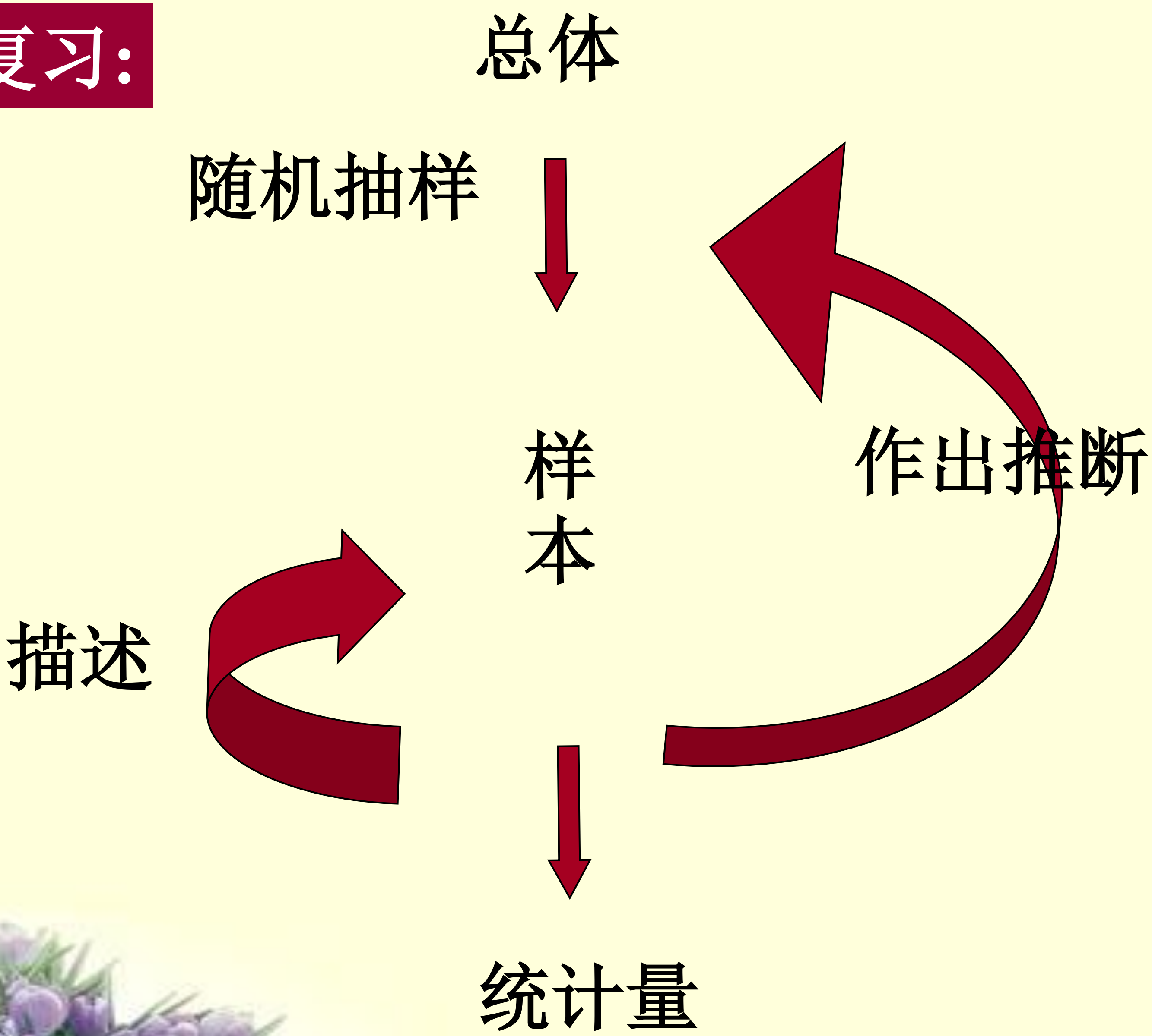
§ 6.3 抽样分布

复习:



抽样分布

统计量既然是依赖于样本的，而后者又是随机变量，故**统计量**也是随机变量，**它的分布**叫做统计量的“抽样分布”．

抽样分布
- 精确抽样分布
- 渐近分布

统计三大分布

χ^2分布

定义: 设 $X_1, X_2, \cdots, X_n$ **相互独立**, 都服从正态分布**$N(0,1)$**, 则称随机变量:

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

所服从的分布为**自由度**为 n 的 χ^2分布.

记为 $\chi^2 \sim \chi^2(n)$

χ^2分布的密度函数为

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

其中伽玛函数 $\Gamma(x)$通过积分

$$\Gamma(x)=\int_0^{\infty}e^{-t}t^{x-1}dt,\quad x>0$$

来定义.

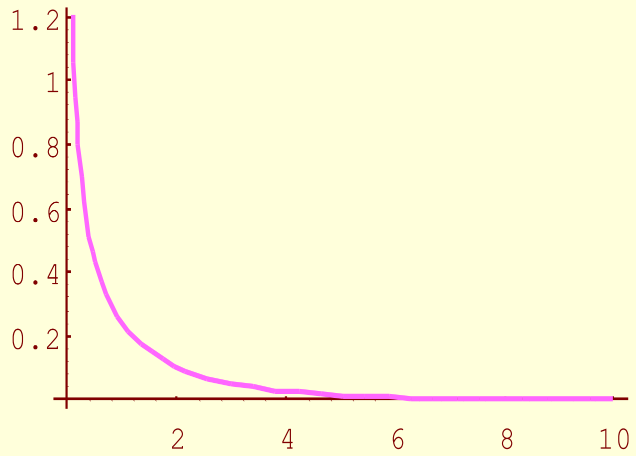
$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt, \quad x > 0$$

$$\Gamma(n+1) = n!$$

$$\Gamma(\frac{n}{2}) = 1 \cdot 3 \cdot 5 \cdots (n-2) \quad 2^{-(n-1)/2} \sqrt{\pi}$$

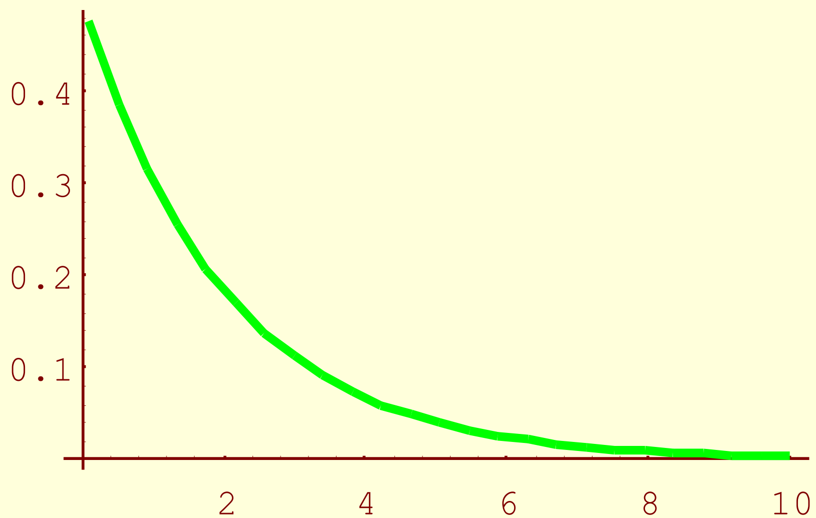
$n=1$ 时, 其密度函数为

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



$n=2$ 时, 其密度函数为

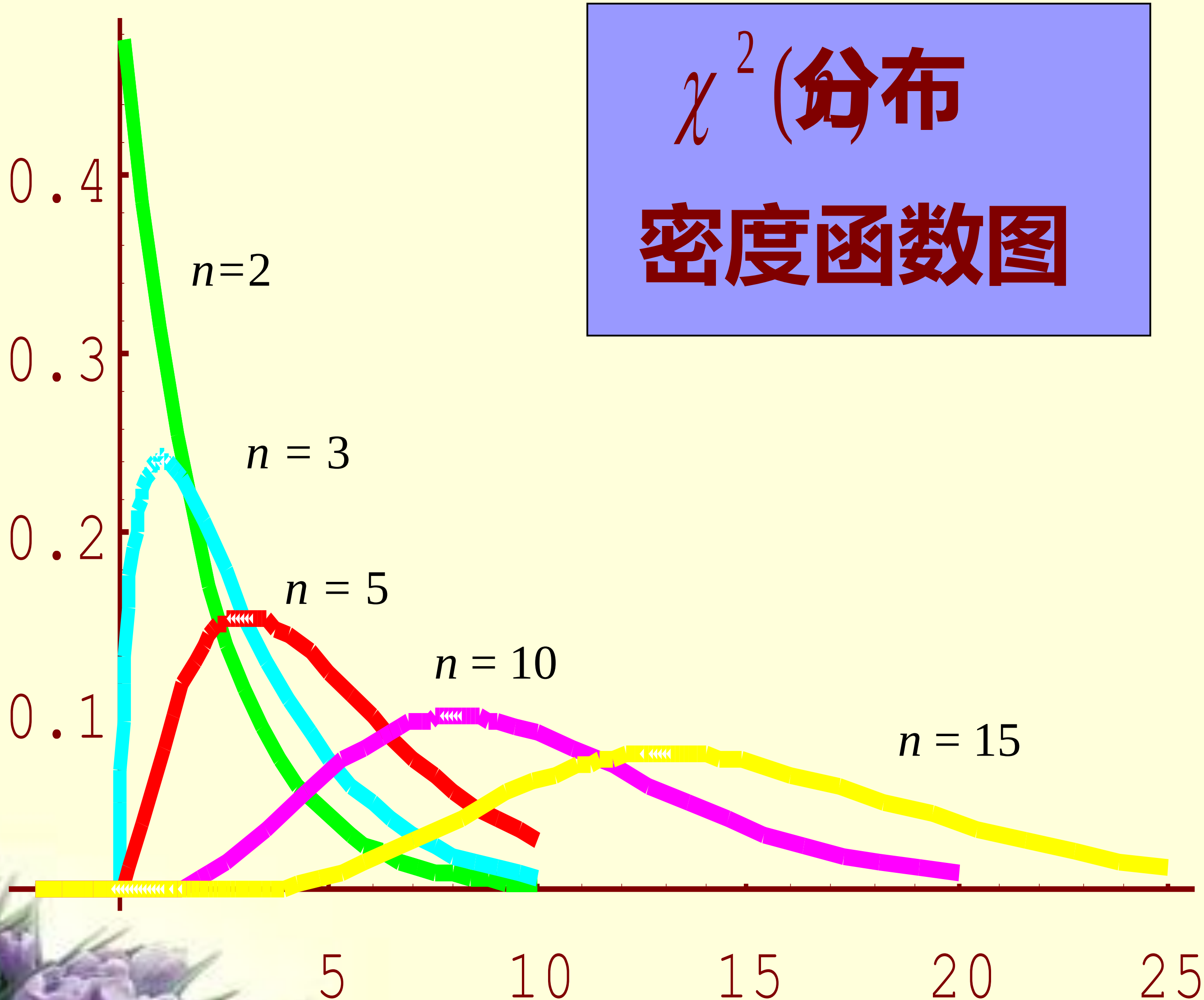
$$f(x)=\begin{cases}\dfrac{1}{2}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$



为参数为1/2的指数分布.

$\chi^2(n)$分布

密度函数图



Karl Pearson
1857-1932

χ^2分布的性质

1 设 $X_1, X_2, \cdots, X_n$ 相互独立，都服从正态分布 $N(\mu, \sigma^2)$，则

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

2 **(可加性)** 设 $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$，且 X_1, X_2 相互独立，则 $X_1 + X_2 \sim \chi^2(n_1 + n_2)$

3 若 $X \sim \chi^2(n)$, 则

$E(X)=n$, $D(X)=2n$

4 若 $X \sim \chi^2(n)$, 则

n充分大时 $\dfrac{X-n}{\sqrt{2n}}$ 近似 $N(0,1)$

性质3的证明：

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

$X_1, X_2, \cdots, X_n$ 相互独立，

则 $E(X_i)=0,\ D(X_i)=1,\ E(X_i^2)=1$

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

$$E(X_i^4)=3$$

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

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

例1：设总体 $X \sim N(0,4)$ **，** $X_1, X_2, \cdots X_{10}$ **为来自总体的样本。令**

$$Y = \left[\sum_{i=1}^{5} X_i\right]^2 + \left[\sum_{j=6}^{10} X_j\right]^2$$

试确定C使CY服从 χ^2 **分布，并指出其自由度。**

解：由已知条件有

$$\sum_{i=1}^{5} X_i \sim N(0,20), \quad \sum_{i=6}^{10} X_i \sim N(0,20),$$

所以 $Y_1 = \dfrac{1}{\sqrt{20}} \sum_{i=1}^{5} X_i \sim N(0,1),$

$$Y_2 = \frac{1}{\sqrt{20}} \sum_{i=6}^{10} X_i \sim N(0,1),$$

且Y_1, Y_1相互独立　故 $\dfrac{1}{20} Y = Y_1^2 + Y_2^2 \sim \chi^2(2)$

t 分布

定义：设X～$N(0,1)$, Y～ $\chi^2(n)$ ，且X与Y相互独立，则称变量

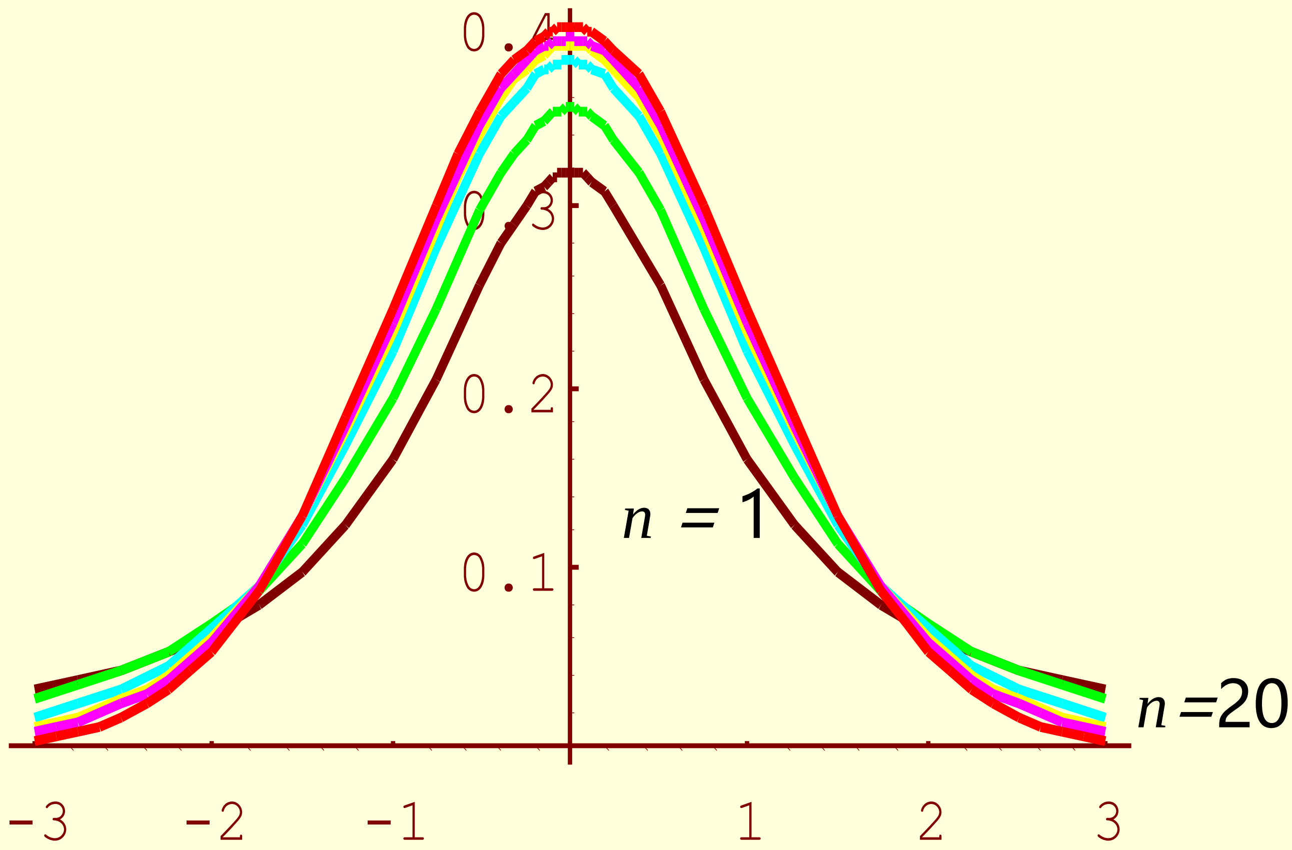
$$T = \frac{X}{\sqrt{Y/n}}$$

所服从的分布为自由度为 n的 t 分布.

记为T～$t(n)$.

T的密度函数为：

$$f(x;n) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$



t 分布的图形(红色的是标准正态分布)

William Gosset
1876-1937

性质

1 *t*分布的密度函数关于*x*=0对称，且

$$\lim_{|x| \to \infty} f(x;n) = 0$$

2 当*n*充分大时，其图形类似于标准正态分布密度函数的图形.

3 T~t(n)数学期望和方差为:

$E(T)=0$; $D(T)=n/(n-2)$, 对 $n>2$

例2 设X与Y相互独立， $X \sim N(0,16)$, $Y \sim N(0,9)$, $X_1, X_2, \ldots, X_9$ 与 $Y_1, Y_2, \ldots, Y_{16}$ 分别是取自 X 与 Y 的简单随机样本, 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 $X_1 + X_2 + \cdots + X_9 \sim N(0, 9 \times 16)$

$$U = \frac{1}{12}(X_1 + X_2 + \cdots + X_9) \sim N(0,1)$$

$$\frac{1}{3}Y_i \sim N(0,1)\quad, i=1,2,\cdots,16$$

$$V=\sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2 \sim \chi^2(16)$$ 并且U,V独立

从而

$$\frac{U}{\sqrt{\frac{V}{16}}}=\frac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_{16}^2}} \sim t(16)$$

3、F分布

定义： 设 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$,**$X$与$Y$相互独立**，则称统计量

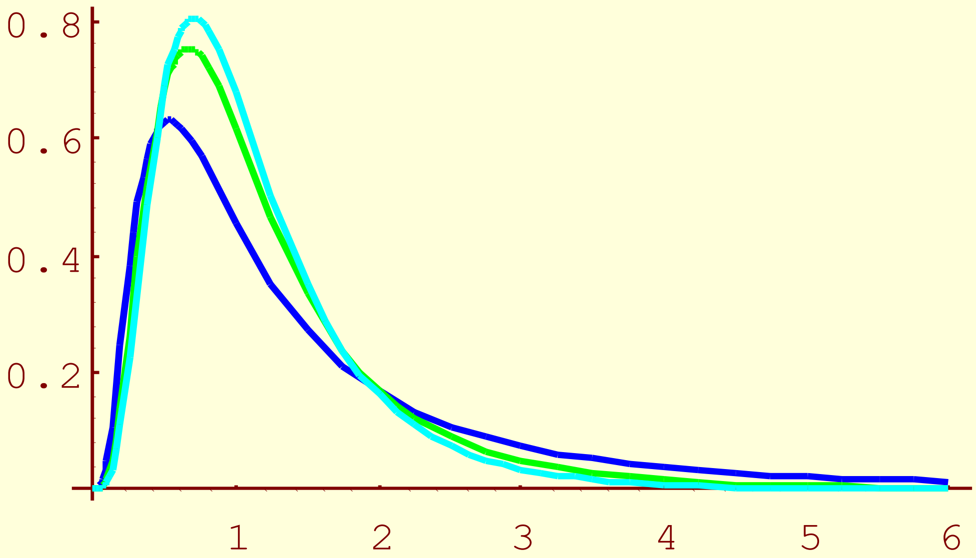
$$F = \frac{X/n_1}{Y/n_2}$$

服从**自由度为n_1及 n_2** 的F分布，记作

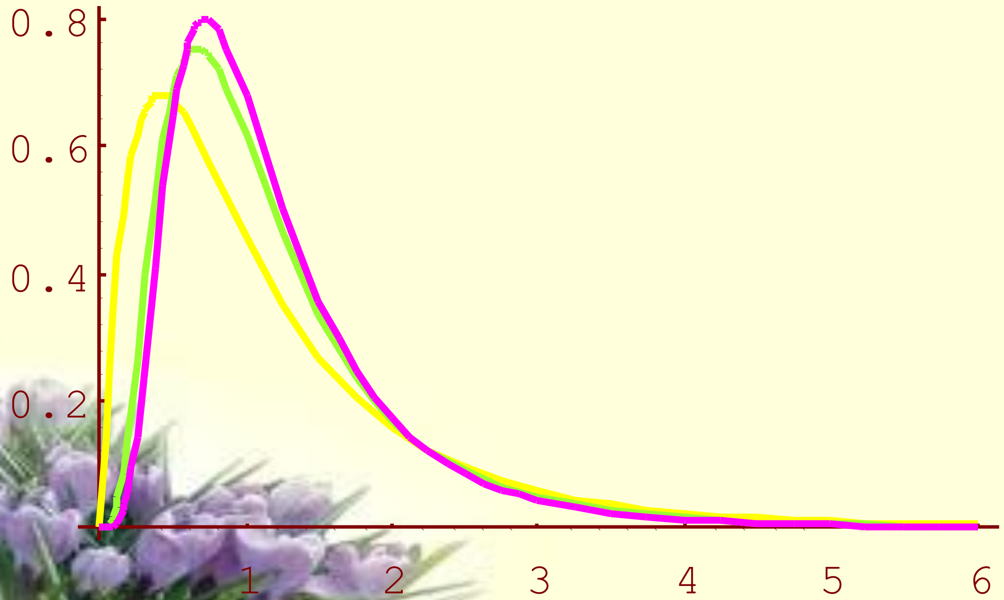
$F \sim F(n_1, n_2)$.

若$X\sim F(n_1,n_2)$， X的概率密度为

$$f(x;n_1,n_2)=\begin{cases}\dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & x\geq 0\\ 0 & x<0\end{cases}$$



$n_1 = 10, n_2 = 4$
$n_1 = 10, n_2 = 10$
$n_1 = 10, n_2 = 15$



$n_1 = 4, \quad n_2 = 10$
$n_1 = 10, \ n_2 = 10$
$n_1 = 15, \ n_2 = 10$

Ronald Fisher
1890-1962

例3

设X_1, X_2, X_3, X_4独立同分布，均服从$N(0,\sigma^2)$,求c

使$P((\frac{X_1+X_2}{X_3-X_4})^2 < c) = 0.9$

概率分布的上侧分位数

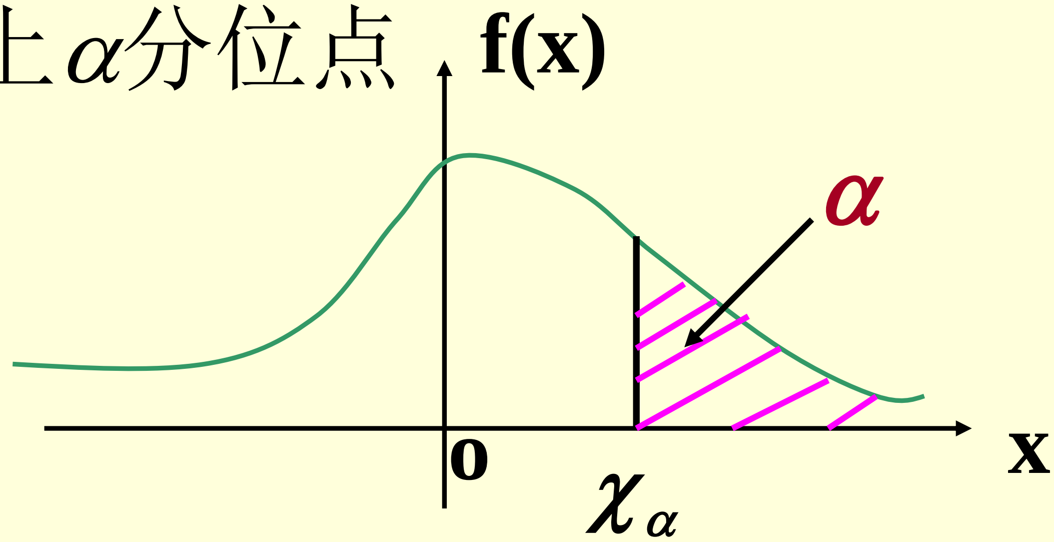
定义

设随机变量*x*的密度函数为*f*(*x*),对给定的

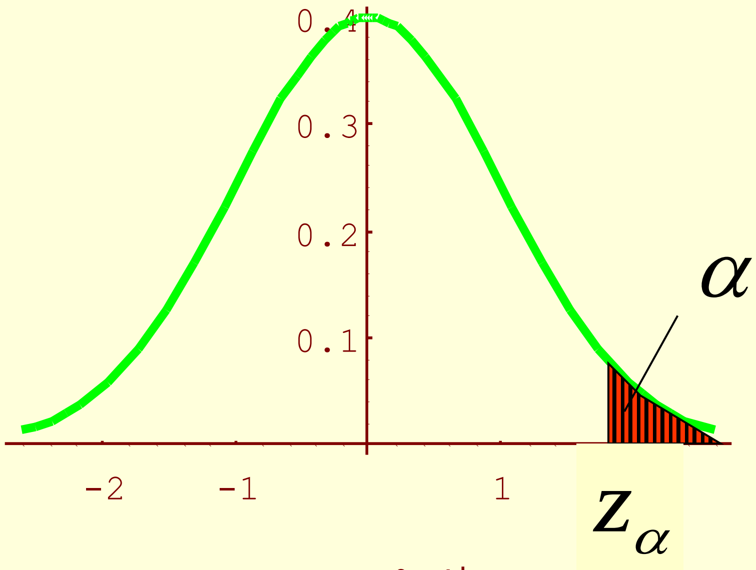
$\alpha(0<\alpha<1)$,称满足条件

$$P(X \geq x_\alpha)=\int_{x_\alpha}^{\infty} f(x)dx=\alpha$$

的实数x_α为X的上α分位点



1 标准正态分布的上α分位数 z_α

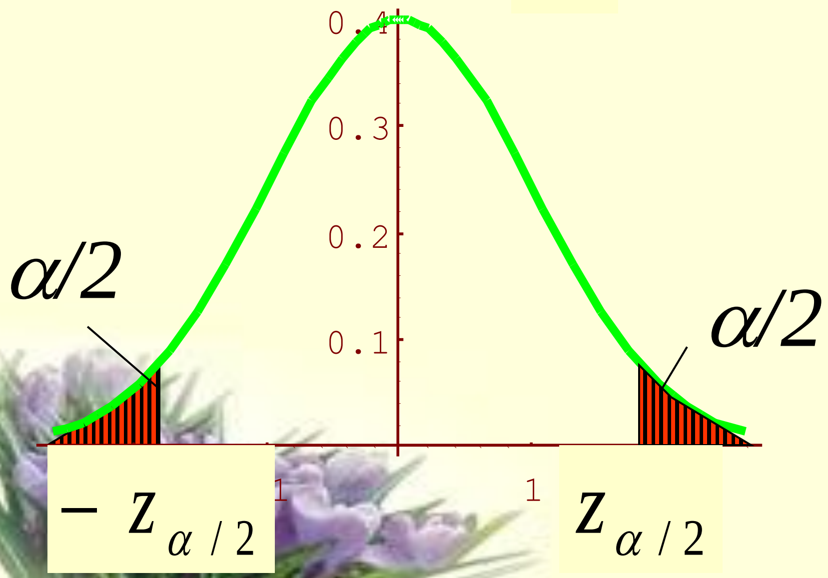


定义 $P(Z > z_\alpha) = \alpha$

例

$$\left.\begin{aligned} z_{0.025} &= 1.96 \\ z_{0.05} &= 1.645 \end{aligned}\right\}$$

常用数字



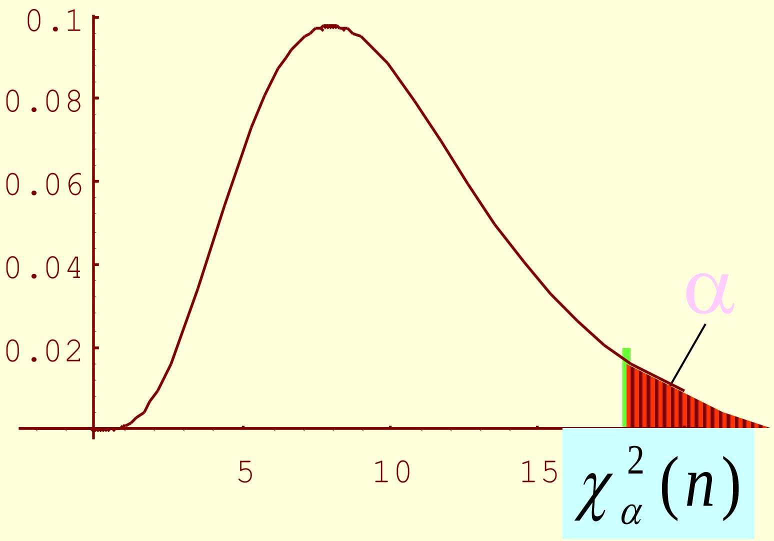
求c,使 $P(|Z| > c) = \alpha$

$c = z_{\alpha/2}$

2 χ^2分布的上α分位点$\chi^2_\alpha(n)$

定义 $P(\chi^2 \geq \chi^2_\alpha(n)) = \alpha$

例 $\chi^2_{0.05}(10) = 18.307$



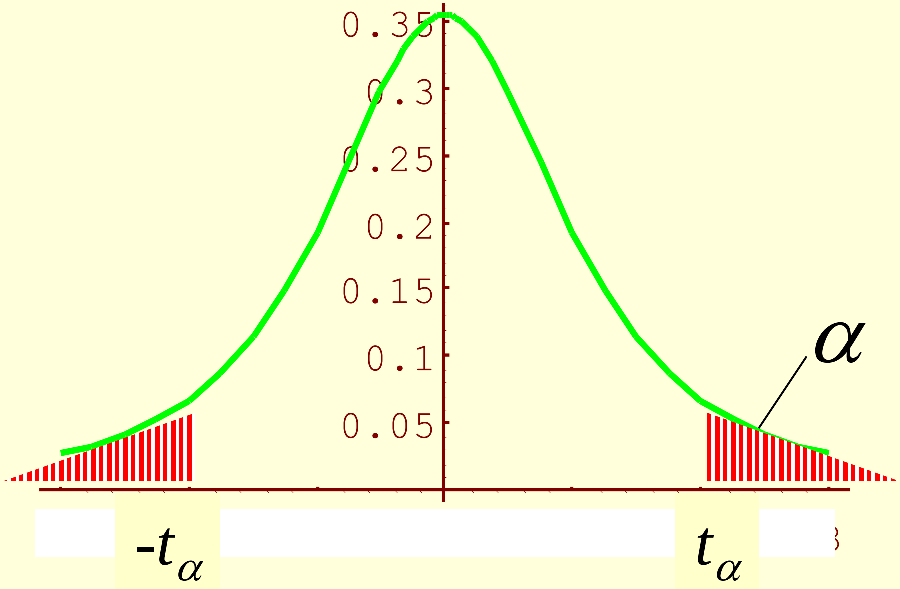
3　t分布的上α分位点$t_\alpha(n)$的性质

定义　$P\left(T > t_\alpha(n)\right) = \alpha$

例　$t_{0.05}(10) = 1.8125$

求c,使　$P(|T| > c) = \alpha$

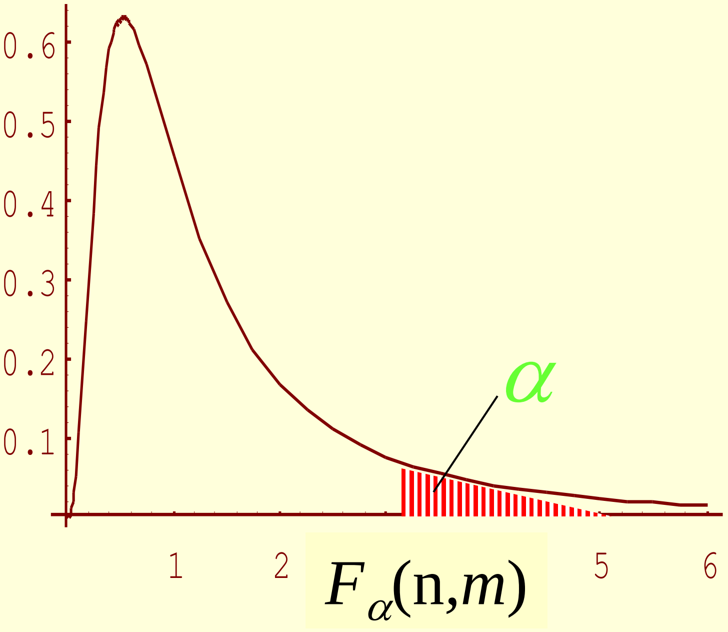
$$c = t_{\alpha/2}(n)$$



4 F分布的上α分位点$F_\alpha(m,n)$的性质

定义 $$P(F > F_\alpha(m,n)) = \alpha$$

例 $$F_{0.05}(4,5) = 5.19$$



几个重要的抽样分布定理

当总体为**正态分布**时，我们给出几个重要的抽样分布定理.

正态分布的重要性质

若 $X_1, X_2, \cdots, X_n$ 相互独立，且 $X_i \sim N(u_i, \sigma_i^2), i = 1 \cdots n,$

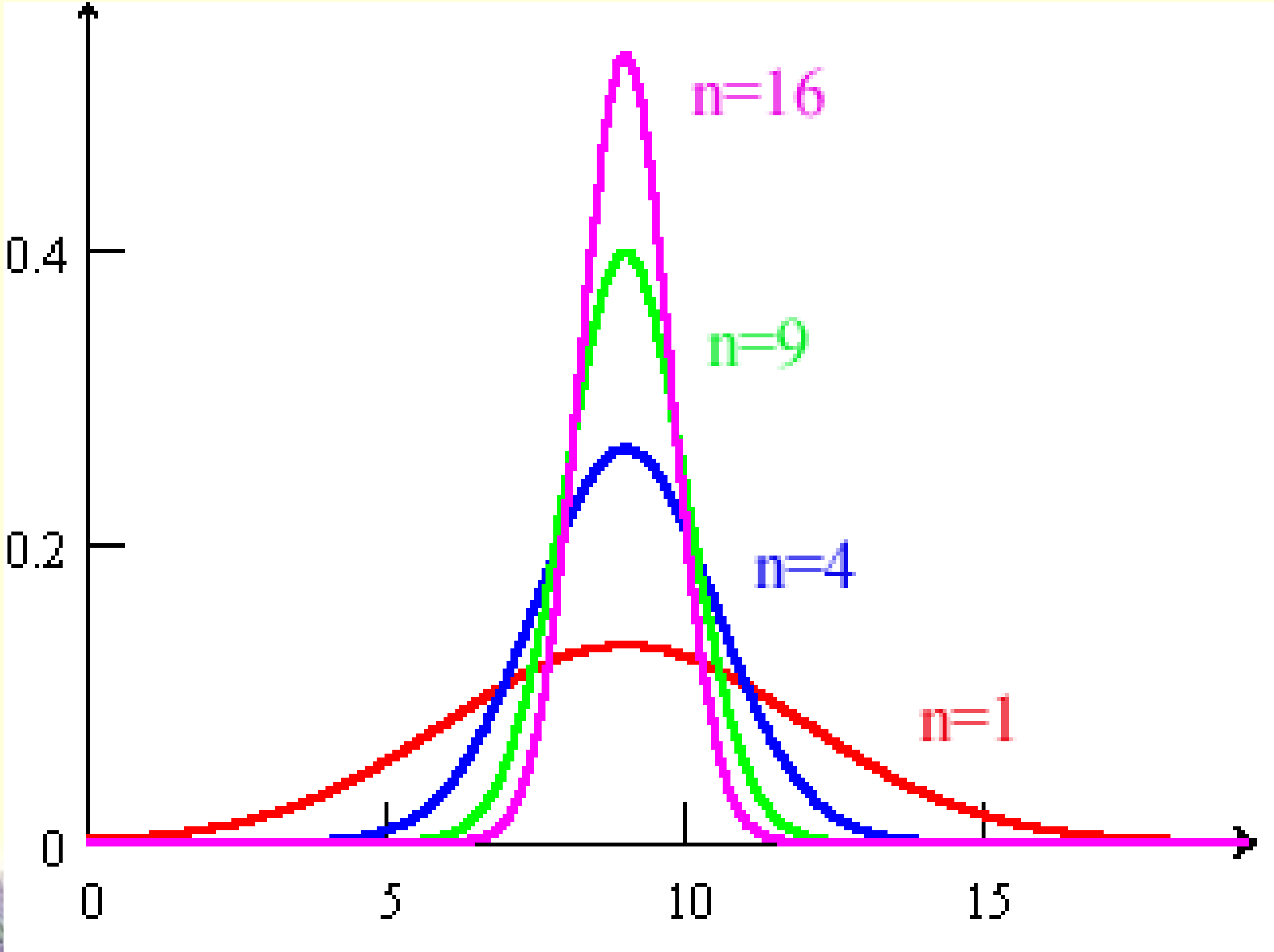
则 $$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

定理 1 (样本均值的分布)

设$X_1, X_2,\ldots, X_n$是取自正态总体$N(\mu,\sigma^2)$的样本，则有

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

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



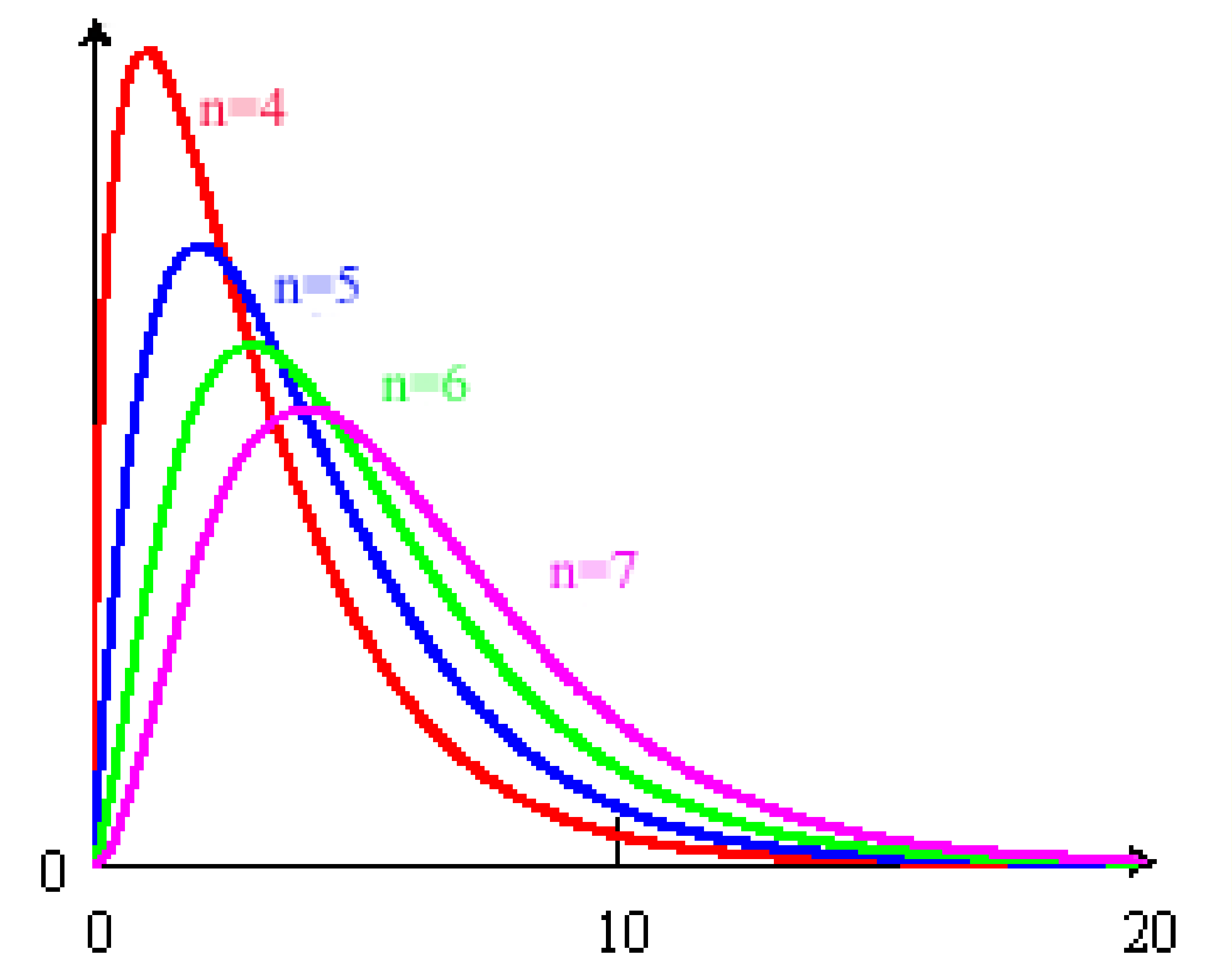
n取不同值时样本均值 $\overline{X}$ 的分布

定理 2 (样本方差的分布)

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$的样本，$\overline{X}$和S^2分别为**样本均值和样本方差,** 则有

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

(2) $\overline{X}$和S^2相互独立.



n取不同值时 $\dfrac{(n-1)S^2}{\sigma^2}$ 的分布

定理 3

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$的样本, $\overline{X}$和S^2分别为**样本均值和样本方差**, 则有

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

证明：由定理1及定理2知

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

且$\overline{X}$和S^2相互独立.

所以

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

定理 4 (两总体样本均值差的分布)

设 $X \sim N(\mu_1,\sigma^2)$, $Y \sim N(\mu_2,\sigma^2)$, 且X与Y独立, $X_1,X_2,\dots,X_{n_1}$ 是取自X的样本, $Y_1,Y_2,\dots,Y_{n_2}$ 是取自Y的样本, $\overline{X}$和 $\overline{Y}$分别是这两个样本的样本均值, S_1^2和 S_2^2分别是这两个样本的样本方差, 则有

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$$

证明: $\overline{X}-\overline{Y} \sim N(\mu_1-\mu_2,(\frac{1}{n_1}+\frac{1}{n_2})\sigma^2)$

$$U=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$$

$$V=\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$$

$\overline{X} \sim N(\mu_1,\frac{\sigma^2}{n_1})$

$\frac{(n_1-1){S_1}^2}{\sigma^2} \sim \chi^2(n_1-1)$

$\overline{Y} \sim N(\mu_2,\frac{\sigma^2}{n_2})$

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

已知U,V独立，所以

$$\frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

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

定理 5 (两总体样本方差比的分布)

设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且$X$与$Y$独立，$X_1, X_2, \ldots, X_{n_1}$是取自$X$的样本，$Y_1, Y_2, \ldots, Y_{n_2}$ 是取自Y的样本，$\overline{X}$和 $\overline{Y}$ 分别是这两个样本的样本均值，S_1^2和 S_2^2 分别是这两个样本的样本方差，则有

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

性质

1 $X \sim F(n_1,n_2)$,则$1/X \sim F(n_2,n_1)$

例4

设$X_1, X_2, \cdots, X_{10}$独立同分布，均服从$N(\mu, \sigma^2)$,

求$\dfrac{1}{\sigma^2}\sum_{i=1}^{5}(X_i-\bar{X}_5)^2+\dfrac{1}{\sigma^2}\sum_{k=6}^{10}(X_k-\mu)^2$

设$S_5^2=\dfrac{1}{4}\sum_{i=1}^{5}(X_i-\bar{X}_5)^2$为$X_1, X_2, \ldots X_5$的样本方差,

$$U=\frac{1}{\sigma^2}\sum_{i=1}^{5}(X_i-\bar{X}_5)^2=\frac{4S_5^2}{\sigma^2}\sim\chi^2(4)$$

$$V=\frac{1}{\sigma^2}\sum_{k=6}^{10}(X_k-\mu)^2\sim\chi^2(5)$$

易知U,V独立,由卡方分布可加性

$$U+V=\frac{1}{\sigma^2}\sum_{i=1}^{5}(X_i-\bar{X}_5)^2+\frac{1}{\sigma^2}\sum_{k=6}^{10}(X_k-\mu)^2\sim\chi^2(9)$$

例5 从正态总体 $X \sim N(\mu, \sigma^2)$ 中，抽取了 $n = 20$的样本$X_1, X_2, \cdots, X_{20}$

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

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

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

$$\overset{\text{查表}}{=} 0.99 - 0.01 = 0.98$$

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

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

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

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

$$= 0.995 - 0.025 = 0.97$$

设$X_1, X_2, \cdots, X_9$独立同分布，均服从$N(0,\sigma^2)$,

（1）$\dfrac{2(X_1+X_2-X_3)^2}{(X_4-X_5+X_6)^2+(X_7+X_8+X_9)^2}$的分布

(2) 求c,使$P\left(\dfrac{(X_1+X_2-X_3)^2}{(X_4-X_5+X_6)^2+(X_7+X_8+X_9)^2}<c\right)=0.9$

作业2：

7,9,10,11,13,16,19