高等学校教材

概率论与数理统计

主　编　韩文忠
副主编　刘官厅　李秀兰　米宇平　斯日古楞

高等教育出版社

高等学校教材

概率论与数理统计

Gailülun yu Shuli Tongji

主　编　韩文忠

副主编　刘官厅　李秀兰　米宇平　斯日古楞

高等教育出版社·北京

内容提要

本书包含十章内容：随机事件及其概率、随机变量及其分布、多维随机变量及其分布、随机变量的数字特征、大数定律和中心极限定理、数理统计的基础知识、参数估计、假设检验、一元回归分析和 Excel 在概率统计中的应用。考虑到不同读者的需要，在内容安排、例题和习题编排等方面尽可能做到由浅入深、易于计算，侧重基本思想、基本概念的理解及基本方法的综合应用。每章附有习题和自测题，收录了较多的历年全国硕士研究生入学统一考试数学试题，书末附有参考答案。

本书可作为高等学校理科、工科、经济、管理等专业本科生的教材或教学参考书，也可作为全国硕士研究生入学统一考试的参考书。

图书在版编目（CIP）数据

概率论与数理统计／韩文忠主编. --北京：高等教育出版社，2015.7

ISBN 978-7-04-042984-8

Ⅰ. ①概… Ⅱ. ①韩… 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（2015）第 134388 号

策划编辑 杨 帆　　责任编辑 杨 波　　封面设计 张 志　　版式设计 童 丹
插图绘制 黄建英　　责任校对 陈旭颖　　责任印制 毛斯璐

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	国防工业出版社印刷厂	网上订购	http://www.landraco.com
开　　本	787 mm×960 mm 1/16		http://www.landraco.com.cn
印　　张	16.5	版　　次	2015 年 7 月第 1 版
字　　数	280 千字	印　　次	2015 年 7 月第 1 次印刷
购书热线	010－58581118	定　　价	26.10 元

物 料 号　42984-00

前　　言

概率论与数理统计是研究随机现象客观规律性的一门数学学科。随着社会的发展、计算机的普及运用及各类研究的不断深入,其应用范围愈来愈广,现已成为社会多领域中不可或缺的工具。它是高等学校理科、工科、经济、管理等专业的基础课。本书是依据该课程教学要求的规定内容,参考全国硕士研究生入学统一考试数学考试大纲,并结合编者多年在大学各专业讲授概率论与数理统计课程积累的经验,博采众家之长,几经修改编写而成的。

本书分三部分,前五章为概率论,第六章至第九章为数理统计,第十章为Excel在概率统计中的应用。为了更好地满足高等学校理工科、经济和管理类复合型高素质人才培养的需要,在本书编写过程中,我们注重了以下几点:

第一,注重基本思想、基本概念、基本定理、基本方法及综合应用,力求体现概念和理论内容的科学性。本书中的基本思想、基本概念大多从具体实际问题中引入,对基本概念、定理和公式进行准确、简洁、通俗的叙述。为了体现概率论与数理统计的应用性,我们编写了一些贴近现实生活的例题和习题,旨在加深对基本思想和基本概念的理解,掌握对基本定理和基本方法的综合应用。

第二,注重不同层次读者的需要。在内容安排、例题和习题编排等方面尽可能做到由浅入深、易于计算,侧重对基本思想、基本概念的准确理解及对常用方法的熟练掌握。另外,考虑到一些读者考研的需要,书中加强了概率性质的应用、二维正态分布的性质和应用、条件分布及其综合应用等内容。每章附有习题和自测题,同时还收录了较多的历年全国硕士研究生入学统一考试试题,力争涵盖每一章内容相应的题型,书末附有参考答案。在附录中还给出了几种常见的概率分布表,便于读者自学。书中加"*"的内容可酌情取舍。

第三,注重概率论与数理统计学习过程中计算机的使用。本书特别编写了Excel在概率统计中的应用,较详细地介绍了Excel的内置函数在常见概率分布中的应用、在描述统计和直方图中的应用、在区间估计和假设检验中的应用以及在一元线性回归中的应用。并在书中首次出现相关问题时,不仅介绍了处理此类问题的传统方法,同时还给出了本书中用Excel处理此类问题的方法所在章节,供读者自学。

本书由韩文忠组织编写、统稿和定稿。其中第一章由刘官厅编写，第二章由米宇平编写，第七、八章由斯日古楞和米宇平合作编写，第三、四、五、六、九、十章由韩文忠编写，校对工作由米宇平完成。另外，参加编写的人员还有毕力格图、刘常彪、李秀兰。

本书在编写过程中参考了多种书籍，引用了一些数据和资料；高等教育出版社的领导和编辑们对本书给予了大力支持和帮助，我们在此一并致谢。

由于编者水平有限，书中难免有不妥之处，恳请专家、同行和读者批评指正。

编　者

2015年3月

目　录

第一章　随机事件及其概率

随机事件及其概率是概率论中的基本概念，是学好概率论的基础. 本章主要讨论随机事件的关系和运算、概率的定义和性质及一些常见的概率模型.

§1.1　随机试验与随机事件

一、随机试验与样本空间

在自然界和人类社会中存在着两类现象，一类是在一定条件下某种现象必定发生或必定不发生，这类现象称为**确定性现象**. 例如，半径为 r 的圆，其面积必定是 πr^2；在标准大气压下，水到 50 ℃ 必定不沸腾. 另一类是在一定条件下某种现象可能发生也可能不发生，这类现象称为**随机现象**，即随机试验描述的现象. 例如，抛掷一枚硬币，观察出现正面或反面的情况；发射一枚炮弹，观察弹着点与目标之间的距离.

这些试验都具有以下的特点：

(1) 可以在相同的条件下重复地进行；

(2) 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；

(3) 进行一次试验之前不能确定会出现哪一个结果.

在概率论中，将具有上述三个特点的试验称为**随机试验**，简称**试验**，记为 E.

随机试验中不能预测每次试验出现什么结果，这反映了随机试验的结果具有随机性. 但如果进行大量重复试验，所出现结果又具有某种规律性. 如射击试验中，弹着点与目标之间的距离可能各不相同，但如果射手技术较好，多次射击击中离靶心近的必定是多数. 这种在大量重复试验中所表现出来的规律性叫做统计规律性. 概率论与数理统计就是研究客观世界随机现象统计规律性的一门数学分支，在科学技术各领域及工农业生产和国民经济各部门中具有非常广泛的应用.

随机试验 E 每一可能的结果称为**样本点**，记作 ω. 样本点全体构成的集合称为**样本空间**，记作 Ω. 例如，掷一颗骰子，所有可能结果为“出现 1 点”，“出现 2

点”,⋯,“出现 6 点”,若用 ω_i 表示“出现 i 点”,则该试验的样本空间 $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$.

样本点和样本空间是概率论中的两个基本概念. 由于对所讨论问题的关注点不同,同一随机试验可以有不同的样本空间,学习时一定要加以注意.

例 1.1 口袋中装有 10 个球,其中有 3 个红球、3 个白球和 4 个黑球,将其进行编号,红球为 1 ~ 3,白球为 4 ~ 6,黑球为 7 ~ 10. 从口袋中任取一球,若关注点为球的编号,用 ω_i 表示“取得第 i 号球”,$i=1,2,\cdots,10$,这时样本空间为

$$\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7,\omega_8,\omega_9,\omega_{10}\}.$$

若只关注球的颜色,用 ω_1' 表示“取得红球”,ω_2' 表示“取得白球”,ω_3' 表示“取得黑球”,这时样本空间为

$$\Omega'=\{\omega_1',\omega_2',\omega_3'\}.$$

例 1.2 记录某网站单位时间的点击次数,用 ω_i 表示“该网站单位时间的点击次数为 i 次”,$i=0,1,2,\cdots$,则该试验的样本空间为

$$\Omega=\{\omega_0,\omega_1,\omega_2,\cdots\}.$$

它包含无限多个样本点,但可按一定的顺序排列起来.

例 1.3 测量某地区的水温(单位℃),则该试验的样本空间 $\Omega=[0,100]$,它是一个连续区间.

在实际应用中,选取合适的样本空间,对解决问题会带来很多方便.

二、随机事件

随机试验 E 的某些可能结果构成的集合称为**随机事件**,简称**事件**. 事件一般用大写英文字母 A,B,C 等表示.

如在例 1.1 中,设

$$A=\{\text{取得红球或白球}\},$$
$$B=\{\text{取得号码小于等于 7 的球}\},$$
$$C=\{\text{取得偶数号码的球}\},$$
$$D=\{\text{取得奇数号码的球}\},$$

则 A,B,C,D 都是随机事件. 若取其样本空间为 Ω 的情形,那么 A,B,C,D 可表示为

$$A=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\},\quad B=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7\},$$
$$C=\{\omega_2,\omega_4,\omega_6,\omega_8,\omega_{10}\},\quad D=\{\omega_1,\omega_3,\omega_5,\omega_7,\omega_9\}.$$

在一次试验中，当且仅当试验结果 $\omega \in A$，则称此次试验中**事件 A 发生**.

显然，样本空间 Ω 可看做一个事件. 因为在每次试验中必然出现 Ω 中的一个样本点，所以 Ω 在每次试验中必定发生. 我们把在每次试验中必然发生的事件，称为试验 E 的**必然事件**，记作 Ω. 类似地，空集$\varnothing$也作为一个事件，它在每次试验中必定不发生. 我们把在每次试验中都不发生的事件，称为试验 E 的**不可能事件**，记作$\varnothing$.

三、随机事件的关系与运算

由于事件是样本点的集合，因此，我们可用集合论的方法来研究事件之间的关系和运算.

1. 事件的关系与运算

(1) 若事件 A 发生必导致事件 B 发生，则称事件 B **包含**事件 A，记作 $A \subset B$ 或 $B \supset A$.

如在例 1.1 中，有 $A \subset B$.

(2) 若 $A \subset B$ 且 $B \subset A$，则称事件 A 与事件 B **相等**，记作 $A = B$.

(3) 表示“事件 A 与事件 B 中至少有一个发生”的事件称为事件 A 与事件 B 的**并事件**(**或和事件**)，记作 $A \cup B$.

如在例 1.1 中，有 $A \cup C = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_8, \omega_{10}\}$.

一般地，表示“n 个事件 $A_1, A_2, \cdots, A_n$ 中至少有一个发生”的事件称为这 n 个事件的**并事件**(**或和事件**)，记作 $A_1 \cup A_2 \cup \cdots \cup A_n$ 或 $\bigcup\limits_{i=1}^{n} A_i$.

类似地，表示“可列无限个事件 $A_1, A_2, \cdots, A_n, \cdots$ 中至少有一个发生”的事件称为这可列无限个事件的**并事件**(**或和事件**)，记作 $A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots$ 或 $\bigcup\limits_{i=1}^{\infty} A_i$.

(4) 表示“事件 A 与事件 B 同时发生”的事件称为事件 A 与事件 B 的**交事件**(**或积事件**)，记作 $A \cap B$ 或 AB.

如在例 1.1 中，有 $A \cap C = \{\omega_2, \omega_4, \omega_6\}$.

一般地，表示“n 个事件 $A_1, A_2, \cdots, A_n$ 同时发生”的事件称为这 n 个事件的**交事件**(**或积事件**)，记作 $A_1 A_2 \cdots A_n$ 或 $\bigcap\limits_{i=1}^{n} A_i$.

类似地，表示“可列无限个事件 $A_1, A_2, \cdots, A_n, \cdots$ 同时发生”的事件称为这可

列无限个事件的**交事件**(或**积事件**),记作 $A_1A_2\cdots A_n\cdots$ 或 $\bigcap\limits_{i=1}^{\infty} A_i$.

(5) 表示"A 不发生"的事件称为事件 A 的**对立事件**(或**逆事件**),记作 $\overline{A}$.

显然有 $A\cup\overline{A}=\Omega, A\overline{A}=\varnothing, \overline{\overline{A}}=A$.

(6) 表示"事件 A 发生且事件 B 不发生"的事件称为事件 A 与事件 B 的**差事件**,记作 $A-B$.

显然有
$$A-B=A\overline{B}=A-AB. \tag{1.1}$$

如在例 1.1 中,有 $B-A=\{\omega_7\}$.

(7) 如果 A 与 B 两事件不可能同时发生,即 $A\cap B=\varnothing$,则称事件 A 与事件 B **互斥**(或**互不相容**),简称事件 A,B 互斥. 若 n 个事件 $A_1,A_2,\cdots,A_n$ 中的任意两个事件都互斥,即 $A_i\cap A_j=\varnothing, i\neq j, i,j=1,2,\cdots,n$,则称这 n 个事件**两两互斥**(或**两两互不相容**). 类似地,两两互斥的概念可以推广到可列无限个事件的情形.

与集合论类似,可借助于文氏图表示事件的关系及运算,如图 1.1 所示,图中用矩形区域表示样本空间 Ω,用圆形区域分别表示事件 A 和 B.

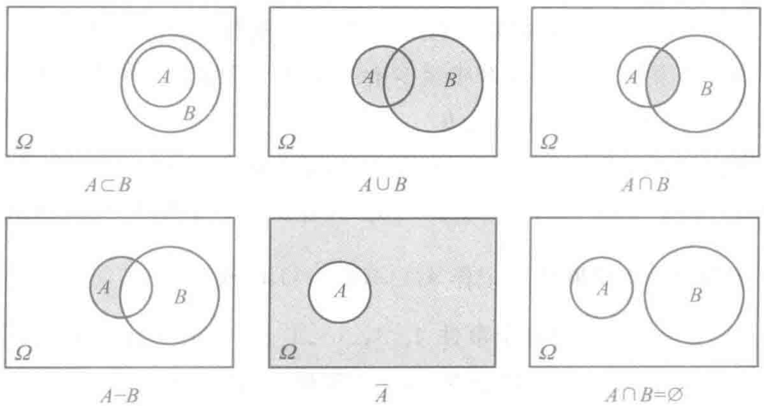

图 1.1

2. 事件的运算性质

事件的运算满足集合论中有关集合运算的一切性质. 设试验 E 的样本空间为 $\Omega, A, B, C, A_i(i=1,2,\cdots,n)$ 为 E 的事件.

(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** $(A\cup B)\cap C=(A\cap C)\cup(B\cap C); (A\cap B)\cup C=(A\cup C)\cap(B\cup C)$.

(4) **对偶律**(**德摩根**(De Morgan)**律**) $\overline{A\cup B}=\overline{A}\cap\overline{B}; \overline{A\cap B}=\overline{A}\cup\overline{B}$.

上述性质可以推广到有限个事件或者可列无限个事件的情形. 如 n 个事件 $A_1,A_2,\cdots,A_n$ 的对偶律为

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}, \quad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}. \tag{1.2}$$

利用上述性质, 可以实现集合论的表示和事件的关系及运算与概率论语言的互相转化, 也可以把复杂事件用简单事件的运算来表示, 请读者在学习时注意体会.

例 1.4 若 A,B,C 是三个事件, 则

"事件 A 与 B 发生, C 不发生"可表示为 $AB\overline{C}$ 或 $AB-C$;

"A,B,C 三个事件中恰有一个发生"可表示为 $A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C$;

"A,B,C 三个事件中恰有两个发生"可表示为 $AB\overline{C}\cup A\overline{B}C\cup\overline{A}BC$;

"A,B,C 三个事件中至少有一个发生"可表示为 $A\cup B\cup C$;

"A,B,C 三个事件同时发生"可表示为 ABC;

"A,B,C 三个事件中不多于两个发生"可表示为 $\overline{A}\cup\overline{B}\cup\overline{C}$;

"A,B,C 三个事件全不发生"可表示为 $\overline{A}\,\overline{B}\,\overline{C}$ 或 $\overline{A\cup B\cup C}$.

§1.2 频率与概率

在一次试验中, 事件 A 可能发生, 也可能不发生, 具有随机性. 但是, 人们在实践中发现, 在相同的条件下, 进行大量的重复试验中, 试验的结果具有某种内在的统计规律性, 即事件发生的可能性大小是可以比较的, 是可以用一个数值进行度量的. 例如, 在投掷一颗均匀的骰子试验中, 显然事件"点数不超过 5"比事件"点数为 1"发生的可能性要大. 我们把事件 A 发生可能性大小的数值称为事件 A 的**概率**, 记作 $P(A)$. 下面首先引入频率, 进而利用频率的稳定性给出概率的统计定义.

一、频率

定义 1.1 若在相同条件下重复进行 n 次试验, 其中事件 A 发生的次数为 n_A, 则称

$$f_n(A) = \frac{n_A}{n} \tag{1.3}$$

为事件 A 在这 n 次试验中发生的**频率**.

易知,频率 $f_n(A)$ 满足下列性质:

(1) 对任一事件 A,有 $f_n(A)\geqslant 0$;

(2) $f_n(\Omega)=1$;

(3) 对任意一组两两互斥的事件 $A_1,A_2,\cdots,A_m$,有

$$f_n\left(\bigcup_{i=1}^{m}A_i\right)=\sum_{i=1}^{m}f_n(A_i).$$

二、概率的统计定义

由频率的定义知,频率不仅与试验次数 n 有关,且在 n 固定时,分别做若干组 n 次试验,各组试验的频率一般也不相同.但是,随着试验次数 n 逐步增大,一般来说,频率 $f_n(A)$ 会逐渐稳定在某个常数 p 附近,且摆动幅度越来越小,这个常数 p 称为频率的稳定值.这种规律称为频率的稳定性.

例如,抛掷一枚均匀硬币,设 $A=\{$出现正面$\}$.历史上有不少学者做过此试验,其结果如表 1.1.

表 1.1 抛掷硬币的试验与结果

试验者	掷硬币次数 n	正面出现次数 n_A	正面出现的频率 $f_n(A)$
德摩根	2 048	1 061	0.518 1
蒲丰(Buffon)	4 040	2 048	0.506 9
皮尔逊(Karl Pearson)	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

由表 1.1 可以看出,随着抛掷硬币次数 n 增大,频率 $f_n(A)$ 呈现出稳定性,即当 n 逐步增大,频率 $f_n(A)$ 稳定于常数 0.5,这里的常数 0.5 就是频率 $f_n(A)$ 的稳定值,它反映了事件 A 发生的可能性大小.我们可以用这个频率的稳定值 0.5 定量地描述事件 A 发生的可能性大小,即事件 A 的概率.

一般地,我们给出如下定义.

定义 1.2 在相同条件下重复进行 n 次试验,若事件 A 发生的频率 $f_n(A)$ 随着试验次数 n 的增大而稳定在某个常数 $p(0\leqslant p\leqslant 1)$ 附近摆动,则称 p 为事件 A 的**概率**,记为 $P(A)$.这一定义叫做**概率的统计定义**.

由概率的统计定义可知,事件 A 的概率 $P(A)$ 是客观存在的,频率 $f_n(A)$ 是随着试验的变化而变化的,且具有稳定性.概率与频率的这种关系就像物体质量