

普通高等教育“十三五”规划教材
全国高等医药院校规划教材

医药数理统计 学习辅导

第4版

汪旭生 曹 敏 主编



科学出版社

普通高等教育“十三五”规划教材
全国高等医药院校规划教材

医药数理统计 学习辅导

第4版

汪旭升 曹 敏 主编
周永治 主审

科学出版社
北 京

· 版权所有 侵权必究 ·

举报电话:010-64030229;010-64034315;13501151303(打假办)

内 容 简 介

本书是普通高等教育“十三五”规划教材、全国高等医药院校规划教材《医药数理统计》(第5版)的配套教材,也是本书的第4版。本书侧重于理论知识的归纳总结及各类习题的分析解法、各种医药实际问题的统计处理。本书针对教材给出了各章的内容提要、基本概念、习题解答和补充习题及解答。除此之外本书还提供了7套试题,对学生进一步的提高能力和备考有很大帮助,也便于学生自学。通过本书的学习,可帮助学生更好地掌握常用统计方法的运用,培养学生归纳总结能力、分析解决问题的能力。

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学习数理统计的参考书。

图书在版编目(CIP)数据

医药数理统计学习辅导 / 汪旭升, 曹敏主编. —4版. —北京: 科学出版社, 2015

普通高等教育“十三五”规划教材·全国高等医药院校规划教材
ISBN 978-7-03-046169-8

I. 医… II. ①汪… ②曹… III. 医用数学-数理统计-医学院校-教学参考资料 IV. R311

中国版本图书馆 CIP 数据核字(2015)第 258188 号

责任编辑:刘 亚 曹丽英 / 责任校对:李 影
责任印制:肖 兴 / 封面设计:大器设计工作室

版权所有, 违者必究。未经本社许可, 数字图书馆不得使用

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

安泰印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2004 年 9 月第 一 版 开本: 787×1092 1/16

2016 年 1 月第 四 版 印张: 8 1/2

2016 年 1 月第十四次印刷 字数: 195 000

定价: 24.90 元

(如有印装质量问题, 我社负责调换)

《医药数理统计学习辅导》(第4版)编写人员

主 编 汪旭升 曹 敏

副主编 钱微微 赵聪俐 沈晓婧 郝 涛

黄爱武 王世钦 赵文峰 沈宗山

主 审 周永治

编 委 (按姓氏笔画排序)

丁敏敏 安徽中医药大学

王 剑 广西中医药大学

王世钦 甘肃中医学院

尹立群 天津中医药大学

吕佳萍 南京中医药大学

李晓红 浙江中医药大学

吴 刚 贵阳中医学院

汪旭升 广西中医药大学

沈宗山 云南中医学院

沈晓婧 南京中医药大学

林剑鸣 广州中医药大学

赵文峰 河南中医药大学

赵聪俐 天津中医药大学

郝 涛 山东中医药大学

钱微微 浙江中医药大学

高 云 山东中医药大学

黄 浩 福建中医药大学

黄 翔 安徽中医药大学

黄爱武 湖南中医药大学

曹 敏 贵阳中医学院

第4版编写说明

《医药高等数学》、《医药数理统计》、《医药数学实验》是全国19所中医院校联合编写的科学出版社2001年4月出版的数学系列教材。相继,2004年8月《医药高等数学》、《医药数理统计》第2版出版,同时由《医药数学实验》转换的《医药高等数学学习辅导》、《医药数理统计学习辅导》第1版出版(辅导教材比配套的理论教材推迟一版);2009年5月出版第3版理论教材与第2版辅导教材;2012年5月出版第4版理论教材与第3版辅导教材(2012年的教材为“十二五”规划教材)。本套教材自2001年出版以来,发行面广,发行量大,在中医院校受到广大师生的欢迎。编写组根据科学出版社对普通高等教育规划教材的具体要求,对前版教材进行全面的分析、总结,认真进行修改、补充,编写成“十三五”规划教材(第5版的《医药高等数学》、《医药数理统计》与第4版的《医药高等数学学习辅导》、《医药数理统计学习辅导》)。本配套教材将更适应医药院校的医药类、管理类、信息类、人文类等专业的需要,定于2016年1月由科学出版社正式出版。

《医药数理统计学习辅导》(第4版)是《医药数理统计》(第5版)的配套教材,相应地也有9章。每章包括四大部分:一、内容提要;二、基本概念;三、习题解答(该章习题的解答过程);四、补充习题及解答(增补一些有代表性有适当难度的习题)。书的最后编入一些院校有代表性的试卷,供学生练习。本辅导教材有利于学生掌握统计方法的原理与步骤,帮助学生学好数理统计同时培养自己分析解决问题的能力,也有利于教师的教学工作。

参加本教材编写的有以下院校:贵阳中医学院、广西中医药大学、广州中医药大学、湖南中医药大学、湖北中医药大学、福建中医药大学、浙江中医药大学、南京中医药大学、安徽中医药大学、天津中医药大学、山东中医药大学、甘肃中医学院、河南中医药大学、陕西中医药大学。

本教材编写过程中得到许多同行专家的关心与支持,在此一并表示感谢。

本教材尚有不足之处,恳请读者与同行批评指正。

编者

2015年11月

目 录

第 4 版编写说明

第一章 事件与概率	(1)
第二章 随机变量的概率分布与数字特征	(9)
第三章 随机抽样和抽样分布	(24)
第四章 总体参数的估计	(32)
第五章 总体参数的假设检验	(40)
第六章 方差分析	(57)
第七章 非参数检验	(73)
第八章 相关与回归	(81)
第九章 正交试验设计	(94)
医药数理统计试题及答案	(111)

事件与概率

一、内容提要

了解样本空间的概念,理解随机事件的概念,掌握事件之间的关系和运算;了解事件频率的概念,了解概率的统计定义,掌握古典型概率的计算;了解条件概率的概念,理解事件独立性的概念,掌握利用事件的独立性进行概率计算.掌握概率的加法公式、乘法公式、全概率公式及贝叶斯公式.

二、基本概念

(一) 随机事件及其关系和运算

1. 随机现象→随机试验→随机事件.
2. 事件的关系和运算.

事件的关系和运算主要有:用简单事件表示复杂事件;化简事件的关系式;证明事件之间的某些等式或不等式.

(1) 四种关系,如表 1-1 所示:

表 1-1

关系	符号	概率论的定义	集合论的含义
包含关系	$A \subset B$	事件 A 发生必然导致 B 发生	A 是 B 的子集
相等	$A=B$	$A \subset B$ 且 $B \subset A$	A 与 B 相等
对立关系	$\bar{A}$	A 所构成的事件不发生	A 的补集
互不相容(或互斥)	$AB=\emptyset$	事件 A 与 B 不能同时发生	A 与 B 没有公共元素

(2) 事件的运算服从下列规律.

- 交换律: $A+B=B+A, AB=BA$;
 结合律: $(A+B)+C=A+(B+C), (AB)C=A(BC)$;
 分配律: $(A+B)(A+C)=A+BC, A(B+C)=AB+AC$;
 吸收律: $A+AB=A, A(A+B)=A$;
 补余律: $A+\bar{A}=\Omega, A\bar{A}=\emptyset$;

De-Morgan 律:对有限个或可列无限个事件 A_i 恒有

$$\overline{\sum_i A_i} = \prod_i \bar{A_i} \Leftrightarrow \overline{\bigcup_i A_i} = \bigcap_i \bar{A_i}, \quad \overline{\bigcap_i A_i} = \bigcup_i \bar{A_i} \Leftrightarrow \overline{\prod_i A_i} = \sum_i \bar{A_i}$$

(二) 事件频率、概率的统计定义、古典型概率的计算

1. 概率的定义:古典概率、几何概率、统计概率、公式化定义.
2. 古典概率计算的要点.

古典概率计算的要点:给定基本事件的总数,然后再计算事件 A 中包含的基本事件数,这就归结为计数问题.计数的基本工具主要有两个:基本原理和排列组合方法.

$$(1) \text{ 排列: } P_n^m = n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!}$$

$$n! = n(n-1)(n-2)\cdots 2 \cdot 1$$

$$(2) \text{ 组合: } C_n^m = \frac{P_n^m}{m!} = \frac{n!}{m!(n-m)!}$$

3. 古典概型概率解题时应注意的若干事项:

(1) 所求中有“至少”的问题,通常用“对立事件”解答较简便.

(2) “任取 k 件”与“无放回地逐件抽取 k 件”,虽然考虑问题的角度不同,但二者所计算出的概率都是相同的.

(3) “任取 k 件”与“有放回地逐件抽取 k 件”,所得概率一般是不同的.

(三) 条件概率、概率的加法公式、乘法公式、全概率公式及贝叶斯公式

在求解较为复杂的条件概率问题时,还需要灵活运用下面三个重要的公式:

1. 如果所求概率是任意 n 个事件 $A_1, A_2, \dots, A_n$ 的交事件的概率且 $P(A_1 A_2 \cdots A_{n-1}) > 0$, 则可应用乘法公式

$$P(A_1 A_2 \cdots A_{n-1} A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \cdots P(A_n | A_1 A_2 \cdots A_{n-1})$$

求解.

2. 如果某一结果(事件 A)是由多种“原因”事件 $B_i (i=1, 2, \dots, n)$ 所引起的,并且作为“原因”的这些事件彼此间互不相容,其和事件恰为必然事件,则结果 A 发生的概率可由全概率公式

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i)$$

求解,其中, $P(B_i) > 0$.

3. 如果某一事件 A 的发生是由多种“原因” $B_i (i=1, 2, \dots, n)$ 所引起的,并且知事件 A 已发生,当需要了解 A 的发生是由某 B_k 所引起的概率有多大时,可按贝叶斯公式

$$P(B_k | A) = \frac{P(B_k)P(A | B_k)}{\sum_{i=1}^n P(B_i)P(A | B_i)}$$

计算,其中, $P(B_i) > 0, P(A) > 0$.

(四) 事件的独立性

独立性是概率论中应用极为广泛的重要概念.就解而言,事件的独立性有助于简化概率计算.

1. 计算相互独立事件的积的概率,可简化为

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

2. 计算相互独立事件的和的概率,可简化为

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - P(\bar{A}_1)P(\bar{A}_2) \cdots P(\bar{A}_n)$$

三、习题一解答

1. 设 A, B, C 为三事件,用 A, B, C 的运算关系表示下列事件:

- (1) A 发生, B 与 C 不发生;
- (2) A 与 B 都发生,而 C 不发生;
- (3) A, B, C 都发生;
- (4) A, B, C 中至少有一个发生;
- (5) A, B, C 都不发生;
- (6) A, B, C 中不多于一个发生;
- (7) A, B, C 中不多于两个发生;
- (8) A, B, C 中至少有两个发生.

解 (1) ABC

(2) ABC

(3) ABC

$$(4) \overline{ABC} + \overline{AB}C + \overline{A}BC + \overline{ABC} + \overline{ABC} + \overline{ABC} + \overline{ABC} \\ = A + B + C$$

$$(5) \overline{A} \overline{B} \overline{C}$$

$$(6) \overline{AB} \overline{C} + \overline{AB}C + \overline{A}BC + \overline{A}B \overline{C} \text{ 或 } \overline{AB} + \overline{BC} + \overline{AC}$$

$$(7) \overline{AB} \overline{C} + \overline{AB}C + \overline{A}BC + \overline{ABC} + \overline{ABC} + \overline{ABC} + \overline{A}B \overline{C} \text{ 或 } \overline{ABC}$$

$$(8) AB + BC + AC$$

2. 对三人做舌诊, 设 $A = \{\text{三人正常}\}$, $B = \{\text{至少一人不正常}\}$, $C = \{\text{只有一人正常}\}$, $D = \{\text{只有一人不正常}\}$ 指出这四个事件中的互斥事件、对立事件, $A+D$, BD 各表示什么意思.

解 A 与 B , A 与 C , A 与 D , C 与 D 是互斥事件.

因为 $A+B = \Omega$, $AB = \emptyset$, 所以 A 与 B 是对立事件.

$$A+D = \{\text{至少有两人正常}\} = \{\text{至多一人不正常}\}$$

$$BD = D = \{\text{只有一人不正常}\} = \{\text{恰有两人正常}\}$$

3. 某市在某年的第一季度出生婴儿的情况为一月份男孩 145 个, 女孩 135 个; 二月份男孩 125 个, 女孩 136 个; 三月份男孩 152 个, 女孩 140 个, 问该季度生男孩的频率是多少?

解 第一季度共出生婴儿数为

$$145 + 135 + 125 + 136 + 152 + 140 = 833$$

该季度出生的男孩数为

$$145 + 125 + 152 = 422$$

因此该季度生男孩的频率为

$$f = \frac{422}{833} = 0.5066$$

4. 40 个药丸中 3 丸已失效, 现任取 5 丸, 求其中有 2 丸失效的概率.

解 $A = \{\text{任取 5 丸, 其中有 2 丸失效}\}$

$$P(A) = \frac{C_{37}^3 \times C_3^2}{C_{40}^5} = \frac{\frac{37 \times 36 \times 35}{3 \times 2 \times 1} \times \frac{3 \times 2}{2 \times 1}}{\frac{40 \times 39 \times 38 \times 37 \times 36}{5 \times 4 \times 3 \times 2 \times 1}} = 0.0354$$

5. 一批针剂共 100 支, 其中, 有 10 支次品, 求

(1) 这批针剂的次品率;

(2) 从中任取 5 支, 全部是次品的概率;

(3) 从中任取 5 支, 恰有 2 支次品的概率.

解 (1) $A = \{\text{次品}\}$

$$P(A) = \frac{10}{100} = 0.1$$

(2) $B = \{\text{任取 5 支, 全部是次品}\}$

$$P(B) = \frac{C_{10}^5}{C_{100}^5} = 0.000003347$$

(3) $C = \{\text{任取 5 支, 恰有两支次品}\}$

$$P(C) = \frac{C_{90}^3 \times C_{10}^2}{C_{100}^5} = 0.07$$

6. 某地居民血型分布为 $P(\text{O 型}) = 50\%$, $P(\text{A 型}) = 14.5\%$, $P(\text{B 型}) = 31.2\%$, $P(\text{AB 型}) = 4.3\%$, 若有一个 A 型血型患者需要输血, 问当地居民任一人可为他输血的概率是多少?

解 $\Omega = \{\text{O 型}\} + \{\text{A 型}\} + \{\text{B 型}\} + \{\text{AB 型}\}$

设

$$A = \{\text{当地居民为一个 A 型血型患者输血}\}$$

则

$$\begin{aligned} P(A) &= P(\{\text{O 型}\} + \{\text{A 型}\}) \\ &= P(\text{O 型}) + P(\text{A 型}) \\ &= 0.5 + 0.145 = 0.645 \end{aligned}$$

7. 药房有包装相同的六味地黄丸 100 盒, 其中, 5 盒为去年产品, 95 盒为今年产品. 现随机发出 4 盒, 求

(1) 有 1 盒或 2 盒陈药的概率;

(2) 有陈药的概率.

解 样本总数为 C_{100}^4 ,

(1) 设 $A = \{\text{有 1 盒或 2 盒陈药}\}$, 则

$$P(A) = \frac{C_{95}^3 \times C_5^1 + C_{95}^2 \times C_5^2}{C_{100}^4} = 0.1879$$

(2) 设 $B = \{\text{有存药}\}$, $\bar{B} = \{\text{无存药}\}$, 则

$$P(\bar{B}) = \frac{C_{95}^4}{C_{100}^4} = 0.8119$$

所以

$$P(B) = 1 - P(\bar{B}) = 1 - 0.8119 = 0.1881$$

8. 从 1, 2, 3, 4, 5 号小白鼠中任取两只做新药试验, 计算所取两只中一只是 4 号小白鼠的概率.

解 设 $A = \{\text{所取两只中一只是 4 号小白鼠}\}$. 样本空间中基本事件的总数为 $n = C_5^2 = 10$, 事件 A 所包含的基本事件数为 $m = C_1^1 \cdot C_4^1 = 4$, 所以

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

9. 某药检所从送检的 10 件药品中先后抽取了两件. 如果 10 件中有三件是次品.

(1) 求第一次检得次品的概率?

(2) 第一次检得次品后, 第二次检得次品的概率?

(3) 两次都检得次品的概率.

解 设 $A_i = \{\text{第 } i \text{ 次所取的药品是次品}\}$, $i = 1, 2$,

$$(1) P(A_1) = \frac{3}{10}$$

$$(2) P(A_2 | A_1) = \frac{3-1}{10-1} = \frac{2}{9}$$

(3) 根据概率的乘法公式得

$$P(A_1 A_2) = P(A_1) \cdot P(A_2 | A_1) = \frac{3}{10} \times \frac{2}{9} = 0.0667$$

10. 某厂生产的产品中, 36% 为一等品, 54% 为二等品, 10% 为三等品, 任取一件产品, 已知它不是三等品, 求它是一等品的概率.

解 设 $A_1 = \{\text{一等品}\}$, $A_2 = \{\text{二等品}\}$, $A_3 = \{\text{三等品}\}$, 则

$$P(A_1) = 0.36, \quad P(A_2) = 0.54, \quad P(A_3) = 0.1$$

设 $B = \{\text{任取一件产品不是三等品, 是一等品}\}$, 由题意知

$$P(B) = \frac{P(A_1)}{P(A_1) + P(A_2)} = \frac{0.36}{0.36 + 0.54} = 0.4$$

11. 经调查, 在 50 个聋耳人中有 4 人色盲, 在 950 个非聋耳人中有 76 人色盲, 试说明聋耳与色盲无关.

解 设 $A = \{\text{色盲}\}$, $B = \{\text{聋耳}\}$, 则

$$P(A) = \frac{80}{1000} = 0.08, \quad P(A | B) = \frac{4}{50} = 0.08$$

可见 $P(A) = P(A | B)$, A 与 B 相互独立, 即聋哑与色盲无关.

12. 假如某人群中患结核病的概率为 0.003, 患沙眼的概率为 0.04, 现从该人群中任意抽查一人, 求下列事件的概率:

(1) 此人患结核病且患沙眼病;

(2) 此人既无结核病又无沙眼病;

(3) 此人至少有这两种病的一种;

(4) 此人只有其中一种病.

解 设 $A = \{\text{患结核病}\}$, $B = \{\text{患沙眼}\}$. 由题意知 A 与 B 是相互独立事件.

$$P(A) = 0.003, \quad P(B) = 0.04$$

$$(1) P(AB) = P(A)P(B) = 0.00012$$

$$(2) P(\overline{A+B}) = 1 - P(A+B) = 1 + P(AB) - P(A) - P(B) \\ = 0.9571$$

$$(3) P(A+B) = P(A) + P(B) - P(AB) \\ = 0.003 + 0.04 - 0.00012 \\ = 0.0429$$

$$(4) P(\overline{AB} + A\overline{B}) = P(\overline{A})P(B) + P(A)P(\overline{B}) \\ = 0.0428$$

13. 设 $A = \{\text{甲市有雨}\}$, $B = \{\text{乙市有雨}\}$, 由以往的气象记录知 $P(A) = 0.3$, $P(B) = 0.4$ 且 $P(AB) = 0.28$,

(1) 说明两市下雨有牵连(非独立);

(2) 求 $P(A|B)$, $P(B|A)$, $P(A+B)$.

(注意: A, B 不互斥也不独立.)

解 (1) 因为

$$P(AB) = 0.28, \quad P(A)P(B) = 0.12$$

所以

$$P(AB) \neq P(A)P(B)$$

$$(2) P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.28}{0.4} = 0.7$$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.28}{0.3} = 0.933$$

$$P(A+B) = P(A) + P(B) - P(AB) = 0.42$$

14. 设某产品进行验收检查, 发现次品率为 0.02.

(1) 今独立地检验 100 件产品, 问至少发现一件产品为次品的概率是多少?

(2) 如保证至少发现一件次品的概率为 0.9, 问应检验多少件产品?

解 (1) 令 $A_i = \{\text{第 } i \text{ 件是次品}\}$, 那么 $\overline{A_i} = \{\text{第 } i \text{ 件是合格品}\}$, $i = 1, 2, \dots, 100$.

$$P(A_i) = 0.02, \quad P(\overline{A_i}) = 1 - P(A_i) = 0.98$$

因为 100 个事件 $\overline{A_1}, \dots, \overline{A_{100}}$ 独立, 所以发现无次品的概率为

$$P\{\text{发现无次品}\} = P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_{100}}) \\ = 0.98^{100} \\ = 0.13$$

$$P\{\text{至少发现一件产品为次品}\} = 1 - P\{\text{发现无次品}\} \\ = 0.8674$$

(2) $P\{\text{至少发现一件产品为次品}\} = 0.9$, 则

$$P\{\text{发现无次品}\} = 0.1$$

即 $0.98^n = 0.1$,

$$n = \frac{\ln 0.1}{\ln 0.98} \approx 114$$

15. 三家工厂生产同一种产品, 每家厂商分别占总产量的 25%, 35%, 40%, 又知每厂的次品率分别为 5%, 4%, 2%, 求从这种产品中取一件, 取到次品的概率.

解 设 $B = \{\text{取到次品}\}$, $A_i = \{\text{取到的产品是属于第 } i \text{ 家工厂生产的}\}$, $i = 1, 2, 3$.

$$P(A_1) = 0.25, \quad P(A_2) = 0.35, \quad P(A_3) = 0.4$$

$$P(B|A_1) = 5\%, \quad P(B|A_2) = 4\%, \quad P(B|A_3) = 2\%$$

$$P(B) = \sum_{i=1}^3 P(A_i)P(B|A_i) \\ = 0.25 \times 0.05 + 0.35 \times 0.04 + 0.4 \times 0.02 \\ = 0.0345$$

16. 仓库里有 10 箱规格相同的产品, 已知其中有 5 箱、3 箱、2 箱依次是甲厂、乙厂、丙厂生产的, 且甲厂、乙

厂、丙厂的产品次品率分别为 $1/10, 1/15, 1/20$, 从这 10 箱中取 1 箱, 再从中任取 1 件产品, 求取得正品的概率.

解 $B = \{\text{次品}\}, A_1 = \{\text{取的箱子是甲厂的}\}, A_2 = \{\text{取的箱子是乙厂的}\}, A_3 = \{\text{取的箱子是丙厂的}\}.$

$$P(A_1) = 0.5, \quad P(A_2) = 0.3, \quad P(A_3) = 0.2$$

$$P(B | A_1) = \frac{1}{10}, \quad P(B | A_2) = \frac{1}{15}, \quad P(B | A_3) = \frac{1}{20}$$

$$P(B) = \sum_{i=1}^3 P(A_i)P(B | A_i) = 0.5 \times \frac{1}{10} + 0.3 \times \frac{1}{15} + 0.2 \times \frac{1}{20} = 0.08$$

$$P(\bar{B}) = 1 - P(B) = 0.92$$

即其概率为 0.92.

17. 把甲乙两种外观一样、数量相等的药片混在一起, 若甲种药片的次品率为 0.05, 乙种药片的次品率为 0.0025, 现从中抽出 1 片发现是次品, 求该药片来自甲、乙种的概率.

解 $A_1 = \{\text{甲种药片}\}, A_2 = \{\text{乙种药片}\}, B = \{\text{次品}\}$ 则

$$P(A_1) = P(A_2) = 0.5$$

$$P(B | A_1) = 0.05, \quad P(B | A_2) = 0.0025$$

$$\begin{aligned} P(B) &= \sum_{i=1}^2 P(A_i)P(B | A_i) \\ &= 0.5 \times 0.05 + 0.5 \times 0.0025 \\ &= 0.02625 \end{aligned}$$

所以

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(B)} = 0.9524$$

$$P(A_2 | B) = \frac{P(A_2)P(B | A_2)}{P(B)} = 0.04761$$

18. 已知一批产品中 96% 是合格品, 检查时, 一个合格品误认为不合格的概率是 0.02, 一个不合格品误认为合格的概率是 0.05, 求在检查合格的产品中确是合格品的概率.

解 设 $A = \{\text{合格}\}, B = \{\text{被判合格}\}$, 则

$$P(A) = 0.96, \quad P(\bar{A}) = 0.04, \quad P(\bar{B} | A) = 0.02$$

$$P(B | A) = 1 - P(\bar{B} | A) = 0.98$$

$$P(B | \bar{A}) = 0.05$$

由贝叶斯公式, 被合格的产品确定有合格产品的概率为

$$\begin{aligned} P(A | B) &= \frac{P(A) \cdot P(B | A)}{P(A) \cdot P(B | A) + P(\bar{A}) \cdot P(B | \bar{A})} \\ &= \frac{0.96 \times 0.98}{0.96 \times 0.98 + 0.04 \times 0.05} \\ &= 0.9979 \end{aligned}$$

19. 用 X 线透视诊断肺结核, 设 $A = \{\text{实有肺结核}\}, B = \{\text{被判有肺结核}\}$. 若某市成人中 $P(A) = 0.001$, 这种检查阳性的正确率 $P(B | A) = 0.95$, 阴性的正确率 $P(\bar{B} | \bar{A}) = 0.998$.

(1) 求该市一人经透视被判有肺结核的概率;

(2) 若一个经透视被判有肺结核, 求他实际患有肺结核的概率.

解 设 $A = \{\text{实有肺结核}\}, B = \{\text{被判有肺结核}\}$. 由题意

$$P(A) = 0.001, \quad P(B | A) = 0.95, \quad P(\bar{A}) = 0.999, \quad P(\bar{B} | \bar{A}) = 0.998, \quad P(B | \bar{A}) = 0.002$$

A 与 $\bar{A}$ 构成互斥完备群,

$$\begin{aligned} (1) P(B) &= P(A) \cdot P(B | A) + P(\bar{A}) \cdot P(B | \bar{A}) \\ &= 0.002948 \end{aligned}$$

$$\begin{aligned} (2) P(A | B) &= \frac{P(A) \cdot P(B | A)}{P(A) \cdot P(B | A) + P(\bar{A}) \cdot P(B | \bar{A})} \\ &= \frac{0.001 \times 0.95}{0.001 \times 0.95 + 0.999 \times 0.002} \\ &= 0.3223 \end{aligned}$$

20. 某电子设备厂所用的晶体管由甲乙丙三家元件制造厂提供. 已知甲乙丙三厂的次品率分别为 0.02, 0.01, 0.03, 又知三个厂提供晶体管的份额分别为 0.15, 0.80, 0.05, 设三个厂的产品是同规格的(无区别标志)且均匀的混合在一起. 求 1. 在混合的晶体管中随机取一支是次品的概率. 2. 现在抽取了一支晶体管次品, 问这支晶体管最可能是哪个厂家生产的?

解 A_i $i=1, 2, 3$ 代表抽到三个厂家生产的晶体管, B 代表抽到的是次品.

$$P(B) = \sum_{i=1}^3 P(A_i)P(B | A_i) = 0.02 \times 0.15 + 0.01 \times 0.80 + 0.03 \times 0.05 = 0.0125$$

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{\sum_{i=1}^3 P(A_i)P(B | A_i)} = 0.24, \quad P(A_2 | B) = \frac{P(A_2)P(B | A_2)}{\sum_{i=1}^3 P(A_i)P(B | A_i)} = 0.64$$

$$P(A_3 | B) = \frac{P(A_3)P(B | A_3)}{\sum_{i=1}^3 P(A_i)P(B | A_i)} = 0.12.$$

所以最有可能是乙厂生产的。

四、补充习题及解答

1. 盒子中盛有 12 个乒乓球, 其中, 9 个是新球, 3 个是旧球. 练球时第一次从盒子中任取 3 个来用(新球用一次后就成为旧球), 用后仍放回盒子中, 第二次再从盒子中任取 3 个球, 试求第二次取出的球都是新球的概率.

解 设 $B = \{\text{第二次取出的球都是新球}\}$, $A_i = \{\text{第一次取出的球中有 } i \text{ 个新球}\}$, $i = 0, 1, 2, 3$, 则 A_0, A_1, A_2, A_3 是一个完备事件组且

$$P(A_0) = \frac{C_3^3}{C_{12}^3} = \frac{1}{220}, \quad P(A_2) = \frac{C_9^2 C_3^1}{C_{12}^3} = \frac{108}{220}$$

$$P(A_1) = \frac{C_9^1 C_3^2}{C_{12}^3} = \frac{27}{220}, \quad P(A_3) = \frac{C_3^3 C_9^0}{C_{12}^3} = \frac{84}{220}$$

$$P(B | A_0) = \frac{C_9^3}{C_{12}^3} = \frac{84}{220}$$

$$P(B | A_1) = \frac{C_8^3}{C_{12}^3} = \frac{56}{220}$$

$$P(B | A_2) = \frac{C_7^3}{C_{12}^3} = \frac{35}{220}$$

$$P(B | A_3) = \frac{C_6^3}{C_{12}^3} = \frac{20}{220}$$

由全概率公式可得

$$P(B) = P(A_0)P(B | A_0) + P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + P(A_3)P(B | A_3)$$

$$= \frac{1}{220} \cdot \frac{84}{220} + \frac{27}{220} \cdot \frac{56}{220} + \frac{108}{220} \cdot \frac{35}{220} + \frac{84}{220} \cdot \frac{20}{220} = \frac{441}{3025} = 0.1458$$

2. 某产品设计长度为 20cm, 规定误差不超过 0.5cm 为合格产品. 今对一批产品尽心测量, 长度如表 1-2 所示:

表 1-2

长度/cm	19.5 以下	19.5~20.5	20.5 以上
件数	5	68	7

计算这批产品的合格率.

解 根据设计要求知长度在 19.5cm 以下和 20.5cm 以上的均为不合格品, 这批产品的合格率为

$$p = \frac{68}{5 + 68 + 7} = 0.85$$

3. 掷三枚硬币,求出现三个正面的概率.

解 掷一枚硬币,记出正面为 H,反面为 T,则掷三枚硬币试验的样本空间为 $\Omega = \{HHH, THT, HTT, TTH, HHT, HTH, THH, TTT\}$, 出现三个正面的事件记为 A, 则 $A = \{HHH\}$, 于是

$$P(A) = \frac{1}{8} = 0.125$$

4. 10 把钥匙中有 3 把能打开门,今任取 2 把,求能打开门的概率.

解法一 随机试验是从 10 把钥匙里任取 2 把,从样本空间 Ω 的样本点总数为

$$n = C_{10}^2 = 45$$

要想把门打开,取出的 2 把钥匙中至少有一把从能把门打开的 3 把钥匙中获得,从而能把门打开这一事件所包含的样本点数为 $m = C_3^2 + C_3^1 C_7^1 = 24$. 故所求概率为

$$p = \frac{m}{n} = \frac{24}{45} = \frac{8}{15} \approx 0.53$$

解法二 随机试验是从 10 把钥匙中任取 2 把,从而样本空间 Ω 的样本点总数为

$$n = C_{10}^2 = 45$$

记事件 $A = \{\text{能把门打开}\}$, 则 $\bar{A} = \{\text{不能把门打开}\}$, 从 7 把不能把门打开的钥匙中任取 2 把, 共有 $C_7^2 = 21$ 种取法, 即事件 $\bar{A}$ 共包含 21 个样本点, 从而

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{21}{45} = \frac{24}{45} \approx 0.53$$

5. 一部 4 卷的文集随便放在书架上,问恰好各卷自左向右或自右向左的卷号为 1, 2, 3, 4 的概率是多少?

解 一部 4 卷的文集随便放在书架上共有 $4 \times 3 \times 2 \times 1 = 24$ 种放法, 而文集恰好各卷自左向右或自右向左的卷号为 1, 2, 3, 4 共有两种放法, 从而所求概率为

$$P = \frac{2}{24} = \frac{1}{12} \approx 0.083$$

6. 100 个产品中有 3 个次品,任取 5 个,求其中次品数分别为 0, 1, 2, 3 的概率.

解 随机试验是从 100 个产品中任取 5 个,样本空间所包含的样本点总数为 $n = C_{100}^5$. 记事件 $A_i = \{\text{取出的 5 个产品中含有 } i \text{ 个次品}\}$, $i = 0, 1, 2, 3$. 若取出的 5 个产品中都没有次品,则取出的 5 个产品都必须从 97 个合格品种获得,从而事件 A_0 所包含的样本点总数为 $m_0 = C_{97}^5$, 故

$$P(A_0) = \frac{m_0}{n} = \frac{C_{97}^5}{C_{100}^5} \approx 0.856$$

同理,若取出的 5 个产品中含有 i 个次品,则 i 个次品必须从 3 个次品中获得, $5-i$ 个合格品必须从 97 个合格品种获得,从而事件 A_i 所包含的样本点数为 $m_i = C_3^i C_{97}^{5-i}$, $i = 1, 2, 3$. 故

$$P(A_1) = \frac{C_3^1 C_{97}^4}{C_{100}^5} \approx 0.138$$

$$P(A_2) = \frac{C_3^2 C_{97}^3}{C_{100}^5} \approx 0.006$$

$$P(A_3) = \frac{C_3^3 C_{97}^2}{C_{100}^5} \approx 0.00006$$

随机变量的概率分布与数字特征

一、内容提要

本章主要介绍用随机变量描述各种随机现象的方法,这是概率论与数理统计的重点.随机变量常见的有两类:离散型与非离散型(主要指连续型的),它们分别用概率函数与概率密度函数描述其概率分布,而又以分布函数的形式统一起来.本章还讲述了几何分布、二项分布、泊松分布、正态分布、对数正态分布、均匀分布、指数分布的概率分布、数字特征及有关概率的计算方法.

二项分布、泊松分布、正态分布是常用的三种分布,而正态分布无论在实际应用还是理论研究中都占有头等重要的地位.

二、基本概念

(一) 随机变量与分布函数

1. 随机变量:设 Ω 是试验 E 的样本空间,对于每一个样本点 $\omega \in \Omega$,都有唯一一个实数 $X(\omega)$ 与之对应且对于任意实数 x

$$\{\omega \mid X(\omega) \leq x\}$$

都是试验 E 的事件,存在着概率,则称 $X(\omega)$ 为随机变量,简记为 X .

注 随机变量是定义域为样本空间的函数,自变量是样本点 ω ,而 ω 未必是数.在每次试验之前,只知道随机变量都可能取哪些值,但不能预知它取什么值;每一次试验,随机变量在某一确定范围中取值的概率是确定的.

随机变量常用 $X, Y, Z, \dots$ 表示.

2. 分布函数.

函数 $F(x) = P(X \leq x) (-\infty < x < +\infty)$ 称为随机变量 X 的分布函数,它具有以下性质:

- (1) $0 \leq F(x) \leq 1, -\infty < x < +\infty$.
- (2) $F(x)$ 是单调不减函数,即若 $x_1 \leq x_2$, 则有 $F(x_1) \leq F(x_2)$.
- (3) $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$.
- (4) $F(x)$ 是右连续的, $\lim_{t \rightarrow x^+} F(t) = F(x+0) = F(x)$.

反之,具有上述性质的函数 $F(x)$ 一定是某随机变量的分布函数.

若已知随机变量 X 的分布函数 $F(x)$, 则

$$\begin{aligned} P(x_1 < X \leq x_2) &= P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1) \\ P(X > x) &= 1 - P(X \leq x) = 1 - F(x) \end{aligned}$$

实际中常遇到的随机变量是离散型和连续型两类随机变量.

(二) 离散型随机变量

1. 离散型随机变量:若随机变量 X 的取值为有限个或可列无限个,则称 X 为离散型随机变量.

2. 概率分布:

(1) 概率函数:设离散型随机变量 X 的取值为 $x_1, x_2, \dots, x_k, \dots$, X 取各个可能值的概率为 $P(X=x_k)=p_k, k=1, 2, \dots$, 则称其为离散型随机变量的概率函数(又称分布率).

(2) 分布列:概率函数用列表的方式给出,如表 2-1 所示:

表 2-1

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P_k	p_1	p_2	$\cdots$	p_i	$\cdots$

该表格称为 X 的分布列,它还可以用矩阵形式来简记

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_i & \cdots \\ p_1 & p_2 & \cdots & p_i & \cdots \end{pmatrix}$$

概率函数满足下列性质:

$$p_k \geqslant 0, k = 1, 2, \cdots, \sum_k p_k = 1$$

(3) 分布函数: X 为离散型随机变量,则有分布函数为

$$F(x) = P(X \leqslant x) = \sum_{x_i \leqslant x} p_i$$

分布函数与概率函数的关系为

$$F(x_i) = P(X \leqslant x_i) = p_1 + p_2 + \cdots + p_i$$

$$p_i = P(X = x_i) = P(X \leqslant x_i) - P(X \leqslant x_{i-1}) = F(x_i) - F(x_{i-1})$$

3. 常见的离散型随机变量的概率分布(简称为分布):

(1) 二项分布:设随机变量 X 的分布为

$$P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, 2, \cdots, n; 0 < p < 1, q = 1 - p$$

则称 X 服从参数为 n, p 的二项分布,记 $X \sim B(n, p)$.

一般地,在 n 重伯努利试验中事件 A 恰好发生 $k (0 \leqslant k \leqslant n)$ 次的概率为 $P_n(k) = P(X = k) = C_n^k p^k q^{n-k}$, $k = 0, 1, 2, \cdots, n$

用 X 表示 n 重伯努利试验中事件 A 发生的次数,则 $X \sim B(n, p)$.

(2) 两点分布:设随机变量 X 的分布为

$$P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad 0 < p < 1$$

则称 X 服从参数为 p 的两点分布. 两点分布又称 1-0 分布,记 $X \sim B(1, p)$.

(3) 泊松分布:设随机变量 X 的分布为

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots, \lambda > 0$$

则称 X 服从参数为 λ 的泊松分布,记 $X \sim P(\lambda)$.

一般地, n 次独立重复试验中, n 很大而 p 很小时,事件 A 恰好发生的次数 X 就服从泊松分布,利用 e^x 的幂级数展开式,容易验证泊松分布的概率值满足

$$\sum_k p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

(4) 几何分布:设随机变量 X 的分布为 $P(X = k) = pq^{k-1} (k = 1, 2, \cdots, n; 0 < p < 1, q = 1 - p)$, 则称 X 服从参数为 p 的几何分布.

一般地,在伯努利试验中,事件 A 首次出现在第 k 次的概率为

$$p_k = pq^{k-1}, \quad k = 1, 2, \cdots$$

若用 X 记事件 A 首发生次数,则 X 服从几何分布.

(5) 超几何分布:设随机变量 X 的分布为

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = k_0, k_0 + 1, k_0 + 2, \cdots, l$$

其中, $n < N - M, k_0 = \max(0, n - m), l = \min(M, n)$, 则称 X 服从超几何分布.

(三) 连续型随机变量

1. 连续型随机变量:对于随机变量 X ,如果存在一个非负可积函数 $f(x) (-\infty < x < \infty)$, 使对任意 $a, b (a < b)$ 都有

$$P(a < x < b) = \int_a^b f(x) dx$$

则称 X 为连续型随机变量, 称 $f(x)$ 为 X 的概率密度函数, 有时称为概率密度或密度函数.

概率密度函数满足下面两个性质:

$$f(x) \geq 0, -\infty < x < +\infty, \int_{-\infty}^{+\infty} f(x) dx = 1$$

反之, 凡满足上面两个条件的函数 $f(x)$, 均可作为某连续型随机变量 X 的密度函数.

另外, 对连续型随机变量 X 有

(1) $P(X = x_0) = 0$, 即连续型随机变量取个别值的概率等于零.

(2) $P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 < X < x_2) = P(x_1 \leq X < x_2)$

$$= \int_{x_1}^{x_2} f(x) dx.$$

(3) $f(x) = F'(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x}$.

2. 常见的连续型随机变量的概率分布:

(1) 均匀分布: 设随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} \frac{1}{(b-a)}, & a \leq x \leq b \\ 0, & \text{其他} \end{cases}$$

则称 X 服从参数为 a, b 的均匀分布.

由定义显然有 $f(x) \geq 0$, $\int_{-\infty}^{+\infty} f(x) dx = \int_a^b \frac{1}{b-a} dx = 1$. 若 X 服从均匀分布, 那么对于任意的 $c, d (a \leq c < d \leq b)$, 按概率密度定义有

$$P(c < X < d) = \int_c^d f(x) dx = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$$

上式表明, X 落在 (a, b) 中任意长度相同为 $(d-c)$ 的子区间内的概率是相同的, 而与小区间的位置无关.

随机变量 X 服从均匀分布, 则其分布函数为

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b \end{cases}$$

(2) 标准正态分布: 若连续型随机变量 X 的概率密度为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty$$

则称 X 服从标准正态分布, 记 $X \sim N(0, 1)$, 其分布函数为

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

为计算方便, 对于不同的 x 值, 有标准正态分布函数值表可查. 此时有

$$\Phi(-x) = 1 - \Phi(x), \quad P(x_1 < X < x_2) = \Phi(x_2) - \Phi(x_1).$$

(3) 正态分布: 若连续型随机变量 X 概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

其中, μ, σ 为常数且 $\sigma > 0$, 则称 X 服从参数为 μ, σ^2 的正态分布, 记 $X \sim N(\mu, \sigma^2)$.

其分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

此时有 $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, 从而可查表计算 $F(x)$. 当 $X \sim N(\mu, \sigma^2)$ 时有

$$P(x_1 \leq X < x_2) = \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

特别地有 $P(|X-\mu| < 3\sigma) = 0.9973$, 称为正态分布的 3σ 规则.

(4) 指数分布: 若连续型随机变量 X 的概率密度为