

Advanced Statistics - Probability Intro

Ralf Becker

January 23, 2023

1 Introducing Probability

So far we have been looking at ways of summarising samples of data drawn from an underlying population of interest. Although at times tedious, all such arithmetic calculations are fairly mechanical and straightforward to apply. To remind ourselves, one of the primary reasons for wishing to summarise data is so assist in the development of inferences about the population from which the data were taken. That is to say, we would like to elicit some information about the mechanism which generated the observed data.

Perhaps you can recall this little schematic which we discussed in the introduction.

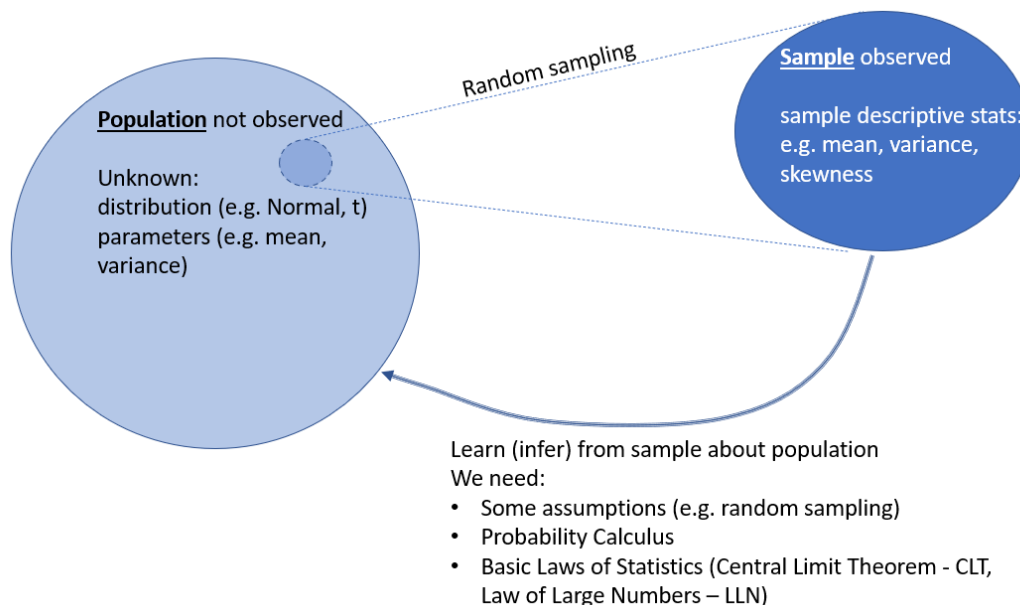


Figure 1: Schematic of Statistical Inference and how it relates to descriptive statistics and probability calculus.

We now start on the process of developing mathematical ways of formulating inferences and this requires the use of *probability*. This becomes clear if we think back to one of the early questions posed in this course: *prior to sampling is it possible to predict with absolute certainty what will be observed?* The answer to this question is *no*; although it would be of interest to know how *likely* it is that certain values would be observed. Or, what is the *probability* of observing certain values?

Before proceeding, we need some more tools:

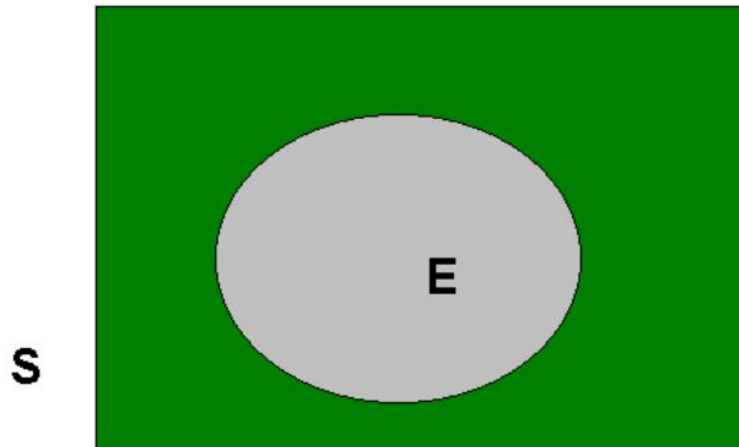
2 Venn diagrams

Venn diagrams (and diagrams in general) are of enormous help in trying to understand, and manipulate probability. We begin with some basic definitions, some of which we have encountered before.

- **Experiment:** any process which, when applied, provides data or an outcome; e.g., rolling a dice and observing the number of dots on the upturned face; recording the amount of rainfall in Manchester over a period of time.
- **Sample Space:** set of possible outcomes of an experiment; e.g., S (or Ω) = $\{1, 2, 3, 4, 5, 6\}$, which is the sample space of rolling a dice. Or $S = \{x; x \geq 0\}$, which is the sample space of an experiment where the outcomes can be any real non-negative number, or ‘*the set of real non-negative real numbers*’.
- **Simple Event:** just one of the possible outcomes on S
- **Event:** a *subset* of S , denoted $E \subset S$; e.g., $E = \{2, 4, 6\}$ (i.e. any even number on a dice) or $E = \{x; 4 < x \leq 10\}$, which means ‘*the set of real numbers which are strictly bigger than 4 but less than or equal to 10*’.

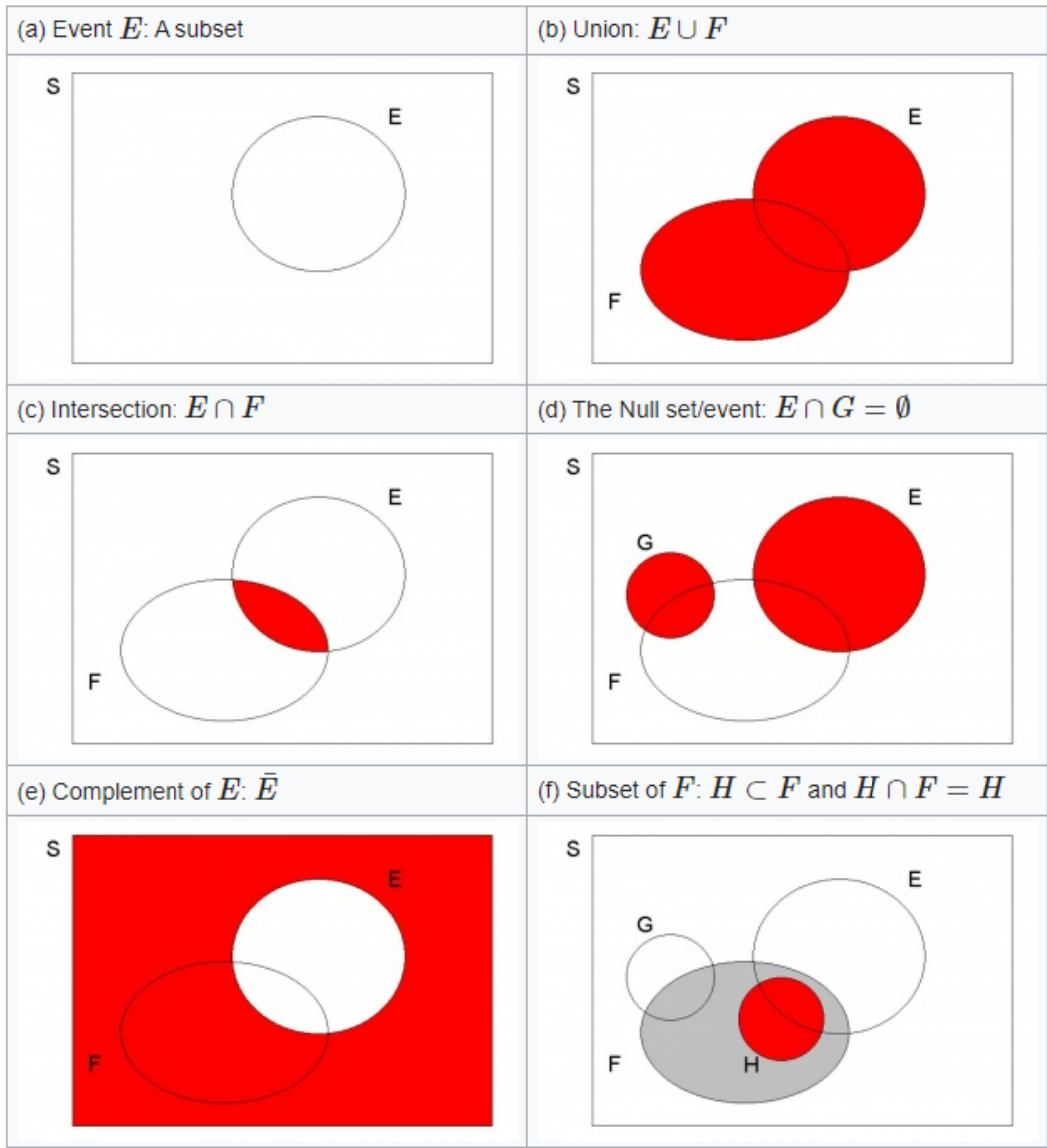
Note that an event, E , is a collection of simple events. You will have to decide what your event of interest is and that depends on the situation.

Such concepts can be represented by means of the following Venn Diagram:



The sample space, S , is depicted as a closed rectangle, and the event E is a closed loop wholly contained within S and we write (in set notation) $E \subset S$.

In dealing with probability, and in particular the probability of an event (or events) occurring, we shall need to be familiar with **unions**, **intersections** and **complements**.



To illustrate these concepts, consider the sample space $S = \{x; x \geq 0\}$. In this example the sample space represents an infinite set of numbers. Don't let the square finite sample space, represented in the next Figure, distract you from that. The following events defined on S , as depicted in the Figure:

$$E = \{x; 4 < x \leq 10\}, F = \{x; 7 < x \leq 17\}, G = \{x; x > 15\}, H = \{x; 9 < x \leq 13\}.$$

Note that the size of the graphical representation does not represent any probabilities here.

(a) Event E : This graphical representation merely illustrates that E is a subset of S , $E \subset S$

(b) Union: $E \cup F$

The *union* of E and F is denoted $E \cup F$, with $E \cup F = \{x; 4 < x \leq 17\}$; i.e., it contains elements (simple events) which are either in E or in F or (perhaps) in both. This is illustrated on the Venn diagram by the dark shaded area in diagram (b).

(c) Intersection: $E \cap F$

The *intersection* of E and F is denoted $E \cap F$, with $E \cap F = \{x; 7 \leq x \leq 10\}$; i.e., it contains elements (simple events) which are common to both E and F . Again this is depicted by the dark shaded area in (c).

(d) The Null set/event: $E \cap G = \emptyset$

If events have no elements in common (as, for example, E and G) then they are said to be *mutually exclusive*, and we can write $E \cap G = \emptyset$, meaning the *null set* which contains no elements. Such a situation is illustrated on the Venn Diagram by events (the two shaded events in (d)) which do not overlap. Notice however that $G \cap F \neq \emptyset$, since G and F have elements in common.

(e) Complement of E : \bar{E}

The *complement* of an event E , say, is everything defined on the sample space which is not in E . This event is denoted \bar{E} , the dark shaded area in (e); here $\bar{E} = \{x; x \leq 4\} \cup \{x; x > 10\}$. An alternative notation that is sometimes used for the complement of an event E is E^c .

(f) Subset of F : $H \subset F$ and $H \cap F = H$

Finally note that H is a sub-set of F ; see (f). It is depicted as the dark closed loop wholly contained within F , the lighter shaded area, so that $H \cap F = H$; if an element in the sample space is a member of H then it must also be member of F . (In mathematical logic, we employ this scenario to indicate that “ H implies F ”, but not necessarily vice-versa.) Notice that $G \cap H = \emptyset$ but $H \cap E \neq \emptyset$.

Exercise

1. Suppose that the sample space S consists of the positive integers from 1 to 10 inclusive. Let X and Y be the following sets:

$$X = \{2, 3, 4\},$$

$$Y = \{3, 4, 5\},$$

$$Z = \{5, 6, 7\},$$

List the members in the following sets

$$\bar{X} \cap Y = \{\underline{\hspace{2cm}}\}$$

$$\bar{X} \cup Y = \{\underline{\hspace{2cm}}\}$$

$$X \cap Z = \{\underline{\hspace{2cm}}\}$$

$$\overline{X \cap (\overline{Y \cup Z})} = \{\text{_____}\}$$

3 Probability

The term *probability* is used in everyday conversation. Watching the daily news, for instance, you will come across weather forecasts which will often include probabilistic statements like "There is a 70% probability of rain for tomorrow." Here is a video from on weather forecasts and an incredibly interesting podcast on the importance of communicating uncertainty in weather forecasts (38 min).

Other areas where we use probabilities are sports (What is the probability of England winning the World Cup) games of chance (e.g. roulette or poker) or economics (What is the probability of a recession?). What you will do now is to learn about a coherent theory of probability; a theory which allows us to combine and manipulate probabilities in a consistent and meaningful manner. You will describe ways of dealing with, and describing, uncertainty. This will involve *rules* which govern our use of terms like probability.

Communicating uncertainty

Communicating probabilities, or uncertainty is one of the great challenges of scientists, economists and data analysts. Dealing in uncertainties is what we do. But often the people we are communicating to do not want to hear uncertainties, they want certainties. Will it rain tomorrow? Yes or No! Will my employer go out of business? Yes or no! Will the Conservatives be able to lead the government after the next election? Yes or no! Will this investment be profitable? Yes or no! Will the Twitter share price go up? yes or no! Despite the thirst for certainty you will have to find ways to communicate uncertainty which both, conveys the right information (including the level of certainty you have), is understandable to the public, and protects your reputation. It turns out that this is a great challenge.

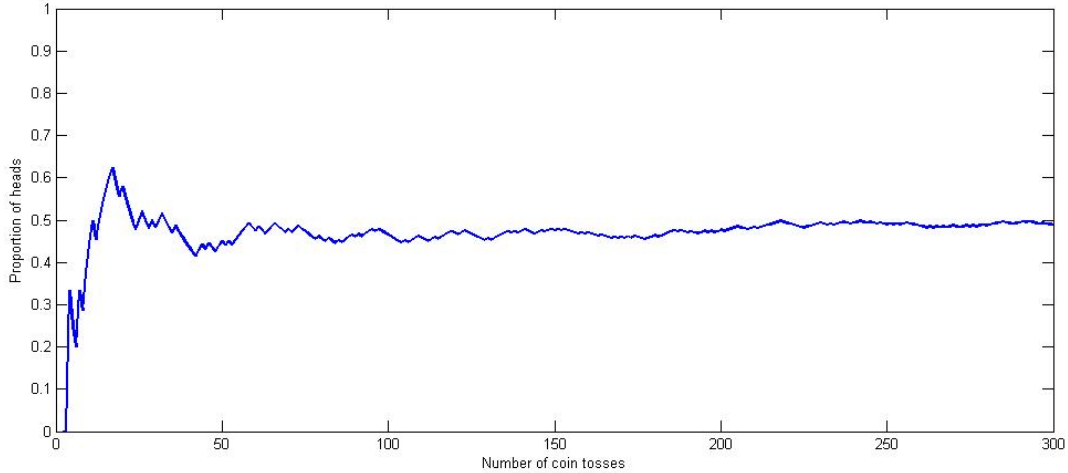
Here is a great document on communicating uncertainty by Full Fact and a lecture on this issue (60 min) by the excellent Lecture by David Spiegelhalter and a website he initiated. This is supplementary and non-assessed information.

There have been a number of different approaches (interpretations) of probability. Most depend, at least to some extent, on the notion of relative frequency as now described:

- Suppose an experiment has an outcome of interest E . The *relative frequency interpretation* of probability says that assuming the experiment can be repeated a large number of times then the relative frequency of observing the outcome E will settle down to a *number*,

denoted $\Pr(E)$, $P(E)$ or $\text{Prob}(E)$, called the **probability** of E .

This is illustrated in the next Figure where the proportion of heads obtained after n flips of a fair coin is plotted against n , as n increases; e.g., of the first 100 flips, 46 were heads (46%). Notice that the plot becomes less ‘wobbly’ after about $n = 140$ and appears to be settling down to the value of $\frac{1}{2}$.



Due to this interpretation of probability, we often use observed sample proportions to approximate underlying probabilities of interest; see, for example, Question 4 of Exercise 2. There are, of course, other interpretations of probability; e.g., the subjective interpretation which simply expresses the strength of one’s belief about an event of interest such as whether Manchester United will win the European Cup! Any one of these interpretations can be used in practical situations provided the implied notion of probability follows a simple set of *axioms* or *rules*.

3.1 The axioms of probability

There are just *three* basic rules that must be obeyed when dealing with probabilities:

1. For any event E defined on S i.e., $E \subset S$, $\Pr(E) \geq 0$; *probabilities are non-negative*.
2. $\Pr(S) = 1$; *having defined the sample space of outcomes, one of these outcomes must be observed*.
3. If events E and F are mutually exclusive defined on S , so that $E \cap F = \emptyset$, then $\Pr(E \cup F) = \Pr(E) + \Pr(F)$. In general, for any set of mutually exclusive events, E_1, E_2, \dots, E_k , defined on S :

$$\Pr(E_1 \cup E_2 \cup \dots \cup E_k) = \Pr(E_1) + \Pr(E_2) + \dots \Pr(E_k)$$

$$\text{i.e., } \Pr\left(\bigcup_{j=1}^k E_j\right) = \sum_{j=1}^k \Pr(E_j).$$

In terms of the Venn Diagram, one can (and should) usefully think of the area of E relative to that of S as providing an indication of probability. (Note, from axiom 2, that the area of S is implicitly normalised to be unity).

Also observe that, contrary to what you may have believed, it is not one of the rules that $\Pr(E) \leq 1$ for any event E . Rather, this is an implication of the 3 rules given:

- **implications:** it must be that for any event E , defined on S , $E \cap \bar{E} = \emptyset$ and $E \cup \bar{E} = S$. By Axiom 1, $\Pr(E) \geq 0$ and $\Pr(\bar{E}) \geq 0$ and by Axiom 3, $\Pr(E) + \Pr(\bar{E}) = \Pr(S)$. So $\Pr(E) + \Pr(\bar{E}) = 1$ by Axiom 2. This implies that

1. $0 \leq \Pr(E) \leq 1$
2. $\Pr(\bar{E}) = 1 - \Pr(E)$

The first of these is what we might have expected from probability (a number lying between 0 and 1). The second implication is also very important; it says that the probability of E not happening is ‘one minus the probability of it happening’. Thus when rolling a die, the probability of getting 6 is one minus the probability of getting either a 1, 2, 3, 4 or 5.

These axioms imply how to calculate probabilities on a sample space of equally likely outcomes. For example, and as we have already noted, the experiment of rolling a fair dice defines a sample space of six, mutually exclusive and equally likely outcomes (1 to 6 dots on the up-turned face). The axioms then say that each of the six probabilities are positive, add to 1 and are all the same. Thus, the probability of any one of the outcomes must be simply $\frac{1}{6}$; which may accord with your intuition. A similar sort of analysis reveals that the probability of drawing a club from a deck of 52 cards is $\frac{13}{52}$ since any one of the 52 cards has an equal chance of being drawn and 13 of them are clubs; i.e., 13 of the 52 are clubs, so the probability of drawing a club is $\frac{13}{52}$. Notice the importance of the assumption of equally likely outcomes here.

In what follows you shall see how these axioms can be used. Firstly, consider the construction of a probability for the *union* of two events; i.e., the probability that *either* E or F or (perhaps) *both* will occur. Such a probability is embodied in the *addition rule of probability*.

3.2 The addition rule of probability

Let us begin with an example.

Example. When rolling a fair dice, let E denote the event of an “odd number of dots” and F the event of the “number of dots being greater than, or equal, to 4”.

$$E = \{1, 3, 5\}, Pr(E) = \frac{3}{6} = 0.5$$

$$F = \{4, 5, 6\}, Pr(F) = \frac{3}{6} = 0.5$$

What is the probability of the event $E \cup F$? To calculate this we can collect together all the mutually exclusive (simple) events which comprise $E \cup F$, and then add up the probabilities (by axiom 3).

$$E \cup F = \{1, 3, 4, 5, 6\}, Pr(E \cup F) = \frac{5}{6}$$

$$E \cap F = \{5\}, Pr(E \cap F) = \frac{1}{6}$$

Each of the simple events has a probability of $\frac{1}{6}$ so the required total probability is: $Pr(E \cup F) = \frac{5}{6}$. Consider carefully how this probability is constructed and note, in particular, that $Pr(E \cup F) \neq Pr(E) + Pr(F)$ since E and F have a simple event in common (namely 5 dots).

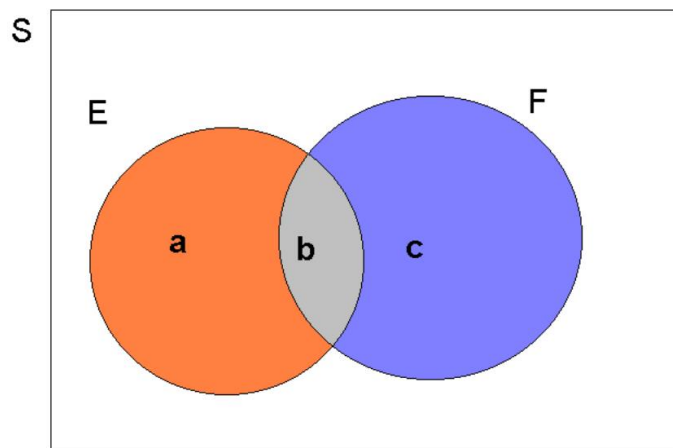
In general, we can calculate the probability of the union of events using the *addition rule of probability*, as follows.

- For any events, $E \subset S$ and $F \subset S$: $Pr(E \cup F) = Pr(E) + Pr(F) - Pr(E \cap F)$. So, in general, $Pr(E \cup F) \leq Pr(E) + Pr(F)$.

We can demonstrate this as follows. Note that

$$E \cup F = (E \cap \bar{F}) \cup (E \cap F) \cup (\bar{E} \cap F)$$

the union of 3 mutually exclusive events. These mutually exclusive events are depicted by the shaded areas **a**, **b** and **c**, respectively, in the next Figure.



Then by Axiom 3, and from the fact that the three events $(E \cap \bar{F})$, $(E \cap F)$ and $(\bar{E} \cap F)$ are mutually exclusive so that the “area” occupied by $E \cup F$ is simply $\mathbf{a} + \mathbf{b} + \mathbf{c}$,

$$\Pr(E \cup F) = \Pr(E \cap \bar{F}) + \Pr(\bar{E} \cap F) + \Pr(E \cap F).$$

But also by Axiom 3, since $E = (E \cap \bar{F}) \cup (E \cap F)$, it must be that $\Pr(E) = \Pr(E \cap \bar{F}) + \Pr(E \cap F)$; similarly, $\Pr(\bar{E} \cap F) = \Pr(F) - \Pr(E \cap F)$. Putting all of this together gives

$$\Pr(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F).$$

This was a lengthy probabilistic argument to make sense of the probability addition formula. You may wonder why the addition formula is not just $\Pr(E \cup F) = \Pr(E) + \Pr(F)$? You can refer back to the image above. If we just added $\Pr(E)$ and $\Pr(F)$ we would have double counted area b. We therefore have to subtract it once. And area b is $\Pr(E \cap F)$.

When E and F are mutually exclusive, there is no overlap, so that $E \cap F = \emptyset$, this rule reveals Axiom 2: $\Pr(E \cup F) = \Pr(E) + \Pr(F)$.

Example. *What is the probability of drawing a Queen (Q) or a Club (C) in a single draw from a pack of cards? Now, 4 out of 52 cards are Queens, so $\Pr(Q) = \frac{4}{52}$, whilst $\Pr(C) = \frac{13}{52}$. The probability of drawing the Queen of Clubs is simply $\frac{1}{52}$; i.e., $\Pr(Q \cap C) = \frac{1}{52}$. What we require is a Club or a Queen, for which the probability is*

$$\begin{aligned} \Pr(Q \cup C) &= \Pr(Q) + \Pr(C) - \Pr(Q \cap C) \\ &= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} \\ &= \frac{16}{52} = \frac{4}{13}. \end{aligned}$$

You will often be presented with frequency (count) tables of the following type.

Region	Health					Sum
	Very good health	Good health	Fair health	Bad health	Very bad health	
North East	11,425	8,795	3,977	1,476	428	26,101
North West	32,413	23,502	9,847	3,842	1,083	70,687
London	20,086	18,302	7,476	2,376	701	52,943
Sum	67,924	50,599	21,300	7,694	2,214	149,731

These are the respondents from three UK regions (North East, North West and London) and their self indicated health status. The data are available in the 2011 Census Microdata Teaching File (see the Data Sets page). For instance there are 8,795 respondents from the North East

that indicated that their health was best categorised as "Good health". Throughout the three regions there were 50,599 respondents that gave "Good health" as their health status.

Now you may be interested whether, for instance, respondents from the North West tend to be more positive about their health, compared to respondents from other regions. For this purpose it is useful to calculate a probability table based on the above frequency table. The way to do that is to divide all frequency values in the above table by the total number of responses, here 149,731. For example, for the above case of someone coming from the North East and indicating "Good health" we would calculate $8795/149731$ which results in 0.0587. This is the probability that, if we randomly drew a person from our 149,731 responses, we would get someone from the North East with "Good health", $Pr("North East" \cap "Good health") = 0.0587$. If you do that for all cells in the above table you get the following:

Region	Health					Sum
	Very good health	Good health	Fair health	Bad health	Very bad health	
North East	0.0763	0.0587	0.0266	0.0099	0.0029	0.1743
North West	0.2165	0.1570	0.0658	0.0257	0.0072	0.4721
London	0.1609	0.1222	0.0499	0.0159	0.0047	0.3536
Sum	0.4536	0.3379	0.1423	0.0514	0.0148	1.000

As another example, the $Pr("London" \cap "Very bad health") = 0.0047$. We call these probabilities joint probabilities and in the above table they are coloured in blue. The probabilities at the margins of the table are also called marginal probabilities. For instance $Pr("London") = 0.3536$, meaning that if we randomly draw a response from the above 149,731 respondents, there is a 35% probability that this response comes from London.

Example. Consider the following frequency table, where 9,900 respondents from two cities have been asked whether they enjoy eating vegetables.

City	Vegetables			Sum
	Enjoy	Neutral	Do not enjoy	
Springfield		1,045	876	2,673
Gotham City	2,893	2,504	1,803	
Sum	3,645		2,706	9,900

Complete the above table:

$$\text{Frequency}("Springfield" \cap "Enjoy") = 752$$

$$\text{Frequency}("Neutral") = 3549$$

$$\text{Frequency}("Gotham City") = 7227$$

Now calculate the probability table:

<i>City</i>	<i>Vegetables</i>			<i>Sum</i>
	<i>Enjoy</i>	<i>Neutral</i>	<i>Do not enjoy</i>	
<i>Springfield</i>	<i>0.0760</i>	<i>0.1056</i>	<i>0.0885</i>	<i>0.2700</i>
<i>Gotham City</i>	<i>0.2922</i>	<i>0.2529</i>	<i>0.1848</i>	<i>0.7300</i>
<i>Sum</i>	<i>0.3682</i>	<i>0.3585</i>	<i>0.2733</i>	<i>1.000</i>

The $\Pr(\text{"Springfield"} \cap \text{"Enjoy"})$ is a marginal/joint probability

The $\Pr(\text{"Gotham City"})$ is a marginal/joint probability

Example. Consider a car journey from Manchester to London via the M6 and M1. Let $E =$ heavy traffic somewhere on route and $F =$ roadworks somewhere on route. It is estimated that $\Pr(E) = 0.8$ and $\Pr(F) = 0.4$, whilst the probability of NOT encountering both is $\Pr(\overline{E \cap F}) = 0.6$. What is the probability of encountering heavy traffic or roadworks?

What is $\Pr(E \cup F)$?

Hint: First calculate $\Pr(E \cup F) = 0.4$

Feedback:

$$\begin{aligned}
 \Pr(E \cup F) &= \Pr(E) + \Pr(F) - \Pr(E \cap F) \\
 &= \Pr(E) + \Pr(F) - (1 - \Pr(\overline{E \cap F})) \\
 &= 0.8 + 0.4 - 1 + 0.6 \\
 &= 0.8 = \Pr(E)
 \end{aligned}$$

Notice that this implies, in this case, that $F \subset E$ (why?). This *model* then implies that when there are roadworks somewhere on route you are bound to encounter heavy traffic; on the other hand, you can encounter heavy traffic on route without ever passing through roadworks.

We can generalise this addition rule to three events, E_1, E_2 and E_3 as

$$\begin{aligned}
 \Pr(E_1 \cup E_2 \cup E_3) &= \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\
 &\quad - \Pr(E_1 \cap E_2) - \Pr(E_1 \cap E_3) - \Pr(E_2 \cap E_3) \\
 &\quad + \Pr(E_1 \cap E_2 \cap E_3).
 \end{aligned}$$

Exercise: You are asked to spin a wheel of fortune (change the wheel size to 12!) with numbers 1 to 12 being equally likely outcomes. The three events you consider are:

- $E_1 = \{ \text{any number larger or equal than 10} \}$
- $E_2 = \{ \text{any number smaller than 5} \}$

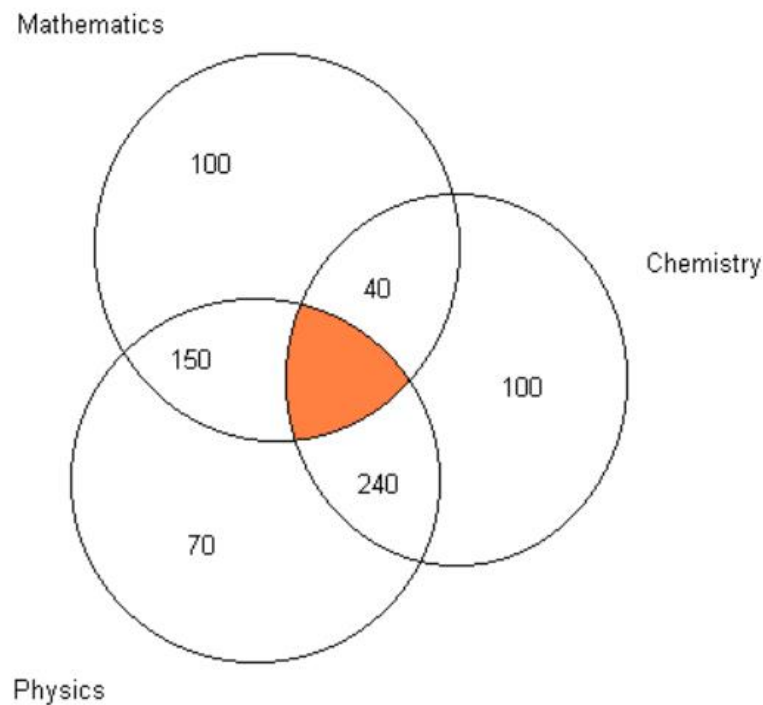
- $E_3 = \{ \text{any even number} \}$

$$\begin{aligned}
 \Pr(E_1 \cup E_2 \cup E_3) &= \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\
 &\quad - \Pr(E_1 \cap E_2) - \Pr(E_1 \cap E_3) - \Pr(E_2 \cap E_3) \\
 &\quad + \Pr(E_1 \cap E_2 \cap E_3) \\
 &= \frac{3}{12} + \frac{4}{12} + \frac{6}{12} - \frac{0}{12} - \frac{2}{12} - \frac{2}{12} + \frac{0}{12} \\
 &= \frac{9}{12} = 0.75
 \end{aligned}$$

Similar concepts apply when manipulating proportions as follows:

- *Example:* A sample of 1000 science undergraduates were asked whether they took either Mathematics (M), Physics (P) or Chemistry (C) at A-level. All of them took at least one of the three subjects at A-level. The following responses were obtained: 100 just took Mathematics; 70 just took Physics; 100 just took Chemistry; 150 took Mathematics and Physics, but not Chemistry; 40 took Mathematics and Chemistry, but not Physics; and, 240 took Physics and Chemistry, but not Mathematics. What proportion took all three?

This can be addressed with the following diagram:



The shaded area contains the number who took all three, which can be deduced from the above information (since the total of the numbers assigned to each part of the Venn diagram must be 1000). The numbers in all areas but for the shaded area sum to 700. The answer is therefore 30% as 300 out of 1000 have to be in the shaded center area.

Exercises To calculate the above solution you can also refer to the previous formula:

$$\Pr(M \cup P \cup C) = \Pr(M) + \Pr(P) + \Pr(C) - \Pr(M \cap P) - \Pr(M \cap C) - \Pr(P \cap C) + \Pr(M \cap P \cap C)$$

Example. What is $\Pr(M \cup P \cup C)$?

Answer: $\Pr(M \cup P \cup C) = 1$

Feedback: $\Pr(M \cup P \cup C) = 1$ is, in the context of this question, the sample space. Note that all students have taken at least one of these three subjects.

The information in the problem says that 70 (out of the 1000) students only took Physics. Which probability is that? Note, as usual, S represents the sample space. [multiple correct answers are possible]

- $\Pr(M \cap P)$
 - $\Pr(P)$
 - $\Pr(\bar{M} \cap \bar{C} \cap P)$ (T)
 - $1 - \Pr(C \cup M)$ (T)
 - $\Pr(P) - \Pr(C \cup M)$
-

Two further results on unions, intersections and complements which are of use (and which are fairly easy to demonstrate using Venn diagrams) are **de Morgan Laws**:

- $\bar{A} \cap \bar{B} = \overline{(A \cup B)}$
- $\bar{A} \cup \bar{B} = \overline{(A \cap B)}$

Exercise

One hundred students were surveyed about their preference between Math and Statistics. The following two-way table displays data for the sample of students who responded to the survey.

Preference	Male	Female	TOTAL
Prefers Math	22	18	40
Prefers Statistics	25	12	37
No preference	8	15	23
TOTAL	55	45	100

1. Find the probability that a randomly selected student prefers Math. Enter your answer as a fraction or decimal.

P(prefers Math) _____

2. In this sample, are the events "prefers Math" and "prefers Statistics" mutually exclusive?

Yes

No

3. Find the probability that a randomly selected student prefers Math or prefers Statistics.

P(prefers Math or Statistics) _____

4. In this sample, are the events "prefers Statistics" and "male" mutually exclusive?

Yes

No

5. Find the probability that a randomly selected student prefers Statistics or is male.

P(prefers Statistics or is male) _____

4 Probability Fair - Online Game

Play a range of different fun probability games by making use of the knowledge of probability. [<https://mrnussbaum.com/probability-fair-online-game#activity-single-content>]

5 Additional resources

Khan Academy

- Basic Probability and Venn Diagram [https://www.khanacademy.org/math/probability/independent-dependent-probability/addition_rule_probability/v/probability-with-playing-cards-and-venn-diagrams]

- Addition Rule [<https://www.khanacademy.org/math/probability/independent-dependent-probability/addition-rule-probability/v/addition-rule-for-probability>]