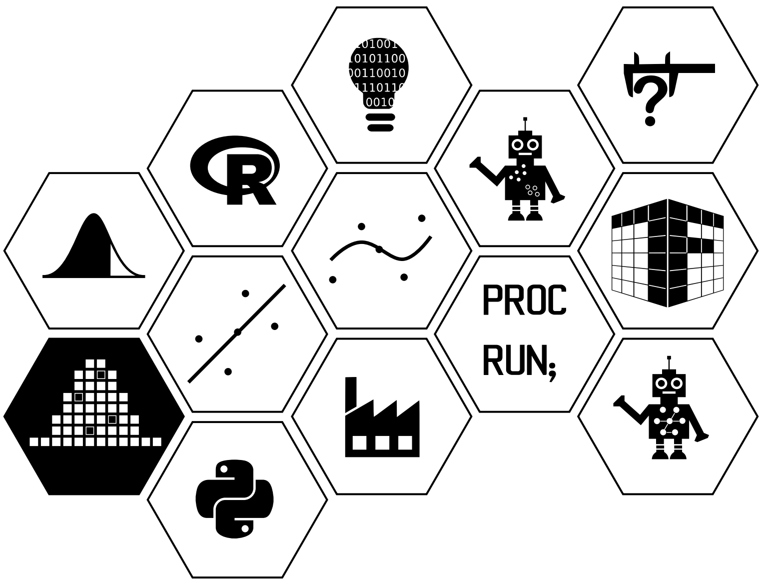
Probability and Sampling Fundamentals

Week 1: Introduction to Probability



DATA ANALYTICS
GLASGOW

Introduction to Probability

Week 1 learning material aims

The material in week 1 covers:

- Definition of a sample space;
- Counting the number of elements in a set;
- The multiplication principle, permutations and combinations;
- Definition of a probability;
- Axioms of probability.

Video

Probability and Sampling Fundamentals - Week 1

Duration 1:38

Introductory example

In this course we will focus on properties of probability and sampling. In order to demonstrate concepts of probability, we will adopt a wide range of everyday examples that will be developed throughout this course.

Example 1

Blood Type

It is known that human blood is classified in four types denoted by O, A,B and AB and each of these groups can be rhesus positive or rhesus negative. This means there are actually **eight blood groups** but we are interested in four (O, A, B, AB) and so we will aggregate positive and negative within each type.

Suppose 100 people donated blood and were classified in these four categories. **How many people would you expect to fall into each of the four categories?**

Suppose you know nothing about blood type, then a reasonable guess may be that $\frac{1}{4}$ of the 100 people fall into each of the four categories. That is, 25 people fall into each of the four categories.

On the other hand, after some investigation (**NHS Scotland**), you discover that in 2019 in Scotland

Type	+/-	Scottish population %	Implied Sample %	Number of sample out of 100
O	+	40.9	40.9	41
	-	9.5	9.5	9
A	+	28.8	28.8	29
	-	6.3	6.3	6
B	+	9.2	9.2	9
	-	2.0	2.0	2
AB	+	2.7	2.7	3
	-	0.6	0.6	1

With this in mind, you expect 29+6=35, 9+2=11, 3+1=4 and 41+9=50 people to fall into each of the four categories A, B, AB, and O respectively.

In this example, there are a few things to consider

1. If you knew nothing about blood type then it is perfectly reasonable to divide the 100 people equally into each category. That is, you have assumed that each of the four categories are equally likely.
2. We discovered that some blood types are more prevalent in Scotland in comparisons to others. That is, there are more people in Scotland with blood type O+ than AB-. In order to derive a number rather than a percentage, we just calculated each percentage out of 100 and rounded to find the number of people in each category (see 'Number of sample out of 100' column in table). We had to round numbers since it's impossible to have 40.9 people,

instead we expected 41. In other words, our answers had to be integers. Given that the question made no mention of positive or negative blood types, it was reasonable to add the '+' and '-' entries for each of the four blood types. For example adding 41 O+ people to 9 O- people to give an expected 50 O people.

3. Using the population percentages, we assumed that these same percentages would apply to our sample of 100 people (see 'Implied Sample Percentage' column in table). Is it reasonable to assume that these same percentages would apply to our sample of 100 people? Under what circumstances would this not be true? If I told you that actually the 100 people who donated were all from the same family, would this have any impact on your 'Implied Sample Percentage' column? We know that our blood type is **inherited from our parents**. That is, two people from the same family are more likely to have similar blood types (under certain conditions) than two unrelated people. In other words, people from the same family don't have independent blood types.

4. We are assuming that everyone *must* fall into one, and only one, of the four categories A,B, AB, and O and there are no unknown blood types. Consequently, the numbers in our final column 'Number of sample out of 100' *must* sum to 100.

5. Notice that we referred to our answer as 'expected' numbers. That is, in a sample of 100 people we would expect 50 of them to have blood type O. In reality, if you conducted this experiment it's likely that not exactly 50 people who donated will have blood type O, however, the number of people will likely be 'around' 50. More over, if you repeated this experiment again, on a different day with a different 100 donations, it's likely you would not have the same numbers as in your first experiment. It's likely that in both samples the number of people with blood type O will be around 50 but not exactly. This is due to random variation or chance. The first set of 100 people will not be exactly the same as the second set of 100 people. Likewise, the proportions of each blood type need not be identical with those in the Scottish population (Scotland's population size in 2019).

6. The question also made no mention of Scotland. Blood type percentages are different in different countries, for example compare to **America**. In addition to country, blood type depends on ethnicity and change across time as populations change. So even within Scotland, the 'Population Percentage' column will be different depending on the year of interest. What we really want to know is do the 100 donations belong to a well defined population and how well do the 100 people represent that population. In other words, how confident are we that the 'Population Percentage' and 'Implied Sample Percentage' columns are actually the same?

Points 1 - 5 neatly lend themselves to fundamental probability concepts whereas point 6 lends itself to sampling theory we will see towards the end of this course.

What is a probability?

Probability in itself is a hard concept to define. If you **looked up probability in a dictionary** you may find several definitions such as

1. a measure of the likelihood that the event will occur,
2. a probable event or circumstance,
3. the chance that a given event will occur, or
4. the relative frequency that an event will occur, as expressed by the ratio of the number of actual occurrences to the total number of possible occurrences.

Definitions 1 - 3 are circular since 'likelihood', 'probable' and 'chance' can be considered as synonyms for probability. Definition 4 seems robust.

Example 2

Rolling a dice

Suppose I roll a dice one time. What is the probability that I roll a 6?

You may think logically that since a standard dice has six sides labeled 1, 2, 3, 4, 5 and 6 and so it is equally likely to land on each of the six sides. Therefore, the probability that I roll a 6 is $\frac{1}{6}$.

This statement makes alto of assumptions. What if I didn't tell you that actually my dice has six sides labelled 1, 2, 3, 4, 6 and 6. Then it's more likely that I roll a six in comparison to a conventional dice (labelled 1, 2, 3, 4, 5 and 6) because there are two possible ways for me to roll a 6 as opposed to just one.

You may then ask me to roll my dice say 100 times before you answer my question. Reasonably, you may then deduce that

$$P(\text{rolling a 6 with my unfair dice}) = \frac{\text{number of times a 6 appears}}{\text{total number of repetitons}} = \frac{\text{number of times a 6 appears}}{100}.$$

Things to note

1. The final probability statement

$$P(\text{rolling a 6 with my unfair dice labelled } 1, 2, 3, 4, 5, 6, 6)$$

is commonly used notation for a probability. We normally write P to mean probability and the event of interest, in this example rolling a 6 with my unfair dice, inside brackets $()$.

2. We can see that probability is a real number. We found $P(\text{rolling a } 6) = \frac{1}{6}$ and $P(\text{rolling a 6 with my unfair dice})$ = number times a 6 appears divided by the total number of repetitions, which we may expected to be $\frac{2}{6}$. Both of these are positive real numbers.

3. As with the **blood type** example, the inferred probability depends on the samples. For instance, how many times should I roll the dice for you to be sure of your answer? 100 times? Or 1000 times? Or even 1,000,000 times? Luckily, we can assume as the number of repetitions increases, the frequency of each number 1, 2, 3, 4 and 6 will stabilize.

Point 3 will become more clear in week 8. Definition 4 above relies on us knowing all possible outcomes from an experiment (like rolling a dice) and the number of actual occurrences of our outcome of interest is stable in the long run (like the number of rolling repetitions).

Definition 1

Sample space and events.

The set of all possible outcomes in an experiment is called the **sample space** and is commonly denoted by S. **Events** are any possible subsets of a sample space.

Sample spaces can fall into one of three categories

1. A space with a finite number elements.
2. A space with an infinite but countable number of elements.
3. A space with an infinite and uncountable number of elements.

Example 3

Finite sample space

Suppose now we did have a dice (with six sides labelled 1, 2, 3, 4, 5 and 6) that we rolled once. Then we necessarily would roll a 1, 2, 3, 4, 5 or 6. Therefore,

$$S = \{1, 2, 3, 4, 5, 6\}.$$

An event E is any possible subset of S, hence we could define

$$
\begin{aligned}
&E = \{1\}, \text{ or} \\
&E = \{1, 3\}, \text{ or} \\
&E = \{2, 4, 5, 6\}, \text{ or} \\
&E = \{1, 2, 3, 4, 5, 6\}, \text{ etc}
\end{aligned}
$$

Example 4

Countable sample space

Suppose I asked you to count the number of people you see in one day including people you know, people you don't know, people you see at a distance, people who don't see you etc.

Here, the outcome would be an integer and the sample space could be all non-negative integers $\mathbb{Z} = \{0, 1, 2, \ldots\}$. Assuming you spend a lot of time outside in public areas, this number will be quite large but still not technically infinite in the sense we could put an upper limit on this sample space. I don't think it's intuitive to what that upper limit should be. Therefore we can safely set

$$S = \mathbb{Z}.$$

An event E is any possible subset of S, hence we could define

$$
\begin{aligned}
&E = \{x \in \mathbb{Z} | x > 10\}, \text{ or} \\
&E = \{1, 3\}, \text{ or} \\
&E = \{x \in \mathbb{Z} | x \leq 100\}, \text{ or} \\
&E = \{x \in \mathbb{Z} | x \geq 5000\}, \text{ etc}
\end{aligned}
$$

Example 5

Uncountable sample space

Suppose I asked you to estimated the average height of people in Scotland. In this case, your answer will be a non-negative real number. This is different to the previous example since the number doesn't have to be an integer, it can be any positive real number ($\mathbb{R}_{>0}$) Again, its reasonable to assume this number will be bounded with a **lower limit** and an **upper limit**. Therefore, we might assume

$$S = \{x \in \mathbb{R}_{>0} | a \le x \le b\}.$$

for some careful chosen a and b measured in centimeter, say. An event E is any possible subset of S, hence we could define

$$\begin{aligned}
E &= \{a \le x \le b \in \mathbb{R}_{>0} | x \le 100\} \\
&= x \text{ is between positive real numbers } a \text{ and } b \text{ but } x \text{ must be less than or equal to } 100 \text{ or} \\
E &= \{a \le x \le b \in \mathbb{R}_{>0} | x \le 200\} \\
&= x \text{ is between positive real numbers } a \text{ and } b \text{ but } x \text{ must be less than or equal to } 200 \text{ or} \\
E &= \{a \le x \le b \in \mathbb{R}_{>0} | x \ge 120\} \\
&= x \text{ is between positive real numbers } a \text{ and } b \text{ but } x \text{ must be greater than or equal to } 120 \text{ etc}
\end{aligned}$$

For more information on set notation and terminology, please refer to the **preliminary maths course**.

Properties of sets

In all examples above, S and E are examples of sets with E a subset of S which we represent with the notation $E \subseteq S$. There are some properties of sets that are worth noting. For some background on set notation and Venn diagrams, please refer to the [**preliminary mathematics course**]

1. A **set** is a well-defined collection of distinct objects. Objects within a set are called elements and elements can be anything that is well-defined. For example numbers, dates, letters, people, any kind of object etc.

2. A set can be finite, infinite, countable, uncountable or empty. The empty set is commonly denoted by $\emptyset$.

3. The union $E \cup F$ of two sets E and F contains all elements that are in E or in F (or in both). In the language of events, the union corresponds to the event that E or F (or both) occur.

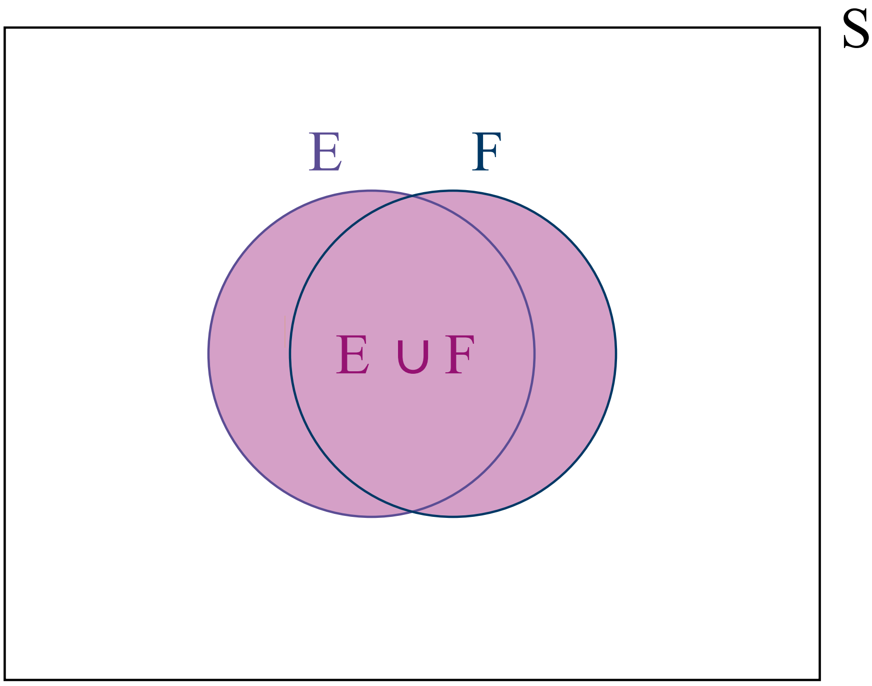


Figure 1

4. The intersection $E \cap F$ of two sets E and F contains all elements that are both in E and in F. In the language of events, the intersection corresponds to the both event E and the event F occurring.

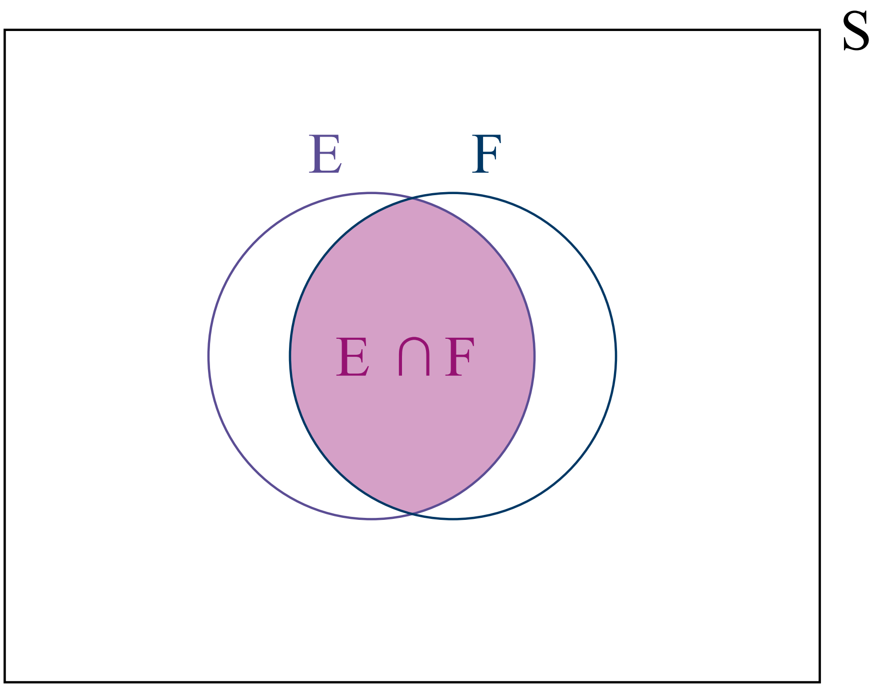


Figure 2

5. The complement of a set E contains all elements that are not in E. The complement is commonly denoted by E'. In the language of events, the complement E' corresponds to the event that E does *not* occur.

6. The difference between two sets, E and F, denoted by $E \backslash F$ is a set that contains all the elements in E that are not in F. In the language of events, this corresponds to the event E occurring, but not F.

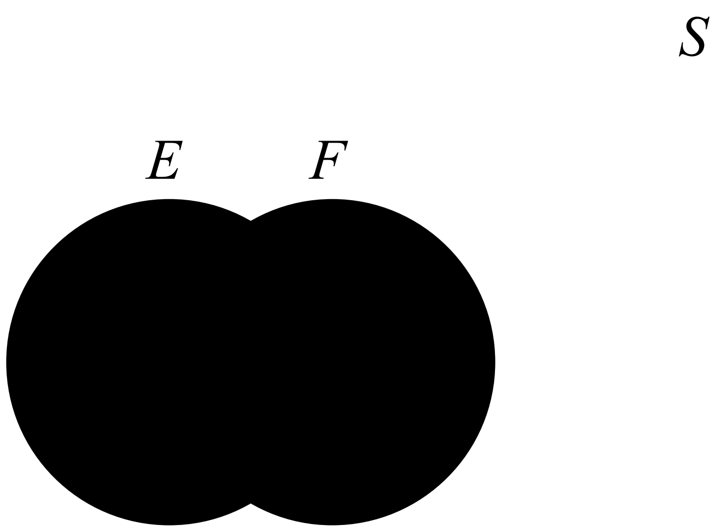


Figure 3

7. Two sets E and F are disjoint if $E \cap F = \emptyset$. In the language of events, this means that the two events are mutually exclusive, i.e. they cannot occur at the same time.

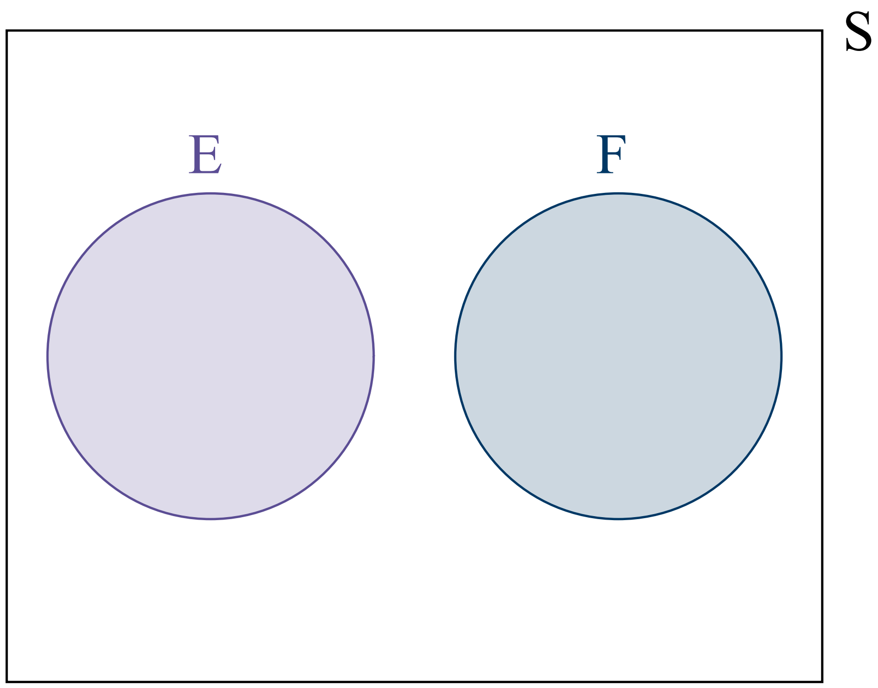


Figure 4

8. For any two sets E and F,

$$E \cap F = F \cap E \text{ and } E \cup F = F \cup E.$$

9. For any sets E, F and H,

$$(E \cap F) \cap H = E \cap (F \cap H)$$

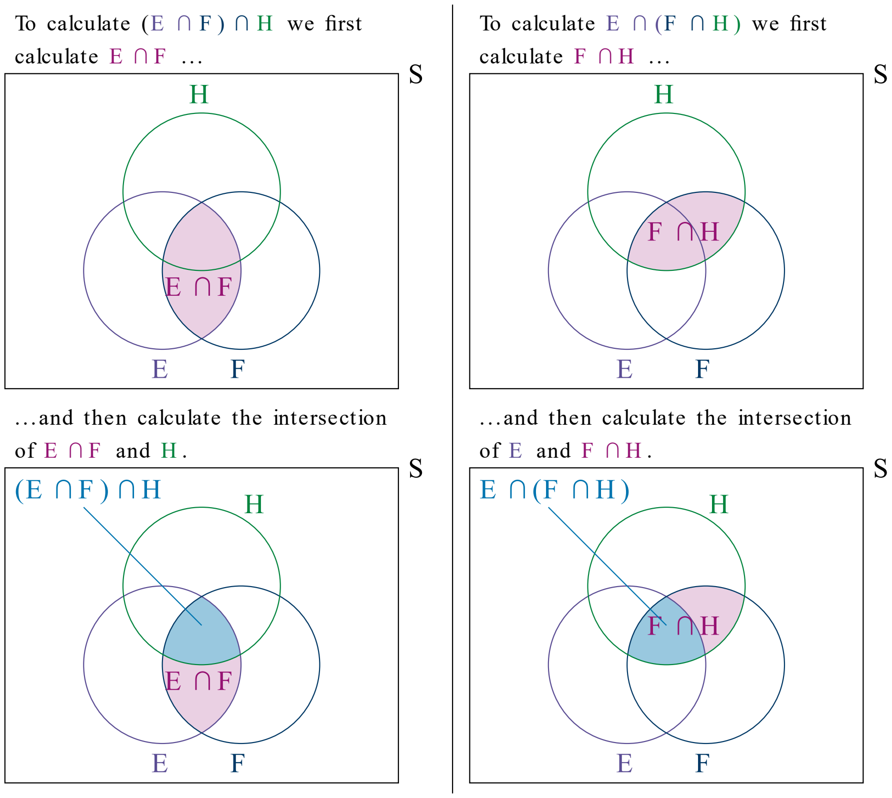


Figure 5

10. Likewise, for any sets E, F and H,

$$(E \cup F) \cup H = E \cup (F \cup H).$$

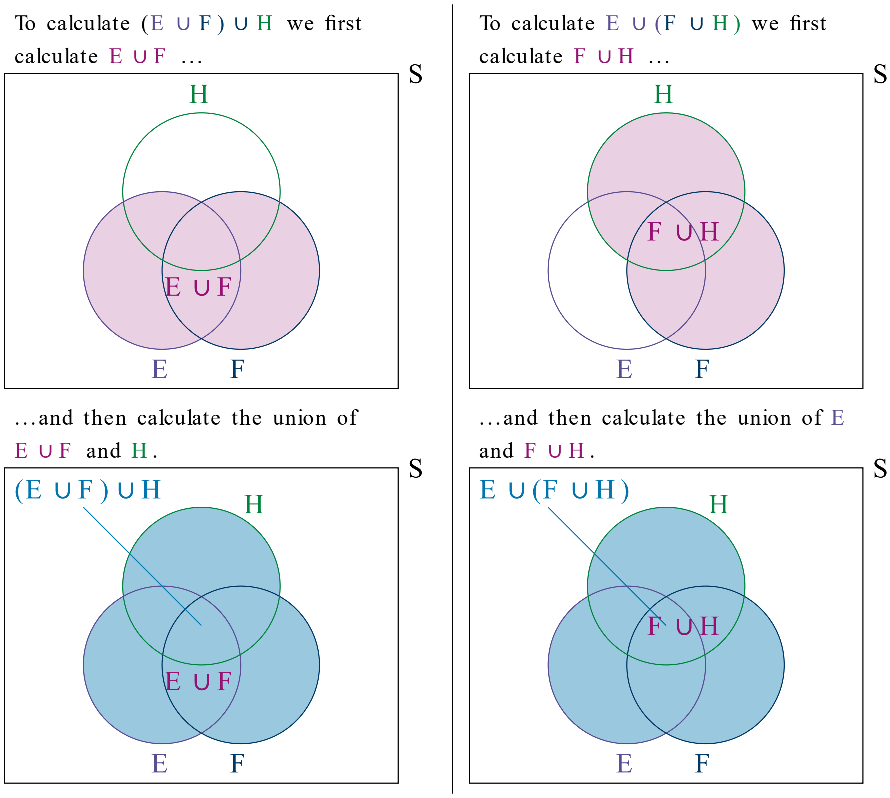


Figure 6

11. The union of sets is distributive over the intersection of sets.

$$E \cup (F \cap H) = (E \cup F) \cap (E \cup H).$$

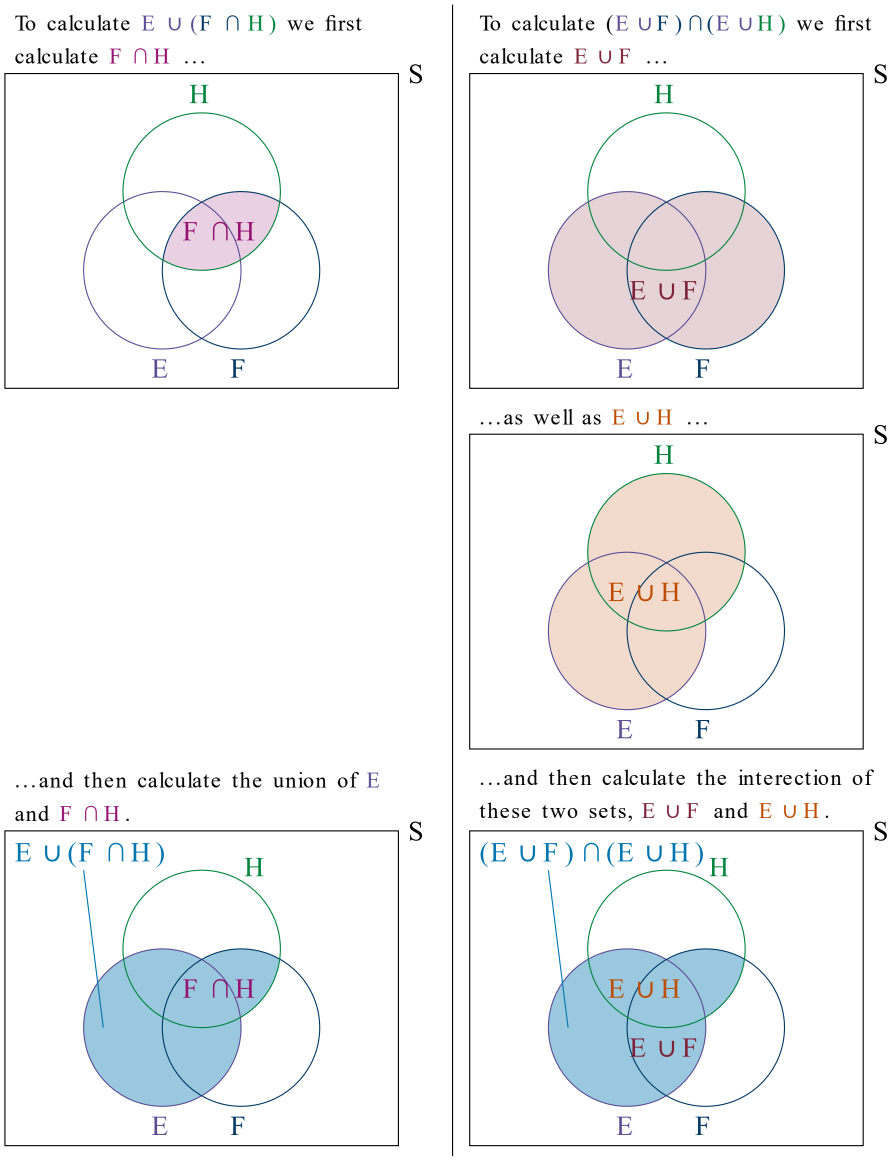


Figure 7

12. Likewise, the intersection of sets is distributive over the union of sets.

$$E \cap (F \cup H) = (E \cap F) \cup (E \cap H).$$

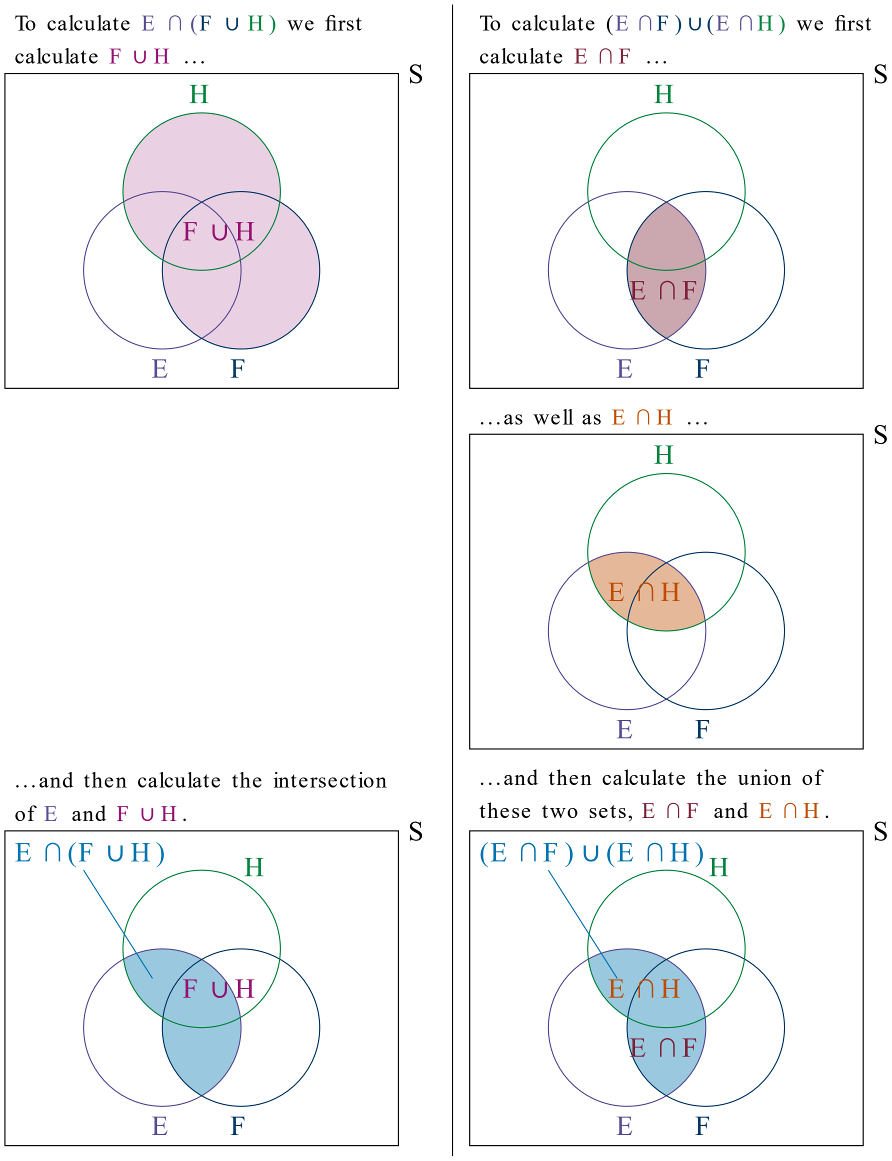


Figure 8

13. **De Morgan's law**. The complement of the union of two sets equals the intersection of their complements, i.e.

$$(E \cup F)' = F' \cap E').$$

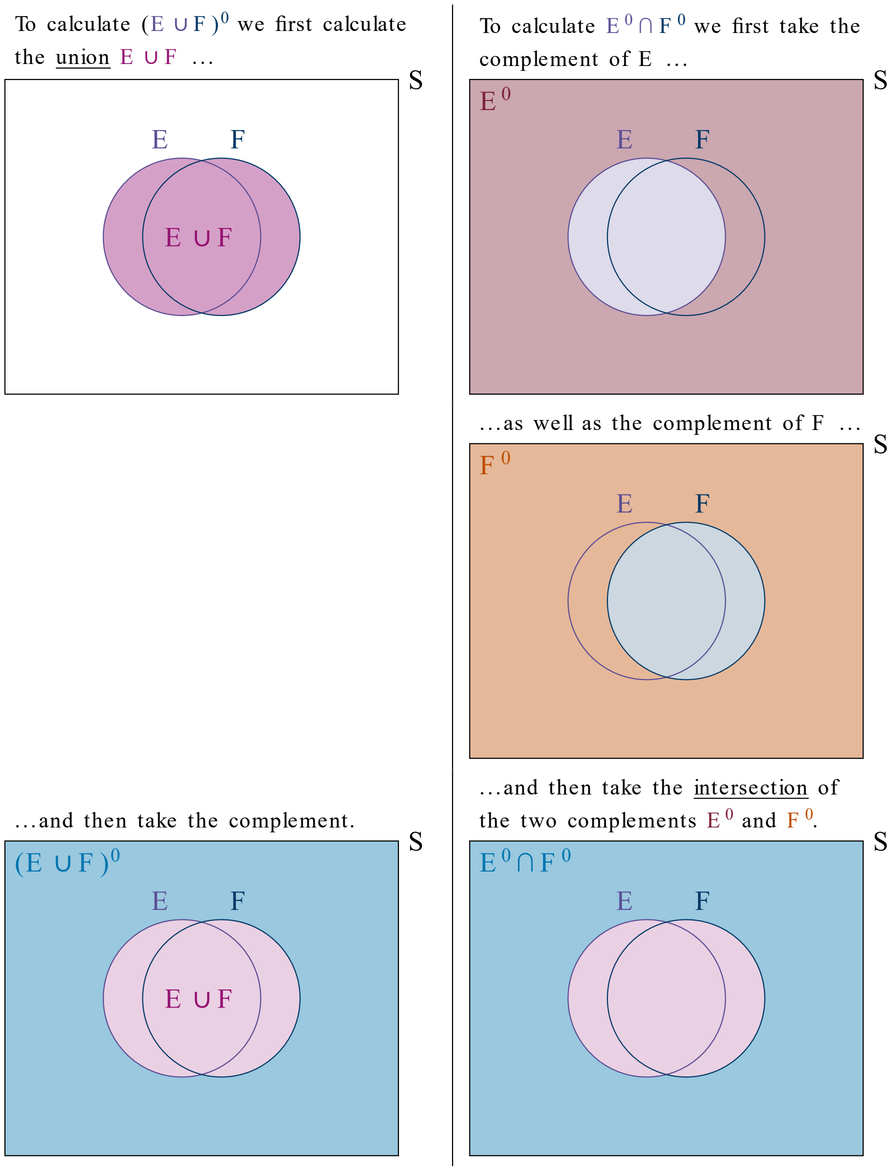


Figure 9

14. De Morgan's laws also state that the complement of the intersection of two sets equals the union of their complements, i.e.

$$(E \cap F)' = E' \cup F'.$$

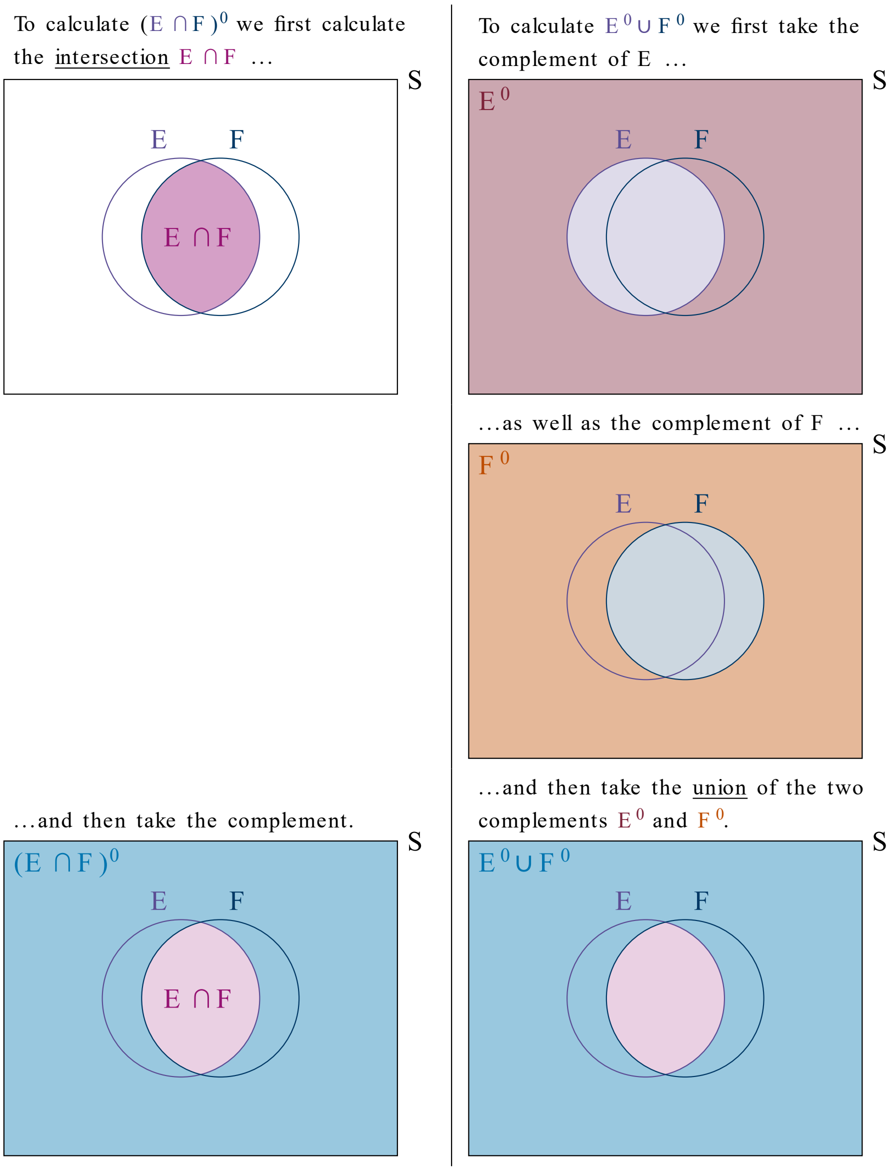


Figure 10

Task 1

Blood Type

Referring back to the blood type example. Write down the sample space S.

Counting samples

Hopefully you can see that in some cases the sample space in an experiment can easily be written down. However, this is not always the case and we need some additional theory in counting samples.

Example 6

Briefcase lock

Suppose you buy a lock for a briefcase that contains three dials and you now want to set the lock. Each dial can be set to a number between 0 and 9 and the sample space is all possible lock combinations. How many possible ways can you set the lock?

You may begin by writing down some possible lock combinations in a systematic way,

$$S = \{000, 001, 002, \ldots, 009, 010, 020, \ldots, 090, \ldots\ldots\ldots, 990, 991, \ldots, 998, 999\},$$

but quickly realise that each dial has ten possible choices: 0, 1, 2, 3, 4, 5, 6, 7, 8 or 9 and the value at one dial has no bearing on the value at another dial. You notice that for each choice of dial 1, there are 10×10 possible combinations of dial 2 and dial 3. Likewise, for each choice of dial 2, there are 10×10 possible combinations of dial 1 and dial 3 and lastly for each choice of dial 3, there are 10×10 possible combinations of dial 1 and dial 2. Putting everything together:

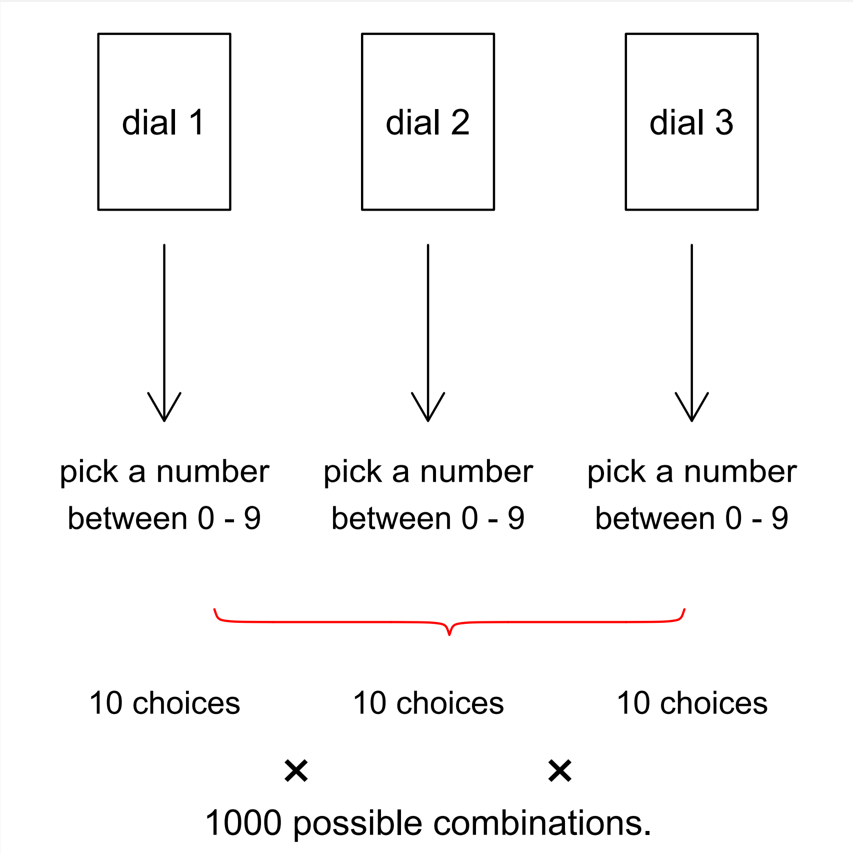


Figure 11

A few notes about this example

1. We came to the conclusion that the value at one dial has no bearing on the value at another dial. That is to say the three dials are independent.
2. Since the three dials were independent and each dial had 10 possible outcomes, we came to the conclusion that there were $10 \times 10 \times 10 = 10^3$possible ways to set the lock.

Both of these observations lead us to formally define independence and the multiplication principle.

Multiplication principle

START HERE

Definition 2

Independent events

Two events E_1 and E_2 are independent if the outcome of E_1 does not affect the outcome of E_2 and the outcome of E_2 does not affect the outcome of E_1.

Independence is an important concept in probability and statistics. It is something we often assume. In the **blood type** example, we used Scotland's population level percentages to infer what we might expect in the 100 donors assuming that the 100 blood donors were independent of each other. That is, knowing the blood type of one donor does not affect the probabilities of the blood type of another donor. However, as noted, if two donors were related then we can no longer assume they are independent since we know the blood type of the first is related to the blood type of the second.

Example 7

Disease outcome

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

What is the probability that a randomly selected person is infected with this disease? Well, there is a 10% chance (probability = 0.1) that any randomly selected person is infected with this disease.

A diagnostic test is available such that

1. 97% of people without the disease will test negatively,
2. 99.5% of people with the disease will test positively.

A randomly selected person tested positively for this disease, what is the probability that they are infected?

Without making any calculations, hopefully you can see that knowing this person has tested postie for this disease increases their probability of actually being infected. It's not completely certain that they are infected with the disease since the test isn't 100% accurate but the chance that they are infected is greater that 10%.

Overall, knowing the test status of a randomly selected person influences the chance of that person being infected with the disease. Therefore, testing positively (or negatively) and being infected (or not) with this disease are not independent events.

Definition 3

Multiplication principle

Suppose we have n independent events $E_1, E_2, \ldots, E_n$. If event E_k has m_k possible outcomes, or elements, for $k = 1, \ldots n$, then there are

$$m_1 \times m_2 \times \ldots \times m_n$$

possible ways for events $E_1, E_2, \ldots, E_n$ to occur.

Example 8

Briefcase lock

The three dials on the lock are independent since knowing the value of dial 1 had no affect on the value of dial 2 or dial 3. Knowing the value of dial 2 had no affect on the value of dial 1 or dial 3. Lastly, knowing the value of dial 3 had no affect on the value of dial 1 or dial 2.

Using the multiplication principle, there are $10 \times 10 \times 10$ possible ways to set the lock.

Example 9

Rolling a dice three times

Suppose we rolled a dice three times. How many possible outcomes are there?

We know if we rolled a (fair) dice one time, there are 6 possibles outcomes: 1, 2, 3, 4, 5 or 6. It's safe to assume that the three rolls are independent therefore using the multiplication principle, there are

$$6 \times 6 \times 6 = 6^3 = 216$$

possible outcomes.

Task 2

Multiple choice exam

Suppose you take a multiple choice exam that contains 20 questions. Each question has 5 possible answers and you are permitted to give only one answer per question.

How many different ways could you randomly answer the 20 questions?

The multiplication principle is useful in the general setting of counting the number of ways to arrange independent events. However, suppose now that we want to set our briefcase lock but we cannot repeat a digit. In other words, we must have dial 1 $\neq$ dial 2 $\neq$ dial 3.

Example 10

Briefcase lock

Suppose you buy a lock for a briefcase that contains three dials and you now want to set the lock. Each dial can be set to a number between 0 and 9, however, each dial number must be set to a unique number such that dial 1 $\neq$ dial 2 $\neq$ dial 3. How many possible ways are there to set the lock?

You begin by writing down some possible lock combinations in a systematic way,

$$S = \{012, 013, 014, \ldots, 091, 092, 093, \ldots, 098, \ldots\ldots\ldots, 980, 981, \ldots, 986, 987\},$$

You notice that if you set the value of dial 1 first, then there are 10 possible values for dial 1 $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Let x_1 denote the value of dial 1.

However, given you have fixed the value at dial 1, x_1, you notice there are now nine possible values for dial 2 $\{0,1,2,3,4,5,6,7,8,9\}\backslash x_1$. Let x_2 denote the value of dial 2 such that $x_1 \neq x_2$.

Given that you fixed the values at dial 1 and dial 2 to x_1 and x_2 respectively, you notice there are now eight possible values for dial 3 $\{0,1,2,3,4,5,6,7,8,9\}\backslash\{x_1, x_2\}$. Let x_3 denote the value of dial 3 such that $x_1 \neq x_2 \neq x_3$.

Therefore, given $x_1 \neq x_2 \neq x_3$, you conclude that there are

$$10 \times 9 \times 8 = 720$$

possible ways to set the lock.

A few notes about this example

1. Instead of 1000 possible ways to set the lock, there are now 720 ways. This is because we are excluding lock combinations such as $000, 001, \ldots, 009, 010, 011, \ldots\ldots, 999$ etc.
2. In other words, what we are really asking is, from a list of 10 elements, how many ways can we choose three unique elements such that the order matters. Or more generally, from a list of n elements, how many ways can we choose r unique elements such that $r \leq n$ and the order matters.
3. The order matters here because if I set the lock combination to 012 then 201 or 021 will not open the lock.

This example leads us onto the definition of a permutation.

Permutation

Definition 4

Factorial

Suppose we have n elements. Then

$$n! = n \times (n-1) \times (n-2) \times (n-3) \times \ldots 2 \times 1.$$

A few notes about this definition

1. $n!$ is common notation for n factorial.

2. For completeness, 0!=1.

Example 11

Factorial

Below are just a few examples of $n!$ for $n = 2, 5$ and 10.

1. $2! = 2 \times 1 = 2$
2. $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
3. $10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800$

Definition 5

Permutation

Suppose we have n elements. The number of ways to select r unique elements such that $r \leq n$ and order matters is

$$\begin{aligned} nPr = P(n,r) &= \frac{n!}{(n-r)!} \\ &= n \times (n-1) \times (n-2) \times \ldots \times (n-r+1) \end{aligned}$$

Both nPr and $P(n,r)$ are common notations for the number of permutations of r elements from n elements. Since the order matters, $P(n,r)$ tells us the number of distinct ways to select r objects from n elements.

Re-visiting the lock example,

Example 12

Briefcase lock

Suppose you buy a lock for a briefcase that contains three dials and you now want to set the lock. Each dial can be set to a number between 0 and 9, however, each dial number must be set to a unique number such that dial 1 $\neq$ dial 2 $\neq$ dial 3. How many possible ways are there to set the lock?

We are choosing 3 numbers from the set of 10 such that the order matters, therefore

$$
\begin{aligned}
P(10,3) &= \frac{10!}{(10-3)!} \\
&= \frac{10!}{7!} \\
&= \frac{10 \times 9 \times 8 \times 7 \times 6 \times \ldots \times 1}{7 \times 6 \times \ldots \times 1} \\
&= 10 \times 9 \times 8 \\
&= 720
\end{aligned}
$$

Again, write this down

Task 3

Winning a race

P(30,3) = 30!/(30-3)! = 24,360

Suppose there are 30 competitors in a race. Those who finish 1st, 2nd and 3rd will be awarded a gold, silver and bronze medal respectively.

How many possible ways could the gold, silver and bronze medals be allocated?

Notice in the definition of a permutation, the order of the r elements was important. For example, in the briefcase lock example it was clearly important to enter the lock combination in the correct way.

Let's be a bit more flexible and suppose that we now want to know the number of ways to select r unique elements such that $r \leq n$ and the order **does not matter**.

Example 13

Briefcase lock

You buy a lock for a briefcase that contains three dials and you now want to set the lock. Each dial can be set to a number between 0 and 9, however, each dial number must be set to a unique number such that dial 1 $\neq$ dial 2 $\neq$ dial 3. How many ways could you select three numbers from $\{0,1,2,3,4,5,6,7,8,9\}$ that may then be used to set the lock?

You begin by writing triplets in a systematic way,

$$
\begin{array}{llllllll}
012, & 013, & 014, & 015, & 016, & 017, & 018, & 019,\\
023, & 024, & 025, & 026, & 027, & 028, & 029, & \\
034, & 035, & 036, & 037, & 038, & 039, & & \\
045, & 046, & 047, & 048, & 049, & & & \\
056, & 057, & 058, & 059, & & & & \\
067, & 068, & 069, & & & & & \\
078, & 079, & & & & & & \\
089, & & & & & & &
\end{array}
$$

$$
\begin{array}{lllllll}
123, & 124, & 125, & 126, & 127, & 128, & 129,\\
134, & 135, & 136, & 137, & 138, & 139, & \\
145, & 146, & 147, & 148, & 149, & & \\
156, & 157, & 158, & 159, & & & \\
167, & 168, & 169, & & & & \\
178, & 179, & & & & & \\
189, & & & & & &
\end{array}
$$

$$
\begin{array}{lll}
567, & 568, & 569,\\
578, & 579, & \\
589, & &
\end{array}
$$

$$
\begin{array}{llllll}
234, & 235, & 236, & 237, & 238, & 239,\\
245, & 246, & 247, & 248, & 249, & \\
256, & 257, & 258, & 259, & & \\
267, & 268, & 269, & & & \\
278, & 279, & & & & \\
289, & & & & &
\end{array}
\begin{array}{lllll}
345, & 346, & 347, & 348, & 349\\
356, & 357, & 358, & 359 & \\
367, & 368, & 369, & & \\
378, & 379, & & & \\
389, & & & &
\end{array}
\begin{array}{llll}
456, & 457, & 458, & 459\\
467, & 468, & 469, & \\
478, & 479, & & \\
489, & & &
\end{array}
\begin{array}{ll}
678, & 679,\\
689, & 789
\end{array}
$$

and count 120 possible ways of choosing three unique numbers that could be used to set the briefcase lock.

Combination

Definition 6

Combination

Suppose we have n elements. The number of ways to select r unique elements such that $r \leq n$ and order does not matter is

$$
\begin{aligned}
nCr = C(n,r) = C_r^n &= \binom{n}{r} \\
&= \frac{n!}{(n-r)!r!}
\end{aligned}
$$

Some points about this definition

1. nCr, $C(n,r)$, C_r^n and $\binom{n}{r}$ (pronounced *n choose r*) are common notations for the number of combinations of r elements from n elements.
2. Notice that the only difference between the **permuation** and combination formulae is the $r!$ in the denominator.

Re-visiting the briefcase lock example,

Example 14

Briefcase lock

You buy a lock for a briefcase that contains three dials and you now want to set the lock. Each dial can be set to a number between 0 and 9, however, each dial number must be set to a unique number such that dial 1 $\neq$ dial 2 $\neq$ dial 3. How many ways could you select three numbers from $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ that may then be used to set the lock?

$$\binom{10}{3} = \frac{10!}{7!3!} = \frac{3628800}{5040 \times 6} = 120.$$

Video

Briefcase lock example

Duration 2:29

In the **rolling a dice three times example**, we noticed that the outcomes of each roll were independent and so there were 216 possible outcomes from the three roles. Taking this one step further, suppose now we were interested in a specific combination of the three rolls.

Example 15

Rolling a dice three times

Suppose we rolled a dice three times. How many possible ways could we roll a 5 exactly two times?

In a single roll, we are interested in whether the outcome of a roll is equal to 5 or not equal to 5. Lets define

$$E = \{5\}$$
$$E' = \{1, 2, 3, 4, 6\}.$$

In the first roll, we can either roll a 5 or not. Therefore we will either be in E or E'. Likewise in the second and third rolls, we can either roll a 5 or not. Putting everything together in a probability tree diagram (displaying all possible outcomes)

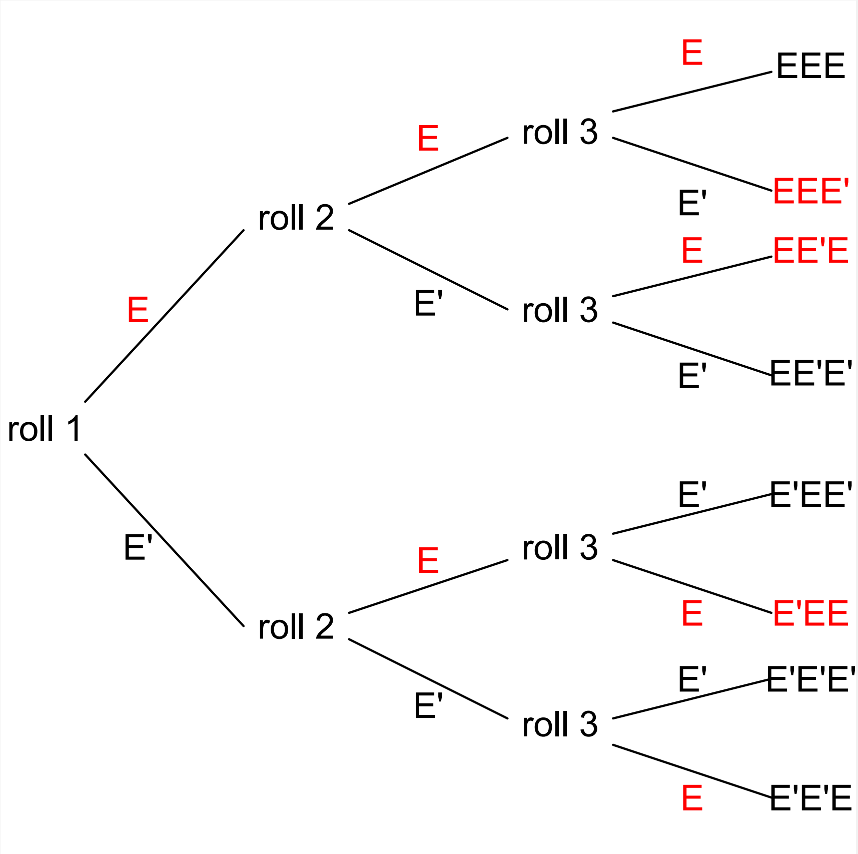


Figure 12

Therefore in this experiment, there are 8 possible outcomes and 3 ways for us to roll a 5 exactly two times.

This example can be re-written in such a way that the combination formula can directly be applied. Instead on thinking in terms of how many possible ways could we roll a 5 exactly two times out of three rolls, lets think about the number of times we roll a five out of a total of three rolls.

Example 16

Rolling a dice three times

Suppose we rolled a dice three times. How many possible ways could we roll a 5 exactly two times?

Let's re-define the experiment in terms of the number of times we roll a 5 out of a total of three rolls. We could roll a 5 either 0, 1, 2 or 3 times. Therefore, lets define our sample space

$$S = \{0, 1, 2, 3\}$$

referring now to the total number of times we roll a 5.

Now, the number of ways of rolling a 5 exactly two times, given that we don't care about the order in which we roll the 5's, is

$$\binom{3}{2} = \frac{3!}{2!1!} = \frac{6}{2 \times 1} = 3.$$

Throughout this section, we have seen three different counting principals; **the multiplication principal**, **permutations** and **combinations**. We can now set out the main differences between the three methods

1. The multiplication principal. Counting the number of ways of selecting n elements with replacement.
2. Permutation. Counting the number of distinct ways of selecting r elements from n without replacement.
3. Combination. Counting the number of ways of selecting r elements from n without replacement.

There are two main distinctions made between the three methods.

1. With permutations, the order does matter whereas in combinations, the order does not matter.
2. Selecting objects with or without replacement. These concepts need to be defined. We will delve deeper into sampling methods in week 9.

Definition 7

Sampling without replacement

Suppose we are selecting r elements from n from . Each element within the sample space has only one change to be selected. Once the element has been selected, it cannot be selected again.

Definition 8

Sampling with replacement

Suppose we are selecting r elements from n. Each element has r chances of being selected. Once the element has been selected, it is placed back in the set and can be selected again.

Some important distinctions between these two definitions

1. When we sample with replacement, each selected value is independent. Practically, this means that what we select for the first draw doesn't affect what we get on the second and so on.

2. When we sample without replacement, selected values aren't independent. Practically, this means that what we select for the first draw affects what we can select for the second and so on.

Data collection

All examples were have seen so far require data in order to answer the question. For example, by rolling a dice we are conducting an experiment with an outcome. On the other hand, in the blood type example, we were told that 100 people donated blood and we can therefore use the blood types of the 100 people to estimate the proportion of people in each of the four blood type categories.

Rolling a dice involves an experiment that requires us to do something whereas blood type categories can be observed using data from the 100 people who donate without us doing anything. Therefore there are different ways in which data are ascertained. It is important to understand how data are collected.

Sampling

The process of selecting representative observational units (for example people) from a population. This requires selection criteria and techniques that will be covered in this course.

Experiment

A scientific procedure conducted under known conditions such that a particular question (or hypothesis) can be tested.

Defining probability

We can now formally define a probability for a given sample space S.

Definition 9

Probability

For a given experiment with sample space S, probability is a real-valued function

$$P : S \to [0,1].$$

For each subset $E \subseteq S$, the function P assigns a number $P(E)$, such that $P(E) \in [0,1]$.

For a given sample space S, there are some important restrictions on the set of probabilities $P(E)$, $E \subseteq S$. The set of probabilities must obey the axioms of probabilities stated by **Kolmogorov**.

Axioms of probability

The axioms of probability are

1. For any event $E \subseteq S$, $0 \le P(E) \le 1$.
2. $P(S) = 1$.
3. For two disjoint events $E \cap F = \emptyset$, $P(E \cup F) = P(E) + P(F)$.
4. More generally, for disjoint events $E_1, E_2, \ldots$

$$P\left(\bigcup_{i \in I} E_i\right) = \sum_{i \in I} P(E_i).$$

Axioms 2 guarantees S must occur. In other words, something must happen!

The axioms imply that probabilities behave like surface areas, which is the reason why Venn diagrams are often used to illustrate probability rules. In more mathematical terms, the axioms state that probability theory is a special case of what is called measure theory.

Rules of probability

Based on these axioms of probability, we can now describe some rules of probability.

1. For any $E \subseteq S$, $P(E) = 1 - P(E')$. Since $E \cup E' = S$ and P(S)=1, then $P(E \cup E') = P(E) + P(E') = 1$.

Assess the extent to which you'll actually have to know this...

2. $P(\emptyset) = 0$.

3. For any $E \subseteq F \subseteq S$, then $P(E) \leq P(F)$. If we suppose $E \subset F$ then we know $F \backslash E \neq \emptyset$ and if $F = E + F \backslash E$, then $P(F) = P(E) + P(F \backslash E)$ with $P(F \backslash E) > 0$. If $F = E$ then $P(F) = P(E)$.

4. For any $E, F \subseteq S$. Then $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

5. For **independent events** $E, F \subseteq S$. Then $P(E \cap F) = P(E)P(F)$.

Task 4

What is the probability that you roll a 5 or a 2 when you roll a dice?

Describing a probability

In this course, we have defined a probability $P(E) \in [0, 1]$ for some $E \subseteq S$. Therefore, we would describe a probability as a real number between 0 and 1, for example $P(E) = 0.5$. There are several other ways of describing probabilities. It is equally valid to write the probability as a fraction $P(E) = \frac{1}{2}$ (as in the **Rolling an unfair dice** example), or as a percentage $P(E) = 50\%$ (as in the **blood type** and **disease outcome** examples) or as a ratio $1:2$.

Interpreting a probability

The probability of an event E describes the level of uncertainly regarding whether E will occur or not. If $P(E) = 1$ then E will surely occur and if $P(E) = 0$ then E will never occur.

There are two main concepts in the interpretation of probability:

2. **Relative frequency**. If we repeat an experiment n times and event E occurs n_E times then the relative frequency of E is $\frac{n_E}{n}$. For instance, the **Rolling an unfair dice** example, we repeatedly rolled the dice and counted the number of times a 6 appeared. We concluded that the inferred probability may depend on the number of repetitions of the experiment. The relative frequency describes the average, or expected, number of times E will occur out of n replications. Relative frequencies rely of experiments that are easily endlessly replicate under identical conditions.

3. **Subjective probability**. We can also treat a probability as a statement about a personal belief. If I roll a dice, then I believe I will roll a 6 with probability $\frac{1}{6}$. This description does not require me to endlessly roll a dice since it's based on my belief. Consequently, my belief may be based on prior knowledge of the experiment, or equally could be based on my ignorance of the experiment. Subjective probability lets us assign probabilities to events

that, by definition, cannot be repeated, such as what is the probability that Scotland will win the football World Cup in 2022 or what is the probability that the UK will leave the EU on October 31st?

Video

Two Die

Duration 10:06

Equally-likely outcomes

Definition 10

Equally-likely outcomes

We say an experiment has equally-likely outcomes if every element in the sample space $S = \{s_1, s_2, \ldots, s_n\}$ is equally likely, i.e. $P(s_i) = \frac{1}{n}$ for $i \in 1, \ldots, n$.

For example, in the **blood type** example, we initially assume that each of the four blood type groupings contained the same number of people, hence

$$P(A) = P(B) = P(AB) = P(O) = \frac{1}{4}$$

, before learning more about specific blood types. If we toss a coin, then we can either land on heads or tails with

$$P(\text{head}) = P(\text{tail}) = 0.5$$

. If we roll a fair dice then

$$P(i) = \frac{1}{6}$$

for all $i = 1, \ldots, 6$. A major problem is that most random experiments do not have this kind of symmetry where all elements are equally likely.

In the equally-likely outcome model the probability of an event E is the number of elements in the event E divided by the number of elements in the sample space S, i.e.

$$P(E) = \frac{|E|}{|S|}.$$

Example 17

Rolling a dice

Suppose we roll a dice once. Then our sample space is

$$S = \{1, 2, 3, 4, 5, 6\}$$

Provided the dice is fair, all outcomes are equally likely.

What is the probability that the score is even? Let's write this as an event

$$E = \{2, 4, 6\},$$

and thus

$$P(E) = \frac{|E|}{|S|} = \frac{3}{6} = \frac{1}{2},$$

i.e. the probability of obtaining an even score is 50%.

Example 18

Rolling a pair of dice

Suppose we roll a pair of dice. Imagine that one dice (suppose it is black) shows a 2 and one dice (suppose it is red, so that we can distinguish the two) shows a 3. We can write this outcome then as a pair (2,3).

Using this notation, our sample space is

$$\begin{aligned} S &= \{(1,1), (1,2), \dots, (1,6), \\ &\quad (2,1), (2,2), \dots, (2,6), \\ &\quad \dots \\ &\quad (6,1), (6,2), \dots, (6,6)\} \end{aligned}$$

Using the multiplication principle, there are $|S| = 6 \times 6 = 36$ possible outcomes, which, if the dice are fair, are equally likely.

What is the probability that the score is even? Let's write this as an event

$$
\begin{aligned}
E \quad = \quad & \{(1,1),(1,3),(1,5), \\
& (2,2),(2,4),(2,6), \\
& \ldots \\
& (6,2),(6,4),(6,6)\}
\end{aligned}.
$$

It contains $|E| = 6 \times 3 = 18$ outcomes and thus

$$P(E) = \frac{|E|}{|S|} = \frac{18}{36} = \frac{1}{2},$$

i.e. the probability of rolling an even score is again 50%.

Suppose we want to know how likely it is that the scores of the two dice are different. Let's consider the event F that the scores are different. Using the multiplication principle $|F| = P(6,2) = 6 \times 5 = 30$, thus the probability of obtaining two different scores is

$$P(F) = \frac{|F|}{|S|} = \frac{30}{36} = \frac{5}{6}$$

We could have calculated this probability also using the complement F' that the scores are the same. We know that $F' = \{(1,1),(2,2),\ldots,(6,6)\}$ and thus

$$P(F') = \frac{|F'|}{|S|} = \frac{6}{36} = \frac{1}{6}$$

which implies that

$$P(F) = 1 - P(F') = 1 - \frac{1}{6} = \frac{5}{6}.$$

Learning outcomes for week 1

By the end of week 1 you should be able to:

- Define a sample space;
- Counting the number of elements in a set;
- Understand the multiplication principle, permutations and combinations;
- Define of a probability;
- List the axioms of probability.

Summary of results from week 1

Sample space and events.

The set of all possible outcomes in an experiment is call the sample space and is commonly denoted by S. Events are any possible subsets of a sample space. Sample spaces can fall into one of three categories

1. A space with a finite number elements.
2. A space with an infinite but countable number of elements.
3. A space with an infinite and uncountable number of elements.

Multiplication principle.

Suppose we have n independent events $E_1, E_1, \ldots, E_n$. If event E_k has m_k possible outcomes, for $k = 1, \ldots n$, then there are

$$m_1 \times m_2 \times \ldots \times m_n$$

possible ways for events $E_1, E_1, \ldots, E_n$ to occur.

Permutation

Suppose we have n elements. The number of ways to select r unique elements such that $r \leq n$ and order matters is

$$nPr = P(n, r) = \frac{n!}{(n-r)!} = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)$$

Combination

Suppose we have n elements. The number of ways to select r unique elements such that $r \leq n$ and order does not matters is

$$nCr = C(n, r) = C_r^n = \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Sampling without replacement

Suppose we are selecting r elements from n. Each element has only one chance to be selected. Once the element has been selected, it cannot be selected again.

Sampling with replacement

Suppose we are selecting r elements from n . Each element has r chances of being selected. Once the element has been selected, it is placed back in the sample space and can be selected again.

Definition of probability

For a given experiment with sample space S, probability is a real-valued function

$$P : S \to [0, 1].$$

For each subset $E \subseteq S$, the function P assigns a number $P(E)$, such that $P(E) \in [0, 1]$.

Axioms of probability

The axioms of probability are

1. For any event $E \subseteq S$, $0 \leq P(E) \leq 1$.
2. $P(S) = 1$.
3. For two disjoint events $E \cap F = \emptyset$, $P(E \cup F) = P(E) + P(F)$.
4. More generally, for disjoint events $E_1, E_2, E_2, \ldots$

$$P\left(\bigcup_{i \in I} E_i\right) = \sum_{i \in I} P(E_i).$$

Answer 1

Video

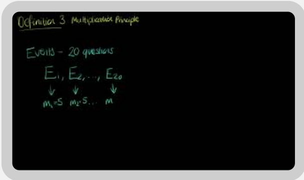
Video model answer

Duration 1:08

$S = \{\text{O, A, B, AB}\}$

Answer 2

Video



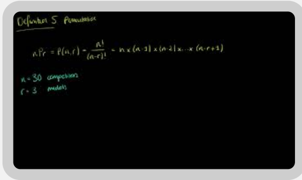
Video model answer

Duration 1:39

We may assume all questions are independent and we randomly select and answer to each. Each of the 20 questions has 5 possible answers and therefore there are 5^{20} possible outcomes.

Answer 3

Video



Video model answer

Duration 1:58

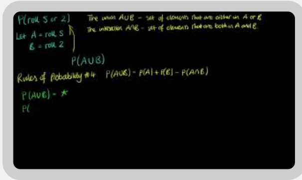
Assuming we know nothing about the individual competitors, and therefore we don't know who is likely to win, we are choosing 3 people from 30 and the order determines the medal received and so order matters. Therefore, the number of possible ways to allocate medals is $P(30, 3) = 24360$.

Answer 4

Video

Video model answer

Duration 4:27



We want to know

$$\begin{aligned} P(\text{rolling a 5 } \cup \text{ rolling a 2}) &= P(\text{rolling a 5 }) + P(\text{rolling a 2 }) - P(\text{rolling a 5 } \cap \text{ rolling a 2}) \\ &= \frac{1}{6} + \frac{1}{6} - 0 \\ &= \frac{1}{3} \end{aligned}$$