1.3 条件概率

北京化工大学数学系

苏贵福

条件概率是概率论中的一个重要而实用的概念. 所考虑的是事件 A已经发生的条件下事件B发生的概率问题.

例1 将一枚硬币投掷两次, 观察其正反面出现的情况. 设事件$A =$"至少有一次为H"; 事件$B =$"两次投掷出同一面". 试求事件A发生的情况下事件B发生的概率.

解 易知样本空间为$U = \{HH, HT, TH, TT\}$, $A = \{HH, HT, TH\}$, $B = \{HH, TT\}$. 由事件A已经发生可知, TT不可能出现. 即试验所有可能的结果构成的集合为A. 而A中 共有3个元素, 其中只有$HH \in B$. 因此

$$P(B|A) = \frac{1}{3}.$$

- $P(B) = \frac{2}{4} \neq \frac{1}{3} = P(B|A)$.
- $P(B|A) = \frac{1}{3} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{P(AB)}{P(A)}$.

一. 条件概率

定义1 设A, B是两个事件, 且$P(A)>0$, 我们称

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

为在事件A已经发生的条件下事件B发生的条件概率.

条件概率$P(\cdot|A)$满足概率定义中的三个条件. 即

- 当$P(A)>0$时, 对任一事件A有$P(B|A)\geq 0$.
- 当$P(A)>0$时, 有$P(U|A)=1$.
- 当$P(A)>0$时, 若$B_1, B_2, \cdots$是两两互斥的事件, 则有

$$P(\bigcup_{i=1}^{\infty} B_i|A)=\sum_{i=1}^{\infty} P(B_i|A)$$

例2 一盒子装有4只产品，其中有3只一等品，1只二等品．现从中取产品两次，每次任取一只，作不放回抽样．设事件$A =$"第一次取到的是一等品"；事件$B =$"第二次取到的是一等品"．试求条件概率$P(B|A)$．

解 易知该问题属于古典概型．为表示方便对产品进行编号，设 1，2，3号为一等品，4号为二等品．以(i, j)表示第一次，第二次分别取到第i号，第j号产品．因而样本空间为

$$U = \{(1,2),(1,3),(1,4),(2,1),(2,3),(2,4),(3,1),(3,2),(3,4),(4,1),(4,2),(4,3$$

同时有

$$A = \{(1,2),(1,3),(1,4),(2,1),(2,3),(2,4),(3,1),(3,2),(3,4)\}$$

$$AB = \{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}$$

于是由条件概率的定义

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{6}{12}}{\frac{9}{12}} = \frac{2}{3}. \quad \blacksquare$$

也可以直接按条件概率的含义来求$P(B|A)$. 我们知道, 当事件A发生后, 试验E的所有可能结果就是集合A. 而A中有9个元素, 其中只有$\{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\} \in B$. 故可得

$$P(B|A) = \frac{6}{9} = \frac{2}{3}.$$

二. 乘法公式

定义2 设$P(A)>0$, 我们称$P(AB)=P(B|A)P(A)$为乘法公式.

- 设A, B, C为事件, 满足$P(AB)>0$, 则有

$$P(ABC)=P(C|AB)P(B|A)P(A).$$

- 设$A_1, A_2, \cdots, A_n$为n个事件, 且 $P(A_1A_2\cdots A_{n-1})>0$, 则有

$$\begin{aligned}P(A_1A_2\cdots A_n) =&P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\\ &P(A_{n-2}|A_1A_2\cdots A_{n-3})\cdots P(A_2|A_1)P(A_1).\end{aligned}$$

例3 设袋中装有r个红球，t个白球．每次从袋中任取一只球，观察其颜色然后放回，并再放入a只与所取出的那只球同色的球．若在袋中连续取球四次．试求第一，二次取到红球且第三，四次取到白球的概率．

解 以A_i $(i=1,2,3,4)$表示事件"第i次取到红球"，则 $\overline{A}_3, \overline{A}_4$分别表示事件"第三，四次取到白球"．因此所求概率为

$$\begin{aligned} P(A_1A_2\overline{A}_3\overline{A}_4) =& P(\overline{A}_4|A_1A_2\overline{A}_3)P(\overline{A}_3|A_1A_2)P(A_2|A_1)P(A_1) \\ =& \frac{t+a}{r+t+3a}\cdot\frac{t}{r+t+2a}\cdot\frac{r+a}{r+t+a}\cdot\frac{r}{r+t}. \end{aligned}$$

例4 对含有百分之五废品的100件产品 进行抽样检查, 整批产品被拒绝接受的条件是在被抽查的5件产品(不放回抽样)中至少有一件是废品. 试问该批产品被拒收的概率是多少?

解 以A_i $(i=1,2,3,4,5)$表示事件"第i次被抽到的产品为合格品",
令$A=A_1A_2A_3A_4A_5$, 则产品被拒收的概率为$P(\overline{A})=1-P(A)$. 而

$$P(A_1)=95/100 \qquad P(A_2|A_1)=94/99 \qquad P(A_3|A_1A_2)=93/98$$

$$P(A_4|A_1A_2A_3)=92/97 \quad P(A_5|A_1A_2A_3A_4)=91/96.$$

因此所求概率为

$$\begin{aligned}P(A)&=P(A_1A_2A_3A_4A_5)\\&=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3)P(A_5|A_1A_2A_3A_4)\\&=\frac{95}{100}\cdot\frac{94}{99}\cdot\frac{93}{98}\cdot\frac{92}{97}\cdot\frac{91}{96}\approx\frac{23}{100}.\end{aligned}$$

●●●●●●

三. 全概率公式与贝叶斯公式

定义3 设U为试验E的样本空间, $B_1, B_2, \cdots, B_n$为E的一组事件. 若

(1) $B_iB_j = \emptyset,\ i \neq j,\ i, j = 1, 2, \cdots, n.$

(2) $B_1 \cup B_2 \cup \cdots \cup B_n = U.$

则称$B_1, B_2, \cdots, B_n$为样本空间U的一个划分.

- 若$B_1, B_2, \cdots, B_n$是样本空间的一个划分, 那么对每次试验E, 事件$B_1, B_2, \cdots, B_n$中必有一个且仅有一个发生.

定理1 设U为试验E的样本空间, A为E的事件, $B_1, B_2, \cdots, B_n$为U的一个划分, 且 $P(B_i) > 0$, $(i = 1, 2, \cdots, n)$. 则

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n).$$

上述公式称为全概率公式.

证明 因为

$$A = AU = A(B_1 \cup B_2 \cup \cdots \cup B_n) = AB_1 \cup AB_2 \cup \cdots AB_n.$$

由假设$P(B_i) > 0$, 且$(AB_i) \cap (AB_j) = \emptyset$. 于是

$$\begin{aligned} P(A) &= P(AB_1) + P(AB_2) + \cdots + P(AB_n) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n). \end{aligned}$$

① 全概率公式旨在将一个随机事件A分成若干互不相容的事件

$$A = AB_1 \cup AB_2 \cup \cdots \cup AB_n$$

使得每一个事件的概率可以通过条件概率公式求得.

② 全概率公式说明事件A发生的概率是它在每一种原因或情况下发生的条件概率的加权平均, 而权重分别为$P(B_1)$, $P(B_2)$, $\cdots$, $P(B_n)$.

例5 某工厂有甲乙丙三个车间生产同一种产品，其产量分别占全厂产量的 $\frac{25}{100}$，$\frac{35}{100}$，$\frac{40}{100}$，相应的次品率分别为 $\frac{5}{100}$，$\frac{4}{100}$，$\frac{2}{100}$. 从全厂产品任取一件产品，求取得次品的概率.

解 以A表示事件"取一件产品为次品"，以B_1, B_2, B_3分别表示"取得甲乙丙 车间生产的产品". 不难发现B_1, B_2, B_3是事件A发生的三种不同情况，故构成 该种试验的样本空间U的一个划分. 于是

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= \frac{5}{100} \cdot \frac{25}{100} + \frac{4}{100} \cdot \frac{35}{100} + \frac{2}{100} \cdot \frac{40}{100} = \frac{3.45}{100}. \end{aligned}$$

定理2 设U为试验E的样本空间, A为E的事件, $B_1, B_2, \cdots, B_n$为U的一个划分, 且 $P(A) > 0$, $P(B_i) > 0$, $(i = 1, 2, \cdots, n)$. 则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}.$$

上述公式称为贝叶斯公式.

♣ 贝叶斯公式主要用于根据已经发生的事情或出现的结果, 反过来推断造成该事情或结果的各种原因的可能性大小.

例6 根据以往的临床记录, 某种诊断癌症的 试验具有如下效果: 若以A表示事件"试验反应为阳性", 以C表示事件"被诊断者患有癌症". 则有$P(A|C)=0.95$, $P(\overline{A}|\overline{C})=0.95$. 现在对自然人群进行普查, 设被试验的人患有癌症的概率为0.005, 即$P(C)=0.005$. 试求$P(C|A)$.

解 已知$P(A|C)=0.95$, 且$P(A|\overline{C})=1-P(\overline{A}|\overline{C})=0.05$.
又$P(C)=0.005$, 则$P(\overline{C})=1-P(C)=0.995$. 从而由贝叶斯公式得

$$\begin{aligned} P(C|A) &= \frac{P(A|C)P(C)}{P(A|C)P(C)+P(A|\overline{C})P(\overline{C})} \\ &= \frac{0.95\times 0.005}{0.95\times 0.005+0.05\times 0.995} = 0.087. \end{aligned}$$

- 本题的结果表明: 若将$P(A|C)$与$P(C|A)$混淆, 会导致医疗误判.