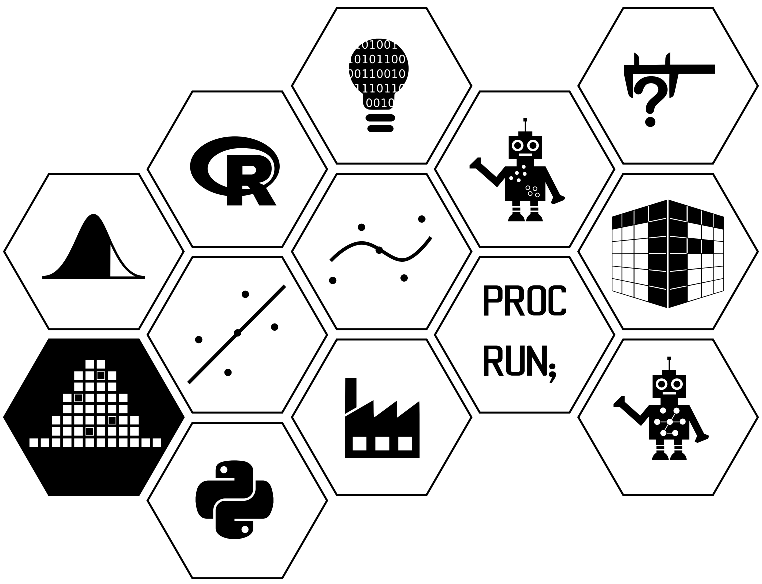
Probability and Sampling Fundamentals

Week 2: Conditional Probability, Independence and Bayes' Theorem



DATA ANALYTICS GLASGOW

Conditional probability, independence and Bayes' theorem

Marginal vs conditional probability

So far in this course we have mainly considered *marginal probabilities* such as $P(E)$ which are probabilities that do not depend on anything other than the event in question, e.g. E. However, events in real life rarely have a simple probability. In general our beliefs about uncertain events change when we get new information. For example, if the weather forecast says there is a 40% chance of rain tomorrow, that probability is based on all of the information that the meteorologists know up until that point, for example a cold front coming to the area. This probability will also be updated if new information is learned, e.g. another front pushing the rain clouds away.

Conditional probability provides a way for us to precisely say how our beliefs change.

Week 2 learning material aims

The material in week 2 covers:

- the concept of a conditional probability;
- how to apply conditional probability;
- the concept of independence of events;
- how to check if two events are independent;
- Bayes' Theorem and its application.

Conditional probability - Motivating example

Consider the questions in the following example.[1]

Example 1

Motivating example

Big bears or swarming bees

Sixty students were asked, "Would you rather be attacked by a big bear or swarming bees?". Their answers, along with their sex were recorded and are presented in the following table.

	Big Bear (B)	Swarming Bees (S)	Total
Female (F)	27	9	36
Male (M)	10	14	24
Total	37	23	60

We would like two find the following two probabilities.

1. What is the probability that a student is male and would rather be attacked by a big bear?
2. What is the probability that a male student would rather be attacked by a big bear?

Before we answer these questions let's first find the sample space.

Here, the sample space contains four possible responses

$$S = \{(F,B),(F,S),(M,B),(M,S)\},$$

where M is male, F is female, B is big bear and S is swarming bees. For example, the first response (F,B) corresponds to a female who has stated she would prefer to be attacked by a big bear.

The figure below shows how each member of the population of 60 students maps to an element of the sample space.

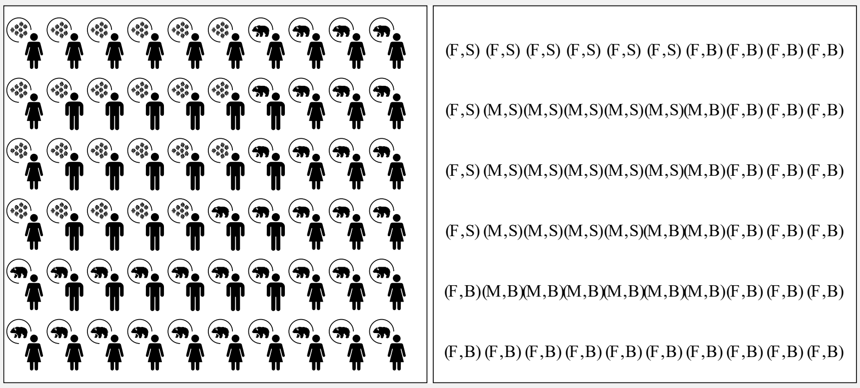


Figure 1

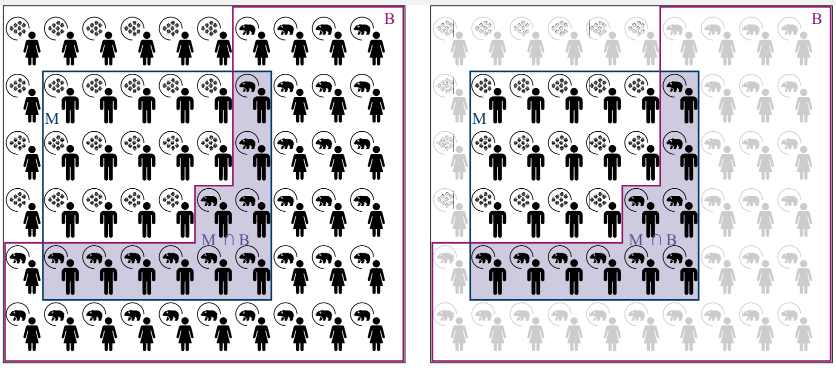
From the table of counts, the probability of each outcome occurring can be calculated as follows:

	Big bear (B)	Swarming bees (S)
Female (F)	$P(F \cap B) = 27/60 = 0.4$	$P(F \cap S) = 9/60 = 0.150$
Male (M)	$P(M \cap B) = 10/60 = 0.167$	$P(M \cap S) = 14/60 = 0.233$

Looking again at the two questions, although they sound similar they ask different things. The probability that a student is male *and* would rather be attacked by a big bear, i.e. $P(M \cap B)$, is $10/60 = 0.167$. However the proportion of male students that would rather be attacked by a big bear is $10/24 = 0.4167$ since there are 24 males and 10 of them would rather be attacked by a big bear. In other words, for the second question, we are calculating the probability that a student would rather be attacked by a big bear *given that* the student is male.

The second probability is an example of *conditional probability*, which is the probability of one event occurring given some information about one or more other events. For example, here the information we know is that the student is male. The probability we have calculated is *conditional* on that information.

We will soon define that for any events A and B, the conditional probability of A given B is denoted as $P(A|B)$. In other words, this is the probability of event A occurring **given that** event B has already occurred. Hence, the probability that a student would rather be attacked by a big bear given that the student is male is $P(B|M) = 0.417$.



The probability $P(M \cap B) = \frac{10}{60}$ relates the number of male students who would prefer to be attacked by a big bear to the number of all students (male and female). It gives the probability that a randomly chosen student (male or female) is both male and would prefer to be attacked by a big bear.

The conditional probability $P(B|M) = \frac{10}{24}$ relates the number of male students who would prefer to be attacked by a big bear to the number of male students. It gives the probability that a randomly chosen male student would prefer to be attacked by a big bear.

Figure 2

We can now state the general definition of conditional probability.

Definition 1

Conditional probability

For any events A and B, the **conditional probability** that event A occurs given that event B happened is

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

where $A|B$ denotes the event that A happens given that B has already happened (provided $P(B) > 0$).

Applying this definition to Example 1 we have, the probability that a student would rather be attacked by a big bear given that the student is male is equal to the probability that a student is male and would prefer to be attacked by a big bear ($P(M \cap B)$) divided by the probability that a student is male ($P(M)$). That is,

$$P(B|M) = \frac{\text{Number of males who prefer a big bear}}{\text{Number of males}} = \frac{P(M \cap B)}{P(M)} = \frac{10/60}{24/60} = \frac{10}{24} = 0.417.$$

If we cancel out the common factor of $1/60$ the calculation just corresponds to the fact that, out of the 24 male students, 10 would prefer to be attacked by a big bear.

We can view the conditional probability $P(A|B)$ as "zooming into" the event B. The (unconditional) probability $P(A)$ tells us how likely the event A is. However, once we know that B has occurred, the conditional probability $P(A|B)$ tells us how likely the event A is in the presence of this additional information. Essentially, if we know that B has occurred we can ignore the entire sample space outside B and just focus only on B.

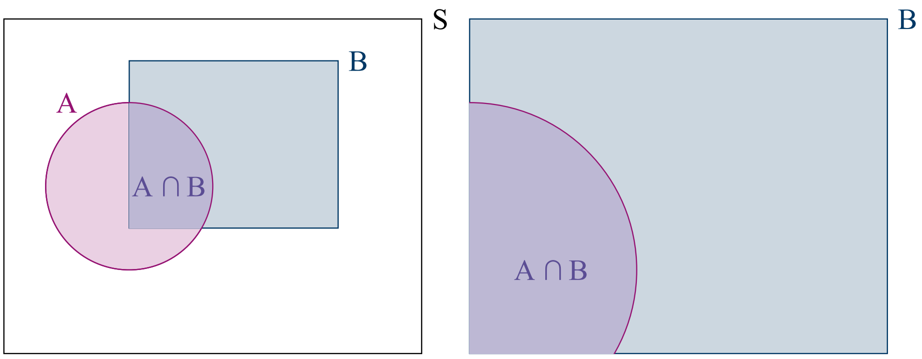


Figure 3

Definition 1 can also be re-expressed as:

$$P(A \cap B) = P(A|B)P(B),$$

by multiplying both sides of the formula in **Definition 1** by $P(B)$.

Also, since $P(A \cap B) = P(B \cap A)$ we can re-express this formula again as

$$P(A \cap B) = P(B|A)P(A).$$

since by **Definition 1**,

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

This gives another way to calculate probabilities of the form $P(A \cap B)$.

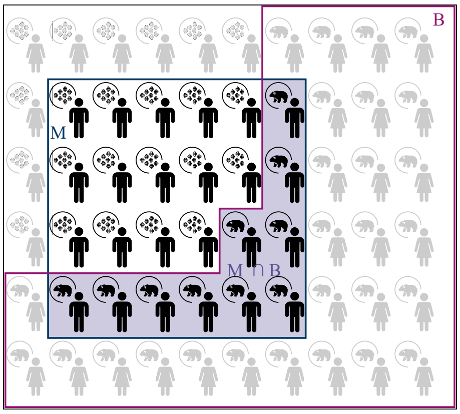
Note: $P(A|B)$ is *not* the same as $P(B|A)$ and should not be used interchangeably. The former is the probability of event A occurring given event B happened, and the latter the probability of event B occurring given event A happened.

For example, we know in **Example 1** that the probability that a student would rather be attacked by a big bear given that they are male is $P(B|M) = 0.417$. This is **not** the same as the probability that a student is male given that the student would rather be attacked by a big bear, which is calculated as,

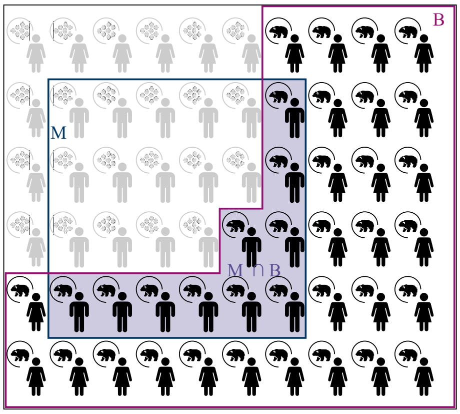
$$P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{10/60}{37/60} = \frac{10}{37} = 0.270.$$

We can see that $P(M|B)$ is just the ratio

$$\frac{\text{Number of males who prefer a big bear}}{\text{Number of students (male or female) who prefer a big bear}}.$$



The conditional probability $P(B|M) = \frac{P(B \cap M)}{P(B)} = \frac{10}{24}$ relates the number of male students who would prefer to be attacked by a big bear to the number of male students. It gives the probability that a randomly chosen male student would prefer to be attacked by a big bear.



The conditional probability $P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{10}{37}$ relates the number of male students who would prefer to be attacked by a big bear to the number of all students (male or female) who would prefer to be attacked by a big bear. It gives the probability that a randomly chosen student who prefers to be attacked by a big bear is male.

Figure 4

Example 2

Conditional probability

Flight

On a flight of 160 people, 80 paid for a hold bag, 60 paid to select a seat, and 40 paid for a hold bag and to select a seat, as shown in the Venn diagram below. If a passenger, chosen at random, paid for a hold bag, what is the probability that they also paid to select a seat?

Answer:

Consider the events

$$H = \{\text{paid for hold bag}\} \qquad \text{and} \qquad S = \{\text{paid for seat selection}\}$$

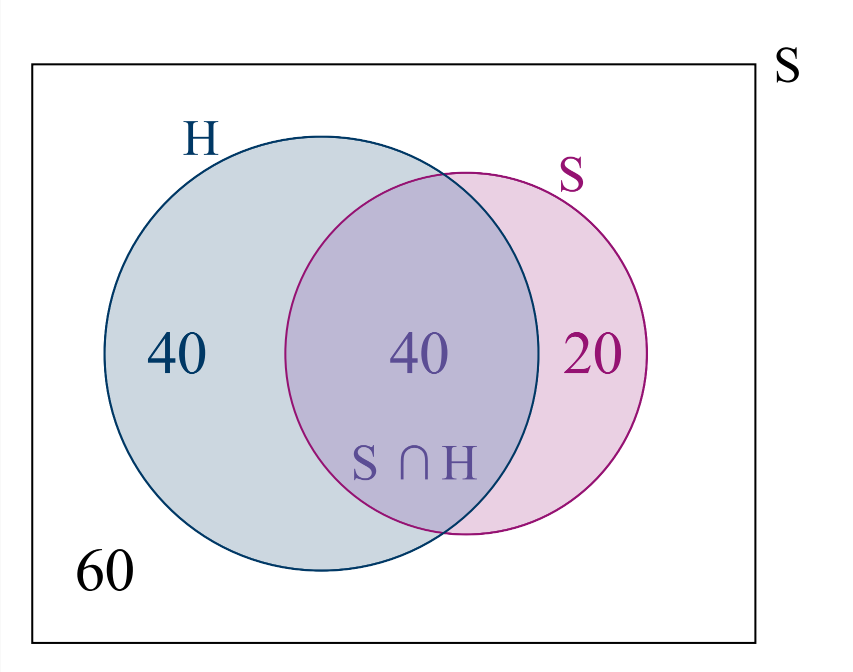


Figure 5

To answer this question first we need to figure out $P(H)$, which is given in the question as $P(H) = \frac{80}{160} = 0.5$.

We are also told $P(H \cap S) = \frac{40}{160} = 0.25$, which is the probability of both events happening.

We can now apply the formula for conditional probability

$$\begin{aligned} P(S|H) &= \frac{P(H \cap S)}{P(H)} \\ &= \frac{0.25}{0.5} \\ &= 0.5. \end{aligned}$$

In other words, the probability that a passenger will purchase a seat **given** that they have purchased a hold bag is 0.5, which we could also write as $\frac{1}{2}$ or 50%.

The equation

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

can be extended to more than two events. For example, for events A, B and C we have:

$$P(A \cap B \cap C) = P(C|A \cap B)P(B|A)P(A).$$

This video provides a recap of conditional probability and how it can be used to answer an interesting real-life question.

Video

Men, Women, and Murder

Duration 6:06

Example 3

written solutions to example in video

Men, Women, and Murder Consider the following data from the Office for National Statistics relating to homicides in England and Wales spanning 11 years from April 2007 to March 2018. The table shows the number of homicides for victims aged 16 and over who were acquainted with the suspect, split by sex of the victim and whether the suspect was a partner or ex-partner or some other acquaintance (e.g. parent, friend, etc.).

	Partner/Ex-Partner	Other	Total
Male (M)	178	2084	2262
Female (F)	929	474	1403
Total	1107	2558	3665

Questions of interest:

1. What is the probability that the suspect is a partner or ex-partner if the victim is female?
2. How does this compare to male victims?

First of all, let's define some terminology.

Let

$$F = \{\text{victim is female}\}, \quad M = \{\text{victim is male}\}, \quad E = \{\text{suspect is partner}/\text{ex-partner}\}.$$

From the table of counts, the probability of each outcome occurring can be calculated as follows:

Sex	Partner or Ex-partner (E)	Other (E')
Male (M)	$P(M \cap E) = 178/3665 = 0.049$	$P(M \cap E') = 2084/3665 = 0.569$
Female (F)	$P(F \cap E) = 929/3665 = 0.253$	$P(F \cap E') = 474/3665 = 0.129$

and

$$P(M) = 2262/3665 = 0.617$$

$$P(F) = 1403/3665 = 0.383$$

Answer

1. To find the conditional probability that the suspect is a partner/ex-partner given that the victim is female, we need to find the following conditional probability, $\mathrm{P}(E|F)$.

$$\begin{aligned}\mathrm{P}(E|F) &= \frac{P(E \cap F)}{P(F)} \\ &= \frac{0.253}{0.383} \\ &= 0.661\end{aligned}$$

2. To find the conditional probability that the suspect is a partner/ex-partner given that the victim is male, we need to find the following conditional probability, $\mathrm{P}(E|M)$.

$$\begin{aligned}\mathrm{P}(E|M) &= \frac{P(E \cap M)}{P(M)} \\ &= \frac{0.049}{0.617} \\ &= 0.079\end{aligned}$$

Task 1

In a population, 70% of people like chocolate ice cream and 35% like chocolate and strawberry. What is the probability that someone who likes chocolate also likes strawberry?

Independence and dependence

Calculating conditional probabilities naturally leads to the concept of **independence** between two events, which arises if the occurrence of the first event does not affect whether the second event happens. This concept has already been introduced in week 1.

Although we typically expect the conditional probability $P(A|B)$ to be different from the probability $P(A)$, this does not have to be the case. When $P(A|B) = P(A)$, the occurrence of the event B has no impact on the likelihood of event A. In other words, whether or not the event A has occurred is **independent** of the event B.

For example, the probability of you running a marathon in less that 4 hours is probably dependent on how much you train. However, it is probably not dependent on what your neighbour's favourite book is. Simply put, your marathon time is **independent** of your neighbour's favourite book. Whereas, if you train well you are more likely to run a sub-4 hour marathon and so your marathon time is **dependent** on how much you train. Therefore, conditional probability can be used to check if two events are independent.

Definition 2

Independence

Events A and B are **independent** if

$$P(A|B) = P(A),$$

or

$$P(A \cap B) = P(A)P(B).$$

Although both of these definitions can be used to test for independence between two events, the former is more intuitive, since events are independent from one another if the probability of one occurring does not impact the probability of the other occurring. The latter definition is known as the *multiplication rule for independent events* and is commonly used as the primary definition of independent events.

Note: Independence of two events means that whether or not one event occurs does not impact the probability of the other event occurring. Hence, if two events, A and B are independent then it follows that the pairs (A, B'), (A', B) and (A', B') are also all independent events.

Supplement 1

Proof that complements of independent events are also independent

Suppose A and B are independent events. Then the definition of independence implies that $P(A \cap B) = P(A)P(B)$. We will now show that A and B' are also independent.

If we look at the Venn diagram below, we can see that we can write A as the disjoint union

$$A = (A \cap B) \cup (A \cap B').$$

In other words, if we know that A happens then either A and B will happen or A and B' will happen.

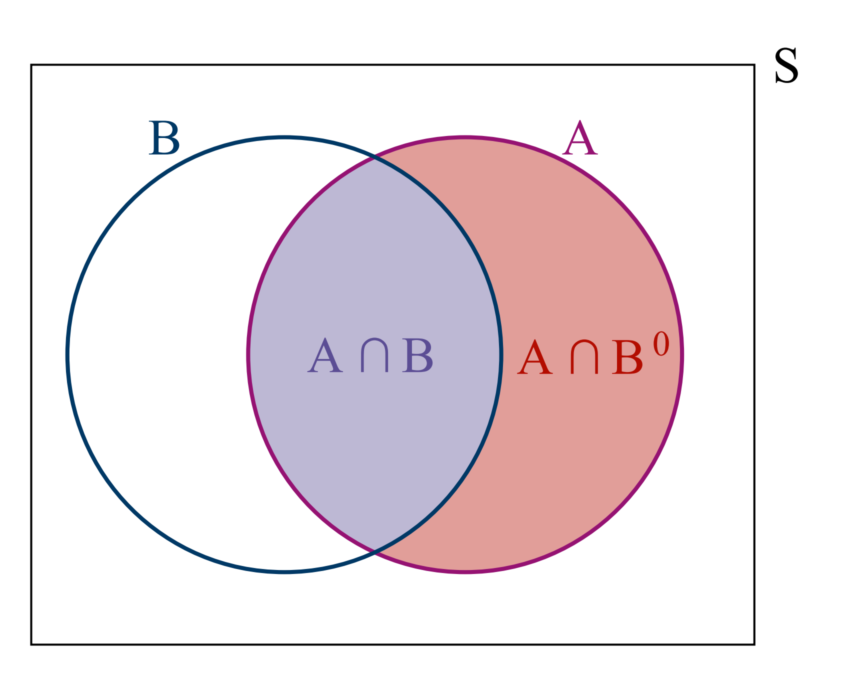


Figure 6

The third axiom of probabilities now implies that

$$P(A) = P(A \cap B) + P(A \cap B').$$

Let's re-arrange this, so that $P(A \cap B')$ is on the left-hand side.

$$\begin{aligned} P(A \cap B') &= P(A) - \underbrace{P(A \cap B)}_{P(A)P(B)} \\ &= P(A) - P(A)P(B) = P(A)\underbrace{(1 - P(B))}_{=P(B')} \\ &= P(A)P(B') \end{aligned}$$

thus A and B' are also independent.

Similarly, one can also show that A' and B and A' and B' are independent if A and B are independent.

The concept of independence can be extended to more that two events.

Definition 3

Mutual independence

Events A,B and C are **mutually independent** if

1. $P(A \cap B \cap C) = P(A)P(B)P(C)$ and
2. $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, and $P(B \cap C) = P(B)P(C)$.

Note: the second condition is known as **pairwise independence**, i.e. all pairs of events are independent. The first condition (1) does not necessarily imply pairwise independence between the three events. In order to check for mutual independence, you need to check that both conditions (1) and (2) hold.

Example 4

Independence

Train

John takes the train to work. Every day there is a 20% chance that his train will be late, a 10% chance that he will be late, and a 2% chance that both he and the train will be late. Let

A be the event that the train is late and B be the event that John is late.

We want to answer the following two questions.

(a) Are events A and B independent? (b) What is the probability that the train is late but John is not?

Answer

(a) From the definition of independence, two events, A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

Here we have

$P(A) = 0.2$, $P(B) = 0.1$ and $P(A \cap B) = 0.02$

since

$$P(A)P(B) = 0.2 \cdot 0.1 = 0.02 = P(A \cap B).$$

Therefore, the events are independent.

Alternatively, we could have calculated

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.02}{0.1} = 0.2,$$

which equals $P(A)$. Hence A and B are independent.

(b) We already know the probability that the train is late, $P(A) = 0.2$.

The event that John is not late is just the *complement* of John being late, i.e. $P(B') = 1 - P(B) = 0.9$.

Since we know that A and B' are independent (from part (a)) then,

$$P(A \cap B') = P(A)P(B') = 0.2 \cdot 0.9 = 0.18.$$

Mutually exclusive or disjoint events

Two events A and B are **mutually exclusive** or **disjoint** if it is **impossible** for them to occur at the same time, i.e. they have no outcomes in common. Compare this to independent events where two events can occur at the same time without affecting each other when they do. Mutually exclusive events exclude each other from occurring and hence affect each other greatly, and are **not** independent.

If the events A and B are mutually exclusive then:

1. $A \cap B = \emptyset$.
2. $P(A \cap B) = 0$.
3. $P(A \cup B) = P(A) + P(B)$.
4. $P(B|A) = P(A|B) = 0$.

Below is a Venn diagram of two mutually exclusive events.

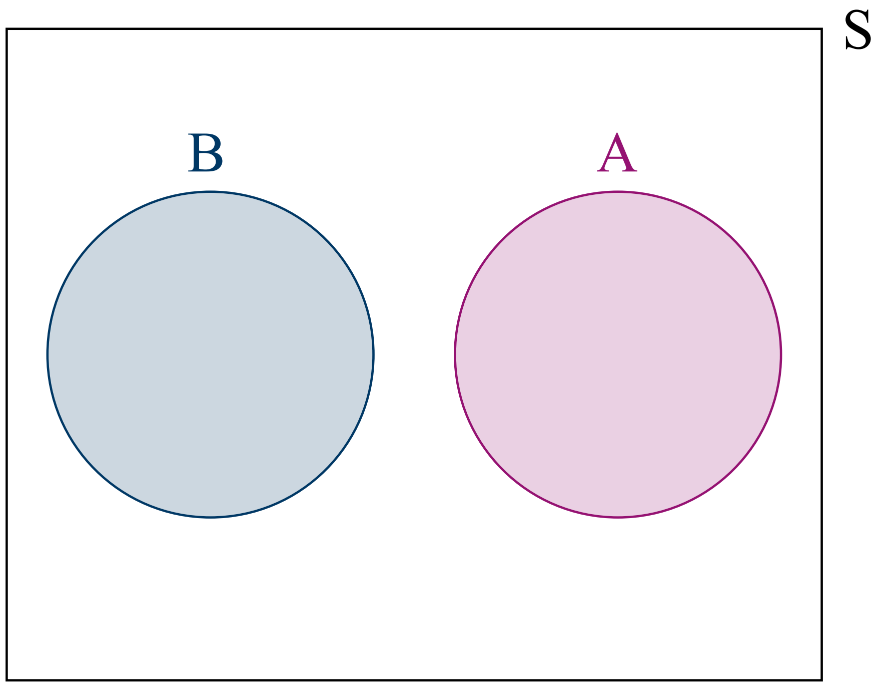


Figure 7

Example 5

Disjoint events

Determine if the following events are disjoint?

(a) $A = \{\text{Drawing a jack from a deck of cards}\}$.

$B = \{\text{Drawing a queen from a deck of cards}\}$.

(b) $A = \{\text{Drawing a jack from a deck of cards}\}$.

$B = \{\text{Drawing a heart from a deck of cards}\}$.

(c) $A = \{\text{Turning left at the next crossroads}\}$.

$B = \{\text{Turning right at the next crossroads}\}$.

Answer

To answer these questions we need to ask ourselves if both of the events listed can occur at the **same** time.

(a) $A = \{\text{Drawing a jack from a deck of cards}\}$.

$B = \{\text{Drawing a queen from a deck of cards}\}$.

Since you cannot draw a card that is a jack and a queen at the same time these two events are **disjoint**.

(b) $A = \{\text{Drawing a jack from a deck of cards}\}$.

$B = \{\text{Drawing a hearts from a deck of cards}\}$.

The card you draw could be the jack of hearts so these two events are **not disjoint**.

(c) $A = \{\text{Turning left at the next crossroads}\}$.

$B = \{\text{Turning right at the next crossroads}\}$.

You cannot turn left and turn right at the same time, so these two events are **disjoint**.

Task 2

A card is drawn at random from a deck of cards. Let $A = \text{the card is a king,}$ $B = \text{the card is a diamond}$ and $C = \text{the card is red.}$

(a) Are the events A and B independent? Are they disjoint?

(b) Are the events A and C independent?

(c) Are the events B and C independent?

Bayes' Theorem & Law of Total Probability

Bayes' theorem is an important formula which allows us to calculate conditional probabilities when limited information is available. However, before we can state this we need to define the law of total probability.

To do this let us refer back to an example introduced in week 1.

Example 6

Law of total probability

Disease outcome continued

Suppose 10% of the population are infected with a disease.

A diagnostic test is available such that

- 97% of people without the disease will test negatively,
- 99.5% of people with the disease will test positively.

What is the probability a randomly selected person tests positively for this disease?

Before we answer this question, we need to develop some terminology. Let, for a randomly select person,

$B = \{\text{person has the disease}\}.$

$A = \{\text{person tests positively for this disease}\}.$

In more formal notation, the two probabilities given above are:

- probability that someone tests negatively given that they don't have the disease = $P(A'|B') = 0.97$;
- probability that someone tests positively given that they have the disease = $P(A|B) = 0.995$.

We can also deduce the following probabilities given the information we have:

- probability that someone tests positively given they don't have the disease as $P(A|B') = 1 - 0.97 = 0.03$, since $P(A'|B') + P(A|B') = 1$;
- probability that someone tests negatively given that they have the disease = $P(A'|B) = 1 - 0.995 = 0.005$, since $P(A|B) + P(A'|B) = 1$.

If we consider examining one randomly selected person, they will either have the disease or not, i.e. either B or B' (but not both) will be true. We say that these events form a **partition** of the sample space for this experiment.

Next we will define a partition in more general.

Definition 4

Partition

A collection of events $\{E_1, E_2, \ldots, E_k\}$ for $k > 1$ form a **partition** of the sample space $\mathcal{S}$ if

- $E_i \cap E_j = \emptyset$ for $i \neq j$ - each pair of events is disjoint;
- $E_1 \cup E_2 \cup \ldots \cup E_k = \mathcal{S}$ - the union of all events covers the whole of $\mathcal{S}$.

In other words, a partition splits the sample space into events, such that *exactly one* of the events in the partition will occur.

The following Venn diagram shows an example of a collection of events that form a partition of the sample space.

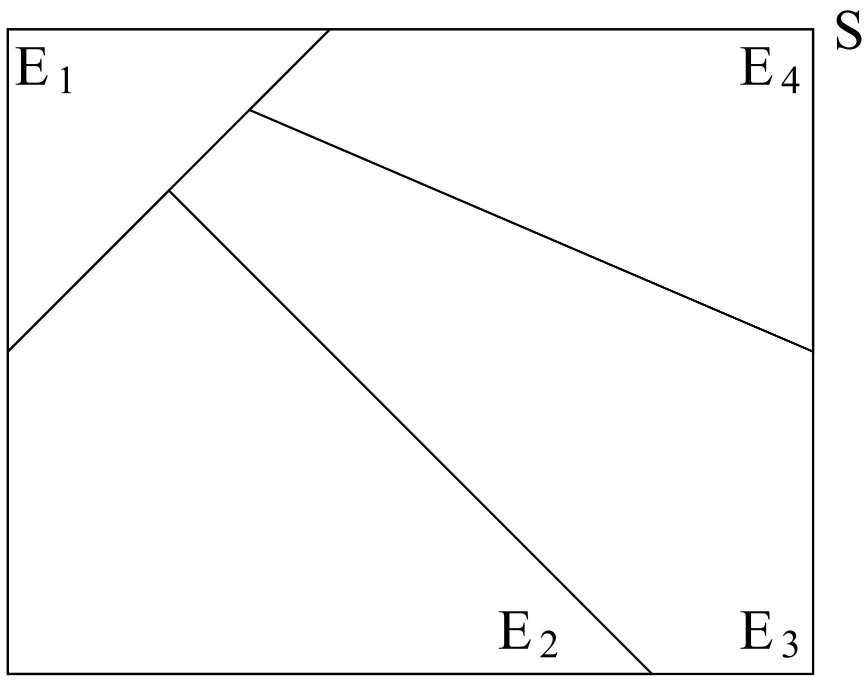


Figure 8

An event B and its complement B' will always form a partition, as exactly one event out of B and B' will occur: B and B' can't both occur at the same time ($B \cap B' = \emptyset$), nor can it be that neither B nor B' will occur (as $B \cup B' = \mathcal{S}$).

Example 7

Partitions

For each of the following examples define a possible partition. Note that partitions may not be unique.

1. Toss a coin - Partition = $\{H_1 = \{\text{Heads}\}, H_2 = \{\text{Tails}\}\}$ (unique partition since only two possible outcomes for a coin toss).
2. Draw one card - Partition = $\{H_1 = \{\text{Spades}\}, H_2 = \{\text{Diamonds}\}, H_3 = \{\text{Hearts}\}, H_4 = \{\text{Clubs}\}\}$, or Partition = $\{H_1 = \{\text{Face cards}\}, H_2 = \{\text{Non-face cards}\}\}$.

Let's now go back to the question from Example 6. We want to find the probability of event \newline $A = \{\text{a randomly selected person tests positively for this disease}\}$. We are told that $P(A|B) = 0.995$ and $P(A|B') = 1 - 0.97 = 0.03$, i.e. we are given conditional probabilities. In order to be able to proceed with our calculations we need to find the unconditional probability $P(A)$. The **law of total probability**, which is sometimes also called the partition theorem, lets us do exactly that.

Theorem 1

Law of total probability

If $\{E_1, E_2, \ldots, E_k\}$ are a partition of $\mathcal{S}$, then for any event A

$$\begin{aligned} P(A) &= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k) \\ &= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k). \end{aligned}$$

We can interpret the law of total probability as a weighted average: we can compute the (unconditional) probability $P(A)$ as a weighted average of the conditional probabilities $P(A|E_i)$, using the $P(E_i)$ as weights.

For the special case of a partition consisting of an event B and its complement B', the law of total probability implies that

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B') \\ &= P(A|B)P(B) + P(A|B')P(B'), \end{aligned}$$

with the second line requiring that $0 < P(B) < 1$.

Example 8

Law of total probability

Disease outcome continued

In Example 6 we can use the law of total probability to calculate the probability that the test returns a positive result, irrespective of whether the person has the disease or not.

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B')P(B') \\ &= (0.995)(0.1) + (0.03)(0.9) \\ &= 0.127 \end{aligned}$$

So, 12.7% of the population would test positively for the disease.

Supplement 2

Proof of the law of total probability

Consider the partition and the event A shown in the Venn diagram below.

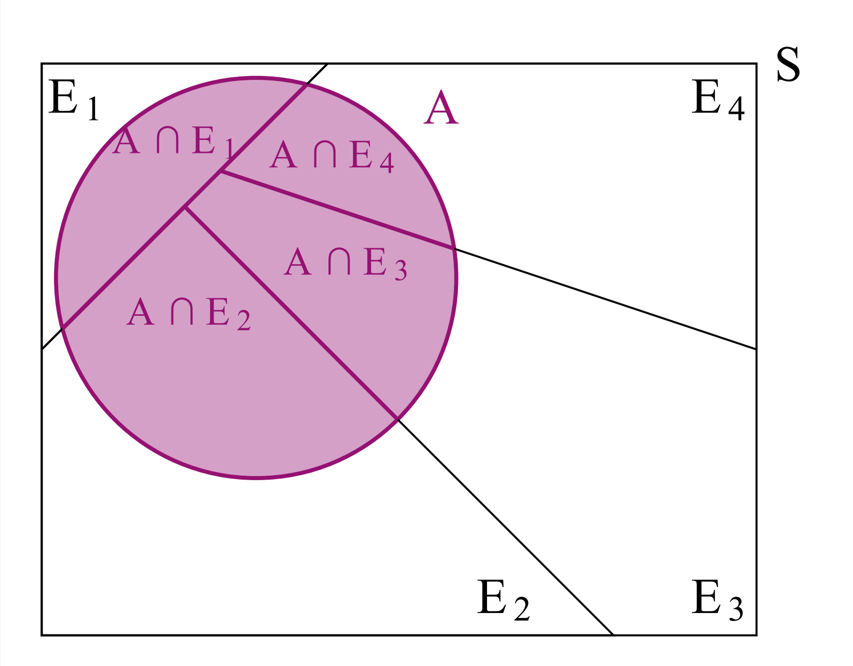


Figure 9

We can use the fact that the partition also splits the event A into events of the form $A \cap E_i$. More formally speaking, we can write A as the disjoint union

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_k).$$

Applying the third axiom of probability gives

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k)$$

For each E_i we can use that $P(A \cap E_i) = P(A|E_i)P(E_i)$, which then yields

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k).$$

Task 3

I have three bags that each contain 100 marbles:

- bag 1 has 75 red and 25 blue marbles;
- bag 2 has 60 red and 40 blue marbles;
- bag 3 has 45 red and 55 blue marbles.

I choose one of the bags at random and then pick a marble, also at random, from the chosen bag. What is the probability that the chosen marble is red?

Now consider another question from Example 6. Say we are now interested in knowing how many of the people that test positively actually have the disease. To show this we need a further result known as **Bayes' Theorem**.

Bayes' Theorem

Bayes' Theorem is one of the most powerful concepts in statistics, not only does it allow us to calculate conditional probabilities when limited information is available but it also allows us to connect the two conditional probabilities, $P(A|B)$ and $P(B|A)$.

Theorem 2

Bayes' theorem

Consider a partition $E_1, E_2, \ldots, E_k$ of $\mathcal{S}$. Then for any event A with $P(A) > 0$ and any E_i

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \cdots + P(A|E_k)P(E_k)}$$

In the special case of a partition consisting of an event B and its complement B' we obtain

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

Supplement 3

Proof of Bayes' theorem

We start the proof with the definitions of the conditional probability $P(E_i|A) = \frac{P(E_i \cap A)}{P(A)}$ and re-arrange the definition of the conditional probability $P(A|E_i) = \frac{P(E_i \cap A)}{P(E_i)}$ to obtain $P(E_i \cap A) = P(A|E_i)P(E_i)$. Putting these to together,

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

To obtain the final expression, we plug in the result from the law of total probability that $P(A) = P(A|E_1)P(E_1) + \cdots + P(A|E_k)P(E_k)$, which yields

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \cdots + P(A|E_k)P(E_k)}.$$

Example 9

Application of Bayes' theorem

Disease outcome

Refer back to Example 6. We are now interested in knowing:

What proportion of people that test positively actually have the disease?

The question "how many of the people that test positively actually have the disease" is really asking for the *conditional* probability that a person has the disease *given* that they have tested positively. Which we can find by applying Bayes' Theorem.

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')} \\ &= \frac{(0.995)(0.1)}{(0.995)(0.1) + (0.03)(0.9)} \\ &= \frac{0.0995}{0.0995 + 0.0270} \\ &\approx 0.79 \end{aligned}$$

In other words, 79% of people who test positively actually have the disease. Therefore, the remaining 21% who test positively do not have the disease. This result may seem surprising given the 99.5% success of detecting the disease when it is present.

There is an intuitive explanation of this. Imagine we test 100 patients. On average 90 of them will be healthy and the remaining 10 will have the disease.

For the 90 healthy patients the test gives the correct result in 97% of the cases. So, on average, it will give a false positive result for approximately 3 patients.

The test has a very low false negative rate (0.5%), so it is very likely that all 10 patients who have the disease will be detected.

The figure below illustrates this. Patients who have the disease are shown with a green unhappy face. Positive test results are indicated by a purple rectangle.

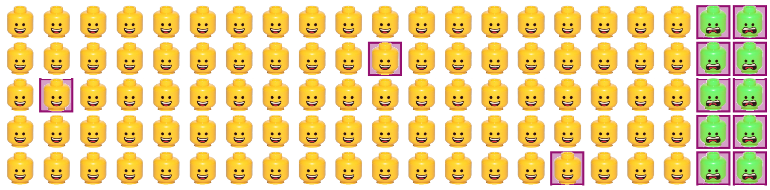


Figure 10

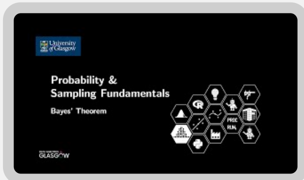
So, on average around 13 patients will have positive test results. On average around 3 of them will be healthy, whereas 10 of them will have the disease. So, if the test result is positive, then the patient actually has the disease with a probability of approximately $\frac{10}{13} \approx 0.8.$

This video works through how we can use the Law of Total Probability and Bayes' Theorem to answer the questions in Example 6.

Video

Bayes' Theorem

Duration 14:14



The intuition behind Bayes' theorem becomes clearer when we state it in terms of odds, rather than in terms of probabilities. Using Bayes' theorem in both the numerator and the denominator we can calculate the conditional ("posterior") odds,

$$\frac{P(B|A)}{P(B'|A)} = \frac{\frac{P(A|B)P(B)}{P(A)}}{\frac{P(A|B')P(B')}{P(A)}} = \frac{P(A|B)P(B)}{P(A|B')P(B')} = \underbrace{\frac{P(B)}{P(B')}}_{\text{prior odds}} \times \underbrace{\frac{P(A|B)}{P(A|B')}}_{\text{likelihood ratio}}.$$

We can see that the conditional odds are a product of two parts:

- the prior odds, $\frac{P(B)}{P(B')}$, which contains the information how likely B is in general, and
- the likelihood ratio, $\frac{P(A|B)}{P(A|B')}$, which compares the probability of observing A under both hypotheses B and B'. The likelihood ratio tells us to what extent observing A provides evidence in favour of B.

In other words, Bayes' theorem provides a principled way of combining prior beliefs with evidence from observed facts. Bayes' theorem and the combination of prior information and likelihood will be central to the course on *Uncertainty Assessment and Bayesian Computation*.

Background 1

Odds

So far, we have expressed the likelihood of an event B happening in terms of the probability $P(B)$. Historically (and nowadays still in gambling) this likelihood is quantified by the the odds in favour of an event B

$$o_B = \frac{P(B)}{P(B')} = \frac{P(B)}{1 - P(B)}$$

The odds are often expressed as a ratio of integers, so for example $P(B) = 0.6$, corresponds to odds

$$o_B = \frac{P(B)}{P(B')} = \frac{0.6}{0.4} = \frac{3}{2} = 1.5$$

which would often be written as odds of $3:2$ in favour of the event B.

Odds can be converted to probabilities and vice versa, with the reverse formula being

$$P(B) = \frac{o_B}{1 + o_B}.$$

The payout odds quoted by bookmakers do not quite correspond to probabilities, as they also include an overround, which yields the profit of the bookmaker.

Odds and odds ratios play an important role in the interpretation of logistic regression models, which will learn more about in *Advanced Predictive Models*.

Example 10

Application of Bayes' theorem in terms of odds

In our medical testing example (Example 6),

$$\frac{P(B|A)}{P(B'|A)} = \frac{P(B)}{P(B')}\,\frac{P(A|B)}{P(A|B')} = \frac{0.1}{0.9}\,\frac{0.995}{0.03} = (0.111)(33.167) = 3.685$$

So, whilst the prior odds (0.111) favour the hypothesis that the patient does not have the disease, the likelihood ratio (33.167) clearly favours the hypothesis that the patient has the disease.

Supplement 4

The Monty Hall Problem

The video below describes a very famous probability puzzle, the Monty Hall problem, and shows how we can use Bayes' theorem to solve this paradoxical problem.

Video

Bayes' Theorem and the Monty Hall Problem

Duration 4:55

Task 4

It is known that any email which arrives in your inbox has a probability of 0.1 of being spam mail. Your email provider has developed an algorithm which tries to classify each email based on whether or not it contains certain high risk words. If it is deemed to be spam, it will be filtered into your spam folder. The algorithm has the following sensitivity and false positive rates

$$P(\text{email contains high risk words}|\text{email is spam}) = 0.95.$$

$$P(\text{email contains high risk words}|\text{email is not spam}) = 0.05.$$

(a) What is the probability that an email is identified as spam? (b) What is the probability that an email that has been identified as spam actually is a spam email?

Coincidences

Coincidences attract our attention because they appear to be strange or unlikely - how often do you find yourself wondering "what are the chances of this happening?". However, from a probabilistic perspective, coincidences are often quite likely to occur and less remarkable than they appear. The concepts introduced in this week's material should provide you with some of the tools to compute the probability of unlikely or strange events occurring.

Example 11

Coincidences

Birthday problem

You are in a room with 30 people and find that there are 2 people who share the same birthday. You may think that this is strange however there is in fact a 70% chance that this would happen so it is actually a likely event.

Why is the probability so high?

Answer

The probability of two people having different birthdays is

$364/365 = 0.997260$

In other words, when comparing one persons birthday to another, in 364 out of 365 scenarios they won't match.

However, we have 30 people in the room so we are now comparing everyone's birthday to everyone else in the room. i.e. we are making many comparisons. To figure out the number of comparisons we are making we can use the **combination** formula that was introduced last week.

$$\begin{aligned}\binom{n}{r} &= \frac{n!}{(n-r)!r!} \\ &= \frac{30!}{28!2!} \\ &= 435\end{aligned}$$

The chance of getting a single miss is high (99.72%), however the chance of missing every single time when we make 435 comparisons is

$$\frac{364}{365}^{435} = 0.3032.$$

Here we have shown that what initially seemed like an unusual event isn't actually so odd after all.

For a more comprehensive explanation, see Dobrow, Robert P. - **Probability: With Applications and R**, Chapter 2.

Learning outcomes for week 2

By the end of week 2, you should be able to:

- know the definition of independence and conditional probability, and how these concepts relate;
- calculate conditional probabilities;
- determine independence of events;
- apply the law of total probability and Bayes' theorem to examples.

A summary of the most important concepts, selected video solutions and written answers to all tasks are provided overleaf.

Week 2 summary

Conditional probability

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

Or equivalently

$$P(A \cap B) = P(A|B)P(B) \quad \text{or} \quad P(A \cap B) = P(B|A)P(A)$$

Independence

Events A and B are **independent** if

$$P(A|B) = P(A) \quad \text{or} \quad P(A \cap B) = P(A)P(B),$$

Mutual independence

Events A,B and C are **mutually independent** if

1. $P(A \cap B \cap C) = P(A)P(B)P(C)$ and
2. $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, and $P(B \cap C) = P(B)P(C)$.

Law of total probability

If $\{E_1, E_2, \ldots, E_k\}$ are a partition of $\mathcal{S}$, then for any event A the **law of total probability** states that

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k).$$

For an event B (with $0 < P(B) < 1$) and its complement,

$$P(A) = P(A|B)P(B) + P(A|B')P(B').$$

Bayes' Theorem

If $\{E_1, E_2, \ldots, E_k\}$ are a partition of $\mathcal{S}$ and A is any event with $P(A) > 0$, then

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \cdots + P(A|E_k)P(E_k)}$$

For an event B (with $0 < P(B) < 1$) and its complement,

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

Answer 1

Ice cream:

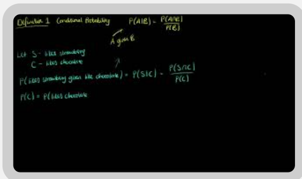
$$\begin{aligned} P(\text{like strawberry}|\text{like chocolate}) &= \frac{P(\text{like chocolate} \cap \text{like strawberry})}{P(\text{like chocolate})} \\ &= \frac{0.35}{0.7} \\ &= 0.5. \end{aligned}$$

Here is a video worked solution.

Video

Week 2 - Task 1

Duration 3:22



Answer 2

Drawing one card from a deck:

(a) We know that $P(A) = 4/52 = 1/13$, $P(A|B) = 1/13$.

We have

$$P(A|B) = P(A)$$

and so events A and B are independent. That is, whether the card is a king is independent of whether the card is a diamond.

Notice that $A \cap B = \{\text{card selected is the king of diamonds}\} \neq \emptyset$ and so the events are not disjoint.

(b) Similarly to (a),

$$P(A|C) = 2/26 = 1/13,$$

and so events A and C are independent. That is, whether the card is a king is independent of whether the card is red.

(c) We have

$$P(B) = 1/4$$

and

$$P(B|C) = 1/2$$

so

$$P(B|C) \neq P(B)$$

and so events B and C are not independent.

This could also have been shown the other way round, i.e.

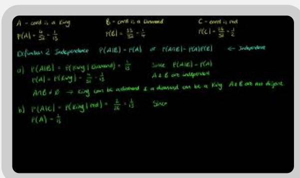
$$P(C) = 1/2$$

and

$$P(C|B) = 1.$$

Here is a video worked solution.

Video



Week 2 - Task 2

Duration 7:36

Answer 3

Let

$$R = \{\text{chosen marble is red}\}$$

and

$$B_i = \{\text{Bag } i \text{ is chosen}\}.$$

We already know that the probability of choosing each bag is equal, i.e.

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

and

$$P(R|B_1) = 0.75,$$

$$P(R|B_2) = 0.60,$$

$$P(R|B_3) = 0.45.$$

Our partition is $\{B_1, B_2, B_3\}$, which is valid since

1. each pair of B_i's are disjoint (only one of them can happen);
2. their union is the entire sample space as one of the bags will definitely be chosen, i.e. $P(B_1 \cup B_2 \cup B_3) = 1$.

Using the law of total probability, we can now calculate the probability that a red marble is chosen as

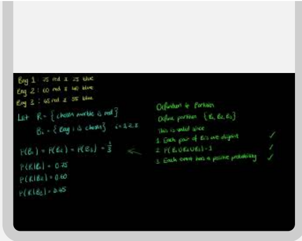
$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\ &= 0.75 \cdot \frac{1}{3} + 0.60 \cdot \frac{1}{3} + 0.45 \cdot \frac{1}{3} \\ &= 0.60 \end{aligned}$$

Here is a video worked solution.

Video

Week 2 - Task 3

Duration 5:08



Answer 4

Spam filtering:

Firstly, let's define our partition. Of all the emails received to your inbox, either an email will be spam or it won't be, so define the partition

$$E_1 = \{\text{email is spam}\},$$

$$E_2 = \{\text{email is not spam}\},$$

and let

$$A = \{\text{email contains high risk words}\}.$$

Then the probabilities we know are:

$$\begin{aligned} P(A|E_1) &= 0.95\,, \quad P(E_1) = 0.1, \\ P(A|E_2) &= 0.05\,, \quad P(E_2) = 1 - P(E_1) = 0.9. \end{aligned}$$

(a) By the law of total probability we have

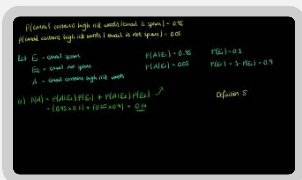
$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) \;=\; 0.95 \cdot 0.1 + 0.05 \cdot 0.9 \;=\; 0.14.$$

(b) Using Bayes' theorem

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1) + P(A|E_2)P(E_2)} = \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.05 \cdot 0.9} \approx 0.68.$$

Here is a video worked solution.

Video



Week 2 - Task 4

Duration 5:29

Footnotes

1. This example was taken from Dobrow, Robert P. Chapter 2, Probability: With Applications and R. https://epdf.pub/queue/probability-with-applications-and-r.html ↩