第六章

样本及抽样分布

§6.1 总体与样本

§6.3 抽样分布

§6.3 抽样分布

一、统计量 样本是进行统计推断的依据. 但在应用时, 往往不是直接使用是样本本身, 而是针对不同的问题构造样本的适当函数, 利用这些样本的函数进行统计推断.

定义1 设$X_1, X_2, \ldots, X_n$是来自总体 X 的一个样本, $g(X_1, X_2, \ldots, X_n)$是$X_1, X_2, \ldots, X_n$函数, 若g 中不含任何未知参数, 则称$g(X_1, X_2, \ldots, X_n)$是一个统计量.

[注] (1) 统计量是一个随机变量;

(2) $(x_1, x_2, \ldots, x_n)$是样本$X_1, X_2, \ldots, X_n$的观察值, 则称 $g(x_1, x_2, \ldots, x_n)$ 是$g(X_1, X_2, \ldots, X_n)$的观察值.

实例1 设 X_1, X_2, X_3 是来自总体 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 为已知，σ^2 为未知，判断下列各式哪些是统计量，哪些不是？

$$T_1 = X_1, \qquad T_2 = X_1 + X_2 e^{X_3},$$

$$T_3 = \frac{1}{3}(X_1 + X_2 + X_3),$$

$$T_4 = \max(X_1, X_2, X_3), \qquad T_5 = X_1 + X_2 - 2\mu,$$

是

$$T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2).$$

不是

从统计量的定义可知，统计量是不含任何未知参数的随机变量.

➢几个常用的统计量

设$X_1, X_2, \ldots, X_n$是来自总体X的一个样本，$(x_1, x_2, \ldots, x_n)$是其观察值.

- 样本均值 $$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i;$$
- 样本方差 $$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)$$
- 样本标准差 $$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$
- 样本k阶(原点)矩 $$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad (k=1,2,\ldots)$$
- 样本k阶中心矩 $$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad (k=2,3,\ldots)$$

其观察值:

样本均值 $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;$$

样本方差 $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}\right)$$

样本标准差 $$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

样本k阶原点矩 $$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \quad (k = 1, 2, \ldots)$$

样本k阶中心矩 $$b_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k \quad (k = 2, 3, \ldots)$$

例1 从一批钢筋中随机抽取10条，测得其直径（单位：mm）为： 24.2, 25.4, 24, 24, 25, 25, 24.4, 24.6, 25.2, 25.2.

(1) 写出总体、样本、样本值、样本容量；

(2) 求样本观测值的均值、方差及二阶原点矩(保留二位).

解 (1) 总体为该批钢筋的直径； 样本为$X_1, X_2, \ldots, X_{10}$

样本值：24.2, 25.4, 24, 24, 25, 25, 24.4, 24.6, 25.2, 25.2.

样本容量：$n=10$；

(2) 样本均值 $$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(24.2+25.2+\ldots+25.2) = 24.68mm$$

样本方差 $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$= \frac{1}{9}\left[\begin{array}{l}(-0.48)^2+(0.72)^2+(-0.68)^2+(-0.68)^2+(0.32)^2 \\ +(0.32)^2+(-0.28)^2+(-0.08)^2+(-0.52)^2+(0.52)^2\end{array}\right] \approx 0.278$$

二阶原点矩 $$A_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{10}\left[24.8^2+25.4^2+\cdots+25.2^2\right] = 610.34.$$

样本矩的性质

若总体 X 的k阶矩 $E(X^k) \xlongequal{\text{记成}} \mu_k$存在,

则当$n \to \infty$时, $A_k \xrightarrow{P} \mu_k$, $k = 1, 2, \cdots$.

证明 因为 $X_1, X_2, \cdots, X_n$ 独立且与 X 同分布,

所以 $X_1^k, X_2^k, \cdots, X_n^k$ 独立且与 X^k 同分布,

故有 $E(X_1^k) = E(X_2^k) = \cdots = E(X_n^k) = \mu_k$.

再根据第五章辛钦定理知

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k, \quad k = 1, 2, \cdots;$$

由第五章关于依概率收敛的序列的性质知

$$\boxed{g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_k),}$$

其中g是连续函数.

以上结论是下一章所要介绍的矩估计法的理论根据.

抽样分布

统计量是样本的函数，它是一个随机变量. 统计量的分布称为抽样分布.

经验分布函数

总体分布函数 $F(x)$ 相应的统计量称为经验分布函数．

经验分布函数的做法如下：

设 $X_1, X_2, \cdots, X_n$ 是总体 F 的一个样本，用 $S(x)\ (-\infty < x < +\infty)$ 表示 $X_1, X_2, \cdots, X_n$ 中不大于 x 的随机变量的个数，

定义经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{1}{n} S(x), \quad -\infty < x < +\infty.$$

对于一个样本值，$F_n(x)$ 的观察值容易求得 .
($F_n(x)$ 的观察值仍以 $F_n(x)$ 表示.)

实例 设总体 F 具有一个样本值 1, 2, 3,

则经验分布函数
$F_3(x)$ 的观察值为

$$F_3(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{3}, & 1 \le x < 2, \\ \frac{2}{3}, & 2 \le x < 3 \\ 1, & x \ge 3. \end{cases}$$

实例 设总体 F 具有一个样本值 $1, 1, 2,$
则经验分布函数 $F_3(x)$ 的观察值为

$$F_3(x) = \begin{cases} 0, & x < 1, \\ \dfrac{2}{3}, & 1 \le x < 2 \\ 1, & x \ge 2. \end{cases}$$

一般地，

设 $x_1, x_2, \cdots, x_n$ 是总体 F 的一个容量为 n 的样本值，

先将 $x_1, x_2, \cdots, x_n$ 按自小到大的次序排列，

并重新编号， $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，

则经验分布函数 $F_n(x)$ 的观察值为

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \dfrac{k}{n}, & x_{(k)} \le x < x_{(k+1)}, \quad k = 1, 2, \cdots, n-1. \\ 1, & x \ge x_{(n)}. \end{cases}$$

格里汶科定理

对于任一实数 x, 当 $n \to \infty$ 时, $F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$, 即

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

对于任一实数 x 当 n 充分大时, 经验分布函数的任一个观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别, 从而在实际上可当作 $F(x)$ 来使用.

证明:利用强大数定律, 易知 $F_n(x)$ 以概1 (几乎处处)收敛于F(x); 利用实数的顺序性, 将R划分成可数个区间证明一致性; 略。

二、常用统计量的分布

(一) χ^2分布

定义： 设$X_1, X_2, \ldots, X_n$是来自总体 $N(0,1)$的样本, 则称统计量

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

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

这里自由度n表示相互独立的随机变量的个数.

注：1. $X_1, X_2, \ldots, X_n$独立同分布且 $X_i \sim N(0,1)$，则

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

2. $X \sim N(0,1)$, 则 $X^2 \sim \chi^2(1)$.

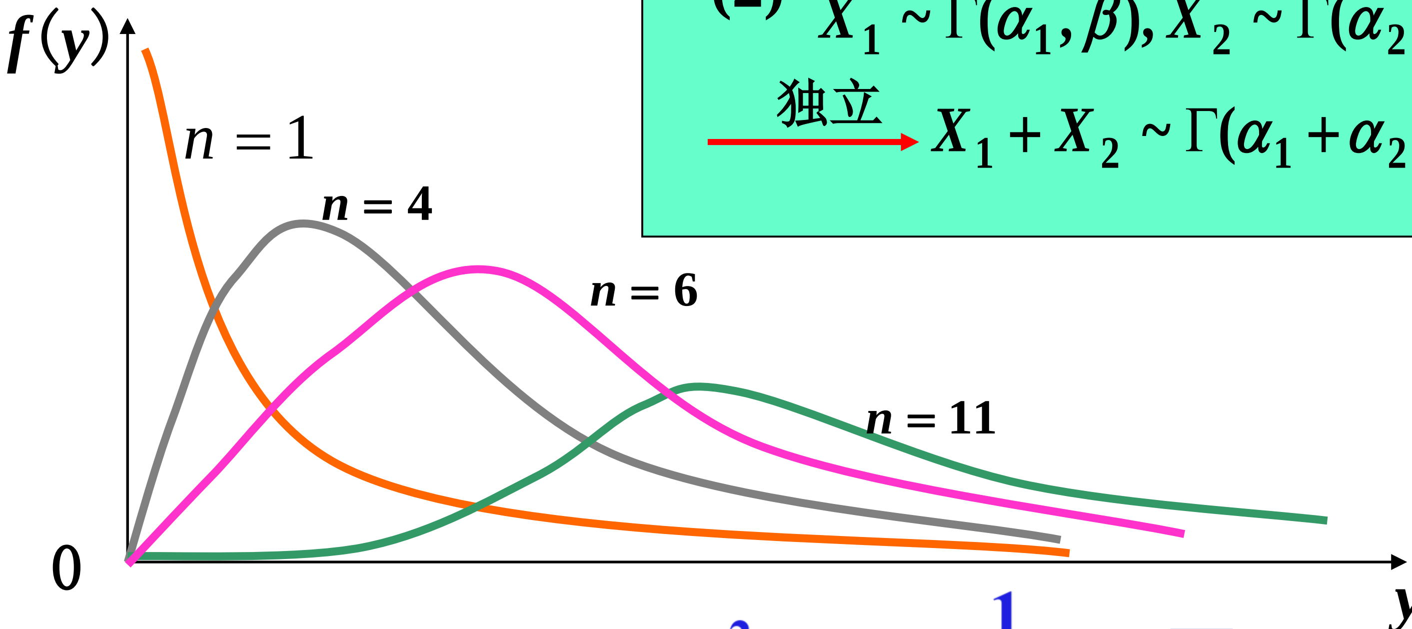
例： $X \sim N(0,2), Y \sim N(0,4)$, X和Y独立，则$\frac{1}{2}X^2 + \frac{1}{4}Y^2 \sim$ $\underline{\chi^2(2)}$.

(1) $X \sim N(0,1)$, $X^2 \sim \Gamma(1/2, 2)$;

(2) $X_1 \sim \Gamma(\alpha_1, \beta), X_2 \sim \Gamma(\alpha_2, \beta)$

$\xrightarrow{\text{独立}} X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \beta)$

☞ $f(y)$ 的图形(与n有关)：



☞ $f(y)$ 的推导： 由定义知， $\chi^2(1) = \Gamma(\frac{1}{2}, 2)$

而$X_i \sim N(0,1)$，由定义$X_i^2 \sim \chi^2(1)$， 即 $X_i^2 \sim \Gamma(\frac{1}{2}, 2)$，

再由$X_1, X_2, \ldots, X_n$的独立性及Γ分布的可加性

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

•Γ函数 $\Gamma(s)=\int_0^{+\infty} e^{-x}x^{s-1}dx \quad (s>0)$

•Γ分布 $f(x)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, & x>0,\\ 0, & \text{其它}.\end{cases}$ $X\sim \Gamma(\alpha,\beta)$

$\chi^2(n)$分布 $f(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}}, & y>0,\\ 0, & \text{其它}.\end{cases}$

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

•结论：若$X\sim N(0,1)$，则 $X^2\sim \Gamma(1/2,2)=\chi^2(1)$

$$f_{Y=X^2}(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}, & y>0,\\ 0, & y\le 0.\end{cases}$$

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

- 伽玛函数的性质

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

2) $\Gamma(1)=1$

3) 对于任何$\alpha>0$，有$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

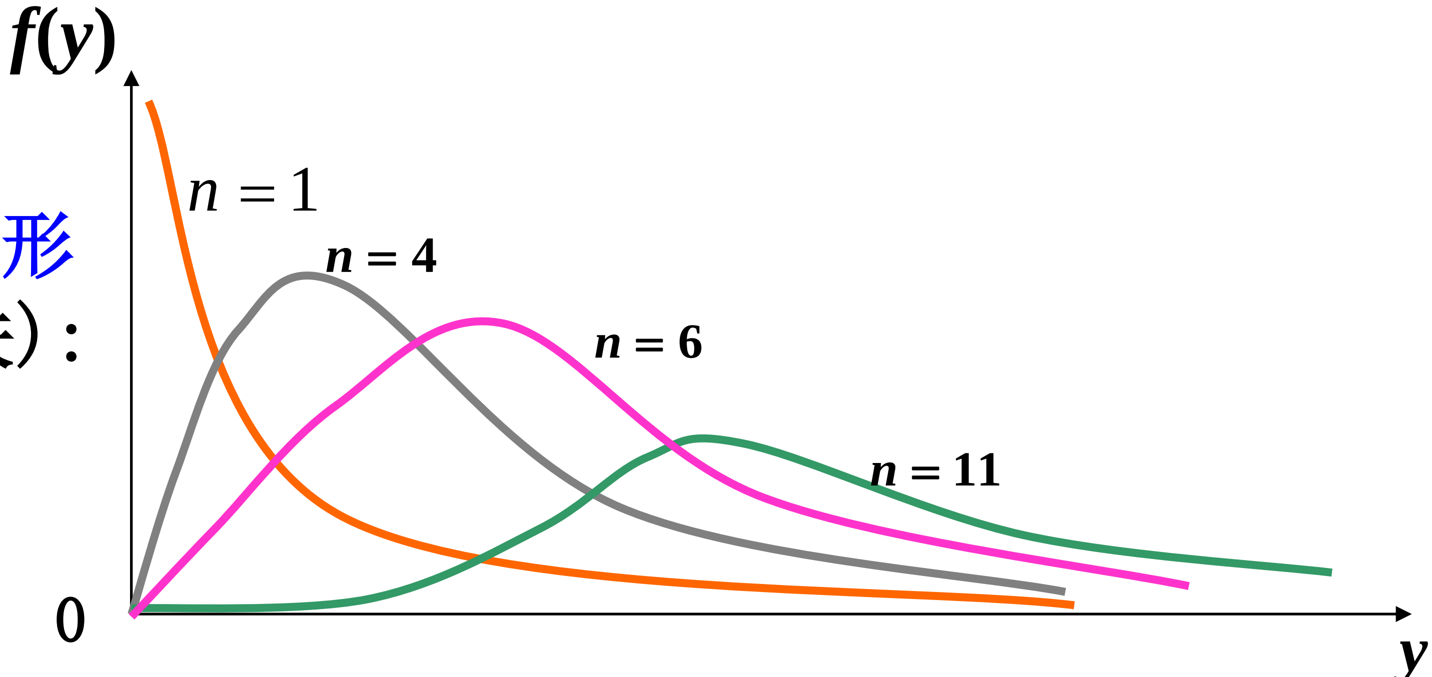
4) 对于任何正整数n，有$\Gamma(n)=(n-1)!$

$\chi^2(n)$ 分布的概率密度为

$$f(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}}, & y>0,\\ 0, & \text{其它}.\end{cases}$$

其中 $\Gamma(\frac{n}{2})=\int_0^{+\infty}e^{-x}x^{\frac{n}{2}-1}dx$.

☞ $f(y)$ 的图形 (与 n 有关):



χ^2 分布的性质

性质1 (χ^2 分布的可加性)

设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 并且 χ_1^2, χ_2^2 独立, 则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

（ 此性质可以推广到多个随机变量的情形. ）

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

例 设$X_1, X_2, \ldots, X_6$是来自总体$X\sim N(0,1)$, 又设

$$Y=(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2$$

试求常数C, 使CY服从χ^2分布.

解 因为$X_1+X_2+X_3\sim N(0,3)$, $X_4+X_5+X_6\sim N(0,3)$,

所以 $\dfrac{X_1+X_2+X_3}{\sqrt{3}}\sim N(0,1)$ $\dfrac{X_4+X_5+X_6}{\sqrt{3}}\sim N(0,1)$

且它们相互独立.于是

$$\left(\frac{X_1+X_2+X_3}{\sqrt{3}}\right)^2+\left(\frac{X_4+X_5+X_6}{\sqrt{3}}\right)^2\sim\chi^2(2)$$

故应取常数 $C=\dfrac{1}{3}$, 于是 $\dfrac{1}{3}Y\sim\chi^2(2)$.

性质2 $(\chi^2$分布的数学期望和方差)

$$\text{若 } \chi^2 \sim \chi^2(n), \text{ 则 } E(\chi^2) = n,\ D(\chi^2) = 2n.$$

证明 因为 $X_i \sim N(0,1)$, 所以 $E(X_i^2) = D(X_i) = 1$,

$$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1 = 2,\ i = 1, 2, \cdots, n.$$

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

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

例: $X \sim \chi^2(5), Y \sim U(0,4)$, X与Y独立,

则 $E(X-Y) =$ __3__, $D(X-Y) =$ __$11\frac{1}{3}$__.

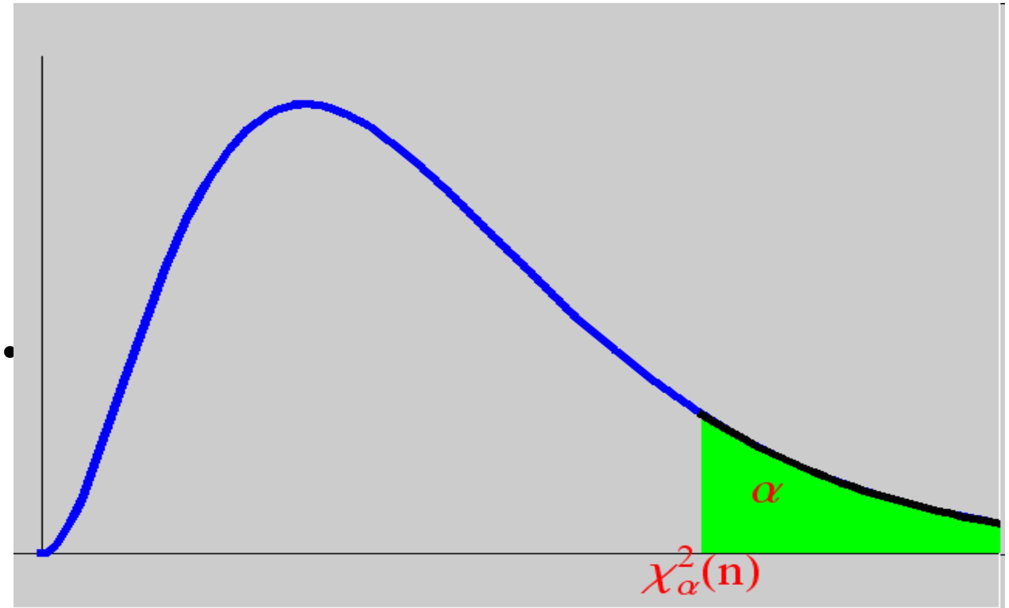
χ^2 分布的分位点

对于给定的正数 α，$0<\alpha<1$，称满足条件

$$P\{\chi^2>\chi_\alpha^2(n)\}=\int_{\chi_\alpha^2(n)}^{\infty} f(y)\mathrm{d}y=\alpha$$

的点 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点.

对于不同的 α，n，可以通过查表求得上 α 分位点的值.



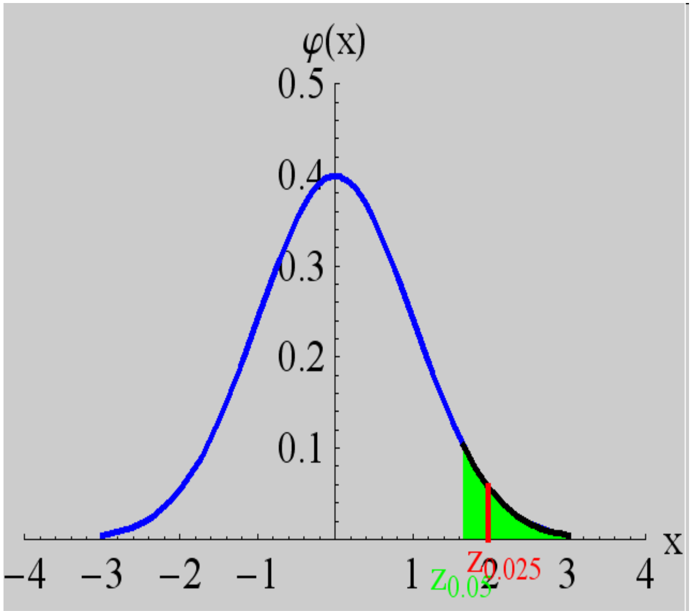
例1 设 X 服从标准正态分布 $N(0,1)$, $N(0,1)$ 的上 α 分位点 z_α 满足 $P\{X > z_\alpha\} = \dfrac{1}{\sqrt{2\pi}}\int_{z_\alpha}^{+\infty} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = \alpha,$ 求 z_α 的值, 可通过查表完成.

$z_{0.05} = 1.645,$

$z_{0.025} = 1.96,$

根据正态分布的对称性知

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



例2 设 $Z \sim \chi^2(n)$, $\chi^2(n)$ 的上 α 分位点满足

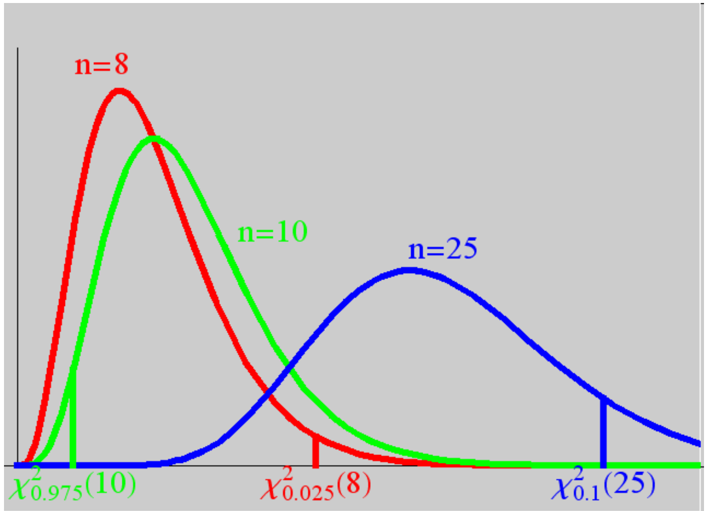
$$P\{Z > \chi^2_\alpha(n)\} = \int_{\chi^2_\alpha(n)}^{+\infty} \chi^2(y;n)\mathrm{d}y = \alpha,$$

求 $\chi^2_\alpha(n)$ 的值，可通过查表完成 .

$\chi^2_{0.025}(8) = 17.535,$

$\chi^2_{0.975}(10) = 3.247,$

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



课本附表5只详列到 n=40为止.

在Matlab中求解

费希尔(R.A.Fisher)证明:

当 n 充分大时, $\chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2.$

其中 z_α 是标准正态分布的上 α 分位点.

利用上面公式,

可以求得 $n > 40$ 时, 上 α 分位点的近似值.

例如 $\chi^2_{0.05}(50) \approx \frac{1}{2}(1.645 + \sqrt{99})^2 = 67.221.$

而查详表可得 $\chi^2_{0.05}(50) = 67.505.$

(二) t 分布

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

称随机变量 $t = \dfrac{X}{\sqrt{Y/n}}$ 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.

$t(n)$ 分布的概率密度函数为

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

其推导由威廉·戈塞（William Sealy Gosset）于1908年首先发表，当时他还在都柏林的健力士酿酒厂工作。因为不能以他本人的名义发表，所以论文使用了学生（Student）这一笔名。之后t检验以及相关理论经由罗纳德·费雪的工作发扬光大，而正是他将此分布称为学生分布。

t 分布的概率密度曲线如图

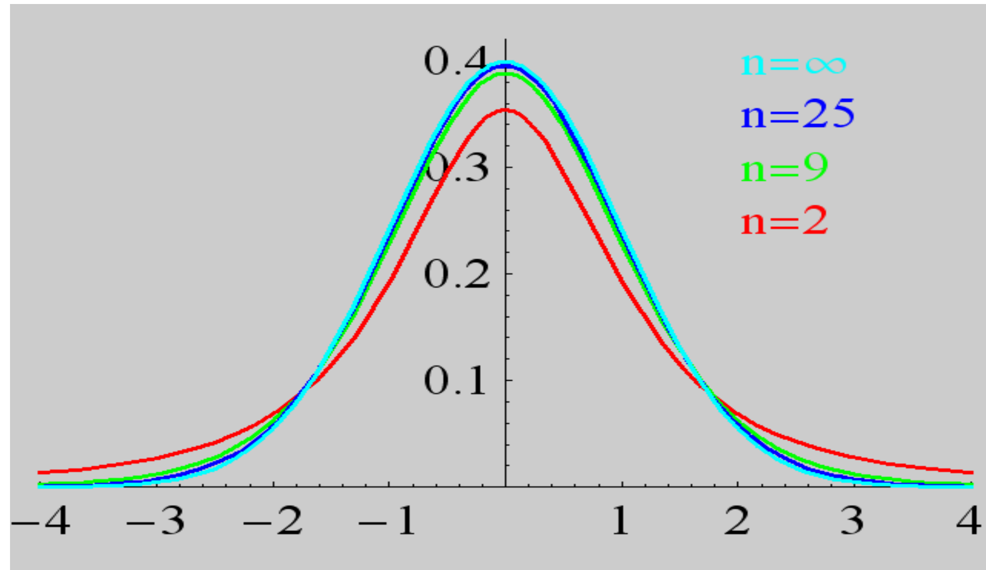
显然图形是关于

$t=0$对称的.

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. 因为$\lim\limits_{n\to\infty} h(t)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$,（考虑 Y/n 的极限？）

所以当n 足够大时 t 分布近似于 $N(0,1)$ 分布,

但对于较小的n, t分布与$N(0,1)$分布相差很大.

t 分布的分位点

对于给定的 α，$0<\alpha<1$，称满足条件

$$P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} h(t)\mathrm{d}t = \alpha$$

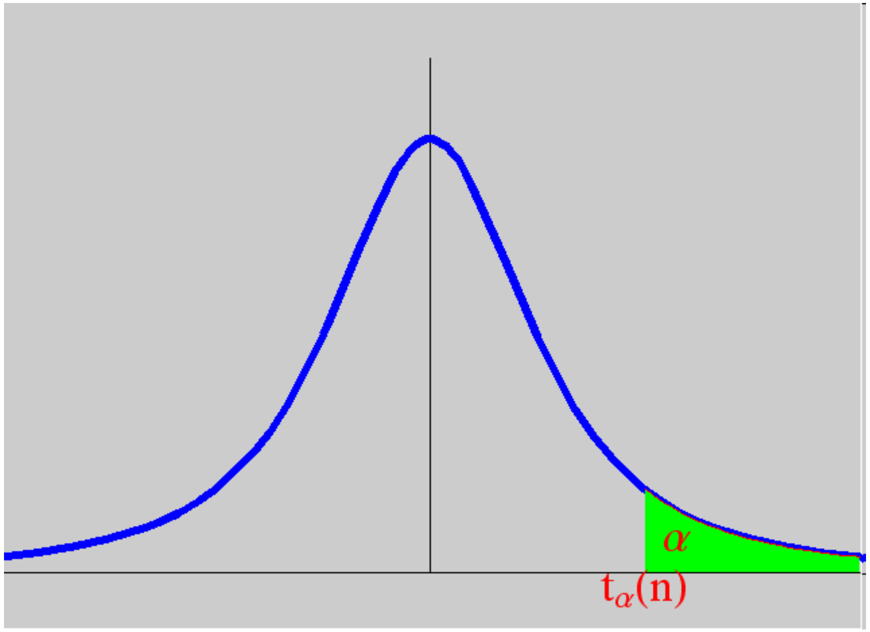
的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点.

可以通过查表求得上α分位点的值.

由分布的对称性知

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

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



例3 设 $T \sim t(n)$, $t(n)$ 的上 α 分位点满足

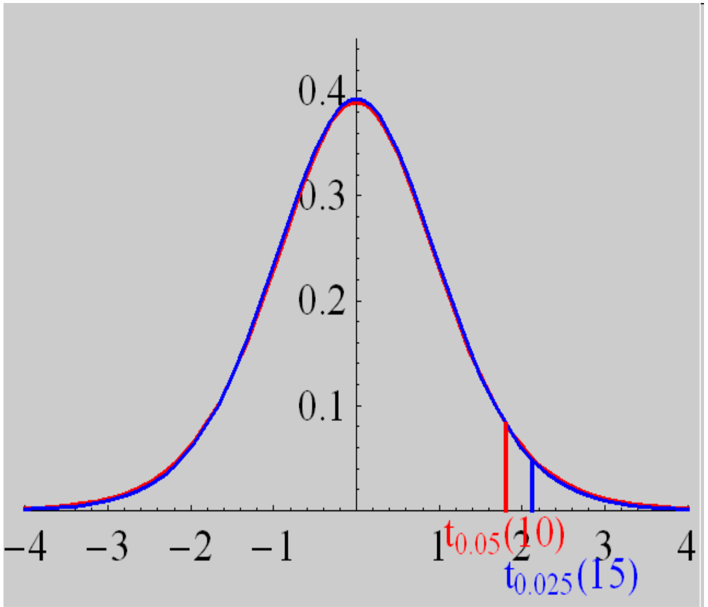
$$P\{T > t_\alpha(n)\} = \int_{t_\alpha(n)}^{+\infty} t(y;n)\mathrm{d}y = \alpha,$$

求 $t_\alpha(n)$ 的值, 可通过查表完成.

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

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

在Matlab中求解



例 设$X\sim N(2,1)$, $Y_1, Y_2, \ldots, Y_4$ 均服从$N(0,4)$，且都相互独立，令

$$T=\frac{4(X-2)}{\sqrt{\sum_{i=1}^{4} Y_i^2}}$$

试求T的分布, 并确定 t_0 的值，使 $P\{|T|>t_0\}=0.01.$

解 因为 $X-2\sim N(0,1)$，$Y_i/2\sim N(0,1)$，$i=1,2,3,4.$

故 $$T=\frac{4(X-2)}{\sqrt{\sum_{i=1}^{4} Y_i^2}}=\frac{X-2}{\sqrt{\sum_{i=1}^{4}\left(\frac{Y_i}{2}\right)^2 \Big/ 4}}\sim t(4)$$

由 $P\{|T|>t_0\}=0.01.$ 查表得:

$$t_0=t_{\alpha/2}(4)=t_{0.005}(4)=4.6041$$

（三）F分布

设$U\sim\chi^2(n_1)$, $V\sim\chi^2(n_2)$,且U 与V相互独立, 则称

$$F=\frac{U/n_1}{V/n_2}$$

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

➢ $F\sim F(n_1,n_2)$分布的概率密度函数为:

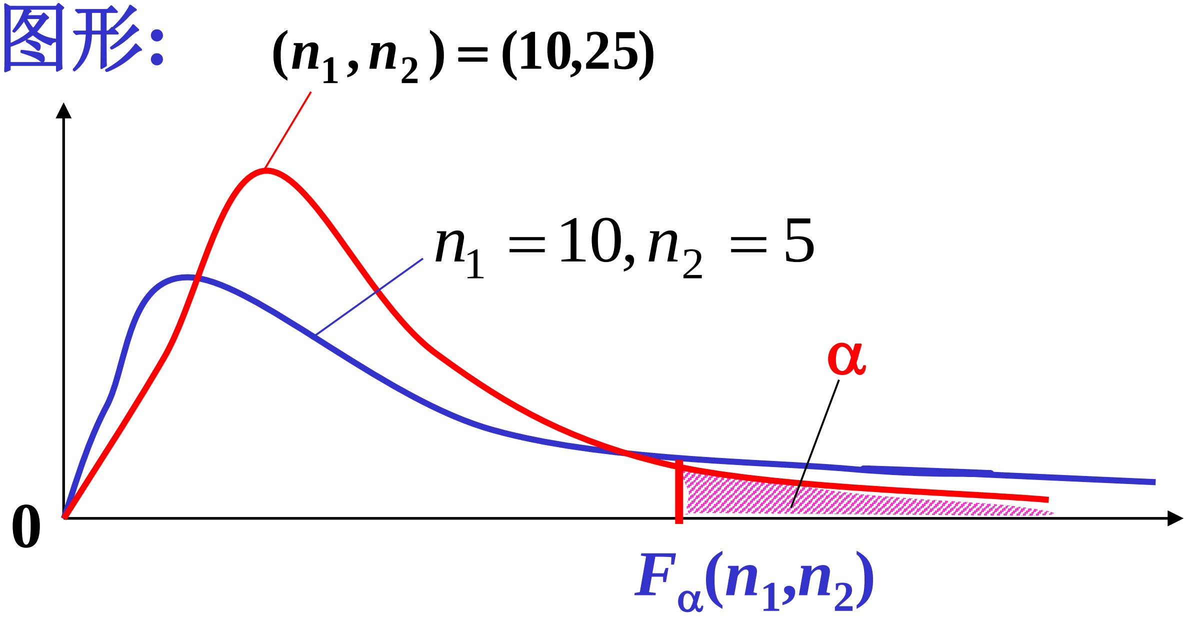
$$\psi(y)=\begin{cases}\dfrac{\boxed{\Gamma[(n_1+n_2)/2]}(n_1/n_2)^{n_1/2}y^{(n_1/2)-1}}{\boxed{\Gamma(n_1/2)\Gamma(n_2/2)}[1+(n_1y/n_2)]^{(n_1+n_2)/2}}, & y>0,\\ 0, & y<0.\end{cases}$$

$$\frac{1}{B(n_1/2,n_2/2)}$$

其中$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

➢ 根据定义可知, 若$F\sim F(n_1,n_2)$, 则 $1/F\sim F(n_2,n_1)$.

➤ψ(y)的图形:



➤F分布的分位点: 对给定α (0<α<1),称满足

$$P\{F > F_\alpha(n_1,n_2)\} = \int_{F_\alpha(n_1,n_2)}^{\infty} \psi(y)dy = \alpha$$

的点$F_\alpha(n_1,n_2)$ 为$F(n_1,n_2)$分布的上α分位点.

例4 设 $F(n_1, n_2)$分布的上α分位点满足

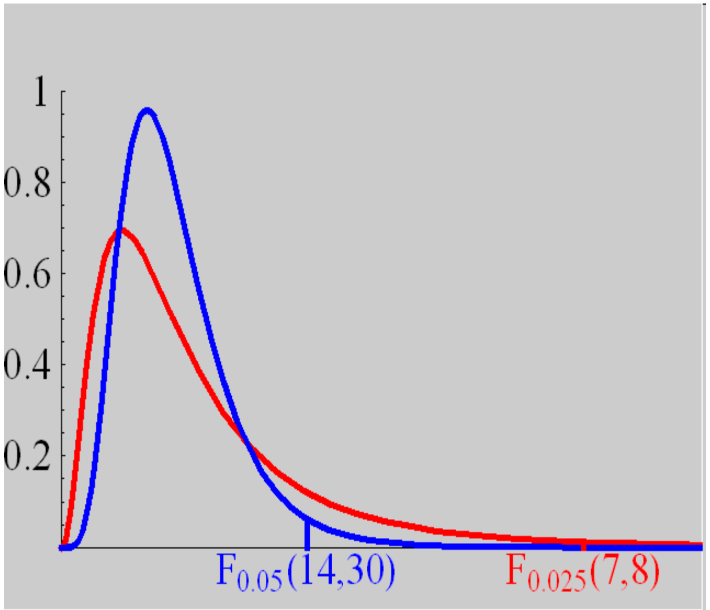
$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1,n_2)}^{+\infty} \psi(y)\mathrm{d}y = \alpha,$$

求 $F_\alpha(n_1, n_2)$ 的值,可通过查表完成.

$F_{0.025}(8,7) = 4.90,$

$F_{0.05}(30,14) = 2.31.$

在Matlab中求解



F 分布的上 α 分位点具有如下性质:

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

证明 因为$F \sim F(n_1, n_2)$,

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

$$= P\left\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} = 1 - P\left\{\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\}$$

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

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

因为 $\dfrac{1}{F} \sim F(n_2, n_1)$, 所以 $P\left\{\dfrac{1}{F} > F_\alpha(n_2, n_1)\right\} = \alpha$,

比较后得 $\dfrac{1}{F_{1-\alpha}(n_1, n_2)} = F_\alpha(n_2, n_1)$,

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

用来求分布表中未列出的一些上 α 分位点.

例 $F_{0.95}(12,9) = \dfrac{1}{F_{0.05}(9, 12)} = \dfrac{1}{2.80} = 0.357$.

（四）正态总体的样本均值与样本方差的分布

设总体X的均值为 μ, 方差为 σ^2, $X_1, X_2, \dots, X_n$是来自总体X 的样本, 则总有

$$\boxed{E(\overline{X}) = \mu, \quad D(\overline{X}) = \frac{\sigma^2}{n}, \quad E(S^2) = \sigma^2.}$$

推导:

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

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

$$E(S^2) = E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] = \sigma^2$$

抽样分布定理（定理1、定理2、定理3 ）

设$X_1, X_2, \dots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，

$\overline{X}, S^2$分别是样本均值和样本方差，则有

定理1： $\boxed{\overline{X} \sim N(\mu, \frac{\sigma^2}{n})}$ $\Longrightarrow$ $\boxed{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}$

定理2：(1) $\boxed{\sum_{i=1}^{n} \dfrac{(X_i - \overline{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1);}$

(2) $\boxed{\overline{X}\text{与}S^2\text{独立}.}$

定理3： $\boxed{\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).}$ $\longleftarrow$ $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}$

定理4 设$X_1, X_2, \ldots, X_{n1}$与$Y_1, Y_2, \ldots, Y_{n2}$分别是来自正态总体$N(\mu_1, \sigma_1^2)$和$N(\mu_2, \sigma_2^2)$的样本, 且这两个样本相互独立. 两个样本的均值和方差分别为$\overline{X}, \overline{Y}, S_1^2, S_2^2$, 则有

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

2° 当$\sigma_1^2 = \sigma_2^2 = \sigma^2$时,

$$\boxed{\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_\varpi \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2),}$$

其中 $$S_\varpi^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad S_\varpi = \sqrt{S_\varpi^2}$$

证明 1° 由定理2 知

$$\frac{(n_1-1)S_1^{\,2}}{\sigma_1^{\,2}} \sim \chi^2(n_1-1), \quad \frac{(n_2-1)S_2^{\,2}}{\sigma_2^{\,2}} \sim \chi^2(n_2-1);$$

两者相互独立，由F 分布定义可知

$$\frac{(n_1-1)S_1^{\,2}}{\sigma_1^{\,2}(n_1-1)} \Big/ \frac{(n_2-1)S_2^{\,2}}{\sigma_2^{\,2}(n_2-1)} \sim F(n_1-1, n_2-1)$$ 化简后即得1° 。

2° 由$\chi^2(n)$的可加性：$\frac{(n_1-1)S_1^{\,2}}{\sigma_1^{\,2}} + \frac{(n_2-1)S_2^{\,2}}{\sigma_2^{\,2}} \sim \chi^2(n_1+n_2-2),$

$$\overline{X}-\overline{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right), \Longrightarrow U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1),$$

$$\xrightarrow{\text{t分布定义}} \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Bigg/ \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2(n_1+n_2-2)}} \sim t(n_1+n_2-2)$$

U 与V 相互独立，
为什么？

化简后即得2°

以上列举的几个重要统计量的分布是数理统计中常用的，它们的密度函数形式都较复杂，对于应用者来说，不要求一一推导，但是查表求上α分位点是统计中经常遇到的，必须熟练掌握.

本节中的四个定理是统计推断的理论依据，要逐步熟悉定理的条件与结论.

例1 设总体$X\sim N(0,1)$, $X_1, X_2, \ldots, X_n$ 为X的样本,

则 $\left(\frac{X_1 - X_2}{X_3 + X_4}\right)^2$ 服从_ $F(1,1)$ _分布。

例2 已知$X\sim t(n)$, 求证 $X^2 \sim F(1, n)$

例3 设总体$X\sim N(\mu, \sigma^2)$, $\overline{X}$, S^2分别是容量为n的样本均值与样本方差,则

$\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2$ 服从 $\chi^2(n-1)$ 分布; $D(S^2) = \frac{2\sigma^4}{n-1}$

小结

常用的统计量

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i;\qquad A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\quad(k=1,2,\dots,n)$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\qquad B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k\quad(k=1,2,\dots,n)$$

三个来自正态分布的抽样分布:

χ^2 分布, t 分布, F 分布.

关于$\overline{X}, S^2$的结果：

$$E(\overline{X}) = \mu, \quad D(\overline{X}) = \frac{\sigma^2}{n}, E(S^2) = \sigma^2$$

来自正态总体$X \sim N(\mu, \sigma^2)$:

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

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

$$\boxed{\overline{X}\text{与 }S^2\text{独 立.}}$$

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

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

$$\boxed{\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_\varpi \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),}$$

数理统计的内容大致分两大类：

(1)数据的收集，包括抽样技术及试验设计的理论和方法的研究，即研究如何队随机现象进行科学的观察和试验，是获得的数据资料及真实又有代表性.

(2)统计推断，即研究如何对已取得的观察之进行整理、分析并做出决策的方法—推断总体的规律性.（我们只讨论统计推断问题）

作业

- **第六章习题2, 3, 6, 7, 8, 9**