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

$$A = \{今天下午6时前不会下雨\}$$

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:

$$甲甲,甲乙,乙甲,乙乙, \tag{1.2}$$

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 $P(E)=1100/3600=11/36$.

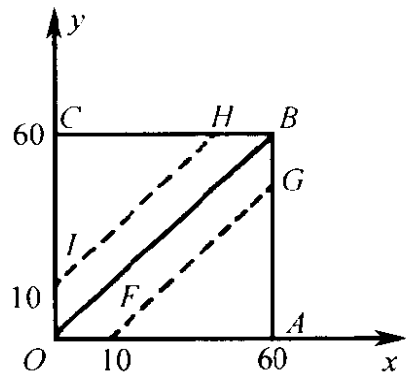

图 1.1

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.

1.2 古典概率计算

1.2.1 排列组合的几个简单公式

按公式(1.1),古典概率计算归结为计算两个数 M 和 N.这种计算大多涉及排列组合.二者的区别在于,排列要计较次序而组合不计较:ab 和 ba 是不同的排列,但是是相同的组合.

1. n 个相异物件取 r 个($1 \leqslant r \leqslant n$)的不同排列总数,为

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) \tag{2.1}$$

因为,从 n 个中取出排列中的第1个,有 n 种取法.在剩下的 $n-1$ 个中取出一个,作为排列中的第2个,有 $n-1$ 种取法…….最后,在剩下的 $n-r+1$ 个中取出一个作为排列中的第 r 个,有 $n-r+1$ 种取法.因此不同的取法数目为 $n, n-1, \cdots, n-r+1$ 这 r 个数之积,从而得出公式(2.1).

例如,从 a, b, c, d 这4个文字中取两个作排列,有 $4 \times 3 = 12$ 种:

$ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, dc$.

特别,若 $n=r$,由(2.1)得

$$P_r^r = r(r-1)\cdots 1 = r! \tag{2.2}$$

$r!$ 读为"r 阶乘",是前 r 个自然数之积.人们常约定把 $0!$ 作为 1.当 r 不是非负整数时,记号 $r!$ 没有意义.

2.n 个相异物件取 r 个($1\leqslant r\leqslant n$)的不同组合总数,为

$$C_r^n = P_r^n/r! = n!/(r!(n-r)!) \tag{2.3}$$

因为,每一个包含 r 物件的组合,可以产生 $r!$ 个不同的排列.故排列数应为组合数的 $r!$ 倍,由此得出公式(2.3).C_r^n 常称为组合系数.

例如,从 a,b,c,d 这 4 个文字中取 2 个作组合.有 $4!/(2!\ 2!)=6$ 种,即 ab,ac,ad,bc,bd,cd.

在有些书籍中把记号 C_r^n 写为 C_n^r.C_r^n 的一个更通用的记号是 $\binom{n}{r}$.我们今后将用 $\binom{n}{r}$ 取代 C_r^n.当 $r=0$ 时,按 $0!=1$ 之约定,由(2.3)算出 $\binom{n}{0}=1$,这可看作一个约定.对组合系数另一常用的约定是:按公式

$$\binom{n}{r} = n(n-1)\cdots(n-r+1)/r!$$

只要 r 为非负整数,n 不论为任何实数,都有意义.故 n 可不必限制为自然数.例如,按上式,有

$$\binom{-1}{r} = (-1)(-2)\cdots(-r)/r! = (-1)^r$$

3.与二项式展开的关系

组合系数 $\binom{n}{r}$ 又常称为二项式系数,因为它出现在下面熟知的二项式展开的公式中:

$$(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^i b^{n-i} \tag{2.4}$$

这个公式的证明很简单:因为,$(a+b)^n=(a+b)\cdot(a+b)\cdots(a+$

b). 为了产生 $a^i b^{n-i}$ 这一项,在这 n 个$(a+b)$中,要从其中的 i 个取出 a,另 $n-i$ 个取出 b. 从 n 个中取出 i 个的不同取法为$\binom{n}{i}$,这也就是 $a^i b^{n-i}$ 这一项的系数.

利用关系(2.4)可得出许多有用的组合公式. 例如,在(2.4)中令 $a=b=1$,得

$$\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}=2^n$$

令 $a=-1,b=1$,则得

$$\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\cdots+(-1)^n\binom{n}{n}=0$$

另一个有用的公式是

$$\binom{m+n}{k}=\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i} \tag{2.5}$$

它是由恒等式$(1+x)^{m+n}=(1+x)^m(1+x)^n$ 即

$$\sum_{j=0}^{m+n}\binom{m+n}{j}x^j=\sum_{j=0}^{m}\binom{m}{j}x^j\sum_{j=0}^{n}\binom{n}{j}x^j$$

比较两边的 x^k 项的系数得到的.

4. n 个相异物件分成 k 堆,各堆物件数分别为 $r_1,\cdots,r_k$ 的分法是

$$n!/(r_1!\cdots r_k!) \tag{2.6}$$

此处 $r_1,\cdots,r_k$ 都是非负整数,其和为 n,又这里要计较堆的次序. 就是说,若有 5 个物体 a,b,c,d,e 分成 3 堆,则(ac),(d),(be)和(be),(ac),(d)是算作两种不同分法.

证明很简单:先从 n 个中取出 r_1 个作为第 1 堆,取法有$\binom{n}{r_1}$种. 在余下的 $n-r_1$ 个中取出 r_2 个作为第 2 堆,取法有$\binom{n-r_1}{r_2}$种,以此类推,得到全部不同的分法为

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-r_2-\cdots-r_{k-1}}{r_k}$$

利用公式(2.3)并注意 $n-r_1-\cdots-r_{k-1}=r_k$,即得(2.6)

(2.6)常称为多项式系数,因为它是$(x_1+\cdots+x_k)^n$ 的展开式中,$x_1^{r_1}\cdots x_k^{r_k}$ 这一项的系数.

1.2.2 古典概率计算举例

例 2.1 一批产品共 N 个,其中废品有 M 个.现从中随机(或说随意)取出 n 个,问"其中恰好 m 个废品"这个事件 E 的概率是多少?

按 1.2.1 所述,从 N 个产品中取出 n 个,不同的取法有 $\binom{N}{n}$ 种.所谓"随机"或"随意"取,是指这 $\binom{N}{n}$ 种取法有等可能性.这是古典概率定义可以使用的前提.所以,从实际的角度言,问题在于怎样保证抽取的方法能满足等可能性这个要求.以下各例中"随机"一词也都是作这种理解.

使事件 E 发生的取法,或者说"有利"于事件 E 的取法,计算如下:从 M 个废品中取 m 个,取法有 $\binom{M}{m}$ 种.从其余 $N-M$ 个合格品中取 $n-m$ 个,取法有 $\binom{N-M}{n-m}$ 种.故有利于事件 E 的取法,共有 $\binom{M}{m}\binom{N-M}{n-m}$ 种.按公式(1.1),得事件 E 的概率为

$$P(E)=\binom{M}{m}\binom{N-M}{n-m}\Big/\binom{N}{n} \tag{2.7}$$

这里要求 $m\leqslant M, n-m\leqslant N-M$,否则概率为 0(因 E 为不可能事件).

例 2.2 n 双相异的鞋共 $2n$ 只,随机地分成 n 堆,每堆 2 只.问"各堆都自成一双鞋"这个事件 E 的概率是多少?

把 $2n$ 只鞋分成 n 堆每堆 2 只的分法,按公式(2.6),有 $N=$

$(2n)!\ /2^n$ 种. 有利于事件 E 的分法可计算如下:把每双鞋各自绑在一起看成一个物体,然后把这相异的 n 个物体分成 n 堆,每堆 1 件. 按公式(2.6),分法有 $M=n!$ 种. 于是

$$P(E) = M/N = n!2^n/(2n)! = 1/(2n-1)!!$$

$a!!$ 这个记号对奇自然数定义:$a!! = 1\cdot3\cdot5\cdots a$,即所有不超过 a 的奇数之积.

另一种算法如下:把这 $2n$ 只鞋自左至右排成一列(排法有 $(2n)!$ 种),然后,把处在 1,2 位置的作为一堆,3,4 位置的作为一堆,等等. 为计算使事件 E 发生的排列法,注意第 1 位置可以是这 $2n$ 只鞋中的任一只,其取法有 $2n$ 种. 第 1 位置取定后,第 2 位置只有一种取法,即必须取与第 1 位置的鞋配成一双的那一只. 依此类推,知奇数位置依次有 $2n, 2n-2, 2n-4, \cdots, 2$ 种取法,而偶数位置则都只有 1 种取法. 所以,有利于事件 E 的排列总数为 $2n(2n-2)\cdots2=2^n n!$,而

$$P(E) = 2^n n!/(2n)!$$

与前面用另外的方法算出的相同.

例 2.3 n 个男孩,m 个女孩($m\leqslant n+1$)随机地排成一列. 问"任意两个女孩都不相邻"这个事件 E 的概率是什么?

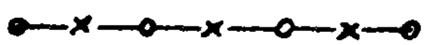

图 1.2

把 $n+m$ 个孩子随意排列,总共有 $N=(n+m)!$ 种不同的排法. 有利于事件 E 发生的排法可计算如下:先把 n 个男孩子随意排成一列,总共有 $n!$ 种方法. 排定以后,每两个相邻男孩之间有一位置,共有 $n-1$ 个;加上头尾两个位置,共 $n+1$ 个位置(图 1.2 画出了 $n=3$ 的情况,"×"表示男孩,4 个"○"表示刚才所指出的 $n+1=4$ 个位置). 为了使两个女孩都不相邻,必须从这 $n+1$ 个位置中取出 m 个放女孩,取法有 $\binom{n+1}{m}$ 种. 取定位置后,m 个女孩子尚可在这 m 个取定位置上随意排列,方法有 $m!$ 种. 由此推出,

有利于事件 E 发生的排列数为 $M = n!\binom{n+1}{m}m!$，因此，

$$P(E) = n!\binom{n+1}{m}m!/(n+m)! = \binom{n+1}{m}\Big/\binom{n+m}{m}$$

如果这 $n+m$ 个孩子不是排成一直线而是排在一圆圈上，则同一事件 E 的概率是多少？初一看以为无所区别，其实不然. 看图 1.2，若以"×"和"○"分别表男、女孩，则在一直线上首尾两女孩并不相邻. 但若把这直线弯成一个圆圈，则首尾两女孩成为相邻了，因此算法略有不同. 我们留给读者去证明：答案为 $\binom{n}{m}\Big/\binom{n+m-1}{m}$.

例 2.4 一个人在口袋里放 2 盒火柴，每盒 n 支. 每次抽烟时从口袋中随机拿出一盒（即每次每盒有同等机会被拿出）并用掉一支. 到某次他迟早会发现：取出的那一盒已空了. 问："这时另一盒中恰好有 m 支火柴"的概率是多少？

解法 1 我们来考察最初 $2n+1-m$ 次抽用的情况，每次抽用时有 2 种方法（抽出甲盒或乙盒）. 故总的不同抽法，有 2^{2n+1-m} 种. 有利于所述事件的抽法可计算如下：先看"最后一次（即第 $2n+1-m$ 次）是抽出甲盒"的情况. 为使所述事件发生，在前 $2n-m$ 次中，必须有 n 次抽用甲盒，实现这一点不同的抽法为 $\binom{2n-m}{n}$. 类似地，"最后一次是抽出乙盒"的抽法也有这么多，故有利于所述事件的全部抽法为 $2\binom{2n-m}{n}$，而事件的概率为

$$2\binom{2n-m}{n}\Big/2^{2n+1-m} = \binom{2n-m}{n}\Big/2^{2n-m} \qquad (2.8)$$

解法 2 因每盒中只有 n 支，最晚到第 $2n+1$ 次抽取时，或在此之前，必发现抽出的盒子已空. 故我们不管结果如何，总把试验做到抽完第 $2n+1$ 次为止，不同的抽法有 2^{2n+1} 种.

现在计算有利于所述事件的抽法. 仍如前，先考虑"先发现甲

盒为空"的抽法有多少.这必然是对某个 $r, r=0,1,\cdots,n-m$,以下情况同时出现:

1° 第 $n+r$ 次抽取时抽出甲盒,而这时甲盒已是第 n 次被抽出;

2° 前 $n+r-1$ 次抽取时,乙盒被抽出 r 次(这不同的抽法有 $\binom{n+r-1}{r}$ 种);

3° 紧接着的 $n-m-r$ 次全是抽出乙盒;

4° 第 $2n-m+1$ 次抽取时抽出甲盒(这时发现它已空,且乙盒恰有 m 支);

5° 最后 m 次抽取结果可以任意(这不同的抽法有 2^m 种).

综合上述,对固定的 r,抽法有 $\binom{n-1+r}{r}2^m$ 种.因此,"有利于事件发生,且先发现甲盒为空"的抽法,有

$$a=\sum_{r=0}^{n-m}\binom{n-1+r}{r}2^m$$

种.类似地,"有利于事件发生,且先发现乙盒为空"的抽法,也有 a 种,故总数为 $2a$,概率为

$$2a/2^{2n+1}=\sum_{r=0}^{n-m}\binom{n-1+r}{r}\Big/2^{2n-m} \tag{2.9}$$

两种方法算出的结果,只能有一个.故比较(2.8)和(2.9),我们得到一个组合恒等式

$$\sum_{r=0}^{n-m}\binom{n-1+r}{r}=\binom{2n-m}{n}$$

当然,你也可以怀疑,这两个解法中有一个不对,因而上式也可能错了.但此式可另行证明.为方便计,将式中的 m 改为 $n-m$,而将该式写为

$$\sum_{r=0}^{m}\binom{n-1+r}{r}=\binom{n+m}{n}=\binom{n+m}{m}$$

而因式易用数学归纳法证明:当 $m=0,1$ 时,直接计算可知其成

立，然后用易证之等式

$$\binom{n+m}{m}+\binom{n+m}{m+1}=\binom{n+m+1}{m+1}$$

去完成归纳证明.

这个例子给人的启发是：适当的考虑得出的简洁解法.第二种解法，把试验做到必然能见分晓的地步，较为自然易懂，但结果则繁复：要不是有(2.8)对照，我们可能停留在(2.9)，而得出不理想的形式.前一解法抓住了这一点：要使所设事件发生，抽取必然是 $2n+1-m$ 次.这一简单的观察导致了远为简洁的解(2.8).

例 2.5 有 21 本不同的书，随机地分给 17 个人.问"有 6 人得 0 本，5 人得 1 本，2 人得 2 本，4 人得 3 本"这个事件 E 的概率是多少？

因为每本书都有 17 种可能的分法，故总的不同分法，有 17^{21} 种.为计算有利于事件 E 的分法，得分两步分析：①按得书本数不同把 17 人分成 4 堆，各堆分别含 6(0 本)、5(1 本)、2(2 本)、4(3 本)人.这不同的分法按公式(2.6)，有 $17!/(6!\ 5!\ 2!\ 4!)$ 种.②把 21 本书按 17 人得书数情况分为 17 堆，各堆数目依次为

$$0,0,0,0,0,0,1,1,1,1,1,2,2,3,3,3,3$$

不同分法有

$$21!/(0!^6 1!^5 2!^2 3!^4)=21!/(2!^2 3!^4)$$

二者相乘，得出有利于事件 E 的分法总数，进而得出 E 的概率为

$$17!\ 21!/(17^{21} 2!^3 3!^4 4!\ 5!\ 6!)$$

以上举的例子都有一定的代表性.古典概率计算实质上就是排列组合计算.但在分析问题时，怎样去选定一个适当的实现随机化的机制(如例 2.4，例 2.5)，怎样去正确计算公式(1.1)中的 M，N，以保证既不重算也不漏算，则需要细心.尤其是：你所设想的机制是否真的实现了等可能性？有时表面上看想当然对，其实是似是而非的.如例 2.3 中，圆圈的情况和直线有所不同——在直线上正确地体现了等可能的做法，在圆圈上却没有.再看下例.

例 2.6 n 本书随机给分甲、乙二人，问"甲、乙各至少得到 1

本"这事件 E 的概率是多少?

n 本书随机地分给 2 人,甲得的本数无非是 $0,1,\cdots,n$,一共有 $n+1$ 种可能性,其中 0 和 n 两种是"全归一人",剩下 $n-1$ 种有利于 E,故 $P(E)=(n-1)/(n+1)$.

这个解法是否对? 不对.问题在于:$0,1,\cdots,n$ 这 $n+1$ 种结果不具有等可能性.凭常识可以推想:若 n 较大,则甲得 $n/2$ 本左右的机会,应比他全得或全不得的机会大一些.正确的解法如下:n 本书分给 2 人,不同的分法有 2^n 种.其中仅有两种是使事件 E 不发生的,故 $P(E)$应为$(2^n-2)/2^n=1-1/2^{n-1}$.

1.3 事件的运算、条件概率与独立性

在实用上和理论上,下述情况常见:问题中有许多比较简单的事件,其概率易于算出或是有了理论上的假定值,或是根据以往的经验已对其值作了充分精确的估计.而我们感兴趣的是一个复杂的事件 E,它通过种种关系与上述简单事件联系起来.这时我们想设法利用这种联系,以便利用这些简单事件的概率去算出 E 的概率.正如在微积分中,直接利用定义可算出若干简单函数的导数,但利用导数所满足的法则,可据此算出很复杂的函数的导数.

例如,向一架飞机射击,事件 E 是"击落这架飞机".设这架飞机有一名驾驶员,两个发动机 G_1 和 G_2.又假定当击中驾驶员,或同时击中两个发动机时,飞机才被击落,记事件

$$E_0 = \text{击中驾驶员}, E_i = \text{击中 } G_i, i = 1,2$$

则 E 与E_0,E_1,E_2 有关,确切地说,E 即由 E_0,E_1,E_2 决定.其关系可通过文字表达如下:

$$E = \{E_0 \text{ 发生或者 } E_1,E_2 \text{ 都发生}\}$$

这种表述很累赘,我们希望通过一些符号来表达,这就是本节要讨论的事件的关系和运算.对事件进行运算,如同对数字作运算一样:对数字进行运算得出新的数,而对事件作运算则得出新的事件.

1.3.1　事件的蕴含、包含及相等

在同一试验下的两事件 A 和 B，如果当 A 发生时 B 必发生，则称 A 蕴含 B，或者说 B 包含 A，记为 $A \subset B$. 若 A，B 互相蕴含，即 $A \subset B$ 且 $B \subset A$，则称 A，B 两事件相等，记为 $A = B$.

例如，掷两粒骰子. 记

$$A = \{\text{掷出的点数之和大于 } 10\}$$
$$B = \{\text{至少有一粒骰子掷出 } 6\}$$

若事件 A 发生，易见 B 非发生不可，故 A 蕴含 B. 一个形象的看法如图 1.3. 向一个方形靶面射击，以 A，B 分别记“命中图中所标出的闭曲线内部”的事件，则命中 A 自意味着命中 B. 这个图形也说明了“B 包含 A”这个说法的来由. 因从图中明白看出，B 这一块包含了 A 这一块.

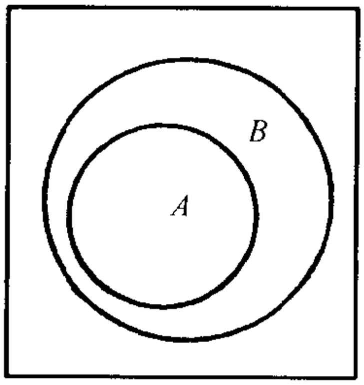

图 1.3

拿“事件是试验的一些结果”(见 1.1.2 段)这个观点去看，如果 A 蕴含 B，那只能是：A 中的试验结果必在 B 中，即 B 这个集合(作为试验结果的集合)要大一些，“包含”一词即由此而来. 实际含义是：若 $A \subset B$(也写为 $B \supset A$)，则 A 和 B 相比，更难发生一些，因而其概率就必然小于或至多等于 B 的概率. “两事件 A，B 相等”无非是说，A，B 由完全同一的一些试验结果构成，它不过是同一件事表面上看来不同的两个说法而已.

例如，掷两个骰子，以 A 记事件“两骰子掷出点数奇偶不同”，B 记事件“掷出点数之和为奇数”. 这两个事件，说法不同，其实则一. 对复杂情况则不必如此一目了然. 证明两事件 A，B 相等的一般方法是：先设事件 A 发生，由此推出 B 发生，再反过来，由假定 B 发生推出 A 发生. 这将在后面举例说明.

1.3.2 事件的互斥和对立

若两事件 A,B 不能在同一次试验中都发生(但可以都不发生),则称它们是互斥的.如果一些事件中任意两个都互斥,则称这些事件是两两互斥的,或简称互斥的.

例如,考虑投掷一个骰子这个试验.记 E_i 为事件"掷出的点数为 i 的倍数",$i=2,3,4$,则 E_3 与 E_4 为互斥.因若 E_4 发生,则只有掷出4点,而它非3的倍数,即 E_3 必不发生.但是,E_2 和 E_3 并非互斥.因若掷出6点,则二者同时发生.简言之,互斥事件即不两立之事件.从"事件是由一些试验结果所构成的"这个观点看,互斥事件无非是说:构成这两个事件各自的试验结果中不能有公共的.

互斥事件的一个重要情况是"对立事件",若 A 为一事件,则事件

$$B = \{A \text{ 不发生}\}$$

称为 A 的对立事件,多记为 $\overline{A}$(读作 $Abar$,也记为 A^c).

例如,投掷一个骰子,事件 $A=\{$掷出奇数点$\}=\{1,3,5\}$ 的对立事件是 $B=\{$掷出偶数点$\}=\{2,4,6\}$.对立事件也常称为"补事件".拿上例来说,事件 A 包含了三个试验结果:1,3和5,而对立事件 B 中所含的三个试验结果2,4和6,正好补足了前面三个,以得到全部试验结果.

1.3.3 事件的和(或称并)

设有两事件 A,B,定义一个新事件 C 如下:

$$C = \{A \text{ 发生,或 } B \text{ 发生}\} = \{A,B \text{ 至少发生一个}\}$$

所谓定义一个事件,就是指出它何时发生,何时不发生.现在这个事件 C 在何时发生呢?只要 A 发生,或者 B 发生(或二者同时发生也可以),就算是 C 发生了,不然(即 A,B 都不发生)则算作 C 不发生,这样定义的事件 C 称为事件 A 与事件 B 的和,记为

$$C = A + B$$

例如,掷一个骰子,以 A 记事件$\{$掷出偶数点$\}=\{2,4,6\}$,B

记事件{掷出 3 的倍数}={3,6},则 $C=A+B=\{2,3,4,6\}$,即当掷出的点为 2,3,4 或 6 时,事件 C 发生,而掷出 1,5 时则不发生. 我们注意到,两事件的和,即把构成各事件的那些试验结果并* 在一起所构成的事件. 如把图 1.4 的正方形视为一个平面靶,A,B 两事件分别表示命中图中所指闭曲线内部,则 $C=A+B$表示"命中由 A,B 两闭曲线的外缘所围成的区域". 这区域比 A,B 都大,它由 A,B 两部分合并而成. 当然,作为集合,重复的部分(图中斜线标出的部分)只须计入一次.

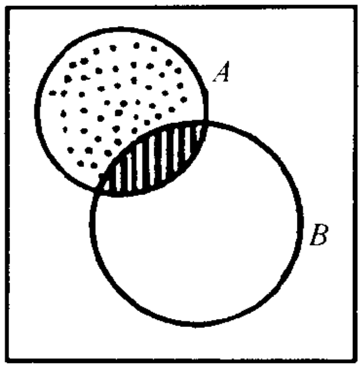

图 1.4

这样,若 $C=A+B$,则 A,B 都蕴含 C,C 包含 A 也包含 B. 经过相加,事件变"大"了(含有更多的试验结果),因而更容易发生了.

事件的和很自然地推广到多个事件的情形. 设有若干个事件 $A_1,A_2,\cdots,A_n$. 它们的和 A,定义为事件

$$\begin{aligned}A &=\{A_1\text{ 发生,或 }A_2\text{ 发生},\cdots,\text{或 }A_n\text{ 发生}\}\\&=\{A_1,A_2,\cdots,A_n\text{ 至少发生一个}\}\end{aligned}$$

且记为 $A_1+A_2+\cdots+A_n$ 或 $\sum_{i=1}^{n}A_i$(也常记为 $\bigcup_{i=1}^{n}A_i$,本书不用这个记号). A 是由把 $A_1,\cdots,A_n$ 所包含的全部试验结果并在一起所得. 和的定义显然地推广到无限个事件的情形.

在此要不厌其烦地重复一点. 有的初学者对事件的运算感到不易理解. 比如,定义事件 A,B 之和为 $C=\{A,B$ 至少发生其一}. 他们问:既然已说 A,B 至少要发生一个,那岂不是对 A,B

---

* 由于这个原因,事件的和也常称为事件的并,和 $A+B$ 也常被记为 $A\cup B$."$\cup$"这个记号有"合并"的含义,由于称呼和书写上的方便,本书中我们一直用"和"与"+"的说法,也有些著作在当 A,B 互斥时才把 $A\cup B$ 写成 $A+B$,本书不采用这个做法.

作了限制？不然，我们不要忘记 1.1 节中所说的“事件不是指已发生了的情况，而是某种情况的陈述”.定义 C 为“A,B 至少发生其一”，当然不是说 A,B 已经或必然发生一个，而是在试验时，若 A,B 至少发生了一个，则算作 C 发生了.在任一次特定的试验中，当然可能 A,B 都不发生，这时 C 也就不发生.理解了这一点就好办，望读者多加留意.

1.3.4 概率的加法定理

定理 3.1 若干个互斥事件之和的概率，等于各事件的概率之和：

$$P(A_1+A_2+\cdots)=P(A_1)+P(A_2)+\cdots \tag{3.1}$$

事件个数可以是有限的或无限的，这定理就称为(概率的)加法定理，其重要条件是各事件必须为两两互斥.

在概率的古典定义和统计定义之下，(3.1)很容易证明.拿古典定义来说，设试验一共有 N 个等可能的结果，而有利于事件 $A_1,A_2,\cdots$ 发生的结果数分别为 $M_1,M_2,\cdots$，则由于互斥性，有利于事件 $A=A_1+A_2+\cdots$ 发生的结果数，应为 $M=M_1+M_2+\cdots$.于是

$$\begin{aligned}P(A)&=(M_1+M_2+\cdots)/N=M_1/N+M_2/N+\cdots\\&=P(A_1)+P(A_2)+\cdots\end{aligned}$$

对统计定义也完全类似地处理.

在概率论书籍中，加法定理往往被称为加法公理，即(3.1)是不加证明而被接受的事实.这条公理就是我们在 1.1.5 段中提到而未加说明的，柯氏公理体系中的第 3 条.

读者可能会问：既然在古典定义、统计定义这样在实用上重要的概率定义之下，(3.1)是可以证明的，那么为什么要把它看作一条公理？问题在于：你可以想像而且也确实可以建立一种概率理论，其中(3.1)不成立.柯氏公理的意思是说：我只考虑那种满足(3.1)的概率理论，而不及其他.正如在几何学中，你可以把“过不

在直线 l 上的任一点只有一条与 l 平行的直线"作为公理,由之建立一套欧氏几何学,也可以废弃这条公理而建立非欧几何学,二者都符合形式逻辑.古典和统计定义之适合(3.1),不过是说明了:它们是柯氏公理体系中的东西.

加法定理(3.1)的一个重要推论如下:

系 3.1 以 $\overline{A}$ 表 A 的对立事件,则

$$P(\overline{A}) = 1 - P(A) \tag{3.2}$$

证明很容易.以 Ω 记必然事件,则按对立事件的定义有 $A+\overline{A}=\Omega$ 且 A 和 $\overline{A}$ 互斥.因 $P(\Omega)=1$.用(3.1)得 $1=P(\Omega)=P(A+\overline{A})=P(A)+P(\overline{A})$,即(3.2).

这个简单公式在概率计算上有用.因为,有时计算 $P(A)$ 不易,而 $P(\overline{A})$ 则易处理些.

1.3.5 事件的积(或称交)、事件的差

设有两事件 A,B,则如下定义的事件 C

$$C = \{A,B \text{ 都发生}\}$$

称为两事件 A,B 之积或乘积,并记为 AB.拿图 1.4 的例子来说,若分别以 A,B 表示"命中图中相应区域"的事件,则 AB 就是事件"命中图中斜线部分".又如骰子试验,分别以 A,B 记"掷出偶数点"和"掷出素数点"之事件,则 AB 就是事件"掷出 2 点".一般,事件 A,B 各是一些试验结果的集合,而 AB 则由同属于这两个集合的那些试验结果组成,即这两个集合的交叉*按积的定义,两个事件 A,B 互斥,等于说 AB 是不可能事件.

多个事件 $A_1,A_2,\cdots$(有限或无限个都可以)的积的定义类似:$A=\{A_1,A_2,\cdots \text{都发生}\}$,记为 $A=A_1A_2\cdots$,或 $\prod_{i=1}^{n} A_i$(事件个数有限)或 $\prod_{i=1}^{\infty} A_i$(事件个数无限).

---

* 由于这个原因,事件的积也常称为事件的交,积 AB 也常记为 $A\cap B$."$\cap$"这个记号有取交的含义.为书写方便,本书一直用 AB 这个记号.

两个事件 A,B 之差，记为 $A-B$，定义为

$$A-B=\{A\text{ 发生},B\text{ 不发生}\}$$

例如，则才提到的掷骰子试验中的两个事件 A 和 B，$A-B=\{4,6\}$. 在图 1.4 中，$A-B$ 就是“命中图中用点标出的区域”这个事件. 一般地，$A-B$ 就是从构成 A 的那些试验结果中，去掉在 B 内的那一些. 很明显

$$A-B=A\overline{B} \tag{3.3}$$

其中 $\overline{B}$ 是 B 的对立事件. 因为，$A\overline{B}$ 无非是说，$A,\overline{B}$ 都发生，即 A 发生 B 不发生. 这样，差可以通过积去定义.

我们对事件引进了和差积等运算，借用了算术中的名词. 但应注意，算术的法则不一定能用于事件运算. 有些规则是成立的，例如，和 $A+B$ 及积 AB 与次序无关：$A+B=B+A$，$AB=BA$，这由定义直接看出. 乘法结合律也成立：$(AB)C=A(BC)$（它们都等于 ABC）. 分配律也对，例如：

$$A(B-C)=AB-AC \tag{3.4}$$

证明如下：设在左边的事件发生，则按积的定义，事件 A 和 $B-C$ 都发生. 按差的定义，B 发生，C 不发生. 因此，A,B 同时发生而 A,C 不同时发生，故 AB 发生而 AC 不发生. 按差的定义，即知 $AB-AC$ 发生. 反过来，若右边的事件发生，则 AB 发生而 AC 不发生. 由前者知 A,B 都发生，由 A 发生及 AC 不发生，知 C 不发生，故 $B-C$ 发生. 因 A 和 $B-C$ 都发生知 $A(B-C)$ 发生，这证明了(3.4).

这就是我们在本节 1.3.1 段末尾处指出的证明事件相等的一般方法之一实例. 读者必须了解，像(3.3)，(3.4)这类的等式，不过是反映了一种逻辑关系，因而必须用上述逻辑思维的方式去验证. 有些关系，看来不习惯，但逻辑上很简单. 例如，$A+A=A$ 而非 $2A$（$2A$ 无意义），$AA=A$ 而非 A^2（A^2 无意义），由 $A-B=\varnothing$（不可能事件），推不出 $A=B$，而只能推出 $A\subset B$. 又如，$(A-B)+B$ 并不是 A 而是 $A+B$（请读者自证），等等.

1.3.6 条件概率

一般讲，条件概率就是在附加一定的条件之下所计算的概率．从广义的意义上说，任何概率都是条件概率，因为，我们是在一定的试验之下去考虑事件的概率的，而试验即规定有条件．在概率论中，规定试验的那些基础条件被看作是已定不变的．如果不再加入其他条件或假定，则算出的概率就叫做"无条件概率"，就是通常所说的概率．当说到"条件概率"时，总是指另外附加的条件，其形式可归结为"已知某事件发生了"．

例如，考虑掷一个骰子的实验．这里，骰子必须为均匀的正立方体，抛掷要有足够的高度等要求，是这试验的固有规定，不作为附加条件．考虑三个事件：A："掷出素数点"，B："掷出奇数点"，C："掷出偶数点"，有

$$A = \{2,3,5\}, B = \{1,3,5\}, C = \{2,4,6\} \qquad (3.5)$$

于是算出 A 的(无条件)概率为 $3/6=1/2$．现若附加上"已知 B 发生"，则可能情况只有三种：1，3，5，其中两种有利于 A 发生，故在这条件下，A 的条件概率，记为 $P(A|B)$，等于 $2/3$．同样，在给定事件 C 发生的条件下，A 的条件概率为 $P(A|C)=1/3$．

让我们在古典概率的模式下来分析一般的情况．设一试验有 N 个等可能结果，事件 A，B 分别包含其 M_1 和 M_2 个结果，它们有 M_{12} 个是公共的，这就是事件 AB 所包含的试验结果数．若已给 B 发生，则我们的考虑由起先的 N 个可能结果局限到现在的 M_2 个，其中只有 M_{12} 个试验结果使事件 A 发生，故一个合理的条件概率定义，应把 $P(A|B)$ 取为 M_{12}/M_2．但

$$M_{12}/M_2 = (M_{12}/N)/(M_2/N) = P(AB)/P(B)$$

由此得出如下的一般定义：

定义 3.1 设有两事件 A，B 而 $P(B)\neq 0$．则"在给定 B 发生的条件下 A 的条件概率"，记为 $P(A|B)$，定义为

$$P(A|B) = P(AB)/P(B) \qquad (3.6)$$

当 $P(B)=0$ 时，(3.6)无意义．在高等概率论中，也要考虑

$P(A|B)$当 $P(B)=0$ 时的定义问题,那要牵涉到高深的数学,超出本书范围之外.在后面我们也会和个别这种情况打交道,那可以用极限的方法去处理.

(3.6)是条件概率的一般定义,但在计算条件概率时,并不一定要有它.有时,直接从加入条件后改变了的情况去算,更为方便.举一个例子.

例 3.1 掷三个均匀骰子.已知第一粒骰子掷出幺点(事件 B).问:"掷出点数之和不小于 10"这个事件 A 的条件概率是多少?

既然第一粒骰子已坐定了 1,则在这一条件下,为使事件 A 发生,第二、三粒骰子掷出点数之和不能小于 9.这一情况有 10 种,即 36,63,45,54,46,64,55,56,65,66.这里"36"表示第二、三粒骰子分别掷出 3 和 6,余类推,这样,得出 $P(A|B)=10/36=5/18$.

此题若直接用公式(3.6)计算,则比上述解法复杂些,读者可一试以证明结果一致.

1.3.7 事件的独立性,概率乘法定理

设有两事件 A,B.A 的无条件概率 $P(A)$ 与其在给定 B 发生之下的条件概率 $P(A|B)$,一般是有差异的.这反映了这两事件之间存在着一些关联.例如,若 $P(A|B)>P(A)$,则 B 的发生使 A 发生的可能性增大了:B 促进了 A 的发生.

反之,若 $P(A)=P(A|B)$,则 B 的发生与否对 A 发生的可能性毫无影响*.这时在概率论上就称 A,B 两事件独立,而由(3.6)得出

---

* 这样说应补充:由 $P(A)=P(A|B)$ 推出 $P(A)=P(A|\overline{B})$,$\overline{B}$ 为 B 的对立事件.事实上,由 $P(A)=P(A|B)$ 及(3.6)知 $P(AB)=P(A)P(B)$.因为 $A=AB+A\overline{B}$ 且 AB,$A\overline{B}$ 互斥,知 $P(A\overline{B})=P(A)-P(AB)=P(A)-P(A)P(B)=P(A)(1-P(B))=P(A)P(\overline{B})$.故 $P(A|\overline{B})=P(A\overline{B})/P(\overline{B})=P(A)$.

$$P(AB)=P(A)P(B) \tag{3.7}$$

拿此式来刻画独立性,比用 $P(A)=P(A|B)$ 更好,因(3.7)不受 $P(B)$ 是否为 0 的制约(当 $P(B)$ 为 0 时(3.7)必成立).因此,我们取如下的定义:

定义 3.2 两事件 A,B 若满足(3.7),则称 A,B 独立.

定理 3.2 两独立事件 A,B 的积 AB 之概率 $P(AB)$ 等于其各自概率之积 $P(A)P(B)$.

这个定理就是(3.7)式,它称为"概率的乘法定理".其实,它就是独立性的定义,我们之所以又将它重复列出并标为一个定理,就是因为这个事实极其重要.

在实际问题中,我们并不常用(3.7)式去判断两事件 A,B 是否独立,而是相反:从事件的实际角度去分析判断其不应有关联因而是独立的,然后就可以用(3.7).例如,两个工人分别在两台机床上进行生产,彼此各不相干,则各自是否生产出废品或多少废品这类事件应是独立的.一城市中两个相距较远的地段是否出交通事故,一个人的收入与其姓氏笔划,这类事凭常识推想,认定为独立的.

由此可知,两事件有独立性多半是在下述情况之下产生的:有两个试验 E_1 和 E_2,其试验结果(各有许多)分别记之以 e_1 和 e_2.考虑一个"大"试验 E,它由 E_1,E_2 两部分构成(故 E 常称为复合试验),可记为 $E=(E_1,E_2)$,其结果可记为 (e_1,e_2).在试验 E 的一个事件,即是牵涉到 (e_1,e_2) 的某一个陈述(见 1.1.2).如果 A_1,A_2 是两个事件,A_1 只牵涉 e_1 而 A_2 只牵涉 e_2,则当两试验结果如果彼此不影响时,A_1,A_2 会有独立性.可以举一个具体例子,设试验 E_1 为掷一个均匀骰子,其试验结果 e_1 有 6 个:1,2,…,6.试验 E_2 为掷一个硬币,其结果 e_2 有两个:"正"和"反".定义两事件 A_1,A_2:

$$A_1=\{\text{掷出 1 点}\},A_2=\{\text{掷出正面}\}$$

这两个事件可看成同一试验 E 下的两个事件,$E=\{E_1,E_2\}$,它包含 12 个可能结果:

$$(1,正),(1,反),(2,正),(2,反),\cdots,(6,正),(6,反)$$

事件 A_1 包含两个可能结果,即$\{(1,正),(1,反)\}$,而 A_2 则包含 6 个可能结果:$\{(1,正),(2,正),\cdots,(6,正)\}$.通过这种方式,我们把两个看来不相干的事件 A_1 和 A_2 统一在一个试验 E 之下,而其独立性就好理解了——即掷骰子和掷硬币彼此不影响而已.这种把若干个不相干的试验统一起来的做法,看起来好像纯粹是一种形式,但在理论上有其方便.

如果试验的内容真是单一的,那么,在这种试验下两事件独立是较少出现的例外.因为,两个事件既然都依赖同一批结果,彼此谅必会有影响.掷两个均匀骰子,以 A_i 记"点数和为 i 的倍数",$i=2,3,5$.通过用(3.7)验证可知,A_2 与 A_3 独立,但这非一般性质,比如,A_2 与 A_5 就不独立.对这种"单一"性试验,(3.7)作为验证独立性的工具,还是有用的.有时,未经周到考虑的直观也可能引入歧途.

例 3.2 再考虑例 3.1,记 $B=\{$至少有一个骰子掷出 1$\}$,而把事件 A 定义为 $A=\{$三个骰子掷出的点数中至少有两个一样(即不全相异)$\}$,问 A,B 是否独立?

初一看使人的倾向于相信 A,B 独立,理由如下:知道 B 发生,即知道掷出的点中有 1,对 A 而言,似与知道掷出的点中有 2(或 3,4,5,6 都可以)一样.故 1 这个数并不相对地更有利于或更不利于 A 发生.经过计算发现不然:A,B 并不独立.这一点看来有些难理解,但是,如按下述分析,则可以信服:考虑$\overline{B}$.若 $\overline{B}$ 发生,则三个骰子都不出么.这样,它们都只有 5 种可能性(2,3,4,5,6),比不知$\overline{B}$发生时可能取的点数 1,2,3,4,5,6 少了一个,在 5 个数中拿 3 个(每个可重复拿),其有两个一样的可能性,自应比在 6 个数中拿 3 个时,有两个一样的可能性要大些.这个分析指出应有 $P(A)<P(A|\overline{B})$,由此推出 $P(A)>P(A|B)$(见习题 15),A,B 不独立.

多个事件独立性的定义,就是两个事件情况的直接推广.

定义 3.3 设 $A_1,A_2,\cdots$为有限或无限个事件.如果从其中任

意取出有限个 $A_{i_1},A_{i_2},\cdots,A_{i_m}$ 都成立.

$$P(A_{i_1}A_{i_2}\cdots A_{i_m}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_m}) \tag{3.8}$$

则称事件 $A_1,A_2,\cdots$ 相互独立或简称独立.

这个定义与由条件概率出发的定义是等价的,后者是说:对任何互不相同的 $i_1,i_2,\cdots,i_m$,有

$$P(A_{i_1}\mid A_{i_2}\cdots A_{i_m}) = P(A_{i_1}) \tag{3.9}$$

即任意事件 A_{i_1} 发生的可能性大小,不受其他事件发生的影响.这更接近于独立性的原义.但是,(3.9)的左边依赖于 $P(A_{i_2}\cdots A_{i_m})>0$,否则无意义,而(3.8)就没有这个问题.另外,定理 3.2 后面说的那段话当然也适用于多个事件的情形:多个事件的独立性往往产生于由多个试验构成的复合试验中,每个事件只与其中一个试验有关.

由独立性定义立即得出下面的概率乘法定理:

定理 3.3 若干个独立事件 $A_1,\cdots,A_n$ 之积的概率,等于各事件概率的乘积:

$$P(A_1\cdots A_n) = P(A_1)\cdots P(A_n) \tag{3.10}$$

乘法定理的作用与加法定理一样:把复杂事件的概率的计算归结为更简单的事件概率的计算,这当然要有条件:相加是互斥,相乘是独立.

由独立性定义可得到下面两条重要推论.

系 3.2 独立事件的任一部分也独立.例如,A,B,C,D 四事件相互独立,则 A,C,或 A,B,D 等,都是独立的.

这一点由独立性定义直接推出.更进一步可推广为:由独立事件决定的事件也独立.举例来说,若事件 $A_1,\cdots,A_6$ 相互独立,则以下三事件

$$B_1 = A_1 + A_2, B_2 = A_3 - A_4, A_3 = A_5A_6 \tag{3.11}$$

也独立.这在直观上很显然,但证明起来很麻烦,因为可以产生的事件很多.在下一章中我们将指出另外的考虑方法(见第二章例3.7).

如果把 B_3 改为 $A_4A_5A_6$，则 B_2, B_3，就不一定独立了. 理由也很明显：二者都与 A_4 有关，因而彼此也就有了关系.

系 3.3 若一列事件 $A_1, A_2, \cdots$ 相互独立，则将其中任一部分改为对立事件时，所得事件列仍为相互独立.

例如，若 A_1, A_2, A_3 相互独立，则 $\overline{A}_1, A_2, A_3$，或 $\overline{A}_1, A_2, \overline{A}_3$，或 $\overline{A}_1, \overline{A}_2, \overline{A}_3$ 等，都是互相独立的.

这一点从直观上也很显然，且对两个事件的情况，已在 27 页的足注中作过证明. 让我们再看一个三个事件的例子. 比如，要证 $\overline{A}_1, A_2, \overline{A}_3$ 独立，要对其验证(3.8)，其中有 $P(\overline{A}_1 A_2 \overline{A}_3) = P(\overline{A}_1)P(A_2)P(\overline{A}_3)$. 为此注意

$$A_2 \overline{A}_3 = A_1 A_2 \overline{A}_3 + \overline{A}_1 A_2 \overline{A}_3$$

且右边两事件互斥，故

$$\begin{aligned} P(\overline{A}_1 A_2 \overline{A}_3) &= P(A_2 \overline{A}_3) - P(A_1 A_2 \overline{A}_3) \\ &= P(A_2) P(\overline{A}_3) - P(A_1 A_2 \overline{A}_3) \end{aligned} \quad (3.12)$$

再利用 $A_1 A_2 = A_1 A_2 A_3 + A_1 A_2 \overline{A}_3$，得

$$\begin{aligned} P(A_1 A_2 \overline{A}_3) &= P(A_1 A_2) - P(A_1 A_2 A_3) \\ &= P(A_1)P(A_2) - P(A_1)P(A_2)P(A_3) \\ &= P(A_1)P(A_2)(1 - P(A_3)) \\ &= P(A_1)P(A_2)P(\overline{A}_3) \end{aligned}$$

以此代入(3.12)，得

$$\begin{aligned} P(\overline{A}_1 A_2 \overline{A}_3) &= P(A_2)P(\overline{A}_3) - P(A_1)P(A_2)P(\overline{A}_3) \\ &= (1 - P(A_1))P(A_2)P(\overline{A}_3) \\ &= P(\overline{A}_1)P(A_2)P(\overline{A}_3) \end{aligned}$$

明所欲证. 可以看出：当涉及众多的事件时，这么处理会很冗长，但并无任何实质困难(可使用数学归纳法，对所含对立事件个数进行归纳).

除了相互独立之外，还有所谓"两两独立"的概念. 一些事件 $A_1, A_2, \cdots$，如果其中任意两个都独立，则称它们两两独立. 由相互

独立必推出两两独立，反过来不一定对．从数学上，这无非是说：由(3.8)对 $m=2$ 及任何 $i_1\neq i_2$ 成立，不必能推出该式当 $m>2$ 时也成立．下面是一个简单的例子：

例 3.3 有四个大小质地一样的球，分别在其上写上数字 1，2，3 和“1，2，3”，即第 4 个球上 1，2，3 这三个数字都有．引进三个事件：

$$A_i = \{\text{随机抽出一球，球上有数字 } i\}, i = 1,2,3$$

所谓随机抽出一球，即每球被抽出的概率都是 1/4．易见 $P(A_1)=P(A_2)=P(A_3)=1/2$．因为，为使事件 A_1 发生，必须抽出第一球或第四球，有 2 种可能．又 $P(A_1A_2)=P(A_1A_3)=P(A_2A_3)=1/4$．因为，要 A_1,A_2 同时发生（抽出的球上既有 1 又有 2），必须抽出第四球．这样，对任一对事件 A_i,A_j，都有 $1/4=P(A_iA_j)=P(A_i)P(A_j)$，而 A_1,A_2,A_3 为两两独立．

但 A_1,A_2,A_3 不是相互独立．因为，易见 $P(A_1A_2A_3)$ 也是 1/4，而 $P(A_1)P(A_2)P(A_3)$ 为 1/8，二者不相等．

在现实生活中，难于想像两两独立而不相互独立的情况．可以这样想：独立性毕竟是一个数学概念，是现实世界中通常理解的那种“独立性”的一种数学抽象，它难免会有些不尽人意的地方．

独立性的概念在概率论中极端重要．较早期（比方说，到上世纪 30 年代止）的概率论发展中，它占据了中心地位．时至今日，有不少非独立的理论发展了起来，但其完善的程度仍不够．而且，独立性的理论和方法也是研究非独立模型的基础和工具．在实用上，确有许多事件其相依性很小，在误差容许的范围内，它们可视为独立的，而方便于问题的解决．

利用本节中引进的事件运算，独立性概念，加法乘法定理，可计算一些较复杂事件的概率．举几个例子．

例 3.4 考虑本节开始处提到的那个“打飞机”的例子．按所作规定，“飞机被击落”这事件 E 可表为

$$E = E_0 + E_1E_2$$

设 E_0,E_1,E_2 三事件独立．这假定从实际角度看还算合理．记 E_0，

E_1, E_2 的概率分别为 p_0, p_1, p_2. 为算 E 的概率 $P(E)$,不能直接用加法定理,因 E_0 与 E_1E_2 并非互斥,考虑 $\overline{E}$,易见 $\overline{E} = \overline{E_0}\ \overline{E_1E_2}$. 因 E_0, E_1, E_2 独立,按系 3.2 后面指出的,$\overline{E_0}$ 和 $\overline{E_1E_2}$ 独立,故

$$P(\overline{E}) = P(\overline{E_0})P(\overline{E_1E_2})$$

有 $P(\overline{E_0}) = 1 - P(E_0) = 1 - p_0$, $P(\overline{E_1E_2}) = 1 - P(E_1E_2) = 1 - P(E_1)P(E_2) = 1 - p_1p_2$. 代入上式得 $P(\overline{E}) = (1 - p_0)(1 - p_1p_2)$,而

$$\begin{aligned} P(E) &= 1 - P(\overline{E}) = 1 - (1 - p_0)(1 - p_1p_2) \\ &= p_0 + p_1p_2 - p_0p_1p_2 \end{aligned}$$

例 3.5 甲、乙二人下象棋,每局甲胜的概率为 a,乙胜的概率为 b,为简化问题,设没有和局的情况,这意味着 $a + b = 1$.

设想甲的棋艺高于乙,即 $a > b$. 考虑到这一点,他们商定最终胜负的规则如下:到什么时候为止甲连胜了三局而在此之前乙从未连胜二局,则甲胜. 反之,若到什么时候为止乙连胜了二局而在此之前甲从未连胜三局,则乙胜. 现要求“甲最终取胜”这事件 A 的概率 $P(A)$,及“乙最终取胜”这事件 B 的概率 $P(B)$.

为方便计,分别以 E 和 F 表甲、乙在特定的一局取胜的事件,有 $P(E) = a$, $P(F) = b$,现考虑“甲取胜”的事件 A,分两种情况.

1. 第一局甲胜而最终甲胜了.

这一情况又可分解为许多子情况:对 $n = 0, 1, 2\cdots$,甲经过 n 个“阶段”后才取胜,每个阶段是 EF 或 EEF,然后接着来一个 EEE. 例如,甲经过 4 个阶段后获胜的一种可能实战结果为

$$\underset{\sim\sim\sim}{EEF}\ \underline{EF}\ \underset{\sim\sim\sim}{EEF}\ EEE$$

即共下了 11 局甲才获胜,其中第 1,2,4,6,7,9,10,11 局甲胜,其余乙胜.

每个阶段不是 EF 就是 EEF,这两种情况互斥,又由独立性,知每个阶段概率为 $ab + aab = ab(1 + a)$. 再由独立性,知“经 n 阶段后甲获胜”的概率,为 $[ab(1 + a)]^n a^3$. n 可以为 $0, 1, 2, \cdots$,不同的 n 互斥. 于是这部分概率总和为

$$p = a^3 \sum_{n=0}^{\infty} [ab(1+a)]^n = a^3 / [1 - ab(1+a)]$$

2. 第一局乙胜而最终甲胜了.

既然第一局为 F 而最终甲胜,第二局必须是 E,故从第二局作起点看.我们回到了情况 1,从而这部分的概率为 bp(请读者注意,这里事实上已用了概率的乘法定理:P(第一局乙胜且最终甲胜)$=P$(第一局乙胜)P(第二局甲胜且最终甲胜),第一项为 b 而后一项为 p.总合两个情况(它们互斥),用加法定理,得

$$P(A) = a^3(1+b)/[1 - ab(1+a)] \tag{3.13}$$

直观上我们觉得,这个竞赛无限期拖下去分不出胜负是不可能的,这意味着 $P(B)=1-P(A)$.可是,上述直观看法仍须证明,不如直接算.方法与算 $P(A)$一样,但须分三种情况:①第一局乙胜.②第一局甲胜,第二局乙胜.③前两局甲胜,我们把具体计算留给读者(习题 16).结果为

$$P(B) = (1 + \overline{a} + a^2)b^2/[1 - ab(1+a)] \tag{3.14}$$

由于 $a+b=1$,极易验证 $P(A)+P(B)=1$.

这个例子值得细心品味.第一,它提供了一个涉及到无限个事件的情况(在甲最终取胜前可以经过任意多的"阶段"),以及在无穷个事件时使用加法定理(3.1).第二,本例告诉我们,在面对一个复杂事件时,主要的方法是冷静地分析以设法把它分拆成一些互斥的简单情况.这里,必须细心确保互斥性又无遗漏,一着不慎,满盘皆非.

例 3.6 设一个居民区有 n 个人,设有一个邮局,开 c 个窗口,设每个窗口都办理所有业务.c 太小,经常排长队;c 太大又不经济.

现设在每一指定时刻,这 n 个人中每一个是否在邮局是独立的,每人在邮局的概率都是 p.设计要求:"在每一时刻每窗口排队人数(包括正在被服务的那个人)不超过 m"这个事件的概率,要不小于 a(例如,$a=0.80,0.90$ 或 0.95).问至少须设多少窗口?

把 n 个人编号为 $1,\cdots,n$,记事件

$E_i=\{$在指定时刻第 i 个人在邮局办事$\}$, $i=1,\cdots,n$ 则在指定时刻,邮局的具体情况可以用形如

$$E_1\overline{E}_2E_3E_4E_5\overline{E}_6\overline{E}_7E_8\cdots\overline{E}_i\cdots E_{n-1}\overline{E}_n \tag{3.15}$$

这种事件去描述之. 为了每个窗口排队人数都不超过 m,在上述序列中,不加"bar"的 E 的个数,至多只能是 cm. 现固定一个 $k\leqslant cm$,来求"在(3.15)中恰有 k 个不加 bar 的 E"这事件 B_k 的概率. 由独立性以及 $P(E_i)=p$, $P(\overline{E}_i)=1-p$,知每个像(3.15)那样的序列且不加 bar 的 E 恰有 k 个时,概率为 $p^k(1-p)^{n-k}$. 但 k 个不加 bar 的位置,可以是 n 个位置中的任何 k 个. 因此,一共有 $\binom{n}{k}$ 个形如(3.15)的序列,其中不加 bar 的 E 恰有 k 个,这样得到 $P(B_k)=\binom{n}{k}p^k(1-p)^{n-k}$. 由于 k 可以为 $0,1,\cdots,cm$,且不同的 k 对应的 B_k 互斥,故得

$$P(\text{每个窗口排队人数不超过 } m)=\sum_{k=0}^{cm}\binom{n}{k}p^k(1-p)^{n-k} \tag{3.16}$$

找一个最小的自然数 c,使上式不小于指定的 a,就是问题的答案.

这是一个有现实意义的例题. 在 n 较大时,可用更方便的近似方法确定 c,参见第三章例 4.1. 当然,实际问题比本例描述的要复杂得多,因为有一个每人服务时间长短的问题. 这时间长短并非固定而是随机的. 这类问题属于排队论,是运筹学的一个分支. 本例是运筹学与概率论有联系的一个例子.

1.3.8 全概率公式与贝叶斯公式

全概率公式

设 $B_1,B_2,\cdots$为有限或无限个事件,它们两两互斥且在每次试验中至少发生一个. 用式表之,即

$$B_iB_j=\varnothing(\text{不可能事件}),\text{当 } i\neq j$$

$$B_1 + B_2 + \cdots = \Omega(\text{必然事件})$$

有时把具有这些性质的一组事件称为一个“完备事件群”.注意,任一事件 B 及其对立事件组成一个完备事件群.

现考虑任一事件 A.因 Ω 为必然事件,有 $A = A\Omega = AB_1 + AB_2 + \cdots$.因 $B_1, B_2, \cdots$两两互斥,显然 $AB_1, AB_2, \cdots$也两两互斥.故依加法定理 3.1,有

$$P(A) = P(AB_1) + P(AB_2) + \cdots \tag{3.17}$$

再由条件概率的定义,有 $P(AB_i) = P(B_i)P(A|B_i)$.代入上式得

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots \tag{3.18}$$

公式(3.18)就称为“全概率公式”.这名称的来由,从公式(3.17)和(3.18)可以悟出:“全部”概率 $P(A)$被分解成了许多部分之和.它的理论和实用意义在于:在较复杂的情况下直接算 $P(A)$不易,但 A 总是随某个 B_i 伴出,适当去构造这一组 B_i 往往可以简化计算.这种思想应用的一个实例是例 3.5 中算“乙最终获胜”这事件 A 的概率.我们在该例中已指出:A 必伴随以下三种互斥情况之一而发生:乙;甲乙;甲甲.只是该例的特殊性使我们可只用加法定理而不必求助于全概率公式.

这公式还可以从另一个角度去理解.把 B_i 看作为导致事件 A 发生的一种可能途径.对不同途径,A 发生的概率即条件概率 $P(A|B)$各各不同,而采取哪个途径却是随机的.直观上易理解:在这种机制下,A 的综合概率 $P(A)$应在最小的 $P(A|B_i)$和最大的 $P(A|B_i)$之间,它也不一定是所有 $P(A|B)$的算术平均,因为各途径被使用的机会 $P(B_i)$各各不同,正确的答案如所预期,应是诸 $P(A|B_i)$, $i=1,2,\cdots$,以 $P(B_i)$, $i=1,2,\cdots$为权的加权平均值.一个形象的例子如下:某中学有若干个毕业班,各班升学率不同.其总升学率,是各班升学率的加权平均,其权与各班学生数成比例.又如若干工厂生产同一产品,其废品率各各不同.若将各厂产品汇总,则总废品率为各厂废品率之加权平均,其权与各厂产量

成比例. 再举一个例.

例3.7 设一个家庭有 k 个小孩的概率为 p_k, $k=0,1,2,\cdots$, 又设各小孩的性别独立. 且生男、女孩的概率各为 1/2. 试求事件 $A=\{$家庭中所有小孩为同一性别$\}$的概率.

引进事件 $B_k=\{$家庭中有 k 个小孩$\}$, 则 $B_0, B_1, \cdots$ 构成完备事件群, $P(B_k)=p_k$, 现考虑 $P(A|B_k)$. 约定当 $k=0$ 时其值为 1. 若 $k\geqslant 1$, 则 k 个小孩性别全同有两种可能: 全为男孩, 概率 $(1/2)^k$; 全为女孩, 概率也是 $(1/2)^k$. 因

$$P(A \mid B_k) = 2(1/2)^k = 1/2^{k-1}, k \geqslant 1$$

由此, 用全概率公式, 得出

$$P(A) = p_0 + \sum_{k=1}^{\infty} p_k/2^{k-1}$$

贝叶斯公式

在全概率公式的假定之下, 有

$$\begin{aligned} P(B|A) &= P(AB_i)/P(A) \\ &= P(B_i)P(A|B_i)/\sum_j P(B_j)P(A|B_j) \qquad (3.19) \end{aligned}$$

这个公式就叫做贝叶斯公式, 是概率论中的一个著名的公式. 这个公式首先出现在英国学者 T. 贝叶斯(1702～1761)去世后的 1763 年的一项著作中.

从形式推导上看, 这个公式平淡无奇, 它不过是条件概率定义与全概率公式的简单推论. 其所以著名, 在其现实以至哲理意义的解释上: 先看 $P(B_1), P(B_2), \cdots$, 它是在没有进一步的信息(不知事件 A 是否发生)的情况下, 人们对诸事件 $B_1, B_2, \cdots$ 发生可能性大小的认识. 现在有了新的信息(知道 A 发生), 人们对 $B_1, B_2, \cdots$ 发生可能性大小有了新的估价. 这种情况在日常生活中也是屡见不鲜的: 原以为不甚可能的一种情况, 可以因某种事件的发生而变得甚为可能, 或者相反. 贝叶斯公式从数量上刻画了这种变化.

如果我们把事件 A 看成"结果", 把诸事件 $B_1, B_2, \cdots$ 看成导致这结果的可能的"原因", 则可以形象地把全概率公式看作成为

"由原因推结果";而贝叶斯公式则恰好相反,其作用在于"由结果推原因":现在有一个"结果"A 已发生了,在众多可能的"原因"中,到底是哪一个导致了这结果? 这是一个在日常生活和科学技术中常要问到的问题.贝叶斯公式说,各原因可能性大小与 $P(B_i|A)$成比例.例如,某地区发生了一起刑事案件,按平日掌握的资料,嫌疑人有张三、李四……等人,在不知道案情细节(事件 A)之前,人们对上述诸人作案的可能性有个估计(相当于 $P(B_1)$, $P(B_2)\cdots$),那是基于他们过去在局子里的记录.但在知道案情细节以后,这个估计就有了变化,比方说,原来以为不甚可能的张三,现在成了重点嫌疑人.

由以上的讨论也不难看出此公式在统计上的作用.在统计学中,是依靠收集的数据(相当于此处的事件 A)去寻找所感兴趣的问题的答案.这是一个"由结果找原因"性质的过程,故而贝叶斯公式有用武之地.事实上,依据这个公式的思想发展了一整套统计推断方法,叫做"贝叶斯统计".在本书后面的章节中将论及贝叶斯统计中的某些方法.

下述简单例子可能有助于理解上述论点.

例 3.8 有三个盒子 C_1, C_2, C_3,各有 100 个球,其中 C_1 盒含白球 80 个,红球 10 个,黑球 10 个;C_2 为白 10、红 80、黑 10;C_3 为白 10,红 10,黑 80.现从这三盒中随机地抽出一个(每盒被抽的概率为 1/3),然后从所抽出的盒中随机抽出一个球(每球被抽的概率为 0.01),结果抽出者为白球.问"该白球是从 C_i 盒中抽出"的可能性有多大? $i=1,2,3$.

记 $B_i=\{$抽出的为 C_i 盒$\}$, $i=1,2,3$; $A=\{$抽出白球$\}$,要求的是条件概率 $P(B_i|A)$.按假定有

$$P(B_1)=P(B_2)=P(B_3)=1/3$$

$$P(A|B_1)=0.8, P(A|B_2)=0.1, P(A|B_3)=0.1$$

代入(3.18),算出

$$P(B_1|A)=0.8, P(B_2|A)=0.1, P(B_3|A)=0.1$$

因为 C_1 盒所含白球最多,故在已知抽出白球的情况下,该球

系来自 C_1 盒的可能性也最大,理所当然.可能仍有读者不完全了然于心,则可以设想这么一个试验:准备两张纸,把例中的试验一次又一次的做下去:每抽出一个盒,在左边的纸上记下其为 C_1 或 C_2 或 C_3(不管从该盒中抽出的球如何),而只有在抽出的球为白球时,才在右边纸上记下该盒为 C_1 或 C_2、C_3.在进行了极大量次数试验后,会发现左边纸上 C_1 的比例很接近 1/3,而在右边纸上 C_1 的比例则很接近 0.8.

例 3.9 设某种病菌在人口中的带菌率为 0.03.当检查时,由于技术及操作之不完善以及种种特殊原因,使带菌者未必检出阳性反应而不带菌者也可能呈阳性反应.假定

$$P(\text{阳性}\mid\text{带菌})=0.99,P(\text{阳性}\mid\text{带菌})=0.01$$

$$P(\text{阳性}\mid\text{不带菌})=0.05,P(\text{阴性}\mid\text{不带菌})=0.95$$

现设某人检出阳性,问"他带菌"的概率是多少?

此问题相当于 $P(B_1)=0.03,P(B_2)=0.97$,且

$$P(A\mid B_1)=0.99,P(A\mid B_2)=0.05$$

所求的概率为 $P(B_1\mid A)$.按公式(3.18)算出

$$(0.03)(0.99)/[(0.03)(0.99)+(0.97)(0.05)]=0.380$$

就是说,即使你检出阳性,尚可不必过早下结论你一定带菌了,实际上这种可能性尚不到百分之四十.

这个例子很值得玩味,且对其"思维定势"中无概率成分的人来说,简直有点难以置信.说穿了,理由简单之极.由于带菌率极低,在全人口中绝大部分不带菌.由于检验方法之不完善,在这大批人中会检出许多呈阳性者.另一方面,带菌者在全人口中很少,即使全检出呈阳性,在这两部分呈阳性者的总和中也只占相对较小的一部分,而大部分属于"虚报"性质.这个例子说明,提高精确度在这类检验中极为重要.

一个不懂概率的人可能会这样推理:由于不带菌时检出阳性的机会才 0.05.我现在呈阳性,说明我有 $1-0.05=0.95$ 的机会带菌.实际不然.大而言之,概率思维是人们正确观察事物而必备的文化修养,这样说也许并不过分!

习　　题

1. 有 5 个事件，$A_1,\cdots,A_5$. 用它们表示以下的事件：

(a) $B_1=\{A_1,\cdots,A_5$ 中至多发生 2 个$\}$

(b) $B_2=\{A_1,\cdots,A_5$ 中至少发生 2 个$\}$

2. 证明：若 A,B 为两事件，则

(a) $A+B=A+(B-A)$，右边两事件互斥；

(b) $A+B=(A-B)+(B-A)+AB$，右边三事件互斥.

3. $(A+B)-(A-B)=?$

4. 把 n 个任意事件 $A_1,\cdots,A_n$ 之和表为 n 个互斥事件之和.

5. 通过把 $A+B+C$ 表为适当的互斥事件之和，以证明

$$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(BC)\\-P(CA)+P(ABC)$$

6. 有没有可能两件事 A,B 又互斥又独立？

7. $P(A-B)=P(A)-P(B)$ 是否必成立？何时成立？

8. 记 $C=\prod_{i=1}^{m}A_i+\prod_{j=1}^{n}B_j$，通过 A_i,B_j 及其对立事件表出 $\overline{C}$.

9. 如果把 $P(A|B)>P(A)$ 理解为"B 对 A 有促进作用"，则直观上似乎能设想如下的结论："由 $P(A|B)>P(A)$ 及 $P(B|C)>P(B)$ 推出 $P(A|C)>P(A)$"（意思是：B 促进 A，C 促进 B，故 C 应促进 A）. 举一简例证明上述直观看法不对.

10. 证明：若 A,C 独立，B,C 也独立，又 A,B 互斥，则 $A+B$ 与 C 独立.

更一般地，若 A,C 独立，B,C 独立，AB,C 也独立，则 $A+B$ 与 C 独立. 说明：上一结论是本结论的特例.

11. （接上题）若除了"A,C 独立，B,C 独立"之外，别无其他条件，则推不出 $A+B$ 与 C 独立，试举一反例以说明之.

12. 若 A,C 独立，B,C 独立，$A+B,C$ 也独立，则 AB 与 C 独立. 但若去掉"$A+B,C$ 也独立"的条件，则结论不再成立. 举一反例以说明之.

13. 办一件事件有 6 个关节，必须：①第 1 个关节要走通，②第 2,3 关节至少通一个，③第 4,5,6 关节至少通 2 个，事情才能办成.

(a) 设置必须的事件，以表出"事情办成"这个事件.

(b) 若各关节独立且每关节走通的机会为 2/3. 求事情能办成的概率.

14. 由 $P(A|B)>P(A)$ 推出 $P(B\ A)>P(B)$. 直观上怎样解释这个事实.

你认为,由 $P(A|B)>P(A)$, $P(A|C)>P(A)$,能否推出 $P(A|BC)>P(A)$? 若认为能,请证明之,若认为不能,请举出反例.

15. 由 $P(A)>P(A|B)$ 推出 $P(A)<P(A|\overline{B})$. 指出一种可能的直观解释.

16. 设 $A_1,A_2,\cdots,A_n$ 独立,而 $B_i=A_i$ 或 $\overline{A}_i$(不同的 i 可以不一样,例如,$B_1=A_1$, $B_2=\overline{A}_2$,等等),$i=1,\cdots,n$. 试用归纳法证明:$B_1,\cdots,B_n$ 也独立.

17. 一个秘书打好 4 封信和相应的 4 个信封. 但她将这 4 封信随机地放入这 4 个信封中,问"每封信都放得不对位"这事件的概率是多少?

18. 一盒内有 8 张空白券,2 张奖券,有甲、乙、丙三人按这个次序和以下的规则,各从此盒中随机抽出一张. 规则如下:每人抽出后,所抽那张不放回但补入两张非同类券(即:如抽出奖券,则放回 2 张空白券,等等). 问甲、乙、丙中奖的概率各有多大?

19. 某作家的全集共 p 卷,现买来 n 套(共 np 本),随机地分成 n 堆,每堆 p 本,问"每堆都组成整套全集"这事件的概率为多少.

20. 在例 1.1 中,把胜负规则改为"谁先胜四局者为胜". 问在甲 2 胜 1 负的情况下中止赌博,应按怎样的比例瓜分赌本才算公平?

21. 把例 3.1 中的事件 B 的定义改为:$B=\{$至少有一个骰子掷出么点$\}$,求该例中事件 A 的条件概率 $P(A|B)$.

直观上看结果应相同,但算出的结果不同,如何解释?

22. 在例 2.3 中,把"排成一列"改为"排成一圆圈". 证明例中所说的事件 A 的概率为 $\binom{n}{m}\Big/\binom{n+m-1}{m}$.

23. 四人打桥牌,问:"至少有一方没有 A"及"至少有一方恰有两个 A"这两个事件的概率.

24. 有一个半径为 1 的圆周 C. 甲、乙二人各自独立地从圆周上随机地取一点,将两点连成一条弦 l,用几何概率的方法计算"圆心到 l 的距离不小于 1/2"这个事件的概率.

25. 把 8 个可以分辨的球随机地放入 7 个可以分辨的盒子中,问"其中有两个盒各得 2 球,一个盒得 3 球,一个盒得 1 球"这事件的概率是多少?

26. 设男女两性人口之比为 51:49. 又设男人色盲率为 2%,女人色盲率

为 0.25% .现随机抽到一个人为色盲,问"该人为男人"的概率是多少?

27. 设有 n 个独立事件 $A_1,\cdots,A_n$,其概率分别为 $p_1,\cdots,p_n$,记 $p=p_1+\cdots+p_n$. 设 $0<p_i<1, i=1,\cdots,n$. 证明:

(a) "$A_1,\cdots,A_n$ 都不发生"这个事件的概率小于 e^{-p}.

(b) "$A_1,\cdots,A_n$ 中至少发生 k 个"这事件的概率小于 $p^k/k!$.

28. 投掷 10 粒均匀骰子,记事件

$$A=\{至少有 2 粒骰子出么点\}$$

$$B=\{至少有 1 粒骰子出么点\}$$

求条件概率 $P(A|B)$.

这个题可不可以这样算:既然已知至少掷出一个么点,不妨(因各骰子地位对称)就设第一粒骰子掷出么点.因而所求的条件概率为:掷 9 粒骰子至少出现一个么的概率,即 $1-(5/6)^9$. 为什么?

29. 假定某种病菌在全人口的带菌率为 10% ,又在检测时,带菌者呈阳、阴性反应的概率为 0.95 和 0.05,而不带菌者呈阳、阴性反应的概率则为 0.01 和 0.99.今某人独立地检测三次,发现 2 次呈阳性反应,1 次呈阴性反应.问:"该人为带菌者"的概率是多少?

30. 甲、乙二人约定了这样一个赌博规则:有无穷个盒子,编号为 n 的盒子中,有 n 红球 1 白球,$n=1,2,\cdots$,然后甲拿一个均匀铜板掷到出现正面为止.若到这时甲掷了 n 次,则甲在编号为 n 的盒子中抽出一个球,如抽到白球算甲胜,否则乙胜.你认为这规则对谁更有利?