Lecture 4

Bayes' Theorem

By
Dr Sean Maudsley-Barton and Abdul Ali

Manchester Metropolitan University

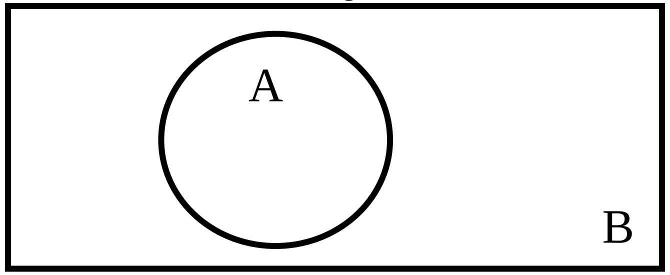
Aims

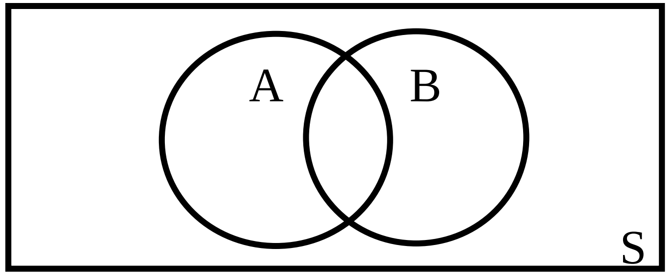
- Understand and apply the theorem of total probability

- Understand and apply Bayes’ theorem

Recap - Conditional Probability

- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$, when $P(B) > 0$





- $\boxed{P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)}$
- $\boxed{P(A^c|B) + P(A|B) = 1}$

Total probability and Bayes' Theorem

Manchester Metropolitan University

Total Probability

We illustrate the concept of total probability in the following theorem,

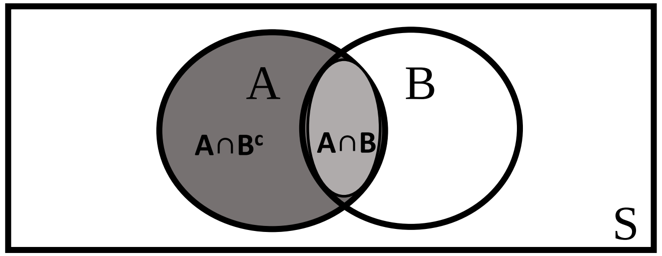
Theorem 4.1 Given A, B $\subset$ S and P(B) > 0, *then*

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$\boxed{= P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

	P(A)	P(A^c)	Total
P(B)	P($A \cap B$)	P($A^c \cap B$)	$\boldsymbol{P(B)}$
P(B^c)	P($A \cap B^c$)	P($A^c \cap B^c$)	$\boldsymbol{P(Bc)}$
Total	$\boldsymbol{P(A)}$	$\boldsymbol{P(Ac)}$	**1**

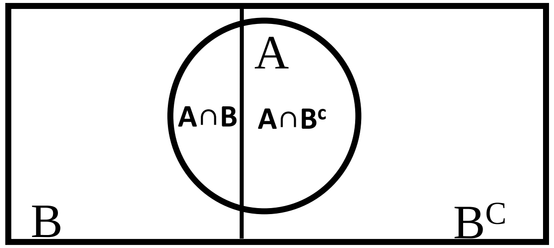
The Proof/Demonstration is:

P(A)



Since $B \cap B^c = \emptyset$, *then* $(A \cap B) \cap (A \cap B^c) = \emptyset$. *Also,*

$$P(A) = P((A \cap B) \cup (A \cap B^c))$$
$$= P(A \cap B) + P(A \cap B^c) \quad \text{\{by Kolmogorov's \textbf{third} axiom\}}$$
$$= P(A \mid B)P(B) + P(A \mid B^c)P(B^c) \quad \text{\{by Conditional formula\}}$$



We can now extend the concept of conditional probability to a general situation

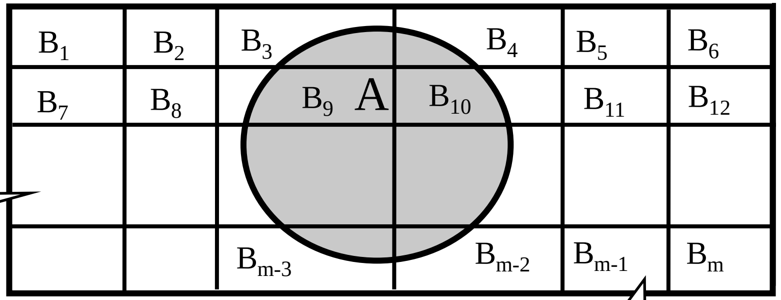
Manchester Metropolitan University

Theorem 4.2 (*Total Probability*)

Let S be the sample space. Suppose the events $B_1, B_2, \ldots, B_m$ are mutually exclusive and exhaustive, that is:

$$B_1 \cup \cdots \cup B_m = S \text{ and } P(B_i \cap B_j) = 0 \text{ for all } i \neq j.$$

Then, for any event A of S,



Exhaustive, Covers the whole space

Mutually exclusive, the Bs do not overlap

$$\boxed{P(A) = P(A|\ B_1)P(B_1) + P(A|\ B_2)P(B_2) + \cdots + P(A|\ B_m)P(B_m)}$$

Or

$$\boxed{P(A) = \sum_{k=1}^{m} P(A|B_k)P(B_k)}$$

The events $B_1, B_2, \ldots, B_m$ are said to form a **partition** of S.

Total probability

Manchester Metropolitan University

Example 4.1 (*Disease prevalence*)

- Suppose **1%** of a population carries a disease
- A blood test, detects the disease **85%** of the time P(+ve test |Carrying the disease)
- It also detects the lack of disease **85%** of the time P(-ve test |Not Carrying the disease)

True positive rate

True negative rate

a) What is the probability that a randomly selected person, will test positive?

b) If a person's test is positive, what is the probability they are a carrier?

Total probability

Manchester Metropolitan University

Example 4.1 (*Disease prevalence*)

a) What is the probability that the blood test of a person selected at random will test positive?

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

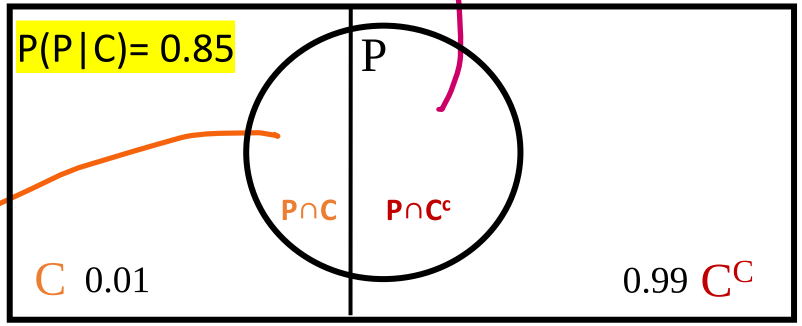
C: Carrier $\quad$ **P(C)** $= 0.01$

P: Positive $\quad$ **P(P|C)** $= 0.85$ TP

N: Negative $\quad$ **P(N|C^c)** $= 0.85$ TN

Probability of being +ve, regardless of them being a carrier or not. i.e. add the two halves that makeup P(P)

$$
\begin{aligned}
P(P) &= P(P|C)P(C) + P(P|C^C)\,P(C^C) \\
&= P(P|C)P(C) + [1 - P(N|C^C)]\,P(C^C) \\
&= (0.85)(0.01) + (1 - 0.85)(1 - 0.01) \\
&= (0.85)(0.01) + (0.15)(0.99) \\
&= 0.0085 + 0.1485 \\
&= 0.157
\end{aligned}
$$



Total probability and Bayes' Theorem

b) If a person's test is positive, what is the probability they are a carrier? $P(C|P)$

C:Carrier   **P(C)** = 0.01

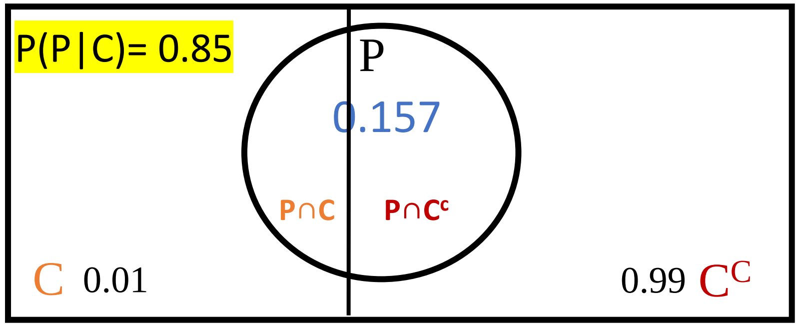
P: Positive   **P(P|C)** = 0.85

N: Negative   **P(N|C)** = 0.85

P(P) = 0.157

Remember:

$P(C|P)P(P) = P(C \cap P) = P(P \cap C) = P(P|C)P(C)$



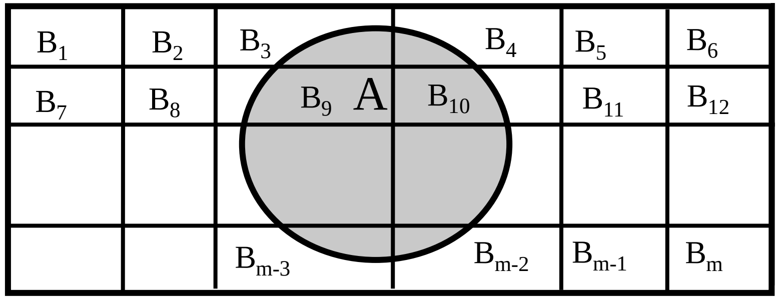
$$P(C|P) = \frac{P(C \cap P)}{P(P)}$$

$$= \frac{P(P|C)P(C)}{P(P)}$$

$$= \frac{(0.85)(0.01)}{(0.157)}$$

$$= 0.054$$

Total probability and Bayes' Theorem



Bayes' Theorem

Calculated the probability of belonging to a partition $P(Bk|A)$

Using the results above, and the definition of **total probability**, we can derive the following theorem,

Theorem 4.3 (*Bayes' Theorem*)

If events $B_1, B_2, . . . , B_m$, are mutually **exclusive** and **exhaustive**, then, for any event A, we have,

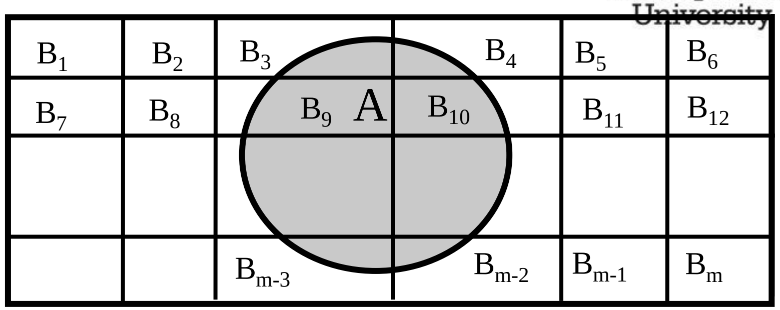
$A \cap B_k$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{k=1}^{m} P(A|B_k)P(B_k)}, \qquad k = 1,2, \dots, m$$

One of the partitions

Total Probability

Total probability and Bayes' Theorem



Bayes' Theorem

The Proof/Demonstration is:

$$P(A \cap B_k) = P(B_k|A)P(A)$$

$$= P(A|B_k)P(B_k)$$

{Hint: $P(A \cap B_k) = P(B_k \cap A)$ }

so that,

$$P(B_k|A)P(A) = P(A|B_k)P(B_k)$$

Hence,

$$\boxed{P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)}}$$

as required.

where $P(A) = \sum_{k=1}^{m} P(A|B_k)P(B_k)$ {using Theorem 4.2.}

Total probability and Bayes' Theorem

Manchester Metropolitan University

Example 4.2 (*Production Faults*)

A company produces electrical components using three shifts. During the first shift **50%** of components are produced with **20%** and **30%** being produced on shifts 2 and 3, respectively. The proportion of defective components produced during shift 1 is **6%.** For shifts 2 and 3 the proportions are **8%** and **12%** respectively.

a) Find the percentage of defective components.

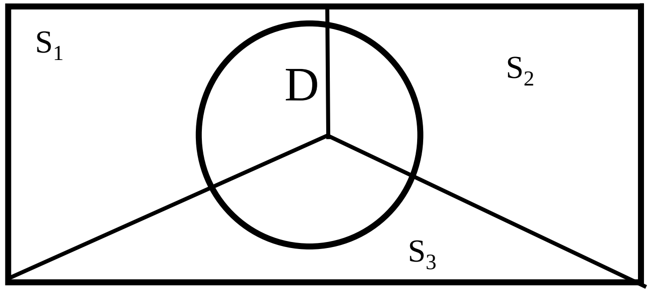
b) If a component is defective, what is the probability it came from shift 3?

Total probability and Bayes’ Theorem

Example 4.2 (*Production Faults*)

A company produces electrical components using three shifts. During the first shift **50%** of components are produced with **20%** and **30%** being produced on shifts 2 and 3, respectively. The proportion of defective components produced during shift 1 is **6%.** For shifts 2 and 3 the proportions are **8%** and **12%** respectively.

a) Find the percentage of defective components.



Let the event D denote that a component is defective and S_1, S_2, S_3 denote that it was produced during shifts 1,2 or 3, respectively.

We use the theorem of total probability.

$$P(D) = P(D|S_1)P(S_1) + P(D|S_2)P(S_2) + P(D|S_3)P(S_3)$$

$$= 0.06 \times 0.5 \quad + 0.08 \times 0.2 \quad + 0.12 \times 0.3$$

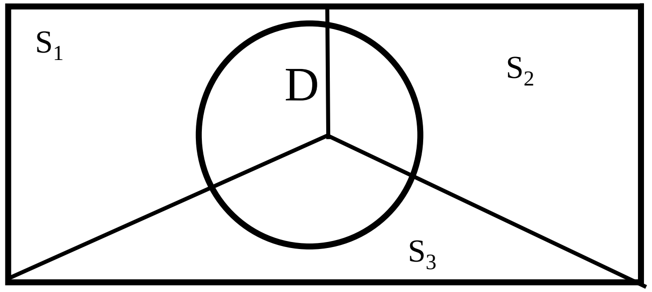
$$= 0.082$$

Total probability and Bayes' Theorem

Example 4.2 (*Production Faults*)

A company produces electrical components using three shifts. During the first shift **50%** of components are produced with **20%** and **30%** being produced on shifts 2 and 3, respectively. The proportion of defective components produced during shift 1 is **6%.** For shifts 2 and 3 the proportions are **8%** and **12%** respectively.

b) If a component is defective, what is the probability it came from shift 3?



Using Bayes' theorem,

$$P(S_3 \mid D) = \frac{P(D \mid S_3)P(S_3)}{P(D)}$$

$$= \frac{0.12 \times 0.3}{0.082} = 0.439$$

Which suggests that almost half of defective items produced are produced during Shift 3.

Total probability and Bayes' Theorem

Prior and Posterior Probabilities:

Bayes' theorem is sometimes referred to as a formula for updating probabilities in the light of new information.

In this sense we talk about **prior** *(Before)* and **posterior** *(after)* probabilities.

- The **prior probabilities** are updated via the evidence represented by the data to produce the **posterior probabilities**.

The next simple example will illustrate this.

Total probability and Bayes’ Theorem

Example 4.3 (*Updating Probabilities*)

A pathologist has narrowed down the cause of death to one of three poisons and believes that the probabilities are 0.6, 0.3 and 0.1 of death being caused by poisons P_1, P_2 and P_3, respectively.

Another test, T, is carried out which is positive with probabilities 0.1, 0.7 and 0.2 if death is due to poisons P_1, P_2 and P_3, respectively.

If the test is positive, how should the probabilities be updated in the light of this new information?

Prior probabilities:

$P(P_1) = 0.6$

$P(P_2) = 0.3$

$P(P_3) = 0.1$

New Test Probabilities :

$P(T|\ P_1) = 0.1$

$P(T|\ P_2) = 0.7$

$P(T|\ P_3) = 0.2$

Total probability and Bayes' Theorem

Example 4.3 (*Updating Probabilities*)

Then, Using Bayes' theorem,

$$P(P_i \mid T) = \frac{P(T \mid P_i)P(P_i)}{\sum_i P(T \mid P_i)P(P_i)}$$

Total probability and Bayes' Theorem

Bayes' Theorem

Using the results above, and the definition of total probability, we can derive the following theorem,

Theorem 4.3 (*Bayes' Theorem*)

If events $B_1, B_2, \ldots, B_m$, are mutually **exclusive** and **exhaustive**, then, for any event A, we have,

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{k=1}^{m} P(A \mid B_k)P(B_k)}, \qquad k = 1,2,\ldots,m$$

Therefore,

$$P(P_1 \mid T)) = \frac{P(T \mid P_1)P(P_1)}{P(T \mid P_1)P(P_1) + P(T \mid P_2)P(P_2) + P(T \mid P_3)P(P_3)}$$

$$= \frac{(0.1)(0.6)}{(0.1)(0.6) + (0.7)(0.3) + (0.2)(0.1)}$$

$$= \frac{0.06}{0.29} = 0.21$$

Total probability and Bayes' Theorem

Example 4.3 (*Updating Probabilities*)

$$\mathrm{P}(\mathrm{P}_2 \mid \mathrm{T})) = \frac{(0.7)(0.3)}{0.29}$$

$$= \frac{0.21}{0.29} = 0.72$$

$$\mathrm{P}(\mathrm{P}_3 \mid \mathrm{T})) = \frac{(0.2)(0.1)}{0.29}$$

$$= \frac{0.02}{0.29} = 0.07$$

Note that the conditional probabilities, $\mathrm{P}(\mathrm{P}_i | \mathrm{T})$, i = 1, . . . , 3, themselves form a probability distribution, known as the **posterior distribution**.

Total probability and Bayes' Theorem

Example 4.4 (*Updating Probabilities - continued*)

Consider the conclusions reached after **Example 4.3** and suppose that a new test, T_2 is available which is positive with the following characteristics,

$$P(T_2 \mid P_1) = 0.05,\ P(T_2 \mid P_2) = 0.95 \text{ and } P(T_2 \mid P_3) = 0.0.$$

Thus, if the new test comes back positive, it almost certainly means it was Poison 2 and totally rules out the possibility of it being Poison 3. We now use the **posterior probabilities as our prior** and update them to find new posterior probabilities in the light of both tests. distribution.

Thus, we obtain,

$$P(P_1 \mid T_2) = \frac{P(T_2 \mid P_1)P(P_1)}{P(T_2 \mid P_1)P(P_1) + P(T_2 \mid P_2)P(P_2) + P(T_2 \mid P_3)P(P_3)}$$

{Using Bayes' theorem}

$$= \frac{(0.05)(0.21)}{(0.05)(0.21) + (0.95)(0.72) + (0.0)(0.07)} = \frac{0.0105}{0.6945} = \mathbf{0.015}$$

Total probability and Bayes' Theorem

Example 4.4 (*Updating Probabilities - continued*)

$$\mathrm{P}(\mathrm{P}_2 \mid \mathrm{T}_2) = \frac{(0.95)(0.72)}{0.6945} = \mathbf{0.985}$$

$$\mathrm{P}(\mathrm{P}_3 \mid \mathrm{T}_2) = \frac{(0.0)(0.07)}{0.6945} = \mathbf{0.000}$$

As might be expected, a positive result from the second test makes the diagnosis of Poison 2 almost conclusive whilst ruling out Poison 3.

In practice *evidence* can be accumulated in this way over any number of steps in order to produce a final posterior distribution.

Applications of Bayes' Theorem

Consider the manner in which juries weigh up evidence against an accused person. We suppose that items of evidence appear independently of each other, obviously an assumption that could be criticised, and that jurors start out with prior beliefs about the guilt or innocence of the accused.

Let, E denote the evidence presented, G denote the guilt of the accused and $\bar{G}$ denote their innocence. Then, using Bayes' theorem

$$P(G \mid E) = \frac{P(E \mid G)P(G)}{P(E)}$$

and

$$P(\bar{G} \mid E) = \frac{P(E \mid \bar{G})P(\bar{G})}{P(E)}$$

We can write the following,

$$\Rightarrow \underbrace{\frac{P(G \mid E)}{P(\bar{G} \mid E)}}_{\text{Posterior Odds}} = \underbrace{\frac{P(G)}{P(\bar{G})}}_{\text{Prior Odds}} \times \underbrace{\frac{P(E \mid G)}{P(E \mid \bar{G})}}_{\text{Likelihood}}$$

Applications of Bayes' Theorem

Clearly, if the evidence consists of m independent factors, i.e. E = $(E_1, E_2, . . . , E_m)$, then we have

$$P(E \mid G) = \prod_{i=1}^{m} P(E_i \mid G) \ etc.$$

Note: The symbol $\prod_{i=1}^{m}$ ***is known as "Pi – product" i.e. product of terms from 1 to m***}

- The prior odds denote the juror's relative belief about the innocence or guilt of the accused before the presentation of the evidence.
- The likelihood measures the contribution of the evidence towards proving the accused's innocence or guilt.
- The posterior odds reflects the juror's updated relative belief about the innocence or guilt of the accused having heard the evidence.

Applications of Bayes' Theorem

Consider two well known phrases concerning the law.

$$\Rightarrow \underbrace{\frac{P(G \mid E)}{P(\bar{G} \mid E)}}_{\text{Posterior Odds}} = \underbrace{\frac{P(G)}{P(\bar{G})}}_{\text{Prior Odds}} \times \underbrace{\frac{P(E \mid G)}{P(E \mid \bar{G})}}_{\text{Likelihood}}$$

Innocent until proven guilty

This suggests that the prior odds, before any evidence is heard, should be a very small value, close to zero.

Sure beyond all reasonable doubt

This suggests that the posterior odds, after hearing all the evidence, should be a very large value. The odds on guilt change, assuming there is sufficient evidence, because of the cumulative effect of the likelihood of the evidence.

Applications of Bayes' Theorem

The Sally Clarke Case (1997):

Recall the calculations carried out by the expert witness,

$$P(\text{One cot death}) = \frac{1}{8500}$$

$$\Rightarrow P(\text{Two cot deaths}) = \frac{1}{8500} \times \frac{1}{8500} \approx \frac{1}{73 \text{ million}}$$

The Office for National Statistics data for 1997 (when the case took place) showed that, of 642093 live births, seven were murdered in the first year of life i.e. $\frac{7}{642093} \approx \frac{1}{91728}$.

Assuming independence, as above, this would give a probability of two babies being murdered of 1 in 8.4 billion i.e. $\left(\frac{1}{91728}\right)^2$ **= 1/84 billion**.

Applications of Bayes' Theorem

The Sally Clarke Case (1997):

E, the evidence before the court is simply that the two babies died. Trivially, the two conditional probabilities $P(E\,|\,G) = P(E\ \bar{G}) = 1$.

Thus, application of Bayes, theorem yields,

$$\Rightarrow \underbrace{\frac{P(G \mid E)}{P(\bar{G} \mid E)}}_{\text{Posterior Odds}} = \underbrace{\frac{P(G)}{P(\bar{G})}}_{\text{Prior Odds}} \times \underbrace{\frac{P(E \mid G)}{P(E \mid \bar{G})}}_{\text{Likelihood}}$$

$$\frac{P(G|E)}{P(\bar{G}|E)} = \frac{P(G)}{P(\bar{G})} \times \frac{P(E|G)}{P(E|\bar{G})}$$

$$= \frac{1/8.4 \text{ billion}}{1/73 \text{ million}} = 0.009$$

i.e. the odds are 9 in 1000 against her being guilty, corresponding to a posterior probability of guilt **0.0089**.

Note that if the odds are denoted by **O**, then the **Posterior probability (Posterior odds)**, **π** say, is given by

$$\boldsymbol{\pi = O/(1 + O).}$$