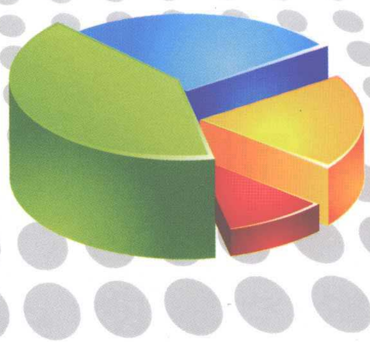
普通高等教育“十二五”规划教材

概率论与数理统计

毛志勇 孙春花◇主 编



科 学 出 版 社

普通高等教育“十二五”规划教材

概率论与数理统计

毛志勇　孙春花　主编

李国晖　陈志芳　韩　猛　米国芳　副主编

科学出版社

北　京

内 容 简 介

本书是按照国家教育部《概率论与数理统计课程教学基本要求》，结合编者多年来的教学经验而编写，本着“少学时、低起点、易理解”的编写思路，内容力求做到简明扼要、清晰易懂。

本书共8章，内容包括随机事件及其概率、随机变量及其分布、随机向量及其分布、随机变量的数字特征、大数定律与中心极限定理、抽样与抽样分布、参数估计、假设检验。

本书可作为少数民族地区普通高等院校经济与管理类本（专）科“概率论与数理统计”课程的教材，也可供从事社会与经济统计的工作者参考。

图书在版编目（CIP）数据

概率论与数理统计/毛志勇，孙春花主编．—北京：科学出版社，2013
（普通高等教育“十二五”规划教材）
ISBN 978-7-03-036673-3

Ⅰ.①概… Ⅱ.①毛… ②孙… 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026781号

责任编辑：王新文 李瑜 / 责任校对：马英菊
责任印制：吕春珉 / 封面设计：东方人华平面设计部

科学出版社出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
新科印刷有限公司印刷
科学出版社发行 各地新华书店经销
*
2013年2月第 一 版 开本：B5（720×1000）
2013年2月第一次印刷 印张：13 3/4
字数：238 000
定价：25.00元
（如有印装质量问题，我社负责调换〈新科〉）
销售部电话 010-62140850 编辑部电话 010-62135517-2037

前　言

概率论与数理统计是研究随机现象中量的规律性的一门学科，是近代数学的组成部分，同时也是近代经济理论应用与研究的重要工具之一。概率论与数理统计在许多科学领域中都有广泛的应用，如近代物理、自动控制、地震预报和气象预报、产品质量控制、农业试验和公用事业等。特别是随着概率论与数理统计方法在经济、金融、管理科学领域的深入应用，已极大地改变了经济、金融、管理科学传统的研究方式，成为研究与分析的有力工具。可以说，概率论与数理统计已成为从事经济数量分析人员必修的基础课程。

为了适应少数民族地区高等财经类院校本（专）科学生学习概率论与数理统计课程的需要，我们在近些年教学讲义的基础上，结合学生学习特点，以及编者在教学中的一些心得与体会，本着“少学时、低起点、易理解”的原则，编写了此书。

本书主体——概率论部分共包含五章内容，第 1 章主要介绍概率论中的基本概念和一些基本结论，并对几类常见的概率模型进行了讨论；第 2、3 章介绍了随机变（向）量引入的方法与目的，着重讨论了随机变（向）量的概率分布及变量之间的相互联系；第 4 章介绍了随机变量的数字特征，同时还引入了特征函数这一重要概念；第 5 章介绍了大数定律与中心极限定理及它们的理论意义。

根据少数民族地区的生源特点，数理统计部分只选择介绍了三章内容。第 6 章主要阐述了统计量的概念及寻求抽样分布的基本方法与相关结论；第 7 章介绍了参数估计的基本概念与统计思想，重点讨论了点估计与区间估计的基本方法和类型；第 8 章介绍了假设检验的统计思想与基本方法和步骤，重点介绍了正态总体参数与总体比例的检验类型，为数理统计的应用提供必要的理论基础知识。

本书第 1 章由李国晖编写，第 2 章由孙春花编写，第 3 章由陈志芳、毛志勇编写，第 4、5 章由毛志勇、韩猛编写，第 6～8 章由毛志勇、李国晖、陈志芳、米国芳共同编写。在编写过程中，各章内容都经过反复讨论和多次修改。另外，刘勇、王娟等老师参加了图表绘制及参考文献的整理工作。

本书的编写自始至终得到了内蒙古财经大学统计与数学学院各位领导和老师及科学出版社的支持与帮助，在此一并表示衷心的感谢。

限于编者的水平，书中难免出现缺点与错误，敬请广大读者对本书提出宝贵意见与建议，对不妥之处给予批评指正。

编　者

2012 年 9 月

目　　录

§1 随机事件及其概率

现代数学分为六大分支，而概率论是这六大分支的其中之一。它是一门研究在自然界普遍存在的随机现象的数量规律性的数学学科。由于随机现象的广泛存在性，概率论逐渐成为近代数学的重要组成部分，同时也是近代和现代经济理论应用与研究的重要数学工具。

在 20 世纪末期，欧洲流行赌博，而且赌法复杂，赌注量大。在这种情况下，一些职业赌徒就开始研究在赌博中取胜的机会。大约在 1653 年，卡尔达诺写成的《掷骰游戏之书》是现存最早的概率论著作。其中阐述了掷骰、打牌等游戏中的数学道理，得到了相当于现在概率论中的幂定理、大数定律等一些基本命题。

大约在 1657 年，荷兰数学家惠更斯的《论赌博中的计算》是目前最早公开发表的概率论专著。该专著成为概率论出现前的代表作，对概率论的建立有较大影响。

而一般认为概率论的鼻祖应当是 17 世纪的法国数学家帕斯卡（Pascal，1623～1662）和费马（Fermat，1601～1665）。他们当时研究的模型较简单，就是现在通称的古典概型，同时建立了概率论的一些基本概念及基本性质和运算法则。1713 年，伯努利《猜度术》的出版，标志着概率论成为数学中的一个独立分支。17 世纪到 19 世纪，伯努利、拉普拉斯、高斯、泊松、切比雪夫、马尔可夫等著名数学家都对概率论的发展做出了杰出的贡献。在这段时间内，概率论的发展简直到了使人着迷的程度。但随着科技的不断发展，在某些科学领域中，概率定义的局限性明显暴露出来，甚至无法解决一般的随机现象。直到 1933 年，苏联数学家柯尔莫哥洛夫发表了《概率论的基本概念》，提出了概率论的公理化定义，成为概率论发展史上的一个里程碑。

概率论在许多科学领域中都有重要的应用，如近代物理、自动控制、地震预报和气象预报、产品质量控制、农业试验和公用事业等。值得一提的是，由于概率论的运用，改变了经济、金融、管理科学的传统的研究方式。目前越来越多的概率论方法被引入经济、金融和管理科学，成为有力工具；甚至可以说，“概率论”成了从事经济数量分析人员的必读课程。

现在，概率论已经发展成为与实际紧密联系的理论严谨的数学学科，其内容丰富，结论深刻，有别开生面的研究课题，又有自己独特的概念和方法，成为近代数学的一个有特色的分支。

§1.1 随机事件及其运算

1.1.1 随机现象

当观察自然和人类社会中各种事物的变化规律时，我们会发现两种不同类型的现象，一种称之为**必然现象**（或决定性现象），它在一定条件下必然会出现某个结果。例如，在标准大气压下，水加热到100℃时必然沸腾；一盒黄色乒乓球中任取一球，这个球一定是黄色的；纯粹自由竞争的市场上，商品供应量的增大，必然引起价格的下落；在没有外力作用下，做匀速直线运动的物体必然继续做匀速直线运动；太阳总是从东方升起；生物总是要经历生长、发育、衰老直至死亡各个阶段，等等。这种在一定条件下必然发生的事情叫做**必然事件**。反之，那种在一定条件下，必然不会发生的事情就称为**不可能事件**。例如，在一个标准大气压下，没有加热到100℃的水是不可能沸腾的。

必然事件和不可能事件虽然表现形式不同，但两者的本质是一样的。必然事件的反面就是不可能事件，而不可能事件的反面就是必然事件。必然事件和不可能事件组成必然现象，自然中的很多现象是必然现象，概率论以外的数学分支研究的就是必然现象的数量规律。

除了必然现象以外，在自然现象和社会现象中还存在着与其有着本质区别的另一类现象。例如，检查某商店各柜台未来一天的营业额，事先无法确定各柜台营业额的大小；金融领域中事先无法断言将来某一时刻某证交所的指数；同一天生产线上用同样的工艺生产出来的灯泡的寿命长短也呈现出偶然性；做种子发芽试验，在同样的土壤和气温、技术等条件下，同一品种的种子有的发芽，有的不发芽，对某一粒种子来说，上述相同条件下“种子发芽”这个现象可能发生也可能不发生；车站候车室同一时间段的候车人数不尽相同。概率论中最经典的例子是向上掷一枚硬币，结果可能是正面也可能是反面，事先无法断定。这些现象预先是不能确定的，也就是说有偶然因素在影响，使得这些现象在一组相同的条件下不可能有确定的结果，这种现象只能用非确定性的判断来表达，我们称之为随机现象。

那么，这些有偶然因素起作用的现象是否有规律呢？透过现象看本质，这类现象也有它的内部规律性。恩格斯指出：“在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现这些规律。”为了揭示偶然性现象蕴含着的内部规律性，常常采用统计方法，即在大量重复同一个实验的情况下就可以发现偶然性现象发生的统计规律性。

1.1.2 样本空间与随机事件

为了揭示随机现象的内部规律性，就要对客观事物进行观察。这种带有明确目的性对随机现象进行观察的过程叫做**随机试验**，简称试验，通常用字母 E 表示。概率论中所讨论的试验具有如下特点：

（1）在相同条件下试验可以重复进行；

（2）每次试验的结果具有多种可能性，但试验之前可以明确试验的所有可能结果；

（3）在每一次试验前，不能准确预知该次试验将出现哪种结果。

【例 1.1.1】 E_1：抛掷一枚硬币，观察正反面情况。这个试验在相同条件下可重复抛掷多次，抛掷的结果有“正面”、“反面”两种可能结果，但在每一次抛掷之前不知该结果是“正面”还是“反面”。

E_2：观察某储蓄所一天的存储户数。这个试验的所有可能结果有 $N+1$ 个，分别是 0 户、1 户、2 户、……、N 户。其中 N 表示某个正整数，但某一天的存款户数在未观察前是不能预知的。

E_3：测量一个灯泡的使用寿命。可能值是 $[0,t]$（t 为某个实数）的任何一个实数，但每一次测量结果是不能预先知道的。

随机试验的每一个可能结果称为样本点，用 ω 表示，样本点全体组成的集合称为样本空间，用 Ω 表示。在具体问题中，给定样本空间是描述随机现象的第一步。下面是几个常见的试验。

（1）掷一枚硬币为一次试验，则有两个可能的试验结果：正面和反面，则样本空间为 $\Omega=\{$正面,反面$\}$。

（2）掷一颗骰子为一次试验，则有六种可能的试验结果：1 点，2 点，3 点，4 点，5 点和 6 点，因此样本空间为 $\Omega=\{$1 点，2 点，3 点，4 点，5 点，6 点$\}$。

（3）一枚硬币掷两次作为一次试验，将两次试验结果排序，则共有四种可能的结果：(反,反),(反,正),(正,反),(正,正)，因此样本空间 $\Omega=\{$(反,反),(反,正),(正,反),(正,正)$\}$。

（4）一颗骰子掷两次作为一次试验，将两个试验结果排序，则共有 36 种可能的结果：

$$\begin{aligned}\Omega=\{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\&(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\&(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\&(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\&(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\end{aligned}$$

$$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$
$$=\{(x,y)\mid x,y=1,2,3,4,5,6\}$$

(5) 测量某一零件的长度，考察其测量结果与真正长度的误差，样本空间 $\Omega=[-m,m]$，其中 m 为最大正误差。如果无法确定这一最大值，则样本空间可描述为 $\Omega=(-\infty,+\infty)$。这里样本空间 $\Omega=[-m,m]$ 含有无穷多个样本点，不可能把样本点一一列出，因此称这样的样本空间所含的样本点数不可列。

(6) 评价某学校小学生的健康状况，需要同时测量小学生的身高、体重和胸围，在这一随机试验中，任一可能结果（样本点）是一个有序数组 (x,y,z)，其中 x，y，z 分别代表被测学生的身高、体重、胸围，因此样本空间为

$$\Omega=\{(x,y,z)\mid 0<x<a,0<y<b,0<z<c\}$$

从以上例子可以看出，随着所考察的随机试验的不同，相应的样本空间可能很简单，也可能很复杂。

在试验 (2) 中，若用 A 表示“出现点数小于等于 3”，则 $A=\{1$ 点，2 点，3 点$\}$。那么 A 是由三个样本点所组成的集合，因此按照这种思路来定义随机事件。

定义 1.1 随机事件就是样本空间的子集，或者说随机事件就是试验结果的集合，通常用大写英文字母 A，B，C，…表示。

注： *并不是样本空间的所有子集都能作为事件，如样本空间的不可测子集。*

【例 1.1.2】 (1) 一枚硬币掷两次试验，事件 $A=$ “至少一次正面朝上”包括三个样本点：(正,反),(反,正),(正,正)。也可以表示为 $A=\{$(正,反),(反,正),(正,正)$\}$。

(2) 一颗骰子掷两次的试验，事件 $B=$ “两次点数相同”，则

$$B=\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

定义 1.2 事件发生是指当且仅当它所包含的某一个样本点出现。

在例 1.1.2 中，如果一颗骰子掷两次出现的点数是 (3,3)，则称事件 B 发生了，如果出现的点数是 (1,3)，则事件 B 没有发生。

为此介绍几个特殊的事件。

(1) 基本事件：只含有一个试验结果（样本点）的事件称为基本事件。

(2) 必然事件：每次试验必然发生的事件，称为必然事件，它是包含样本空间 Ω 所有元素的事件，用 Ω 表示。

(3) 不可能事件：每次试验一定不会发生的事件，称为不可能事件，它是不包含任何元素的空集，用 Φ 表示。

【例 1.1.3】 掷一颗骰子的试验中：A_1 表示“出现的点数是 2”这个事件，A_2 表示“出现的点数是 4”，A_3 表示“出现的点数是 6”，它们都是基本事件；B 表示“出现偶数点”，B 是由 A_1，A_2，A_3 复合而成的随机事件，称之为复合事

件。C 表示“出现的点数小于 7”这个事件，显然它是一个必然事件。D 表示“出现的点数大于等于 7”这个事件，它是不可能事件。

注：（1）必然事件和不可能事件有着紧密的联系，必然事件的反面事件一定是不可能事件。例如，“出现的点数小于 7”的反面事件为“出现的点数不小于 7”。

（2）无论是必然事件还是不可能事件，还是随机事件，都是相对于一定的试验条件而言的，如果试验条件改变了，则事件的性质也就发生了变化。

例如，掷骰子试验中，观察“点数之和小于 7”这个事件，当掷一颗骰子时它是必然事件；掷两颗骰子时，它是随机事件；掷八颗骰子时，它是不可能事件。概率论所研究的都是随机事件，为讨论问题方便起见，把必然事件和不可能事件作为随机事件的两个极端情况纳入随机事件中去研究。

1.1.3　事件间的关系及其运算

1. 用集合表示事件

事件有复合事件和基本事件，所以事件之间有一定的关系和运算。为了研究它们的关系和运算，应用点集的概念和图示的方法就比较直观，容易理解。

事件与集合的对应关系如表 1-1 所示。

表 1-1　事件与集合的对应关系

事　件	集合	记号	事　件	集合	记号
样本空间，必然事件	全集	Ω	样本点	元素	ω
不可能事件	空集	Φ	事件	子集	A

为了直观，经常使用图示来表示事件。一般地，在一个平面上用某个矩形区域表示必然事件或者整个样本空间 Ω，其中的一个子区域表示一具体的事件（简称 Venn 图），如图 1-1 所示。

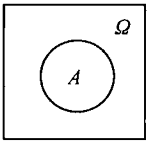


图 1-1

2. 事件的关系及运算

（1）事件的包含。

如果事件 A 发生必然导致事件 B 发生，即属于 A 的每一个样本点都属于 B，则称事件 B 包含事件 A 或称事件 A 包含于事件 B，记作 $B\supset A$ 或 $A\subset B$。等价的说法是，如果 B 不发生则 A 也不会发生。对于任何事件 A，有 $\Phi\subset A\subset\Omega$。

（2）事件的相等。

如果事件 A 包含事件 B，同时事件 B 也包含事件 A，则称事件 A 与 B 相等，即 A 与 B 中的样本点完全相同，记作 $A=B$。

（3）事件的并（和运算）。

两个事件 A、B 中至少有一个发生，称为事件 A 与 B 的并事件或和运算。它是属于 A 和 B 的所有样本点构成的集合，记作 $A+B$ 或 $A\cup B$。

可推广到多个随机事件的情形。设 n 个事件 A_1，A_2，…，A_n 中至少有一个发生是一个事件，称这个新事件为 n 个事件 A_1，A_2，…，A_n 的和，记作 $A_1+A_2+\cdots+A_n$ 或 $A_1\cup A_2\cup\cdots\cup A_n$。

设可列个事件 A_1，A_2，…，A_n，…中至少有一个发生，也称为事件 A_1，A_2，…，A_n，…的和，记作 $\sum\limits_{i=1}^{\infty}A_i$ 或 $\bigcup\limits_{i=1}^{\infty}A_i$。

（4）事件的交（积运算）。

两个事件 A 与 B 同时发生，称为事件 A 与 B 的交事件或积运算。它是由 A 与 B 的所有公共样本点构成的集合，记作 AB 或 $A\cap B$。也可推广到多个随机事件的情形。

（5）事件的差运算。

事件 A 发生而事件 B 不发生是一个事件，称为事件 A 与 B 的差。它是由属于 A 但不属于 B 的那些样本点构成的集合，记作 $A-B$，显然 $A-B=A-AB$ 。

（6）互不相容事件（或称互斥关系）。

如果事件 A 与 B 不能同时发生，即 $AB=\Phi$，则称事件 A 与 B 互不相容（或称事件 A 与 B 互斥）。因此，互不相容事件 A 与 B 没有相同的样本点。显然，基本事件间是互不相容的。

（7）对立事件。

事件“非 A”称为 A 的对立事件。它是由样本空间中所有不属于 A 的样本点组成的集合，记作 $\overline{A}$。显然有 $A\overline{A}=\Phi$，$A+\overline{A}=\Omega$，$\overline{A}=\Omega-A$，$\overline{\overline{A}}=A$。

（8）完备事件组。

若事件 A_1，A_2，…，A_n 为两两互不相容事件，并且 $A_1+A_2+\cdots+A_n=\Omega$，则称 A_1，A_2，…，A_n 构成一个完备事件组或构成一个划分。

可推广到可列个随机事件的情形。

事件的关系及运算如图 1-2 所示。

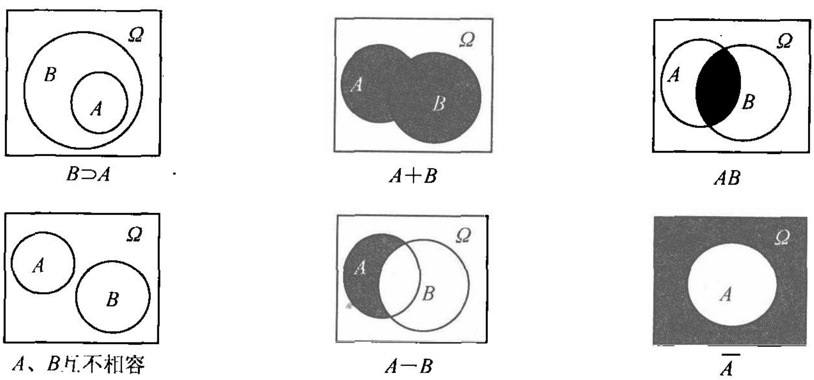


图 1-2

事件的运算和集合的运算有类似的性质。设 A，B，C 是随机事件，则有以下运算定律。

（1）交换律：$A\cap B=B\cap A$，$A\cup B=B\cup A$。

（2）结合律：$(A\cap B)\cap C=A\cap(B\cap C)$，$(A\cup B)\cup C=A\cup(B\cup C)$。

（3）分配律：$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$，$(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$。

（4）对偶律：$\overline{A\cup B}=\overline{A}\cap\overline{B}$，$\overline{A\cap B}=\overline{A}\cup\overline{B}$。

分配律和对偶律可以推广到任意个有限事件或无限多个事件上去。例如，$(\bigcup A_i)C=\bigcup A_iC$，$(\bigcap A_i)C=\bigcap A_iC$；$\overline{\bigcup A_i}=\bigcap\overline{A_i}$，$\overline{\bigcap A_i}=\bigcup\overline{A_i}$。

【例 1.1.4】 掷一颗骰子的试验，观察出现的点数，A 表示“奇数点”，B 表示“点数小于 5”，C 表示“小于 5 的偶数点”。用集合的列举表示法表示下列事件：Ω，A，B，C，$A+B$，$A-B$，$B-A$，AB，AC，$\overline{A}$。

解：$\Omega=\{1,2,3,4,5,6\}$，$A=\{1,3,5\}$，$B=\{1,2,3,4\}$，$C=\{2,4\}$，$A+B=\{1,2,3,4,5\}$，$A-B=\{5\}$，$B-A=\{2,4\}$，$AB=\{1,3\}$，$AC=\Phi$，$\overline{A}=\{2,4,6\}$。

【例 1.1.5】 从一批产品中每次取出一个产品进行检验（每次取出的产品不放回），事件 A_i 表示第 i 次取到的是合格品（$i=1,2,3$）。试用事件的运算符号表示下列事件：

（1）三次都取到了合格品；

（2）三次中至少有一次取到合格品；

（3）三次中恰有两次取到合格品；

（4）三次中最多有一次取到合格品。

解：（1）三次都取到合格品：$A_1A_2A_3$。

（2）三次中至少有一次取到合格品：$A_1+A_2+A_3$。

（3）三次中恰有两次取到合格品：$A_1A_2\overline{A}_3+A_1\overline{A}_2A_3+\overline{A}_1A_2A_3$。

（4）三次中最多有一次取到合格品：$\overline{A}_1\overline{A}_2+\overline{A}_1\overline{A}_3+\overline{A}_2\overline{A}_3$。

【例 1.1.6】 如果 x 表示一个沿数轴做随机运动的质点的位置，试说明下列各事件的关系：

$A=\{x\mid x\leqslant 20\}$，$B=\{x\mid x>3\}$，$C=\{x\mid x<9\}$，$D=\{x\mid x<-5\}$，$E=\{x\mid x\geqslant 9\}$

解：由图 1-3 可见，$A\supset C\supset D$；$B\supset E$；D 与 B，D 与 E 互不相容；C 与 E 为对立事件；B 与 C，B 与 A，E 与 A 相容；显然 A 与 C，A 与 D，C 与 D，B 与 E 也是相容的。

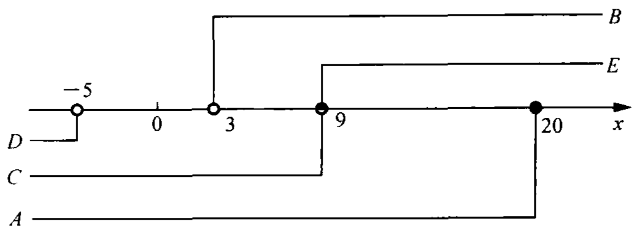


图 1-3

【例 1.1.7】 在统计系学生中任选一名学生，令事件 A 表示被选的学生是男生，事件 B 表示该生是三年级学生，事件 C 表示该生是运动员。

(1) 试述事件 $AB\overline{C}$ 的意义；

(2) 在什么条件下 $ABC=C$ 成立?

(3) 什么时候关系式 $B\supset C$ 是正确的?

解：(1) $AB\overline{C}=$ "选出的学生是三年级男生但不是运动员"。

(2) 在运动员都是三年级男生的条件下，$ABC=C$。

(3) 在运动员全是三年级学生时，$B\supset C$。

§1.2 随机事件的概率

概率论研究的是随机现象的规律性。因此，仅仅知道试验可能发生哪些结果是不够的，还必须对这些结果（事件）发生的情况进行量的描述。为此引入概率这一最基本的概念，来描述事件发生的可能性大小。

每一个事件都有它发生的概率，记作 $P(A)$或 $P\{A\}$。但怎样获得切合实际的一个事件的概率呢?

1.2.1 概率的统计定义

在 n 次重复试验中，如果事件 A 发生了 m 次，则 m/n 称为事件 A 发生的频率。同样若事件 B 发生了 k 次，则 k/n 为 B 发生的频率。如果 A 是必然事件，有 $m=n$，即必然事件的频率是 1，当然不可能事件的频率为 0。如果 A 与 B 互不相容，则事件 $A+B$ 的频率为 $(m+k)/n$，它恰好等于两个事件的频率之和，这称之为频率的可加性。

频率是否能够确切地描述事件发生的可能性大小呢? 经验告诉我们，多次重复同一试验，随机事件的频率具有稳定性。试验次数 n 很大时，事件 A 的频率 m/n 接近于某一个常数。这种频率的稳定性就是定义概率的经验基础。这就是说，频率不能确切地描述事件发生的可能性大小，但频率的稳定值可以描述。

为了说明频率的稳定性，先来看一些著名的例子。

【例 1.2.1】 抛掷一枚硬币，可能出现正面也可能出现反面，事先做出确定的判断是不可能的。但假如硬币是均匀的，那么我们有理由认为出现正面和反面的可能性应该一样，即在大量试验中出现正面和反面的频率都接近于 0.5，为了验证这一点，历史上曾有不少人做过试验，如表 1-2 所示。

表 1-2 历史上的掷硬币试验统计表

试验者	抛掷次数 n	正面出现的次数 m	正面出现的频率 m/n
德·摩根	2 048	1 061	0.518
蒲丰	4 040	2 048	0.506 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5
维尼	30 000	14 994	0.499 8

从上述数据可以看出：当抛掷次数较小时，频率的随机波动较大，但随着抛掷次数的增大，频率呈现出稳定性，即当 n 逐渐增大时，频率总是在 0.5 附近摆动而逐渐稳定于 0.5。

【例 1.2.2】 对于新生婴儿的性别情况，很多人做了大量的统计。法国数学家拉普拉斯在《概率论哲学探讨》一文中指出，全法国男婴出生率为 22/43；1935 年，瑞典数学家克拉美统计出瑞典的男婴出生率为 0.518；我国的几次人口普查，也得到这样的结论：男婴与女婴的比率为 1.04∶1。在这个例子中，对于同一现象，不同时期、不同国家及不同的调查人得到了几个非常接近的数字，这表明，现象具有一定的规律性。

上述种种事实表明，随机现象有其偶然性的一面，也有其必然性的一面。这种必然性表现为大量试验中随机事件出现的频率的稳定性，即一个事件出现的频率常在某个固定的常数附近摆动，这种规律性我们称之为统计规律性。频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的、不随人们意志改变的一种客观属性，因此可以对它进行度量。

定义 1.3 （概率的统计定义）在不变的条件下，重复进行 n 次试验，事件 A 发生的频率稳定地在某一常数附近摆动，且一般说来，n 越大，摆动幅度越小，则称这个常数为事件 A 的概率，记作 $P(A)$。

数值 $P(A)$ 就是一次试验中事件 A 发生的可能性大小的数量描述，它指出了事件概率的存在性，但并不能用这个定义计算事件的概率。实际中，人们常常采取大量试验的频率或用一系列频率的平均值作为事件概率的近似值。这样定义的概率并不是严格的概率定义，而只是描述了一个大数定律，这是概率论中极限理论的雏形，它的严格讨论将在本书的第 5 章给出。

1.2.2 概率的古典定义

古典方法是通过对被考察事件发生的可能性进行逻辑分析得出该事件的概率。这种方法简单、直观，不需做试验，但只能在一类特定的随机现象中使用。这类试验的特点如下。

（1）每次试验只有有限种可能的试验结果。

（2）每次试验中，各基本事件出现的可能性完全相同。

具有这两个特点的试验称为**古典概型试验**，简称古典概型。

在古典概型试验中，如果总共有 n 个可能的试验结果，那么每个基本事件发生的概率为 $1/n$，假设事件 A 包含 m 个基本事件，则事件 A 发生的概率为 m/n。

定义 1.4 若试验结果一共由 n 个基本事件 A_1，A_2，…，A_n 组成，并且这些事件的发生具有相同的可能性，而事件 A 由其中某 m 个基本事件 A_1，A_2，…，A_m 组成，则事件 A 的概率可以用下式计算：

$$P(A)=\frac{A\text{ 所包含的基本事件个数}}{\text{样本空间所包含的基本事件个数}}=\frac{m}{n}$$

【例 1.2.3】 掷一枚硬币的试验，基本事件为正面和反面，而且由于硬币的对称性，因此出现正面和反面的概率一样，都是 1/2。

【例 1.2.4】 掷一次骰子的试验，基本事件有 6 个，因此每个基本事件的概率为 1/6，则 $P\{\text{偶数点}\}=3/6=1/2, P\{\text{小于 }5\}=4/6=2/3$ 等。

古典概型试验有着多方面的应用，产品抽样检查就是其中之一。

产品抽样检查技术，在各个生产部门中被广泛采用。许多大工厂产量很高，每天生产的产品数以万计，对这些产品的质量，如果要进行全面的逐件检验通常是不可能或不经济的；另外在有些情况下，产品的检验带有破坏性（如灯泡寿命检验和棉纱强度试验），因此，最适宜的检验方法是采用抽样检查，即从产品中随机地抽查若干件来检验，并根据检验结果来判断整批产品的质量。关于产品的质量，可以有多种多样的衡量标准，这里只考虑最简单的情形，即把产品分为合格品和次品。假如产品的好坏从外形上看不出来，而且我们采用的是随机抽样，那么任何一件产品被抽到的可能性都一样，这正是古典概型。

【例 1.2.5】 为了检验某厂生产的产品质量，从该厂生产的产品中随机抽取 1000 件产品进行检验，经检验发现有 5 件次品，若 A 表示产品是次品，则基本事件有 1000 个，A 所包含的样本点数有 5 个，那么可以认为该厂生产的产品的次品率为 $5/1000=0.005$。

在研究古典概型试验时，大部分问题都能形象地用摸球模型来描述。例如，一个袋子中有 a 个白球，b 个黑球，大小形状相同，从这个袋子中任摸一球，则所有的球被摸到的可能性相同。若把黑球作为废品，白球作为正品，则这个摸球

模型就可以描述产品的抽样问题。如果产品分为多个等级，则可用多种颜色球的摸球模型来描述。

古典概型试验中许多概率的计算相当困难而又富有技巧。计算的要点是给定样本点并计算它的总数，然后计算有利场合的数目。在这些计算中，经常要用到一些排列与组合公式。

排列与组合的分析公式基于下列两条原理。

（1）**乘法原理**：如果一件事情可以分为两步做，第一步有 n 种选择，第二步有 m 种选择，则整件事情共有 $m\times n$ 种选择。

特别地，如果有 k 件事情 A_1，A_2，…，A_k有待完成，而完成 A_1有 n_1种方法，完成 A_2有 n_2种方法，完成 A_k有 n_k种方法，那么一次完成这 k 件事情就有 $n_1\times n_2\times\cdots\times n_k$种不同的方法。这个计数原理很重要，它不但可以解决不少具体问题，同时也是推导排列组合公式的基础。

（2）**加法原理**：如果一件事情进行 A_1过程有 n 种选择，进行 A_2过程有 m 种选择，假定 A_1过程和 A_2过程是并行的，则整件事情共有 $m+n$ 种选择。

排列可看做是从元素总体中随机选取若干个且考虑出现顺序的选取方法，可分为两类：第一类是有重复排列（放回抽样），即每次选取都在全体元素中进行，同一个元素可被重复选中；第二类是无重复排列（不放回抽样），这时一个元素一旦被取出便从总体中除去，因此每个元素最多被选中一次。下面分别举例介绍这两种情况。

（1）有重复排列的例子如下。

【例 1.2.6】 假设一副牌有 52 张，将它们编号为 1，2，…，52。每次取出一张观察后再放回去（这样下一次这张牌仍有机会被取到），这称为有放回抽样。假设这样取了 5 次，共有多少种可能的取法？

第一次有 52 种取法，第二次又有 52 种取法……因此取 5 次共有 $52\times52\times52\times52\times52=52^5$种取法。

【例 1.2.7】 把 m 个球随机地放到 n 个盒子中（每个盒子中可以放多个球），共有多少种放法？

第一个球可以放进 n 个盒子的任意一个盒子中，所以共有 n 种放法；第二个球也可以放进 n 个盒子的任意一个盒子中，所以也有 n 种放法……因此共有 $n\times n\times\cdots\times n=n^m$ 种放法。

一般地，从 n 个元素中进行 m 次放回抽样，则共有 n^m种不同的放法。

（2）无重复排列的例子如下。

【例 1.2.8】 在例 1.2.6 中，从 52 张牌中每次抽取一张，但不放回，则第二次抽取时只有 51 张牌，第三次就只有 50 张牌。如果这样抽取 5 次，就共有 $52\times51\times50\times49\times48=52!/47!$ 种不同的取法。