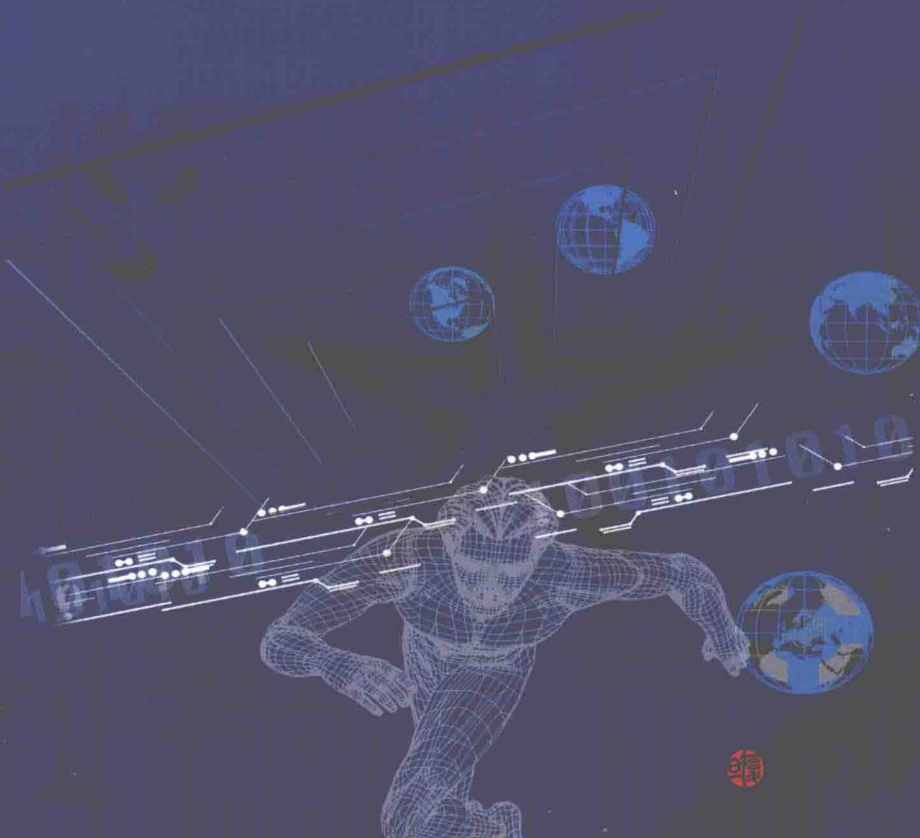
新锐丛书

高等学校经管类数学系列教材

概率论与数理统计

GAILÜLUN YU SHULITONGJI

主　编　李亚琼　黄立宏



新锐丛书

高等学校经管类数学系列教材

概率论与数理统计

李亚琼　黄立宏　主编

復旦大學出版社

图书在版编目(CIP)数据

概率论与数理统计/李亚琼,黄立宏主编. —上海:复旦大学出版社,2007.8
(新锐丛书)
高等学校经管类数学系列教材
ISBN 978-7-309-05683-9

Ⅰ. 概…　Ⅱ. ①李…②黄…　Ⅲ. ①概率论-高等学校-教材②数理统计-高等学校-教材　Ⅳ. 021

中国版本图书馆 CIP 数据核字(2007)第 122632 号

概率论与数理统计
李亚琼　黄立宏　主编

出版发行 复旦大学 **出版社**　上海市国权路 579 号　邮编 200433
86-21-65642857(门市零售)
86-21-65100562(团体订购)　86-21-65109143(外埠邮购)
fupnet@ fudanpress. com　http://www. fudanpress. com

责任编辑 白国信
总 编 辑 高若海
出 品 人 贺圣遂

印　　刷 浙江临安市曙光印务有限公司
开　　本 787×960　1/16
印　　张 26
字　　数 453 千
版　　次 2007 年 8 月第一版第一次印刷

书　　号 ISBN 978-7-309-05683-9/O・403
定　　价 35.00 元

如有印装质量问题,请向复旦大学出版社发行部调换。

内 容 简 介

本书根据高等学校经管类专业概率论与数理统计课程的教学大纲及考研大纲编写而成. 内容包括:概率论的基本概念、随机变量及其分布、随机向量及其分布、随机变量的数字特征、大数定律和中心极限定理、数理统计的基础知识、参数估计、假设检验、计量经济学初步,以及随机过程初步等. 每章均配有大量的例题和习题,其中有些例题有明确的经济背景. 书后附有习题的答案,既便于教学,又利于考试复习.

本书可作为高等学校经管类专业的教材,也可供从事相关专业的科研及教学人员参考.

前 言

本书根据高等学校经管类专业概率论与数理统计课程的教学大纲及考研大纲编写而成．内容包括概率论、数理统计、计量经济学初步和随机过程初步四部分．本书的编写不仅注重对学生基础理论功底的培养，而且也注意更加广泛、密切地与实际经济相结合，注重培养经管类学生应用经济数学方法，特别是随机方法解决实际经济问题的能力．特别值得一提的是，本书的内容为后续课程和知识的拓广奠定了一定的基础，本书将传统教材编写中的回归分析改为计量经济学初步，按照计量经济学的体系进行了编写，介绍了相关软件的使用以及和本门课程相关知识的关系．书中每章都配有大量的例题和习题，其中有些例题有明确的经济背景．对于加“*”号的内容可以不作教学要求，教师和学生可灵活选择．

本书由李亚琼和黄立宏主编，参加讨论和编写的人员有李亚琼、黄立宏、汪端阳、晏木荣、方涛、战学秋、吉小东、焦红伟、闫慧臻．在本书的成书过程中得到了湖南大学数学与计量经济学院的大力支持，同时得到了复旦大学出版社的热心指导和精心编辑，在此表示衷心的感谢．

尽管作者在编书过程中非常认真和努力，但教材中难免会有不妥之处，希望使用本教材的教师和学生提出宝贵意见．

编者

2007 年 3 月

目　　录

第一章　概率论的基本概念

概率论与数理统计是从数量化的角度来研究现实世界中的不确定现象或者随机现象及规律性的一门应用数学学科. 20 世纪以来，它已经广泛应用于工业、国防、国民经济，以及工程技术等各个领域. 它也是将来经济类的专业核心课程——计量经济学的重要基础之一.

当我们多次观察自然现象和社会现象后，会发现许多现象在一定条件下必然会发生. 例如，在没有外力作用下作等速直线运动的物体必然继续作等速直线运动；又如，在 1 个标准大气压下，水加热到 100 ℃时必然会沸腾；等等. 这类现象称为**确定性现象**.

但是，在自然现象和社会现象中也还广泛存在着与确定性现象有着本质区别的另一类现象. 例如，在相同条件下抛同一枚硬币，其结果可能是正面朝上，也可能是反面朝上，并且在每次抛掷之前无法肯定抛掷的结果是什么；用同一门炮向同一目标射击，各次弹着点不尽相同，在一次射击之前无法预测弹着点的确切位置. 对于在一定条件下，可能出现这样的结果，也可能出现那样的结果，而在试验或观察之前不能预知确切的结果，这种现象我们称之为**不确定现象**. 但人们经过长期实践并深入研究之后，发现这类现象在大量重复试验或观察下，它的结果都呈现出某种规律性. 例如，多次重复抛一枚硬币得到正面朝上大致有一半，同一门炮射击同一目标的弹着点按照一定规律分布等. 这种不确定现象在大量重复试验或观察中所呈现出的固有规律性，就是我们今后所说的**统计规律性**.

这种在个别试验中其结果呈现出不确定性，在大量重复试验中其结果又具有统计规律性的现象，我们称之为**随机现象**.

第一节　样本空间　随机事件

一、随机试验

在这里，我们把对自然现象的一次观察或进行的一次科学试验统称为一

个**试验**,下面给出一些试验的例子.

E_1:抛一枚硬币,观察正面 H,反面 T 出现的情况.

E_2:将一枚硬币抛掷 3 次,观察正面 H,反面 T 出现的情况.

E_3:将一枚硬币抛掷 3 次,观察出现正面的次数.

E_4:抛一颗骰子,观察出现的点数.

E_5:记录某电话交换台一分钟内接到的呼叫次数.

E_6:在一批同型号灯泡中任意抽取一只测试它的寿命.

E_7:记录某地一昼夜的最高温度和最低温度.

上面举出了七个试验的例子,它们有着共同的特点.例如,试验 E_1 有两种可能结果,出现 H 或者出现 T,但在抛掷之前不能确定出现 H 还是出现 T,这个试验可以在相同条件下重复地进行.又如,试验 E_6,我们知道灯泡的寿命(以小时计) $t \geqslant 0$,但在测试之前不能确定它的寿命有多长,这一试验也可以在相同的条件下重复地进行.概括起来,这些试验具有以下的特点:

(1) 一个试验可以在相同条件下重复进行(**可重复性**).

(2) 每次试验的可能结果不止一个,但事先能明确试验的所有可能结果(**确定性**).

(3) 进行一次试验之前不能确定哪一个结果会出现(**随机性**).

在概率论中,我们将具有上述三个特点的试验称为**随机试验**,简称试验,用字母 E 表示.

二、事件

随机试验的每个可能的结果称为该试验的**随机事件**,简称为**事件**,一般用大写字母 A, B, C, …来表示.

例如,上述试验 E_1 中,"出现 H"这是试验 E_1 的一个可能结果,故它是 E_1 的一个随机事件.在试验 E_4 中"出现的点数是偶数"及"出现的点数是 6 点"都是试验 E_4 的一个可能结果,故都是 E_4 的一个随机事件.

对于一个试验来说,在每次试验中必然要发生(出现)的结果称为此试验的**必然事件**,记为 Ω;在每次试验中必然不发生(出现)的结果称为此试验的**不可能事件**,记为 $\varnothing$.例如,在试验 E_4 中,"出现的点数小于 7"是一个必然事件,"出现的点数为 2.5"则是一个不可能事件.

三、样本空间

对试验 E,我们把其最简单的不能再分的事件称为该试验的**基本事件**;由

若干个基本事件组合而成的事件称为**复合事件**.

例如,在试验 E_4 中,"出现的点数为 i"($i=1, 2, \cdots, 6$)均为 E_4 的基本事件;而"出现的点数为偶数"为 E_4 的复合事件.

由所有基本事件组成的集合称为该试验的**样本空间**,记为 Ω;样本空间 Ω 中的元素称为**样本点**,用 ω 表示,因而基本事件又称作样本点.

下面给出第一章第一节中试验 E_k ($k=1, 2, 3, \cdots, 7$) 的样本空间 Ω_k:

Ω_1:$\{H, T\}$;

Ω_2:$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$;

Ω_3:$\{0, 1, 2, 3\}$;

Ω_4:$\{1, 2, 3, 4, 5, 6\}$;

Ω_5:$\{0, 1, 2, 3, \cdots\}$,这个样本空间含有无穷多个样本点,但这些样本点可以依照某种次序排列出来,以后我们将称它的点数为可列个;

Ω_6:$\{t \mid t \geqslant 0\}$;

Ω_7:$\{(x, y) \mid T_0 \leqslant x \leqslant y \leqslant T_1\}$,这里 x 表示最低温度,y 表示最高温度,并设这一地区的温度不会小于 T_0 也不会大于 T_1.

Ω_6 与 Ω_7 这两个样本空间都包含有无穷多个样本点,它们充满一个区间,不是一个可列集.

注意 样本空间的元素是由试验的目的所确定的. 例如,在 E_2 和 E_3 中,同时将一枚硬币连抛 3 次,由于试验的目的不一样,其样本空间也不一样.

我们也可将随机事件定义为样本点的某个集合,称某事件发生当且仅当它所包含的某一个样本点出现.

我们把样本空间 Ω 也作为一个事件,因为在每次试验中必然出现 Ω 中的某个样本点,也即 Ω 必然发生,所以常称样本空间 Ω 为必然事件. 必然事件 Ω 与不可能事件 $\varnothing$ 可以说不是随机事件,但为了今后研究的方便,我们还是把必然事件与不可能事件作为随机事件的两个极端情形来处理.

在今后讨论中,经常把样本空间认为是预先给定的,这是必然的抽象,这种抽象使我们能更好地把握随机现象的本质,而且得到的结果能广泛地应用. 事实上,一个样本空间可以概括各种实际内容大不相同的问题. 例如,只包含两个样本点的样本空间既能作为掷硬币出现正、反面的模型,也能用于产品检验中出现"正品"及"废品",又能用于气象中"下雨"与"不下雨",以及公用事业排队现象中"有人排队"与"无人排队",等等. 尽管问题的实际内容如此不同,但有时却能归结为相同的概率模型.

只含有限个或可列无穷个样本点的样本空间,称为**离散样本空间**. 对于离散样本空间,它的任何子集都是事件.

四、事件间的关系与运算

在一个样本空间中，显然可以定义不止一个事件. 在实际问题中，往往要求我们同时考察几个在同样条件下的事件及它们之间的联系. 详细地分析事件之间的关系，将帮助我们更深刻地认识事件的本质.

下面讨论事件间的关系及事件的运算. 事实上，由于引进了样本空间，并建立了事件和集合间的联系，于是事件的关系和运算完全可以运用集合间的关系和运算来处理.

设试验 E 的样本空间为 Ω，而 A, B, $A_k(k=1, 2, 3, \cdots)$ 是 Ω 的子集.

(1) 若 $A\subset B$，则称事件 B 包含事件 A. 这指的是事件 A 发生必导致事件 B 发生，如图 1-1 所示.

若 $A\subset B$，且 $B\subset A$，记为 $A=B$，称事件 A 与事件 B 相等，或称事件 A 与 B 等价.

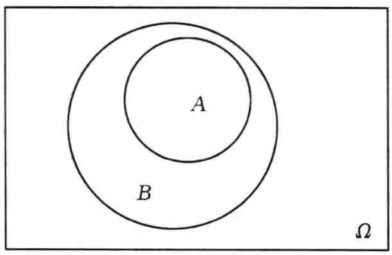


图 1-1 $A\subset B$

(2) 事件 $A\cup B=\{x\mid x\in A 或 x\in B\}$，称为事件 A 与事件 B 的**和事件**. 当且仅当 A, B 中至少有一个发生时，事件 $A\cup B$ 发生，如图 1-2 所示.

类似地，称 $\bigcup\limits_{k=1}^{n}A_k$ 为 n 个事件 A_1, A_2, $\cdots$, A_n 的和事件；称 $\bigcup\limits_{k=1}^{\infty}A_k$ 为可列个事件 A_1, A_2, $\cdots$的和事件.

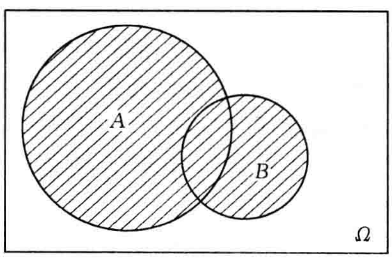


图 1-2 $A\cup B$

(3) 事件 $A \cap B = \{x \mid x \in A 且 x \in B\}$，称为事件 A 与事件 B 的**积事件**. 当且仅当 A，B 同时发生时，事件 $A \cap B$ 发生，$A \cap B$ 也记作 AB，如图 1-3 所示.

类似地，称 $\bigcap\limits_{k=1}^{n} A_k$ 为 n 个事件 A_1，A_2，…，A_n 的积事件；称 $\bigcap\limits_{k=1}^{\infty} A_k$ 为可列个事件 A_1，A_2，…的积事件.

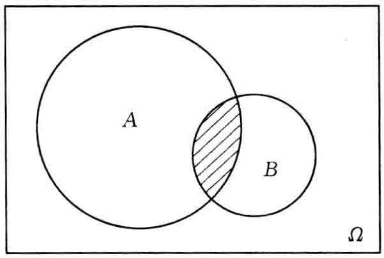
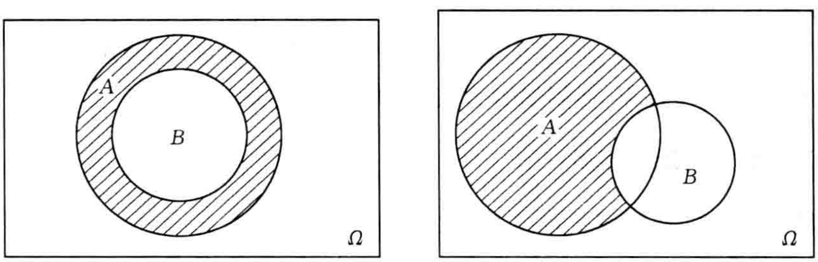


图 1-3　$A \cap B$

(4) 事件 $A - B = \{x \mid x \in A 且 x \notin B\}$，称为事件 A 与事件 B 的**差事件**. 当且仅当 A 发生而 B 不发生时，事件 $A-B$ 发生，如图 1-4 所示.



(a)　(b)

图 1-4　$A-B$

(5) 若 $A \cap B = \varnothing$，则称事件 A 与 B 是互不相容的，或互斥的. 这指的是事件 A 与事件 B 不能同时发生，如图 1-5 所示.

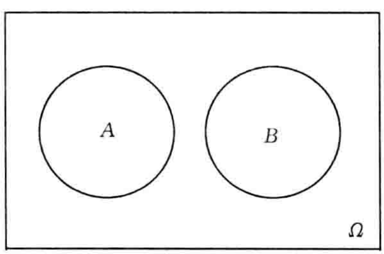


图 1-5　$A \cap B = \varnothing$

基本事件是两两互不相容的.

(6) 若 $A \cup B = \Omega$ 且 $A \cap B = \varnothing$，则称事件 A 与事件 B **互为逆事件**，又称事件 A 与事件 B **互为对立事件**. 即对每次试验而言，事件 A，B 中必有一个发生，且仅有一个发生. A 的对立事件记为 $\overline{A}$，$\overline{A} = \Omega - A$，如图 1-6 所示.

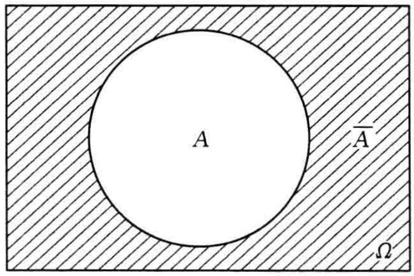


图 1-6　$\overline{A}$

我们已经用图 1-1～图 1-6 直观地表示以上事件之间的关系与运算. 例如，在图 1-1 中矩形表示样本空间 Ω，圆 A 与圆 B 分别表示事件 A 与事件 B，事件 B 包含事件 A. 又如，在图 1-2 中矩形表示样本空间 Ω，圆 A 与圆 B 分别表示事件 A 与事件 B，而阴影部分表示和事件 $A \cup B$.

五、事件的运算性质

随机事件的运算满足以下基本性质.

1. 交换律

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

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

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

4. 德·摩根律

$$\overline{A \cup B} = \bar{A} \cap \bar{B},$$
$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

分配律和德·摩根(De Morgan)律均可推广到有限个或可列无穷多个事件的情形,如

$$(\bigcup_i A_i) \cap B = \bigcup_i A_i B;$$

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

例 1　若 A, B, C 是三个事件,则

(1) A 发生而 B 与 C 都不发生,可以表示为

$$A\bar{B}\bar{C} \text{ 或 } A-B-C \text{ 或 } A-(B\cup C);$$

(2) A 与 B 都发生而 C 不发生,可以表示为

$$AB\bar{C} \text{ 或 } AB-C \text{ 或 } AB-ABC;$$

(3) 所有三个事件都发生,可以表示为

$$ABC;$$

(4) 这三个事件恰好发生一个,可以表示为

$$A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C;$$

(5) 这三个事件恰好发生两个,可以表示为

$$AB\bar{C} \cup A\bar{B}C \cup \bar{A}BC;$$

(6) 这三个事件至少发生一个,可以表示为

$$A\cup B\cup C$$

或 $$A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C \cup AB\bar{C} \cup A\bar{B}C \cup \bar{A}BC \cup ABC.$$

例 2　甲、乙、丙三人各进行一次试验,A_1, A_2, A_3 分别表示甲、乙、丙成功,说明下列各式表示的事件:

(1) $\bar{A}_1$;

(2) $A_1 \cup A_2$;

(3) $\overline{A_2 A_3}$;

(4) $A_1 A_2 A_3$.

解　(1) 甲未成功.

(2) 甲与乙至少有一人成功.

(3) 乙与丙至少有一人未成功.

(4) 甲、乙、丙均成功.

第二节　随机事件的概率

对于一个事件(除必然事件和不可能事件外),它在一次试验中可能发生,也可能不发生,但我们常常希望知道某些事件在一次试验中发生的可能性究竟有多大.例如,为了确定水坝的高度,就要知道河流在造水坝地段每年最大洪水达到某一高度这一事件发生的可能性大小.为此,首先引入频率,它描述了事件发生的频繁程度,进而引出概率的统计定义和公理化定义.

一、频率与概率的统计定义

定义 1　在相同条件下,进行了 n 次试验,在这 n 次试验中,事件 A 发生的次数 n_A 称为事件 A 发生的**频数**,比值 $\frac{n_A}{n}$ 称为事件 A 发生的**频率**,并记为 $f_n(A)$,即

$$f_n(A) = \frac{n_A}{n}.$$

不难验证,频率具有如下性质:

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

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

(3) 若事件 $A_1, A_2, \cdots, A_m$ 互不相容,则有

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

例 1　考虑"抛硬币"这个试验,历史上有数学家做过这样的试验,得到的结果如表 1-1 所示.

表 1-1

实验者	n	n_A	$f_n(A)$
德·摩根	2 048	1 061	0.518 1
蒲　丰	4 040	2 048	0.506 9
费　勒	10 000	4 979	0.497 9
皮尔逊	24 000	12 012	0.500 5

从上述数据可以看出:(1)频率具有随机**波动性**.即对于同样的 n,所得的

$f_n(A)$不尽相同.(2)抛硬币次数 n 较小时,频率 $f_n(A)$随机波动的幅度较大,但随着 n 增大,频率 $f_n(A)$呈现出**稳定性**.即当 n 逐渐增大时,$f_n(A)$总是在0.5附近摆动,而逐渐稳定于0.5.

例 2　考察英语中特定字母出现的频率.当观察字母的个数 n(试验次数)较小时,频率有较大幅度的随机波动.但当 n 增大时,频率呈现出稳定性.表 1-2 所示就是一份英文字母频率的统计表(由 G. Dewey 统计了约 438 023 个字母得到).

表 1-2

字母	频率	字母	频率	字母	频率
E	0.126 8	L	0.039 4	P	0.018 6
T	0.097 8	D	0.038 9	B	0.015 6
A	0.078 8	U	0.028 0	V	0.010 2
O	0.077 6	C	0.026 8	K	0.006 0
I	0.070 7	F	0.025 6	X	0.001 6
N	0.070 6	M	0.024 4	J	0.001 0
S	0.063 4	W	0.021 4	Q	0.000 9
R	0.059 4	Y	0.020 2	Z	0.000 6
H	0.057 3	G	0.018 7		

例 3　著名的数学家拉普拉斯(Laplace)观察了伦敦、彼得堡、柏林、法兰西在 10 年间新生儿的性别,发现男婴出生的频率总摆动于$\frac{22}{43}$的数值左右.

从上述例子我们认识到,频率的稳定性是通过大量的试验所得到的随机事件的规律性,因此这种规律性称为**统计规律性**.由此我们给出概率的统计定义.

定义 2　在相同的条件下,重复做 n 次试验,n_A 是 n 次试验中事件 A 发生的次数,当试验次数 n 很大时,如果频率 $f_n(A)$稳定地在某一数值 p 的附近摆动,且随着试验次数的增多,摆动的幅度越来越小,则称数值 p 为事件 A 在这个条件下发生的**概率**,记作

$$P(A)=p.$$

由以上定义给出的事件 A 的概率 p 就称为**统计概率**.

由于统计概率是由频率来定义的,所以它也具有类似于频率的三个基本性质: