古典模型与概率空间

第1章

1.1 概率的古典模型与对概率认识的经验概括

1.1.1 引言

概率统计(probability and statistics)是介绍概率与统计基本概念的入门教材,在我国也称为随机数学,其内容在于阐明不确定的**随机现象**(Random phenomenon)的规律性.这种不确定的随机现象的特点是:在一组固定的条件下,某个事件既可能发生,又可能不发生,而其概率就是描述这种随机事件发生的可能性大小的百分比.建立在概率概念构架上的统计学,是从观测数据出发,对于随机现象进行分析、判断,从而帮助人们在面对随机现象时,给出较为合理的决策.

概率现象的早期启蒙性研究,常出自博弈与机会游戏.早在17世纪,人们已经用经验的等可能性分析,来计算博弈与机会游戏中一些随机事件出现概率的大小.18世纪以后,人们注意到在天文观测、人口统计、误差理论、产品检查等科学问题和社会问题中的许多现象,与以上的机会游戏有相似之处,于是,由机会游戏起源的概率论,就逐渐发展成为数学的一个重要分支.

瑞士数学家 Bernoulli 一家,建立了概率论中第一个大数定律,阐述了随机事件发生的频率稳定于其发生的概率这一个重要的统计事实.由此奠定了从统计的角度认识概率的基础,并将频率这种感性认识与等可能性分析的结果联系了起来.De Moivre 和 Laplace 用数学推导,得到中心极限定理的初始形

式，揭示了由大量独立的随机因素积累的结果，所呈现的瘦身、轻尾的钟形分布，即所谓正态分布，在概率论中处于核心位置. 此后，从 19 世纪末到 20 世纪初，Chebyshev，Markov，Lyapunov，Lévy，Khinchine 等用数学分析的方法，建立了大数定律及中心极限定理的一般形式，由此解释了在实际中遇到的许多随机变量都近似服从正态分布这个经验事实. 大数定律和中心极限定理是早期概率论研究的两个主要推动力.

到了 20 世纪 30 年代，前苏联数学家 Kolmogorov 用集合论与测度论的思想，归纳总结了随机事件及其概率所蕴含的最基本的规律和性质，创建了概率论的公理体系，使概率论完全地成为一个严格的数学学科. 另一方面，在 20 世纪初，英国统计学家 Fisher 在农业与生物学的应用中，开创并奠定了统计的理念和基础. 此后，统计的思想与方法的研究蓬勃发展，并逐渐成为随机建模的基本工具.

1.1.2　概率的古典模型——等可能性分析

概率模型的核心思想出自于对只有有限个等可能结果的随机试验的经验研究与理论扩展. 只有有限个等可能结果的随机环境，是人们最早注意到的随机对象. 在生产和生活实践中，人们不可避免地需要对某些具有随机性的事件发生的可能性的大小进行估量，即对其概率性能进行了解.

随机试验的一个可能结果，称为一个**基本事件**（simple event）（或称为**样本点**（sample point）），通常用小写的希腊字母 ω 泛记一个基本事件.

全体基本事件的集合，通常用大写的希腊字母 Ω 表示，称为**样本空间**（sample space）.

在样本空间 Ω 中，如果只包含有限个基本事件，且基本事件间是平等地（即对称地）出现的情形，称为**古典模型**（classical model）. 古典模型是人们认识概率现象，并发展成概率学科的基础.

古典模型情形所感兴趣的对象，常常是样本空间 Ω 的某一个部分，即需要知道这一部分中的基本事件出现可能性的大小. Ω 的一个部分，即 Ω 的某个子集，称为**随机事件**（random event），通常用一个大写的英文字母泛记. 在概率统计中还常常简称为**事件**（event）.

1. 古典模型中随机事件的概率分析——等可能性分析

一个随机试验如果具备了以下两个基本特征，就称为古典模型：

(1) 随机试验的全部可能结果（全部基本事件）一共只有有限个，即 Ω 是一个有限集；

(2) 各个基本事件的出现是彼此对称的，即平等的.

掷均匀的骰子、投均匀的硬币等简单的随机现象，是古典模型的最简单的特例.因此，它们常用作教学中的范例.

例 1.1 掷一颗各个面均匀的骰子，一共可以有 6 种不同结果，即掷到的点数为 1,2,3,4,5, 6.因此，对于“掷一个各个面均匀的骰子”这个随机试验而言，共有 6 个基本事件，从而样本空间是 $\Omega=\{1,2,3,4,5,6\}$.而随机事件“掷到奇数点”，就是掷到 1,3 或 5，它就是 Ω 的子集合$\{1,3,5\}$.假如在掷骰子的游戏中，甲乙两人约定：掷到 1 或 2 则甲赢，而掷到 5 或 6 则乙赢，而其余情况须重掷，那么

$$“甲赢”=\{1,2\},\quad “乙赢”=\{5,6\},\quad “须重掷”=\{3,4\}.$$

如果试验掷到的结果是 5，那么{掷到素数点}、{掷到奇数点}及{乙赢}这三个随机事件都发生了，而“甲赢”、“须重掷”这两个随机事件都没有发生.

从此例中，我们可以归纳出一个一般的认识：若在一次随机试验(random experiment)中得到的结果为 ω，那么，如果某个事件 A 包含 ω，我们就认为在该次随机试验中随机事件 A 发生了，否则就说随机事件 A 没有发生.

需要提醒的是，Ω 也是它自己的一个子集，因此它也是一个随机事件.又因为无论试验结果 ω 是什么，总有 $\omega\in\Omega$，所以，我们称 Ω 为**必然事件**.而其完全相反的情形是空集，即不含任何基本事件的集合，我们将它记为$\varnothing$，我们也将它作为 Ω 的子集，又因为无论试验结果 ω 是什么，总有 $\omega\notin\varnothing$，所以我们称$\varnothing$为**不可能事件**(null event).

现在，我们来分析例 1.1 中的随机试验所涉及的各个随机事件的概率.在这个试验中的骰子是均匀的，因此 6 个基本事件是彼此对称的，所以它们的概率彼此相同，故而每个基本事件发生的比例都是 1/6，也就是概率为 1/6.而“掷到素数点”$=\{2,3,5\}$，在 6 份中占了 3 份，故其概率应该是 $3/6=1/2$.而“乙赢”的概率应为 $2/6=1/3$.

注意，同一个随机试验也可以用不同的样本空间来考察.例如，在例 1.1 中，如果我们只感兴趣于出现点数的奇偶，那么，可以只考虑以下形式的两种结果，并将它们分别作为基本事件：

$$\omega_1=出现奇数点,\quad \omega_2=出现偶数点.$$

此时的样本空间应考虑为 $\Omega'=\{\omega_1,\omega_2\}$.再如，我们考虑的是点数分为大(出现 5 或 6 点，记为 B)，中(出现 3 或 4 点，记为 M)，小(出现 1 或 2 点，记为 S)，那么样本空间应考虑为 $\Omega''=\{B,M,S\}$.无论是样本空间 Ω'，还是样本空间 Ω''，它们分别都构成古典模型.但是，在掷一颗各面均匀的骰子这个随机试验中，样本空间 Ω，样本空间 Ω'，样本空间 Ω''能起的作用是不同的.直观看来，Ω'最粗放，而 Ω 最精细.例如，随机事件“乙赢”可以在样本空间 Ω 或样本空间

Ω''中考虑，却无法在样本空间 Ω' 中考虑，因为它在样本空间 Ω' 中并不是随机事件. 我们说样本空间 Ω 最精细，是相对而言，它能容纳的随机事件最多，包含的信息最丰富. 样本空间 Ω 可以看成是样本空间 Ω' 或样本空间 Ω'' 的细分. 然而对于一个具体的随机试验而言，在研究与其有关的某个具体概率问题时，并非考虑越细的样本空间越好，有时考虑适当粗细的样本空间，可以节省计算量.

2. 人们对可能性的认识的经验概括——古典模型的概率的经验计算公式

若 $A\subset\Omega$(即 A 是 Ω 的一个子集，也就是 A 是一个随机事件)，通常用 $\# A$ 记事件 A 中所含基本事件的数目，并用 $P(A)$ 表示在随机试验中，事件 A 发生的概率.

在古典模型中，样本空间 Ω 中只有有限个基本事件，且各个基本事件的出现是彼此平等的，于是我们有

$$P(A)=\frac{\# A}{\# \Omega}.$$

例 1.2　先后掷两颗均匀的骰子. 在这种情形“做一次试验”，就是“先后掷两颗均匀的骰子”. 如果将掷一颗均匀的骰子视为一个简单的试验，那么“先后掷两颗均匀的骰子”，实际上就是连续地做两个简单的试验，称为复合试验. 在实际问题中，所遇见的试验往往都是复合试验. 在“先后掷两颗均匀的骰子”这一试验(指复合试验)中，每一个可能出现的结果都可以用一个有序对(ω_1,ω_2)来表示，其中 ω_1,ω_2 都代表 1 至 6 之间的一个数，分别表示第一个骰子出现的点数与第二个骰子出现的点数. 例如，(1,2)表示第一个骰子出现的点数为 1，而第二个骰子出现的点数为 2 这一结果. 注意，此时做一次完整的随机试验就是先后掷两颗均匀的骰子，所以一共有 $6\times6=36$ 个可能的结果，即有 36 个基本事件，而且由于骰子是均匀的，故此 36 个可能结果是彼此平等的，从而属于古典模型.

此随机试验的样本空间为

$$\begin{aligned}\Omega&=\{(\omega_1,\omega_2):\omega_1,\omega_2=1,2,3,4,5,6\}\\&=\{(1,1),(1,2),\cdots,(1,6),(2,1),\cdots,(6,1),\cdots,(6,5),(6,6)\}.\end{aligned}$$

这里展示了两种不同的表达方法. 前一种为注释性表达式，它比后一种具体表达式更为抽象和严格. 对于元素很多的集合，前一种表达方式可使问题表达得简洁而准确.

这时，事件 $A=\{$两颗骰子的点数成对$\}=\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$. 而事件 $B=\{$两颗骰子出现的点数之和为 8$\}=\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$.

此外,{两颗骰子出现的点数之和不超过 12}是必然事件,而{掷两颗骰子出现的点数之和超过 12}则是不可能事件. 于是我们有

$$P(A)=\frac{\# A}{\# \Omega}=\frac{6}{36}=\frac{1}{6}, \quad P(B)=\frac{\# B}{\# \Omega}=\frac{5}{36}.$$

在计算有关的概率时,首先要判断是否属于古典模型,因此重要的是弄清以下情形:

(1) 怎样才是做完了一次完整的随机试验;

(2) 此随机试验的所有可能的结果是否总数有限? 全部基本事件是什么,是否彼此对称(即是否等可能)?

例 1.3 将一枚均匀的硬币(正面记为 H,方面记为 T)连掷 3 次. 求以下各事件的概率:

$$A=\{\text{恰有一次出现正面 } H\}, \quad B=\{\text{至少有一次出现正面 } H\}.$$

解 此时,做一次试验是"将一个均匀硬币连掷 3 次",这是一个复合试验. 此试验的样本空间可考虑为

$$\Omega=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}, \quad \# \Omega=8.$$

而

$$A=\{HTT, THT, TTH\}, \quad \# A=3, \quad P(A)=\frac{3}{8}.$$

$$B=\{HHH, HHT, HTH, HTT, THH, THT, TTH\},$$

$$\# B=7, \quad P(B)=\frac{7}{8}.$$

再则,事件 B 与事件"全出现反面 T"合起来正好反映了全部可能,而后者只含一个基本事件"TTT",即其概率为 1/8,故而显然有 $P(B)=1-1/8=7/8$. 此种考虑常常可能简化计算.

例 1.4 有一批产品共有 N 件,其中有 M 件次品. 假定从中任取 n 件,计算其中恰有 m 件次品的概率.

解 这里做一次试验就是"任取 n 件",于是样本空间中的基本事件的总数就是从 N 个产品中任取 n 件的组合数,即 $\# \Omega=\mathrm{C}_N^n$. 各个产品之间是平等的,从而对于"任取 n 件"的不同取法之间是等可能的. 因此这个随机试验属于古典模型.

令随机事件

$$A_m=\{\text{有 } m \text{ 件次品}\},$$

其中的任意一个基本事件,就是在 M 件次品中任取 m 件(一共有 C_M^m 种不同

取法)，并在 $N-M$ 件合格品中任取其余的 $n-m$ 件(一共有 C_{N-M}^{n-m} 种取法)的一种取法. 所以 A_m 中的基本事件的个数为它们的乘积，即 $\# A_m = C_M^m C_{N-M}^{n-m}$. 由古典模型计算概率的公式，我们得到

$$P(A_m) = \frac{\# A_m}{\# \Omega} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

注　对于例 1.4 中的随机试验，也可以用一个排列作为一个基本事件，这时所涉及的仍然是古典模型，而不同之处在于，这时的样本空间(记为 Ω_1)有

$$\# \Omega_1 = P_N^n = n! C_N^n,$$

而

$$\# A_m = C_n^m P_M^m P_{N-M}^{n-m},$$

其中的 C_n^m 反映了次品可以处在取出的 n 个产品的排列中的任意位置. 于是

$$P(A_m) = \frac{\# A_m}{\# \Omega_1} = \frac{C_n^m P_M^m P_{N-M}^{n-m}}{P_N^n} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

可见对此概率，采用更为精细的样本空间并未带来更多的信息.

例 1.4 涉及两个品种. 自然地可推广为有 3 个品种的情形，如下例.

例 1.5　袋中分别有红、黄、蓝色的球，各有 N_1, N_2, N_3 个. 从中任取 n 个，假定 $n=n_1+n_2+n_3$. 那么，其中红、黄、蓝色球的个数分别恰为 n_1, n_2, n_3 的概率为

$$\frac{C_{N_1}^{n_1} C_{N_2}^{n_2} C_{N_3}^{n_3}}{C_{N_1+N_2+N_3}^{n_1+n_2+n_3}}.$$

读者可以进一步将此推广到有 k 个品种的情形.

以上我们看到在古典模型中，随机事件概率的计算，常常涉及某个集合中元素的个数的计数. 而计数的方法除了用到中学课程中的排列、组合等技巧以外，在离散数学中还有差分方程方法、母函数方法等，但是这些方法只是在计算某些具体随机事件的概率时有用. 而它们在概率统计中，相对于理解概率论的基本思维与方法而言，只是枝节问题. 在学习概率论时，我们不应过多地关注计数的技巧. 另外，用计数计算概率，也不是概率计算的惟一途径. 相对地随机事件间的关系常常能提供很有用的信息，可以用来计算概率. 下面来分析随机事件间的关系和随机事件的运算，以便由已知概率的随机事件，演绎出更为复杂的随机事件的概率.

1.1.3　事件的运算及古典模型中概率的加法法则

因为一个随机事件是 Ω 的一个子集. 我们可以通过集合的运算与包含等关系，用一些随机事件来表达更为复杂的随机事件及其关系，以化简随机事

件的概率的计算.

假设 A,B 为随机试验中的两个随机事件，它们都是 Ω 的子集.

集 A 与集 B 的**并集**(union)$A\cup B$，表示由 A 中或 B 中的所有基本事件合在一起所组成的随机事件，称为随机事件 A 与 B 的**和事件**. 从而“在某次试验中事件 $A\cup B$ 发生”，就是在该次试验中，随机事件 A 与 B 至少有一个发生. 在例 1.2 中有

$$A\cup B=\{(1,1),(2,2),(3,3),(4,4),(5,5),\\ (6,6),(2,6),(3,5),(5,3),(6,2)\}.$$

类似地，**交集**(joint)$A\cap B$ 表示既属于 A，又属于 B 的所有基本事件组成的随机事件，称为随机事件 A 与 B 的**交事件**. 在例 1.2 中，$A\cap B=\{(4,4)\}$. $A\cap B$ 也常常简记为 AB.

A 的**余集**(complement)$\Omega\backslash A$，用 A^c 表示，它就是样本空间 Ω 中那些不属于 A 的所有基本事件的全体所组成的随机事件，也称为 A 的**对立事件**(complement event).

于是，A^c 发生当且仅当 A 不发生. 在例 1.2 中有

$$A^c=\{(\omega_1,\omega_2):(\omega_1,\omega_2)\notin A\}=\{(\omega_1,\omega_2):\omega_1\neq\omega_2\},$$

即随机事件{两颗骰子的点数不相等}.

我们用记号 $A-B$ 表示 $A\cap B^c$，它即是**差事件**

$$A-B=A\cap B^c=A-AB.$$

下面给出随机事件之间相互关系的一些概念.

若 $A\cap B=\varnothing$，即随机事件 A 与随机事件 B 中没有共同的基本事件. 这表示在一次试验中 A 与 B 不可能都发生，所以这时称 A 与 B **互不相容**(mutually exclusive，或**互斥**).

我们又称随机事件 $A_1,A_2,\cdots,A_n$ 互不相容，如果它们两两互不相容.

若 $A\subset B$，即随机事件 A 中的基本事件都包含在随机事件 B 中，于是，随机事件 A 发生必然蕴含着(implies)随机事件 B 的发生，此时称为随机事件 A 包含于(contained)随机事件 B. A 包含于 B 也可说成为 B 包含了(contains)A.

如果 A 包含于 B，同时 B 包含于 A，那么 A 与 B 必是同一个随机事件，记为 $A=B$.

随机事件关系可以表示为一个示意图，称为随机事件的 **Venn 图**(Venn diagram). 利用 Venn 图可以对事件的运算，事件的包含关系有一个直观的理解，从而可以从直观的角度，推导或简化一些随机事件. Venn 图表示可见图 1.1.

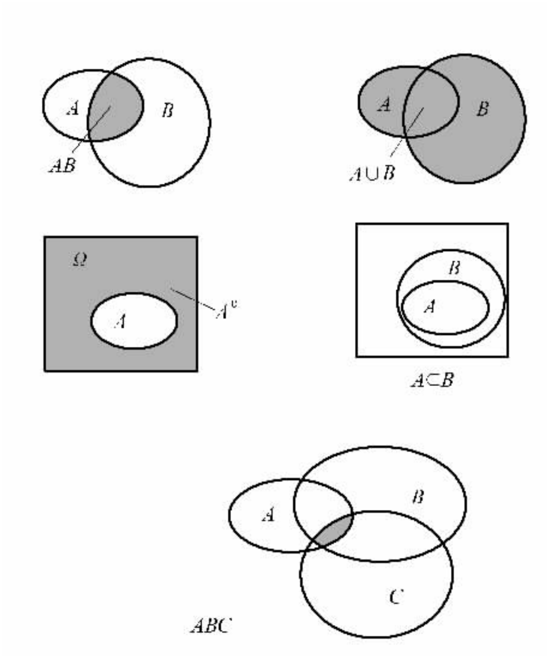

图 1.1　集合的运算和关系的 Venn 图

可见，随机事件间的运算与关系，就是集合的运算与关系，从而集合运算的全部规则对事件运算完全适用，主要有以下几条：

(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

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

(4) 对偶律(也称为 De Morgan 律)

$$(A\cup B)^c=A^c\cap B^c,\quad (A\cap B)^c=A^c\cup B^c.$$

用集合的观点来处理随机事件，对于理解随机事件间的关系至关重要. 这是走向概率的公理化的关键一步.

在古典模型中，如果随机事件 A 与 B 互不相容，则它们的和事件 $A\cup B$ 中所含的基本事件数，正是它们各自所含的基本事件数的和，从而它们各自的概率之和应该等于它们和的概率. 再利用归纳法可知，这个规则对多个事

件也同样适用,即有下面的法则.

概率的加法法则　如果事件 $A_1,A_2,\cdots,A_n$互不相容,则

$$P(A_1\cup A_2\cup\cdots\cup A_n)=P(A_1)+P(A_2)+\cdots+P(A_n).$$

1.1.4　古典模型的条件概率

例 1.6　研究所有恰有两个小孩的家庭,其孩子的性别的分配共有四种情况,于是我们的样本空间为

$$\Omega=\{(男,男),(男,女),(女,男),(女,女)\}.$$

假定男女的出生是等可能的,那么这是古典模型. 以 A 表示事件“随机地选择一个家庭有两个男孩”,以 B 表示事件“随机地选择一个家庭至少有一个男孩”. 显然有

$$P(A)=\frac{1}{4},\quad P(B)=\frac{3}{4}.$$

假如我们只在至少有一个男孩的家庭中选取,那么,样本空间就改变了,变为新的样本空间

$$\Omega_B=\{(男,男),(男,女),(女,男)\}.$$

任取一个这样的家庭,那么其中恰有两个男孩的概率应为 1/3. 如果我们还坚持在原来的样本空间 Ω 中考虑,则所涉及的就是“在已知所取的一个家庭中至少有一个男孩的条件下,此家庭的两个孩子都是男孩”的条件概率,称为 B 发生条件下,事件 A 发生的**条件概率**(conditional probability),并记为 $P(A|B)$.

注意条件概率 $P(A|B)=1/3$,它与无条件概率 $P(A)=1/4$ 是不同的. 这是因为随机事件 B 的发生改变了样本空间. 前者由于知道了“B 发生”这一个附加的信息,这时原来的基本事件(女,女)就被排除在新的样本空间之外,故而条件概率与概率当然未必相同. 可见,条件概率 $P(A|B)$的实质就是缩减了样本空间,把原有的样本空间 Ω 缩减为 Ω_B. 这样,在 Ω 中计算随机事件 A 的概率就是 $P(A)$,而在 Ω_B 中计算的就是条件概率 $P(A|B)$.

为了便于与通常的概率作比较,有时也用记号 $P_B(A)$代替记号 $P(A|B)$.

条件概率的概念归结于样本空间的缩减. 然而,在计算条件概率时,并不必要事事都用样本空间缩减的方法,在某些场合下,这种方法甚至很难实行. 所以,我们还要给出在原概率空间 Ω 上计算条件概率的公式.

在原来的样本空间 Ω 中,随机事件 $A,B,A\cap B$ 中的基本事件的个数分别为 $\#A,\#B,\#(A\cap B)$. 新的样本空间 Ω_B 的基本事件总数 $\#\Omega_B=\#B$,而在新的样本空间 Ω_B 中,事件“恰有两个男孩”所含的基本事件数

$$\#\{恰有两个男孩\}=\#(A\cap B).$$

于是按定义就有

$$P(A \mid B)=\frac{\#(A\cap B)}{\# B}=\frac{\dfrac{\#(A\cap B)}{\#\Omega}}{\dfrac{\# B}{\#\Omega}}=\frac{P(A\cap B)}{P(B)}.$$

这样我们就得到从普通概率计算条件概率的如下公式：

$$P(A \mid B)=\frac{P(A\cap B)}{P(B)}.$$

例 1.7　袋中装有 n 个白球和 m 个黑球，随机地一次取出 $k(k\leqslant n,m)$ 个球，发现都是同一种颜色的. 求它们是黑色的概率.

解　此时做一次随机试验就是无放回地随机取 k 个球，共有 C_{n+m}^{k} 个等可能的基本事件，即 $\#\Omega=\mathrm{C}_{n+m}^{k}$. 记事件

$$A=\{\text{取出的球为同色}\},\quad B=\{\text{取出的球全为黑色}\},$$

则

$$\# A=\mathrm{C}_{n}^{k}+\mathrm{C}_{m}^{k},\quad \#(A\cap B)=\mathrm{C}_{m}^{k},$$

故由条件概率计算公式得到

$$P(B \mid A)=\frac{P(A\cap B)}{P(A)}=\frac{\mathrm{C}_{m}^{k}}{\mathrm{C}_{n}^{k}+\mathrm{C}_{m}^{k}}.$$

1.1.5　频率与其加法法则

除了古典模型以外，在带有随机现象的环境中，人们通过长期的实践活动，从经验中认识到，众多的大家关心的事件的发生与否，是具有一个客观的百分比的(古典模型就是在特殊情形下，对这种存在性的认识的佐证)，人们理解事件发生的百分比就是这个事件发生的可能性，自然称之为该事件发生的概率. 然而，人们对概率的朴素认识，并不仅仅局限于等可能性分析. 实践经验是我们认识概率的重要途径. 例如，当我们抛掷一枚硬币 100 次，如其中出现 70 次正面向上，我们就会怀疑这个硬币两面是否均匀. 因为人们对于均匀的实践认识是，正面向上与反面向上的可能性应该差不多. 这种直观的思维，就是经验中估计概率的频率方法，即一个事件发生的可能性的大小，是可以通过多次重复试验，由该事件在其中出现的频繁程度(频率)来估量的. 这就是最朴素的统计思想.

在 n 次重复的独立试验中，分别记随机事件 $A,B,\cdots$ 发生的次数(称为频数)为 $n_A,n_B,\cdots$，我们称

$$f_n(A)=\frac{n_A}{n},\quad f_n(B)=\frac{n_B}{n},\quad \cdots$$

为在这 n 次重复试验中随机事件 $A,B,\cdots$ 出现的**频率**(frequency). 在给定的 n