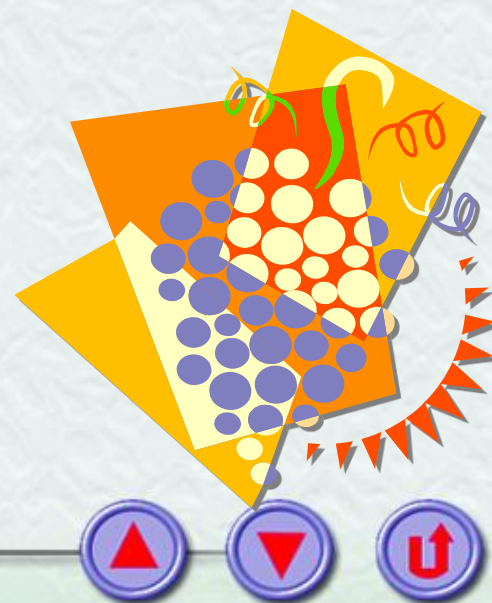


# 第五章 数理统计的基本概念

第一节 总体与样本

第二节 统计量

第三节 抽样分布



# 什么是统计学(Statistics)

陈希孺先生《数理统计简史》(2002)提到:

统计学是收集和分析数据的科学与艺术。

——不列颠百科全书(2000-2007年)

Statistics, the science of collecting, analyzing, presenting, and interpreting data.

统计学是收集、分析、展示和解释数据的科学。

——不列颠百科全书(2008-2020年)

<https://www.britannica.com/science/statistics>



一般来说，人们将统计分为两类：

➤ 描述性统计 (Descriptive Statistics)

运用制表和分类，图形以及计算概括性数据来描述数据特征，主要包括数据的频数分析、集中趋势分析、离散程度分析、分布以及一些基本的统计图形。

➤ 数理统计 (Mathematical Statistics)

本课程主要研究数理统计。





# 什么是统计学(Statistics)

陈希孺先生《数理统计简史》(2002) 提到:

统计学是收集和分析数据的科学与艺术。

用有效的  
方法

带有  
随机性  
的数据

统计思维  
统计思想  
统计方法

统计分析都有一定的虽很小但仍属可能的犯错误的机会

## § 5.1 总体与样本

### 一、总体和个体

总体 (population) —— 研究对象的全体

个体 —— 每一个对象

**例如** 研究某企业生产的一批电视机显象管的平均使用寿命, 那么这一批显象管的全体就组成一个总体, 其中每一只显象管就是一个个体。

**例如** 研究某大学一年级学生的身高情况, 这时一年级大学生的全体就是总体; 每个大学生就是一个个体。



在实际中我们真正所关心的是总体的某种数量指标，例如显象管的寿命指标 $X$ ，学生的身高指标 $Y$ ，它们都是 $r.v.$ （意思是：从中任取一只显象管，其寿命是不能预先确定的，可看作是 $X$ 的可能取值）。称这样的 $r.v.$ 为**表征总体的随机变量**。

为了方便起见，我们就将表征总体的随机变量定义为总体。

总体  $\longleftrightarrow$   $r.v. X$

若 $X$ 的分布函数为 $F(x)$ ，则称总体的分布函数为 $F(x)$ 。





## 有限总体和无限总体 (按总体所含个体的数量分)

当有限总体包含的个体的总数很大时, 可近似地将它看成是无限总体.

**K维总体:** 所研究的数量指标有K个



对总体进行研究时，对总体中每个个体逐一进行考察，这在实际中往往是行不通的，一是试验具有破坏性，二是需花费大量的人力物力；

常用的方法是：从总体中随机地抽取若干个个体，根据对这部分个体的研究结果推断总体某方面的特征。

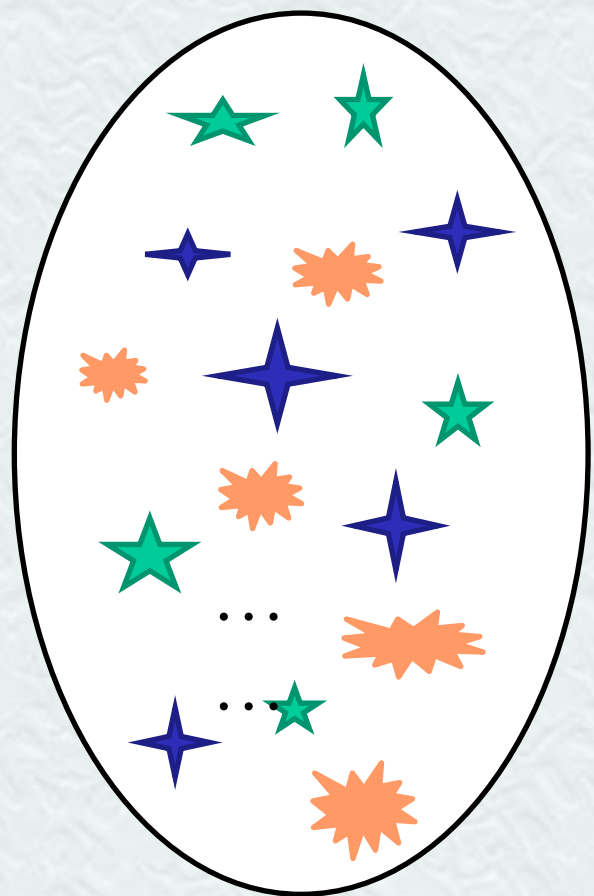
## 二、样本 (sample)

**定义** 从总体 $X$ 中随机地抽取 $n$ 个个体，称之为总体 $X$ 的一个**样本容量**为 $n$ 的**样本**。



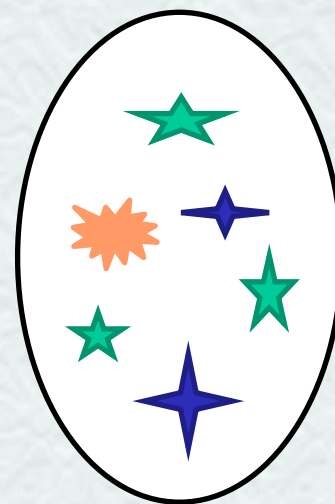


总体(未知)



研究对象的全体

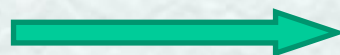
样本(已知)



分析



收集



具有代表性的部分个体



从总体中抽取若干个个体的过程称为**抽样**

假设抽样满足下述两个条件：

- (1) **随机性** 为了使样本具有充分的代表性，抽样必须是随机的，应使总体中的每一个个体都有同等的机会被抽取到。
- (2) **独立性** 各次抽样必须是相互独立的，即每次抽样的结果既不影响其它各次抽样的结果，也不受其它各次抽样结果的影响。

这种随机的、独立的抽样方法称为**简单随机抽样**，由此得到的样本称为**简单随机样本**。本课程中凡是提到抽样与样本，都是指简单随机抽样与简单随机样本。



**例如** 总体 $X$ 是一批显象管的使用寿命，现从总体 $X$ 中抽取 $n$ 个显象管， $X_i$ 表示抽到的第 $i$ 个显象管的使用寿命， $i=1, 2, \dots, n$ ；由于抽取的随机性，显然，每一个 $X_i$ 都是随机变量，并且有着和总体 $X$ 相同的分布。另外，由于抽取的独立性，

$X_1, X_2, \dots, X_n$  相互独立。

**记**  $X_1, X_2, \dots, X_n$  为总体 $X$ 的一个样本容量为 $n$ 的样本。其中 $X_i$ 表示第 $i$ 个个体的某个数量指标，是一个r.v.。

且  $X_1, X_2, \dots, X_n$  独立同分布（与总体 $X$ 同分布）。





从总体 $X$ 中抽取一个个体，就是对 $X$ 进行一次试验（或观测），得到一个试验数据（或观测值）。因此对于一次具体的抽样观测结果，我们将得到一组数据，记作  $x_1, x_2, \dots, x_n$ ，称之为**样本的一次观测值（样本值）**。

**例如** 从某厂生产的显象管中随机抽取10个显象管，测得寿命如下（单位千小时）：

4.8, 3.4, 5.2, 4.7, 5.5, 4.2, 4.5, 3.9, 5.0, 4.9

这十个数据就是样本容量为10的样本  $X_1, X_2, \dots, X_{10}$  的一组观测值  $x_1, x_2, \dots, x_{10}$ 。



样本可看作 $n$ 维随机变量 $(X_1, X_2, \dots, X_n)$ , 且

$X_1, X_2, \dots, X_n$ 独立同分布(与总体 $X$ 同分布), 则

(1) 当总体 $X$ 是离散型随机变量, 若记其分布律为  $P(X = x) = p(x)$ , 则样本 $(X_1, X_2, \dots, X_n)$ 的联合分布律为:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \underline{p(x_1, x_2, \dots, x_n)} \\ &= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n) \\ &= \underline{p(x_1) p(x_2) \cdots p(x_n) = \prod_{i=1}^n p(x_i)} \end{aligned} \quad (1)$$



(2) 当总体  $X$  是连续型随机变量, 且具有概率密度函数  $f(x)$  时, 则样本  $(X_1, X_2, \dots, X_n)$  的联合概率密度函数为

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^n f(x_i) \quad (2)$$





**例1** 设总体  $X$  服从参数为  $\lambda$  ( $\lambda > 0$ ) 的指数分布,  $(X_1, X_2, \dots, X_n)$  是来自总体的样本, 求样本  $(X_1, X_2, \dots, X_n)$  的概率密度.

**解** 总体  $X$  的概率密度为  $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$

所以  $(X_1, X_2, \dots, X_n)$  的概率密度为

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n f(x_i) \\ &= \begin{cases} \prod_{i=1}^n \lambda e^{-\lambda x_i}, & x_i > 0, \\ 0, & \text{其他.} \end{cases} = \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, & x_i > 0, \\ 0, & \text{其他.} \end{cases} \end{aligned}$$



**例2** 设总体  $X$  服从两点分布  $b(1, p)$ , 其中  $0 < p < 1$ ,  $(X_1, X_2, \dots, X_n)$  是来自总体的样本, 求样本  $(X_1, X_2, \dots, X_n)$  的分布律.

**解** 总体  $X$  的分布律为  $P\{X = x\} = p^x (1-p)^{1-x}$

所以  $(X_1, X_2, \dots, X_n)$  的分布律为 ( $x = 0, 1$ )

$$\begin{aligned} p(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n p(x_i) = \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \end{aligned}$$

其中  $x_1, x_2, \dots, x_n$  在集合  $\{0, 1\}$  中取值.



## § 5.2 统 计 量

### 一、统计量的定义 (定义5.2.1)

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. 统计量的分布称为**抽样分布**.

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.





**例1** 设  $X_1, X_2, X_3$  是来自总体  $X \sim N(\mu, \sigma^2)$  的一个样本, 其中  $\mu$  为已知,  $\sigma^2$  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

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

$$T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2). \quad T_7 = X + 2\mu.$$

不是



## 二、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  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k;$

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

5. 样本k阶中心矩  $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

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



**例2** 从某初中一年级男生中任意抽取12名，测得他们的身高如下（单位：cm）：171，165，174，175，168，164，173，178，168，170，172，173  
试估计该年级男生的平均身高，并估计其方差和标准差

**解：**

$$\bar{x} = \frac{1}{12} (171 + 165 + 174 + \cdots + 173) \approx 170.92$$
$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$
$$= \frac{1}{11} (171^2 + 165^2 + \cdots + 173^2 - 12 \times 170.92^2) \approx 16.99$$
$$s = \sqrt{s^2} \approx 4.12$$



## 6. 经验分布函数 (Empirical cdf)

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

$$F_n(x) = \frac{1}{n} S(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x), \quad -\infty < x < +\infty$$

经验分布函数是为了估计总体分布函数  $F(x)$  而构造的统计量, 也称样本分布函数.





一般地,

设  $x_1, x_2, \dots, x_n$  是总体  $F$  的一个容量为  $n$  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

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



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.)

**例3** 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

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



**例4** 设总体  $F$  具有一个样本值 1, 1, 2,

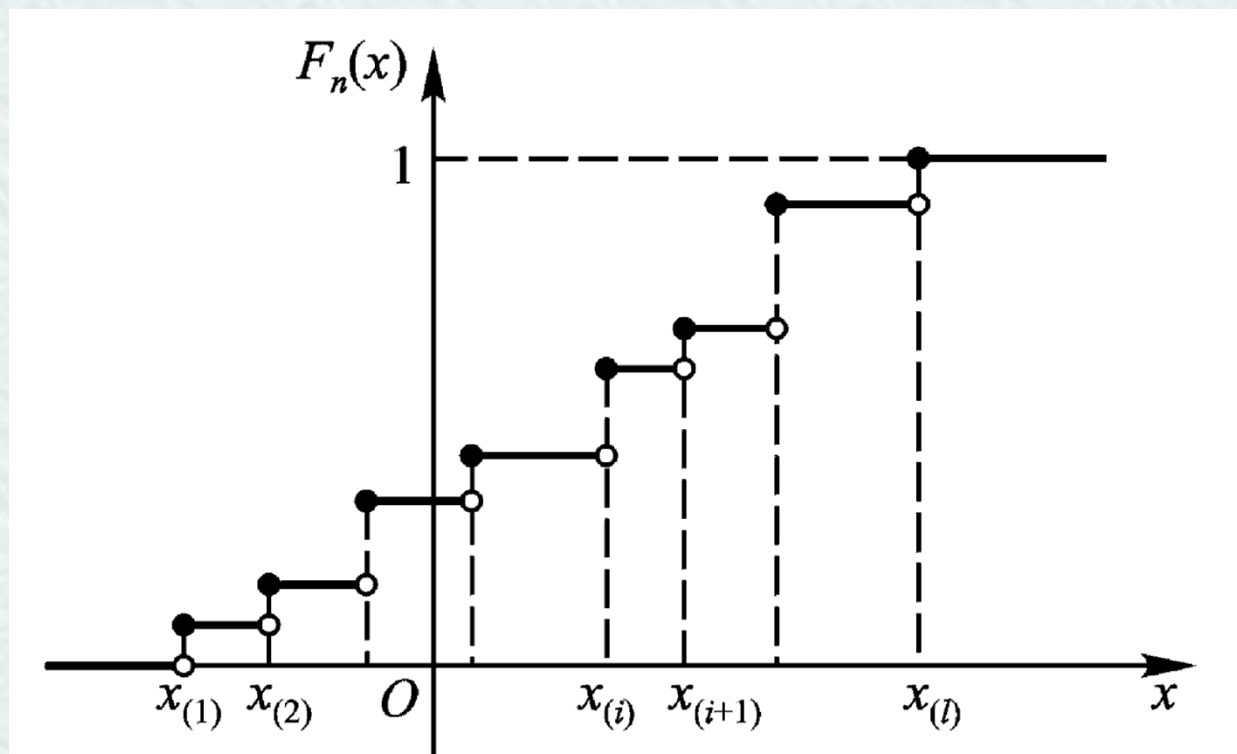
则经验分布函数  $F_3(x)$  的观察值为

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

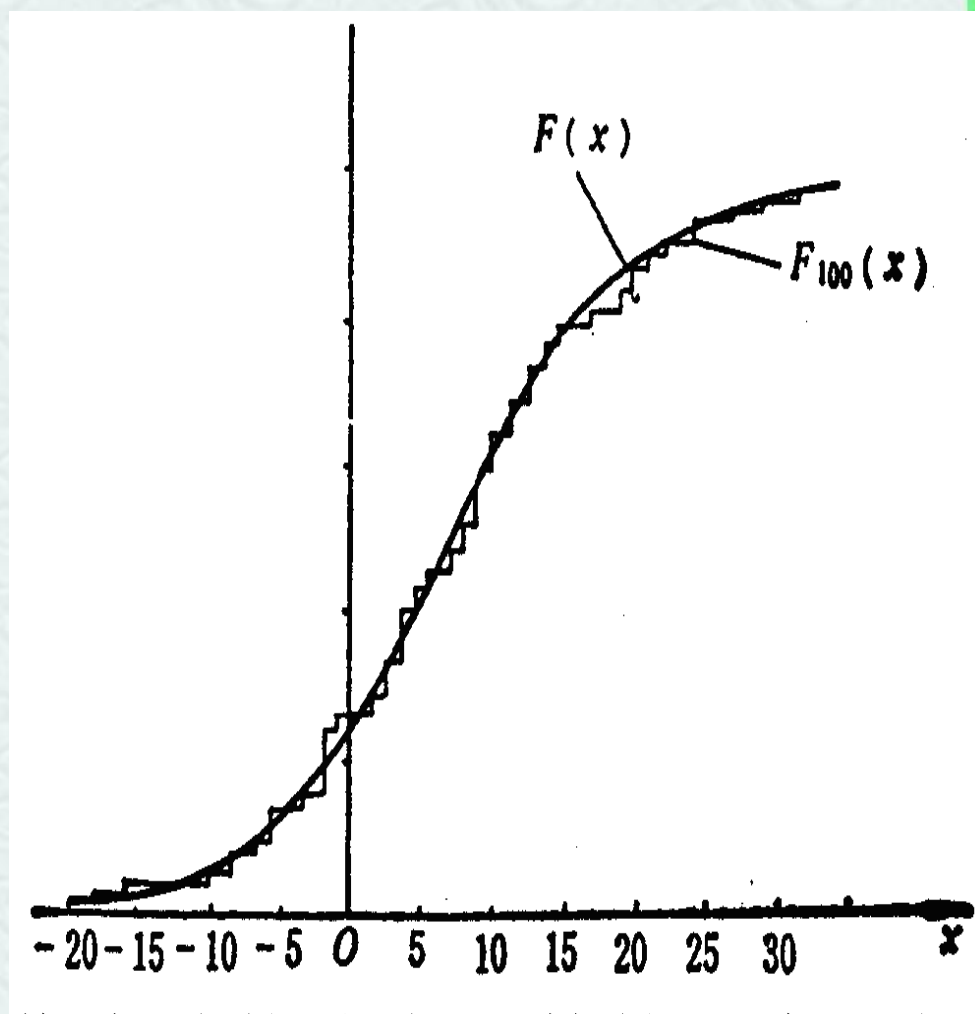




样本分布函数  $F_n(x)$  的图形如图所示



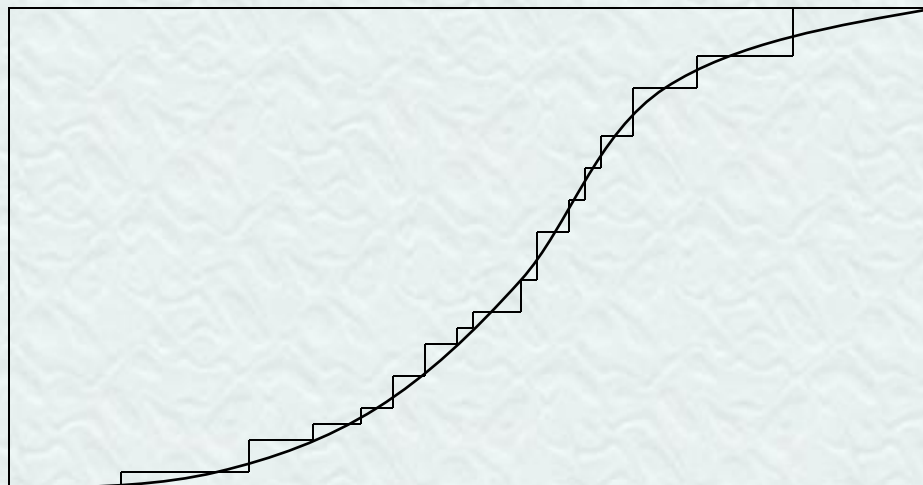
例



$F_n(x)$  可作为总体分布函数的一个近似， $n$  越大，近似得越好。

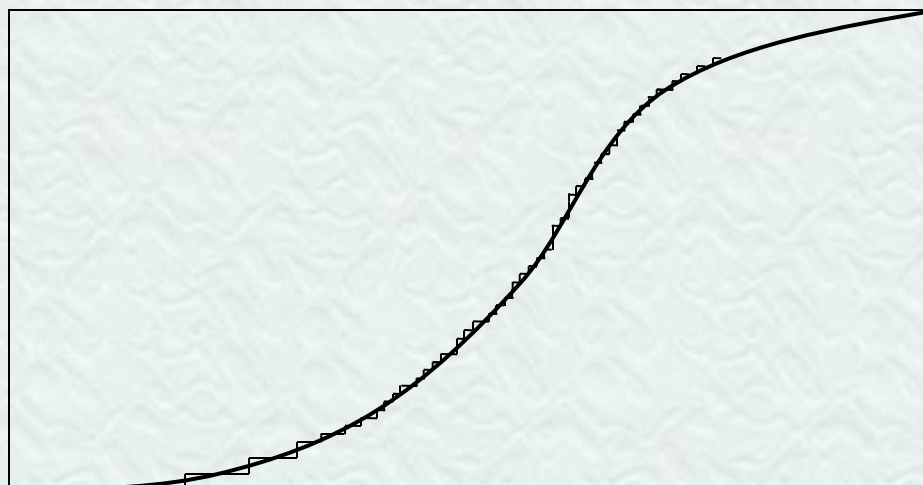


(a)



$n$ 较小

(b)



$n$ 较大





可以看出，对于不同的样本，得到的样本分布函数也不尽相同。随着 $n$ 的增大，样本分布函数与总体分布函数  $F(x)$  越来越接近。

有定理表明，样本分布函数实际上将近似地对等于总体的分布函数。这就是我们用样本推断总体的理论依据。



由伯努利大数定律和中心极限定理，可得：

对任意固定的  $x$ ，

$$\lim_{n \rightarrow \infty} P \left\{ |F_n(x) - F(x)| > \varepsilon \right\} = 0,$$

$$\text{即 } F_n(x) \xrightarrow{P} F(x), \quad n \rightarrow \infty$$

$$\sqrt{n} (F_n(x) - F(x)) \xrightarrow{d} N(0, F(x)[1 - F(x)])$$



## 格里汶科定理

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, 从而在实际中可当作  $F(x)$  来使用.





## 7. 次序统计量

**定义** 设 $(X_1, X_2, \dots, X_n)$ 是从总体 $X$ 中抽取的一个样本,  
 $(x_1, x_2, \dots, x_n)$ 是其一个观测值,将观测值按由小到  
大的次序重新排列为

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

当 $(X_1, X_2, \dots, X_n)$ 取值为 $(x_1, x_2, \dots, x_n)$ 时,定义  
 $X_{(k)}$ 取值为 $x_{(k)}$  ( $k = 1, 2, \dots, n$ ),由此得到

$$(X_{(1)}, X_{(2)}, \dots, X_{(n)})$$

称其为样本 $(X_1, X_2, \dots, X_n)$ 的次序统计量.对应的  
 $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ 称为其观测值.



## 特别的

$X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$  称为最小次序统计量.

$X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$  称为最大次序统计量.

## 说明

由于每个  $X_{(k)}$  都是样本  $(X_1, X_2, \dots, X_n)$  的函数, 所以,  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  也都是随机变量, 并且它们一般不相互独立.



## $X_{(n)}$ 及 $X_{(1)}$ 的分布

设  $X_1, X_2, \dots, X_n$  是总体  $F$  的一个样本,

则有  $F_{X_{(n)}}(x) = P\{X_{(n)} \leq x\} = P\{X_1 \leq x, \dots, X_n \leq x\}$

$$\stackrel{\text{独立}}{=} P\{X_1 \leq x\} \cdots P\{X_n \leq x\} \stackrel{\text{同分布}}{=} [F(x)]^n$$

$$F_{X_{(1)}}(x) = P\{X_{(1)} \leq x\} = 1 - P\{X_{(1)} > x\}$$

$$= 1 - P\{X_1 > x, \dots, X_n > x\}$$

$$\stackrel{\text{独立}}{=} 1 - P\{X_1 > x\} \cdots P\{X_n > x\} \stackrel{\text{同分布}}{=} 1 - [1 - F(x)]^n$$





## § 5.3 抽样分布 —— 统计量的分布

### 一、三个重要的抽样分布



卡尔·皮尔逊 (1857-1936)



戈塞特 (1876-1937)



费希尔 (1890-1962)

数理统计三剑客

## § 5.3 抽样分布 —— 统计量的分布

### 一、三个重要的抽样分布

#### 1. $\chi^2$ 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  $\chi^2$  分布, 记为  $\chi^2 \sim \chi^2(n)$ .

自由度:

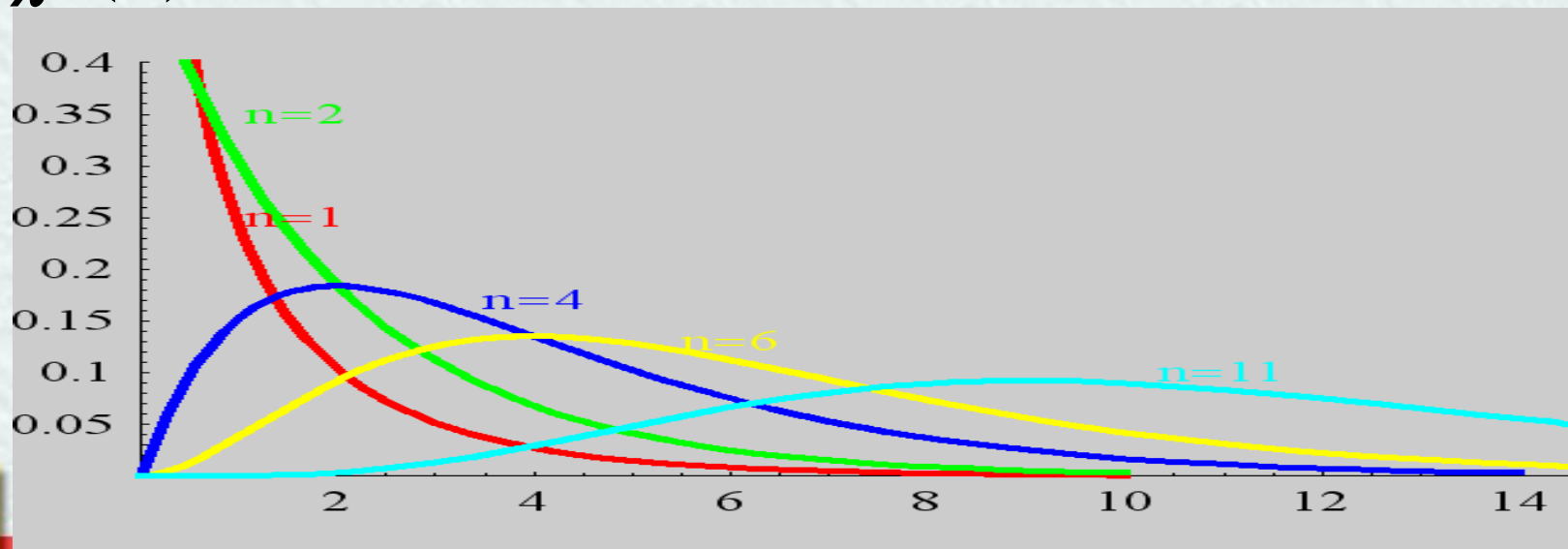
指  $\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$  中右端包含独立变量的个数.



$\chi^2(n)$ 分布的概率密度为

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

$\chi^2(n)$ 分布的概率密度曲线如图.





## 练习册 § 5.1-5.3

8. 设  $X_1, X_2, X_3$  是取自正态总体  $X \sim N(0, 0.25)$  的样本, 若  $a(X_1^2 + X_2^2 + X_3^2) \sim \chi^2(3)$ , 则  $a =$ \_\_\_\_\_.

例5.3.1 设总体  $X \sim N(0,4)$ ,  $X_1, X_2, X_3, X_4$

是来自该总体的一个样本, 若

$$a(2X_1 - X_2)^2 + b(3X_3 + 4X_4)^2 \sim \chi^2(2)$$

求常数  $a, b$  的值.

$$2X_1 - X_2 \sim N(0, 20), \quad \frac{2X_1 - X_2}{\sqrt{20}} \sim N(0, 1)$$

$$3X_3 + 4X_4 \sim N(0, 100), \quad \frac{3X_3 + 4X_4}{\sqrt{100}} \sim N(0, 1)$$

$$\therefore a = \frac{1}{20}, \quad b = \frac{1}{100}$$



## 二 计算题

1. 设  $X_1, X_2, X_3, X_4$  是取自正态总体  $X \sim N(0, 4)$  的一组样本,

$$Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$$

问  $a, b$  分别为何值时, 统计量  $Y$  服从  $\chi^2$  分布, 其自由度是多少?



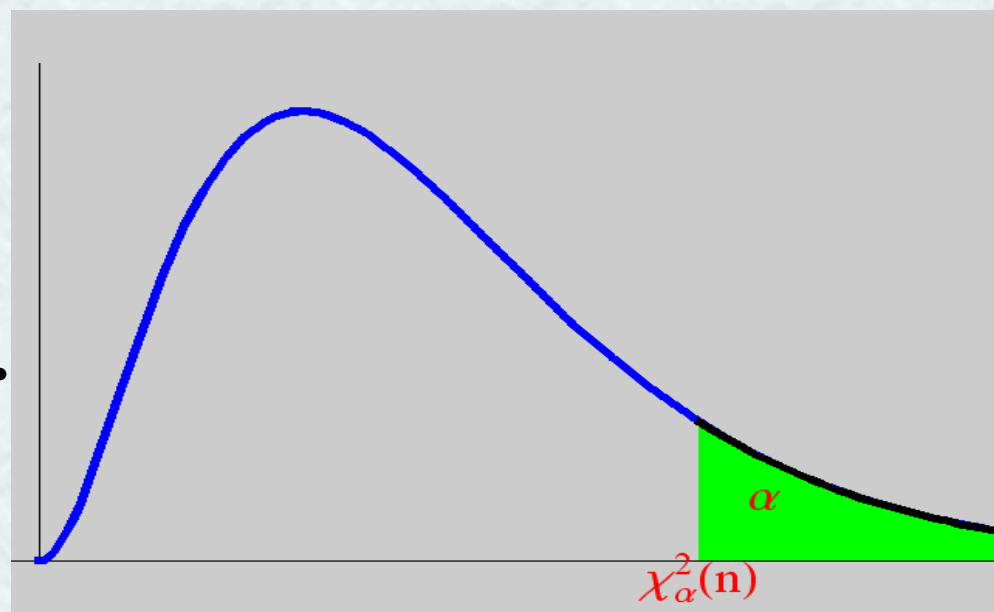
## $\chi^2$ 分布的分位点

对于给定的正数  $\alpha$ ,  $0 < \alpha < 1$ , 称满足条件

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

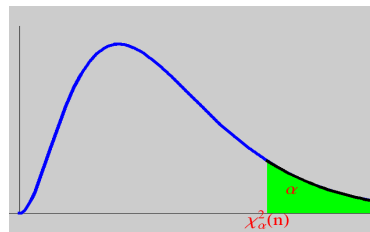
的点  $\chi_{\alpha}^2(n)$  为  $\chi^2(n)$  分布的上  $\alpha$  分位点.

对于不同的  $\alpha, n$ ,  
可以通过查表求  
得上  $\alpha$  分位点的值.



# 附录四 $\chi^2$ 分布表

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



统计

$n \backslash \alpha$	0.995	0.99	0.975	0.95	0.90	0.75	0.25	0.10	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	14.845	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	9.299	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	20.489	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	13.675	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	22.718	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	24.935	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	17.240	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	18.137	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	28.241	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	19.939	29.339	34.382	37.652	40.646	44.314	46.928



**例1**  $\chi_{0.975}^2(10) = 3.247,$

$$\chi_{0.1}^2(25) = 34.382$$

**例2** 设  $\chi^2 \sim \chi^2(8)$ , 试确定  $\lambda_1, \lambda_2$  的值, 使之满足:

$$P(\chi^2 > \lambda_1) = 0.05, \quad P(\chi^2 < \lambda_2) = 0.05.$$

**解**  $\lambda_1 = \chi_{0.05}^2(8) = 15.507,$

$$\lambda_2 = \chi_{0.95}^2(8) = 2.733,$$





## 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  $T \sim t(n)$ .

$t$  分布又称**学生氏(Student)分布**.

$t(n)$  分布的概率密度函数为

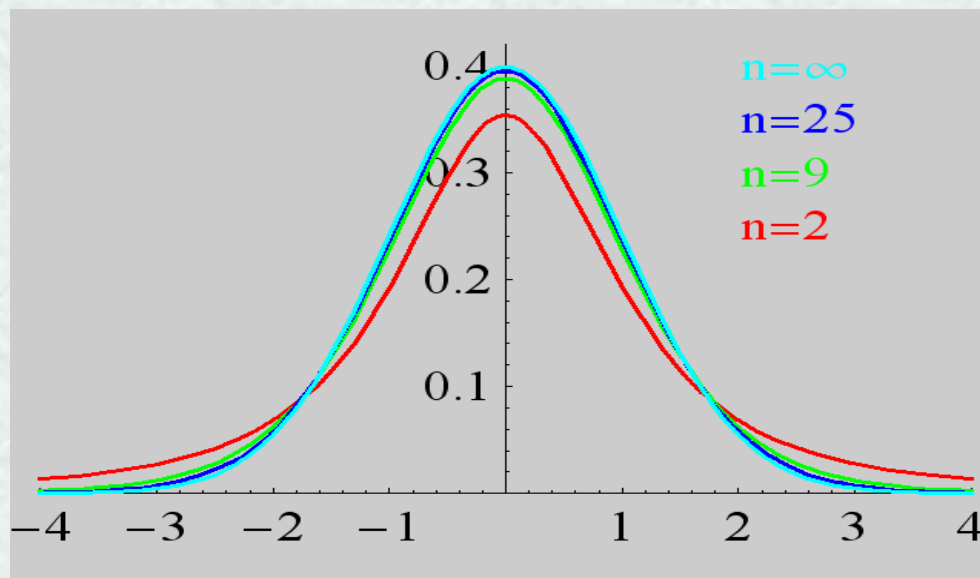
$$h(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty$$



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

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.



**例5.3.2** 设  $X_1, X_2, X_3, X_4, X_5$  是来自标准正态总体  $X \sim N(0,1)$  的一组样本，试求下列随机变量的分布：

$$(1) \quad \frac{2X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2 + X_5^2}} = \frac{X_1}{\sqrt{\frac{X_2^2 + X_3^2 + X_4^2 + X_5^2}{4}}} \sim t(4)$$

$$(2) \quad \frac{X_2}{|X_1|} = \frac{X_2}{\sqrt{X_1^2}} \sim t(1)$$





## $t$ 分布的分位点

对于给定的  $\alpha$ ,  $0 < \alpha < 1$ , 称满足条件

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

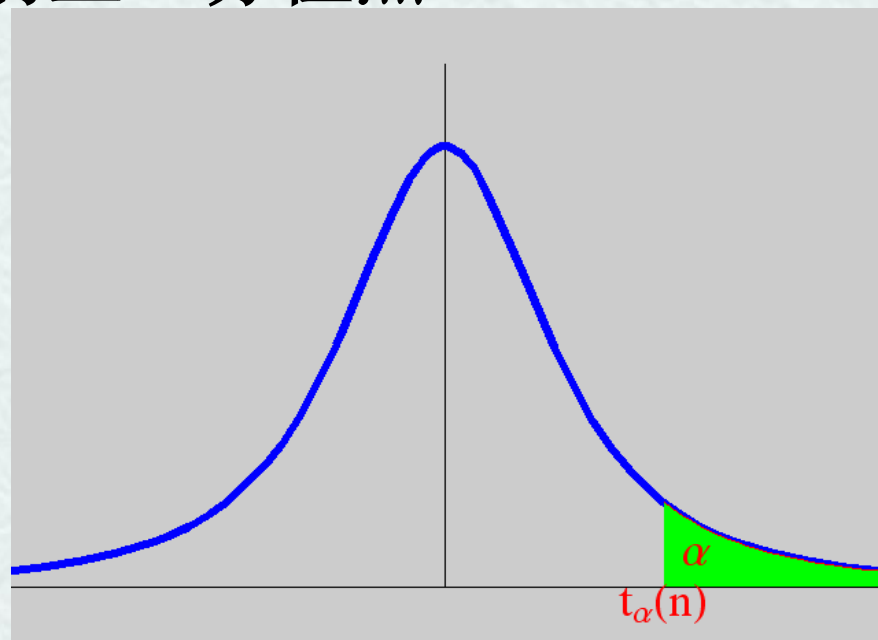
的点  $t_{\alpha}(n)$  为  $t(n)$  分布的上  $\alpha$  分位点.

可以通过查表求得  
上  $\alpha$  分位点的值.

由分布的对称性知

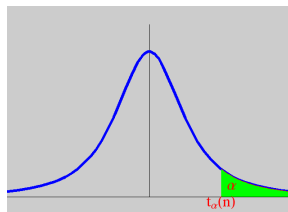
$$t_{1-\alpha}(n) = -t_{\alpha}(n).$$

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



# 附录三 $t$ 分布表

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



统计

$\alpha \backslash n$	0.20	0.15	0.10	0.05	0.025	0.01	0.005
1	1.376	1.963	3.0777	6.3138	12.7062	31.8205	63.6567
2	1.061	1.386	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.978	1.250	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.941	1.190	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.920	1.156	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.906	1.134	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.896	1.119	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.889	1.108	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.883	1.100	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.879	1.093	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.876	1.088	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.873	1.083	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.870	1.079	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.868	1.076	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.866	1.074	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.865	1.071	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.863	1.069	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.862	1.067	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.861	1.066	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.860	1.064	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.859	1.063	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.858	1.061	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.858	1.060	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.857	1.059	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.856	1.058	1.3163	1.7081	2.0595	2.4851	2.7874



**例3**  $t_{0.05}(10) = 1.8125,$

$$t_{0.975}(15) = -t_{0.025}(15) = -2.1314.$$

**例4** 设  $T \sim t(14)$  , 试确定  $\lambda_1, \lambda_2$  的值, 使之满足:

$$P(T > \lambda_1) = 0.05, \quad P(|T| < \lambda_2) = 0.95.$$

**解**  $\lambda_1 = t_{0.05}(14) = 1.7613,$

$$P(T > \lambda_2) = \frac{0.05}{2} = 0.025$$

$$\Rightarrow \lambda_2 = t_{0.025}(14) = 2.1448,$$





## 二、正态总体的样本均值和样本方差的分布

**定理一** 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

$$(1) \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \text{ 即 } U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1);$$

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

(3)  $\bar{X}$  与  $S^2$  独立;

$$(4) \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1);$$

$$(5) \quad \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n).$$



证明: (1)  $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$

根据性质: 独立正态分布r.v.的线性组合  
仍然服从正态分布;

$$\text{又 } E(\bar{X}) = \mu, \quad D(\bar{X}) = \frac{\sigma^2}{n},$$

$$\text{故 } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\text{标准化: } U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$



## 证明 (4)

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

且两者独立, 由  $t$  分布的定义知

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





## 练习册 § 5.1-5.3

7. 设  $X_1, X_2, \dots, X_{25}$  是取自正态总体  $X \sim N(20, 16)$  的一组样本,  $\bar{X} = \frac{1}{25} \sum_{i=1}^{25} X_i$ ,

则  $P(\bar{X} < 21) =$  \_\_\_\_\_,  $P\{\sum_{i=1}^{16} X_i - \sum_{i=17}^{25} X_i \leq 155\} =$  \_\_\_\_\_.

2. 设  $X_1, X_2, \dots, X_n$  是来自总体  $U(0,1)$  的样本, 求:

(1)  $P\{X_1 + X_2 + X_3 \leq 1\}$ ; (2)  $P\{X_1^2 + X_2^2 + X_3^2 \leq 1\}$ ; (3)  $P\{X_1 + \dots + X_{1200} > 605\}$