21世纪高等学校大学数学系列教材

# 概率论与数理统计

陈生安　钟绍军　彭　娟　主编

科学出版社

21世纪高等学校大学数学系列教材

# 概率论与数理统计

陈生安　钟绍军　彭　娟　主编



科学出版社
北　京

## 内 容 简 介

本书分为两部分共9章，1～5章为概率论部分，主要介绍随机事件及其概率、随机变量及其分布、随机变量的数字特征、多维随机变量、大数定律与中心极限定理等；6～9章为数理统计部分，主要介绍抽样分布、参数估计、假设检验、方差分析与回归分析等.

本书可作为高等学校本、专科理科(非数学专业)，工科各专业概率论与数理统计课程的教材，也可以供科技工作者和学习应用数学知识者参考.

**图书在版编目(CIP)数据**

概率论与数理统计/陈生安，钟绍军，彭娟主编. —北京：科学出版社，2013.2
21世纪高等学校大学数学系列教材
ISBN 978-7-03-036667-2

Ⅰ.概… Ⅱ.①陈…②钟…③彭… Ⅲ.①概率论—高等学校—教材②数理统计—高等学校—教材 Ⅳ.O21

中国版本图书馆CIP数据核字(2013)第026785号

责任编辑：吉正霞/责任校对：董艳辉
责任印制：彭 超/封面设计：苏 波

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
武汉市新华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
*
开本：B5(720×1000)
2013年2月第 一 版 印张：14 3/4
2013年2月第一次印刷 字数：284 000

**定价：28.00元**
(如有印装质量问题，我社负责调换)

# 《概率论与数理统计》编委会

**主　编**　陈生安　钟绍军　彭　娟

**副主编**　库在强　桂咏新　周涌

**编　委**　（按姓氏笔画为序）

任全玉　华　锐　刘君娥　吴　敏

库在强　陈　敏　陈生安　周　涌

钟绍军　桂咏新　彭　娟

# 前 言

概率论与数理统计是数学的一个有特色且又十分活跃的分支,一方面,它有别开生面的研究课题,有自己独特的概念和方法,内容丰富,结果深刻;另一方面,它与其他学科又有紧密的联系,是近代数学的重要组成部分.由于它近年来突飞猛进的发展与应用的广泛性,目前已发展成为一门独立的一级学科.概率论与数理统计的理论与方法已广泛应用于工业、农业、军事和科学技术中,如预测和滤波应用于空间技术和自动控制,时间序列分析应用于石油勘测和经济管理,马尔可夫(Markov)过程与点过程统计分析应用于地震预测等;同时它又向基础学科、工科学科渗透,与其他学科相结合发展成为边缘学科,这是概率论与数理统计发展的一个新趋势.

本书致力于介绍概率论与数理统计的基本思想与方法,内容上按照教育部现时使用的《概率论与数理统计课程基本要求》所规定的广度与深度设置,可作为高等学校本、专科理科(非数学专业),工科各专业概率论与数理统计课程的教材,也可以供各类有关专业技术人员参考.

本书分为两部分共 9 章,1～5 章为概率论部分,主要介绍随机事件及其概率、随机变量及其分布、随机变量的数字特征、多维随机变量、大数定律与中心极限定理等;6～9 章为数理统计部分,介绍抽样分布、参数估计、假设检验、方差分析与回归分析等.在内容的叙述上,本书侧重于剖析解决问题的思想与方法,强调基本知识和技巧的应用.在全书的结构上,注重内容之间的条理性与逻辑性,力求说理简明,思路清晰.

为反映实践教学的需要,作为本书特色之一,设置了 4 个实验:随机数的模拟(第 1 章)、概率分布图像的绘制(第 2 章)、假设检验(第 8 章)和方差分析(第 9 章),介绍 Excel 软件的相关应用,以加深读者对相关知识点的理解.

同时为了让读者在学习中体会概率论与数理统计这门课程应用性强的特点,本书的例题与习题力求源自生活中的实际问题,同时强调抽象思维与建模思想在解决问题中的应用,藉此希望能给读者带来更多的启发.另外,为满足读者不同层次的需要,本书除在每小节配有习题之外,在每章末尾还安排了复习题,这些复习题部分选自历年考研真题,一方面可供读者检验自己各章的掌握情况,另一方面也可供读者复习备考使用.

本书章节中标有“*”的内容,各专业可根据学时情况由教师酌情取舍.

本书的1～5章由彭娟编写，第6章由周涌编写，第7章由桂咏新编写，第8、9章两章由钟绍军编写. 全书由陈生安负责编审；陈敏、吴敏等老师参与了本书的审阅工作，并提出了许多宝贵的意见，编撰组全体成员对他们表示衷心的感谢.

由于编者的水平有限，本书难免存在疏漏，今后将在工作中予以纠正，诚请广大读者批评指正.

编　者

2012.11.20

# 目　　录

# 第1章　随机事件及其概率

## 1.1　随机事件

### 1.1.1　必然现象与随机现象

**必然现象**　也称确定性现象，它是一类可事先预言的、结果肯定的现象. 例如，抛掷一枚硬币它必定会落下；匀速行驶的列车，能根据它的速度确定它到达目的地的时间；水稻从生长到收获必定要经历发芽、育秧、长叶、吐穗、扬花、结实这几个阶段. 早期的科学研究都围绕必然现象展开，形成了一系列的经典学科.

**随机现象**　也称偶然性现象，相对于必然现象，它是指在相同条件下重复进行观察或试验，每次结果不能预知、不可确定的现象. 在日常生活中，我们所关心的随机现象有这样一些：明天是否会下雨，购买的彩票能否中奖，喜爱的球队能否获胜等. 虽然概率论成为一门严谨数学学科的历史并不算早，但是人类对随机现象的观察研究却是由来已久. 例如，古埃及人通过长期观察获得了尼罗河的涨落规律，我国古代很早就发现新生婴儿中男婴和女婴的比例接近1∶1等.

在对随机现象的长期观察和研究中人们发现，随机现象并非毫无规律可循，而是表现出一种统计规律性. 所谓**统计规律性**是指：在一定条件下进行观察或试验，有多种可能结果，事前不能预知哪一个结果会发生，但大量重复观察或试验时，所得的结果却呈现某种规律性. 在现实生活中，统计规律性时有应用但人们并不知道. 例如，餐厅的老板根据以往的经验确定某日的食材储备量，他的“经验”正是建立在统计规律性之上：该日的食材消耗量并不能预知，但是由过往同期的销售资料得到的规律却能对当日购进多少食材提供指导. 这样的例子比比皆是. 人类能够揭开随机现象的神秘面纱，正是因为“统计规律性”的存在. 而概率论与数理统计正是研究随机现象的统计规律性的一门数学学科.

### 1.1.2　随机试验与样本空间

随机现象是一个具有普遍意义的概念，我们具体要研究的是其中的一个或者一类. 为了表述的方便与严谨，我们对研究的对象做一个清晰的界定，这就是随机试验的概念. 对自然界中某个现象的一次观察或一次科学试验称为**随机试验**

(random experiment),如果它满足下述条件:

(1) 能够在特定条件下重复进行;

(2) 试验的所有可能结果已知;

(3) 每次试验进行之前不能预知哪个结果出现.

下面,我们将用字母 $E$ 表示一个随机试验,并且将其简称为试验.

下面是一些随机试验的例子.

**例 1.1.1** 抛掷一枚质地均匀的硬币,观察哪一面朝上.

这个随机试验包含两个可能结果,它是最简单也是最基本的随机试验,在后面的多个场合,我们都会再次提到这个试验.

**例 1.1.2** 抛掷两颗质地均匀的骰子,观察出现的点数.

**例 1.1.3** 观察售票处单位时间内排队等候的人数.

**例 1.1.4** 超市任选一天,调查该日的销售额.

**例 1.1.5** 向飞镖盘发射一只飞镖,假设飞镖不会落到盘外,考虑飞镖的落点.

不难验证例 1.1.1 ~ 例 1.1.5 即之前提到的随机试验,这些随机试验包含有限或无限个可能结果.我们不能事先预知哪个结果会发生,但是通过大量重复试验能发现所有可能结果以何种规律出现.为了更加简洁地叙述这些可能结果,我们引入样本空间这个概念.如果令每个可能结果与一个元素对应,那么所有的可能结果就构成了一个集合,这个集合称为试验 $E$ 的**样本空间**(sample space),通常用 $S$ 表示.而样本空间的每一个元素,也就是试验的每一个结果,称为一个**样本点**(sample point).

下面尝试写出例 1.1.1 ~ 例 1.1.5 的样本空间.对于例 1.1.1,可以直接用"正"、"反"表示两个不同的结果,也可以更为简单地用"1"代表"正","0"代表"反";对于例 1.1.2 ~ 例 1.1.4,可以直接用对应的数字来表示结果;对于例 1.1.5,以盘心为原点建立二维平面的直角坐标,设盘的半径为 $r$,那么飞镖的所有可能落点包含在一个半径为 $r$ 的圆内.如此处理之后,这些随机试验的样本空间便抽象成如下一些集合:

$$S_1:\{0,1\}$$

$$S_2:\{(1,1),(1,2),\cdots,(1,6),(2,1),(2,2),\cdots,(6,6)\}$$

$$S_3:\{0,1,2,3,\cdots\}$$

$$S_4:\{x \mid x \geqslant 0\}$$

$$S_5:\{(x,y) \mid x^2+y^2 \leqslant r^2\}$$

### 1.1.3 随机事件

对于一些随机试验,我们关心的也许并不是每个可能结果.如例 1.1.2,对于

掷骰子的人而言，他们可能对两个骰子的和是多少更感兴趣；对于例 1.1.4，超市关心的并不是销售额是否为某个具体数值，而是它是否达到某个范围. 这些“事件”可以看成样本空间的一个子集. 所以，我们就将随机试验 $E$ 的样本空间 $S$ 的子集称为 $E$ 的**随机事件**，简称**事件**，用大写字母 $A,B,C,\cdots$ 来表示. 如果在一次试验中，事件中包含的某个样本点出现，我们就称该事件发生.

样本空间 $S$ 有两个特殊的子集，一个是其自身 $S$，另一个是空集 $\varnothing$. 我们不难找到它们与事件之间的对应关系：$S$ 包含了试验的所有可能结果，它是必然发生的，称为**必然事件**；而 $\varnothing$ 不包含任何一个可能结果，在任何一次试验中都不可能发生，称为**不可能事件**.

**例 1.1.6**　将下面一些事件用集合表示出来：

(1) 在例 1.1.1 的试验中，出现正面；

(2) 在例 1.1.2 中，两个骰子点数之和为 7；

(3) 在例 1.1.4 中，销售额超过 10 万.

**解**　用 $A_1,A_2,A_3$ 分别表示这三个事件，有

$$A_1=\{1\},\quad A_2=\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\},\quad A_3=\{x\mid x>100000\}$$

在上面的例子中，有的事件不可再分，如 $A_1$，这样的事件称为**基本事件**；而 $A_2,A_3$ 由若干基本事件构成，这样的事件称为**复合事件**. 要注意的是，一个事件是否为基本事件是相对于试验目的而言的. 例如，考虑某个地区某个时间段的降雨量，该试验的样本空间 $S$ 可以表示为 $\{x\mid x\geqslant 0\}$，它的基本事件有无穷个. 但是如果试验目的是为了考虑是否发布暴雨预警，此时根据暴雨蓝色、黄色、橙色、红色预警的降雨量要求，基本事件就只有 5 个(包括无需预警的情况).

实际生活中一些较为复杂的事件大多可以分解为基本事件，这对于今后计算事件的概率有很大帮助，但前提是我们必须熟知事件之间的结构方式，这就是事件的关系与运算.(读者大概已经注意到了事件与集合之间的紧密联系，这一点在我们后续的叙述之中，会有更明显的体现.)

设试验 $E$ 的样本空间为 $S$；$A,B,A_i,B_i\ (i=1,2,\cdots)$ 是 $E$ 的事件.

(1) 若事件 $A$ 发生必然导致事件 $B$ 发生，则称事件 $A$ 包含于事件 $B$，或称事件 $B$ 包含事件 $A$，记为 $A\subset B$. 例如，在例 1.1.2 中令 $A$ 表示事件“两个骰子的点数相等”，$B$ 表示事件“两个骰子的点数之和为偶数”，则有 $A\subset B$. 容易验证对任一事件 $A$，都有 $\varnothing\subset A\subset S$.

特别地，若 $A\subset B$ 且 $B\subset A$，则称 $A$ 与 $B$ 相等，记为 $A=B$.

(2)“$A$ 与 $B$ 至少有一个发生”这一事件称为 $A$ 与 $B$ 的**和**(**或并**)，记为 $A\cup B$，注意和事件也可以等价表述为“$A$ 发生或者 $B$ 发生”.

(3)“$A$ 与 $B$ 同时发生”这一事件称为 $A$ 与 $B$ 的**积**(**或交**)，记为 $A\cap B$ 或 $AB$，积事件可以等价表述为“$A$ 发生且 $B$ 发生”.

事件的和与事件的积都可以很容易地推广到有穷多个乃至无穷可列个的情形：

“事件 $A_1, A_2, \cdots, A_n$ 至少有一个发生”这 $n$ 个事件的和可以表示为 $\bigcup\limits_{i=1}^{n} A_i$；

“事件 $A_1, A_2, \cdots, A_n$ 同时发生”这 $n$ 个事件的积可以表示为 $\bigcap\limits_{i=1}^{n} A_i$；

“事件 $A_1, A_2, \cdots$ 至少有一个发生”这无穷可列个事件的和可以表示为 $\bigcup\limits_{i=1}^{\infty} A_i$；

“事件 $A_1, A_2, \cdots$ 同时发生”这无穷可列个事件的积可以表示为 $\bigcap\limits_{i=1}^{+\infty} A_i$.

(4) “事件 $B$ 发生而事件 $A$ 不发生”这一事件称为 $B$ 与 $A$ 的差，记为 $B-A$. 特别地，若 $B=S$，事件 $S-A$ 称为事件 $A$ 的逆事件，记为 $\overline{A}$，此时事件 $A$ 与 $B$ 称为**对立事件或互逆事件**. 注意事件 $B-A$ 与事件 $B \cap \overline{A}$ 是等价的，这一关系在进行事件的运算时时有应用.

(5) 若两个事件 $A$ 与 $B$ 不能同时发生，即 $A \cap B = \varnothing$，则称 $A$ 与 $B$ **互不相容**，或**互斥**.

（读者可能注意到了我们在表述事件的关系时，用到的记号与集合的关系符号完全一致. 这样处理的合理性可以根据事件与集合的关系自行验证. 我们也可以如同描述集合的关系一样用图形描述事件的关系）.

在下面的图 1.1.1 ～ 图 1.1.6 中，矩形表示样本空间 $S$，圆 $A$ 与 $B$ 分别表示事件 $A$ 与 $B$，结合图形，我们可以更直观地理解事件 $A$ 与 $B$ 的关系.

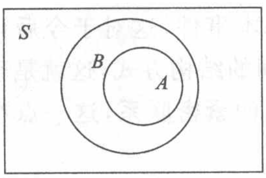


图 1.1.1　$A \subset B$

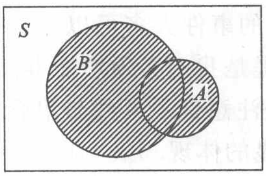


图 1.1.2　$A \cup B$

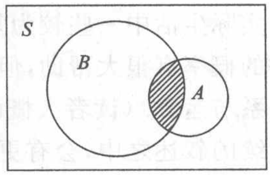


图 1.1.3　$A \cap B$

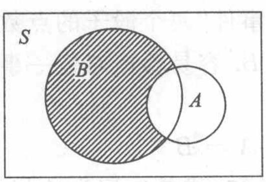


图 1.1.4　$B-A$

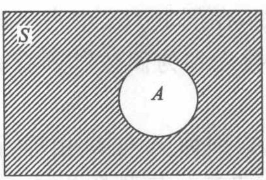


图 1.1.5　$\overline{A}$

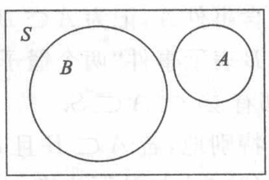


图 1.1.6　$A \cap B = \varnothing$

图 1.1.1、图 1.1.6 分别表示了事件的包含关系与互斥关系的直观状态，图 1.1.2 ～ 图 1.1.5 的阴影部分分别表示两个事件的并交、差、逆.

事件的运算也与集合一致，满足如下关系：

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

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

**德摩根律**　$\overline{A\cup B}=\overline{A}\cap\overline{B},\ \overline{A\cap B}=\overline{A}\cup\overline{B}$

对有穷或无穷可列个事件 $A_i$，德摩根律依然成立，即

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

**例 1.1.7**　设 $A,B,C$ 是三个事件，试用事件的关系与运算表示下面一些事件：

(1) $A$ 与 $B$ 发生但 $C$ 不发生；

(2) $A,B,C$ 都不发生；

(3) $A,B,C$ 至少有一个发生；

(4) $A,B,C$ 恰有一个发生；

(5) $A,B,C$ 至多一个发生；

(6) $A,B,C$ 至少两个发生；

(7) $A,B,C$ 至多两个发生.

**解**　用 $D_i$ $(i=1,2,\cdots,7)$ 表示上述 7 个事件.

(1) $A$ 与 $B$ 发生即它们同时发生，这一事件可表示为 $AB$，但 $C$ 不发生，根据事件的差的概念，$D_1$ 可表示为 $AB-C$ 或 $AB\overline{C}$；

(2) $A,B,C$ 都不发生可以叙述为“$A$ 不发生且 $B$ 不发生且 $C$ 不发生”，它可以看成三个用“且”连接的事件，表明三者是“积”的关系；一个事件不发生即其逆事件发生，所以 $D_2$ 可表示为$\overline{A}\,\overline{B}\,\overline{C}$；

(3) 根据事件的和的概念，直接得到

$$D_3=A\cup B\cup C$$

(4) $A,B,C$ 恰有一个发生意味着只有 $A$ 发生或者只有 $B$ 发生或者只有 $C$ 发生；只有 $A$ 即 $A$ 发生但 $B$ 不发生且 $C$ 不发生，该事件可表示为 $A\overline{B}\,\overline{C}$，类似地，只有 $B$ 发生与只有 $C$ 发生可分别表示为$\overline{A}B\overline{C}$，$\overline{A}\,\overline{B}C$，而这三个事件用“或者”连接，表明它们是“和”的关系，所以

$$D_4=A\overline{B}\,\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\,\overline{B}C$$

(5) $A,B,C$ 至多一个发生可以分解为前面提到的两个事件 $D_2$ 与 $D_4$，不难得到 $D_5=D_2\cup D_4$，但如果将其改写成 $A,B,C$ 的关系则这种表达过于复杂. 如果我们从反面考虑，它们至多一个发生意味着三个事件至少有两个不发生，即$\overline{A}\,\overline{B}$，$\overline{A}\,\overline{C}$，$\overline{B}\,\overline{C}$ 至少有一个发生，所以 $D_5$ 可以更为简洁地表示为$\overline{A}\,\overline{B}\cup\overline{A}\,\overline{C}\cup\overline{B}\,\overline{C}$；

(6) 与前面提到的“三个事件至少有两个不发生”类似，有

$$D_6 = AB \cup AC \cup BC$$

(7) 根据(5) 的经验，我们也从反面考虑这个事件. 至多两个发生意味着至少一个不发生，这样很容易得到 $D_7 = \overline{A} \cup \overline{B} \cup \overline{C}$，由德摩根律它也可以表示为$\overline{ABC}$.

如果对事件 $A,B,C$ 赋予现实背景，类似于上面的求解过程，我们可以将许多实际中的复杂事件用一些“基本”的事件表示出来. 例如，某人有三把形状很接近的钥匙，此人深夜回家，周围一片漆黑，他凭借触摸无法确定哪个是开门钥匙. 他试着用其中一把钥匙开门，直到门被打开为止. 假定他第一次尝试是从三把钥匙中任选一把，而第二次尝试从剩下的两把钥匙中任选一把，他最有可能尝试几次将门打开？要解决这个问题，仅利用本节的知识并不足够，但是我们不妨先将这个问题用本节学到的知识来表达. 后面我们将看到，这其实是解决现实中概率问题必须掌握的基本功.

要解决上面提出的问题，就是要比较“尝试 $i$ 次打开门”($i = 1,2,3$) 这三个事件(分别记为 $D_1,D_2,D_3$) 发生的可能性哪个最大，所以我们先将这三个事件用基本事件来表示. 门能否打开仅与每次尝试能否选中正确的钥匙有关，所以用 $A_1$，$A_2$，$A_3$ 分别表示第一次、第二次、第三次选中正确的钥匙. 类似于例 1.1.7 的分析，不难得到

$$D_1 = A_1, \qquad D_2 = \overline{A}_1 A_2, \qquad D_3 = \overline{A}_1 \overline{A}_2 A_3$$

这个问题其实与现实中的抽奖问题本质一样. 所谓抽奖问题可以简述为：奖券箱中有 $n$ 张奖票，其中只有一张是大奖，先抽后抽中奖的机会是否有差别？读者可以自行尝试将“第 $i$ 次抽中大奖”($i = 1,2,\cdots,n$) 这 $n$ 个事件用基本事件来表示，在后面的 1.3 节，我们将计算这些事件的概率，从而解决这个问题.

## 习　题　1.1

1. 记者从某节车厢的乘客中任选一人进行访问，令 $A,B,C$ 分别表示事件“选中乘客为女性”、“选中乘客年龄不超过 30 岁”、“选中乘客乘车目的为旅游”，试叙述下列事件的意义：

(1) $\overline{A}BC$；　　(2) $AB \cup C$；

(3) $A \cap (B \cup C)$；　　(4) $C - (A \cup B)$.

2. 在本节例 1.1.5 中提到的飞镖运动于 15 世纪兴起于英格兰，如今在英国、法国、美国已是非常普及的大众运动. 镖盘是一个 20 等分的圆盘(图 1.1.7)，外圈的数字表示该扇形区域对应的得分值. 镖盘有几个特殊的得分区：外圈的狭窄圆环是该区域分值的双倍区，内圈狭窄圆环为该区域分值的三倍区，镖盘正中的外中心圆分值为 25 分，内中心圆为 50 分. 用 $A_i$($i = 1,2,\cdots,20$) 表示“飞镖落到 $i$ 分得

分区(非特殊区)",$B_i$ ($i=1,2,\cdots,20$) 表示"飞镖落到 $i$ 分双倍区",$C_i$($i=1,2,\cdots,20$) 表示"飞镖落到 $i$ 分三倍区",$D$ 表示"飞镖落到外中心圆",$F$ 表示"飞镖落到内中心圆". 试用 $A_i,B_i,C_i,D,F$ 表示下列事件:

(1)"在一次投掷中得到 20 分";

(2)"在一次投掷中得分不超过 10 分";

(3)"在一次投掷中得分超过了 50 分";

(4)"在一次投掷中得分为奇数".

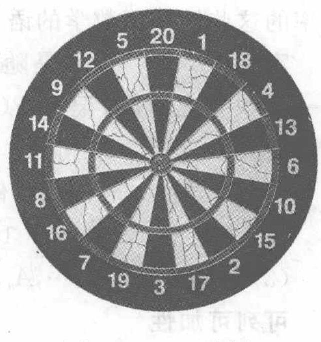


图 1.1.7　飞镖盘

3. 在题 2 中,假设某人投掷飞镖 $n$ 次,$G_i$ ($i=1,2,\cdots,n$) 表示此人第 $i$ 次射中内中心圆,试用 $G_i$ 表示下列事件:

(1) 没有一次投中内中心圆;

(2) 至少投中一次内中心圆;

(3) 恰好有一次投中内中心圆.

# 1.2　随机事件的概率

## 1.2.1　概率的定义与性质

概率也称几率,与概率相关的问题在现实生活中可以找出许多. 例如,前面提到的掷硬币试验,我们都知道正面出现的概率是 $\frac{1}{2}$;篮球比赛中可以根据以往的统计资料推断某个球队的胜率是 70%;某个超市举行了转盘抽奖的活动,根据转盘的布局,我们大致能知道中奖的概率是多少等. 而究竟什么是概率呢?从直观上,我们可以认为概率是随机事件发生可能性大小的一种度量. 这种提法对于我们理解概率的现实意义是合适的,但是对于挖掘概率的本质,形成严谨的理论却显得乏力. 因为根据这样的概念,我们并不能解决"如何度量随机事件发生的可能性大小"的问题.

而概率的本质到底是什么?既然概率是一种"度量",不妨看看我们熟知的度量 —— 长度有什么性质. 度量物体的长度首先需要一个标尺;其次,长度有它的取值范围(非负);另外,长度是可加的,如果尺不够长,我们可以通过量取物体的各个部分得到它的总长度. 根据长度的这些特点,从理论上我们可以度量出任何物体的长度. 那么概率这个度量应该具有什么样的性质呢?概率要度量的对象是"随机事件",度量所需的标尺可以认为是基本事件的概率;根据日常习惯,概率的取值范围应限定在[0,1];复合事件的概率应该能够通过基本事件的概率"相加"得到. 将