数理统计

Mathematical Statistics

钟　波　刘琼荪　刘朝林　黄光辉　编著

高等教育出版社

数理统计

SHULI TONGJI

Mathematical Statistics

钟　波　刘琼荪　刘朝林　黄光辉　编著

高等教育出版社·北京

内容提要

本书共七章，内容包括概率论概述、数理统计基础知识、参数估计、假设检验、回归分析、方差分析与正交设计、多元统计分析。为便于学生更好地学习和掌握各章经典的统计内容，在每章中特别配有应用案例、章节总结和内容扩展。为了拓展学生的知识面，在内容扩展小节中引入了一些前沿性的现代统计知识。全书习题选材丰富，书末附有部分习题参考答案。

本书讲解简明扼要，深入浅出，图文并茂，便于自学。既可作为高等学校非数学类专业研究生的数理统计教材，也可作为数学类、统计学类专业高年级本科生用书，还可作为相关技术人员的参考书或工具书。

图书在版编目（CIP）数据

数理统计／钟波等编著. --北京：高等教育出版社，2015.2

ISBN 978-7-04-041283-3

Ⅰ. ①数… Ⅱ. ①钟… Ⅲ. ①数理统计-研究生-教材 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字（2014）第 240454 号

策划编辑 兰莹莹　责任编辑 张晓丽　封面设计 李卫青　版式设计 余　杨
插图绘制 黄建英　责任校对 胡美萍　责任印制 赵义民

出版发行	高等教育出版社	咨询电话	400-810-0598
社　　址	北京市西城区德外大街 4 号	网　　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　　刷	北京京科印刷有限公司	网上订购	http://www.landraco.com
开　　本	787mm×960mm　1/16		http://www.landraco.com.cn
印　　张	26	版　　次	2015 年 2 月第 1 版
字　　数	480 千字	印　　次	2015 年 2 月第 1 次印刷
购书热线	010-58581118	定　　价	39.80 元

物 料 号　41283-00

前 言

数理统计是随机数学领域中一门应用性极强的基础课程，是各个学科不可缺少的数学理论基础和数据分析工具。随着计算机技术的迅速发展，数理统计不仅在工程技术、自然科学等领域中发挥着越来越重要的作用，而且以空前的广度和深度向社会、经济、管理、金融、生物、医学、环境、交通等各个领域渗透，它已经成为当代高新技术的重要组成部分。时代呼唤我们的学生学习一些数据分析方法，掌握定性与定量分析有机结合的研究技能，本书正是适应这一需要而编写的。

重庆大学研究生的数理统计课程被列为首批重庆市级优质课程，是重点建设的公共基础课程之一。本书结合重庆大学十多年来数理统计课程积累的教学经验，并在参考国内外同类优秀教材的基础上编写而成。编写本书的宗旨是：突出教学内容的现代化，重视统计思想的介绍，把学科前沿的研究成果引入书中，开阔读者的视野；结合应用案例介绍统计方法，强调统计应用和以统计软件为支撑的计算，适应现代统计教育的特点以及时代发展的新要求。

本书具有以下特色：(1) 以问题为导向引出统计方法，用图文并茂的直观形式解释各种现象的统计规律和结果，基于寓意深刻的背景内容阐释一些理论结果，以清晰的语言系统地介绍数理统计的经典内容。(2) 在介绍经典内容的同时，开辟了扩展知识栏目，适当引入一些前沿的研究成果，如 Copula 函数、贝叶斯统计、非参数估计、似然比检验、回归诊断等内容。诱导和启发学生思考，开阔学生的视野，以满足学生进一步学习的需求。(3) 为了培养学生分析问题、解决问题的能力，在例题、应用案例和习题中选择了大量丰富的实际案例，这些案例来源于经济管理、医学卫生、生物、工学、社会学等学科领域，有利于分类分层教学。增设的应用案例呈现了“提出问题→建立统计模型→运用统计方法→软件求解（模拟）→结果分析”这种解决实际问题的全过程。学生通过应用案例的学习，并完成三个层面（基础层、提高层和拓展层）的习题的教学实践，能有效地掌握多种统计模型和统计方法。

本书适合高等学校非数学类专业一年级的研究生学习，也可作为数学类、统计学类专业本科生的教材，所需的预备知识是微积分、线性代数和概率论基础。

本书共七章，具体分工是：第一、二章由刘朝林编写，第三、四章由钟波编写，第五、六章由刘琼荪编写，第七章由黄光辉编写。在编写过程中得到了各级领导和数理统计课程组教师的积极支持，尤其是课程组的荣腾中老师补充了部分C组的习题，并对教材提出了建设性的修改意见。同时，本书的编写也得到了部分研究生同学的热情帮助，在此对他们表示真诚的感谢！

由于编者水平有限，不当之处在所难免，恳请国内外同行及广大读者不吝赐教。

编者

2014年5月26日于重庆大学

郑重声明

目　录

第1章 概率论概述

概率论(Probability Theory)、数理统计(Mathematical Statistics)是研究随机现象及其统计规律的两门数学学科,是随机数学领域中的基础部分. 在概率论与数理统计的基础上又形成了许多随机数学分支,如随机过程、试验设计、随机运筹学、可靠性数学、随机分析等,在实际中这些分支都有着广泛的应用.

概率论的起源可以追溯到17世纪中叶. 著名法国数学家帕斯卡(Pascal)和费马(Fermat)受法国数学家兼赌徒德米尔(De Mere)之托,对赌博游戏中的输赢概率进行了研究,为古典概率论的形成做出了贡献. 而建立在严格逻辑基础上的概率论的研究却持续了3个多世纪,直至20世纪初,勒贝格(Lebesgue)测度与积分理论的完成以及随后发展的抽象测度和积分理论,为概率公理化体系的建立提供了理论方法. 在这种背景下,苏联杰出的数学家柯尔莫哥洛夫(Kolmogorov)1933年在他的《概率论基础》一书中首次给出了基于测度论的概率定义和一套严密的公理化体系,为现代概率论的建立奠定了基础. 概率论作为理论严谨、应用广泛的数学学科日益受到人们的重视,并随着社会和科学技术的发展而发展. 21世纪以来,概率论理论与方法已渗透到各个学科和不同的领域. 例如:

◆ 基于概率论的复杂系统可靠性研究. 如,电力系统、核能系统、通信系统、商用或军用飞机、软件系统、桥梁等的可靠度计算、可靠性分析与设计.

◆ 建立在概率论基础上的当代亚微粒子物理学理论.

◆ 基于随机性的复杂网络理论,包括基因动态调控模型、干细胞等重要生物体的演化、计算机病毒在互联网或邮件网络中的传播、计算机黑客对计算机网站进攻的演化过程,以及SARS、禽流感和艾滋病等恶性传染病在人群构成的复杂网络中的传播、信用风险与非法资金在金融机构形成的复杂网络中的传播与扩散等内容.

◆ 概率是现代计算机科学核心理论不可缺少的一部分.如随机化算法、算法的概率分析、概率组合构造等,其应用遍及组合优化、机器学习、通信网络以及安全协议等诸多领域,越来越高级、越来越复杂的概率技术已应用于富有挑战性的问题.

◆ 大气边界层湍流的概率分布研究.如大气湍流下自由空间光通信中断的概率分析.

◆ 保险精算学科中的保险模型、破产理论、分红理论、风险分析以及决策与风险控制理论等都高度依赖概率理论.

◆ 物流供应链中货物的运输与随机化存储模型.

这样的例子,举不胜举. 总之,科学研究已经将随机性视为建模和分析中的基本组成部分.

1.1 随机事件及概率

1.1.1 随机事件

将可能出现也可能不出现的现象称为随机事件(random event),简称为事件,一般用符号 $A,B,C,\cdots$ 或带下标的 $A_k,B_k,C_k,\cdots(k=1,2,\cdots)$ 表示. 随机事件是概率论的研究对象.在概率论中,随机事件都是针对某个随机试验而言的.将具有下列三个特点的一次观测或测验称为随机试验,简称为试验:

(1) 可重复性:观测可以在相同的条件下重复进行.

(2) 多样性与明确性:观测结果不止一个,但观测前所有可能出现的结果是明确的.

(3) 不确定性:每次观测前,都不能确定哪一个结果会出现.

随机试验中最简单的、不能再分解的事件称为样本点(sample point),记为 ω 或带下标的 $\omega_i(i=0,1,\cdots)$. 所有样本点构成的集合称为样本空间(sample space),记为 Ω. 例如,观测一天内进入某超市的人数. 显然,该试验的样本点是“一天内有 i 个人进入该超市”,记为 $\omega_i(i=0,1,\cdots)$,则样本空间 $\Omega=\{\omega_i \mid i=0,1,\cdots\}$. 若 A 表示“一天内进入该超市的人数至少有 200 人”的事件,则与 A 有关的全部样本点为 $\omega_i,i=200,201,\cdots$,将这些样本点组成集合 $\{\omega_i \mid i=200,201,\cdots\}$. 易见,该集合是样本空间 Ω 的一个子集,且具有如下性质:一方面,集合 $\{\omega_i \mid i=200,201,\cdots\}$ 中任何一个样本点在试验中出现,则意味着事件 A 必然出现;另一方面,当事件 A 在试验中出现时,则集合 $\{\omega_i \mid i=200,201,\cdots\}$ 中某个样本点一定出现. 于是,可以用集合 $\{\omega_i \mid i=200,201,\cdots\}$ 代表事件 A. 这样,随机试验中的任何一个事件都对应样本空间 Ω 的一个子集,研究随机事件的问题就可以通过“集合”转化为研究样本空间的子集的问题. 显然,样本空间 Ω 代表在每次随机试验时都要出现的现象,称之为必然事件(certain event).空集 $\varnothing$ 代表在每次随机试验时都不会出现的现象,称之为不可能事件(impossible event).

对于复杂事件,一般处理方式是转化为简单事件或已知的事件.而这个转化

可以通过事件的运算或关系来实现. 因此,需要研究事件之间的关系与运算. 由于可以用集合表示事件,所以,借助集合的运算与关系定义事件的运算与关系. 设 A、B、C、…, A_1、A_2、… 是某个随机试验中的事件,那么,这些符号也表示相应的样本空间 Ω 的子集.

(1) 若子集 A、B 满足: $A \subset B$, 则称事件 A 为事件 B 的子事件,表示"事件 A 发生必然导致事件 B 发生". 如果事件 A 与 B 互为子事件,则称事件 A 与 B 相等,记为 $A = B$.

(2) 集合 $A \cup B$ 称为事件 A 与 B 的和事件,表示"事件 A 和 B 至少有一个发生". 类似地,集合 $A_1 \cup A_2 \cup \cdots \cup A_n$ 称为事件 $A_1, A_2, \cdots, A_n$ 的和事件,表示"事件 $A_1, A_2, \cdots, A_n$ 中至少有一个发生",简记为 $\bigcup\limits_{i=1}^{n} A_i$. 还可以进一步类似地定义 $\bigcup\limits_{i=1}^{+\infty} A_i$.

(3) 集合 $A \cap B$ 称为事件 A 与 B 的积事件,简记为 AB, 表示"A 与 B 同时发生". 类似地,集合 $A_1 \cap A_2 \cap \cdots \cap A_n$ 称为事件 $A_1, A_2, \cdots, A_n$ 的积事件,表示"事件 $A_1, A_2, \cdots, A_n$ 同时发生",简记为 $\bigcap\limits_{i=1}^{n} A_i$ 或 $A_1 \cdots A_n$. 同理,可定义 $\bigcap\limits_{i=1}^{+\infty} A_i$.

(4) 如果子集 A、B 满足: $AB = \varnothing$, 则称 A 与 B 为互斥事件(也称为互不相容事件),其含义为:事件 A 与 B 不可能在一次试验中同时发生. 对两个互斥事件 A 与 B, $A \cup B \xlongequal{\text{记为}} A + B$. 如果事件 A 与 B 满足 $AB = \varnothing$ 和 $A \cup B = \Omega$, 则称 A、B 为互为对立的事件. 将 A 的对立事件 B 记为 $\overline{A}$, 于是有 $A + \overline{A} = \Omega$, $A\overline{A} = \varnothing$. 如果有事件 $A_1, A_2, \cdots, A_n$, 满足 $A_iA_j = \varnothing$, $i \neq j (i, j = 1, 2, \cdots, n)$, 则称事件组 $A_1, A_2, \cdots, A_n$ 是两两互斥的,这时, $\bigcup\limits_{i=1}^{n} A_i \xlongequal{\text{记为}} \sum\limits_{i=1}^{n} A_i$. 若对两两互斥的事件组 $A_1, A_2, \cdots, A_n$, 有 $\bigcup\limits_{i=1}^{n} A_i = \Omega$, 则称 $A_1, A_2, \cdots, A_n$ 为完备事件组.

(5) 集合 $A - B$ 称为 A 与 B 的差事件,表示"事件 A 发生而事件 B 不发生". 易知, $A - B = A\overline{B}$ 或 $A - AB$.

易验证,上述事件的运算满足以下规律

(1) 交换律: $A \cup B = B \cup A$, $A \cap B = B \cap A$.

(2) 结合律: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$.

(3) 分配律: $A \cap (B \cup C) = AB \cup AC$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

(4) 德摩根定律: $\overline{\bigcup\limits_{i=1}^{n} A_i} = \bigcap\limits_{i=1}^{n} \overline{A_i}$, $\overline{\bigcap\limits_{i=1}^{n} A_i} = \bigcup\limits_{i=1}^{n} \overline{A_i}$.

1.1.2　概率

1. 概率的公理化定义

对事件 A 在一次随机试验中发生的几率的数量度量,称为事件 A 的概率(probability),记为 $P(A)$. 概率的公理化定义如下:

定义 1.1.1　设随机试验 E 的样本空间为 Ω, $\mathscr{F}$是 Ω 中所有事件组成的集合,称之为事件域, P 是定义在 $\mathscr{F}$上的实值集合函数(称为测度),即 $P:\mathscr{F}\to\mathbf{R}$. 若 $P(\cdot)$ 满足:

(1) 规范性: $P(\Omega)=1$.

(2) 非负性:对 $\mathscr{F}$中的任意一个事件 A, 有 $P(A)\geqslant 0$.

(3) 可列可加性:对无穷可数个两两互斥的事件组 $A_1,A_2,\cdots,A_n,\cdots$, 有

$$P\Big(\sum_{i=1}^{+\infty}A_i\Big)=\sum_{i=1}^{+\infty}P(A_i). \tag{1.1.1}$$

则称实函数 $P(\cdot)$ 为概率测度(probability measure), $P(A)$ 称为事件 A 的概率, $(\Omega,\mathscr{F},P)$ 称为概率空间(probability space).

易证,公理化定义下的概率,具有以下性质:

性质 1.1.1　(1) 不可能事件的概率为零,即 $P(\varnothing)=0$.

(2) 有限可加性:设 $A_1,A_2,\cdots,A_n$ 为两两互斥的 n 个事件,则

$$P\Big(\sum_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i). \tag{1.1.2}$$

(3) 对任意一个事件 A, $P(\overline{A})=1-P(A)$.

(4) 减法公式: $P(A-B)=P(A)-P(AB)$.

(5) 单调性:若 $A\subset B$, 则 $P(A)\leqslant P(B)$.

(6) 有界性:对任意一个事件 A, $0\leqslant P(A)\leqslant 1$.

(7) 加法公式: $P(A\cup B)=P(A)+P(B)-P(AB)$.

对三个事件和的情形,有

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(AC)+P(ABC).$$

更一般的加法公式为

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)=\sum_{i=1}^{n}P(A_i)-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P(A_iA_j)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n). \tag{1.1.3}$$

例 1.1.1　把分别标有 $1,2,\cdots,n$ 的 n 个球随机地放入分别标有 $1,2,\cdots,n$ 的 n 个盒子中,每个盒子只能放一个球. 求盒子上的号码与盒子中球的号码都

不一致的概率.

解　设 $A_i(i=1,2,\cdots,n)$ 表示"第 i 号球放入第 i 号盒子"的事件,则所求概率为 $P\left(\bigcap\limits_{i=1}^{n}\overline{A}_i\right)$. 由德摩根定律以及性质 1.1.1 中的(3)和(7),得

$$P\left(\bigcap_{i=1}^{n}\overline{A}_i\right)=P\left(\overline{\bigcup_{i=1}^{n}A_i}\right)=1-P\left(\bigcup_{i=1}^{n}A_i\right)$$

$$=1-\left[\sum_{i=1}^{n}P(A_i)-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}P(A_iA_j)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)\right].$$

而

$$P(A_i)=\frac{(n-1)!}{n!}=\frac{1}{n}=\frac{1}{\mathrm{P}_n^1},\quad i=1,2,\cdots,n,$$

$$P(A_iA_j)=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)}=\frac{1}{\mathrm{P}_n^2},\quad i\neq j,$$

$$P(A_iA_jA_k)=\frac{(n-3)!}{n!}=\frac{1}{\mathrm{P}_n^3},\quad i\neq j\neq k,$$

$$\cdots\cdots\cdots\cdots$$

$$P(A_1A_2\cdots A_n)=\frac{1}{n!}=\frac{1}{\mathrm{P}_n^n}.$$

所以,

$$P\left(\bigcap_{i=1}^{n}\overline{A}_i\right)=1-\left(\mathrm{C}_n^1\frac{1}{n}-\mathrm{C}_n^2\frac{1}{\mathrm{P}_n^2}+\cdots+(-1)^{n-1}\mathrm{C}_n^n\frac{1}{\mathrm{P}_n^n}\right)$$

$$=1-\left(1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}\right)$$

$$=\frac{1}{2!}-\frac{1}{3!}+\cdots-(-1)^{n-1}\frac{1}{n!}.$$

2. 条件概率

条件概率 $P(B\mid A)$ 描述的是在事件 A 已经发生的情况下事件 B 发生的概率,也可以理解为是对 $P(B)$ 的一种修正概率或是在掌握了新信息 A 时对事件 B 发生的概率的新认识. 当 $P(A)>0$ 时,有下面的条件概率公式

$$P(B\mid A)=\frac{P(AB)}{P(A)}.\tag{1.1.4}$$

易验证,条件概率具有如下的性质.

性质 1.1.2　(1) 非负性:对任意事件 B, 有 $P(B\mid A)\geqslant 0$.

(2) 规范性: $P(\Omega\mid A)=1$.