

第一节 随机样本

一、总体与个体

二、随机样本的定义

三、小结

一、总体与个体

1. 总体

试验的全部可能的观察值称为总体。

2. 个体

总体中的每个可能观察值称为个体。

实例1 在研究2 000名学生的年龄时, 这些学生的年龄的全体就构成一个总体, 每个学生的年龄就是个体。



3.容量

总体中所包含的个体的个数称为总体的容量.

4.有限总体和无限总体

容量为有限的称为有限总体.

容量为无限的称为无限总体.

实例2 某工厂10月份生产的灯泡寿命所组成的总体中,个体的总数就是10月份生产的灯泡数,这是个有限总体;而该工厂生产的所有灯泡寿命所组成的总体是一个无限总体,它包括以往生产和今后生产的灯泡寿命.

实例3 在考察某大学一年级男生的身高这一试验中,若一年级男生共2 000人, 每个男生的身高是一个可能观察值,所形成的总体中共含2 000个可能观察值,是一个有限总体.

实例4 考察某一湖泊中某种鱼的含汞量,所得总体也是有限总体.



有些有限总体，它的容量很大，我们可以认为它是一个无限总体.



实例5 考察全国正在使用的某种型号灯泡的寿命所形成的总体，由于可能观察值的个数很多，就可以认为是无限总体.



5. 总体分布

实例6 在2 000名大学一年级学生的年龄中, 年龄指标值为“15”, “16”, “17”, “18”, “19”, “20”的依次有 9, 21, 132, 1207, 588, 43 名, 在总体中所占比率依次为

$$\frac{9}{2000}, \frac{21}{2000}, \frac{132}{2000}, \frac{1207}{2000}, \frac{588}{2000}, \frac{43}{2000},$$

即学生年龄的取值有一定的分布.

补充例题



一般地, 我们所研究的总体, 即研究对象的某项数量指标 X , 其取值在客观上有一定的分布, X 是一个随机变量.

总体分布的定义

我们把数量指标取不同数值的比率叫做总体分布.

如实例6中, 总体就是数集 $\{15, 16, 17, 18, 19, 20\}$.

总体分布为

年龄	15	16	17	18	19	20
比率	$\frac{9}{2000}$	$\frac{21}{2000}$	$\frac{132}{2000}$	$\frac{1207}{2000}$	$\frac{588}{2000}$	$\frac{43}{2000}$

X 的分布函数和数字特征就称为总体的分布函数和数字特征. 今后将不区分总体与相应的随机变量.

例如, 我们检验自生产线出来的零件是次品还是正品, 以0表示产品是正品, 以1表示产品为次品. 设出现次品的频率为 p (常数), 那么总体是由一些 “0” 和一些 “1” 所组成. 这一总体对应于一个具有参数为 p 的 (0-1) 分布:

$$P\{X = x\} = p^x (1 - p)^{1-x}, \quad x = 0, 1$$

的随机变量.



在数理统计中，人们都是通过从总体中抽取一部分个体，根据获得的数据来对总体分布得出判断的。被抽出的部分个体叫做总体的一个样本。

所谓从总体抽取一个个体，就是对总体 X 进行一次观察并记录其结果。

当 n 次观察一经完成，我们就得到一组实数 $x_1, x_2, \dots, x_n$ ，它们依次是随机变量 $X_1, X_2, \dots, X_n$ 的观察值，称为样本值。



二、随机样本的定义

1. 样本的定义

设 X 是具有分布函数 F 的随机变量, 若 $X_1, X_2, \dots, X_n$ 是具有同一分布函数 F 、相互独立的随机变量, 则称 $X_1, X_2, \dots, X_n$ 为从分布函数 F (或总体 F 、或总体 X) 得到的容量为 n 的简单随机样本, 简称样本.

它们的观察值 $x_1, x_2, \dots, x_n$ 称为样本值, 又称为 X 的 n 个独立的观察值.



2. 简单随机抽样的定义

获得简单随机样本的抽样方法称为简单随机抽样.

若 $X_1, X_2, \dots, X_n$ 为 F 的一个样本则 $X_1, X_2, \dots, X_n$ 相互独立,且它们的分布函数都是 F ,所以 $(X_1, X_2, \dots, X_n)$ 的分布函数为

$$F^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i).$$

又若 X 具有概率密度 f , 则 $(X_1, X_2, \dots, X_n)$ 的概率密度为

$$f^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i).$$



三、小结

基本概念: 个体 总体 $\left\{ \begin{array}{l} \text{有限总体} \\ \text{无限总体} \end{array} \right.$ 随机样本

说明1 一个总体对应一个随机变量 X , 以后将不区分总体和相应的随机变量, 统称为总体 X .

说明2 在实际中遇到的总体往往是有限总体, 它对应一个离散型随机变量; 当总体中包含的个体的个数很大时, 在理论上可认为它是一个无限总体.

第三节 抽样分布

一、基本概念

二、常见分布

三、小结

一、基本概念

1. 统计量的定义

设 $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本, $g(X_1, X_2, \dots, X_n)$ 是 $X_1, X_2, \dots, X_n$ 的函数, 若 g 中不含未知参数, 则称 $g(X_1, X_2, \dots, X_n)$ 是一个统计量.

设 $x_1, x_2, \dots, x_n$ 是相应于样本 $X_1, X_2, \dots, X_n$ 的样本值, 则称 $g(x_1, x_2, \dots, x_n)$ 是 $g(X_1, X_2, \dots, X_n)$ 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?

$$T_1 = X_1,$$

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

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

是

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

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

不是



2. 几个常用统计量的定义

设 $X_1, X_2, \dots, X_n$ 是来自总体的一个样本，
 $x_1, x_2, \dots, x_n$ 是这一样本的观察值。

(1) 样本平均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i;$

其观察值 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$

(2) 样本方差

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



其观察值

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

(3) 样本标准差

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

其观察值

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



(4) 样本 k 阶(原点)矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, k = 1, 2, \dots;$

其观察值 $\alpha_k = \frac{1}{n} \sum_{i=1}^n x_i^k, k = 1, 2, \dots.$

(5) 样本 k 阶中心矩

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, k = 2, 3, \dots;$$

其观察值 $b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, k = 2, 3, \dots.$



由以上定义得下述结论:

若总体 X 的 k 阶矩 $E(X^k)$ 记成 μ_k 存在,
则当 $n \rightarrow \infty$ 时, $A_k \xrightarrow{P} \mu_k, k = 1, 2, \dots$.

证明 因为 $X_1, X_2, \dots, X_n$ 独立且与 X 同分布,
所以 $X_1^k, X_2^k, \dots, X_n^k$ 独立且与 X^k 同分布,
故有 $E(X_1^k) = E(X_2^k) = \dots = E(X_n^k) = \mu_k$.

再根据第五章辛钦定理知

辛钦定理

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



由第五章关于依概率收敛的序列的性质知

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

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.



3. 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 $X_1, X_2, \dots, X_n$ 是总体 F 的一个样本, 用 $S(x) (-\infty < x < +\infty)$ 表示 $X_1, X_2, \dots, X_n$ 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

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



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

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



一般地,

设 $x_1, x_2, \dots, x_n$ 是总体 F 的一个容量为 n 的样本值,
先将 $x_1, x_2, \dots, x_n$ 按自小到大的次序排列,
并重新编号, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$,
则经验分布函数 $F_n(x)$ 的观察值为

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, \quad k = 1, 2, \dots, n-1. \\ 1, & x \geq x_{(n)}. \end{cases}$$



格里汶科定理

格里汶科资料

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

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

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.

二、常见分布

统计量的分布称为抽样分布.

1. χ^2 分布

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

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

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

自由度是指上式右端包含的独立变量的个数.



$\chi^2(n)$ 分布的概率密度为

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

证明 因为 $\chi^2(1)$ 分布即为 $\Gamma\left(\frac{1}{2}, 2\right)$ 分布,

又因为 $X_i \sim N(0, 1)$, 由定义 $X_i^2 \sim \chi^2(1)$,

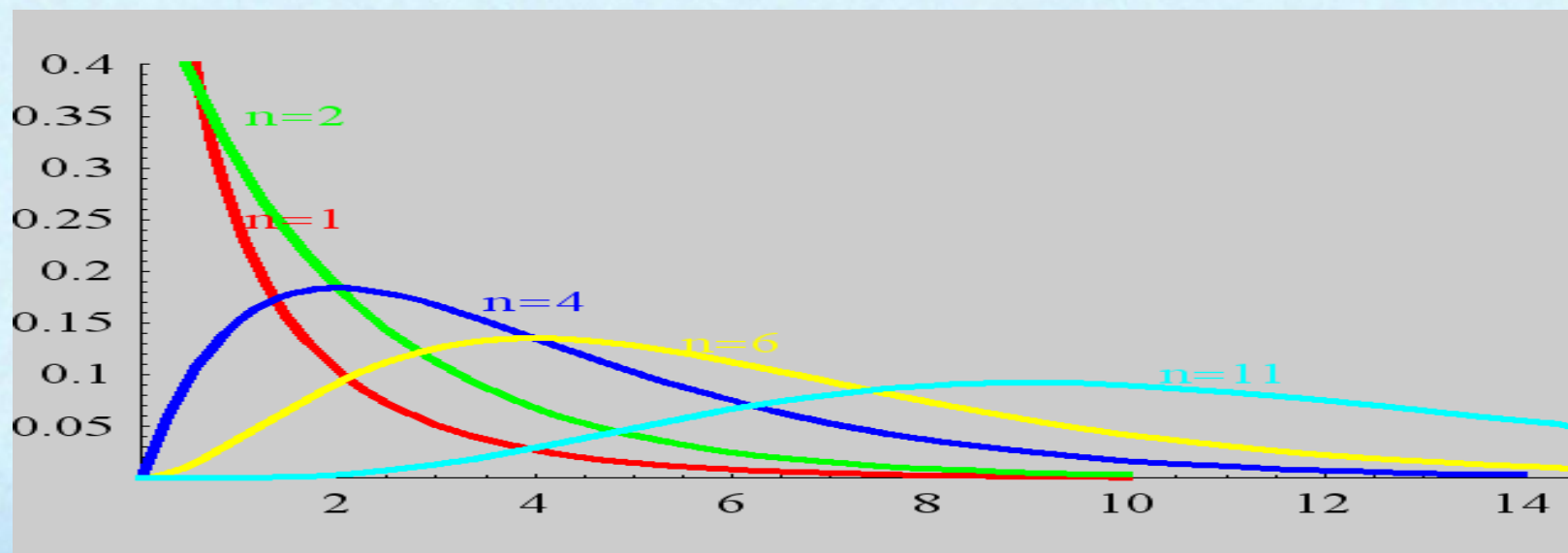
即 $X_i^2 \sim \Gamma\left(\frac{1}{2}, 2\right)$, $i = 1, 2, \dots, n$.



因为 $X_1, X_2, \dots, X_n$ 相互独立,
所以 $X_1^2, X_2^2, \dots, X_n^2$ 也相互独立,

根据 Γ 分布的可加性知 $\chi^2 = \sum_{i=1}^n X_i^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$.

$\chi^2(n)$ 分布的概率密度曲线如图.



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

性质2 (χ^2 分布的数学期望和方差)

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

证明 因为 $X_i \sim N(0, 1)$, 所以 $E(X_i^2) = D(X_i) = 1$,
 $D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 2 = 1$, $i = 1, 2, \dots, n$.

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

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



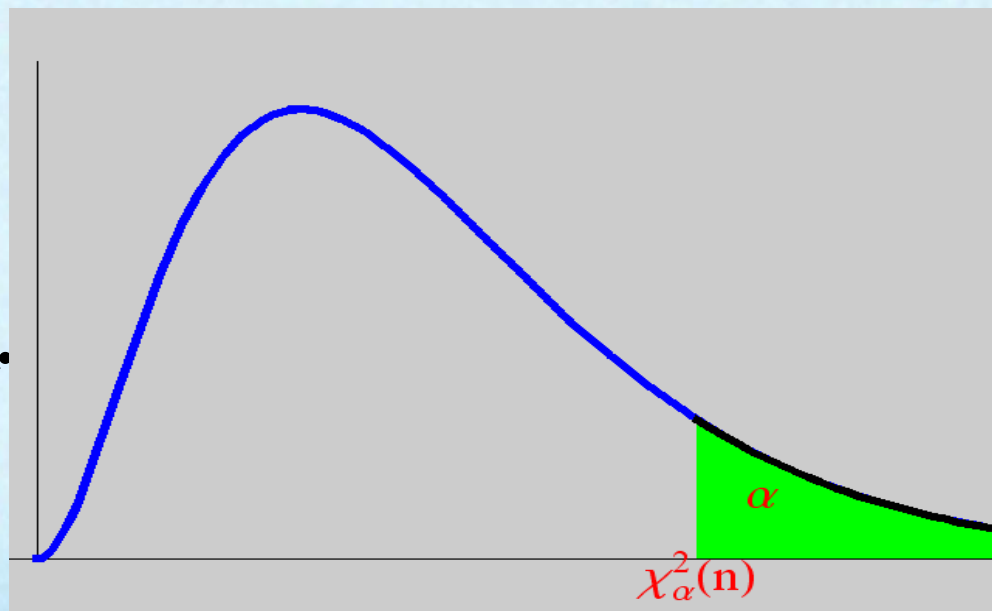
χ^2 分布的分位点

对于给定的正数 α , $0 < \alpha < 1$, 称满足条件

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

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 $P\{X > z_\alpha\} = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx = \alpha$, 求 z_α 的值, 可通过查表完成.

$$z_{0.05} = 1.645,$$

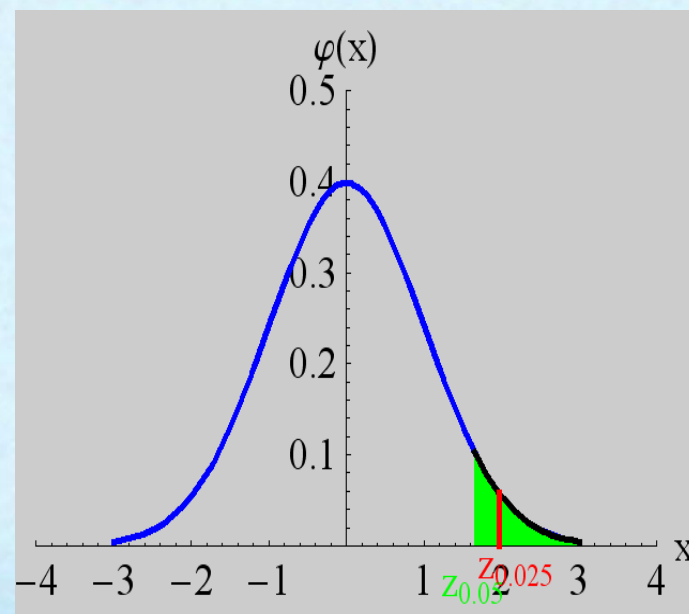
附表1-1

$$z_{0.025} = 1.96,$$

附表1-2

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

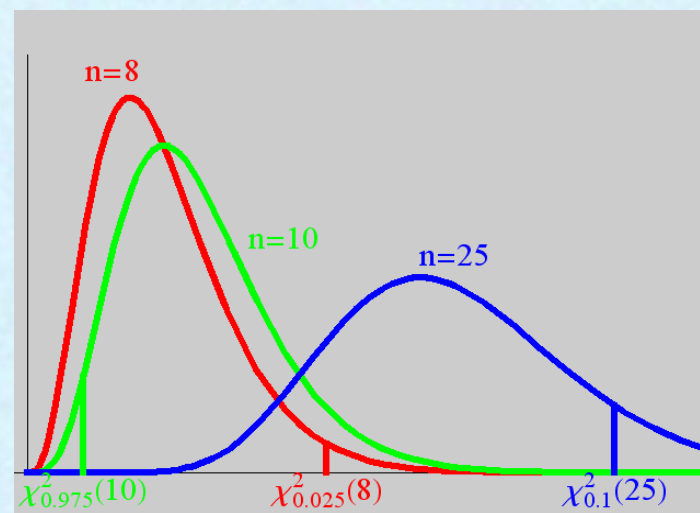
$$P\{Z > \chi_{\alpha}^2(n)\} = \int_{\chi_{\alpha}^2(n)}^{+\infty} \chi^2(y; n) dy = \alpha,$$

求 $\chi_{\alpha}^2(n)$ 的值, 可通过查表完成.

$$\chi_{0.025}^2(8) = 17.535, \quad \text{附表2-1}$$

$$\chi_{0.975}^2(10) = 3.247, \quad \text{附表2-2}$$

$$\chi_{0.1}^2(25) = 34.382. \quad \text{附表2-3}$$



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:

费希尔资料

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

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.

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

$$\text{而查详表可得 } \chi_{0.05}^2(50) = 67.505.$$

2. 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 $t = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$.

学生氏资料

t 分布又称学生氏(Student)分布.

随机数演示

$t(n)$ 分布的概率密度函数为

分布函数与密度函数演示

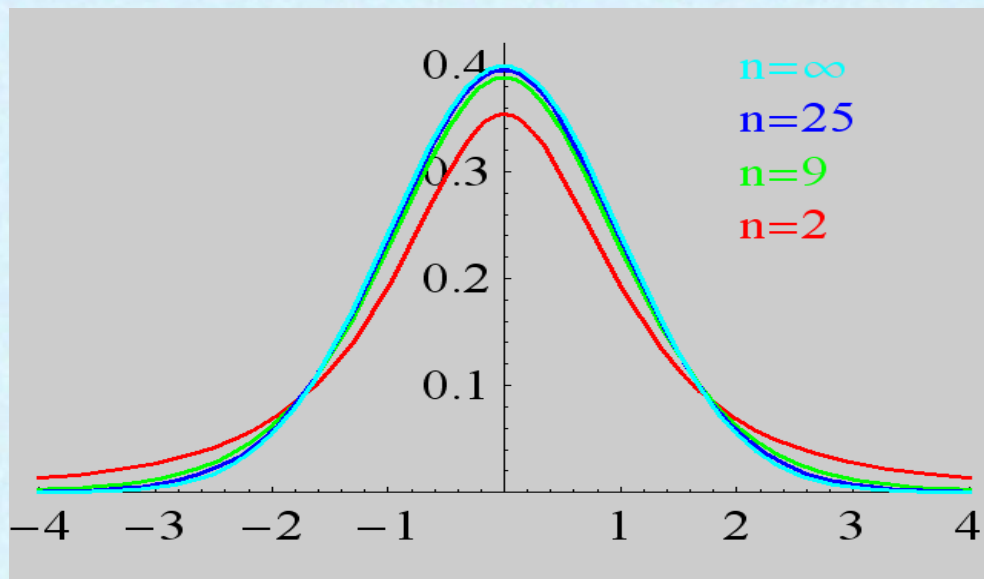
$$h(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty$$

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.



形. 因为 $\lim_{n \rightarrow \infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$,

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

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

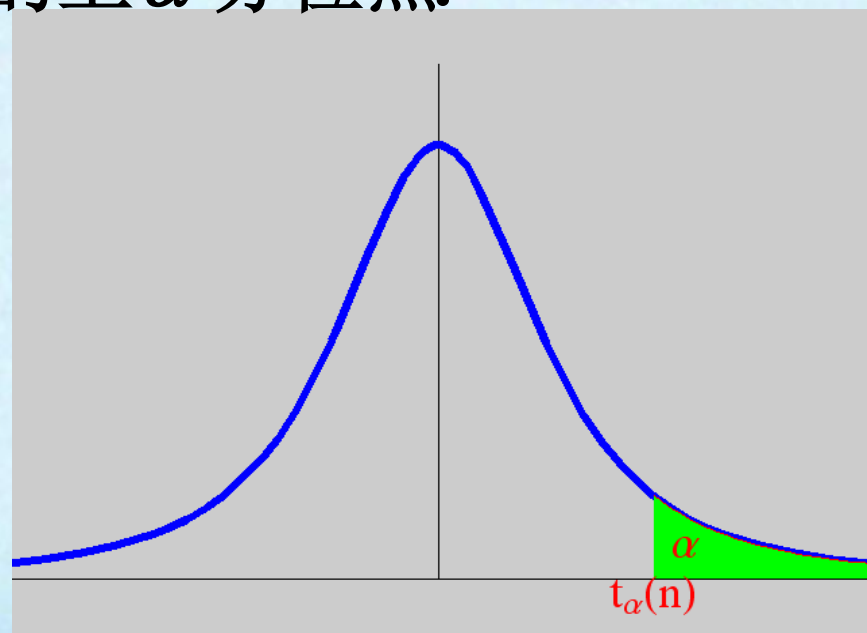
的点 $t_{\alpha}(n)$ 为 $t(n)$ 分布的上 α 分位点.

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

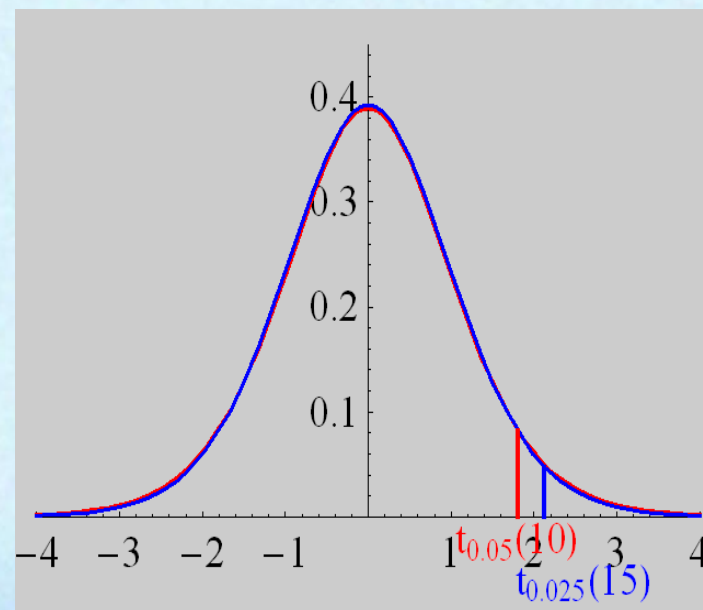
$$P\{T > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{+\infty} t(y; n) dy = \alpha,$$

求 $t_{\alpha}(n)$ 的值, 可通过查表完成.

$$t_{0.05}(10) = 1.8125, \quad \text{附表3-1}$$

$$t_{0.025}(15) = 2.1315. \quad \text{附表3-2}$$

在Matlab中求解



3. F 分布

设 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 且 U, V 独立, 则称随机变量 $F = \frac{U/n_1}{V/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布, 记为 $F \sim F(n_1, n_2)$.

随机数演示

分布函数与密度函数演示



$F(n_1, n_2)$ 分布的概率密度为

$$\psi(y) = \begin{cases} \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right) \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} y^{\frac{n_1}{2} - 1}}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right) \left[1 + \left(\frac{n_1 y}{n_2}\right)\right]^{\frac{n_1 + n_2}{2}}}, & y > 0, \\ 0, & \text{其他.} \end{cases}$$

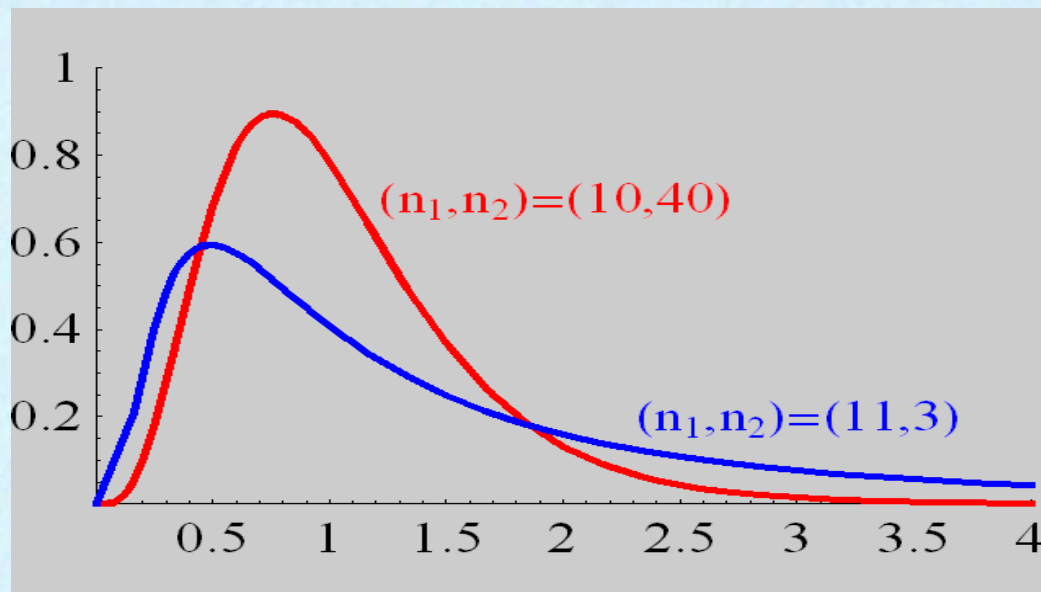
F 分布的概率密度曲线如图

根据定义可知,

若 $F \sim F(n_1, n_2)$,

则 $\frac{1}{F} \sim F(n_2, n_1)$.

F 分布的分位点



对于给定的 α , $0 < \alpha < 1$, 称满足条件

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

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

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

求 $F_\alpha(n_1, n_2)$ 的值, 可通过查表完成.

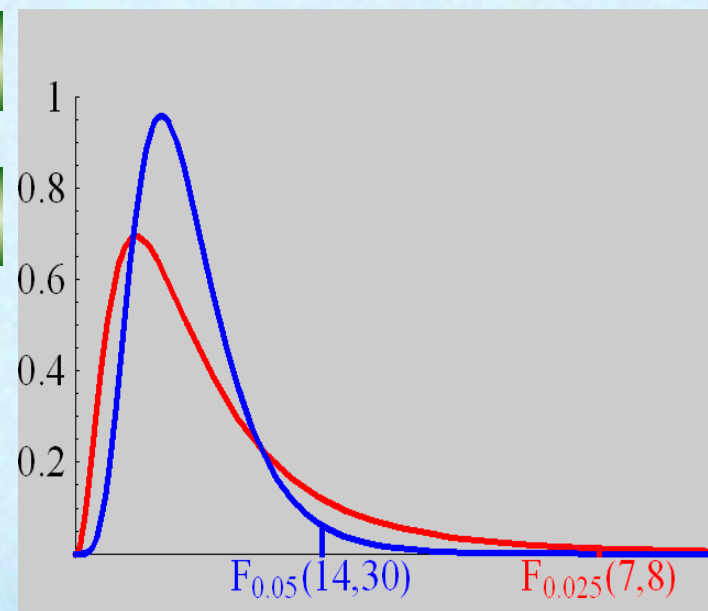
$$F_{0.025}(7, 8) = 4.90,$$

附表4-1

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

附表4-2

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

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

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



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

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

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 $F_{0.95}(12, 9) = \frac{1}{F_{0.05}(9, 12)} = \frac{1}{0.28} = 0.357$.



4. 正态总体的样本均值与样本方差的分布

定理一

设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, $\bar{X}$ 是样本均值, 则有

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

正态总体 $N(\mu, \sigma^2)$ 的样本均值和样本方差有以下两个重要定理.



定理二

设 $X_1, X_2, \dots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, $\bar{X}$, S^2 分别是样本均值和样本方差, 则有

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

(2) $\bar{X}$ 与 S^2 独立.



定理三

设 $X_1, X_2, \dots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, $\bar{X}, S^2$ 分别是样本均值和样本方差, 则有

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

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

且两者独立, 由 t 分布的定义知

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} \sim t(n-1).$$



定理四

设 $X_1, X_2, \dots, X_{n_1}$ 与 $Y_1, Y_2, \dots, Y_{n_2}$ 分别是具有相同方差的两正态总体 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ 的样本, 且这两个样本互相独立, 设 $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$,

$\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$ 分别是这两个样本的均值,

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

分别是这两个样本的方差, 则有

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

(2) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

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



证明 (1) 由定理二

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

由假设 S_1^2, S_2^2 独立, 则由 F 分布的定义知

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

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



$$(2) \text{ 因为 } \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

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

$$\text{由 } \frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1), \quad \frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1),$$

且它们相互独立, 故由 χ^2 分布的可加性知

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

由于 U 与 V 相互独立, 按 t 分布的定义.

$$\begin{aligned} & \frac{U}{\sqrt{V/(n_1 + n_2 - 2)}} \\ &= \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2). \end{aligned}$$



三、小结

两个最重要的统计量:

样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

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

三个来自正态分布的抽样分布:

χ^2 分布, t 分布, F 分布.

