4.5. Bayes' theorem:

[Intuition of Bayes' Theorem]:

Prior Probability $\mathrm{P(A)}$	➡	New Information (Event B)	➡	Application of Bayes' Theorem	➡	Posterior Probability $P(A|B)$

Example 14:

B: test (positive), B^c: test (negative)

A: no AIDS, A^c: AIDS.

From past experience and records, we know

$$P(A) = 0.99, P(B|A) = 0.03, P(B|A^c) = 0.98.$$

That is, we know the probability of a patient having no AIDS, the conditional probability of test positive given having no AIDS (wrong diagnosis), and the conditional probability of test positive given having AIDS (correct diagnosis). Our object is to find $P(A|B)$, i.e., we want to know the probability of a patient having not AIDS even known that this patient is test positive.

Example 15:

A_1: the finance of the company being good.

A_2: the finance of the company being O.K.

A_3: the finance of the company being bad.

B_1: good finance assessment for the company.

B_2: O.K. finance assessment for the company.

B_3: bad finance assessment for the company.

From the past records, we know

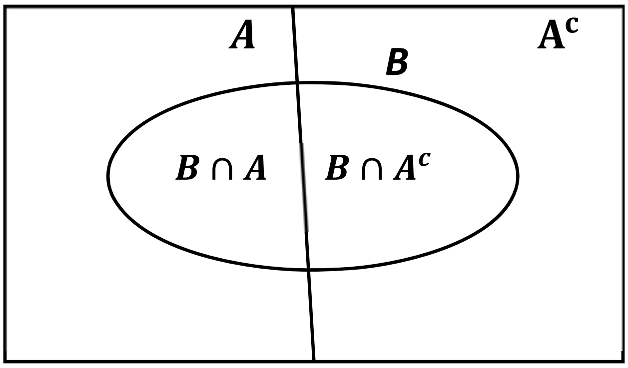
$$P(A_1) = 0.5, P(A_2) = 0.2, P(A_3) = 0.3,$$
$$P(B_1|A_1) = 0.9, P(B_1|A_2) = 0.05, P(B_1|A_3) = 0.05.$$

That is, we know the chances of the different finance situations of the company and the conditional probabilities of the different assessments for the company given the finance of the company known, for example, $P(B_1|A_1) = 0.9$ indicates 90% chance of good finance year of the company has been predicted correctly by the finance assessment. Our objective is to obtain the probability $\mathrm{P}(A_1|B_1)$, i.e., the conditional probability that the finance of the company being good in the coming year given that good finance assessment for the company in this year. To find the required probability in the above two examples, the following Bayes's theorem can be used.

Bayes' theorem (two events):

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

[Derivation of Bayes's theorem (two events)]:



Since $P(B \cap A) = P(A)P(B|A)$ **and**

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

thus

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

Example 14 (continue):

$P(A^c) = 1 - P(A) = 1 - 0.99 = 0.01$. **Then, by Bayes' theorem,**

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{0.99 \cdot 0.03}{0.99 \cdot 0.03 + 0.01 \cdot 0.98}$$
$$= 0.7519.$$

⇒ A patient with test positive still has high probability (0.7519) of no AIDS.

Bayes' Theorem (general):

Let $A_1, A_2, \cdots, A_n$ **be mutually exclusive events and** $A_1 \cup A_2 \cup \cdots \cup A_n = S$**, then**

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

for $i = 1, 2, \cdots, n$.

[Derivation of Bayes' theorem (general)]:

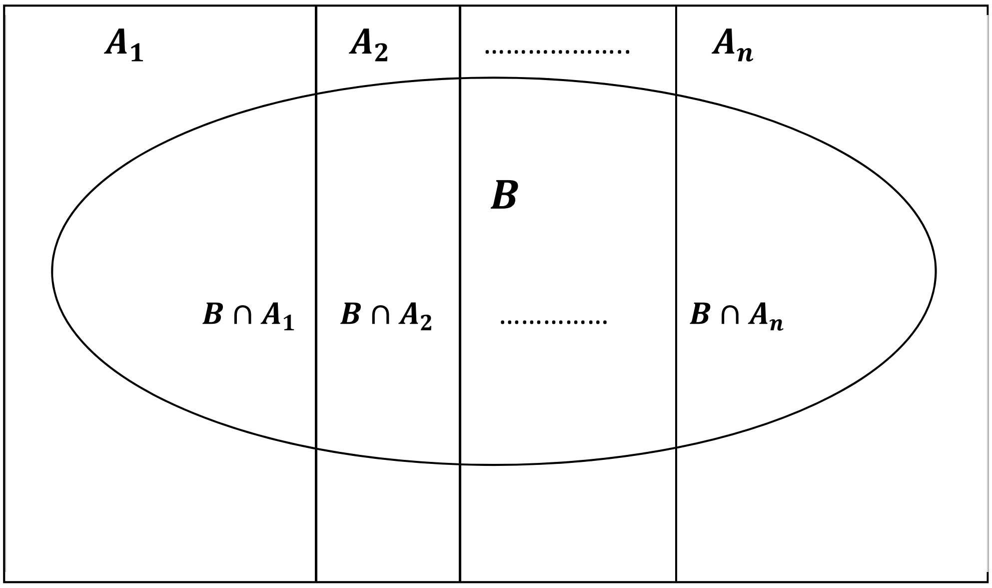
Since $P(B \cap A_i) = P(A_i)P(B|A_i)$ **and**

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

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

thus

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



Example 15 (continue):

$$P(A_1|B_1) = \frac{P(A_1)P(B_1|A_1)}{P(A_1)P(B_1|A_1) + P(A_2)P(B_1|A_2) + P(A_3)P(B_1|A_3)}$$

$$= \frac{0.5 \cdot 0.9}{0.5 \cdot 0.9 + 0.2 \cdot 0.05 + 0.3 \cdot 0.05} = 0.95.$$

$\Longrightarrow$ A company with good finance assessment has very high probability (0.95) of good finance situation in the coming year.

Example 16:

In a recent survey in a Statistics class, it was determined that only 60% of the students attend class on Thursday. From past data it was noted that 98% of those who went to class on Thursday pass the course, while only 20% of those who did not go to class on Thursday passed the course.

(a) What percentage of students is expected to pass the course?

(b) Given that a student passes the course, what is the probability that he/she attended classes on Thursday.

[Solution:]

A: the students attend class on Thursday

A^c: the students do not attend class on Thursday

B: the students pass the course

B^c: the students do not pass the course

Then, $\mathrm{P}(A) = 0.6, P(A^c) = 0.4, P(B|A) = 0.98, P(B|A^c) = 0.2$.

(a)

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(A)P(B|A) + P(A^c)P(B|A^c) \\ &= 0.6 \cdot 0.98 + 0.4 \cdot 0.2 \\ &= 0.668 \end{aligned}$$

(b) By Bayes' theorem,

$$\begin{aligned} P(A|B) &= \frac{P(B \cap A)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\ &= \frac{0.6 \cdot 0.98}{0.6 \cdot 0.98 + 0.4 \cdot 0.2} \\ &= 0.854 \end{aligned}$$

Example 17:

Assume that a patient is believed to have one of two diseases, denoted D_1 and D_2 with $P(D_1) = 0.3$ and $P(D_2) = 0.7$. Suppose that, given diseases D_1 and D_2, the probabilities that the patients will have symptoms S_1, S_2, S_3, and S_4 are as follows.

	S_1	S_2	S_3	S_4
D_1	$P(S_1\|D_1) = 0.8$	$P(S_2\|D_1) = 0.05$	$P(S_3\|D_1) = 0.05$	$P(S_4\|D_1) = 0.03$
D_2	$P(S_1\|D_2) = 0.15$	$P(S_2\|D_2) = 0.6$	$P(S_3\|D_2) = 0.05$	$P(S_4\|D_2) = 0.05$

Compute the posterior probabilities of each disease given the patient has symptom S_2 and determine which disease the patient might have.

[Solution:]

$$\begin{aligned} P(D_1|S_2) &= \frac{P(D_1 \cap S_2)}{P(S_2)} = \frac{P(D_1)P(S_2|D_1)}{P(D_1)P(S_2|D_1) + P(D_2)P(S_2|D_2)} \\ &= \frac{0.3 \cdot 0.05}{0.3 \cdot 0.05 + 0.7 \cdot 0.6} \\ &= 0.0344 \end{aligned}$$

and

$$P(D_2|S_2) = \frac{P(D_2 \cap S_2)}{P(S_2)} = \frac{P(D_2)P(S_2|D_2)}{P(D_1)P(S_2|D_1) + P(D_2)P(S_2|D_2)}$$

$$= \frac{0.7 \cdot 0.6}{0.3 \cdot 0.05 + 0.7 \cdot 0.6}$$
$$= 0.9656.$$

$\Longrightarrow$ **the patient might have disease** D_2.

Note: $P(D_2|S_2) = 1 - P(D_1|S_2) = 1 - 0.0344 = 0.9656.$

Example 18:

An oil company has purchased an option on land in Alaska. Preliminary geologic studies have assigned the following prior probabilities.

$$P(high - quality\ \ oil) = 0.2, P(medium - quality\ \ oil) = 0.3,$$
$$P(no\ \ oil) = 0.5.$$

After 200 feet of drilling on the first well, a soil test is taken. The probabilities of finding particular type of soil identified by the test follow.

$$P(soil|high - quality\ \ oil) = 0.6, P(soil|medium - quality\ \ oil) = 0.8,$$
$$P(soil|no\ \ oil) = 0.1.$$

What is the probability of finding oil given finding particular type of soil ?

[Solution:]

A_1: high quality oil;

A_2: medium quality oil;

A_3: no oil;

$B =$ soil test.

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)}$$
$$= \frac{0.5 \cdot 0.1}{0.2 \cdot 0.6 + 0.3 \cdot 0.8 + 0.5 \cdot 0.1} = \frac{5}{41}$$

and hence

$$P(A_1 \cup A_2|B) = 1 - P(A_3|B) = \frac{36}{41}.$$