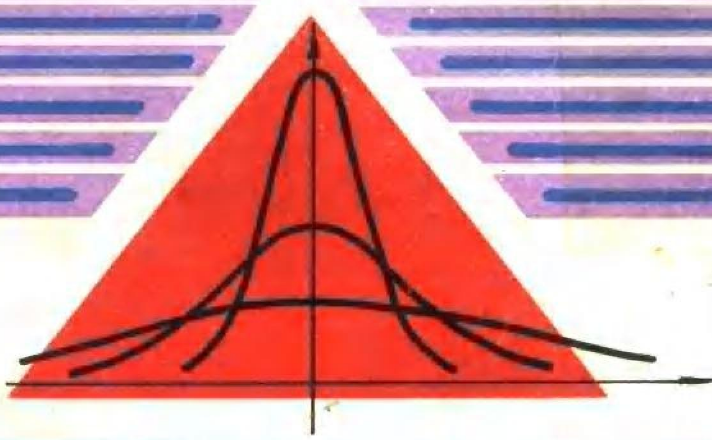
概率论与数理统计

伊亨云　朱金明　孙荣恒　黄雯莹　编



重庆大学出版社

概率论与数理统计

伊亨云　朱金明
孙荣恒　黄雯莹　编

重庆大学出版社

内容提要

本书是根据全国高等工科院校工程数学基本要求，结合编者多年从事教学的实践经验编写的. 本书 1—6 章为概率论部分，内容包括随机事件及其概率、条件概率、事件的独立性，一维随机变量及其分布、二维随机变量及其分布、随机变量的数字特征、极限定理；7—12 章为数理统计部分，内容包括：数理统计的基本概念、参数估计、假设检验、回归分析、方差分析、正交试验设计.

本书对主要概念及方法的实际背景及应用去向进行了详细论述，对容易混淆的概念和不易掌握的部分进行了详细阐述和举例分析. 每章后配有适当习题，书末附有参考答案. 本书可作为高等工科院校本科各专业概率论或数理统计课程教材，经适当删减后也可供高等工科院校专科、夜大学、函授大学工科各专业学生使用.

概率论与数理统计

伊亨云　朱金明
孙荣恒　黄雯莹　编

责任编辑　刘茂林

*

重庆大学出版社出版发行

新　华　书　店　经　销

重庆电力印刷厂印刷

*

开本：787×1092　1/16　印张：14　字数：348 千

1995年9月第1版　1995年9月第1次印刷

印数：1—6000

ISBN　7-5624-1089-5/O·123　定价：11.00 元

(川)新登字 020 号

编 者 的 话

本书是根据全国高等工科院校数学教材会议制订的工程数学基本要求(概率论与数理统计部分)编写的.本书是在1991年出版的同名书的基础上重新编写出版的.本书保留了原书的紧贴教学基本要求、便于组织教学和自学等特色,对原书中的错误进行了仔细的修改和订正.

本书1—6章为概率论部分,教学时数为36～40学时.7—12章为数理统计部分,教学时数为30～36学时.本书可作为高等工科院校本科各专业概率论或数理统计课程的教材.经适当删减后也可作为高等工科院校夜大学、函授大学工科各专业本、专科相应课程的教材.亦可作为工程技术人员自学的参考书.

本书每章后均附有习题,教学时可选择使用,书后附有各章习题答案.

本书第1、2、3章由朱金明编写,第4、5、6、9章由孙荣恒编写,第7、8章由黄雯莹编写,第10、11、12章由伊亨云编写.伊亨云任主编,对全书作了统一审定.

由于编者水平所限,错误在所难免,敬请读者和使用该书的师生批评指正.

编者

1995年9月

目　　录

第 1 章　随机事件及其概率

1-1　随 机 事 件

1-1-1　随机现象

概率论是研究随机现象的统计规律性的一门数学分支. 那末，什么是随机现象呢？

自然界和人类社会所发生的现象，虽然是各种各样的，但归纳起来，大体上可分为三类：

第一类是必然现象. 例如，带电体之间，总是同性相斥异性相吸；在理想条件下，自由落体总是遵循着 $S=\frac{1}{2}gt^2$ 的运动规律；在一个标准大气压下，把纯水加热到 80℃ 时，水必然不沸腾. 像这类在一定条件下，重复进行试验，每次结果必然相同的现象，称为必然现象或确定性现象. 研究这类现象所用的数学工具有代数、几何、微积分、高等代数、数理方程等所谓的经典数学.

第二类是模糊现象. 例如，这辆汽车开得飞快；那个孩子长得真漂亮；人民大会堂建得很雄伟；我刚才看见一个大胡子的高个子男人. 其中，"飞快"、"真漂亮"、"很雄伟"、"刚才"、"大胡子"、""高个子"等均是不精确的模糊概念. 像这类在一定条件下，发生具有不精确性(或模糊性)的事件的现象，称为模糊现象. 研究模糊现象的数学，即称之为模糊数学，它诞生于 1965 年，创始人是美国自动控制专家查德(L. A. Zadeh)教授. 并且查德教授在 1975 年指出："如果深入研究人类的认识过程，我们发现人类能运用模糊概念这将是一笔巨大的财富而不是负担. 这一点，是理解人类智能和机器智能之间深奥区别的关键". 由于客观上需要对人的主观、心理因素进行研究，提高计算机的"活性"以及增加机器的"智能"，这一切都离不开所谓的模糊性，也孕育着模糊数学的产生，并促使其迅速发展. 所以，模糊数学的产生反映了信息革命的迫切需要，并将为信息革命提供一种继经典数学、统计数学之后新的富有魅力的数学工具和手段.

第三类是随机现象. 例如，同一门大炮向同一目标射击，各次射击的弹着点不尽相同，且每次射击之前无法预言弹着点的确切位置；抛一枚硬币，可能国徽向上，可能国徽向下；新生的婴儿可能是男孩也可能是女孩等. 这类现象有一个共同的特点：在同样的条件下，每次试验或观察会得不完全相同的结果，有多种可能结果，而且每次试验或观察之前无法预知会出现什么结果，这类现象叫做随机现象或偶然性现象. 起初，人们把这种现象称为"偶然的"、"不正常的"、"原因不明"的现象，并且认为这些现象是无规律的现象. 但是，后来人们通过长时期的观察和实践，发现这类现象虽然就每次观察的结果来说有不确定性，但是经过大量重复观察，便呈现出某种规律性. 例如，大量重复抛一枚硬币，得国徽向上的次数大致是抛的总次数的 $\frac{1}{2}$；对各国出生婴儿统计中，可得男婴数大致等于婴儿总数的一半；同一射手射击同一目标，弹着点按一定的规律分布等. 这种在大量地重复试验或观察中所呈现的固有的规律，就是所谓的统计规律性. 概率论就是研究和揭示随机现象统计规律性的数学学科.

1-1-2 随机试验

为了掌握随机现象的统计规律，就必须对随机现象进行大量观测或实验．例如

例 1-1 抛硬币试验 E_1：抛一枚硬币，观察国徽向上的情况；

例 1-2 摸球试验 E_2：一个盒子有 10 个完全相同的球，球上分别标有号码 1，2，…，10，从中任摸一球，观察球上所标的号码；

例 1-3 旋转陀螺试验 E_3：在一个均匀的陀螺圆周上，均匀地刻上区间 $[0,3)$ 上的诸数字，旋转陀螺．当它停下时，观察陀螺圆周上与桌面接触点的刻度；

例 1-4 抽样试验 E_4：100 件产品中有 4 件次品，任抽 3 件，观察所抽 3 件中含的次品件数．

以上这些试验，概括起来都具有以下三个特点：

(1) 试验可以在相同条件下重复进行；

(2) 试验有多种可能结果，并且试验的所有可能结果事先是知道的；

(3) 每次试验到底出现哪 一个结果，是无法预言的．

在概率论中，将具有以上三个特点的试验称为**随机试验**，简称试验，常用 E 来表示．以后所说的试验都是指随机试验．

注意 随机试验是研究随机现象统计规律性的手段，即一般是通过研究随机试验来研究随机现象的．

1-1-3 随机事件

在随机试验中，可能发生也可能不发生的结果称为随机事件，简称事件，常用大写字母 A，B，C，D…表示，有时也以{……}或"……"表示事件．

例如，在摸球试验 E_2 中，"摸到球的标号为 1"，"摸到球的标号为 10"，"摸到球的标号为偶数"，"摸到球的标号大于 5"等，都是可能发生也可能不发生的结果，所以都是随机事件．

只包含一个可能的试验结果的事件称为**基本事件**；由两个或两个以上基本事件复合而成的事件称为复合事件．

例如，在摸球试验 E_2 中，事件"摸到球的标号为 i"$(1\leqslant i\leqslant 10)$是基本事件，$E_2$ 共有 10 个基本事件．而事件"摸到球的标号为偶数"是复合事件，因为，此复合事件是由摸到球的标号为 2、4、6、8、10 这五个基本事件复合而成的事件．

在每次试验中，必然发生的事件，称为必然事件，记为 Ω；必然不发生的事件，称为不可能事件，记为 Φ．

例如，在摸球试验 E_2 中，"摸到球的标号大于 0"是必然事件；"摸到球的标号小于 0"是不可能事件．

注意 (1) 必然事件与不可能事件本来没有不确定性，也就是说，它们不是随机事件，但是为了今后讨论方便起见，把它们看作特殊的随机事件．正如高等数学中把常量看作特殊变量一样，也是为了方便；(2) 随机事件与随机试验的关系是：随机事件是随机试验的结果．

1-1-4 样本空间

为了用数学方法来描述与随机试验有关的问题，使概率论建立在严密的理论基础上，下面

从集合论的角度来描述随机事件. 为此，首先引入样本空间的概念.

对于随机试验，尽管在每次试验之前不能预知试验结果，但试验的所有可能结果组成的集合是已知的. 随机试验的每一个最简单的不能再分解的可能结果称为一个样本点，习惯上用 ω 表示. 由随机试验的所有样本点组成的集合称为**样本空间**或基本事件空间，记作 Ω. 只含一个样本点的集合即为基本事件.

例如，在摸球试验 E_2 中，若用 ω_i 表示"摸到标号为 i 的球"$(i=1,2,\cdots,10)$，则 $\omega_1,\omega_2,\cdots,\omega_{10}$是 E_2 的样本点，E_2 的样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_{10}\}$.

注意 (1) 样本空间可以是各种对象的集合，即可以是数集也可以不是数集；(2) 样本空间的这种抽象表示，实际上是抓住了随机现象的本质，从而使研究的结果能广泛适合于那些表面上不同而本质上一致的各种随机现象. 例如，只包含两个样本点的样本空间，既能作为抛硬币出现"国徽向上"、"国徽向下"的模型，又能用于描述产品检验中出现"合格品"、"不合格品"的模型，以及用于描述气象中"下雨"、"不下雨"的模型等等. 由于这样的原因，今后要用抽象化的随机模型作为同一类随机试验的代表.

在引入了样本空间的概念后，就可以从集合论的观点，说明随机事件. 为此还是从考察一个例子开始.

例 1-5 某盒中装有 2 只白球和 2 只黑球，考虑"不放回依次从中摸出两球"所可能发生的事件. 若对球进行编号，2 只白球分别编为 1、2 号，2 只黑球编为 3、4 号. 用数对(i,j)表示第一次摸到 i 号球，第二次摸到 j 号球，则所有样本点是

$(1,2),(1,3),(1,4),(2,1),(2,3),(2,4),(3,1),(3,2),(3,4),(4,1),(4,2),(4,3)$

一共有 12 个样本点，它们构成了这个试验的样本空间.

同时，也可以研究这个试验的另外一些事件：

$A=$"第一次摸出白球；"

A 事件发生当且仅当下列样本点之一出现：

$$(1,2),(1,3),(1,4)$$
$$(2,1),(2,3),(2,4)$$

$B=$"第二次摸出黑球"；

B 事件发生当且仅当下列样本点之一出现：

$$(1,3),(2,3),(4,3)$$
$$(1,4),(2,4),(3,4)$$

$C=$"两次都摸出白球".

C 事件发生当且仅当下列样本点之一出现：

$$(1,2),(2,1)$$

从集合论的观点可以看出最简单的随机事件(基本事件)是只有一个样本点的集合，而随机事件 A、B、C 分别是含有 6 个样本点的集合和含有 2 个样本点的集合，且都是样本空间的子集. 进一步可以说，随机事件是样本空间的子集.

随机事件 A 是对应随机试验的样本空间 Ω 的子集. 在每次试验中，当且仅当这 A 中所包含的样本点之一出现，则称 A 事件发生.

样本空间 Ω 包含所有的样本点，它是 Ω 自身的子集，在每次试验中它总是发生的，所以样本空间是必然事件. 空集 Φ 不包含任何样本点，它也是样本空间 Ω 的子集，它在每次试验中都

不发生，所以空集是不可能事件. 由一个样本点组成的单点集是基本事件.

1-2 事件间的关系与运算

在一个随机试验中，可以观测到很多事件，它们各有其特点，彼此之间又有一定的联系. 而且其中有些事件比较简单，有些事件比较复杂，如何用一些简单事件来表示复杂的事件?因此，就需要研究事件之间的关系及运算.

例如，检验圆柱形产品，要求它的长度与直径都合格时才算合格. 这时，就要考虑“产品合格”、“产品不合格”、“长度合格”、“长度不合格”、“直径合格”、“直径不合格”、“直径合格且长度不合格”、“直径与长度均不合格”等等事件，显然这些事件之间存在关系，须找出它们之间存在的关系和运算. 通过将复杂事件化为简单事件的关系和运算，将有利于对复杂事件统计规律性的叙述和研究.

下面讨论事件之间的几个主要关系和运算.

1-2-1 子事件

定义 如果事件 A 发生必然导致事件 B 发生，则称事件 A 是事件 B 的**子事件**，或称事件 A 包含在事件 B 中，记为

$$A \subset B \text{ 或 } B \supset A$$

例如，“直径不合格”⇒“产品不合格”

故 “直径不合格”⊂“产品不合格”.

从样本空间的子集表示事件的观点来看，A 是 B 的子事件意味着所有属于 A 的样本点一定都是属于 B 的. 其直观形象可用 *Venn* 图表示，如图 1-1。

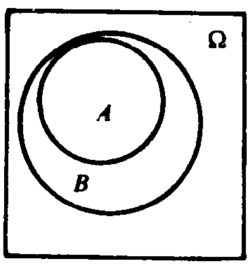


图 1-1

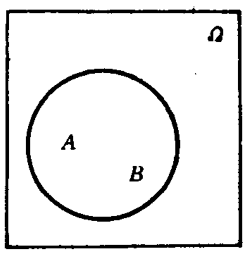


图 1-2

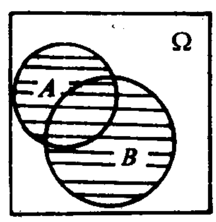


图 1-3

显然，任一事件A总是必然事件的子事件，不可能事件Φ为任一事件A的子事件，即$\Phi \subset A \subset \Omega$.

1-2-2 两事件相等

定义 如果对事件 A 和 B，同时有 $A \subset B$ 和 $B \subset A$ 成立，则称事件 A 与事件 B **相等**，记作

$$A = B$$

$A=B$ 实际上表示 A 和 B 为同一事件，它们所包含的样本点完全相同，如图 1-2 所示.

例如，“产品合格”=“长度与直径都合格”.

事件的包含及等价关系有下列性质：

如果 $A\subset B$ 且 $B\subset C$，则 $A\subset C$；

如果 $A=B$ 且 $B=C$，则 $A=C$.

1-2-3 和事件

定义 C 事件为由事件 A 与 B 至少有一个发生所导致的事件，称 C 为 A、B 的**和事件**，或称 C 为 A 与 B 的并. 记为

$$C=A\cup B$$

例如，“产品不合格”=“长度不合格”$\cup$”直径不合格”

从集合论的观点看，A 与 B 的和事件 C 是由事件 A 和 B 所有样本点构成的集合. 如图 1-3 所示，阴影部分即为 $A\cup B$.

类似地，由 n 个事件 $A_1,A_2,\cdots,A_n$ 中至少一个发生所导致的事件 C，称为是 $A_1,A_2,\cdots,A_n$ 的和事件，记为

$$C=A_1\cup A_2\cup\cdots\cup A_n \quad 或 \quad C=\bigcup_{i=1}^{n}A_i$$

由可列个事件 $A_1,A_2,\cdots,A_n,\cdots$ 中至少一个发生所导致的事件 C，称为是可列个事件 $A_1,A_2,\cdots,A_n,\cdots$ 的和事件，记为

$$C=A_1\cup A_2\cup\cdots\cup A_n\cup\cdots \quad 或 \quad C=\bigcup_{i=1}^{\infty}A_i$$

1-2-4 积事件

定义 由事件 A 与 B 同时发生所导致的事件 D 称为 A 与 B 的**积事件**. 记为

$$D=AB \quad 或 \quad D=A\cap B$$

例如，“产品合格”=“长度合格”$\cap$“直径合格”.

A 与 B 的积事件 $A\cap B$ 是由 A 与 B 所有公共样本点构成的集合. 如图 1-4 所示.

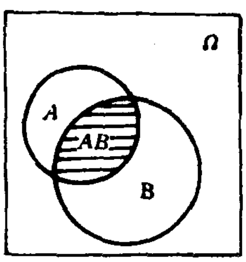


图 1-4

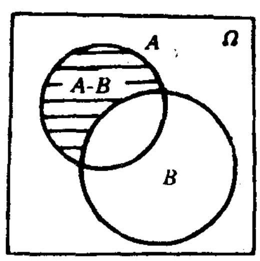


图 1-5

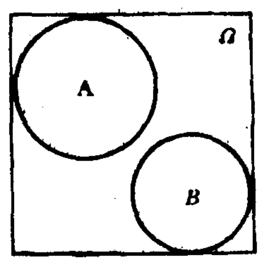


图 1-6

类似地，由事件 $A_1,\cdots,A_n$ 都同时发生所导致的事件 D 称为 $A_1,\cdots,A_n$ 的积事件，记为

$$D=A_1\cdots A_n \quad 或 \quad D=\bigcap_{i=1}^{n}A_i$$

由可列个事件 $A_1,\cdots,A_i,\cdots$ 都同时发生所导致的事件 D，称为可列个事件 $A_1,\cdots,A_i,\cdots$ 的积事件. 记为

$$D=A_1\cdots A_i\cdots \quad 或 \quad D=\bigcap_{i=1}^{\infty}A_i$$

1-2-5 差事件

定义 由事件 A 发生而事件 B 不发生所导致的事件 C 称为 A 与 B 的**差事件**，记为

$$C=A-B$$

例如,“长度合格而直径不合格”=“长度合格”−“直径合格”.

从集合论观点看,$A-B$ 表示包含在事件 A 中而不包含在事件 B 中的样本点的集合.如图 1-5 所示.

1-2-6 互斥事件

定义 若事件 A 与 B 不可能同时发生,即 A 与 B 的积事件为不可能事件 $AB=\Phi$,则称 A 与 B 为**互斥事件**,或称 A 与 B 互不相容.

例如,“产品合格”∩“直径不合格”$=\Phi$,故“产品合格”与“直径不合格”是互斥事件.

从集合论观点看,A 与 B 互斥,即 $AB=\Phi$,表示 A 与 B 没有公共的样本点.如图 1-6 所示.

显然,基本事件(只有一个样本点的集合)之间是互不相容.

两两互斥 若事件 $A_1,\cdots,A_n$(或 $A_1,\cdots,A_n,\cdots$)中任意两个事件都互斥,则称事件组 $A_1,\cdots,A_n$(或 $A_1,\cdots,A_n,\cdots$)两两互斥.

当 $AB=\Phi$ 时,则 $A\cup B$ 可记为 $A+B$;

同样,当 $A_iA_j=\Phi(i\neq j,\ i,j=1,2,\cdots,n$ 或 $i,j=1,2,\cdots)$时,则 $\bigcup\limits_{i=1}^{n}A_i$ 可记为 $\sum\limits_{i=1}^{n}A_i$,$\bigcup\limits_{i=1}^{\infty}A_i$ 可记为 $\sum\limits_{i=1}^{\infty}A_i$.

1-2-7 逆事件

定义 如果两事件 A 与 B 中必有一个发生,而又不能同时发生,即 $A\cup B=\Omega$ 且 $AB=\Phi$ 同时成立,则称 B 是 A 的**逆(对立)事件**,也称 A,B 是**互逆事件**,记为

$$B=\overline{A} \quad \text{或} \quad A=\overline{B}$$

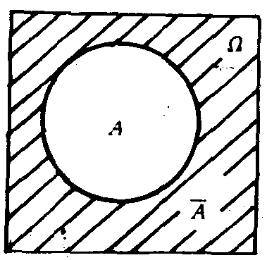


图 1-7

例如,“直径合格”∪“直径不合格”$=\Omega$,且“直径合格”∩“直径不合格”$=\Phi$,则“直径合格”与“直径不合格”是互逆事件,或互相对立事件.

从集合论观点看,A 的逆事件 $\overline{A}$ 是由样本空间中所有不属于 A 的那些样本点组成的集合.如图 1-7 所示.

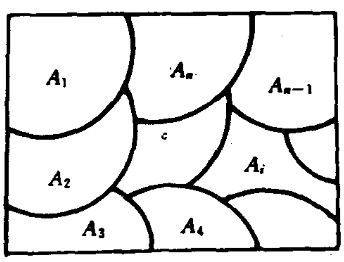


图 1-8

注意 互斥事件与互逆事件是既有联系又有区别的两个不同的概念.互逆的两事件必然互斥,因为都满足两事件不能同时发生;但是互斥的两事件不一定是互逆的,因为互斥的两事件的和事件不一定是必然事件.

例如,“产品合格”与“直径不合格”是互斥事件,但不是互逆事件,而“产品合格”与“产品不合格”才是互逆事件,当然必是互斥事件.

完备事件组 若事件 $A_1,\cdots,A_n$ 为两两互斥,且 $\sum\limits_{i=1}^{n}A_i=\Omega$,则称 $A_1,\cdots,A_n$ 构成一个完备事件组.

构成完备事件组的事件 $A_1,\cdots,A_n$ 的直观形象如图 1-8 所示，它们表现为任意两个事件都不含公共的样本点，而这 n 个事件的全体恰好拼成样本空间 Ω.

显然，随机试验 E 的所有基本事件构成一个完备事件组.

从以上的讨论可以看出，概率论中事件之间的运算关系与集合论中集合之间相应运算关系是完全一致的. 为使用方便起见，现将事件之间的关系与集合之间的关系对照列于表 1-1：

表 1-1

记　号	概　率　论	集　合　论
Ω	样本空间、必然事件	空间
Φ	不可能事件	空集
$\omega\in\Omega$	样本点	Ω 中的点(或元素)
$\{\omega\}$	基本事件	单点集
$A\subset\Omega$	事件 A	Ω 的子集
$A\subset B$	A 是 B 的子事件	A 是 B 的子集
$A=B$	事件 A 与 B 相等	集合 A 与 B 相等
$A\cup B$	事件 A 与 B 的和	集合 A 与 B 的并(或和)集
$A\cap B$	事件 A 与 B 的积	集合 A 与 B 的交(或积)集
$A-B$	事件 A 与 B 的差	集合 A 与 B 的差集
$AB=\Phi$	事件 A 与 B 互斥	集合 A 与 B 无公共元素
$\overline{A}$	A 的逆事件	A 的余集

1-2-8 事件的运算规律

由于随机事件是样本空间的子集，所以从集合的运算规律可得事件的运算规律如下：

(1) 交换律：$A\cup B=B\cup A,\ AB=BA$ (1-1)

(2) 结合律：$(A\cup B)\cup C=A\cup(B\cup C)$ (1-2)

$(AB)C=A(BC)$ (1-3)

(3) 分配律：$(A\cup B)\cap C=AC\cup BC$ (1-4)

$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$ (1-5)

(4) 德摩根(De Morgan)定理(对偶律)：

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A}_i \tag{1-6}$$

$$\overline{\bigcap_{i=1}^{n}A_i}=\bigcup_{i=1}^{n}\overline{A}_i \tag{1-7}$$

(5) 差化积：$A-B=A\overline{B}$ (1-8)

(6) 吸收律：如果 $A\subset B$，则 $A\cup B=B,AB=A$ (1-9)

由于事件就是集合，所以关于事件进行混合运算时的先后顺序，与集合进行混合运算时的先后顺序是一样的，即先进行求逆的运算，再进行积的运算，最后按出现的先后顺序进行和与差的运算；如果要求和与差的运算在先，则应加上括号；在式子中带有括号时，则应先按上述的

运算顺序完成括号内的运算，再按上述运算顺序完成全部运算.

以上规律是不难证明的.这里用集合论的语言证明(1-6)、(1-8)式，请读者把它们“翻译”成概率论的语言，并证明其余各式，以作练习.

(1-6)式的证明：

设 $\omega\in\overline{\bigcup_{i=1}^{n}A_i}$，即 $\omega\overline{\in}A_1\cup A_2\cup\cdots\cup A_n$，这表明 ω 不属于 $A_1,A_2,\cdots,A_n$ 中任一个，也就是 ω 既不属于 A_1，不属于 A_2，不属于 A_3，…，也不属于 A_n，

即 $\qquad \omega\overline{\in}A_1,\omega\overline{\in}A_2,\cdots,\omega\overline{\in}A_n$

也就是 $\qquad \omega\in\overline{A}_1,\omega\in\overline{A}_2,\cdots,\omega\in\overline{A}_n$

同时成立，所以

$$\omega\in\bigcap_{i=1}^{n}\overline{A}_i$$

于是

$$\overline{\bigcup_{i=1}^{n}A_i}\subset\bigcap_{i=1}^{n}\overline{A}_i$$

成立.

反过来，设 $\omega\in\bigcap_{i=1}^{n}\overline{A}_i$，即同时有

$$\omega\in\overline{A}_1,\omega\in\overline{A}_2,\cdots,\omega\in\overline{A}_n$$

从而同时有

$$\omega\overline{\in}A_1,\omega\overline{\in}A_2,\cdots,\omega\overline{\in}A_n$$

这意味着 ω 不属于 $A_1,A_2,\cdots,A_n$ 中任一个，即

$$\omega\overline{\in}\bigcup_{i=1}^{n}A_i$$

也就是

$$\omega\in\overline{\bigcup_{i=1}^{n}A_i}$$

这就说明

$$\overline{\bigcup_{i=1}^{n}A_i}\supset\bigcap_{i=1}^{n}\overline{A}_i$$

成立.因而有

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A}_i.$$

(1-8)式的证明：

设 $\omega\in(A-B)$，即 $\omega\in A$ 且 $\omega\overline{\in}B$，也就是

$$\omega\in A,\omega\in\overline{B}$$

同时成立，所以

$$\omega\in A\overline{B}$$

于是

$$(A-B)\subset A\overline{B}$$

反过来，设 $\omega\in A\overline{B}$，即同时有

$$\omega\in A,\omega\in\overline{B}$$

从而同时有 $\qquad \omega\in A,\omega\overline{\in}B$

这意味着 ω 只属于 A 而不属于 B，即

$$\omega\in(A-B)$$

说明

$$(A-B)\supset A\overline{B}$$

成立，因而有

$$A-B=A\overline{B}.$$

在许多场合，用集合的术语和表达方法表示事件之间的关系和运算，显得十分简练，并且也易于理解. 但是，对于初学概率论者来说，重要的是学会用概率论的语言来解释集合间的关系及运算，并能运用它们.

例 1-6 在 100 件产品中有 4 件次品，不放回抽取 3 次，每次任抽 1 件的抽样试验中，用 A_i 表示事件"第 i 次抽到正品"$(i=1,2,3)$试用 A_1,A_2,A_3 表示下列各事件：

(1) 抽到的三个产品中，没有一个产品是次品；

(2) 抽到的三个产品中，至少有一个产品是次品；

(3) 抽到的三个产品中，只有一个产品是次品.

解 (1) 事件"没有一个产品是次品"意味着所抽三个产品全是正品，即第一、二、三次同时抽到正品，则可表示为

$$A_1A_2A_3$$

(2) 事件"至少有一个产品是次品"意味着"第一次抽到次品"，"第二次抽到次品"，"第三次抽到次品"至少一个发生所导致的事件，即可表示为

$$\overline{A}_1\cup\overline{A}_2\cup\overline{A}_3$$

(3) 事件"只有一个产品是次品"意味着所抽三个产品中只有一个产品是次品，而其他两个产品是正品，而这一个次品可能是第一次抽到，或是第二次抽到，或是第三次抽到，因此可表示为

$$\overline{A}_1A_2A_3\cup A_1\overline{A}_2A_3\cup A_1A_2\overline{A}_3$$

注意 按(3)的解题思路，(2)中的"至少一个产品是次品"事件，又可表示为

$$\overline{A}_1A_2A_3\cup A_1\overline{A}_2A_3\cup A_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3\cup A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2\overline{A}_3$$

这说明有的事件可有多种表示. 事实上，从事件相等的定义不难证明事件"至少一个产品是次品"的两种表示是等价的.

例 1-7 如果 $A\subset B$，试证 $A\cup B=B$.

分析 根据事件相等的定义，只需证明

$$A\cup B\subset B \quad 且 \quad B\subset A\cup B.$$

也就是说，只需证明 $A\cup B$ 的样本点都属于 B，B 的样本点也都属于 $A\cup B$.

证明 设 $\omega\in A\cup B$，那末 $\omega\in A$ 或 $\omega\in B$，

因为 $A\subset B$，因此不论 $\omega\in A$ 还是 $\omega\in B$，必有 $\omega\in B$，于是 $A\cup B\subset B$.

又设 $\omega\in B$，因为 $A\subset B$，有 $\omega\in A$ 或 $\omega\in B$，即 $\omega\in A\cup B$，于是 $B\subset A\cup B$.

从而 $A\cup B=B$.

注意 以上证法通常用于证明事件的运算规律，至于其它事件等式的证明，以运用事件间的关系和运算规律为宜.

例 1-8 试证：

(1) $A-B=A-AB$；

(2) $A=AB\cup A\overline{B}$

(3) $(A-AB)\cup B=A\cup B$

证明 (1) $A-B=A\overline{B}$ （差化积）

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

$$=A(\overline{B}\cup\overline{A}) \quad \text{(分配律)}$$

$$=A(\overline{AB}) \quad \text{(对偶律)}$$

$$=A-AB \quad \text{(差化积)}$$

(2) $A=A\Omega=A(B\cup\overline{B})$ （吸收律）

$$=AB\cup A\overline{B} \quad \text{(分配律)}$$

(3) $(A-AB)\cup B=A\,\overline{AB}\cup B$ （差化积）

$$=A(\overline{A}\cup\overline{B})\cup B \quad \text{(对偶律)}$$

$$=A\overline{A}\cup A\overline{B}\cup B \quad \text{(分配律)}$$

$$=A\overline{B}\cup B=A\overline{B}\cup AB\cup B \quad \text{(吸收律)}$$

$$=A(\overline{B}\cup B)\cup B \quad \text{(分配律)}$$

$$=A\Omega\cup B=A\cup B \quad \text{(吸收律)}$$

1-3 随机事件的频率 概率的统计定义

前面介绍了作为概率论研究对象的随机现象，以及作为随机现象具体体现的随机试验可能结果的随机事件的概念. 下一步工作是研究随机现象的统计规律性，具体地说，随机事件的发生的可能性大小有没有规律性？若有，又如何用"数"来"度量"随机事件发生的可能性大小呢？

为了回答以上问题，首先引入随机事件的频率概念，它描述了随机事件发生的频繁程度，进而通过实例引出度量随机事件在一次试验中发生的可能性大小的数——概率.

1-3-1 随机事件的频率

定义 在相同的条件下，进行了 n 次试验，在这 n 次试验中，事件 A 发生的次数为 r，则比值 $\frac{r}{n}$ 称为事件 A 发生的**频率**，记为 $f_n(A)$. 即

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

由定义，易见频率具有下述基本性质：

(1) 非负有界性：对任意事件 A，有

$$0\leqslant f_n(A)\leqslant 1;$$

(2) 规范性：$f_n(\Omega)=1$；

(3) 有限可加性：若 $A_1,A_2,\cdots,A_m$ 是两两互斥事件，则

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

1-3-2 事件频率的稳定性

由于事件 A 发生的频率是它发生的次数与试验总次数之比，其大小表示事件 A 发生的频繁程度. 频率愈大，事件 A 发生愈频繁，这意味着事件 A 在一次试验中发生的可能性愈大. 因而，直观的想法是用频率来表示 A 在一次试验中发生的可能性的大小. 这是否可行呢？这还要

取决于当试验重复多次时，事件 A 发生的频率是否具有一定的稳定性，即当随机试验次数增大时，事件 A 发生的频率在[0,1]上某个固定的数附近摆动. 如果具有这种稳定性，则该固定的数才能表示 A 在一次试验中发生的可能性的大小. 为此，考察以下几个例子.

例 1-9 抛硬币试验. 抛一枚均匀的硬币 5 次，50 次、500 次，各做 10 遍. 得到数据如表 1-2 所示(其中 A 表示"国徽向上").

表 1-2

实验序号	$n=5$		$n=50$		$n=500$	
	r	$f_n(A)$	r	$f_n(A)$	r	$f_n(A)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	253	0.506
4	5	1.0	25	0.50	256	0.512
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

这种试验，历史上曾有不少人作过，其结果如表 1-3 所示的数据.

表 1-3

实验者	n	r	$f_n(A)$
德・摩根	2048	1061	0.5181
蒲　丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
维　尼	30000	14994	0.4998

从以上数据可以看出：(1)频率有随机波动性，即对相同的 n，其 $f_n(A)$不尽相同；(2)n 较小时，$f_n(A)$随机波动幅度较大，但随着n增大，$f_n(A)$呈现出稳定性. 即随n逐渐增大时，$f_n(A)$总是在 0.5 附近摆动，而逐渐稳定于 0.5.

例 1-10 英文字母使用频率统计表如表 1-4. 对任一篇英文文章中出现的英文字母进行统计，当统计的字母总数 n 较小时，各字母使用频率有较大幅度的随机波动. 但随 n 增大时，频率呈现出稳定性. 下面是 Dewey 统计了一本小说 438023 个字母得到的英文字母使用频率表.

研究字母使用频率，对于打字机键盘的设计(在便于操作的地方安排使用频率较高的字母键)、信息的编码(常用字母用较短的码)、密码的破译(利用密码符号频率与字母频率对照，再运用文字结构和语法规则来破译)都有极为重要的作用.