概率论与数理统计

第三十讲

抽样分布

数据收集 ⇨ 样本、样本观测值

—包含了总体的有用信息

数据整理 ⇨ 统计量

—提炼数据中包含的信息

统计量 $g(X_1, X_2, \ldots, X_n)$ 是随机变量

- **确定统计量的分布是数理统计的基本问题之一**

定义1 统计量 $g(X_1, X_2, \ldots, X_n)$ 的分布称为

抽样分布.

本讲主要介绍与标准正态总体相关

的抽样分布：

χ^2 - 分布　　t - 分布　　F - 分布

一、χ^2-分布

定义2 设 X_1，X_2，…，X_n 相互独立，且都服从标准正态分布 $N(0,\ 1)$，则称随机变量

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

服从自由度为 n 的 χ^2 分布，记为

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

$\chi^2(n)$ 分布的密度函数

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

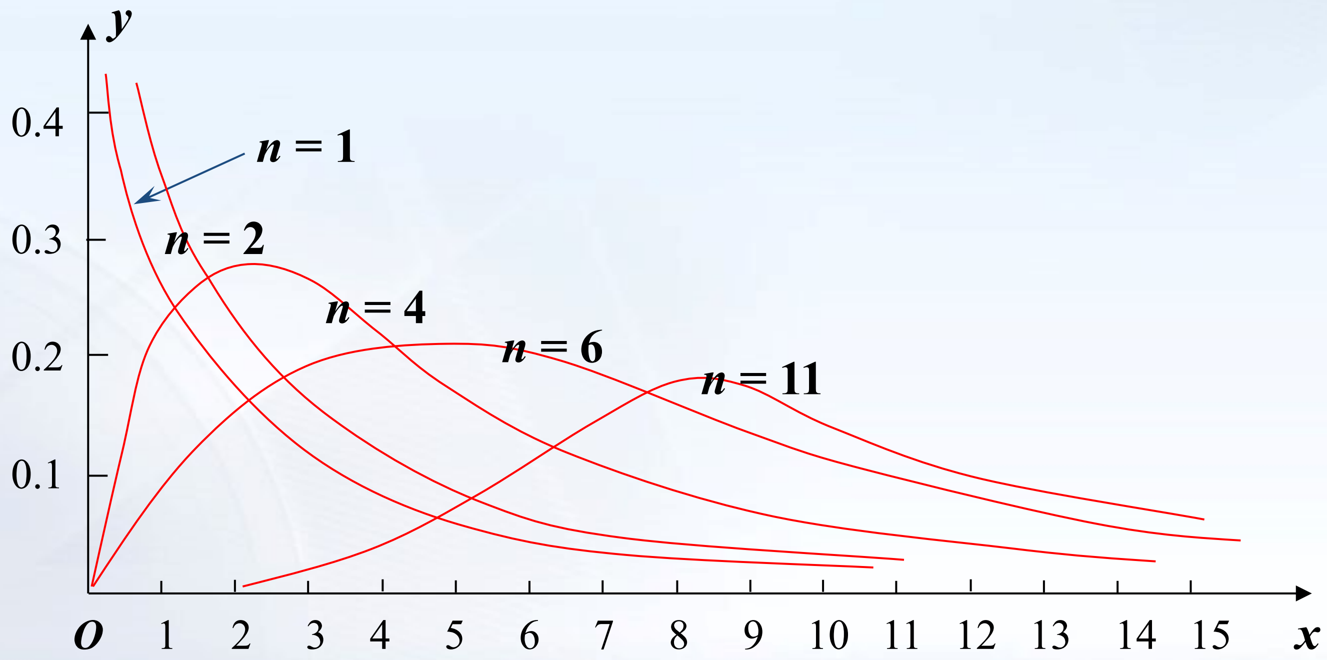
其中

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

称为 Γ 函数，具有性质

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

$\chi^2(n)$ 分布密度函数的图形



χ^2 -分布的性质

1、可加性

若 $Y_1 :\ \chi^2(n)$, $Y_2 :\ \chi^2(m)$，且 Y_1 与 Y_2 相互独立，则有

$$Y_1 + Y_2 :\ \chi^2(n+m)$$

推广：若 $Y_1, Y_2, \mathrm{L}\ , Y_k$ 相互独立，且

$$Y_i :\ \chi^2(n_i), \quad i = 1, 2, \mathrm{L}\ , k$$

则有 $$\sum_{i=1}^{k} Y_i :\ \chi^2\left(\sum_{i=1}^{k} n_i\right)$$

2、数字特征

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

证明：存在 X_1，X_2，$\cdots$，X_n i.i.d. ，服从 $N(0，1)$，

使得

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

由此有

$$E(Y) = E(X_1^2 + X_2^2 + \cdots + X_n^2)$$

$$= E(X_1^2) + E(X_2^2) + \cdots + E(X_n^2) = nE(X_1^2) = n$$

$$D(Y)=D(X_1^2+X_2^2+\cdots+X_n^2)$$

$$=D(X_1^2)+D(X_2^2)+\cdots+D(X_n^2)=nD(X_1^2)$$

$$D(X_1^2)=E(X_1^4)-[E(X_1^2)]^2=E(X_1^4)-1$$

$$E(X_1^4)=\int_{-\infty}^{+\infty}x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x=3\int_{-\infty}^{+\infty}x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x$$

$$=3E(X_1^2)=3$$

$$D(Y)=nD(X_1^2)=n(3-1)=2n$$

二、t - 分布

定义3 设 $X \sim N(0,\ 1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，则称随机变量

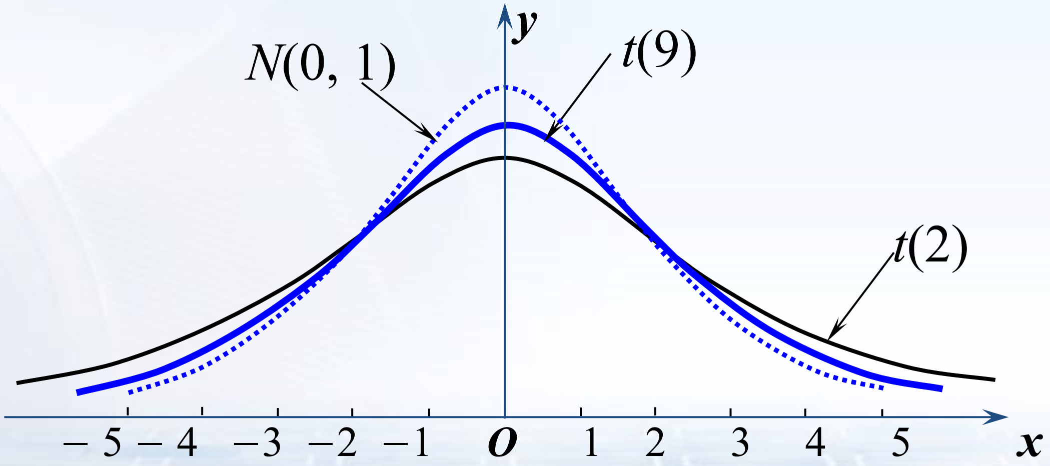
$$T = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布，记为

$$T \sim t(n)$$

$t(n)$ 分布的密度函数及其图形

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



三、F-分布

定义4 设 $X \sim \chi^2(m), Y \sim \chi^2(n)$，且 X 与 Y 相互独立，则称随机变量

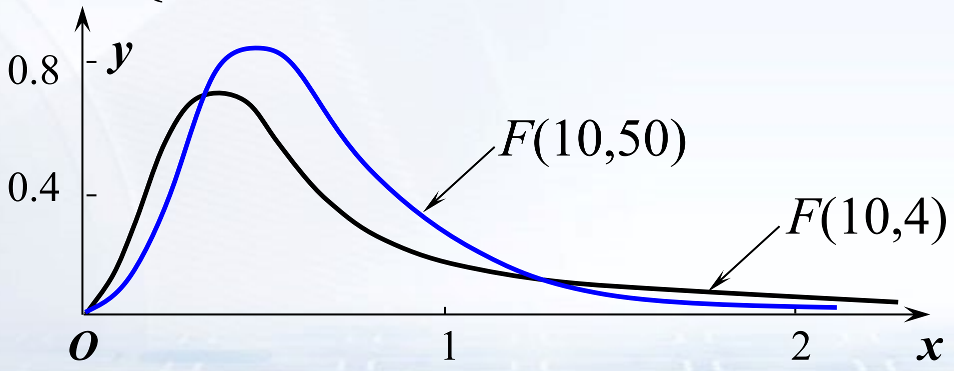
$$T = \frac{X/m}{Y/n}$$

服从自由度为 (m, n) 的 F 分布，记为

$$F \sim F(m,n)$$

$F(m,n)$ 分布的密度函数及其图形

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



F–分布的一个重要性质

若 $F\sim F(m,n)$，则有 $\frac{1}{F}\sim F(n,m)$.

事实上，因 $F\sim F(m,n)$，所以由 F 分布定义知

存在 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立,

使得
$$F=\frac{X/m}{Y/n}$$

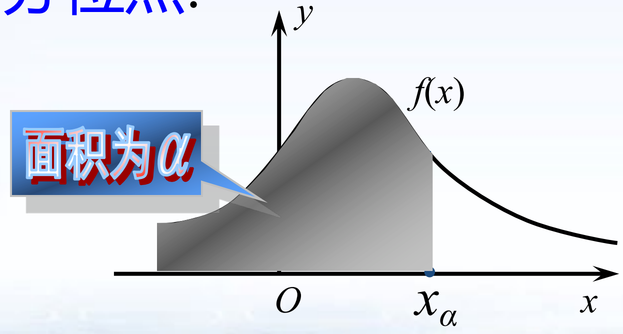
所以有
$$\frac{1}{F}=\frac{Y/n}{X/m}\sim F(n,m)$$

四、分布点

定义5 设连续型随机变量 $X \sim f(x)$，对给定的 α，$0<\alpha<1$，存在一个实数 x_α，使得

$$P(X \le x_\alpha) = \int_{-\infty}^{x_\alpha} f(x)\mathrm{d}x = \alpha$$

则称 x_α 为密度函数 $f(x)$的 α **分位点**.



1. 标准正态分布 $N(0\,,\,1)$ 的 α 分位点记为 u_α .

$$\Phi(u_\alpha) = \int_{-\infty}^{x_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x = \alpha$$

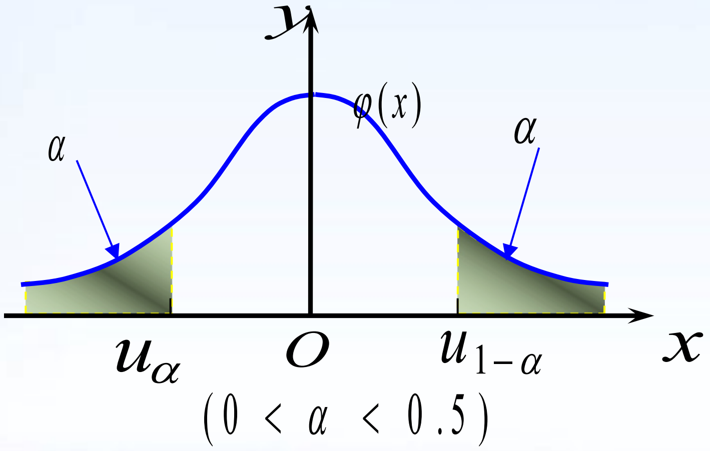
$$\boxed{u_\alpha = -u_{1-\alpha}}$$

查标准正态分布表

$u_{0.975} = 1.96$

$u_{0.95} = 1.645$

$u_{0.05} = -1.645$

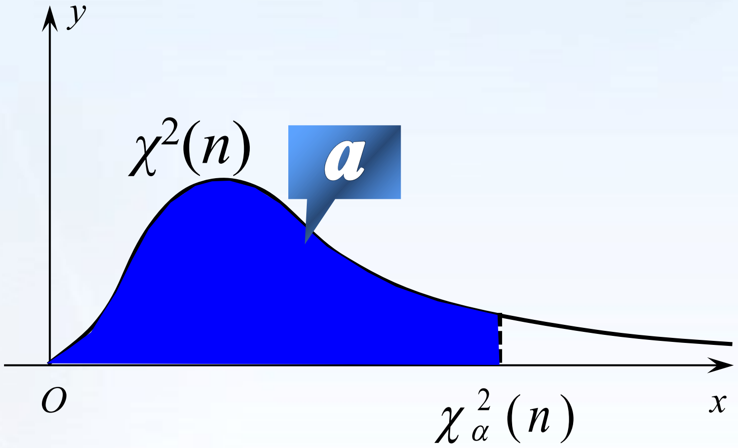


2. $\chi^2(n)$ 的 α 分位点记为 $\chi^2_\alpha(n)$.

查 χ^2 分布表

$$\chi^2_{0.05}(10) = 3.940$$

$$\chi^2_{0.95}(10) = 18.307$$



3. $t(n)$ 分布的α 分位点记为 $t_\alpha(n)$.

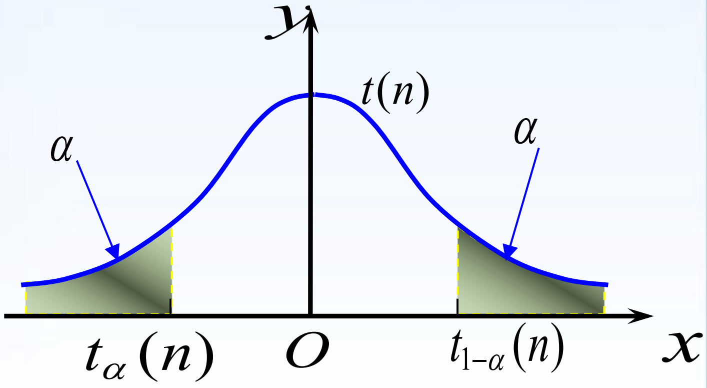
$$\boxed{t_\alpha(n) = -t_{1-\alpha}(n)}$$

查 t 分布表

$$t_{0.975}(10) = 2.2281$$

$$t_{0.95}(18) = 1.7341$$

$$t_{0.05}(20) = -t_{0.95}(20) = -1.7341$$



$(0 < \alpha < 0.5)$

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

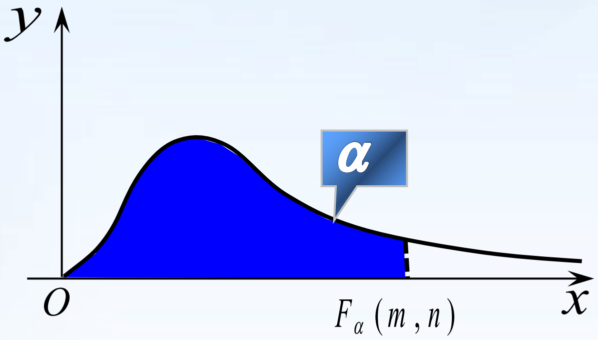
4. $F(m, n)$分布的 α 分位点记为 $F_{\alpha}(m,n)$.

查 $F(m, n)$ 分布表

$$F_{0.95}(10,20)=2.35$$

$$F_{0.05}(15,10)=\frac{1}{F_{0.95}(10,15)}$$

$$=\frac{1}{2.54}$$



若 $F\sim F(m,n)$，则

$$\frac{1}{F}\sim F(n,m) \Longrightarrow \boxed{F_{\alpha}(m,n)=\frac{1}{F_{1-\alpha}(n,m)}}$$