條件機率(Conditional Probability)與貝氏定理(Bayes' theorem)

條件機率(Conditional Probability)

> 在某個事件確實已經發生的前提下，另一個事件發生的機率，所以也稱為事後機率(Posterior probability)

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

分子:A和B兩個事件同時發生的機率

分母:B事件發生的機率，也稱為事前機率(Prior probability)

例子(條件機率)：

假設有個群體總數100人，男生有40人，女生有60人，而男生中長髮的人數為10人，短髮為30人，女生長髮有40人，短髮有20人，假設A事件是長髮，B事件是女生，那麼當你隨機從這100人抽樣1人，以下機率是如何？

1. 是長髮的機率?
2. 是女生的機率?
3. 是女生且為長髮的機率?
4. 發現他是長髮，那他是女生的機率?
5. 發現他是女生，那他是長髮的機率?

計算：

可以先將資訊做成cross table方便計算，table如下

事件	男生	女生	總數
短髮	30	20	50
長髮	10	40	50
總數	40	60	100

1. 是長髮的機率?

$$P(A) = \frac{50}{100} = 0.5$$

1. 是女生的機率?

$$P(B) = \frac{60}{100} = 0.6$$

1. 是女生且為長髮的機率?

$$P(A\ \cap\ B) = \frac{40}{100} = 0.4$$

1. 發現他是長髮，那他是女生的機率?

$$P(B|A) = \frac{P(A\ \cap\ B)}{P(A)} = \frac{0.4}{0.5} = 0.8$$

1. 發現他是女生，那他是長髮的機率?

$$P(A|B) = \frac{P(A\ \cap\ B)}{P(B)} = \frac{0.4}{0.6} = \frac{2}{3}$$

貝氏定理(Bayes' theorem)

> 和條件機率一樣是表達在某個事件發生的前提下，另一個事件發生的機率，但又在做更近一層的推廣

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

更一般化的表達是

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots P(B|A_n)P(A_n)}$$

$$= \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n}[P(B|A_k)P(A_k)]}$$

Remark:

1. 由上式可看出，原本要求B發生的前提下A_i發生的機率，計算上可以轉換成A_i發生的前提下發生B的方法計算
2. 分母是一種全機率的展開，也就是事件B發生的機率是所有$A_i, i = 1 \sim n$發生的前提下發生B的機率總和，其中$P(A_i\ \cap\ A_j)$在i≠j的情況下是空集合，完整的表達如下：

$$\Omega = [A_1, A_2, \cdots A_n]$$

$$P(A_i\ \cap\ A_j) = \emptyset$$

$$P(A_1) + P(A_2) + \cdots P(A_n) = P(A_1\ \cup\ A_2\ \cup\ \cdots A_n) = P(\Omega) = 1$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots P(B|A_n)P(A_n) =$$

$$\sum_{k=1}^{n}[P(B|A_k)P(A_k)]$$

例子1(貝氏定理)：同上題4

假設有個群體總數100人，男生有40人，女生有60人，而男生中長髮的人數為10人，短髮為30人，女生長髮有40人，短髮有20人

1. 發現他是長髮，那他是女生的機率?

計算：

事件	男生(B_2)	女生(B_1)	總數
短髮(A_2)	30	20	50
長髮(A_1)	10	40	50
總數	40	60	100

假設

A_1 =長髮， A_2 =短髮;

B_1 =女生， B_2 =男生

1. 發現他是長髮，那他是女生的機率?

$$P(B_1|A_1) = \frac{P(A_1|B_1)P(B_1)}{P(A_1|B_1)P(B_1) + P(A_1|B_2)P(B_2)}$$

$$P(B_1) = \frac{60}{100}$$

$$P(A_1|B_1) = \frac{P(A_1 \cap B_1)}{P(B_1)} = \frac{40/100}{60/100} = \frac{2}{3}$$

$$P(B_2) = \frac{40}{100}$$

$$P(A_1|B_2) = \frac{P(A_1 \cap B_2)}{P(B_2)} = \frac{10/100}{40/100} = \frac{1}{4}$$

最終計算結果如下：

$$P(B_1|A_1) = \frac{\frac{2}{3}\frac{60}{100}}{\frac{2}{3}\frac{60}{100} + \frac{1}{4}\frac{40}{100}} = 0.8$$

例子2(貝氏定理):

假設某個群體有流感的人數比例為0.05，而某快篩試劑廠商宣稱當受測者有流感時，驗出陽性的機率為0.95，當受測者為沒有流感時，驗出陰性的機率為0.9，那麼隨機抽樣一人，驗出陽性時，他也真的有流感的機率是多少？那麼快篩檢測的正確率(TP+TN)又是多少？

計算：

假設

A_1=有流感

A_2=無流感

B_1=陽性

B_2=陰性

那麼廠商宣稱的效果如下：

$$P(B_1|A_1) = 0.95 -> P(B_2|A_1) = 0.05$$

$$P(B_2|A_2) = 0.9 -> P(B_1|A_2) = 0.1$$

最後所求為：

$$P(A_1|B_1) = \frac{P(B_1|A_1)P(A_1)}{P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2)}$$

$$= \frac{0.95 \times 0.05}{0.95 \times 0.05 + 0.1 \times 0.95} = \frac{0.0475}{0.0475 + 0.095} = 0.33$$

若將正確率定義成，患者有流感的情況能檢測出陽性(True Positive,TP)+患者沒有流感的情況檢測為陰性(True Negative,TN)，則計算總和為

$$P(A_1 \cap B_1) + P(A_2 \cap B_2) =$$

$$P(B_1|A_1)P(A_1) + P(B_2|A_2)P(A_2) = 0.95 \times 0.05 + 0.9 \times 0.95 = 0.9025$$

也就是有**九成的正確率**