第三节　条件概率

一、条件概率与乘法公式

1. 条件概率

所谓条件概率是指在某事件 B 发生的条件下，求另一事件 A 发生的概率，记为 $P(A|B)$．它与 $P(A)$ 是不同的，下面先看一个例子.

例 1 考察有两个小孩的家庭，其样本空间为{（男男），（男女），（女男），（女女）}.

（1）$A=$“家中至少有一个女孩”，则$P(A)=\dfrac{3}{4}$.

（2）若已知事件$B=$“家中至少有一个男孩”发生，再求事件A发生的概率为$P(A|B)=\dfrac{2}{3}$.

（3）$AB=$“家中有一男一女两个孩子”，则$P(AB)=\dfrac{2}{4}$，且有$P(A|B)=\dfrac{2/4}{3/4}=\dfrac{P(AB)}{P(B)}$.

定义 1 设 A, B 是两个事件，且 $P(B)>0$，称

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

为在事件 B 发生的条件下事件 A 发生的**条件概率**.

注 1： $P(\overline{A}|B)=1-P(A|B)$.

条件概率的计算方法

（1）公式法：$P(A|B)=\dfrac{P(AB)}{P(B)}$.

(2）融入法：将条件直接代入到计算过程中去计算.

例 2 设袋中有十只球，其中六只红球和四只白球，现从中不放回地任取两只球，求已知在第一次取得红球的条件下，第二次取得白球的概率.

解法 1（公式法）设事件 A 表示第一次取得红球，B 表示第二次取得白球，则 $P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{\frac{6\times 4}{10\times 9}}{\frac{6}{10}}=\dfrac{4}{9}$.

解法 2（融入法）当第一次取得红球时，袋中还剩下九只球，其中五只红球和四只白球（注意：样本空间已经发生变化）.此时再从袋中任取一只球，则该球为白球的概率为 $\dfrac{4}{9}$.

例 3 在肝癌普查中发现，某地区的自然人群中，每十万人内平均有 40 人患有原发性肝癌，有 34 人甲胎球蛋白高含量，有 32 人既患原发性肝癌又出现甲胎球蛋白高含量．从这个地区的居民中任取一人，若他患有原发性肝癌则记为事件 A，甲胎球蛋白高含量记为事件 B，这时．

$$P(A)=0.0004,\quad P(B)=0.00034,\quad P(AB)=0.00032,$$

于是

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{0.00032}{0.0004}=0.8,$$

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{0.00032}{0.00034}=0.9412.$$

思考：以上两个概率的医学意义.

通过计算得知，患原发性肝癌的人有80%其甲胎球蛋白呈现高含量，而甲胎球蛋白的测定大大有助于发现原发性肝癌的患者；另一方面，若出现甲胎球蛋白高含量，则有94%以上的概率对患原发性肝癌作出正确诊断.

由于事件 B 的发生，使事件 A 发生的概率由 0.0004 一下子上升到 0.9412，可见，事件发生的概率与条件有关，即与信息有关.

2. 乘法公式

定理 1 设 $P(B)>0$，则 $P(AB)=P(B)P(A|B)$.

推论 1 设 $P(A)>0$，则 $P(AB)=P(A)P(B|A)$.

推论 2 设 $P(A_1A_2\cdots A_n)>0$，$n\geq 2$，则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1}).$$

注 2: 一般来说，如果 $A_1,A_2,\cdots,A_n$ 之间有依赖关系，有先后发生的顺序时，计算 $P\left(A_1A_2\cdots A_n\right)$ 用乘法公式，等式右边的条件概率由融入法计算.

例 5 设50个晶体管中有2个次品，每次从中任取一个测试，测试后不放回.

（1）求2个次品分别在第2次测试和第4次测试时出现的概率；

（2）问最少应抽检到多少个晶体管，才能使至少发现一个次品的概率超过0.6？

解 设 A_i 表示第 i 次取得次品，$i=1,2,\cdots,50$.

$$(1)\ P(\overline{A_1}A_2\overline{A_3}A_4)=P(\overline{A_1})P(A_2|\overline{A_1})P(\overline{A_3}|\overline{A_1}A_2)P(A_4|\overline{A_1}A_2\overline{A_3})$$

$$=\frac{48}{50}\times\frac{2}{49}\times\frac{47}{48}\times\frac{1}{47}=\frac{1}{1225}.$$

(2)设抽检到n个晶体管，且不难得$n>1$．则n个晶体管全为正品的概率为

$$P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\cdots\overline{A_n})$$

$$=P(\overline{A_1})P(\overline{A_2}\mid\overline{A_1})P(\overline{A_3}\mid\overline{A_1}\,\overline{A_2})\cdots P(\overline{A_n}\mid\overline{A_1}\,\overline{A_2}\cdots\overline{A_{n-1}})$$

$$=\frac{48}{50}\times\frac{47}{49}\times\cdots\times\frac{48-n+1}{50-n+1}=\frac{(50-n)(49-n)}{50\times 49}.$$

由题意知，$1-\dfrac{(50-n)(49-n)}{50\times 49}>0.6$，解得$n>18$，所以最少应抽检到19个晶体管时，才能使至少发现一个次品的概率超过0.6．

二、全概率公式与贝叶斯公式

1. 完备事件组

定义 2 如果事件组 $A_1, A_2, \cdots, A_n$ 两两互不相容，且 $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$，就称事件组 $A_1, A_2, \cdots, A_n$ 构成样本空间 Ω 的一个完备事件组，简称完备组.

例如，A 和 $\overline{A}$ 为一完备事件组.

样本空间 Ω 的完备事件组实际上就是将 Ω 分解为若干个互不相容事件的并，从而将一个复杂的问题化为几个简单问题去解决.

2. 全概率公式

定理 2 设事件组 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个完备事件组，且 $P(A_i) > 0$， $i = 1, 2, \cdots, n$， 则对任何事件 B，有 $P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$，此式也称为全概率公式.

特别，当完备事件组为 A 和 $\overline{A}$ 时，全概率公式为

$$P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}).$$

例 5（摸彩模型） 设在 n 张彩票中有一张奖券，求第二人摸到奖券的概率.

解 设事件 A_i 表示“第 i 人摸到奖券”， $i=1,2,\cdots,n$.

$P(A_2|A_1)=0, P(A_2|\overline{A_1})=\dfrac{1}{n-1}$，由全概率公式

$$P(A_2)=P(A_1)P(A_2|A_1)+P(\overline{A_1})P(A_2|\overline{A_1})$$

$$=\frac{1}{n}\times 0+\frac{n-1}{n}\times\frac{1}{n-1}=\frac{1}{n}.$$

这表明摸到奖券机会与先后次序无关．因后者可能处于“不利状况”（前者已摸到奖券），但也可能处于“有利状况”（前者没摸到奖券，从而增加后者摸到奖券的机会），两种状况用全概公式综合（加权平均）所得结果（机会均等）既全面又合情理．用类似的方法可得

$$P(A_3)=P(A_4)=\cdots=P\left(A_n\right)=\frac{1}{n}.$$

例 6 已知甲、乙两箱中装有同种产品，其中甲箱中装有3 件合格品和 3 件次品，乙箱中仅装有 3 件合格品．从甲箱中任取 3 件产品放入乙箱后，求从乙箱中任取一件产品是次品的概率．

解 设事件 A_i 表示从甲箱中任取 3 件产品放入乙箱后乙箱中有 i 个次品，$i=0,1,2,3$，又设 B 表示从乙箱中任意取出的一件产品是次品．则 $P(A_i)=\dfrac{C_3^i C_3^{3-i}}{C_6^3}$, $P(B|A_i)=\dfrac{i}{6}$，$i=0,1,2,3$．

由全概率公式

$$P(B)=\sum_{i=0}^{3}P(A_i)P(B|A_i)=\sum_{i=0}^{3}\frac{C_3^i C_3^{3-i}}{C_6^3}\times\frac{i}{6}=\frac{1}{4}.$$

3. 贝叶斯公式

定理 3 设事件组 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个完备事件组，且 $P(A_i) > 0$，$i = 1, 2, \cdots, n$，B 为一随机事件，且 $P(B) > 0$，则

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}, \quad j = 1, 2, \cdots, n.$$

此公式称为贝叶斯公式或逆概率公式.

在试验之前，所求概率 $P(A_1), P(A_2), \cdots, P(A_n)$ 称为先验概率．当试验结束后，发现 B 发生了，此信息有助于探讨事件 $A_1, A_2, \cdots, A_n$ 发生的原因，故所求条件概率

$$P(A_1|B), P(A_1|B), \cdots, P(A_n|B)$$

称为后验概率．

贝叶斯公式可以分为两个步骤计算

第一步：先用全概率公式求出 $P(B)$；

第二步：计算 $P(A_j|B) = \dfrac{P(A_j)P(B|A_j)}{P(B)}$ $j = 1, 2, \cdots, n$.

例 7 假定用血清甲蛋白法诊断肝癌．设事件 A 表示被检验者已患有肝癌，事件 B 表示被检验者被诊断出患有肝癌．又在自然人群中调查得知 $P(A)=0.0004$, $P(B|A)=0.95$，$P(\overline{B}|\overline{A})=0.90$．现有一人被此检验法诊断为患有肝癌，求此人已患有肝癌的概率．

解 由于 A 和 $\overline{A}$ 构成完备组，故由贝叶斯公式

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}$$

$$=\frac{0.0004\times0.95}{0.0004\times0.95+0.9996\times(1-0.90)}=0.0038.$$

思考：既然检验法相当可靠，为什么用该方法诊断为肝癌的人真正患有肝癌的概率却如此之小呢？

主要是先验概率$P(A)$很小，对自然人群来讲，肝癌毕竟是一种罕见病. 诊断告诉我们被此检验法诊断为患有肝癌时，此人真正患有肝癌的可能性并不大.

另一方面，$P(A|B)=0.0038$是$P(A)=0.0004$的9.5倍，表明一旦被诊断为患有肝癌时，千万不可麻痹.