医药应用概率统计

周怀梧　　主编
倪永兴　　副主编

百家出版社

医药应用概率统计

周怀梧　主　编
倪永兴　副主编

百　家　出　版　社

医药应用概率统计

周怀梧　主编

倪永兴　副主编

百家出版社出版

新华书店上海发行所发行　　群众印刷厂印刷

上海沪江电脑科技排印公司排版

开本 850×1156　1／32　印张 14.5　字数 357,000

1990 年 2 月第 1 版　1990 年 2 月第 1 次印刷

印数 1—7000 册

ISBN　7-80576-039-x／R · 02　　定价：6.00 元

医药应用概率统计编写组成员

第 二 军 医 大 学 薛祉绥
山 东 医 科 大 学 郭怀兰 虞孝珍
山 西 医 科 大 学 刘庆欧
哈尔滨医科大学 张 仲
贵 阳 医 学 院 严 燕
浙 江 医 科 大 学 徐绍英 周怀梧
上 海 医 科 大 学 王 珍 孙伟民
广 东 医 药 学 院 饶裕珍 陈昭华
中 国 药 科 大 学 倪永兴

前　　言

概率论和数理统计都是研究随机现象的数量规律性的学科，在医学、药学及卫生科技工作中有着广泛的应用。我国高等医药院校中的药学、中药、生物医学工程，以及部分临床医学、卫生、检验、口腔等专业，已为本科生或研究生开设了概率统计课程。我们九所院校正是鉴于教学工作的需要而协作编写了这本教材。讲授本书全部内容约需72学时，其中基本部分（不打*号的）约需50学时。

本书根据医学、药学、卫生及生物医学工程科研工作的需要，结合医药科技的实际背景，叙述概率论的基本知识及数理统计的常用方法，包括：概率及其计算，随机变量和随机向量，抽样分布和抽样估计，假设检验与方差分析，正交试验的设计和分析，相关与回归分析，质量管理常用统计方法。编写中充分注意系统性、实用性和可接受性，除了可供高等医药院校各专业作教材使用以外，也可供医药卫生科技工作者自学或参考。

限于编者的教学经验和学识水平，书中或有不妥乃至错误之处，尚盼读者批评指正。

1989年6月

目　　录

第一章 随机事件与概率

第一节　随机事件及其运算

一、随机试验

自然现象和社会现象大致可分为两类：一类为确定性现象，另一类为随机性现象。

为了研究客观现象的规律性，我们在一定条件下对该现象进行观察、调查或实验(统称试验)。若在相同条件下试验的结果是确定的，则此类现象称为**确定性现象**。例如，电流通过导线，使导线周围产生磁场，这是确定性现象。另一种现象则不然。虽然在相同条件下进行试验，其结果可能是这样，也可能是那样，呈现偶然性；此类现象称为**随机性现象**。人们经过长期实践，发现在大量试验中随机性现象也具有明显的规律性。

下面举一些随机性现象的例子：

抛一枚硬币，出现正面、反面的情况；

某种药品对一种疾病的治疗效果；

某电话交换台一分钟内接到的呼唤次数；

电视显像管的使用寿命(小时)。

上面这些随机性现象在试验中具有下列三个特性：

1. 可在相同的条件下重复进行；
2. 出现多个(两个或两个以上)可能的结果；
3. 每次试验之前，不能肯定将会出现哪个结果。

具有上述三个特性的试验称为**随机试验**(random trial)，简称试验。我们就是通过随机试验来研究随机性现象内在的规律性的。

在一随机试验中，它的每一个可能出现的结果称为**基本事件**，或称**样本点**(sample point)。为了研究一随机试验 E，首先应明确该试验究竟有哪些不同的结果。随机试验 E 的基本事件(样本点）所组成的集合称为 E 的**完备事件组**，或称 E 的**样本空间**(sample space)，记为 U。

U 中的元素就是试验 E 的各个基本事件。

上面所举四个试验的样本空间可依次表为：

$U_1=\{$正面，反面$\}$

$U_2=\{$治愈，有效，无效$\}$

$U_3=\{0, 1, 2, 3, \cdots\cdots\}$

$U_4=\{x \mid 0 \leqslant x < \infty\}$

在具体问题中，构造样本空间是描述随机性现象的第一步。

二、随机事件

定义　随机事件 (random event，简称事件)A 是指样本空间 U 中的某个子集 A。称事件 A 发生，当且仅当子集 A 中的一个样本点出现。

例如在上述交换台的例子中，事件 $A=\{$一分钟内呼唤的次数不超过 4 次$\}$是由五个样本点：0，1，2，3，4 组成的。又如事件 $B=\{$一分钟内呼唤的次数多于 1 次少于 5 次$\}$是由样本点：2，3，4 组成的。

事件用大写字母 A，B，C，……表示。我们把样本空间 U 也作为一个事件，因为在每次试验中必然出现 U 中的某个样本点，也即 U 必然发生，所以常称 U 为**必然事件**。类似地，我们

把空集 V 也作为一个事件，它在每次试验中都不会发生，故称为**不可能事件**。

三、事件间的关系和运算

基本事件也称**简单事件**，一般所说的事件是指样本空间的某个子集，也称**复杂事件**。为了研究事件的性质，需要讨论事件的关系和运算。

设试验 E 的样本空间为 U，A，B，$A_k(k=1, 2, \cdots\cdots)$是 E 的事件。

1. 包含关系　若事件 A 发生必然导致事件 B 发生，则称事件 B **包含**事件 A，记为 $B\supset A$，或 $A\subset B$。

这可用图 1.1 来直观说明，图中长方形表示样本空间 U，圆 A 与圆 B 分别表示事件 A 与事件 B，事件 B 包含 A。这种示意图称为**文(Venn)图**。

2. 相等关系　若事件 B 包含事件 A，事件 A 也包含事件 B，即 $B\supset A$ 且 $A\supset B$，则称事件 A 与事件 B **相等**，记为 $A=B$。

3. 事件的并　事件 A 与事件 B 至少有一个发生，这一事件称为事件 A 与事件 B 的**并**(或**和**)，记为 $A\bigcup B$(或 $A+B$)。

图 1.2 中阴影部分即表示事件 A 与事件 B 的并。

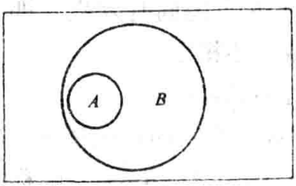

图 1.1　$A\subset B$

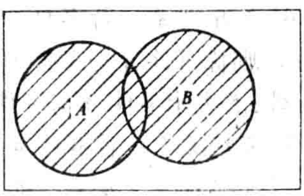

图 1.2　$A\bigcup B$

类似地，事件 A_1，A_2，…，A_n 中至少有一个发生，这一事件称为这 n 个事件的并(或和)，记为 $A_1 \bigcup A_2 \bigcup \cdots \bigcup A_n$（或 $A_1+A_2+\cdots+A_n$)，简记为 $\bigcup\limits_{k=1}^{n} A_k \left(或 \sum\limits_{k=1}^{n} A_k \right)$。

4. 事件的差　事件 A 发生而事件 B 不发生，这一事件称为事件 A 与事件 B 的**差**，记为 $A-B$（图 1.3)。

5. 事件的交　若事件 A 与事件 B 同时发生，则这一事件称为事件 A 与事件 B 的**交**（或积)，记为 $A \bigcap B$ 或 $A \cdot B$ 或 AB (图 1.4)。

事件 A_1，A_2，…，A_n 同时发生，这一事件称为这 n 个事件的交（或积)，记为 $A_1 \bigcap A_2 \bigcap \cdots \bigcap A_n$，(或 $A_1 A_2 \cdots A_n$)，简记为 $\bigcap\limits_{k=1}^{n} A_k \left(或 \prod\limits_{k=1}^{n} A_k \right)$。

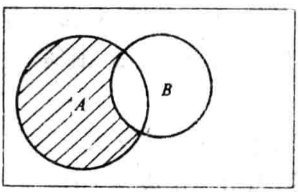

图 1.3　$A-B$

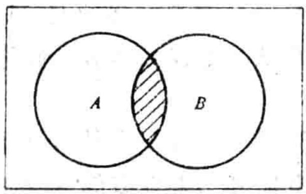

图 1.4　$A \bigcap B$

6. 互斥关系　若事件 A 与事件 B 不能同时发生，即 $A \bigcap B=V$，则称事件 A 与事件 B **互斥**(或**互不相容**)。

图 1.5 表示事件 A 与事件 B 是互不相容的。

7. 互逆关系　若事件 A 与事件 B 互斥，而且在任何一次试验中二者必有一个发生，即 $A \bigcap B=V$ 且 $A \bigcup B=U$，则称事件 A 与事件 B **互逆**(或**互相对立**)。A 的对立事件记为 $\overline{A}$(图 1.6)。

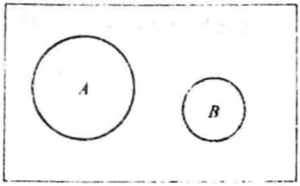

图 1.5 $A \cap B = V$

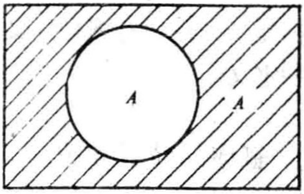

图 1.6 $\overline{A}$

从上面的讨论可以看到，概率论中事件间的关系和运算与集合论中集合间的关系和运算是一致的，见表 1.1。

表 1.1

记号	概率论	集合论
U	样本空间，必然事件	全集
V	不可能事件	空集
e	基本事件	元素
A	事件	子集
$\overline{A}$	A 的对立事件	A 的余集
$A \subset B$	事件 A 发生导致事件 B 发生	A 是 B 的子集
$A = B$	事件 A 与事件 B 相等	A 与 B 相等
$A \cup B$	事件 A 与事件 B 至少有一个发生	A 与 B 的并集
$A \cap B$	事件 A 与事件 B 同时发生	A 与 B 的交集
$A - B$	事件 A 发生而事件 B 不发生	A 与 B 的差集
$AB = V$	事件 A 与事件 B 互不相容	A 与 B 没有相同的元素

第二节　概率的定义

上面已指出，随机试验虽然是在一定条件下进行观察和实验，而一个随机事件却可能发生也可能不发生。那么能否对随机事件发生的可能性大小作出度量呢？

所谓事件的概率，粗糙地说，就是事件发生的可能性大小的一种度量。概率有多种定义方式，下面是科学技术工作中常用的两种定义：

一、概率的统计定义

1、频率

从频率来引出概率的统计定义如下：

定义 设在 n 次随机试验中，事件 A 出现 m 次，比值

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

称为事件 A 在 n 次试验中出现的**频率**(frequency)，m 称为**频数**。

医疗工作中所说的发病率、死亡率、治愈率等都是频率，常用百分数表示。显然，频率具有如下性质：

$$0\leqslant f_n(A)\leqslant 1 \tag{1.2}$$

例 1 在某地区 40 岁以上的居民中进行心血管疾病调查，结果在 4000 名男子中发现冠心病患者 132 人；在 1800 名女子中发现患该病的有 60 人。所以，该地区 40 岁以上男子患冠心病(记为 M) 的频率为

$$f\ (M)=\frac{132}{4000}=3.30\%$$

40 岁以上女子中患冠心病(记为 W)的频率为

$$f\ (W)=\frac{60}{1800}=3.33\%$$

2、频率的稳定性

例 2 掷一枚质量均匀的硬币，出现正面与反面的机会是相等的，即在大量重复试验中，出现正面的频率应接近于 0.5。为了验证这一点，历史上曾有几位数学家做过试验，结果见表 1.2。

表 1.2

实验者	掷硬币次数	出现正面次数	频率
狄摩根	2048	1061	0.5181
蒲　丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

例 3 瑞典 1935 年官方统计资料显示该年各个月份出生的男婴和女婴人数，以及生女频率见表 1.3。

表 1.3

月份	总　数	男　婴	女　婴	生女频率
1	7280	3743	3537	0.486
2	6957	3550	3407	0.490
3	7883	4017	3866	0.490
4	7884	4173	3711	0.471
5	7892	4117	3775	0.478
6	7609	3944	3665	0.482
7	7585	3964	3621	0.477
8	7393	3797	3596	0.486
9	7203	3712	3491	0.485
10	6903	3512	3391	0.491
11	6552	3392	3160	0.482
12	7132	3761	3371	0.473
全年	88273	45682	42591	0.4825

由表 1.3 可以发现一个有趣的现象，即各个月份的生女频率摆动于 0.4825 左右。