

# 数理统计

---

高等医药院校用

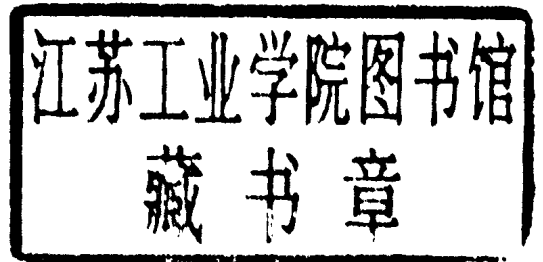
山东教育出版社

# 数 理 统 计

(高等医药院校用)

周怀梧 主编

一九八五年六月十日



山东教育出版社

一九八四年·济南

## 参加编写者

上海第一医学院	王 珍
南京药学院	倪永兴
上海中医学院	王家本 杨仁德
第二军医大学	薛祉绥
浙江医科大学	周怀梧 徐绍英 毛宗秀

## 数 理 统 计

(高等医药院校用)

周怀梧 主编

\*

山东教育出版社出版

(济南经九路胜利大街)

山东省新华书店发行 山东新华印刷厂德州厂印刷

\*

787×1092毫米16开本 17.75印张 4插页 412千字

1984年5月第1版 1984年5月第1次印刷

印数 1—4,000

书号 13275·20 定价 1.35元

# 前 言

数理统计在医药学中的应用十分普遍。随着我国医药卫生事业的发展，人们感到只学习一些医学（或卫生）统计方法已经不够了。当前，在高等医药院校中的药学（包括中药）专业已将数理统计列为必修课；部分院校也为医学专业学生和研究生开设了这门课；同时从事医学（或卫生）统计及医药研究的人员进修或自学概率论和数理统计的要求日益迫切。鉴于这些需要，我们编写了这本书。

本书除供高等医药院校数理统计课作教材用外，也可供广大医药学工作者自学或参考。讲授全部内容约需72学时，其中基本部分（不打\*号者）约需50学时。

在编写过程中，我们努力做到下述三点要求：

（1）根据医药、卫生专业教学和科研的实际需要，着重讲述有关概率论、实验设计、抽样及数理统计的基本概念和主要方法，以便使读者能正确理解和熟练运用概率、统计方法。

（2）紧密联系医药实际，适当反映概率、统计方法在医药研究中的应用的新进展，如：疾病的计量鉴别诊断、威布尔分布的应用、正交试验设计与分析、药物动力学数据处理、以及质量管理等，以便使读者尽快地掌握这些新方法。

（3）有较强的系统性和逻辑性，各章配备适量的习题并附答案，既便于教，又便于学。

本书承山东医学院张馥、郭怀兰、虞孝珍等老师审阅并提供宝贵意见，浙江医科大学陆琦、景荣荣等老师协助验做了大部分习题，在此一并致谢。对本书存有错误和缺点，恳请读者指正。

编 者

1983年12月

# 目 录

绪 论	1
第一章 随机事件与概率	3
第一节 事件及其运算	3
第二节 频率与概率	6
第三节 古典概型	7
第四节 概率的基本运算法则	8
*第五节 全概率公式	12
*第六节 贝叶斯公式与鉴别诊断	13
第七节 独立重复试验概型	16
习题一	17
第二章 随机变量及其分布	20
第一节 离散型随机变量及其分布	20
第二节 连续型随机变量及其分布	23
第三节 随机变量的数字特征	26
第四节 二项分布和泊松分布	33
第五节 正态分布和对数正态分布	39
*第六节 威布尔分布	44
第七节 大数定律和中心极限定理	46
习题二	49
第三章 实验设计、抽样及抽样分布	52
第一节 实验设计概论	52
第二节 抽样的基本概念和方法	54
第三节 样本的数字特征	56
第四节 样本分布图	59
*第五节 概率纸及其应用	62
第六节 $\chi^2$ 分布、 $t$ 分布和 $F$ 分布	68
第七节 样本均数和样本方差的分	71
习题三	72
第四章 抽样估计	75
第一节 抽样估计的概念	75
第二节 最大似然估计法	78
第三节 总体均数的区间估计	81

*第四节	正态总体方差的区间估计	84
第五节	总体率的区间估计	85
*第六节	泊松分布参数 $\lambda$ 的区间估计	88
习题四		91
<b>第五章</b>	<b>假设检验</b>	93
第一节	基本原理与一般步骤	93
第二节	单个总体均数的假设检验	94
第三节	配对比较和成组比较	99
第四节	两个正态总体方差的比较	105
第五节	总体率的假设检验	107
*第六节	序贯概率比检验	111
第七节	非参数检验	116
*第八节	参照单位 (Ridit) 分析与应用	120
第九节	适合性检验和独立性检验	124
习题五		131
<b>第六章</b>	<b>方差分析</b>	135
第一节	基本思想与一般步骤	135
第二节	单因素试验的方差分析	137
第三节	多组均数间的两两比较	142
*第四节	方差分析中的数据变换	144
习题六		146
<b>第七章</b>	<b>正交试验设计与分析</b>	148
第一节	正交试验的基本思想	148
第二节	正交试验计划的拟订	151
第三节	正交试验的直观分析	152
第四节	考虑交互作用的试验分析	159
*第五节	正交试验的方差分析	162
习题七		165
<b>第八章</b>	<b>相关与回归</b>	167
第一节	相关与回归的概念	167
第二节	相关系数	168
第三节	一元线性回归	172
第四节	一元非线性回归	179
*第五节	药物动力学分析中的剩余法	182
*第六节	计算半数致死量的概率单位法	185
习题八		191
<b>*第九章</b>	<b>质量管理常用统计方法</b>	193
第一节	质量管理的概念	193

第二节	排列图法和因果分析图法	194
第三节	直方图法	197
第四节	分层法	201
第五节	控制图法	201
第六节	抽样检查	215
	习题九	221
	习题答案	225
	附表	229
1.	泊松(Poisson)分布表	229
2.	正态分布的密度函数表	234
3.	正态分布表	235
4.	正态分布的双侧分位数(Na)表	237
5.	随机数表	238
6.	$t$ 分布的双侧分位数表	240
7.	$\chi^2$ 分布的上侧分位数表	241
8.	二项分布参数 $P$ 的置信区间表	242
9.	$\varphi = 2\arcsin\sqrt{p}$ 数值表	246
10.	泊松分布参数 $\lambda$ 的置信区间表	248
11.	$F$ 检验的临界值表	249
12.	符号检验表	254
13.	秩和检验表	254
14.	游程总数检验表	255
15.	多重比较中的 $q$ 表	256
16.	多重比较中的 $S$ 表	257
17.	正交表	258
18.	检验相关系数 $\rho = 0$ 的临界值表	266
19.	百分率与概率单位对照表	267
20.	概率单位与权重系数对照表	267
21.	作业用概率单位之极小值、极大值及全距	268
22.	样本容量字码	269
	索引	271
	参考文献	276

# 绪 论

## 一、数理统计的研究对象和内容

客观世界中出现的种种现象，大体上可分为两类：一类是在一定条件下必然发生或决不可能发生的，称为确定性现象。例如，在标准大气压下把纯水加热到  $100^{\circ}\text{C}$  时必然沸腾；人坐着的高度决不可能超过身高，等等。另一类是在一定条件下可能发生也可能不发生的，或者说可能出现这个结果也可能出现那个结果的，称为随机现象。例如，观察一种新药治疗某种疾病的疗效，对一个病人来说，可能有效也可能无效；在射击训练中，靶子上的弹着点可能出现在这里，也可能出现在那里，等等。

实践表明，对同一个随机现象作大量次数的观察，常可发现某种确定的数量规律性，即所谓“统计规律性”。例如，为评价一种新药的疗效，通过足够多个病例的试用和观察，可以对其有效率作出客观的估计。又如，对某个靶子连续射击，当射击次数不多时，弹着点的分布杂乱无章，没有什么明显的规律性。但射击次数足够多时，可发现靶子上的弹着点对某个中心近似于对称分布，越靠近中心越密，越离开中心越稀，并且这种疏密的变化服从确定的规律即所谓正态分布律（详见第二章第五节）。数理统计（*Mathematical statistics*）同概率论一样，都是研究随机现象的数量规律性的数学分支。

数理统计以概率论为其理论基础，它应用概率论的结果深入地研究试验的设计及进行统计资料的分析。虽然从理论上说，只要对随机现象进行大量次数的观察，就能发现一定的统计规律性，但是实际所允许的观察往往是有限的，甚至是少量的。因此，必需研究怎样合理地安排试验和观察，以便获得完整的、正确的、足够多的资料；还必需研究怎样有效地利用所获得的资料揭示随机现象内在的规律性，对实践中提出的问题作出尽可能精确、可靠的结论。这些就是数理统计所研究的基本内容，前者称为试验设计，后者称为统计推断。

## 二、数理统计的发展简史

数理统计同其他自然科学一样，是应社会的需要而产生，随生产和科学技术的发展而发展的。早在十七世纪中叶，人们便开始研究随机现象的数量规律性，典型的课题是赌博中的机遇问题。由于赌博为随机现象提供最简单、直观、又便于研究的模型，所以至今在概率论中仍广泛采用掷钱币、掷骰子、打扑克、摸球等例子。在这些研究中建立了概率、随机变量、数学期望等基本概念。之后，在解决有关人口统计、生命保险、测量误差、射击理论等实际问题中，数理统计的理论和方法获得了迅速的发展。皮尔逊（*Karl Pearson*, 1857—1936）将数理统计应用于生物科学研究，提出了著名的  $\chi^2$  分布，创办了“生物统计”（*Biometrika*）杂志及数理统计学校，为数理统计的发展作出了卓越的贡献。他的学生哥赛特（*W. Gosset*, 1876—1937）以“学生”（*Student*）



的笔名提出了著名的 $t$ 检验。费歇 (*R·A·Fisher*, 1890—1962) 受到皮尔逊和哥赛特的影响, 和他的学生为数理统计方法在许多学科中的应用, 尤其是在农业科学、生物学和遗传学中的应用作出了重要的贡献, 他于 1925 年所著的《研究工作者用统计方法》(*Statistical Method for Research Workers*) 一书, 标志着数理统计作为一门独立学科的开端。

由于近代生物学、遗传学、工业产品质量控制、农业试验、计算机科学技术发展的推动, 数理统计更加迅速地发展, 出现了一些新分支, 如质量控制、试验设计、序贯分析、非参数统计推断理论及多元分析等。数理统计的应用范围大大地扩张了, 已经成了研究自然现象和社会现象不可缺少的工具。

### 三、数理统计在医药工作中的应用

数理统计作为认识随机现象数量规律性的有效工具, 在医药工作中已经被经常的广泛的应用。现在, 无论在基础医学、临床医学和预防医学各个方面的科学研究中, 无论在疾病防治工作计划的拟订和效果的正确评价、以及在药物生产中处方的选择、工艺的改进、质量的控制等方面, 都必须合理地设计试验 (包括调查、实验、观察), 正确地收集和整理资料, 并进行恰当的统计分析。

医药学的研究对象主要是人 (健康人和病人), 而人的个体差异相当大。比如, 同是二十岁健康的男青年, 他们的身长、体重、血压、脉搏、白细胞等数值往往不同, 而且可能相差悬殊。又如, 让同一种病人口服一定剂量的某种药物, 可能有的有效, 有的无效。既然如此, 我们就不可仅凭对个别人的观察结果, 贸然作出一般性的结论, 而应通过多次的观察去揭示统计规律性。这是数理统计在医药工作中被广泛应用的基本原因之所在。

随着现代医药科学技术的进展, 尤其是电子计算机在医药研究中的推广使用, 使得数理统计在医药工作中的应用更加广泛、更加深刻。因此每一个医药学工作者都应当不同程度地掌握数理统计的基本原理和方法。

# 第一章 随机事件与概率

## 第一节 事件及其运算

### 一、随机试验

为了研究随机现象的规律性，需要进行足够多次的试验、实验、调查或观测，我们把这些工作统称为试验。概率论中所说的试验是指随机试验 (*Random trials*)，它具有下列三个特性：

1. 可在相同的条件下重复进行；
2. 有多个（两个及以上）可能的结果；
3. 每次试验之前，不能肯定将会出现哪个结果。

**例 1** 抛一枚硬币，观察字面、徽面出现的情况。

**例 2** 临床上试验一种复方抗结核片对肺结核的疗效——治愈、有效或无效。

**例 3** 调查某地正常成年男子尿汞的含量。

上面列举的都是随机试验的例子。应该注意，医药学研究中的试验，尤其是临床试验，影响试验结果的因素往往很多，而且不易控制。换句话说，要使试验在相同的条件下重复进行，有时不易办到。因此，要特别重视试验设计（见第三章），否则尽管表面上运用概率统计方法没有错误，但作出的结论则不一定可靠。

### 二、随机事件

随机试验的结果称为随机事件 (*Random events*)，简称事件。由于在一次试验中，可能出现这个结果、也可能出现那个结果，所以，指定的一个随机事件便可能发生，也可能不发生。对于例 1，若用  $A$ 、 $B$  分别表示“出现字面”和“出现徽面”，则  $A$ 、 $B$  都是随机事件。对于例 3，用  $X$  表示尿汞的含量（微克/升），若记  $E$  为“ $X = 1.8$ ”， $F$  为“ $X \leq 3.5$ ”， $G$  为“ $X > 3.5$ ”……则它们都是随机事件。

如果在每次试验中，某个结果必定出现，或者必定不出现，则分别称为必然事件和不可能事件。为讨论方便起见，把它们也当作随机事件，并把必然事件特记为  $U$ ，把不可能事件特记为  $V$ 。例如，在例 1 中，“出现字面或徽面”是必然事件  $U$ ，而“既不出现字面也不出现徽面”是不可能事件  $V$ 。

### 三、事件间的关系和运算

同一随机现象下的各个随机事件，有的简单，有的比较复杂。例如，从野外移栽三株中草药——“一枝黄花”到花盆里，其成活情况是一个随机现象，“活二株”是一个简单的事件，而“至少活一株”则是较为复杂的事件。为了研究各种事件发生的可能性大小，需要讨论事件之间的关系。

1. 事件的包含与相等。

设有事件 $A$ 与 $B$ ，若事件 $A$ 发生，则事件 $B$ 必发生，则称事件 $B$ 包含事件 $A$ ，记作 $A \subset B$ 或 $B \supset A$ 。对于上述例子，用 $A$ 表示“活二株”，用 $B$ 表示“至少活一株”，显然有 $B \supset A$ 。

若事件 $A$ 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(或等价)，记作 $A = B$ 。

### 2. 事件的和与差。

设有事件 $A$ 与 $B$ ，若两者中至少有一个发生，这样所组成的复杂事件“ $A$ 或 $B$ ”称为事件 $A$ 与 $B$ 的和(或并)，记作 $A + B$ 或 $A \cup B$ 。

若事件 $A$ 发生而事件 $B$ 不发生，这样所组成的事件称为事件 $A$ 与 $B$ 的差，记作 $A - B$ 。

**例4** 抛两枚硬币，用 $A$ 表示事件“正好一个字面朝上”， $B$ 表示事件“正好两个字面朝上”， $C$ 表示事件“至少一个字面朝上”，则事件 $C$ 为事件 $A$ 与 $B$ 的和事件，即 $C = A + B$ ；而事件 $A$ 为事件 $C$ 与 $B$ 的差事件，即 $A = C - B$ 。

### 3. 事件的积与互不相容性。

设有事件 $A$ 与 $B$ ，若 $A$ 与 $B$ 同时发生，这样所组成的事件“ $A$ 且 $B$ ”称为事件 $A$ 与 $B$ 的积(或交)，记作 $AB$ 或 $A \cap B$ 。

对于例4，我们有： $AB = V$ (不可能事件)； $AC = A$ 。

若事件 $A$ 与 $B$ 不可能同时发生，即 $AB = V$ ，则称 $A$ 与 $B$ 是互不相容(或互斥)事件。在例4中， $A$ 与 $B$ 是互不相容事件。

### 4. 对立事件。

设有事件 $A$ 与 $B$ ，若两者中必有一个发生且仅有一个发生，即同时满足条件

$$A + B = U \text{ 及 } AB = V$$

则称 $A$ 与 $B$ 为互相对立(或互逆)事件。事件 $A$ 的对立事件记作 $\bar{A}$ 。则上式常被写成

$$A + \bar{A} = U \text{ 及 } A\bar{A} = V$$

这里我们可以看出，两个互逆事件一定是互斥事件，但两个互斥事件不一定是互逆事件。因为互斥事件只具备“两个事件中只能发生一个”的性质，而不一定具备“两个事件中必然发生一个”的性质。例如从工厂的产品中抽取两件，则“其中有一件废品”与“两件都是废品”这两个事件是互斥事件，但不是互逆事件，因为很可能抽取的两件产品都是合格品。

对于例4， $\bar{C}$ 表示事件“没有一个字面朝上”，亦即“正好两个字面朝上”。

为直观起见，设事件 $U$ 是随机点落在某正方形内，事件 $A$ 是随机点落在其中的一个小圆内，事件 $B$ 是落在一个大圆内，则事件 $U$ 、 $A$ 、 $\bar{A}$ 、 $A + B$ 、 $A - B$ 、 $AB$ 便可分别以阴影表示出(如图1·1)。容易想到， $B \supset A$ 相应于 $A$ 圆完全包含在 $B$ 圆内； $A$ 与 $B$ 互不相容，则相应于两圆互不相交。这种表示法称为文(Venn)图。

最后指出，事件的和与积的概念可推广到有限个或可列个事件注的情形。对于 $n$ 个事件 $A_1, A_2, \dots, A_n$ ，用 $A_1 + A_2 + \dots + A_n$ 或 $\sum_{i=1}^n A_i$ 表示事件 $A_1, A_2, \dots, A_n$ 中至少发生

注：可按自然数顺序编号的无穷多个事件序列称为可列个事件。

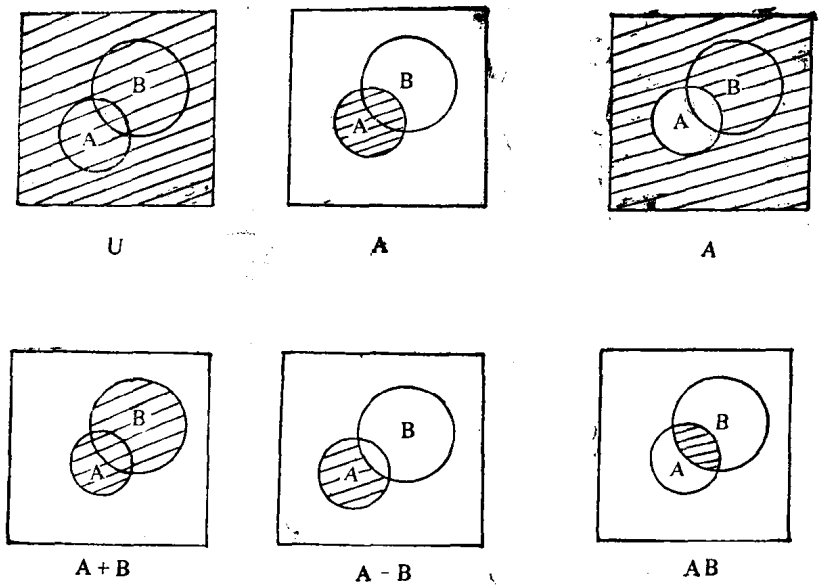


图1.1

一个，称为这 $n$ 个事件的和，而用  $A_1A_2\cdots A_n$  或  $\prod_{i=1}^n A_i$  表示事件  $A_1, A_2, \dots, A_n$  同时发生，称为这 $n$ 个事件的积。

**例 5** 有个电子系统由 $n$ 个元件  $e_1, e_2, \dots, e_n$  并联或串联而成（如图 1.2）。

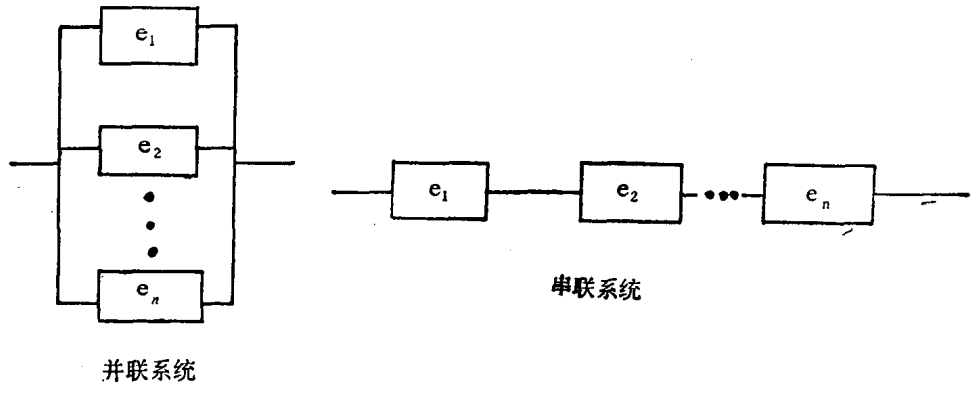


图1.2

今用  $E_i$  表示事件“元件  $e_i$  工作可靠” ( $i = 1, 2, \dots, n$ )，用  $S$  表示事件“系统工作可靠”，那么，对于并联系统，有

$$S = \text{“}e_1, e_2, \dots, e_n \text{中至少有一个元件工作可靠”} = \sum_{i=1}^n E_i$$

对于串联系统，则有

$$S = \text{“}e_1, e_2, \dots, e_n \text{全部元件工作可靠”} = \prod_{i=1}^n E_i$$

## 第二节 频率与概率

### 一、频率的概念

**定义** 设在 $n$ 次试验中, 事件 $A$ 发生 $m$ 次, 比值

$$f_n(A) = \frac{m}{n} = \frac{A \text{ 发生的试验次数}}{\text{试验的总次数}} \quad (1.1)$$

称为事件 $A$ 在这 $n$ 次试验中出现的频率 (Frequency), 称 $m$ 为频数。

医药工作中通常所说的发病率、病死率、治愈率等都是频率, 常用百分数表示。显然, 频率具有下列性质:

$$0 \leq f_n(A) \leq 1$$

**例 1** 为调查某地区居民患结核病的情况, 在该地抽查了400人, 发现有6人患结核病, 所以, 该地区居民患结核病的频率为

$$f = \frac{6}{400} \times 100\% = 1.5\%$$

### 二、频率的稳定性

**例 2** 今有甲、乙、丙、丁四人分别作投掷硬币的试验, 结果见表 1.1。

表1.1

试 验 者	甲	乙	丙	丁
试验次数 ( $n$ )	1500	2800	4800	8500
出现字面的次数 ( $m$ )	739	1405	2395	4252
出现字面的频率 ( $f$ )	0.4927	0.5018	0.4990	0.5002

每次掷币结果可能出现字面也可能出现徽面。由表 1.1 可见, 投掷次数愈多, 出现字面的频率愈来愈接近0.5。

**例 3** 打开任意一本英文书, 查阅某行的某字母, 它可能是26个字母中的一个, 也可能是“空格”(或标点)。经过大量统计, 发现各个字母被使用的频率相当稳定, 其值见表 1.2。

字母使用频率的研究对于打字机键盘的设计, 印刷铅字的铸造, 信息的编码及密码的破译等方面都十分有用。

实践表明, 在重复试验中, 事件 $A$ 出现的频率, 随着试验次数的逐渐增多, 总的趋势是越来越接近一个确定的常数, 明显地呈现出稳定于该常数的特性, 这就是通常所说的频率的稳定性。

### 三、概 率

频率的稳定性充分说明随机事件出现的可能性是事物本身固有的一种客观属性, 因此可以对它进行度量。

**定义** 在大量重复试验中, 如果事件 $A$ 出现的频率稳定地在某一常数 $p$ 的附近摆动,

表1.2

字母	空格(或标点)	E	T	O	A	N	I	R
频率	0.2	0.105	0.072	0.0654	0.063	0.059	0.055	0.054
字母	S	H	D	L	C	F	U	M
频率	0.052	0.047	0.035	0.029	0.023	0.0225	0.0225	0.021
字母	P	Y	W	G	B	V	K	X
频率	0.0175	0.012	0.012	0.011	0.0105	0.008	0.003	0.002
字母	J	Q	Z					
频率	0.001	0.001	0.001					

便称常数 $p$ 为事件 $A$ 的概率 (Probability), 记作 $P(A) = p$ .

这一定义通常称为概率的统计定义. 容易看出, 频率一般为变数, 概率则为常数; 当试验次数足够多, 频率相当稳定时, 便可把频率作为概率的近似估计, 即

$$P(A) \approx f_n(A) \quad (1.2)$$

由于频率总是介于0和1之间, 因而根据概率的定义可知概率有下列性质:

(1) 对于任何事件 $A$ , 有

$$0 \leq P(A) \leq 1 \quad (1.3)$$

(2) 对于必然事件 $U$ , 有

$$P(U) = 1 \quad (1.4)$$

(3) 对于不可能事件 $V$ , 有

$$P(V) = 0 \quad (1.5)$$

如某事件的概率几乎接近于0, 说明这个事件即使在大量次数的试验中出现的次数也非常少, 由此推知这种小概率事件虽不是不可能事件, 但在一次试验中可认为实际上是不会发生的. 由此引伸出所谓“小概率原理”, 它是统计假设检验中的一条重要原理 (见第五章).

### 第三节 古典概型

按概率的统计定义, 为了确定一个随机事件的概率, 就得进行大量重复试验. 可是, 在一些情况下依据事物本身所具有的某种“对称性”, 就可直接计算某事件的概率.

**例1** 掷一颗骰子, 出现3点或4点的概率是多少?

**解:** 由于骰子是一个正六面体, 各点出现的可能性一样, 可认为都是 $\frac{1}{6}$ , 所以, 在一次投掷中, 出现3点或4点的概率为 $\frac{2}{6}$ .

**例 2** 在 1、2、3、4、5 这五个数字中任取两个，取得的两数之和为偶数的概率是多少？

**解：**从五个数字中任取两个，共有  $C_5^2 = 10$  种不同的取法，它们是 (1, 2)，(1, 3)，(1, 4)，(1, 5)，(2, 3)，(2, 4)，(2, 5)，(3, 4)，(3, 5)，(4, 5)。既然是“任取”，那么这 10 种结果出现的机会都一样，而事件“两数之和为偶数”（记为 A）的出现，相当于 (1, 3)，(1, 5)，(2, 4)，(3, 5) 这四个结果之一出现，于是所求的概率为

$$P(A) = \frac{4}{10} = 0.4$$

这类简单的随机现象具有下列三个特征：

(1) 可能结果只有有限个，记为  $A_1, A_2, \dots, A_N$ ，它们出现的机会相等（等可能性）；

(2) 在任一次试验中， $A_1, A_2, \dots, A_N$  至少有一个出现（完备性）；

(3) 在任一次试验中， $A_1, A_2, \dots, A_N$  至多有一个出现（互不相容性）。

一般把具有上述特征的随机现象的数学模型称为古典概率模型，简称古典概型。因为这类随机现象在概率论发展的初期是主要的研究对象。

我们把具有上述特征的事件组  $A_1, A_2, \dots, A_N$  称为基本事件组，其中任一事件  $A_i$  ( $i = 1, 2, \dots, N$ ) 称为基本事件。如果试验时某一基本事件的出现导致事件 A 的出现，则称该基本事件是有利于事件 A 的。例如，在例 2 里 10 种不同的结果就构成基本事件组，有利于事件 A 的是其中四个基本事件，即 (1, 3)，(1, 5)，(2, 4)，(3, 5) 这四个结果。

**定义** 设试验的全部可能结果可表为由  $N$  个互不相容且等可能的事件构成的基本事件组，其中  $M$  个基本事件有利于事件 A，则事件 A 的概率为

$$P(A) = \frac{M}{N} = \frac{\text{有利于 A 的基本事件数}}{\text{基本事件的总数}} \quad (1.6)$$

这就是概率的古典定义。

**例 3** 瓶中装有 30 片药，其中有 6 片已失效，今从瓶中任取 5 片，求其中有 2 片失效的概率。

**解：**记 A 为任取 5 片中有 2 片失效这一事件，按题意可知

基本事件总数  $N = C_{30}^5 = 142506$

有利于 A 的基本事件数  $M = C_6^2 \cdot C_{24}^3 = 30360$

$$\therefore P(A) = \frac{C_6^2 \cdot C_{24}^3}{C_{30}^5} = \frac{30360}{142506} = 0.2130$$

## 第四节 概率的基本运算法则

### 一、概率加法定理

**定理 1** 两个互不相容事件 A 与 B 之和的概率，等于这二事件的概率之和，即

$$P(A+B) = P(A) + P(B) \quad (1.7)$$

**证:** 就概率的古典定义来作出证明。设试验的可能结果是  $N$  个基本事件构成的事件组，其中  $M_1$  个有利于事件  $A$ ， $M_2$  个有利于事件  $B$ 。由于  $A$  与  $B$  互不相容，因而有利于  $A$  的基本事件与有利于  $B$  的应该完全不相同，所以，有利于  $A+B$  的基本事件共有  $M_1+M_2$  个。于是，有

$$P(A+B) = \frac{M_1+M_2}{N} = \frac{M_1}{N} + \frac{M_2}{N} = P(A) + P(B)$$

这一定理称为概率加法定理，不难推广到有限多个事件的情形。设事件  $A_1, A_2, \dots, A_n$  两两互不相容，则有

$$P(A_1 + A_2 + \dots + A_n) = \sum_{i=1}^n P(A_i) \quad (1.8)$$

利用概率加法定理可推出以下定理：

**定理 2** 设  $\bar{A}$  是  $A$  的对立事件，则有

$$P(A) = 1 - P(\bar{A}) \quad (1.9)$$

**证:** 因  $A + \bar{A} = U$ ， $A \bar{A} = V$ ，故由 (1.4) 式有

$$P(A + \bar{A}) = P(U) = 1$$

由定理 1 又有

$$P(A + \bar{A}) = P(A) + P(\bar{A})$$

于是有

$$P(A) + P(\bar{A}) = 1$$

从而 (1.9) 式得证。

**定理 3** 设有事件  $A$  与  $B$ ，其中  $A \supset B$ ，则有

$$P(A - B) = P(A) - P(B) \quad (1.10)$$

**证:** 因  $A \supset B$ ，故等式  $A = (A - B) + B$  成立，又由于  $(A - B)$  与  $B$  互不相容，所以由定理 1 有

$$P(A) = P(A - B) + P(B)$$

从而 (1.10) 式得证。

**定理 4** 设  $A$ 、 $B$  为任意两个事件，则有

$$P(A + B) = P(A) + P(B) - P(AB) \quad (1.11)$$

**证:** 由  $A + B = A + (B - AB)$ ， $B \supset AB$ ，且事件  $A$  与  $(B - AB)$  互不相容，所以由定理 1 和 3 有

$$P(A + B) = P(A) + P(B - AB) = P(A) + P(B) - P(AB)$$

**例 1** 设 50 支针剂中有 3 支不合格，今从中任取 4 支，求其中有不合格品的概率。

**解:** 取出的 4 支针剂中有不合格品这一事件记为  $A$ ，有 1 支、2 支、3 支不合格品的事件分别记为  $A_1$ 、 $A_2$ 、 $A_3$ 。显然， $A_1$ 、 $A_2$ 、 $A_3$  互不相容，且  $A = A_1 + A_2 + A_3$ 。首先，由概率的古典定义，有

$$P(A_1) = \frac{C_{47}^3 \cdot C_3^1}{C_{50}^4} = 0.2112,$$



$$P(A_2) = \frac{C_{47}^2 \cdot C_3^2}{C_{50}^4} = 0.0141,$$

$$P(A_3) = \frac{C_{47}^1 \cdot C_3^3}{C_{50}^4} = 0.0002$$

然后, 根据概率加法定理, 得

$$P(A) = P(A_1) + P(A_2) + P(A_3) = 0.2255$$

本题也可利用定理 2 来求解。显然,  $A$  的对立事件  $\bar{A}$  表示取出的 4 支针剂中没有不合格品 (即全部为合格品), 其概率为

$$P(\bar{A}) = \frac{C_{47}^4}{C_{50}^4} = 0.7745$$

从而

$$P(A) = 1 - P(\bar{A}) = 1 - 0.7745 = 0.2255$$

## 二、条件概率和事件的独立性

先考虑一个例子。设一袋中有质量和大小都相同的红球 6 只, 白球 4 只, 从中任意摸取二次, 每次摸出一球, 不再放回。摸出红球记为  $A$ , 摸出白球记为  $B$ 。第一次摸出的可能是红球 (事件  $A$ ), 也可能是白球 (事件  $B$ )。显然,  $P(A) = \frac{6}{10}$ ,  $P(B) = \frac{4}{10}$ 。

但在第一次摸出红球的条件下, 第二次摸出白球的概率便不再是  $\frac{4}{10}$ , 而是  $\frac{4}{9}$  了。

**定义** 如果在事件  $A$  已发生的条件下计算事件  $B$  的概率, 则这种概率称为事件  $B$  在事件  $A$  已发生的条件下的条件概率 (Conditional probability), 记作  $P(B|A)$ 。

在上述例子里,  $P(B|A) = \frac{4}{9}$ , 这说明事件  $A$  发生与否对事件  $B$  的概率是有影响的, 这时我们说事件  $B$  对事件  $A$  不是独立的。倘若第一次摸出的红球仍然放回、混匀后再作第二次摸球, 则  $P(BA)$  就等于  $P(B) = \frac{4}{10}$  了; 在这个试验中, 我们说事件  $B$  对事件  $A$  是独立的。

**定义** 若事件  $B$  的概率  $P(B)$  等于在事件  $A$  已发生的条件下的条件概率  $P(B|A)$ , 即

$$P(B) = P(B|A) \quad (1 \cdot 12)$$

则称事件  $B$  对事件  $A$  是独立的 (Independent)。

## 三、概率乘法定理

**定理 5** 设  $A$ 、 $B$  为两个随机事件, 其概率  $P(A)$  和  $P(B)$  不等于 0, 则

$$P(AB) = P(A)P(B|A) = P(B)P(A|B) \quad (1 \cdot 13)$$

**证:** 就概率的古典定义来作出证明。设试验的可能结果是  $N$  个基本事件构成的事件组, 其中有利于事件  $A$  的有  $M_1$  个, 有利于事件  $B$  的有  $M_2$  个, 而有利于事件  $AB$  的有  $M$  个 (显然,  $M \leq M_1$ ,  $M \leq M_2$ ), 如果已知事件  $A$  发生, 则表明有利于  $A$  的  $M_1$  个基本事件中必有一个发生。这时, 有利于事件  $B$  的基本事件有而且仅有  $M$  个, 即有利于事件  $AB$  的那些基本事件。于是, 我们有