第3篇 信息纽带

在第 1 篇和第 2 篇中我们引入并讨论了数据的概念，并将它与物质和能量并列为构成人类世界的三个基本要素。同时，我们给出了关于数据的严格定义、普遍形式和度量方法，并且提出和讨论了数据所遵循的三个基本法则，分别对应和描述数据所具有的客观性、物理性和生物性。所谓“客观性”，是指可观测的世界是可以被数据化的，而数据化过程是由生物特别是人类所特有的自然数据功能以及所创造的人工数据系统来实现的。在生物世界出现之后，生物特别是人类作为自然的数据机器便开启了对世界进行数据化的过程。所谓“物理性”，是指数据的存在和运动离不开物理世界的物质与能量，并受到所遵循的物理规律的制约。同时，人类遵循这些物理规律，通过科技发明创造不断降低每个比特数据所需要的物质和能量，在全球范围内建立了强大的数据系统和技术生态。所谓“生物性”，是指数据是人类相互交流的媒介和认识世界的工具。通过对世界的高度数据化，人类开始建立一个数据无处不在、无时不在的数字虚拟世界。这个世界既是客观物理和生物世界的反映，更是人类主观精神世界的扩展和升华。数字世界给人类带来了更大的生存空间和更多的生命自由，也给物理和生物世界带来了新的问题与挑战。正如农业革命和工业革命引发了人类对地球和宇宙物质和能量资源的开发和利用，当前正在爆发的信息革命开启了对数据的开发与利用的新纪元，推动和代表人类发展的一个崭新的阶段。

3.1　信息的定义

数据是人类和其他智能物体与客观世界相互作用过程中产生的一种原始符号，它是事物特性与变化的反映、记录和展现。数据自诞生之日便开始脱离所反映的客体独立在世界上存在和运动，并经历一系列的变换和改变，导致它与其反映物之间的关系变得错综复杂、扑朔迷离，从而产生了多样性、动态性和不确定性。数据作为一种人类和智能物体相互交流合作的媒介以及认识改造世界的工具，在被生物和人工数据处理系统存储、传输和处理之后，不仅形式结构会发生变化，特定含义和预期效用也会产生变异。即使数据的形式结构没有发生变化，信息接收者对其特定含义和预期效用也可能产生歧义。所谓数据的“形式结构”，是指符号的表现方式与规则，如中文和英文具有不同的语言的单词、语句的结构规则等，在语言学中统称为语法(grammar)。所谓“特定含义”，是指数据符号与所代表事物的关系，如“老鼠”和“MOUSE”均代表一种哺乳纲啮齿目的动物，但这种关系不一定是唯一确定的，如英文的 MOUSE 还可以指计算机等设备上使用的鼠标。同一单词在不同的上下文和情境下也可能有不同的含义。在语言学中符号的含义叫语意(semantics)。关于“效用”的概念比较模糊和抽象，一般是指数据的形式和含义之外的东西，或者是数据所产生的效果、影响或价值等，在语言学中称为语用(pragmatics)。总之，数据对于其接收和使用者(自然或人工数据系统)来说，最基本的特性就是它的“不确定性”或“随机性”。而另一方面，作为媒介和工具的数据必须具有更高级的功能(如结构、含义和效用等)才能够长期存在和不断发展。正是这些具有结构、含义和效用的数据才构成了我们所说的信息。从这个意义上，数据中的信息才是真正连接世界与人类以及

人类之间的纽带。

信息(information)也许是目前社会上最流行的“热词”,但关于信息定义却众说纷纭。维基百科关于信息的条目是:“信息,又称情报,是一个严谨的科学术语,其定义不统一,是由它的极端复杂性决定的”。百度百科的解释是:“信息,指音讯、消息、通信系统传输和处理的对象,泛指人类社会传播的一切内容。人通过获得、识别自然界和社会的不同信息来区别不同事物,得以认识和改造世界。在一切通信和控制系统中,信息是一种普遍联系的形式”。显然,这些似是而非、含糊其辞的说法对我们信息领域的专业人士来讲,没有多少实际的指导意义。在阅读和研究了目前学术界关于信息的各种不同定义之后,我们采用以下对信息的定义:

信息是具有一定结构形式、特定含义和预期效用的数据。

根据这个定义,信息一定是数据,但数据却不一定是信息,两者不能混为一谈。数据升华到信息需要具有一定的结构形式、特定含义和预期效用,如图 3-1 所示。客观世界的事物通过与数据系统的相互作用产生了相对应的符号即数据。当这些数据与自然或人工系统发生作用时,需要确定它所具有的结构形式、特定含义和预期效用。因为信息的结构形式、特定含义和预期效用均是相对于信息观察者的主观世界来判断与衡量的,所以存在一定的不确定性。换句话讲,这些代表信息的数据是随机的。只有那些对观察者而言在形式、含义和效用方面确定的数据才成为信息。从这个意义上讲,数据中信息才是联系客观世界和主观世界之间的纽带。

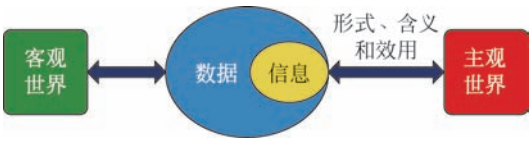


图 3-1　信息的定义

为了更好地说明数据与信息的区别与联系,观察和讨论以下几个例子。

图 3-2 给出一组“无规则随机”二进制数字序列。因为没有给出关于这些数据结构形式、特定含义和预期效用的编码规则,即使我们耗尽精力研究和猜测这些数字所呈现的结构规律,也可能还是毫无头绪。如果想再进一步搞清这些数字所代表的含义和所预期的效用,即它们到底反映、记录和展现了客观世界中什么事物和期望达到什么效用等,则更是难上加难。总之,这些数据对一般的观察者来说太不确定,很难转化为真正意义上的信息。

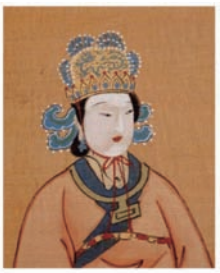
图 3-2　无规则的随机数字

第二个例子是一句话(图 3-3):“天亮已走,母病危,速转院!”。这组数据是一段中文叙述,语法结构清晰严谨,具有汉语文字知识的观察者均能够明白字面的意思。这句话所表达的意思是一个人天亮的时候已经离开,但母亲病危,需要马上转院。如果我们对此句话的历史背景和故事不了解,以上这种判断也许是合理的。实际上,这句话是当年担任国民党中统机要秘书的地下党员钱壮飞发给上海党组织的经过加密编码的暗语。根据这套编码规则,“天亮”即黎明,是中共叛徒顾顺章的化名;“已走”指的是叛变;“母病危”意思是中共地下党将面临巨大危险;“速转院”即立即转移。在熟悉中文但不清楚编码规则的情况下,不同的观察者对同样的数据含义的理解可能不同。只有了解对此数据的“编码”规则才能够得到其原始的含义。毫无疑问,当时收到此密文的上海地下党领导,不仅明白电文内容的含义,而且马上对此作出反应和行动。这些信息的作用和意义对发送者和接收者来讲生命攸关。正如周恩来后来所说:“要不是钱壮飞等同志,我们这些人是要死在国民党反动派手上的。”

图 3-3　钱壮飞发给上海地下党的密文

第三个例子请见图 3-4(a)的石碑。这是位于陕西省乾县唐乾陵的一块石质巨碑,是为武周皇帝武则天所立。与历代皇帝的纪念碑不同,武则天的碑却是一块“无字碑”,因最初碑上未刻一个字而得名。你若注视这样一位中国历史上绝无仅有的女皇帝所留下的没有任何碑文的丰碑,又会有何感想呢?虽然没有任何文字,但武则天和她所处时代的同人使用了什么形式的数据、试图传递什么含义和达到什么预期效用呢?目前至少有三种不同的学说试图“解码”武则天“无字碑”的含义和用意,但结果却是相互矛盾,难以判断真伪。武则天的“无字碑”的“无”,却给我们留下了无穷的想象和猜测,真可谓此处无字胜有字啊!

(a) 武则天的“无字碑”



(b) 武则天(624—705)

图 3-4　武则天的无字碑

最后一个例子如图 3-5(a)所示。这是人类基因中 DNA 序列的一个片段。对这组数据稍作观察,你就会发现它是由 4 个字母 T、C、A 和 G 组成的。如果你对 DNA 的

化学结构有一定了解，则知道它们分别代表组成 DNA 碱基对的 4 种核糖核酸。这些字母的某种组合需要按某种方式代表 20 种不同的氨基酸，但具体组合（即编码）的方式是怎样的呢？显然，用 1 或 2 个字母代表 1 种氨基酸只能有 4 和 $4^2=16$ 种组合，无法表示 20 种氨基酸。若用 3 个字母代表一种氨基酸则可以有 $4^3=64$ 种组合，完全可以涵盖 20 种氨基酸。我们称由这 4 个字母组成的三字母元素为密码子（Condon）。实验证明，这 64 种组合中，有 3 种组合并不对应任何氨基酸；余下 61 个密码子对应 20 种不同的氨基酸。这种对应关系由图 3-5(b)的密码子编码表表述，其中一种氨基酸可能由多个密码子来代表，这种现象称为密码子简并。可以看到密码子最高简并度为 6（对应 3 种氨基酸），其次为 4（对应 5 种氨基酸）、3（对应 1 种氨基酸和停止功能）和 2（对应 9 种氨基酸）。没有简并的密码子只有 3 个（对应 2 种氨基酸和开启功能）。这种安排虽然看起来有些浪费，但实际上可以带来一些好处，如避免编码出错等。美国生化学家马歇尔·尼伦伯格（Marshall Nirenberg，1927—2010）在 1961—1964 年破解了 DNA 中核糖核酸碱基对组合对应氨基酸的编码规则，并因此在 1968 年获得诺贝尔生理学或医学奖。与图 3-2 中完全无序和没有含义的数据不同，图 3-5 的数据不仅具有符合一定规则的结构，即由 4 个字母构成的 2 连体密码子，并且具有特定的含义，即每个密码子对应一种氨基酸或其他 DNA 功能（如“开启”或“停止”）。最后，根据生物学著名的“中心法则”，DNA 通过 RNA 将这些原先存储在 DNA 双螺旋长链上的基因密码经过“转录”和“翻译”最终生成对应由氨基酸构成的另外一个长链，即蛋白质。所以，信息的“效用”也是明确的。当然，并不是所有的人面对这组基因数据都能对它的结构、含义和效用有同样的观察、分析和结论。这是因为这必须借助信息观察者所具有的“先验知识”（prior knowledge）才可能得到。所以，即使我们同时注视这段数据，因我们所具有的背景知识不同，所得到的数据结构、含义和效用的结果（即信息）却可能有天壤之别。

DNA中的基因数据

This sequence represents the first exon of the Human Ras DNA sequence. an important gene in cell signalling and human disease:

ATGACGGAATATAAGCTGGTGGTGGTGGGCGCCGGCGGTGTGGGCAAGAGTGCGCTGACC
ATCCAGCTGATCCAGAACCATTTTGTGGACGAATACGACCCCACTATAGAGGATCCTACA-
CAGCTGGTGGTCGGTGTGGGACCCCAGGCAAGATGGACGAATACGACCCCACTATA
AGCTGATCCAGAACCATTTTGTGGAGATAAGCTGGTGGTGGTGGGCGCCGGGGGCAAGAT-
GGACGAATACGACCCTGGTGGGCGCCGGCGGTGTGGGACCCCACTATAGAGGATCCTACA
TGATCCAGAACCATTTTGTGGTGGTGGTGGTGGGCGCCGGCGGTGTCGGCGGTGTGGGAC-
CCCACTATCGGTGTGGGACCCCACGGTGTGGGACCCCACGGTGTGGGACCCCACGCCGGC
AGCTGATCCAGAACCATTTTGTGGAGATAAGCTTGGACGAATACGAACCATTTTGTGGACG
AATACCACGGTGTGGGACCCCACGGTGTGGGACCCCACGAGCTGATCCAGAACCATTTTGT
GGAGATAAGCTAGCTGATCCAGAACCATCCTGA

(a)

基因→氨基酸编码表

Amino acid	DNA codons					
Alanine	GCT	GCC	GCA	GCG		
Arginine	CGT	CGC	CGA	CGG	AGA	AGG
Asparagine	AAT	AAC				
Aspartic acid	GAT	GAC				
Cysteine	TGT	TGC				
Glutamic acid	GAA	GAG				
Glutamine	CAA	CAG				
Glycine	GGT	GGC	GGA	GGG		
Histidine	CAT	CAC				
Isoleucine	ATT	ATC	ATA			
Leucine	CTT	CTC	CTA	CTG	TTA	TTG
Lysine	AAA	AAG				
Methionine	ATG					
Phenylalanine	TTT	TTC				
Proline	CCT	CCC	CCA	CCG		
Serine	TCT	TCC	TCA	TCG	AGC	AGT
Threonine	ACT	ACC	ACA	ACG		
Tryptophan	TGG					
Tyrosine	TAT	TAC				
Valine	GTT	GTC	GTA	GTG		
Start(CI)	ATG					
Stop(CT)	TAA	TAG	TGA			

(b)

图 3-5　生物基因 DNA 序列的片段和密码子编码表

3.2 概率基础知识

为了更好地描述和解释数据中存在的不确定或随机现象，我们首先介绍描述事件发生机会或可能性的数学理论，即“概率论”的基础知识。法国数学家拉普拉斯（Pierre-Simon Laplace，1749—1827）曾经说过：“概率论只不过是把常识用数学公式表达了出来”。关于概率统计的系统理论，读者可以通过更加专业的教科书获得。需要指出的是，概率知识对于学习与掌握数据与智能科学的重要性怎样强调都不过分。

对于概率的完整定义，即严格意义上满足“必要且充分”条件的定义，目前学术界还没有一个普遍接受和满意的答案。作为一个数学函数，概率必须首先满足一些基本的必要条件。我们知道，概率函数 $P(A)$ 是对随机事件 A 发生可能性的度量，所以至少应该限制在 0～1，即 $0 \leqslant P \leqslant 1$。$P$ 越接近 1，说明该事件 A 越可能发生；P 越接近 0，则越不可能发生。另外，概率还必须满足可加性的条件。这些性质是关于概率的基本数学公理，也是我们对于概率理论的基本假设和必要条件。但这些基本数学公理却不是充分条件，并没有给出如何计算或确定概率的方法。

概率方法：

现实中，至少有两种定义和方法可以用来确定一个事件发生的概率。方法 A 是古典法，即对事件的样本进行分析，找出对事件发生有利的样本。假定所有有利样本发生的概率相同，则事件的概率等于有利样本数与总样本数之比。

$$P(A)=\frac{\text{对事件 } A \text{ 发生有利的样本数}}{\text{所有样本的总数}}$$

方法 B 是频率法，即对事件进行前后相互独立的 N 次随机试验，针对每次试验记录下相对频率值 A。

$$P(A)=\frac{\text{事件 } A \text{ 发生的次数}}{\text{所观察事件发生的总数}}$$

随着试验次数 N 的增加，相对频率值趋向某个极限值，这个极限值称为统计概率。很显然，用此方法所得到的概率值与所做的试验（或观察事件发生）的总数有关，一般需要有足够的试验观察才能得到一个可靠和精确的答案。

为了比较以上两类方法，下面我们讨论一个例子。

问题：掷两次相同的骰子，出现数字之和为 7 的概率是多少？

解法 A：古典概率方法

每个骰子有 6 个平面，每个平面有一个数字，分别为 1～6，所以有 6 种可能性。2 个骰子共有 $6\times6=36$ 种可能性。如图 3-6 所示，两次试验得到数字之和为 7 的次数为 6。所以古典概率方法得到的结果是 $6/36=16.7\%$。

解法 B：频率概率方法

若使用频率概率的方法，我们需要在尽量理想的条件下做掷骰子的试验并统计结果。试验的统计结果如图 3-7 所示，50 次试验中有 13 次出现了两个骰子数字之和为 7

的结果，所以统计概率为 13/50＝26％。

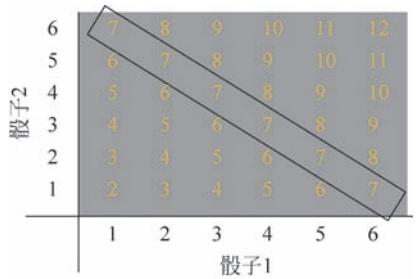


图 3-6　古典概率方法的例子

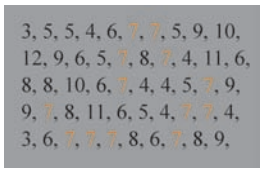


图 3-7　频率概率方法的例子

显然两种方法得到的结果是不同的。也许我们在频率方法中所做的试验次数不够多？若试验次数足够多，是否可以得到与古典概率相同的结果？对此我们无法证明，也不能做这样的假设。虽然我们不能确定频率法所得的概率结果是否能够收敛到古典概率，但却知道一个随机试验结果的平均值随着试验次数增加最终会收敛到古典概率所预测的平均值，这就是著名的“大数定理”或“大数法则”。

如每掷 6 次相同的骰子，最可能出现的平均数是多少？用古典概率方法计算，平均值为 $N_{古典}=(1+2+3+4+5+6)/6=3.5$，用统计的方法，大量实现的最终结果趋向于古典概率的结果（图 3-8）。

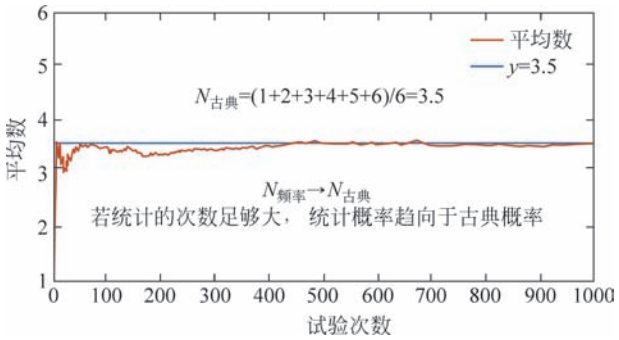
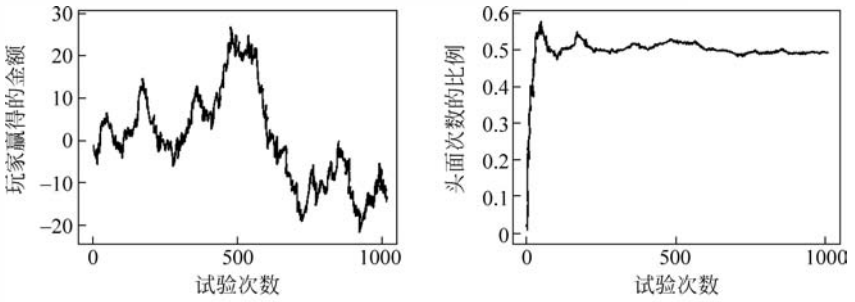


图 3-8　大数定理的例子

“大数法则”所揭示的规律是指当试验次数增加，古典概率和频率概率所得到的关于事件某些特性的平均值将趋于同一数值。有些人由此提出了所谓“平均定律”，声称某些事件一个极端情景的发生将伴随另一个极端情景的发生以保持事件发生的平均值与大数定律所预测的平均值相符。如一个人在赌场经历了一系列“坏运气”之后，一定会时来运转而获得“好运气”等。其实这个结论是不正确的。这是因为任何一次试验均是一次独立的偶然事件，对历史上所发生的事情并没有记忆或关联，对未来将要发生的事情也无法预测或影响。为了说明这一点，我们引用了一个掷硬币的模拟试验。游戏规则是玩家在头面（Head）出现时将获得 1 美元，而尾面（Tail）出现将失去 1 美元。1000 次试验的结果[图 3-9(a)]表明玩家的净收益（定义为赢得金额—输掉金额）波动较大，并不循序所

谓“平均定律”。但头面出现比例的平均数却趋于 50%[图 3-9(b)],完全符合“大数定律”。这说明我们不应将每次事件发生的偶然性与多次事件的统计平均的必然性混为一谈。概率所描述的事件发生的不确定性对于某一次或几次独立的事件是没有意义的,只有在同样的事件不断重复发生的情况下才真正有实际意义。即使小概率的事情,若做得太多,任何偶然也都会变为必然。这就是所谓“常在河边走,哪能不湿鞋”所揭示的道理。当然,这里的假设是所有事件均是相互独立的。在现实中,与赌场的“老虎机”不同,许多事件并不一定是独立的,所以前期事件中的“坏运气”可能为后期事件积累经验教训,从而改进“玩家”取胜的概率,带来“好运气”。



(a) 事件变量: 短期的偶然性　　(b) 事件平均: 长期的必然性

图 3-9　平均谬误和大数定律

古典概率是“先验”概率(a priori),即在没有做任何试验之前便已经知晓的概率。同时,它又是一个理想的情况,本身存在一些限制性的假设:①各个事件发生的概率相同;②各个事件的发生相互独立;③所有可能发生的事件总数已知且有限。这些限制条件在很大程度上制约了古典概率在实际中的应用。尽管如此,古典概率为我们分析和预测现实中的随机事件提供了一个简单有效的工具,可以对实际情况产生有用的洞见。我们可以将古典概率推广到更普遍的“先验概率”而不局限于等概率的假设。在这种情况下,先验概率就成为观察者的一种主观的“确信度”。这些确信度的基础可能是基于科学知识、实际经验、某种信仰甚至迷信等。而频率概率是“后验”概率(a posterior),需要通过一系列试验结果统计才能得到。这种方法不需要受等概率假设的限制,但仍需要假设各个试验相互独立且样本有限。现实中如何确定试验的条件(如试验次数等),以及对试验结果的合理(科学)解释和对未来事件的预测等是必须谨慎回答和解决的问题。

一个更为基本的问题是概率本身的客观性和主观性问题。关于概率客观性的观点认为,随机性是客观世界事物运动与变化规律的反映,与观察者无关,概率是对客观世界这种随机性的描述和预测。支持这种观点的论据是世界的许多事情发生的方式完全是随机的,特别是一些概率很小的“黑天鹅”事件,我们从根本上无法精确地预测这些事件的发生。关于概率主观性的观点则认为,随机性不是客观世界实际运动和变化规律的反映,而是观察者对客观世界缺乏信息和知识的表现,概率是观察者根据所具有的信息和知识等对事件发生的可能性(或随机性)的确信度。

严格意义上讲,以上两种观点均有一定道理。关于客观世界的运动和变化规律,我

们将在“知识升华”一篇中做更为详细和深入的讨论。假定现实中所有的事件本质上具有确定性，那么我们所讨论的随机性则是我们自身认识局限性或信息不对称所致。我们之所以对某些事件发生的可能性用概率的方法做预测，是因为我们对这些事情发生的条件和规律缺乏信息和知识。从这个意义上讲，概率本质上是主观的。概率的方法中，不同的观察者可能根据自身的知识、经验等主观因素对同一事件发生的可能性（或随机性）赋予不同的概率。即使对于同一观察者，随着对该事件发生的相关因素了解得更多，或者其他主观的因素驱使，也可能对这一事件发生的概率有不同的赋值。由此可以理解为什么数据和信息对于我们认识、理解和预测世界中事物的运动和变化具有极其关键的作用。

为了说明概率的“主观性”问题，我们以赌场中一种比较流行的游戏轮盘赌为例加以说明[图 3-10(a)]。美国的轮盘赌转盘上均匀分布着 0～36 以及 00，共 38 个数字。当转盘转起来后，会有一个小球在转盘内滚动，同时轮盘本身朝小球相反的方向转动。最终小球会落到某个数字对应的小槽里，从而产生“中奖数字”。游戏的赔率是 1∶35，即若赢了，1 元可以变成 36 元。因轮盘上一共有 38 个数字，赢钱概率只有 1/38，所以庄家相对于玩家有平均 5.26%的优势（注：赌场商业模式的核心机制之一）。所以，在游戏的结果完全随机的前提下，进行足够多次的赌博，最终的结果一定是庄家赢。

图 3-10　美国轮盘赌的故事

但小球和转盘的运动并不是随机的，而是严格遵循牛顿力学定律。只要小球的初始位置和速度已知并且小球运动过程中与其他物体相互作用的方式已知，理论上讲可以精确预测小球最终所达到的位置。这里的关键是如何在真实赌场情境下确定小球的初始条件和计算小球的轨迹。1961 年 MIT 的教授爱德华·索普（Edward Thorp，1932— ）和香农不仅开发了预测小球运动轨迹的数学模型和算法，还制造了世界上第一台“穿戴式”计算机[图 3-10(c)]。利用这台由 12 个晶体管和当时先进的电子元件构成的装置，他们在赌场实践中创造了大大超出庄家的成绩。当时由于计算和通信技术水平所限，他们设备的可靠性和实用性（特别是隐蔽性）均较差，所以索普后来只专注于另外一种不需要借助任何设备而只利用数学算法的游戏——21 点扑克牌（Black Jack），并进军华尔街证券金融市场，建立了全球第一个量化对冲基金，取得了巨大的成功。此后，他还出版了一本关于 21 点扑克牌的畅销书《击败庄家：21 点的有利策略》。预测轮盘赌结果的数学方法和穿戴技术在 20 世纪 70 年代初被当时正在加州大学攻读博士学位的多伊恩·法默

(J. Doyne Farmer,1952—)和同学诺尔曼·派克(Norman Packard,1954—)等改进,获得比赌场庄家多20%的取胜概率优势[图3-10(b)下图]。但由于硬件可靠性问题和对赌场暴力的恐惧,他们所具有的技术优势并没有在现实中兑现。对这些科学家来讲,轮盘赌上的小球不再是完全随机地运动;他们虽然还不能完全精确地预测小球的最终位置,却提高了预测的概率,减小了事件的不确定性。一旦优势超过庄家,他们便可稳操胜券。当然,做到这一点的前提是他们的技术必须安全可靠。同时,他们还必须成为"赌徒",因为概率的统计意义必须在大量试验中才能够体现。这些年轻时的恶作剧最终并没有持续多久。法默和派克后来在复杂性数学理论、经济学理论与应用的方面做出了独特的学术贡献,并且成功创办了一家高科技公司。法诺现在是英国牛津大学数学教授,而派克则在科技界连续创办了几家公司,成为一名职业企业家。

条件概率:

前面讨论的概率理论假定样本空间事件发生是相互完全独立的。在现实中,一个随机事件的发生,对另一个相关的事件发生的概率是会发生影响的。为此,我们引入"条件概率"的定义,即假设 A 和 B 是两个事件,在事件 B 已经发生的条件下,事件 A 发生的概率。基于古典概率的模型,若所有事件发生的概率相同,事件 A 的概率便是事件 A 的样本与总样本数之比,即 $P(A)=n(A)/n(\Omega)$,$n(A)$ 和 $n(\Omega)$ 分别为事件 A 的样本数和总样本数,也可用相对应的面积形象地表示(图3-11)。同样的推导也适合于事件 B 的概率。若事件 A 和 B 相关联,即两者的事件空间有交集,则事件 A 发生后,事件 B 发生的概率 $P(A|B)$ 则变为 A 和 B 同时发生的样本数 $P(AB)$ 除以事件 B 的样本数 $P(B)$,即

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

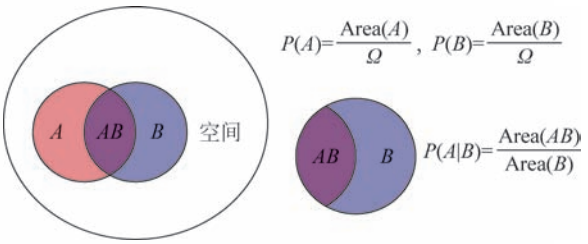


图3-11　条件概率公式的推导

关于条件概率有两个特例:①若事件 B 是必然事件,$P(B)=1$,则对事件 A 的概率没有贡献,条件概率 $P(A|B)=P(A)$;②若事件 B 不会发生,则 $P(B)=0$,则关于事件 B 的信息量无穷大,将无法确定事件 A 发生的概率。另外,观察 A 和 B 两个事件各自的条件概率,可以证明 $P(A|B)$ 与 $P(B|A)$ 一般是不相等的,即条件概率对于事件 A 和 B 是不对称的。

将条件概率的公式重新写成对称的形式,便得到了贝叶斯定理(Bayes Theorem),即事件 B 发生的概率乘以事件 B 发生条件下事件 A 发生的概率等于事件 A 发生的概率乘以事件 A 发生条件下事件 B 发生的概率。