现代物理基础丛书 45

实验数据分析

（上 册）

朱永生 著

科 学 出 版 社

内 容 简 介

本书介绍实验和测量数据分析中涉及的概率和数理统计及相关的数学知识，内容包括概率论、经典数理统计、贝叶斯统计、蒙特卡罗方法、极小化方法和去弥散方法六个部分. 特别讨论了数据统计处理中的一些困难问题和近期国际上发展起来的新方法. 书中分析了取自普通物理、核物理、粒子物理和工程技术问题的许多实例，注重物理问题与数学方法的结合，具体阐述了概率和数理统计及相关的数学方法在实际问题中的应用. 书末附有详尽的数理统计表，可供本书涉及的几乎所有数据分析问题之需要，而无需查阅专门的数理统计表书籍.

本书可供实验物理工作者和大专院校相关专业师生、理论物理研究人员、工程技术人员以及从事自然科学和社会科学的数据测量和分析研究人员参考.

图书在版编目（CIP）数据

实验数据分析. 上册/朱永生著. —北京：科学出版社，2012
(现代物理基础丛书；45)
ISBN 978-7-03-034731-2

I. ①实… II. ①朱… III. ①物理学-实验数据-分析 IV. ①O4-33

中国版本图书馆 CIP 数据核字(2012) 第 122005 号

责任编辑：钱 俊 鲁永芳／责任校对：朱光兰
责任印制：钱玉芬／封面设计 陈 敬

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

源海印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2012 年 6 月第 一 版 开本 B5(720 × 1000)
2012 年 6 月第一次印刷 印张 25 3/4
字数 504 000

定价：88.00 元

前　言

自然科学和社会科学的许多领域, 诸如粒子物理和核物理、粒子和核天体物理、宇宙学和宇宙线物理、核工程和核医学、遗传学、人口统计、经济学、交通运输等等, 存在大量随机现象; 即使是对确定性现象的研究, 由于测量工具和仪器的测量误差通常具有随机性质, 所获得的数据也带有随机性. 因此, 实验或测量数据的分析几乎必定需要随机性数学及相关的数学分支的知识.

本书介绍实验或测量数据分析中所涉及的概率和数理统计及相关的数学知识, 内容包括概率论、经典数理统计、贝叶斯统计、蒙特卡罗方法、极小化方法和去弥散方法六个部分. 其中第 1~5 章和第 6~12 章分别阐述概率论和经典数理统计的基本内容, 第 13 章则专门介绍在现代统计学中具有重要影响的贝叶斯学派的观点与理论, 第 14 章讨论应用日益广泛的蒙特卡罗方法的基本概念, 第 15 章介绍的极小化 (或最优化) 方法是求解许多数理统计问题的重要工具 (例如, 极大似然法、最小二乘法等), 最后第 16 章介绍去弥散方法, 处理从观测数据和测量仪器的分辨函数反演出原分布的问题 (第 12~16 章见本书下册).

一般的数学书籍侧重于数学上的严密和确切, 但对于以数学为工具的研究人员而言, 更为迫切的是了解相关数学的基本内容和方法, 并正确地运用它们. 因此, 本书力求数学与物理问题相结合, 避免过于抽象和过于数学化的讨论. 重点是介绍基本概念、基本原理和方法, 阐明方法的应用及适用条件, 而不是对定理作严格的证明和推导; 有些定理或结论只是直接引用, 但与数据处理直接相关的内容则予以充实. 本书中相当数量的例子取自普通物理、核物理、粒子物理和工程技术问题的数据分析问题的实际需要, 以期达到加深理解基本原理和正确运用它们的目的. 本书还力求反映国际上近期发展起来的处理数据分析中一些困难问题的新的概念和方法, 这方面的例子有第 11 章介绍的 “小信号测量的区间估计”、12.5 节讨论的 “信号的统计显著性”、8.8 节和 9.10 节讨论的 “多个实验结果的合并” 等内容. 书末附有详尽的数理统计表, 可供本书涉及的许多数据分析问题之需要, 而无需查阅专门的数理统计表书籍.

作者根据近年来在科研实践 (北京正负电子对撞机、北京谱仪的粒子物理实验研究) 和教学实践 (中国科学院研究生院讲授 “实验物理中的概率和统计” 课程) 中的体会, 认为有必要对科学出版社 2006 年出版的拙作《实验物理中的概率和统计 (第二版)》的内容作必要的修改和扩充. 增加的两章是近几十年来影响日益重要的贝叶斯统计, 以及在粒子物理和核物理、粒子和核天体物理、宇宙线物理和宇宙学、

光学和医学成像重建等领域得到广泛应用的去弥散方法. 在原有的章节中则增补了原书没有讨论但具有相当重要性的一些内容和概念, 如贝塔分布、对数正态分布和朗道分布, 费希尔信息, 区间估计的大样本法, 有约束的极大似然估计, 斯米尔诺夫-克拉美-冯 • 迈希斯检验, 多维随机变量分量的独立性检验, 相关性检验等. 此外, 还补充了与新增内容相对应的数理统计表, 增加了一些从实践中提炼出来的有典型意义的实例.

本书介绍了六部分共 16 章比较广泛的数学内容, 掌握这些知识能够大体应对数据分析对于随机性数学知识的基本需求, 至少可以提供进一步学习的基础. 事实上, 其中的每一部分都是一个重要的数学分支, 可以并且应当按照需要进行更为深入的学习. 对出于实用目的阅读本书的读者而言, 可以有选择地阅读部分章节. 书中第 1~5 章、第 6~10 章和第 12 章阐述概率论和经典数理统计的基本内容, 是利用随机性数学处理实际问题的基础; 第 11 章和第 13~16 章则可以根据需要加以选择.

本书的出版得到中国科学院科学出版基金的资助. 作者感谢中国工程院院士、中国高等科学技术中心学术主任叶铭汉研究员, 中国科学院研究生院郑阳恒教授, 中国科学院院士、中国科学院理论物理研究所张肇西研究员对于本书的热情鼓励和积极推介, 以及科学出版社钱俊等同志的大力支持和辛勤、细致的工作.

限于本人水平, 疏漏及不妥之处在所难免, 诚恳欢迎专家和读者批评指正.

朱永生

2011 年 9 月于北京

目　　录

前言
第 1 章　概率论初步 …… 1
　1.1　随机试验, 随机事件, 样本空间 …… 1
　1.2　概率 …… 4
　1.3　条件概率, 独立性 …… 7
　1.4　概率计算举例 …… 9
　1.5　边沿概率, 全概率公式, 贝叶斯公式 …… 13
第 2 章　随机变量及其分布 …… 17
　2.1　随机变量 …… 17
　2.2　随机变量的分布 …… 18
　2.3　随机变量函数的分布 …… 21
　2.4　随机变量的数字特征 …… 24
　2.5　随机变量的特征函数 …… 29
　2.6　离散随机变量的概率母函数 …… 33
第 3 章　多维随机变量及其分布 …… 36
　3.1　二维随机变量的分布, 独立性 …… 36
　3.2　条件概率分布 …… 39
　3.3　二维随机变量的数字特征 …… 41
　3.4　二维随机变量的函数的分布 …… 45
　3.5　多维随机变量, 向量和矩阵记号 …… 54
　3.6　多维随机变量的联合特征函数 …… 59
　3.7　多维随机变量的函数的分布 …… 61
　3.8　线性变换和正交变换 …… 64
　3.9　误差传播公式 …… 68
第 4 章　一些重要的概率分布 …… 73
　4.1　伯努利分布和二项分布 …… 73
　4.2　多项分布 …… 82
　4.3　泊松分布, 泊松过程 …… 85
　4.4　泊松分布与其他分布的相互联系 …… 91
　4.5　复合泊松分布 …… 95

4.6 几何分布, 负二项分布, 超几何分布 ······ 97
4.7 均匀分布 ······ 100
4.8 指数分布 ······ 102
4.9 伽马分布 ······ 104
4.10 贝塔分布 ······ 107
4.11 正态分布 ······ 108
4.12 二维正态分布 ······ 114
4.13 多维正态分布 ······ 120
4.14 对数正态分布 ······ 124
4.15 柯西分布 ······ 125
4.16 朗道分布 ······ 127
4.17 χ^2 分布 ······ 129
4.18 t 分布 ······ 136
4.19 F 分布 ······ 140
4.20 实验分布 ······ 145
4.20.1 实验分辨函数 ······ 145
4.20 2 探测效率 ······ 152
4.20 3 复合概率密度 ······ 154
第 5 章 大数定律和中心极限定理 ······ 158
5.1 大数定律 ······ 158
5.2 中心极限定理 ······ 161
第 6 章 子样及其分布 ······ 166
6.1 随机子样, 子样分布函数 ······ 166
6.2 统计量及其数字特征 ······ 168
6.3 抽样分布 ······ 175
6.3 1 子样平均值的分布 ······ 175
6.3.2 服从 χ^2 分布的统计量, 自由度 ······ 177
6 3 3 服从 t 分布和 F 分布的统计量 ······ 180
6.3 4 正态总体子样偏度、子样峰度、子样相关系数的分布 ······ 181
6.4 抽样数据的图形表示, 频率分布 ······ 182
6 4.1 一维散点图和直方图, 频率分布 ······ 182
6.4 2 二维散点图和直方图 ······ 185
第 7 章 参数估计 ······ 189
7.1 估计量, 似然函数 ······ 189
7.2 估计量的相合性 ······ 191

7.3　估计量的无偏性……192
7.4　估计量的有效性和最小方差……195
7.5　估计量的充分性, 信息……203
7.5.1　充分统计量……203
7 5 2　充分性与信息……211
7.6　区间估计……213
7 6 1　枢轴变量法……214
7.6.2　大样本法……218
7.7　正态总体均值的置信区间……221
7.8　正态总体方差的置信区间……225
7.9　正态总体均值和方差的联合置信域……229
第 8 章　极大似然法……231
8.1　极大似然原理……231
8.2　正态总体参数的极大似然估计……237
8.3　极大似然估计量的性质……239
8.3 1　参数变换下的不变性……240
8 3 2　相合性和无偏性……240
8 3.3　充分性……241
8 3.4　有效性……242
8 3 5　唯一性……245
8.3 6　渐近正态性……246
8.4　极大似然估计量的方差……248
8 4 1　方差估计的一般方法……249
8 4 2　充分和有效估计量的方差公式……251
8 4.3　大子样情形下的方差公式……254
8.5　极大似然估计及其误差的图像确定……258
8.5 1　总体包含单个未知参数……258
8.5 2　总体包含两个未知参数……262
8.6　利用似然函数作区间估计, 似然区间……264
8.6 1　单个参数的似然区间……266
8.6 2　由巴特勒特函数求置信区间……268
8 6 3　两个参数的似然域……271
8.6 4　多个参数的似然域……276
8.7　极大似然法应用于直方图数据……278
8.8　极大似然法应用于多个实验结果的合并……280

8.8.1 正态型似然函数 ······ 280
8.8.2 非正态型似然函数 ······ 283
8.9 极大似然法应用于实验测量数据 ······ 289
8.10 有约束的极大似然估计 ······ 291
第 9 章 最小二乘法 ······ 295
9.1 最小二乘原理 ······ 295
9.2 线性最小二乘估计 ······ 297
9.2.1 正规方程 ······ 298
9.2.2 线性最小二乘估计量的性质 ······ 302
9.2.3 线性最小二乘估计举例 ······ 303
9.2.4 一般多项式和正交多项式拟合 ······ 306
9.3 非线性最小二乘估计 ······ 310
9.4 最小二乘拟合 ······ 319
9.4.1 测量拟合值和残差 ······ 319
9.4.2 线性模型中 σ^2 的估计 ······ 323
9.4.3 正态性假设, 自由度 ······ 325
9.4.4 拟合优度 ······ 326
9.5 最小二乘法应用于直方图数据 ······ 328
9.6 最小二乘法应用于实验测量数据 ······ 333
9.7 线性约束的线性最小二乘估计 ······ 335
9.8 非线性约束的最小二乘估计 ······ 342
9.8.1 拉格朗日乘子法 ······ 342
9.8.2 误差估计 ······ 347
9.8.3 一般最小二乘拟合的自由度 ······ 349
9.9 最小二乘法求置信区间 ······ 350
9.9.1 单个参数的误差和置信区间 ······ 351
9.9.2 多个参数的误差和置信域 ······ 352
9.10 协方差矩阵未知的多个实验结果的合并 ······ 353
第 10 章 矩法, 三种估计方法的比较 ······ 359
10.1 简单的矩法 ······ 359
10.2 一般的矩法 ······ 361
10.3 举例 ······ 363
10.4 矩法、极大似然法和最小二乘法的比较 ······ 366
10.4.1 反质子极化实验的模拟 ······ 367
10.4.2 不同估计方法的应用 ······ 367

10.4.3 讨论 ······ 372
第 11 章　小信号测量的区间估计 ······ 376
11.1 经典方法 ······ 378
11.1.1 正态总体 ······ 379
11.1.2 泊松总体 ······ 381
11.2 似然比顺序求和方法 ······ 382
11.2.1 泊松总体 ······ 383
11.2.2 正态总体 ······ 384
11.3 改进的似然比顺序求和方法 ······ 385
11.4 考虑系统误差时泊松总体的区间估计 ······ 387
参考文献 ······ 389
《现代物理基础丛书》已出版书目 ······ 398

第 1 章　概率论初步

1.1　随机试验, 随机事件, 样本空间

自然界存在着在一定条件下必然发生的现象. 例如, 两个点电荷之间必定有相互作用力; 高处的重物必定落向地面; 水在一个大气压、100°C 条件下必然沸腾, 等等. 这些现象称为必然现象, 它们的过程和后果是完全确定的, 可以唯一地用一定的物理规律给以精确的描述. 如点电荷之间的作用力服从库仑定律, 真空中物体的下落过程服从自由落体规律.

但自然界还存在另一类性质不同的现象, 即使在"完全相同"的条件下对同一事物做多次测量或试验, 我们发现, 试验的结果并不一样, 一次单独的试验结果是不确定的, 因此无法用任何数学公式计算出来. 尽管每次试验的结果看来似乎杂乱无章, 但如做大量重复试验, 其结果却呈现出某种规律性. 我们来举例说明.

投掷一枚均匀硬币, 其结果或者是正面朝上, 或者是反面朝上. 我们无法预言任何一次投掷中硬币的哪一面朝上, 但当投掷次数很多时, 则正面朝上的次数约占 $1/2$.

掷一个骰子, 骰子的六个面分别刻有 1, 2, 3, 4, 5, 6 等数字. 每扔一次得到的点数是 1~6 中的哪一个数无法确定, 但在大量投掷中, 每一个点数的出现次数占总投掷数的 $1/6$ 左右.

上述两例的共同特征是: 个别试验中的结果是不确定的, 但大量重复试验的结果会出现某种规律性. 这类现象称为**随机现象**, 这种规律性称为**统计规律性**. 揭示随机现象的统计规律性的数学工具是概率论和数理统计.

扔骰子、扔硬币的试验有以下特性: 试验可以在"相同条件"下重复进行; 试验的结果不止一个, 但所有结果都已明确地知道; 每次试验结果究竟是其中的哪一种则无法确定. 具有这些性质的试验称为**随机试验**, 简称**试验**. 将某种随机试验 E 重复进行 n 次, 若各次试验的结果互不影响, 则称 n 次试验是**互相独立**的. 随机试验中可能出现的各种结果称为**随机事件**, 简称**事件**. 随机试验中每一种可能出现的结果是最简单、最基本的事件, 称为**基本事件**. 如扔骰子试验中, 每扔一次即是一次随机试验; "出现 1 点"、"出现 2 点"……"出现 6 点" 是 6 个基本事件; "出现大于 4 的点"、"出现偶数点" 是事件, 但不是基本事件. 试验中必定发生的事件叫**必然事件**, 不会发生的事件叫**不可能事件**. 如 "点数大于 0" 是必然事件, "点数大

于 6” 是不可能事件.

随机试验 E 的所有基本事件组成的集合称为 E 的**样本空间**, 记为 S. S 的元素是试验 E 的所有基本事件, 元素也称**样本点**. 例如, 扔硬币和扔骰子试验的样本空间可记为 $S_{\text{硬币}}$: {正面, 反面}, $S_{\text{骰子}}$: $\{1, 2, 3, 4, 5, 6\}$. 引入样本空间的概念后, 可以看到事件是样本空间的一个**子空间或子集**. 如 “点数大于 4” 是子集$\{5, 6\}$, “偶数点” 是子集$\{2, 4, 6\}$. 必然事件就是样本空间 S 的全域; 不可能事件是空集, 用 $\varnothing$ 表示.

现在我们来规定事件之间的关系及运算. 设随机试验 E 的样本空间为 S, 事件 $A, B, A_k(k=1, 2, \cdots)$ 为 E 的事件, 我们用下述符号表示它们之间不同的关系.

$A \subset B$(或 $B \supset A$) 称为事件 B **包含**事件 A, 表示事件 A 的发生必然导致事件 B 的发生. 这可用图 1.1 加以说明, 图中长方形表示样本空间 S, 圆 A 和圆 B 表示事件 A 和 B 的子集, 子集 A 含于子集 B 内.

$A = B$ 称为事件 A 与事件 B **相等**, 表示事件 A 包含事件 B 且事件 B 包含 A, 即 $B \supset A$ 且 $A \supset B$.

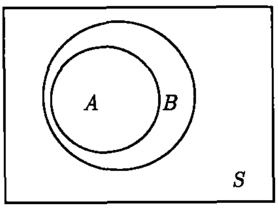


图 1.1 $A \subset B$

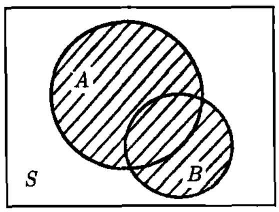


图 1.2 $A \cup B$

$A \cup B$ 称为事件 A 与事件 B **之和**, 表示事件 A 或事件 B 至少有一个发生. 图 1.2 中斜线部分表示 $A \cup B$. 类似地, $A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots \equiv \bigcup_{k=1}^{\infty} A_k$ 称为 $A_1, A_2, \cdots$ 之和, 表示这些事件中至少有一个发生.

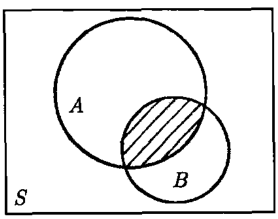


图 1 3 $A \cap B$

$A \cap B$ 或 AB 称为事件 A 与事件 B **之积**, 表示事件 A 和事件 B 同时发生. 图 1.3 中斜线部分表示 AB.

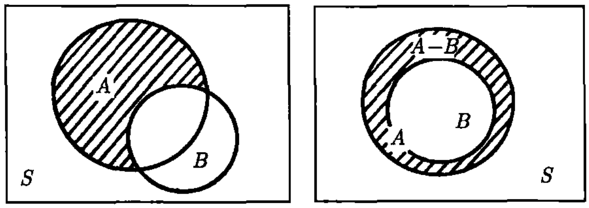
类似地, $A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots \equiv \bigcap_{k=1}^{\infty} A_k$ 为事件 $A_1, A_2, \cdots$ 之积, 表示这些事件同时发生.

$A - B$ 称为事件 A 与事件 B **之差**, 表示事件 A 发生而事件 B 不发生.

$A - B$ 如图 1.4 中斜线部分所示.

$AB = \varnothing$ 称为事件 A 与事件 B **互不相容**, 表示事件 A 与事件 B 不可能同

时发生. 图 1.5 是互不相容的两个事件 A 和 B 的图示. 基本事件之间是互不相容的.



(a) (b)

图 1.4 $A-B$

$A=\bar{B}$ 或 $B=\bar{A}$ 称事件 A 与事件 B **互逆**, 或 A, B 互为**对立事件**, 表示事件 A 和 B 中必有且仅有一个发生, 也即 $A\cup B=S$, $AB=\varnothing$ 图 1.6 中斜线部分为事件 B 的对立事件 $A=\bar{B}$. 由此规定可知, 互逆事件一定互不相容.

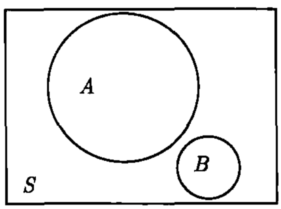


图 1.5 $AB=\varnothing$

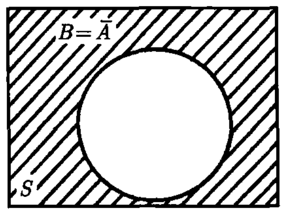


图 1.6 $A=\bar{B}$

样本空间的**划分**是十分有用的一个概念. 设 S 为随机试验 E 的样本空间, E 的一组事件 B_1, B_2, $\cdots$, B_n 两两互不相容, 且 B_1, B_2, $\cdots$, B_n 之和等于样本空间的全域, 即满足

$$\begin{cases} B_iB_j=\varnothing, \quad i\neq j,\ i,\ j=1,2,\cdots,n, \\ B_1\cup B_2\cup\cdots\cup B_n=S, \end{cases} \tag{1.1.1}$$

则称 B_1, B_2, $\cdots$, B_n 为样本空间 S 的一个划分. 图 1.7 是样本空间 S 的一个划分的图示. 显然样本空间的所有元素构成它的一个划分; 对立事件也是样本空间的一个划分.

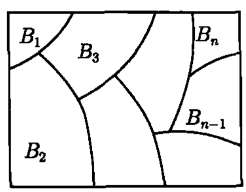


图 1.7 样本空间的划分

以扔骰子为例, 骰子面朝上的点数作为随机试验, 其样本空间是 $S=\{1,2,3,4,5,6\}$. 设一组事件 $B_1=\{1,2\}$、$B_2=\{3,4\}$, $B_3=\{5,\}$, 则 B_1,B_2,B_3 构成 S 的一个划分. 事件组 $C_1=\{1,2\}$, $C_2=\{2,3\}$, $C_3=\{4,5\}$ 不是 S 的划分, 它不满足式 (1.1.1) 的要求.

1.2 概　　率

所谓随机事件的概率, 指的是随机试验中该随机事件发生的可能性大小的数值表示. 历史上出现过几种不同的概率模型.

(1) 统计 (频率) 概型

重复进行一种随机试验, 共作了 N 次, 其中事件 A 出现 n 次 (称为事件 A 的频数), 比值 n/N 称为事件 A 在 N 次试验中出现的频率. 随着试验次数 N 的增加, 频率 n/N 的值将逐渐稳定于某个常数. 事件 A 的概率定义为试验次数 N 趋向无穷大的极限情形下的频率:

$$P(A) = \lim_{N\to\infty} \frac{n}{N}. \tag{1.2.1}$$

这样定义的概率称为**统计概率**或**频率概率**. 由以上定义可见, 事件的概率是随机试验中该事件发生的可能性大小的数量表述.

概率的上述定义相当直观, 但数学上不够严格, 而且无穷多次试验事实上无法实行.

(2) 古典概型

假设一种随机试验的样本空间包含有限个元素, 每个基本事件出现的可能性相等. 即随机试验的样本空间为 $S = \{e_1, e_2, \cdots, e_n\}$, 每个基本事件的概率相等, 则有

$$1 = P(S) = \sum_{i=1}^{n} P(e_i) = nP(e_i),$$

即

$$P(e_i) = \frac{1}{n}, \quad i = 1, 2, \cdots, n. \tag{1.2.2}$$

若事件 A 包含 k 个基本事件, 则事件 A 的概率为

$$P(A) = k/n. \tag{1.2.3}$$

这样的概率模型称为**等可能概型**或**古典概型**, 前面提到的掷硬币和扔骰子试验都属于古典概型.

(3) 几何概型

古典概型要求随机试验的样本空间只包含有限个可能性相等的元素 (或称样本点), 若随机试验的样本空间包含无限个可能性相等的样本点, 则不能按古典概型计算, 而需要用几何概型求解. 设随机试验 E 的样本空间 S 可用 $m(m = 1, 2, \cdots)$ 维空间中的一有界区域 Ω 表示, 其样本点具有均匀分布性质 (类似于古典概率中基本事件的等可能性), 事件 $A(A \subset S)$ 发生的可能性用区域 ω_A 表示, 则事件 A 发生的概率为

$$P(A)=\frac{\omega_A}{\Omega}, \tag{1.2.4}$$

称为**几何概率**. 它可以应用于无限可列个元素、甚至无限不可列个元素的情形 (含义见下文).

概率的上述模型, 每种定义都是针对不同的随机试验而设计的, 都很浅显直观, 但都存在一定的局限性, 数学上也不够严密. 1933 年前苏联科学家柯尔莫哥洛夫综合了前人的成果, 提出了概率的公理化定义, 从此, 概率论才成为一个严密的数学分支.

概率的公理化定义可简述如下.

设 S 为一随机试验 E 的样本空间, 对于 E 的任一事件 A, 满足如下条件的一个非负实函数 $P(A)$ 称为事件 A 的概率:

(1) $0 \leqslant P(A) \leqslant 1$, 对一切 $A \subset S$. (1.2.5)

(2) $P(S)=1$. (1.2.6)

(3) 对两两不相容的事件 $A_k(k=1,2,\cdots)$ 有

$$P\left(A_1 \cup A_2 \cup \cdots \cup A_n\right)=\sum_{k=1}^{n} P(A_k) \tag{1.2.7}$$

或

$$P\left(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots\right)=\sum_{k=1}^{\infty} P(A_k). \tag{1.2.8}$$

式 (1.2.7) 和式 (1.2.8) 分别称为概率的**有限可加性**和**可列可加性**, 它们分别适用于样本空间含有有限个元素和**无限可列个**元素的情形. 所谓无限可列个, 指满足两个条件: 有无限个元素, 但可以与自然数列 1, 2, 3, $\cdots$ 建立起一一对应的关系. 式 (1.2.6) 也称为样本空间概率的归一性, 它表示随机试验整个样本空间的概率和恒等于 1. 式 (1.2.5)~ 式 (1.2.8) 表明了概率的定义可以简单地归结为: 非负性、归一性和可加性.

概率的公理化定义在数学上是严密的, 但只规定了概率应满足的条件, 而没有给出计算事件 A 的概率 $P(A)$ 的方法, 因而对同一个样本空间, 只要符合这三个公理化条件, 概率可以有多种不同的定义. 例如, 前面讨论过的统计概率、古典概率和几何概率定义不同, 但都符合概率的公理化定义的要求; 在第 13 章贝叶斯统计中使用的贝叶斯概率, 其定义与这三者皆不同, 但同样满足公理化定义的要求.

在第 5 章中, 大数定律将证明, 在相当广泛的情形中, 当试验次数 N 趋向无穷时, 事件 A 的频率 n/N 与其概率 $P(A)$ 的严格定义值十分接近. 在实际使用时, 只要试验次数 N 充分大, 可用频率 n/N 作为概率 $P(A)$ 的近似值.

根据概率的定义, 立即可推导出概率的如下性质:

(1) 若 $A, \bar{A}$ 为一随机试验的互逆事件, 则有

$$P(A) + P(\bar{A}) = 1. \tag{1.2.9}$$

(2) 不可能事件的概率为 0, 即

$$P(\varnothing) = 0. \tag{1.2.10}$$

(3) 若事件 A 包含事件 B, 则

$$P(A) \geqslant P(B). \tag{1.2.11}$$

(4) 若 $A_1, \cdots, A_n$ 为一随机试验样本空间 S 的一个划分, 则由式 (1.2.6) 和式 (1.2.7) 立即得到

$$\sum_{i=1}^{n} P(A_i) = 1, \tag{1.2.12}$$

样本空间的所有基本事件的概率和等于 1. 式 (1.2.9) 可视为本式的特例.

(5) 若 $A \supset B$, 则

$$P(A - B) = P(A) - P(B). \tag{1.2.13}$$

(6)
$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.2.14}$$

由图 1.2 和图 1.3 可知, $A \cup B = A + B - AB$, 故得上式. 该公式也称为概率的加法定理. 推广到 n 个事件的一般情况, 设 $A_1, A_2, \cdots, A_n$ 是随机试验 E 的 n 个事件, 则有

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) =& P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) \\ &- P(A_2 A_3) - P(A_1 A_3) + P(A_1 A_2 A_3). \end{aligned} \tag{1.2.15}$$

利用数学归纳法, 令

$$\begin{aligned} S_1 &= \sum_{i=1}^{n} P(A_i), \\ S_2 &= \sum_{i<j=2}^{n} P(A_i A_j), \\ S_3 &= \sum_{i<j<k=3}^{n} P(A_i A_j A_k)^{①}, \\ &\vdots \end{aligned}$$

① 这里记号的意义 $\sum\limits_{i<j<k=3}^{4} P(A_i A_j A_k) = P(A_1 A_2 A_3) + P(A_1 A_2 A_4) + P(A_1 A_3 A_4) + P(A_2 A_3 A_4)$

则有

$$P\left(\bigcup_{i=1}^{n} A_i\right) \equiv P(A_1 \cup A_2 \cup \cdots \cup A_n)$$
$$= S_1 - S_2 + S_3 - \cdots - (-1)^n S_n. \tag{1.2.16}$$

如果式 (1.2.14) 中的事件 A 和事件 B 互不相容, 则加法定理的形式特别简单

$$P(A \cup B) = P(A) + P(B). \tag{1.2.17}$$

推广到 n 个互不相容的事件 $A_1, A_2, \cdots, A_n$, 则有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i). \tag{1.2.18}$$

1.3 条件概率, 独立性

设 A, B 为一随机试验的两个事件, 事件 A 的概率为 $P(A)$, 则在事件 A 发生的条件下事件 B 发生的概率称为**条件概率**, 表示为

$$P(B|A) = \frac{P(AB)}{P(A)}. \tag{1.3.1}$$

我们用图 1.3 来说明条件概率的含义. 图中斜线部分的面积现在表示事件 $A \cap B$ 的概率 $P(AB)$, 区域 A, B 的面积表示事件 A 和事件 B 的概率. 事件 A 发生条件下事件 B 发生的概率, 即条件概率 $P(B|A)$ 为事件 A 和事件 B 的共有区域与事件 A 的区域之面积比, 因此有式 (1.3.1). 类似地, 事件 B 发生条件下事件 A 发生的概率可表示为

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{1.3.2}$$

由上面两式立即得到

$$P(AB) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B), \tag{1.3.3}$$

该式称为概率的**乘法定理**.

可以证明, 条件概率具有概率的一般性质: 非负性、归一性和可加性, 即

$$0 \leqslant P(A|B) \leqslant 1, \tag{1.3.4}$$

$$P(S|B) = 1, \tag{1.3.5}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} P(A_i|B). \tag{1.3.6}$$