

$$\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cdots = \{x \mid x \in S_n \text{ 对某个 } n \text{ 成立}\},$$

$$\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cdots = \{x \mid x \in S_n \text{ 对一切 } n \text{ 成立}\}.$$

有时候我们需要考虑几个甚至无穷个集合的并和交的问题. 例如, 如果每一个正整数 n 都确定一个集合 S_n , 则

$$\bigcup_{n=1}^{\infty} S_n = S_1 \cup S_2 \cdots = \{x \mid x \in S_n \text{ 对某个 } n \text{ 成立}\},$$

$$\bigcap_{n=1}^{\infty} S_n = S_1 \cap S_2 \cdots = \{x \mid x \in S_n \text{ 对一切 } n \text{ 成立}\}.$$

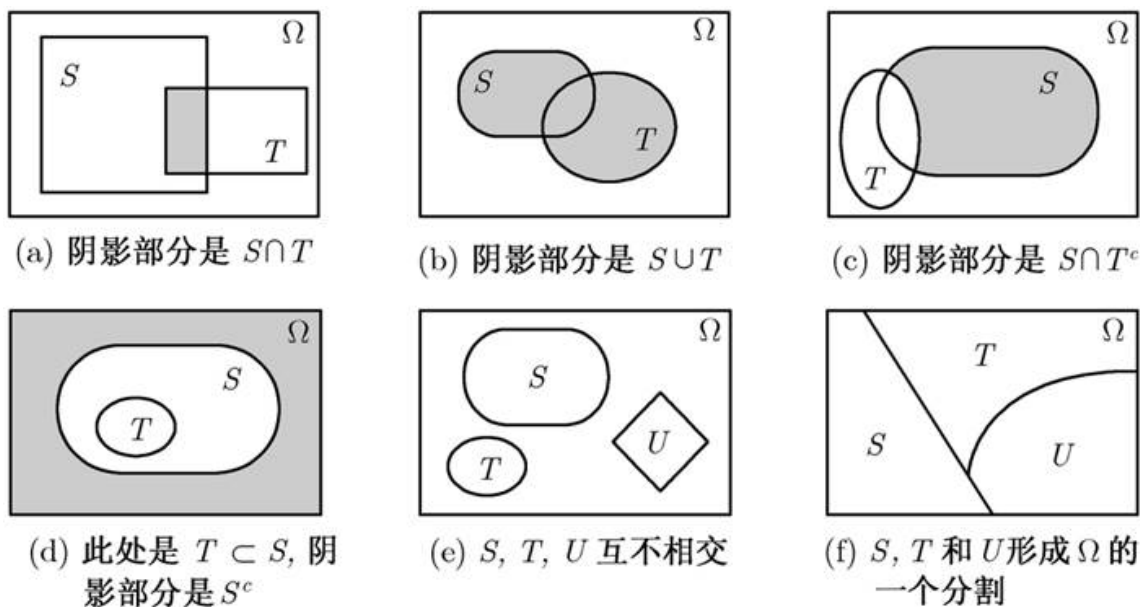


图 1.1 维恩图的例子

两个集合称为不相交的, 如果它们的交集为空集. 更一般地, 几个集合称为互不相交的, 如果任何两个集合没有公共元素. 一组集合称为集合 S 的分割, 如果这组集合中的集合互不相交, 并且它们的并为 S .

设 x 和 y 为两个研究对象, 我们用 (x, y) 表示 x 和 y 的有序对. 我们用 \mathbf{R} 表示实数集合, 用 \mathbf{R}^2 表示实数对的集合, 即二维平面, 用 \mathbf{R}^3 表示三维实数向量的集合 (三维空间). 集合及其运算可用维恩图形象化表示, 见图1.1.

1.1.2 集合的代数

集合运算具有若干性质, 这些运算性质可由运算的定义直接证得, 举例如下:

$$\begin{array}{ll} S \cup T = T \cup S, & S \cup (T \cup U) = (S \cup T) \cup U, \\ S \cap (T \cup U) = (S \cap T) \cup (S \cap U), & S \cup (T \cap U) = (S \cup T) \cap (S \cup U), \\ (S^c)^c = S, & S \cap S^c = \emptyset, \\ S \cup \Omega = \Omega, & S \cap \Omega = S. \end{array}$$

下面给出的两个公式就是著名的德摩根定律:

$$\left(\bigcup_n S_n \right)^c = \bigcap_n S_n^c, \quad \left(\bigcap_n S_n \right)^c = \bigcup_n S_n^c.$$

现在证明第一个公式. 设 $x \in (\bigcup_n S_n)^c$, 这说明 $x \notin \bigcup_n S_n$, 即对一切 n , $x \notin S_n$. 因而, 对每一个 n , x 属于 S_n 的补集, 即 $x \in \bigcap_n S_n^c$. 这样, 我们得到 $(\bigcup_n S_n)^c \subset \bigcap_n S_n^c$. 反过来包含关系的证明, 只需将我们的论证从后面往前推即可. 而第二个公式的证明完全类似.

1.2 概率模型

概率模型是对不确定现象的数学描述. 为了与本节讨论的基本框架保持一致, 下面列出了它的两个基本构成, 并用图1.2做了形象阐释.

概率模型的基本构成

- 样本空间 Ω , 这是一个试验的所有可能结果的集合.
- 概率律, 概率律为试验结果的集合 A (称为事件) 确定一个非负数 $P(A)$ (称为事件 A 的概率). 而这个非负数刻画了我们对事件 A 的认识或所产生的信念的程度. 稍后将指出概率律必须满足的某些性质.

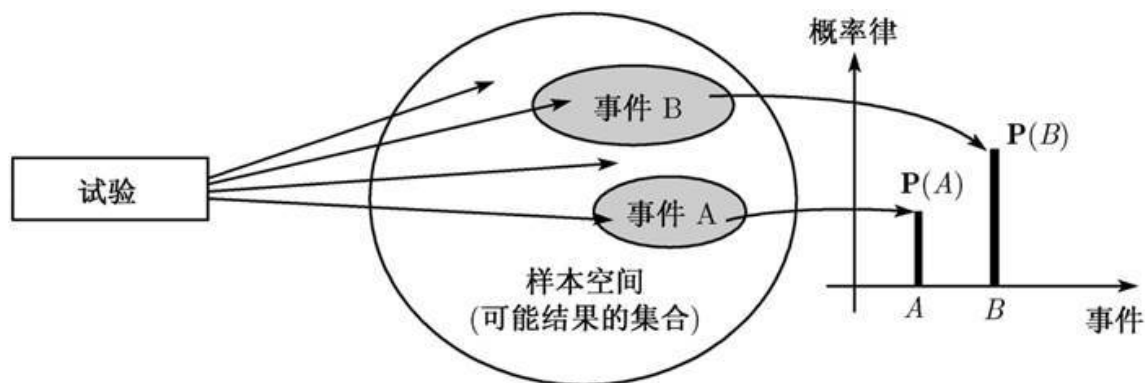


图 1.2 概率模型的基本构成

1.2.1 样本空间和事件

每一个概率模型都关联着一个试验, 这个试验将产生一个试验结果. 该试验的所有可能结果形成样本空间, 用 Ω 表示样本空间. 样本空间的子集, 即某些试验结果的集合, 称为事件¹. 一个试验由什么组成, 并没有什么限制. 例如, 可以抛掷一次硬币, 也可以抛掷三次硬币, 或持续地、无限次地抛掷硬币. 然而我们所讨论的概率模型的问题中, 只涉及一个试验. 所以连续抛掷三次硬币的试验, 只能作为一次试验, 不能认为是三次试验.

¹任意一个可能结果的集合, 包括样本空间 Ω 本身和它的补集 \emptyset , 都可能作为事件. 当然, 严格来讲, 在一个具有不可数无限多个试验结果的样本空间中, 有些子集不可能定义有意义的概率. 这涉及测度论的数学知识. 但实际上我们一般不会遇到这种特殊的情况, 因此我们不必考虑这种特殊问题.

样本空间可由若干个试验结果组成, 也可由无限多个试验结果组成. 从数学上和概念上来看, 有限样本空间比较简单. 实际应用中, 具有无限多个结果的样本空间也是很常见的. 例如, 往一个方形目标上掷飞标, 可将每个可能的弹着点作为试验的结果.

1.2.2 选择适当的样本空间

在确定样本空间的时候,不同的试验结果必须是相互排斥的,这样,在试验过程中只可能产生唯一的一个结果.例如,当试验是掷一枚骰子的时候,不能把“1 或 3”定为一个试验结果,同时又把“1 或 4”也定义为一个结果.如果这样定义了,那么当掷得1点的时候,就不知道得到的是什么结果了.

对同一个试验,根据我们的兴趣可以确定不同模型.但是确定模型时,我们不能遗漏其样本空间中的任何一个结果.也就是说在试验过程中不管发生什么情况,我们总能够得到样本空间中的一个结果.另外,在建立样本空间的时候,要有足够的细节区分我们感兴趣的事件,同时要避免不必要的烦琐.

例 1.1 考虑两个不同的游戏,它们都涉及连续抛掷10次硬币.

游戏1: 每次抛掷硬币的时候,只要出现正面向上,我们就赢1元钱.

游戏2: 每次抛掷硬币时,我们都赢1元钱,直到出现第一次正面向上(包括这一次).以后每次抛掷硬币时我们赢2元钱,一直到第二次出现正面向上.每次抛掷得到正面向上的时候,以后每次抛掷硬币所赢的钱数比以前每次抛掷硬币所赢得的钱数加倍.

在游戏1中,我们赢的钱数只与10次抛掷中正面向上的次数有关;而在游戏2中,我们的赢钱数不仅与正面出现的次数有关,也与正反面出现的顺序有关.这样在游戏1中,样本空间可由11个(即 $0, 1, 2, \dots, 10$)试验结果组成,而在游戏2中,样本空间由所有的长度为10的正、反序列组成.

1.2.3 序贯模型

许多试验本身具有序贯的特征.例如,连续抛掷一枚硬币,一共抛三次,或者连续观察一只股票,共观察5天,又或者在一个通信接收设备上接收8位数字.常用序贯树形图来刻画样本空间中的试验结果,如图1.3所示.

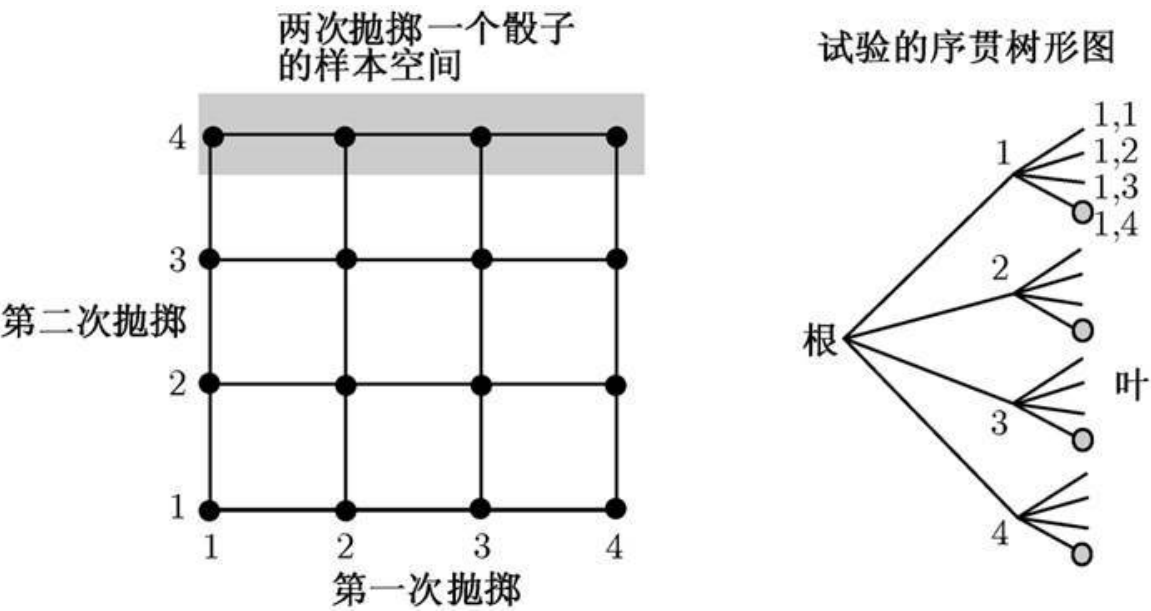


图 1.3 序贯树形图示例. 设所考虑的试验连续两次抛掷有4个面的骰子², 其样本空间有两种等价的刻画方法. 在这个试验中, 可能的结果是全体有序对 (i, j) , 其中 i 表示第一次抛掷骰子得到的数字, j 表示第二次抛掷骰子得到的数字. 试验结果可用左图中的2维格子点表示, 也可以用右图中的序贯树形图表示, 后者的优点是可以表示试验的序贯特征. 在序贯树形图中, 每个可能的试验结果可以用一个末端的树叶表示, 或等价地用与树叶相联系的由根部到树叶的一个路径表示³. 左图中阴影部分代表事件 $\{(1, 4), (2, 4), (3, 4), (4, 4)\}$, 它表示第2次抛掷得到4. 同一个事件可以在右图中用空心圆点标示的叶子集合表示. 注意序贯树形图中的每一个结点可以代表一个事件, 这个事件就是由这个结点出发的所有的叶子构成的事件. 例如, 在序贯树形图中用1标示的结点代表事件 $\{(1, 1), (1, 2), (1, 3), (1, 4)\}$, 即第一次抛掷得1的事件

²当抛掷的骰子有6个试验结果的时候, 就是指抛掷常见的正六面体. 此处可理解为抛掷正四面体, 当落在桌面时, 只有一面与桌面接触. 本书中的骰子都可以如此解释. ——译者注

³用路径表示更能显示试验的序贯特征. ——译者注

1.2.4 概率律

假定我们已经确定了样本空间 Ω 以及与之联系的试验, 为了建立一个概率模型, 下一步就是要引进概率律的概念. 直观上, 它确定了任何结果或者任何结果的集合(称为事件)的似然程度. 更精确一点说, 它给每一个事件 A , 确定一个数 $P(A)$, 称为事件 A 的概率. 它满足下面的几条公理.

概率公理

(1) (非负性) 对一切事件 A , 满足 $P(A) \geq 0$.

(2) (可加性) 设 A 和 B 为两个互不相交的集合(概率论中称为互不相容的事件), 则它们的并满足

$$P(A \cup B) = P(A) + P(B).$$

更一般地, 若 A_1, A_2, \dots 是互不相容的事件序列, 则它们的并满足

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots.$$

(3) (归一化) 整个样本空间 Ω (称为必然事件) 的概率为1, 即 $P(\Omega) = 1$.

为了将概率律形象化, 可以把样本空间中的试验结果看成质点, 每一个质点有一个质量. $P(A)$ 就是这个质点集合的总质量, 而全空间的总质量为1. 这样, 概率律中的可加性公理就变得很直观了: 不相交的事件序列的总质量等于各个事件的质量之和.

概率的更具体的解释是频率. $P(A) = 2/3$ 表示: 在大量重复的试验中事件 A 出现的频率约为 $2/3$. 这样的解释虽然不总是合适的, 但有时却很直观易懂. 第5章将会重新讨论这种解释.

概率律有许多重要的性质并没有包含到上述公理系统中, 原因很简单, 它们可以从公理系统中**推导**出来. 例如, 由可加性和归一化公理可得到

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset),$$

由这个性质可知空事件(称为不可能事件)的概率为0, 即

$$P(\emptyset) = 0.$$

现在推导另一个性质, 令 A_1 、 A_2 和 A_3 为互不相容的事件, 重复利用可加公理, 可得到

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1 \cup (A_2 \cup A_3)) \\ &= P(A_1) + P(A_2 \cup A_3) \\ &= P(A_1) + P(A_2) + P(A_3). \end{aligned}$$

类似的推导可以得到: 有限多个互不相容的事件的并的概率等于它们各自的概率之和. 后面将讨论更多的性质.

1.2.5 离散模型

现在以实例说明构造概率律的方法. 通常我们根据实际试验中的一些常识性假设构造概率律.

例 1.2 考虑抛掷一枚硬币. 一共有两种结果, 正面向上 $\{H\}$ 和反面向上 $\{T\}$. 样本空间为 $\Omega = \{H, T\}$, 事件为

$$\{H, T\}, \{H\}, \{T\}, \emptyset.$$

若硬币是均匀的, 即我们相信在抛掷硬币的时候, 两面具有相同的机会出现, 应该确定两个结果的概率是相等的, 即 $P(\{H\}) = P(\{T\})$. 由可加性公理和归一性公理可知

$$P(\{H, T\}) = P(\{H\}) + P(\{T\}) = 1,$$

由此可推导得概率律

$$P(\{H, T\}) = 1, \quad P(\{H\}) = 0.5, \quad P(\{T\}) = 0.5, \quad P(\{\emptyset\}) = 0.$$

显然, 所建立的概率律满足三条公理.

考虑另一个试验, 依次抛掷三枚硬币. 试验结果是由正面和反面组成的长度为3的序列. 样本空间为

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

假定上述8种结果的可能性是相同的, 即每个结果的概率为1/8. 现在利用三条公理建立概率律. 例如事件

$$A = \{\text{两个正面向上, 一个反面向上}\} = \{HHT, HTH, THH\}.$$

利用概率律的可加性公理, 事件 A 的概率等于组成该事件的试验结果的概率之和:

$$\begin{aligned} P(\{HHT, HTH, THH\}) &= P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{3}{8}. \end{aligned}$$

相似地, 任何事件的概率等于 $1/8$ 乘上该事件中包含的结果的个数. 所建立的概率律满足三条公理.

利用概率律的可加性公理以及前面例子中的推理方法, 可以得到下面的结论.

离散概率律

设样本空间由有限个可能的结果组成, 则事件的概率可由组成这个事件的试验结果的概率所决定. 事件 $\{s_1, s_2, \dots, s_n\}$ 的概率是 $P(s_i)$ 之和, 即

$$P(\{s_1, s_2, \dots, s_n\}) = P(s_1) + P(s_2) + \dots + P(s_n).$$

此处用简单的记号 $P(s_i)$ 表示事件 $\{s_i\}$ 的概率, 而不用正式的记号 $P(\{s_i\})$. 本书后面都按这个简化的约定书写.

现在设样本空间为 $\Omega = \{s_1, s_2, \dots, s_n\}$, 并且每个试验结果是等概率的. 利用归一化公理可知 $P(s_i) = 1/n, i = 1, 2, \dots, n$, 得到以下定律.

离散均匀概率律(古典概型)

设样本空间由 n 个等可能性的试验结果组成, 因此每个试验结果组成的事件(称为基本事件)的概率是相等的. 由此得到

$$P(A) = \frac{\text{含于事件 } A \text{ 的试验结果数}}{n}.$$

现在进一步讨论一些例子.

例 1.3 考虑连续两次抛掷一个有4个面的骰子(见图1.4). 现在假定这些骰子是均匀的, 这个假定意味着16种可能的试验结果是等可能的, 即16种结果的每一种可能的结果 (i, j) 出现的概率为 $1/16 (i, j = 1, 2, 3, 4)$. 这是一个古典概型. 在计算一个事件的概率的时候必须数清楚这个事件所包含的试验结果数(基本事件数), 将这个结果数除以16(基本事件总数)便得到这个事件的概率. 下面几个事件概率就是用这种方法计算得到的.

$$P(\{\text{两次点数总和为偶数}\}) = 8/16 = 1/2,$$

$$P(\{\text{两次点数总和为奇数}\}) = 8/16 = 1/2,$$

$$P(\{\text{第一次点数与第二次点数相同}\}) = 4/16 = 1/4,$$

$$P(\{\text{第一次点数比第二次点数大}\}) = 6/16 = 3/8,$$

$$P(\{\text{至少有一次点数等于 } 4\}) = 7/16.$$

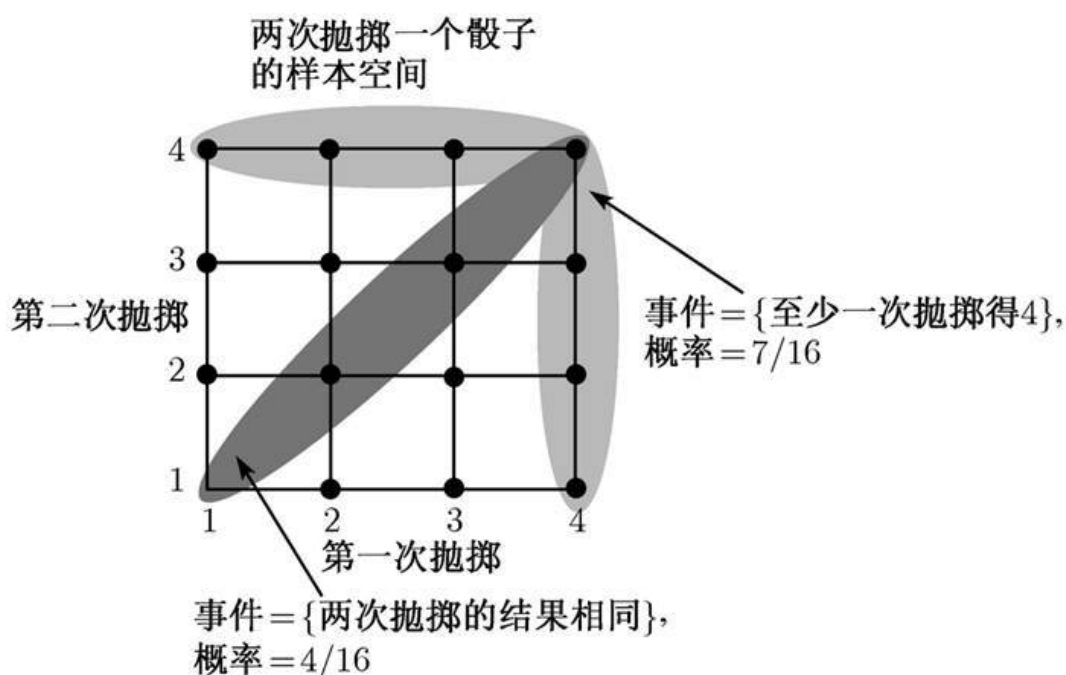


图 1.4 在连续两次抛掷有4个面的骰子的过程中的若干事件和它们的概率, 计算依据离散均匀概率律

1.2.6 连续模型

若试验的样本空间是一个连续集合, 其相应的概率律与离散情况有很大的差别. 在离散情况下, 用基本事件的概率就可以确定概率律, 但连续情况却不同. 下面是一个例子. 这个例子将离散模型中的均匀概率律推广到连续的情况.

例 1.4 在赌场中有一种称为幸运轮的赌具. 在轮子上均匀连续地刻度, 刻度范围为0到1. 当转动的轮子停止时, 固定的指针会停留在刻度上. 这样, 产生的试验结果是 $[0, 1]$ 中的一个数, 指针所指向的位置的刻度. 因此样本空间是 $\Omega = [0, 1]$. 假定轮子是均匀的, 因此可以认为轮子上的每一个点在试验中都是等可能的. 但一个单点在试验中出现的可能性有多大呢? 它不可能是正数, 否则的话, 若单点出现的概率为正, 利用可加性公理, 可导致某些事件的概率大于1的荒谬结论. 因此单个点所组成的事件的概率必定为0.

在本例中,可定义子区间 $[a, b]$ 的概率为 $b-a$.更复杂的集合的概率可以定义为这个集合的长度.⁴这样定义的概率满足概率律的三条公理,因此本例中定义的概率是符合要求的概率律.

⁴ $[0, 1]$ 的一个子集 S 的长度定义为 $\int_S dt$,对于比较简单的子集,可利用通常的微积分计算这个积分.对于某些不寻常的集合,这个积分可能没有合适的定义,这些事情属于更高深的数学处理的问题.顺便指出,用长度刻画概率律的合法性取决于单位区间是一个不可数无限集.不然的话,由于每个单点的概率为零,可导致 $[0, 1]$ 的概率为0的结论,这与概率的归一化公理相矛盾.

例 1.5 罗密欧和朱丽叶约定在某时刻见面,而每个人到达约会地点的时间都会有延迟,延迟时间在 $0 \sim 1$ 小时.第一个到达约会地点的人会在那儿等待15分钟,等了15分钟后若对方还没有到达约会地点,先到者会离开约会地点.问他们能够相会的概率有多大?

考虑直角坐标系的单位正方形 $\Omega = [0, 1] \times [0, 1]$.正方形中的每个点的两个坐标分别代表他们可能的延迟时间.每个点都可以是他们的延迟时间,而且是等可能的.由于等可能性的特点,我们将 Ω 的子集出现的概率定义为这个子集的面积.这个概率律满足三条概率公理.罗密欧和朱丽叶能够相会的事件可用图1.5中阴影部分表示.它的概率等于 $7/16$.

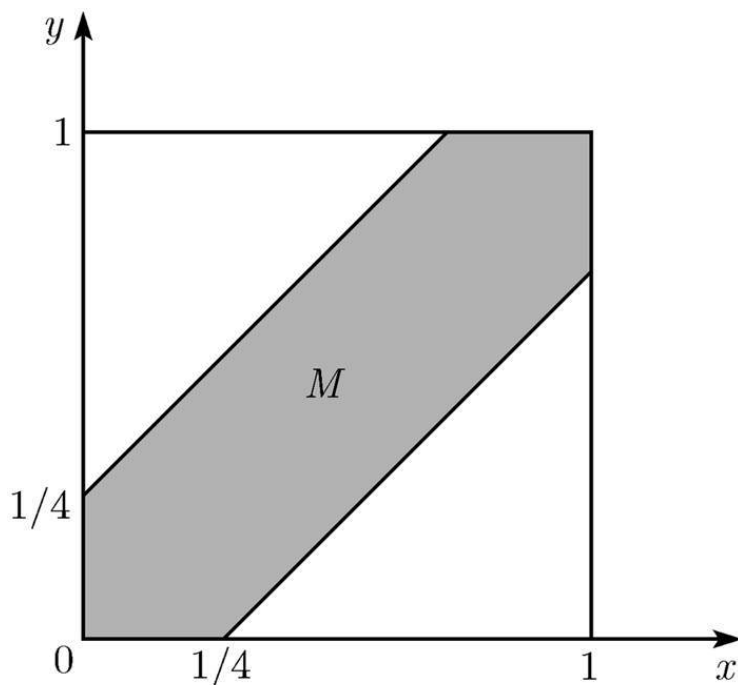


图 1.5 事件 M 代表罗密欧和朱丽叶的相互等待时间不超过15分钟, M 是图中阴影部分(见例 1.5),即

$$M = \{(x, y) \mid |x - y| \leq 1/4, 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

M 的面积等于1减去两个没有阴影的三角形的面积之和,即 $1 - (3/4) \cdot (3/4) = 7/16$.因此,他们能够相会的概率为 $7/16$

1.2.7 概率律的性质

由概率公理可以推导出很多性质, 下面列举若干性质.

概率律的若干性质

考虑一个概率律, 令 A 、 B 和 C 为事件.

(a) 若 $A \subset B$ 则 $P(A) \leq P(B)$.

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(c) $P(A \cup B) \leq P(A) + P(B)$.

(d) $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.

这些性质以及其他类似的性质, 都可以形象化地用维恩图证明 (见图1. 6). 注意, 性质 (c) 可以推广成

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^n P(A_i).$$

现在证明这个推广的结果. 将性质 (c) 用于事件 A_1 和 $A_2 \cup \cdots \cup A_n$, 得到

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

进一步将性质 (c) 用于事件 A_2 和 $A_3 \cup \cdots \cup A_n$, 得到

$$P(A_2 \cup A_3 \cup \cdots \cup A_n) \leq P(A_2) + P(A_3 \cup \cdots \cup A_n).$$

如此继续下去, 最后将诸不等式相加, 便得到所需结果.

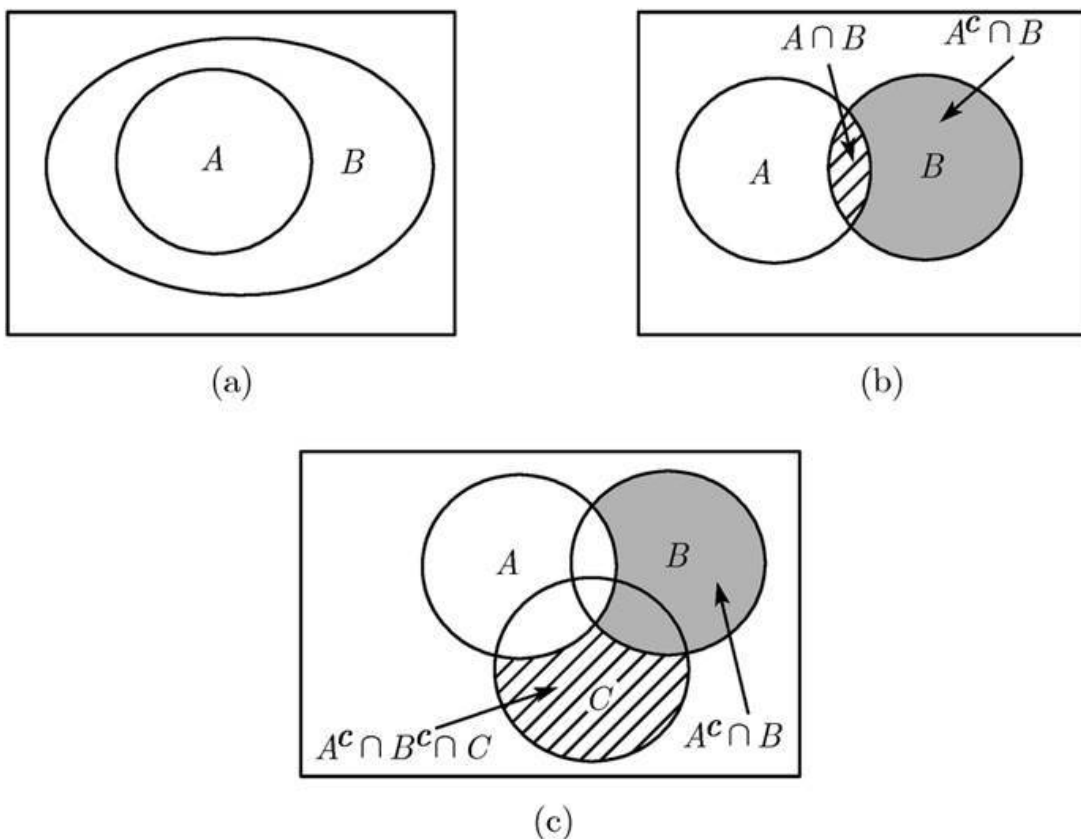


图 1.6 利用维恩图直观地验证概率律的性质(a)、(b)、(c)和(d). 设 $A \subset B$, 则 B 是两个不相容的事件 A 和 $A^c \cap B$ 的并(见图(a)). 利用可加性公理得

$$P(B) = P(A) + P(A^c \cap B) \geq P(A),$$

其中不等式利用了概率的非负性公理. 性质(a)得证.

由图(b), 可将事件 $A \cup B$ 和 B 分解成不相容的事件之并:

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

利用可加性公理, 得到

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

第一式减去第二式并移项得到 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, 即性质(b)成立. 利用概率的非负性公理得到 $P(A \cap B) \geq 0$, 从而 $P(A \cup B) \leq P(A) + P(B)$ 成立, 性质(c)得证.

由图(c), 可以看出事件 $A \cup B \cup C$ 可以分解成三个互不相容的事件的并:

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

重复利用可加性公理可得到性质(d).

1.2.8 模型和现实

概率理论可以用来分析现实世界的许多不确定现象. 这个过程通常分成两个阶段.

(a) 第一阶段, 在一个适当的样本空间中给出概率律, 从而建立概率模型. 在这个阶段, 没有关于建立模型的一般规则, 只要你建立的概率律符合概率的三条公理就行. 有些人会怀疑所建立模型的真实性的. 有时, 人们宁愿使用“错误”的模型, 其理由是“错误”的模型比“正确”的模型简单且易于处理. 这种处理问题的态度在科学和工程学中很普遍. 在实际工作中, 选择的模型往往既要准确、简单又要兼顾易操作性. 此外, 统计学家还依据历史数据和过去相似试验的结果, 利用统计方法确定模型. 这将在第8章和第9章讨论.

(b) 在第二阶段, 我们将在完全严格的概率模型之下进行推导, 计算某些事件的概率或推导出一些十分有趣的性质. 第一阶段的任务是建立现实世界与数学的联系, 而第二阶段则是严格限制在概率公理之下的逻辑推理. 在后一阶段, 如果涉及的计算很复杂或概率律的陈述不简明, 推理和理解就会遇到困难. 但是所有的问题将会有有一个准确的答案, 不会产生歧义. 只要有足够高的能力, 所有的困难都将化为乌有.

在概率论中充满这样的“悖论”: 对同一个问题, 不同的计算方法似乎会得到不同的结论. 在这种模棱两可的概率律中选定模型会导致结论的不确定. 贝特朗悖论是一个著名的例子(见图1.7).

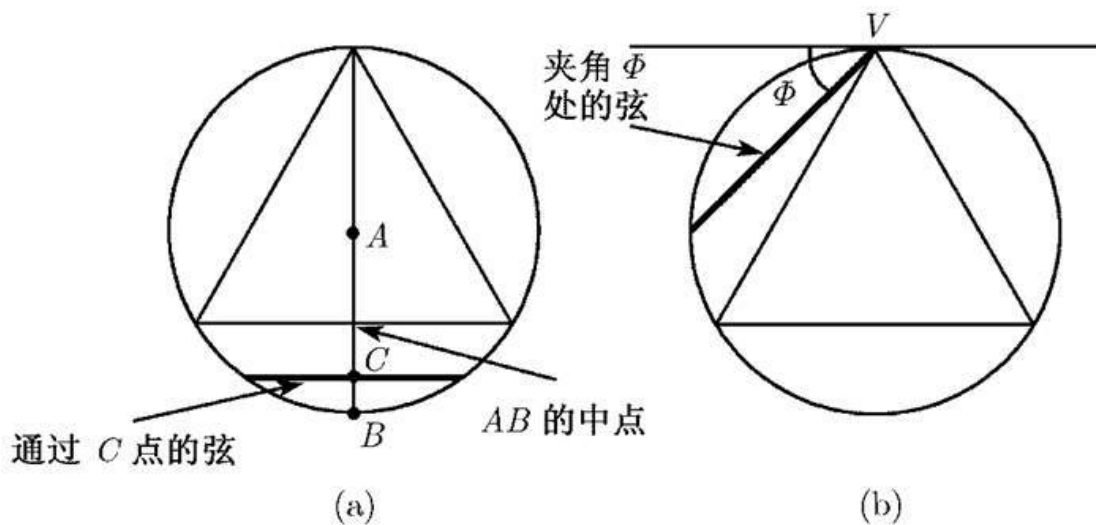


图 1.7 贝特朗悖论. 该例子由贝特朗于1889年提供, 它说明这样一个原理: 解决一个实际问题的时候, 必须建立无歧义的概率模型. 设在一个圆内有一个正三角形, 内接于圆周. 现在随机地选定一个弦, 问其长度大于内接正三角形的边的概率等于多少? 其解答依赖于“随机地选定”的确切定义. 图中(a)和(b)的两种方法导致相互矛盾的结论.

在图(a)中, 取一半径 AB , 在 AB 上随机地取一个点 C . 所谓随机地取点是指 AB 上所有的点具有相同的机会被取到. 通过点 C , 作一条弦垂直于 AB . 由初等几何的知

识可知,当 C 点的位置恰巧在 AB 的中点时,弦的长度刚好等于三角形的边长,而远离圆心时,弦的长度减小.这样弦的长度大于内接正三角形的边长的概率等于 $1/2$.

在图(b)中,圆周上取一点 V 作为顶点.通过 V 先画一条切线,然后随机地画一条通过 V 的直线.记直线与切线的夹角为 ϕ .由于这条直线是随机地画的,可以认为夹角 ϕ 是在 $(0, \pi)$ 上均匀地分布的.现在考虑这条直线割圆得到的弦的长度.由初等几何的知识可知,当 ϕ 处于 $(\pi/3, 2\pi/3)$ 的范围内,弦的长度大于三角形的边长.由于 ϕ 取值于 $(0, \pi)$,故这根弦大于内接正三角形边长的概率是 $1/3$

概率论发展简史

- 公元前.在古希腊和古罗马时期,机会游戏十分盛行.但是这个时期关于游戏的理论还没有发展起来.究其原因,那时候希腊的数字系统不能提供代数运算发展的机会.在科学分析基础上的概率论一直等到印度和阿拉伯发明了现代算术系统(第一个千年的后半叶),以及文艺复兴时期产生了大量的科学思想,才有机会发展.
- 16世纪.卡尔达诺,一个光彩夺目同时富有争议的意大利数学家出版了第一本关于机会游戏的书,书中给出了掷骰子和扑克游戏中随机事件的概率的正确计算方法.
- 17世纪.费马和帕斯卡之间的通信中提及几个十分有趣的概率问题,推动了这个领域的研究热潮.
- 18世纪.雅各布·伯努利研究了重复投币试验序列并引入了第一条大数定律.这条大数定律为联系理论概率与经验事实打下了基础.后面的数学家,诸如丹尼尔·伯努利、莱布尼茨、贝叶斯、拉格朗日等人对理论概率论的发展和实际应用也作出了巨大贡献.棣莫弗引入了正态分布并证明了第一个中心极限定理.
- 19世纪.拉普拉斯在他的一本很有影响的书中确立了概率论在定量研究领域中的重要地位.同时他本人对概率论作出了许多原创性的贡献,包括推导了更一般形式的中心极限定理.勒让德和高斯将概率论应用到天文预测,并且应用了最小二乘法,他们的工作大大地拓展了概率论的应用领域.泊松出版了一本很有影响的书,其中包括了很多原创性的成果,以他的姓名命名的泊松分布也在其中.切比雪夫和他的学生马尔可夫、李雅普诺夫等研究了极限定理,在这个领域内提高了数学的严格性的标准.在此时期,概率论被认为是自然科学的一部分,它的主要任务是解释物理现象.在这种思想的主导之下,概率被解释为重复试验中相对频率的极限.
- 20世纪.现在已经不再以相对频率作为概率论的基础概念.代之以由科尔莫戈罗夫引入的普遍适用的概率论公理系统.与数学的其他分支一样,在公理系统的基础上发展起来的概率论只依赖于逻辑的正确性,而与实际物理现象的联系无关.然而,由于概率论能够描述和解释现实世界中绝大部分的不确定性现象,因而在科学和工程中,概率论得到广泛应用.

1.3 条件概率

条件概率是在给定部分信息的基础上对试验结果一种推断. 下面是一些例子.

(a) 在连续两次抛掷骰子的试验中, 已知两次抛掷的点数的总和为9, 第一次抛掷的点数为6的可能性有多大?

(b) 在猜字游戏中, 已知第一个字母为 t , 第二个字母为 h 的可能性有多大?

(c) 在查体时, 为检查是否患某种疾病需要检测某项指标. 已知某人的该项指标为阴性, 问这个人得病的可能性有多大?

(d) 在雷达显示屏上出现一个点, 这个点代表远处有一架飞机的可能性有多大?

用更确切的话说, 给定一个试验、与这个试验相对应的样本空间和概率律, 假设我们已经知道给定的事件 B 发生了, 而希望知道另一个给定的事件 A 发生的可能性. 因此, 我们要构造一个新的概率律, 它顾及了事件 B 已经发生的信息, 求出任何事件 A 发生的概率. 这个概率就是给定 B 发生之下事件 A 的条件概率, 记作 $P(A|B)$.

这个新的条件概率必须是合格的概率律, 即满足三条概率公理. 同时当原来的概率律为等概率模型时, 其相应的条件概率也应当与直观相符合. 例如, 在抛掷骰子的试验中一共有6种等概率的试验结果. 如果我们已经知道试验的结果是偶数, 即2, 4, 6这三种结果之一发生. 而这三种结果发生的可能性应该是相等的. 这样, 得到

$$P(\text{试验结果是 } 6 \mid \text{试验结果是偶数}) = \frac{1}{3}.$$

从这个结果的推导过程看出, 对于等概率模型的情况, 下面的关于条件概率的定义是合适的, 即

$$P(A|B) = \frac{\text{事件 } A \cap B \text{ 的试验结果数}}{\text{事件 } B \text{ 的试验结果数}}.$$

将这个结果推广, 我们得到下面的条件概率定义:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

其中假定 $P(B) > 0$. 如果 B 的概率为0, 相应的条件概率是没有定义的. 总而言之, $P(A|B)$ 是事件 $A \cap B$ 的概率与事件 B 的概率的比值.

1.3.1 条件概率是一个概率律

对于给定的事件 B , 条件概率 $P(A|B)$ 形成了样本空间上的一个概率律, 即条件概率满足三条概率公理. 非负性是明显的. 又由于

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

说明归一化公理也是满足的. 现在验证可加性. 设 A_1 和 A_2 是任意两个不相容的事件,

$$\begin{aligned} P(A_1 \cup A_2|B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\ &= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} \\ &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \\ &= P(A_1|B) + P(A_2|B), \end{aligned}$$

此处第三个等式利用了事件 $A_1 \cap B$ 和 $A_2 \cap B$ 的不相容性和无条件概率的可加性. 可数个互不相容事件的可加性的验证是类似的.

由于我们已经证实了条件概率是一个合格的概率律, 所有关于概率律的性质对于条件概率都是成立的. 例如将 $P(A \cup C) \leq P(A) + P(C)$ 转变成条件概率的性质, 变成

$$P(A \cup C|B) \leq P(A|B) + P(C|B).$$

注意到 $P(B|B) = P(B)/P(B) = 1$, 条件概率完全集中在 B 上, 这样, 我们也可以将 B 以外的结果排除掉, 并将 B 看成新的样本空间.

现在将条件概率的性质加以总结.

条件概率的性质

- 设事件 B 满足 $P(B) > 0$, 则给定 B 之下, 事件 A 的条件概率由下式给出

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

这个条件概率在同一个样本空间 Ω 上给出了一个新的(条件)概率律. 凡是现有的概率律的所有性质对这个条件概率都是适用的.

- 由于条件概率所关心的事件都是事件 B 的子事件, 可以把条件概率看成 B 上的概率律, 即把事件 B 看成全空间或必然事件.
- 当试验的 Ω 为有限集, 并且所有试验结果为等可能的情况下, 条件概率律可由下式给出

$$P(A|B) = \frac{\text{事件 } A \cap B \text{ 的试验结果数}}{\text{事件 } B \text{ 的试验结果数}}.$$

例 1.6 在连续三次抛掷一个两面均匀的硬币的试验中, 我们希望找到 $P(A|B)$, 其中 A 和 B 由下式给出

$$A = \{\text{正面出现的次数多于反面出现的次数}\}, \quad B = \{\text{第一次抛掷得到正面}\}.$$

样本空间由下列8个试验结果组成:

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

由于硬币两面的均匀性, 可以假定这8个试验结果是等可能的. 事件 B 由4个试验结果 HHH 、 HHT 、 HTH 、 HTT 组成, 因此

$$P(B) = \frac{4}{8}.$$

而事件 $A \cap B$ 由结果 HHH 、 HHT 、 HTH 组成, 其概率

$$P(A \cap B) = \frac{3}{8}.$$

这样, 得到

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4}.$$

由于所有的试验结果是等概率的, 我们也可用简化的算法计算 $P(A|B)$. 不必计算 $P(B)$ 和 $P(A \cap B)$, 而直接计算事件 $A \cap B$ 和 B 中的基本事件个数(分别等于3和4), 相比即得 $3/4$.

例 1.7 在连续两次抛掷一个均匀的有4个面的骰子的试验中, 假定所有16种试验结果是等可能的, 分别记 X 和 Y 为第一次和第二次抛掷的结果. 现在希望计算条件概率 $P(A|B)$, 其中

$$A = \{\max(X, Y) = m\}, \quad B = \{\min(X, Y) = 2\},$$

而 $m=1, 2, 3, 4$. 像上一个例子一样, 可以有两种计算方法. 一种方法是首先计算 $P(A \cap B)$ 和 $P(B)$, 然后按条件概率的定义计算 $P(A|B)$. 而 $P(A \cap B)$ 和 $P(B)$ 的计算方法是: 数清楚这些事件中的试验结果的个数, 再除以16. 另一种方法是直接将 $A \cap B$ 中的试验结果的个数除以 B 中的试验结果的个数(见图1.8).

所有结果都具有等可能的概率1/16

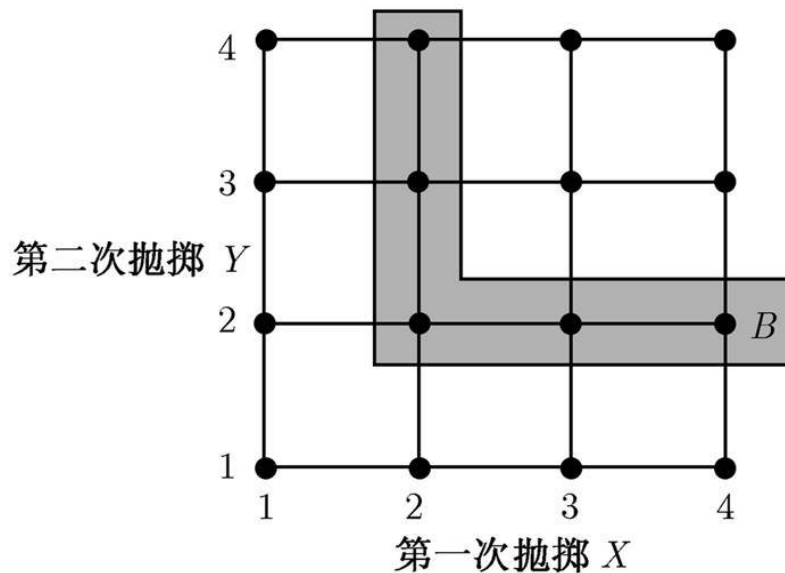


图 1.8 例1.7的图示. 试验的样本空间由连续两次抛掷有4个面的骰子的所有可能结果组成, 刻画条件的事件 $B = \{\min(X, Y) = 2\}$ 由5个阴影部分的点所代表的试验结果组成. 事件 $A = \{\max(X, Y) = m\}$ 与事件 B 的相交部分当 $m = 3$ 或4时有两个试验结果, 当 $m = 2$ 时只有一个试验结果, 当 $m = 1$ 时没有公共元素. 这样, 我们得到

$$P(\{\max(X, Y) = m\} | B) = \begin{cases} 2/5, & m = 3 \text{ 或 } 4, \\ 1/5, & m = 2, \\ 0, & m = 1. \end{cases}$$

例 1.8 有两个设计团队, 一个比较稳重, 记作 C , 另一个具有创新性, 记作 N . 要求他们分别在一个月内做一个新设计. 从过去的经验知道:

- (a) C 成功的概率为2/3;
- (b) N 成功的概率为1/2;
- (c) 两个团队中至少有一个成功的概率为3/4.

已知两个团队中只有一个团队完成了任务. 问这个任务是 N 完成的概率有多大?

现在共有4种可能的结果,

SS : 双方成功

FF : 双方失败

SF : C 成功, N 失败

FS : C 失败, N 成功

现在将(a)、(b)和(c)写成概率等式

$$P(SS)+P(SF) = \frac{2}{3}, \quad P(SS)+P(FS) = \frac{1}{2}, \quad P(SS)+P(SF)+P(FS) = \frac{3}{4}.$$

结合归一化公理

$$P(SS) + P(SF) + P(FS) + P(FF) = 1,$$

得到

$$P(SS) = \frac{5}{12}, \quad P(SF) = \frac{1}{4}, \quad P(FS) = \frac{1}{12}, \quad P(FF) = \frac{1}{4}.$$

所求的条件概率为

$$P(FS|\{SF, FS\}) = \frac{\frac{1}{12}}{\frac{1}{4} + \frac{1}{12}} = \frac{1}{4}.$$

1.3.2 利用条件概率定义概率模型

在为实验建立具有序贯特征的概率模型的时候,通常很自然地首先确定条件概率,然后确定无条件概率.在这个过程中,经常使用的是条件概率公式 $P(A \cap B) = P(B)P(A|B)$.

例 1.9(雷达探测器) 有一台雷达探测设备在工作,若在某区域有一架飞机,雷达以99%的概率探测到并报警.若该地区没有飞机,雷达会以10%的概率虚假报警.现在假定一架飞机以5%的概率出现在该地区.问飞机没有出现在该地区而雷达虚假报警的概率有多大?飞机出现在该地区而雷达没有探测到的概率有多大?

可以用图1.9的序贯树形图表达这些事件.记

$$A = \{\text{飞机出现}\},$$

$$B = \{\text{雷达报警}\},$$

而它们的补集为

$$A^c = \{\text{飞机不出现}\},$$

$$B^c = \{\text{雷达未报警}\}.$$

题中给出的概率记录在图1.9中描述样本空间的序贯树的相应枝条上.每个试验结果可用树形图的叶子表示,它的概率等于由根部到树叶的枝条上显示的数据的乘积.所求的