第一章　事件的概率

1.1　概率是什么

概率，又称或然率，几率，是表示某种情况(事件)出现的可能性大小的一种数量指标，它介于 0 与 1 之间.

这个概念笼统地说起来很容易理解，但若从理论或者说从哲学的高度去分析，就可以提出一大堆的问题. 虽然在本课程范围内我们不必去深入讨论这些问题的各个方面，但仍希望，通过下文的叙述，使读者对"什么是概率"这个问题，有一个较为全面的理解.

1.1.1　主观概率

甲、乙、丙、丁四人一早进城去办事，要傍晚才能回来. 为了决定是否带伞，各自在出发前，对

$$A = \{\text{今天下午 6 时前不会下雨}\}$$

这个情况或事件发生的可能性大小作个估计. 设根据个人的经验和自信，甲、乙、丙、丁分别把这一可能性估计为 0,0.2,0.7 和 1. 这意味着甲认为事件 A 不可能出现，丁认为必然出现，乙认为 A 出现的可能性是有的，但很小，而丙认为 A 有相当大的可能性出现，但并非必然. 这些数字反映了他们四个人对一种情况的主观估计，故称为主观概率. 其实际后果是，例如，甲、乙决定带伞而丙、丁则否.

主观概率可以理解为一种心态或倾向性. 究其根由大抵有二：一是根据其经验和知识. 拿上例来说，若某人在该城市住了 30 年，又是一个有些气象知识的人，他在作出可能性大小的估计时，多半会使用这些经验和知识，这将会使他的估计较易为人所相信. 从这一点说，所谓主观概率也可有其客观背景，终究不同于信口雌黄.

二是根据其利害关系. 拿上例来说,若对某人而言下雨并不会造成多大问题而带伞又增加不少麻烦,则其心态将倾向于去把 A 的可能性高估一些.

主观概率的特点是:它不是在坚实的客观理由基础上为人们所公认的,因而看来应被科学所否定(科学是以探讨客观真理为任务的). 本书作者说不清楚这问题该如何全面地去理解,但不同意简单的全盘否定的态度. 理由有三:①这个概念有广泛的生活基础. 我们几乎无时不在估计种种情况出现的可能性如何,而不同的人很少能在"客观"的基础上达成一致. ②这可能反映认识主体的一种倾向性,而有其社会意义. 例如,"若问三年后经济形势会得到根本改善"的可能性大小怎样,则不同经济状况、社会地位以至政治倾向的人,会作出有差异的估计. 就个别估计而言可能谈不上多大道理,但从总体而言,则反映了社会上广大群众对长远发展的信心如何. 对社会学家乃至决策者来说,这是很有用的资料. ③在涉及利益(经济和其他的)得失的决策问题中,处于不同地位和掌握情报多少不同的人,对某事件可能性大小要参照这些情况及可能的后果去作衡量. 适合于某人的决策,虽则风险较小,不必适合于另一个人,因对他而言,这一决策可能风险仍太大. 因此,主观概率这个概念也有其实用基础. 事实上,许多决策都难免要包含个人判断的成分,而这就是主观概率.

1.1.2 试验与事件

前面我们已经提到了"事件"这个名词. 事件是什么? 在通常的意义下,它往往是指一种已发生的情况,例如某某空难事件,1941 年日本偷袭珍珠港的事件之类. 在概率论中则不然,事件不是指已发生了的情况,而是指某种(或某些)情况的"陈述". 它可能发生,也可能不发生,发生与否,要到有关的"试验"有了结果以后,才能知晓.

拿前例而言,事件 A"陈述"了这样一种情况:下午 6 时前不会下雨. 我们当然并未说这已发生了. 它是否发生,要等试验结果,

这个试验,就是对到下午 6 时前的天气情况进行观察.

推而广之,我们就不难明白:在概率论中,"事件"一词的一般含义是这样的:

1. 有一个明确界定的试验.试验一词,有人为主动的意思,而像上例那样,人只处在被动地位,只是记录而并不干预气象过程.这类情况一般称为"观察".在统计学中,这一分别有时有实际含义,但对目前的讨论不重要,可以把试验一词理解为包含了观察.

2. 这个试验的全部可能结果,是在试验前就明确的,拿上例来说,试验的全部可能结果只有两个:其一是 A,另一是 $\overline{A}=\{$今天下午 6 时前会下雨$\}$.为此,可把这试验写为 $(A,\overline{A})$.不必等到试验完成(不必到下午 6 时)就知道:非 A 即 $\overline{A}$,必居其一.又如,投掷一个赌博用的骰子这个试验,虽无法预卜其结果如何,但总不外乎是"出现 1 点",……,"出现 6 点"这 6 个可能结果之一,因而不妨把这试验简记为 $(1,2,\cdots,6)$.

在不少情况下,我们不能确切知道一试验的全部可能结果,但可以知道它不超出某个范围.这时,也可以用这个范围来作为该试验的全部可能结果.如在前例中,若我们感兴趣的不止在于下午 6 时前是否下雨,而需要记录下午 6 时前的降雨量(如以毫米为单位),则试验结果将是非负实数 x.我们无法确定 x 的可能取值的确切范围,但可以把这范围取为 $[0,\infty)$,它总能包含一切可能的试验结果,尽管我们明知,某些结果,如 $x>10000$,是不会出现的.我们甚至可以把这范围取为 $(-\infty,\infty)$ 也无妨.这里就有了一定的数学抽象,它可以带来很大的方便,这一点在以后会更清楚.

3. 我们有一个明确的陈述,这个陈述界定了试验的全部可能结果中一确定的部分.这个陈述,或者说一确定的部分,就叫做一个事件.如在下雨的例中,A 是全部可能结果 $(A,\overline{A})$ 中确定的一部分.在掷骰子的例中,我们可以定义许多事件,例如

$$E_1=\{\text{掷出偶数点}\}=(2,4,6)$$

$$E_2=\{\text{掷出素数点}\}=(2,3,5)$$

$$E_3=\{\text{掷出 3 的倍数点}\}=(3,6)$$

等等，它们分别明确地界定了全部试验结果的集合(1,2,…,6)中的一个相应的部分.

如果我们现在把试验做一次，即把这骰子投掷一次.则当投掷结果为2，或为4，或为6时，我们说事件 E_1“发生了”，不然就说事件 E_1“不发生”.因此，我们也可以说：事件是与试验结果有关的一个命题，其正确与否取决于试验结果如何.

在概率论上，有时把单一的试验结果称为一个“基本事件”.这样，一个或一些基本事件并在一起，就构成一个事件，而基本事件本身也是事件.在掷骰子的例中，有1,2,…,6等6个基本事件.事件 E_2 则由2,3,5这三个基本事件并成.

设想你处在这样一种情况：投掷一个骰子，若出现素数点，则你将中奖.则在骰子投掷之前你会这样想：我能否中奖，取决于机遇.因此，在概率论中，常称事件为“随机事件”或“偶然事件”.“随机”的意思无非是说，事件是否在某次试验中发生，取决于机遇.其极端情况，是“必然事件”(在试验中必然发生的事情，例如，{掷一个骰子，其出现点数不超过6})和“不可能事件”(在试验中不可能发生的事件).这两种情况已无机遇可言，但为方便计，不妨把它们视为随机事件的特例，正如在微积分中，常数可视为变量的特例.

可以把必然事件和不可能事件分别等同于概率为1和概率为0的事件.从严格的理论角度而言这二者有所区别，但这种区别并无实际的重要性.

本段讲的概念虽很浅显，但是很重要，特别提醒读者区别“事件”一词的日常及在概率论中的不同含义.

1.1.3 古典概率

承接上一段.假定某个试验有有限个可能的结果 $e_1, e_2, \cdots, e_N$.假定从该试验的条件及实施方法上去分析，我们找不到任何理由认为其中某一结果，例如 e_i，比任一其他结果，例如 e_j，更具有优势(即更倾向于易发生)，则我们只好认为，所有结果 $e_1, \cdots, e_N$ 在试验中有同等可能的出现机会，即 $1/N$ 的出现机会.常常把这

样的试验结果称为“等可能的”.

拿掷骰子的例子而言,如果①骰子质料绝对均匀.②骰子是绝对的正六面体.③掷骰子时离地面有充分的高度,则一般人都会同意,其各面出现的机会应为等可能.当然,在现实生活中这只能是一种近似,何况,在骰子上刻上点数也会影响其对称性.

在“等可能性”概念的基础上,很自然地引进古典概率的定义.

定义 1.1 设一个试验有 N 个等可能的结果,而事件 E 恰包含其中的 M 个结果,则事件 E 的概率,记为 $P(E)$,定义为

$$P(E) = M/N \tag{1.1}$$

本定义所根据的理由很显然.按前面的分析,由等可能性的含义,每个结果的概率同为 $1/N$.今事件 E 包含 M 个结果,其概率理应为 $1/N$ 的 M 倍,即 M/N.古典概率是“客观”的.因为,如果等可能性是基于客观事实(例如在骰子绝对均匀且为严格正六面体时)而非出于主观设想,则看来除按(1.1)式外,别无其他的合理定义法.因此在等可能性的前提下,(1.1)式应为大家所公认.这样,关键就在于保证这等可能性成立无误.在开奖时要设计适当的方法并设置公证人,这些措施都是为了保证所用方法导致等可能的结果.

设有一个坛子,其中包含 N 个大小和质地完全一样的球,M 个为白球,$N-M$ 个为黑球.将这 N 个球彻底扰乱,蒙上眼睛,从中抽出一个.则人们都能接受:“抽到白球”这个事件的概率,应取为 M/N.这个“坛子模型”看起来简单却很有用:它是在一切概率的讨论中,唯一的一个易于用形象的方法加以体现的情况.日常习用的按“抽签”来保证机会均等的做法,就是基于这一模型.有了这一模型,我们可以把一些难于理解的概率形象化起来而获得感性.如在“下雨”那个例中,说乙估计事件 A 的概率为 0.20,这听起来不甚了然和不好理解.但如乙说“我认为 A 发生的机会,正如在 4 黑球 1 白球中,抽出白球的机会”,则人们就感到顿时领悟了他的意思.

古典概率的计算主要基于排列组合,将在下一节举一些例子

来说明. 这个名称的来由是远自 16 世纪以来，就有一些学者研究了使用骰子等赌具进行赌博所引起的“机会大小”的问题，由此结晶出概率论的一些最基本的概念，如用(1.1)式定义的概率(赌博中各种结果自应公认为等可能的)及数学期望(见下章)等. 其中一个著名的问题是“分赌本问题”. 在下面已简化了的例中，我们来看看，使用古典概率的概念，如何使这个问题达到一个公正的解决.

例 1.1 甲、乙两人赌技相同，各出赌注 500 元. 约定：谁先胜三局，则谁拿走全部 1000 元. 现已赌了三局，甲二胜一负而因故要中止赌博，问这 1000 元要如何分，才算公平?

平均分对甲欠公平，全归甲则对乙欠公平. 合理的分法是按一定比例而甲拿大头. 一种看来可以接受的方法是按已胜局数分，即甲拿 2/3，乙拿 1/3. 仔细分析，发现这不合理，道理如下：设想继续赌两局，则结果无非以下四种情况之一：

$$\text{甲甲},\text{甲乙},\text{乙甲},\text{乙乙}, \tag{1.2}$$

其中“甲乙”表示第一局甲胜第二局乙胜，余类推. 把已赌过的三局与(1.2)中这四个结果结合(即甲、乙赌完五局)，我们看出：对前三个结果都是甲先胜三局，因而得仟元，只在最在一个结果才由乙得仟元. 在赌技相同的条件下，(1.2)中的四个结果应有等可能性. 因此，甲、乙最终获胜可能性大小之比为 3:1. 全部赌本应按这比例分，即甲分 750 元，乙分 250 元，才算公正合理.

这个例子颇给人启发，即表面上看来简单自然的东西，经过深入一层的分析而揭示了其不合理之处. 这个例子还和重要的“数学期望”的概念相关，见第二章.

古典概率的局限性很显然：它只能用于全部试验结果为有限个，且等可能性成立的情况. 但在某些情况下，这概念可稍稍引申到试验结果有无限多的情况，这就是所谓“几何概率”. 举一个例子.

例 1.2 甲、乙二人约定 1 点到 2 点之间在某处碰头，约定先到者等候 10 分钟即离去. 设想甲、乙二人各自随意地在 1—2 点之间选一个时刻到达该处，问“甲乙二人能碰上”这事件 E 的概率是

多少?

以1点钟作原点,一分为单位,把甲、乙到达时间 x,y 构成的点(x,y)标在直角坐标系上.则图1.1中的正方形 $OABC$ 内每个点都是一个可能的试验结果,而这个正方形就是全部可能的结果之集."甲、乙二人各自随意地在1—2点之间选一个时刻到达该处"一语,可以理解为这正方形内任一点都是等可能.按约定,只有在点(x,y)落在图中的多边形 $OFGBHI$ 内时,事件 E 才发生.因正方形内包含无限个点,古典概率定义(1.1)无法使用.于是,我们把"等可能性"这概念按本问题特点引申一下:正方形内同样的面积有同样的概率.全正方形的面积为 $60^2=3600$,而易算出上述多边形的面积为1100.按上述引申了原则,算出事件 E 的概率为 $P(E)=1100/3600=11/36$.

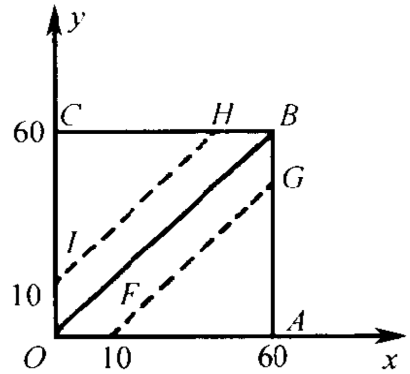

图1.1

这样算出的概率称为"几何概率",因它是基于几何图形的长度、面积、体积等而算出的.就本例而言,重要之点在于把等可能性解释或引申为"等面积,等概率".其他一些可用几何概率处理的问题,都需要作类似的引申.在某些较复杂的问题中,几种引申看来都可接受,由此可算出不同的结果.这并无矛盾可言,因为每一种不同的引申,意味着对"等可能性"的含义作不同的解释.问题在于哪一种解释最符合你的问题的实际含义.

1.1.4 概率的统计定义

从实用的角度看,概率的统计定义无非是一种通过实验去估计事件概率的方法.拿"掷骰子"这个例子来说,若骰子并非质地均匀的正方体,则投掷时各面出现的概率不必相同.这时,"出现幺点"这个事件 E_1 的概率有多大,已无法仅通过一种理论的考虑来确定.但我们可以做实验:反复地将这骰子投掷大量的次数,例如

n 次.若在这 n 次投掷中么共出现 m_1 次,则称 m_1/n 是 E_1 这个事件在这 n 次试验(每次投掷算作一个试验)中的"频率".概率的统计定义的要旨是说,就拿这个频率 m_1/n 作为事件 E_1 的概率 $P(E_1)$ 的估计.这个概念的直观背景很简单:一事件出现的可能性大小,应由在多次重复试验中其出现的频繁程度去刻画.

一般的情况与此毫无区别,只须在上文的叙述中,把"掷骰子"改换成某个一般的试验,而把"出现么点"这事件 E_1 改换成某个指定的事件即可.要点在于:该试验必须能在同样条件下大量次数重复施行,以便我们有可能观察该事件的频率.

读者恐怕已注意到上述定义中的不足之处,即频率只是概率的估计而非概率本身.形式上可以用下面的说法来解脱这个困难.把事件 E 的概率定义为具有如下性质的一个数 p:当把实验重复时,E 的频率在 p 的附近摆动,且当重复次数增大时,这摆动愈来愈小.或者干脆说:概率就是当试验次数无限增大时频率的极限.要这样做,就必须回答下述问题:你怎样去证明具有上述性质的数 p 存在,抑或 p 的存在只是一个假定?

依本书作者的观点,"概率的统计定义"的重要性,不在于它提供了一种定义概率的方法——它实际上没有提供这种方法,因为你永远不可能依据这个定义确切地定出任何一个事件的概率.其重要性在于两点:一是提供了一种估计概率的方法,这在上文已谈到了,这种应用很多.例如在人口的抽样调查中,根据抽样的一小部分人去估计全部人口的文盲比例;在工业生产中,依据抽取的一些产品的检验结果去估计产品的废品率;在医学上依据积累的资料去估计某种疾病的死亡率等.二是它提供了一种检验理论正确与否的准则.设想根据一定的理论、假定等等算出了某事件 A 的概率为 p,这理论或假定是否与实际相符?我们并无把握.于是我们可诉诸实验,即进行大量重复的试验以观察事件 A 的频率 m/n.若 m/n 与 p 接近,则认为实验结果支持了有关理论,若相去较远,则认为理论可能有误.这类问题属于数理统计学的一个重要分支——假设检验,将在本书第五章中讨论.

1.1.5 概率的公理化定义

数学上所说的“公理”，就是一些不加证明而承认的前提.这些前提规定了所讨论的对象的一些基本关系和所满足的条件，然后以之为基础，推演出所讨论的对象的进一步的内容.几何学就是一个典型的例子.

成功地将概率论实现公理化的，是现代前苏联大数学家柯尔莫哥洛夫，时间在1933年.值得赞赏的不止在于他实现了概率论的公理化，还在于他提出的公理为数很少且极为简单，而在这么一个基础上建立起了概率论的宏伟大厦.

在第1.1.2段中我们曾指出：事件是与试验相连，试验有许多可能的结果，每个结果叫做一个基本事件.与此相应，在柯氏的公理体系中引进一个抽象的集合 Ω，其元素 ω 称为基本事件.我们又曾指出：一个事件是由若干基本事件构成.如在掷骰子的试验中，“掷出素数点”这个事件，由2,3,5这三个基本事件构成.与此相应，在柯氏公理体系中考虑由 Ω 的子集(包括 Ω 本身及空集 $\varnothing$)构成的一个集类 $\mathscr{F}$. $\mathscr{F}$不必包括 Ω 的一切可能的子集，且必须满足某种我们在此不必仔细说明的条件. $\mathscr{F}$中的每个成员就称为“事件”.事件有概率，其大小随事件而异，换句话说，概率是事件的函数.与此相应，在柯氏公理体系中，引进了一个定义在 $\mathscr{F}$上的函数 P.对 $\mathscr{F}$中任一成员 A，$P(A)$之值理解为事件 A 的概率.柯氏公理体系对这个函数 P 加上了几条要求(即公理)：①$0 \leqslant P(A) \leqslant 1$ 对 $\mathscr{F}$任何成员 A，这相应于要求概率在0,1之间.②$P(\Omega)=1$，$P(\varnothing)=0$.这相应于说必然事件有概率1，不可能事件有概率0.③加法公理.这一条将在1.3节中解释.

我们举一个简单例子来说明柯氏公理的实现，就是那个“掷骰子”的例子.在本例中，集合 $\Omega=\{1,2,3,4,5,6\}$，由6个元素构成，反映掷骰子试验的6个基本结果.作为 $\mathscr{F}$，在本例中包含 Ω 的一切可能的子集，故 $\mathscr{F}$ 一共有64个成员.至于概率函数 P 的定义，则要考虑骰子的具体情况，若骰子是均匀的正立方体，则 P 定义为

$$P(A) = A \text{ 中所含点数 } /6$$

若骰子非均匀,则每面的出现概率 $p_1, \cdots, p_6$ 可不同.这时,先定出上面这6个数,然后对每个 A,把其中所含点相应的 p 值加起来作为 $P(A)$.例如,若 $A = \{2,3,5\}$,则 $P(A) = p_2 + p_3 + p_5$.

由这个例子我们也看出:柯氏公理只是介定了概率这个概念所必须满足的一些一般性质,它没有也不可能解决在特定场合下如何定出概率的问题.拿后一例子而言,如何以足够的精确度定出 $p_1, \cdots, p_6$,那是要作大量艰苦的工作的.柯氏公理的意义在于它为一种普遍而严格的数学化概率理论奠定了基础.例如,刚才讨论过的这个例子可用于任何一个只有6个基本结果的试验,而无须过问这试验是掷骰子或其他.这就是数学的抽象化.正如我们可说 $1 + 2 = 3$,而不必要去讨论一只牛加二只牛等于三只牛之类的东西.