高等医药院校教材

供药学类专业用

医药数理统计方法

第二版

方积乾 主编

人民卫生出版社

高等医药院校教材

供药学类专业用

医药数理统计方法

第二版

方积乾 主编

编者（按姓氏笔画为序）
王 珍（上海医科大学）
方积乾（北京医科大学）
刘定远（华西医科大学）
李柏新（中国药科大学）
宋学源（沈阳药科大学）
姚金华（北京医科大学）

秘书 张 俊（北京医科大学）

人民卫生出版社

医药数理统计方法

第 二 版

方积乾 主编

人民卫生出版社出版

（北京市崇文区天坛西里 10 号）

北京市房山区印刷厂印刷

新华书店北京发行所发行

787×1092毫米16开本 21 $\frac{1}{2}$ 印张 5 插页 496千字

1987年10月第1版 1995年9月第2版第9次印刷

印数：29 231—50 280

ISBN 7-117-00199-2/R·200 定价：16.80元

说　　明

这套"普通高等教育医药类规划教材"是卫生部组织编写的规划教材。初版始于1978年，迄1983年出齐。1985年至1989年进行了第二轮修订。这次第三轮修订工作是1990年开始的。由于出版单位和课程设置的变动，故新版教材的版次略有不同，多数为第三版，少数为二版和一版，请读者注意。本教材紧密结合药学专业培养目标要求，着重基础理论基本知识，亦反映本学科的新发展。本教材可供药学及相关专业选用。全套教材现为19种，均经卫生部聘任的全国药学专业教材评审委员会审定。教材名录如下：

1	《高等数学》（第二版）	方积乾　主编
2	《医药数理统计方法》（第二版）	方积乾　主编
3	《物理学》（第二版）	王鸿儒　主编
4	《物理化学》（第三版）	鲁纯素　主编
5	《无机化学》(第二版)	王　夔　主编
6	《分析化学》（第三版）	孙毓庆　主编
7	《有机化学》（第三版）	廖清江　主编
8	《人体解剖生理学》(第三版)	钱梓文　主编
9	《微生物学》（第三版）	王道若　主编
10	《生物化学》（第三版）	陈琼华　主编
11	《药理学》（第三版）	竺心影　主编
12	《药物分析》（第三版）	安登魁　主编
13	《药用植物学》（第二版）	沈联德　主编
14	《生药学》（第二版）	徐国钧　主编
15	《药物化学》（第三版）	李正化　主编
16	《天然药物化学》（第二版）	姚新生　主编
17	《药剂学》（第三版）	奚念朱　主编
18	《中医学基础》（第三版）	李向中　主编
19	《药事管理学》	吴　蓬　主编

以上教材均由人民卫生出版社出版，新华书店总店科技发行所发行。

序　言

本书系卫生部组织编写的全国高等医药院校规划教材。第一版于1987年由人民卫生出版社出版。得力于卫生部教育司教材办公室和专家委员会的指导和编审小组全体成员的配合，顺利完成了修订再版工作。

本版前三章的内容大致如前，仅将原属后面章节的正常值范围和概率坐标纸二项内容提前。第四、五、六章删减并充实了前一版的第四、五章，着重区间估计和假设检验的基本概念及常用统计方法，其中贯穿了关于区间估计的两种解释和关于假设检验的P值方法、临界值方法和置信区间方法，以帮助读者从多个侧面感受与理解统计学思维逻辑，并适应文献中不同作者的不同表达方式。第七章方差分析，增加了两因素试验的多重比较。第八章回归在先，相关在后，并删去了第一版中关于失拟情况的检验。第九章非参数检验充实了原理的解释，并增添了颇为实用的符号秩检验、秩相关和游程检验。

为了满足现代药物科研和生产的实际需要，进一步提高读者对前述统计学基本概念与方法的理解和应用的能力，本版增加了两部分内容：一是统计推断在专业工作中的四方面应用；二是临床试验设计的原理与方法。在第十章中，介绍了比较性研究的样本量预算、产品验收抽样的原理与方案制订、生产过程的控制图和异常值的检验，一则密切联系专业工作，二则生动地应用前面的内容。在第十一章中，除原有的正交试验设计外，增加了临床试验设计的内容，旨在吸引读者顺应国际潮流，在药物研制过程中重视临床试验的设计与分析。

此外，本版将原每节之后的"练习"改为"思考与讨论"，着重引导读者讨论概念和统计学思想，鼓励将习题课逐渐过渡为课堂讨论；课文中主要的统计学术语附有相应的英文，并在书后汇总成汉英对照表(以汉语拼音为序)，帮助读者熟悉有关的英文词汇，以利于日后的文献阅读和写作；每章之末各有一简短小结，或概括要点，或纵横联系，或提请警戒；每章之后的习题有所调整，书后仍附答案；附表也有增删；目录中附有星号*的章、节或段可视学时多寡选用。

为使本版更适应培养目标的需要，更便于教学，对于极为重要的概念与方法，尽量深入浅出地多方解释与应用；对于相对次要的枝节，或略或删，以引入专业工作中普遍有用的新知识。衷心希望使用本书的师生以及从事药物科学教育、科研和生产的同道多提宝贵建议。

最后，感谢北京医科大学生物数学与生物统计学教研室的同事。张侠老师除担任秘书工作外，还参与了全部讨论，提出了许多有益的建议。杜彩霞技师热情为本书精心绘制了全部新增插图，贺东奇、王纪宪、王艺、栾建安和杨淑英老师帮助誊清了大部份书稿，阮小群老师协助做了部分秘书工作。此外，还感谢人民卫生出版社的有关工作人员。曾经关心和参与本书再版工作的还有各院校的负责人、教务部门以及评审委员会全体成员，在此一并鸣谢。

方积乾

1992年 5月31日

目　　录

第一章　随机事件的概率及其计算

概率统计是研究随机现象的规律性的学科，在医药学领域中，有非常广泛的应用，是医药工作者必备的知识。本章从随机事件出现的统计规律出发，阐明概率的实际背景与定义，进而讨论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，有一类现象在一定条件下必然发生。例如，地球上，在标准大气压下，100℃的水必然沸腾；上抛的石子必然下落。这类现象称为确定性现象。过去我们学过的微积分就是研究这类现象的一种数学。

还有一类现象，在同一条件下有不确定的结果。例如，投掷一枚硬币，结果可能是正面（图案面），也可能是反面（币额面）。对一次投掷来说，究竟出现哪一种结果，在投掷前是不能预先确定的；在同一条件下生产的一批针剂中，有的是合格品，有的是不合格品；某种疾病的患者，服相同剂量的一种药物后，有的痊愈，有的无效，有的显效而未痊愈。这类现象的共同特点是：在相同条件下进行实验或观察，每次结果不尽相同，对一次实验或观察来说，究竟出现哪一种结果，事前是不能确定的。这种在同一条件下进行实验观察具有不确定结果的现象，就是所谓随机现象。对随机现象进行实验或观察称为**随机试验**（random trial），以后简称试验。随机试验的各种可能结果称为**随机事件**（random event），简称**事件**。常用大写字母 A，B，C，…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。这两种事件也可当作随机事件的特例。

对于一个随机试验，最关心的是它的各种可能结果（随机事件）出现的可能性大小例如服用某一药物，最关心的就是服药后出现痊愈或有效等的可能性大小。

二、频率和概率

随机事件是一种可能发生，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，当我们在同一条件下进行大量试验时，就会呈现某种规律性。

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

$$f_n(A)=m/n \tag{1}$$

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

单独投掷一次，出现正面还是反面，预先是不能确定的。因此“出现正面”是一个随机事件。表 1·1 的统计资料告诉我们，随着试验次数的大量增加，出现正面的频率稳定地接近 0.5。

表 1·1 硬币投掷试验

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

【例 2】 某地 1927 年至 1932 年婴儿出生数如表 1·2。试考察生男孩的频率的规律。

对单独一次生育，生男生女是不能预先确定的。因此，生男孩是一个随机事件。表 1·2 的统计资料告诉我们，当试验次数充分大时，生男孩的频率在 0.517 附近摆动。

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

由此可见，当试验次数 n 充分大时，随机事件的频率总是围绕着某一确定值稳定地摆动。这个确定值正好代表这个随机事件出现的可能性大小。这就是概率概念的客观背景。

定义 1 设在同一条件下，重复进行 n 次独立试验，随机事件 A 出现 m 次，若试验次数 n 充分大，频率 m/n 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 $P(A)=p$。

这个定义通过频率的稳定性，不仅说明了概率的客观性，而且还指出了估计概率的方法，即经过大量试验后，可用事件 A 的频率作为所求概率的近似值

$$P(A)=p\approx m/n \tag{2}$$

【例 3】 某医院用新药治疗老年性气管炎，其疗效如下：

治疗结果	临床治愈	明显好转	症状缓解	无效	合计
例数(m)	83	180	117	23	403(n)
频率(m/n)	0.206	0.477	0.290	0.057	1.00

这里，感兴趣的问题应该是临床治愈率（概率）是多少，而不是403例的治愈频率。由概率的定义知道，当n比较大时，频率可以作为概率的近似值。所以，分别算出各治疗结果的频率，以估计其相应的概率。当然，治疗例数越多，这个近似值就会越说明问题。

任何事件 A 的概率满足

$$0 \leqslant P(A) \leqslant 1 \tag{3}$$

必然事件U的概率为 1，不可能事件V的概率为 0，即

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

三、古典概型

有些事件的概率不用进行大量重复的试验也能确定，像例 1 中的事件可以这样推断其概率：投掷一次，可分为两个可能事件，出现正面和出现反面。因为硬币是比较均匀的，可以认为出现正面和反面具有同等的可能性，而且每次能且只能出现其中一个结果，所以出现正面的可能性大小是½，即概率 $P(\text{正})=½$。这就是说，可以用划分等可能事件的方法求得事件的概率。

【例4】 设一个口袋里装有大小相等，质量相同的各色球 16 个，其中白球 2 个，红球 3 个，黄球 5 个，黑球 6 个。从中任摸一球，问摸得白、红、黄、黑球及彩色球的概率分别是多少？

解 从袋内摸球，共有 16 种等可能结果，每次能且只能出现一种结果，所以总共可以划分出 16 种等可能事件，每次只能出现其中的一种。

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。因为事件A_1包含两个等可能事件，所以它的概率是 $P(A_1)=2/16=1/8$。同样，因为 A_2、A_3、A_4 分别包含 3 个、5 个、6 个等可能事件，所以，它们的概率分别为

$$P(A_2)=3/16, \quad P(A_3)=5/16, \quad P(A_4)=6/16=3/8$$

令 $B=\left\{\text{摸得彩色球}\right\}$，则$B$包含 14 个等可能事件，于是

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

依照例 4 的思路，我们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，…，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，每一事件 e_i 称为等概基本事件：

(1) n个事件中，每一个出现的概率是相等的（等可能性）；

(2) 经一次试验，n个事件中必然出现一个（完备性）；

(3) 经一次试验，n个事件中只能出现一个（互不相容性）。

如果在等概基本事件完备组的 n 个等概基本事件中，事件A包含m $(m \leqslant n)$ 个等概基本事件，则事件A发生的概率为

$$P(A)=\frac{\text{A所包含的等概基本事件个数}}{\text{等概基本事件总数}}=\frac{m}{n} \tag{4}$$

这种用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5】 瓶中装有50片药，其中有3片次品。今自瓶中任取5片，求所取5片中有2片是次品的概率。

解 自50片中任取5片，总共有C_{50}^5种机会均等的取法，每次能且只能是一种取法，其中，5片中有2片次品和3片正品的情况占$C_3^2 \cdot C_{47}^3$种，故所求概率为

$$\frac{C_3^2 \cdot C_{47}^3}{C_{50}^5}=\frac{9}{392}=0.0230$$

【例6】 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 由遗传学知识，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），母亲为（X^*，X）时，其子代有四种可能情形：

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是½，正常的概率是½。女性患遗传病的概率是½，为遗传病携带者的概率是½，为正常人的概率是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，上面的概率定义有较大的局限性，更一般的是概率的公理化定义，但是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**。

思考与讨论

1. 某事件的频率和概率有何区别？
2. 什么是古典概型？如何用古典概型计算概率？举例说明。
3. 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 事件间的相互关系

运用古典概型计算概率是比较难的，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A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所谓简单事件。因此，弄清事件之间的相互关系对于把复杂事件的概率分解成若干简单事件的概率来算是很必要的。为此先定义事件间的一些基本关系，然后讨论事件

的运算规律。

一、事件间的基本关系

1. **包含与相等** 两个事件A和 B，如果事件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$A\subseteq B$。

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$A=B$。

例如，在第一节例 4 中，摸得彩色球的事件一定包含摸得红球这一事件，记作$B\supseteq A_2$。

2. **事件的和** 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之和，记作$A+B$或$A\cup B$。

一般地若事件 A_1，A_2，…，A_n中至少有一个发生，这一事件被称为事件 A_1，A_2，…，A_n的和，记作 $A_1+A_2+\cdots+A_n$ 或 $A_1\cup A_2\cup\cdots\cup A_n$。

例如，在第一节例 4 里，摸得彩色球就是摸得红、黄、黑三种球中至少一种球的事件，$B=A_2+A_3+A_4$。

事件A发生且事件B不发生，这一事件称为事件A与事件B的差，记作 $A-B$。

例如，在第一节的例 4 里，摸得红球可以说成是摸得彩色球但不是黄球也不是黑球。故 $A_2=B-(A_3+A_4)$。

3. **事件的积** 事件A与事件B同时发生，这一事件称为事件A与事件B的积。记作 AB 或 $A\cdot B$ 或 $A\cap B$。

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

类似可定义：有限个事件 A_1，A_2，…A_n同时发生，这一事件称为 n 个事件的积，记作 $A_1A_2\cdots A_n$ 或$\prod\limits_{i=1}^{n}A_i$ 或 $A_1\cap A_2\cap\cdots\cap A_n$。

4. **互不相容事件** 若事件A与事件B不可能同时发生，则称这两个事件互不相容。如果 n 个事件是两两互不相容的，则称这 n 个事件互不相容。

例如，在第一节例 4 里，摸得彩色球的事件B所包含的 3 个事件：摸得红球（A_2），摸得黄球（A_3），摸得黑球（A_4），它们是两两互不相容的。显然，摸得白球（A_1）及 A_2，A_3，A_4 也是两两互不相容的。

若一组事件 A_1，A_2…，A_n两两互不相容，且它们的和是必然事件，则称该事件组为**互不相容完备事件组**。特别，若 AB 为不可能事件，$A+B$ 为必然事件，则称事件B为事件A的**对立事件**，记作 $\bar{A}$。显然，A与 $\bar{A}$ 互为对立事件，即 $\bar{\bar{A}}=A$。

例如，在第一节例 4 里，摸得的不是白球就一定是彩色球，即 $\bar{A}_1=B$。当然，摸得的不是彩色球就一定是白球，即 $\bar{B}=\bar{\bar{A}}_1=A_1$。

【例1】 一张含有某种特殊细胞的片子，在显微镜下观察时，细胞出现在某一位置上是随机的，若在A圈内找到该种细胞用事件A表示，在B圈内找到该种细胞用事件 B表示。试用几何图形表示事件$A\subseteq B$，$A+B$，$A-B$，AB，$\bar{A}$ 以及A与B互不相容。