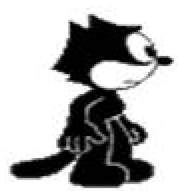
第五章 样本及抽样分布

主讲教师 邓小艳



第五章　样本及抽样分布

§5-1　随机样本

§5-2　描述统计

§5-3　统计量

§5-4　抽样分布

抽样分布——统计量的分布

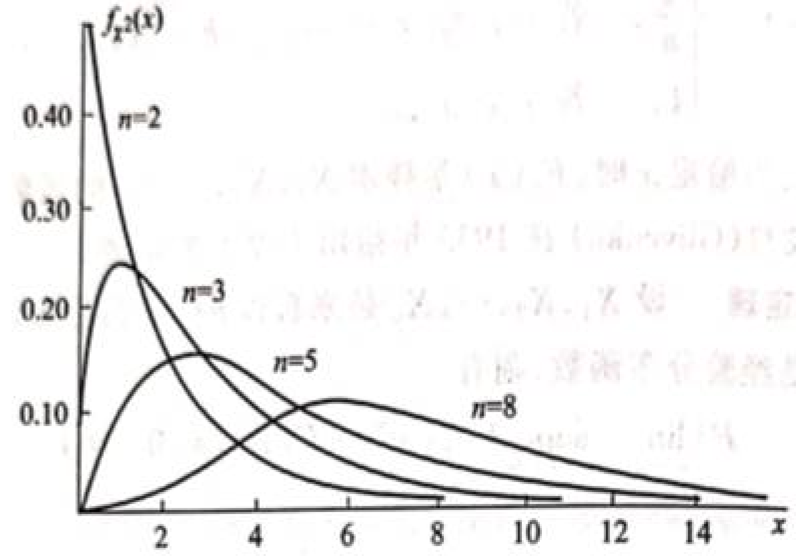
一、χ^2 分布

1.定义：若R.V. $X_1, X_2, \cdots, X_n$ 独立同服从N(0，1)，则

称 $\chi^2=\sum_{i=1}^{n} X_i^2$ 服从自由度为n的 χ^2分布，记为 $\chi^2 \sim \chi^2(n)$。

2. 概率密度

$$f(x;n)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x\geq 0\\ 0 & x<0\end{cases}$$



例1： 设 $X_1, X_2, \cdots, X_n$ 是总体$X \sim N(\mu, \sigma^2)$的一个样本，则 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim$ ________________。

分析： $X_1, X_2, \cdots, X_n$ 独立同服从正态分布 $N(\mu, \sigma^2)$

$\therefore \frac{X_1-\mu}{\sigma}, \frac{X_2-\mu}{\sigma}, \cdots, \frac{X_n-\mu}{\sigma}$ 独立同服从正态分布 $N(0,1)$

根据 χ^2 分布的定义

$$\therefore \boxed{\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)}$$

例2： 设X_1，X_2，X_3，X_4 是总体 $X \sim N(0,4)$ 的一个样本，

而$X = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$

当$a =$__________，$b =$__________时，$X \sim \chi^2(2)$

分析： X_1，X_2，X_3，X_4 独立同服从正态分布 $N(0\ ,4)$

$$\therefore E(X_i) = 0, D(X_i) = 4$$

根据 χ^2 分布的定义

$$\begin{cases} \sqrt{a}(X_1 - 2X_2) \sim N(0,1) \\ \sqrt{b}(3X_3 - 4X_4) \sim N(0,1) \end{cases} \Rightarrow \begin{cases} D[\sqrt{a}(X_1 - 2X_2)] = 1 \\ D[\sqrt{b}(3X_3 - 4X_4)] = 1 \end{cases}$$

$$\Rightarrow \begin{cases} a[D(X_1) + 4D(X_2)] = 20a = 1 \\ b[9D(X_3) + 16D(X_4)] = 100b = 1 \end{cases} \boxed{\Rightarrow \begin{cases} a = 1/20 \\ b = 1/100 \end{cases}}$$

3.性质

(1) 可加性

若 $\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，且$\chi_1^2$和$\chi_2^2$相互独立，则$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

(2) 期望与方差

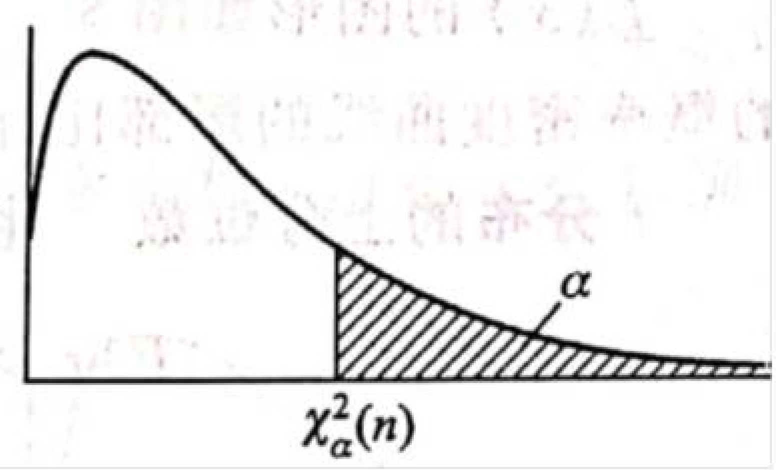
设$X^2 \sim \chi^2(n)$，则 $E(\chi^2) = n, D(\chi^2) = 2n.$

4. χ^2 分布的上α分位数

定义：对 $\forall\alpha\ (0<\alpha<1)$，若

$$P\{\chi^2>\chi_\alpha^{\ 2}(n)\}=\int_{\chi_\alpha^2(n)}^{+\infty}f(x)dx=\alpha$$

则称 $\chi_\alpha^{\ 2}(n)$为该 $\chi^2(n)$ 的上**α** 分位数(点)。

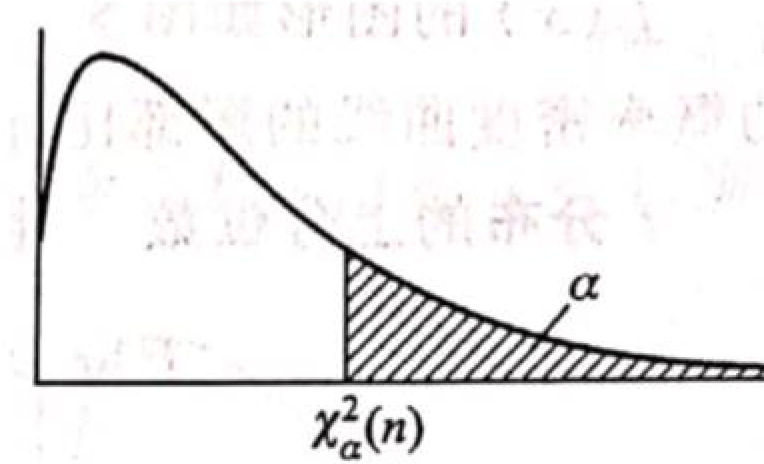


② 上 α 分位点的求法

$a.\ n < 40$时，查附表5

$b.\ n > 40$时

$$\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$$



其中是Z_α标准正态分布的上 α 分位点

例4：求$\chi_{0.1}^2(25)$，$\chi_{0.05}^2(30)$，$\chi_{0.05}^2(50)$

解：$\chi_{0.1}^2(25)=34.381$，$\chi_{0.05}^2(30)=43.773$

$$\chi_{0.05}^2(50) \approx \frac{1}{2}(z_{0.05} + \sqrt{99})^2 = \frac{1}{2}(1.645 + \sqrt{99})^2 = 67.221$$

例3： 设 $V_1, V_2, \cdots, V_6$ 是总体 $N(2,3)$ 的一个样本，当

$b =$ ________，$P\{\sum_{i=1}^{6}(V_i-2)^2 \le b\} = 0.95$

分析： $V_1, V_2, \cdots, V_6$ 独立同服从正态分布 $N(2,3)$

$\therefore \frac{V_1-2}{\sqrt{3}}, \frac{V_2-2}{\sqrt{3}}, \cdots, \frac{V_6-2}{\sqrt{3}}$ 独立同服从正态分布 $N(0,1)$

根据 χ^2 分布的定义 $\sum_{i=1}^{6}\left(\frac{V_i-2}{\sqrt{3}}\right)^2 \sim \chi^2(6)$

要使 $P\left\{\sum_{i=1}^{6}(V_i-2)^2 \le b\right\} = 0.95 \Leftrightarrow P\left\{\sum_{i=1}^{6}\left(\frac{V_i-2}{\sqrt{3}}\right)^2 \le \frac{b}{3}\right\} = 0.95$

即需 $P\left\{\sum_{i=1}^{6}\left(\frac{V_i-2}{\sqrt{3}}\right)^2 > \frac{b}{3}\right\} = 0.05$ 查表得：$\boxed{\frac{b}{3} = 12.592 \Rightarrow b = 37.776}$

$\chi^2_{0.05}(6)$

二、t分布

1.定义：设$X \sim N(0, 1)$，$Y \sim \chi^2(n)$，且X与Y相互独立，

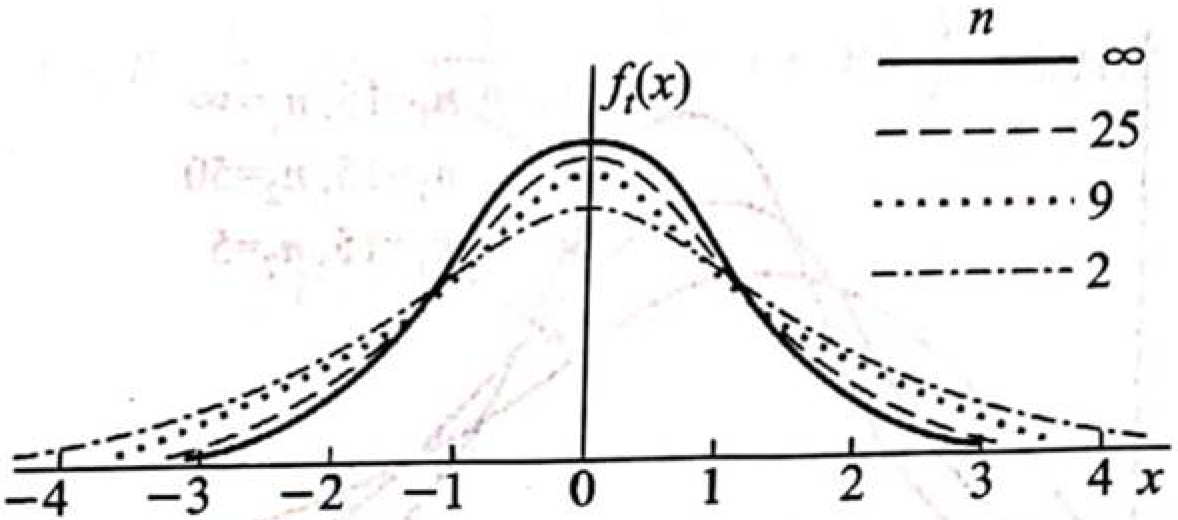
则 $t = \dfrac{X}{\sqrt{Y/n}}$ 服从的自由度为n的 t 分布，记作 $t \sim t(n)$.

2.概率密度

$$h(t) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < t < \infty$$

3. *t*分布的性质

$$\lim_{n\to\infty} t(n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \Phi(x)$$



例5： 设$R.V.X$和Y相互独立且同服从$N(0,9)$，X_1，$X_2,\cdots$，X_9 和 Y_1，$Y_2,\cdots$，Y_9 是分别来自总体X和Y 的一个简单随机样本，则$U=\dfrac{X_1+X_2+\cdots X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_9^2}}\sim$ ________________。

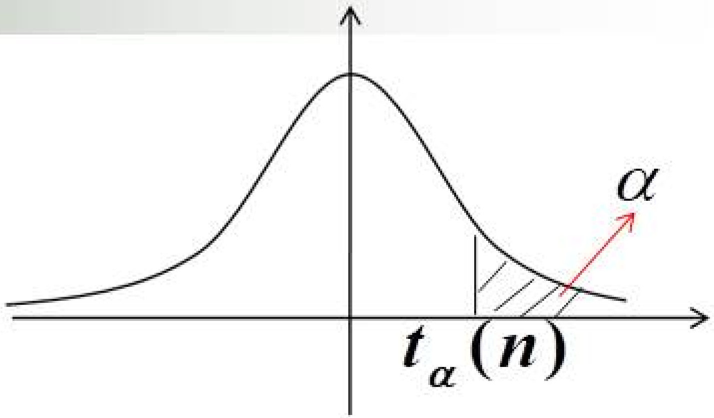
分析： $\bar{X}=\dfrac{1}{9}\sum_{i=1}^{9}X_i\sim N(0,1),$

$$\frac{Y_i}{3}\sim N(0,1)\ \Rightarrow Y=\sum_{i=1}^{6}\left(\frac{Y_i}{3}\right)^2=\frac{1}{9}\sum_{i=1}^{6}Y_i^2\sim\chi^2(9)$$

$\because\ \bar{X}$与Y相互独立

$$\therefore U=\frac{\bar{X}}{\sqrt{Y/9}}=\frac{\frac{1}{9}\sum_{i=1}^{9}X_i}{\sqrt{\frac{1}{9}\sum_{i=1}^{9}Y_i^2/9}}=\frac{\sum_{i=1}^{9}X_i}{\sqrt{\sum_{i=1}^{9}Y_i^2}}\sim t(9)$$

4. t 分布的上 α 分位数



(1) 定义： 对 $\forall\alpha\ (0<\alpha<1)$，若

$$P\{t>t_\alpha(n)\}=\int_{t_\alpha(n)}^{+\infty}h(t)dt=\alpha$$

则称 $t_\alpha(n)$ 为 $t(n)$ 分布的上α分位数(点)。

(2) 上 α 分位数的求法

a. 查附表4

b. 当α较大 $t_{1-\alpha}(n)=-t_\alpha(n)$

c. n >45时，$t_\alpha(n)\approx z_\alpha$，其中是$z_\alpha$标准正态分布的上$\alpha$分位数

例如： $t_{0.025}(20)=2.086$ $\quad t_{0.95}(10)=-t_{0.05}(10)=-1.8125$

$$t_{0.025}(60)\approx z_{0.025}=1.96$$

三、F分布

1.定义： 设$X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$,且X与Y相互独立,则

$$F = \frac{X / n_1}{Y / n_2}$$

服从自由度(n_1, n_2)的 F-分布，记为：$F \sim F(n_1, n_2)$

2.概率密度

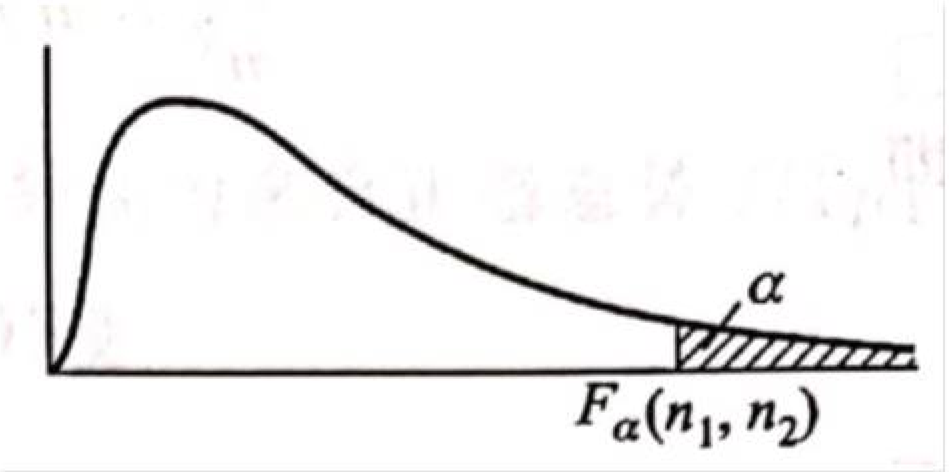
$$\varphi(y) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} (y)^{\frac{n_1}{2}-1} \left(1+\frac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}} & y > 0 \\ 0 & y \le 0 \end{cases}$$

3. F 分布的上 α 分位数

(1) 定义：对 $\forall\alpha$ $(0<\alpha<1)$，若

$$P\{F>F_{\alpha}(n_1,n_2)\}=\int_{F_{\alpha}(n_1,n_2)}^{+\infty}f(x)dx=\alpha$$

则称 $F_{\alpha}(n_1,n_2)$为$F(n_1,n_2)$分布的上α分位数(点)。



(2) 性质

a. 若$F \sim F(n_1, n_2)$，则$1/F \sim F(n_2, n_1)$

b. $F_{1-\alpha}(n_1, n_2) = 1/F_{\alpha}(n_2, n_1)$

① (5.4.11) 式的证明如下：若 $F \sim F(n_1, n_2)$，按定义

$$1-\alpha = P\{F > F_{1-\alpha}(n_1, n_2)\} = P\left\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\}$$

$$= 1 - P\left\{\frac{1}{F} \geqslant \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} = 1 - P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\},$$

于是

$$P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} = \alpha.$$

再由$\frac{1}{F} \sim F(n_2, n_1)$知

$$P\left\{\frac{1}{F} > F_{\alpha}(n_2, n_1)\right\} = \alpha.$$

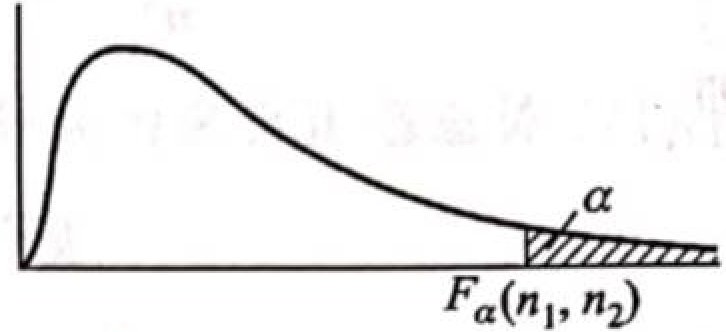
比较(1)、(2)两式得

$$\frac{1}{F_{1-\alpha}(n_1, n_2)} = F_{\alpha}(n_2, n_1),$$

即

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}.$$

(3) 上α 分位数的求法



$a.$ 查附表6

$b.$ 当 α较大时，$F_{1-\alpha}(n_1,n_2)=1/F_\alpha(n_2,n_1)$

例6： 设$F\sim F(24,15)$，求满足下列各式的 λ

$(1)P\{F>\lambda\}=0.025$ $(2)P\{F<\lambda\}=0.025$

$(3)P\{F<\lambda\}=0.95$

解： $(1)\lambda=F_{0.025}(24,15)=2.7$

$$(2)\lambda=F_{0.975}(24,15)=\frac{1}{F_{0.025}(15,24)}=\frac{1}{2.44}$$

$(3)\lambda=F_{0.05}(24,15)=2.29$

四、正态总体的 $\bar{X}$ 和 S^2 的分布

1. $\bar{X}$ 与 S^2 的期望与方差

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本，若

$E(X)=\mu, D(X)=\sigma^2$，则：

$\boxed{E(X)}$ $\boxed{D(X)/n}$

(1) $E(\bar{X})=\mu, D(\bar{X})=\dfrac{\sigma^2}{n}$

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

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

(2) $E(S^2)=\sigma^2 \rightarrow \boxed{D(X)}$

$$
\begin{aligned}
E(S^2) &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2)\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right] = \sigma^2.
\end{aligned}
$$

注意： 无论总体服从什么分布，样本均值的期望等于总体的期望，样本均值的方差等于总体方差的$1/n$，样本方差的期望等于总体方差.

例7： 设$(X_1, X_2, \cdots, X_n)$是来自泊松分布$\pi(\lambda)$的一个样本，$\overline{X}$、S^2分别为样本均值、样本方差，求 $E(\overline{X})$、$D(\overline{X})$、$E(S^2)$.

解： $E(\bar{X})=E(X)=\lambda$

$$D(\bar{X})=D(X)/n=\lambda/n$$

$$E(S^2)=D(X)=\lambda$$

2.正态总体中的抽样分布

(1) $\bar{X}$ 的 分 布

Th1 设$X \sim N(\mu,\sigma^2)$，$X_1, X_2,\cdots, X_n$ 是来自总体X

的样本，则：$\bar{X} \sim N(\mu,\sigma^2/n)$ $\boxed{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}$

(2) S^2的 分 布

Th2 设$X \sim N(\mu,\sigma^2)$，$X_1, X_2,\cdots, X_n$是来自总体 X的

样本，则：

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

Th3 设 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，则：

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

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

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

$$\Rightarrow \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}} \sim t(n-1).$$

Th4 设$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，$(X_1, X_2, \cdots, X_m)$和$(Y_1, Y_2, \cdots, Y_n)$分别为来自总体X和Y的样本，且两样本相互独立，设$\overline{X}$、$\overline{Y}$分别为两样本的样本均值，$S_1^2$、$S_2^2$分别为两样本的样本方差，则有

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

分析：$\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$,

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1).$$

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

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

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

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

例8： 设 $X \sim N(80,100)$，在总体中随机抽取容量n=100的样本，求：

(1) 样本均值与总体均值差的绝对值小于1的概率是多少？

(2)若要求样本均值与总体均值的偏差小于2的概率不小于0.95，问样本容量n至少等于多少？

解： (1) 按题意需要求概率 $P\{|\bar{X}-\mu|<1\}$. 现在样本容量 $n=100$, $\sigma^2=100, \sigma/\sqrt{n}=1$, 即有

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

$$\begin{aligned}
P\{|\bar{X}-\mu|<1\} &= P\{-1<\bar{X}-\mu<1\} \\
&= P\left\{\frac{-1}{\sigma/\sqrt{n}}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{1}{\sigma/\sqrt{n}}\right\} \\
&= \Phi\left(\frac{1}{\sigma/\sqrt{n}}\right)-\Phi\left(\frac{-1}{\sigma/\sqrt{n}}\right) \\
&= 2\Phi(1)-1 = 0.6826.
\end{aligned}$$

(2) 按题意要求

$$P\{|\bar{X}-\mu|<2\}=P\{-2<\bar{X}-\mu<2\}\geqslant 0.95.$$

上式即为

$$P\left\{\frac{-2}{\sigma/\sqrt{n}}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{2}{\sigma/\sqrt{n}}\right\}=\Phi\left(\frac{2}{\sigma/\sqrt{n}}\right)-\Phi\left(-\frac{2}{\sigma/\sqrt{n}}\right)$$

$$=2\Phi\left(\frac{2}{\sigma/\sqrt{n}}\right)-1\geqslant 0.95.$$

亦即

$$\Phi\left(\frac{2}{\sigma/\sqrt{n}}\right)\geqslant 0.975=\Phi(1.96).$$

因此要求

$$\frac{2}{\sigma/\sqrt{n}}\geqslant 1.96,\quad 即\quad \frac{\sqrt{n}}{5}\geqslant 1.96,\ n\geqslant 96.04.$$

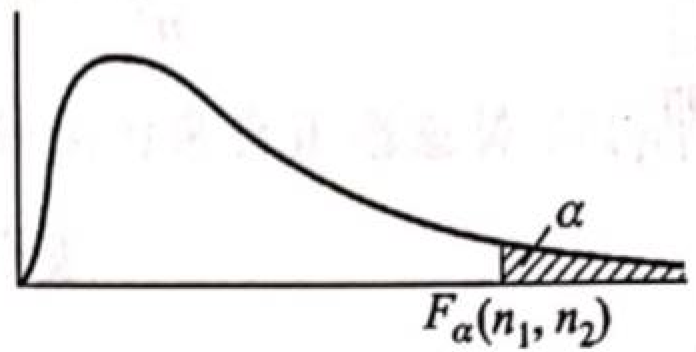
取 $n=97$,即可使 $P\{|\bar{X}-\mu|<2\}\geqslant 0.95$.

例9 正态总体X，Y的方差分别为：$\sigma_1^2=12$，$\sigma_2^2=18$，在总体X，Y中分别取出样本容量为 $n_1=61$，$n_2=31$ 的样本，两样本相互独立，样本方差分别为 S_1^2、S_2^2，求：

$$P\left\{S_1^2/S_2^2>1.16\right\}$$

分析：

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



$$P\{S_1^2/S_2^2>1.16\}=P\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}>\frac{1.16}{12/18}\right\}=P\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}>1.74\right\}$$

$$F_\alpha(n_1-1,n_2-1)$$

例9 正态总体X，Y的方差分别为：$\sigma_1^2=12$，$\sigma_2^2=18$，在总体X，Y中分别取出样本容量为 $n_1=61$，$n_2=31$ 的样本，两样本相互独立，样本方差分别为 S_1^2、S_2^2，求：

$$P\left\{S_1^2/S_2^2>1.16\right\}$$

解： 由定理 4 知

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}=\frac{S_1^2/S_2^2}{12/18}\sim F(60,30).$$

而

$$P\{S_1^2/S_2^2>1.16\}=P\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}>\frac{1.16}{12/18}\right\}=P\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}>1.74\right\}.$$

查表知，$F_{0.05}(60,30)=1.74$，故有

$$P\{S_1^2/S_2^2>1.16\}=0.05.$$

本章小结

一、简单随机样本的定义及性质

二、常用的统计量：① ~⑤

三、抽样分布

① χ^2 分布

② t 分布

③ F 分布

④正态总体的 $\bar{X}$ 和 S^2 的分布（*Th*1~*Th*4）