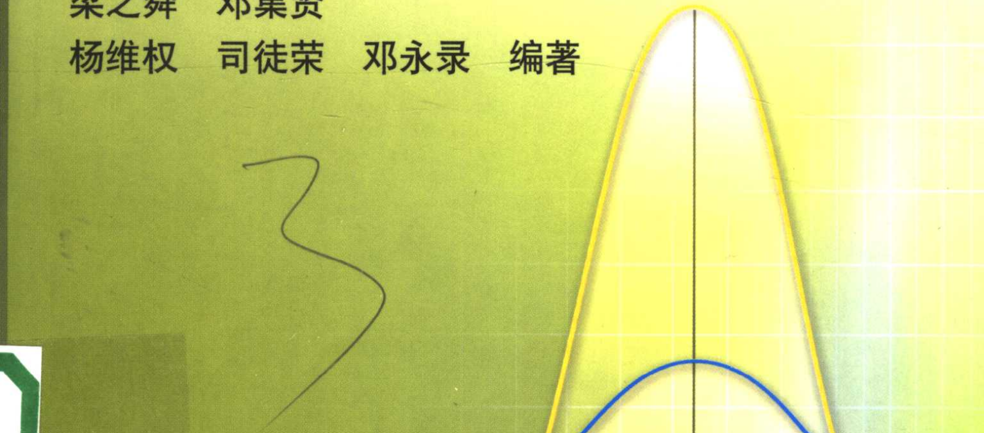
高等学校教材

概率论及数理统计

第3版（下册）

中山大学统计科学系
梁之舜 邓集贤
杨维权 司徒荣 邓永录 编著

3



高等教育出版社
HIGHER EDUCATION PRESS

概率论及数理统计

（下　册）

（第3版）

中山大学统计科学系
梁之舜　　邓集贤
杨维权　　司徒荣　　邓永录
编　著

图书在版编目(CIP)数据

概率论及数理统计. 下册/梁之舜等编著;中山大学统计科学系.—3版.—北京:高等教育出版社,2005.11
ISBN 7-04-015957-0

Ⅰ.概... Ⅱ.梁... Ⅲ.①概率论-高等学校-教材②数理统计-高等学校-教材 Ⅳ.O21

中国版本图书馆 CIP 数据核字(2005)第 020677 号

策划编辑 李蕊　**责任编辑** 丁鹤龄　**封面设计** 王凌波
责任绘图 吴文信　**版式设计** 王莹　**责任校对** 杨雪莲
责任印制 孔源

出版发行	高等教育出版社	购书热线	010-58581118
社　　址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	北京蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	北京明月印务有限责任公司		
		版　　次	1980年7月第1版
开　　本	850×1168　1/32		2005年11月第3版
印　　张	12.5	印　　次	2005年11月第1次印刷
字　　数	310 000	定　　价	14.60元

物料号　15957-00

内容提要

本书第三版是在中山大学数学系梁之舜等五人编著的《概率论及数理统计》(1988 年第二版)的基础上修订而成的,现将中山大学数学系的署名改为中山大学统计科学系,五位编者不变。第三版与第二版相比有不少小的修改,将原第四章与第五章合并为新的第四章,增设新的第五章勒贝格－斯蒂尔切斯积分与新的第十章统计决策及贝叶斯统计,全书共分设十二章,仍分上、下两册出版。

学习本书只要求读者具有高等数学(微积分、高等代数)的基础知识,因此本书具有适应面广、便于自学的特点。

本书可作为综合大学、师范院校及其他高等院校的数学与应用数学、信息计算科学、统计学等专业的教材,也可作为其他有关专业的教学参考书。

第三版说明

本书(上、下册)作为全国高等院校通用教材于1980年出版。1988年第二版发行至今又经历了十五年时光,在这期间本书(上、下册)又共重印发行十余万册,仍然受到全国高等院校师生及广大读者的欢迎。高等教育出版社根据当前高等院校教学的具体情况和需要,要求我们对本书重新修订作为第三版继续出版发行。本书编者们多次讨论了高等教育出版社对本书修订的建议及第三版修订的基本原则。此外,我们听取了曾多次使用过本书的有关院校老师们的意见,并同中山大学统计科学系何远江、区景琪、尹小玲、余锦华等教师进行了讨论。我们还注意到当前高等院校不同专业对"概率论及数理统计"课程的教学要求有所不同。为了使本书具有广泛的适用性,教师在教学中使用本书时更好地操作,因此将第三版编写的基本原则与具体措施体现在下述三个方面:

(1) 第三版仍保持第一版提出的基本原则,即本书只要求读者具有高等数学的数学基础;

(2) 对第二版作了删繁就简的修改,如将第二版的第四、五两章合并为第三版的第四章,删去纯分析和偏难的内容(大部分定理和性质的证明过程)。又如对第二版的第七、八、九各章按删繁就简原则作了较大的修改。其他各章也都作了部分修改,并适当增加一些实际应用的例子;

(3) 我们注意到有许多读者希望进一步学习概率论,因此在第三版中增设新的第五章,主要介绍勒贝格－斯蒂尔切斯积分的有关知识及其在概率论中的一些应用。又考虑到数理统计学的一个学派,即贝叶斯统计学在理论与应用方面有较成熟的发展,于是在第三版中增设第十章,介绍统计决策及贝叶斯统计的基本知识。

我们认为概率论及数理统计课程的基本内容是本书第三版(上册)第一、二、三、四及(下册)第六、七、八、九共计八章。书中有些章、节、段或定理的证明打*号或用小字编排,供教学参考选用。不同专业按其教学计划的要求,可以从本书(上、下册)中适当选讲有关章节内容。

本书的第一、四、十一各章由邓集贤执笔;第二、三两章由司徒荣执笔;第五章由梁之舜(区景琪、何远江协助)执笔;第六、七、八、九、十各章由杨维权执笔;第十二章由邓永录执笔。本书第三版能顺利地发行,编者们感谢广大读者的厚爱和高等教育出版社的大力支持和鼓励,并感谢中山大学统计科学系何远江教授的协助。再版后本书可能还会有许多不妥之处,希望读者及专家批评指正。

编　者

2004 年 7 月 1 日于广州中山大学

郑重声明

目　　录

第六章　抽样分布

前述各章,是概率论的引论,讲述了概率论的基本内容,为数理统计学建立了重要的数学基础.从本章起的接连五章,是数理统计学的初步,主要讲述估计与检验等原理、回归分析与方差分析、贝叶斯统计等统计方法.

数理统计学是运用概率论的基本知识,对要研究的随机现象进行多次观察或试验,研究如何合理地获得数据资料,建立有效的数学方法,根据所获得的数据资料,对所关心的问题作出估计与检验.数理统计学的重要分支有统计推断、多元统计分析、试验设计等,其具体方法甚多,应用相当广泛,已成为各学科从事科学研究及生产、经济等部门进行有效工作的必不可少的数学工具.

在这一章中,我们从数理统计学的基本概念讲起,讨论抽样分布及其重要的定理.这些抽样分布及其几个重要的定理,在前述各章中尚未提到,而在后述各章中却经常要用到它们.

§6.1　基本概念

一、总体、个体、简单随机样本

总体、个体、样本是数理统计学中三个最基本的术语.我们把对某一个问题的研究对象的全体称为总体(或母体),组成总体的每个基本单元称为个体,从总体中随机抽取的 n 个个体称为容量为 n 的样本.例如,把某月的整批产品视为一总体,则每个产品为个体;又如把某批灯泡视为一总体,则每个灯泡为个体;某地在某季度内每天的日平均气温的全体视为一总体,则其中某天的日平均气温为个体.在数理统计学中,我们是对总体成员的一个或者若干个数量表征进行研究.如日平均气温的度数用 ξ 表示,灯泡的

使用寿命(小时)用 η 表示,这样对总体的研究就归结为讨论随机变量 ξ 或 η 的分布函数及其主要数字特征(如数学期望 $E(\xi)$、方差 $D(\xi)$等)的研究.又如某月的整批产品为一总体,若只要检查产品的质量,可用数量表征 ζ 来反映.产品为一等品,令 ζ 取值为1;二等品,ζ 取值为2;不合格品,ζ 取值为3.我们研究这批产品的质量规律,就归结为讨论随机变量 ζ 的分布函数及其主要数字特征来反映.

今后我们常用"总体 ξ 服从什么分布"这样的术语,它是指总体的某个具体数量表征 ξ 服从什么分布规律.我们说对总体进行 n 次独立的重复试验或 n 次独立的观察,就是从总体中随机地抽取容量为 n 的样本,对应的数量表征以 $\xi_1,\xi_2,\cdots,\xi_n$ 表示,不难理解,ξ 及 $\xi_1,\xi_2,\cdots,\xi_n$ 都是随机变量,$(\xi_1,\xi_2,\cdots,\xi_n)$是 n 维随机向量.例如我们说整批灯泡的寿命分布如何?是指灯泡的寿命 ξ 服从什么分布.从整批灯泡中随机抽取容量为 n 的样本,是指随机抽取 n 个灯泡,观察这 n 个灯泡的寿命,用 $\xi_1,\xi_2,\cdots,\xi_n$ 表示,其中 ξ_i 表示随机抽取到的第 i 个灯泡的寿命,显然 ξ_i 也是一随机变量$(i-1,2,\cdots,n)$.为区别总体与样本的概念,我们今后常用诸如"总体 ξ 服从什么分布"、"从总体 ξ 中抽取样本 $\xi_1,\xi_2,\cdots,\xi_n$"等这类术语.

我们在对总体进行研究中,当总体相当大时(指个体相当多时)或当对个体进行试验具有破坏性(如炮弹能否引发爆炸)或费时耗资大时,只可能从总体中抽取部分个体来作研究,即抽取容量为 n 的样本来作研究.我们要从样本的观察或试验结果的特性来对总体的特性作出估计与推断,一方面自然要研究应该怎样从总体中抽取样本,使得样本在尽可能大的程度上反映总体的特性;同时必须建立一整套的方法,使能根据所选取的样本的性质,来对总体的特性进行估计与推断.因此,我们在抽取样本来对总体作出估计与推断时,从总体中抽取样本必须是随机的,即每一个个体都有

同等概率被抽取(当总体中的个体是有限个时,要用有返回抽取方式).其具体要求为两个方面:独立性,是指 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立;代表性,是指 $\xi_1,\xi_2,\cdots,\xi_n$ 中每一个都与总体 ξ 有相同分布.

定义 6.1.1(简单随机样本) 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为来自总体 ξ 的容量为 n 的样本,如果 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立且每一个都是与总体 ξ 有相同分布的随机变量,则称 $\xi_1,\xi_2,\cdots,\xi_n$ 为总体 ξ 的容量为 n 的简单随机样本,简称为简单样本或样本①.

在这一章中,我们所讲的样本 $\xi_1,\xi_2,\cdots,\xi_n$,无特别声明的话,都是指简单样本,即 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立且每一个都与总体 ξ 有相同的分布.引入简单随机样本,是基于要从样本所获得的概率统计特性来对总体的概率统计特性作出估计与推断.对于简单随机样本,我们可以应用概率论中对独立随机变量的情形所建立的许多重要的定理,这些重要的结论为数理统计学提供了必要的基础.

二、统计量

样本 $\xi_1,\xi_2,\cdots,\xi_n$ 也可用 n 维随机向量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 表示.记 x_i 为 ξ_i 的一次观察值,并称 $(x_1,x_2,\cdots,x_n)$ 为样本的一次观察值.样本 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的所有可能取值的全体称为样本空间,记作 $\mathscr{X}$,它是 n 维空间.样本的一次观察值 $(x_1,x_2,\cdots,x_n)$ 就是样本空间 $\mathscr{X}$ 中的一个点,即 $(x_1,x_2,\cdots,x_n)\in\mathscr{X}$.

定义 6.1.2(统计量) 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为总体 ξ 的样本,T 为样本空间 $\mathscr{X}$ 中点 $(x_1,x_2,\cdots,x_n)$ 的实值函数,作样本的函数 $T=T(\xi_1,\xi_2,\cdots,\xi_n)$,$T$ 的取值记为 $t=T(x_1,x_2,\cdots,x_n)$.若 $T(\xi_1,\xi_2,\cdots,\xi_n)$ 也为一随机变量②,且不带未知参数,则称 T 或 $T(\xi_1,\xi_2,\cdots,\xi_n)$ 为统计量.

---

① 是指若 ξ 为定义在概率空间 $(\Omega,\mathscr{F},P)$ 上的随机变量,则 $\xi_1,\xi_2,\cdots,\xi_n$ 为概率空间 $(\Omega,\mathscr{F},P)$ 上 n 个相互独立的并同 ξ 具有相同分布的随机变量.

② 具体要求就是 $T(\xi_1,\xi_2,\cdots,\xi_n)$ 是 n 元可测实函数,也就是在第二章中所述的 n 元随机变量,它满足对每一 $t\in\mathbf{R}_1$,有 $\{T(\xi_1,\xi_2,\cdots,\xi_n)\leqslant t\}\in\mathscr{F}$.

我们在用样本 $\xi_1,\xi_2,\cdots,\xi_n$ 获得的信息来对总体 ξ 作出估计与推断时,是按不同的统计问题的要求而规定样本的各种函数.在这本教材中,所涉及到的样本的各种函数 $T(\xi_1,\xi_2,\cdots,\xi_n)$, $T(x_1,x_2,\cdots,x_n)$一般都是多维连续函数,因而 $T(\xi_1,\xi_2,\cdots,\xi_n)$ 都是随机变量.

例 6.1.1 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为总体 ξ 的样本,其容量为 n.记

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2,$$

则 $\bar{\xi}$ 及 S^2 都是统计量,称 $\bar{\xi}$ 及 S^2 分别为样本 $\xi_1,\xi_2,\cdots,\xi_n$ 的平均值及方差.样本的观察值为 $x_1,x_2,\cdots,x_n$,$\bar{\xi}$ 及 S^2 的观察值分别记作

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

今后,大写的 S^2 表示统计量,小写的 s^2 表示统计量 S^2 的观察值.

定义 6.1.3(顺序统计量) 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为总体 ξ 的样本,今由样本建立 n 个函数:

$$\xi_k^* = \xi_k^*(\xi_1,\xi_2,\cdots,\xi_n), \qquad k = 1,2,\cdots,n,$$

其中 ξ_k^* 为这样的统计量,它的观察值为 x_k^*,x_k^* 为样本 $\xi_1,\xi_2,\cdots,\xi_n$ 的观察值 $x_1,x_2,\cdots,x_n$ 中由小至大排列(即 $x_1^* \leqslant \cdots \leqslant x_k^* \leqslant \cdots \leqslant x_n^*$)后的第 k 位数值,则称 $\xi_1^*,\xi_2^*,\cdots,\xi_n^*$ 为顺序统计量.

易见,$\xi_1^* = \min\{\xi_1,\xi_2,\cdots,\xi_n\}$,$\xi_n^* = \max\{\xi_1,\xi_2,\cdots,\xi_n\}$.称 ξ_1^* 为最小项统计量,ξ_n^* 为最大项统计量.若 n 为奇数,则称 $\xi_{\frac{n+1}{2}}^*$ 为样本的中值;若 n 为偶数,则称 $\xi_{\frac{n}{2}+1}^*$ 为样本的中值.

定义 6.1.4(极差) 设 $\xi_1,\xi_2,\cdots,\xi_n$ 为总体 ξ 的样本,则称统计量 $D_n^* = \xi_n^* - \xi_1^*$ 为样本的极差.

极差是样本中最大值与最小值之差,反映了样本观察值的波动幅度.它同方差一样是反映观察值离散程度的数量指标,而且计

算方便.

例 6.1.2 设 $\xi_1,\xi_2,\cdots,\xi_5$ 为 ξ 的容量为 5 的样本,今对这个样本作了三次观察,其值如表 6.1.1 所示,试求 $\overline{\xi}$、S^2 及 D_n^* 的观察值.

表 6.1.1

x \ ξ	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
1	3	1	10	5	6
2	2	6	7	2	8
3	8	3	9	10	5

解 计算结果列于表 6.1.2.

表 6.1.2

ξ_1^*	ξ_2^*	ξ_3^*	ξ_4^*	ξ_5^*	$\bar{x}$	s^2	D_n^*
1	3	5	6	10	5	9.2	9
2	2	6	7	8	5	6.4	6
3	5	8	9	10	7	6.8	7

三、小样问题与大样问题

统计量是我们对总体 ξ 的分布函数或数字特征进行估计与推断最重要的基本概念,求出统计量 $T(\xi_1,\xi_2,\cdots,\xi_n)$的分布函数是数理统计学的基本问题之一.统计量的分布,称为抽样分布.

设总体 ξ 的分布函数表达式已知,对于任一自然数 n,如能求出给定统计量 $T(\xi_1,\xi_2,\cdots,\xi_n)$的分布函数,这分布称为统计量 T 的精确分布.求出统计量 T 的精确分布,这对于数理统计学中的所谓小样问题(即在样本容量 n 比较小的情况下所讨论的各种统计问题)的研究是很重要的.