Statistics and the Evaluation of Evidence for Forensic Scientists,
Third Edition. Colin Aitken, Franco Taroni and Silvia Bozza.

1

Uncertainty in Forensic Science

1.1 INTRODUCTION

The purpose of this book is to discuss the statistical and probabilistic evaluation of scientific evidence for forensic scientists. A formal definition of *evidence* is given in Section 1.4. For the most part the evidence to be evaluated will be the so-called *transfer* or (physical) *trace* evidence, but the general logic also applies to other types of evidence, such as digital, pattern, or testimonial evidence.

There is a well-known principle in forensic science known as *Locard's principle*, which states that every contact leaves a trace (Locard 1920).

> *[. . .] tantôt le malfaiteur a laissé sur les lieux les marques de son passage, tantôt, par une action inverse, il a emporté sur son corps ou sur ses vêtements, les indices de son séjour ou de son geste. (p. 139)*

This is translated as (Inman and Rudin 2001)

> *[. . .] either the wrong-doer has left signs at the scene of the crime, or, on the other hand, has taken away with him – on his person (body) or clothes – indications of where he has been or what he has done. (p. 93)*

The principle was reiterated using different words in Locard (1929). This has been translated by the same author in 1930. Locard (1930) wrote

> *For the microscopic debris that covers our clothes and bodies are the mute witnesses, sure and faithful, of all our movements and of all our encounters. (p. 276)*

Transfer evidence and Locard's principle may be illustrated as follows. Suppose a person gains entry to a house by breaking a window and assaults the man of the house, during which assault blood is spilt by both victim and assailant. The criminal may leave traces of their presence at the crime scene in the form of bloodstains from the assault and fibres from his clothing. This evidence is said to be transferred from the criminal to the scene of the crime. The criminal may also take traces of the crime scene away with them. These could include bloodstains from the assault victim, fibres of their clothes, and fragments of glass from the broken window. Such evidence is said to be transferred to the criminal from the crime scene. A *person of interest*[1] (PoI) is soon identified, at a time at which

[1] In earlier editions the term *suspect* was used in this context. It is now felt that the term *person of interest* is a more accurate description of the status of the person. The term *suspect* will still be used if use of 'person of interest' would be clumsy. The term *defendant* will only be used if the context is clearly that of a court.

they will not have had the opportunity to change their clothing. The forensic scientists examining the PoI's clothing find similarities amongst all the different types of evidence: blood, fibres, and glass fragments. They wish to *evaluate this evidence*. It is hoped that this book will enable them so to do.

However, for evaluation, it is not only similarity that is important but also the rarity of the characteristics of interest. Hence, quantitative issues relating to the distribution of these characteristics will be discussed. However, there will also be discussion of qualitative issues such as the choice of a suitable population against which variability in the measurements of the characteristics of interest may be compared. Also, a brief history of statistical aspects of the evaluation of evidence is given in Chapter 3.

1.2 STATISTICS AND THE LAW

The book does not focus on the use of statistics and probabilistic thinking for legal decision making, other than by occasional reference. Also, neither the role of statistical experts as expert witnesses presenting statistical assessments of data nor their role as consultants preparing analyses for counsel is discussed. There is a distinction between these two issues (Fienberg 1989, Tribe 1971). The main focus of this book is on the assessment of evidence for forensic scientists, in particular for *identification* purposes. The process of addressing the issue of whether or not a particular item came from a particular source is most properly termed *individualization*. 'Criminalistics is the science of individualization' (at p. 236) as defined by Kirk (1963) but established forensic and judicial practices have led to it being termed *identification*. The latter terminology will be used throughout this book. An *identification*, however, is more correctly defined as 'the determination of some set to which an object belongs or the determination as to whether an object belongs to a given set' (Kingston 1965a). Further discussion is given by Kwan (1977), Evett et al. (1998), and Champod et al. (2016b). For a critical discussion of individualisation as a decision, see Cole (2014), Biedermann et al. (2008a), and Saks and Koehler (2008). More details are given in Section 2.5.9.

For example, in a case involving a broken window, similarities may be found between the refractive indices of fragments of glass found on the clothing of a PoI and the refractive indices of fragments of glass from the broken window. The assessment of this evidence, in consideration of the association or otherwise of the PoI with the scene of the crime, is part of the focus of this book.

For those interested in the issues of statistics and the law beyond those of forensic science, in the sense used in this book, there are several books available and some of these are discussed briefly.

'The evolving role of statistical assessments as evidence in the courts' is the title of a report, edited by Fienberg (1989), by the Panel on Statistical Assessments as Evidence in the Courts formed by the Committee on National Statistics and the Committee on Research on Law Enforcement and the Administration of Justice of the United States, and funded by the National Science Foundation. Through the use of case studies, the report reviews the use of statistics in selected areas of litigation, such as employment discrimination, antitrust litigation, and environment law. One case study is concerned with identification in a criminal case. Such a matter is the concern of this book, and the ideas relevant to this case study, which involves the evidential worth of similarities amongst human head hair samples, will be discussed in greater detail later (Section 3.5.5). The report makes various recommendations, relating to the role of the expert witness, pretrial discovery, the provision of statistical resources, the role of court-appointed experts, the enhancement of the capability of the fact-finder, and statistical education for lawyers.

Two books that take the form of textbooks on statistics for lawyers are Vito and Latessa (1989) and Finkelstein and Levin (2015). The former focusses on the presentation of statistical concepts commonly used in criminal justice research. It provides criminological examples to demonstrate the calculation of basic statistics. The latter introduces rather more advanced statistical techniques and again uses case studies to illustrate the techniques.

Historically, the particular area of discrimination litigation is covered by a set of papers edited by Kaye and Aickin (1986). This starts by outlining the legal doctrines that underlie discrimination litigation. In particular, there is a fundamental issue relating to discrimination in hiring. The definition of the relevant market from which an employer hires has to be made very clear. For example, consider the case of a man who applies, but is rejected, for a secretarial position. Is the relevant population the general population, the representation of men amongst secretaries in the local labour force, or the percentage of male applicants? The choice of a suitable reference population is also one with which the forensic scientist has to be concerned. This is discussed at several points in this book, see, for example, Sections 5.5.3.4 and 6.1.

Another textbook, which comes in two volumes, is Gastwirth (1998a,b). The book is concerned with civil cases and 'is designed to introduce statistical concepts and their proper use to lawyers and interested policy makers . . . ' (volume 1, p. *xvii*). Two areas are stressed, which are usually given less emphasis in most statistical textbooks. The first area is concerned with measures of relative or comparative inequality. These are important because many legal cases are concerned with issues of fairness or equal treatment. The second area is concerned with the combination of results of several related statistical studies. This is important because existing administrative records or currently available studies often have to be used to make legal decisions and public policy; it is not possible to undertake further research. Gastwirth

(2000) has also edited a collection of essays on statistical science in the courtroom, some of which are directly relevant for this current book and will be referred to as appropriate.

A collection of papers on Statistics and Public Policy has been edited by Fairley and Mosteller (1977). One issue in the book, which relates to a particularly infamous case, the *Collins* case, is discussed in detail later (Section 3.4). Other articles concern policy issues and decision making.

Of further interest is a book (Kadane 2008) explicitly entitled 'Statistics and the Law', which considers the question 'how can lawyers and statisticians best collaborate in a court of law to present statistics in the most clear and persuasive manner?'. With the use of case studies that refer to employment, jury behaviour, fraud, taxes, and aspects of patents, there is clarification of what a statistician and what a lawyer should know for a fruitful collaboration.

Other recent publications on the interaction between law and statistics are, for example, Dawid et al. (2014), Kadane (2018a,b), Kaye (2017a,b), and Gastwirth (2017).

The remit of this book is one that is not covered by these others in great detail. The use of statistics in forensic science in general is discussed in a collection of essays edited by Aitken and Stoney (1991). The remit of this book is to describe statistical procedures for the evaluation of evidence for forensic scientists. This will be done primarily through a Bayesian approach, the principle of which was described in Peirce (1878). It was developed further in the work of I.J. Good and A.M. Turing as code-breakers at Bletchley Park during World War Two. A brief review of the history was given in Good (1991). A history of the Bayesian approach for a lay audience was given in Bertsch McGrayne (2011). An essay on the topic of probability and the weighing of evidence was written by Good (1950). This also referred to entropy (Shannon 1948), the expected amount of information from an experiment, and Good remarked that the expected weight of evidence in favour of a hypothesis H and against its complement $\bar{H}$ is equal to the difference of the entropies assuming H and $\bar{H}$, respectively. A brief discussion of a frequentist approach and the problems associated with it is given in Section 3.6 (see also Taroni et al. 2016). A general review of the Bayesian approach was given by Fienberg (2006).

It is of interest to note that a high proportion of situations involving the so-called objective presentation of statistical evidence uses the frequentist approach with tests of significance (Fienberg and Schervish 1986).[2] However, Fienberg and Schervish go on to say that the majority of examples cited for the use of the Bayesian

[2]There is a recent critical discussion on the use of tests of significance. See, for example, Wasserstein et al. (2019), Amrhein et al. (2019), Ioannidis (2019), Haaf et al. (2019), and Johnson (2019). Note that such a discussion dates back to 1986 (Kaye 1986a) with a later update (Kaye 2017c).

approach are in the area of identification evidence. It is this area that is the main focus of this book, and it is Bayesian analyses that will form the basis for the evaluation of evidence as discussed here. Examples of the applications of such analyses to legal matters include Cullison (1969), Finkelstein and Fairley (1970, 1971), Fairley (1973), Lempert (1977), Lindley (1975, 1977b,c), Fienberg and Kadane (1983), Lempert (1986), Redmayne (1995, 1997), Friedman (1996), Redmayne (2002), Anderson et al. (2005), Robertson et al. (2016), and Adam (2016).

Another approach that will not be discussed here is that of Shafer (1976, 1982). This concerns so-called *belief functions*, see Section 3.1. The theory of belief functions is a very sophisticated theory for assessing uncertainty that endeavours to answer criticisms of both the frequentist and Bayesian approaches to inference. Belief functions are non-additive in the sense that belief in an event A [denoted Bel(A)] and belief in the opposite of A [denoted Bel($\bar{A}$)] do not sum to 1. See also Shafer (1978) for a historical discussion of non-additivity. Further discussion is beyond the scope of this book. Practical applications are few. One such, however, is to the evaluation of evidence concerning the refractive index of glass (Shafer 1982). More recent developments of the role of belief functions in law for burdens of proof and in forensic science with a discussion of the island problem (Section 6.1.6.3) and parental identification are given in Nance (2019) and Kerkvliet and Meester (2016).

It is very tempting when assessing evidence to try to determine a value for the probability of the so-called *probandum* of interest (or the ultimate issue) such as the true guilt of a PoI (as distinct from a verdict of guilty, which may or may not be correct), or a value for the odds in favour of guilt and perhaps even to reach a decision regarding the PoI's guilt. However, this is the role of the jury and/or judge. It is not the role of the forensic scientist or statistical expert witness to give an opinion on this (Evett 1983). It is permissible for the scientist to say that the evidence is 1000 times more likely, say, if the PoI is the offender than if he is not the offender. It is not permissible to interpret this to say that, because of the evidence, it is 1000 times more likely that the PoI is the offender than is not the offender. Some of the difficulties associated with assessments of probabilities are discussed by Tversky and Kahneman (1974) and are further described in Section 2.5. An appropriate representation of probabilities is useful because it fits the analytic device most used by lawyers, namely, the creation of a story. This is a narration of events 'abstracted from the evidence and arranged in a sequence to persuade the fact-finder that the story told is the most plausible account of "what really happened" that can be constructed from the evidence that has been or will be presented' (Anderson and Twining 1998, p. 166). Also of relevance is Kadane and Schum (1996), which provides a Bayesian analysis of evidence in the *Sacco–Vanzetti* case (Sacco 1969) based on subjectively determined probabilities and assumed relationships amongst evidential events. A similar approach is presented in Section 2.9.

1.3 UNCERTAINTY IN SCIENTIFIC EVIDENCE

Scientific evidence requires considerable care in its interpretation (Evett 2009). Emphasis needs to be put on the importance of asking the question: what do the results mean in this particular case? (Jackson 2000). Kirk and Kingston (1964) emphasised:

> *Suppose that the fibres do match – what does it mean? Suppose that there is a defined degree of similarity in the bullet marking, or the handwriting, does it prove identity of origin, or does it merely give a sometimes controversial basis for making a decision as to the identity of origin? (p. 439)*

Scientists and jurists have to '[. . .] abandon the idea of absolute certainty so that a fully objective approach to the problem can be made. [. . .] If it can be accepted that nothing is absolutely certain then it becomes logical to determine the degree of confidence that may be assigned to a particular belief' (Kirk and Kingston 1964, p. 435). On the same line of reasoning, the authors (Kingston and Kirk 1964) expressed themselves on *uncertainty* and they emphasised:

> *A statistical analysis is used when uncertainty must exist. If there were a way of arriving to a* certain *answer to a*

> *problem, statistical methods would not be used. But when uncertainty does exist, and a statistical approach is possible, then this approach is the best one available since it offers an index on the uncertainty based upon a precise and logical line of reasoning. [. . .] It is undoubtedly true that serious errors have been made in applying incorrect statistical methods to the evaluation of physical evidence, but such misuse does not support the generalisation that statistics cannot be properly used in criminalistics at all.* (p. 516)

There are various kinds of problems concerned with the random variation naturally associated with scientific observations. There are problems concerned with the definition of a suitable reference population against which concepts of rarity or commonality may be assessed. There are problems concerned with the choice of a measure of the value of the evidence.

The effect of the random variation can be assessed with the appropriate use of probabilistic and statistical ideas. There is variability associated with scientific observations. Variability is a phenomenon that occurs in many places. People are of different sexes, determination of which is made at conception. People are of different height, weight, and intellectual ability, for example. The variation in height and weight is dependent on a person's sex. In general, females tend to be lighter and shorter than males. However, variation is such that there can be tall, heavy females and short, light males. At birth, it is uncertain how tall or how heavy the baby will be as an adult. However, at birth, it is usually known whether the baby is a boy or a girl. This knowledge affects the uncertainty associated with the predictions of adult height and weight.

People are of different blood groups. A person's blood group does not depend on the age or sex of the person but does depend on the person's ethnicity. The refractive index of glass varies within and between windows. Observation of glass as to whether it is window or bottle glass will affect the uncertainty associated with the prediction of its refractive index and that of other pieces of glass, which may be thought to come from the same origin.

It may be thought that, because there is variation in scientific observations, it is not possible to make quantitative judgements regarding any comparisons between two sets of observations. The two sets are either different or they are not and there is no more to be said. However, this is not so. There are many phenomena that vary but they vary in certain specific ways. It is possible to represent these specific ways mathematically. Various probability distributions to represent variation are introduced in Appendix A. It is then possible to assess differences quantitatively and to provide a measure of uncertainty associated with such assessments.

It is useful to recognise the distinction between *statistics* and *probability*. Probability is a deductive process that argues from the general to the particular. Consider a fair coin, i.e. one in which when tossed the probability of a head landing uppermost equals the probability of a tail landing uppermost equals 1/2. A fair coin is tossed 10 times. Probability theory enables a determination to be made of the probability that there are three heads and seven tails, say. The general concept of a fair coin is used to determine something about the outcome of the particular case in which it was tossed 10 times.

On the other hand, statistics is an inductive process that argues from the particular to the general. Consider a coin that is tossed ten times and there are three heads and seven tails. Statistics enables the question as to whether the coin is fair or not to be addressed. The particular outcome of three heads and seven tails in ten tosses is used to determine something about the general case of whether the coin was fair or not.

Fundamental to both statistics and probability is *uncertainty*. Given a fair coin, the number of heads and tails in ten tosses is uncertain. The probability associated with each outcome may be determined but the actual outcome itself cannot be predicted with certainty. Given the outcome of a particular sequence of 10 tosses, information is then available about the fairness or otherwise of the coin. For example, if the outcome were 10 heads and no tails, one may believe that the coin is double-headed but it is not certain that this is the case. There is still a non-zero probability (1/1024) that 10 tosses of a fair coin will result in 10 heads. Indeed this has occurred in the first author's own experience. A class of some 130 students were asked to each toss a coin 10 times. One student tossed 10 consecutive heads from what it is safe to assume was a fair coin. The probability of this happening is $1 - (1 - 1/1024)^{130} \simeq 130/1024 = 0.13$. Probability is therefore the measure of choice for the quantification of uncertainty. It is therefore important to define probability carefully. This point is clarified in Section 1.7. At present, it suffices to mention a brief definition by de Finetti (1968).

> *[A probability is]* subjective *[and it] means the degree of belief (as actually held by someone, on the ground of his whole knowledge, experience, information) regarding the truth of a sentence or event, E (a fully specified single event or sentence, whose truth or falsity is, for whatever reason, unknown to that person). (p. 45)*

1.3.1 The Frequentist Method

Consider a consignment of compact disks (CDs), containing N disks. The consignment is said to be of size N. It is desired to make inferences about the proportion θ ($0 \leq \theta \leq 1$) of the consignment which is pirated. It is not practical to inspect the whole consignment so a sample of size n, where $n < N$ is inspected.

The frequentist method assumes that the proportion θ of the consignment that is pirated is unknown but fixed. The data, that is the number of CDs in the sample that are pirated, are variable. A so-called *confidence interval* is calculated. The name *confidence* is used since no probability can be attached to the uncertain event that the interval contains θ. These ideas are discussed further in Chapter 4.

The frequentist method derives its name from the relative frequency definition of probability. The probability that a particular event, A, say, occurs is defined as the relative frequency of the number of occurrences of event A compared with the total number of occurrences of all possible events, over a long run of observations, conducted under identical conditions of all possible events. The limitations of such a definition are presented in Section 1.7.4.

For example, consider tossing a coin n times. It is not known if the coin is fair. The outcomes of the n tosses can be used as information from which the probability of a head occurring on an individual toss may be assigned. There are two possible outcomes, heads (H) and tails (T). Let $n(H)$ be the number of H and $n(T)$ be the number of T such that $n(H) + n(T) = n$. Then the probability of tossing a head on an individual toss of the coin is defined as the limit as $n \to \infty$ of the fraction $n(H)/n$. The frequentist approach relies on a belief in the long-run repetition of trials[3] under identical conditions. This is an idealised situation, seldom, if ever, realised in practice. More discussion on the interpretation of such a result is given in Section 3.6.

The way in which statistics and probability may be used to evaluate evidence is the theme of this book. Care is required. Statisticians are familiar

[3]The use of the word *trial* here is a statistical one and is not to be confused with the legal use. In a statistical use, a trial is a particular event, such as the toss of a coin. More details are in the Appendix A.2.2.

with variation, as are forensic scientists who observe it in the course of their work. Lawyers, however, prefer certainties. A defendant is found *guilty* or *not guilty* (or also, in Scotland, *not proven*). The scientist's role is to testify to the worth of the evidence, the role of the statistician and this book is to provide the scientist with a quantitative measure of this worth. It is shown that there are few forms of evidence that are so definite that statistical treatment is neither needed nor desirable. It is up to other people (the judge and/or the jury) to use this information as an aid to their deliberations. It is for neither the statistician nor the scientist to pass judgement (Kind 1994).

The use of these ideas in forensic science is best introduced through the discussion of several examples. These examples will provide a constant theme throughout the book. Consideration in detail of populations from which the criminal may be thought to have come, to which reference is made in the following text, are discussed in Section 6.1.1 where they are called *relevant populations*. The value of evidence is measured by a statistic known as the *likelihood ratio* and its logarithm. These are introduced in Sections 2.3 and 2.4.

1.3.2 Stains of Body Fluids

Example 1.1. A crime is committed. A bloodstain is found at the scene of the crime. All innocent explanations for the presence of the stain are eliminated. A PoI is found. Their DNA profile is established and found to correspond to that of the crime stain. What is the evidential value of this correspondence? This is a very common situation yet the answer to the question provides plenty of opportunity for discussion of the theme of this book.

Certain other questions need to be addressed before this particular one can be answered. Where was the crime committed, for example? Does it matter? Does the value of the evidence of the bloodstain change depending on where the crime was committed?

Apart from their DNA profile, what else is known about the criminal? In particular, is there any information, such as ethnicity, which may be related to their DNA profile? What is the population from which the criminal may be thought to have come? Could they be another member of the family of the PoI?

Questions such as these and their effect on the interpretation and evaluation of evidence will be discussed in greater detail. First, consider only the evidence of the DNA profile in isolation and one particular locus, *LDLR*. It is no longer used in forensic genetics; it is used here for ease of explanation. Assume the crime was committed in Chicago and that there is eyewitness evidence that the criminal was a Caucasian. Information is available to the investigating officer about the genotypic distribution for the *LDLR* locus in Caucasians in Chicago and is given in Table 1.1. The information about the location of the crime and

Table 1.1 Genotypic relative frequencies for locus *LDLR* amongst Caucasians in Chicago based on a sample of size 200.

Genotype	*AA*	*BB*	*AB*
Relative frequency (%)	18.8	32.1	49.1

Source: From Johnson and Peterson (1999). Reprinted with permissions of ASTM International.

the ethnicity of the criminal is relevant. Genotypic population proportions vary across locations and amongst ethnic groups. A PoI is arrested and a swab is taken for a comparative genetic test. For locus *LDLR* the genotype of the crime stain and that of the PoI correspond. The investigating officer knows a little about probability and works out that the probability of two people chosen at random and unrelated to the suspect having corresponding alleles, using the figures in Table 1.1 as estimates of population proportions, is

$$0.188^2 + 0.321^2 + 0.491^2 = 0.379, \qquad (1.1)$$

(see Section 3.5.1). They are not too sure what this result means. Is it high and is a high value incriminating for the PoI? Is it low and is a low value incriminating? In fact, a low value is more incriminating than a high value.

They think a little more and remembers that, not only do the genotypes correspond, but also that they are both of type *BB*. The proportions of genotypes *AA* and *AB* are not relevant. He works out the probability that two people chosen at random both have genotype *BB* as

$$0.321^2 = 0.103,$$

(see Section 3.5). He is still not too sure what this means but feels that it is more representative of the information available to him than the previous probability, since it takes account of the actual genotypes of the crime stain and the PoI.

The genotype of the crime stain for locus *LDLR* is *BB*. The genotype of the PoI is also *BB* (if it were not they would not be a PoI). What is the value of this evidence? The discussion earlier suggests various possible answers.

(1) The probability that two people chosen at random have the same genotype for locus *LDLR*. This is 0.379.

(2) The probability that two people chosen at random have the same, pre-specified, genotype. For genotype *BB* this is 0.103.

(3) The probability that one person, chosen at random, has the same genotype as the crime stain. If the crime stain is of group *BB*, this probability is 0.321, from Table 1.1.

The relative merits of these answers will be discussed in Section 3.5 for (1) and (2) and Section 2.4.5 for (3).

The phrase *at random* is taken to include the caveat that the people chosen are unrelated to the PoI. In practice, the phrase *at random* is not considered in its scientific usage (a person chosen accordingly to a randomising device). Note that, as expressed by Balding and Steele (2015),

> *It is important to keep in mind that in any crime investigation,* random man *is pure fiction: nobody was actually chosen at random in any population, and so probabilities calculated under an assumption of randomly sampled suspects have no direct bearing on evidential weight in actual cases. (p. 160)*

A comment on randomness is also made by Kingston and Kirk (1964) (see pp. 515–516). A discussion about the extension of the concept of 'random man' to the one of 'unrelated person' to the PoI, when dealing with DNA evidence evaluation, is discussed in Milot et al. (2020).

1.3.3 Glass Fragments

Section 1.3.2 discussed an example of the interpretation of the evidence of DNA profiling. Consider now an example concerning glass fragments and the measurement of the refractive index of these.

Example 1.2. As before, consider the investigation of a crime. A window has been broken during the commission of the crime. A PoI is found with fragments of glass on their clothing, similar in refractive index to the broken window. Several fragments are taken for investigation and their refractive index measurements taken.

Note that there is a difference here from Example 1.1, where it was assumed that the crime stain had come from the criminal and been transferred to the crime scene. In Example 1.2 glass is transferred from the crime scene to the criminal. Glass on the PoI need not have come from the scene of the crime; it may have come from elsewhere and by perfectly innocent means. This is an asymmetry associated with this kind of scenario. The evidence is known as *transfer evidence*, as discussed in Section 1.1, because evidence (e.g. blood or glass fragments) has been transferred from the criminal to the scene or *vice versa*. Transfer from the criminal to the scene has to be considered differently from evidence transferred from the scene to the criminal. A full discussion of this is given in Chapters 5 and 6.

Comparison in Example 1.2 has to be made between the two sets of fragments on the basis of their refractive index measurements. The evidential value of the outcome of this comparison has to be assessed. Notice that it is assumed that none of the fragments has any distinctive features and comparison is based only on the refractive index measurements.

Methods for evaluating such evidence were discussed in many papers in the late 1970s and early 1980s Evett (1977, 1978), Evett and Lambert (1982, 1984, 1985), Grove (1981, 1984), Lindley (1977c), Seheult (1978), and Shafer (1982). These methods will be described as appropriate in Chapters 3 and 7. Knowledge-based computer systems have been developed. See Curran and Hicks (2009) and Curran (2009) for a review of practices in the forensic evaluation of glass and DNA evidence. As an aside, sophisticated systems have been developed to deal with DNA, notably DNA mixtures complexities (i.e. number of donors, peaks heights). Examples are presented and evaluated in Bright et al. (2016), Alladio et al. (2018), and Bleka et al. (2019).

Evett (1977) gave an example of the sort of problem that may be considered and developed a procedure for evaluating the evidence that mimicked the interpretative thinking of the forensic scientist of the time. The case is an imaginary one. Five fragments from a suspect are to be compared with 10 fragments from a window broken at the scene of a crime. The values of the refractive index measurements are given in Table 1.2. The procedure developed by Evett is a two-stage one. It is described here briefly. It is a rather arbitrary and hybrid procedure. While it follows the thinking of the forensic scientist, there are interpretative problems, which are described here, in attempting to provide due value to the evidence. An alternative approach that overcomes these problems is described in Chapter 7.

The first stage is known as the *comparison* stage. The two sets of measurements are compared. The comparison takes the form of the calculation of a statistic, D, say. This statistic provides a

Table 1.2 Refractive index measurements.

Measurements from the window	1.518 44	1.518 48	1.518 44	1.518 50	1.518 40
	1.518 48	1.518 46	1.518 46	1.518 44	1.518 48
Measurements from the PoI	1.518 48	1.518 50	1.518 48	1.518 44	1.518 46

measure of the difference, known as a *standardised* difference, between the two sets of measurements that takes account of the natural variation there is in the refractive index measurements of glass fragments from within the same window. If the absolute value of D is less than (or equal to) some pre-specified value, known as a *threshold* value, then the two sets of fragments are deemed to be similar and the second stage is implemented. If the absolute value of D is greater than the threshold value, then the two sets of fragments are deemed to be dissimilar. The two sets of fragments are then deemed to have come from different sources and the second stage is not implemented. (Note the use here of the word *statistic*, which in this context can be thought of simply as a function of the observations.) A classic example of such an approach is the use of the Student t-test or the modified Welch test for the comparison of means (Welch 1937; Walsh et al. 1996; Curran et al. 2000).

The second stage is known as the *significance* stage. This stage attempts to determine the significance of the finding from the first stage that the two sets of fragments were similar. The significance is determined by calculating the probability of the result that the two sets of fragments were found to be similar, under the assumption that the two sets had come from different sources. If this probability is very low then this assumption is deemed to be false. The fragments are then assumed to come from the same source, an assumption that places the PoI at the crime scene. This assumption says nothing about how the fragments came to be associated with the PoI. This may have occurred in an innocent manner. See Section 5.3.2 for further discussion of this point in the context of activity level propositions.

The procedure can be criticised on two points. First, in the comparison stage the threshold provides a qualitative step that may provide very different outcomes for two different pairs of observations. One pair of sets of fragments may provide a value of D, which is just below the threshold, whereas the other pair may provide a value of D just above the threshold. The first pair will proceed to the significance stage, the second stage will not. Yet, the two pairs may have measurements, which are close together. The difference in the consequences is greater than the difference in the measurements merits (such an approach is called a *fall-off-the-cliff* effect; see Evett (1991) who attributed this term to Ken Smalldon. Criticisms have been developed in Robertson and Vignaux (1995b). They wrote:

> *This sudden change in decision due to crossing a particular line is likened to falling off a cliff, one moment you are safe, the next dead. In fact, rather than a cliff we have merely a steep slope. Other things being equal, the more similar the samples the stronger the evidence that they had a common origin, and the less similar the samples the stronger the evidence that they came from different sources.* (p. 118)

A related problem is that of cut-off where a decision is taken dependent on whether a statistic is above or below a certain value (see Section 7.7.5).

A better approach, as suggested in the quotation from Robertson and Vignaux (1995b) above and that is described in Section 7.3, provides a measure of the value of the evidence that decreases as the distance between the two sets of measurements increases, subject, as explained later, to the rarity or otherwise of the measurements.

The second criticism is that the result is difficult to interpret. Because of the effect of the comparison stage, the result is not just simply the probability of the evidence, assuming the two sets of fragments came from different sources. A reasonable interpretation, as will be explained in Section 2.4, of the value of the evidence is the effect that it has on the odds in favour of the true guilt of the PoI. In the two-stage approach this effect is difficult to measure. The first stage discards certain sets of measurements, which may have come from the same source and does not discard other sets of measurements which may have come from different sources. The second stage calculates a probability, not of the evidence but of that part of the evidence for which D was not greater than the threshold value, assuming the two sets came from different sources. It is necessary, as is seen later, to compare this probability with the probability of the same result, assuming the two sets came from the same source. There is also an implication in the determination of the probability in the significance stage that a small probability for the evidence, assuming the two sets came from different sources, means that there is a large probability that the two sets came from the same source. This implication is unfounded; see Section 2.5.1.

A review of the two-stage approach and the development of a Bayesian approach is provided by Curran et al. (2000) and Curran and Hicks (2009).

As with DNA profiling, there are problems associated with the definition of a suitable population from which probability distributions for refractive measurements may be obtained; see, for example, Walsh and Buckleton (1986).

These examples have been introduced to provide a framework within which the evaluation of evidence may be considered. In order to evaluate evidence, something about which there is much uncertainty, it is necessary to establish a suitable terminology and to have some method for the assessment of uncertainty. First, some terminology will be introduced followed by a method for the measurement of uncertainty. This method is *probability*. The role of uncertainty, as represented by probability, in the assessment of the value of scientific evidence will form the basis of the rest of this chapter. A commentary on so-called *knowledge management*, of which this is one part, has been given by Evett (1993b, 2015).

1.4 TERMINOLOGY

It is necessary to have clear definitions of certain terms. First note that the term *evidence* is generally used in the literature and in practice rather than terms such as *finding*, *outcome*, or *material*. The term *evidence* may be confusing. In some legal contexts it can refer to a judicial qualification of a finding. Forensic scientists are interested in the probative value of an observation before it qualifies as *evidence* in a trial. The European Networks of Forensic Science Institutes (ENFSI) Guideline for evaluative reporting (ENFSI 2015) provides definitions of both terms (pp. 19–20). Despite the risk of confusion, the term *evidence* will be used for the material examined by the scientist and for which the value is required.

The crime scene and suspect (associated with a PoI) materials have fundamentally different roles. The assignment of a probability for a correspondence between two randomly chosen sets of materials is not the important issue. One set of materials, crime scene or suspect, can be assumed to have a known source. It is then required to assess the probability of the corresponding material, suspect or crime scene, corresponding in some sense, to the known set of materials, under two competing hypotheses. Examples 1.1 and 1.2 serve to illustrate this.

Example 1.1. (continued) A crime is committed. A bloodstain is found at the scene of the crime.

All innocent explanations for the presence of the stain are eliminated. A PoI is found. Their DNA profile is found to match that of the crime stain. The crime scene material is the crime stain. The suspect material is a swab with saliva. The finding to evaluate is the observed correspondence between the DNA profiles of the crime scene and suspect materials.

Example 1.2 (continued) As before, consider the investigation of a crime. A window has been broken during the commission of the crime. Several fragments are taken for investigation and measurements made of their refractive indices. These fragments, as their origin is known, are sometimes known as *control fragments* and the corresponding measurements are known as *control measurements*. A PoI is found. Fragments of glass are found on their person and measurements of the refractive indices of these fragments are made. These fragments are sometimes known as *recovered fragments*. Their origin is not known. They may have come from the window broken at the crime scene but need not necessarily have done so.

The crime scene material is the fragments of glass and the measurements of refractive index of these at the scene of the crime. The suspect material is the fragments of glass found on the PoI and their refractive index measurements.

Evidence where the source is known will be known as *source* evidence. These fragments of glass will be known as source fragments and the corresponding measurements will be known as source measurements, as their source is known. An alternative term for this type of evidence was bulk source evidence (Stoney 1991a) but this terminology appears to have fallen into disuse.

A PoI is identified. Fragments of glass are found on their person and measurements of the refractive indices of these fragments are made. Evidence such as this where the evidence has been received and is in particulate form will be known as *transferred particle* evidence. The fragments of glass in this example will be known as *transferred particle* fragments. Their origin is not known. They have been received from somewhere by the PoI. They are particles that have been transferred to the PoI from somewhere. They may have come from the window broken at the crime scene but need not necessarily have done so.

There will also be occasion to refer to the location at which, or the person on which, the evidence was found. Material found at the scene of the crime will be referred to as *crime* evidence. Material found on the suspect's clothing or in the suspect's natural environment, such as their home, will be referred to as *suspect* evidence. Note that this does not mean that the evidence itself is of a suspect nature!

Locard's principle (see Section 1.1) is that every contact leaves a trace. In the earlier examples the contact is that of the criminal with the crime scene. In Example 1.1, the trace is the bloodstain at the crime scene. In Example 1.2, the trace is the fragments of glass that would be removed from the crime scene by the criminal (and, later, hopefully, be found on their clothing).

The evidence in both examples is *transfer evidence* (see Section 1.1) or sometimes *trace evidence*. Material has been transferred between the criminal and the scene of the crime. In Example 1.1 blood has been transferred *from* the criminal *to* the scene of the crime. In Example 1.2 fragments of glass *may* have been transferred *from* the scene of the crime *to* the criminal. Note that the direction of transfer in these two examples is different. Also, in the first example the blood at the crime scene has been identified as coming from the criminal. Transfer is known to have taken place. In the second example it is not known that glass has been transferred from the scene of the crime to the criminal. The PoI has glass fragments on his clothing but these need not necessarily have come from the scene of the crime. Indeed if the PoI is innocent and has no connection with the crime, the fragments will not have come from the crime scene.

The term *control evidence* has been used to indicate the material whose origin is known. Similarly, the term *recovered* has been used to indicate the material whose origin is unknown.

Alternatively, *questioned* has been used for ‘recovered’. See, for example, Brown and Cropp (1987). Also Kind et al. (1979) used the word *crime* for material known to be associated with a crime and *questioned* for material thought to be associated with a crime. All these terms are ambiguous. The need to distinguish the various objects or persons associated with a crime was pointed out by Stoney (1984).

Definitions given in the particular context of fibre evidence are provided also by Champod and Taroni (1999). The object or person on which traces have been recovered is defined as the *receptor*, and the object or person that could be the source (or one of the sources) of the traces, and which is the origin of the material defined as *known material*, is defined as the *known source*.

Material will be referred to as *control* form where appropriate and to *recovered* or *transferred* particle form where appropriate. In the previous examples, there are two possibilities for the origin of the material which is taken to be known: the scene of the crime and the PoI. One or other is taken to be known, the other to be unknown. The two sets of material are compared. Two probabilities for what is assumed known are determined. One depends on an assumption of common source. The other depends on an assumption of different sources. The two possibilities for the origin of the material that is taken to be known are called *scene-anchored* and *suspect-anchored*, where the word 'anchored' refers to that which is assumed known (Stoney 1991a). The distinction between scene-anchoring and suspect-anchoring is important when determining relevant probabilities (Section 5.3.1.4); it is not so important in the determination of likelihood ratios or Bayes' factors (see Section 2.3.2). Reference to form (source or receptor) is a reference to one of the two parts of the evidence. Reference to anchoring (scene or suspect) is a reference to a perspective for the evaluation of the evidence.

It is sometimes useful to refer to material found at the scene of a crime as the *crime scene item* and to material found on or about a PoI as the *suspect item*. This terminology reflects the site at which the material was found. It does not indicate the kind of material (bulk or transferred particle form) or the perspective (scene – or suspect – anchored) by which the evidence will be evaluated.

1.5 TYPES OF DATA

A generic name for observations that are made on items of interest, such as bloodstains, refractive indices of glass, etc. is *data*. There are different types of data and some terminology is required to differentiate amongst them. For example, consider shoe types. The observations of interest are the shoe types as observable on surveillance camera recordings and those observable in possession of a PoI. These shoe types are not quantifiable. There is no numerical significance that may be attached to these. The shoe type is a qualitative characteristic. As such, it is an example of so-called *qualitative data*, sometimes known as *categorical data*. The observation of interest is a quality, the shoe type, which has no numerical significance. The different shoe types are sometime known as *categories*. The assignation of a shoe to a particular category is called a *classification*. A shoe may be said to be assigned to one of several categories (see the discussion on the definition of *identification* in Section 2.5.9). Other examples of categorical data include types of firearms and makes of cars.

It is not possible to order shoe types and say that one type is larger or smaller than another. However, there are other qualitative data that do have a natural ordering, such as the level of burns on a body. There is not a numerical measure of this but the level of burns may be classified as first, second, third degree, for example. Qualitative data that have no natural ordering are known as *nominal* data. Qualitative data to which a natural ordering may be attached are known as *ordinal* data. An ordinal characteristic is one in which there is an underlying order even though it is not quantifiable. Pain is another such characteristic; level of trauma may be ordered as none, slight, mild, severe, or very severe. The simplest case of nominal data arises when an observation may be classified into one of only two possible categories. For example, consider the magnetism of toner present on printed documents. Some toners are magnetic, whereas others are not. Such data are known as *binary*. Alternatively, the variable of interest, here magnetism, is known as *dichotomous*; it is either present or absent (Biedermann et al. 2016c).

Other types of data are known as *quantitative* data. These may be either counts (known as *discrete* data, since the counts take discrete, integer, values) or measurements (known as *continuous* data, since the measurements may take any value on a continuous interval).

A violent crime involving several people, victims and offenders, may result in much blood being spilt and many stains from each of several DNA profiles being identified. Then the numbers of stains for each of the different profiles are examples of discrete, quantitative data. Other examples of discrete quantitative data are the number of glass fragments found on a PoI, or the number of gunshot residue particles on hands.

The refractive indices and elemental concentrations of glass fragments are examples of continuous measurements. In practice, variables are rarely truly continuous because of the limits imposed by the sensitivity of the measuring instruments. For example, refractive indices may be measured only to a certain number of decimal places.

Observations, or data, may thus be classified as qualitative or quantitative. Qualitative data may be classified further as nominal or ordinal, and quantitative data may be classified further as discrete or continuous.

1.6 POPULATIONS

'Who is "random man"?' This is the title of a paper by Buckleton et al. (1991). In order to evaluate evidence, it is necessary to have some idea of the variability or distribution of the evidence under consideration within some population. This population will be called the *relevant* population (and a more formal definition will be given later in Section 6.1.1) because it is the population that is deemed relevant to the evaluation of the evidence. Variability is important because if the PoI did not commit the crime and is, therefore, assumed innocent it is necessary to be able to determine the probability of associating the evidence with them when they are innocent. Surveys of populations are required in order to obtain this information. Surveys for reference data are regularly published in forensic science or legal medicine journals (e.g. *Forensic Science International*, *Science & Justice*, *Journal of Forensic Sciences*, *International Journal of Legal Medicine*); they are widely available to the scientific community.

Care has to be taken when deciding how to choose the relevant population. Buckleton et al. (1991) describe two situations and explain how the relevant population is different for each.

The first situation is one in which there is transfer from the criminal to the crime scene as in Example 1.1 and discussed in greater detail in Section 5.3.1.4. In this situation, the details of any PoI are irrelevant under H_d, the proposition that the PoI was not present at the scene of the crime. Consider a bloodstain at the crime scene that, from the background information I, it is possible to assume is of blood from the criminal.

If the PoI was not present, then clearly some other person must have left the stain. There is no reason to confine attention to any one group of people. In particular, attention should not be confined only to any group (e.g. ethnic group) to which the PoI belongs. However, if there is some information that might cause one to reconsider the choice of population, then that choice may be modified. Such information may come from an eyewitness, for example, who is able to provide information about the offender's ethnicity. This would then be part of the background information I. Further comments on the role of background information I may be found in Section 2.4.4. In general, though, information about DNA profile frequencies would be required from a survey, which is representative of all possible offenders. For evidence of DNA profiles, it is known that age is not a factor affecting a person's profile but that ethnicity is. It is necessary to consider the racial composition of the population of possible offenders (not persons of interest). Normally, it is necessary to study a general population since there will be no information available to restrict the population of possible criminals to any particular ethnic group or groups.

The second situation considered by Buckleton et al. (1991) is possible transfer from the crime scene to the PoI or criminal and discussed further in Section 5.3.2.4. The details of the PoI are now relevant even assuming they were not present at the crime scene. Consider the situation where there is a deceased victim who has been stabbed numerous times. A PoI, with a history of violence, has been apprehended with a heavy bloodstain on their jacket that is not of their own blood. What is the evidential value in itself, and not considering possible DNA evidence, of the existence of such a heavy bloodstain, not of the blood of the PoI? The probability of such an event (the existence of a heavy bloodstain) if the PoI did not commit the crime needs to be considered.

The PoI may offer an alternative explanation. The jury can then assign a probability to the occurrence of the evidence, given that explanation. The two propositions to be considered would then be

H_p: the blood was transferred during the commission of the crime;

H_d: the explanation of the PoI is true,

and the jury could assess the evidence of the existence of transfer under these two propositions. Evaluation of the evidence of the DNA profile frequencies would be additional to this. The two parts could then be combined using the technique described in Section 5.3.2.

In the absence of an explanation from the PoI, the forensic scientist could conduct a survey of persons as similar as possible to the PoI in whatever are the key features of their behaviour or lifestyle. The survey would be conducted with respect to the PoI since it is of interest to learn about the transfer of bloodstains for people with their background. In a particular case, it may be that a survey of people of a violent background is needed. One example is that of Briggs (1978) in which 122 suspects who were largely vagrants, alcoholics, and violent sexual deviants were studied. The nature and lifestyle of the PoI determine the type of population to survey. Buckleton et al. (1991) reported also the work of Fong and Inami (1986) in which clothing items from persons of interest, predominantly in offences against the person, were searched exhaustively for fibres that were subsequently grouped and identified.

The idea of a relevant population is a very important one and is discussed further in Section 6.1.1 following the development proposed by Champod et al. (2004). Consider the example of offender profiling, one which is not strictly speaking forensic science but which is still pertinent during an investigation. Consider the application to rape cases. Suppose the profiler is asked to comment on the offender's lifestyle, such as age, marital status, existence and number of previous convictions, and so on, which the profiler may be able to do. However, it is important to know something about the distribution of these in some general population. The question arises here, as in Buckleton et al. (1991) described earlier, as to what is the relevant population. In rape cases, it may not necessarily be the entire male population of the local community. It could be argued that it might be the population of burglars, not so much because rapists are all burglars first but rather burglars are a larger group of people who commit crimes involving invasion of someone else's living space. Information from control groups is needed, regarding both the distribution of observed traits amongst the general non-offending population and the distribution of similar offences amongst those without the observed traits. Discussions on relevant population have also been published in legal journals, see, for example, Lempert (1991).

1.7 PROBABILITY

1.7.1 Introduction

The interpretation of scientific evidence may be thought of as the assessment of a comparison. The comparison is that between the recovered material (denote this by M_r) and the control material (denote this by M_c). Denote the combination by $M = (M_r, M_c)$. As a first example, consider the bloodstains of Example 1.1. The crime stain is M_r, the recovered evidential material (i.e. evidential material whose source is unknown), and M_c is the genotype of biological material (e.g. blood, saliva swab) taken from the suspect under controlled conditions (i.e. so-called control material whose source is known). From Example 1.2, suppose glass is broken during the commission of a crime. M_c would be the fragments of glass (the control material) found at the crime scene, M_r would be fragments of glass (the recovered material) found on the clothing of a suspect, and M would be the two sets of fragments.

Qualities, such as genotypes, or measurements, such as the refractive indices of glass fragments, are taken from M. Comparisons are made of the measurements made on recovered and control material. Denote these by E_r and E_c, respectively, and let $E = (E_r, E_c)$ denote the combined set. Comparison of E_r and E_c is to be made and the assessment of this comparison has to be quantified. The totality of the evidence is denoted Ev and is such that $Ev = (M, E)$.

Statistics has developed as a subject in which one of its main concerns is the quantification of the assessments of comparisons. The performance of a new treatment, drug, or fertiliser has to be compared with that of an old treatment, drug, or fertiliser, for example. Two sets of materials, control and recovered, are to be compared. It seems natural that statistics and forensic science should come together, and this has been happening over the last 40 years after strong criticisms from some outstanding quarters. Recall Kirk and Kingston (1964). They remarked that

> *When we claim that criminalistics is a science, we must be embarrassed, for no science is without some mathematical background, however meagre. This lack must be a matter of primary concern to the educator [...]. Most, if not all, of the amateurish efforts of all of us to justify our own evidence interpretations have been deficient in mathematical exactness and philosophical understanding. (pp. 435–436)*

They concluded by affirming that

> *It can be fairly stated that there is no form of evidence whose interpretation is so definite that statistical treatment is not needed or desirable. (p. 437)*

As discussed in Section 1.2, there have been several books describing the role of statistics in the law. Until the first edition of this book, there had been none concerned with statistics and the evaluation of scientific evidence. Two factors may have been responsible for this.

First, there was a lack of suitable data from relevant populations. There was a consequential lack of a baseline against which measures of typicality of any characteristics of interest may be determined. One exception are the reference data that have been available for many years on allele frequencies for DNA analysis amongst certain populations. Not only has it been possible to say that the DNA of a PoI corresponded[4] to that of a stain found at the scene of a crime, but also that this profile is only present in, say, 0.01% of the population. Now these have been superceded by results of surveys of allele frequencies in various populations. Announcements of population data are published regularly in peer-reviewed journals such as *Forensic Science International: Genetics* and the *International Journal of Legal Medicine.*

[4]The term *correspond* is used here rather than the more commonly used *match*. The term 'match' suggests certainty or identity which is not always the case with trace evidence. Use of the term 'correspond' emphasises this and is a reminder to be careful with interpretation. See Section 2.5.11 and Friedman (1996) for further comment.

Also, data collections exist for the refractive index of glass fragments found at random on clothing and for transfer and persistence parameters linked to glass evidence; see, for example, Curran et al. (2000), O'Sullivan et al. (2011), and Jackson et al. (2013). Contributions towards characterising the rarity of different fibre types have also been published since the late 1990s; for a review, see Palmer (2016).

Secondly, the approach adopted by forensic scientists in the assessment of their evidence has been difficult to model. The approach has been one of comparison and significance. Characteristics of the control and recovered items are compared. If the examining scientists believe them to be similar, the typicality, and hence the significance of the similarity, of the characteristics is then assessed. This approach is what has been modelled by the two-stage approach of Evett (1977), described briefly in Section 1.3.3 and in fuller detail in Chapter 3. However, interpretation of the results provided by this approach is difficult.

Then, in a classic paper, Lindley (1977c) described an approach that was easy to justify, to implement, and to interpret. It combined the two parts of the two-stage approach into one statistic and is discussed in detail in Section 7.4.3. The approach compares two probabilities, the probability of the evidence, assuming one proposition to be true (e.g. that a PoI is the source of the evidence), and the probability of the evidence, assuming another, mutually exclusive, proposition to be true (e.g. that the PoI is not the source of the evidence). Note that some people use the term *hypothesis* rather than *proposition*; the authors will endeavour to use the term *proposition* as they believe this reduces the risk of confusion of their ideas with the ideas of *hypothesis testing* associated with the alternative term. A proposition is interpreted here as an assertion or statement that, for example, a particular outcome has occurred or a particular state of nature occurs.

This approach implies that it is not enough for a prosecutor to show that there is a low probability to observe the evidence if a PoI is innocent. It should also be more probable to observe the evidence if the PoI is truly guilty. Such an approach has a good historical pedigree (Good, 1950, and also Good, 1991, for a review) yet it had received very little attention in the forensic science literature, even though it was clearly proposed at the beginning of the twentieth century (Taroni et al. 1998; Champod et al. 1999), and earlier by Peirce (1878). It is also capable of extension beyond the particular type of example discussed by Lindley, as will be seen by the discussion throughout this book, for example, in Chapters 6 and 7.

However, in order to proceed it is necessary to have some idea about how uncertainty can be measured. This is best done through probability (Lindley 1991, 1985, 2014). This central role for probability in evidence evaluation is supported by the ENFSI. In the ENFSI Guideline for evaluative reporting in forensic science,[5] is reported, at page 6 (under point 2.3), that:

> *Evaluation of forensic science findings in court uses probability as a measure of uncertainty. This is based upon the findings, associated data and expert knowledge, case specific propositions and conditioning information.*

where the term 'findings' denotes 'evidence' in our usage.

1.7.2 A Standard for Uncertainty

An excellent description of probability and its role in forensic science has been given by Lindley (1991). Lindley's description starts with the idea of a standard for uncertainty. He provides an analogy using the concept of balls in an urn. Initially, the balls are of two different colours, black and white. In all other respects, size, weight, texture, etc., they are identical. In particular, if one were to pick a ball from the urn, without looking at its colour, it would not be possible to tell what colour it was. The two colours of balls are in the urn in proportions b and w for black and white balls, respectively, such that $b + w = 1$. For example, if there were 10 balls in the urn of which 6 were black and 4 were white, then $b = 0.6, w = 0.4$, and $b + w = 0.6 + 0.4 = 1$.

[5]The 2015 Guidelines for evaluative reporting can be found at http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

The urn is shaken up and the balls thoroughly mixed. A ball is then drawn from the urn. Because of the shaking and mixing, it is assumed that each ball, regardless of colour, is equally likely to be selected. Such a selection process, in which each ball is equally likely to be selected, is known as a random selection, and the chosen ball is said to have been chosen *at random*.

The ball, chosen at random, can be either black, an event that will be denoted B, or white, an event that will be denoted W. There are no other possibilities; one and only one of these two events has to occur. The uncertainty of the event B, the drawing of a black ball, is related to the proportion b of black balls in the urn. If b is small (close to zero), B is unlikely. If b is large (close to 1), B is likely. A proportion b close to 1/2 implies that B and W are about equally likely. The proportion b is referred to as the probability of obtaining a black ball on a single random drawing from the urn. In a similar way, the proportion w is referred to as the probability of obtaining a white ball on a single random drawing from the urn.

Notice that on this simple model probability is represented by a proportion. As such it can vary between 0 and 1. A value of $b = 0$ occurs if there are no black balls in the urn, and it is, therefore, impossible to draw a black ball from the urn. The probability of obtaining a black ball on a single random drawing from the urn is zero. A value of

$b = 1$ occurs if all the balls in the urn are black. It is certain that a ball drawn at random from the urn will be black. The probability of obtaining a black ball on a single random drawing from the urn is one. All values between these extremes of 0 and 1 are possible (by considering very large urns containing very large numbers of balls).

A ball has been drawn at random from the urn. What is the probability that the selected ball is black? The event B is the selection of a black ball. Each ball has an equal chance of being selected. The colours black and white of the balls are in the proportions b and w, respectively. The proportion, b, of black balls corresponds to the probability that a ball, drawn in the manner described (i.e. at random) from the urn is black. It is then said that the probability a black ball is drawn from the urn, when selection is made at random, is b. Some notation is needed to denote the probability of an event. The probability of B, the drawing of a black ball, is denoted $\Pr(B)$ and similarly $\Pr(W)$ denotes the probability of the drawing of a white ball. Then it can be written that $\Pr(B) = b$ and $\Pr(W) = w$. Note that

$$\Pr(B) + \Pr(W) = b + w = 1.$$

This concept of balls in an urn can be used as a reference for considering uncertain events. The methodology has been described as follows (Lindley 2006):

> *Your*[6] probability *of the uncertain event of rain tomorrow is the fraction of [black] balls in an urn from which the withdrawal of a [black] ball at random is an event of the same uncertainty for you as that of the event of rain. [. . .] You are invited to compare that event with the standard, adjusting the number of [black] balls in the urn until you have the same beliefs in the event and in the standard. Your probability for the event is then the resulting fraction of [black] balls. (p. 35)*

Another example concerns a hypothetical sporting event. Let R denote the uncertain event that the England football team will win the next major international football championship. Let B denote the uncertain event that a black ball will be drawn from the urn. A choice has to be made between R and B, and this choice has to be ethically neutral. If B is chosen and a black ball is drawn from the urn then a prize is won. If R is chosen and England do win the championship the same prize is won. The proportion b of black balls in the urn is known in advance. Obviously, if $b = 0$ then R is the better choice, assuming, of course, that England do have some non-zero probability of winning the championship. If $b = 1$ then B is the better choice. Somewhere in the interval $[0, 1]$, there is a value of b, b_0 say, where the choice does not matter to You. You are indifferent as to whether R or B is chosen. If B

[6] Note the use of the capitalised words 'You' and 'Your' in this quotation. This is a rhetorical device to help readers keep in mind that probabilities are their own degrees of belief based on the information they have at the time they make the judgement.

is chosen $\Pr(B) = b_0$. Then it said that $\Pr(R) = b_0$ also. In this way, the uncertainty in relation to any event can be measured by a probability b_0, where b_0 is the proportion of black balls, which leads to indifference between the two choices, namely, the choice of drawing a black ball from the urn and the choice of the uncertain event in whose probability one is interested.

Notice, though, that there is a difference between these two probabilities. By counting, the proportion of black balls in the urn can be determined precisely. Probabilities of other events such as the outcome of the toss of a coin or the roll of a die are also relatively straightforward to determine, based on assumed physical characteristics such as fair coins and fair dice. Let H denote the event that when a coin is tossed it lands head uppermost. Then, for a fair coin, in which the outcomes of a head H or a tail T at any one toss are considered as equally likely, the probability the coin comes down head uppermost is $1/2$. Let F denote the event that when a die is rolled it lands 4 uppermost. Then, for a fair die, in which the outcomes $1, 2, \ldots, 6$ at any one roll are equally likely, the probability the die lands 4 uppermost is $1/6$.

Probabilities relating to the outcomes of sporting events, such as football matches or championships or horse races, or to the outcome of a civil or criminal trial, are rather different in nature. It may be difficult to decide on a particular value for b_0. The value may change as evidence accumulates such as the results of particular matches and the fitness or otherwise of particular players, or the fitness of horses, the identity of the jockey, the going of the race track, etc. Also, different people may attach different values to the probability of a particular event.

These kinds of probability – as briefly specified before in Section 1.3 – are sometimes known as *subjective* or *personal* probabilities; see de Finetti (1933), Savage (1954), Good (1959), DeGroot (1970), and the more recent publication by Kadane (2011). Another term is *measure of belief* since the probability may be thought to provide a measure of one's belief in a particular event. A philosophical discussion on the use of those terms is given in Lucena-Molina (2016, 2017). Despite these difficulties the arguments concerning probability still hold. Given an event R whose outcome is uncertain, the probability that R occurs, $\Pr(R)$, is defined as the proportion of black balls b_0 in the urn such that if one had to choose the outcome B (the event that a black ball was chosen) where $\Pr(B) = b_0$ and the outcome R then one would be indifferent to which one was chosen. There are difficulties but the point of importance is that a standard for probability exists. An extended comment on subjective probabilities is given in Sections 1.7.5–1.7.7.

A use of probability as a measure of belief is described in Section 1.7.5 where it is used to represent *relevance*. The differences and similarities in the two kinds of probability discussed earlier and their ability to be combined have been referred to as a duality (Hacking 1975).

It is helpful also to consider two quotes concerning the relationship amongst probability, logic and consistency, both from Ramsey (1931).

> *We find, therefore, that a precise account of the nature of partial beliefs reveals that the laws of probability are laws of consistency, an extension to partial beliefs of formal logic, the logic of consistency. They do not depend for their meaning on any degree of belief in a proposition being uniquely determined as the rational one; they merely distinguish those sets of beliefs which obey them as consistent ones. (p. 182)*

> *We do not regard it as belonging to formal logic to say what should be a man's expectation of drawing a white or black ball from an urn; his original expectations may within the limits of consistency be any he likes; all we have to point out is that if he has certain expectations he is bound in consistency to have certain others. This is simply bringing probability into line with ordinary formal logic, which does not criticise premises but merely declares that certain conclusions are the only ones consistent with them. (p. 189)*

In brief, a person is entitled to their own measures of belief, but must be consistent with them. Ramsey's remarks relate to the appropriateness of a set of probabilities held by a particular individual. This appropriateness needs to be checked. Probability values need to be expressed in an operational way that will also make clear what coherence means and what coherent conditions are. De Finetti (1976) framed the operational perspective as follows:

> *However, it must be stated explicitly how these subjective probabilities are defined, i.e. in order to give an operative (and not an empty verbalistic) definition, it is necessary to indicate a procedure, albeit idealised but not distorted, an (effective or conceptual) experiment for its measurement. (p. 212)*

Therefore, one should keep in mind the distinction between the *definition* and the *assessment* of probability. A description of de Finetti's perspective has been published by Dawid and Galavotti (2009).

One way in which these expressions can be checked is to measure probabilities maintained by an individual in terms of bets the individual is willing to accept. An alternative to consideration of balls in an urn is to consider two lotteries. An individual probability can be determined using a process known as *elicitation*. In this context, elicitation is the comparison of two lotteries of the same price. Consider a situation in which it is of interest to determine a probability for rain tomorrow. This example can be found in Winkler (1996). There are two lotteries:

- Lottery A: Win £100 with probability 0.5 or win nothing with probability 0.5.
- Lottery B: Win £100 if it rains tomorrow or win nothing if it does not rain tomorrow.

In this situation, it is reasonable to assume a person would choose that lottery which, in their opinion, presents the greater probability of winning the prize. If lottery B is preferred, then this indicates that one considers the probability of rain tomorrow to be greater than 0.5. Similarly, a choice of lottery A implies the probability of rain tomorrow is less than 0.5. Additionally, in a case in which one is indifferent between the two lotteries, one's probability for rain tomorrow equates with the probability of winning the prize in lottery A. Therefore, a procedure can be devised in which the probability of winning lottery A is adjusted so that the individual, whose probability for a proposition of interest is to be elicited, is indifferent with respect to lotteries A and B. In a similar manner, the personal probability of an individual for any event of interest can be elicited.

The possibility that subjective degrees of belief may be represented in terms of betting rates in lotteries or in the relative frequency of balls in an urn is often put forward as support for an argument that requires subjective degrees of belief to satisfy the laws of probability. This requirement is satisfied with the notion of *coherence* that has the normative role of forcing people to be honest and to make the best assessment of their own measure of belief.

De Finetti (1931a) showed that coherence, a simple economic behavioural criterion, implies that a given individual should avoid a combination of probability assignments that is guaranteed to lead to loss. All that is needed to ensure such avoidance is for uncertainty to be represented and manipulated using the theory of probability. In this context, the possibility of representing subjective degrees of belief in terms of betting odds is often forwarded as part of a line of argument to require that subjective degrees of belief should satisfy the laws of probability. This line of argument takes two parts. The first is that betting odds should be coherent, in the sense that they should not be open to a sure-loss contract. The second part is that a set of betting odds is coherent if and only if it satisfies the laws of probability. The Dutch Book argument encompasses both parts: the proof that betting odds are not open to a sure loss contract if and only if they are probabilities is called the *Dutch book theorem*. Thus, if an individual translates their state of knowledge in such a manner that the assigned probabilities, as a whole, do not respect the laws of probability (standard probability axioms), then their assignments are not coherent. In this context, such incoherence is also called *logical imprudence*. An example can be found in Section 1.7.6.

1.7.3 Events

The outcome of the drawing of a ball from the urn was called an event. If the ball was black, the event was denoted B. If the ball was white, the event was denoted W. It was not certain which of the two events would happen: would the ball be black, event B, or white, event W? The degree of uncertainty of the event (B or W) was measured by the proportion of balls of the appropriate colour (B or W) in the urn and this proportion was called the probability of the event (B or W). In general, for an event R, $\Pr(R)$ denotes the probability that R occurs.

As underlined by Lindley (2014), events may have happened (past events), may be relevant at the present time (present events), or may happen in the future (future events).

> *There are some things that you [. . .] know to be true, and others that you know to be false; yet, despite this extensive knowledge that you have, there remain many things whose truth or falsity is not known to you. We say that you are* uncertain *about them. You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. (p. xi)*

So, a person may be uncertain about each of these three types of events. Such uncertainty can be expressed by probability.

- *Past event*: A crime is committed and a bloodstain with a particular DNA profile is found at the crime scene. A PoI is found. The event of interest is that the suspect is the source of the stain at the crime scene. Though the PoI either is or is not the source of the stain the knowledge of it is incomplete and hence there is uncertainty about this event. This uncertainty can be expressed by probability.
- *Present event*: A PoI is identified. The event of interest is that they have a particular DNA profile (e.g. Y-STR haplotype). Again, before the result of a DNA analysis is available, this knowledge is incomplete.

- *Future event*: The event of interest is that it will rain tomorrow.

All of these events are uncertain and have probabilities associated with them. Notice, in particular, that even if an event has happened, its actual outcome may be unknown so that knowledge about it is incomplete. The probability the PoI is the source of the stain at the crime scene requires consideration of many factors, including the possible location of the PoI at the crime scene and the properties of transfer of blood from a person to a site. With reference to the gene expression, consideration has to be given to the proportion of people in some population with that gene expression. Probabilistic statements are common with weather forecasting. Thus, it may be said, for example, that the probability it will rain tomorrow is 0.8 (though it may not always be obvious what this means).

1.7.4 Classical and Frequentist Definitions of Probability and Their Limitations

The *classical* definition of probability defines it as the ratio of the number of favourable cases to the total number of possible cases, provided that all cases are equally probable. There is an obvious circularity to this definition. The statement does not define probability, it only offers a way by which it may be evaluated.

The *frequentist* definition of probability is the limit of the relative frequency of a target event that has occurred in a large number of trials, as the number of trials increases to infinity, with the important and unrealistic assumption that the trials are repeated under identical conditions. This definition limits the range of applications since if frequency is to be used as a measure of probability, it must be possible to repeat the underlying experiment a large number of times under identical conditions. Consider a scenario in which a coin is tossed. This definition of probability is equivalent to the assessment of the probability of a head (or tail) by imagining that the act of tossing a coin is able to be repeated a large number of times under identical conditions (e.g. with the same force). The number of heads is observed and the ratio of heads to the total number of tosses is taken as an estimate of the probability of a head for that coin. The frequentist definition of probability is inconceivable operationally for applications in forensic science. A well-known challenge to the frequentist view in the context of criminal law is given by Lindley (1991).

> *There is nothing wrong with the frequency interpretation or chance. It has not been used in this treatment because it is often useless. What is the chance that the defendant is guilty? Are we to imagine a sequence of trials in which the judgements, 'guilty' or 'not guilty', are made and the frequency of the former found? It will not work because it confuses the judgement of guilt, but, more importantly, because it is impossible to conceive of a suitable sequence.*

> *Do we repeat the same trial with a different jury; or with the same jury bur different lawyers; or do we take all Scottish trials; or only Scottish trials for the same offence? The whole idea of chance is preposterous in this context. (p. 48)*

The example makes clear that a definition of probability based on the long-run relative frequency of an event is inapplicable in many situations arising in real life. There are implicit assumptions that must apply in each of the classical and frequentist definitions. These assumptions are, first, that according to our state of knowledge, all cases are equally likely, and, second, that it is theoretically possible to perform an experiment a large number of times under identical conditions. Use of these assumptions to assign a numerical value to a probability implies a judgement that these assumptions are satisfied. A definition of probability that seeks to avoid subjectivity is based on an acceptance of assumptions that are inherently subjective. The frequentist view presumes the possibility of the performance of a long sequence of experiments under identical conditions, with each experiment being physically independent of all other experiments. These assumptions are typically unachievable in many different applied contexts such as history, law, economics, medicine, and, especially, forensic science. In these contexts, the events of interest are usually not the result of repetitive or replicable processes. On the contrary, they are unique.

This aspect has been explicitly underlined by Kingston and Kirk (1964) in the area of forensic science. The authors wrote:

> *There are many philosophically oriented fundamental ideas of probability. Perhaps the most practical basic approach to the subject lies in the concept of frequency, in which statements of probability express the relative frequencies of repeated events. [. . .] For practical use in criminalistics, it is of little interest what might happen in a long series of trials; the crime is committed only once. Of what use, then, is the above frequency concept? The answer to this lies in another way of looking at probability, which is to consider it as a degree of belief. (p. 514)*

Such complications do not arise with the subjective interpretation of probability because that interpretation does not consider probability as a feature of the external world. Instead probability is understood as a notion that describes the relationship between a person (e.g. You, the reader) who makes a statement of uncertainty and the real world to which that statement relates and in which the person acts. With the subjective concept of probability, it is therefore very reasonable to assign a probability to events that are not repeatable, for example, as in a given judicial context. An extended discussion on the limitations of the classical and frequentist definitions of probability can be found in Taroni et al. (2018b).

1.7.5 Subjective Definition of Probability

In forensic science, it is often emphasised that there is a real paucity of numerical data, so that the numerical evaluation of evidence (Section 2.3.1) is sometimes very difficult. Examples of this difficulty are the numerical assessments of parameters such as transfer or persistence probabilities (see Sections 6.2.3 and 6.2.4) or even the relevance of a piece of evidence (see Section 6.3). The Bayesian approach considers probabilities as measures of belief (also called subjective probabilities) since such probabilities may be thought of as measures of one's belief in the occurrence of a particular event. The approach allows scientists to assign their probabilities, not only by certified knowledge and experience, but also by any data relevant for the event of interest, such as knowledge of an event that is often available in terms of a relative frequency. This specific relationship between relative frequency and probability is discussed in Section 1.7.6. Note that *frequency* is a term that relates to data and *probability* is a term that relates to personal belief.

Jurists are also interested in probabilistic reasoning using subjective probabilities, notably probabilities related to the credibility of witnesses and the conclusions that might be drawn from their testimony.

Any kind of uncertainty is assessed in the light of the knowledge possessed at the time of the assessment. This idea is not new. The Italian mathematician Bruno de Finetti (de Finetti, 1931a) defined probability – the measure of uncertainty – as a degree of belief, insisting that probability is conditional on the status of information of the subject who assesses it. So, if a given person, *S*, say, is interested in the probability of an event, E, say, that person's probability $\Pr(E)$ should be written as $\Pr(E \mid I_{s,t})$ where $I_{s,t}$ is the information available to person S at time t.

Subjective probability may be found in many scientific areas (Press and Tanur 2001). In physics, for example, Schrödinger (1947) wrote

> *Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence and thus different numerical probabilities may be attached to the same event. (p. 53)*

He then added that

> *Thus, whenever we speak loosely of the probability of an event, it is always to be understood: probability with regard to a certain given state of knowledge. (p. 54)*

The same perspective is expressed by de Finetti in his famous sentence 'Probability does not exist' (in things) (de Finetti 1975, p. x). Probability is not something that can be known or not known, probabilities are states of mind, not states of nature. This aphorism can also be found in previous statistical and philosophical literature (i.e. de Morgan 1838; Jaynes 2003; Jevons 1913; Maxwell 1990). For example, Jevons (1913) wrote

> *Probability belongs wholly to the mind. This is proved by the fact that different minds may regard the very same event at the same time with widely different degrees of probability [. . .] Probability thus belongs to our mental condition, to the light in which we regard events, the occurrence or non-occurrence of which is certain in themselves. (p. 198)*

Probability is a fact about one's state of mind, not a fact about a phenomenon.

In summary, a person's assessment of their degree of belief (subjective probability) in the truth of a given statement or in the occurrence of an event (i) depends on information, (ii) may change as the information changes, and (iii) may differ from the assessment of others because different individuals may have different information or assessment criteria. In Savage's words (Savage 1954)

> *Probabilistic views hold that probability measures the confidence that a particular individual has in the truth of a particular proposition, for example, the probability that it will rain tomorrow. These views postulate that the individual concerned is in some way 'reasonable', but they do not deny the possibility that two reasonable individuals faced with the same information may have different degrees of confidence in the truth of the same proposition. (p. 3)*

The only constraint in the assessment – as noted in Section 1.7.2 – is that it must be coherent. Coherence may be understood through consideration of subjective probability in terms of betting, for example, on the outcome of a horse race. For the probabilities on winning for each horse in a race to be coherent, the sum of the probabilities over all the horses must be 1. This property characterises a 'reasonable individual'. An example is presented in Section 1.7.6.

For a historical and philosophical discussion of subjective probabilities and a commentary on the work of de Finetti and Savage in the middle of the twentieth century, see Lindley (1980), Lad (1996), Taroni et al. (2001), Dawid (2004), Dawid and Galavotti (2009), Galavotti (2016, 2017), and Zynda (2016).

Savage, like de Finetti, viewed a personal probability as a numerical measure of the confidence a person has in the truth of a particular proposition. This opinion is viewed with scepticism today and was viewed with scepticism then (Savage 1967), as illustrated by Savage (1954).

> *I personally consider it more probable that a Republican president will be elected in 1996 than it will snow in Chicago sometime in the month of May, 1994. But even this late spring snow seems to me more probable than that Adolf Hitler is still alive. Many, after careful consideration, are convinced that such statements about probability to a person mean precisely nothing or, at any rate, that they mean nothing precisely. At the opposite extreme, others hold the meaning to be so self-evident [. . .]. (p. 27)*[7]

1.7.6 The Quantification of Probability Through a Betting Scheme

The introduction of subjective probability through a betting scheme is straightforward. The concept is based on hypothetical bets (Scozzafava 1987):

> *The force of the argument does not depend on whether or not one actually intends to bet, yet a method of evaluating probabilities making one a sure loser if he had to gamble (whether or not he really will act so) would be suspicious and unreliable for any purposes whatsoever. (p. 685)*

[7]Part of this sentence is also reported by Kadane and Schum (1996, p. 160).

Consider a proposition E that can only take one of two values, namely, 'true' and 'false'. There is a lack of information on the actual value of E and an operational system is needed for the quantification of the uncertainty about E imparted by the lack of information. A value $p = \Pr(E)$ is regarded as an amount to be paid to bet on E with the conditions that a unit amount will be paid if E is true and nothing will be paid if E is false. In other words, p is the amount to be paid to obtain an amount equal to the value of E, that is associating the value 1 with 'true' and the value 0 with 'false'. This idea was expressed by de Finetti (1940) in the following terms.

> *The probability of event E is, according to Mr NN, equal to 0.37, meaning that if the person was forced to accept bets for and against event E, on the basis of the betting ratio p which he can choose as he pleases, this person would choose $p = 0.37$. (p. 113)*[8]

Coherence, as briefly described in Section 1.7.2, is defined by the requirement that the choice of p does not make the player a certain loser or a certain winner. Denote an event which is certain, sometimes known as a *universal set*, as Ω and an event which is impossible, sometimes known as the *empty set*, as ϕ so that if $E \neq \Omega$ and $E \neq \phi$ the two possible gains are

$$G_1 = (-p + 1) \text{ if } E \text{ occurs;}$$

$$G_2 = -p \text{ if } E \text{ does not occur.}$$

[8]English version of the paper reprinted in Monari and Cocchi (1993).

When $E = \Omega$ or $E = \phi$, there is no uncertainty in the outcome of the corresponding bet and so the coherence (in the absence of uncertainty) requires the respective gains to be zero. The values of the gains are therefore

$$G(\Omega) = -p + 1 = 0 \text{ and } G(\phi) = -p = 0.$$

This happens when $p = 1$ for $E = \Omega$ and $p = 0$ for $E = \phi$. Therefore if the subjective probability of E, that represents our degree of belief on E, is defined as an amount $p = \Pr(E)$, which makes a personal bet on the event or proposition E coherent, then the probability $\Pr(E)$ satisfies two conditions.

(1) $0 \leq \Pr(E) \leq 1$;
(2) $\Pr(\phi) = 0, \Pr(\Omega) = 1$.

Consider the case of n possible bets on events $E_1, \ldots, E_n$ that partition Ω; i.e. $E_1, \ldots, E_n$ are mutually exclusive and exhaustive (Scozzafava 1987, p. 686). Let $\Pr(E_i), i = 1, \ldots, n$, be the amount paid for a coherent bet on E_i. These bets can be regarded as a single bet on Ω with amount $\Pr(E_1) + \cdots + \Pr(E_n)$. Another condition may be specified from the requirement of coherence, namely

(3) $\Pr(E_1) + \cdots + \Pr(E_n) = 1$.

These conditions are the axioms of probability. Further details are given by de Finetti (1931b) and in

Section 1.7.8. An example of a *Dutch book* is given to examine if a given person assigns subjective probabilities coherently. Consider a horse race with three horses, A, B, and C. A bookmaker offers probabilities of 1/4, 1/3, and 1/2, respectively, for these horses to win. Note that these probabilities add up to more than 1 and so violate condition 3. The corresponding odds are 3 to 1 against winning, 2 to 1 against winning and 'evens'.

The relationship between odds and probability is described briefly here with fuller details given in Chapter 2. An event with probability p of occurring has odds O of happening where $O = p/(1 - p)$. Conversely, an event that has odds of O to 1 of happening has a probability of $O/(O + 1)$ of happening and an event that has odds of O to 1 of not happening has a probability of $1/(O + 1)$ of happening. Odds of 'evens' correspond to $O = 1$ or $p = 1/2$.

Suppose the odds offered by the bookmaker are accepted by the person. Thus, their beliefs do not satisfy the additivity law of probability (condition 3). If any single bet is acceptable, they can all be accepted. This is equivalent to a bet on the certain event that one of A, B, or C wins the race. The individual should therefore expect to break even on the outcome of the race; their winnings will equal their initial stake. Of course, it does not make sense to bet on the certain event as there should then be nothing to win or lose. This assumes the odds are fixed to satisfy condition 3. However, in this example, the odds do not satisfy condition 3 and the person will not break even. Suppose the following bets are placed: £3000 on A to win, £4000 on B to win, and £6000 on C to win; i.e. £13 000 in total. If A wins the bookmaker pays out £12 000, the original £3000 bet and another £9000 in accordance with the odds of 3 to 1 against. If B wins, the bookmaker also pays out £12 000, the original £4000 bet and another £8000 in accordance with the odds of 2 to 1 against. If C wins the bookmaker again pays out £12 000, the original £6000 bet and another £6000 in accordance with the odds of evens. Regardless of which horse has won the race, the individual has paid out £13 000 and receives £12 000 in winnings, thus incurring a loss of £1000. This situation is known as a Dutch book. The odds quoted did not satisfy condition 3. Conversely, if the set of odds determine probabilities that add up to less than 1, then the bookmaker will lose money. It would be incoherent for such odds to be set.

Judgements are required in all aspect of scientific investigation. The elicitation of probability distributions for uncertain quantities represents a challenging work for scientists and decision-makers. O'Hagan (2019) recently wrote:

> *Subjective expert judgments play a part in all areas of scientific activity, and should be made with the care, rigour, and honesty that science demands. (p. 80)*

A discussion can be found in Section 1.7.7.

1.7.7 Probabilities and Frequencies: The Role of Exchangeability

It is not uncommon for subjective (or personal) probabilities to be considered as a synonym for arbitrariness. This is not so; the use of subjectivism does not mean the use of acquired knowledge that is often available for consideration of relative frequencies is neglected. The main source of misunderstanding is concerned with the relationship between *frequencies* and *beliefs*. The two terms are, unfortunately, often regarded as equivalent since frequency data can be used to inform probabilities (Lindley 1991) but they are not equivalent. Dawid and Galavotti (2009, p. 100) quoted de Finetti's view:

> *every probability evaluation essentially depends on two components: (1) the objective component, consisting of the evidence of known data and facts; and (2) the subjective component, consisting of the opinion concerning unknown facts based on known evidence.*

As emphasised more recently by D'Agostini (2016)

> *It is a matter of fact that relative frequency and probability are somehow connected within probability theory, without the need for identifying the two concepts. (p. 13)*

It is reasonable to use relative frequencies to inform measures of belief and the relationship takes the form of a mathematical theorem, de Finetti's *Representation theorem*. According to the theorem, the convergence of one's personal probability towards the value of observed frequencies, as the number of observations increases, is a logical consequence of Bayes' theorem if a condition called *exchangeability* is satisfied by the degrees of belief prior to the observations (Dawid 2004).

As an illustration of the connection between frequency and probability, consider again an urn containing a certain number of balls, indistinguishable except by their colour, which is either white or black, and the number of balls of each colour being known. The extraction of a ball from this urn defines an experiment having two and only two possible outcomes that are generally denoted as success (say, the withdrawal of a white ball) or failure (say, the withdrawal of a black ball). Let W denote the event 'a white ball is extracted'. Under the circumstances that balls are all indistinguishable from each other except for the colour, the subjective probability to extract a white ball can be assessed as the known *proportion* θ of white balls, that is, $\Pr(W \mid \theta) = \theta$. Assuming the urn contains a large number of balls, so that the extraction of a few balls does not alter its composition substantially, individual draws (i.e. sampling[9]) will be considered as with

[9]Note that the use of the term 'sample' in this context is one of a purely technical nature in statistics and has nothing to do with the widespread but inappropriate use of the same term for designating physical trace material recovered or collected in a forensic science context. In particular, the seizure (e.g. at crime scenes) and

replacement and the probability of extracting a white ball at subsequent withdrawals will still be θ, independently on previous observations. In this way one realises a series of so-called *Bernoulli trials* (Section A.2.1), where the outcome of each trial has a constant probability independent from previous outcomes.

Suppose now the observer does not know the absolute value of balls present, nor the proportion that are of each colour. De Finetti (1931a) showed that every series of experiments having two and only two possible outcomes that can be taken as exchangeable (i.e. the probability assigned to the outcomes of a sequence of trials is invariant to permutation) can be represented as random withdrawals from an urn of unknown composition. If one can assess one's uncertainty in such a way that labelling of the trials is not relevant, then it can be proved that as the number of observations increases the relative frequencies of successes (i.e. the relative frequency of white balls) tend to a limiting value that is the proportion θ of white balls. A *subjective* assessment about the outcome of a sequence of Bernoulli trials is equivalent to placing a prior distribution on θ. According to this, one only needs to model a prior distribution $\Pr(\theta)$ for the possible values that θ might take: personal beliefs concerning the colour of the next

analysis of trace material has to deal with the material as it is, irrespective of its condition; there is no such thing as randomisation, for example.

ball extracted can be computed as

$$\begin{aligned}\Pr(W) &= \int_\theta \Pr(W \mid \theta)\Pr(\theta)\, d\theta \\ &= \int_\theta \theta \Pr(\theta)\, d\theta. \qquad (1.2)\end{aligned}$$

The introduction of a prior probability distribution modelling personal belief about θ may seem, at first sight, in contradiction with statements that probability is a single number. One can have probabilities for events, or probabilities for propositions, but not probabilities of probabilities, otherwise one would have an infinite regression (de Finetti 1976). Confusion may arise from the fact that parameter θ is generally termed as *'probability of success'*. However, it is worth noting that, although it is effectively a probability, it represents a *chance* rather than a *belief*.

A set of observations $x_1, \ldots, x_n$ is said to be exchangeable – for you, given a knowledge base – if their joint distribution is invariant under permutation. A formal definition is as follows (Bernardo and Smith 2000):

> *The random quantities* $x_1, \ldots, x_n$*, are said to be judged exchangeable under a probability measure* Pr *if the implied joint degree of belief distribution satisfies* $\Pr(x_1, \ldots, x_n) = \Pr(x_{\pi(1)}, \ldots, x_{\pi(n)})$ *for all permutations* π *defined on the set* $\{1, \ldots, n\}$*. (p. 169)*

Practically, consider the following hypothetical case example. A laboratory receives a consignment of discrete items whose attributes may be relevant within the context of a criminal investigation. The laboratory is requested to conduct analyses in order to gather information that should allow an inference to be drawn, for example about the proportion of items in the consignment that are of a certain kind (e.g. counterfeit products). The term 'positive' is used here to refer to the presence of an item's property that is of interest (e.g. counterfeit); otherwise the result of the analysis is termed 'negative'. This allows the introduction of a random variable X that takes the value 1 (i.e. success) if the analysed unit is positive and 0 (i.e. failure) otherwise. This is a generic type of case that applies well to many situations, such as surveys or, more generally, sampling procedures conducted to infer the proportion of individuals or items in a population who share a given property or possess certain characteristics (e.g. that of being counterfeit). Suppose now that $n = 10$ units are analysed, so that there are $2^n = 1024$ possible outcomes. The forensic scientist should be able to assign a probability to each of the 1024 possible outcomes. At this point, if it was reasonable to assume that only the observed values $x_1, x_2, \ldots, x_n$ matter and not the order in which they appear, the forensic scientist would have a sensibly simplified task. In fact, the total number of probability assignments would reduce from 1024 to 11, since it is assumed that all sequences are assigned the same probability if they have the same number of 1's, (i.e. successes). This is possible if it is thought that all the items are indistinguishable in the sense that it does not matter which particular item produced a success (e.g. a positive response) or a failure (e.g. a negative response). Stated otherwise, this means that one's probability assignment is invariant under changes in the order of successes and failures. If the outcomes were permuted in any way, assigned probabilities would be unchanged. For a coin-tossing experiment, Lindley (2014) has expressed this as follows:

> *One way of expressing this is to say that any one toss, with its resulting outcome, may be exchanged for any other with the same outcome, in the sense that the exchange will not alter your belief, expressing the idea that the tosses were done under conditions that you feel were identical. (p. 148)*

The role of exchangeability in the reconciliation of subjective probabilities and frequencies in forensic science is developed in Taroni et al. (2018b). It is possible to give relative frequency an explicit role in probability assignments but this does not mean that probabilities can only be given when relative frequencies are available.

The existence of relative frequencies is not a necessary condition for the assignment of probabilities. Typically, relative frequencies are not available in the case of single (not replicable) events. Other methods of elicitation, such as scoring rules, can be implemented to deal with such situations. An extended discussion on elicitation is given by O'Hagan et al. (2006).

The use of scores for the assessment of forecasts is described in DeGroot and Fienberg (1983). The association of scores for the assessment of forecasts and the use of scores for the assessment of the performance of methods for evidence evaluation will be made clear later in Section 8.4.3. A score is used to evaluate and compare forecasters who present their predictions of whether or not an event will occur as a subjective probability of the occurrence of that event. A common use for forecasts is that of weather from one day to the next. Let x denote a forecaster's prediction of rain on the following day. Let p be the forecaster's actual subjective probability of rain for that day. Let an arbitrary function $g_1(x)$ be the forecaster's score if rain occurs and let another arbitrary function $g_2(x)$ be their score if rain does not occur. With an assumption that the forecaster wishes to maximise their score, assume that $g_1(x)$ is an increasing function of x and $g_2(x)$ is a decreasing function of x. For a prediction of x and an actual subjective probability of p, the expected score of the forecaster is

$$p\, g_1(x) + (1 - p)g_2(x). \tag{1.3}$$

A *proper scoring rule* is one for which (1.3) is maximised when $x = p$. A *strictly proper scoring rule* is one for which $x = p$ is the *only* value of x that maximises (1.3).

One of the earliest scoring rules, proposed for meteorological forecasts, is the *quadratic scoring rule* (Brier 1950). This score has the property that the forecaster will minimise their subjective expected Brier score on any particular day with a stated prediction x of their actual subjective probability p of rain on that day. The expected Brier score is then

$$p\,(x-1)^2 + (1-p)\,x^2. \tag{1.4}$$

This is minimised uniquely when $x = p$. The negative of the Brier score is a strictly proper scoring rule with $g_1(x) = -(x-1)^2$ and $g_2(x) = -x^2$ (minimisation of a function corresponds to maximisation of the negative of the function).

The notion of exchangeability is illustrated with the following example of selection without replacement of items of a particular type, say, Q, from a small population. As an example of what Q might be, consider tablets in a consignment of drugs; the tablets may be either illicit (Q) or licit. The descriptor 'small' for the population size is used to indicate that removal of a member from the population, as in selection without replacement, effects the probability of possession of Q when the next member is selected for removal.

Denote the population size by N. Of the items in the population, R possess Q and $(N-R)$ do not and R is not known. A sample of size $n(< N)$ is taken. The probability the first item selected from the population is of type Q is R/N. If the first member selected from the population possesses Q, the probability the next member selected also possesses Q is $(R-1)/(N-1)$. The population size N is sufficiently small that $(R-1)/(N-1)$ cannot be approximated meaningfully by R/N. Successive draws from the consignment are not independent in that knowledge of the outcome of one draw affects the probability of a particular outcome at the next draw.

Let X be the number of members of the sample of size n that possess Q. The probability distribution for X is the hypergeometric distribution (Section 4.3.2 and Appendix A.2.5) and

$$\Pr(X = x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}.$$

This distribution does not depend on the order in which the n members are drawn from the population, only on the number x which possess Q and the number $(n - x)$ which do not. The property that the distribution is independent of the order is that of exchangeability.

As R is not known, it is not possible to determine $\Pr(X = x)$. However, it is possible given values for n, N, and x to make inferences about R. A comparison of the frequentist and Bayesian approaches to this small consignment sampling problem is given in Section 4.3.2 and Aitken (1999).

Probabilities based on frequencies may be thought of as objective probabilities. They are considered objective in the sense that there is a well-defined set of circumstances for the long-run repetition of the trials, such that the corresponding probabilities are well-defined and that one's personal or subjective views will not alter the value of the probabilities. Each person considering these circumstances will provide the same values for the probabilities. The frequency model relates to a relative frequency obtained in a long sequence of trials, assumed to be performed in an identical manner, physically independent of each other. Such a circumstance has certain difficulties. This point of view does not allow a statement of probability for any situation that does not happen to be embedded, at least conceptually, in a long sequence of events giving equally likely outcomes. However, note the following words of Lindley (2004):

> *Objectivity is merely subjectivity when nearly everyone agrees. (p. 87)*

1.7.8 Laws of Probability

There are several laws of probability that describe the values that probability may take and how probabilities may be combined as it has been discussed already in Section 1.7.6. These laws are given here, first for events that are not conditioned on any other information and then for events which are conditioned on other information.

The first law of probability, has already been suggested implicitly.

First Law of Probability

Probability can take any value between 0 and 1, inclusive, and only one of those values. Let R be any event and let $\Pr(R)$ denote the probability that R occurs. Then $0 \leq \Pr(R) \leq 1$. For an event that is known to be impossible, the probability is zero. Thus if R is impossible, $\Pr(R) = 0$. For an event that is known to be certain, the probability is one. Thus, if R is certain, $\Pr(R) = 1$. This law is sometimes known as the *convexity rule* (Lindley 1991).

Consider the hypothetical example of the balls in the urn of which a proportion b are black and a proportion w white, with no other colours present, such that $b + w = 1$. Proportions lie between 0 and 1; hence $0 \leq b \leq 1$, $0 \leq w \leq 1$. For any event R, $0 \leq \Pr(R) \leq 1$. Consider B, the drawing of a black ball. If there are no black balls in the urn, this event is impossible then $b = 0$. This law is sometimes strengthened to say that a probability can *only* be 0 when the associated event is known to be impossible.

The first law concerns only one event. The next two laws, sometimes known as the second and third laws of probability, are concerned with combinations of events. Events combine in two ways. Let R and S be two events. One form of combination is to consider the event 'R and S', the event that occurs if and only if R and S both occur, sometimes denoted RS. This is known as the *conjunction* of R and S.

Consider the roll of a six-sided fair die. Let R denote the throwing of an odd number. Let S denote the throwing of a number greater than 3 (i.e. a 4, 5, or 6). Then the event 'R and S' denotes the throwing of a 5.

Secondly, consider rolling two six-sided fair die. Let *R* denote the throwing of a six with the first die. Let *S* denote the throwing of a six with the second die. Then the event '*R* and *S*' denotes the throwing of a double 6.

The second form of combination is to consider the event '*R* or *S*', the event that occurs if *R* or *S* (or both) occurs. This is known as the *disjunction* of *R* and *S*.

Consider again the roll of a single six-sided fair die. Let *R*, the throwing of an odd number (1, 3, or 5), and *S*, the throwing of a number greater than 3 (4, 5, or 6), be as before. Then '*R* or *S*' denotes the throwing of any number other than a 2 (which is both even and less than 3).

Secondly, consider drawing a card from a well-shuffled pack of 52 playing cards, such that each card is equally likely to be drawn. Let *R* denote the event that the card drawn is a spade. Let *S* denote the event that the card drawn is a club. Then the event '*R* or *S*' is the event that the card drawn is from a black suit.

Second Law of Probability

The second law of probability concerns the disjunction '*R* or *S*' of two events. Events are called *mutually exclusive* when the occurrence of one excludes the occurrence of the other. For such events, the conjunction '*R* and *S*' is impossible. Thus $\Pr(R \text{ and } S) = 0$.

If R and S are mutually exclusive events, the probability of their disjunction 'R or S' is equal to the sum of the probabilities of R and S. Thus, for mutually exclusive events,

$$\Pr(R \text{ or } S) = \Pr(R) + \Pr(S). \tag{1.5}$$

Consider the drawing of a card from a well-shuffled pack of cards with R defined as the drawing of a spade and S the drawing of a club. Then $\Pr(R) = 1/4$, $\Pr(S) = 1/4$, $\Pr(R \text{ and } S) = 0$ (a card may be a spade, a club, neither but not both). Thus, the probability that the card is drawn from a black suit, $\Pr(R \text{ or } S)$ is $1/2$, which equals $\Pr(R) + \Pr(S)$.

Consider the earlier example, the rolling of a single six-sided fair die. Then the events R and S are not mutually exclusive. In the discussion of conjunction it was noted that the event 'R and S' denoted the throwing of a 5, an event with probability 1/6. The general law, when $\Pr(R \text{ and } S) \neq 0$, is

$$\Pr(R \text{ or } S) = \Pr(R) + \Pr(S) - \Pr(R \text{ and } S).$$

This rule can be easily verified in this case where $\Pr(R) = 1/2$, $\Pr(S) = 1/2$, $\Pr(R \text{ and } S) = 1/6$, $\Pr(R \text{ or } S) = 5/6$.

Before discussing the third law of probability for the conjunction of two events, it is necessary to introduce the ideas of dependence and independence.

1.7.9 Dependent Events and Background Information

Consider, one roll of a fair die with R, the throwing of an odd number as before, and S, the throwing of a number greater than 3, as before. Then, $\Pr(R) = 1/2, \Pr(S) = 1/2, \Pr(R) \times \Pr(S) = 1/4$ but $\Pr(R$ and $S) = \Pr(\text{throwing a } 5) = 1/6$. Event R and S are said to be *dependent*.

The third law of probability for dependent events was first presented by Bayes (1763) (see also Barnard 1958; Pearson and Kendall 1970; Poincaré 1912). It is the general law for the conjunction of events. Before the general statement of the third law is made, some discussion of dependence is helpful.

It is useful to consider that a probability assessment depends on two things: the event R whose probability is being considered and the information I available when R is being considered. The probability $\Pr(R \mid I)$ is referred to as a *conditional probability*, acknowledging that R is conditional or dependent on I. Note the use of the vertical bar $|$. Events listed to the left of it are events whose probability is of interest. Events listed to the right are events whose outcomes are known and which may affect the probability of the events listed to the left of the bar, the vertical bar having the meaning ‘given’ or ‘conditional on’.

Consider a defendant in a trial who may or may not be truly guilty. Denote the event that they are truly guilty by G. The uncertainty associated with their true guilt, the probability that they are truly guilty, may be denoted by $\Pr(G)$. It is a subjective probability. The uncertainty will fluctuate during the course of a trial. It will fluctuate as evidence is presented. It depends on the evidence. Yet neither the notation, $\Pr(G)$, nor the language, the probability of true guilt, makes mention of this dependence. The probability of true guilt at any particular time depends on the knowledge (or information) available at that time. Denote this information by I. It is then possible to speak of the probability of true guilt given, or conditional on, the information I available at that time. This is written as $\Pr(G \mid I)$. If additional evidence E is presented this then becomes, along with I, part of what is known. What is taken as known is then 'E and I', the conjunction of E and I. The revised probability of true guilt is $\Pr(G \mid E \text{ and } I)$. If the information concerns individual S at time t as in Section 1.7.5 the probability can be written as $\Pr(G \mid E, I_{s,t})$.

All probabilities should be thought of as conditional probabilities. Personal experience informs judgements made about events. For example, judgement concerning the probability of rain the following day is conditioned on personal experiences of rain following days with similar weather patterns to the current one. Similarly, judgement concerning the value of evidence or the guilt of a PoI is conditional on many factors. These include other evidence at the trial but may also include a factor to account for the perceived reliability of the evidence. There may be eyewitness evidence that the PoI was seen at the scene of the crime but this evidence may be felt to be unreliable. Its value will then be lessened.

The value of scientific evidence will be conditioned on the background data relevant to the type of evidence being assessed. Evidence concerning frequencies of different DNA profiles will be conditioned on information regarding ethnicity of the people concerned for the values of these frequencies. Evidence concerning distributions of the refractive indices of glass fragments will be conditioned on information regarding the type of glass from which the fragments have come (e.g. building window, car headlights etc.). The existence of such conditioning events will not always be stated explicitly. However, they should not be forgotten. As stated above, all probabilities may be thought of as conditional probabilities. The first two laws of probability can be stated in the new notation, for events R, S and information I as:

First law of probability for dependent events

$$0 \leq \Pr(R \mid I) \leq 1. \tag{1.6}$$

If I is known, $\Pr(I \mid I) = 1$ and $\Pr(\text{not } I \mid I) = 0$.

Second law of probability for dependent events

$$\Pr(R \text{ or } S \mid I) = \Pr(R \mid I) + \Pr(S \mid I) - \Pr(R \text{ and } S \mid I). \qquad (1.7)$$

Events R and S are said to be dependent if the knowledge that R has occurred affects the probability that S will occur, and *vice versa*. For example, let R be the outcome of a draw of a card from a well-shuffled pack of 52 playing cards. This card is not replaced in the pack so there are now only 51 cards in the pack. Let S be the draw of a card from this reduced pack of cards. Let R be the event '*an Ace is drawn*'. Thus $\Pr(R) = 4/52 = 1/13$. (Note here the conditioning information I that the pack is well-shuffled, with its implication that each of the 52 cards is equally likely to be drawn has been omitted for simplicity of notation; explicit mention of I will be omitted in many cases but its existence should never be forgotten.) Let S be the event '*an Ace is drawn*' also. Then $\Pr(S \mid R)$ is the probability that an Ace was drawn at the second draw, given that an Ace was drawn at the first draw (and given everything else that is known, in particular that the first card was not replaced). There are 51 cards at the time of the second draw of which 3 are Aces. (Remember that an Ace was drawn the first time which is the information contained in R.) Thus $\Pr(S \mid R) = 3/51$. It is now possible to formulate the third law of probability for dependent events.

Third law of probability for dependent events

$$\Pr(R \text{ and } S \mid I) = \Pr(R \mid I) \times \Pr(S \mid R \text{ and } I). \tag{1.8}$$

Thus in the example of the drawing of the Aces from the pack, the probability of drawing two Aces is

$$\begin{aligned}\Pr(R \text{ and } S \mid I) &= \Pr(R \mid I) \times \Pr(S \mid R \text{ and } I)\\ &= \frac{4}{52} \times \frac{3}{51}.\end{aligned}$$

Example 1.3. A study of the brains of 120 road accident fatalities given in Pittella and Gusmäo (2003, Table 2), reproduced in Lucy (2005) observed the numbers of diffuse vascular injuries (DVI) and diffuse axonal injuries (DAI) with the results presented in Table 1.3.

Denote the presence of DVI by R and the presence of DAI by S. Then various probabilities

Table 1.3 Presence and absence of diffuse vascular injuries (DVI) and diffuse axonal injuries (DAI) in 120 road accident fatalities.

	DAI		
DVI	Present	Absent	Total
Present	14	0	14
Absent	82	24	106
Total	96	24	120

Source: From Pittella and Gusmäo (2003). ©ASTM International. Reprinted with permissions of ASTM International.

for the incidences of the two types of injuries in the population of road accident fatalities can be estimated from this sample of 120 fatalities. Thus $\Pr(R)$ is estimated by 14/120, the total number of DVI divided by the total number of fatalities. Similarly $\Pr(S)$ is estimated by 96/120, the number of DAI divided by the total number of fatalities.

The third law of probability for dependent events (1.8) can be verified using Table 1.3. For example,

$$\frac{14}{120} = \Pr(R \text{ and } S) = \Pr(R) \times \Pr(S \mid R)$$
$$= \frac{14}{120} \times \frac{14}{14}.$$

Alternatively

$$\frac{14}{120} = \Pr(R \text{ and } S) = \Pr(S) \times \Pr(R \mid S)$$
$$= \frac{14}{96} \times \frac{96}{120}.$$

Thus, for dependent events, R and S, the third law of probability, (1.8) may be written as

$$\begin{aligned}\Pr(R \text{ and } S) &= \Pr(S \mid R) \times \Pr(R)\\ &= \Pr(R \mid S) \times \Pr(S),\end{aligned} \tag{1.9}$$

where the conditioning on I has been omitted.

1.7.9.1 Independence

If two events R and S are such that, given background information I,

$$\Pr(R \mid I) = \Pr(R \mid S, I)$$

they are said to be *independent*. Uncertainty about R is independent of the knowledge of S. From (1.9) it can be seen that

$$\Pr(RS \mid I) = \Pr(R \mid I) \times \Pr(S \mid I).$$

Independent events are exchangeable. It is not necessarily the case that exchangeable events are independent. See Taroni et al. (2018b) for a discussion. Also, two events which are mutually exclusive cannot be independent. As an example of independence, consider the rolling of two six-sided fair dice, A and B say. The outcome of the throw of A does not affect the outcome of the throw of B. If A lands 6 uppermost, this result does not alter the probability that B will land 6 uppermost. The same argument applies if one die is rolled two or more times. Outcomes of earlier throws do not affect the outcomes of later throws. Similarly, with the drawing of two cards from a pack of 52 cards, if the first card drawn is replaced in the pack, and the pack shuffled, before the second draw, the outcomes of the two draws are independent. The probability of drawing two aces is $4/52 \times 4/52$. This can be compared with the probability $4/52 \times 3/51$ if the first card drawn was not replaced.

Third law of probability for independent events

The third law, assuming R and S independent, and conditional on I is

$$\Pr(R \text{ and } S \mid I) = \Pr(R \mid I) \times \Pr(S \mid I). \qquad (1.10)$$

Notice that the event I appears as a conditioning event in *all* the probability expressions. The laws are the same as before but with this simple extension.

Consider Table 1.3 again. If DVI and DAI were independent then the probability of both occurring in a road accident fatality would be the product of the probability of each happening separately. Thus

$$\begin{aligned}\Pr(R \text{ and } S) = \Pr(R) \times \Pr(S) &= \frac{14}{120} \times \frac{96}{120}\\ &= 0.094,\end{aligned}$$

However, it is not the case that 9.4% of road accident fatalities have both injuries. An examination of Table 1.3 illustrates that this is not so. From Table 1.3 it can be seen that $14/120 = 0.12$ or 12% of fatalities have both injuries. In such a situation where $\Pr(R \text{ and } S) \neq \Pr(R) \times \Pr(S)$ it can be said that DVI and DAI are not independent.

As another example of the use of the ideas of independence, consider a diallelic system in genetics in which the alleles are denoted A and a, with $\Pr(A) = p, \Pr(a) = q; \Pr(A) + \Pr(a) = p + q = 1$. This gives rise to three genotypes that, assuming Hardy–Weinberg equilibrium to hold, are expected to have the following probabilities

- p^2 (homozygotes for allele A),
- $2pq$ (heterozygotes),
- q^2 (homozygotes for allele a).

The genotype probabilities are calculated by simply multiplying the two allele probabilities together on the assumption that the allele inherited from one's father is independent of the allele inherited from one's mother. The factor 2 arises in the heterozygous case because two cases must be considered, that in which allele A was contributed by the mother and allele a by the father, and *vice versa*. Both of these cases have probability pq because of the assumption of independence (see Table 1.4). Note that $p^2 + 2pq + q^2 = (p + q)^2 = 1$. The particular locus under consideration is said to be in Hardy–Weinberg equilibrium when the two parental alleles are considered as independent.

This law may be generalised to more than two events. Consider n events $S_1, S_2, \ldots, S_n$. If they are mutually independent then

$$\Pr(S_1 \text{ and } S_2 \text{ and } \ldots \text{ and } S_n) =$$
$$\Pr(S_1) \times \Pr(S_2) \times \cdots \times \Pr(S_n) = \prod_{i=1}^{n} \Pr(S_i).$$

Table 1.4 Genotype probabilities, assuming Hardy–Weinberg equilibrium, for a diallelic system with allele probabilities p and q.

Allele from mother	Allele from father	
	A (p)	a (q)
A (p)	p^2	pq
a (q)	pq	q^2

1.7.10 Law of Total Probability

Events $S_1, S_2, \ldots, S_n$ are said to be *mutually exclusive and exhaustive* if one of them has to be true and only one of them can be true; they exhaust the possibilities and the occurrence of one excludes the possibility of any other. Alternatively, they are called a *partition*. The event (S_1 or ... or S_n) formed from the conjunction of the individual events $S_1, \ldots, S_n$ is certain to happen since the events are exhaustive and exclusive. Thus, it has probability 1 and

$$\Pr(S_1 \text{ or } \ldots \text{ or } S_n) = \Pr(S_1) + \cdots + \Pr(S_n) = 1, \quad (1.11)$$

a generalisation of the second law of probability, (1.7), for exclusive events. Consider as an example allelic distributions at a locus, e.g. locus *TPOX*. There are five alleles, $8, 9, 10, 11$, and 12, and these are mutually exclusive and exhaustive.

Consider $n = 2$ for events S_1 and S_2. Let R be any other event. The events 'R and S_1' and 'R and S_2' are exclusive. They cannot both occur. The event "'R and S_1' *or* 'R and S_2' " is simply R. For example, let S_1 be male, S_2 be female, R be left-handed. Then

- 'R and S_1' denotes a left-handed male,
- 'R and S_2' denotes a left-handed female.

The event ' "R and S_1" or "R and S_2" ' is the event that a person is a left-handed male or a left-handed female, which implies the person is left-handed (R).

Thus,

$$\begin{aligned}\Pr(R) &= \Pr(R \text{ and } S_1) + \Pr(R \text{ and } S_2)\\ &= \Pr(R \mid S_1)\Pr(S_1) + \Pr(R \mid S_2)\Pr(S_2).\end{aligned}$$

The argument extends to any number of mutually exclusive and exhaustive events to give the law of total probability.

Law of Total Probability

If $S_1, S_2, \ldots, S_n$ are n mutually exclusive and exhaustive events,

$$\Pr(R) = \Pr(R \mid S_1)\Pr(S_1) + \cdots + \Pr(R \mid S_n)\Pr(S_n). \tag{1.12}$$

This is sometimes known as the *extension of the conversation* (Lindley 1991)

An example for blood types and paternity cases is given by Lindley (1991). Consider two possible groups, S_1 (Rh−) and S_2 (Rh+) for the father, so here $n = 2$. Assume the relative frequencies of the two groups are p and $(1 - p)$, respectively. The child is Rh− (event R) and the mother is also Rh− (event M). The probability of interest is the probability a Rh− mother will have a Rh− child, in symbols $\Pr(R \mid M)$. This probability is not easily derived directly but the derivation is fairly straightforward if the law of total probability is invoked to include the father.

$$\begin{aligned}\Pr(R \mid M) = \Pr(R \mid M \text{ and } S_1)\Pr(S_1 \mid M)\\ + \Pr(R \mid M \text{ and } S_2)\Pr(S_2 \mid M).\end{aligned} \tag{1.13}$$

This is a generalisation of the law to include information M. If both parents are Rh−, event (M and S_1), then the child is Rh− with probability 1, so $\Pr(R \mid M \text{ and } S_1) = 1$. If the father is Rh+ (the mother is still Rh−), event S_2, then $\Pr(R \mid M \text{ and } S_2) = 1/2$. Assume that parents mate at random with respect to the Rhesus quality. Then $\Pr(S_1 \mid M) = p$, the relative frequency of Rh− in the population, independent of M. Similarly, $\Pr(S_2 \mid M) = 1 - p$, the relative frequency of Rh+ in the population. These probabilities can now be inserted in (1.13) to obtain

$$\Pr(R \mid M) = 1(p) + \frac{1}{2}(1 - p) = (1 + p)/2,$$

for the probability that a Rh− mother will have a Rh− child. This result is not intuitively obvious, unless one considers the approach based on the law of total probability.

An example using DNA profiles is given in Evett and Weir (1998). According to the 1991 census, the New Zealand (NZ) population consists of 83.47% Caucasians, 12.19% Maoris, and 4.34% Pacific Islanders; denote the event that a person chosen at random from the 1991 NZ population is Caucasian, Maori, or Pacific Islander as *Ca*, *Ma*, and *Pa*, respectively. The probabilities of finding the same YNH24 genotype g (event G) in a crime sample for a Caucasian, Maori, or Pacific Islander are 0.012, 0.045, and 0.039, respectively. These values are the assessments for the following three conditional probabilities: $\Pr(G \mid \text{Ca})$, $\Pr(G \mid \text{Ma})$,

$\Pr(G \mid \mathrm{Pa})$. Then the probability of finding the YNH24 genotype, G, in a person taken at random from the whole population of New Zealand is

$$\begin{aligned}\Pr(G) &= \Pr(G \mid \mathrm{Ca})\Pr(\mathrm{Ca}) + \Pr(G \mid \mathrm{Ma})\Pr(\mathrm{Ma}) \\ &\quad + \Pr(G \mid \mathrm{Pa})\Pr(\mathrm{Pa}) \\ &= 0.012 \times 0.8347 + 0.045 \times 0.1219 \\ &\quad + 0.039 \times 0.0434 \\ &= 0.017.\end{aligned}$$

A further extension of this law to consider probabilities for combinations of genetic marker systems in a racially heterogeneous population has been given by Walsh and Buckleton (1988). Let C and D be two genetic marker systems with realisations C and $\bar{C}$, D, and $\bar{D}$. Let S_1 and S_2 be two mutually exclusive and exhaustive subpopulations such that a person from the population belongs to one and only one of S_1 and S_2. Let $\Pr(S_1)$ and $\Pr(S_2)$ be the probabilities that a person chosen at random from the population belongs to S_1 and to S_2, respectively. Then $\Pr(S_1) + \Pr(S_2) = 1$. Within each subpopulation C and D are independent so that the probability an individual chosen at random from one of these subpopulations is of type CD is simply the product of the individual probabilities. Thus

$$\begin{aligned}\Pr(CD \mid S_1) &= \Pr(C \mid S_1) \times \Pr(D \mid S_1), \\ \Pr(CD \mid S_2) &= \Pr(C \mid S_2) \times \Pr(D \mid S_2).\end{aligned}$$

However, such a so-called *conditional independence* result does not imply unconditional independence (i.e. that $\Pr(CD) = \Pr(C) \times \Pr(D)$). The probability that an individual chosen at random from the population is CD, without regard to his subpopulation membership, may be written as follows

$$\begin{aligned}\Pr(CD) &= \Pr(CDS_1) + \Pr(CDS_2)\\ &= \Pr(CD \mid S_1) \times \Pr(S_1)\\ &\quad + \Pr(CD \mid S_2) \times \Pr(S_2)\\ &= \Pr(C \mid S_1) \times \Pr(D \mid S_1) \times \Pr(S_1)\\ &\quad + \Pr(C \mid S_2) \times \Pr(D \mid S_2) \times \Pr(S_2).\end{aligned}$$

This is not necessarily equal to $\Pr(C) \times \Pr(D)$ as is illustrated in the following example. Let $\Pr(C \mid S_1) = \gamma_1, \Pr(C \mid S_2) = \gamma_2, \Pr(D \mid S_1) = \delta_1, \Pr(D \mid S_2) = \delta_2, \Pr(S_1) = \theta$, and $\Pr(S_2) = 1 - \theta$. Then

$$\begin{aligned}\Pr(CD) &= \gamma_1\delta_1\theta + \gamma_2\delta_2(1-\theta),\\ \Pr(C) &= \gamma_1\theta + \gamma_2(1-\theta),\\ \Pr(D) &= \delta_1\theta + \delta_2(1-\theta).\end{aligned}$$

The product of $\Pr(C)$ and $\Pr(D)$ is not necessarily equal to $\Pr(CD)$. Suppose, for example that $\theta = 0.40$, $\gamma_1 = 0.10$, $\gamma_2 = 0.20$, $\delta_1 = 0.15$, and $\delta_2 = 0.05$. Then

$$\begin{aligned}\Pr(CD) &= \gamma_1\delta_1\theta + \gamma_2\delta_2(1-\theta)\\ &= 0.10 \times 0.15 \times 0.4 + 0.20 \times 0.05 \times 0.6\\ &= 0.0120.\end{aligned}$$

$$\Pr(C) = \gamma_1\theta + \gamma_2(1-\theta) = 0.04 + 0.12 = 0.16.$$

$$\Pr(D) = \delta_1\theta + \delta_2(1-\theta) = 0.06 + 0.03 = 0.09.$$

$$\Pr(C) \times \Pr(D) = 0.0144 \neq 0.0120 = \Pr(CD).$$

1.7.11 Updating of Probabilities

Notice that the probability of true guilt is a subjective probability, as mentioned before (Section 1.7.4). Its value will change as evidence accumulates. Also, different people will have different values for it. The following examples, adapted from similar ones in DeGroot (1970), illustrate how probabilities may change with increasing information. The examples have several parts and each part has to be considered in turn without information from a later part.

Example 1.4.

(a) Consider four events S_1, S_2, S_3, and S_4. Event S_1 is that the area of Lithuania is no more than 50 000 km^2, S_2 is the event that the area of Lithuania is greater than 50 000 km^2 but no more than 75 000 km^2, S_3 is the event that the area of Lithuania is greater than 75 000 km^2 but no more than 100 000 km^2, and S_4 is the event that the area of Lithuania is greater than 100 000 km^2. Assign probabilities to each of these four events. Remember that these are four mutually exclusive events and that the four probabilities should add up to 1. Which do you consider the most probable and what probability do you assign to it? Which do you consider the least probable and what probability do you assign to it?

(b) Now, consider the information that Lithuania is the 25th largest country in Europe (excluding Russia). Use this information to reconsider your probabilities in part (a).

(c) Consider the information that Estonia, which is the 30th largest country in Europe, has an area of 45 000 km^2, and use it to reconsider your probabilities from the previous part.

(d) Consider the information that Austria, which is the 21st largest country in Europe has an area of 84 000 km^2, and use it to reconsider your probabilities from the previous part.

The area of Lithuania is given at the end of the chapter.

Example 1.5.

(a) Imagine you are on a jury. The trial is about to begin but no evidence has been led. Consider the two events: H_p the defendant is truly guilty; and H_d, the defendant is innocent. What are your probabilities for these two events?

(b) The defendant is a tall Caucasian male. An eyewitness says he saw a tall Caucasian male running from the scene of the crime. What are your probabilities now for H_p and H_d?

(c) A bloodstain at the scene of the crime was identified as coming from the criminal. A partial DNA profile has been obtained, with proportion 2% in the local Caucasian population. What are your probabilities now for H_p and H_d?

(d) A window was broken during the commission of the crime. Fragments of glass were found on the defendant's clothing of a similar refractive index to that of the crime window. What are your probabilities now for H_p and H_d?

(e) The defendant works as a demolition worker near to the crime scene. Windows on the demolition site have refractive indices similar to the crime window. What are your probabilities now for H_p and H_d?

This example is designed to mimic the presentation of evidence in a court case. Part (a) asks for a prior probability of guilt before the presentation of any evidence. It may be considered as a question concerning the understanding of the dictum 'innocent until proven guilty'. See Section 2.7 for further discussion of this with particular reference to the logical problem created if a prior probability of zero is assigned to the event that the suspect is guilty.

Part (b) involves two parts. First, the value of the similarity in physical characteristics between the defendant and the person running from the scene of the crime, assuming the eyewitness is reliable, has to be assessed. Secondly, the assumption that the eyewitness is reliable has to be assessed.

In part (c) it is necessary to check that the defendant has the same profile. It is not stated that he has but if he has not he should never have been a defendant. Secondly, is the local Caucasian population the relevant population? The evaluation of evidence of the form in (c) is discussed in Chapter 5.

The evaluation of refractive index measurements mentioned in (d) is discussed in Chapter 7. Variation both within and between windows has to be considered. Finally, how information about the defendant's lifestyle may be considered is discussed in Chapter 6.

It should be noted that the questions asked initially in Example 1.5 are questions that should be addressed by the judge and/or jury. The forensic scientist is concerned with the evaluation of their evidence, not with probabilities of guilt or innocence. These probabilities are the concern of the jury. The jury combines the evidence of the scientist with all other evidence and uses its judgement to reach a verdict. The theme of this book is the evaluation of evidence. Discussion of issues relating to guilt or otherwise of PoIs will not be very detailed.

As a tail piece to this chapter, the area of Lithuania is 65 301 km^2.