



全国统计教材编审委员会“十二五”规划教材

概率论与数理统计 三十三讲

第三版

魏振军 编著

 中国统计出版社
China Statistics Press



· 014009843

021
255-3

全国统计教材编审委员会“十二五”规划教材

概率论与数理统计 三十三讲

第三版

魏振军 编著



中国统计出版社
China Statistics Press



北航 C1695760

021
255-3

048000410

图书在版编目(CIP)数据

概率论与数理统计三十三讲 / 魏振军编著. — 北京:
中国统计出版社, 2013. 8

全国统计教材编审委员会“十二五”规划教材

ISBN 978—7—5037—6839—2

I. ①概… II. ①魏… III. ①概率论—高等学校—教材②数理统计—高等学校—教材 IV. ①O21

中国版本图书馆 CIP 数据核字(2013)第 133257 号

概率论与数理统计三十三讲 (第三版)

作 者/魏振军

责任编辑/徐 颖

装帧设计/上智博文

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/http://csp. stats. gov. cn

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/710×1000mm 1/16

字 数/380 千字

印 张/22. 5

印 数/1—3000 册

版 别/2013 年 10 月第 3 版

版 次/2013 年 10 月第 1 次印刷

定 价/42. 00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区
以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

国家统计局

全国统计教材编审委员会

顾 问 罗 兰 袁 卫 冯士雍 吴喜之
方积乾 王吉利 庞 皓 李子奈

主 任 徐一帆

副主任 严建辉 田鲁生 邱 东 施建军
耿 直 徐勇勇

委 员(按姓氏笔划排序)

丁立宏	万崇华	马 骏	毛有丰	王兆军
王佐仁	王振龙	王惠文	丘京南	史代敏
龙 玲	刘建平	刘俊昌	向书坚	孙秋碧
朱 胜	朱仲义	许 鹏	余华银	张小斐
张仲梁	张忠占	李 康	李兴绪	李宝瑜
李金昌	李朝鲜	杨 虎	杨汭华	杨映霜
汪荣明	肖红叶	苏为华	陈 峰	陈相成
房祥忠	林金官	罗良清	郑 明	柯惠新
柳 青	胡太忠	贺 佳	赵彦云	赵耐青
凌 亢	唐年胜	徐天和	徐国祥	郭建华
崔恒建	傅德印	景学安	曾五一	程维虎
蒋 萍	潘 璠	颜 虹		

出版说明

全国统计教材编审委员会是国家统计局领导下的、全国统计教材建设工作的最高指导机构和咨询机构,自1988年成立以来,分别组织编写和出版了“七五”至“十一五”全国统计规划教材。

“十二五”时期,是我国全面实施素质教育,全面提高高等教育质量,深化教育体制改革,推动教育事业科学发展,提高教育现代化水平的时期。“十二五”伊始,统计学迎来了历史性的重大变革和飞跃。2011年2月,在国务院学位委员会第28次会议通过的新的《学位授予和人才培养学科目录(2011)》(以下简称“学科目录”)中,统计学从数学和经济学中独立出来,成为一级学科。这一变革和飞跃将对中国统计教育事业产生巨大而深远的影响,中国统计教育事业将在“十二五”时期发生积极变化。

正是在这一背景下,全国统计教材编审委员会制定了《“十二五”全国统计教材建设规划》(以下简称“规划”)。根据“学科目录”在统计学下设有数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科的构架,“规划”对“十二五”全国统计规划教材建设作了全面部署,具有以下特点:

第一,打破以往统计规划教材出版学科单一的格局。全面发展数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科规划教材的出版,使“十二五”全国统计规划教材涵盖5个二级学科,形成学科全面并平衡发展的出版局面。

第二,打破以往统计规划教材出版层次单一的格局。在编写出版好各学科本科生教材的基础上,对研究生教材出版进行深入研究,出版一批高水平高层次的研究生教材,为我国研究生教育、尤其是应用统计研究生教育提供教学服务。同时,积极重视统计专科教材出版,联合各专科院校,组织编写和出版适应统计专科教学和学习的优秀教材。

第三,打破以往统计规划教材出版品种单一的格局。鼓励内容创新,联系统计实践,具有教学内容和教学方法特色的、各高校自编的相同内容选题的精品教材出版,促进统计教学向创新性、创造性和多样性发展。

第四,重视非统计专业的统计教材出版。探讨对非统计专业学生的统计教学问题,为非统计专业学生组织编写和出版概念准确、叙述简练、深入浅出、表达方式活泼、练习题贴近社会生活的统计教材,使统计思想和统计理念深入非统计专业学生,以达到统计教学的最大效果。

第五,重视配合教师教学使用的电子课件和辅助学生学习使用的电子产品的配套出版,促进高校统计教学电子化建设,以期最后能形成系统,提高统计教育现代化水平。

第六,重视对已经出版的统计规划教材的培育和提高,本着去粗存精、去旧加新、与时俱进的原则,继续优化已经出版的统计教材的内容和写作,强化配套课件和习题解答,使它们成为精品,最后锤炼成为经典。

“十二五”期间,编审委员会将本着“重质量,求创新,出精品,育经典”的宗旨,组织我国统计教育界专家学者,编写和编辑出版好本轮教材。本轮教材出版后,将能够形成学科齐全、层次分明、品种多样、配套系统的高质量立体式结构,使我国统计规划教材建设再上新台阶,这将对推动我国统计教育和统计教材改革,推动我国统计教育事业科学发展,提高我国统计教育现代化水平产生积极意义。

让教师的教学和学生的学习事半功倍,并使学生在毕业之后能够学以致用,是本轮教材的追求。编审委员会将努力使本轮教材好教、好学、好用,尽力使它们在内容上和形式上都向国外先进统计教材看齐。限于水平和经验,在教材的编写和编辑出版过程中仍会有不足,恳请广大师生和社会读者提出批评和建议,我们将虚心接受,并诚挚感谢!

国家统计局

全国统计教材编审委员会

2012年7月

第三版前言

——足迹、思考、交流

《概率论与数理统计三十三讲》第一版由中国统计出版社出版发行迄今已 13 个年头,值此第三版付梓之际,作者先说些并非题外的话。

辟蹊径顺势而为

20 世纪 70 年代,作者曾是一名大学基础数学课教师。伴随着国家改革开放,军校迎来大发展机遇期,概率统计在国防教育中地位日益显著。崭新的领域、独特的思维方式、广泛的应用,把作者引入求索之路。

军校得天独厚的条件,使作者在 1970 年代就结识了大型计算机。进入 80 年代,个人计算机问世后,兄弟院校在数学教学中进行了模拟试验的尝试,使作者看到计算机的神奇功能,触发了作者用计算机技术辅助概率统计教学的想法。

此后的二十几年间,由作者主持编制的 DOS 版的《概率统计演示系统》、Windows 版的《教学系统》、基于平台的《网络课程》相继问世;《概率统计电子教案》、《模拟试验》等连续推出。

开新篇独树一帜

DOS 版的计算机辅助教学软件《概率论与数理统计演示系统》1991 年 4 月在北京通过专家评审。北京师范大学教授(时任中国数学会副理事长)严士健等看过耳目一新的软件后曾为之欢欣鼓舞。

评审会后,作者先后受邀与著名统计学家张尧庭、茆诗松合作著书;1990年代中期作者主持研发的《概率论与数理统计》教学系统被批准列入国家教委“九五”国家重点科技攻关项目(计算机辅助教学软件研制开发与应用);1997年9月国家教委数学力学指导委员会对“教学系统”组织了鉴定。此后,教学系统软件在几百所院校使用过程中受到用户普遍欢迎。有的用户提出了修改意见,有的则建议作者编写一本与软件配套的教材。

接受热情用户的意见,结合多年的教学体会,吸收合作著书的一些成果,作者赶在世纪之交写出第一版《概率论与数理统计三十三讲》。张尧庭教授认真审阅书稿后,认为该书“是国内一本很具特色的教材”,他归纳了四个特色:

1. 是结合实际的教材。这一点从书中引用的大量实例可以看出,它重点教人怎么用。

2. 与计算机结合紧密。国内已有的教材中,没有一本能结合得如此好。

3. 图文并茂。使教材可读性大大增加。

4. 文字语言生动而有启发性。这只要看几页就会感受到的,不是一般教科书那种死板的文字。

基于此,他认为“这是一本优秀的大学教材”。

师生情没齿难忘

一路走来,近30年的教改岁月和三版次成书历程,作者对前辈学者的悉心指导常怀感恩之情,对本单位给予过支持帮助的同志深表谢意。

作者的老师,1980年代分别任院、系领导的蒋维汉、赵人杰教授对作者所进行的改革曾给予充分肯定与支持;1990年代严士健教授满腔热情地向教育界、出版界推荐,使作者的教改成果迅速扩大了影响;张尧庭教授两度主持会议,对作者的成果进行鉴定,合作编写教材时,专程到郑州当面指导。目前,严士健、蒋维汉、赵人杰都已是一抔黄土,张尧庭教授则在六年前驾鹤西去。

值此新版问世之际,作者特别怀念早逝的张教授。严格地说,作者与张教授并无直接师生关系,推进概率统计教学改革的心愿

景使我们超过寻常师生关系,在学术上他热情指导,关键时刻他仗义执言,让我感受到大学问家的风范,使他成为作者心目中的“恩师”。严士健教授与作者不属同一单位,他倾尽全力支持作者的工作,让作者没齿难忘。

致读者

从第一版到第三版的13年间,国内外情况都发生了许多重大变化。计算机技术在各层次教学中得到广泛应用,新知识、新内容、新方法不断涌现;国内基础数学教育改革取得阶段性成果,部分概率统计内容已经出现在中小学数学课本中。网络环境催生了新的统计软件,也使教学手段发生一些变化,案例教学、情景教学等被普遍采用,高等院校的概率统计教学迈上了新的台阶。新版教材秉持与时俱进的原则,在内容上进行了一些更新。

从粗线条看,章节结构基本未变,统计内容有所加强。第一章(基础概率)由原来的9讲浓缩为7讲,在第四章增加了条件数学期望1讲,第九章增加了方差分析初步1讲。全书更加突出了概率统计的思维方式及其实用性和趣味性。除第三章多维随机变量及其分布改动较小外,其他各章的写法都有变化。第八章假设检验一讲作了较大的改进,比前版更突出介绍基本思想,明确给出统计显著性的概念,在判定是否拒绝原假设时主要采用 p 值法。

此次改版,对前版的思考与练习题也作了部分修改,增加了联系实际的问题。

为了使读者更好理解所讲述的基本概念、原理和方法,在教材中经常引领读者参与一些交互试验,这些试验都链接在与教材配套的电子教案中,希望读者在学习时不要忽视这一环节。此外,在教材讲述中适当穿插的“停下来想一想”的问题,若能认真思考讨论,也会加深对相关知识的理解。

尽管本书的电子教案链接了与教材内容相关的大部分模拟试验和演示,但作者仍希望有条件的读者在计算机上自主编写程序,把理论与实际相结合。目前流行的统计软件有多个,本书着重推荐的是免费的R。因此在本书附录一中,结合教材相关内容,作者简要介绍了如何用R语言进行统计计算及计算机模拟,同时在教材的部

分案例中使用 R 语言编写的程序。利用 R 在模拟、计算、作图和数据处理方面的强大功能,相信会给本课程的学习注入新的活力。也为读者进一步学习应用 R 打下一个初步基础。

配合第三版教材,作者重新制作了《电子教案》,电子教案更为生动活泼,利于学生自主学习和教师备课时参考。需要说明的是,教案链接的试验需要安装必要的插件,下载包中包含有安装程序,执行其中的 setup. exe 即可顺利安装。对装有 win7 版本(正版)的计算机还须再安装支持 VB 程序的软件才能运行这些试验。可独立运行的自测题以选择题的形式出现,供学生复习基本概念使用。

以上资源(包括与教材配套的几个数据文件)放在统计出版社网站(网址:<http://csp.stats.gov.cn>),读者可以下载。

在本书的编写过程中,作者曾参阅过大量国内外教材和资料,凡出处明确的都在参考文献中一一列出。作者非常感谢众多同行、读者对作者的鼓励和支持。国家统计局教育培训中心和中国统计出版社的同志对作者的工作给予了很多帮助。本科毕业生于本校的刘璐博士曾认真细致地校阅书稿,帮助改正了一些疏漏之处。在此一并向他们表示诚挚的谢意。受本人水平所限,新版教材及电子教案错谬之处难免,作者恳请各位老师和同学给予批评指正。

魏振军

2013 年 7 月

目 录

第一章 基础概率	(1)
第一讲 概率论的研究对象	(1)
思考与练习(1)	(6)
第二讲 随机事件及其概率	(6)
思考与练习(2)	(12)
第三讲 确定概率的几种方法	(13)
思考与练习(3)	(25)
第四讲 概率的公理化定义与性质	(26)
思考与练习(4)	(33)
第五讲 条件概率和乘法公式	(34)
思考与练习(5)	(39)
第六讲 全概率公式和贝叶斯公式	(40)
思考与练习(6)	(44)
第七讲 事件的独立性	(45)
思考与练习(7)	(49)
第二章 随机变量及其分布	(50)
第八讲 随机变量的概念	(50)
思考与练习(8)	(53)
第九讲 离散随机变量及其概率函数	(53)
思考与练习(9)	(57)
第十讲 连续随机变量及其概率密度函数	(58)
思考与练习(10)	(64)
第十一讲 分布函数	(65)
思考与练习(11)	(70)
第十二讲 随机变量函数的分布	(71)
思考与练习(12)	(75)
第十三讲 二项分布	(76)
思考与练习(13)	(81)

第十四讲	泊松分布	(82)
	思考与练习(14)	(88)
第十五讲	正态分布	(89)
	思考与练习(15)	(96)
第三章 多维随机变量及其分布		(97)
第十六讲	随机向量、联合分布和边缘分布	(97)
	思考与练习(16)	(105)
第十七讲	随机变量的独立性	(106)
	思考与练习(17)	(109)
第十八讲	条件分布	(111)
	思考与练习(18)	(115)
第十九讲	随机向量函数的分布	(117)
	思考与练习(19)	(125)
第四章 随机变量的数字特征		(127)
第二十讲	随机变量的数学期望	(127)
	思考与练习(20)	(138)
第二十一讲	随机变量的方差	(140)
	思考与练习(21)	(146)
第二十二讲	协方差与相关系数	(147)
	思考与练习(22)	(155)
第二十三讲	条件数学期望	(156)
	思考与练习(23)	(163)
第五章 大数定律与中心极限定理		(164)
第二十四讲	大数定律	(164)
	思考与练习(24)	(168)
第二十五讲	中心极限定理	(169)
	思考与练习(25)	(176)
第六章 数理统计的基本概念		(178)
第二十六讲	总体、样本和统计量	(179)
	思考与练习(26)	(190)
第二十七讲	抽样分布	(191)
	思考与练习(27)	(201)

第七章 参数估计	(203)
第二十八讲 参数点估计	(203)
思考与练习(28)	(217)
第二十九讲 区间估计	(219)
思考与练习(29)	(234)
第八章 假设检验	(237)
第三十讲 假设检验	(237)
思考与练习(30)	(259)
第三十一讲 拟合优度的 χ^2 检验	(262)
思考与练习(31)	(269)
第九章 方差分析与回归分析初步	(270)
第三十二讲 方差分析初步	(270)
思考与练习(32)	(281)
第三十三讲 回归分析初步	(282)
思考与练习(33)	(296)
附录一 用 R 软件进行统计计算和计算机模拟	(298)
附录二 排列与组合	(310)
附录三 常见概率分布表	(314)
附表 1 泊松分布表	(316)
附表 2 标准正态分布函数数值表	(318)
附表 3 t 分布分位数表	(319)
附表 4 χ^2 分布分位数表	(320)
附表 5 F 分布分位数表	(322)
总复习题	(324)
各章部分思考与练习参考解答	(331)
参考书目	(345)

第一章 基础概率

内容提要 作为开篇,本章从生活中的随机现象开始,引入随机事件、概率、条件概率、独立性等概念;介绍了常用的几种确定概率的方法、概率的公理化定义和性质;给出了计算概率的加法公式、乘法公式、全概率公式和贝叶斯公式。

第一讲 概率论的研究对象

概率论是研究机会的数学.最初是对掷骰子等赌博游戏中机遇问题的研究,而现在已经成为一门理论严谨、方法独特的充满活力的数学学科.在工农业生产、通信、气象、国防、生物、经济和金融等诸多领域都得到了广泛的应用,并渗透到现代生活的方方面面.正如 19 世纪法国著名数学家拉普拉斯所说:

“生活中最重要的问题,绝大部分其实只是概率问题.你可以说几乎我们所掌握的所有知识都是不确定的,只有一小部分我们能确定地了解.甚至数学科学本身,归纳法、类推法和发现真理的首要手段都是建立在概率论的基础之上.因此,整个人类知识系统是与这一理论相联系的……”.

概率论诞生以来,科学家们都从概率这门学问里寻找处理我们这个不确定的世界的工具,计算某事件发生的概率,就好象窥视未来一样.让我们携手徜徉在五光十色的世界,探寻其中的奥秘.

先看一个短片“**概率论的研究对象**”(在电子教案中).

一、随机现象

人们所关心的一场足球赛就要开始了,这是一场实力相当的比赛,可能甲队赢,也可能乙队赢,事先我们不能准确地预言哪个球队能取胜.

在十字交叉口,每天都要通过许多人和车辆,但是我们无法事先预测每天确切的人数及车辆数.

某城市一天中发生交通事故的次数,可能是 1 次,2 次,3 次,……,连交警也无法准确预言明天将会发生几起交通事故.

你家买回了一台电视机,使用多少时间后它会出故障呢?也许多于 10000 小时,也许少于 10000 小时.事先谁也无法准确预言.

你考上了大学,即将到新的班级生活,在报到前你能准确地预知新同学中的最大身高是多少吗?

1986 年 1 月 28 日,震惊世界的美国“挑战者”号航天飞机升空不久便发生爆炸,2003 年 2 月 1 日,“哥伦比亚”号航天飞机即将返回地面时发生爆炸.这两件令世界震惊的事件事先谁也无法料到.

好吧!让我们来做一个试验.

这里有十个签,其中只有一张是彩票,也就是说抽到这张签的人可以获得许多的好处.

小张:“当然啦,这类活动总是少不了我的,而且到目前为止,我一直运气不赖,不信您瞧!”

(没抽到彩票)“真倒霉!运气太坏了!”

上述这些现象,在一定条件下,并不总是出现相同的结果.我们称之为随机现象.随机现象有什么特点呢?

1. 相同条件下的结果不止一个;
2. 人们事先无法确知出现哪一个结果.

显然,当我们不能预测某件事的结果时,这件事就和随机性联系起来了.不难发现,我们的生活和随机现象结下了不解之缘.

二、随机现象的统计规律性

很多现象看起来偶然,实际上有一定的必然性.

让我们来看几个例子:

例 1.1 在掷一枚质地均匀的硬币时,既可能出现正面,也可能出现反面,但事先我们无法知道到底会出现哪一面.但是在大量的重复试验后,人们发现,出现正面与出现反面的次数竟然几乎各占一半.

法国科学家蒲丰做过试验,他掷硬币 4040 次,结果出现正面的频率为 50.69%.

英国科学家皮尔逊,通过两组上万次重复试验,进一步得出,出现正面的频率分别为 50.16%和 50.05%.

事实告诉我们,抛掷次数越多,一般来说,出现正面的频率就越接近于 $1/2$.

例 1.2 在医院里,新生婴儿按他们出生次序登记.男孩和女孩 (B 和 G) 一个接一个没有明显规律,如

GBBGBGBBGGBBBGG

虽然我们不能预言这个随机序列的细节,即在任何个别情况下,我们不能预言新生婴儿的性别.但是,对大量新生婴儿的性别作统计后可以相当准确地预测每年新生婴儿中生男生女的比例.

法国数学家拉普拉斯(Laplace,1749~1827)在他的名著《概率论的哲学探讨》中研究了男婴出生的频率,他对伦敦、彼得堡、柏林和全法国的大量人口资料进行研究,发现男婴出生率总在一个数左右波动,这个数大约是 22/43.

另一位统计学家克拉美在他的名著《统计学数学方法》中引用瑞典 1935 年的官方统计资料,该资料表明,女婴出生的频率稳定在 0.482 左右,男婴出生的频率为 0.518.

资料表明,在不受干扰的情况下,世界各国出生婴儿性别比是基本一致的,一般在(103~107):100 的范围内,即每出生 100 个女婴,对应的男婴出生数为 103~107 个.人口学家把这个指标称为出生婴儿性别比的恒定值.由于男婴死亡率略高于女婴,这样到婚龄期时男女性别比例就基本持平.

但是,若受到其他因素干扰,就很可能偏离这个比例.如我国 2000 年第五次和 2010 年第六次全国人口普查数据表明,全国出生婴儿男女性别比分别为 117:100 和 118.06:100.针对此异常情况,国家正在采取综合措施治理,争取尽早回归人口出生性别正常比值.

例 1.3 让我们随意选择一本英文书,莎士比亚的剧作也行,华莱士的侦探小说也好.任意选择其中的一句话,每个字母出现的次数是随机的,事先无法精确估计.但是就整本书或其中的一大段进行统计,却发现某些字母的出现次数远远高于另一些字母.在进行了更深入地研究之后,人们还发现各个字母被使用的频率相当稳定.下面是英文字母使用频率的一份统计表.

字母	频率	字母	频率	字母	频率
E	0.1268	L	0.0394	P	0.0186
T	0.0978	D	0.0389	B	0.0156
A	0.0788	U	0.028	V	0.0102
O	0.0776	C	0.0268	K	0.006
I	0.0707	F	0.0256	X	0.0016
N	0.0706	M	0.0244	J	0.001
S	0.0634	W	0.0214	Q	0.0009
R	0.0594	Y	0.0202	Z	0.0006
H	0.0573	G	0.0187		

你可知道,著名的侦探福尔摩斯在破译一份密码时,恰恰是运用了字母的这种规律!

一次,福尔摩斯受理破译一份密码(见图 1.1),这份密码与以往的方式完全不同,上面画的全是些会跳舞的小人.

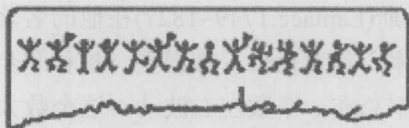


图 1.1 第一张纸条

怎样破译这份密码呢?请看福尔摩斯的回答:

“只要看出了这些符号是代表字母的,再应用字母的规律来分析,就不难找到答案.”

字母的规律是什么呢?

人们通过大量的试验发现,在英文文章中,每个字母出现的频率是相当稳定的.比如,据一份英文频率的统计表(不同的统计表可能有些差异,但大致相同):字母 *E* 出现的频率最高,是 12.7%,*T* 为 9.78%,*A* 为 7.88%,*O* 为 7.76%,*I* 为 7.07%,*N* 为 7.06%,字母 *Z* 出现的频率最小,只有 0.6%.

福尔摩斯恰恰就是运用了这种规律,破出了这份神秘的密码,抓住了凶手.

让我们从《福尔摩斯探案全集》中找出有关的内容:

“第一张纸条上的话很短,我只能稍有把握假定 $\times$ 代表 *E*,*E* 是英文中最常见的字母,它出现的次数即使在一个短的句子中也是最常见的.第一张纸条上的 15 个符号,其中有 4 个完全一样,因此把它估计为 *E*.这些图形中,有的还带一面小旗,有的没有小旗.从小旗的分布来看,带旗的图形可能是用来把这个句子分成一个一个的单词.我把这看作一个可以接受的假设,同时记下 *E* 是用 $\times$ 来代表的.”

“可是,现在最难的问题来了,因为,除了 *E* 之外,英文字母出现次数的顺序并不很清楚,这种顺序,在平常一页印出的文字里和一个短句子里,可能正相反.一般来说,字母按出现次数排列的顺序是

T, A, O, I, N, S, H, R, D, L, ...

但是 *T, A, O, I* 出现的次数几乎不相上下.要是把每一种组合都试一遍,直到得出一个意思来,那会是一项无止境的工作.所以我们只好等来了新材料再说.”

之后,又有几张纸条来到福尔摩斯手中(见图 1.2),它们是这样的: