



“十二五”普通高等教育本科国家级规划教材

概率论 与数理统计

第二版

何书元

010 010
010100
0 10010 10
0101101010 1
011 0 100 1 10100 10100
0110 01 101 10 0 1
011 01 01
01 010 00
010 0 010
100 0 00 101
1001101010
010 0101 010
010100 01

 高等教育出版社
HIGHER EDUCATION PRESS



“十二五”普通高等教育本科国家级规划教材

概率论 与数理统计

第二版

Gailülun yu Shuli Tongji

何书元



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

本书较系统地介绍了概率论和数理统计的基本内容,内容丰富,富有时代特色.书中有许多新的简明讲法,帮助学生更好地理解所学内容和加深对问题本质的理解.

本书有许多反映现代科技和现代生活特点的例子,包括赌博问题、运气问题、求职问题、医药疗效问题、敏感问题调查、碳 14 方法、钾氩比方法等.本书讲授的微分法,是计算随机变量和随机向量函数分布的简捷新方法.条件分布和边缘分布的计算方法也都简单易行,较大程度地降低了数学难度.在判断随机变量的独立性方面,也有十分简单的新方法.

为了帮助读者更快地掌握计算机的使用,本书以工程技术和科学研究中普遍使用的 MATLAB 为例,在相关章节后面介绍有关的 MATLAB 调用命令.

本书的内容和习题难度适中,适合作为理工科大学、师范和财经院校非数学类专业本科生概率论与数理统计课程的教材或教学参考书.学习本书的先修课程是高等数学.

图书在版编目(CIP)数据

概率论与数理统计 / 何书元编. —2 版. —北京:
高等教育出版社, 2013. 3

ISBN 978 - 7 - 04 - 036939 - 7

I. ①概… II. ①何… 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 022958 号

策划编辑 李蕊 责任编辑 胡颖 封面设计 张楠
版式设计 童丹 责任印制 韩刚

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街 4 号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.landaco.com
印 刷	高教社(天津)印务有限公司		http://www.landaco.com.cn
开 本	787mm × 960mm 1/16	版 次	2006 年 6 月第 1 版
印 张	21.5		2013 年 3 月第 2 版
字 数	400 千字	印 次	2013 年 3 月第 1 次印刷
购书热线	010 - 58581118	定 价	29.30 元
咨询电话	400 - 810 - 0598		

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 36939 - 00

作者简介

何书元博士现任首都师范大学教授、北京大学数学科学学院兼职教授，曾任教育部数学与统计学教学指导委员会副主任委员、统计学教学指导分委员会主任委员、北京大学数学科学学院副院长和概率统计系主任。从事概率论与数理统计的教学和科研工作，主讲的课程有概率论、数理统计、应用随机过程、应用时间序列分析等，是北京大学 2006 年国家精品课程数理统计的课程主持人和主讲教师，1997—2009 年任北京大学教授。

第二版前言

本书的第一版被评为 2007 年度普通高等教育精品教材。此事给我带来不安，因为每次使用此书时常感到还有值得改进的地方。和第一版比较，本书有下列改动：

1. 在保证主要内容不少的情况下，降低了数学难度。特别体现在一些主要结论的证明上。比如抽签原理（见 §1.2 例 2.4）的证明，原来的归纳法证明的确值得学习，但是现在的证明更能帮助读者理解抽签原理的实质。

2. 利用微分法计算随机向量函数的概率密度运算简单，能把许多复杂的积分运算省略，第一版列在了附录中。现在有了这一方法的数学证明，所以放入正文。微分法的叙述更加简单，使用和记忆更加方便。

3. 在实际问题中，是选用双边假设还是单边假设、在单边原假设中的不等号如何设计往往困扰初学者。本书通过举例把这一问题交待得更加清楚。

4. 新增了分布函数检验的 Q-Q 图方法。这一方法在现实中使用最多也最方便，而许多概率统计书中缺少这一内容。

5. 新增了列联表的假设检验，因为相应问题也常在实际中遇到。

6. 增加了方差分析一章。

7. 新增了应用举例一章。这一章能帮助读者进一步理解概率统计在实际中的应用。认真阅读这一章，哪怕略去数学的内容，都会有收获。

8. 将章节进一步划分。减少了每节的内容，降低了各节的难度。

9. 例子更加贴近现实生活，方便读者联系现实。例子和习题也有了较多的改进。

可以用 χ^2 分布的密度曲线大致描述本书的难度轮廓。从简单开始逐步增加，约在五分之二处到顶，以后逐渐降低。

编者的本意要进一步压缩概率的内容，特别是随机向量的内容。但是这部分内容又常出现在考研考题中，所以只好作罢。

本书的前十二章是针对一个学期的课程（每周四学时）设计的。每周三学时的课程可略去带 * 的内容和第十一章。

借此机会，向使用过本书第一版的老师和同学们表示感谢。你们的任何意见对编者都很有帮助。

本书的再版得到教育部 2010 年度国家级教学团队项目和北京市本科数学基

础课程教学团队项目的资助，特致谢意。

需要授课使用的简易 PDF 文件时，可向 syhe@math.pku.edu.cn 或 syhe@cnu.edu.cn 索取。

由于作者水平有限，书中难免不妥之处，请读者不吝指正。

何书元

2012 年 11 月于北京田村丹青府

第一版前言

概率是描述随机事件发生的可能性的度量. 概率论通过对简单随机事件的研究, 逐步进入复杂随机现象规律的研究, 是研究复杂随机现象规律的有效方法和工具. 概率论还是学习数理统计的基础.

作为统计学的基础, 数理统计已经渗透到理、工、农、医、经济管理与人文社会科学领域, 并由此产生许多新的交叉学科, 改变了原有单一学科发展的思路, 对各学科的发展具有极大的支撑作用. 统计学没有自己的基于试验的专门研究对象, 但是可以为物理学家、化学家、医生、社会学家、心理学家等提供一套研究他们的问题的有效方法. 这套方法可以帮助各个领域的研究工作者更快地获得成功. 因此, 统计学被称为 21 世纪最有发展前途的学科之一.

多做习题是学好概率论和数理统计的基础. 本书列出了较多的难易适中的习题供学生选择. 习题配有答案, 对于技巧性较高的题目还给出提示. 使用本书的同学应当尝试完成每一道习题.

在北京大学, 编者多次为历史学系、经济学院、生命科学学院、化学与分子工程学院、物理学院、地球与空间科学学院等讲授概率统计课程, 参考了较多的国内外优秀教材, 通过素材的积累, 逐步形成了本书的框架. 本书的内容选择, 包括例子和习题的选择较多地考虑了以后使用或继续学习统计学的需要. 许多例子和习题的结论在今后的学习甚至工作中都是有用的. 除了概率论和数理统计的基本内容外, 本书还力图通过较多的例子和习题介绍概率论与数理统计的众多应用领域.

为了帮助学生在学习中掌握计算机的使用, 本书以工程技术和科学研究中广泛使用的 MATLAB 为例, 介绍和本书内容相关的 MATLAB 调用命令. 参考这些简单的调用命令, 同学们就可以处理和本书内容相关的实际数据了. 但是这些内容属于同学们的课外兴趣部分, 不属于课程的要求范围.

根据经验, 使用本书作为教材时的授课进度可大致如下 (仅供参考).

第一章 12% 学时, 基本内容是古典概型、事件的独立性、条件概率、全概率公式与贝叶斯公式.

第二章 8% 学时, 基本内容是随机变量及其独立性、概率分布、连续型随机变量的概率密度.

第三章 10% 学时, 基本内容是随机向量及其联合分布、随机向量函数的分

布.

第四章 12% 学时, 基本内容是数学期望及其性质、方差、协方差、相关系数、协方差矩阵的计算.

第五章 8% 学时, 基本内容是多元正态分布、弱强大数律和中心极限定理.

第六章 6% 学时, 介绍总体和参数、随机抽样的重要性用样本描述总体、随机对照试验.

第七章 15% 学时, 基本内容是矩估计、最大似然估计、区间估计.

第八章 15% 学时, 基本内容是假设检验的概念、正态均值和方差的假设检验、总体分布的假设检验.

第九章 14% 学时, 基本内容是数据的相关性、一元线性回归和多元线性回归介绍.

全书是针对一个学期的课程 (每周四学时) 设计的. 每周三学时的课程应略去带 * 的内容. 不带 * 的部分是本书的主要内容.

除了编者写作的内容外, 本书的部分例子和习题参考了书后所列的参考书. 作者在这里对这些书的作者表示感谢.

由于作者水平有限, 书中难免不妥之处, 请读者不吝指教.

何书元

2006 年 1 月于北京海淀蓝旗营

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第一章 概率的定义	(1)
§1.1 有限样本空间	(1)
A. 有限样本空间及其事件	(2)
B. 用等可能性定义概率	(4)
§1.2 古典概率模型	(5)
§1.3 概率的公理化	(8)
§1.4 概率与频率	(10)
习题一	(11)
第二章 概率公式	(13)
§2.1 加法公式	(13)
§2.2 事件的独立性	(15)
§2.3 条件概率和乘法公式	(17)
§2.4 全概率公式	(20)
§2.5 贝叶斯公式	(23)
习题二	(27)
第三章 随机变量	(30)
§3.1 随机变量及其分布函数	(30)
A. 随机变量	(30)
B. 分布函数和概率密度	(30)
§3.2 离散型随机变量	(32)
§3.3 连续型随机变量	(39)
§3.4 随机变量函数的分布	(45)
用 MATLAB 产生随机数	(50)
用 MATLAB 计算概率分布函数和密度	(50)
习题三	(51)
第四章 随机向量	(54)
§4.1 随机向量	(54)
§4.2 离散型随机向量	(56)
§4.3 连续型随机向量	(60)

A. 联合密度	(60)
B. 边缘密度	(62)
C. 独立性	(63)
§4.4 随机向量函数的分布	(67)
A. 离散型随机向量的函数	(67)
B. 连续型随机向量函数的分布	(69)
§4.5 随机向量函数的联合密度	(71)
§4.6 二维正态分布	(73)
§4.7 条件分布	(75)
习题四	(80)
第五章 数学期望和方差	(82)
§5.1 数学期望	(82)
A. 数学期望的定义	(83)
B. 数学期望的统计含义	(86)
§5.2 常用的数学期望	(86)
§5.3 数学期望的计算	(89)
§5.4 数学期望的性质	(91)
§5.5 随机变量的方差	(94)
A. 常用的方差	(95)
B. 方差的性质	(98)
§5.6 协方差和相关系数	(100)
A. 内积不等式	(100)
B. 协方差和相关系数	(101)
C. 协方差矩阵	(103)
§5.7 正态分布的参数计算	(104)
习题五	(105)
第六章 大数律和中心极限定理	(109)
§6.1 强大数律	(109)
§6.2 切比雪夫不等式	(110)
§6.3 中心极限定理	(112)
习题六	(115)
第七章 描述性统计	(117)
§7.1 总体和参数	(117)

A. 总体、个体和总体均值	(117)
B. 样本与估计	(118)
§7.2 抽样调查	(120)
A. 抽样调查的必要性	(120)
B. 随机抽样	(121)
C. 随机抽样的无偏性	(122)
D. 分层抽样方法	(124)
E. 系统抽样方法	(125)
§7.3 用样本估计总体分布	(126)
A. 频率分布表	(126)
B. 频率分布直方图	(128)
C. 频率折线图	(129)
D. 数据茎叶图	(130)
§7.4 众数和中位数	(132)
A. 众数	(132)
B. 中位数	(133)
§7.5 随机对照试验	(135)
用 MATLAB 计算样本均值、样本标准差, 绘制直方图	(140)
习题七	(140)
第八章 参数估计	(143)
§8.1 样本均值和样本方差	(143)
A. 样本均值	(144)
B. 样本方差	(145)
C. 样本标准差	(146)
§8.2 矩估计	(147)
§8.3 最大似然估计	(150)
A. 离散分布的情况	(150)
B. 连续分布的情况	(153)
C. 矩估计和 MLE 的比较	(156)
习题八	(157)
第九章 参数的区间估计	(160)
§9.1 一个正态总体的区间估计	(160)
A. 已知 σ 时, μ 的置信区间	(160)
B. 未知 σ 时, μ 的置信区间	(164)

C. 方差 σ^2 的置信区间	(166)
D. 单侧置信限	(169)
§9.2 两个正态总体的区间估计	(172)
A. 均值差 $\mu_1 - \mu_2$ 的置信区间	(172)
B. 方差比 σ_1^2/σ_2^2 的置信区间	(174)
*§9.3 非正态总体的置信区间	(176)
A. 正态逼近法	(176)
B. 比例 p 的置信区间	(177)
C. 样本量的确定	(178)
§9.4 置信区间小结	(180)
置信区间表	(182)
用 MATLAB 计算置信区间	(182)
用 MATLAB 计算上 α 分位数	(183)
习题九	(184)
第十章 正态总体的显著性检验	(187)
§10.1 假设检验的概念	(187)
§10.2 正态均值的显著性检验	(190)
A. 已知 σ 时, μ 的检验	(190)
B. 未知 σ 时, μ 的检验	(192)
C. 未知 σ 时, μ 的单边检验	(194)
§10.3 均值比较的显著性检验	(197)
A. 已知 σ_1^2, σ_2^2 时, μ_1, μ_2 的检验	(198)
B. 已知 $\sigma_1^2 = \sigma_2^2$ 时, $\mu_1 - \mu_2$ 的检验	(199)
C. 成对数据的假设检验	(201)
D. 未知 σ_1^2, σ_2^2 时, μ_1, μ_2 的大样本检验	(202)
§10.4 方差的显著性检验	(203)
正态总体的显著性检验表	(207)
习题十	(209)
第十一章 总体分布和比例的假设检验	(212)
§11.1 总体分布的假设检验	(212)
A. Q-Q 图	(212)
B. 拟合优度检验	(214)
§11.2 非正态总体的显著性检验	(218)
A. 比例 p 的假设检验	(219)

B. 两个总体比例的比较	(220)
§11.3 列联表的独立性检验	(224)
A. 2×2 列联表	(224)
B. $k \times l$ 列联表	(228)
*§11.4 P 值检验和验收检验	(230)
A. P 值检验	(230)
B. 验收检验	(233)
正态逼近法的假设检验法列表	(236)
习题十一	(236)
第十二章 线性回归分析	(240)
§12.1 数据的相关性	(240)
A. 样本相关系数	(241)
B. 相关性检验	(244)
§12.2 回归直线	(247)
§12.3 一元线性回归	(250)
A. 最大似然估计和最小二乘估计	(251)
B. 平方和分解公式	(252)
C. 斜率 b 的检验	(254)
D. 预测的置信区间	(256)
E. 应用案例	(258)
习题十二	(262)
第十三章 方差分析	(264)
§13.1 单因素方差分析	(264)
A. 平方和分解	(265)
B. 检验方法	(267)
C. 方差分析表	(268)
D. 参数估计	(271)
§13.2 双因素方差分析	(273)
A. 双因素方差分析模型	(273)
B. 检验方法	(275)
§13.3 无重复试验的双因素方差分析	(279)
习题十三	(282)
*第十四章 应用举例	(284)
§14.1 运气问题	(284)

A. 游程数的概率分布	(285)
B. 游程长度的概率分布	(286)
§14.2 求职问题	(287)
§14.3 医生用药问题	(290)
§14.4 基因遗传问题	(291)
§14.5 猜奖问题	(292)
§14.6 惊人的预测	(295)
§14.7 高考作文的评分质量控制	(297)
附录 A 组合公式与微积分	(299)
附录 B 常见分布的均值、方差、母函数和特征函数	(302)
附录 C1 标准正态分布表	(303)
附录 C2 标准正态分布的上 α 分位数表	(305)
附录 C3 t 分布的上 α 分位数表	(306)
附录 C4 χ^2 分布的上 α 分位数表	(307)
附录 C5 F 分布的上 α 分位数表	(309)
部分习题参考答案和提示	(315)
名词索引	(323)
符号说明	(327)
参考书目	(328)

第一章 概率的定义

在判断一个事件是否会发生的时候，人们实际上是在关心该事件发生的可能性的大小。在概率论中，我们用概率 (probability) 衡量一个未来事件发生的可能性的。本章介绍在某些特定情况下，如何计算一个未来事件发生的概率。为了学习概率，需要学习事件和概率空间 (probability space) 的概念。

§1.1 有限样本空间

通常把按照一定的想法去做的事情称为试验。下面都是试验的例子：掷一个硬币，掷两枚骰子 (分别标有 $1, 2, \dots, 6$ 的正六面体)，在一副扑克牌中随机抽取两张，进行概率论课程的期中考试，观测明天的最高气温。

作试验的目的是考察试验出现的可能结果。掷一枚硬币时，用 H (head) 表示硬币正面朝上，用 T (tail) 表示硬币反面朝上。本试验的可能结果是 H 和 T 。在概率论中，称 H 和 T 是样本点，称样本点的集合 $\Omega = \{H, T\}$ 为试验的样本空间。

掷一枚骰子时，用 1 表示掷出点数 1 ，用 2 表示掷出点数 2 ， \dots ，用 6 表示掷出点数 6 。本试验的可能结果是 $1, 2, 3, 4, 5, 6$ ，每个数都是试验的样本点。称样本点的集合

$$\Omega = \{\omega \mid \omega = 1, 2, \dots, 6\}$$

是试验的样本空间。

向区间 $(0, 1)$ 中投掷一个质点，用 ω 表示落点，则 ω 是样本点，试验的样本空间是 $\Omega = (0, 1)$ 或

$$\Omega = \{\omega \mid \omega \in (0, 1)\}.$$

为了叙述的方便和明确，下面把一个特定的试验称为试验 S 。称试验 S 的可能结果为 **样本点** (sample point)，用 ω 表示。称试验 S 的样本点 ω 构成的集合为 **样本空间** (sample space)。用 Ω 表示样本空间时，有

$$\Omega = \{\omega \mid \omega \text{ 是试验 } S \text{ 的样本点}\}.$$

A. 有限样本空间及其事件

如果样本空间 Ω 只有有限个样本点, 即

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\},$$

则称 Ω 是有限样本空间.

投掷一枚骰子的样本空间是

$$\Omega = \{\omega \mid \omega = 1, 2, \dots, 6\}.$$

这是一个有限样本空间. 用集合 $A = \{3\}$ 表示掷出 3 点, 则 A 是 Ω 的子集. 以后称 A 是事件. 如果掷出 3 点, 则称事件 A 发生, 否则称事件 A 不发生. 用集合 $B = \{2, 4, 6\}$ 表示掷出偶数点, B 是 Ω 的子集, B 也是事件. 如果掷出偶数点, 则称事件 B 发生, 否则称事件 B 不发生. 事件 B 发生和掷出偶数点是等价的.

如果试验 S 的样本空间 Ω 是有限集合, 则称 Ω 的子集为 **事件**(event). 如果试验的结果 $\omega \in A$, 则称事件 A 发生, 否则称 A 不发生.

还可以用集合的语言描述有限样本空间: 试验 S 的样本空间 Ω 是一个全集, Ω 的元素 ω 是样本点, 样本点是试验的可能结果. Ω 的子集是事件. 对于 $A \subset \Omega$, 如果元素 (试验的结果) $\omega \in A$, 则称事件 A 发生, 否则称 A 不发生.

本书对于子集符号 “ \subset ” 和 “ \subseteq ” 不加区分, 统一使用 “ \subset ”. $A \subset B$ 只表示 A 是 B 的子集, 不表示是真子集.

按照上面的规定, 事件是样本空间 Ω 的子集. 通常用大写字母 A, B, C, D 或 $A_1, A_2, \dots, B_1, B_2, \dots$ 等表示事件.

对集合可以进行集合运算, 其结果仍然是集合. 由于事件是子集, 所以可以对事件进行集合的运算, 其结果仍然是事件. 用 $\bar{A} = \Omega - A$ 表示集合 A 的余集, 则事件 A 发生和试验结果 $\omega \in A$ 是等价的, 事件 A 不发生和试验结果 $\omega \in \bar{A}$ 是等价的.

例 1.1 掷一个骰子, 用 A 表示掷出奇数点, 用 B 表示掷出偶数点. 写出试验的样本空间, 事件 A 和事件 B .

解 样本空间是 $\Omega = \{j \mid 1 \leq j \leq 6\}$,

$$A = \{1, 3, 5\}, \quad B = \{2, 4, 6\}.$$

例 1.2 投掷两枚硬币, 写出试验的样本点和样本空间.