石油大学成人教育教材


# 概率论与数理统计

常兆光 王清河 施宝正 编

石油工业出版社

石油大学成人教育教材

# 概率论与数理统计

常兆光　王清河　施宝正　编

石油工业出版社

## 内 容 提 要

本书是按照普通高等理工院校成人教育《概率论与数理统计》课程基本要求编写的，内容包括随机事件与概率、随机变量及分布、随机变量的数字特征、大数定律和中心极限定理、数理统计的基本概念、参数估计、假设检验、回归分析、方差分析与正交试验设计等。每章均配适量练习。

本书可作为普通高等理工院校成人、网络教育《概率论与数理统计》课程的教材与教学参考书，也可作为工程技术人员的参考书。

**图书在版编目(CIP)数据**

概率论与数理统计/常兆光，王清河，施宝正编．
北京：石油工业出版社，2003.9
石油大学成人教育教材
ISBN 7-5021-4388-2

Ⅰ．概… Ⅱ．①常…②王…③施… Ⅲ．①概率论—高等教育：成人教育—教材②数理统计—高等教育：成人教育—教材 Ⅳ．021

中国版本图书馆 CIP 数据核字(2003)第 082342 号

石油工业出版社出版
（100011 北京安定门外安华里二区一号楼）
汉魂图文设计公司排版
石油工业出版社印刷厂印刷
新华书店北京发行所发行
*
850×1168 毫米 32 开本 9.25 印张 236 千字 印 5301—10300
2003 年 9 月北京第 1 版 2005 年 1 月北京第 2 次印刷
ISBN 7-5021-4388-2/O·12
定价：17.00 元

# 前　言

《概率论与数理统计》是研究和揭示随机现象统计规律性的数学学科，是高等成人教育本、专科各专业的一门重要基础理论课，是基础数学与工程专业课程的桥梁，在各行各业都有广泛的应用。本书是按照普通高等理工院校成人教育《概率论与数理统计》课程基本要求编写的。只要具备高等数学、线性代数的知识，就能顺利地学好其中的内容，达到该课程的基本要求。由于该课程内容比较抽象、思维方式比较独特、知识面涉及广，使自学具有一定的困难，因此本书的编写力求重点突出、深入浅出、通俗易懂，具有较强的实用性。对基本概念、重要公式和定理，注意其实际意义的解释和说明，讲清其实质，并在每节中配有较多的例题。通过各种类型例题的分析、运算，进一步加深和巩固所学的理论与概念，培养学生分析与解决随机现象问题的能力，培养学生的创造性思维能力。

本书可作为普通高等理工院校成人、网络教育《概率论与数理统计》课程的教材与教学参考书，也可作为工程技术人员的参考书。

在教材处理上，部分章、节打 * 号的内容可不作为基本要求，留给读者自学。建议第一章至第五章的参考学时为 40 学时左右。在例题的选择上也分两种类型，一种例题是比较简单且易于为人们所理解的，这种例题主要是用来便于引入或解释概念、定义和定理；另一种例题是提高型的，主要用来培养分析问题和解决问题的能力，进一步加深对概念、定义和定理的理解和掌握。个别打 * 号的例题不作为基本要求。习题中打 * 号的题目是为拓宽知识面之用，也不作为基本要求。

在本书编写过程中，石油大学成人（网络）教育学院、应用数学系

给予了大力支持；王才经教授在百忙之中对本书初稿进行了详细认真的审阅，并提出了许多宝贵意见，在此一并表示感谢。

由于编者水平所限，错误与不足之处在所难免，敬请读者批评指正。

编　者

2003 年 5 月

# 目　　录

第一章　随机事件与概率 …… 1
第一节　随机试验与随机事件 …… 1
第二节　频率与概率 …… 7
第三节　古典概型(等可能概型) …… 12
第四节　几何概型* …… 16
第五节　条件概率 …… 17
第六节　事件的独立性 …… 25
第七节　贝努利概型与二项概率公式 …… 29
小结 …… 31
习题 …… 34
第二章　随机变量及其分布 …… 37
第一节　随机变量及其分布函数 …… 38
第二节　离散型随机变量及其概率分布 …… 40
第三节　连续型随机变量及其概率分布 …… 47
第四节　随机向量及其分布 …… 56
第五节　随机变量的独立性 …… 71
第六节　随机变量函数的分布 …… 73
小结 …… 83
习题 …… 90
第三章　随机变量的数字特征 …… 95
第一节　数学期望 …… 96
第二节　方差 …… 105
第三节　相关系数与相关阵* …… 115

小结…………………………………………………………………… 123
习题…………………………………………………………………… 126

第四章　大数定律与中心极限定理………………………………… 130
第一节　大数定律………………………………………………… 130
第二节　中心极限定理…………………………………………… 135
小结…………………………………………………………………… 139
习题…………………………………………………………………… 140

第五章　数理统计初步……………………………………………… 142
第一节　数理统计的基本概念…………………………………… 143
第二节　参数估计………………………………………………… 153
第三节　假设检验………………………………………………… 176
小结…………………………………………………………………… 186
习题…………………………………………………………………… 189

第六章　回归分析…………………………………………………… 193
第一节　一元线性回归…………………………………………… 194
第二节　多元线性回归…………………………………………… 203
第三节　可线性化的非线性回归………………………………… 212
小结…………………………………………………………………… 215
习题…………………………………………………………………… 218

第七章　方差分析与正交试验设计………………………………… 219
第一节　单因素方差分析………………………………………… 219
第二节　多因素方差分析………………………………………… 225
第三节　正交试验设计…………………………………………… 232
小结…………………………………………………………………… 244
习题…………………………………………………………………… 248

附录…………………………………………………………………… 250
附表 1　标准正态分布表 ………………………………………… 250

附表 2　泊松分布表 …………………………………………… 251
附表 3　$t$ 分布表 …………………………………………… 253
附表 4　$\chi^2$ 分布表 …………………………………………… 254
附表 5　$F$ 分布表 …………………………………………… 258
附表 6　正交表 …………………………………………… 270

参考文献 …………………………………………… 284

# 第一章　随机事件与概率

随机事件的概率是概率论研究的基本内容。本章将主要介绍随机事件、随机事件的概率、概率的基本性质、条件概率以及计算概率常用到的几个重要公式。通过本章的学习，应初步掌握处理随机现象的基本思想和方法，学会运用概率方法分析和解决实际问题。基本要求如下：

(1) 理解随机事件和样本空间的概念，掌握事件之间的关系与基本运算。

(2) 理解频率的概念和概率的公理化定义及古典概率定义，会进行简单古典概率的计算。

(3) 掌握概率的基本性质并会应用这些性质进行概率的计算。

(4) 理解条件概率的概念，掌握概率的乘法公式、全概率公式和贝叶斯(Bayes)公式以及应用这些公式进行概率计算。

(5) 理解事件的独立性的概念，会应用事件的独立性进行概率计算。掌握贝努利(Bernoulli)概型和二项概率的计算方法。

## 第一节　随机试验与随机事件

### 一、必然现象与随机现象

在生产实践、科学实验和日常生活中，人们观察到的现象一般可分为两种类型，一类是确定性现象，即在一定条件下，某些事情一定发生或一定不发生的现象，也称为必然现象。例如：纯净水在一个标准大气压下，加热到100℃必然沸腾；上抛的物体必然下落等等。早期的科学就是研究这一类现象的规律性，所应用的数学工具如高等数学、几何、代数等是大家熟悉的。但人们逐渐发现另一类现象，它是事前不可预言的。即在一定条件下，可能发生也可能不发生的现

象，这一类现象我们称之为偶然现象或随机现象。例如：抛掷一枚均匀硬币，是正面朝上还是反面朝上；从一批含有次品的产品中任取一件，取得的产品是正品还是次品等，这些都是事前不能肯定的。类似的例子还可以举出很多。

必然现象遵循必然性规律，人们根据已知的事实推断它将发生的结果。随机现象具有明显的不确定性（随机性），就一次试验（或观察）而言，其结果难以确定；但若进行大量重复试验，其结果就会呈现出某种规律性，即所谓统计规律。概率论与数理统计的任务就是要研究和揭示随机现象的这种统计规律性。

**二、随机试验**

在工农业生产、科学实验和现实生活中，我们遇到过各种各样的试验。在概率论中，我们把试验作为一种广泛的术语，它包括各种各样的科学试验，甚至对某一事物的某一特征的观察也可认为是一种试验。例如（以 $E$ 或 $E$ 加一个下标表示试验）：

$E_1$：掷一枚均匀硬币，观察正、反面出现的情况（正面记为 $H$，反面记为 $T$）；

$E_2$：将一枚均匀硬币抛两次，观察正、反面出现的情况；

$E_3$：将一枚均匀硬币抛两次，观察正面出现的次数；

$E_4$：掷一颗均匀骰子，观察出现的点数；

$E_5$：一名射手进行射击，直到击中为止，观察射击次数；

$E_6$：从一批灯泡中任意抽取一只，测其寿命。

不难看出，以上几个试验具有下面三个共同特征：

(1) 试验可以在相同条件下重复进行；

(2) 每次试验的可能结果不止一个，但能事先明确试验的所有可能结果；

(3) 进行一次试验之前，不能确定哪一个结果会出现。

在概率论中，我们把具有以上三个特征的试验称为随机试验，简称为试验，记作 $E$。

### 三、随机事件与样本空间

我们将试验 $E$ 的所有可能出现的结果组成的集合称为 $E$ 的样本空间，记作 $\Omega$。$\Omega$ 中的每个元素（可能结果）称为样本点。例如上述试验 $E_i(i=1,2,3,4,5,6)$ 的样本空间 $\Omega_i$ 分别为：

$$\Omega_1=\{H,T\}$$

$$\Omega_2=\{HH,HT,TH,TT\}$$

$$\Omega_3=\{0,1,2\}$$

$$\Omega_4=\{1,2,3,4,5,6\}$$

$$\Omega_5=\{1,2,3,4,\cdots\}$$

$$\Omega_6=\{t\mid t\geqslant 0\}$$

需要指出的是，样本空间中的样本点是由试验目的所确定的。例如 $E_2$ 和 $E_3$ 同是将一枚均匀硬币抛两次，由于试验目的不同，其样本空间也就不一样。

样本空间包含了试验 $E$ 的所有可能结果，我们将每一个可能的结果称为随机事件（简称为事件），通常用大写字母 $A$、$B$、$C$ 等表示。只包含一个样本点的事件称为基本事件。例如掷骰子试验中，每一个可能出现的点数都是基本事件。而由两个或两个以上的基本事件（样本点）组成的事件称为复合事件。例如将一枚均匀硬币连续掷两次，观察正、反面出现的情况，其样本空间 $\Omega=\{HH,HT,TH,TT\}$，则两次都出现正面为一基本事件，记作 $A=\{HH\}$；至少有一次出现正面为一复合事件，记作 $B=\{HH,HT,TH\}$。可以看出，$B$ 事件是由三个类似于事件 $A$ 的基本事件所组成的，即这三个事件只要有一个发生就认为 $B$ 事件发生。

在每次试验中，一定发生的事件叫做必然事件，记作 $\Omega$；而一定不发生的事件叫做不可能事件，记作 $\phi$。

需要指出的是，无论是必然事件、随机事件还是不可能事件，都是相对“一定条件”而言的。条件发生变化，事件的性质也发生变化。例如抛掷两颗骰子，“出现的点数之和为 5 点”及“出现的点数之和大于 5 点”都是随机事件。若同时抛掷 6 颗骰子，“出现的点数之和为

5 点”则是不可能事件了；而“出现的点数之和大于 5 点”则是必然事件了。为了以后讨论问题方便，通常将必然事件和不可能事件看成是特殊的随机事件。

**四、事件及其运算关系**

由样本空间的定义知，它是随机试验中所有可能出现的结果(样本点)组成的集合，因此随机事件又可理解为样本空间中具有某种特性的样本点构成的集合，而基本事件可看成此集合的元素。这样一来，集合论中集合之间的运算均可推广到事件之间的运算中来。

若记 $\Omega$：样本空间(必然事件)；$\phi$：不可能事件；$e$：基本事件；$A$，$B$，$A_1$，$A_2$，…，$A_n$：随机事件，则事件之间的运算关系如下：

(1) 包含关系：若 $A \subset B$，则称事件 $B$ 包含事件 $A$，表示事件 $A$ 发生必导致事件 $B$ 发生，如图 1—1 所示。

对任一事件 $A$，都有 $\phi \subset A \subset \Omega$。

(2) 相等关系：若 $A \subset B$ 且 $B \subset A$，则称事件 $A$ 等于事件 $B$，记作 $A = B$。

(3) 和事件：事件 $A \cup B$ 称为事件 $A$ 与事件 $B$ 的和事件，表示事件 $A$ 与事件 $B$ 至少有一个发生，如图 1—2 所示。

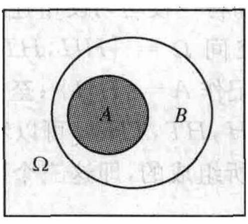


图 1—1

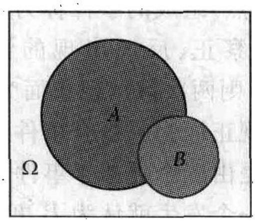


图 1—2

事件 $A_1 \cup A_2 \cup \cdots \cup A_n \triangleq \bigcup_{i=1}^{n} A_i$ 称为有限个事件的和事件，表示事件 $A_1$，$A_2$，…，$A_n$ 至少有一个发生。

类似地，事件 $A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots \triangleq \bigcup_{i=1}^{\infty} A_i$ 称为无限多个事件的和事件，表示事件 $A_1$，$A_2$，…，$A_n$，… 至少有一个发生。

(4) 积事件：事件 $A \cap B$ 称为事件 $A$ 与事件 $B$ 的积事件，简记为 $AB$，表示事件 $A$ 与事件 $B$ 同时发生，如图 1—3 所示。

事件 $A_1 \cap A_2 \cap \cdots \cap A_n \triangleq \bigcap_{i=1}^{\infty} A_i$ 称为有限个事件的积事件，表示事件 $A_1, A_2, \cdots, A_n$ 同时发生。

类似地，事件 $A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots \triangleq \bigcap_{i=1}^{\infty} A_i$ 称为无限多个事件的积事件，表示事件 $A_1, A_2, \cdots, A_n, \cdots$ 同时发生。

(5) 差事件：事件 $A-B$ 称为事件 $A$ 与事件 $B$ 的差事件，表示事件 $A$ 发生而事件 $B$ 不发生，如图 1—4 所示。

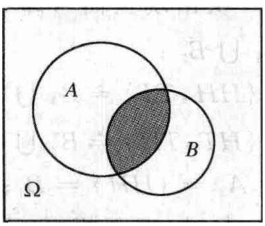


图 1—3

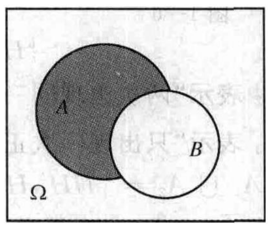


图 1—4

(6) 互不相容(互斥)事件：若 $AB = \phi$，则称事件 $A$ 与事件 $B$ 互不相容，表示事件 $A$ 与事件 $B$ 不同时发生，如图 1—5 所示。基本事件是互不相容的。

类似地，若 $\forall i \neq j, A_iA_j = \phi(i、j = 1, 2, \cdots)$，，则称 $A_1, A_2, \cdots, A_n$，…为两两互不相容的。

若 $A_1, A_2, \cdots, A_n$ 两两互不相容，且在每次试验中事件 $A_1, A_2, \cdots, A_n$ 必发生其中之一，即 $\bigcup_{i=1}^{n} A_i = \Omega$，且 $A_iA_j = \phi(i \neq j)$，则称 $A_1, A_2, \cdots, A_n$ 为互不相容事件完备组。

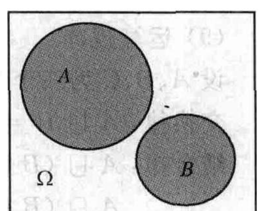


图 1—5

(7)对立事件(逆事件)：若 $AB = \phi$ 且 $A \cup B = \Omega$，则称事件 $A$

与事件 $B$ 互逆，又称事件 $A$ 为事件 $B$ 的对立事件（或事件 $B$ 为事件 $A$ 的对立事件），记作 $A=\overline{B}$（$\overline{B}$ 表示 $B$ 不发生），$B=\overline{A}$，如图 1—6 所示。

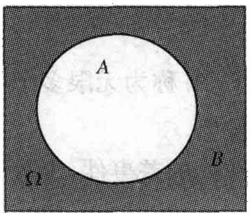


图 1—6

需要指出的是，由(6)、(7)知，若事件 $A$ 与事件 $B$ 互逆，则事件 $A$ 与事件 $B$ 一定互不相容，但反之则不一定。

**例 1—1** 在试验 $E_2$ 中，若记 $B_1=\{HH\}$，$B_2=\{HT\}$，$B_3=\{TH\}$，$B_4=\{TT\}$，则

$A_1$ 表示"第一次出现正面"，即 $A_1=\{HH,HT\}\triangleq B_1\cup B_2$；

$A_2$ 表示"两次出现同一面"，即 $A_2=\{HH,TT\}\triangleq B_1\cup B_4$；

$A_3$ 表示"只出现一次正面"，即 $A_3=\{HT,TH\}\triangleq B_2\cup B_3$；

那么，$A_1\cup A_2=\{HH,HT,TT\}$；$A_1\cap A_2=\{HH\}=B_1$；$A_1-A_2=A_1\overline{A}_2=B_2$。由于 $A_2A_3=\phi$，故 $A_2$ 与 $A_3$ 互不相容；又由于 $A_2\cup A_3=\Omega$，所以 $A_2$ 与 $A_3$ 互逆。

(8)德·摩根定律(对偶原理)：

$$\overline{AUB}=\overline{A}\cap\overline{B},\overline{A\cap B}=\overline{A}\cup\overline{B}$$

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A_i},\quad \overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A_i}$$

(9) 运算规律：

设 $A$、$B$、$C$ 为三个事件，则有

交换律：$A\cup B=B\cup A,A\cap B=B\cap A$

结合律：$A\cup(B\cup C)=(A\cup B)\cup C$，

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

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

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

**例 1—2** 设 $A$、$B$、$C$ 为三个事件，试用事件之间的运算关系表示下列事件：

(1) $A$ 发生而 $B$ 与 $C$ 都不发生；

(2) $A$、$B$、$C$ 恰有一个发生；

(3) $A$、$B$、$C$ 至少有一个发生；

(4) $A$、$B$、$C$ 至多有两个事件发生。

解：(1)$A$ 发生而 $B$ 与 $C$ 都不发生，也就是 $A$、$\bar{B}$、$\bar{C}$ 同时发生，即 $A\cap\bar{B}\cap\bar{C}$ 或 $A-B-C$；类似可得

(2) $(A\cap\bar{B}\cap\bar{C})\cup(\bar{A}\cap B\cap\bar{C})\cup(\bar{A}\cap\bar{B}\cap C)$；

(3) $A\cup B\cup C$；

(4) $\bar{A}\cup\bar{B}\cup\bar{C}$ 或 $\overline{ABC}$ 。

## 第二节　频率与概率

一个随机试验有许多可能的结果，而在许多情况下，我们想知道的往往是随机事件(可能结果)发生的可能性大小。如建造水坝时，为确定坝高，需要知道建造水坝地段每年最大洪水到达某高度的可能性大小。在概率论中，将描述随机事件 $A$ 发生的可能性大小的数记为 $P(A)$，称 $P(A)$ 为随机事件 $A$ 的概率。那么如何确定随机事件 $A$ 的概率呢？一种方法是通过反复试验来确定，为此先来讨论频率的概念。

### 一、频率

**引例 1—1**　将一枚均匀硬币在同样条件下连续掷 $n$ 次，用 $A$ 表示出现正面这一事件，$n(A)$表示 $n$ 次重复试验中 $A$ 出现的次数，则 $\frac{n(A)}{n}$ 在一定程度上能反映事件 $A$ 发生的可能性大小，将其记为

$$f_n(A)\triangleq\frac{n(A)}{n}$$

试验表明(见表 1—1 及表 1—2)，随着 $n$ 的增大，$\frac{n(A)}{n}$ 在 $\frac{1}{2}$ 附近波动的幅度越来越小，逐渐稳定于 $\frac{1}{2}$，$\frac{1}{2}$ 这个值称为 $f_n(A)$ 的稳定值，通常把这个稳定值称为事件 $A$(出现正面)的概率。$f_n(A)$称

为 $n$ 次重复试验中 $A$ 出现的频率，$n(A)$ 称为 $n$ 次重复试验中 $A$ 出现的频数。下面给出频率的一般定义。

**表 1—1**

| 实验序号 | $n=5$ | | $n=50$ | | $n=500$ | |
|---|---|---|---|---|---|---|
| | $n(A)$ | $f_n(A)$ | $n(A)$ | $f_n(A)$ | $n(A)$ | $f_n(A)$ |
| 1 | 2 | 0.4 | 22 | 0.44 | 251 | 0.502 |
| 2 | 3 | 0.6 | 25 | 0.50 | 249 | 0.498 |
| 3 | 1 | 0.2 | 21 | 0.42 | 256 | 0.512 |
| 4 | 5 | 1.0 | 25 | 0.50 | 253 | 0.506 |
| 5 | 1 | 0.2 | 24 | 0.48 | 251 | 0.502 |
| 6 | 2 | 0.4 | 21 | 0.42 | 246 | 0.492 |
| 7 | 4 | 0.8 | 18 | 0.36 | 244 | 0.488 |
| 8 | 2 | 0.4 | 24 | 0.48 | 258 | 0.516 |
| 9 | 3 | 0.6 | 27 | 0.54 | 262 | 0.524 |
| 10 | 3 | 0.6 | 31 | 0.62 | 247 | 0.494 |

**表 1—2**

| 实验者 | $n$ | $n(A)$ | $f_n(A)$ |
|---|---|---|---|
| 蒲丰 | 4040 | 2048 | 0.5070 |
| K. 皮尔逊 | 12000 | 6019 | 0.5015 |
| K. 皮尔逊 | 24000 | 12012 | 0.5005 |

**定义 1—1** 设 $E$ 为随机试验，$A$ 为其中任一事件，$n(A)$ 为事件 $A$ 在 $n$ 次重复试验中出现的次数，则称比值 $\frac{n(A)}{n}$ 为 $n$ 次试验中 $A$ 出现的频率，记为

$$f_n(A) \triangleq \frac{n(A)}{n} \tag{1—1}$$

其中，$n(A)$ 称为事件 $A$ 在 $n$ 次重复试验中出现的频数。当 $n$ 增大时，$f_n(A)$ 逐渐稳定于某一个确定值 $P(A)$，称 $P(A)$ 为频率的稳定值。

由频率的定义不难看出，$f_n(A)$ 具有以下性质：

(1) 非负性：$0 \leqslant f_n(A) \leqslant 1$；

(2) 规范性：$f_n(\Omega) = 1$；

(3) 可加性：若事件 $A$ 与事件 $B$ 互不相容，则

$$f_n(A \cup B) = f_n(A) + f_n(B)$$

进一步，若 $A_1, A_2, \cdots, A_m$ 两两互不相容，则

$$f_n(\bigcup_{i=1}^{m} A_i) = \sum_{i=1}^{m} f_n(A_i)$$

由掷硬币试验可见，用频率来刻画事件 $A$ 发生的可能性大小比较直观，但它有随机波动的缺陷，即在一定条件下做重复试验，其结果可能是不一样的。因此，用频率的稳定值来刻画事件 $A$ 发生的可能性大小是比较恰当的，从而得出概率的统计性定义如下。

**定义 1—2** 在不变条件下做大量重复试验，称在重复试验中事件 $A$ 发生的频率的稳定值 $p$ 为事件 $A$ 的概率，记为 $P(A)$。

尽管由频率的稳定值可得概率 $P(A)$，但我们不可能对每个事件都通过做反复试验以求得 $P(A)$。为此，我们以频率的性质和频率的稳定值 $P(A)$ 为背景，采用抽象化方法给出概率 $P(A)$ 的一般定义。

## 二、概率的公理化定义

**定义 1—3** 设 $E$ 为随机试验，$\Omega$ 为它的样本空间，对 $E$ 中的每一个事件 $A$ 都赋予一个实数，记为 $P(A)$，且满足

(1) 非负性：$0 \leqslant P(A) \leqslant 1$；

(2) 规范性：$P(\Omega) = 1$；

(3) 可加性：若 $A_1, A_2, \cdots, A_n \cdots$，两两互不相容，有

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1—2}$$

则称 $P(A)$ 为事件 $A$ 的概率。

由定义得概率 $P(A)$ 的性质如下：

(1) $P(\phi) = 0$；

(2) 若 $A_1, A_2, \cdots, A_n$ 两两互不相容，则