

《概率论与数理统计》课程教案

主讲教师_____ 所在单位_____

授课班级_____ 专业_____ 撰写时间_____

教案编号	15-0601	教案内容	6.1 随机样本 6.2 直方图	学时	2
教学目标	基本要求	(1) 会表述总体、个体、抽样、样本及简单随机样本等概念； (2) 理解直方图的概念，了解箱线图的概念 (3) 掌握从总体中获得样本的方法 (4) 掌握使用样本数据绘制直方图的方法			
	能力要求	1. 培养能力要求： a) 掌握概率论和数理统计中的基本概念和性质并能够运用到复杂工程问题的适当表述之中；			
教学重点	总体、总体容量、有限与无限总体、个体、放回抽样与不放回抽样、样本等概念的理解与掌握				
教学难点	同教学重点				
教学方法	讲授、板书、PPT				
工具仪器	多媒体教具、教材、教案、教学课件、考勤表、平时成绩登记表				
教学安排	考勤、复习相关知识点、新课内容概述、组织教学、布置作业、课后小结				
教学过程	教学组织、具体教学内容及教学方法、时间分配及其它说明				备 注
第一部分：旧知识点复习和新课内容概述 （5 分钟） 在前面的五章中，我们学习了概率论的基本内容，从第六章开始我们将讲述数理统计方面的内容。我们首先讲述一下概率论与数理统计之间到底有何区别与联系： 概率论和数理统计的关系 概率论：分析和解决随机现象统计规律的基本理论和方法。 在概率论中，通常研究的是随机变量的概率分布已知情况下的性质、特点和规律性。 然而实际应用中往往是通过大量测量和收集得到的数据来研究随机现象，数理统计知识就是用来解决这方面问题的。 数理统计：以概率论为基本理论，根据试验或观察得到的数据来研究随机现象，对客观规律性作出合理的估计和判断，以解决实际问题。 在数理统计中，随机变量的分布是未知的，或不能完全知道的，人们通过对所研究的随机变量的重复独立的观察，得到许多观察值，对这些数据进行分析，进而对随机变量的分布作出种种推断。					讲授、板书，回顾概率论知识，引出数理统计概念

第二部分：随机样本的概念（60 分钟）

2.1° 总体和样本

定义 在数理统计研究中，常常关心（一批）研究对象的某项数量指标，为此，考虑与这一数量指标联系的随机试验，对这一数量指标进行试验或观察，则将试验的**全部可能的观察值称为总体**。

这些值不一定都不相同，数目上也不一定是有限的。

把每一个可能的观察值称为一个**个体**。总体中所包含的个体的个数称为**总体的容量**。容量为有限的称为有限总体。容量为无限的称为无限总体。

样本空间和总体的区别

样本空间是一次随机试验中的所有可能结果，它不一定是数量的，对一个对象的观察就是一个随机试验，所有可能出现的数量指标 X 满足一定分布。

总体是数理统计中研究大量对象的相关概念（有一堆硬币），研究对象是大量对象的某一个数量指标，一批对象中，每一对象的数量指标值拿过来构成集合称为总体。

如果把样本空间映射到一个随机变量 X ，多数情况下总体是 X 的部分取值的集合（其中取值可以重复），总体中对象取值的情况与 X 的分布相关。

例：考察某大学一年级男生的身高这一试验

研究对象：某大学一年级男生（设 2000 人）

关心的数量指标：身高 H

个 体：每一个男生的身高是一个可能的观察值，形成个体，它是存在的

总 体：所有男生的身高，共含有 2000 个可能观察值，是有限总体

样本空间：任意挑选一男生，其所有可能的身高构成样本空间， X

例：

测量一湖泊中鱼的含汞量——有限总体（对象(鱼)的个数有限）

测量一湖泊中任一地点的深度——无限总体（连续的）

2.2° 总体与随机变量

总体中的每一个个体是一个随机试验的观察值，因此它是某一随机变量 X 的值（对一个对象观察试验而来），**这样一个总体对应于一个随机变量 X** 。

对总体的研究就是对相应的随机变量 X 的研究， X 的分布函数和数字特征就是总体的分布函数和数字特征。

含义：总体的分布不是这些个体构成的空间的分布（这些取值往往是确定存在的），而是指每一个个体的取值所来自的随机变量 X 的分布

今后将不区分总体和相应的随机变量，笼统称为总体 X

板书、讲授

基本概念

难点，区分样本空间和总体的概念

在总体与随机变量之间建立联系

数理统计

例：一批灯泡的寿命构成一个有限总体，一个灯泡的寿命可以用一个符合指数分布的随机变量 X 来描述，而这批灯泡中每一个灯泡的寿命的取值即个体是来自于符合指数分布的随机变量 X 的取值

例：检查生产的一批零件是正品还是次品问题，正品表示为 0，次品表示为 1，那么对应于所有生产出来的零件的取值构成总体，其中有若干个零件为正品 0，若干个为次品 1。现在生产一个零件是正品或次品的情况可用一个符合(0—1)分布的随机变量 X 来描述，设次品率为 p ，则 X 的分布为

X	0	1
p_k	$1-p$	p

可知这批零件中每一个零件的关于正品或次品的取值来自于随机变量 X ，而且在这批总体中，取 1 的个体的个数与总体的个数之比，当总体容量很大时应接近 p ，这时的总体可近似看作无限总体。

因此无限总体的特性更接近 X 的分布，无限总体是人们对具体事务的抽象，分布形式较为简明，便于数学处理，是研究的主要对象

需要说明的是在实际中大量的总体都是有限总体，这与人们通常考察某一范围内的个体有关，但如果数量足够大，可以近似为无限总体，或抽象为无限总体的情况。因为个体少量的有限总体与其所对应的 X 的真实分布一般相差很大，比如一个总数为 2 的总体（灯泡的寿命），很难从中看出它与指数分布有什么关系。

2.3° 样本

实际中，总体的分布往往是未知的，或只知道该分布具有某种形式，但包含未知参数。比如我们知道总体符合正态分布，但不知道期望和方差两个参数。在数理统计中要解决的问题之一就是如何推断总体的分布和性质。而且即使我们掌握了所有个体的取值情况，那么由于这些取值可能仅是对应的随机变量 X 的部分取值，很难准确给出判定（比如连续型变量）。

在数理统计中人们通常从总体中抽取一部分个体，根据所获得的数据来对总体的分布和性质得出推断的。被抽取的部分个体叫做总体的一个样本。

特别是对于以下情况更有意义：

无限总体

有破坏性的试验，如灯泡寿命测试，炮弹的可靠性等

有时间或空间上的限制，如观察或测试耗时太多，可能使工作无意义了等情况，采用样本来考察总体的办法是十分必要和有效的。

2.4° 与样本相关的问题

样本的描述：从总体中抽取一个个体，就是对总体 X 进行一次观察，并记录其结果。在相同条件下对总体 X 进行 n 次重复的、独立的观察，将 n 次观察结果按试

与概率论
区别的一个表现

PPT

在实际中
研究总体

讲授、板
书

引入样本
概念的意义

讲授、板

验次序记为 $X_1, X_2, \dots, X_n$。 对于每一次观察，其可能结果构成的随机变量 X_i 与总体 X 是相同的，比如考察一个灯泡的寿命值，该考察可能出现的结果对应的随机变量就是总体 X。 $X_1, X_2, \dots, X_n$ 都是对 X 的观察的结果，且各次观察在相同条件下独立进行，所以 $X_1, X_2, \dots, X_n$ 相互独立且与总体 X 具有相同分布的随机变量。这样 $X_1, X_2, \dots, X_n$ 称为来自总体 X 的一个简单随机样本，n 称为这个样本的容量。 n 次观察一经完成，得到一组实数 $x_1, x_2, \dots, x_n$，它们依次为随机变量 $X_1, X_2, \dots, X_n$ 的观察值，称为样本值。 简单随机样本的获得： 对于有限总体，采用放回抽样（相互独立），就能得到简单随机样本，但不方便，每次要放回搅匀。 当个体总数 N 比要得到的样本的容量 n 大得多时，可将不放回抽样近似当作放回抽样处理。 对于无限总体，因抽取一个个体不影响它的分布，所以总是用不放回抽样。 例：在生产中每隔一定时间抽取一个个体，抽取 n 个就得到一个简单随机样本。 试制新产品得到的样品的质量指标，也常被认为是样本。 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 的、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，他们的观察值 $x_1, x_2, \dots, x_n$ 称为样本值，又称为 X 的 n 个独立的观察值。 样本看成是一个随机向量，写成 $(X_1, X_2, \dots, X_n)$，此时样本值相应的写成 $(x_1, x_2, \dots, x_n)$，一个样本可以有多个不同的样本值。 由定义得：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 $f^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$ 第 3 部分：直方图（20 分钟） 为了研究总体分布的性质，人们通过试验得到许多观察值，一般来说这些数据是杂乱无章的。为了利用它们进行统计分析，将这些数据加以整理，还常借助于表格或图形对它们加以描述。本节将通过例子对连续型随机变量 X 引入“频率直方	书 实际中获得样本的方法 样本的数学定义 有样本性质导出的样本分布函数和概率密度（分布律）函数 PPT 虽不为重
--	--

图”，简称“直方图”。

例 下面列出了 84 个伊特拉斯坎(Etruscan)人男子的头颅的最大宽度(mm)，现在来画这些数据的“频率直方图”。

141 148 132 138 154 142 150 146 155 158 150 140 147
148
144 150 149 145 149 158 143 141 144 144 126 140 144
142
141 140 145 135 147 146 141 136 140 146 142 137 148
154
137 139 143 140 131 143 141 149 148 135 148 152 143
144
141 143 147 146 150 132 142 142 143 153 149 146 149
138
142 149 142 137 134 144 146 147 140 142 140 137 152
145

这些数据杂乱无章，先要将它们进行整理。最小值和最大值分别是 126，158，即所有数据都落在区间[126，158]上，现取区间[124.5，159.5]能覆盖上述区间，并将[124.5，159.5]等分为 7 个小区间，等分区间长度不宜过小，以免小区间内频率为 0， n 较大时等分区间数 k 取 10 到 20， $n < 50$ 时取 5 到 6。小区间的长度记为 Δ ， $\Delta = (159.5 - 124.5) / 7 = 5$ 。 Δ 称为组距，小区间的端点称为组限。数出落在每个小区间内的数据的频数 f_i ，算出频率 f_i/n ($n=84$ ， $i=1, 2, \dots, 7$) 如下表：

组限	频数 f_i	频率 f_i/n	累积频率
124.5~129.5	1	0.0119	0.0119
129.5~134.5	4	0.0476	0.0595
134.5~139.5	10	0.1191	0.1786
139.5~144.5	33	0.3929	0.5715
144.5~149.5	24	0.2857	0.8572
149.5~154.5	9	0.1071	0.9643
154.5~159.5	3	0.0357	1

现在自左至右依次在各个小区间上作以 $(f_i/n)/\Delta$ 为高的小矩形。如图 6—1 所示这样的图形叫频率直方图。显然这种小矩形的面积就等于数据落在该小区间的频率 f_i/n 。由于当 n 很大时，频率就接近于概率，因而一般来说，每个小区间上的小矩形面积接近于概率密度曲线之下该小区间上的曲边梯形的面积。于是，一般来说，直方图的外廓曲线接近于总体 X 的概率密度曲线。从本例的直方图看(6-1)，它有一个峰，中间高，两头低，比较对称，看起来很像来自于某一正态总体 X 。从直方图上还可以估计 X 落在某一区间的概率，例如从图上看到有 51.2% 的人最大头颅宽度在区间

点，但在实际应用中很重要。箱线图有兴趣可自学

讲授、板书

(134.5, 144.5)之间, 最大头颅宽度小于 129.5 的仅占 1.1%等等。

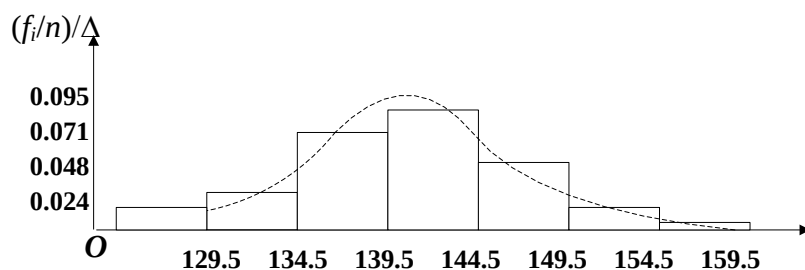


图 6—1

课后作业

P147: 2, 3

课后小结

数据统计：对总体（随机变量）的分布做出推断的数学工具。

总体、个体的概念。实际中大多为有限总体，当总体容量足够大时可近似为无限总体。

总体和样本空间是有区别的，总体中个体取值可相同。

由于总体与变量的对应关系，在数理统计中研究对象从概率论中的变量变为总体。

获得样本必须确定各个样本点是独立同分布的。

直方图在实际中通常用来近似获得总体（随机变量）的概率密度。

《概率论与数理统计》课程教案

主讲教师_____ 所在单位_____

授课班级_____ 专业_____ 撰写时间_____

教案编号	16-0602	教案内容	6.3 抽样分布	学时	2
教学目标	基本要求	(1) 理解统计量的概念，熟练掌握样本均值、样本方差、样本原点矩、样本中心矩的概念和性质； (2) 理解抽样分布的定义，能表述三大抽样分布的概念和简单性质； (3) 熟练表述正态总体的样本均值和样本方差的分布。			
	能力要求	1. 培养能力要求： a) 掌握概率论和数理统计中的基本概念和性质并能够运用到复杂工程问题的适当表述之中； b) 能够根据试验或观察得到的数据来研究随机现象，运用参数估计和假设检验等基本的统计方法，对客观规律性做出合理的估计和判断，以解决实际问题。			
教学重点	统计量的概念、样本均值、样本方差、样本矩的概念 样本矩（的连续函数）依概率收敛于总体矩（的连续函数） 常用统计量的分布： χ^2 、t、F 统计量的构造及分布 上 α 分位点，正态总体样本均值与样本方差及其函数的分布				
教学难点	同教学重点				
教学方法	讲授、板书、PPT				
工具仪器	多媒体教具、教材、教案、教学课件、考勤表、平时成绩登记表				
教学安排	考勤、复习相关知识点、新课内容概述、组织教学、布置作业、课后小结				
教学过程	教学组织、具体教学内容及教学方法、手段、时间分配及其它说明				备 注
第一部分：旧知识点复习与新课内容概述（3 分钟） 上一节课我们在了解总体、个体的基础上引出样本的概念，在实际中为了对某个总体的分布做成推断，首先需要获得样本。 利用样本来进行统计推断或估计，往往不是直接使用样本本身，而是针对不同的问题构造样本的适当函数，利用这些样本函数进行统计推断。样本的函数及其分布就成为进行统计推断的重要工具，这正是本节课的内容。 第二部分：统计量（50 分钟） 2.1 统计量的概念： 定义 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$ 是 $X_1, X_2, \cdots, X_n$ 的函数，					若含有未

若含有未

若 g 中不含未知参数, 则称 $g(X_1, X_2, \dots, X_n)$ 是一统计量。

显然统计量是一随机变量, 因其是随机变量 $X_1, X_2, \dots, X_n$ 的函数。

设 $x_1, x_2, \dots, x_n$ 是相应于 $X_1, X_2, \dots, X_n$ 的一组样本值, 则称 $g(x_1, x_2, \dots, x_n)$ 是 $g(X_1, X_2, \dots, X_n)$ 的观察值。

一组常用的统计量:

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

定义:

$$(1) \text{ 样本平均值 } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{相应的观察值为 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

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

$$\text{相应的观察值为 } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

$$(3) \text{ 样本标准差 } S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{观察值 } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$(4) \text{ 样本 } k \text{ 阶 (原点) 矩 } A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k=1, 2, \dots, \quad \text{观察值 } a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k=1, 2, \dots$$

$$(5) \text{ 样本 } k \text{ 阶中心矩 } B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k=2, 3, \dots, \quad b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \quad k=2, 3, \dots$$

注意在中心矩中, $k=1$ 时矩为 0, 没有意义, 观察值仍分别称为样本均值, 样本方差, 样本标准差, 样本 k 阶矩, 样本 k 阶中心矩。

下一章内容矩估计法的理论根据:

若总体 X 的 k 阶矩 $E(X^k)$ 记成 μ_k 存在, 则由辛钦定理, 当 $n \rightarrow \infty$ 时, $A_k \xrightarrow{P} \mu_k, \quad k=1, 2, \dots,$

这是因为 $X_1, X_2, \dots, X_n$ 相互独立且与 X 同分布, 所以 $X_1^k, X_2^k, \dots, X_n^k$ 相互独立且与 X^k 同分布, 固有 $E(X_1^k) = E(X_2^k) = \dots = E(X_n^k) = \mu_k$

而样本 k 阶矩 A_k 为 n 个独立同分布的随机变量 $X_1^k, X_2^k, \dots, X_n^k$ 的算术平均值, 由辛钦定理依概率收敛于 X^k 的数学期望 μ_k 即

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

知参数则其分布不确定

除以 $n-1$ 是为了确保作为总体方差估计的无偏性 (下一章内容)

样本 2 阶中心矩不等于样本方差!

样本矩 (的连续函数) 依概率收敛于总体矩 (的连续函数)

由依概率收敛的性质知道 $g(A_1, A_2, \dots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \dots, \mu_k)$, g 为连续函数

2.2 与总体分布函数 $F(x)$ 相应的统计量——经验分布函数

设 $X_1, X_2, \dots, X_n$ 是总体 F 的一个样本, 用 $S(x)$ 表示 $X_1, X_2, \dots, X_n$ 中不大于 x 的随机变量的个数, 定义经验函数为

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

这与分布函数的含义是相似的, 不超过 x 的变量的个数占变量总数的概率, 就对应于随机变量取值落在 $-\infty$ 到 x 内的概率

对于一个样本值, 经验分布函数 $F_n(x)$ 的观察值(仍以 $F_n(x)$ 表示)是很容易得到的, 例如:

设 X_1, X_2, X_3 是来自于总体 F 的一个简单随机样本

(1) 设样本的一组样本值为 1, 2, 3, 则经验分布函数 $F_3(x)$ 的观察值为

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

(2) 设得到的一组样本值为 1, 1, 2, 则经验分布函数 $F_3(x)$ 的观察值为

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

一般, 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. 设为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, 则经验分布函数 $F_n(x)$ 的观察值为

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

对于经验分布函数 $F_n(x)$, 格里汶科(Glivenko)在 1933 年证明了以下的结果

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

对于任意实数 x , 当 n 充分大时经验分布函数的任一观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别, 从而在实际上可以当作 $F(x)$ 来使用

作为 $F(x)$ 的统计量 $F_n(x)$ 其取值的数学期望等于 $F(x)$

这是因为, 对于任意的 x , $-\infty < x < \infty$, $S(x)$ 是随机变量落在 $-\infty$ 到 x 的次数, 引入随机变量 X_i , 当取值落在 $-\infty$ 到 x 内时为 1, 否则为 0, 则由于总体 X 的分布函数知 X_i 满足 0—1 分布, 参数为 $F(x)$

略降其概念, 了解即可

$S(x)=X_1+X_2+\cdots+X_n$ 正好满足二项分布 $\sim b(n, F(x))$

于是 $E(F_n(x))=1/nE(S(x))=1/n(nF(x))=F(x)$

可见这种估计的合理性。

2.3 几个常用统计量的分布

统计量是随机样本的函数，因此也是一个随机变量，统计量的分布称为抽样分布，在使用统计量进行统计推断时需要知道它的概率分布。

当总体分布函数已知时抽样分布是确定的，要求出统计量的精确分布往往是困难的。本节介绍来自正态总体的几个常用统计量的分布。

(一) χ^2 分布

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

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

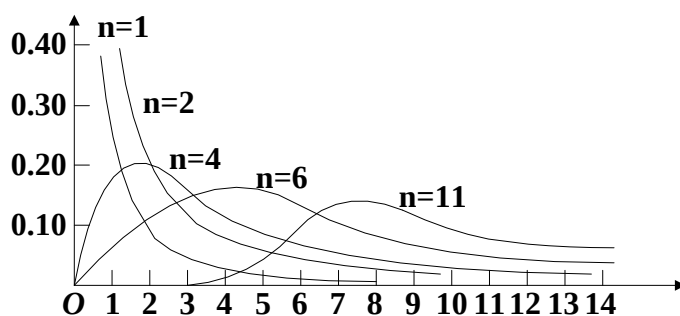
服从自由度为 n 的 χ^2 分布，记为 $\chi^2 \sim \chi^2(n)$

自由度是指右端独立变量的个数

$\chi^2(n)$ 分布的概率密度为

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

图形如下



现在来推导概率密度函数

1° 先计算 $\chi^2(1)$ 的分布即 $\chi^2(1) = X_1^2$ 的分布

此时 X_1 来自正态总体，所以 $X_1 \sim N(0,1)$

由第二章 § 5 例 3，64 页上方，当 $X \sim N(0,1)$ 时， $Y=X^2$ 的概率密度函数为

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

此即 $\chi^2(1)$ 的概率密度函数

又由第三章 § 5 例 3, 97 页下方, 满足 Γ 分布的随机变量: $X \sim \Gamma(\alpha, \beta)$

$$f_X(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & \text{其它} \end{cases} \quad \alpha > 0, \beta > 0$$

Γ 分布的可加性: $X \sim \Gamma(\alpha_1, \beta), Y \sim \Gamma(\alpha_2, \beta)$, X 和 Y 相互独立

则 $Z = X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$

在 $\chi^2(1)$ 分布中, $\Gamma(1/2) = \sqrt{\pi}$ 代入并变形得到概率密度函数为

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

比较两式, 此即参数为 $(1/2, 2)$ 的 Γ 分布, 即 $\chi^2(1) \sim \Gamma(1/2, 2)$

即对任意的 $X_i^2 \sim \Gamma(1/2, 2)$

2° 自由度大于 1 时

由于 $X_1, X_2, \dots, X_n$ 相互独立, 所以 $X_1^2, X_2^2, \dots, X_n^2$ 也相互独立, 从而由 Γ 分布的可

加性, $\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2 \sim \Gamma(n/2, 2)$

即 $\chi^2(n)$ 分布是 Γ 分布的特例

χ^2 分布的性质

1) 可加性: 由于 χ^2 分布是一种特殊的 Γ 分布, 因而具有可加性

即设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 并且 χ_1^2 和 χ_2^2 独立

则 $\chi^2 = \chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$

2) χ^2 分布的数学期望和方差

若 $\chi^2 \sim \chi^2(n)$, 则有

$$E(\chi^2) = E\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n E(X_i^2) = \sum_{i=1}^n [D(X_i) + E^2(X_i)] = \sum_{i=1}^n [1 + 0^2] = n$$

其中 $X_i \sim N(0, 1)$, 且有方差计算的恒等式变形

而由各独立变量之间的独立性

$$D(\chi^2) = D\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n D(X_i^2) = \sum_{i=1}^n [E(X_i^4) - E^2(X_i^2)]$$

而 $E(X_i^4) = \int_{-\infty}^{\infty} x_i^4 \varphi(x_i) dx_i = \int_{-\infty}^{\infty} x_i^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} dx_i = 3$, $E(X_i^2) = D(X_i) + E^2(X_i) = 1$

χ^2 分布的
独立可加
性在证明
题中十分
有用

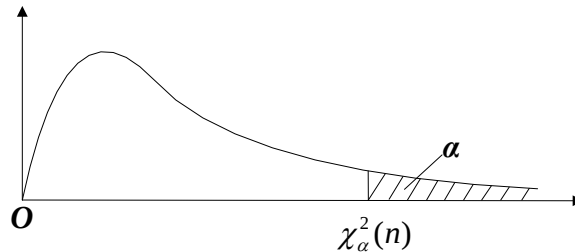
代入上式得 $D(\chi^2)=2n$

要求记忆

3) χ^2 分布的上 α 分位点

对于给定正数 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 , 点 $\chi_\alpha^2(n)$ 的值可查表,

如: $\alpha=0.1$, $n=25$ 时, $\chi_{0.1}^2(25) = 34.382$

表中仅给出了 $n \leq 45$ 时的情况, 费舍尔(P.A.Fisher)曾经证明, 当 n 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, 由此可求得 $n > 45$ 时 $\chi^2(n)$ 分布上 α 分

位点, 例如 $\chi_{0.05}^2(50) \approx \frac{1}{2}(z_{0.05} + \sqrt{2 \times 50 - 1})^2 = \frac{1}{2}(1.645 + \sqrt{99})^2 = 67.221$

由详表查得 $\chi_{0.05}^2(50) = 67.505$, 非常接近

(二) t 分布, 也称学生氏 (Student) 分布

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

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

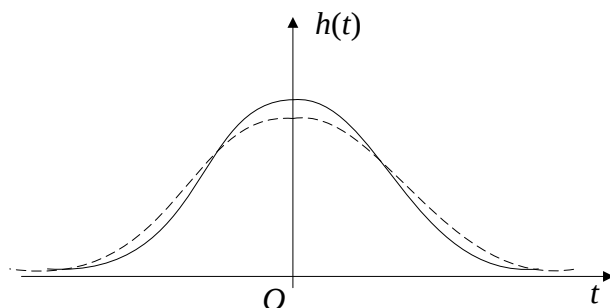
服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$

$t(n)$ 分布的概率密度为

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

图形如下

t 分布的
pdf 不要
求记忆



$t=0$ 为对称轴, 当 $n \rightarrow \infty$ 时, 图形类似于标准正态分布概率密度函数的图形,

实际上, 利用 r 函数的性质有, $\lim_{n \rightarrow \infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, 当 n 足够大时, t 分布近似于

$N(0,1)$

从另一个角度, 我们看 $t = \frac{X}{\sqrt{Y/n}}$, 由于 Y/n 相当于平方和的算术平均值,

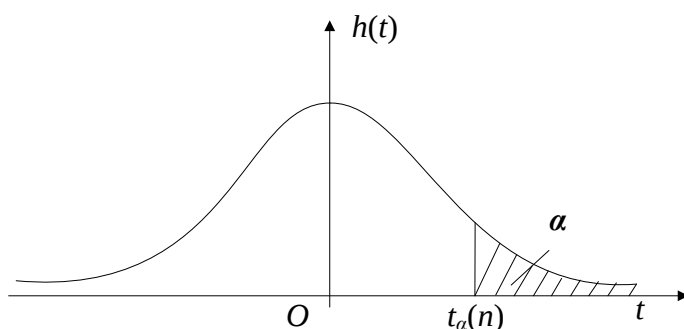
当 n 充分大时, 该值趋于一个稳定的常数, 即 Y/n 的数学期望, 由 Y 的分布特性值 Y/n 的期望为 1, 即当 n 充分大时分母趋于一个稳定的常数 1, 从而 t 无限接近 $X \sim N(0,1)$, 当 n, t 较小时, 两种分布相差较大

概率密度的求解方法, 先求解 $U = \sqrt{Y/n}$ 的分布函数, 再求解 $t = X/U$ 的分布函数, 比较繁琐。

t 分布的上 α 分位点

对于给定正数 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



由 t 分布上 α 分位点的定义及 $h(t)$ 图形的对称性, 它有和标准正态分布的上 α 分位点相同的性质, 即 $t_{1-\alpha}(n) = -t_\alpha(n)$

当 $n > 45$ 时, 对于常用的 α 值就用标准正态分布的上 α 分位点来近似

即 $t_\alpha(n) \approx z_\alpha$

(三) F 分布

设 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 且 U 和 V 独立, 则称随机变量

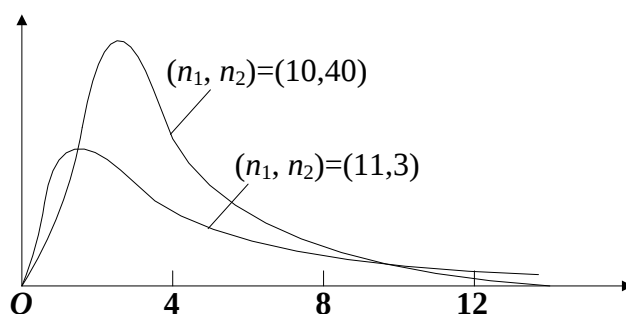
$$F = \frac{U/n_1}{V/n_2}$$

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

$F(n_1, n_2)$ 分布的概率密度为

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

图形如下



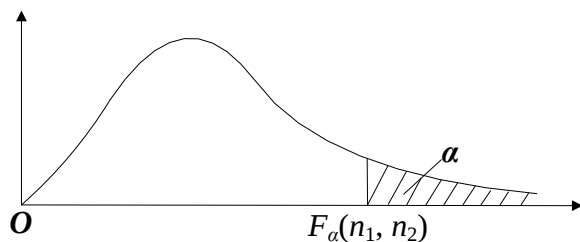
由定义可知, 若 $F \sim F(n_1, n_2)$, 则

$$\frac{1}{F} \sim F(n_2, n_1),$$

F 分布的上 α 分位点

对于给定正数 $\alpha, 0 < \alpha < 1$, 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, 有表可查



F 分布的上 α 分位点有如下重要性质

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

常用来求 F 分布表中未列出的常用上 α 分位点

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

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

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

再由 $1/F \sim F(n_2, n_1)$ 知 $P\{1/F > F_\alpha(n_2, n_1)\} = \alpha$

Pdf 不要
求记忆

该性质在
解题中很
重要

比较两式的得证。

例： $F_{0.95}(12,9)=1/F_{0.05}(9,12)=1/2.80=0.357$

第三部分 正态总体的样本均值与样本方差及其函数的分布（32 分钟）

设任意总体 X (不管服从什么分布, 只要均值和方差存在) 的均值为 μ , 方差为 σ^2 , $X_1, X_2, \dots, X_n$ 是来自 X 的一个样本, $\bar{X}$, S^2 是样本均值和样本方差, 则总有

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \mu, \quad D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \sigma^2/n,$$

注意到 $X_1, X_2, \dots, X_n$ 是相互独立的情况, 不难证明

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right) = \frac{1}{n-1} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2)\right) = \sigma^2 \\ \text{即 } E(S^2) &= \sigma^2 \end{aligned}$$

进而设总体 $X \sim N(\mu, \sigma^2)$, 由于 $X_1, X_2, \dots, X_n$ 相互独立, 均服从同一正态分布, 而相互独立的正态变量的任意线性变换仍符合正态分布, 则 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 恰是 n 个独立正态

变量的线性组合, 也服从正态分布, 此即

定理一 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本, $\bar{X}$ 是样本均值, 则有 $\bar{X} \sim N(\mu, \sigma^2/n)$

对于正态总体 $N(\mu, \sigma^2)$ 的样本均值 $\bar{X}$ 和样本方差 S^2 有以下两个重要定理:

定理二 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本, $\bar{X}$, S^2 分别是样本均值和样本方差, 则有

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

$$2^\circ \bar{X} \text{ 与 } S^2 \text{ 独立。}$$

证明见附录。

定理三 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本, $\bar{X}$, S^2 分别是样本均值和样本方差, 则有

定理 2 对一般分布是不成立的, 证明中用到正态变量相互独立与两两不相关等价的

性质，可
自学

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

证：由定理一及定理二知

$\bar{X} \sim N(\mu, \sigma^2/n)$ ，所以 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ ， $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ ，且两者独立，由 t 分布

的定义 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} / \sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)} \sim t(n-1)$ 化简左式即得证明

对于两个正态总体的样本均值和样本方差有以下定理

定理四 设 $X_1, X_2, \dots, X_{n_1}$ 与 $Y_1, Y_2, \dots, Y_{n_2}$ 分别来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 的样本，且这两个样本相互独立。

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样本均值；

$S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$ ， $S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$ 分别是这两个样本的样本方差，则

有

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

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

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

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

证：1° 由定理二

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

由假设知 S_1^2 ， S_2^2 相互独立，则由 F 分布的定义知

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

化简后即所证结果

2° 由定理一，易知

$\bar{X} \sim N(\mu_1, \sigma^2/n_1)$, $\bar{Y} \sim N(\mu_2, \sigma^2/n_2)$, 于是它们的线性组合

$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2/n_1 + \sigma^2/n_2)$, 则其标准化变量

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

又由给定条件知

$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1)$, $\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$, 且它们相互独立, 由 χ^2 分布的

可加性知, $V = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$, 由附录 2° U, V 独立, 从而按 t

分布的定义知

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

证毕。

例: (习题五) 已知 $X \sim t(n)$, 求证 $X^2 \sim F(1, n)$

证: 由 $t(n)$ 分布定义, $X = \frac{U}{\sqrt{V/n}}$, 其中 $U \sim N(0, 1)$, $V \sim \chi^2(n)$, U, V 独立, 从而

$X^2 = \frac{U^2}{V/n}$, 由于 $U^2 \sim \chi^2(1)$, 而 $V \sim \chi^2(n)$,

所以 $X^2 = \frac{U^2}{V/n} = \frac{U^2/1}{V/n} \sim F(1, n)$

证毕

课后作业

P147: 4、5、6、9

课后小结

统计量仅仅是样本的函数, 不能包含未知参数。

样本方差与样本二阶中心矩的区别。

样本矩 (的连续函数) 依概率收敛于总体矩 (的连续函数)。

χ^2 、 t 、 F 统计量的构造及其分布, 上 α 分位点。

正态总体样本均值与样本方差及其函数的分布中的 4 个定理是随后章节参数估计与假设检验的基础。

