西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第五章 数理统计的基本概念与抽样分布

第一节 基本概念

第二节 常用统计分布

第三节 抽样分布

第二节　常用统计分布

- 一、常见分布
- 二、概率分布的分位数

一、常见分布

1. χ^2分布

(1) 定义

定义5.6 设$X_1, X_2, \cdots, X_n$是来自总体$N(0,1)$的样本，则称统计量$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$服从自由度为$n$的$\chi^2$分布. $\sim \chi^2(n)$

自由度n:

指 $\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 中右端包含独立变量的个数.

(2) χ_n^2分布的概率分布

χ_n^2分布的概率密度：

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x>0\\ 0 & \text{其它}\end{cases}$$

证

$$\Gamma(\alpha,\beta)\text{分布}:p(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}$$

$$\therefore \Gamma(\frac{n}{2},\frac{1}{2})=\chi^2(n)$$

其中：$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx \quad (\alpha>0)$

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

$$\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi}$$

特别的$X\sim N(0,1)$, 则$X^2\sim\chi^2(1)=\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$.

$\Rightarrow$ $\chi^2(n)=\Gamma(\frac{n}{2},\frac{1}{2})$ $\qquad$ $\Gamma(1,\beta)=Exp(\beta)$

$$\therefore\chi^2(2)=\Gamma(1,\frac{1}{2})=Exp(\frac{1}{2})$$

(3) χ^2 分布的性质

性质5.3 (χ^2 分布的可加性)

设 $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, 并且 Y_1, Y_2 独立, 则 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$.

(此性质可以推广到多个随机变量的情形)

设 $Y_i \sim \chi^2(n_i)$, 并且 Y_i $(i = 1, 2, \cdots, m)$ 相互独立, 则 $\sum_{i=1}^{m} Y_i \sim \chi^2(n_1 + n_2 + \cdots + n_m)$.

性质5.4 （χ^2分布的数学期望和方差）

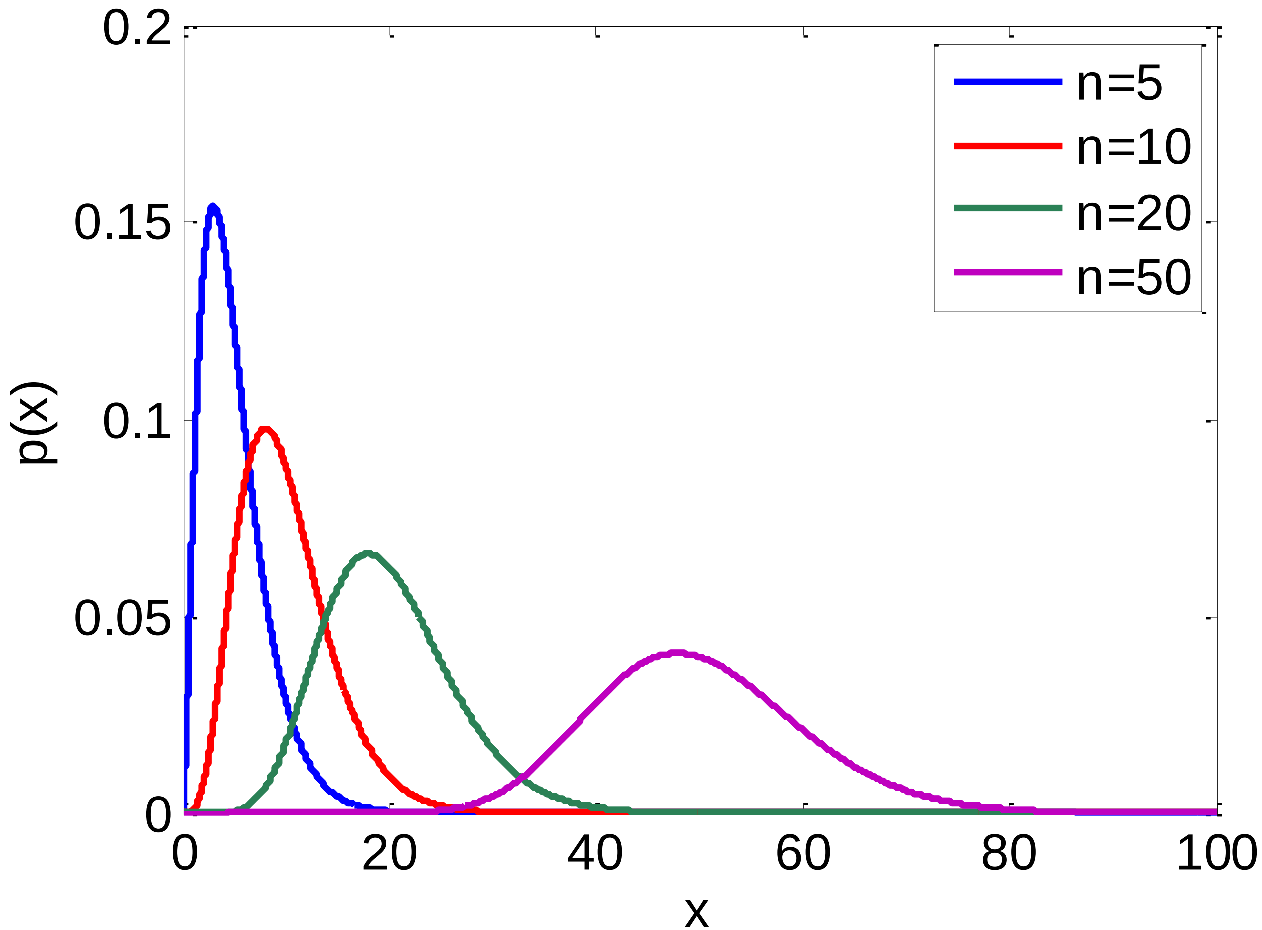
若 $\chi_n^2 \sim \chi^2(n)$，则 $E(\chi_n^2) = n,\quad D(\chi_n^2) = 2n.$

性质5.5 设$\chi_n^2 \sim \chi^2(n)$,则对任意x,有

$$\lim_{n\to\infty} P\left\{\frac{\chi_n^2 - n}{\sqrt{2n}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

即$\chi_n^2 \sim AN(n, 2n)$ (开方分布的极限分布为正态分布)

$$X \sim \chi^2(n)$$

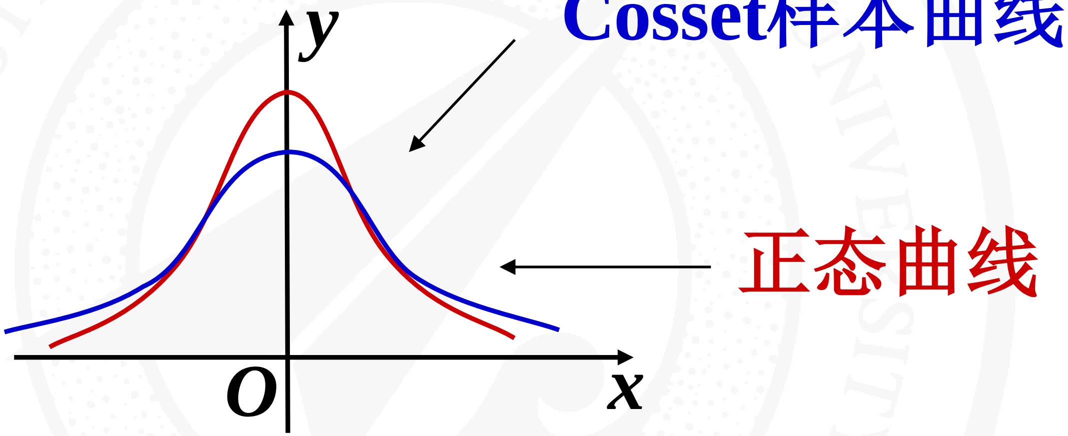


2. *t* 分布

历史上，正态分布由于其广泛的应用背景和良好的性质，曾一度被看作是“**万能分布**”，我们知道在总体均值和方差已知情况下，**样本均值的分布将随样本量增大而接近正态分布, 即** $\overline{X} \sim AN(\mu, \frac{1}{n}\delta^2).$

在这样的背景下，十九世纪初英国一位年轻的酿酒化学技师**Cosset. WS**, 他在酒厂从事试验数据分析工作，对数据误差有着大量感性的认识，

但是Cosset在实验中遇到的样本容量仅有5~6个，在其中他发现实际数据的分布情况与正态分布有着较大的差异.



于是Cosset怀疑存在一个不属于正态的其他分布，通过学习终于得到了新的密度曲线，并在1908年以“Student”笔名发表了此项结果，后人称此分布为“*t* 分布”或“学生氏”分布.

(1) 定义

定义5.7 设 $X \sim N(0,1),\ Y \sim \chi^2(n)$, 且 X, Y 独立,则称随机变量

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

服从自由度为 n 的 t 分布, 记为 $T \sim t(n)$.

t 分布又称**学生氏**(**Student**)**分布**.

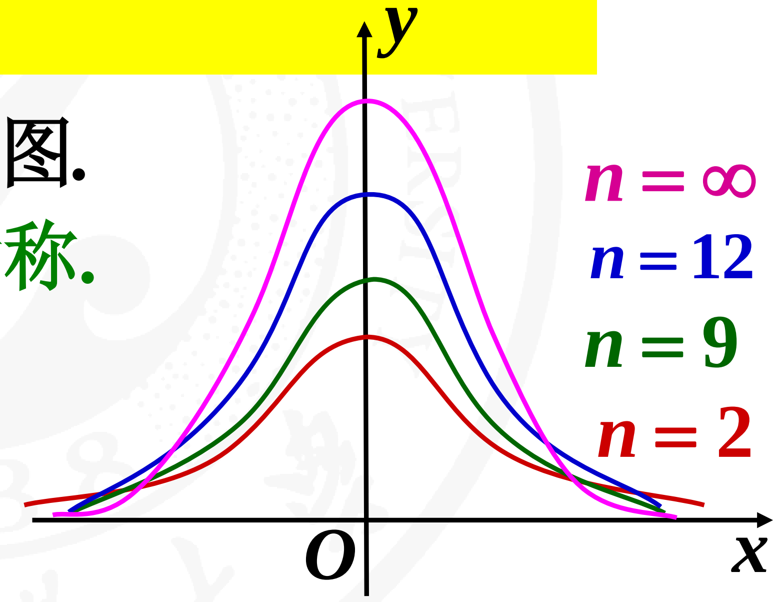
(2) $t(n)$ 分布的概率密度函数为

$$p(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<x<+\infty$$

t分布的概率密度曲线如图.

显然图形是关于 $t=0$ 对称.

当n充分大时，其图形类似于标准正态变量概率密度的图形.

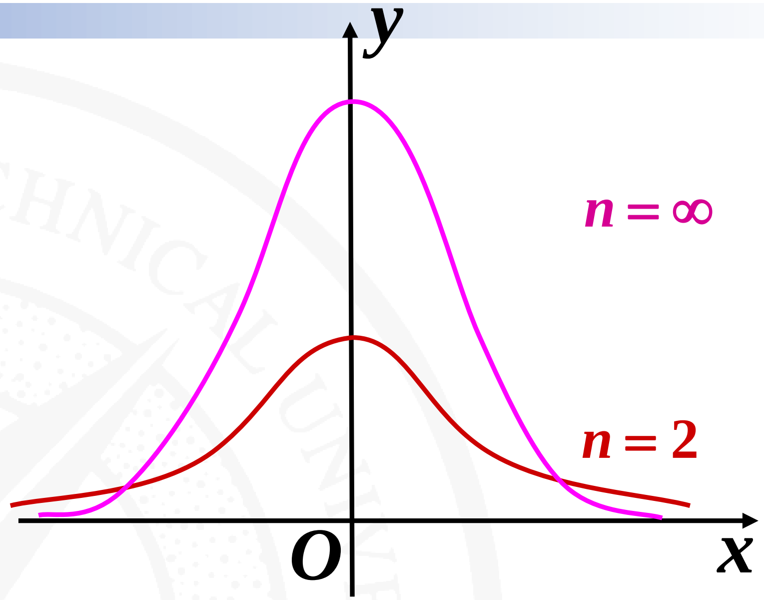


(3) t分布的性质

性质5.6 若$T \sim t(n)$，则

$$E(T) = 0,$$

$$D(T) = \frac{n}{n-2} \qquad (n > 2).$$



性质5.7 若$T \sim t(n)$，则

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

即$T \sim AN(0,1)$分布

但对于较小的n，t分布与$N(0,1)$分布相差很大，

具有厚尾性。

3. F分布

(1) 定义

定义5.8 设$X \sim \chi^2(n_1)$,$Y \sim \chi^2(n_2)$,且X,Y相互独立,

则称随机变量

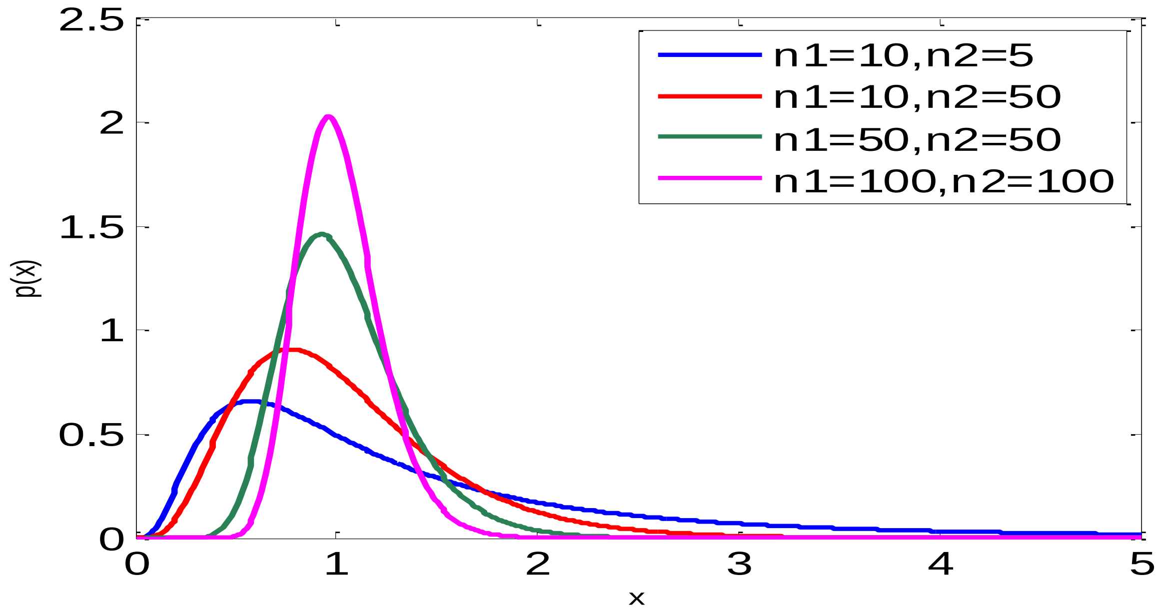
$$F = \frac{X/n_1}{Y/n_2}$$

服从自由度为(n_1,n_2)的F分布，记为

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

(2) $F(n_1, n_2)$分布的概率密度为

$$p(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}\left[1+\left(\dfrac{n_1}{n_2}x\right)\right]^{-\frac{n_1+n_2}{2}}, & x>0\\ 0, & \text{其它}\end{cases}$$



(3) F分布有以下性质

1) $$E(F)=\frac{n_2}{n_2-2},\quad (n_2>2),$$

$$D(F)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)},\quad (n_2>4)$$

2) **性质5.8** 若$F\sim F(n_1,n_2)$, 则$\dfrac{1}{F}\sim F(n_2,n_1)$.

3) 设$F\sim F(n_1,n_2)$,则当$n_2>4$时,对任意x有

$$\lim_{n_1\to\infty}P\left\{\frac{F-E(F)}{\sqrt{D(F)}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

这说明F分布极限分布也是正态分布.

$T \sim t(n)$，则$T^2 \sim F(1,n)$.

证 因为$T \sim t(n)$，由定义5.7有

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

其中$X \sim N(0,1), Y \sim \chi^2(n)$，且$X, Y$独立，

从而 $X^2 \sim \chi^2(1), Y \sim \chi^2(n)$，且$X^2$与$Y$独立，

$\therefore$ 由定义5.8，有 $T^2 = \dfrac{X^2}{Y/n} \sim F(1,n).$

例1 设$X \sim N(0,4), Y \sim \chi^2(2)$，且$X,Y$相互独立，试求解$\dfrac{X^2}{4}+Y$的概率分布.

解 因为$X \sim N(0,4)$且X,Y相互独立，所以

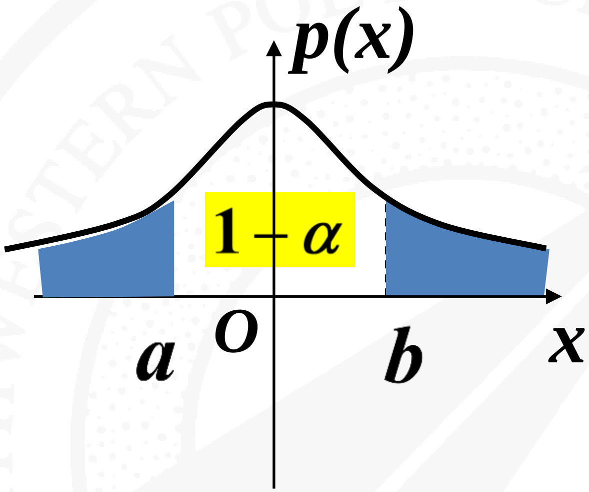
$$\frac{X}{2} \sim N(0,1)$$

且$\dfrac{X^2}{4}$与Y相互独立

又因为$\dfrac{X^2}{4} \sim \chi^2(1)$，由开方分布的可加性得

得 $$\frac{X^2}{4}+Y \sim \chi^2(3).$$

二、概率分布的分位数



已知X的分布 $\Leftrightarrow P(a < X \le b) = 1-\alpha$

(1)给定区间范围a,b,求概率 $P(a < X \le b) = ?$

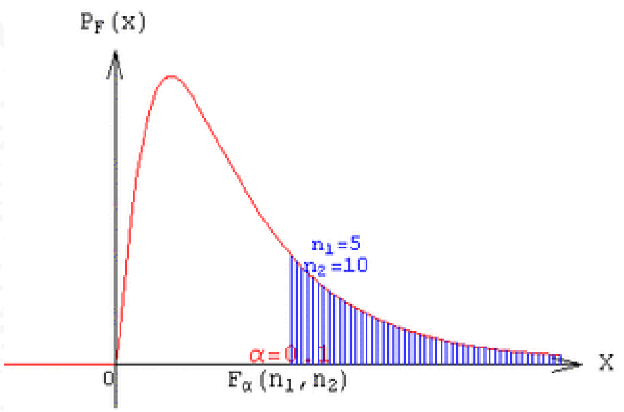
(2)给定概率α求区间范围$[a,b]$

1. 定义

定义5.9 对于总体X和给定的$\alpha(0<\alpha<1)$，若存在x_α，使

$$P\{X > x_\alpha\} = \alpha$$

则称x_α为X的分布的上侧α分位数.



$$\Rightarrow F(x_\alpha) = P\{X \le x_\alpha\} = 1-\alpha$$

2. 常用分布的上侧分位数记号

分布	$N(0,1)$	$\chi^2(n)$	$t(n)$	$F(n_1,n_2)$
记号	u_α	$\chi^2_\alpha(n)$	$t_\alpha(n)$	$F_\alpha(n_1,n_2)$

3. 查表法

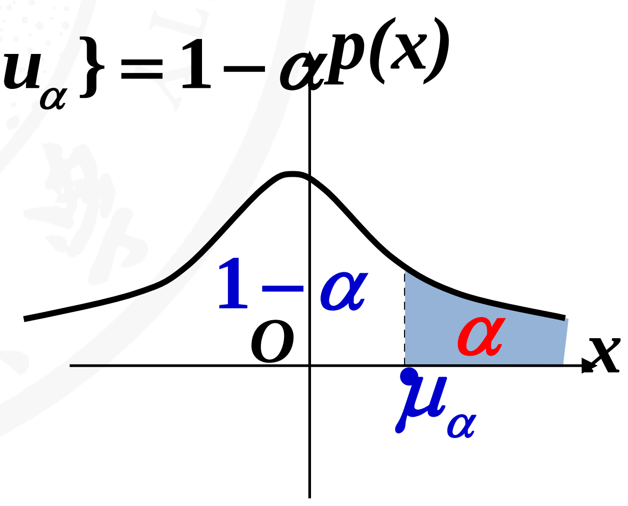
(1) 若X的分布密度关于y轴对称，则

1) 正态分布的上侧分位数u_α:

设$X \sim N(0,1)$,则其上侧分位数u_α满足 $P\{X > u_\alpha\} = \alpha$

$$\Phi(u_\alpha) = P\{X \le u_\alpha\} = 1 - P\{X > u_\alpha\} = 1 - \alpha$$

即 $\Phi(u_\alpha) = 1 - \alpha$



给定α,由附表2可查得u_α的值.

$$\boxed{\Phi(u_\alpha) = P(X \le u_\alpha) = 1-\alpha}$$

$u_{0.05} = 1.645,$ $\boxed{\Phi(1.645) = 0.95}$ $(\alpha = 0.05)$

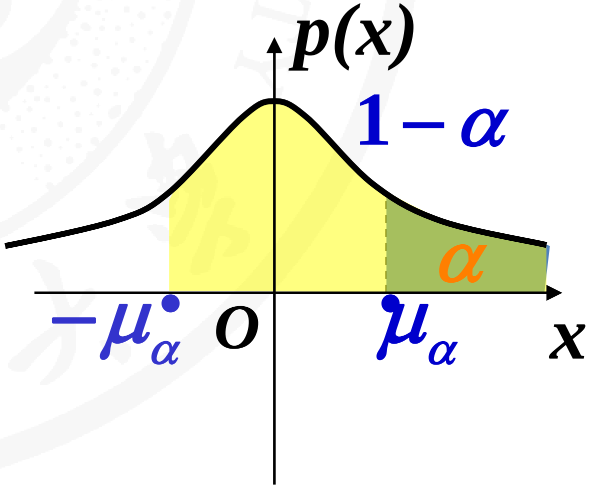
$u_{0.025} = 1.96,$ $\boxed{\Phi(1.96) = 0.975}$ $(\alpha = 0.025)$

根据正态分布的对称性知 $u_{1-\alpha} = -u_\alpha.$

$\therefore$ 当 $\alpha > 0.5$ 无法查表

$$u_\alpha = -u_{1-\alpha}.$$

例： $u_{0.95} = -u_{0.05} = -1.645.$



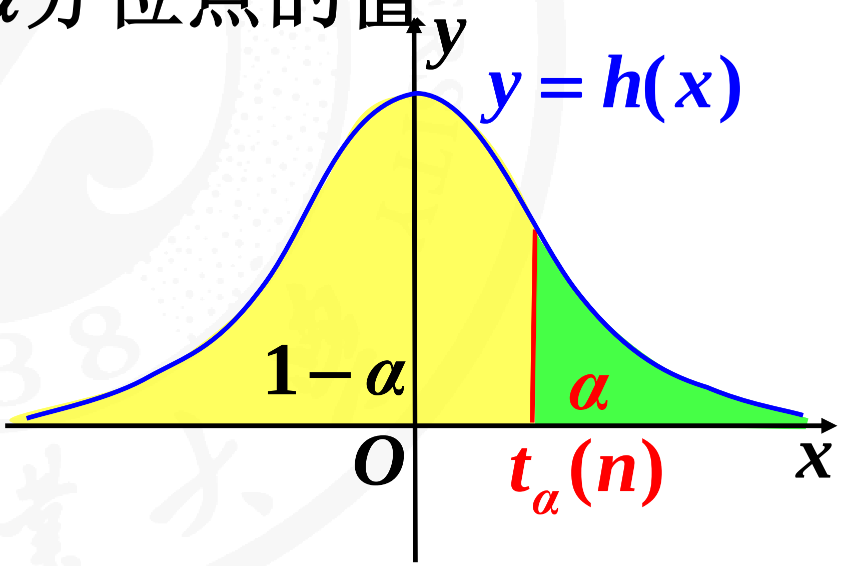
2) t分布的上侧分位数$t_\alpha(n)$:

设$T\sim t(n)$,则其上侧分位数t_α满足

$$P\{T>t_\alpha(n)\}=\alpha$$

可以通过查表求得上α分位点的值

$(\alpha\le 0.25, n\le 45)$.



$t_{0.05}(10)=1.8125,$

$t_{0.025}(15)=2.1315.$

当$\alpha > 0.25, n \le 45,$ 由分布的对称性知

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

$$t_{0.95}(10) = -t_{0.05}(10) \quad = -1.8125,$$

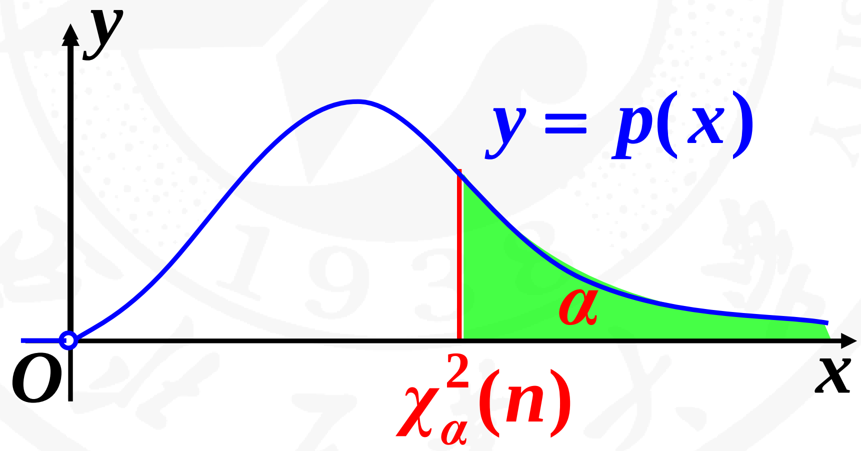
当$n > 45$时，$t_{\alpha}(n) \approx u_{\alpha}.$

$$t_{0.05}(100) \approx \mu_{0.05} \quad = 1.645$$

(2) X的分布密度无对称性的情形

1)设$\chi^2 \sim \chi^2(n)$,则其上侧分位数$\chi^2_\alpha(n)$满足

$$P\{\chi^2 > \chi^2_\alpha(n)\} = \alpha$$

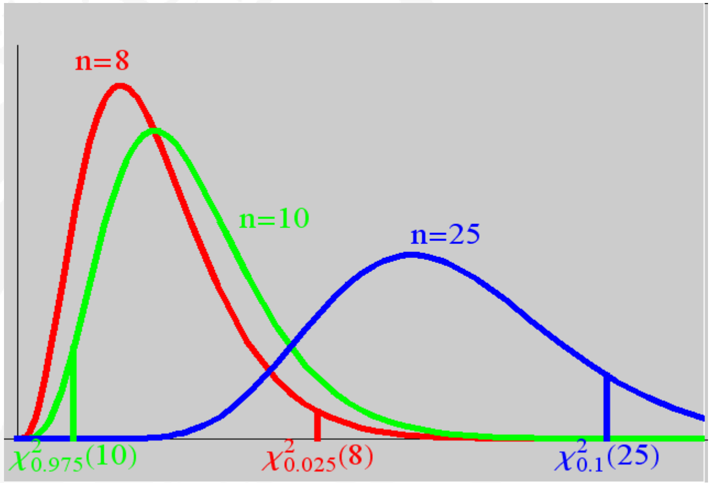


当$n \le 60$时，可查表4(表4只详列到 n=60 为止).

$$\chi^2_{0.025}(8) = 17.5,$$

$$\chi^2_{0.975}(10) = 3.25,$$

$$\chi^2_{0.1}(25) = 34.382.$$



2) F分布的上侧分位数$F_\alpha(n_1,n_2)$:

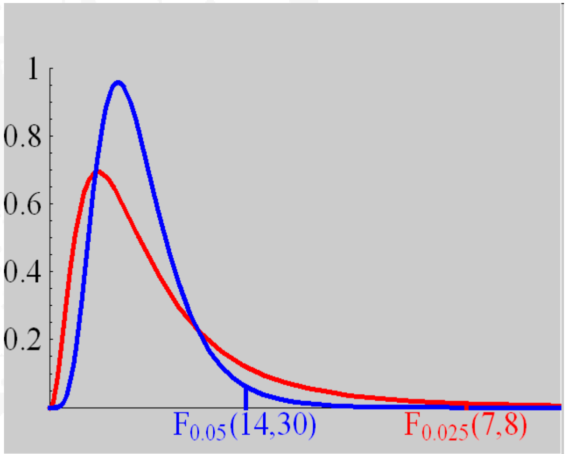
对于 $\alpha = 0.01,\ 0.025,\ 0.05,\ 0.1$ 等，可直接查表5 ~ 8.

$F_{0.05}(14,30) = 2.04.$ $F_{0.025}(7,8) = 4.53,$

当α为其它值，可利用关系式

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

由F_α求得$F_{1-\alpha}$.



如： $F_{0.95}(12,9) = \dfrac{1}{F_{0.05}(9, 12)} = \dfrac{1}{2.8} = 0.357.$

内容小结

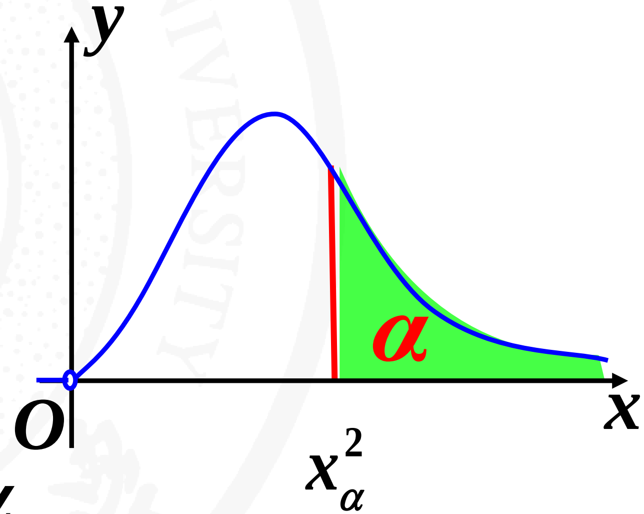
1.三大抽样分布:

χ^2 分布, t 分布, F 分布

的定义,性质.

2.概率分布的分位数概念. x_α

$$P\{X > x_\alpha\} = \alpha$$



$$\Rightarrow F(x_\alpha) = P(X \le x_\alpha) = 1 - \alpha$$

$$u_{1-\alpha} = -u_\alpha; \quad t_{1-\alpha}(n) = -t_\alpha(n) \quad F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}.$$

(1)χ^2 分布　$X_1, X_2, \cdots, X_n \sim N(0,1)$且$X_i$相互独立，

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$$

$$\Rightarrow E(\chi_n^2) = n, \quad D(\chi_n^2) = 2n$$

(2)t 分布　$X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X,Y独立,

$$\text{则} T = \frac{X}{\sqrt{Y/n}} \sim t(n) \quad \Rightarrow E(T) = 0,\ D(T) = \frac{n}{n-2}(n > 2).$$

(3)F 分布

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

$$\text{则} F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)\text{。}$$

若$T \sim t(n)$, 则　$T^2 \sim F(1,n)$

分布	统计量	上侧分位数	性质
$N(0,1)$	U	μ_α	$\mu_{1-\alpha}=-\mu_\alpha$, $\Phi(\mu_\alpha)=1-\alpha$
$\chi^2(n)$	$\chi_n^2=\sum_{i=1}^{n}U_i^2$	$\chi_\alpha^2(n)$	$\chi_\alpha^2(n)\approx n+\sqrt{2n}\mu_\alpha$ $(n>60)$
$t(n)$	$T=\dfrac{U}{\sqrt{\chi_n^2/n}}$	$t_\alpha(n)$	$t_{1-\alpha}(n)=-t_\alpha(n)$, $t_\alpha(n)\approx\mu_\alpha(n>45)$
$F(n_1,n_2)$	$F=\dfrac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2}$	$F_\alpha(n_1,n_2)$	$F_\alpha(n_1,n_2)=\dfrac{1}{F_{1-\alpha}(n_2,n_1)}$

西北工業大學

NORTHWESTERN POLYTECHNICAL UNIVERSITY

5-2 常用统计分布

Thank You !