§ 6.2　抽样分布

确定统计量的分布—— 抽样分布, 是数理统计的基本问题之一. 采用求随机向量的函数的分布的方法可得到抽样分布.

由于正态总体是最常见的总体，故本节介绍的几个抽样分布均基于正态总体.

统计中常用分布

(1) 正态分布

若 $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma_i^2)$

则 $$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

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

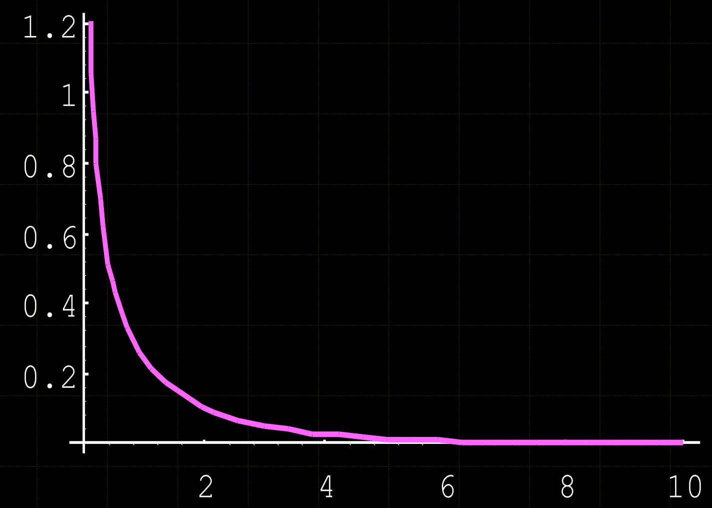
(2) $\chi^2(n)$分布 （ n为自由度 ）

定义 设 $X_1, X_2, \cdots, X_n$ 相互独立,

且都服从标准正态分布$N(0, 1)$, 则 $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

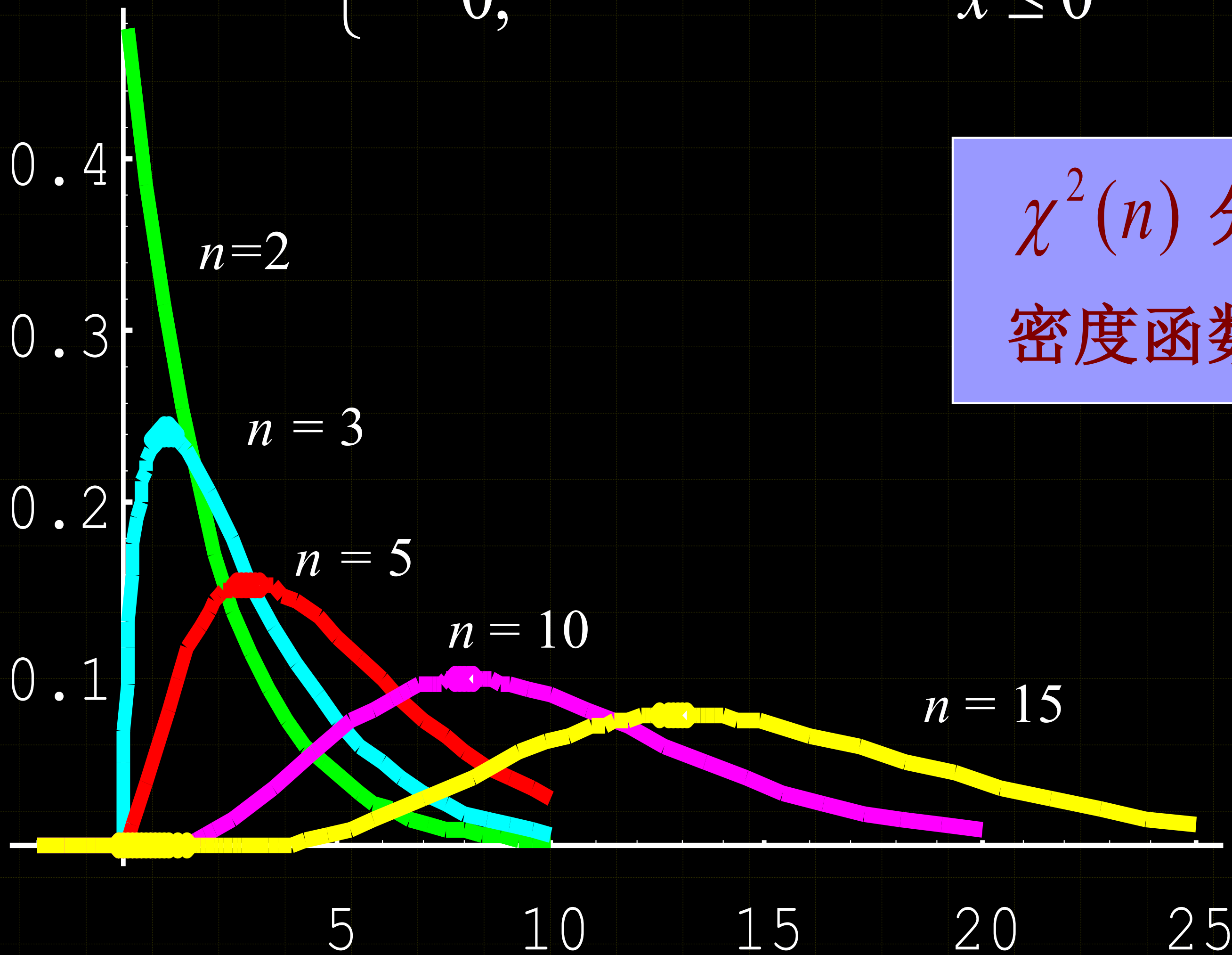
$n = 1$ 时, 其密度函数为

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$



一般地，自由度为 n 的 $\chi^2(n)$ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{x}{2}}x^{\frac{n}{2}-1}, & x>0\\[2ex] 0, & x\le 0\end{cases}$$



$\chi^2(n)$ 分布

密度函数图

$\chi^2(n)$ 分布的性质

1° $E\left(\chi^2(n)\right) = n,\ D\left(\chi^2(n)\right) = 2n$

2° 若$X_1 = \chi^2(n_1), X_2 = \chi^2(n_2), X_1, X_2$相互独立，

则 $X_1 + X_2 \sim \chi^2(n_1 + n_2)$

3° $n \to \infty$时，$\chi^2(n) \to$ 正态分布

证 1° 设 $\chi^2(n)=\sum_{i=1}^{n}X_i^2 \quad X_i\sim N(0,1)\ \ i=1,2,\cdots,n$

$X_1,X_2,\cdots,X_n$ 相互独立,

则

$$E(X_i)=0,\ \ D(X_i)=1,\ \ E(X_i^2)=1$$

$$E\left(\chi^2(n)\right)=E\left(\sum_{i=1}^{n}X_i^2\right)=n$$

$$E(X_i^4)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^4e^{-\frac{x^2}{2}}\mathrm{d}x=3$$

$$D(X_i^2)=E(X_i^4)-E^2(X_i^2)=2$$

$$D\left(\chi^2(n)\right)=D\left(\sum_{i=1}^{n}X_i^2\right)=2n$$

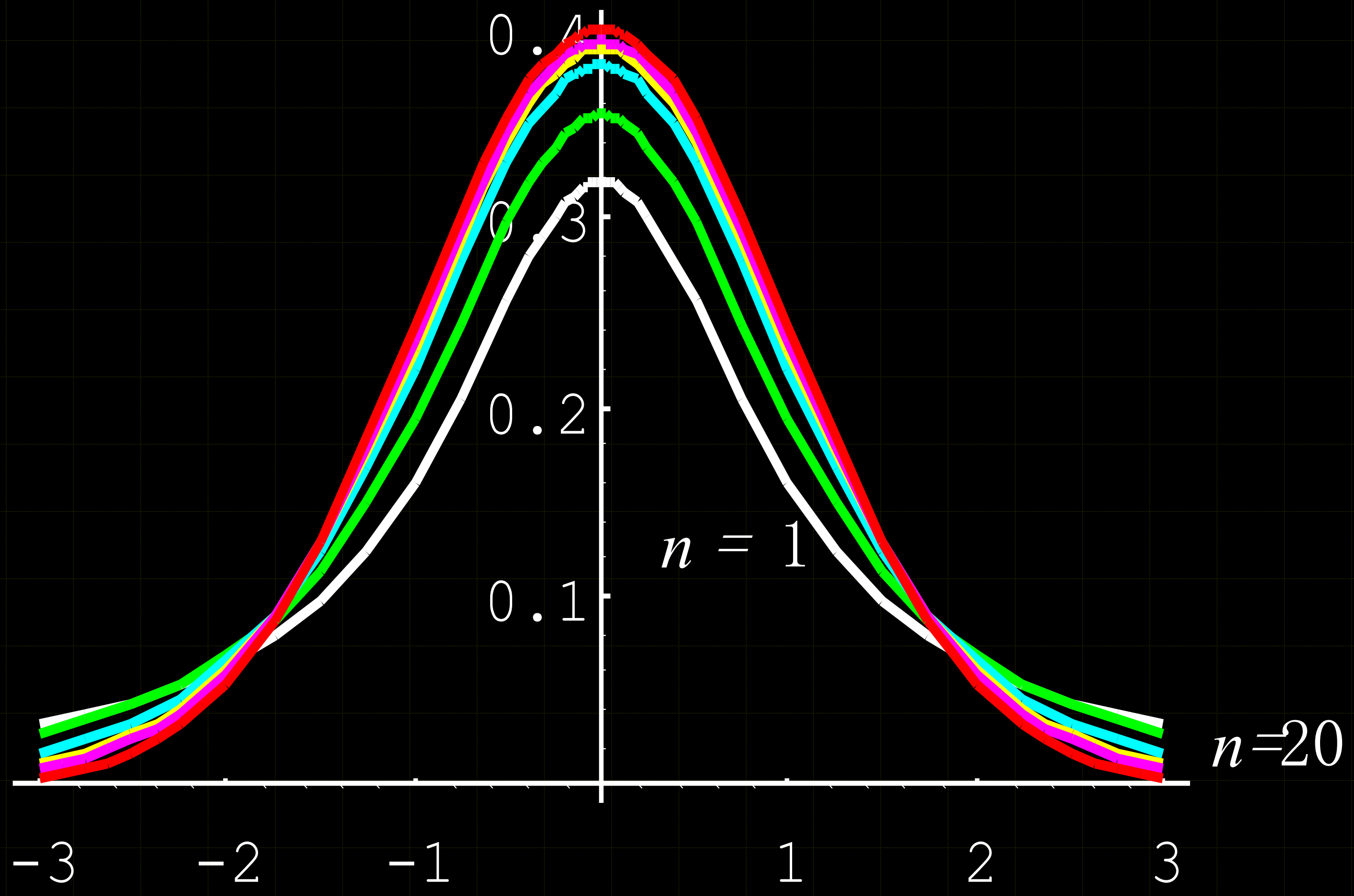
(3) t 分布（Student 分布）

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

$$T = \frac{X}{\sqrt{Y/n}}$$

则T所服从的分布称为自由度为 n 的T分布其密度函数为

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < t < \infty$$



t 分布的图形(红色的是标准正态分布)

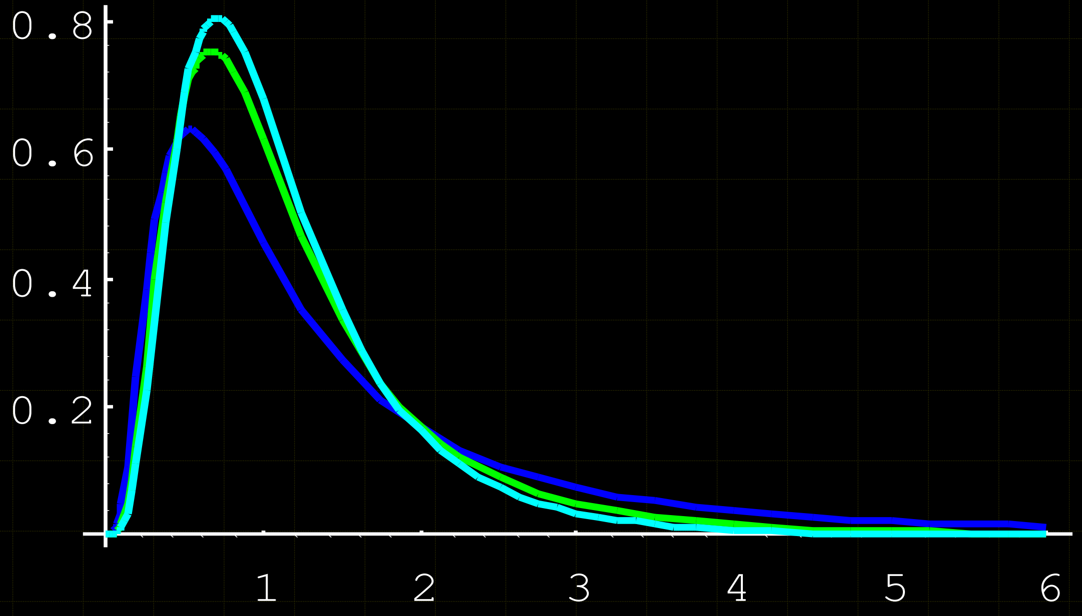
(4) F 分布

定义 设 $X \sim \chi^2(n), Y \sim \chi^2(m), X, Y$ 相互独立,

令 $$F = \frac{X/n}{Y/m}$$

则F 所服从的分布称为**第一自由度为n, 第二自由度为 m 的F 分布,** 其密度函数为

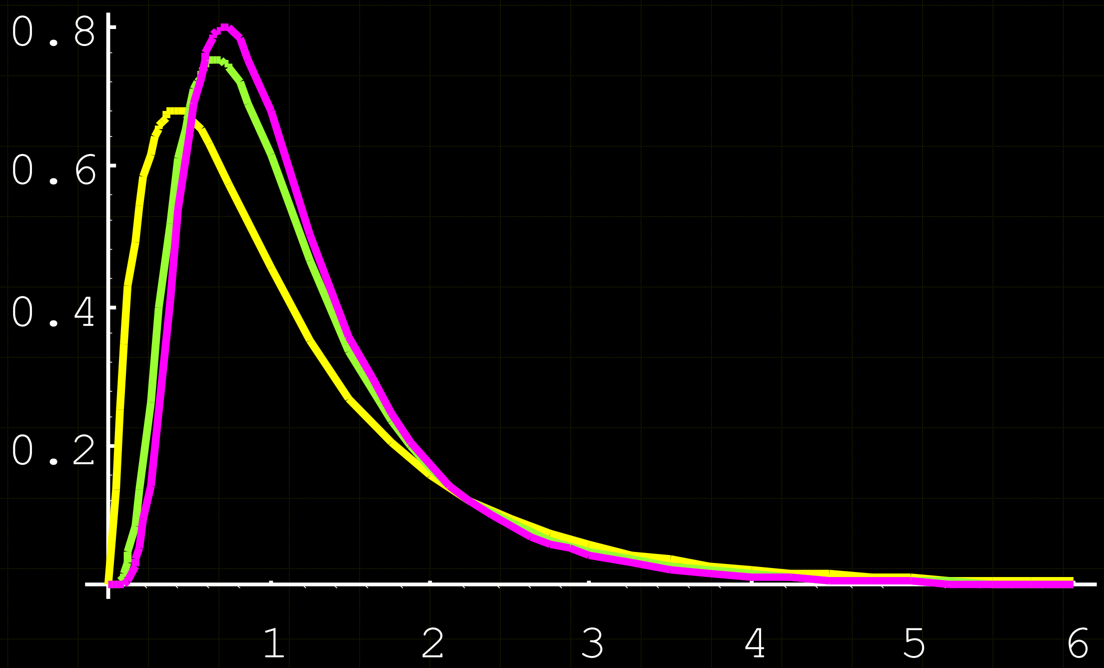
$$f(t,n,m) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{\frac{n}{2}} t^{\frac{n}{2}-1}\left(1+\dfrac{n}{m}t\right)^{-\frac{n+m}{2}}, & t > 0 \\ 0, & t \le 0 \end{cases}$$



$m = 10, n = 4$

$m = 10, n = 10$

$m = 10, n = 15$



$m = 4, n = 10$

$m = 10, n = 10$

$m = 15, n = 10$

F 分布的性质 若$F \sim F(n,m)$，则$\dfrac{1}{F} \sim F(m,n)$

- 分位数

抽样本分布的某些结论

（Ⅰ）单正态总体抽样分布定理

设 $X \sim N(\mu,\sigma^2)$ $E(X)=\mu$, $D(X)=\sigma^2$

总体的样本为（ $X_1, X_2, \cdots, X_n$ ），则

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

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

$\frac{(n-1)S^2}{\sigma^2}$ 与 $\overline{X}$ 相互独立 (1)

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \div \frac{S}{\sigma} = \boxed{\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)} \quad (2)$$

（II）双正态总体

设 $X_1, X_2, \cdots, X_n$ 来自正态总体 $X \sim N(\mu_1, \sigma_1^2)$
$Y_1, Y_2, \cdots, Y_m$ 来自正态总体 $Y \sim N(\mu_2, \sigma_2^2)$
相互独立

令 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad \overline{Y} = \dfrac{1}{m}\sum_{j=1}^{m} Y_j$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad S_2^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2$$

则 $\dfrac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1) \qquad \dfrac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$

$$\boxed{\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)} \quad \text{(3)}$$

例1 设总体 $X \sim N(72,100)$ ，为使样本均值大于70 的概率不小于 90% ，则样本容量

$n =$ <u>42</u> .

解 设样本容量为 n， 则 $\overline{X} \sim N(72,\frac{100}{n})$

故 $P(\overline{X} > 70) = 1 - P(\overline{X} \le 70) = \Phi\left(0.2\sqrt{n}\right)$

令 $\Phi\left(0.2\sqrt{n}\right) \ge 0.9$ 查表得 $0.2\sqrt{n} \ge 1.29$

即 $n \ge 41.6025$ 所以取 $n = 42$

例2 $X_1, X_2, \cdots, X_{20}$ 来自总体 $X \sim N(\mu, \sigma^2)$

(1) 求 $P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.76\sigma^2\right)$

(2) 求 $P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \mu\right)^2 \le 1.76\sigma^2\right)$

解 (1) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 即 $\frac{19S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \sim \chi^2(19)$

故 $P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.76\sigma^2\right)$

$$= P\left(7.4 \le \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 35.2\right)$$

$$= P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 7.4\right) - P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 35.2\right)$$

查表

$$= 0.99 - 0.01 = 0.98$$

(2) $$\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(20)$$

故 $$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}(X_i-\mu)^2 \le 1.76\sigma^2\right)$$

$$= P\left(7.4 \le \sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \le 35.2\right)$$

$$= P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 7.4\right) - P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 35.2\right)$$

$$= 0.995 - 0.025 = 0.97$$

例3 X 与 Y 相互独立， $X\sim N(0,16)$, $Y\sim N(0,9)$,
$X_1, X_2,\dots, X_9$ 与 $Y_1, Y_2,\dots, Y_{16}$ 分别来自 X 与 Y, 求统 计量

$$\frac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_{16}^2}}$$ 所服从的分布 。

解 $X_1+X_2+\cdots+X_9\sim N(0,9\times 16)$

$$\frac{1}{3\times 4}(X_1+X_2+\cdots+X_9)\sim N(0,1)$$

$$\frac{1}{3}Y_i\sim N(0,1)\quad ,i=1,2,\cdots,16 \qquad \sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2\sim\chi^2(16)$$

从而 $$\frac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_{16}^2}}=\frac{\frac{1}{3\times 4}(X_1+X_2+\cdots+X_9)}{\sqrt{\dfrac{\sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2}{16}}}\sim t(16)$$

例7 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体$N(\mu,\sigma^2)$的简单随机样本，$\overline{X}$ 是样本均值，

$$S_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,\qquad S_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

$$S_3^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2,\qquad S_4^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2,$$

则服从自由度为$n-1$的 t 分布的随机变量为:

(A) $\dfrac{\overline{X}-\mu}{S_1}\sqrt{n-1}$　　(B) $\dfrac{\overline{X}-\mu}{S_2}\sqrt{n-1}$

(C) $\dfrac{\overline{X}-\mu}{S_3}\sqrt{n}$　　(D) $\dfrac{\overline{X}-\mu}{S_4}\sqrt{n}$

解 $\dfrac{\dfrac{\overline{X}-\mu}{\sigma}}{\sqrt{n}} \sim N(0,1)$ $\quad \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2(n-1)$

$$\frac{\dfrac{\dfrac{\overline{X}-\mu}{\sigma}}{\sqrt{n}}}{\sqrt{\dfrac{\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}}} = \frac{\sqrt{n(n-1)}(\overline{X}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \sim t(n-1)$$

故应选(B)

作业　习题6

A组：10，11，12，13

B组：3