

高等院校统计学精品课教材系列

概率论与数理统计

(第四版)

苏均和 主编



格致出版社
上海人民出版社

高等院校统计学精品课教材系列

概率论与数理统计

苏均和 主编

清华大学出版社

图书在版编目(CIP)数据

概率论与数理统计:第4版/苏均和主编. —上海:
格致出版社:上海人民出版社,2011
高等院校统计学精品课教材系列
ISBN 978-7-5432-2019-5

I. ①概… II. ①苏… III. ①概率论-高等学校-教材
②数理统计-高等学校-教材 IV. ①021

中国版本图书馆 CIP 数据核字(2011)第 203706 号

丛书策划 谷 雨
责任编辑 王韵霏
装帧设计 钱宇辰

高等院校统计学精品课教材
概率论与数理统计(第四版)
苏均和 主编

出 版 格致出版社
世纪出版集团 www.hibooks.cn
www.ewen.cc 上海人民出版社

(200001 上海福建中路193号24层)



编辑部热线 021-63914988
市场部热线 021-63914081

发 行 世纪出版集团发行中心
印 刷 浙江省临安市曙光印务有限公司
开 本 787×1092 毫米 1/16
印 张 20.5
插 页 1
字 数 414,000
版 次 2011年12月第1版
印 次 2011年12月第1次印刷
ISBN 978-7-5432-2019-5/F·478
定 价 39.00 元

作者简介

苏均和，副教授，硕士生导师，1987年毕业于上海交通大学应用数理统计专业，现任教于上海财经大学统计与管理学院，长期从事数理统计和数量经济的教学和科研工作。在《宏观经济研究》、《中国金融》等核心刊物发表论文三十余篇，编写专著和教材共六本。

高等院校统计学精品课教材编委会

主任

徐国祥 上海财经大学教授、博导

委员(以姓氏笔画排列)

王振龙 西安财经学院教授
史代敏 西南财经大学教授、博导
艾春荣 上海财经大学统计系教授、博导
美国佛罗里达大学教授
刘建平 暨南大学教授、博导
刘洪 中南财经政法大学教授
向书坚 中南财经政法大学教授、博导
纪宏 首都经济贸易大学教授、博导
许鹏 湖南大学教授
余思勤 上海海事大学教授
李宝瑜 山西财经大学教授、博导
李金昌 浙江工商大学教授、博导
杨灿 厦门大学教授、博导
肖红叶 天津财经大学教授、博导
苏卫华 浙江工商大学教授、博导
邱东 中央财经大学教授、博导
庞皓 西南财经大学教授、博导
林洪 广东商学院教授
罗良清 江西财经大学教授
金勇进 中国人民大学教授、博导
贺铿 中央财经大学教授、博导
袁卫 中国人民大学教授、博导
曾五一 厦门大学教授、博导
蒋萍 东北财经大学教授、博导
谢邦昌 台湾辅仁大学统计咨询系教授、博导
韩兆州 暨南大学教授、博导
雷钦礼 暨南大学教授、博导

前 言

随着我国经济的快速发展,教育体制以及课程改革的不断深入,学习方式也从传统的“被动性、依赖性、统一性、虚拟性、认同性”向现代的“主动性、独立性、独特性、体验性、问题性”方向转变。为了使概率论与数理统计这门课程适应这种改革的需要,我们编写了此教材。此教材是在《概率论与数理统计》(第三版)的基础上改编完善而成,在内容体系、案例编排上都做了较大幅度的改变。本教材更完善、更成熟,也更适应时代发展的需要。

概率论与数理统计是研究随机现象并找出其统计规律的一门学科,是广泛应用于社会、经济、科学等各个领域的定量和定性的科学体系。我们要求学生通过学习该课程掌握概率、统计的基本概念,熟悉数据处理、数据分析、数据推断的各种基本方法,并能用所掌握的方法解决社会经济所遇到的各种具体问题。

本书共分10章。第1章至第5章介绍概率论,主要讲述概率基础知识,包括随机事件和概率、一维随机变量及其分布、二维随机变量及其分布、大数定律和中心极限定理等内容。第6章至第10章介绍数理统计,阐述数理统计的基础概念、参数估计、假设检验、非参数统计推断和回归分析等内容。

本教材有以下特点:

- (1) 具有财经特色,所选的案例尽可能与社会经济相结合。
- (2) 内容全面、深入浅出,编排更加细致,易于学生打好基础。
- (3) 适用面广,本教材既可作为统计专业本科的教材,又可作为各类读者的统计入门教材。
- (4) 理论和方法相结合,本教材不仅注重基础理论的阐述,而且更重视各类方法的具体使用。

最后要感谢为此书编写做过许多工作的张龙武、冯凤霞、王琪琛和邹东方。

编 者

目 录

前言

第 1 章 随机事件和概率	1
第一节 随机事件及其运算	1
第二节 随机事件的概率及其性质	7
第三节 条件概率、乘法公式、全概率公式与贝叶斯公式	17
第四节 相互独立的随机事件与独立试验概型	22
本章小结	26
习题	27
第 2 章 一维随机变量及其分布	31
第一节 随机变量及其分布函数	31
第二节 离散型随机变量	35
第三节 连续型随机变量	44
第四节 随机变量函数的分布	54
本章小结	57
习题	57
第 3 章 多维随机变量及其分布	61
第一节 二维随机变量及其分布	61
第二节 条件分布	68
第三节 随机变量的独立性	72
第四节 二维随机变量函数及多维随机变量的分布	75
本章小结	82
习题	83
第 4 章 随机变量的数字特征和特征函数	87
第一节 随机变量的数学期望	87
第二节 随机变量的方差、协方差与相关系数	97
第三节 随机变量的特征函数	112

本章小结	119
习题	120
第 5 章 大数定律和中心极限定理	125
第一节 大数定律	125
第二节 中心极限定理	130
本章小结	139
习题	141
第 6 章 数理统计的基本概念	143
第一节 总体与样本	143
第二节 统计量与样本矩	145
第三节 抽样分布	147
本章小结	157
习题	158
第 7 章 参数估计	160
第一节 参数的点估计	160
第二节 点估计量的优良标准	169
第三节 参数的区间估计	178
本章小结	188
习题	188
第 8 章 假设检验	193
第一节 假设检验的基本概念	193
第二节 单个正态总体下参数的检验	196
第三节 两个正态总体下参数的比较检验	202
第四节 非正态总体下的大样本参数检验	208
本章小结	209
习题	210
第 9 章 非参数统计推断	213
第一节 分布的 χ^2 -检验	213
第二节 柯尔莫哥洛夫和斯米尔诺夫检验	222
第三节 偏度、峰度检验方法与正态概率纸方法	227

第四节	符号检验与中位数检验	230
第五节	游程检验与秩和检验	235
	本章小结	240
	习题	240
第 10 章	回归分析	243
第一节	概述	243
第二节	一元回归分析和相关分析	244
第三节	多元线性回归	253
第四节	非线性回归	265
第五节	逐步回归	270
	本章小结	272
	习题	272
附录一	习题参考答案及提示	277
附录二	关于上、下、双侧分位数	293
附录三	统计用表	295
附表 1	随机数字表	295
附表 2	正态分布双侧临界值表	296
附表 3	二项分布表	297
附表 4	泊松分布表	299
附表 5	正态分布函数 $N(0, 1)$ 的数值表	301
附表 6	t -分布双侧临界值表	303
附表 7	χ^2 -分布的上侧临界值 χ_{α}^2 表	304
附表 8	F -分布上侧临界值表	306
附表 9	柯尔莫哥洛夫检验的临界值 ($D_{n\alpha}^*$) 表	314
附表 10	秩和检验表	315
附表 11	相关系数检验表	316
附表 12	符号检验表	316
附表 13	游程数检验表	317
参考文献		318

随机事件和概率

第一节 随机事件及其运算

一、必然现象和随机现象

在自然界、生产实践、科学试验和日常生活中,人们观察到的现象大体可以归结为两种类型。一类是在一定条件下必然出现(或者不出现)某种结果的现象。这类现象的一个共同特点就是可以事前预言,即在准确地重复某些条件下,它的结果总是可以肯定的,或是根据它过去的状态,在相同的条件下完全可以预言将来的发展,我们把这类现象称之为确定性现象或者必然现象。例如,月亮绕地球转;直角三角形的三边必然满足勾股定理;在一批合格的产品中任意取一件,必定不是废品等。几何、微积分、线性代数都是研究确定性现象的数学工具。另一类现象是在相同的条件下可能得到多种不同的结果,出现具体哪一种结果是不可预测的,即在相同的条件下,未来的发展事前不能肯定。例如,将一枚硬币抛向空中,无论如何控制硬币的抛掷条件,我们在每次抛掷前总是无法肯定硬币自由下落后,我们可以看见硬币的哪一面;又如,某射手向同一目标多次射击,各次弹着点不尽相同,并且无论事前怎样控制射击条件,在一次射击之前都无法肯定弹着点的确切位置。这类现象的共同特点就是:可以在相同条件下重复进行试验或者观察,而每次试验或者观察的可能结果不止一个,且事前不能预知确切结果,即试验结果呈现出不确定性。人们经过长期实践并深入研究之后,发现这类现象虽然在个别试验或观察中的结果呈现出不确定性,但是在大量观察或多次重复试验后,其结果往往呈现出某种客观规律,并且这种客观规律性是可以认识的。例如,多次重复抛掷同一枚均匀硬币,正面朝上的次数大致占抛掷总次数的一半;某射手向同一目标射击的弹着点按照一定的规律分布等。这种在大量重复试验或者观察中所呈现出事物的固有规律称为“统计规律性”。我们把在个别试验或者观察中呈现

的不确定性,在大量重复试验或者观察中又具有统计规律的现象称为随机现象(或者偶然现象)。

概率论与数理统计就是研究随机现象统计规律性的一门数学学科,但是各有侧重。概率论侧重于理论上的研究,介绍随机现象反映的基本概念,建立相应的定理和公式,找出计算统计规律的方法;而数理统计是以概率论为理论依据,研究如何设计试验并对试验结果进行整理和统计分析。

概率论的研究开始于意大利文艺复兴时期,源于赌博。数理统计学是伴随着概率论而发展的,最早见于国家的人口统计中。当今,概率论和数理统计的理论和方法几乎遍及所有的科学技术领域、工农业生产和国民经济各个部门之中。在经济科学中,它是经济管理、质量控制、保险理论、系统论、经济预测和决策、计量经济学等学科的理论基础,是各类经济、管理专门人才不可缺少的数学工具。

二、随机试验、随机事件与样本空间

为了叙述方便,我们把对某种自然现象作一次观察或者进行一次科学试验,统称为一个试验。如果这个试验具有下列特性:

- (1) 试验可以在相同的条件下重复进行;
- (2) 每次试验的可能结果不止一个,并且能在试验之前明确知道所有可能的结果;
- (3) 每次试验之前不能肯定这次试验会出现哪一个结果,但是可以肯定每次试验总会出现并且仅出现这些可能结果中的一个。

那么,我们将具有上述三个特性的这种试验称为随机试验,简称试验。在随机试验中,我们所关心的是某种结果是否出现,我们把试验的每一个可能结果称之为随机事件,简称事件。

通过试验观察随机现象时,假定试验 E 中一切可能的结果是已知的,并且每次试验都必定是其中一个结果出现,我们把这些结果组成的集合称为样本空间或者基本事件空间,记做 Ω 。 Ω 的元素(即试验中的一个可能结果)称为样本点或基本事件,记做 ω 。

在研究随机现象时,不仅考虑基本事件(即不可能再分的事件),而且还要考虑一些由基本事件组成的复合事件。一般把事件定义为样本空间 Ω 的子集,用大写字母 A 、 B 等来表示。在试验中,当且仅当事件 A 中所包含的某一样本点出现时,称事件 A 发生。

下面举一些样本空间和事件的例子。

[例 1-1] 做掷一枚均匀硬币的试验 E ,用“正”表示“出现正面”,用“反”表示“出现反面”。试验的可能结果只有两个: $\omega_1 = \text{“正”}$, $\omega_2 = \text{“反”}$,试验 E 的样本空间 $\Omega = \{\omega_1, \omega_2\}$ 由两个基本事件组成。

[例 1-2] 设试验 E 为同时掷两枚可以区别的硬币,试验的可能结果共有四个:

$\omega_1 = \{\text{正, 正}\}, \omega_2 = \{\text{正, 反}\}, \omega_3 = \{\text{反, 正}\}, \omega_4 = \{\text{反, 反}\}$, 试验 E 的样本空间 $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ 。

在这个试验中, 还可以考虑它的某些复合事件。例如,

A: “两枚硬币出现同一面”的事件, 则事件 A 由 ω_1 和 ω_4 组成, 即 $A = \{\omega_1, \omega_4\}$ 。这时, 当且仅当 ω_1 或者 ω_4 之一出现时, 事件 A 就发生了。

B: “恰有一枚硬币出现正面”的事件, 则 $B = \{\omega_2, \omega_3\}$ 。

C: “至少有一枚硬币出现正面”的事件, 则 $C = \{\omega_1, \omega_2, \omega_3\}$, 这时, 当且仅当 ω_1 、 ω_2 或者 ω_3 之一出现时, 事件 C 就发生了。

[例 1-3] 某路口, 在单位时间内出现交通事故的次数是 $0, 1, 2, \dots$, 用 ω_i 表示发生 i 次事故的事件, 则样本空间 $\Omega = \{0, 1, 2, \dots\}$ 这一随机试验中, 我们可以考虑 $A = \{i \mid i \leq 2, i \text{ 为自然数}\}, B = \{i \mid i \leq 4, i \text{ 为自然数}\}$ 等复合事件。

[例 1-4] 在一批灯泡中, 任意取一只测试其寿命 $t(\text{h})$, 则样本空间为 $\Omega = \{t \mid t > 0\}$ 。

在这一试验中, 可以考虑 $A = \{t \mid t > 500\}, B = \{t \mid t < 2000\}$ 等这样一些复合事件。

由上面例子可知, 对于一个具体的试验来说, 样本空间是根据试验的具体内容来决定的。样本空间可以相当简单, 也可以相当复杂。但是在概率论的研究讨论中, 一般认为样本空间是给定的(或者是事前已经明确了试验的所有可能的结果)。这种必要的抽象, 其目的是为了能够更好地把握随机现象的本质, 而使所得的理论性的结论得到更加广泛的应用。当然, 对一个实际问题来讲, 样本空间如何描述应进行具体的分析和研究。此外, 一个抽象的样本空间经常可以被赋予各种不同的解释(即可以概括各种内容不同的问题)。例如, 仅仅包含两个样本点的样本空间, 它既可以作为掷硬币出现“正面”或者出现“反面”的模型, 也能用于检验产品的“合格”或者“不合格”; 用于公用事业中排队现象的“有人排队”或“无人排队”; 以及出生婴儿性别的“男”或“女”; 等等。这说明尽管问题的实际内容不同, 但是有时候却能够归结为相同的概率模型。因此, 我们常常以抛掷硬币、摸球等这样的一些既典型又形象且易于理解的例子来阐明一些问题, 以便使问题阐述得更加明确, 使问题的本质更为突出。

为了研究的方便, 把样本空间 Ω 也作为一个事件。因为在每次试验时, 必然出现 Ω 中的某个样本点, 即 Ω 必然发生, 所以也称 Ω 为必然事件。类似地, 空集 Φ 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、事件的关系

(一) 事件的包含与相等

如果任一属于 A 的样本点一定属于 B, 则称事件 B 包含事件 A, 记为 $B \supset A$, 或者 $A \subset B$ 。显然, $B \supset A$ 的含义是: 事件 A 发生, 必然导致事件 B 发生。

例如,以 A 表示“灯泡使用寿命大于 5 000 小时”,以 B 表示“灯泡使用寿命大于 2 000 小时”,则 $B \supset A$ 。

如果 $B \supset A$ 与 $A \supset B$ 同时成立,则 A 与 B 相等,记作 $A = B$ 。

(二) 事件的并

由属于 A 或者属于 B 的所有样本点组成的集合,称为 A 与 B 的并(或者和),记为 $A \cup B$ 。显然, $A \cup B$ 表示事件 A 与事件 B 至少有一个发生。

(三) 事件的交

由属于 A 同时又属于 B 的所有样本点组成的集合,称为 A 与 B 的交(或者积),记为 $A \cap B$ 或者 AB 。显然, $A \cap B$ 表示事件 A 与事件 B 同时发生。

(四) 事件的差

由属于 A 但是不属于 B 的所有样本点组成的集合,称为 A 与 B 的差,记为 $A - B$ 。显然, $B - A$ 表示事件 B 发生而事件 A 不发生。

例如,在投掷一颗骰子的试验中,记事件 $A = \{\text{出现偶数点}\} = \{2, 4, 6\}$,记事件 $B = \{\text{出现点数超过 } 3\} = \{4, 5, 6\}$ 。则 A 与 B 的并集为 $A \cup B = \{2, 4, 5, 6\}$, A 与 B 的交集 $A \cap B = \{4, 6\}$, A 对 B 的差 $A - B = \{2\}$, B 对 A 的差 $B - A = \{5\}$ 。

(五) 对立事件

样本空间 Ω 与 A 的差 $\Omega - A$ 称为事件 A 的对立事件(或者逆事件),记为 \bar{A} ,事件 \bar{A} 表示事件 A 不发生。

例如,“灯泡使用寿命大于 5 000 小时”的事件是“灯泡使用寿命不大于 5 000 小时(小于等于 5 000 小时)”事件的对立事件。

显然, A 与 \bar{A} 互为对立事件,即有 $\bar{\bar{A}} = A$ 。

对立事件 A 和 \bar{A} 间的关系可以表示为: $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \Phi$ 。

(六) 互不相容事件

如果 $A \cap B = \Phi$,则称事件 A 和事件 B 互不相容,也称 A 、 B 是互斥的。

基本事件是互不相容的。

事件的并与事件的交的概念可以推广到有限多个事件或者可列多个事件的场合。

下面先介绍一些定义: $\bigcup_{k=1}^n A_k$ 表示 $A_1 \cup A_2 \cup \cdots \cup A_n$, $\bigcap_{i=1}^n A_i$ 表示 $A_1 \cap A_2 \cap \cdots \cap A_n$;称 $\bigcup_{k=1}^n A_k$ 为事件 A_1, A_2, \cdots, A_n 的并,通常表述为“ A_1, A_2, \cdots, A_n 至少有一个发生”;称 $\bigcup_{k=1}^{\infty} A_k$ 为事件 $A_1, A_2, \cdots, A_n, \cdots$ 的并,通常表述为“ $A_1, A_2, \cdots, A_n, \cdots$ 至少有一个发生”;称 $\bigcap_{k=1}^n A_k$ 为事件 A_1, A_2, \cdots, A_n 的交,通常表述为“ A_1, A_2, \cdots, A_n 同时发生”;称 $\bigcap_{k=1}^{\infty} A_k$ 为事件 $A_1, A_2, \cdots, A_n, \cdots$ 的交,通常表述为“ $A_1, A_2, \cdots, A_n, \cdots$ 同时发生”。

由上面的定义可以看出事件的关系和运算与集合的相应关系和运算一致,因此也常常用韦恩图(Venn)来直观地表示,见图 1-1。

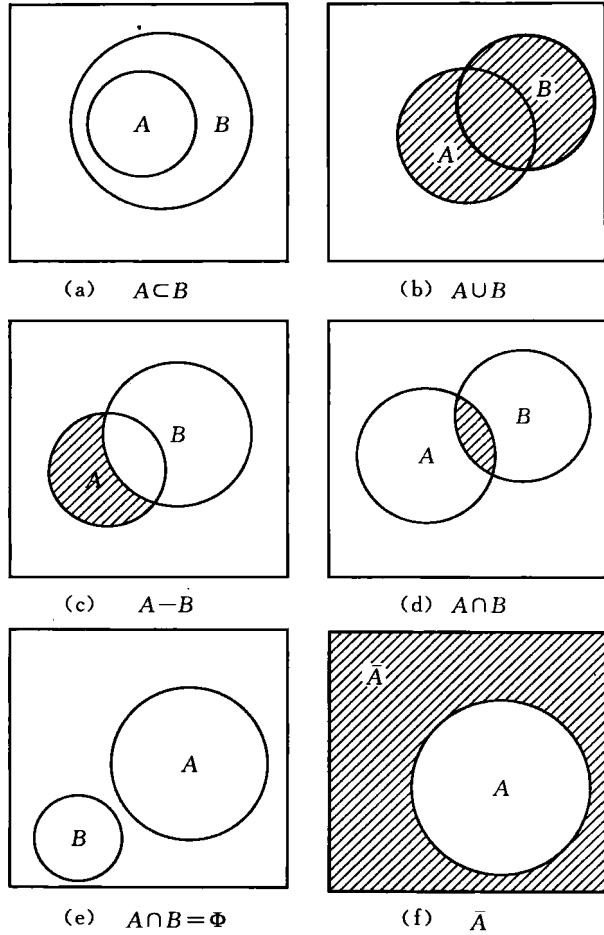


图 1-1 事件的关系与运算

四、事件的运算法则

由图 1-1 可以验证,事件的运算满足下列运算法则:

(1) 交换律: $A \cup B = B \cup A, A \cap B = B \cap A$

(2) 结合律: $A \cup (B \cup C) = (A \cup B) \cup C$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

(3) 分配律: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

(4) 德莫根(Demorgan)公式: $\overline{A \cup B} = \bar{A} \cap \bar{B}, \overline{A \cap B} = \bar{A} \cup \bar{B}$

分配律和德莫根公式可以推广到有限个或者可列多个事件的情形,即有:

$$A \cap (\bigcup_i A_i) = \bigcup_i (A \cap A_i)$$

$$A \cup (\bigcap_i A_i) = \bigcap_i (A \cup A_i)$$

$$\overline{\bigcup_i A_i} = \bigcap_i \bar{A}_i, \quad \overline{\bigcap_i A_i} = \bigcup_i \bar{A}_i$$

上述运算法则为基本运算法则,另外还有:

(5) 补元律: $A\bar{A} = \emptyset, A \cup \bar{A} = \Omega$

(6) 还原律: $\bar{\bar{A}} = A$

(7) 分解律: 若 $A \subset B$, 则 $B = A \cup \bar{A}B$

(8) 差积转换律: $A - B = A\bar{B} = A - AB$

(9) 吸收律: 若 $A \subset B$, 则 $AB = A, A \cup B = B$

[例 1-5] 一个货箱中装有 12 只同类型的产品,其中 3 只是一等品,9 只是二等品。从中随机地抽取两次,每次任取 1 只, $A_i (i = 1, 2)$ 表示第 i 次抽取的是一等品,试用字母及事件间的关系表示下列事件:

- (1) 两只都是一等品;
- (2) 两只都是二等品;
- (3) 一只是一等品,另一只是二等品;
- (4) 第二次抽取的是一等品。

解:由题意,用 \bar{A}_i 表示第 i 次抽取的是二等品 ($i = 1, 2$), 则

(1) 两只都是一等品: $A_1 \cap A_2$;

(2) 两只都是二等品: $\bar{A}_1 \cap \bar{A}_2$;

(3) 一只是一等品,另一只是二等品: $\bar{A}_1 A_2 \cup A_1 \bar{A}_2$;

(4) 第二次抽取的是一等品: $\bar{A}_1 A_2 \cup A_1 A_2$ 。

[例 1-6] 在一批产品中有若干正品与次品,设事件 A_i 表示第 i 次取到次品 ($i = 1, 2, 3$), 试用文字叙述下列事件:

$$A_1 \cup A_2, A_1 \cup A_2 \cup A_3, A_1 A_2 A_3, \bar{A}_2, A_3 - A_2, \overline{A_1 \cup A_2}, \bar{A}_2 \cup \bar{A}_3, A_1 A_2 A_3 \cup \bar{A}_1 A_2 A_3 \cup A_1 \bar{A}_2 A_3 \cup A_1 A_2 \bar{A}_3。$$

解: $A_1 \cup A_2$: 前两次抽取中至少有一次取到次品;

$A_1 \cup A_2 \cup A_3$: 三次抽取中至少有一次取到次品;

$A_1 A_2 A_3$: 三次抽取都取到次品;

\bar{A}_2 : 第二次抽到正品;

$A_3 - A_2 = A_3 \bar{A}_2$: 第三次抽到次品,而第二次抽到正品;

$\overline{A_1 \cup A_2} = \bar{A}_1 \bar{A}_2$: 前两次都取到正品;

$\bar{A}_2 \cup \bar{A}_3 = \overline{A_2 A_3}$: 后两次抽取中至少有一次是正品;

$A_1 A_2 A_3 \cup \bar{A}_1 A_2 A_3 \cup A_1 \bar{A}_2 A_3 \cup A_1 A_2 \bar{A}_3$: 三次抽取中至少有一次抽到次品,或三次抽取中至多有一次抽到正品。

第二节 随机事件的概率及其性质

一、概率的直观意义

我们观察一个随机试验的各种事件,一般会发现,并非所有的事件出现的可能性都相等,有些事件出现的可能性大些,有些事件出现的可能性小些。

研究随机现象不仅要知道它可能出现哪些事件,更加重要的是要研究各种事件出现可能性的大小,揭示出现这些事件的内在统计规律,会有利于我们认识世界和改造世界。例如,知道了某电话总机在 24 小时内出现呼叫次数可能性的大小,就可以根据要求配置一定的线路设施和管理人员等。

如果只能大概地比较事件发生的可能性大小,要想进行确切的推断、得出精确的结论是不可能的。我们要求一个刻画事件发生可能性大小的数量指标,这个数量指标至少应该满足两个要求:

(1) 它应该具有一定的客观性,不能够随意改变,而且理论上可以通过“在相同的条件下”大量的重复试验予以识别和检验。

(2) 它必须符合一般常情。例如,事件发生可能性大的,它的值就大;事件发生可能性小的,它的值就小;必然事件的值最大;不可能事件的值最小且等于零。

我们把刻画事件发生的可能性大小的数量指标叫做事件的概率。事件 A 的概率以 $P(A)$ 表示,并且规定 $0 \leq P(A) \leq 1$ 。

二、随机事件的频率与概率的统计定义

对于随机事件来说,虽然在一次试验中发生与否是不确定的,但是在相同的条件下大量重复进行这一试验就会发生:一个随机事件发生的可能性是确定的,而且不同事件发生的可能性有大小之分。例如,在掷一颗骰子的试验中,“出现偶数点”这个事件与“出现奇数点”的可能性是相同的,而“出现偶数点”这个事件要比“出现 4 点”的可能性要大。这种可能性的大小是事件本身固有的一种属性,为定量描述随机事件的这种属性,在这里我们首先介绍频数和频率的概念。

定义 1.1 在相同的条件下,重复 n 次试验,随机事件 A 在 n 次试验中出现的次数 m 称为频数,比值 $\frac{m}{n}$ 称为事件 A 发生的频率,记为 $f_n(A) = \frac{m}{n}$ 。

可以验证,当试验次数 n 固定时,事件 A 发生的频率 $f_n(A)$ 有如下的性质:

$$(1) 0 \leq f_n(A) \leq 1;$$

$$(2) f_n(\Omega) = 1;$$

$$(3) \text{若 } A_1, A_2, \dots, A_k \text{ 互不相容,则 } f_n\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k f_n(A_i)。$$

关于这些性质的证明,前两个是显然的,而第三个我们也可以通过数学归纳法来证明,在此略去。

在历史上,曾经有人做过大量抛硬币的试验,表 1-1 是试验的部分结果。

表 1-1 抛硬币试验的部分结果

试验者	抛币次数	出现正面次数	出现正面频率
德莫根	2 048	1 061	0.515 8
浦 丰	4 040	2 048	0.506 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

由表 1-1 可以知道,“出现正面”的频率总是在 0.5 附近摆动,而且随着试验次数的增加逐渐稳定在 0.5。人类大量的试验证明,当在同样条件下大量重复一个试验时,随机事件的频率具有随着随机试验次数增加而趋向稳定的性质。而所谓的事件发生的可能性的,就是“频率”的稳定值。下面给出概率的统计定义。

定义 1.2 在相同的条件下,重复进行 n 次试验,事件 A 发生的频率 $f_n(A)$ 在某个常数值 P 附近摆动,则称频率的稳定值 P 为事件 A 发生的概率,简称事件 A 的概率,记作 $P(A)$ 。一般来说, n 越大,摆动幅度越小。

数值 P (即 $P(A)$) 就是在一次试验中对事件 A 发生的可能性大小的数量描述。例如,用 0.5 来描述掷一枚硬币“正面”出现的可能性。

对于任何事件 A ,均有 $0 \leq \frac{m}{n} \leq 1$,故由定义 1.2 知 $0 \leq P(A) \leq 1$ 。特别地,因为必然事件 Ω 的发生的频率恒为 1,所以 $P(\Omega) = 1$,不可能事件 Φ 发生的频率恒为 0,所以 $P(\Phi) = 0$ 。实际上,由于频率与统计概率的这种联系,统计概率满足:

$$(1) 0 \leq P(A) \leq 1;$$

$$(2) P(\Omega) = 1, P(\Phi) = 0;$$

$$(3) \text{若 } A_1, A_2, \dots, A_n \text{ 互不相容,则 } P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

虽然事件的频率与概率都是事件出现可能性大小的度量,但是频率是试验值,依赖于试验次数,即使试验次数相同,频率也可能取得不同的值,频率具有随机性。概率是先于试验而客观存在的理论值,它是一个确定的常数,其大小取决于事件本身固有的规律性。定义 1.2 是描述性定义,不能直接用来计算事件 A 的概率。在实际应用中,常常通过做大量重复的试验得到事件 A 发生的频率,并且以它作为 $P(A)$ 的“近似值”或者“估计值”。例如,一名篮球运动员在 100 次投篮中,命中的次数是 40 次,则他投中篮的概率是 0.4;又如,在产品检测中,抽取 1 000 件产品,发现了 10 件是次品,那么我们可以认为次品率为 1%。这些例子都是用近似值代替概率的例子,类似的例子在实际生活中还有很多。