

21

世纪

高等医药院校教材

张春华
严云良 主编

医药 数理统计



科学出版社

21 世纪高等医药院校教材

医药数理统计

张春华 严云良 主编

科学出版社

2001

内 容 简 介

本教材是根据卫生部 1982 年以来颁布的有关数学教学计划的要求、1998 年 10 月教育部高等教育司举办的“数学教育在大学教育中作用研讨会”的会议精神、以及目前教学改革的新情况,由全国 18 所中医药院校长期从事数学教学和科研工作的教师编写的数学系列教材(医药高等数学、医药数理统计、医药数学实验和线性代数)之一。全书分 10 章,内容主要包括数理统计所需的概率论基本知识、统计的几个重要概念及分布、医药学中常用的统计方法及正交试验设计和均匀分布试验设计。重点介绍方法和在医药学中的应用,而不是数学的推导证明。

本书可供医药院校各专业各层次的学生使用,也可作为医药工作者做统计处理的参考书。

图书在版编目(CIP)数据

医药数理统计/张春华,严云良主编.-北京:科学出版社,

2001.6

21 世纪高等医药院校教材

ISBN 7-03-009227-9

I. 医… II. ①张… ②严… III. 数理统计-应用-医药学-医学院校-教材 IV. R311

中国版本图书馆 CIP 数据核字(2001)第 07044 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001 年 6 月第 一 版 开本: 850×1168 1/16

2001 年 6 月第一次印刷 印张: 18

印数:1—12 000 字数: 363 000

定 价:22.50 元

(如有印装质量问题,我社负责调换〈新欣〉)

《医药数理统计》编写人员

主 编：张春华 严云良

副主编：王建生 别雪君 马志庆 于鹤丹 何 雁
封 峰 李秀昌 刘明芝

主 审：周永治

编 委：(按姓氏笔画为序)

马光菊	王世钦	王淑媛	邓 超	闫雪隐
李和伟	汪旭升	苏 越	陈 勇	范 颖
周介南	周 喆	金善玉	赵文峰	高敏艳
钱微微	崔红新	崔相学	黄 浩	

编写说明

数理统计方法是研究随机现象统计规律的一门学科,它所建立的方法广泛应用于自然科学,社会科学和工农业生产,特别在医药学中它已成为一种不可缺少的工具,从而使医药数理统计成为中医药专业的一门重要基础课。

本教材是根据卫生部 1982 年以来颁布的有关数学教学计划的要求、1998 年 10 月教高司举办的“数学教育在大学教育中作用研讨会”的会议精神以及目前教学改革的新情况,由全国 18 所中医院校长期从事数学教学和科研的教师编写的数学系列教材(医药高等数学、医药数理统计、医药数学实验和线性代数)之一。

全书共分 10 章,内容主要包括数理统计所需的概率论基本知识、统计的几个重要概念及分布、医药学中常用的统计方法及正交试验设计和均匀分布试验设计。重点介绍方法和在医药学中的应用,而不是数学的推导证明。

本书可作为高等医药院校的“医药数理统计”教材,也可作为医药人员自学统计方法的用书,对缺乏微积分知识的读者,可略去有关数学证明而直接使用结论,并可通过示范性的例子掌握方法的应用。

本教材约需 60 学时,如删去(*)号的内容及概率论知识,40~50 学时也能讲授。书后配有习题答案,供教师和读者参考选用。

参加本教材编写的有:黑龙江中医药大学、长春中医学院、辽宁中医学院、甘肃中医学院、天津中医学院、北京中医药大学、河南中医学院、山东中医药大学、上海中医药大学、南京中医药大学、浙江中医学院、江西中医学院、福建中医学院、湖北中医学院、湖南中医学院、成都中医药大学、广西中医学院、贵阳中医学院。

由于我们水平有限,书中定有不少不当与错误之处,恳请读者与同行批评指正。

编者
2001 年



编写说明

第一章 事件与概率

§ 1-1 随机事件及其运算	(1)	1-3.2 条件概率、概率的乘法定理	(10)
1-1.1 随机事件	(1)	§ 1-4 全概率与逆概率公式	(12)
1-1.2 事件之间的关系及运算	(2)	1-4.1 全概率公式	(12)
§ 1-2 事件的概率	(4)	1-4.2 逆概率公式(贝叶斯公式)	(13)
1-2.1 频率与概率(统计定义)	(5)	习题一	(15)
1-2.2 古典概率	(6)		
§ 1-3 概率的运算	(7)		
1-3.1 加法定理	(8)		

第二章 随机变量的概率分布与数字特征

§ 2-1 离散型随机变量及其概率分布	(17)	2-3.1 均数(数学期望)	(35)
2-1.1 随机变量	(17)	2-3.2 方差和标准差	(38)
2-1.2 离散型随机变量的概率分布	(18)	2-3.3 变异系数	(41)
2-1.3 二项分布、泊松分布及其他常见的离散型变量的分布	(20)	§ 2-4 三种重要分布的渐近关系	(42)
§ 2-2 连续型随机变量及其概率分布	(27)	2-4.1 二项分布的泊松近似	(42)
2-2.1 连续型变量的概率分布	(27)	2-4.2 二项分布的正态近似	(42)
2-2.2 正态分布及其他常见的连续型变量的分布	(29)	2-4.3 泊松分布的正态近似	(44)
§ 2-3 随机变量的数字特征	(35)	§ 2-5 大数定律及中心极限定理	(45)
		2-5.1 大数定律	(45)
		2-5.2 中心极限定理	(47)
		习题二	(48)

第三章 随机抽样和抽样分布

§ 3-1 随机抽样	(50)	3-3.4 F 分布	(58)
3-1.1 总体与样本	(50)	§ 3-4 样本分布图	(59)
3-1.2 随机抽样	(51)	3-4.1 样本的直方图	(59)
§ 3-2 样本的数字特征	(51)	3-4.2 经验分布图	(61)
3-2.1 统计量	(51)	§ 3-5 概率纸及其应用	(62)
3-2.2 样本的数字特征	(52)	3-5.1 正态概率纸	(62)
§ 3-3 抽样分布	(53)	3-5.2 对数正态概率纸	(64)
3-3.1 样本均数的分布	(54)	3-5.3 威布尔概率纸	(65)
3-3.2 χ^2 分布	(54)	习题三	(68)
3-3.3 t 分布	(56)		

第四章 连续型随机变量的参数估计与检验

§ 4-1 参数估计	(70)	4-3.1 单个正态总体均数 μ 的假设检验	(86)
4-1.1 点估计及其性质	(70)	4-3.2 单个正态总体方差的假设检验	(91)
4-1.2 区间估计的概念	(72)	§ 4-4 两个正态总体的参数检验	(93)
4-1.3 正态总体均数 μ 的区间估计	(73)	4-4.1 配对比较两个正态总体均数的差异	(93)
4-1.4 正态总体方差 σ^2 的区间估计	(80)	4-4.2 成组比较两个正态总体均数的差异	(95)
§ 4-2 假设检验	(83)	4-4.3 方差齐性检验(方差齐性与非齐性)	(99)
4-2.1 什么是假设检验	(83)	习题四	(101)
4-2.2 假设检验的基本思想	(84)		
4-2.3 假设检验中的两类错误	(85)		
§ 4-3 单个正态总体的参数检验	(85)		

第五章 方差分析

§ 5-1 单因素方差分析	(105)	5-2.1 q 检验法(HSD 法)	(112)
5-1.1 方差分析的原理与步骤	(106)	5-2.2 S 检验法	(113)
5-1.2 单因素方差分析的计算	(108)	§ 5-3 两因素试验的方差分析	(116)
§ 5-2 两两间多重比较的检验法	(112)	5-3.1 无重复试验	(116)

5-3.2 重复试验的双因素分析 … (119)	习题五 …………… (120)
--------------------------	-----------------

第六章 离散型变量的参数估计与检验

§ 6-1 总体率的区间估计 …… (123)	独立性检验 …………… (128)
6-1.1 查表法及其原理 …… (123)	6-3.2 $R \times C$ 列联表中独立性的检验 …………… (130)
6-1.2 正态近似法(大样本) …… (125)	
§ 6-2 总体率的假设检验 …… (126)	§ 6-4 参照单位法 …………… (132)
6-2.1 单个总体率的假设检验 … (126)	6-4.1 Ridit 分析 …………… (132)
6-2.2 两个总体率的假设检验 … (127)	6-4.2 用置信区间作显著性检验 …………… (133)
§ 6-3 列联表中独立性的检验 … (128)	习题六 …………… (135)
6-3.1 2×2 列联表(四格表)中的	

* 第七章 非参数检验

§ 7-1 配对符号秩和检验(Wilcoxon 配对法)…………… (137)	§ 7-4 配伍组设计多个样本比较 的秩和检验(Friedman 秩 和检验)…………… (146)
7-1.1 配对比较的符号秩和检验 …………… (137)	§ 7-5 两两比较的秩和检验 …… (148)
7-1.2 样本中位数与总体中位数比 较的符号秩和检验 …… (140)	7-5.1 多个样本间两两比较的秩和 检验 …………… (148)
§ 7-2 完全随机设计两样本比较的 秩和检验(Wilcoxon 两样本 比较法)…………… (141)	7-5.2 配伍组设计两两比较的秩和 检验 …………… (149)
7-2.1 原始数据的两样本比较 … (141)	7-5.3 多个实验组分别与一个对照 组比较的秩和检验 …… (151)
7-2.2 频数表资料的两样本比较 …………… (142)	§ 7-6 中位数检验法和游程检验 …………… (151)
§ 7-3 完全随机设计多样本比较的 秩和检验(H 检验法) …… (144)	7-6.1 中位数检验法 …………… (151)
7-3.1 原始资料多样本比较的秩和 检验 …………… (144)	7-6.2 游程检验 …………… (153)
7-3.2 频数表资料的多样本比较秩 和检验 …………… (145)	§ 7-7 等级相关分析(Spearman 法) …………… (156)
	习题七 …………… (157)

第八章 相关与回归

§ 8-1 相关	(160)	8-2.3 预测与控制	(168)
8-1.1 散点图	(160)	8-2.4 多元线性回归与一元非线性 回归的简介	(170)
8-1.2 相关系数的概念	(161)	§ 8-3 ED ₅₀ 和 LD ₅₀ 估计	(173)
8-1.3 相关系数的检验	(163)	8-3.1 概率单位法	(174)
§ 8-2 线性回归方程	(164)	*8-3.2 序贯法(上下法)	(176)
8-2.1 一元线性模型	(165)	习题八	(178)
8-2.2 线性回归方程	(165)		

第九章 正交试验设计

§ 9-1 基本概念	(180)	9-3.1 综合加权评分法	(192)
9-1.1 因素、水平、指标	(180)	9-3.2 综合平衡法	(194)
9-1.2 正交表、交互作用	(182)	§ 9-4 有交互作用的试验设计	(194)
§ 9-2 用正交表安排试验	(183)	§ 9-5 试验结果的方差分析	(201)
9-2.1 二水平试验	(184)	9-5.1 无重复试验的方差分析	(201)
9-2.2 三水平试验	(186)	9-5.2 有重复试验的方差分析	(205)
9-2.3 不等水平试验	(188)	习题九	(212)
9-2.4 试验结果的直观分析	(189)		
§ 9-3 多指标试验	(191)		

第十章 均匀试验设计

§ 10-1 基本概念	(215)	10-2.2 均匀设计表的选择及试验 方案的安排	(217)
10-1.1 概述	(215)	10-2.3 均匀设计试验的数据分析	(218)
10-1.2 均匀设计表及其使用表	(216)	10-2.4 均匀设计的注意事项	(220)
§ 10-2 均匀设计的步骤	(217)	习题十	(221)
10-2.1 均匀设计的步骤	(217)		

习题答案	(223)
------------	-------

附表	(228)
----------	-------

附表 1 二项分布累积概率 $P(X \geq k)$ 值表	(228)
--------------------------------------	-------

附表 2 泊松分布累积概率 $P(X \geq k)$ 值表	(230)
--------------------------------------	-------

附表 3 标准正态概率密度 $\varphi(x)$ 值表	(236)
-------------------------------------	-------

附表 4 标准正态分布函数 $\Phi(x)$ 值表	(237)
附表 5 标准正态分布的临界值表	(239)
附表 6 χ^2 分布的临界值表	(239)
附表 7 t 分布的临界值表	(241)
附表 8 F 分布的临界值表	(242)
附表 9 多重比较中的 q 表	(247)
附表 10 多重比较中的 S 表	(250)
附表 11 二项分布参数 p 的置信区间表	(251)
附表 12 泊松分布参数的置信区间表	(255)
附表 13 相关系数临界值表	(255)
附表 14 百分率与概率单位换算表	(256)
附表 15 配对比较符号秩和检验用 T 界值表	(258)
附表 16 两样本比较秩和检验用 T 界值表	(258)
附表 17 三样本比较秩和检验用 H 界值表	(259)
附表 18 配伍组试验秩和检验用 M 界值表($P=0.05$)	(260)
附表 19 游程个数检验用 r 界值表配伍组试验秩和检验用 M 界值表($P=0.05$)	(260)
附表 20 Spearman 等级相关系数 r_s 界值表	(261)
附表 21 常用正交表	(261)
附表 22 常用均匀设计表与使用表	(269)

第一章

事件与概率

数理统计方法是以概率论为理论基础,通过一定的设计来收集数据和进行整理分析,以部分资料推断总体的一种方法论.用它去研究大量随机现象的规律性,由于概率和随机事件是联系在一起的,因此事件和概率都是数理统计中最基本的概念.

本章将介绍随机事件,事件的概率及其运算.

§ 1-1 随机事件及其运算

1-1.1 随机事件

当我们多次观察自然现象和社会现象后,会发现许多事情在一定条件下必然会发生或者必然不会发生,例如,纯净的水,在一个大气压下,温度是 0°C 时必然结冰,在 20°C 时必然不会结冰,在 100°C 时必然沸腾,在 80°C 时必然不会沸腾;再比如,附子里必然含有乌头碱,黄连里必然提不出青霉素.这种完全可以预言其结果的现象,是一种确定性现象,叫**必然现象**.

另一类现象,在一定条件下都不可能事前完全准确地预言其结果,也就是它有多种可能发生的结果,是一种不确定性现象,这类现象称为**偶然现象**,例如,抛起一枚硬币究竟哪一面落地时朝上?从一批针剂中抽取一支来检验,其结果可能是正品,也可能是次品,在抽取之前是无法肯定的.偶然现象也称为**随机现象**.

对各种现象的“观察”称为试验,对随机现象的“观察”就称为**随机试验**.随机试验具有下列特征:

- (1) 在相同条件下,可以重复进行;
- (2) 各次试验结果不一定相同,而且每次试验之前不能预先判断哪一个结果发生;

(3) 所有可能的试验结果是预先可以明确的,并且在每一次试验中必有其中一个结果出现.

对某种现象的“观察”试验而得到的结果就称为**事件**.在一定条件下,试验结果中必然出现的事件,称为**必然事件**,记为 Ω .例如, {纯净的水,在一个大气压下,加热到 100°C 沸腾} = Ω , {附子含有乌头碱} = Ω .反之,那种在一定条件下,试验结果中必然不出现的事件,称为**不可能事件**,记为 Φ .例如, { $x^2 + 1 = 0$ 有实数解} = Φ ; {黄连内提出青霉素} = Φ 等.

随机试验观察的是随机现象,在一定条件下,试验结果中可能出现,也可能不出现的事件,称为**随机事件**,简称**事件**.随机事件一般用大写字母 A, B, C 等表示.例如,投掷一个硬币,这个随机试验中有两个事件 $A = \{\text{出正面}\}$ 和 $B = \{\text{出反面}\}$.必然事件与不可能事件可以说不是随机事件,但为了研究方便起见,把必然事件与不可能事件作为随机事件的两个极端来统一处理.

1-1.2 事件之间的关系及运算

在各种现象中,往往要求我们同时考察几个在同样条件下的事件及它们之间的联系,下面我们就来讨论事件的关系及运算:

一、包含

设有事件 A 及 B ,如果事件 A 发生必然导致事件 B 发生,则称事件 A 包含于事件 B 或事件 B 包含事件 A .并记为 $A \subset B$ 或 $B \supset A$.

二、等价

若事件 A 包含事件 B ,事件 B 也包含事件 A ,即 $A \supset B$ 且 $B \supset A$,就称事件 A 与 B 等价(或称相等),记作 $A = B$.

三、并事件

若事件 $C = \{A \text{ 或 } B \text{ 中至少有一个发生}\}$,则称 C 为 A, B 两事件的并事件,记

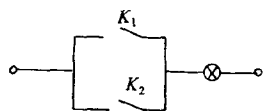


图 1-1

为 $C = A + B$. n 个事件的并事件记为 $A = \sum_{i=1}^n A_i$.

如把两个开关 K_1 和 K_2 并联后接入电路(图 1-1),设 $A = \{\text{电路接通}\}$, $A_1 = \{K_1 \text{ 闭合}\}$, $A_2 = \{K_2 \text{ 闭合}\}$,那么 $A = A_1 + A_2$.

四、交事件

若事件 $C = \{A \text{ 与 } B \text{ 同时发生}\}$,则称 C 为 A, B 两事件的交事件,记为 $C = AB$. n 个事件的交事件记为 $A = \prod_{i=1}^n A_i$.

如把两个开关 K_1 和 K_2 串联后接入电路(图 1-2),
那么 $A = A_1 A_2$.

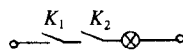


图 1-2

五、互不相容事件

若事件 A 与事件 B 不能同时发生,则称 A 与 B 为**互不相容事件**,记作 $AB = \Phi$.互不相容事件也称为**互斥事件**. n 个事件互斥,是指它们两两互斥.

若 n 个互斥事件的并事件是必然事件,即 $A_i A_j = \Phi (1 \leq i < j \leq n)$,且 $\sum_{i=1}^n A_i = \Omega$,则称这 n 个事件构成**互斥完备群**.

例如,治疗某种疾病,其疗效标准分为四个等级:痊愈、显效、微效和无效.那么,就一次试验(治疗一个患者的结果)而言,事件{痊愈}、{显效}、{微效}、{无效}是互斥事件,而且这四个事件构成互斥完备群.

六、对立事件

若在任一次试验中,事件 A 与事件 B 二者必有一个发生,且仅有一个发生,亦即 A, B 同时满足 $A + B = \Omega$ 及 $AB = \Phi$ 两个条件,也就是互斥完备群仅由两事件 A 与 B 构成,则称**事件 A 与事件 B 对立**,或**事件 B 是事件 A 的对立事件**,当然**事件 A 也是事件 B 的对立事件**. A 的对立事件记作 \bar{A} ,那么就有 $\bar{\bar{A}} = A$,或 $A = \bar{B}$.

如果治疗某种疾病,只考虑有效和无效两个等级,那么事件{有效}与{无效}就是对立事件.

不难理解,对立事件必为互斥事件,而互斥事件不一定对立.

例如,投掷一枚骰子,事件{出 1 点}与{出 2 点}互斥,但不对立.而事件{出偶数点}与{出奇数点}对立且互斥.

对任一事件 A ,都有 $A\bar{A} = \Phi, A + \bar{A} = \Omega, \bar{\bar{A}} = A$.

事件之间的这些关系,读者可以通过熟知的韦恩图作直观理解,图 1-3 给出几

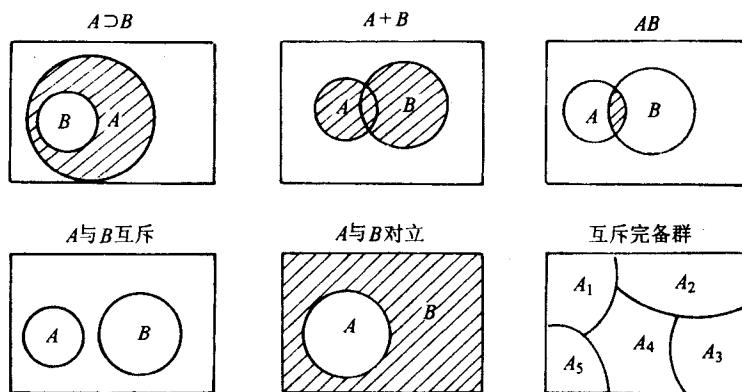


图 1-3 韦恩图

种常见情况.

对于事件的运算方法我们不但要学会用概率论的语言来解释这些关系及运算,并且应当逐步熟悉在不甚复杂的情况下,把一个事件化成其他事件的并、交或混合形式以及找出构成互斥完备群的全部事件,因为这是讨论事件间关系进而施行运算的重要途径.

例 大黄、黄连、黄芩是组成泻心汤的主要成分,依次检查它们的质量算一次试验,令 $A = \{\text{大黄质量合格}\}$, $B = \{\text{黄连质量合格}\}$, $C = \{\text{黄芩质量合格}\}$. 试用三个事件表示下列在一次试验中出现的事件:

- (1) 只有大黄质量合格;
- (2) 只有一种成分质量合格;
- (3) 每一种成分质量都不合格;
- (4) 至少有一种成分质量合格;
- (5) 构成互斥完备群的全部事件.

解 设 $\bar{A} = \{\text{大黄质量不合格}\}$, $\bar{B} = \{\text{黄连质量不合格}\}$, $\bar{C} = \{\text{黄芩质量不合格}\}$.

(1) $\{\text{只有大黄质量合格}\} = \{\text{大黄质量合格且黄连、黄芩质量不合格}\} = A\bar{B}\bar{C}$.

(2) 因为 $\{\text{只有大黄质量合格}\} = A\bar{B}\bar{C}$;

$\{\text{只有黄连质量合格}\} = \bar{A}B\bar{C}$;

$\{\text{只有黄芩质量合格}\} = \bar{A}\bar{B}C$;

所以, $\{\text{只有一种成分合格}\} = A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{B}C$.

(3) $\{\text{每种成分质量都不合格}\} = \bar{A}\bar{B}\bar{C}$.

(4) $\{\text{至少有一种成分质量合格}\} = A + B + C$,

或者 $\{\text{至少有一种成分质量合格}\} = A\bar{B}\bar{C} + \bar{A}B\bar{C} + \bar{A}\bar{B}C + \bar{A}BC + \bar{A}\bar{B}C + \bar{A}\bar{B}C + ABC$.

(5) 构成互斥完备群的全部事件有八个, 即 $\Omega = \bar{A}\bar{B}\bar{C} + \bar{A}\bar{B}C + \bar{A}B\bar{C} + \bar{A}BC + \bar{A}\bar{B}C + \bar{A}\bar{B}C + \bar{A}\bar{B}C + ABC$.

注意, 把一个事件化成若干个事件的并事件时, 必须明白是有交并(不互斥), 还是无交并(互斥), 这对后面的计算尤为重要. 如本例的问题(2)表示成三个事件的无交并; 问题(4)可表示成三个事件的有交并, 或者七个事件的无交并; 问题(5)构成互斥完备群的八个事件当然是无交并.

§ 1-2 事件的概率

通俗的说, 所谓概率是某一随机事件在试验中发生的可能性大小的数值表示, 通常用 $P(A)$ 来表示事件 A 的概率. $P(A)$ 越大, 说明事件 A 发生的可能性越大. 下面给出概率论中概率的两个定义, 从中可以了解概率的特性和计算方法.

1-2.1 频率与概率(统计定义)

人们在长期的实践中认识到:随机事件发生的可能性大小可以在大量重复试验中显示出来.

若进行条件相同的 n 次试验,事件 A 出现 m 次,则称 m 为事件 A 的**频数**,称比值 m/n 为事件 A 的**频率**,记为

$$f(A) = \frac{m}{n} \quad (1-1)$$

显然,事件的频率是通过特定的试验获得的,每作 n 次试验,所得到的频率可以各不相同.但经验证明,在同一条件下进行多次重复试验时,事件出现的频率会在某一常数附近左右摆动,这种性质,叫做**频率的稳定性**.

在历史上,这种频率的稳定性是在人口统计方面最先注意到的.如世界上一些国家通过多年观察,发现男婴的出生率稳定在 $22/43$ 附近,而女婴的出生率稳定在 $21/43$ 附近.

再如著名的投币试验,表 1-1 列出试验记录,容易看出,投掷次数逐渐增多时,|出现正面|这个事件的频率 $\frac{m}{n}$ 总是在 0.5 这个数附近摆动而逐渐稳定于 0.5. 所以 0.5 这个数能反映事件|出现正面|发生的可能性的.

表 1-1

试验者	投掷次数 n	正面次数 m	频率 m/n
德莫根	2048	1061	0.5181
蒲 丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

由此可见,频率的稳定性充分说明随机事件发生的可能性大小是事件本身固有的一种客观属性,并为我们衡量一个随机试验中随机事件发生的可能性提供了客观的基础.

概率的统计定义 在条件相同的 n 次试验中事件 A 发生 m 次,如果加大 n 时, A 的频率 m/n 逐渐稳定在一个常数附近,就把这个常数称为事件 A 的概率.在此定义下有

$$0 \leq P(A) \leq 1, \quad P(\Omega) = 1, \quad P(\Phi) = 0 \quad (1-2)$$

概率的统计定义实际上给出了一个近似的计算随机事件的概率的方法,即当试验次数 n 足够大时,一个事件的频率与概率应充分接近,所以用事件的频率作为概率的近似值.在医药学中,这种估计经常用到.需要注意的是不要把频率和概率相混淆.频率是我们已经进行的试验的结果,其数值随着试验次数的不同而变

化,具有偶然性,而概率是一种客观存在,是个确定的数值,具有必然性.

例 1 在某地区40岁以上的男子中进行心血管疾病的调查,结果在4000名男子中发现患高血压的有132人,求该地40岁以上男子患高血压病的概率.

解 由于 $n = 4000$, 这个数字比较大,故可以用频率作为患病(记为 A) 的概率,于是得:

$$P(A) \approx \frac{132}{4000} = 3.30\%$$

1-2.2 古典概率

如果构成互斥完备群的各个事件具有等可能性,则这些事件称为**基本事件**.所以今后凡说基本事件,就当然地具有等可能性而且它们的全部构成互斥完备群.

我们再次研究投掷一枚硬币出现正面的概率.由于掷币实验全部可能的结果,即总的基本事件只有两个: {正面} 和 {反面}, 而且硬币的物理和几何对称性导致出现正面和反面具有等可能性.因此不必按统计定义去作大量试验,掷币一次就可以直接判断出现正面的概率是 $1/2$.

长期的经验使人们注意到有一类特殊的随机现象:每次试验,基本事件只有有限个,而且各基本事件出现的可能性相等,我们称它为等可能概型.经过归纳(从个别到一般),在概率论的早期就提出了关于概率的第二个定义:

概率的古典定义 如果互斥完备群由有限的 n 个基本事件构成,而事件 A 包含 m 个基本事件,那么事件 A 的概率

$$P(A) = \frac{m}{n} \tag{1-3}$$

且有 $0 \leq P(A) \leq 1$, $P(\Omega) = 1$, $P(\Phi) = 0$

一般把古典定义下关于事件概率的数学模型称为古典概型.古典概型的大部分问题都能形象化地归结为抽球问题.

例 2 瓶中装有50片药,其中有3片次品,求

- (1) 一次取一片,取得次品的概率;
- (2) 一次取5片,5片中有2片是次品的概率.

解 (1) 50片药中取一片,其可能结果有50个基本事件(每片药被取到的可能性相等),即 $n = 50$.

设 $A = \{\text{取到次品}\}$, 则 A 包含3个基本事件,即 $m = 3$, 由古典定义得

$$P(A) = \frac{m}{n} = \frac{3}{50} = 0.06$$

(2) 50片药中任取5片,其可能结果有 C_{50}^5 个基本事件(C_{50}^5 种机会均等的取法),即 $n = C_{50}^5$.

设 $B = \{\text{5片中有2片次品}\}$, 则事件 B 包含的基本事件数 $m = C_3^2 C_{47}^3$, 故所

求概率

$$P(B) = \frac{C_3^2 C_{47}^3}{C_{50}^5} = \frac{9}{392} = 0.023$$

例 3 袋中有 2 个白球和 8 个黑球, 现在无放回地一个个抽出来, 求第 k 次抽到的是白球的概率 ($1 \leq k \leq 10$).

解法一 把 10 个球当作是有区别的, 即设想把它们按 1, 2, \dots , 10 进行编号, 若将抽出的球依次排成一排, 则全部可能的结果相当于把 10 个元素进行全排列, 即全部基本事件数为 $10!$.

第 k 次抽到白球, 即排在第 k 号位置上的那一个白球, 只能在 2 个白球中取得, 故有 2 种抽法. 而另外 9 次抽的球, 可在余下的 9 个中任取, 故有 $9!$ 种抽法. 所以事件 | 第 k 次抽到白球 | 包含的基本事件数为 $2 \times 9!$, 故第 k 次抽到白球的概率 $p = \frac{2 \times 9!}{10!} = \frac{2}{10}$.

解法二 把 2 个白球看成一样, 8 个黑球看成一样, 把抽出的球仍依次放在 10 个位置上, 由于白球看成一样, 黑球看成一样, 所以当白球位置选好, 其他位置必放黑球, 故总的排法即总的基本事件数为 C_{10}^2 , 而事件 | 第 k 次抽到白球 | 所包含的基本事件数为 $C_{10-1}^2 = C_9^1$ (因为 2 个位置中已有 1 个位置即第 k 号位置固定放了白球) 所以

$$p = \frac{C_9^1}{C_{10}^2} = \frac{2}{10}$$

两种解法结果一样, 抽到白球的概率 $p = 2/10$, 与次数无关, 这正好说明广泛应用于生产和生活中的抽签方法是公平合理的, 先抽后抽都一样, 机会均等.

需要说明的是, 无论是概率的统计定义, 还是古典定义, 都在概率计算中起一定的作用, 但又有着各自的局限性. 古典概率是以试验的所有可能结果只有有限个且具有等可能性为基础, 实际上这种条件很难满足, 至于统计概率, 则要求试验次数 n 充分大, 并以事件频率的稳定值近似地作为该事件的概率, 这里的 n 大到什么程度, 稳定值是什么都是不确切的, 因此, 人们需要对概率有个严格的定义, 使之能适用于一般的随机试验. 经过人们的不断地探索和总结, 终于在 1933 年, 由前苏联数学家柯尔莫戈洛夫提出了概率公理化结构, 明确定义了基本概念, 使概率论成为严谨的数学分支. 关于公理化体系的内容, 有兴趣的读者可参阅概率论专著, 此处不再赘述.

§ 1-3 概率的运算

把复杂事件的概率分解成简单事件的概率来计算, 可以借助于概率的运算法则.