^T$，$Z=(Z_1,Z_2\cdots,Z_n)^T$，则上述正交变换可简记为 $Z=CY$；注意到 $C^{-1}=C^T$，

则有 $\sum_{i=1}^{n}Z_i^2=Z^TZ=\left(CY\right)^T\left(CY\right)=Y^TY=\sum_{i=1}^{n}Y_i^2$；

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

$$=\frac{\sqrt{n}}{\sigma}\left\{\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\cdot n\mu\right\}=\frac{\sqrt{n}}{\sigma n}\left\{\sum_{i=1}^{n}X_i-n\mu\right\}$$

$$=\frac{1}{\sigma\sqrt{n}}\left\{\sum_{i=1}^{n}(X_i-\mu)\right\}=\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}\right\}$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i=Z_n\sim N(0,1)\;;$$ 即证.

证明(2) $$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2$$

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

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

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

$$=\sum_{i=1}^{n}Z_i^2-Z_n^2=\sum_{i=1}^{n-1}Z_i^2\sim\chi^2(n-1)$$

注意到 $\sum_{i=1}^{n}X_i=n\bar{X}$

证明(3) 由前面的证明可知 $S^2(S^{*2})$ 只和 $Z_1,\cdots,Z_{n-1}$ 有关,

而 $\overline{X}$ 只与 Z_n 有关, 因此 $\overline{X}$ 与 $S^2(S^{*2})$ 相互独立.

推论1.1 设$(X_1, X_2, \cdots, X_n)$是取自正态总体$N(\mu, \sigma^2)$的简单随机样本，则$\sqrt{n}\dfrac{\bar{X}-\mu}{S^*} = \sqrt{n-1}\dfrac{\bar{X}-\mu}{S} \sim t(n-1)$.

证明：由定理1.9的(1)(3)(4)知

$$\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1), \quad \frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2(n-1),$$

且 $\bar{X}$ 与 S^{*2} 相互独立，因此

$$\frac{\sqrt{n}\dfrac{\bar{X}-\mu}{\sigma}}{\sqrt{\dfrac{(n-1)S^{*2}}{\sigma^2(n-1)}}} = \sqrt{n}\frac{\bar{X}-\mu}{S^*} \sim t(n-1).$$

例22　设总体 $X \sim N(40, 5^2)$，试求解下列各题:

(1)抽取容量为 36 的样本，求概率 $P(38 \le \bar{X} \le 43)$；

(2)抽取容量为 64 的样本，求概率 $P(|\bar{X} - 40| < 1)$；

(3)问样本容量 n 多大时，才能使 $P(|\bar{X} - 40| < 1) = 0.95$ ？

解　(1)由 $X \sim N(40, 5^2)$，$n = 36 \Rightarrow \bar{X} \sim N(40, \frac{25}{36})$，

所以 $P(38 \le \bar{X} \le 43) = \Phi(\frac{43-40}{5/6}) - \Phi(\frac{38-40}{5/6})$

$$= \Phi(3.6) - 1 + \Phi(2.4) \approx 0.9918.$$

(2)抽取容量为 **64** 的样本，求概率 $P(|\bar{X}-40|<1)$；

由 $X \sim N(40,25),\, n=64 \Rightarrow \bar{X} \sim N(40,\dfrac{25}{64})$，

即 $\dfrac{\bar{X}-40}{5/8} \sim N(0,1)$，所以

$$P(|\bar{X}-40|<1)=P\left(\left|\frac{\bar{X}-40}{5/8}\right|<\frac{1}{5/8}\right)$$

$$=P\left(\left|\frac{\bar{X}-40}{5/8}\right|<1.6\right)=2\Phi(1.6)-1\approx 0.8904.$$

(3)问样本容量 n 多大时，才能使 $P(|\bar{X}-40|<1)=0.95$?

由 $X \sim N(40,25) \Rightarrow \sqrt{n}\cdot\dfrac{\bar{X}-40}{5} \sim N(0,1)$，因此

$$0.95 = P(|\bar{X}-40|<1) = P\left(\left|\frac{\bar{X}-40}{5/\sqrt{n}}\right| < \frac{1}{5/\sqrt{n}}\right)$$

$$= P\left(\left|\frac{\bar{X}-40}{5/\sqrt{n}}\right| < \frac{\sqrt{n}}{5}\right) = 2\Phi\left(\frac{\sqrt{n}}{5}\right)-1 \Rightarrow \Phi\left(\frac{\sqrt{n}}{5}\right) = 0.975$$

$$\Rightarrow \frac{\sqrt{n}}{5} = u_{0.975} = 1.96 \Rightarrow n = 96.04,$$

所以，取 $n = 97$.

例23　设$(X_1, X_2, \cdots, X_n)$是取自正态总体$N(0,1)$的样本，

记$T = \overline{X}^2 - \dfrac{1}{n}S^{*2}$，其中$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i, S^{*2} = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$，

试计算方差 $D(T)$.

解　由题意知$\sqrt{n}\,\overline{X} \sim N(0,1)$，即有 $n\overline{X}^2 \sim \chi^2(1)$. 又

$(n-1)S^{*2} \sim \chi^2(n-1)$，且$\overline{X}$与$S^{*2}$相互独立，故

$$D(T) = \frac{1}{n^2}D(n\overline{X}^2) + \frac{1}{n^2(n-1)^2}D((n-1)S^{*2})$$

$$= \frac{1}{n^2}\cdot 2 + \frac{2(n-1)}{n^2(n-1)^2} = \frac{2}{n(n-1)}.$$

2.次序统计量的分布

定理1.3 设总体 X 具有分布函数 $F(x)$ 及密度函数 $f(x)$，则最小次序统计量具有密度函数

$$f_1(y) = n\left[1 - F(y)\right]^{n-1} f(y);$$

最大次序统计量具有密度函数

$$f_n(y) = nF(y)^{n-1} f(y).$$

记 $U=\max(X_1,X_2,\cdots,X_n)=X_{(n)}$,

则 $F_U(y)=P(U\le y)=P(\max(X_1,\cdots,X_n)\le y)$

$=P(X_i\le y,i=1,\cdots,n)=P(X_1\le y)\cdots P(X_n\le y)$

$=F_{X_1}(y)\cdots F_{X_n}(y)=F^n(y)$

求导即得 $X_{(n)}\sim f_n(y)=nF^{n-1}(y)f(y)$;

记 $V=\min(X_1,X_2,\cdots,X_n)=X_{(1)}$,

$$F_V(y) = P(V \le y) = P(\min(X_1, \cdots, X_n) \le y)$$

$$= 1 - P(\min(X_1, \cdots, X_n) > y)$$

$$= 1 - P(X_i > y, i = 1, \cdots, n)$$

$$= 1 - P(X_1 > y) \cdots P(X_n > y)$$

$$= 1 - (1 - P(X_1 \le y)) \cdots (1 - P(X_n \le y))$$

$$= 1 - (1 - F_{X_1}(y)) \cdots (1 - F_{X_n}(y)) = 1 - (1 - F(y))^n$$

求导即得 $X_{(1)} \sim f_1(y) = n(1 - F(y))^{n-1} f(y)$.

例24 设 $X_1,\cdots,X_n$ 是来自均匀总体 $R(0,\theta)$ 的样本，则

(1) $y\in(0,\theta)$ 时，$F_{X_{(n)}}(y)=(\frac{y}{\theta})^n$，故

$$f_n(y)=\begin{cases}\dfrac{ny^{n-1}}{\theta^n}, & 0<y<\theta,\\ 0, & \text{其余}.\end{cases}$$

(2) $y\in(0,\theta)$ 时，$F_{X_{(1)}}(y)=1-(1-\frac{y}{\theta})^n$，故

$$f_1(y)=\begin{cases}\dfrac{n(\theta-y)^{n-1}}{\theta^n}, & 0<y<\theta,\\ 0, & \text{其余}.\end{cases}$$

3.双正态总体下的抽样分布:

设$(X_1, X_2 \cdots, X_m)$是取自正态总体 $N(\mu_1, \sigma_1^2)$ 的一个样本, $(Y_1, Y_2 \cdots, Y_n)$ 是取自正态总体 $N(\mu_2, \sigma_2^2)$ 的另一个样本, 且 $(X_1, X_2, \cdots, X_m)$ 与 $(Y_1, Y_2, \cdots, Y_n)$ 相互独立;

记 $\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$;

$$S_1^{*2} = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, S_2^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2;$$

$$S_w^2 = \frac{1}{m+n-2}[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2]$$

$$= \frac{m-1}{m+n-2}S_1^{*2} + \frac{n-1}{m+n-2}S_2^{*2}$$

(1) 若 σ_1^2, σ_2^2 已知, 则 $G \triangleq \dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$;

证　由 $\bar{X} \sim N(\mu_1, \dfrac{\sigma_1^2}{m}), \bar{Y} \sim N(\mu_2, \dfrac{\sigma_2^2}{n})$, 且 $\bar{X}, \bar{Y}$ 相互独立,

可推出 $\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n})$, 标准化后即得.

(2) 若 $\sigma_1^2=\sigma_2^2 \triangleq \sigma^2$ 未知，则

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

证　因为 $\bar{X}-\bar{Y}$ 与 S_w^2 相互独立，且由卡方分布可加性、

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

再结合(1)的结论由 t-分布定义即得.

(3) 若 μ_1, μ_2 已知，则 $G \triangleq \dfrac{\sum_{i=1}^{m}(X_i-\mu_1)^2/m\sigma_1^2}{\sum_{i=1}^{n}(Y_i-\mu_2)^2/n\sigma_2^2} \sim F(m,n)$;

证　因为 $\sum_{i=1}^{m}\left(\dfrac{X_i-\mu_1}{\sigma_1}\right)^2 \sim \chi^2(m)$, $\sum_{i=1}^{n}\left(\dfrac{Y_i-\mu_2}{\sigma_2}\right)^2 \sim \chi^2(n)$,

且两个总体相互独立，由 **F**-分布的定义即得上述结论.

(4) 若 μ_1,μ_2 未知，则 $G \triangleq \dfrac{S_1^{*}/\sigma_1^2}{S_2^{*}/\sigma_2^2} \sim F(m-1,n-1)$.

证　因为 $\dfrac{m-1}{\sigma_1^2}S_1^{*2} \sim \chi^2(m-1)$，$\dfrac{n-1}{\sigma_2^2}S_2^{*2} \sim \chi^2(n-1)$，

由独立性及 **F**-分布的定义即得.