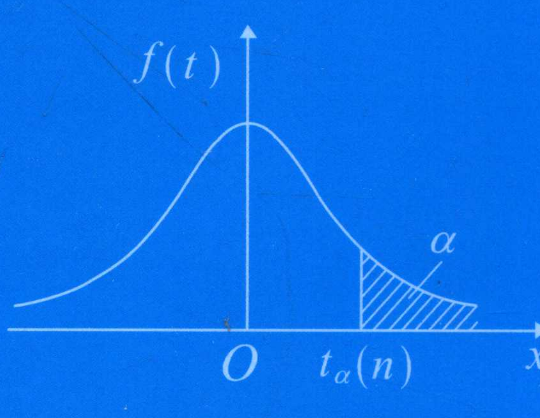
普通高等教育“十二五”规划教材
新世纪新理念高等院校数学教学改革与教材建设精品教材

丛书主编：朱长江 彭双阶
执行主编：何 穗

数理统计

SHULI TONGJI

陈应保 龙兵 周涌◎主编



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

华中师范大学出版社

普通高等教育“十二五”规划教材
新世纪新理念高等院校数学教学改革与教材建设精品教材

数理统计

主　编:陈应保　龙　兵　周　涌

华中师范大学出版社

内 容 提 要

《数理统计》是研究随机现象数量规律的一门学科。全书共分5章，内容包括：数据整理与抽样分布、参数估计、假设检验、统计决策的基本理论、回归分析与方差分析。书中各章配有适量的习题，本书基本概念叙述清晰，循序渐进，内容全面，应用性强，符合大学本科对本门课程的教学要求和实际需要。

本书可作为数学类各专业的本科教材，也可以供科研人员以及工程技术人员参考使用。

新出图证(鄂)字10号

图书在版编目(CIP)数据

数理统计/陈应保 龙兵 周涌 主编. —武汉：华中师范大学出版社，2014.4

(普通高等教育"十二五"规划教材/新世纪新理念高等院校数学教学改革与教材建设精品教材)

ISBN 978-7-5622-6519-1

Ⅰ.①数… Ⅱ.①陈… ②龙… ③周… Ⅲ.①数理统计—高等学校—教材 Ⅳ.①O212

中国版本图书馆CIP数据核字(2014)第018442号

数理统计

©陈应保 龙兵 周涌 主编

编辑室：第二编辑室 **电话**：027－67867362

责任编辑：张方毅 袁正科 **责任校对**：易 雯 **封面设计**：胡 灿

出版发行：华中师范大学出版社

社址：湖北省武汉市珞喻路152号 **邮编**：430079

销售电话：027－67863426/67863280(发行部) 027－67861321(邮购) 027－67863291(传真)

网址：http://www.ccnupress.com **电子信箱**：hscbs@public.wh.hb.cn

印刷：湖北新华印务有限公司 **督印**：章光琼

开本：787 mm×1092 mm 1/16 **印张**：14 **字数**：318千字

版次：2014年5月第1版 **印次**：2014年5月第1次印刷

印数：1—1 500 **定价**：26.00元

欢迎上网查询、购书

普通高等教育“十二五”规划教材
新世纪新理念高等院校数学教学改革与教材建设精品教材

丛书编写委员会

丛书总序

未来社会是信息化的社会，以多媒体技术和网络技术为核心的信息技术正在飞速发展，信息技术正以惊人的速度渗透到教育领域中，正推动着教育教学的深刻变革。在积极应对信息化社会的过程中，我们的教育思想、教育理念、教学内容、教学方法与手段以及学习方式等方面已不知不觉地发生了深刻的变革。

现代数学不仅是一种精密的思想方法、一种技术手段，更是一个有着丰富内容和不断向前发展的知识体系。《国家中长期教育改革和发展规划纲要(2010—2020年)》指明了未来十年高等教育的发展目标："全面提高高等教育质量"、"提高人才培养质量"、"提升科学研究水平"、"增强社会服务能力"、"优化结构办出特色"。这些目标的实现，有赖于各高校进一步推进数学教学改革的步伐，借鉴先进的经验，构建自己的特色。而数学作为一个基础性的专业，承担着培养高素质人才的重要作用。因此，新形势下高等院校数学教学改革的方向、具体实施方案以及与此相关的教材建设等问题，不仅是值得关注的，更是一个具有现实意义和实践价值的课题。

为推进教学改革的进一步深化，加强各高校教学经验的广泛交流，构建高校数学院系的合作平台，华中师范大学数学与统计学学院和华中师范大学出版社充分发挥各自的优势，由华中师范大学数学与统计学学院发起，诚邀华中和周边地区部分颇具影响力的高等院校，面向全国共同开发这套"新世纪新理念高等院校数学系列精品教材"，并委托华中师范大学出版社组织、协调和出版。我们希望，这套教材能够进一步推动全国教育事业和教学改革的蓬勃兴盛，切实体现出教学改革的需要和新理念的贯彻落实。

总体看来，这套教材充分体现了高等学校数学教学改革提出的新理念、新方法、新形式。如目前各高等学校数学教学中普遍推广的研究型教学，要求教师少

讲、精讲，重点讲思路、讲方法，鼓励学生的探究式自主学习，教师的角色也从原来完全主导课堂的讲授者转变为学生自主学习的推动者、辅导者，学生转变为教学活动的真正主体等。而传统的教材完全依赖教师课堂讲授、将主要任务交给任课教师完成、学生依靠大量的被动练习应对考试等特点已不能满足这种新教学改革的推进。如果再叠加脱离时空限制的网络在线教学等教学方式带来的巨大挑战，传统教材甚至已成为教学改革的严重制约因素。

基于此，我们这套教材在编写的过程中注重突出以下几个方面的特点：

一是以问题为导向、引导研究性学习。教材致力于学生解决实际的数学问题、运用所学的数学知识解决实际生活问题为导向，设置大量的研讨性、探索性、应用性问题，鼓励学生在教师的辅导、指导下于课内课外自主学习、探究、应用，以加深对所学数学知识的理解、反思，提高其实际应用能力。

二是精选内容、逻辑清晰。整套教材在各位专家充分研讨的基础上，对课堂教学内容进一步精炼浓缩，以应对课堂教学时间、教师讲授时间压缩等方面的变革；与此同时，教材还在各教学内容的结构安排方面下了很大的功夫，使教材的内容逻辑更清晰，便于教师讲授和学生自主学习。

三是通俗易懂、便于自学。为了满足当前大学生自主学习的要求，我们在教材编写的过程中，要求各教材的语言生动化、案例更切合生活实际且趣味化，如通过借助数表、图形等将抽象的概念用具体、直观的形式表达，用实例和示例加深对概念、方法的理解，尽可能让枯燥、繁琐的数学概念、数理演绎过程通俗化，降低学生自主学习的难度。

当然，教学改革的快速推进不断对教材提出新的要求，同时也受限于我们的水平，这套教材可能离我们理想的目标还有一段距离，敬请各位教师，特别是当前教学改革后已转变为教学活动“主体”的广大学子们提出宝贵的意见！

朱长江

于武昌桂子山

2013 年 7 月

前　言

统计学作为一门重要的应用学科，目前已受到国家广泛的重视。在中学数学课程中，概率和统计的最基础的内容已成为中学教学的重要平台之一，而在高等教育中统计学已提升到与数学平行的一级学科。本书以概率论为基础，同时它又是学习统计学其他课程的必备知识，因此数理统计是本科统计专业、数学专业必修的专业基础课程，也是各高校各专业广泛开设的重要课程。可以说数理统计虽脱胎于数学，但其本身具有丰富的研究内容、独特的研究方法和广泛的实际应用背景。

本书作为统计学专业和数学专业的本科教材，介绍数理统计的基本概念和基础知识，结合数理统计具有广泛实际应用的背景，关注问题的实际背景和解决不同问题所提供的统计方法和思想与合理性，力图对介绍的数理统计的基本概念和基本方法，首先从直观上解释清楚其合理性，然后厘清其理论基础，因此作者认为学习数理统计这门课程应有别于其他数学课程的学习，而应更关注其统计思想、统计方法的来源和应用，掌握一套分析处理实际问题的技术和方法。

全书分为 5 章，前 3 章是其核心内容。第 1 章介绍数理统计的基本概念和初步统计方法；第 2 章较详细地讨论了参数估计的基本理论，围绕估计方法、估计方法的评价及如何寻找“好”的估计量展开；第 3 章讨论假设检验问题，围绕假设检验的基本思想、检验方法和检验方法的评价展开；第 4 章介绍统计决策的基本概念和基本思想；第 5 章简要介绍了回归分析和方差分析的基本概念和基本思想，它们可看成统计推断的两种最基本形式——估计的检验在特定模型下的应用。

全书第 1、3 章由陈应保编写，第 2、4 章由龙兵编写，第 5 章由周涌编写。同时感谢张方毅对文稿的校订及对习题的验算工作，感谢华中师范大学出版社袁正科对编写工作的大力支持。

由于作者水平有限，恳请大家赐教。

编者
2013 年 12 月

目　录

第 1 章

数据整理与抽样分布

我们已经学习了概率论的基础知识，在概率论中，主要讨论随机变量的分布或概率模型已知的条件下，计算各种各样的概率，讨论确定随机变量分布的性质和数字特征等。而在实际问题中，我们常常遇到的是随机变量的分布未知而且正是我们希望通过观察和试验去了解的，这就是数理统计的任务。数理统计是运用概率论的基本知识，对要研究的对象进行多次观察或试验，研究如何获得有效的数据资料，建立科学的数学方法，依据所获得的数据资料对所关心的问题作出统计推断。

本章讨论数理统计的基本概念、样本数字特征及其分布、数据整理，重点讨论抽样分布定理。

1.1 基本概念

1.1.1 总体、个体、简单随机样本

总体、个体、样本是数理统计的三个最基本的概念。所谓**总体**（或**母体**）即研究对象的全体。组成总体的每个最基本的单元称为**个体**。为研究未知总体，从中抽取的 n 个个体称为容量为 n 的**样本**（或**子样**）。例如，若需要考察某批产品，则整批产品构成总体，每个产品就是一个个体。考察某校学生的状况，则全校学生构成研究的总体，每个学生为一个个体。在数理统计的研究中，对构成总体的每个个体我们并不关心其整体状况，或每个个体所具有的特殊属性，而关注的只是其反映总体状况一个或若干个数量方面的特征及其分布。这些数量特征的表现形式为一随机变量或随机向量，记为 X，其分布称为**总体分布**。例如，若关注新生婴儿的性别，男婴记为“0”，女婴记为“1”，则新生婴儿性别 X 为一个取两个值的随机变量，记 p 为新生婴儿为女孩的概率，则 X 服从两点分布。若我们关心的是新生婴儿的身高（或体重）X，则 X 的可能取值为一有限区间（或无穷区间），其分布可认为服从正态分布。

今后我们不再区别总体和我们所关注的指标的取值空间。由于数理统计中所研究的总体分布是未知的，为了对未知的总体分布进行统计推断，就需要做试验、做调查。调查的方式有**全面调查**和**抽样调查**。当总体中所含个体数相当大时，全面调查很难实现，有时也没有意义，特别当试验具有破坏性（如产品寿命，炮弹能否引爆）或耗资太大时，只可能从总体中抽取部分个体来做研究。从总体中抽取 n 个个体来研究，就得到一个容量为 n 的样本，记为 $X_1,X_2,\cdots,X_n$。抽样的目的是为了研究未知总体，因此样本中应尽可能包含总体

的未知信息，这里对样本 $X_1,X_2,\cdots,X_n$ 有两个要求：(1) **代表性**，要求 $X_1,X_2,\cdots,X_n$ 与我们所关注的总体指标 X 有相同的分布，这只要按照随机性原则抽取样本就可实现；(2) **简单性**，为了数学上处理的方便，要求 $X_1,X_2,\cdots,X_n$ 相互独立。

定义 1.1 （简单随机样本）设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的容量为 n 的样本，如果 $X_1,X_2,\cdots,X_n$ 相互独立且与总体 X 有相同的分布，则称 $X_1,X_2,\cdots,X_n$ 为总体 X 的容量为 n 的**简单随机样本**，简称为**简单样本或样本**，记为 $X_1,X_2,\cdots,X_n \mathrm{i\cdot i\cdot d}X$。

今后我们所提到的样本 $X_1,X_2,\cdots,X_n$，在无特别声明时都是指的简单随机样本，数理统计的任务就是要基于简单随机样本对总体分布进行统计推断，而对于简单随机样本我们就可以应用概率论中独立随机变量情形所建立的许多重要定理和结论，它们为数理统计学提供了必要的基础。

1.1.2 统计量

设 $X_1,X_2,\cdots,X_n$ 是来自于总体 X 的简单随机样本，它也可用 n 维随机向量 $(X_1,X_2,\cdots,X_n)$ 表示。若 x_i 为 X_i 的一次观察值，则称 $x_1,x_2,\cdots,x_n$ 为样本 $X_1,X_2,\cdots,X_n$ 的一次**观察值**。样本 $X_1,X_2,\cdots,X_n$ 中包含有总体未知分布的信息，但是这些信息是分散在样本中的，依据样本 $X_1,X_2,\cdots,X_n$ 对总体的未知分布进行统计推断，首先需要根据我们统计推断的内容，集中样本中我们所关心的信息，扮演这一角色的就是下面我们定义的统计量。

定义 1.2 设 $X_1,X_2,\cdots,X_n$ 为来自于总体 X 的简单随机样本，若 $T(X_1,X_2,\cdots,X_n)$ 是样本 $X_1,X_2,\cdots,X_n$ 的函数，且不含任何未知参数。即当样本取值 $x_1,x_2,\cdots,x_n$ 给定后，$T(X_1,X_2,\cdots,X_n)$ 的取值完全确定，则称 $T(X_1,X_2,\cdots,X_n)$ 为**统计量**，$t=T(x_1,x_2,\cdots,x_n)$ 为统计量 $T(X_1,X_2,\cdots,X_n)$ 的取值。

本书中所涉及的统计量 $T(X_1,X_2,\cdots,X_n)$ 都是随机变量。

例 1 设 $X_1,X_2,\cdots,X_n$ 为总体 X 的容量为 n 的样本，记

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\qquad S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

作为样本的函数，当样本的取值给定时，$\overline{X}$ 与 S_n^2 的取值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$s_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 也完全确定，因此 $\overline{X}$ 与 S_n^2 都是统计量。称 $\overline{X}$ 和 S_n^2 分别为样本 $X_1,X_2,\cdots,X_n$ 的**样本均值**及**样本方差**，$\bar{x}$ 为 $\overline{X}$ 的观察值，s_n^2 为 S_n^2 的观察值。

定义 1.3 （次序统计量）设 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的简单随机样本，由样本建立 n 个函数

$$X_{(k)}=X_{(k)}(X_1,X_2,\cdots,X_n)\quad(k=1,2,\cdots,n),$$

当样本取值为 $x_1,x_2,\cdots,x_n$ 时，将它们由小到大排列成 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，则 $X_{(k)}(x_1,x_2,\cdots,x_n)$ 的取值为 $x_{(k)}$，于是由统计量的定义可知，$X_{(k)}(X_1,X_2,\cdots,X_n)$ $(k=1,2,\cdots,n)$ 均为统计量，且满足 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 称之为**次序统计量**，显然

$$X_{(1)}=\min\{X_1,X_2,\cdots,X_n\},\ X_{(n)}=\max\{X_1,X_2,\cdots,X_n\},$$

称 $X_{(1)}$ 为**极小统计量**，$X_{(n)}$ 为**极大统计量**。当 n 为奇数时，称 $X_{(\frac{n+1}{2})}$ 为**样本中值**，当 n 为偶数时，称 $\frac{1}{2}(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)})$ 为**样本中值**。

定义 1.4 （极差）设 $X_1, X_2, \cdots, X_n$ 为来自于总体 X 的简单随机样本，称统计量 $R_n = X_{(n)} - X_{(1)}$ 为**样本极差**。

极小统计量 $X_{(1)}$ 与极大统计量 $X_{(n)}$ 取值反映了样本取值的变化范围，而极差 R_n 的取值反映了样本取值的波动幅度，它与样本方差都是反映了样本观察值的离散程度的数量指标，且计算要更为方便。

例 2 设 X_1, X_2, X_3, X_4 为总体 X 的容量为 4 的简单随机样本，对其做了三次观察，其值如表 1-1 所示，求 $\overline{X}$、S_n^2 及 R_n 的观察值。

表 1-1

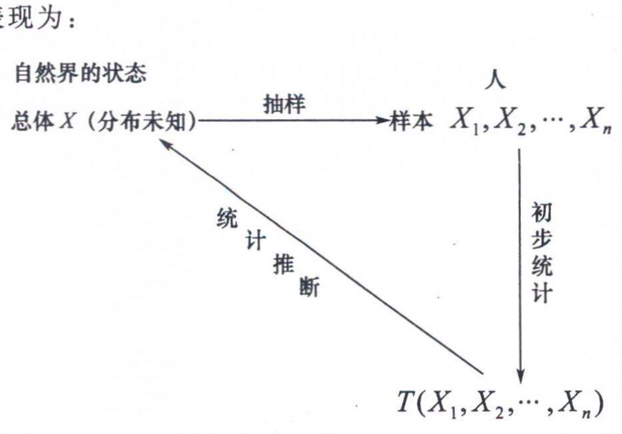
样本号 \ 样本	X_1	X_2	X_3	X_4
1	5	1	8	6
2	3	6	7	2
3	8	4	6	10

解 计算结果如表 1-2 所示：

表 1-2

样本号	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$\overline{X}$	S_n^2	R_n
1	1	5	6	8	5	6.5	7
2	2	3	6	7	4.5	4.25	5
3	4	6	8	10	7	5	6

统计量是统计推断的基础，而统计推断的最基本形式为估计和假设检验。当总体 X 的分布函数形式已知，所未知的只是有限个参数（记为 $F(x;\theta_1,\theta_2,\cdots,\theta_l)$）时，对总体分布及其特征的统计推断等价于对未知参数的统计推断，因此统计推断的两种最基本形式，估计又可分为参数估计和非参数估计，假设检验又可分为参数假设检验和非参数假设检验，统计的实质可表现为：



上循环图体现了人类对自然界的一个认识⟶提高⟶再认识⟶再提高的认识过程。

1.2 样本数字特征及其分布

1.2.1 经验分布函数及格列汶科定理

设总体 X 的分布函数为 $F(x)$，$X_1,X_2,\cdots,X_n$ 为其简单随机样本，实际问题中 $F(x)$ 是未知的，我们关心的问题是依据样本 $X_1,X_2,\cdots,X_n$ 给出 $F(x)$ 的估计。

定义 1.5 （经验分布函数）设 $X_1,X_2,\cdots,X_n$ 为来自于总体 X 的容量为 n 的简单随机样本，$X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为其顺序统计量，当样本取值 $x_1,x_2,\cdots,x_n$ 给定时，顺序统计量 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 的取值 $x_{(1)},x_{(2)},\cdots,x_{(n)}$ 也给定，对任意实数 x，定义函数 $F_n^*(x)$ 为

$$F_n^*(x)=\begin{cases}0, & x\leqslant x_{(1)},\\ \dfrac{k}{n}, & x_{(k)}<x\leqslant x_{(k+1)}\ (k=1,2,\cdots,n-1),\\ 1, & x>x_{(n)},\end{cases}$$

称 $F_n^*(x)$ 为总体 X 的**经验分布函数**。

注：上述定义对 $x_{(1)}<x_{(2)}<\cdots<x_{(n)}$ 时是明确的，而对 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$ 时，$F_n^*(x)$ 更一般的定义为 $F_n^*(x)=\dfrac{v_n}{n}$，其中 v_n 为 $x_1,x_2,\cdots,x_n$ 中小于 x 的个数。

对于经验分布函数 $F_n^*(x)$ 可以做如下两种解释：

首先，当样本观察值 $x_1,x_2,\cdots,x_n$ 给定时，$F_n^*(x)$ 作为 x 的函数具有单调非降、左连续和正则性的性质，因此 $F_n^*(x)$ 是一分布函数，它同时也具有分布函数的其他性质。

其次，对每一个给定的 x，当样本取值 $x_1,x_2,\cdots,x_n$ 给定时，$F_n^*(x)$ 的取值完全确定，因此 $F_n^*(x)$ 是样本 $X_1,X_2,\cdots,X_n$ 的函数，它是一个统计量，其可能取值为 $0,\dfrac{1}{n},\cdots,\dfrac{n-1}{n},1$。由于 $X_1,X_2,\cdots,X_n$ 相互独立且有相同的分布函数 $F(x)$，事件 $\left\{F_n^*(x)=\dfrac{k}{n}\right\}$ 等价于 $X_1,X_2\cdots,X_n$ 中小于 x 的个数为 k，这又等价于 n 重贝努利试验中事件 $\{X<x\}$ 恰好发生 k 次，因此

$$P\left\{F_n^*(x)=\frac{k}{n}\right\}=C_n^k\,(F(x))^k\,(1-F(x))^{n-k},$$

其中 $F(x)=P\{X<x\}$ 为总体 X 的分布函数，也即事件 $\{X<x\}$ 发生的概率，而 $F_n^*(x)$ 为 n 重贝努利试验中事件 $\{X<x\}$ 发生的频率。由概率论中贝努利场合的强大数定律，有

$$P\left\{\lim_{n\to+\infty}F_n^*(x)=F(x)\right\}=1,$$

而 1933 年格列汶科(Glivenko)给出了下面更强的结论。

定理 1.1 ［格列汶科(Glivenko-cantelli)定理］设总体 X 的分布函数为 $F(x)$，经验

分布函数为 $F_n^*(x)$，记

$$D_n = \sup_{-\infty<x<+\infty} \left| F_n^*(x) - F(x) \right|,$$

则

$$P\left\{ \lim_{n\to+\infty} D_n = 0 \right\} = 1。$$

由上面已经知道，$F_n^*(x)$ 为统计量，因而 D_n 为一随机变量，它衡量 $F_n^*(x)$ 同 $F(x)$ 之间在所有 x 值上最大的差异程度。格列汶科定理证明了 D_n 以概率1收敛于零，但并未给出随机变量 D_n 的分布或极限分布，柯尔莫哥洛夫定理和斯米尔诺夫定理给出了 D_n 极限分布的表达式。

1.2.2　样本数字特征

1. 样本矩

设 $X_1, X_2, \cdots, X_n$ 是来自于总体 X 的容量为 n 的简单随机样本，$F_n^*(x)$ 为总体 X 的经验分布函数，记

$$A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r,$$

$$B_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^r,$$

称 A_r 及 B_r 分别为 r **阶样本原点矩**和 r **阶样本中心矩**，且 $A_1 = \overline{X}, B_2 = S^2$，由斯蒂阶积分的性质，有

$$A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r = \int_{-\infty}^{+\infty} x^r \mathrm{d}F_n^*(x),$$

$$B_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^r = \int_{-\infty}^{+\infty} (x - \bar{x})^r \mathrm{d}F_n^*(x),$$

即 A_r 和 B_r 为经验分布函数 $F_n^*(x)$ 对应的 r 阶原点距和 r 阶中心距，它们都是样本的函数而不含任何未知参数，因而都是统计量。

2. 样本协方差与样本相关系数

设 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ 为二维总体 (X, Y) 的简单随机样本，记

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_1^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2,$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad S_2^2 = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \overline{Y})^2,$$

$$S_{12} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})。$$

称统计量 S_{12} 为**样本协方差**，统计量 $R = \dfrac{S_{12}}{S_1 \cdot S_2}$ 为**样本相关系数**。

1.2.3　抽样分布

统计量是我们依据样本对总体进行统计推断的基础，求出统计量 $T(X_1, X_2, \cdots, X_n)$

的分布是数理统计所研究的基本问题之一，统计量的分布称为**抽样分布**，求统计量分布的问题也称为抽样分布问题。

若总体 X 的分布函数形式已知，未知的只是有限个参数，对于任一正整数 n，都能求出统计量 $T(X_1,X_2,\cdots,X_n)$ 的分布函数，这个分布称为统计量 $T(X_1,X_2,\cdots,X_n)$ 的**精确分布**。求出统计量 $T(X_1,X_2,\cdots,X_n)$ 的精确分布，对数理统计中所谓小样本问题（样本容量 n 比较小的情况下所讨论的各种统计问题）的研究很重要，但能够求出统计量的精确分布并不容易，只是对一些重要的特殊情形下可以做到，如正态总体时，我们给出的 t- 统计量、χ^2- 统计量及 F- 统计量，构成数理统计的三个最重要的分布，它们在统计推断问题的研究中起着主要作用。

若不能求出统计量 $T(X_1,X_2,\cdots,X_n)$ 的精确分布或即使能求出其精确分布，但其表达形式很复杂而不便于实际使用，此时我们可以退而求其次，讨论 $n\to\infty$ 时统计量的极限分布，由此当样本容量 n 很大时，就可以用极限分布近似代替统计量的精确分布。极限分布的讨论在数理统计中所谓大样本问题（即在样本容量 n 比较大时的各种统计问题）的研究中有重要意义。下面介绍样本均值和次序统计量的相关分布。

1. 样本均值的分布

设 $X_1,X_2\cdots,X_n$ 是来自于总体 X 的简单随机样本，总体 X 的特征函数为 $\varphi(t)$，由特征函数的性质，立即可得 $\sum\limits_{i=1}^{n}X_i$ 的特征函数为 $[\varphi(t)]^n$，$\frac{1}{n}\sum\limits_{i=1}^{n}X_i=\overline{X}$ 的特征函数为

$$\varphi_1(t)=\left[\varphi\left(\frac{t}{n}\right)\right]^n,$$

根据特征函数与分布函数的一一对应关系，由特征函数就可求出其分布函数。

例 1 设 $X_1,X_2,\cdots,X_n$ 来自于正态总体 $X\sim N(\mu,\sigma^2)$，求样本均值 $\overline{X}$ 的分布。

解 由概率论中的结论知，总体 X 的特征函数为

$$\varphi(t)=\exp\left\{\mathrm{i}\mu t-\frac{\sigma^2}{2}t^2\right\},$$

从而 $\overline{X}$ 的特征函数为

$$\begin{aligned}\varphi_1(t)&=\left\{\exp\left[\mathrm{i}\mu\frac{t}{n}-\frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2\right]\right\}^n\\&=\exp\left\{\mathrm{i}\mu t-\frac{\sigma^2t^2}{2n}\right\},\end{aligned}$$

而正态分布 $N\left(\mu,\frac{\sigma^2}{n}\right)$ 的特征函数亦为 $\varphi_1(t)$，所以 $\overline{X}$ 服从正态分布 $N\left(\mu,\frac{\sigma^2}{n}\right)$。

例 2 设 $X_1,X_2,\cdots,X_n\ \mathrm{i}\cdot\mathrm{i}\cdot\mathrm{d}X$ 服从参数为 λ 的泊松分布，求 $\overline{X}$ 的分布。

解 总体 X 的特征函数为

$$\varphi(t)=\exp\{\lambda(\mathrm{e}^{\mathrm{i}t}-1)\},$$

$\sum\limits_{i=1}^{n}X_i$ 的特征函数为

$$[\varphi(t)]^n = \exp\{n\lambda(e^{it} - 1)\},$$

这恰好是参数为 $n\lambda$ 的泊松分布随机变量的特征函数，所以 $\sum_{i=1}^{n} X_i$ 服从参数为 $n\lambda$ 的泊松分布，从而

$$P\left\{\overline{X} = \frac{k}{n}\right\} = P\left\{\sum_{i=1}^{n} X_i = k\right\} = \frac{(n\lambda)^k}{k!}e^{-n\lambda} \quad (k = 0,1,2,\cdots)。$$

上述两例说明了特征函数作为工具在求统计量精确分布中的应用。

2. 次序统计量的分布

设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本，总体 X 的分布函数为 $F(x)$；$X_{(1)}$，$X_{(2)}, \cdots, X_{(n)}$ 是其次序统计量，记极大统计量 $X_{(n)}$ 的分布函数为 $F_n(x)$，极小统计量 $X_{(1)}$ 的分布函数为 $F_1(x)$，则有

$$\begin{aligned}
F_n(x) &= P\{X_{(n)} < x\} = P\{X_1 < x, \cdots, X_n < x\} = [F(x)]^n, \\
F_1(x) &= P\{X_{(1)} < x\} = 1 - P\{X_{(1)} \geqslant x\} \\
&= 1 - P\{X_1 \geqslant x, \cdots, X_n \geqslant X\} \\
&= 1 - [1 - F(x)]^n。
\end{aligned}$$

进一步，若总体 X 为连续型随机变量，其密度函数为 $f(x)$，则有 $X_{(1)}, X_{(n)}$ 仍为连续型随机变量，其密度函数分别为

$$\begin{aligned}
f_1(x) &= [F_1(x)]' = n[1 - F(x)]^{n-1} f(x), \\
f_n(x) &= [F_n(x)]' = n[F(x)]^{n-1} f(x)。
\end{aligned}$$

例 3　设 $X_1, X_2, \cdots, X_n \, \mathrm{i \cdot i \cdot d} X \sim U[0,\theta]$，求 $X_{(1)}, X_{(n)}$ 的密度函数。

解　由条件知，总体 X 的分布函数和密度函数分别为

$$F(x) = \begin{cases} 0, & 0 \leqslant 0, \\ \dfrac{x}{\theta}, & 0 < x \leqslant \theta, \\ 1, & x > \theta; \end{cases} \quad f(x) = \begin{cases} \dfrac{1}{\theta}, & 0 \leqslant x \leqslant \theta, \\ 0, & 其他。\end{cases}$$

则 $X_{(1)}$ 的分布函数和分布密度函数为

$$F_1(x) = \begin{cases} 0, & x \leqslant 0, \\ 1 - \left(1 - \dfrac{x}{\theta}\right)^n, & 0 < x \leqslant \theta, \\ 1, & x > \theta; \end{cases} \quad f_1(x) = \begin{cases} n\left(1 - \dfrac{x}{\theta}\right)^{n-1} \dfrac{1}{\theta}, & 0 \leqslant x \leqslant \theta, \\ 0, & 其他。\end{cases}$$

$X_{(n)}$ 的分布函数和分布密度函数为

$$F_n(x) = \begin{cases} 0, & x \leqslant 0, \\ \left(\dfrac{x}{\theta}\right)^n, & 0 < x \leqslant \theta, \\ 1, & x > \theta; \end{cases} \quad f_n(x) = \begin{cases} n\dfrac{x^{n-1}}{\theta^n}, & 0 \leqslant x \leqslant \theta, \\ 0, & 其他。\end{cases}$$