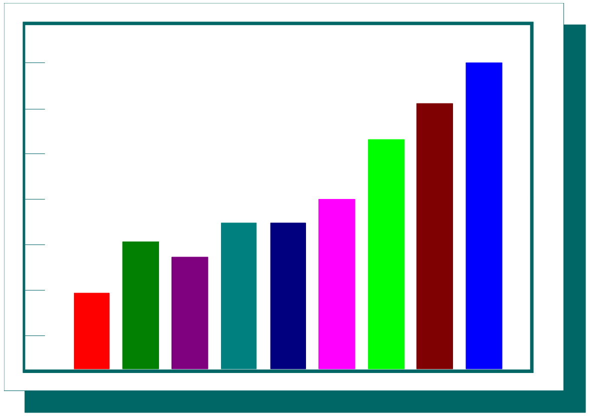
第六章 样本及抽样分布

§1 随机样本

§2 抽样分布

数理统计学是一门应用性很强的学科. 它是研究怎样以有效的方式收集、 整理和分析带有随机性的数据，以便对所考察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议.



计算机的诞生与发展，为数据处理提供了强有力的技术支持，数理统计与计算机的结合是必然的发展趋势.

学习统计无须把过多时间花在计算上，可以更有效地把时间用在基本概念、方法原理的正确理解上. 国内外著名的统计软件包： *SAS*，*SPSS*，*STAT*等，都可以让你快速、简便地进行数据处理和分析.

概率论与数理统计是两个有密切联系的学科，它们都以随机现象的统计规律为研究对象.

但在研究问题的方法上有很大区别：

概率论 —— 已知随机变量服从某分布，寻求分布的性质、数字特征、及其应用；

数理统计 —— 通过对试验数据的统计分析，寻找所服从的分布和数字特征，从而推断总体的规律性.

数理统计的核心问题——由样本推断总体

《概率论》中待解决的问题

一． **在实际问题中，如何验证随机现象的等可能性.**

例如，“生男生女一样吗”，“骰子均匀吗”.

二． **在实际问题中，如何确定随机变量的分布类型和分布中的参数.**

例如，“学生成绩服从正态分布吗”，“人的寿命服从指数分布吗”，“交通事故数服从泊松分布吗”.

什么是数理统计?

数理统计学以简单随机样本为依据,

利用统计量或枢轴量的抽样分布,

对所研究的问题的总体分布做统计推断。

§1 随机样本

- 总体
- 个体
- 样本

一、总体和个体

1）总体：研究对象的某项数量指标的值的全体。

2）个体：总体中的每个元素为个体。

例如：某工厂生产的灯泡的寿命是一个总体，每一个灯泡的寿命是一个个体；某学校男生的身高的全体一个总体，每个男生的身高是一个个体。

如研究某批灯泡的寿命时，关心的数量指标就是寿命，那么，此总体就可用描述其寿命的随机变量 X 或用其分布函数 $F(x)$ 表示.

再如，若研究某地区中学生的营养状况时，关心的数量指标是身高和体重。

我们用X和Y分别表示身高和体重，那么此总体就可用二维随机变量(X,Y)或其联合分布函数$F(x, y)$来表示.

总体概念的要旨： **总体就是一个随机变量或一个概率分布！**

一、总体及总体分布

直观地说：全部可能的考察值称为总体；每一个可能的考察值称为个体.

随机变量的一个实现值

随机变量

抽像地说，我们把用来描述研究对象的某些属性或特征的随机变量或随机向量称为总体，并将该变量或向量的分布称为总体分布.

定义：在统计学中，称随机变量为总体，并把将该变量的分布称为总体分布.

注 意

<u>总体的三层含义</u>：

- 研究对象的全体；
- 数据；
- 分布

例5.1.1 考察某厂的产品质量，以0记合格品，以1记不合格品，则

总体 = {该厂生产的全部合格品与不合格品}

= {由0或1组成的一堆数}

若以 p 表示这堆数中1的比例（不合格品率），则该总体可由一个二点分布表示：

X	0	1
P	$1-p$	p

比如：两个生产同类产品的工厂的产品的总体分布：

X	0	1
p	0.983	0.017

X	0	1
p	0.915	0.085

一、总体及总体分布

两个数量指标，
如身高和体重

说明：

（1）**总体**---也可以是随机向量，但是只考虑随机变量；

（2）**个体**---其特征不是数量指标，需转化成数量指标

如：工厂的
品和非残

（3）**总体分布**---表示总体的随机变量X的分布，分布未知、或是类型已知但参数不知

例1 如果想研究“生男生女的机会是否一样”, 请问如何设定总体？ 总体 $X \sim B(1, p),\ p = P\{\text{男孩出生}\}.$

例2 如果想研究“中国城乡收入是否一样”, 请问如何设定总体？

$$\text{令}X=\text{家庭收入}, Y=\begin{cases}1, & \text{城市家庭}\\ 0, & \text{农村家庭}\end{cases}, \text{则总体为}(X,Y).$$

例5.1.2　在二十世纪七十年代后期，美国消费者购买日产SONY彩电的热情高于购买美产SONY彩电，原因何在？

原因在于总体的差异上！

- 1979年4月17日日本《朝日新闻》刊登调查报告指出$N(m, (5/3)^2)$，日产SONY彩电的彩色浓度服从正态分布，

  而美产SONY彩电的彩色浓度服从$(m-5\ ,\ m+5)$上的均匀分布。

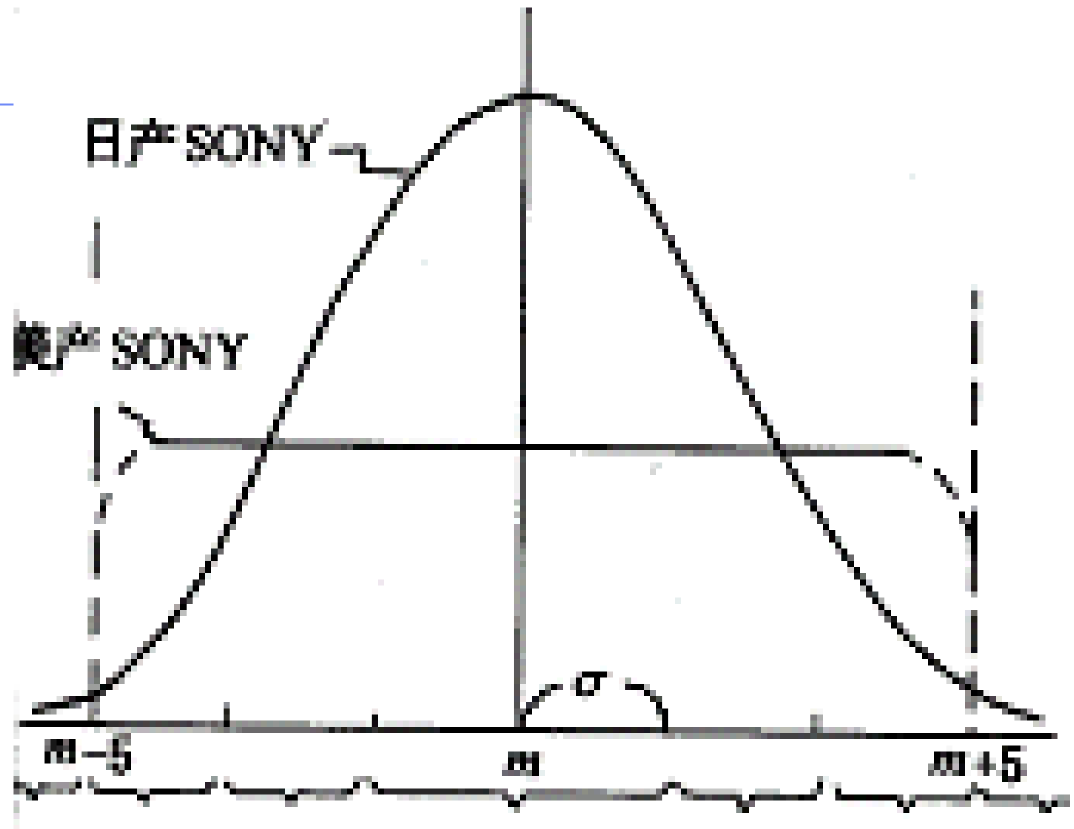


图5.1.1　SONY彩电彩色浓度分布图

表5.1.1 各等级彩电的比例(%)

等级	I	II	III	IV
美产	33.3	33.3	33.3	0
日产	68.3	27.1	4.3	0.3

二、样本及样本分布

如，抽查的部分居民的年收入

直观地说，**样本**就是按一定规定从总体中抽出的部分个体.

其中所谓“一定的规定”通常是指每个个体都有**同等地被抽出**的机会，而且各次抽出之间**相互没有影响**.

抽出的某一个个体---随机变量，记为 X_i

抽取的某一个个体的数量指标 观察值，记为 x_i

一次抽样的观察值，记为 $(x_1, x_2, \cdots, x_n)$

抽出的个体—随机向量，记为 $(X_1, X_2, \cdots, X_n)$

约定：大写的英文字母表示随机变量，相应的小写字母表示他的观察值

统计学的主要任务

借助这组观察值对未知的总体分布进行合理的推断

二、样本

定义：设 X 是具有分布函数 F 的随机变量，若 $X_1,\cdots X_n$ 是具有同一分布函数 F 的相互独立的随机变量，则称 $X_1,\cdots X_n$ 为从总体 X 中得到的容量为 n 的简单随机样本，简称为样本，其观察值 $x_1,\cdots x_n$ 称为样本值。

由定义知：若 $X_1,\cdots,X_n$ 为 X 的一个样本，则 $(X_1,\cdots,X_n)$ 的联合分布函数为：

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

若设X的概率密度为$f(x)$，则$(X_1,\cdots,X_n)$的联合概率密度为：

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

若设X的分布率为$P\{X=x\}=p(x)$，则$(X_1,\cdots,X_n)$的联合分布率为：

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i)$$

例1 若$X_1,\cdots,X_n$是正态总体$X\sim N(1,4)$的样本，则

$EX_1X_n=$ __1__， $D(X_1-2X_2)=$ __20__.

5.1.2 样本

样品、样本、样本量： 样本具有两重性

- 一方面，由于样本是从总体中随机抽取的，抽取前无法预知它们的数值，因此，样本是随机变量，用大写字母 $X_1, X_2, \ldots, X_n$ 表示；
- 另一方面，样本在抽取以后经观测就有确定的观测值，因此，样本又是一组数值。此时用小写字母 $x_1, x_2, \ldots, x_n$ 表示是恰当的。

简单起见，无论是样本还是其观测值，样本一般均用 $x_1, x_2, \ldots x_n$ 表示，应能从上下文中加以区别。

例5.1.3　啤酒厂生产的瓶装啤酒规定净含量为640克。由于随机性，事实上不可能使得所有的啤酒净含量均为640克。现从某厂生产的啤酒中随机抽取10瓶测定其净含量，得到如下结果：

641, 635, 640, 637, 642, 638, 645, 643, 639, 640

这是一个容量为10的样本的观测值，
对应的总体为该厂生产的瓶装啤酒的净含量。

这样的样本称为完全样本。

例5.1.4　考察某厂生产的某种电子元件的寿命，选了100只进行寿命试验，得到如下数据：

表5.1.2 100只元件的寿命数据

寿命范围	元件数	寿命范围	元件数	寿命范围	元件数
(0 24]	4	(192 216]	6	(384 408]	4
(24 48]	8	(216 240]	3	(408 432]	4
(48 72]	6	(240 264]	3	(432 456]	1
(72 96]	5	(264 288]	5	(456 480]	2
(96 120]	3	(288 312]	5	(480 504]	2
(120 144]	4	(312 336]	3	(504 528]	3
(144 168]	5	(336 360]	5	(528 552]	1
(168 192]	4	(360 184]	1	>552	13

表5.1.2中的样本观测值没有具体的数值，只有一个范围，这样的样本称为分组样本。

样本的要求：简单随机样本

要使得推断可靠，对样本就有要求，使样本能很好地代表总体。通常有如下两个要求：

- 随机性: 总体中每一个个体都有同等机会被选入样本 －－ x_i 与总体X有相同的分布。
- 独立性: 样本中每一样品的取值不影响其它样品的取值 －－ $x_1, x_2, \ldots, x_n$ 相互独立。

用简单随机抽样方法得到的样本称为简单随机样本，也简称样本。

于是，样本 $x_1, x_2, \ldots, x_n$ 可以看成是独立同分布(iid) 的随机变量，其共同分布即为总体分布。

设总体X具有分布函数$F(x)$，$x_1, x_2, \ldots, x_n$为取自该总体的容量为n的样本，则样本联合分布函数为

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

总体分为有限总体与无限总体

实际中总体中的个体数大多是有限的。当个体数充分大时，将有限总体看作无限总体是一种合理的抽象。

对无限总体，随机性与独立性容易实现，困难在于排除有意或无意的人为干扰。

对有限总体，只要总体所含个体数很大，特别是与样本量相比很大，则独立性也可基本得到满足。

例5.1.5 设有一批产品共N个，需要进行抽样检验以了解其不合格品率p。现从中采取不放回抽样抽出2个产品，这时，第二次抽到不合格品的概率依赖于第一次抽到的是否是不合格品，如果第一次抽到不合格品，则

$$P(x_2 = 1 \mid x_1 = 1) = (Np-1)/(N-1)$$

而若第一次抽到的是合格品，则第二次抽到不合格品的概率为

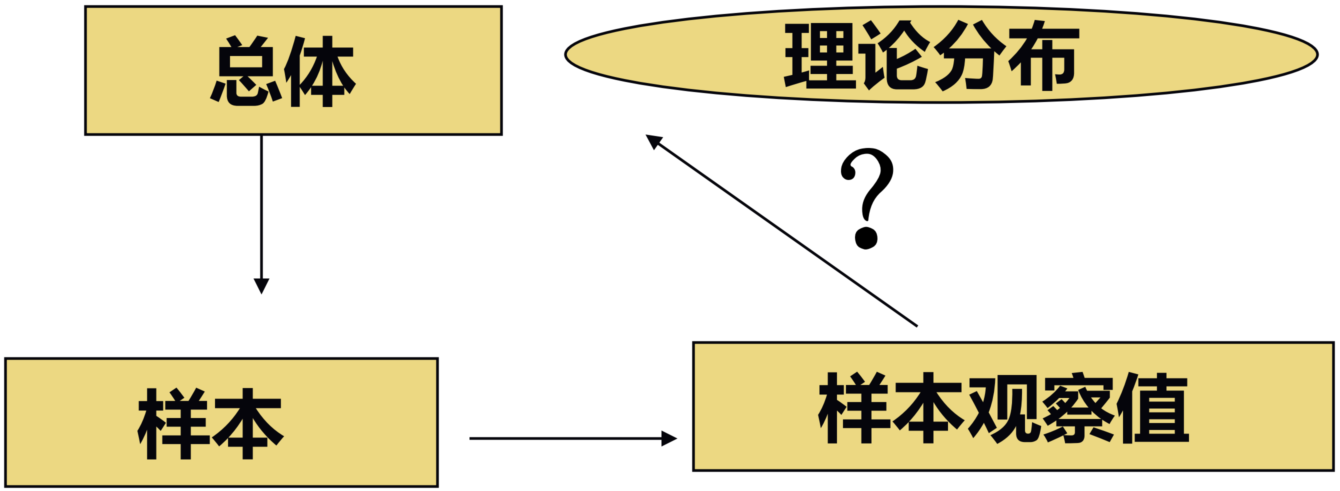
$$P(x_2 = 1 \mid x_1 = 0) = (Np)/(N-1)$$

显然，如此得到的样本不是简单随机样本。但是，当N很大时，我们可以看到上述两种情形的概率都近似等于p。所以当N很大，而n不大（一个经验法则是 $n/N \le 0.1$）时可以把该样本近似地看成简单随机样本。

思考:

若总体的密度函数为$p(x)$，则其样本的（联合）密度函数是什么?

总体、样本、样本观察值的关系



统计是从手中已有的资料——样本观察值，去推断总体的情况——总体分布。样本是联系两者的桥梁。总体分布决定了样本取值的概率规律，也就是样本取到样本观察值的规律，因而可以用样本观察值去推断总体

例2 若$X_1,\cdots,X_n$是总体$X\sim B(1,p)$的样本，求$(X_1,\cdots,X_n)$的联合分布率.

解： 总体 X 的分布率为

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

所以（$X_1,\cdots,X_n$）的联合分布率为

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i)$$

$$=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}$$

$$x_i=0,1,\ \ i=1,\cdots,n.$$

应用背景：工厂生产产品的废品率的估计

例3 若$X_1,\cdots,X_n$是参数为λ的泊松分布总体X的样本，求$(X_1,\cdots,X_n)$的联合分布率.

解：总体 X 的分布率为

$$p(x)=P\{X=x\}=\frac{\lambda^x}{x!}e^{-\lambda},x=0,1,\cdots$$

所以（$X_1,\cdots,X_n$）的联合分布率为

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i)$$

$$=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda},\quad x_i=0,1,\cdots,\ i=1,\cdots,n.$$

例4 若$X_1,\cdots,X_n$是总体$X\sim N(\mu,\sigma^2)$的样本，求$(X_1,\cdots,X_n)$的联合概率密度.

解: 总体 X 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty.$$

所以（$X_1,\cdots,X_n$）的联合概率密度为

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

$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi)^{-\frac{n}{2}}(\sigma)^{-n}e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

$$-\infty<x_i<\infty,i=1,\cdots,n.$$

应用背景：钢筋强度的估计、随机误差的估计

出

统计推断的关键：
构造基于样本的适当的函数

三、统计推断问题简述

统计推断问题：借助总体X的一个样本（$X_1, X_2, \cdots, X_n$），对总体X的未知分布进行推断的问题

即

借助样本**构造样本的适当的函数**：

利用这些函数所**反映的分布的信息**

对总体分布所属的**类型**，或总体分布中所含的**未知参数做出统计推断**

§2 抽样分布

统计量

χ^2-分布

$t-$分布

$F-$分布

正态总体的样本均值与样本方差的分布

一、统计量

1） 定义：设 $X_1,\cdots X_n$ 为来自总体X的一个样本，g是 $X_1,\cdots X_n$的函数，若g是连续函数，且g中不含任何未知参数，则称$g(X_1,\cdots X_n)$是统计量。

样本的函数；

随机向量的函数；

一个随机变量

设$(x_1,\cdots,x_n)$是相应于样本$(X_1,\cdots X_n)$的样本值。

则称$g(x_1,\cdots x_n)$是$g(X_1,\cdots X_n)$的观察值。

三、 统计量

函数构造的结果

直观地说，统计量就是根据问题需要对样本进行加工所得到的量.

例如, 设$(X_1, X_2, \cdots X_n)$是正态总体 $X \sim N(\mu, \sigma^2)$ 的一个样本,则 $\frac{1}{n}\sum_{i=1}^{n} X_i$ 就是一个统计量.

抽象地说, 一个统计量是样本$(X_1, X_2, \cdots, X_n)$的一个函数$g(X_1, X_2, \cdots, X_n)$且该函数不含任何未知参数.

练习: 设$(X_1, X_2, \cdots X_n)$是正态总体 $X \sim N(\mu, \sigma^2)$ 的一个样本,则 $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ 是否是一个统计量.

注：统计量是随机变量。

常用统计量

用于估计总体特征值

设 $(X_1, X_2, \cdots, X_n)$ 为总体 X 的一个样本，

1. 样本均值

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

2. 样本方差

未修正的样本方差：$S_0^2=\frac{1}{n}\sum_{i=1}^{n}(X_i$

修正的样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i$

为了保证无偏性；

修正样本方差是总体方差的无偏估计

3. 样本标准差

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

4. 样本原点矩 $$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\qquad k=1,2,\cdots$$

5. 样本中心矩 $$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\ k=1,2,\cdots$$

例如，五十年的最大降雨量；
居民收入的中间水平

6. 顺序统计量

设$(X_1,X_2,\cdots,X_n)$为总体 X 个样本, 将其排列成 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$,则$(X$ $), X_{(2)},\cdots,X_{(n)})$称为顺序统计量, $X_{(i)}$ 称为样本的第 i 个顺序统计量.

$X_{(1)}=\min(X_1,\cdots,X_n)$和$X_{(n)}=\max(X_1,\cdots,X_n)$ 分别称为样本极小值和样本极大值, $X_{(n)}-X_{(1)}$称为极差.

例1 设 $X_1,\cdots X_n$ 为来自总体 $X \sim N(\mu,\sigma^2)$ 的一个样本，其中 μ 未知，σ^2 已知，问下列随机变量中那些是统计量

$$\min(X_1,X_2,\cdots,X_n);\quad \frac{X_1+X_n}{2};\quad \frac{X_1+\cdots+X_n}{n}-\mu;$$

$$\frac{(X_1+X_n)^2}{\sigma^2};\quad \frac{(X_1+\cdots+X_n)-n\mu}{\sqrt{n}\sigma}.$$

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**。

2） 常用的统计量

样本均值 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i,$ 证明：

样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{1}{n-1}[\sum_{i=1}^{n}X_i{}^2-n\overline{X}^2]$

样本标准差 $S=\sqrt{S^2}=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

样本k阶原点矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i{}^k\quad k=1,2,\cdots$

样本k阶中心矩 $B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k\quad k=1,2,\cdots$

它们的观察值分别为：
$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}[\sum_{i=1}^{n}x_i^{\,2}-n\bar{x}^2]$$

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

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^{\,k},\quad k=1,2\cdots$$

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

分别称为样本均值、样本方差、样本标准差、样本k阶原点矩、样本k阶中心矩的观察值。

例2 在某工厂的轴承中随机取10只，测得其中量（以kg计）为

2.36 2.42 2.38 2.34 2.40

2.42 2.39 2.43 2.39 2.37

求样本均值，样本方差和样本标准差.

解：

$$\bar{x} = \frac{2.36 + 2.42 + \cdots + 2.37}{10} = 2.39(kg);$$

$$s^2 = \frac{1}{10-1}(2.36^2 + 2.42^2 + \cdots + 2.37^2 - 10 \times 2.39^2)$$

$$= 0.0008222(kg^2);$$

$$s = \sqrt{0.0008222} = 0.02867(kg).$$

3）结论：设 $X_1,\cdots X_n$ 为来自总体X 的一个样本，

$$EX=\mu,\ DX=\sigma^2,$$

则 $$E\overline{X}=\mu,\ D\overline{X}=\frac{\sigma^2}{n},\ ES^2=\sigma^2.$$

请记熟此结论！

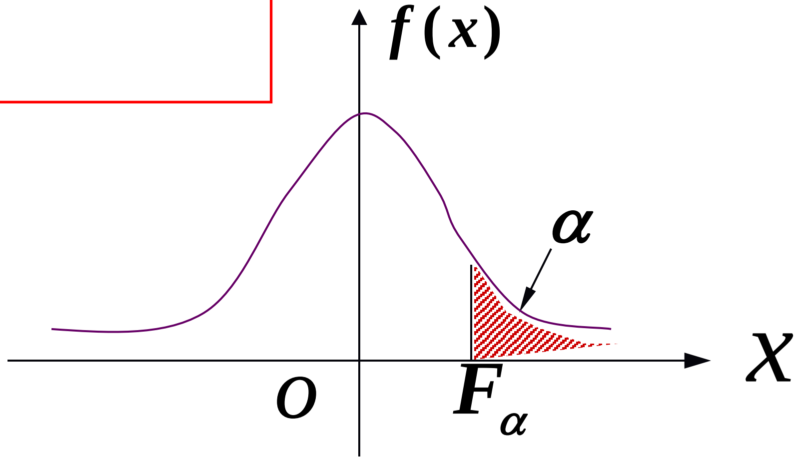
一、常用统计分布的分位数

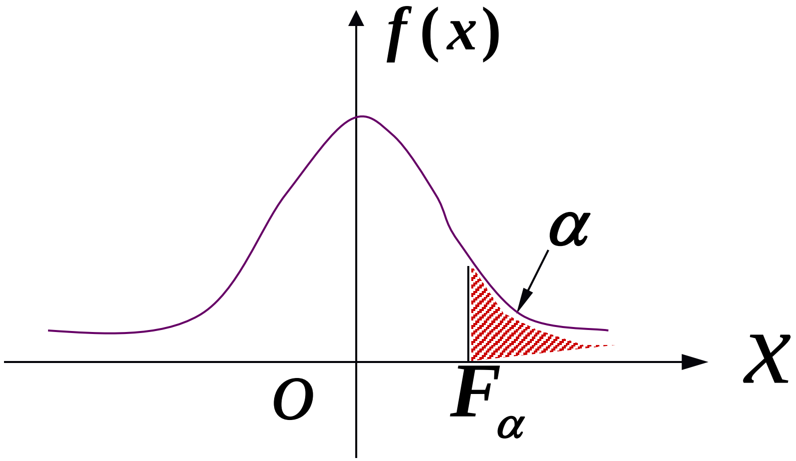
设随机变量X 的分布函数为$F(x)$，对于给定的正数α (0<α<1),若数 F_α 满足

$$P\{X > F_\alpha\} = \alpha,$$

则称 F_α 是此分布的$\alpha-$水平的上侧分位数。

https://www.zhihu.com/question/292047444





上侧分位数与分布函数的关系

$$P\{X \le F_{1-\alpha}\} = \alpha$$

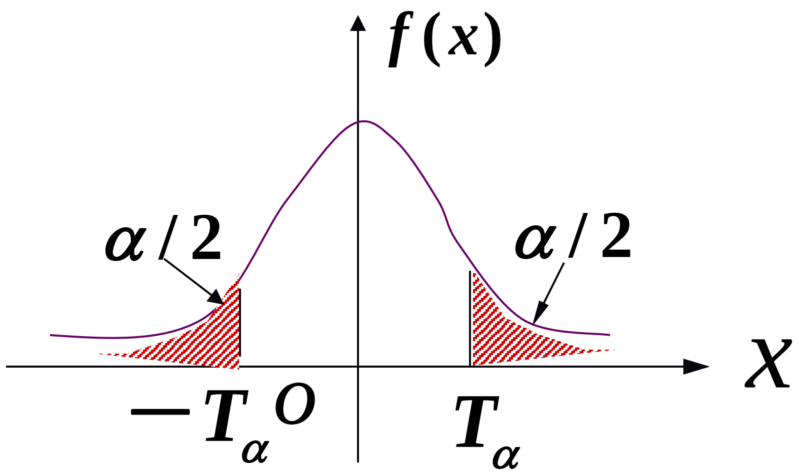
$$P\{F_{1-\alpha/2} < X \le F_{\alpha/2}\} = 1 - \alpha$$

定义

设 X 具有对称密度函数(即该密度函数关于 y 轴对称), 分布函数为 $F(x)$, 对于给定的实数 $0<\alpha<1$, 如果正实数 T_α 满足

$$P\{|X|>T_\alpha\}=\alpha$$

则称 T_α 为 X 的分布 $\alpha-$水平的双侧分位数。



$$F(T_\alpha)=1-\alpha/2$$

$$P\{X>T_\alpha\}=\alpha/2$$

$$T_\alpha=F_{\alpha/2}$$

$$F_\alpha=-F_{1-\alpha}$$

二、 常用统计量的分布

1) χ^2－分布

设$(X_1,\cdots X_n)$为来自于正态总体$N(0,1)$的样本，

则称统计量：$\chi^2 = X_1^2 + \cdots + X_n^2$

所服从的分布为自由度是n的χ^2分布。

记为 $\chi^2 \sim \chi^2(n)$

独立
标准正态分布

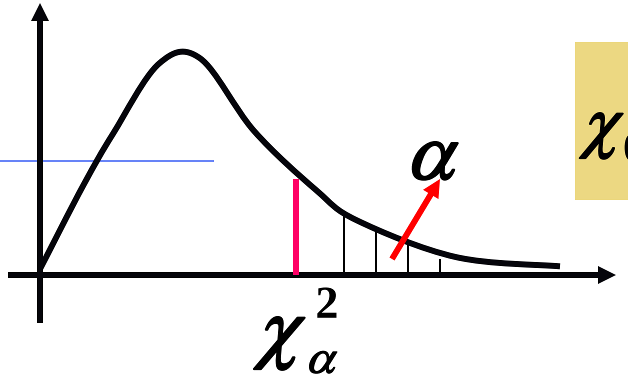
χ^2分布的性质：

1^0若 $X \sim \chi^2(m), Y \sim \chi^2(n)$,且 X，Y 独立，则有

$$X + Y \sim \chi^2(m+n)$$

2^0若 $\chi^2 \sim \chi^2(n)$,则 $E\chi^2 = n$，$D\chi^2 = 2n$.

https://zhuanlan.zhihu.com/p/637865899



例如

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

而查详表可得 $\chi^2_{0.05}(50) = 67.505$.

对于给定的 $\alpha(0 < \alpha < 1)$，称满足条件：

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

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

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

z_α 是标准正态分布的上 α 分位点。

例3 设$(X_1,\cdots X_n)$为来自于正态总体$N(\mu,\sigma^2)$的样本，

则 $\sum\limits_{i=1}^{n}\dfrac{(X_i-\mu)^2}{\sigma^2}\sim$ $\underline{\quad\chi^2(n)\quad}$.

解： $\dfrac{X_i-\mu}{\sigma}\sim N(0,1),\ i=1,\cdots,n,$ 且它们独立.

则 $\sum\limits_{i=1}^{n}\dfrac{(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n).$

例4 $\chi^2_{0.05}(8)=\underline{\ 15.507\ }$, $\chi^2_{0.95}(8)=\underline{\ 2.733\ }$.

$\chi^2_{0.05}(50)=\underline{\ 67.22\ }$, $\chi^2_{0.05}(100)=\underline{\ 124.06\ }$.

3) F - 分布

若 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$, X, Y 独立,

则称随机变量 $F = \dfrac{X/n_1}{Y/n_2}$

所服从的分布为自由度

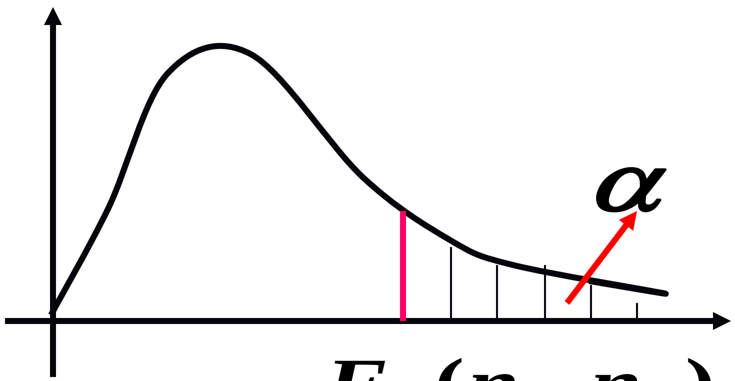
是n_1, n_2 的 F - 分布，记作$F \sim F(n_1, n_2)$.

定理：若 $F \sim F(n_1, n_2)$, 则 $1/F \sim F(n_2, n_1)$.

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

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

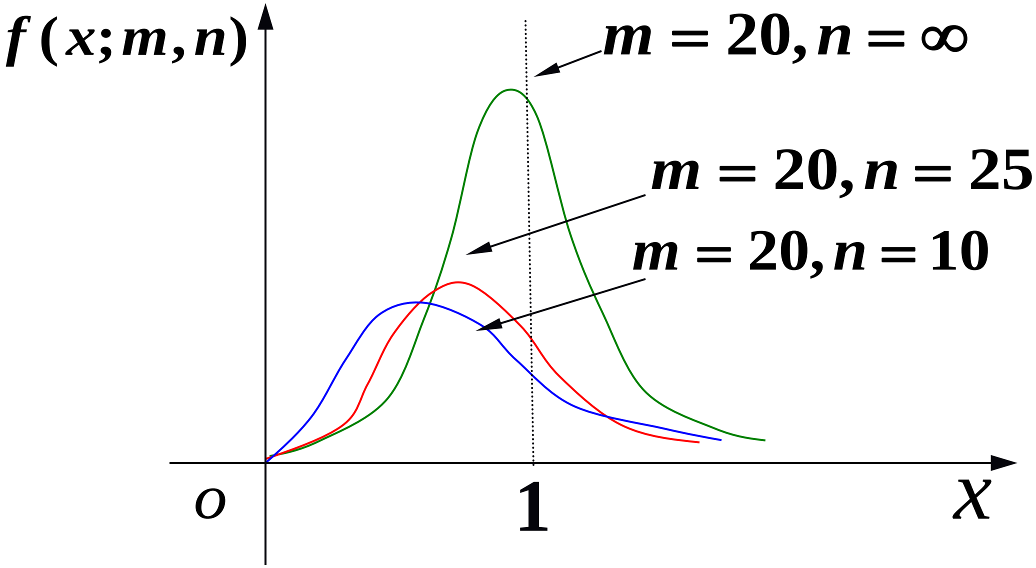
的点$F_{\alpha}(n_1,n_2)$为F分布的上α分位点。



结论：$F_{1-\alpha}(n_1,n_2)=1/F_{\alpha}(n_2,n_1)$

证明：

图像



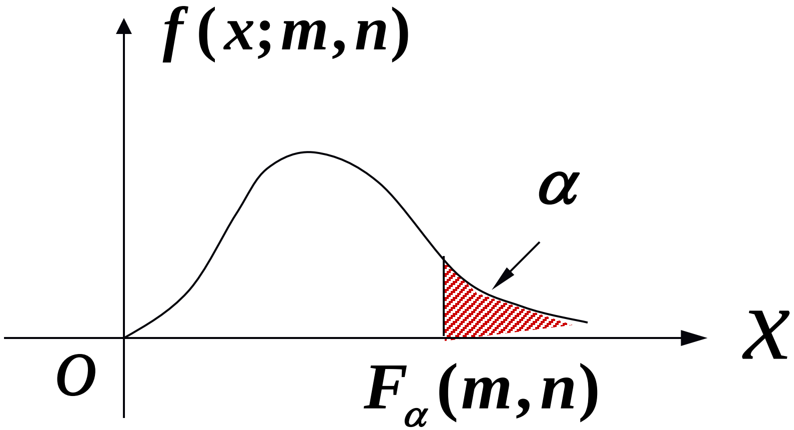
例2

$F(m,n)$ 分布的水平α的上侧分位数记作 $F_\alpha(m,n)$，

即 $P\{F(m,n) > F_\alpha(m,n)\} = \alpha$

$$P\{X > F_{\alpha/2}(m,n)\} \cup P\{X < F_{1-\alpha/2}(m,n)\} = \alpha$$

$$P\{F_{1-\alpha/2}(m,n) < X < F_{\alpha/2}(m,n)\} = 1-\alpha$$



由F分布的定义有 $F_\alpha(m,n) = \dfrac{1}{F_{1-\alpha}(n,m)}$.

例7 $F_{0.05}(8,6) = \underline{4.15}$, $F_{0.95}(8,6) = \underline{\dfrac{1}{F_{0.05}(6,8)} = 0.28}$.

$F_{0.1}(4,7) = \underline{2.96}$, $F_{0.9}(4,7) = \underline{\dfrac{1}{F_{0.1}(7,4)} = 0.25}$.

例8 若 $X \sim F(10,6)$, 且 λ_1 使 $P\{X > \lambda_1\} = 0.05$, 则

$\lambda_1 = \underline{F_{0.05}(10,6) = 4.06}$.

若 $X \sim F(9,8)$, 且 λ_2 使 $P\{X < \lambda_2\} = 0.05$, 则

$\lambda_2 = \underline{F_{0.95}(9,8) = \dfrac{1}{F_{0.05}(8,9)} = 0.31}$.

2) $t-$分布

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

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

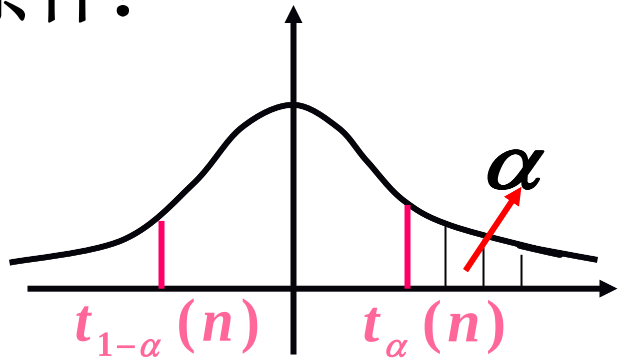
所服从的分布为自由度是n的 $t-$分布，记作$t \sim t(n)$.

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

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}}.$$

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

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



的点$t_\alpha(n)$为t分布的上α分位点。

由概率密度的对称性知： $t_{1-\alpha}(n) = -t_\alpha(n)$

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

例5 $t_{0.05}(8) = \underline{1.8595}$, $t_{0.95}(8) = \underline{-1.8595}$.

$t_{0.05}(50) = \underline{1.645}$, $t_{0.95}(50) = \underline{-1.645}$.

例6 若 $X \sim t(10)$,且 λ_1 使 $P\{X > \lambda_1\} = 0.05$,则

$\lambda_1 = \underline{t_{0.05}(10) = 1.8125}$.

若 $X \sim t(9)$,且 λ_2 使 $P\{X < \lambda_2\} = 0.05$,则

$\lambda_2 = \underline{t_{0.95}(9) = -1.8331}$.

回顾：三大统计分布

1. χ^2 分布

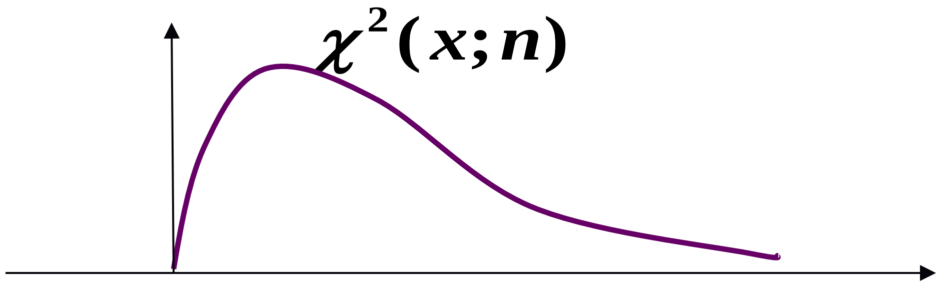
定理1 设 $X_1, X_2, \cdots, X_n$ 独立同服从 $N(0,1)$, 则

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

推论:

1. 设 $X \sim \chi^2(n)$,则 $EX = n, DX = 2n$;

2. 设$X \sim \chi^2(n), Y \sim \chi^2(m)$且独立, 则$X + Y \sim \chi^2(n+m)$.

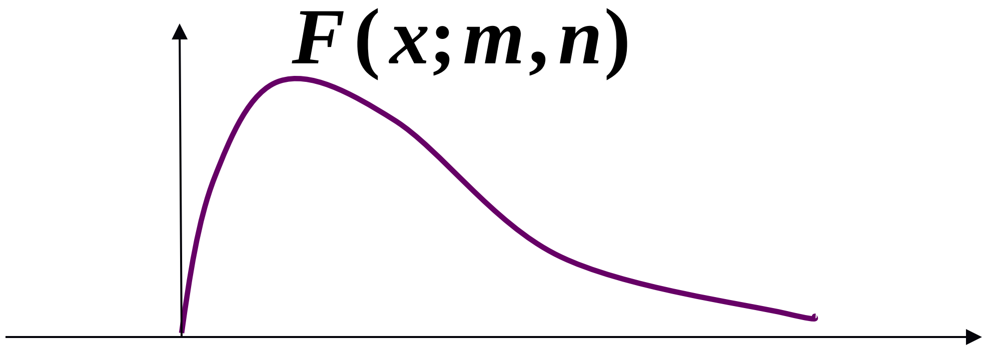


2. F 分布

定理2 设 $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$ 且相互独立, 则

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

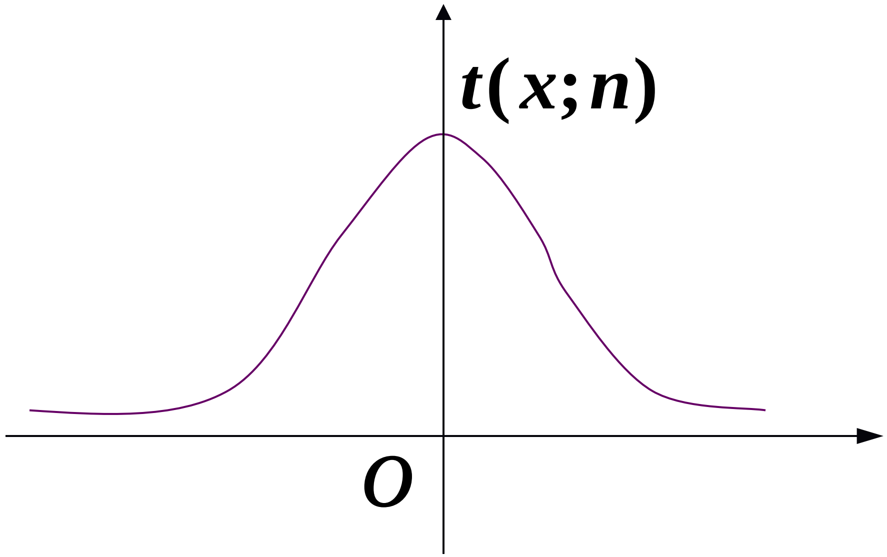
推论 设 $X \sim F(m,n)$, 则 $\frac{1}{X} \sim F(n,m)$.



3. t 分布

定理3 设 $X \sim N(0,1), Y \sim \chi^2(n)$ 且独立, 则

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$



思考？？

$$T^2 = \frac{X^2}{Y/n}$$ 服从什么分布？

$$T^2 = \frac{X^2}{Y/n} \sim F(1, n)$$

例9 已知 $X \sim t(n)$, 试证 $X^2 \sim F(1,n)$.

解: 由于 $X \sim t(n)$, 所以 $X = \dfrac{Y}{\sqrt{Z/n}}$,

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

则 $X^2 = \dfrac{Y^2/1}{Z/n}$,

由 F 分布的定义知 $X^2 \sim F(1,n)$.

作业：

- 4版　147页　6，8；5，7，补充题

4. 设 $(X_1, X_2, \ldots, X_n)$ 是来自总体 X 的简单随机样本，总体方差 $\sigma^2 \neq 0$，$\bar{X}$ 为样本均值，求 $D(X_1 - \bar{X})$．?

5. 设 (X_1, X_2, X_3, X_4) 是来自正态总体 $N(1, \sigma^2)$ 的简单随机样本，试确定随机变量 $\left(\dfrac{X_1 - X_2}{X_3 + X_4 - 2}\right)^2$ 的分布．