21世纪高等教育规划教材

概率论与数理统计

黄建雄　等编著

$$P\{X=k\}=\binom{n}{k}p^kq^{n-k}$$

中国物资出版社

概率论与数理统计

21 世纪高等教育规划教材

概率论与数理统计

黄建雄　等编著

中国物资出版社

图书在版编目(CIP)数据

概率论与数理统计/黄建雄等编著. —北京:中国物资出版社,2009.3

21世纪高等教育规划教材

ISBN 978-7-5047-3023-7

Ⅰ.概… Ⅱ.黄… 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(2009)第004124号

策划编辑 钱 瑛

责任编辑 钱 瑛

责任印制 何崇杭

责任校对 孙会香 梁 凡

中国物资出版社出版发行

网址:http://www.clph.cn

社址:北京市西城区月坛北街25号

电话:(010)68589540 邮编:100834

全国新华书店经销

中国农业出版社印刷厂印刷

开本:710 mm×1000 mm 1/16 印张:16.25 字数:309千字

2009年3月第1版 2009年3月第1次印刷

书号:ISBN 978-7-5047-3023-7/O·0038

定价:26.00元

内容提要

本书主要内容包括：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，大数定律与中心极限定理，数理统计的基本概念，参数估计，假设检验，方差分析与回归分析初步。

本书可作为高等学校理、工、管理、经济类各专业数学课程教材或教学参考书，也可供广大工程技术人员或企业管理人员业务学习参考。

编审说明

《概率论与数理统计》是高等院校理、工、管理、经济类专业必修课，同时也是许多专业研究生入学考试内容。

从概率统计学科的发展来看，“概率论与数理统计”是应用性很强的学科，概率论起源于博弈问题研究，但现在概率统计方法已广泛应用于国民经济的各个领域及人民生活的方方面面。例如天气预报，水文预报及地震预报，产品的抽样检验，试验方案的制定，系统的可靠性评估，提高通信系统的效率，自动控制系统控制方案的设计，保险险种设计及费率计算，彩票方案的设计，期权定价问题等，甚至在确定性问题研究中也应用随机性方法，如计算机对确定性问题的计算中，可采用随机算法来大幅度提高运算效率等。

同时，概率论作为一门数学学科，其公理化体系形成较晚。大约在20世纪30年代，由于苏联数学家柯尔莫哥罗夫(Kolmogrov)的研究，概率论的公理化工作才得以完成。因此，学习概率论的研究方法，对提高广大科学技术人员的科学素养也有较大的帮助。

考虑到概率论的公理化思想的重要性，本书从事件的运算封闭性和概率的公理化定义两方面，对概率论的公理化思路进行简要的阐述。

概率论与数理统计包含两个部分——概率论和数理统计技术。概率论是对随机现象统计规律进行演绎的研究，而数理统计技术则是对随机现象的统计规律进行归纳的研究。两者在方法上有着明显的不同，但两者却相互联系，相互渗透。本书前5章内容为概率论部分，后4章内容为数理统计部分。

概率论与数理统计是一门应用性很强的学科，同时其抽象性和技巧程度也非常高，本书在编排上力求做到简繁得当，通俗易懂。对现实生活中大量应用的概率方法进行了较多的例题讨论，力求既能提高

读者的兴趣，又能使读者从中加深对基本概念的理解及对基本方法的掌握。本书在概率论部分对几个重要分布，如0—1分布、二项分布、泊松分布、均匀分布、正态分布的来源与应用均有大量的讨论，望读者加以关注（泊松过程的讨论超出了工科类专业概率论与数理统计教学的基本要求，但考虑到它在应用上的重要性，本书也将其讨论过程列出，供读者选用）。

本书每章均列有基本要求，其大体符合工科学生概率论与数理统计课程研究生入学考试要求。

本书编写策划工作由黄建雄、李泽榕、李康弟完成。参加本书编写的有黄建雄（第1、2章），钱道翠、蔡文康（第3、4、5章），蒋书法、李泽榕（第6、7、8、9章），于娜参加了本书的例题和习题的编选工作，并对稿件进行了校订。全书最后由黄建雄、李泽榕、李康弟进行统稿和定稿。

限于编者水平，加之时间仓促，书中错误和不妥之处在所难免，敬请广大读者不吝批评指正。

21世纪高等教育规划教材编审指导委员会

2009年3月

目 录

第1章　随机事件及其概率

本章基本要求

1. 了解样本空间(基本事件空间)的概念,理解随机事件的概念,掌握事件的关系与运算。

2. 理解概率、条件概率的概念,掌握概率的基本性质,会计算古典型概率和几何型概率,掌握概率的加法公式、乘法公式、减法公式、全概率公式以及贝叶斯公式。

3. 理解事件的独立性的概念,掌握用事件独立性进行概率计算;理解独立重复试验的概念,掌握计算有关事件概率的方法。

自然界中有两类现象:一类为确定性现象,例如上抛的石子必定落下,同性电子必定相斥等,这类现象的结果是确定的。另一类为随机性现象,例如掷硬币,有可能出现正面,也有可能出现反面,掷骰子可能出现的点数为“1,2,3,4,5,6”中的一个,但在抛掷之前不可能知道究竟出现哪一种结果,还有炮弹瞄准射击的弹着点问题,每一颗炮弹在射击前不可能知道其弹着点的确切位置,这些现象称为随机性现象。这些在相同条件下的试验有一个共同点:在试验之前不可能预先知道出现哪个结果,有可能出现这样或那样的结果,但大量的试验结果分析,则呈现出一种统计规律性。如掷硬币时,出现正面的次数大致占到总次数的一半,掷骰子时,出现点数“6”的次数大致占到总次数的1/6,瞄准射击时,目标附近的弹着点较为密集等。

这种在个别试验中,试验结果呈现不确定性,但在大量重复试验中,试验结果呈现出统计规律性的随机现象,就是概率论与数理统计的主要研究对象。概率论与数理统计是一门研究随机现象的内在统计规律性的学科。现在概率统计的方法已广泛应用于国民经济的许多领域及人民生活的很多方面,如用于天气预报、水文预报、地震预报、产品的抽样检验、确定试验方案、对系统的可靠性进行评估、在通信中用于提高通信系统的效率、用于自动控制系统的控制方案的设计、进行保险险种设计及费率计算、彩票销售方案的设计、股票市场的期权定价问题研究。甚至在确定性现象的研究中也广泛地应用随机方法,如计算机对确定

性问题的计算中，用随机算法来大幅度提高运算效率等。

§1.1 随机试验与随机事件

在科学和工程技术研究中，时常要在相同条件下做试验，并对结果进行观察，下面举一些试验的例子。

【例1】 口袋中有编号为 $1,2,\cdots,n$ 的球，从袋中任取一只，观察球的编号。

【例2】 掷骰子一次，观察出现的点数。

【例3】 将一枚硬币掷两次，观察第一次、第二次出现正面、反面的情况。

【例4】 将一枚硬币掷两次，观察两次中正面出现的次数。

【例5】 在 7:00 ～ 9:00，观察通过道口的车辆数。

【例6】 在一批电视机中，任取一台，测试其寿命(小时)。

【例7】 对一个圆柱体构件，测量其直径。

上述举出的 7 个试验的例子，它们有如下共同的特点：

(1) 试验可在相同的条件下进行；

(2) 每次试验的可能结果不止一个，但能事先知道所有可能的结果；

(3) 试验之前并不知道哪一种结果会出现。

具有上述三个性质的试验称为**随机试验**，以后将随机试验简称为**试验**。

对于随机试验，尽管试验之前不能预知哪一个结果出现，但所有可能的结果是知道的，我们将所有可能结果组成的集合称为随机试验的**样本空间**，记为 S，每一个可能的结果，即样本空间的元素称为**样本点**。

【例8】 在例 1～ 例 7 的试验中，分别写出相应的样本空间 $S_k(k=1,2,\cdots,7)$。

解 在例 1 的试验中，$S_1=\{1,2,\cdots,n\}$；

在例 2 的试验中，$S_2=\{1,2,\cdots,6\}$；

在例 3 的试验中，将每次掷硬币时出现正面的结果记为 1，出现反面的结果记为 0，则 $S_3=\{(1,0),(1,1),(0,1),(0,0)\}$；

在例 4 的试验中，$S_4=\{0,1,2\}$；

在例 5 的试验中，$S_5=\{0,1,2,\cdots\}$；

在例 6 的试验中，$S_6=\{t\mid t\geqslant 0\}=[0,+\infty)$；

在例 7 的试验中，$S_7=\{x\mid a\leqslant x\leqslant b\}$，这里 a,b 分别表示可能测到的最小值和最大值。

注意到事实，由于试验的观察目的不相同，因此样本空间对同一种试验也可能不一样，如例 3 和例 4 的试验中，样本空间就不相同。

在实际应用中，我们关心的是某一类试验结果是否发生，如在例 2 的试验中，若我们关心的是是否出现较大的点子，即结果 4,5,6 是否出现，这些结果组成集合 $A=\{4,5,6\}$，当且仅当集合 $A=\{4,5,6\}$ 中的某一样本点出现时，掷骰子就掷出了较大的点子；在例 6 的试验中，若我们关心的是电视机寿命不小于 5 000 小时，满足这一条件的样本点组成集合 $B=\{t \mid t \geqslant 5\,000\}=[5\,000,+\infty)$，当且仅当 B 中的某一样本点出现时，电视机寿命不小于 5 000 小时。

通常，样本空间 S 的一个子集称为随机试验的一个**随机事件**，简称**事件**。当且仅当这一子集中的样本点出现时，称事件发生。随机事件通常用大写字母 A，B 或 A_1, A_2, B_1, B_2 等表示。

由样本点组成的单点集，称为**基本事件**。如例 2 的试验中，基本事件有 6 个，即 $\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}$；例 5 的试验中，基本事件有多个，即 $\{0\}$，$\{1\},\{2\},\cdots$

S 是样本空间 S 的子集，它包含了所有可能的试验结果，在每次试验中必定发生，因此称 S 是**必然事件**。空集也是样本空间的子集，它不含任何的试验结果，在试验中必定不发生，因此称空集 $\varnothing$ 是**不可能事件**。下面举几个事件的例子。

【例 9】 (1) 在例 1 的试验中，摸到号码不超过 $\frac{n}{2}$ 的球的事件为 $A_1=\left\{1,2,\cdots,\left[\frac{n}{2}\right]\right\}$，摸到号码为自然数的球的事件为 S；

(2) 在例 3 的试验中，首次掷出反面的事件为 $A_2=\{(0,0),(0,1)\}$，恰好掷出一次正面的事件为 $A_3=\{(1,0),(0,1)\}$，掷出 3 次正面的事件是不可能出现的，记为 $\varnothing$；

(3) 在例 5 的试验中，通过车辆数超过 1 000 辆的事件为 $A_4=\{1\,001, 1\,002,\cdots\}$；

(4) 在例 6 的试验中，若电视机的寿命小于 5 000 小时是次品，则电视机是次品的事件为 $A_5=\{t \mid 0 \leqslant t < 5\,000\}=[0,5\,000)$。

§1.2　事件的关系与运算

上节中，利用集合的表示方法，我们用样本空间的子集来表示随机事件，因

此集合的关系与运算可用到事件的关系与运算中，下面就给出事件的关系和运算在概率论中的提法。

一、包含关系

如果事件 A 发生必定导致事件 B 发生，则称 B **包含** A 或称 A 是 B 的**子事件**，记为 $B \supseteq A$ 或 $A \subseteq B$(图 1-1)。若记 A 为事件"通过车辆数超过 1 000 辆"，B 为事件"通过车辆数超过 100 辆"，则 $A \subseteq B$。显然对任何事件 A，都有 $\varnothing \subseteq A \subseteq S$。

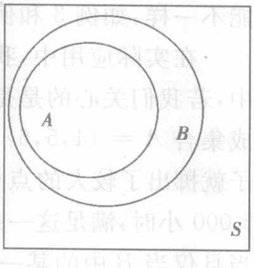

图 1-1

二、相等关系

如果事件 A 发生必定导致事件 B 发生，并且事件 B 发生必定导致事件 A 发生，即 $A \subseteq B$ 且 $B \subseteq A$，则称事件 A 与事件 B **等价**，或称事件 A 与事件 B **相等**，记为 $A = B$。

三、和事件

如果事件 A，B 至少一个发生，就称 A 与 B 的**和事件**发生，A 与 B 的和事件记为 $A \cup B$，和事件的集合表示为 $A \cup B = \{x \mid x \in A$ 或 $x \in B\}$(图 1-2)。

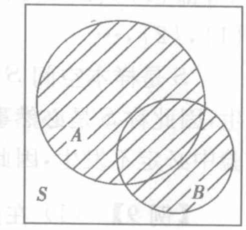

图 1-2

类似地，称 $\bigcup\limits_{k=1}^{n} A_k$ 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的和事件，事件 $\bigcup\limits_{k=1}^{n} A_k$ 发生，当且仅当事件 $A_1, A_2, \cdots, A_n$ 至少有一个发生。同理，称 $\bigcup\limits_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \cdots$ 的和事件。

四、积事件

如果事件 A 与 B 同时发生，就称 A 与 B 的**积事件**发生，A 与 B 的积事件记为 $A \cap B$(或记为 AB)，积事件的集合表示为 $A \cap B = \{x \mid x \in A$ 且 $x \in B\}$(图 1-3)。

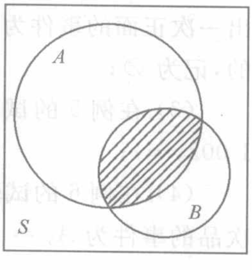

图 1-3

类似地，称 $\bigcap\limits_{k=1}^{n} A_k$(或记为 $A_1 A_2 \cdots A_n = \prod\limits_{k=1}^{n} A_k$) 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的积事件，事件 $\bigcap\limits_{k=1}^{n} A_k$ 发生，当且仅当事件 $A_1, A_2, \cdots, A_n$ 同时发生。同理，称 $\bigcap\limits_{k=1}^{\infty} A_k$(或记为 $\prod\limits_{k=1}^{\infty} A_k = A_1 A_2 \cdots$) 为可列个事件 $A_1, A_2, \cdots$ 的积事件。

五、差事件

如果事件 A 发生且 B 不发生，则称 A 与 B 的**差事件** $A-B$ 发生。差事件的集合表示为 $A-B=\{x \mid x \in A$ 且 $x \notin B\}$（图 1-4）。

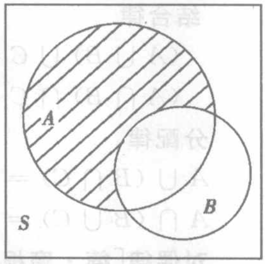

图 1-4

六、互斥事件

如果事件 A 与 B 不能同时发生，则称事件 A,B 是**互斥的**或称**互不相容的**，其集合表示为 $A \cap B=\varnothing$（图 1-5）。基本事件是两两互斥的。

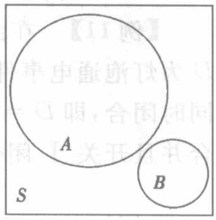

图 1-5

七、对立事件

如果事件 A,B 至少一个发生，且不能同时发生（即 $A \cup B=S$ 且 $A \cap B=\varnothing$），则称事件 A 与 B 互为**逆事件**，或称事件 A 与 B 互为**对立事件**，记 $B=\overline{A}$ 或 $A=\overline{B}$，事件 A 的逆事件 $\overline{A}$ 也可表示为 $\overline{A}=S-A$，其集合表示为 $\overline{A}=\{x \mid x \notin A\}$（图 1-6），显然 $\overline{\overline{A}}=A$。

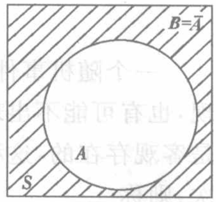

图 1-6

通常用矩形内的图形（Venn 图）表示事件的关系和运算，此方法比较容易掌握（图 1-1 ～ 图 1-6）。

【例 10】　设一口袋中有白球若干只，红球若干只，从袋中取两次球，记第一次取白球的事件为 A，第二次取白球的事件为 B，第三次取白球的事件为 C，则：

(1) 第一次取到红球的事件为 $\overline{A}$；

(2) 第一次取到红球且第二次取到白球的事件为 $\overline{A} \cap B=\overline{A}B$；

(3) 第一次取到白球且第二次取到红球的事件为 $A \cap \overline{B}=A\overline{B}=A-B$；

(4) 3 次全取到红球的事件为 $\overline{A} \cap \overline{B} \cap \overline{C}=\overline{A \cup B \cup C}$；

(5) 3 次中至少取到一只白球的事件为 $A \cup B \cup C=\overline{\overline{A} \cap \overline{B} \cap \overline{C}}$；

(6) 第三次才取到白球的事件为 $\overline{A} \cap \overline{B} \cap C$。

类似于集合的关系与运算，事件的关系与运算也经常要用到下列的规律。设 A,B,C 是 3 个事件，则有下列运算规律成立。

交换律

$$A \cup B=B \cup A$$

$$A \cap B=B \cap A$$

结合律

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

分配律

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

对偶律[德·摩根(De Morgan)律]

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

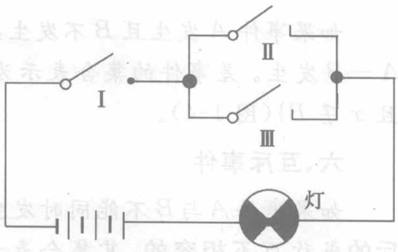

图 1-7

【例 11】 在图 1-7 的电路中,分别记 A,B,C 为开关 Ⅰ,Ⅱ,Ⅲ 闭合的事件,D 为灯泡通电事件。则灯泡通电可表示为开关 Ⅰ,Ⅱ 同时闭合或者开关 Ⅰ,Ⅲ 同时闭合,即 $D = AB \cup AC$。同时,灯泡通电可表示为开关组 Ⅱ,Ⅲ 至少一个闭合并且开关 Ⅰ 闭合,即 $D = A \cap (B \cup C)$。

§ 1.3 事件的频率与概率

一个随机事件 A(除必然事件和不可能事件外)在一次试验中,它有可能出现,也有可能不出现,具有偶然性,但在大量的重复试验中,其出现的统计规律性是客观存在的,这种性质称为频率稳定性。若随机事件 A 在 n 次试验中发生 n_A 次,则称

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

为随机事件 A 在 n 次试验中出现的**频率**。下面对历史上几个著名随机试验进行讨论。

一、几个著名试验

历史上有不少人对随机事件出现的频率进行了观察,下面列举一些试验的结果。

在掷硬币的试验中,正面在一次试验中有可能出现,也可能不出现,预先作出确定的预测是不可能的。记 A 为出现正面事件,下面的表格中列出了几个实验结果:

实验者	试验次数 n	正面出现次数 n_A	频率 $f_n(A)$
德·摩根	2 048	1 061	0.518 1
布丰	4 040	2 048	0.506 9
K.皮尔逊	12 000	6 019	0.501 6
K.皮尔逊	24 000	12 012	0.500 5

从上述表格可以看出，在大量的重复试验中，出现正面的频率接近 50%，和直觉上出现正面的机会很近似。

又如，在英文中某些字母出现的频率远远高于另外一些字母，人们经过深入研究之后，发现字母的使用频率相当稳定。下面是字母使用频率的统计表：

字　母	空　格	E	T	O	A	N	I	R	S
频　率	0.2	0.105	0.072	0.065 4	0.063	0.059	0.055	0.054	0.052
字　母	H	D	L	C	F	U	M	P	Y
频　率	0.047	0.035	0.029	0.023	0.022 5	0.022 5	0.021	0.017 5	0.012
字　母	W	G	B	V	K	X	J	Q	Z
频　率	0.012	0.011	0.010 5	0.008	0.003	0.002	0.001	0.001	0.001

字母使用频率的研究，对打字机键盘的设计（方便的地方安排使用频率较高的字母键），信息编码（常用字母使用较短的编码），密码的破译等方面有重要的意义。

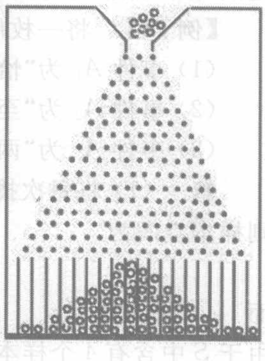

图 1-8

另一个著名验证频率稳定性的试验是由英国的生物统计学家高尔顿(Galton)做出的，试验用具如图 1-8 所示。在上端放入一小球，任其下落，下落过程中，小球碰到钉子，经过一系列的碰撞后，小球落入底板中的某格子中，一只小球将落入哪个格子事先无法确定。但实验证明，重复进行放入大量小球的试验后，则小球最后所呈现的曲线几乎都是一样的。这说明，小球落入各个格子的频率具有稳定性。

人们在长期的实践活动中，观察到随机事件 A 出现的频率 $f_n(A)$ 有如下特点，在试验次数 n 较小时，频率 $f_n(A)$ 在 0 至 1 之间波动较大，但当试验次数 n 逐渐增大时，频率 $f_n(A)$ 逐渐接近于某一常数。因此，用频率的这一稳定值