2007 Journal for Juridical Science 32(1): 1-43

Bruce Brown & Lirieka Meintjes-van der Walt

The use and misuse of statistical evidence in criminal proceedings*

Summary

This article explains how statistical analysis of evidence can be used by lawyers for interpreting data in criminal justice proceedings. It elucidates basic statistical terms which need to be understood by both prosecutors and defence lawyers in presenting their cases to the court. It also shows that such knowledge is essential in cross-examining an expert in court and how statistical argument should be used correctly to avoid arriving at fallacious deductions. It is impossible in a single article to fully explain the knowledge and technical skills required to become an expert in statistics, however, the legal professional needs to be able to interpret and evaluate statistical concepts and results which are commonly used in criminal law. The use of statistical evidence in the legal system is ever increasing and obligates lawyers as well as expert witnesses to obtain a strong background in the physical sciences including a basic understanding of statistics in order better to make expert testimony intelligible to the decision-maker.

Opsomming

Die gebruik en misbruik van statistiese getuienis in strafverrigtinge

Die artikel verduidelik hoe statistiese ontleding van getuienis deur regsgeleerdes in kriminele hofsake gebruik kan word om gegewens te interpreteer. Basiese statistiese terme wat aanklaers sowel as verdedigers moet verstaan ten einde hul sake in die hof te kan stel word toegelig. Die artikel toon ook aan dat sodanige kennis noodsaaklik is om wetenskaplike deskundiges in die hof te kan ondervra en hoe statistiese bewysgronde korrek gebruik moet word om to voorkom dat valse gevolgtrekkings gemaak word. Dit is nie moontlik om in een enkele artikel volledig te verduidelik watter kennis en tegniese vaardighede nodig is om ‘n deskundige in statistiese metodes te word nie. Dit is egter nodig vir ‘n regsgeleerde om in staat te wees om statistiese konsepte wat alledaags in kriminele reg gebruik word te vertolk en te evalueer. Die gebruik van statistiese getuienis in hofsake is aan die toeneem en maak dit noodsaaklik vir regsgeleerdes en deskundige getuies om ‘n stewige kennis van natuurwetenskappe te bekom. Hierby ingesluit is ‘n basiese begrip van statistiek sodat dit moontlik sal wees om deskundige getuienis meer verstaanbaar vir die hof te maak.

* This article is informed by research carried out by the authors while engaged as consultants to the South African Unit for Expert Evidence and Forensic Skills, Rhodes University for the project “Science for Prosecutors,” conducted for the National Prosecuting Authority. We gratefully acknowledge financial support from the NPA. We are indebted to the peer reviewers for valuable direction during the preparation of this article.

Bruce Brown, Lecturer, Department of Education, Rhodes University, Grahamstown 6140 & Lirieka Meintjes-van der Walt, Adjunct Professor of Law, University of Fort Hare, 21 Prince Alfred Street, Grahamstown 6139.

1. Introduction

As early as the nineteenth century the great American lawyer Oliver Wendell Holmes[1] contended:

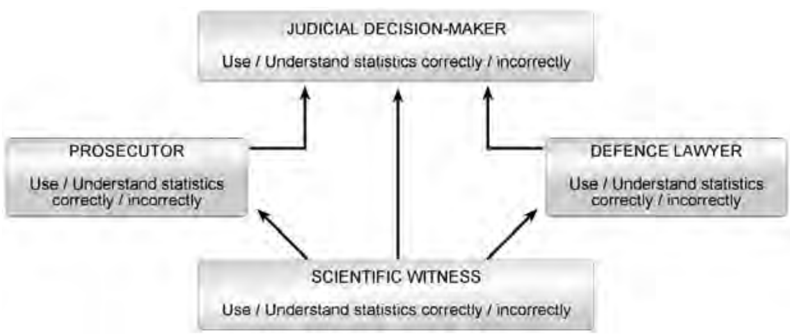
> For the rational study of the law the black letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.

In the 21st century, this statement is more valid than ever. Statistics is the science of quantification of uncertainty and includes methods for collecting data, analysing data and drawing inferences from data.[2] Despite the fact that lawyers often feel alienated when confronted with the field of statistics, the two disciplines have a number of issues in common. Both deal with interpreting evidence, analyzing data, testing hypotheses and making decisions under uncertainty.

The aim of this article is to introduce lawyers to methods of statistical analysis used in legal disputes in criminal proceedings. It sets out to illustrate areas of the criminal law in which statistics have played a role or have purported to do so and in this process illustrates a variety of ways to reason quantitatively. In particular, the article explains basic statistical terminology, the nature of statistical thinking, and how to make appropriate inferences from statistical data.

A further issue that is discussed is how statistical evidence can be presented in such a way that it is properly understood by the court. Ultimately, the lawyer must be in a position to convince the presiding officer of the validity of the statistical techniques involved and the results that they produce.

The following illustration indicates how statistical argument can be used in criminal litigation, and how each of these players may use it either correctly or incorrectly.



1 Holmes 1897:139. Anderson and Twining 1991:387 comment "To update Holmes: 'the trial lawyer of today needs to be a person of statistics.'"

2 Kaye and Freedman 2000:1.

2. Proof and probabilities

It is important for lawyers to note that probabilities in a legal sense differ from the mathematical concept. In England and Wales, lawyers have been reminded of this in a number of judgments:[3]

> The concept of "probability" in the legal sense is certainly different from the mathematical concept; indeed it is rare to find a situation in which the two usages co-exist although when probability may be calculated theoretically (under certain assumptions) for many different situations, it can not be directly measured. Interpreting the theoretical concept of probability also poses a number of difficult problems. These difficulties are manifest in the fact that there are a number of different ways to interpret probability, two of which have broad acceptance in the discipline of statistics. Probability is a theoretical concept used by statisticians to quantify the chances of occurrences of an uncertain event.

Probability is a highly controversial concept, both in law and statistics. William Twining[4] distinguishes four different theories of probability. The first theory is the so-called *classical doctrine of chances*, which can only apply where all outcomes are equiprobable e.g. when throwing an unweighted dice. *Statistical* or *actuarial* probabilities are based on determinations of relative frequency within a given class. The third category of probability judgements can be found in subjective expressions of *degrees of confidence* in some particular proposition, such as "I think there is a likelihood of rain today." As Twining explains, these kind of judgements may be "based to greater or lesser extent on evidence or experience or argument or intuition or irrational beliefs or pure guesswork or a combination of such factors." The subjective probabilities may be Pascalian i.e. expressed in mathematical terms, but need not be. In the final instance, there are inductive or *Baconian* probabilities, encompassing judgments of probability based on reasoning which in principle is non-mathematical. Baconian probabilities are, according to Cohen,[5] *The probable and the provable* (1977) judgments which are based on rational arguments and based on weight or cogency of evidence supporting a particular hypothesis.

One of the main areas of the so-called "new evidence scholarship"[6] is the application of probability theory to arguments about facts in legal cases. With the increased use of statistical evidence in courtrooms, the legal scholarship has given more attention to formal theories of uncertainty. The article, therefore, devotes section 7 to the use of theories that have their origin in mathematics and statistics as a way of approaching legal probabilities and inference in litigation.

---

3 *JS (a minor), Re* [1981] Fam 22 at 29, [1980] 1 All ER 1061 at 1066 and *R v Shepherd* (1988) 85 ALR 387: "Degrees of probability and degrees of proof with which juries are concerned are rarely capable of expressions in mathematical terms."

4 Twining 1980:51 note 31.

5 Cohen 1977:144.

3. Law: lies, damn lies and statistics

In many different fields of expertise, knowledge is expressed as data that is appraised statistically. Expert witnesses testifying in their diverse fields of expertise make use of statistical tools for description and inference.[7] There is also the tendency among experts from other disciplines, among lawyers and even courts, to use statistical presentations in litigation.[8] However, the manner in which statistics is used in litigation is not always correct. This section identifies some of the problems that can arise. In the next section, some basic statistical concepts and methodologies will been explained. Thereafter, we will discuss some solutions and approaches that can be used in respect of the problems identified.

3.1 The relationship to the underlying science

In the forensic context the essential problem is the uncertainty about the relationship between a suspect and trace evidence. All experts who examine traces have to quantify in one way or another the evidential power of their findings.

Thornton and Peterson,[9] indicate that "[b]ehind every individualizing opinion rendered by a forensic examiner there is a statistical basis." Faigman *et al*[10] explain that although this may be so, "[w]e may not know what that basis is, and we may have no feasible means of developing an understanding of that basis, but it is futile to deny that one exists." With the exception of DNA profiling and certain aspects of gunshot residue models, the other forensic sciences do not have the data to submit to conventional statistics.

When using scientific evidence in litigation, the statistical arguments presented need to be justified by an appeal to the underlying science, not by purely statistical reasons. For this reason, the statistical relationships given are expressions of the underlying assumptions on which the scientific analysis is based. When using these relationships to describe a specific situation, it is thus important to ensure that the underlying assumptions do indeed hold in a particular case.

In the context of DNA evidence, it is a prerequisite for the expert to show the statistical significance of a match of DNA pattern.[11] Without valid statistics DNA evidence is meaningless and scientifically unsound. From a statistical viewpoint it has been said[12] that the scientific foundation for fingerprint individuality is weak. Fingerprint evidence does not meet the prerequisite of statistical significance.[13] Mears and Day[14] explain this problem as follows:

---

6 Richard Lempert invented the label "new evidence scholarship." See Lempert 1986:439.
7 DNA evidence is an example.
8 Finkelstein and Levin 2001:x.
9 Thornton and Peterson in Faigman *et al* (2002): vol 1 para 1-5.4 at 15.
10 Faigman *et al supra* 2002: vol 3:169.
11 Epstein 2002:631.
12 Kaye *et al* 2004:455-457.
13 Mears and Day 2002:731 (n48).
14 Mears and Day 2002:731 (n48).

> Due to the lack of testing, a latent fingerprint technician cannot, with consistent scientific accuracy and verifiable reliability, correctly determine whether the ridge characteristics are in common in the two prints under comparison. Given the existence of those matching characteristics, the examiner has no verifiable basis to give an opinion of the probability that the two prints were actually made by the same finger.

Being not only a forensic science scholar, but also a fingerprint examiner, David Stoney[15] points to this current problem:

> [T]here is no justification for fingerprint identifications based on conventional science: no theoretical model, statistics or an empirical validation process.

3.2 Identification and match probabilities

Statistics are often used to interpret the significance of a "match" between crime scene evidence and a reference sample (such as a suspect sample). The random[16] match probabilities that are seen in many laboratory results are a statistical estimate of the expected frequency of a particular profile in a population. These match probabilities are not necessarily probative. For example, if there were a "match" between the blood type (for example, type O positive) of a crime scene sample, and the O positive blood of a suspect, this match would not be particularly probative, since a large percentage of the population has type O positive blood. Champod[17] in an essay on identification and individualisation states:

> The conclusion of a positive identification is then an opinion, a statement of probability[18] expressing the view that the chance of observing on Earth another object or person presenting the same characteristics is zero.

Unfortunately, as lawyers and judicial officers are not trained in basic statistical thinking, the danger exists that inappropriate assumptions are used or that the law applies erroneous statistical calculations or methodology.[19]

3.3 Fallacious calculations by expert witnesses

Even practicing forensic scientists may poorly comprehend the various significant statistical laws that are appropriate to the forensic examinations carried out by them.

---

15 Stoney 2002:386.

16 The chief importance of randomness in research is that, by using it, researchers increase the probability that their conclusions will be valid.

17 Champod in Siegel *et al* 2000:1077; Champod and Evett 2001:103 "[T]he process of identification … is essentially inductive and … probabilistic."

18 Vogt 1993:178 "Probability — likelihood that a particular event or relationship will occur."

19 Finkelstein and Levin *supra* note 8 at x.

In *R v Sally Clark*,[20] Clark was wrongly convicted of murdering two of her babies. Statistical evidence was given by a paediatrician, Sir Roy Meadow, who drew on a published study to obtain an estimate of the frequency[21] of Sudden Infant Death Syndrome (SIDS or "cot death") in families having the same characteristics as the Clark family. The witness went on to square this estimate to obtain a value of 1 in 73 million for the frequency of two cases of SIDS in such a family.

It was later shown that the calculation leading to 1 in 73 million was invalid. It would only be valid if SIDS cases arose independently[22] within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong reasons for supposing that the assumption is false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely than would be the case in another, apparently similar, family. As expert, Meadow stated that: "One sudden infant death is a tragedy, two is suspicious and three is murder, unless proven otherwise."[23]

A separate concern is that the characteristics used to classify the Clark family were chosen on the basis of the same data as was used to evaluate the frequency for that classification. This double use of data is well recognized by statisticians as perilous, since it can lead to subtle yet important biases.

For these reasons, the 1 in 73 million figure cannot be regarded as statistically valid. The Court of Appeal recognized flaws in the calculation, but seemed to accept it as establishing "… a very broad point, namely the rarity of double SIDS."[24] However, not only is the error in the 1 in 73 million figure likely to be very large, it is almost certain to point in one particular direction — against the accused.

3.4 Sampling used incorrectly

Another legal example is provided by *People v Collins*[25] where statistics were incorrectly used by the prosecutor and expert witness. This case concerned the prosecution of an African-American and his Caucasian wife for robbery. The victim testified that her purse was snatched by a girl with a blond ponytail. A second witness testified that he saw a blond girl, ponytail flying, entering a yellow convertible driven by an African-American male with a beard and moustache. Neither of the witnesses could identify the suspects directly. In an attempt to

20 *R v Sally Clark* 2003 WECA Crim 1020. See also http://www.sallyclark.org.uk/judgment03.html (accessed on 17 November 2004). Referred by the Criminal Case Review Commission under Section 9 of the *Criminal Appeal Act* 1995.

21 Vogt 1993:94 "Frequency — the number of times a particular type of event occurs."

22 The concept independence will be discussed below.

23 http://newsvote.bbc.co.uk/mpapps/pagetools/print/news.bbc.co.uk/1/hi/health/443227 (accessed on 31 October 2006).

24 *R v Sally Clark* 2003 WECA 1020 note 20, above para 115.

25 *People v Collins* 68 Cal 2d 319 438 P 2d 33, 66 Cal Rptr 497 (1968).

prove that the accused were in fact the persons who had committed the crime, the prosecutor called a college instructor of mathematics to establish that, assuming the robbery was perpetrated by a Caucasian female with a blond ponytail who left the scene in a yellow Lincoln accompanied by a African-American with a beard and a moustache, there was overwhelming probability that the crime was committed by a couple who had such distinctive characteristics. The California Supreme Court found in *Collins* that the lawyer invented the probabilities without any foundational evidence.[26] This, by itself, should have invalidated any statistical evidence based on these probabilities.

3.5 The defence's and prosecutor's fallacies

3.5.1 The defence lawyer's fallacy

An error likely to be made by defence lawyers assigns probabilities of guilt from transfer evidence in a manner that over-estimates the chances of innocence. Interpreting probabilities in terms of frequency will yield the expected number of people in the population from whom the evidence could have originated, under the assumption that each member of the population was in the position to carry out the offence. The defence's fallacy occurs when the lawyer over-estimates the size of the population of people who were in a position to carry out the offence.

3.5.2 The prosecutor's or lawyers fallacy

An example of the prosecutor's fallacy is to equate the rarity of a DNA profile to the likelihood of innocence. This over-estimates the chances of guilt. Expressing a statistical conclusion in the wrong terms may mislead the judicial decision-maker. The following statement by an expert witness provides an example of the phenomenon:

> The chances of finding the matching profiles if this semen (in the crime stain) had originated from a man in the general population other than and unrelated to the suspect is 1 in 5 million.

Should the statement above be translated into any of the following propositions, such propositions would be misleading and would require evidence other than the scientist's findings of a match:

- The likelihood that the accused is guilty is 5 million to 1;
- The likelihood that the accused is innocent is 5 million to 1;
- The semen is 5 million times more likely to have come from the accused than any other man;
- It is 5 million to 1 against that a man other than the accused left the semen.

26 The lawyer could have satisfied this requirement by having the statistician conduct sampling to establish the probability. The expert (statistician) could also have relied on the probabilities set out in standard, authoritative texts falling within the learned treatise hearsay exception.

One of the first significant cases manifesting the prosecution's or lawyer's fallacy in England is *Doheny and Adams v The Queen*.[27] In this case the court stressed the need to avoid evidence that is compromised by the "prosecution's fallacy":

1) Only one person in a million will have a DNA profile which matches that of the crime stain;

2) The accused has a DNA profile which matches the stain;

3) Ergo there is a million to one probability that the accused left the crime stain and is guilty of the crime.

The error in the lawyer's fallacy arises out of the confusion of two conditional probabilities:

> (1) The probability that a DNA sample taken from an innocent person matches that found at the murder scene GIVEN THAT the person is innocent.
>
> (2) The probability that a person is innocent GIVEN THAT their DNA sample matches that found at the scene of the crime.
>
> These two probabilities are NOT the same and it is clearly the second one that is of interest in determining whether a guilty verdict should be returned or not The First probability is the 1 in 50 000 that is given by the prosecution's expert witness. He is saying that 1 in 50 000 people have a DNA sample like that found at the scene of the crime. Thus if an innocent person is tested there is a one in 50 000 chance that their DNA sample will match!
>
> Notice that in the lawyer's summing up he in effect claims that the probability that the defendant is innocent is 1 in 50 000. In other words, he is claiming that the SECOND probability described above is 1 in 50 000. This is NOT the case.[28]

The error was also committed in *R v Deen*[29] where the expert, after stating that "the likelihood of this being any other man than [the accused] is one in 3 million," concluded by saying "My conclusion is that the semen originated from [the accused]." Meintjes-van der Walt[30] explains that the expert committed an error which was accepted by the jury. The expert confused two questions, namely:

> (i) What is the probability of finding the evidence, given that the accused is innocent?
>
> (ii) What is the probability that the accused is innocent, given the evidence?
>
> The difference between the two questions is succinctly illustrated by Redmayne who considers two different questions which have the same logical structure:

27 [1997] 1 Cr App R 369.

28 The Lawyer's Fallacy http://www.colchsfc.ac.uk/maths/dna/discuss.htm (accessed on 23 April 2005).

29 *The Times* 10 January 1994.

30 Meintjes-van der Walt 2001:216-217.

(i) What is the probability that an animal has four legs given that it is a cow?

(ii) What is the probability that an animal is a cow, given that it has four legs?

4. Quantitative reasoning, mathematics and statistics

At a fundamental level, the power of statistics derives from the use of numbers to lend precision to statistical arguments. Reasoning that makes use of numbers in such a fashion is called **quantitative reasoning**. The use of numbers to precisely describe properties is called **quantification**. To properly understand quantitative reasoning, it is critical to understand the extent of the weight and precision that is lent to an argument by the use of numerical quantities.

A quantity is not merely a number. It is a number that has a carefully formulated interpretation in terms of measurable properties of specific phenomena. The quantity will lend no weight at all to an argument based on a different interpretation of the quantity. Properly used, the value of the quantity is irrelevant to such an argument.

Implicit in the definition of any quantity, is the degree of precision that it is possible to achieve when measuring or calculating the quantity. This provides an absolute limit for the precision of any argument using the quantity. Arguments that attribute a higher degree of precision than can be obtained for a quantity are fallacious, because they attribute too much weight to the quantity. On the other hand, a conclusion based on the assumption that a quantity is less precise than it is, may also be fallacious, because it attributes too little weight to the quantity. This is particularly so when the value of the quantity supports a conflicting conclusion.

With these issues in mind, it may be useful to start with a brief discussion of the process of quantification and its relation to statistics.

4.1 Quantification

Statistics is one of the mathematical sciences. Quantification is a defining feature of these sciences, and one from which they derive much of their power. Instead of describing properties of things in terms of adjectives or metaphors, they construct measurement procedures which make it possible to use numbers to describe these properties as quantities.

According to the Oxford English Dictionary,[31] a **quantity** is:

> In the most abstract sense, esp. as the subject of mathematics: That property of things which is involved in the questions "how great?" or "how much?" and is determinable, or regarded as being so, by measurement of some kind.

31 *Oxford English Dictionary* (2004).

Quantities allow great precision and flexibility when describing properties:

- The unboundedness of the number system means that quantities are able to accurately describe a property, no matter how small or large it is.
- The use of rational numbers (numbers that are not simple counts and are represented by whole numbers, fractions, or decimals) makes it possible to capture subtle variations of a property.

The numerical value of many quantities may be determined through direct measurement. To properly understand the meaning or significance of the value of the quantity, it is important to interpret this value in terms of the process of measurement and so relate it back to the property it describes.

Also, any process of measurement imposes some limit on the accuracy, or precision, of the measurement. Describing or interpreting the measured value of the quantity to a greater degree of precision would be misleading. For example, when measuring distances with a desk ruler, the best we could hope for would be to describe the distance correct to half a millimetre. If we measured a distance of 53.5mm, then a conclusion that is was different from another length measured as 53.4mm would be faulty. Because a difference of 0.1mm would not be evident when measuring with a desk ruler.

In the legal use of statistics, many common quantities are counts of the frequencies of some occurrence. For instance, Aitken and Taroni[32] refer to the frequency of occurrence of some set of properties among glass fragments, or the frequency of occurrence of a DNA profile in a particular population. Forensics also deals with many other quantities that are not simple counts and so are measured using rational numbers. Examples from Aitken and Taroni[33] are the medullar width of hair, or the refractive index of glass fragments.

4.2 Theoretical quantities

As well as allowing precise description, the numbers representing quantities may be related and combined using different mathematical operations. This makes it possible to develop mathematical models, that form the basis for theories describing the interrelationships between the measured quantities. Such models are the fundamental theoretical tools of mathematical sciences and, in particular, of statistics.

A number of quantities arise from successful (accepted) models or theories. These may be termed theoretical quantities. When a theoretical quantity has an important place in the model or theory, scientists attempt to develop an interpretation for the quantity, to relate it directly to the situation being modelled. In many cases, the interpretation of the theoretical quantity may be in terms of the direct measurement of an objective property.

32 Aitken and Taroni 2004:30.
33 Aitken and Taroni 2004:30.

In other cases, the quantity describes a relationship between, or an effect on, directly measurable quantities. This yields an indirect measurement of the quantity. The assertion that such a relationship or effect exists, is a theoretical statement. It follows that such a measurement only makes sense if the theory or model is valid. That is, indirectly measured quantities can only be reliably interpreted within the boundaries set by the theory.

An example of this may be found in psychological assessment. The Millon Clinical Multiaxial Inventory[34] is a personality inventory which may be used in psychological assessment for the purposes of deciding whether a person is fit to stand trial, or is criminally responsible. Scoring this inventory yields a number of personality indices, including indices for aggressiveness and for antisocial tendencies. These indices may only be validly interpreted on the basis of a statistical model of complete response profiles for a given population, informed by a clinical interview carried out by a competent psychologist. Outside this context, these indices are meaningless.

Another example is the refractive index (RI) of glass, which relates the angle of a light ray entering the glass to the angle of the path it follows through the glass.[35] This depends on the fundamental theoretical assertion that such a relation depends only on the glass. Although indirect measurement of the RI is simple and accurate, interpreting the results may still be difficult. The following extract comes from an article by Meintjes-van der Walt:[36]

> Many different kinds of glass are manufactured for different purposes, and each kind, depending on the manufacturer and the type of glass, has a specific refractive index (RI). The forensic scientist needs only a particle with a diameter no greater than that of a human hair, in order to measure the refractive index of the glass sample. The glass particle is placed on a microscope slide and covered with either silicone oil or paraffin-like liquids. The refractive index of these fluids at room temperature rises. Identification is possible because the relationship between temperature and the refractive index of the liquid is known.
>
> It has been noted that the distribution of RI values for sheet glass, beginning roughly in the 1960s, became narrower than that observed previously. Some authors have used this observation to support their contention that more discriminating methods of comparison, such as elemental composition, should be used in glass examination. These assertions are reliant on comparisons of recently acquired databases with older databases, more often than not, those compiled by the FBI Laboratory in Washington, DC, and the Forensic Science Service in the United Kingdom. Kroons points out that it cannot be determined from these articles whether the purported changes in RI distributions reflect changes in glass manufacture or are an artefact of the limited source distribution in small geographical areas of individual studies. Although it is useful to bear all of this in mind it must be noted that the refractive index (RI) is still probably the most useful measurement for discrimination of glass particles.

34 Schutte 2001:251.
35 Bueche 1969:636; See in general Curran, Hicks and Buckleton 2000.
36 Meintjes-van der Walt 2005:61-62.

The question as to whether certain glass particles of the same refractive index as that of a sample of broken glass from a crime or accident scene are likely to be found on clothing at random, is of obvious importance when interpreting glass particles. Thus:

> What is the chance of such a coincidence occurring? The following question requires to be answered: If one takes control fragments from a glass source other than the crime scene, what is the chance they would match the recovered glass found on the accused using the same comparison test. The smaller the chance the greater the value of the observed match would be. This is what Evett and Lambert call the coincidence probability.[37]

A theoretical quantity that is of fundamental importance to statistics, is probability. This quantity will be discussed in detail below.

5. Developing statistical concepts

5.1 Statistics and statistical reasoning

Two concepts are fundamental to statistics: data and chance. Garfield[38] defines statistical reasoning as 'the way people reason with statistical ideas and make sense of statistical information'. So, statistical reasoning involves the collection, representation and interpretation of data. To make valid inferences about data, it is necessary to understand randomness and chance, and to be able to use statistical tools and techniques to reason about chance. Probability is a fundamental tool for inferential statistics, because it provides a quantification of chance.

5.2 Probability — quantifying chance

Probability is a theoretical quantity that is not easily interpreted as a direct measurement. It is used in statistics in relation to sequences of events that are random. A random sequence is not deterministic. That is, before the unfolding of the sequence, it is not possible to predict with certainty what will occur. But they are also not arbitrary. For, we could describe a number of possible sequences that may unfold and then talk with reasonable confidence about the chance of occurrence of each sequence. Probability arises in statistics as a quantification of the chances of occurrence of such random sequences of events.

In section 2 above, it was noted that legal probability differs from statistical probability. But that discussion also underscores the complexity of both the interpretation and the measurement of probability. Because, of the four theories of probability identified by William Twining,[39] the first three relate to different ways

37 Meintjes-van der Walt 2005:61-62.

38 Garfield 2002:31.

39 Twining 1980: 51 note 31.

of interpreting and measuring statistical probability. The *classical doctrine of chances* relates to the frequentist interpretation of probability. *Statistical or actuarial probabilities* relate to the measurement of probability, in that they allow us to empirically estimate probabilities for a chosen population. And *degrees of confidence* are fundamental to the Bayesian[40] interpretation of probability.

In this article, we shall start by taking the classical approach and introduce the concept of statistical probability in a simple form, by means of games of chance.

Games of chance

In the Lotto draw, the winning combination is obtained using a machine to select six numbered balls from a set of 49. For this introduction, we shall simplify the situation by looking at only a single selection. That is, we set out to describe the chance that a single ball (say the '5') will be chosen in one selection. The machine is carefully designed and regularly tested to make sure that each ball has the same chance of being selected. Because of this we may describe this chance by counting the possible outcomes of selection. We count:

- The number of outcomes which select ball '5'. This is 1, because there is only one '5' ball.
- The number of possible outcomes of a single selection. This is 49, because there are 49 balls which may possibly be selected.

Because each ball has the same chance of being selected, comparing these counts is a good way of describing the chances of the '5' being selected. That is: the '5' ball has 1 chance out of 49 possible chances of being selected.

Note that this description involves two numbers (counts), but what is really important is the comparison between the two numbers. It will be technically much easier to work with a single number which describes this comparison. To do this, we note that 1 out of 49 is very much in the same form as the rational (ratio) comparison that we use when working with fractions. For example 1/49 is the fraction used to describe one part out of 49. So, we shall use the corresponding fraction as the single number to describe the comparison. That is:

- The probability of the machine selecting the '5' ball is 1/49.

This is how the theoretical concept of probability is defined and calculated for this specific situation. This process also allows us to interpret the probability in terms of possible outcomes. Thus:

- If the probability of some event happening is 1/49, we may interpret this in terms of equal chances as: The event has 1 chance out of 49 possible chances of happening.

It is important to note that both the definition and the interpretation only make sense when each of the possible outcomes has exactly the same chance of occurring.

40 Vogt 1993:18: "Bayes' theorem — a method of evaluating the conditional probability of an event."

Defining and interpreting the theoretical concept of probability becomes a lot more complex in more general cases where different outcomes have different chances of occurring, or where it is not possible to count all the different possible outcomes. But, even in these cases the basic idea in the simple interpretation given above can often function with some adaptation as a useful underlying metaphor.

Because the use of probability is so frequent, we shall introduce a more compact notation for statements of this type. We shall start by introducing a few terms.

A **random experiment** is any situation that involves uncertainty and which may result in any one of a number of possible outcomes.

> E.g. Carrying out a single selection is a random experiment. This experiment has 49 possible outcomes — each being the selection of a specific numbered ball.

Often we are not concerned with the occurrence of one specific outcome, but are interested in the occurrence of any one of a chosen group of the possible outcomes. To deal with this, a collection (set) of possible outcomes of a random experiment is called an **event**.

> E.g. Selecting a ball with number less than 6 is an event made up of five different outcomes. Selecting the '5' ball is an event which consists of a single outcome.

It is useful to introduce a symbolic name for the event of interest.

> E.g. 'A' is the event that the '5' ball has been selected by the machine.
>
> Then the probability of the event A occurring is denoted by P(A).
>
> In our example, we have: P(A) = 1/49.

A comparative concept closely related to probability is that of "**odds**". This concept is useful when one of two mutually exclusive events must occur. Odds involve a comparison of the chances of each of the events occurring. For example, if there are 4 chances of the first event occurring and 10 chances of the second event, then the odds of the first event occurring are 4 to 10 (or, in a more concise form, 4:10).

Odds are often used when comparing the chances of an event occurring, with the chances of the same event not occurring. In this case the odds are:

- The chances of the event occurring: The chances of the event not occurring.

Dividing each of these chances by the same value will yield the same comparison. Because the probability of each of these occurrences is obtained by dividing by the number of possible chances, the odds can then also be expressed as:

- The probability of the event occurring: The probability of the event not occurring.

This is particularly relevant to criminal cases where one would like to compare the chances that the accused may be guilty with the chances that they may not be guilty.

Generally both chances are given when describing the odds, but in some situations it may be technically more convenient to express the odds as a fraction. We shall see an example of this later when we discuss a Bayesian approach to hypothesis testing.

5.3 Interpreting statistical probabilities

5.3.1 The non-deterministic nature of probabilistic conclusions

Probability was developed to describe and analyse non-deterministic processes. When interpreting probability, it is important to bear in mind the non-deterministic relationship between the probability of an event occurring in a given situation and the actual occurrence of the event in the situation.

Because of our tendency to think in a deterministic fashion, this non-determinism may at times yield conclusions which appear rather counter-intuitive. This is particularly relevant when dealing with the following two issues.

i. The relevance of the history of previous outcomes

Returning to our simple example, recall that the probability of the '5' ball being selected by the lottery machine is 1/49. But this does not mean that the '5' ball will appear exactly once in every 49 selections. It may appear more than once, or it may not appear at all. If we look at our development of the concept, at no stage did we consider repeated selections. The probability merely specifies that for a single selection, the appearance of the '5' is one outcome out of 49 equally likely possible outcomes. The history of previous selections has no bearing on the outcome of the next selection.

The idea that the history of previous outcomes has an effect on the outcome of the current experiment, is a very common misinterpretation of probability often called the "gambler's fallacy". That it is a fallacy appears quite obvious when only a few selections have been made. But when a large number of selections have been made, it is very easy to make this mistake. For example, if 48 selections have been made and the '5' has not appeared, then it is not the case that the '5' will appear at the next selection — the chances of the '5' appearing remains 1 in 49. In fact, even if 500 selections have been made and the '5' has not appeared, the chances of the '5' appearing at the next selection are still only 1 in 49.

Saks and Kidd[41] give an example of this fallacy:

> After observing three consecutive red wins, a group of people playing roulette start to switch their bets to black. After red wins on the fourth and fifth spins, more and more players switch to black, and they are

41 Saks and Kidd 1980-1981:123.

> increasingly surprised when the roulette wheel produces a red win the sixth, and then the seventh time. In actuality, on each spin the odds of a red win remain constant at 1:1. The shifting of bets to black was irrational, as was the strong subjective sense that after each successive red win, black became more likely.[42]

Lay people err in believing that a small local sequence of events will be representative of an infinite sequence. Although compelling to human intuition, common-sense judgements such as "lightning will not strike twice in the same place" or that a cricket player who has not had a hit in some time is "due" for one, are nevertheless wrong.[43]

The question of what happens when repeated selections are made is a very important question in probability theory and statistics for most substantive questions in statistics are of this form. Providing accurate quantitative answers to such questions requires the development of many of the technical skills of the science of statistics.

ii. The occurrence of events with low and high probabilities

Probability is expressed as a fraction between 0 and 1 and describes the chances of an event occurring in an experiment. The lower the probability (closer to 0), the smaller the chances of occurrence of the event. And the higher the probability (closer to 1), the greater the chances of the event occurring.

It follows that a very low probability means that the event has almost no chance of occurring. But this understanding has to be balanced against the fact that the experiment will indeed yield an outcome. This outcome may have a high probability of occurrence, but it may also have an extremely low probability of occurrence. For as long as an outcome has a non-zero probability of occurrence, there is some chance that it may occur.

The lottery provides a good example of this. Let's look at the full selection process for the lottery. In this case, an outcome of the selection is a set of six numbers from 1 to 49. The probability that any specific outcome will occur can be calculated as approximately

$$0.000\,000\,072$$

This is a very small probability, but with every draw, a single combination is chosen, even though its probability of being chosen is so small.

On the other hand, if an event has a probability close to 1 (a high probability), then it has a very good chance of occurring. But again, this does not mean that it certainly will occur, for an alternative outcome (which will have a very low chance of occurring) may well be chosen. Again, using the lottery as an example, the probability of making a selection that includes at least one number

42 Saks and Kidd 1980-1981:127.

43 Saks and Kidd 1980-1981:127. Understanding the fact that the history of previous outcomes does not influence a statistical result, a fundamental concept to statisticians, makes Sir Roy Meadow's "law": "one sudden infant death is a tragedy, two is suspicious and three is murder" so much more atrocious.

greater than 15, is approximately 0.9998. But this does not mean that every selection will have at least one number greater than 15. In fact, on 31 October 2001, the selection drawn was 7, 5, 3, 14, 11, 4, which has all numbers less than 15.

Next we shall look at the two approaches to interpreting statistical probability, viz. the frequentist or objective approach and the Bayesian or subjective approach of probability.

5.3.2 The frequentist, classical, or objective interpretation of probability

It is also important for lawyers to know whether the expert giving evidence is using the frequentist or Bayesian approach of an eclectic combination of both to the interpretation of probability.

Frequentists see probability as a mathematical tool for analysing uncertain objective processes. The frequentist interpretation of probabilities is based on the fundamental result that when a large number of trials are carried out, the relative frequency of an outcome provides a good approximation to the probability of the occurrence of the outcome.

The **relative frequency** of an outcome is the frequency of the outcome, expressed as a fraction of the total number of trials. That is:

Relative frequency = Frequency / (Total number of trials).

Note that this approach to interpreting probabilities is quite similar to our use of “chances” to calculate probabilities theoretically. In fact, if we were sure beforehand that each outcome had the same chance of occurring, and if we could itemise each possible outcome, then we could calculate the probabilities by means of the relative frequencies of the event outcomes as compared to the total number of possible outcomes. The frequentist interpretation takes the importance of relative frequencies beyond calculations in the theoretical realm, by asserting that measurements of relative frequencies can be used empirically to estimate probabilities.

It is important to note that according to the strict frequentist approach, probability only applies to uncertain events. However, an event that has already occurred is no longer uncertain. As a result, this interpretation of probabilistic analysis is not well suited to decisions about matters of fact such as in a criminal trial. For example, according to the frequentist interpretation, the question (in a sexual offence case) “What is the probability that this is the accused’s semen?” has only two answers: it either is or it isn’t. The probability is therefore, either one or zero.

i. Using relative frequencies to describe probabilities

Relative frequencies are often viewed as a manner of describing probabilities that is more accessible to the non-expert. For example, the random match probability for a DNA sample may be described as follows:

> With the probability of 1 in 7 million it can be said that the blood on the clothing of the suspect could be that of someone other than the victim.[44]

Statements of this form need to be interpreted bearing in mind the fact that this relative frequency represents a probability and as such describes a non-deterministic possible outcome of a single experiment. It does not describe the combined outcomes of many possible experiments.

So, if the community in question had less than 7 million members, the above example does not mean that no other match could be found. Or, if the community had 14 million members, this does not mean that 2 matches would be found. Rather, this statement should be interpreted as expressing the conclusion that if blood were taken from a single person other than the victim, the chance of that blood providing a match is low (1 in 7 million).

ii. Defence's fallacy

The defence's fallacy was already brought to your attention under 3.5.

The defence fallacy assumes that in a given population, anyone with an identical DNA profile as the forensic sample is as likely to have deposited the forensic sample as the accused. For example, if 150 persons in a given area are believed to have the same DNA as the forensic sample, it would be a fallacy on the part of the defence to argue that the probability that the accused contributed the forensic sample is only 1/150. This is because usually other evidence would have been attributed to the accused, that did not exist for the other 149 other persons expected to have the same DNA profile.[45]

5.3.3 The Bayesian, or subjective interpretation of probability

Bayesians take a more subjective interpretation of probability, using it to describe a person's 'degree of belief' in a proposition or hypothesis. That is, Bayesians assign probabilities to propositions or hypotheses that are uncertain. The probability assigned to a proposition is then interpreted as the degree of validity that a rational person would assign to the proposition. When the proposition asserts something about the outcomes of an objective process, it is assumed that the degree of validity assigned to the proposition is directly related to the chances that the outcome stated in the proposition does indeed occur.

Relative frequencies of occurrence are still important for the Bayesian interpretation. But when setting probabilities, Bayesian analysts also allow for the inclusion of prior probabilities, which are not necessarily determined through objective experimentation. For this reason, the Bayesian interpretation is often called the subjective interpretation of probability.

44 *Wike vs State*, transcript, 147-148, given in JJ Koehler, "DNA matches and statistics: important questions, surprising answers." *Judicature* 76:222-229. *Wike v State*. 596 So 2nd 1020, Fla S. Ct. 1992 81.

45 *Commonwealth v McNickles* 753 NE 2d 131 (Mass 2001).

The Bayesian interpretation relies for its applicability on a technical result known as Bayes' theorem. This result is discussed later in this article, after which the use of the Bayesian interpretation in legal settings is investigated.

5.4 Some basic analytical tools in statistics

Almost all analytical processes in statistics depend on some underlying assumptions. The result of the analysis is only valid if the underlying assumptions are satisfied in the specific case being investigated. To understand and feel confident of the analysis thus requires two things:

- understanding the calculations and the technical details of the analytical process; and
- understanding the assumptions underlying the analysis and being confident that they are indeed valid when interpreted in terms of the case being considered.

Developing a detailed understanding of the analytical tools of statistics requires extensive training and is beyond the scope of this article. It aims instead to develop a general understanding of some of the more important assumptions that undergird statistical analysis and ways that these assumptions may be interpreted in terms of case specifics. To do this, some of the basic analytical tools of statistics are explored and the assumptions underlying these tools are discussed.[46]

5.4.1 Sampling issues

Most statistical analyses are based on an underlying probability model for the situation of interest. That is, the statistician starts by describing how the probabilities of the occurrence of different events are distributed between the events. In some cases, the model may be chosen theoretically, on the basis of structural similarities between the situation of interest and another well-researched situation, which has a generally accepted probability model. In other cases, the model of the probability distribution needs to be determined empirically. This is generally done by carrying out a survey, which investigates a sample of the different populations relevant to the situation and counts the frequencies of different events in the sample. The probability distribution is then set on the basis of these frequencies.

46 The authors have done tried to simplify the statistics as far as possible for the average lawyer or judge to understand, without reflecting a skewed view of statistics and thereby doing injustice to both disciplines. The approach followed by Buckleton *et al* 2005:29 with regard to interpreting DNA evidence informed our approach: "It is important that the following discussion is read with fear of the more mathematical approaches as the fear wrongly pressures some commentators to advocate simpler approaches. It is probably fair for a jury to prefer a method for the reason of mathematical simplicity, but it would be a mistake for a scientist to do so. Would you like your aircraft designer to use the best engineering models available or one that you can understand without effort?"

Carrying out such a survey and setting these probabilities is not a simple undertaking. It is generally not possible to investigate all the members of the population and as a result, the survey may only be carried out on some sample of the population. A sample is said to be representative if the probability estimates it yields provide an accurate representation of the probabilities for the entire population. If the sample is not representative, the probabilities it yields will not accurately represent those of the population and this will introduce errors and inaccuracies into any inferences or conclusions drawn from this probability distribution.

Whether or not a sample is representative, depends on both the size of the sample, and the manner in which it is chosen. Random sampling involves choosing the members of the sample randomly from the population as a whole. A random sample of sufficient size will generally be representative, although the size may need to be rather large if the population is not homogeneous. Fox and Levin[47] provide more details on the statistics involved in random sampling and how a random sample may be used to estimate population means and frequencies.

If the population is not homogeneous, but has a well-understood structure, stratified random sampling[48] will ensure that the structure of the population is preserved in the sample, while preserving randomness within each structural grouping. This can greatly reduce the size of the sample of a structured population, while ensuring that it is still representative.

Creating random or stratified random samples is generally a difficult and expensive task. In contrast to this, convenience sampling involves surveying a sample which has been created for some other reason, and as such is convenient for the purposes of the survey.[49] Convenience samples[50] are generally not representative because the sample will be constituted as a result of some criteria for choice, which in all probability will introduce a bias into the result. If a survey is based on a convenience sample, it will be important to analyse the sample for possible bias.

A forensic laboratory would not take a random, representative, stratified sample of individuals to determine a DNA match probability, but will use a convenience[51] sample.

47 Fox and Levin 2005:33.

48 Vogt 1993:223: "Stratified random sampling — Random or probability samples drawn from particular categories (or "strata") of the population being studied."

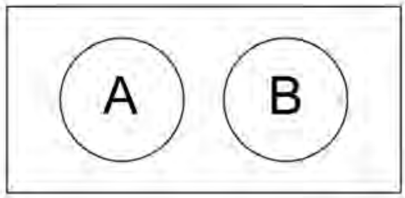
49 Kaye and Freedman *supra* note 2.

50 Vogt: 1993:48: "Convenience sample — A sample of subjects selected for a study not because they are representative but because it is convenient to use them e.g. a professor studying his own students."

51 This is also the practice followed by the South African Police Service — Forensic Science Laboratory.

5.4.2 Mutually exclusive events

Two events in a random experiment are mutually exclusive if they cannot both occur. This can be illustrated by the following Venn diagram:



Mutually exclusive events are quite useful for calculating probabilities, because the probability (P) of either event occurring is equal to the sum of the probabilities of each event occurring.[52] That is, if the events are labelled A and B,

P(A or B) = P(A) + P(B).

Evett and Weir[53] illustrate how certain events can be mutually exclusive. They use the example of throwing a dice. Given that all the conditions to which we can agree for a fair tossing are present, then we also agree that the probability of any one score is 1/6:

> If we ask the probability of the event that the dice will show an even number, few people will hesitate before replying 1⁄2 and this is the result of adding together the probabilities for a 2, a 4 and a 6. There is also one very important feature of these last three events that we must emphasize: they are ***mutually exclusive***. If any one of them occurs, then none of the others has occurred. (Emphasis in original.)[54]

The result may be generalised in a straightforward fashion to deal with a number of events which are pairwise mutually exclusive (each pair of events is mutually exclusive).

Using this, we may calculate the probability of occurrence of an event of interest, by decomposing it into a number of mutually exclusive events, where the probability of occurrence of each event is relatively simple to calculate.

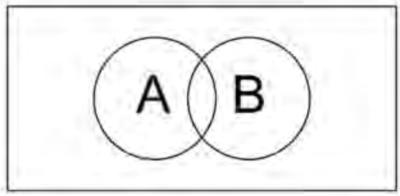
The above relationship (illustrated by the above Venn diagram) is only valid when the two events are mutually exclusive.[55] If it is possible for both events to occur, then the combined probability is not the sum of the individual probabilities. In effect, the probability of both events occurring has been counted

52 Good 2001:54.
53 Evett and Weir 1998:10.
54 Evett and Weir 1998:10.
55 Vogt 1993:148 “For example, subjects in a study cannot be both female and male, nor can they be both Protestant and Catholic, for those are mutually exclusive categories. They could, however, be both female and Protestant because those are not mutually exclusive.”

twice and so the straightforward sum overestimates the combined probability. In this case, the combined probability may be calculated as follows:

P(A or B) = P(A) + P(B) – P(A and B).

Events that are not mutually exclusive can be illustrated by the following Venn diagram:



Example 1:

Returning to our earlier legal example about the random match probability of 1 in 7 million, recall that the probability of blood from a single person matching that on the clothing of the suspect was 1 in 7 million.

To calculate the probability of a match if blood is taken from 2 randomly selected people (call them person A and person B) we could describe two events:

- Event A: Person A yields a match.
- Event B: Person B yields a match.

These events are not mutually exclusive, because there is a possibility of both people yielding a match. So, we would need to calculate the probability of at least one match using the formula

P(A or B) = P(A) + P(B) – P(A and B).

Calculating the probability P(A and B) will be dealt with in section 5.4.3 below.

Alternatively, we could define mutually exclusive events as follows:

i. No match is found.

ii. Event A1: Person A yields a match and person B does not.

iii. Event B1: Person B yields a match and person A does not.

iv. Event C: Both people yield a match.

Each of the options (ii) (iii) and (iv) involve at least one match. So, to calculate at least one match, we could calculate the probability of each of these occurrences and then add them to get the final probability (because they are mutually exclusive).

P(A1 or B1 or C) = P(A) + P(B) + P(C).

5.4.3 The product rule and independence issues

A further important rule for calculating probabilities deals with independent events. Two events are said to be statistically independent if the occurrence of either one of the events does not affect the probability of occurrence of the other event.

If two events are statistically independent, then the probability of the occurrence of both events may be calculated using the product rule. That is,

P(A and B) = P(A)P(B),

the probability of both occurring is the product of the individual probabilities.

In most realistic situations, the events of interest are made up as the co-occurrence of a number of simpler events. In such cases, the product rule is a very useful tool for calculating the probability of the composite event.

A simple example of the product rule can be given in the case of the lottery draw. Let us look at two selections in the draw and ask about the probability of balls '5' and '6' being drawn. First note that this event corresponds to two different draws, one with the '5' first and the other with the '6' first. These draws are mutually exclusive and so the probability of the event of interest is the sum of the probabilities of each of these draws. We can write this symbolically as

P(5,6) = P(5 then 6) + P(6 then 5)

To calculate the probability of the draw '6 then 5', we note that this is a composite event which is made up of an outcome of '6' on the first selection followed by an outcome of '5' on the second selection. We have already calculated the probability for the first selection. It is 1/49 — the probability of choosing 1 ball from 49 (each with an equal chance of being chosen). For the second selection, one ball (the '6') has been removed from the pool, and so the probability of selecting a '5' is now 1/48. To investigate independence, note that **no matter what ball was drawn first**, the second draw involves the selection of 1 ball from a pool of 48, where each ball has an equal chance of selection. That is, for any possible draw of two balls, the outcome of the first draw has no effect on the probability of the outcome of the second draw. So the two events are statistically independent and we can use the product rule, giving

P(6 then 5) = (1/49) x (1/48) = 1/(49x48).

Exactly the same argument could be used for the draw of '5 then 6' and so we get the final probability

P(5,6) = 2/(49x48).

Note that the factor of 2 in the numerator corresponds to the number of different ways the two balls could be ordered. Using this insight, it is a simple matter to calculate the probability of any winning 6 ball selection. It is

(6x5x4x3x2) / (49x48x47x46x45x44) $\approx$ 0.000 000 072.

Example 1:

Returning again to our legal example where the probability of blood from a single person matching that on the clothing of the suspect was 1 in 7 million (or approximately 0.00000014).

Using the above technical results, we may calculate the probability of at least one match being found in a group of more than 1 person. First, as described in section 5.4.2, calculate the probability of a match if blood is taken from 2 randomly selected people.

To calculate the probability of both people yielding a match, we first note that the two people were selected randomly. Because of this, the probabilities of each person yielding a match are independent. Thus, we may use the product rule to calculate the combined probability:

$$
\begin{aligned}
P(A \text{ and } B) &= P(A)P(B) \\
&= (1/7000000)(1/7000000) \\
&\approx (0.00000014)^2 \\
&\approx 0.000\,000\,000\,000\,02
\end{aligned}
$$

So, the probability of at least one match is

$$
\begin{aligned}
P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B). \\
&= 2(1/7000000) - (1/7000000)^2 \\
&\approx (0.00000028) - 0.000\,000\,000\,000\,02 \\
&\approx (0.00000028)
\end{aligned}
$$

and the desired probability is approximately 28 in 100 million or (again approximating) 1 in 4 million.

Note that this is very close to the probability of 2 in 7 million, which would be obtained by merely doubling the probability of a match from one person. But doubling the match probability for one person is not quite correct for determining the probability of at least one match from two people. Examining the separation into four different cases yields some insight into why this is the case. Viewed informally, simply adding the probabilities for each single match does not take into account the possibility that both people may yield a match. We would expect that such an overlap would result in the final probability being slightly less than the simple sum.

From this, we would expect that the probability of at least one match in a larger group (say it consisted of n people) would be less than n times the individual match probability. This can be verified by calculating the probability that there is at least one match in a population of seven hundred thousand people. Carried out as before, this is quite tedious, because it will involve calculating the probability for every possible result. We can simplify our work, by examining the complementary case (option i) and using the fact that the probabilities must sum to 1. That is:

P(at least one match) = 1 – P(no match).

To calculate the probability of no match being found, we may use the product rule because, as discussed before, the probabilities of each person being a match are independent.

$$P(\text{no match}) = (0.999999857)^{700000}$$

$$\approx 0.999999714$$

$$\approx 0.905$$

And so, finally

$$P(\text{at least one match}) = 1 - (0.999999857)^{700000}$$

$$\approx 1 - 0.905 \text{ }^{a}\text{ } 0.095 \text{ (or 9.5\%)}$$

This is a probability of 95 in 1000, which, as expected, is less than the 100 in 1000 obtained from a simple addition.

Examining the above figures brings out the second issue of importance when attempting to relate the single match probability to the chances of obtaining a match in a larger population. For we see that, in this example, there is a chance of over 9% of obtaining a match in a population of 700000.

Just like the original match probability, this does not provide a definite conclusion as to whether a match will be found in the population. Instead, it is a quantity which describes uncertainty inherent in finding such a match. As with interpreting all probabilities, to reach a definite conclusion, it is necessary to make a conceptual leap to a conclusion, which cannot be proved with certainty. Most experts who take a frequentist view of statistics would draw a definite conclusion if they judge that the probability of being in error is insignificant. The common levels used to judge a probability as insignificant are at most 5%, and so these experts would not regard the chances of a match in this population as insignificant. Alternatively, experts who follow a Bayesian approach would defer the decision as long as possible, by incorporating this probability into a prior probability. More detail on both these procedures will be provided below, and in the section on hypothesis testing.

Example 2:

Returning to the *Sally Clark*[56] case, the invalid calculation of the probability of repeated events was one of the issues of contention brought up in the first appeal. More specifically, a probability of 1 in 8543 was calculated as the probability of a single cot death in a family such as the Clark's. Due to the problems with sampling mentioned in the introduction to this article, it is questionable whether this probability is correct. But even if the probability were correct, the calculation of the probability of two cot deaths in one family of this type is based on the assumption that the probability of a second cot death is independent of the

56 *R v Sally Clark supra* note 20.

occurrence of the first cot death in the family.[57] Under this assumption, the probability of two cot deaths in the family would be:

$$(1/8543)\ (1/8543) = (1/72982849) \approx (1/73000000).$$

5.4.4 Conditional probabilities and Bayes' rule

When two events are not independent, the probability of occurrence of one event will depend on whether or not the other has occurred. The conditional probability describes the probability of the one event if the other has occurred. To be more precise, let the two events be labelled A and B. Then, the conditional probability of B, given A is the probability of B occurring on condition that A has occurred. This is denoted by P(B | A). Note that the two events are statistically independent if and only if

$$P(B) = P(B \mid A) \qquad \text{and} \qquad P(A \mid B) = P(A).$$

Conditional probabilities allow the product rule to be generalised to non-independent events. In this case, it becomes

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

In a game of *Monopoly* if we ask `what is the probability that the next throw will be a 5?' Our answer will be conditioned on the following: the dice has 6 sides numbered 1 to 6; the dice has been loaded and the tossing is designed to maximize uncertainty. Only then can everyone agree on the answer 1/16.[58]

It is an interesting fact that the two different conditional probabilities P(B | A) and P(B | A) are related through the generalised product rule. For

$$P(B)P(A \mid B) = P(A \text{ and } B) = P(A)P(B \mid A)$$

and so

$$P(A \mid B) = P(B \mid A)\ P(A)\ /\ P(B).$$

This relationship is known as Bayes' rule or Bayes' theorem.

The Bayesian interpretation is based on this relationship. According to this interpretation, the subjective probability a person ascribes to an event A will depend on the events described by the observations, B, carried out by the person. The probability that the person ascribes to the event A as a result of these observations, will then correspond to the conditional probability P(A | B). It is called the **posterior probability** because it takes account of the event described by the observations. However, this probability does not only depend

57 However, recent research into Sudden Infant Death Syndrome shows that it is likely that genetically inherited conditions contribute to cot deaths. This would imply that repeated cot deaths are not independent events, as the first death could indicate the existence of such a genetic factor. Even though the existence of genetic factors is not a fully accepted scientific conclusion, this research definitely shows that the assumption of independence is also not scientifically accepted, which invalidates the above calculation.

58 Evett and Weir *supra* note 53: 7.

on the observation events, it also depends on the **prior probability**, P(A), which is the subjective probability the person ascribes to the event before the observations. The observation events modify the prior probability through the probability of the observation event P(B), and the **likelihood function** P(B | A). Note that the likelihood function is the likelihood of the observations, given the event A.

Bayes' theorem is a formula for calculating conditional probabilities and the weight of evidence is usually expressed in terms of likelihood ratios.[59] To illustrate the application of Bayesianism in the context of expert evidence the following example can be used: blood that has been found at the scene of a murder, is analyzed and found to contain characteristics shared by one in a thousand members of the population. The blood of the accused is also found to contain these characteristics. What should the relevance be to the fact-finder of the 1:1000 frequency? The Bayesian approach uses a likelihood ratio to assess the probability of finding the blood of the accused at the scene of the murder in the light of two hypotheses. On a Bayesian analysis, an expert would provide the fact-finder with a likelihood ratio which represents a comparison between:

- the probability of the evidence[60] given that the accused committed the murder P(E | M)[61]
- the probability of the evidence given that the accused did not commit the murder P(E | nM).

The Bayes' model as a means of presenting forensic science and other scientific evidence to court has been investigated over the years.[62] The Bayes' approach which has been criticised by Tribe[63] and Kind[64] in a different context, is discussed in Section 7.

Evett and Weir, however, maintain that Bayes' theorem is "the best available model for understanding the interpretation of scientific evidence".[65] The Bayesian model lays down three ground rules for the forensic scientist:

i) To evaluate the uncertainty of any given proposition it is necessary to consider at least one alternative;

ii) Scientific interpretation is based on questions of the kind: what is the probability of the evidence given the proposition?

iii) Scientific interpretation is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated.[66]

---

59 Redmayne 1995:467 note 21.

60 Example — finding the blood.

61 Notation: P = probability; M = committed murder; nM = did not commit murder; E = expert evidence. The symbol | means "given," therefore P(M | E) can be read as "the probability that the accused is guilty given the evidence."

62 Finkelstein and Fairley 1970:83; Robertson and Vignaux 1995; Aitken 1995.

63 Tribe 1971:1329-1393.

64 Kind 1973:155-164.

65 Evett and Weir *supra* note 53:29.

66 Friedman 1996:1810-1838.

5.5 Hypothesis testing

As with all sciences, statistics is not able to assert a hypothesis about what happened in a situation with absolute certainty, for it is impossible to take into account all the different possibilities that might have occurred. To make statistical testing possible, it is necessary to formulate the hypothesis in statistical terms. This involves specifying a probability distribution to model the situation in such a way that the hypothesis to be tested corresponds to certain values of the parameters of the distribution. The frequentist interpretation allows the testing of this hypothesis against the alternative that the parameters do not have the specified values. Alternatively, Bayesian interpretation allows for comparison of two different hypotheses which do not necessarily cover all the alternative values of the parameters.

An hypothesis test will generally involve relating a number of hypotheses to a number of observed events. Great care needs to be taken when describing or interpreting the probabilities involved in such a test. The notation for conditional probabilities is often useful to make these descriptions precise, even though some statisticians may question the validity of this usage. To demonstrate this, it will be sufficient to examine the simplest case of a single hypothesis, denoted by H, and a single event, denoted by E. Two different probabilities are of interest. They are:

- $P(H \mid E)$ i.e. the probability that the hypothesis H is true, given that the event E has occurred.
- $P(E \mid H)$ i.e. the probability that the event E occurs, given that the hypothesis H is true.

5.5.1 Hypothesis testing — a Bayesian view

In one form, statistical hypothesis testing may be carried out in the same fashion as above, except that the results are interpreted in a Bayesian fashion. That is, the probabilities describe the degree of validity that the tester ascribes to the hypotheses. It is important to note that in this interpretation, these probabilities are not seen as objective measures of the truth of the hypothesis, rather they are a measure of the espoused belief of the tester as a result of the observations on which the test is based.

However, this process in not very common in Bayesian analysis because it does not take into account the prior probability — the belief of the tester before the test is carried out.

It is more common in Bayesian analysis to base the testing of the hypothesis on Bayes' rule. That is, if the hypothesis is denoted by H and the observed event by A, then the conditional probability of the hypothesis given the observed event is

$$P(H \mid A) = P(A \mid H)\, P(H) / P(A).$$

Here the likelihood function, or conditional probability $P(A \mid H)$, is the probability of observing the event A given the validity of the hypothesis H. The probability

P(H) is the prior probability of the hypothesis — the probability that is believed to hold before the observation. This is determined on the basis of the other facts known about the situation that the hypothesis describes. In legal cases, this may often be chosen in a fitting manner on the basis of the other facts of the case. The most difficult aspect of this manner of testing an hypothesis is determining the probability, P(A), of the observed event occurring in general.

A noteworthy simplification of the process occurs when two hypotheses are compared against each other. Let the second hypothesis be denoted by K. Then the probability of the second hypothesis, given the observed event A, is

$P(K \mid A) = P(A \mid K)\, P(K) / P(A).$

If we now examine the quotient of these two probabilities, we find that

$$\frac{P(H \mid A)}{P(K \mid A)} = \frac{P(A \mid H)\, P(H)}{P(A \mid K)\, P(K)}$$

Note that the general probability P(A), which appears in both expressions, is cancelled out in the quotient. Thus, comparing two hypotheses does away with the factor which is difficult to determine. Note also that if one of the two hypotheses H and K must be satisfied, then the posterior and prior ratios are the odds of the hypothesis H expressed as a fraction. In this expression, we find that the posterior odds are related to the prior odds, through multiplication by the factor.

This factor is termed the **likelihood ratio** because it is the ratio of the two likelihood functions.

When evaluating DNA evidence when a stain has been left at the scene, people using this approach wish to determine the posterior odds ratio:

$$\frac{\text{The probability that the suspect left the crime stain}}{\text{The probability that the suspect did not leave the crime stain.}}$$

To effectively formulate the second probability, it is necessary to identify the population of interest. It may then be specified as:

The probability that a random member of the identified population left the crime stain.

Using conditional probabilities, this may be expressed as

$$\frac{P(H_p \mid E, I)}{P(H_d \mid E, I)}$$

where Hp is the 'prosecution hypothesis' that the suspect did leave the crime stain, Hd is the 'defence hypothesis' that a random person in the identified population left the crime stain, E is the DNA evidence and I is the prior evidence.

Using Bayes' theorem as shown above, to relate this to the probabilities of the evidence, gives

$$\frac{P(H_p \mid E,I)}{P(H_d \mid E,I)} = \frac{P(E \mid H_p,I)P(H_p \mid I)}{P(E \mid H_d,I)P(H_d \mid I)}$$

When the identified population is a large group, the likelihood ratio

$$\frac{P(E \mid H_p, I)}{P(E \mid H_d, I)}$$

is often called the **random occurrence ratio**. The random occurrence ratio relates the prior odds to the posterior odds.

5.5.2 Hypothesis testing — a frequentist view

On the basis of their interpretive approach, frequentists do not take a probabilistic view of hypotheses. An hypothesis is not an uncertain event, but rather a proposition and as such it is either objectively true, or false. The uncertainty in an hypothesis arises from our uncertain knowledge of whether the hypothesis is valid or not. This is a subjective uncertainty, which cannot be interpreted in terms of frequencies and so will not fit with a frequentist interpretation of probability. Thus, according to the frequentist interpretation, statistical hypothesis testing involves using statistical evidence as grounds for making a choice as to whether to assert the hypothesis as true, or as false.

In the deterministic setting, any attempt to use an observation as evidence to provide certain proof of an hypothesis, is doomed to failure. This is the well known problem of induction, which arises due to the lack of symmetry in logic. No amount of supporting evidence will prove the hypothesis with logical certainty, while a single counter-example will disprove the hypothesis. In response to this, an indirect argument based on an attempt to disprove the hypothesis is often used for hypothesis testing. The frequentist approach to hypothesis testing is based on a similar, indirect argument.

This is the reason why the authors of this article emphasise that lawyers should be aware of whether the experts in a particular trial subscribe to a frequentist or Bayesian approach in the interpretation of probability.

6. Law: statistics, damn statistics and lies revisited

In view of the discussion above, this section briefly touches on ways in which some problems identified at the outset can be averted.

6.1 The relationship to the underlying science

David Stoney[67] contrasts the use of statistics with regard to DNA typing and forensic methods such as fingerprint pattern comparisons:

67 Stoney 2002:386.

> Efforts to assess the individuality of DNA blood typing make an excellent contrast. There has been intense debate over which statistical models are to be applied and how one should quantify increasingly rare events. To many, the absence of adequate statistical modelling, or the controversy regarding calculations, brings the admissibility of the evidence into question. Woe to fingerprint practice were such criteria applied!
>
> Much of the discussion of fingerprint practices in this and preceding sections may lead the critical reader to the question "Is there any scientific basis for an absolute identification?" It is important to realize that an absolute identification is an opinion, rather than a conclusion based on scientific research. The functionally equivalent scientific conclusion (as seen in some DNA evidence) would be based on calculations showing that the probability of two different patterns being indistinguishably alike is so small that it asymptotes with zero … The scientific conclusion, however, must be based on tested probability models. These simply do not exist for fingerprint pattern comparisons.

Despite these scientific findings and legal debate,[68] in the United States of America and most common law jurisdictions, including South Africa, fingerprint identification has been legally and culturally considered the classic example of incontrovertible expert evidence.[69]

There have been a dozen or so statistical models proposed in connection with fingerprint identification.[70] These vary considerably in their complexity, but in general there is much speculation and little data using these models. Champod's recent work[71] is the exception, bringing forth the first realistic means to predict frequencies of occurrence of specific combinations on ridge minutiae. Scientifically, the next step would be to assess the accuracy of the predictions. Currently no such work is being done. Champod's work does support rejection of simple minutiae counts as a realistic summary of fingerprint individuality. This is because the specific portion of the finger that the print comes from, and the specific nature of each minutiae, have a highly significant effect on the observed frequencies of occurrence.

6.2 Identification and match probabilities

In this regard, as indicated above it is essential that lawyers and judges are award of the danger of inappropriate assumptions and erroneous statistical calculations or methodology so that were necessary lawyer should call upon their own experts. The time might be ripe for legal reform that considers the use of court-appointed experts or assessors.

---

68 *United States v Llera Plaza (to be referred to as Llera Plaza I)* 179 F Supp 2d 492, 504 (ED Pa Jan (2002) vacated Cr No. 98-362-10, 2002 WL 389163 (ED Pa Mar 13 2002) (to be referred to as *Llera Plaza II*).

69 *S v Nala* 1965 (4) SA 360 (A); *S v Malindi* 1983 (4) SA 99 (T); *S v Gumede* 1982 (4) SA 561 (T); *S v Blom* 1992 (1) SASV 649 (OK); *S v Nyathe* 1988 (2) SA 211 (O); *S v Legote* 2001 (1) SACR 179 (HHA), *S v Nzimande* 2003 (1) SACR 280 (O).

70 Galton 1892:100-113.

71 Champod and Margot 1996:39.

6.3 Fallacious calculations by expert witnesses

Returning to the 1 in 73 million figure at the original trial of *R v Clark*,[72] the expert used a figure of about 700,000 UK births per year to conclude that "… by chance that happening will occur every 100 years." This conclusion is fallacious, not only because of the invalidity of the 1 in 73 million figure, but also because the 1 in 73 million figure relates only to families having some characteristics matching the family of the accused. **This error seems not to have been recognized by the Appeal Court**, which cited it without critical comment.[73] While the Appeal Court expressed the view that the accused had not been given a fair trial and that the conviction was unsafe and had to be quashed, it nevertheless failed to apply its mind to the problems surrounding the statistics testified to by Sir Roy Meadow.

Subsequently, Meadow was removed from the role of medical practitioners.[74] The Attorney General is reported to have intervened in an appeal, after a ruling that Professor Sir Roy Meadow should not have been struck off the roll for giving mistaken evidence in the Sally Clark cot-death trial.[75]

One of the mistaken inferences in the Clark case was that the probability of two cot deaths in a family could be calculated as the product of the probability of two single cot deaths. However, it is not an accepted scientific fact that cot deaths within a family occur independently of each other. A cot death in a family may indicate a medical or genetic vulnerability to this condition in the family and thus substantially increase the probability of a second cot death occurring in the same family. In fact, more recent research into this issue[76] suggests that there is indeed a genetic factor which increases the probability of cot deaths occurring in a family, and this would result in the hypothesis of independence becoming accepted as scientifically false.

Leaving aside the matter of validity, figures such as the 1 in 73 million are very easily misinterpreted. The jury should have been instructed to weigh up two competing explanations for the babies' deaths: SIDS or murder. The fact that two deaths by SIDS are quite unlikely is, taken alone, of little value. Two deaths by murder may well be even more unlikely. What matters is the relative likelihood of the deaths under each explanation, not just how unlikely they are under one explanation.

72 *R v Sally Clark* 2003 WECA Crim 1020. See also http://www.sallyclark.org.uk/judgment03.html accessed on 2004/11/17. Referred by the Criminal Case Review Commission under Section 9 of the *Criminal Appeal Act* 1995.

73 *R v Sally Clark* 2003 WECA Crim 1020 para 114.

74 *The Independent* 20 May 2006.

75 *The Independent* 20 May 2006.

76 Hofmann *et al* 1997:1835-1846.

6.4 Statistics incorrectly used by prosecutor and expert witness

This problem was illustrated in the *People v Collins*[77] case set out in section 3.4 above. In this case, the prosecutor, without presenting any statistical evidence in support of the probabilities for the factors selected, proceeded to assume probability factors for the various characteristics that he deemed to be shared by the guilty couple and all other couples answering to these distinctive characteristics.

The multiplication rule[78] was also used incorrectly by the prosecution. There was no proof that the probabilities were independent: one factor included persons with beards while another included people with moustaches. These two probabilities are clearly interdependent issues.

Expanding on what he had suggested as an hypothesis, the prosecutor offered what the court described as the completely unfounded and improper testimonial attestation that in his opinion, the factors he assigned were conservative estimates and that, in reality, the chances that anyone other than the accused was at the scene approached one in a billion. On appeal, the appellants contended that the introduction of evidence pertaining to the mathematical theory of probability and the use of it by the prosecution during the trial constituted prejudicial error.

Phillip Good[79] a mathematical statistician states the following:

> From a statistician's point of view, the prosecutor's errors are twofold. First, estimates were used instead of facts. The prosecutor suggested the odds were one in four that a girl in San Pedro would be blond — but he provided absolutely no fact in support of this simple allegation — not even a survey of local hairdressers. Similarly, he "suggested" that only one in ten cars in the area was yellow, again without any supporting evidence.
>
> Second, the prosecutor's probability calculations were in error. … the product rule applies only if the events are independent. African-Americans with beards and African-Americans with moustaches represent overlapping categories, although they were treated as separate and independent by the prosecution's expert witness. Beards, yellow convertibles, and a girl friend of a different race all seem associated with a certain type of flamboyant personality rather than being independent traits. The appeals court, commenting on the failure of the witnesses to make a positive identification, wondered whether, "the guilty couple might have included a light-skinned Negress with bleached hair rather than a Caucasian blond; or the driver of the car might have been wearing a false beard as a disguise ..."

77 *People v Collins* 68 Cal 2d 319 438 P 2d 33, 66 Cal Rptr 497 (1968).
78 See Multiplication rule below.
79 Good 2001:66.

6.5 The defence's and prosecutor's/lawyer's fallacies

Ways to overcome these thinking problems were discussed in section 3.5 and 5.3.2.2.

7. The applicability of Bayes' theorem to judicial decision-making

The probabilistic nature of judicial proof has long been emphasized in legal scholarship.[80] The 'new evidence scholarship' has generated a large body of literature on 'trial by mathematics' and statistical evidence.[81]

In "Decision-makers dilemma: Evaluating expert evidence,"[82] Meintjes-van der Walt poses the question whether Bayes' Theorem might be used as a tool by decision-makers to evaluate all the evidence before court.

> This idea was developed by Robertson and Vignaux,[83] who suggest that fact-finders can find guidance from the mathematical formula known as Bayes' theorem in the process of combining different probabilities. Bayes' theorem is used by them as a tool for hypothesis comparison: `[it] is logically meaningless to suggest that any piece of evidence has any value in itself as support for any particular hypothesis. Its value depends upon its ability to discriminate between one hypothesis and another'.[84] Bayes' theorem is a formula for calculating conditional probabilities and the weight of evidence is usually expressed in terms of likelihood ratios.[85] (Footnotes refer to original.)

This is vividly illustrated by the facts of *R v G Adams*[86] and in the English Appeal Court's approach to the argument in the case. The facts are briefly the following: In 1991 Ms X was walking to her home in Hemel Hempstead, England. A male stranger asked her the time. She only saw his face for a few seconds, before looking at her watch. The next moment he attacked her from behind and then raped her. Her description to the police was that her attacker was Caucasian, clean-shaven, with a local accent and between 20 and 25 years old. It was only in 1994 that Adams was apprehended in connection with another incident and his DNA profile was found to match that from the semen recovered from the rape in 1991. In the subsequent identity parade, Ms X, failed to identify him. Adams was 37 at the time, but Ms X described him as around 40 to 42 years old. At the subsequent committal proceedings, she said that Mr Adams did not look like the man who attacked her. Adams also had an alibi.

80 Tillers 1993:1466.
81 Tillers 1993:1466.
82 Meintjes-van der Walt 2000:339-341.
83 Robertson and Vignaux 1995:41.
84 Robertson and Vignaux 1995:220.
85 Redmayne 1995:467 note 54.
86 *R v G Adams* [1996] 2 Cr App 467.

The case for the prosecution had been that DNA samples from Adams and the crime scene sample had been compared and a visual match within one percent could be declared. Computer calculations showed that the chance of a randomly chosen unrelated man matching the DNA profile was 1 in 297 million. This was brought down in the interests of conservatism to 200 000 000, the calculation based on nine bands of DNA identified by the profile.[87]

The defence, in seeking to logically present the non-scientific and the DNA evidence, called a statistician, Professor Donnelly, to explain how this could be done. The statistical expert explained to the jurors how they could use Bayes' Theorem to combine each of the other items of evidence in the case with each other. Adams was found guilty, but an appeal was granted.

Lord Justice Rose of the Court of Appeal found that the difficulties in the case were of the defence's making because it introduced Bayes' theorem.[88] The Court of Appeal made a number of observations enumerated below in italics:

i) *the theorem can only operate by giving to each separate piece of evidence a numerical percentage representing the ratio between the probability of circumstance A and the probability of circumstance B granted the existence of that evidence.*[89]

In the last part of the sentence, the court transposed the conditional, the so-called lawyer's fallacy or the prosecutor's fallacy.[90] A likelihood ratio is the ratio between the probability *of the evidence* given circumstance A and circumstance B. It describes the strength of the evidence in distinguishing between proposition A and proposition B. What the Court of Appeal described here was the posterior odds which the Court was trying to assess.

ii) *The percentages chosen are matters of judgment: that is inevitable. But the apparently objective numerical figures used in the theorem may conceal the element of judgement on which it entirely depends.*[91]

Robertson and Vignaux[92] comment that:

> One would hope that these judgements would be made in the same way that all assessments of probability should be made, rationally and by reference to the evidence. The figures given are merely expressions of strength of belief. Any system for expressing strength of belief must comply with some simple rules such as:
>
> if I believe that A is more likely to be true than B and that B is more likely to be true than C, I must believe that A is more likely to be true than B and C, and equivalent levels of belief are expressed equivalently and divergent levels of belief expressed divergently. Not only are numbers

87 Freckelton and Selby 2002:533.
88 *R v G Adams supra* note 86 at 481.
89 *R v G Adams supra* note 86 at 481.
90 See section 3.5.2 above.
91 *R v G Adams supra* note 70 at 481.
92 Robertson and Vignaux *supra* note 67.

> a convenient way of achieving this, but it can be shown that any system of expressing strength of belief which complies with these rules can be reduced to numbers.

(iii) the theorem's methodology requires, as we have described, that items of evidence be assessed separately according to their bearing on the accused's guilt, before being combined in the overall formul'.[93]

For purposes of legal decision-making in non-jury trials, this view is only partially correct. In the assessment of evidence process, items of evidence must be examined separately and in the context of the evidence already considered:

> Doubt about one aspect of the evidence led in a trial may arise when that aspect is viewed in isolation. Those doubts may be set at rest when it is evaluated again together with all the other available evidence. There is no substitute for a detailed and critical examination of each and every component in a body of evidence. But once that has been done, it is necessary to step back a pace and consider the mosaic as a whole.[94]

(iv) That [Bayes' approach] in our view is far too rigid an approach to evidence of the type that a jury characteristically has to assess, where the cogency of (for instance) identification evidence may have to be assessed, at least in part, in the light of the strength of the chain of evidence of which it forms part.'[95]

The fact that a Bayesian approach may be too difficult for a jury, may be an argument that can be considered in England and Wales, but Huygen[96] explains that Bayesian belief networks can be a suitable tool for lawyers to analyse evidence in the Dutch legal system where judges are the ultimate decision-makers.

After the Court of Appeal quashed Adams' conviction, he was retried. Despite the Court of Appeal's disapproval of the use of Bayes' Theorem during the original trial, the defence expert presented the jury with the Bayes' Theorem and how to use it to combine items of evidence in the case. Both the judge and prosecution agreed that this be allowed. The jury was provided with a questionnaire which asked them to make a number of probability assessments and they were also provided with a formula for the combination of probabilities. Adams was convicted.[97]

In 1971 Professor Laurence Tribe,[98] argued in an article that the usefulness of mathematical methods in the trial process is greatly exaggerated. Contrary to Tribe, Saks and Kidd[99] contend that while certain errors and harm may be inherent even in the proper use of probabilistic tools, even more harm may

93 *R v G Adams supra* note 70 at 481.
94 *S v Mbuli* 2003 (1) SACR 97 (SCA) 99h-i.
95 *R v G Adams supra* note 70 at 481.
96 Huygen Use of Bayesian Belief Networks in Legal Reasoning. http://www.bileta.ac.uk/02papers/huygen.html (accessed on 1 April 2005).
97 Freckelton and Selby note 71 *supra*.
98 Tribe 1971:1329.
99 Saks and Kidd (1980-1981):123.

be inherent in not using them. In making this point, Saks and Kidd first apply the research findings of behavioral decision theorists to challenge Tribe's assumptions from an empirical point of view. They conclude that explicit calculation of probabilities will, in most cases, lead a trier of fact closer to the correct conclusion than will reliance on intuitive, commonsense judgments. This is because most lay decision-makers employ a number of simplifying strategies, known as "heuristics" to reduce comples (*sic*) information to a point where they can make a decision.[100]

Section 5.3.1.1 above refers to the "gambler's fallacy" and the Saks and Kidd discussion around the proneness of humans, even that of judicial decision-makers, to commit these errors.

Saks and Kidd also used the following problem to analyse the phenomenon of heuristics: A man was selected at random from a group consisting of 70 lawyers and 30 engineers. The man selected was called John. He is 39 years of age, married and has two children. He is active in local politics and enjoys rare book collecting. He is competitive, argumentative and articulate.

A large group of respondents was requested to estimate the probably that John is a lawyer rather than an engineer. This group median probability estimate that John was an engineer was .95. Another group of respondents was asked the same question, but they were first told that the group from which John was selected was composed of 30 lawyers and 70 engineers. This group's median probability estimate that John was a lawyer was also .95. Information about the composition of the second group from which John was selected should logically have affected the probability estimate, but in fact had no effect at all on the judgment of the decision-makers. Decision-makers only become sensitive to the information about group composition when extremes in the compositions are reached such as where the group approaches 100 lawyers and 0 engineers (or the converse).

This example illustrates that human decision-making is inclined to be insensitive to base rates when case-specific information is available. If only the group base rates of 30 lawyers:70 engineers are given, people will rely heavily on this information when making their judgments. They are correct to say that the probability is .30 that the person selected is a lawyer. When they are also given case-specific information the respondents tend to ignore the numerical base and rely instead on the degree to which the description of John corresponds with their stereotype of lawyers. Respondents base their estimate of the probability that John is a lawyer on the degree of agreement between his description and their stereotype of lawyers as argumentative, competitive and politically aware. Based on the data given in this example, it is 5.44 times as likely that John is a lawyer when the group from which he is selected consists of 70 lawyers and 30 engineers than when the opposite membership distribution applies.

100 Note: The use of mathematical and probabilistic proof in the trial process. http://-www.law.harvard.edu/publications/evidenceiii/articles/notes1.htm (accessed 31 July 2004).

These problems illustrate that judicial decision-making should no longer be a matter based on "experience and intuition rather than analysis"[101] and a more mathematical approach may not be as far-fetched as previously thought.

8. Statistics and courtroom skills

Faigman *et al*[102] discuss four ways in which statistical testimony can be improved. Only three of these methods are applicable to South Africa.

They suggest that the expert:

(i) maintains his/her professional autonomy and objectivity;

(ii) discloses other analyses (statisticians usually analyze data by using various statistical models and methods);

(iii) discloses data and analytical methods before trial (Pre-trial conferences are recommended to debate the issues and to narrow the trial aspects).

As in the case of other expert evidence, proper preparation by lawyers is essential. Imwinkelried,[103] for instance, indicates that if cross-examination has not been carefully planned and where the cross-examiner is not certain what he or she wants to achieve, the cross-examination may merely give the witness the opportunity to repeat his/her direct-examination and strengthen that party's case.

There are cases in which an expert's reliance on erroneous statistical data as explained above must be attacked under cross-examination.[104] Such instances include, but are not limited to, when the following takes place:

- prosecutor's fallacy;
- defense's fallacy;
- the statistical methodology used by the expert renders his/her testimony unreliable;
- an expert is not knowledgeable about statistics and demonstrates a lack of understanding of the statistical basis of the opinion; or
- the premise of the expert's opinion is data that, although analyzed with correct methodology, is nevertheless invalid.[105]

An expert who relies on data that is not statistically significant or, although purporting to be statistically significant, is invalid because of flaws in the study design should be challenged. The key to challenging the evidence is being able

101 Hoffman & Zeffertt 1988:525. It is encouraging to note that the successor to this text, *The South African law of evidence (formerly Hoffman and Zeffertt)* by Zeffert DT, Paizes AP and Skeen A St Q. does no longer make mention of this advice.

102 Faigman *et al* 2002:162-164.

103 Imwinkelried 1997:10.

104 Parker and Vittoria 1999:48-49.

105 Parker and Vittoria 1999:48-49.

to demonstrate that the statistical methodology used by the expert was inappropriate, or that fundamental flaws in the study design render the data invalid.

In some cases, experts rely on others to analyze their data statistically. These experts are susceptible to an effective cross-examination on the statistical errors in their data. If the error is such that it invalidates the data, the expert's inability to defend the data may cause judicial decision-makers to question his/her qualifications.

Finally, experts who mislead the court by confusing concepts of statistical significance (95 percent probability) and the burden of proof with no mathematical percentage attached to it must be attacked under cross-examination.

Darrell Huff[106] offers the following five simple, yet effective, questions for lawyers to ask themselves before accepting statistical data:

> "**Who says so?**" Look for bias, both conscious and unconscious. Is the proponent of the data biased or is there bias in the manner in which the data is presented? Was unfavourable data withheld? Does the witness possess the statistical knowledge to do the analysis?
>
> "**How does he know?**" Was there bias in the sample or in the way the data was collected? Was the sample large enough for the result to have any meaning? Is a claimed correlation large enough to be important?
>
> "**What is missing?**" Statistics, such as percentages, are generally meaningless without raw data. Claimed correlations between two variables should not be taken seriously if the standard error (SE) or standard deviation (SD) of the estimate has not been given. Was the best measure of the "average" chosen to explain the data?
>
> "**Did someone change the subject?**" Look to see if the raw data has been switched in the conclusion. For example, are reported changes simply due to redefining what is being reported (i.e. crime) rather than a true change? Surveys are often misinterpreted.
>
> "**Does it make sense?**" Is a statistic based on an unreasonable and/or unproven assumption?

An increased confidence in statistical knowledge and understanding statistical jargon should improve counsel's ability to find the weaknesses and errors in statistical data relied on by an opposing expert.[107]

Cross-examination can commence with foundational questions regarding the importance of statistics. This serves at least two purposes. First, it gives counsel an idea whether the expert appears to be uneasy about responding to statistical questions. The expert's responses will suggest whether more sophisticated questions might be productive. Conversely, if the expert is evasive and argumentative these factors suggest that further questioning may not be productive.

However, when the expert's statistical error is fundamental and critical, counsel should proceed with the statistical cross-examination.

106 Huff 1954:123-142.
107 Parker and Vittoria *supra* note 104:56-57.

Keep the statistical cross-examination short and simple.[108] One danger for lawyers who develop expertise in scientific disciplines is a tendency to demonstrate their knowledge by engaging in cross-examination that is understood, at best, only by an expert. While demonstrating one's proficiency in science is important to establish credibility with the court, and with the opposing expert, clear communication with the court remains of paramount importance.

Use examples to explain statistical concepts to the bench by using examples relevant to their lives. Rolls of the dice, and the Lotto, for example, can be incorporated into a cross-examination to assist the court. Teaching by analogy is effective as it allows counsel to make difficult subjects more understandable.

9. Conclusion

Statistics is a science and to become an expert in this field requires extensive training. Fully to understand and work with the technical results of this science would entail the development of a large number of technical skills. To elucidate these in a single article will not be possible. This exposition only deals with a few of the simpler technical skills. The most important task that a legal professional would need to be able to carry out in this context, would be to interpret, and evaluate the interpretation of statistical results in a legal setting. In keeping with this, the article focuses mostly on issues relevant to the interpretation of statistical concepts and results that are commonly used in criminal law.

As in the case of other fields of expertise, the ever increasing use of statistical evidence by the legal system requires lawyers and experts using statistical evidence to become competent consumers of statistics: "Now, more than ever, the onslaught of technology obligates the criminalist to draw on a strong background of physical sciences, including an understanding of statistics and logic."[109] The knowledgeable lawyer,[110] equipped with a basic understanding of statistics or with the help of his or her own expert statistical witness will be in a better position to bolster a case with appropriate statistics, expose failings in statistical models and methodology used by the opposing party, or to make expert testimony intelligible to the decision-maker.

108 Parker and Vittoria *supra* note 104:56-57.

109 Inman and Rudin 2000:302.

110 Hawkins and Hawkins 'Lawyers and likelihoods' http://www.swin.edu.au/maths/icots5/sess4.html (accessed on 30 November 2004) who indicate that lawyers are indeed susceptible to a number of statistical and probabilistic misconceptions.

Bibliography

AITKEN CGG
1995. *Statistics and the evaluation of evidence for forensic scientists*. New York: J. Wiley.

AITKEN CGG AND TARONI F
2004. *Statistics and the evaluation of evidence for forensic scientists*. 2nd edition. New York: John Wiley and Sons.

ANDERSON T AND TWINING W
1991. *Analysis of evidence. How to do things with facts*. London, Weidenfeld and Nicholson.

BALDING DJ AND DONNELY P
1994. The prosecutor's fallacy and DNA evidence. *Criminal Law Review*:711.

BUCKLETON J, TRIGGS CM AND WALSH SJ
2005. *Forensic DNA evidence interpretation*. Boca Raton: CRD Press.

BUECHE F
1969. *Introduction to physics for scientists and engineers*. New York: McGraw-Hill.

COHEN J
1977. *The probable and the provable*. Oxford: Clarendon Press.

CHAMPOD C
2000. Identification in individualisation: Overview and meaning of ID. In Siegel JA, Saukko P and Knupher G (eds) *Encyclopeadia of Forensic Science*. London: Academic Press: 1077-1085.

CHAMPOD C AND EVETT IW
2001. A probabilistic approach to fingerprinting evidence. *Journal of Forensic Identification* 51:101-122.

CHAMPOD C AND MARGOT P
1996. Computer assisted analysis of minutiae occurrences on fingerprints. In Almog J and Springer E (eds) *Proceedings of the International Symposium on Fingerprint Detection and Identification*. Jerusalem: Hemed Press: 305-318.

CURRAN JM, HICKS TN AND BUCKELTON JS
2000. *Forensic interpretation of glass evidence*. Boca Raton, CRC Press.

EPSTEIN R
2002. Fingerprints meet *Daubert*: The myth of fingerprint "science" is revealed. *California Law Review* 75: 605-657.

EVETT IW AND WEIR BS
1998. *Interpreting DNA evidence statistical genetics for forensic scientists*. Sunderland: Sinauer Associates.

FAIGMAN DL, KAYE DH, SAKS MJ AND SANDERS J
2002. *Modern scientific evidence: The law and science of expert testimony*. Vol 3. St Paul, Minnesota: West Group.

FINKELSTEIN MO AND FAIRLEY WB
1970. A Bayesian approach to identification evidence. *Harvard Law Review*: 489-517.

FINKELSTEIN MO AND LEVIN B
2001. *Statistics for lawyers*. 2nd edition. New York: Springer.

FOX JA AND LEVIN J
2005. *Elementary statistics in criminal justice research: The essentials*. Boston: Allyn & Bacon/Longman.

FRECKELTON I AND SELBY H
2002. *Expert evidence: Law, practice, procedure and advocacy*. Pyrmont: Lawbook Co.

FRIEDMAN RD
1996. Assessing evidence. *Michigan Law Review* 94:1810-1838.

GALTON F
1892. *Finger prints*. London: Macmillan.

GARFIELD J
2002. The challenge of developing statistical reasoning. *Journal of Statistical Education* 10(3):1-51.

GOOD PI
2001. *Applying statistics in the courtroom. A new approach for attorneys and expert witnesses*. Baco Raton: Chapman & Hall/CRC.

GOODWIN J AND MEINTJES-VAN DER WALT L
1997. The use of DNA evidence in South Africa: A powerful tool or prone to pitfalls. *The South African Law Journal*: 114-141.

HAWKINS J AND HAWKINS H
2004. Lawyers and likelihoods. http://www.swin.edu.au/maths/icots5/sess4.html (accessed on 30 November 2004)

HENDERSON JP
2002. The use of DNA statistics in criminal trials. *Forensic Science International* 128:183.

HOFFMAN LH AND ZEFFERTT DT
1988. *The South African law of evidence*. 4th edition. Durban: Butterworths.

HOFMANN S, JAKSCH M, BEZOLD R, MERTENS S, AHOLT S, PAPROTTA A AND GERBITZ KD
1997. Population genetics and diseases susceptibility characterization of central European genetics on disease susceptibility; characterization of central European haplo groups by M+ DNA gene mutations, correlation with D loop variants and association with disease. *Human Molecular Genetics* 6(11):1835-1846.

HOLMES OW
1897. The path of law. *Harvard Law Review* 10:457-530.

HUFF D
1954. *How to lie with statistics*. New York: WW Norton & Company Incorporated.

HUYGEN PEM
Use of Bayesian belief networks in legal reasoning. http://www.bileta.ac.uk/02papers/huygen.html (accessed on 1 April 2005).

IMWINKELRIED EJ
1997. T*he methods of attacking scientific evidence*. 4th edition. Charlottesville: Lexis Law Publishing.

INMAN K AND RUDIN N
2000. *Principles and practice of criminalistics: The profession of forensic science*. Boca Raton: CRC Press.

KAYE DH AND FREEDMAN DA
2000. *Reference guide on statistics*. Washington DC: Federal Judicial Centre.

KAYE DH, BERNSTEIN DE, MNOOKIN JL AND FRIEDMAN RD
2004. *The new wigmore: A treatise on evidence expert-evidence*. New York: Aspen.

KIND SS
1973. Crime investigation and the criminal trial: A three chapter paradigm of evidence. *Journal of Forensic Science Society* 155-156.

LEMPERT R
1986. The new evidence scholarship: Analyzing the process of proof. *Boston University Law Review* 66:439-477.

LINDLEY V
1977. A problem in forensic science' *Biometrika* 64:207.

MEARS M AND DAY TM
2003. The challenge of fingerprint comparisons in the defence of a criminally charged client. *Georgia State University Law Review* 19:705-761.

MEINTJES-VAN DER WALT L
2000. Decision-makers' dilemma: Evaluating expert evidence. *South African Journal of Criminal Justice* 13:319-434.

2001. *Expert evidence in the criminal justice process: A comparative perspective*. Amsterdam: Rozenberg Publishers.

2005. Shattering the silence: How glass evidence speaks. *South African Journal of Criminal Justice* 18(1):56-65.

PARKER BR AND VITTORIA AF
1999. Debunking junk science: Techniques for effective use of biostatics. *Defence Counsel Journal* 48:33-58.

REDMAYNE M
1995. Doubts and burdens: DNA evidence, probability and the courts. *Criminal Law Review*:464-482.

ROBERTSON B AND VIGNAUX GA
1995. *Interpreting evidence: Evaluating forensic science in the courtroom*. Chichester: Wiley.

SAKS M AND KIDD R
1980-1981. Human information processing and adjudication: Trial by heuristics. *Law & Society Review* 15:123-160.

SCHUTTE JW
2001. Using the MCMI III in forensic evaluations. *American Journal of Forensic Psychology* 19:5-20.

STONEY M
2002. Fingerprint identifications. In Faigman DL, Kaye DH, Saks MJ and Sanders J (eds) *Modern scientific evidence: The law and science of expert testimony*. Vol 3. St. Paul, Minnesota: West Group:368-399.

THE INDEPENDENT
2006. The Attorney General steps into the Meadow Case. http://news.independent.co.uk./uk/legal/article548938.ece (accessed on 20 May 2006).

The use of mathematical and probabilistic proof in the trial process. http://www.law.harvard.edu/publications/evidenceiii/articles/notes1.htm (accessed on 31 July 2004).

THORNTON JI AND PETERSON JL
2002. The general assumptions and rationale of forensic identification. In Faigman DL, Kaye DH, Saks MJ and Sanders J (eds). *Science in the law: Forensic science issues*. St. Paul, Minnesota: West Group, paras 1-54 15.

TILLERS P
1993. Intellectual history, probability, and the law of evidence. *Michigan Law Review* 91(6):1465-1490.

TRIBE LH
1971. Trial by mathematics: Precision and ritual on the legal process. *Harvard Law Review* 84:1329.

TWINING W
1980. Debating probabilities. *The Liverpool Law Review* 2:51-64.

VOGT WP
1993. *Dictionary of statistics and methodology. A nontechnical guide for the social sciences*. London: Sage Publications.

ZEFFERT DT, PAIZES A AND SKEEN A ST Q
2003. *The South African Law of evidence* (formerly Hoffman and Zef fert) Durban: LexisNexis/Butterworth.