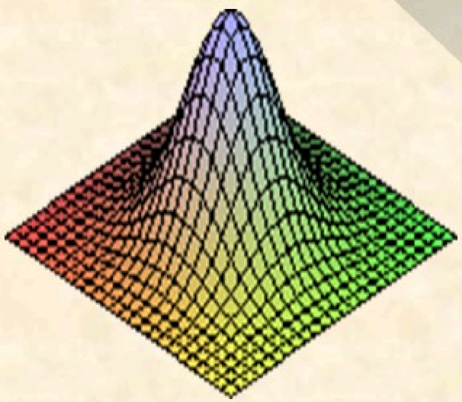
概率论与数理统计



制作人：叶鹰　吴娟

主讲人：吴 娟

wujuan@hust.edu.cn

§6.1 总体与样本

一、总体

研究对象的全体 $\xrightarrow{\text{量化}}$ 指标集 $\xrightarrow{\text{规律}}$ R.V. X

总体： 研究对象的某项数量指标的值的全体.

个体： 总体中的每个元素为个体.

考察某大学一年级学生的年龄　总体 X 的概率分布是：

年龄	18	19	20	21	22
比例	0.5	0.3	0.1	0.07	0.03

$$\begin{pmatrix} 18 & 19 & 20 & 21 & 22 \\ 0.5 & 0.3 & 0.1 & 0.07 & 0.03 \end{pmatrix}$$

二、样本

总体的部分个体： $X_1, X_2, \ldots, X_n$ 独立同分布于$F(x)$

试验前： $X_1, X_2, \ldots, X_n$ 为 R.V.

试验后： $x_1, x_2, \ldots, x_n$ 为样本观察值（实数）

n： 样本容量（样本大小）

基本思想：

由样本对总体的分布（特征）进行合理地**推断**.

三、理论分布与经验分布函数

1．理论分布函数 $F(x)$

对总体 $F(x)$： 样本的联合分布 $F(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} F(x_i)$

离散型总体： $P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X = x_i)$

如： $X_1, X_2, \dots, X_n$ 为取自总体 $B(1, p)$ 的样本，则其联合分布率

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}$$

连续型总体： $f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$

如： $X_1, X_2, \dots, X_n$ 为取自总体 $N(0,1)$ 的样本，则其联合密度函数

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = (2\pi)^{-n/2} \exp(-\tfrac{1}{2} \sum_{i=1}^{n} x_i^2)$$

2．经验分布函数 $F_n(x)$

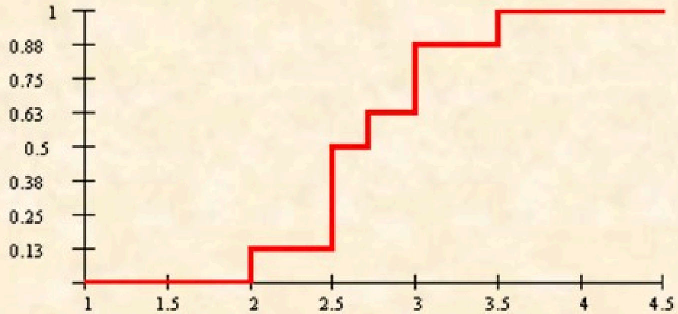
(1) $x_1, x_2, \cdots, x_n \to x_1^* \le x_2^* \le \cdots \le x_n^*$

(2) $$F_n(x) = \begin{cases} 0 & x < x_1^* \\ k/n & x_k^* \le x < x_{k+1}^* \quad k = 1,2,\cdots,n-1 \\ 1 & x \ge x_n^* \end{cases}$$

例 随机地观测总体X得8个数据：2.5, 3, 2.5, 3.5, 3, 2.7, 2.5, 2, 试求X的一个经验分布函数.

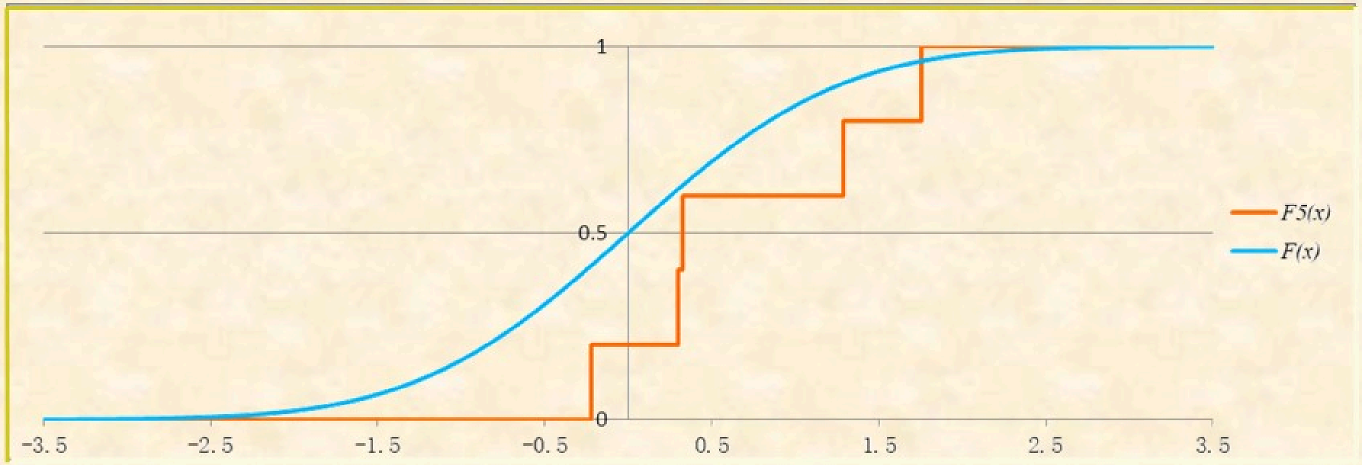
解: $2 < 2.5 = 2.5 = 2.5 < 2.7 < 3 = 3 < 3.5$

$$F_8(x) = \begin{cases} 0 & x < 2 \\ 1/8 & 2 \le x < 2.5 \\ 4/8 & 2.5 \le x < 2.7 \\ 5/8 & 2.7 \le x < 3 \\ 7/8 & 3 \le x < 3.5 \\ 1 & x \ge 3.5 \end{cases},$$

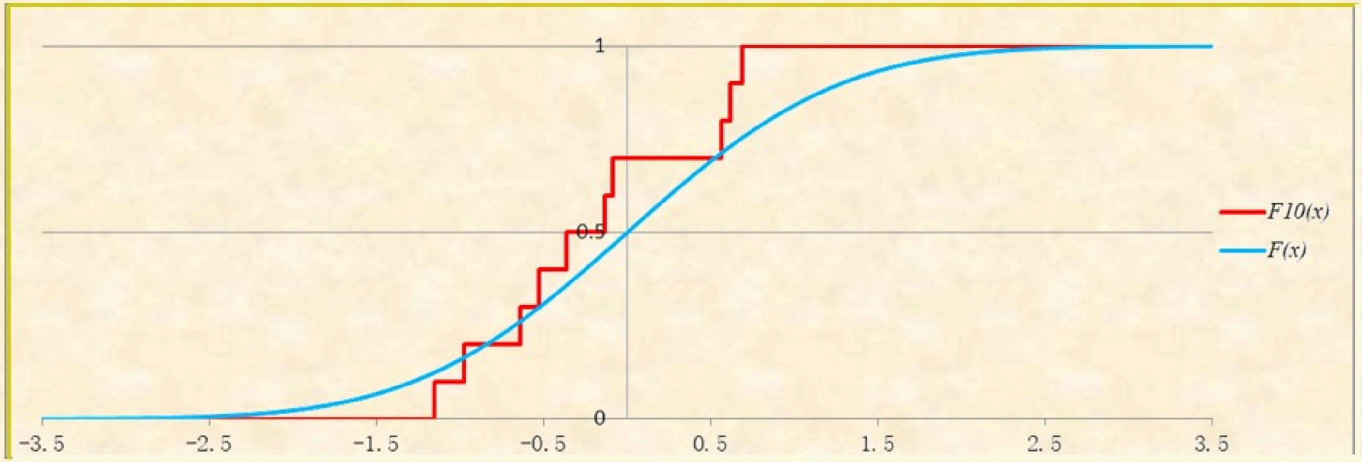


X	2	2.5	2.7	3	3.5
P	1/8	3/8	1/8	2/8	1/8

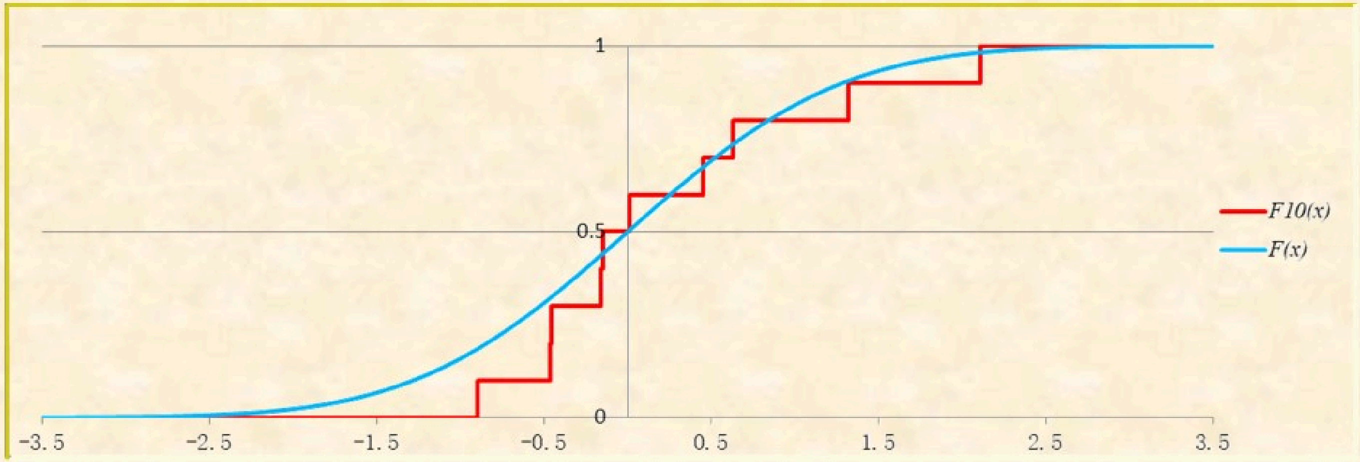
格列汶科定理 $P(\lim_{n\to\infty}\sup_{x}|F_n(x)-F(x)|=0\ )=1$



格列汶科定理 $P(\lim_{n\to\infty}\sup_{x}|F_n(x)-F(x)|=0\ )=1$

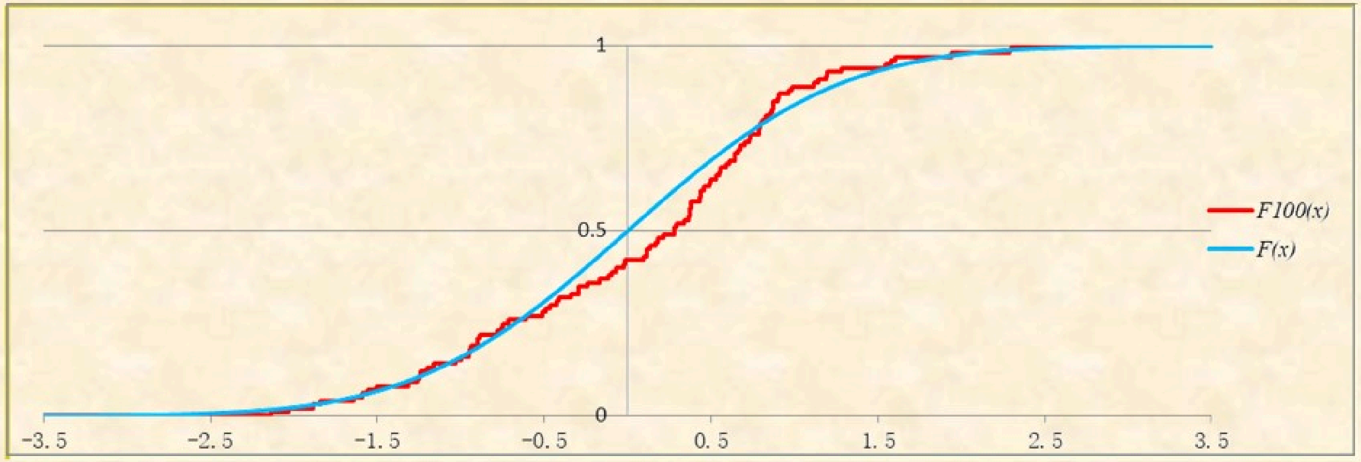


格列汶科定理 $P(\lim_{n\to\infty}\sup_{x}|F_n(x)-F(x)|=0\ )=1$

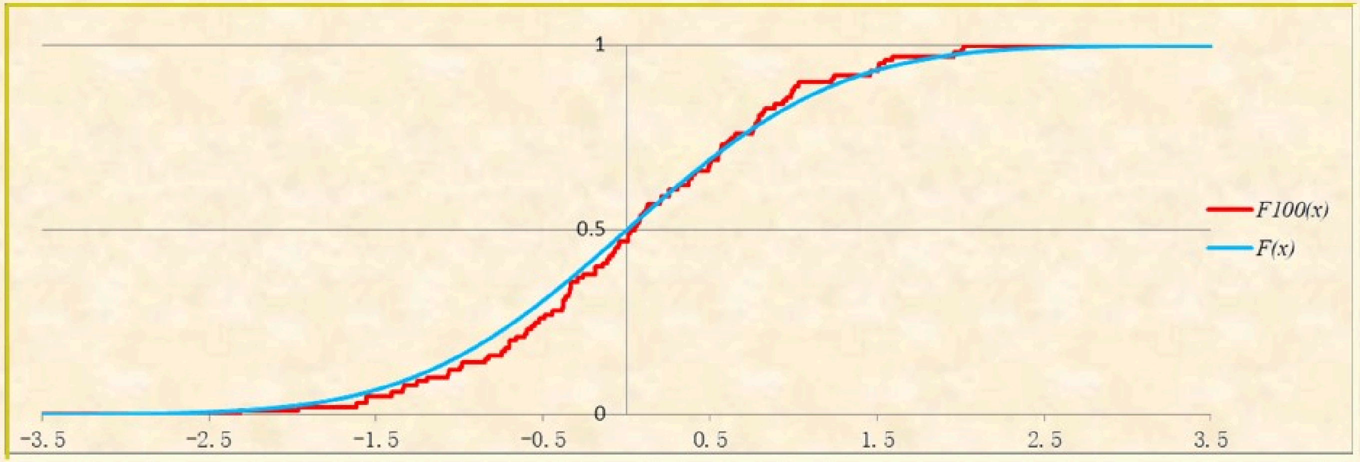


格列汶科定理

$$P(\lim_{n\to\infty}\sup_{x}|F_n(x)-F(x)|=0\ )=1$$



格列汶科定理 $P(\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0\ ) = 1$



四、统计量

定义 设$X_1, X_2, \ldots, X_n$为来自总体X的一个样本，若

（1）$T = g(X_1, X_2, \cdots, X_n)$ 连续;

（2）$T = g(X_1, X_2, \cdots, X_n)$ 中不含有关总体的未知参数.

则称 $T = g(X_1, X_2, \cdots, X_n)$ 为统计量，称 $t = g(x_1, x_2, \cdots, x_n)$ 为统计量观察值.

统计量是随机变量

单选题 100分

设X_1，X_2，…，X_n为来自总体$N(\mu,\sigma^2)$的样本，其中μ已知，σ^2未知，哪个不是统计量？

A $X_1 + X_n$

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

C $\sum_{i=1}^{n} |X_i|/\sigma$

D $\min_{1\le i\le n}\{X_i\}$

常用统计量

1、样本均值 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \rightarrow \quad E(X)$

2、样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \rightarrow \quad D(X)$

样本标准差 $S=\sqrt{S^2}$

3、样本k阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad \rightarrow \quad E(X^k)$

$$A_1=\bar{X}$$

4、样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad \rightarrow \quad E(X-EX)^k$

$$B_2=\frac{n-1}{n}S^2=\tilde{S}^2$$

常用统计量

5、顺序统计量 $X_1, X_2, \cdots, X_n \to X_1^* \le X_2^* \le \cdots \le X_n^*$

次数	X_1	X_2	X_3	X_4	X_5
1	3	1	10	5	6
2	2	6	7	2	8
3	8	3	9	10	5

次数	X_1^*	X_2^*	X_3^*	X_4^*	X_5^*
1	1	3	5	6	10
2	2	2	6	7	8
3	3	5	8	9	10

中位数 $$\tilde{X} = \begin{cases} X_{m+1}^* & n = 2m+1 \\ \frac{1}{2}(X_m^* + X_{m+1}^*) & n = 2m \end{cases} \quad \to \quad E(X)$$

极差 $$R = X_n^* - X_1^* \quad \to \quad D(X)$$

例　设有一容量n=8的样本观察值为(8,6,7,5,7,8,9,6)，求样本均值及样本方差的观察值.

解：
$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

$$\bar{x}=\frac{1}{8}(8+6+7+5+7+8+9+6)=7$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)$$

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

$$=\frac{1}{7}(404-392)=\frac{12}{7}$$

§6.2　抽样分布

一、χ^2分布

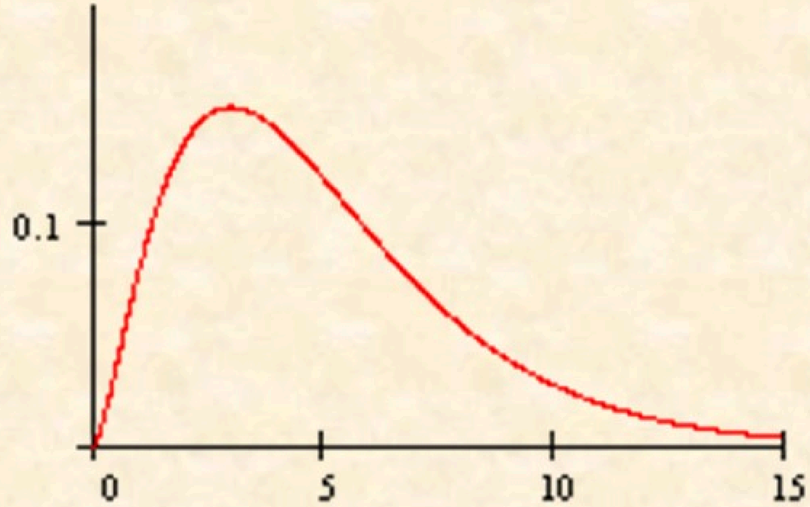
设X_1，X_2，…，X_n独立同分布于N（0，1），则称

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

服从自由度为 n 的χ^2分布，记$\chi^2 \sim \chi^2(n)$.

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

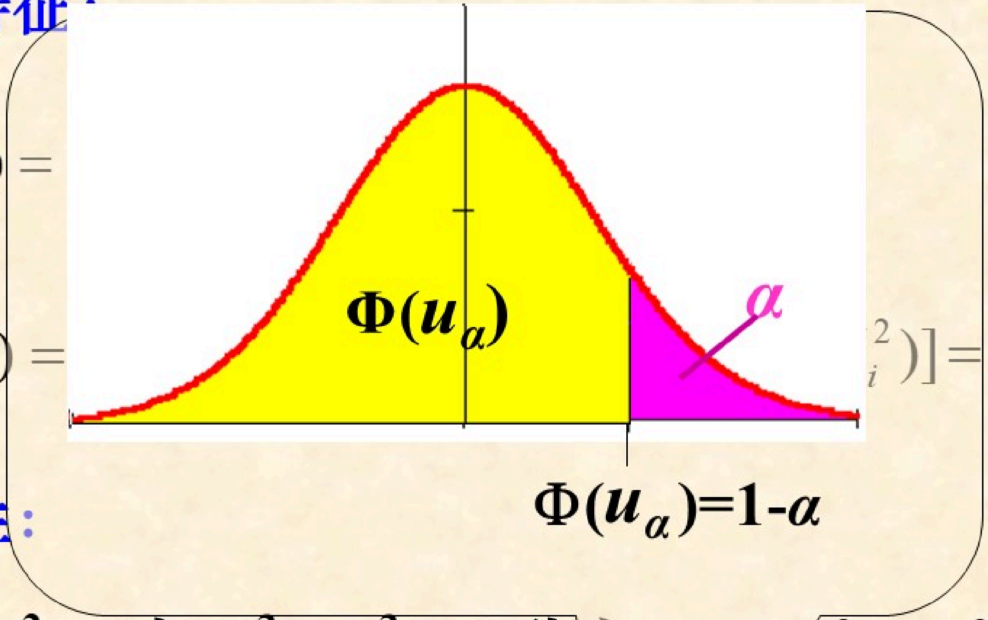
$$\Gamma(\alpha)=\int_0^{+\infty} x^{\alpha-1}e^{-x}dx$$



•数字特征:

$E(\chi^2) =$

$D(\chi^2) =$ $\ldots{}_i^2)] = \sum_{i=1}^{n}(3-1) = 2n$



•可加性:

$\chi_1^2 \sim \chi^2(n_1)$ 与 $\chi_2^2 \sim \chi^2(n_2)$ 独立 $\Longrightarrow$ $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1+n_2)$

•上侧分位点:

$P(\chi^2 > \chi_\alpha^2(n)) = \alpha$ $\qquad \chi_{0.01}^2(20) = 37.566$

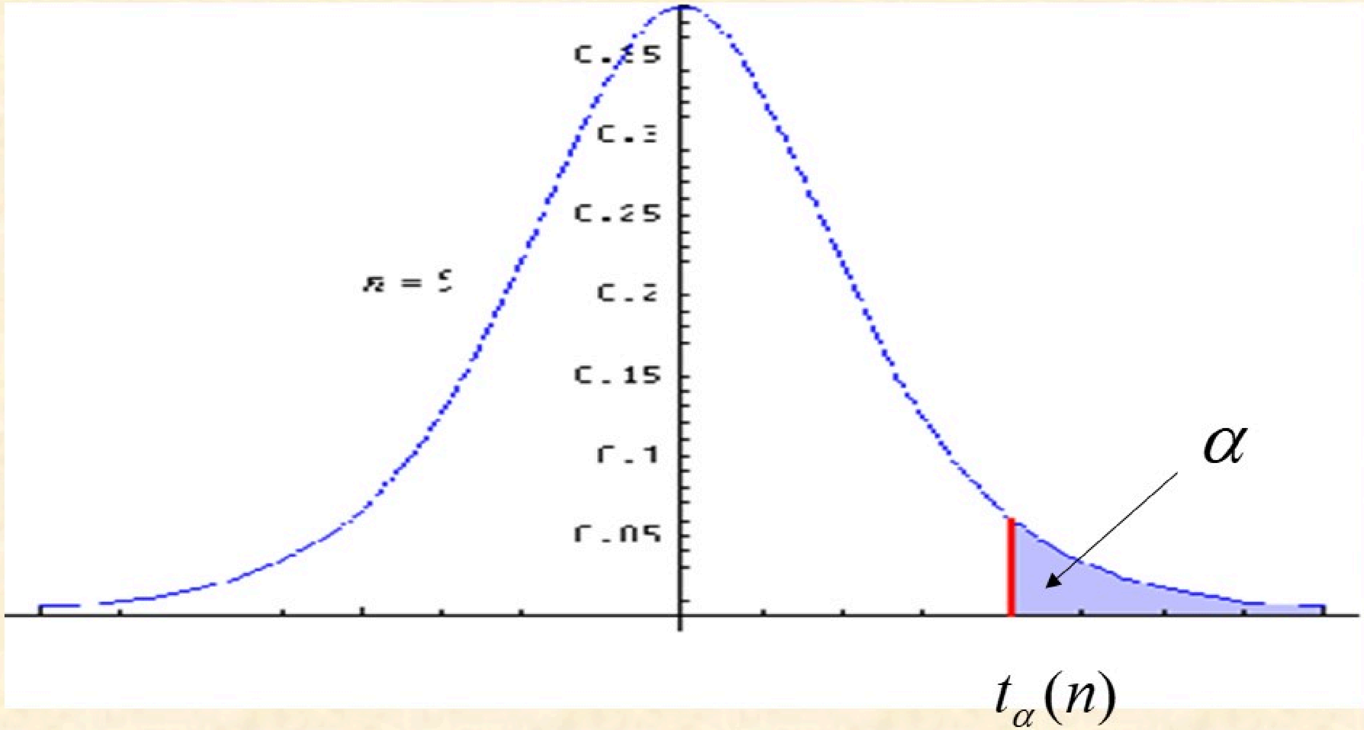
当$n>45$时，有 $\chi_\alpha^2(n) \approx \frac{1}{2}(u_\alpha + \sqrt{2n-1})^2$

查表

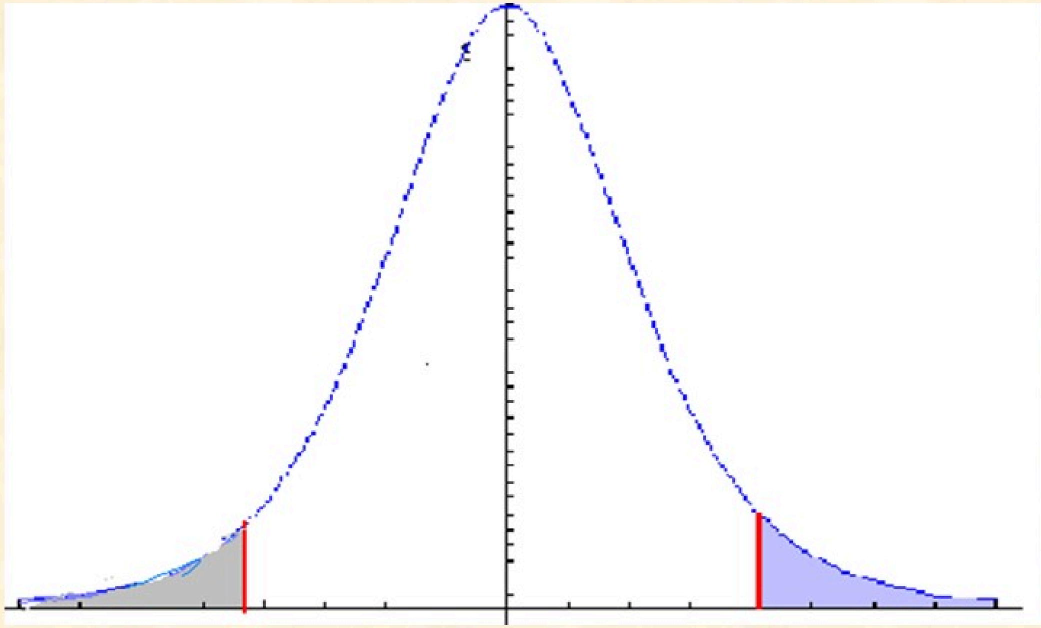
二、*t* 分布

设 $X \sim N(0,1)$ 与 $Y \sim \chi^2(n)$ 独立，则称 $T = \dfrac{X}{\sqrt{Y/n}}$

服从自由度为 n 的 t 分布，记 $T \sim t(n)$.



渐进正态　　$P(T > t_\alpha(n)) = \alpha$



$t_{1-\alpha}(n)$　　　　　　　　$t_{\alpha}(n)$

$$t_{1-\alpha}(n) = -t_{\alpha}(n)$$

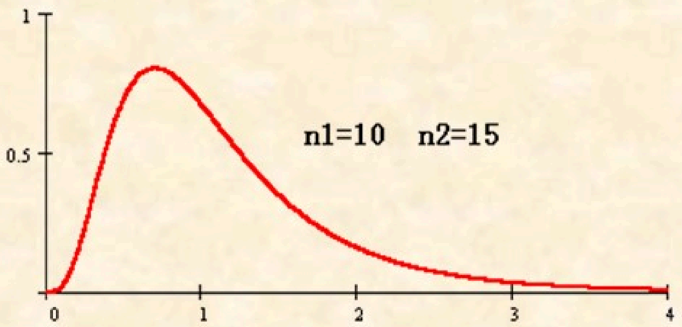
三、 F分布

设$X \sim \chi^2(n_1)$与$Y \sim \chi^2(n_2)$相互独立，则称

$$F = \frac{X/n_1}{Y/n_2} \qquad \frac{1}{F} \sim F(n_2, n_1)$$

服从自由度为（n_1，n_2）的**F分布**，记$F \sim F(n_1, n_2)$。

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



•上侧分位点：$P(F > F_\alpha(n_1, n_2)) = \alpha$

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

查表