1.5.4 贝叶斯公式（逆概公式）

定理 1.11 设一列事件 $B_1, B_2, \cdots$ 为样本空间 Ω 的一个划分，则对任一事件 A，若 $P(A) > 0$，则有

$$P(B_i|A) = \frac{P(AB_i)}{P(A)} = \begin{cases} 0, & \text{若 } P(B_i) = 0, \\ \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}, & \text{若 } P(B_i) > 0. \end{cases}$$

思想：（1） 先验与后验：把 B_i 看成各种情况，$P(B_i)$ 可以视为各情况 B_i 发生的先验概率；如果知道 A 发生了，条件概率 $P(B_i|A)$ 作为得到 A 发生这一附加信息后，对事件 B_i 发生的概率的重新认识，称为事件 B_i 的后验概率。贝叶斯公式从数量上刻画了从先验到后验的这种变化。

（2） 逆向思维：直接计算 $P(B_i|A)$ 不好算时，可以通过 $P(A|B_i)$ 间接计算。

条件概率是概率论学习中的一个重要台阶，稍微复杂一点的随机现象分析，都会碰到条件概率。以后会碰到条件分布，基础是条件概率，思想相类似。

例 1.18 某种病的发病率为 *0.0004*，健康人在体检之后有 *0.1%* 的概率结果呈阳性（假阳性），病者在体检之后有 *1%* 的概率结果呈阴性（假阴性）。如果一个人体检之后结果显示为阳性，那么他有多大概率患病？

解： “化验呈阳性”为事件 A，“患病”为事件 B，则 $P(A|B) = 0.99$，$P(\bar{A}|B) = 0.01$，$P(A|\bar{B}) = 0.001$，$P(\bar{A}|\bar{B}) = 0.999$。

$$P(A) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B}) = 0.0004 \times 0.99 + (1 - 0.0004) \times 0.001 = 0.13956\%$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{0.0004 \times 0.99}{0.13956\%} = \frac{3.96}{13.956} \approx 28.4\%$$

在本例中可以看到，在化验呈阳性的人中，真患该病的人不足 *30%*。稍加分析就可以知道，由于人群中该病的发病率很低，绝大部分人没有患病，所以虽然假阳性出现的概率很低，但仍然有很多假阳性结果；而病人很少，所以真阳性结果也就相对比较少。可以分析在 *10000* 人中，*9996* 不患病，如果有 *0.1%* 假阳性，就是 *9.996* 呈假阳性；*4* 人患病，*99%* 呈阳性，则有 *3.96* 呈阳性；在这些阳性结果中，真患病的只有 $\frac{3.96}{9.996+3.96} \approx 28.4\%$。

既然在体检的阳性结果中大部分是假阳性，那么应该如何提高待检人群中真阳性的概率？一个办法是通过对首次检查得到阳性结果的人进行复查。

解一：定义事件 B 表示“首查为阳性的人真患病”，A 表示“第二次化验呈阳性”，则

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \\ &= \frac{0.284 \times 0.99}{0.284 \times 0.99 + (1 - 0.284) \times 0.001} \approx 99.7\% \end{aligned}$$

解二：以事件 A_1 表示“第一次体检呈阳性”，A_2 表示“第二次体检呈阳性”，则

$$\begin{aligned}P(B|A_1A_2) &= \frac{P(B|A_1)P(A_2|BA_1)}{P(B|A_1)P(A_2|BA_1)+P(\bar{B}|A_1)P(A_2|\bar{B}A_1)}\\ &= \frac{0.284\times 0.99}{0.284\times 0.99+(1-0.284)\times 0.001}\approx 99.7\%\text{（条件分布的贝叶斯公式）}\end{aligned}$$

这是贝叶斯公式的递归运用，揭示了第一次检验与第二次检验的递推关系。

解三：也可以直接使用贝叶斯公式

$$P(B|A_1A_2)=\frac{P(B)P(A_1A_2|B)}{P(B)P(A_1A_2|B)+P(\bar{B})P(A_1A_2|\bar{B})}\approx 99.7\%$$

例 1.19 在一次智力答题节目中，主持人亮出一道单选题目以及 A、B、C 三个选项，选手毫无把握，随机选择了 A，主持人看到选手一脸茫然，想给选手一个提示，但他又不能直接告诉选手的选择是对还是错，所以他随机地划掉了另外两个选项中错误的那个（比如说 B），主持人说请做出最后的选择，在这种情况下，选手是选择坚持原来的选择，还是应该改变选择？

解:（1） 有一部分人认为，一开始选手选 A 答对的概率为 $\frac{1}{3}$，$P(A\text{ 对})=P(B\text{ 对})=P(C\text{ 对})=\frac{1}{3}$，主持人提示后，剩两个选项，有一个是对的，选 A 答对的可能性为 $\frac{1}{2}$。这样的逻辑是不对的。

（2） 若坚持原来的选择，答对的可能性有多大？

考虑条件概率

$$\begin{cases} A\text{ 对}\begin{cases}\text{主持人划掉 } B\to P=\frac{1}{6}\\ \text{主持人划掉 } C\to P=\frac{1}{6}\end{cases}\\ B\text{ 对，主持人划掉 } C\to P=\frac{1}{3}\\ C\text{ 对，主持人划掉 } B\to P=\frac{1}{3}\end{cases}$$

$$\begin{aligned}P(A\text{ 对}\mid\text{选手选 }A\text{，划掉 }B) &= \frac{P(A\text{ 对}\cap\text{划掉 }B)}{P(\text{划掉 }B)}\\ &= \frac{P(A\text{ 对})P(\text{划掉 }B|A\text{ 对})}{P(A\text{ 对})P(\text{划掉 }B|A\text{ 对})+P(B\text{ 对})P(\text{划掉 }B|B\text{ 对})+P(C\text{ 对})P(\text{划掉 }B|C\text{ 对})}\\ &= \frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times 0+\frac{1}{3}\times 1}=\frac{1}{3}\end{aligned}$$

1.6 事件的独立性

1.6.1 两个事件的独立性

定义 1.14 对于两个事件 A 和 B，若 $P(AB)=P(A)P(B)$，则称 A 与 B 相互独立。

两个事件的独立性是指一个事件的发生不影响另一个事件的可能性。例如掷两枚骰子，第一枚出现的点数不会影响第二枚的点数。从概率角度看，已知 A 发生的条件下，B 发生的概率 $P(B|A)$ 一般来说不等于 B 发生的无条件概率 $P(B)$，如果事件 A 对事件 B 有某种影响，A 发生通常会改变 B 发生的概率。如果 $P(B|A) = P(B)$，说明事件 A 发生不影响事件 B 发生的可能性，这一条件等价于 $P(AB) = P(A)P(B)$，$P(A|B) = P(A)(P(B) > 0)$。

注 1： 独立性也可以用条件概率判断，若 $P(A) > 0$，则 A 与 B 独立等价于 $P(B|A) = P(B)$。

注 2： 概率为 0 的事件与任何事件相互独立；概率为 1 的事件与任何事件相互独立。

注 3： 4 对事件 $\begin{cases} A, B \\ A, \bar{B} \\ \bar{A}, B \\ \bar{A}, \bar{B} \end{cases}$ 中有一对相互独立，则另外 3 对也相互独立。

注 4： 应该如何判断两个事件是否独立呢？一种方法是根据经验，根据题意判断两个事件是否相互影响；另一种方法就是通过定义和性质来计算。

1.6.2 多个事件的独立性

定义 1.15 *如果 3 个事件 A、B、C 满足* $\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{cases}$ *，称 A、B、C 两两独立。*

定义 1.16 *如果 3 个事件 A、B、C 满足* $\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{cases}$ *，称 A、B、C 相互独立，或 A、B、C 彼此独立，或简称 A、B、C 独立。*

注 1： 一组事件两两独立不能保证相互独立。

注 2： 称 n 个事件是相互独立的，如果 $\forall \{i_1, \cdots, i_k\} \subset \{1, \cdots n\}, k = 2, \cdots, n$，成立
$P(A_{i_1}, \cdots, A_{i_k}) = P(A_{i_1}) \cdots\ P(A_{i_k})$

注 3： 如果无限个事件的任意有限的子集都是相互独立的，则称这无限个事件相互独立。

1.6.3 条件独立性

定义 1.17 *若 $P(C) > 0$，$P(AB|C) = P(A|C)P(B|C)$，则称在给定 C 下，A 与 B（条件）独立。*

例 1.20 随机从班里抽取一名同学，记事件 M“性别为男”，事件 H“留短发”，事件 T“身高超过 170cm”。考虑：事件 H 与 T 是否独立？在给定 M 条件下，H 与 T 是否独立？

解： 即考虑 $P(H)=P(H|T)$?，$P(H|M)=P(H|MT)$?

1.6.4 试验的独立性

随机试验由一系列子试验组成，例如连续投掷一枚硬币，每掷一次就可看做一个子试验。

称一系列子试验是相互独立的，如果第 1 次子试验的任一结果（事件）$A_1\subset\Omega_1$，第 2 次子试验的任一结果（事件）$A_2\subset\Omega_2$，$\cdots$，是相互独立的事件序列。

如果各个子试验还彼此相同（即各个子试验有相同的样本空间和相同的概率测度），则称为 n 重独立重复试验；进一步，如果每个子试验的可能基本结果只有两个（A 和 $\bar{A}$），则称为 n 重贝努利试验。

注： n 次随机试验的样本空间可表示为 $\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n$。

例 1.21 连续掷一枚硬币，求首次正面出现在 k 次的概率？

这种情况属于独立重复试验，“首次正面出现在 k 次” = “第 1 次反面”$\cap$“第 2 次反面”$\cap\cdots\cap$“第 k 次正面”，

则 $P(\text{首次正面出现在 } k \text{ 次})=P(\text{第 1 次反面})P(\text{第 2 次反面})\cdots P(\text{第 } k \text{ 次正面})=\frac{1}{2^k}$。