面向21世纪课程教材

Textbook Series for 21st Century

概率论与应用数理统计

刘光祖 主编

高等教育出版社

HIGHER EDUCATION PRESS

478

C21-42
L-1

面 向 2 1 世 纪 课 程 教 材
Textbook Series for 21st Century

概率论
与应用数理统计

刘光祖　主编

高 等 教 育 出 版 社
HIGHER EDUCATION PRESS

内容简介

本书是教育部"高等教育面向21世纪教学内容和课程体系改革计划"的研究成果，是面向21世纪课程教材.本书主要内容包括：随机事件及其概率、随机变量的分布及其数字特征、多维随机变量的分布及其数字特征、大数定律与中心极限定理、参数估计、假设检验、方差分析和回归分析.本书在编写中注意联系农林院校实际，选用了大量的结合农林、经济等专业的例题和习题，针对性强.

本书可作为高等农林院校各专业教材，也可供其他院校非数学专业选用.

图书在版编目(CIP)数据

概率论与应用数理统计/刘光祖主编 —北京：高等教育出版社，2000(2001重印)
ISBN 7-04-008988-2

Ⅰ.概… Ⅱ.刘… Ⅲ.①概率论②数理统计
Ⅳ.021

中国版本图书馆CIP数据核字(2000)第62094号

概率论与应用数理统计
刘光祖 主编

出版发行	高等教育出版社		
社址	北京市东城区沙滩后街55号	邮政编码：	100009
电话：	010-64054588	传真	010-64014048
网址	http://www.hep.edu.cn		
经销	新华书店北京发行所		
印刷	北京民族印刷厂		
开本	797×960 1/16	版次	2000年8月第1版
印张	26.25	印次	2001年7月第2次印刷
字数	480 000	定价	22.10元

前 言

《概率论与应用数理统计》是根据原国家教委“面向21世纪高等农林院校本科数学(含生物统计)教学内容和课程体系的改革与实践”立项项目的教材编写规划而编写的,它是高等数学系列教材之一.这套教材的立意就是要通过教材的系列来实现培养学生的数学素质和提高学生应用数学的能力.本书就是根据这一要求构思和作内容安排的.希望它能与其他教材一起,为新世纪农林院校的数学教学产生较好的整体效应.

概率论与数理统计是从数量方面研究随机现象统计规律性的数学分支.从17世纪的巴斯卡和费尔马研究“得点”问题开始,到贝叶斯、拉普拉斯和科尔莫戈罗夫等人建立概率的公理化体系,把概率的思想、方法提高为处理一般科学问题的原理,这为概率成为数学学科奠定了基础.概率理论的自身发展以及它与其他数学分支的结合,极大地影响、促进了数理统计的发展.K.皮尔逊和R.A.费歇是推动数理统计发展的标志性人物.由于大批概率统计学家先后不懈的努力和杰出贡献,使得当今概率统计的理论更加丰富,应用更加广泛.随着现代科学技术的发展,概率统计方法已广泛地用于工业、农业、医学、经济等各个领域.

本教材共8章.前4章是概率论内容,其中包括随机事件及其概率、随机变量的分布及其数字特征、多维随机变量的分布及其数字特征、大数定律与中心极限定理;后4章是数理统计内容(含生物统计部分内容),其中包括参数估计、假设检验、方差分析和回归分析.本书在编写过程中力求体现下述特点:

1.在保持必要的数学严谨性和逻辑思维科学性的基础上,注意以简洁明快的方式引入概率统计的概念和介绍概率统计的方法.对有的抽象概念,通过例题中的具体问题的直观描述予以说明;对有的定理、性质的繁难证明和方法的推导适当弱化,对有些重要例题略去了证明或计算过程,仅给出了相应的结论,但对于相关知识解决什么问题,以及如何应用相关知识都作了较为明确的解释.

概率部分渗透了统计思想,如通过作直方图来说明什么是概率密度.统计部分的某些难点适当分散,如:将χ^2分布、t分布、F分布的定义以及分位数概念在多维随机变量的分布及其数字特征一章给出,而与它们相关的统计量分布分散在参数估计和假设检验两章.引进了某些新统计方法,如广义LS估计等.

2.书中各章分为主干内容和总结与举例两部分,便于学生学习、掌握以及应用.通过各章主干内容的教学,学生阅读总结,可使他们把既有联系,又相对独立的知识贯穿起来,较好地掌握该章的重点知识线.各章后的举例是主干内容的

拓展与延伸，它们大都可以利用前面所学的概率统计知识研究解决．这些例题可以讲解，也可以作为课外阅读．为学生留有较大的自由选学空间．

3．加强数学建模，重视本课程对学生初涉科学研究的基础训练．书中除主干内容有一定数量的数学建模举例外，后面的举例部分，还建立了一些概率模型和回归模型．与此同时，注意了对数学模型的数学性态分析和它的实际意义的解释，例如生物学意义的解释．使学生看到数学与其他学科结合的重要作用，从而增强学数学、用数学的意识．在举例部分，也有少量例题展示的统计方法是主干内容中没有介绍过的，编入这类例题旨在让学生了解研究自己尚不清楚的某些科学问题，应有科学的态度和科学的方法，用自己既有的知识研究解决就是创新．编写中重视了试验资料的收集、整理、查阅及统计分析，这样做，会使学生受到一定程度的科学研究训练．

本书是为农林各专业编写的概率统计教材，学习本书内容大约需要70学时．各专业根据需要，可以选学其中的部分内容，或全部内容．本书也可供其他院校非数学专业选用，还可作为农林科研人员的参考书．

参加本书编写的有刘光祖、李新有、史美英、刘瀛州、李惠东和吴素萍．刘光祖为主编，李新有、史美英、李惠东为副主编．郑瑶为本书计算机绘图．全书由刘新平教授主审，他认真细致地对书稿进行了审查，并且提出了许多宝贵的意见和建议，为提高本书的质量起到了重要的作用．

编写面向21世纪高等农林院校的概率统计教材，为我们提供了通过教学要件——教材的建设来探索和尝试提高教学质量和培养学生素质的学习机会，由于编者水平所限，书中可能仍有疏漏或不妥之处，恳切希望专家和读者批评指正．

编　者

2000年3月

第一章　随机事件及其概率

随机事件的概率是概率论研究的基本内容.本章将主要介绍随机事件,随机事件的概率,概率的基本性质,条件概率以及计算概率常用到的几个重要公式.

第一节　随 机 事 件

1.1　随机试验与随机事件

1. 随机试验

试验,通常是指对现象的观察.在自然界与人类社会中,人们所能观察到的现象是各种各样的.但归纳起来,它们大体可以分为两类:一类是确定性现象,另一类是随机现象.

所谓确定性现象,即每当试验(或观察)具备一定条件,就能够在试验结束之前预知它的确切结果.例如:没有水分,种子不会发芽;在 101325Pa 大气压下,水加热到 100℃ 会沸腾,等等.

所谓随机现象,即每次试验有多种可能结果,但试验结束之前不能预知出现哪一种确切结果.例如:一穴播两粒种子,可能两粒都出苗,也可能一粒出苗,另一粒不出苗,还可能两粒都不出苗;一批新产品投入市场,可能畅销,也可能滞销;抛掷一枚硬币,可能出现币值面向上,也可能出现徽花面向上,等等.随机现象所呈现出的这种不确定性,只表现在一次或少数几次试验中.如果在相同条件下进行大量重复的试验,随机现象就会呈现出某种规律性,并且称这种规律性为随机现象的**统计规律性**.

研究随机现象的统计规律性,是概率论的一个基本任务.为了方便研究某一特定的随机现象的统计规律性,往往需要了解发生这一随机现象的试验背景.

如果试验可以在相同(或基本相同)的条件下重复进行;且每次试验有多种可能结果;在每次试验结束之前明确试验的所有可能结果,但不能预知出现哪一个确切结果,则称这样的试验为**随机试验**,简称**试验**.

2. 随机事件

试验的结果,叫做事件.事件可以分为三类:随机事件,必然事件和不可能事件.

(1) 随机事件　在随机试验中可能发生也可能不发生的事件,叫做**随机事件**,简称**事件**.随机事件通常用字母 A,B,C 等表示.

例如:播 3 粒种子,用 A 表示 3 粒全出苗,B 表示 2 粒出苗,1 粒不出苗,它们可分别记为 $A=\{3$ 粒全出苗$\}$,$B=\{2$ 粒出苗,1 粒不出苗$\}$. 显然 A,B 都是随机事件.

(2) 必然事件　在每次试验中都发生的事件,叫做**必然事件**. 必然事件通常用字母 Ω 表示.

(3) 不可能事件　在任何一次试验中都不发生的事件,叫做**不可能事件**. 不可能事件通常用字符 $\varnothing$ 表示.

显然,必然事件、不可能事件都是确定性事件. 为了今后讨论问题的方便,也可将它们看作是两个特殊的随机事件.

例如:$A=\{$同性电荷互斥$\}$是必然事件 Ω;$B=\{$没有水分,种子会发芽$\}$是不可能事件 $\varnothing$.

1.2　样本空间

设 E 是一随机试验,根据试验目的和要求所确定的 E 的一个最基本的结果 ω,称为 E 的一个**样本点**. 由一个样本点构成的单点集称为**基本事件**。

基本事件是随机试验中不可再分的事件,而由若干个基本事件可以构成某个较复杂的随机事件.

样本空间　由试验 E 的所有样本点构成的集合$\{\omega\}$,叫做 E 的**样本空间**,记作 Ω,即 $\Omega=\{\omega\}$.

例 1　在适宜的条件下,每穴种 3 粒西瓜种子,观察一穴的西瓜苗株数. 观察的所有可能结果有 4 种:0 株,1 株,2 株,3 株. 设 ω_i 表示该穴有 i 株西瓜苗,即 $\omega_i=\{$该穴有 i 株西瓜苗$\}$,$i=0,1,2,3$. 则试验的全部样本点是 $\omega_0,\omega_1,\omega_2,\omega_3$. 故试验的样本空间为

$$\Omega=\{\omega_0,\omega_1,\omega_2,\omega_3\}.$$

在例 1 中,如果设 $i=\omega_i=\{$该穴有 i 株西瓜苗$\}$,则试验的样本空间还可以表示为

$$\Omega=\{0,1,2,3\}.$$

例 2　测量某块地中的小麦株高(以厘米为单位),设 a 表示株高的下界,b 表示株高的上界,则介于 a,b 之间的每个实数 x 都可能是测量的结果,即测量的样本点. 故测量(试验)的样本空间 Ω 是实数轴的一个区间:$a\leqslant x\leqslant b$. 它还可表示为 $\Omega=\{x\mid a\leqslant x\leqslant b\}$.

例 3　在完全相同的条件下接连进行两次射击,观察击中目标的情况,有 4 种可能的结果:$\omega_1=\{0,0\}$,$\omega_2=\{1,0\}$,$\omega_3=\{0,1\}$,$\omega_4=\{1,1\}$. 其中$\{0,0\}$表示两次均未击中目标;$\{1,0\}$表示第一次击中目标而第二次未击中目标;$\{0,1\}$表示

第一次未击中目标而第二次击中目标；{1,1}表示两次都击中目标.故试验的样本空间为

$$\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}.$$

在例 3 中，如果试验的目的是观察击中目标的次数，且设 $\omega_i=\{$击中 i 次$\}$，则试验有 3 种可能结果：$\omega_0=\{$击中 0 次$\}$，$\omega_1=\{$击中 1 次$\}$，$\omega_2=\{$击中 2 次$\}$.故试验的样本空间为

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

由例 3 及其后继问题的讨论可以看出，试验的样本点及样本空间是根据试验的目的与要求来确定的.若试验要求相同，而试验目的不同，则由试验所确定的样本点和样本空间也就不同.

有了随机事件和样本空间的概念后，对于随机试验 E 中的任何一个事件 A，则都可以理解为该试验的样本空间 Ω 的子集，即 $A\subset\Omega$.现在，从集合论的观点对前述的一些事件再作如下说明：在随机试验 E 中，随机事件是以 E 中样本点为元素的集合；基本事件是只包含一个样本点的单元素集合；必然事件是由 E 中全体样本点组成的集合，即 E 的样本空间 Ω，亦即全集 Ω；不可能事件是不包含 E 中任何一个样本点的集合，即空集 $\varnothing$.而且，所谓事件 A 发生，是指当且仅当 A 中所包含的某个样本点在试验 E 中出现.

例 4　掷一枚骰子(骰子是均匀的小正六面体，其各面分别有 1 点，2 点，…，6 点的标识)一次，观察出现的点数.设 $\omega_i=\{$出现 i 点$\}$，$i=1,2,\cdots,6$，则试验的样本空间为

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_6\}.$$

若设事件 A,B 分别表示出现偶数点和出现奇数点，即设 $A=\{$出现偶数点$\}$，$B=\{$出现奇数点$\}$，则

$$A=\{\omega_2,\omega_4,\omega_6\},\quad B=\{\omega_1,\omega_3,\omega_5\}.$$

显然事件 A 和 B 都是样本空间 Ω 的子集.

对于例 4 中的事件 A 来说，若在一次试验中，出现 4 点(或 2 点，或 6 点)，则 A 发生；反之，若 A 发生，则在一次试验中出现了某个偶数点.同样，对于事件 B 有类似的相应说法.

1.3　事件之间的关系及运算

为了方便以后讨论或计算随机事件的概率，下面定义一些事件之间的主要关系及运算.

1. 事件的包含

如果事件 A 发生必然导致事件 B 发生，则称**事件 B 包含事件 A**，记作 $A\subset B$ 或 $B\supset A$.

例如：在10粒种子的发芽试验中，设事件 $A=\{$发芽种子数是9$\}$，$B=\{$发芽种子数不小于8$\}$，则 $A\subset B$.

2. 事件的相等(等价)

设 A,B 为二事件，如果 $A\subset B$，且 $B\subset A$，则称事件 A 与事件 B 相等(或等价)，记作 $A=B$.

例如：从编号为 $1,2,\cdots,10$ 的十套考题中，任意抽取一套题准备某次考试用. 设事件 $A=\{$抽取到的一套考题编号为4$\}$，事件 $B=\{$抽取到的一套考题是编号大于3且小于6的偶数号考题$\}$，则事件 A 与事件 B 等价，即 $A=B$.

3. 事件的和(并)

事件 A 与事件 B 至少有一个发生所构成的事件，称为事件 A 与事件 B 的和(或并)，记作 $A\cup B$.

事件 A 与 B 的和事件 $A\cup B$，也常常表述为 A 发生或 B 发生.

例如：一穴播两粒种子甲和乙，设 $A=\{$甲粒出苗$\}$，$B=\{$乙粒出苗$\}$，则 $A\cup B=\{$至少有一粒种子出苗$\}$.

对于 n 个事件 $A_1,A_2,\cdots,A_n$ 而言，类似地有，事件 $A_1,A_2,\cdots,A_n$ 中至少有一个发生所构成的事件，称为这 n 个事件 $A_1,A_2,\cdots,A_n$ 的和(或并)，记作

$$A_1\cup A_2\cup\cdots\cup A_n, \text{或} \bigcup_{i=1}^{n} A_i.$$

同样，可列个事件 $A_1,A_2,\cdots$ 至少有一个事件发生所构成的事件，称为这可列个事件 $A_1,A_2,\cdots,A_n$ 的和(或并)，记作

$$A_1\cup A_2\cup\cdots\cup A_n\cup\cdots, \text{或} \bigcup_{i=1}^{\infty} A_i.$$

4. 事件的差

事件 A 发生而事件 B 不发生所构成的事件，称为事件 A 与事件 B 的差，记作 $A-B$.

例如：设事件 $A=\{0,1,2,\cdots,9\}$，$B=\{6,7,\cdots,15\}$，则事件 $A-B=\{0,1,2,3,4,5\}$，$B-A=\{10,11,12,13,14,15\}$.

5. 事件的积(交)

事件 A 与事件 B 同时发生所构成的事件，称为事件 A 与事件 B 的积(或交)，记作 AB 或 $A\cap B$.

类似地，事件 $A_1,A_2,\cdots,A_n$ 同时发生所构成的事件，称为事件 $A_1,A_2,\cdots,A_n$ 的积(或交)，记作

$$A_1A_2\cdots A_n,\ \text{或}\ \bigcap_{i=1}^{n} A_i,\ \text{或}\ A_1\cap A_2\cap\cdots\cap A_n.$$

同样,可列个事件 $A_1,A_2,\cdots,A_n,\cdots$ 同时发生所构成的事件,称为这可列个事件的积(或交),记作

$$A_1A_2\cdots A_n\cdots,\ \text{或}\ A_1\cap A_2\cap\cdots A_n\cap\cdots,\ \text{或}\ \bigcap_{i=1}^{\infty} A_i.$$

例如:在产品质量检验中,设事件 $A=\{$某厂生产的产品$\}$,$B=\{$某厂生产的一级品$\}$,则事件 $AB=\{$某厂生产的一级品$\}$.显然,这里 $AB=B$.

6. 互不相容事件(互斥事件)

事件 A 与事件 B 不能同时发生,即 $AB=\varnothing$,则称事件 A 与事件 B 为互不相容(或互斥)事件.

例如:事件{第 i 粒种子发芽}与事件{第 i 粒种子不发芽}是互不相容事件.

7. 互逆事件(对立事件)

如果事件 A 与事件 B 必发生其一,但又不能同时发生,即 $A\cup B=\Omega$,$AB=\varnothing$,则称事件 A 与事件 B 为**互逆事件**(或**对立事件**).

例如:在例 4 中,A 与 B 是互逆事件,因为 $A\cup B=\Omega$,$AB=\varnothing$.

事件 A 与 B 互逆,也常常说成 A 是 B 的逆事件,或 B 是 A 的逆事件.A 的逆事件还常用 $\bar{A}$ 表示,若 B 是 A 的逆事件,则 $\bar{A}=B$.

这里应该注意,互逆事件与互不相容事件之间的关系是既有联系,又有区别,即互逆事件一定是互不相容事件,而互不相容事件却未必是互逆事件.

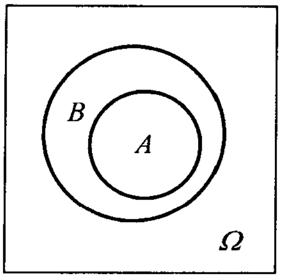
例如:在 3 粒种子的发芽试验中,设事件 $A=\{$恰有两粒种子发芽$\}$,$B=\{3$ 粒种子全发芽$\}$,则 $AB=\varnothing$,即 A 与 B 为互不相容事件,但不是互逆事件.因为 $A\cup B\neq\Omega$,即 A 与 B 的和事件不是必然事件.

8. 完备事件组

在试验 E 中,如果事件 $A_1,A_2,\cdots,A_n$ 必发生其一,且 $A_1,A_2,\cdots,A_n$ 两两互不相容,即 $A_1\cup A_2\cup\cdots\cup A_n=\Omega$,且 $A_iA_j=\varnothing(i,j=1,2,\cdots,n$,但 $i\neq j)$,则称事件组 $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$ 为**完备事件组**.

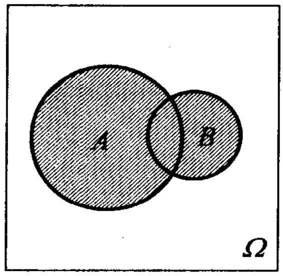
例如:向一目标连开两枪,观察击中目标的次数,设事件 $\omega_0=\{$击中 0 次$\}$,$\omega_1=\{$击中 1 次$\}$,$\omega_2=\{$击中 2 次$\}$,则 $\omega_0,\omega_1,\omega_2$ 是完备事件组.因为 $\omega_0\cup\omega_1\cup\omega_2=\Omega$,且 $\omega_i\omega_j=\varnothing(i\neq j;i,j=0,1,2)$.

事件之间的关系与运算可以用图形直观表示,这和集合论中集合的关系与运算的图示法相同.如图 1-1 表示 $A\subset B$,图 1-2 和图 1-4 的阴影部分分别表示和事件 $A\cup B$ 与差事件 $A-B$,图 1-3 和图 1-5 的阴影部分分别表示积事件 AB 与逆事件 $\bar{A}$,图 1-6 表示 A 与 B 为互不相容事件.除此之外,集合论中关于集合的关系与运算法则的若干规定对于概率论中事件的关系与运算也都是适用的.



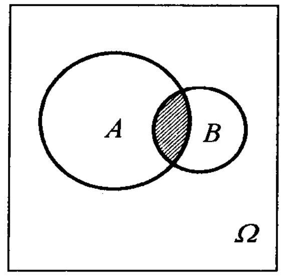
$A \subset B$

图 1-1



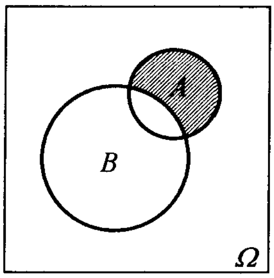
$A \cup B$

图 1-2



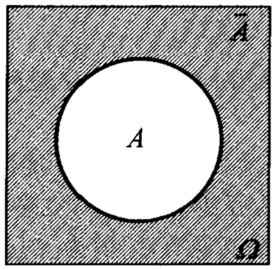
AB

图 1-3



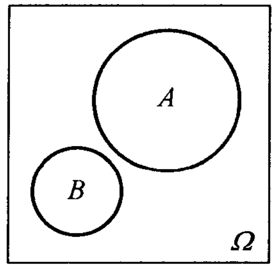
$A-B$

图 1-4



$\overline{A}$

图 1-5



$AB=\varnothing$

图 1-6

例 5　将三种耐旱性较强的半干旱地区作物引入干旱地区试验种植，观察它们的适应性，设事件 $A_i=\{$第 i 种作物适应旱区种植$\}$，$i=1,2,3$，试用 A_1，A_2，A_3 表示下列事件：(1) 只有第三种作物适应旱区种植；(2) 只有一种作物适应旱区种植；(3) 至少有一种作物适应旱区种植.

解　(1) 事件{只有第三种作物适应旱区种植}，意味着第一种作物、第二种作物都不适应旱地种植，所以它可表示为

$$\overline{A}_1\overline{A}_2A_3.$$

(2) 事件{只有一种作物适应旱区种植}，由于它未指明是哪一种作物适应旱区种植，所以可能是{只有第一种作物适应旱区种植}，或{只有第二种作物适应旱区种植}，或{只有第三种作物适应旱区种植}，故它可表示为

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

这里，$A_1\overline{A}_2\overline{A}_3$，$\overline{A}_1A_2\overline{A}_3$ 和 $\overline{A}_1\overline{A}_2A_3$ 是两两互不相容事件.

(3) 事件{至少有一种作物适应旱区种植}可表示为

$$A_1\cup A_2\cup A_3.$$

第二节　随机事件的概率

随机事件的概率(几率),简言之是指随机事件发生的可能性的大小,它可由介于0和1之间的某个数来表示.

概率论的基本任务,是通过随机试验来研究随机现象的统计规律性,其中一个最基本的问题就是给随机事件的概率一个科学的测度.

这一节将在三种不同的背景下,分别给出概率的古典定义、统计定义和公理化定义.

2.1　概率的古典定义

1. 古典概型

具有以下特征的随机试验模型,称为**古典概型**(概型是概率模型的简称).

(1) 试验的样本空间只有有限个样本点;

(2) 试验中每个样本点出现的可能性相同.

古典概型也叫做**等可能概型**.在概率论的产生和发展过程中,它是最早且最常用到的一种概率模型.

2. 概率的古典定义

定义1　设随机试验 E 为古典概型,它的样本空间 Ω 包含 n 个样本点,随机事件 A 包含 k 个样本点,则称比值 $\frac{k}{n}$ 为随机事件 A 的**概率**,记作 $P(A)$. 即

$$P(A)=\frac{k}{n}=\frac{A\text{ 中的样本点数}}{\Omega\text{ 中的样本点数}}. \tag{1.1}$$

定义1称为**概率的古典定义**.显然,概率的古典定义是以有限样本空间及其每个样本点出现的等可能性作为基础条件的.

例1　一次掷两枚均匀的硬币,观察出现正面与反面的情况,并分别求两枚都出现正面、一枚出现反面、两枚都出现反面的概率.

解　以 H 和 T 分别表示抛掷一枚硬币出现正面和出现反面,则掷两枚硬币的这个试验样本空间为 $\Omega=\{(H,H),(H,T),(T,H),(T,T)\}$,其中 (H,H) 表示两枚都出现正面,(H,T) 表示第1枚出现正面、第2枚出现反面,(T,H) 表示第1枚出现反面、第2枚出现正面,(T,T) 表示两枚都出现反面.

设事件 $A=\{$两枚都出现正面$\}$,$B=\{$只有一枚出现正面$\}$,$C=\{$两枚都出现反面$\}$,则由于 Ω 中包含有4个样本点,且这4个样本点出现的可能性相同,故试验属古典概型;又因事件 A 只包含1个样本点 (H,H),事件 B 包含2个样本点 (H,T) 和 (T,H),事件 C 包含1个样本点 (T,T),故由概率的古典定义得

$$P(A)=\frac{1}{4},\quad P(B)=\frac{2}{4}=\frac{1}{2},\quad P(C)=\frac{1}{4}.$$

例 2 某地区电话号码由 7 位数字组成,且每一位数字可以是 0,1,2,…,9 中的任意一个数.求由 7 个不同数字组成的电话号码的概率.

解 由于该地区的电话号码由 7 位数字组成,而每一位数可以是 0,1,2,…,9 中的任意一个数,所以每一位数不同数字的排法有 10 种,7 位数不同数字的排法共有 10^7 种;将 7 位数字的每一种排法看成 1 个样本点,则它的所有排法 10^7 种就是样本空间 Ω 包含的样本点总数:$n = 10^7$.

设 $B=\{$由 7 个不同数字组成的电话号码$\}$,由于由 7 个不同数字组成的电话号码有 A_{10}^7 个(A_{10}^7 表示从 10 个元素中任取 7 个元素的选排列.以后若遇到类似记号,则有相应含意),所以 B 包含的样本点数为 $k = A_{10}^7$,故由概率的古典定义得

$$P(B) = \frac{A_{10}^7}{10^7} = \frac{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4}{10^7} = 0.060\ 48,$$

即由 7 个不同数字组成的电话号码的概率为 0.060 48.

例 3 一箱中有 50 件产品,其中有 5 件次品.从箱中任取 10 件产品,求恰有两件次品的概率.

解 从 50 件产品中任取 10 件,有 C_{50}^{10} 种不同取法,将每一种取法看作 1 个样本点,则样本空间 Ω 共含 $n = C_{50}^{10}$ 个样本点.

设 $A=\{$取到的 10 件中恰有两件次品$\}$,则由于这两件次品是从 5 件次品中抽取的,其不同取法有 C_5^2 两种;其余 8 件正品是从 45 件正品中抽取的,其不同取法有 C_{45}^8 种.故取到的 10 件中恰有两件次品的不同取法共有 $C_5^2C_{45}^8$ 种,即 A 中包含的样本点数为 $k = C_5^2C_{45}^8$.于是,由概率的古典定义得

$$P(A) = \frac{C_5^2C_{45}^8}{C_{50}^{10}} = \frac{5!}{2!\ 3!}\cdot\frac{45!}{8!\ 37!}\Big/\frac{50!}{10!\ 40!} = 0.209\ 8$$

为所求的概率.

一般地,设有 N 件产品,其中有 M 件次品,若从 N 件产品中任意抽取 n 件,则抽到的 n 件中恰有 $m(m\leqslant M)$ 件次品的概率为

$$p = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}. \tag{1.2}$$

例 4 设某小麦研究中心随机地将 9 个小麦新品种平均分配到三个基地去种植,在这 9 个新品种中有 3 个是常规品种,6 个是杂交品种.问:(1) 每个基地各分到一个常规品种的概率是多少? (2) 3 个常规品种分配到同一基地的概率是多少?

解 将 9 个小麦新品种平均分到三个基地,其分法为:从 9 个新品种中任取

3 个分到第一个基地，分法为 C_9^3 种；从其余 6 个新品种中任取 3 个分到第二个基地，分法有 C_6^3 种；剩余的 3 个品种分到第三个基地，分法有 C_3^3 种. 故总的分法是 $C_9^3C_6^3C_3^3$ 种，即样本空间的总样本点数为

$$n = C_9^3C_6^3C_3^3 = \frac{9!}{(3!)^3}.$$

(1) 设 $A=\{$每个基地各分到一个常规品种$\}$. 由于将 3 个常规品种分配到三个基地，使每个基地都有一个常规品种，分法有 $3\cdot2\cdot1=3!$ 种；其余 6 个品种平均分到三个基地的分法有 $C_6^2C_4^2C_2^2=\frac{6!}{(2!)^3}$种. 因此每个基地各分到一个常规品种的分法有$\frac{3!\cdot6!}{(2!)^3}$种，即 A 包含 $k=\frac{3!\cdot6!}{(2!)^3}$个样本点. 故所求概率为

$$P(A)=\frac{\dfrac{3!\cdot6!}{(2!)^3}}{\dfrac{9!}{(3!)^3}}=\frac{9}{28}.$$

(2) 设 $B=\{3$ 个常规品种分配到同一基地$\}$. 由于将 3 个常规品种分配到同一基地的不同分法有 3 种，其余 6 个品种分到另外两个基地的不同分法有 $C_6^3\cdot C_3^3$ 种，因此 3 个常规品种分到同一基地的不同方法共有 $3C_6^3\cdot C_3^3=60$ 种，即 B 包含 $k=60$ 个样本点. 故所求概率为

$$P(B)=\frac{60}{\dfrac{9!}{(3!)^3}}=\frac{1}{28}.$$

由概率的古典定义，可以得到它的如下性质：

(1) 设 A 为E 中任一事件，则 $0\leqslant P(A)\leqslant1$；

(2) 对必然事件 Ω、不可能事件$\varnothing$，有 $P(\Omega)=1,P(\varnothing)=0$；

(3) 设 E 中事件$A_1,A_2,\cdots,A_m$ 两两互不相容，则

$$P(A_1\cup A_2\cup\cdots\cup A_m)=P(A_1)+P(A_2)+\cdots+P(A_m)=\sum_{i=1}^{m}P(A_i).$$

证　(1) 由于任一事件 A 包含的基本事件数k 满足 $0\leqslant k\leqslant n$，所以 $0\leqslant P(A)=\frac{k}{n}\leqslant1$，即

$$0\leqslant P(A)\leqslant1.$$

(2) 必然事件 Ω 由所有n 个基本事件组成，显然 Ω 包含的基本事件数$k=n$，从而 $P(\Omega)=\frac{n}{n}=1$，即

$$P(\Omega)=1.$$

不可能事件$\varnothing$不包含任何基本事件,即可认为$\varnothing$包含的基本事件数 $k=0$,从而 $P(\varnothing)=\frac{0}{n}=0$,即

$$P(\varnothing)=0.$$

(3) 设 A_i 含有 $k_i(k_i\leqslant n)$个基本事件,$i=1,2,\cdots,m$,则 $P(A_i)=\frac{k_i}{n}$,$i=1,2,\cdots,m$.

由于各 A_i 互不相容,所以 $A_1\cup A_2\cup\cdots\cup A_m$ 中含有 $k_1+k_2+\cdots+k_m$ 个不相同的基本事件.故

$$\begin{aligned}P(A_1\cup A_2\cup\cdots\cup A_m)&=\frac{k_1+k_2+\cdots+k_m}{n}\\&=\frac{k_1}{n}+\frac{k_2}{n}+\cdots+\frac{k_m}{n}\\&=P(A_1)+P(A_2)+\cdots+P(A_m),\end{aligned}$$

即 $P(A_1\cup A_2\cup\cdots\cup A_m)=P(A_1)+P(A_2)+\cdots+P(A_m)=\sum_{i=1}^{m}P(A_i)$.

2.2 概率的统计定义

在古典概型中,定义或计算随机事件的概率,是以试验中基本事件等可能发生作为一个重要条件.但对于一般的随机试验而言,这样的条件并不是都能具备.例如:在一批某品种的种子发芽试验中,所抽取到的各粒种子发芽或不发芽的可能性不一定相同,即定义 1 中基本事件等可能发生的条件不具备.在这样的情形下,如何确定这批种子的发芽率,或者说如何确定从该批种子中任取一粒发芽的概率?由此,我们可以联想到一个带有普遍性的问题,即在一般的随机试验中,如何确定随机事件的概率?为了解决这个问题,下面先引入事件的频率的概念.

频率 如果在 n 次重复试验中,事件 A 发生了 μ 次,则称比值$\frac{\mu}{n}$是事件 A 在这 n 次试验中发生的频率(简称$\frac{\mu}{n}$是事件 A 的频率),记作 $f_n(A)$,即

$$f_n(A)=\frac{\mu}{n}.$$

频率的概念不仅在科学研究中会用到,而且在平时的工作及日常生活中也会用到.例如,某大学学生去年参加全国英语四级考试的通过率为

$$\frac{\text{该校考生及格人数}}{\text{该校考生总人数}},$$

这个通过率就是该校考生及格的频率，它反映了该校大学生学习和掌握英语的整体水平．又如，检查某商店一批产品的质量，得到该批产品的合格率为

$$\frac{\text{抽查到的产品合格数}}{\text{抽查到的产品总数}},$$

这个合格率就是抽查到产品合格的频率，它反映了该商店所销售的这类产品的质量水平．

例 5 为考察一批某品种的小麦种子发芽情况，从该批种子中分别抽取 10 粒、50 粒、100 粒、300 粒、500 粒、800 粒，在相同条件下作发芽试验，观察结果及种子发芽频率列入表 1－1．

表 1－1

种子总粒数 n	10	50	100	300	500	800
发芽种子数 μ	10	43	92	271	452	721
种子发芽频率 $\frac{\mu}{n}$	1.000	0.860	0.920	0.903	0.904	0.901

这里，如果将对任意一粒种子发芽与否的观察看作一次试验，在这次试验中若种子发芽，则说事件 A 发生，显然 $A=\{$任取 1 粒种子发芽$\}$．对 n 粒种子发芽与否的观察看作 n 次重复试验，则由表 1－1 可以看出，对于不同的试验次数（即种子总粒数）n 来说，种子发芽数 μ 随着 n 的不同在变化，从而种子发芽的频率即 A 的频率 $\frac{\mu}{n}$ 也在变化．但是，我们注意到：随着试验次数 n 的增大，事件 A 的频率 $f_n(A)=\frac{\mu}{n}$ 仅在 0.9 附近有微小变化．这说明频率具有稳定性．习惯上称 0.9 是频率 $f_n(A)$ 的**稳定值**，或者说 0.9 表示任取 1 粒种子发芽的概率，即 $P(A)=0.9$，亦即 0.9 是这批种子的发芽率．

例 6 检查某工厂一批产品的质量，从中分别抽取 10 件、20 件、50 件、100 件、150 件、200 件、300 件检验，检查结果及次品频率列入表 1－2．

表 1－2

抽取产品总件数 n	10	20	50	100	150	200	300
次品数 μ	0	1	3	5	7	11	16
次品频率 $\frac{\mu}{n}$	0	0.050	0.060	0.050	0.047	0.055	0.053

由表 1－2 看出，在抽出的 n 件产品中，次品数 μ 随着 n 的不同而取不同

值,从而次品频率$\frac{\mu}{n}$仅在 0.05 附近有微小变化.所以 0.05 是次品频率的稳定值.如果设事件 A 表示从该批产品中任取 1 件是次品,即 $A=\{$任取 1 件是次品$\}$,则 A 的概率 $P(A)=0.05$ 就是这批产品的次品率.

于是,关于事件的概率有下面的定义 2.

定义 2　设在相同条件下重复进行同一试验,如果随着试验次数 n 的增大,事件 A 发生的频率 $f_n(A)=\frac{\mu}{n}$仅在某个数 p 附近有微小变化,则称数 p 是**事件 A 的概率**,即 $P(A)=p$.

定义 2 称为**概率的统计定义**.因为该定义中的数 p 习惯上叫做事件 A 的频率$\frac{\mu}{n}$的稳定值,所以概率的统计定义可以简单地说成:**事件 A 的频率稳定值叫做事件 A 的概率**.

事实上,概率的统计定义中的数 p 是不易精确得到的,因为我们没有理由说 $n+1$ 次重复试验中 A 的频率一定比 n 次试验中 A 的频率更接近于 p.但在**处理实际问题中,常常是用试验次数较多(即 n 较大)时的频率$\frac{\mu}{n}$作为 A 的概率 $P(A)$ 的近似值,即$\frac{\mu}{n}\approx P(A)$.**

在理解事件 A 的频率与其概率的概念时,应注意下面两方面的问题:一方面,事件 A 的频率与事件 A 的概率密切相关,这表现在 n 次重复试验中,当 A 的概率(出现的可能性)$P(A)$愈大时,它的频率 $f_n(A)$就愈大;当 A 的频率 $f_n(A)$愈大时,它的概率 $P(A)$也愈大.另一方面,频率具有稳定性,即 A 的频率仅在某个数 p 附近有微小变化,则 A 的频率的稳定值就是概率 $P(A)$.

例 7　从某鱼池中取 100 条鱼,做上记号后再放入该鱼池中.现从该池中任意捉来 40 条鱼,发现其中两条有记号,问池内大约有多少条鱼?

解　设池内有 n 条鱼,则从池中捉到一条有记号鱼的概率为$\frac{100}{n}$,它近似于捉到有记号鱼的频率$\frac{2}{40}$,即

$$\frac{100}{n}\approx\frac{2}{40},$$

解之,得 $n\approx2000$,故池内大约有 2000 条鱼.

概率的统计定义有类似于概率的古典定义的前述三条性质,即

(1) 设 A 为试验 E 中任一事件,则 $0\leqslant P(A)\leqslant1$;

(2) 对必然事件 Ω、不可能事件$\varnothing$,有 $P(\Omega)=1, P(\varnothing)=0$;

(3) 设试验 E 中的事件 $A_1,A_2,\cdots,A_m$ 两两互不相容,则