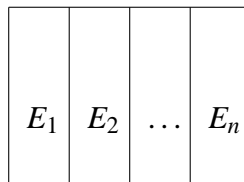## The Theorem of Total Probability

Sometimes we are faced with a situation where we do not know the probability of an event A, but we know what its probability would be if we were sure that some other event had occurred.

Example An ice-cream seller has to decide whether to order more stock for the Bank Holiday weekend. He estimates that, if the weather is sunny, he has a 90% chance of selling all his stock; if it is cloudy, his chance is 60%; and if it rains, his chance is only 20%. According to the weather forecast, the probability of sunshine is 30%, the probability of cloud is 45%, and the probability of rain is 25%. (We assume that these are all the possible outcomes, so that their probabilities must add up to 100%.) What is the overall probability that the salesman will sell all his stock?

This problem is answered by the *Theorem of Total Probability*, which we now state. First we need a definition. The events E_1, E_2, ..., E_n form a *partition* of the sample space if the following two conditions hold:

(a) the events are pairwise disjoint, that is, $E_i \cap E_j = \emptyset$ for any pair of events E_i and E_j;

(b) $E_1 \cup E_2 \cup \cdots \cup E_n = \mathcal{S}$.

Another way of saying the same thing is that every outcome in the sample space lies in exactly one of the events E_1, E_2, ..., E_n. The picture shows the idea of a partition.

Now we state and prove the Theorem of Total Probability.

Theorem of Total Probability Let E_1, E_2, ..., E_n form a partition of the sample space with $P(E_i) \neq 0$ for all i, and let A be any event. Then

(i) $P(A) = P(E_1) \times P(A \mid E_1) + P(E_2) \times P(A \mid E_2) + \cdots + P(E_n) \times P(A \mid E_n)$.

(ii) Furthermore, if $P(A) \neq 0$ then

$$P(E_1 \mid A) + P(E_2 \mid A) + \cdots + P(E_n \mid A) = 1.$$

Proof Consider the events $A \cap E_1$, $A \cap E_2$, ..., $A \cap E_n$. These events are pairwise disjoint; for any outcome lying in both $A \cap E_i$ and $A \cap E_j$ would lie in both E_i and E_j, and by assumption there are no such outcomes. Moreover, the union of all these events is A, since every outcome lies in one of the E_i. So, by Axiom 3a, we conclude that

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i). \qquad (**)$$

But $P(A \cap E_i) = P(E_i) \times P(A \mid E_i)$. Substituting into $(**)$ gives

$$P(A) = \sum_{i=1}^{n} P(E_i) \times P(A \mid E_i),$$

which is (i).
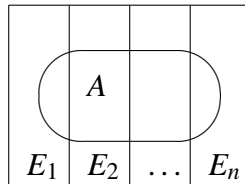
Now, dividing $(**)$ by $P(A)$ gives

$$1 = \sum_{i=1}^{n} \frac{P(A \cap E_i)}{P(A)} = \sum_{i=1}^{n} P(E_i \mid A),$$

giving (ii). ∎

One special case of the Theorem of Total Probability is very commonly used, and is worth stating in its own right. For any event A, the events A and A' form a partition of $\mathcal{S}$. To say that both A and A' have non-zero probability is just to say that $P(A) \neq 0, 1$. Thus we have the following corollary:

Corollary 1 Let A and B be events, and suppose that $P(A) \neq 0, 1$. Then

$$P(B) = P(A) \times P(B \mid A) + P(A') \times P(B \mid A').$$

You will now realise that the Theorem of Total Probability is really being used when you calculate probabilities by tree diagrams. It is better to get into the habit of using it directly, since it avoids any accidental assumptions of independence.

More about conditional probability

Example (Ice-cream: part 2) Consider the ice-cream salesman in our example. Let A_1 be the event 'it is sunny', A_2 the event 'it is cloudy', and A_3 the event 'it is rainy'. Then A_1, A_2 and A_3 form a partition of the sample space, and we are given that

$$P(A_1) = 0.3, \quad P(A_2) = 0.45, \quad P(A_3) = 0.25.$$

Let B be the event 'the salesman sells all his stock'. The other information we are given is that

$$P(B \mid A_1) = 0.9, \quad P(B \mid A_2) = 0.6, \quad P(B \mid A_3) = 0.2.$$

By the Theorem of Total Probability,

$$P(B) = (0.9 \times 0.3) + (0.6 \times 0.45) + (0.2 \times 0.25) = 0.59.$$

Example (Eye-colour: part 2) We know that $P(D) = 1/3$ and $P(D') = 2/3$, where D is the event 'John's genes are BB'. Now John marries Jill, who has blue eyes. What is the probability that their first child has brown eyes?

Solution Jill's genes must be bb, so the gene that the child gets from Jill must be b. If John's genes are BB then the child will get a B gene from him, so the child will be Bb and have brown eyes. If John is Bb then the gene that the child gets from him is equally likely to be B or b, so the child is equally likely to be Bb or bb; in the first case the child has brown eyes, in the second case blue. Thus if we put $C =$ 'the child has brown eyes' then $P(C \mid D) = 1$ and $P(C \mid D') = 1/2$. Now

$$
\begin{aligned}
P(C) &= P(D)P(C \mid D) + P(D')P(C \mid D') \\
&= \frac{1}{3} \times 1 + \frac{2}{3} \times \frac{1}{2} = \frac{2}{3}.
\end{aligned}
$$

3

Example A shop buys loaves of bread from three bakeries:

30% from bakery A	10% of A's loaves are underweight
50% from bakery B	2% of B's loaves are underweight
20% from bakery C	15% of C's loaves are underweight

A loaf is chosen at random in the shop. What is the probability that the loaf is underweight?

Solution Let U be the event that the loaf is underweight. Then

$$
\begin{aligned}
P(U) &= P(A)P(U\,|\,A) + P(B)P(U\,|\,B) + P(C)P(U\,|\,C) \\
&= 0.3 \times 0.1 + 0.5 \times 0.02 + 0.2 \times 0.15 = 0.07.
\end{aligned}
$$

Sampling revisited

Suppose that we sample from N objects without replacement, and that m of the objects are red. Let A be 'first object is red' and B be 'second object is red'. Some people think that $P(B)$ is obviously the same as $P(A)$, while others are very suspicious about it. Here is the argument, using the Theorem of Total Probability.

$$
P(A) = \frac{m}{N} \qquad P(A') = \frac{N-m}{N}
$$

$$
P(B\,|\,A) = \frac{m-1}{N-1} \qquad P(B\,|\,A') = \frac{m}{N-1}
$$

So

$$
\begin{aligned}
P(B) &= P(A) \times P(B\,|\,A) + P(A') \times P(B\,|\,A') \\
&= \frac{m}{N} \times \left(\frac{m-1}{N-1} \right) + \left(\frac{N-m}{N} \right) \times \left(\frac{m}{N-1} \right) \\
&= \frac{m}{N(N-1)}[m-1+N-m] \\
&= \frac{m(N-1)}{N(N-1)} = \frac{m}{N}.
\end{aligned}
$$

Bayes' Theorem

There is a very big difference between $P(A \mid B)$ and $P(B \mid A)$.

For example, in a murder trial, an expert witness who has carried out a DNA test is called. He testifies that, given that the suspect is innocent, the chance of getting a positive test result is one in a million. Rather than say 'he must be guilty then', we have to realise that we are not interested in this value, but in the probability that the suspect is innocent given the positive test result, which is completely different. [Regrettably, the House of Lords has ruled that juries are unable to understand this difference, and it must not be explained, for fear of confusing them.]

Another place where the difference occurs is in medicine. Suppose that a new test is developed to identify people who are liable to suffer from some genetic disease in later life. Of course, no test is perfect; there will be some carriers of the defective gene who test negative, and some non-carriers who test positive. So, for example, let A be the event 'the patient is a carrier', and B the event 'the test result is positive'.

The scientists who develop the test are concerned with the probabilities that the test result is wrong, that is, with $P(B \mid A')$ (the probability of a false positive result on someone who doesn't have the gene) and $P(B' \mid A)$ (the probability of a false negative result for someone who does have it).

However, a patient who has taken the test has different concerns. If I tested positive, what is the chance that I have the disease? If I tested negative, how sure can I be that I am not a carrier? In other words, $P(A \mid B)$ and $P(A' \mid B')$.

These conditional probabilities are related by *Bayes' Theorem*:

Bayes' Theorem Let A and B be events with non-zero probability. Then

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)}.$$

The proof is not hard. We have

$$P(B \mid A) \times P(A) = P(A \cap B) = P(A \mid B) \times P(B),$$

using the definition of conditional probability twice. (Note that we need both A and B to have non-zero probability here.) Now divide this equation by $P(A)$ to get the result.

If $P(B) \neq 0, 1$ and $P(A) \neq 0$, then we can use the corollary to the Theorem of Total Probability to write this as

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A \mid B) \times P(B) + P(A \mid B') \times P(B')}.$$

Bayes' Theorem is often stated in this form.

Example (Ice-cream: part 3) Consider the ice-cream salesman. Given that he sold all his stock of ice-cream, what is the probability that the weather was sunny? (This question might be asked by the warehouse manager who doesn't know what the weather was actually like.) In the notation that we used before, A_1 is the event 'it is sunny' and B the event 'the salesman sells all his stock'. We are asked for $P(A_1 \mid B)$. We were given that $P(B \mid A_1) = 0.9$ and that $P(A_1) = 0.3$, and we calculated that $P(B) = 0.59$. So by Bayes' Theorem,

$$P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B)} = \frac{0.9 \times 0.3}{0.59} = 0.46$$

to 2 d.p.

Example (Bakery: part 2) If a randomly chosen loaf is found to be underweight, what is the probability that it came from bakery C?

This is just $P(C \mid U)$, and Bayes' Theorem gives

$$P(C \mid U) = \frac{P(U \mid C) \times P(C)}{P(U)} = \frac{0.15 \times 0.2}{0.07} = \frac{3}{7}.$$

Example (Eye-colour: part 3) Now suppose that John and Jill have four children, all brown-eyed. What is our new assessment of the probability that John's genes are BB?

Solution Let F be the event that all four children are brown-eyed. The children inherit genes independently of each other, so

$$P(F \mid D) = 1$$

and

$$P(F \mid D') = \left(\frac{1}{2}\right)^4 = \frac{1}{16}.$$

The Theorem of Total Probability gives:

$$P(F) = P(D)P(F \mid D) + P(D')P(F \mid D') = \frac{1}{3} \times 1 + \frac{2}{3} \times \frac{1}{16} = \frac{3}{8}.$$

Then Bayes' Theorem gives

$$P(D \mid F) = \frac{P(F \mid D)P(D)}{P(F)} = \frac{1/3}{3/8} = \frac{8}{9}.$$

We can interpret this conditional probability as a measure of our belief that John is BB given all the information available.

Example Consider the clinical test described earlier. Suppose that 1 in 1000 of the population is a carrier of the disease. Suppose also that the probability that a carrier tests negative is 1%, while the probability that a non-carrier tests positive is 5%. (A test achieving these values would be regarded as very successful.) Let A be the event 'the patient is a carrier', and B the event 'the test result is positive'. We are given that $P(A) = 0.001$ (so that $P(A') = 0.999$), and that

$$P(B \mid A) = 0.99, \qquad P(B \mid A') = 0.05.$$

(a) A patient has just had a positive test result. What is the probability that the patient is a carrier? The answer is

$$
\begin{aligned}
P(A \mid B) &= \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A')P(A')} \\
&= \frac{0.99 \times 0.001}{(0.99 \times 0.001) + (0.05 \times 0.999)} \\
&= \frac{0.00099}{0.05094} = 0.0194.
\end{aligned}
$$

(b) A patient has just had a negative test result. What is the probability that the patient is a carrier? The answer is

$$
\begin{aligned}
P(A \mid B') &= \frac{P(B' \mid A)P(A)}{P(B' \mid A)P(A) + P(B' \mid A')P(A')} \\
&= \frac{0.01 \times 0.001}{(0.01 \times 0.001) + (0.95 \times 0.999)} \\
&= \frac{0.00001}{0.94906} = 0.00001.
\end{aligned}
$$

So a patient with a negative test result can be reassured; but a patient with a positive test result still has less than 2% chance of being a carrier, so is likely to worry unnecessarily.

Of course, these calculations assume that the patient has been selected at random from the population. If the patient has a family history of the disease, the calculations would be quite different.

Example 2% of the population have a certain blood disease in a serious form; 10% have it in a mild form; and 88% don't have it at all. A new blood test is developed; the probability of testing positive is 9/10 if the subject has the serious form, 6/10 if the subject has the mild form, and 1/10 if the subject doesn't have the disease.

I have just tested positive. What is the probability that I have the serious form of the disease?

Solution Let A_1 be 'has disease in serious form', A_2 be 'has disease in mild form', and A_3 be 'doesn't have disease'. Let B be 'test positive'. Then we are given that A_1, A_2, A_3 form a partition and

$$P(A_1) = 0.02 \qquad P(A_2) = 0.1 \qquad P(A_3) = 0.88$$
$$P(B \mid A_1) = 0.9 \quad P(B \mid A_2) = 0.6 \quad P(B \mid A_3) = 0.1$$

Thus, by the Theorem of Total Probability,

$$P(B) = 0.9 \times 0.02 + 0.6 \times 0.1 + 0.1 \times 0.88 = 0.166,$$

and then by Bayes' Theorem,

$$P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B)} = \frac{0.9 \times 0.02}{0.166} = 0.108$$

to 3 d.p.

Proposition 6 Let E_1, E_2, ..., E_n be pairwise disjount events whose union is the whole sample space. Let A and B be events with $P(B \cap E_i) > 0$ for $i = 1, \ldots, n$. Then

$$P(A \mid B) = \sum_{i=1}^{n} [P(A \mid B \cap E_i) \times P(E_i \mid B)].$$

Proof

$$
\begin{aligned}
P(A \mid B) &= \frac{P(A \cap B)}{P(B)}, & \text{by the definition of conditional probability,} \\[2mm]
&= \sum_{i=1}^{n} \frac{P(A \cap B \mid E_i) \times P(E_i)}{P(B)}, & \text{by the Theorem of Total Probability for } A \cap B, \\[2mm]
&= \sum_{i=1}^{n} \left[\frac{P(A \cap B \cap E_i)}{P(E_i)} \times \frac{P(E_i)}{P(B)} \right], & \text{by the definition of conditional probability,} \\[2mm]
&= \sum_{i=1}^{n} \left[\frac{P(A \cap B \cap E_i)}{P(B)} \right] \\[2mm]
&= \sum_{i=1}^{n} \left[\frac{P(A \cap B \cap E_i)}{P(B)} \times \frac{P(B \cap E_i)}{P(B \cap E_i)} \right], & \text{because multiplication by 1 changes nothing!} \\[2mm]
&= \sum_{i=1}^{n} \left[\frac{P(A \cap B \cap E_i)}{P(B \cap E_i)} \times \frac{P(B \cap E_i)}{P(B)} \right], \\[2mm]
&= \sum_{i=1}^{n} [P(A \mid B \cap E_i) \times P(E_i \mid B)]. \quad \blacksquare
\end{aligned}
$$

8

Example Two channels, C and D, are used to send electronic information as a sequence of 0s and 1s. The probability of correct transmission is

	C	D
1 sent	0.95	0.99
0 sent	0.98	0.91

In each channel, $2/3$ of the signals are 1s. Also, channel C sends $3/5$ of the traffic. Find the probability that a 1 is received when a single item is sent.

Solution Put

$$S_1 = \text{``1 is sent''} \qquad S_0 = \text{``0 is sent''}$$
$$R_1 = \text{``1 is received''} \quad R_0 = \text{``0 is received''}$$

Then we want $P(R_1)$.

Use S_0 and S_1 as the partition. Then

$$
\begin{aligned}
P(R_1 \mid C) &= P(R_1 \mid S_1 \cap C)P(S_1 \mid C) + P(R_1 \mid S_0 \cap C)P(S_0 \mid C) \\
&= 0.95 \times \frac{2}{3} + 0.02 \times \frac{1}{3} = 0.64
\end{aligned}
$$

and

$$
\begin{aligned}
P(R_1 \mid D) &= P(R_1 \mid S_1 \cap D)P(S_1 \mid D) + P(R_1 \mid S_0 \cap D)P(S_0 \mid D) \\
&= 0.99 \times \frac{2}{3} + 0.09 \times \frac{1}{3} = 0.69,
\end{aligned}
$$

so

$$
\begin{aligned}
P(R_1) &= P(R_1 \mid C)P(C) + P(R_1 \mid D)P(D) \\
&= 0.64 \times \frac{3}{5} + 0.69 \times \frac{2}{5} = 0.66.
\end{aligned}
$$

Can you find what is the probability that the signal is sent correctly if a 1 is received?