卫生部规划教材

全国高等医药院校教材

供药学类专业用

医药数理统计方法

第三版

刘定远 主编

人民卫生出版社

全国高等医药院校教材

供药学类专业用

医药数理统计方法

第三版

刘定远　主编

编　者　（按姓氏笔画为序）
毕育学　（西安医科大学）
刘定远　（华西医科大学）
刘艳杰　（沈阳药科大学）
闵心畅　（华西医科大学）
张　侠　（北京医科大学）
施立人　（上海医科大学）
倪永兴　（中国药科大学）

人民卫生出版社

图书在版编目（CIP）数据

医药数理统计方法/刘定远主编．－3版：－北京：人民卫生出版社，1999
ISBN 7-117-03272-3

Ⅰ.医…　Ⅱ.刘…　Ⅲ.医用数学：数理统计-高等学校：医学院校-教材　Ⅳ.R311

中国版本图书馆 CIP 数据核字（1999）第 31933 号

医药数理统计方法
第三版
刘定远　主编
人民卫生出版社出版发行
（100078 北京市丰台区方庄芳群园 3 区 3 号楼）
三河市宏达印刷厂印刷
新 华 书 店 经 销

787×1092　16 开本　17.75 印张　403 千字
1987 年 10 月第 1 版　1999 年 11 月第 3 版第 12 次印刷
印数：40 481— 55 480
ISBN 7-117-03272-3/R·3273　定价：16.00 元
（凡属质量问题请与本社发行部联系退换）

全国高等医药院校药学专业
第四轮规划教材修订说明

为适应我国高等药学教育的改革和发展，在总结前三轮药学专业教材编写经验的基础上，卫生部教材办公室于1996年9月决定进行第四轮教材修订，根据药学专业的培养目标，确定了第四轮教材品种和修订的指导思想，药学本科教育的培养对象是从事一般药物制剂、鉴定及临床合理用药等工作的药师，教材修订应紧紧围绕培养目标，突出各学科的基本理论、基本知识，同时又反映学科的新进展。该套教材可供药学及相关专业选用。全套教材共22种，均经卫生部聘任的全国药学专业教材评审委员会审定。教材目录如下：

1．高等数学（第三版） 毛宗秀 主编
2．医药数理统计方法（第三版） 刘定远 主编
3．物理学（第三版） 王鸿儒 主编
4．物理化学（第四版） 侯新朴 主编
5．无机化学（第三版） 许善锦 主编
6．分析化学（第四版） 孙毓庆 主编
7．有机化学（第四版） 倪沛洲 主编
8．人体解剖生理学（第四版） 龚茜玲 主编
9．微生物学与免疫学（第四版） 李明远 主编
10．生物化学（第四版） 吴梧桐 主编
11．药理学（第四版） 李 端 主编
12．药物分析（第四版） 刘文英 主编
13．药用植物学（第三版） 郑汉臣 主编
14．生药学（第三版） 郑俊华 主编
15．药物化学（第四版） 郑 虎 主编
16．药剂学（第四版） 毕殿洲 主编
17．天然药物化学（第三版） 姚新生 主编
18．中医学基础（第四版） 李向中 主编
19．药事管理学（第二版） 吴 蓬 主编
20．生物药剂学与药代动力学 梁文权 主编
21．分子生物学基础 史济平 主编
22．药学英语（第二版） 胡廷熹 主编

以上教材均由人民卫生出版社出版。

卫生部教材办公室

全国药学专业教材第二届评审委员会

前言

根据全国高等医药院校药学专业教材评审委员会对第四轮规划教材的编写要求，在总结了前两版使用经验的基础上，为加强学生统计推断分析能力的培养及教学的方便，本书对第二版教材的结构和内容进行了部分修改。

本书以统计推断为中心、概率为基础、统计分析为应用的结构方式安排教材内容。其中，第一至三章为统计的概率基础；第四至八章为统计推断方法；第九、十两章为方差和回归模型的统计分析；删去了第二版教材中的第十章和第十一章，其中估计样本容量的内容分别并入相关的参数估计和参数检验；正交表与试验设计并入第九章；将第二版第三章的概率纸及其应用和第六章的离散资料的 χ^2 检验，加上正态性的偏态峰态检验组成第七章拟合优度检验；第八章非参数检验增加了多重比较的秩和检验及两个样本的游程检验，以增强非参数检验的应用范围。另外，在统计的概率基础内，还增加了超几何分布、中位数和众数等基本概念。

鉴于医药数理统计教材需要兼顾医药统计的实践性和数理统计的理论性。因此，在编写本书时，对统计概念的数学定义着力阐明其医药背景；对统计方法，一方面充分利用直观和数学推导相结合的方法分析其数学原理，另一方面通过对医药实例的分析处理，突出分析思路和计算步骤，增强应用统计方法的可操作性。这样，不仅便于学习和掌握，而且更有利于在实践中正确选择统计方法，准确进行统计计算，合理应用分析结果，提高对医药问题的统计处理能力。

对于统计假设检验中的三种判断方法，虽然都作了讨论，但是，为了便于掌握，本书在各种统计推断分析中，重点介绍的是当前普遍采用的 P 值方法。为帮助学生更好地应用统计推断分析方法，了解统计的计算机技术，书后还编有计算机“常用统计分析软件简介”的附录。为更好发挥学生学习的积极性和主动精神，本版各章不再编写小结。

本书引用了前版教材中的一些资料，在此对前版主编和编者表示谢意。同时，感谢药学专业教材评审委员会、卫生部教材办、华西医科大学和上海医科大学药学院的领导和工作人员对本书编写、出版的帮助和支持。

限于编者的水平，书中定有不少缺点、错误，恳请使用本书的师生和广大读者批评指正。

刘定远

1998 年 6 月于华西医科大学

目　录

第一章 随机事件及其概率

数理统计是研究和揭示随机现象数量规律性的数学学科，在医药学及其生产领域中，有着极其广泛的应用，为医药工作者所必备的知识。本章从随机事件出现的统计规律出发，阐明概率的含义、实际背景与定义，并进而探讨概率计算的一些基本方法。

第一节 随机事件的概率

一、随机事件

自然界里有各种现象，常见的有两类。一类在一定条件下必然发生或必然不发生。例如，在标准大气压下，100℃的纯水必然沸腾；同性电荷必不相互吸引，等等。这类现象称为确定性现象。微积分学和线性代数等就是研究这类现象的数学。另一类是在一定条件下有不确定的结果，即可能出现这种结果，也可能出现那种结果。例如投掷一枚硬币，可能是正面(图案面)，也可能是反面(币额面)；在同一条件下生产的一批针剂中，有的是合格品，有的是次品；某种疾病的患者，服用相同剂量的同种药物后，有的痊愈，有的无效，有的显效而未痊愈；投掷一颗骰子，观察出现的点数，等等。这类现象称为随机现象。为了研究随机现象的规律，需要进行大量重复的调查、实验、测试等，所有这些统称为试验，数理统计中所说的试验是指**随机试验**(random trial)，它具有下列三个特性：

1. 可以在相同的条件下重复进行。
2. 每次试验有且都有两个及其以上的可能结果，且能事先明确所有可能结果。
3. 试验前无法预测其结果。

上面列举的投掷硬币等均属随机试验。随机现象就是通过随机试验来研究的。

随机试验的部分可能结果组成的集合称为**随机事件**(random event)，简称**事件**，常用 A，B，C 等大写字母表示。在试验中必然发生的事件称为**必然事件**，记作 U；必然不发生的事件称为**不可能事件**，记作 V。这两种事件可当作随机事件的特例。

例如，一批产品中，有正品也有次品，从中任取1件，“取得正品”与“取得次品”都是随机事件，显然取得“正品”的可能性一般应比“取得次品”的可能性大。可见随机事件出现的可能性大小，是研究随机现象的一个主要问题。

二、频率与概率

随机事件是在一定条件下可能发生，也可能不发生的事件，看起来似乎没有什么规律可循，其实不然，当我们在相同条件下进行大量重复试验时，就会出现一定的统计规律性。

若随机事件 A 在 n 次重复试验中出现了 m 次，则称比值 m/n 为事件 A 出现的**频率**(frequency)，记作

$$f_n(A)=m/n \tag{1.1}$$

【例 1】 作投掷硬币试验，结果如表 1.1。试考察其出现正面的频率的规律性。

表 1.1 投掷硬币试验

试验者	投掷次数(n)	出现正面次数(m)	频率(m/n)
Demorgen	2046	1061	0.5186
Buffon	4040	2048	0.5096
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

投掷一次，出现正面或反面，都是随机事件，预先无法确定。但由表 1.1 可知，随着投掷次数的增加，出现正面的频率稳定地接近 0.5。

【例 2】 某地 1927 年至 1932 年婴儿出生数如表 1.2。试考察生男孩的规律性。

表 1.2 某地婴儿出生数与频率

出生年份	出生数			频率	
	男孩(m)	女孩(k)	合计($m+k$)	f_n(男)	f_n(女)
1927	496544	462189	958733	0.518	0.482
1928	513654	477339	990993	0.518	0.482
1929	514765	479336	994101	0.518	0.482
1930	528072	494739	1022811	0.516	0.484
1931	496986	467587	964573	0.515	0.485
1932	482431	452232	934663	0.516	0.484
合计	3032452	2833422	5865874	0.517	0.483

分娩一次，生男生女均为随机事件，无法预测。但由表 1.2 可知，当分娩次数充分大时，男孩的出生率在 0.517 附近摆动。

种种事实表明，在大量重复试验中，随机事件的频率总是围绕着某一确定的值稳定地摆动，这是一种统计规律，称为频率的**稳定性**。频率的稳定性说明随机事件发生的可能性大小，是其本身固有的客观属性，这就是概率的含义与实际背景，也是概率定义的理论依据。

定义 1(概率的统计定义) 设在相同的条件下，进行大量重复试验，若事件 A 的频率稳定地在某一确定值 p 的附近摆动，则称 p 为事件 A 的**概率**(probability)，记作 $P(A)=p$。

应强调指出，频率和概率都是刻划随机事件发生的可能性大小的数量指标。前者就局部而言，对于同种而不同的试验和试验次数的不同，其值不尽相同，具有偶然性；后者就全体而言，是个确定值，是客观存在的，具有必然性。由频率认识概率的过程，体现了由偶然认识必然，由局部认识全体的过程，这是概率统计思维方法的重要特点。

概率的统计定义提供了计算概率的近似方法。当试验次数 n 充分大，频率相当稳定时，可把频率作为概率的近似值

$$P(A)=p\approx m/n \tag{1.2}$$

由定义可知，任何事件 A 的概率满足

$$0\leqslant P(A)\leqslant 1 \tag{1.3}$$

必然事件的概率为 1，不可能事件的概率为 0，即

$$P(U)=1,\ P(V)=0$$

三、古 典 概 型

有些事件的概率不必进行大量重复试验就能确定。如例 1 中的掷币试验。投掷一次，有且只有两个可能结果，即出现正面或反面。由于硬币较均匀，可以认为出现这两种结果具有同等的可能性，又每次能且只能出现其中的一个结果，所以出现正面的可能性大小为 1/2，即 P(正面)$=1/2$。这就是说，可以利用划分等可能事件的方法求得事件的概率。

【例 3】 一个口袋里装有大小相等、质量相同的球 16 个，其中白球 2 个，红球 3 个，黄球 5 个，黑球 6 个。从中任摸 1 球，问摸得白、红、黄、黑及彩色球的概率各为多少？

解 从袋内任摸一球，共有 16 种等可能结果，每次能且只能出现其中一种结果。以 A_1，A_2，A_3，A_4 分别表示摸得白球、红球、黄球、黑球等事件，则 A_1，A_2，A_3，A_4 分别包含 2 个、3 个、5 个、6 个等可能事件，所以它们的概率依次为

$$P(A_1)=2/16=\frac{1}{8};\quad P(A_2)=3/16;$$

$$P(A_3)=5/16;\quad P(A_4)=6/16=3/8;$$

令 $B=\{$摸得彩色球$\}$，则 B 包含 14 个等可能事件，于是

$$P(B)=14/16=7/8$$

据例 3 的思路，可以总结出一种求概率的方法。为此，先给出等概基本事件完备组的定义。

定义 2 如果一组事件 $e_1,e_2,\cdots,e_n$ 满足以下三个条件，则称该事件组为等概基本事件完备组，其中每一事件 e_i 称为等概基本事件。

(1) $e_1,e_2,\cdots,e_n$ 出现的概率是相等的(等可能性)。

(2)每次试验，n 个事件必然出现其中一个(完备性)。

(3)每次试验，n 个事件只能出现其中一个(互不相容性)。

据此，如果把试验的结果划分为 n 个等概基本事件的完备组，事件 A 包含其中 m $(m\leqslant n)$个等概基本事件，则事件 A 发生的概率为

$$P(A)=\frac{A\text{ 所包含的等概基本事件个数}}{\text{等概基本事件总数}}=\frac{m}{n} \tag{1.4}$$

这种借助分析等概基本事件的个数来计算概率的模型称为**古典概型**。

【例 4】 箱中装有 100 件产品，其中有 3 件次品。为检查产品质量，从中任取 5 件，求所取的 5 件中恰有 1 件次品的概率。

解 自 100 件中任取 5 件，共有 C_{100}^5 种等可能取法，每次能且只能是其中的一种取法，而 5 件中恰有 1 件次品(A)的取法占其中的 $C_3^1C_{97}^4$ 种，故所求概率为

$$P(A)=\frac{C_3^1C_{97}^4}{C_{100}^5}=0.138$$

【例 5】 设某遗传病反映在性染色体 X 上(记为 X^*)。已知父亲的染色体为(X^*, Y),母亲的为(X^*,X)。试讨论其子代中遗传病患者或遗传病携带者的概率。

解 根据遗传学知识,性染色体是(X^*,Y)的男人和性染色体是(X^*,X^*)的女人为该遗传病的表现者,性染色体为(X^*, X)的女人为遗传病携带者。在精卵结合时,父母各一染色体被结合在一起。因此,当父亲的染色体为(X^*, Y),母亲的染色体为(X^*,X)时,其子代有四种可能:

母	父	
	X^*	Y
X^*	(X^*,X^*)	(X^*,Y)
X	(X,X^*)	(X,Y)

可见,子代中男性患遗传病的概率为1/2,正常的概率为1/2。女性患遗传病的概率为1/2,为遗传病携带者的概率为1/2,为正常人的概率为0。

由此可见,从优生的意义讲,此种婚配不甚合适。如果已婚配,一定要通过产前诊断,选取其所怀子代为正常男性(X, Y)。

最后指出两点:第一,上述的概率定义有较大局限性,更一般的是概率的公理化定义,但这已超出了本书的范围;第二,通常一个随机试验的所有可能结果只具有完备性和互不相容性,而不一定具有等可能性。我们称随机试验所有互不相容的可能结果的集合为**结果空间或样本空间**(sample space),记作 Ω。所有互不相容的可能结果中的每一个结果都称为样本空间的**基本事件或样本点**(sample point)。

思考与讨论

1. 什么是随机事件?它有什么特点?举例说明。
2. 事件的频率和概率有何区别?
3. 什么是古典概型?如何用古典概型计算概率?举例说明。
4. 如果父母中至少有一人的性染色体有问题时,试讨论其子代中遗传病或遗传病携带者的概率。

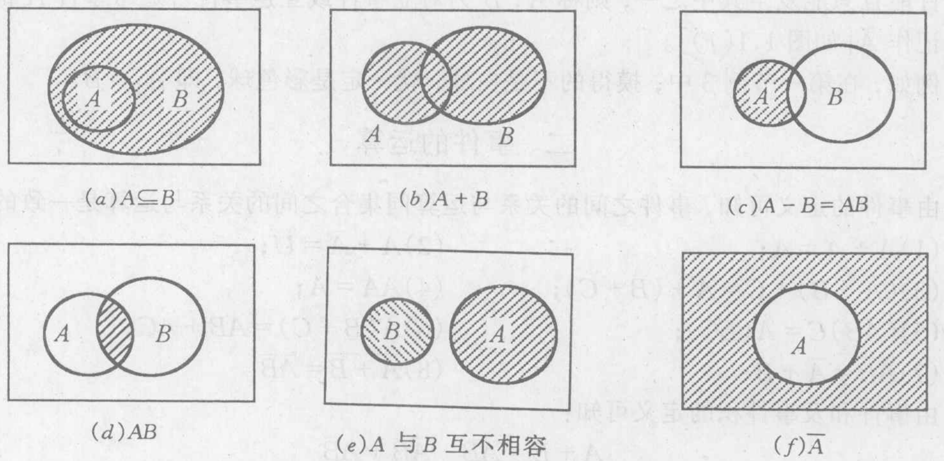
第二节 事件间的关系和运算

运用古典概型计算概率是较难的,借以作较复杂事件的概率计算就更难了。但是我们已经看到,一个事件常常包含若干个等概基本事件;一个比较复杂的事件常常包含若干个所谓简单事件。因此,弄清事件间的相互关系和运算规律对于把复杂事件的概率分解成若干简单事件的概率来计算是十分必要的。

一、事件间的关系

1. 包含与相等 若事件 A 的发生必然导致事件 B 的发生,则称事件 B 包含事件 A。记作 $B \supseteq A$ 或 $A \subseteq B$。事件间的关系还可用韦恩(Ven)图作直观理解。如图 1.1 中的(a)即为事件 B 包含事件 A 的图示。

若事件 A 包含事件 B,同时事件 B 又包含事件 A,即 $A \supseteq B$ 且 $A \subseteq B$,则称事件 A 与事件 B 相等,记作 $A=B$。



(a) $A \subseteq B$　(b) $A+B$　(c) $A-B=A\overline{B}$

(d) AB　(e) A 与 B 互不相容　(f) $\overline{A}$

图 1.1　事件 A 与 B 间几种关系的示意图

例如，在第一节例 3 中，摸得彩色球的事件一定包含摸得红球这一事件，即 $B \supseteq A_2$。

2. 事件的和　若事件 A 与事件 B 中至少有一个发生，这样的事件称为事件 A 与事件 B 的和或并，记作 $A+B$ 或 $A \cup B$，如图 1.1(b)。

一般地，若有限个事件 A_1，A_2，…，A_n 中至少有一个发生，这样的事件称为这 n 个事件的和或并，记作 $A_1+A_2+\cdots+A_n$ 或 $A_1 \cup A_2 \cup \cdots \cup A_n$，简记为 $\sum\limits_{i=1}^{n} A_i$ 或 $\bigcup\limits_{i=1}^{n} A_i$。

例如，在第一节例 3 中，摸得彩色球的事件就是摸得红、黄、黑三种球中至少一种球的事件，即 $B=A_2+A_3+A_4$。

若事件 A 发生且事件 B 不发生，这样的事件称为事件 A 与事件 B 的差，记作 $A-B$[如图 1.1(c)]。

例如，在第一节例 3 中，摸得红球可以说成摸得彩色球但不是黄球也不是黑球。即 $A_2=B-(A_3+A_4)$。

3. 事件的积　若事件 A 与事件 B 同时发生，这样的事件称为事件 A 与事件 B 的积或交，记作 AB 或 $A \cap B$[如图 1.1(d)]。

例如，投掷两枚硬币，分别以 A，B，C 表示事件"第一枚为正面"，"第二枚为反面"，"第一枚为正面且第二枚为反面"，则 $C=AB$ 或 $C=A \cap B$。

一般地，若有限个事件 A_1，A_2，…，A_n 同时发生，这样的事件称为这 n 个事件的积或交，记作 $A_1A_2\cdots A_n$ 或 $A_1 \cap A_2 \cap \cdots \cap A_n$，简记为 $\prod\limits_{i=1}^{n} A_i$ 或 $\bigcap\limits_{i=1}^{n} A_i$。

4. 互不相容事件　若事件 A 与事件 B 不能同时发生，则称这两个事件互不相容或互斥，记作 $AB=V$[如图 1.1(e)]。

如果 n 个事件两两互不相容，则称这 n 个事件互不相容。

例如，在第一节的例 3 中，摸得白球(A_1)，摸得红球(A_2)，摸得黄球(A_3)和摸得黑球(A_4)是两两互不相容的。

若一组事件 A_1，A_2，…，A_n 两两互不相容，且它们的和为必然事件，则称该事件组为**互不相容完备事件组**。特别，若 A，B 两事件满足：$AB=V$，$A+B=\cup$，即 A，B