



## STATISTICS IN FORENSIC SCIENCE. PART II. AN AID TO THE EVALUATION OF EVIDENCE

Colin G. G. AITKEN

*School of Mathematics, The University of Edinburgh, Edinburgh, United Kingdom*

### Abstract

This is the second of two papers describing roles for statistics in forensic science. Two procedures for the evaluation of evidence are compared. It is shown that a procedure based on significance testing has several unsatisfactory characteristics which can be overcome using a procedure which compares the probability of evidence under each of two propositions. Some comments on interpretation and on possible future directions for work in this area are given.

### Key words

Evidence evaluation; Significance testing; Likelihood ratio; Interpretation of evidence.

*Received 16 April 2006; accepted 15 May 2006*

### 1. Introduction

There are two main roles for statistics in forensic science. The first role is taken during the investigatory stage of a crime before a suspect has been identified and statistics can be used to assist in the investigation. The second role is taken during the trial stage. There is a defendant and statistics can be used to assist in the evaluation of the evidence. These two roles are discussed in two separate papers in this volume. This is the second of the two papers and discusses the role of statistics as an aid to the evaluation of evidence. The first paper [1] discussed the role of statistics as an aid to investigation.

### 2. Significance tests

A so-called classical approach is one which attempts to mimic the approach of a forensic scientist in the comparison of evidence from the crime scene and from the environment of a suspect. The approach has

been called “classical”, yet it is based on significance testing, an idea which was developed in the first half of the twentieth century. A later method advocated by many statisticians interested in forensic science is based on consideration of Bayes’ theorem. This theorem, however, was developed in the eighteenth century, well ahead of the “classical” approach. Thus, the classical approach is based on ideas over a hundred years younger than the approach based on that of Bayes. The phrase “environment of a suspect” is used to summarise the idea that evidence may be found from the suspect themselves (e.g., a DNA profile), from their clothes (e.g., glass fragments) or footwear (e.g., trademarks on shoes) or their domestic surroundings (e.g. paint fragments on their car).

The approach is a two-stage one. It mimics the approach of the forensic scientist in that there is a comparison stage and then a rarity stage. In the first stage, evidence from the crime is compared with evidence from the environment of the suspect. There is a binary outcome to this comparison. The two sets of evidence may be deemed similar or dissimilar. If they are

deemed similar, the second stage of the analysis is conducted. If they are deemed dissimilar then the analysis is stopped and it is decided to act as if the two sets of evidence came from different sources. For those sets of evidence deemed similar, the second stage is the assessment of the rarity of the evidence. Similarity of two sets of evidence which are rare in the characteristics of similarity will be deemed stronger evidence of a common source for the two sets than a similarity of two sets of evidence which are common in the characteristics of similarity. For example, consider eyewitness evidence of the number of arms possessed by the criminal. Consider, first, that the eyewitness says the criminal had two arms and a suspect is later identified and discovered to have two arms. The number of arms of the suspect is similar to (actually identical to) the number of arms of the criminal. However, this outcome is a match of a very common characteristic, the possession of two arms, and the eyewitness evidence concerning the number of arms is not valuable. Now, consider, second that the eyewitness says the criminal had only one arm and a suspect is later identified and discovered to have only one arm. The number of arms of the suspect is similar to (actually identical to) the number of arms of the criminal. This outcome is a match of a very rare characteristic, the possession of one, and only one, arm, and the eyewitness evidence concerning the number of arms is valuable.

Consider a crime in which a window is broken and glass fragments from the scene are compared with fragments found on the clothing of a suspect. Continuous measurements are made of various characteristics, refractive index and chemical compositions. The data are multivariate and a multivariate significance test may be developed. The data are determined to be dissimilar if the difference is determined to be more than two standard errors, corresponding to a 5% significance level.

There are several problems with this approach. First, there is an effect which has been likened to falling off a cliff. At one point, one is safely positioned on the edge of a cliff yet a very small step will take you over the edge and a fall to the bottom of the cliff. A very small step has a very large effect. Similarly, for a 5% significance test, the difference between a comparison which is significant at the 4.9% level and a comparison which is significant at the 5.1% level has a very large effect on the suspect. In the first case, the evidence is taken to the second stage and is still considered evidence against the suspect. In the second case, the evidence is discounted and is not considered evidence against the suspect. An associated problem is that there is a subjective choice in the cut-off point (or edge of the cliff) for the significance test.

The second problem is that it is not easy to combine the effect of other factors, such as transfer and persistence, with evidence evaluated by use of a significance test. Finally, the third problem is that the burden of proof is placed on the defence rather than the prosecution. The significance test is based on an assumption that the evidence from the crime scene and from the environment of the suspect are from the same source, unless there is a difference between characteristics of the evidence which is significant at some pre-specified level, say 5%. It is the responsibility of the defence to show that the difference is significant at this level. Thus, there is an assumption that the suspect is associated with the crime unless demonstrated otherwise.

There is one possible apparent advantage with this approach. It is possible to compare the two sets of evidence using only the data in those two sets. For example, if the evidence is in the form of measurements of the refractive indices of glass fragments, then these may be compared using a two-sample *t*-test. However, such a comparison ignores information of the variability of refractive indices which may be available from other windows or glass objects.

### 3. Evidence evaluation – the comparison of two propositions

It is possible to develop an approach which resolves these problems. The approach treats the position of the prosecution and the position of the defence as equal. There is not a presumption in favour of the proposition's case as in the approach based on significance testing described above. The approach compares the likelihood of the evidence if the prosecution's case is true with the likelihood of the evidence if the defence case is true.

#### 3.1. Propositions

Evidence is evaluated by the comparison of two probabilities, the probability of the evidence if the prosecution proposition is true and the probability of the evidence if the defence proposition is true.

Propositions need not only be those of guilt and innocence. Table I lists five pairs of propositions with levels [1] noted for the first three pairs.

The court is interested in the truth or otherwise of these propositions. The scientist is interested in the probability of the evidence in the cases when the prosecution proposition is true and when the defence proposition is true. Evidence is evaluated by its effect on

the odds in favour of the prosecution's proposition (or odds against the defence proposition).

TABLE I. PAIRS OF PROPOSITIONS FOR COMPARISON OF EVIDENCE

	Prosecution ( $H_p$ )	Defence ( $H_d$ )
a	Guilt	Innocence
b	Suspect at crime scene	Suspect not at crime scene
c	Suspect's DNA at crime scene	Suspect's DNA not at crime scene
d	Alleged father is true father	Alleged father is not true father
e	Identified person on database is brother of criminal	Identified person on database is unrelated to criminal

a crime level, b activity level, c source level, d paternity example, e familial DNA as described in Aitken [1].

It is helpful to introduce some notation. First  $Pr$  indicates probability,  $H_p$  and  $H_d$  denote the prosecution and defence propositions,  $E$  denotes the evidence to be evaluated and  $I$  denotes all other information. The vertical bar  $|$  denotes conditioning of events. The event to the left of the bar is the event whose probability is of interest. The event or events to the right are those whose outcomes are assumed known. Then the odds form of Bayes' theorem

$$\frac{Pr(H_p|E, I)}{Pr(H_d|E, I)} = \frac{Pr(E|H_p, I)}{Pr(E|H_d, I)} \frac{Pr(H_p|I)}{Pr(H_d|I)} \quad \{1\}$$

gives the relationship amongst the propositions, the evidence and all other information. As an example of the interpretation of the vertical bar consider the numerator of the ratio on the left-hand-side. The event to the left of the bar is  $H_p$  and that is the event whose probability is of interest. The events to the right,  $E$  and  $I$  are those whose outcomes are assumed known. The numerator is short-hand for the probability of the prosecution's proposition being true, given the evidence under consideration and all other information.

The second ratio on the right-hand-side of  $\{1\}$   $\frac{Pr(H_p|I)}{Pr(H_d|I)}$ , is the prior odds in favour of  $H_p$ . This is prior to consideration of the evidence but assuming all other information known. The ratio on the left-hand-side of  $\{1\}$ ,  $\frac{Pr(H_p|E, I)}{Pr(H_d|E, I)}$ , is the posterior odds in fa-

vour of  $H_p$ . This is posterior to consideration of the evidence.

The factor which converts the prior odds to posterior odds is the first ratio on the right-hand-side and it is known as the likelihood ratio. It measures the relative likelihoods of the evidence under each of the two propositions. Consider a likelihood ratio which has a value  $V$ . The evidence may be said to be  $V$  times more likely if the prosecution proposition is true than if the defence proposition is true.

### 3.2. Likelihood ratio

The likelihood ratio,  $\frac{Pr(E|H_p, I)}{Pr(E|H_d, I)}$ , may be thought

of as the value of the evidence. It provides a measure of the change in the odds in favour of the prosecution's proposition arising from consideration of the evidence. As an example of how it may work in practice, consider a case in which the prior odds of guilt say are 1 in 1000 against. Perhaps this is thought to be a measure of what it means to be "innocent until proven guilty". It may also be that the person is not to be found guilty unless the posterior odds are 1000 to 1 in favour. The likelihood ratio is the ratio of the posterior odds to the prior odds which in this example is 1 million. Thus to convert prior odds of 1 in 1000 to 1000 to 1 the evidence has to be 1 million times more likely if the prosecution proposition is true than if the defence proposition is true.

Several properties of the relationship  $\{1\}$  are worth noting. A likelihood ratio whose value is greater (less) than one implies the evidence supports the prosecution (defence) proposition. It takes values between 0 and infinity; it is not a probability. The posterior odds for one piece of evidence becomes the prior odds for the next piece of evidence, but care has to be taken that the previous evidence is part of the conditioning for the next piece of evidence. The likelihood ratio is not a statement concerning the probability of either of the propositions. The approach provides a logical approach for the consideration of evidence in that it emphasises that it is not enough to consider only the probability of the evidence if  $H_d$  is true. Also, probabilities may be subjective or personal or objective based on data. As a result of this, it is possible to take into account ideas about the probabilities of transfer, persistence and relevance.

A further illustration of the use of the likelihood ratio is given using an example of so-called transfer evidence. This is evidence which is transferred from the crime scene to the criminal or vice-versa. An example

of the former transfer would be glass fragments from a window broken at the crime scene and transferred to clothing of the criminal. An example of the latter would be a blood stain from a cut sustained by the criminal in breaking the window. That part of the evidence whose source is known is called control evidence. This would be the glass fragments at the crime scene or a blood sample from a suspect (who may or may not be the criminal). That part of the evidence whose source is unknown is called recovered evidence. This would be the glass fragments found on a suspect (who may or may not be the criminal) or the blood stain from the crime scene, which is assumed to come from the criminal but the criminal has yet to be identified.

The likelihood ratio combines assessments of similarity and rarity in one statistic, unlike the two-stage process involved with significance testing. A simplistic illustration of how this may work in practice is given in Aitken and Taroni [2] using a formula developed by Evett [6].

A fragment of glass (the recovered fragment) is found on the clothing of a suspect for a crime in which a window has been broken. Its refractive index is  $y$ . A sample of  $m$  fragments are taken from the crime scene window and the mean of the measurements of the refractive index of these fragments is  $x$ . The propositions to be compared are at the source activity:

- $H_p$ : the recovered fragment is from the crime scene window;
- $H_d$ : the recovered fragment is not from the crime scene window.

The distribution of refractive index measurements within a window are assumed to be normally distributed with variance  $\sigma^2$  and, simplistically, the mean refractive index of the crime scene window is taken to be  $x$ . Then, if  $H_p$  is true, the distribution of the refractive index of a fragment from the crime scene window is taken to be normally distributed with mean  $x$  and variance  $\sigma^2$ .

The distribution of refractive indices over the population of relevant windows is taken to be normal with mean  $\mu$  and variance  $\sigma^2$ . The variance in refractive index between windows  $\sigma^2$  is much greater than the variance within windows,  $\sigma^2$ , by a factor of  $10^4$ , say, such that the ratio  $\sigma^2 / \sigma^2 = 100$ . The value of the evidence of the measurement  $y$  of the refractive index of the fragment of glass on the suspect is then the ratio of two normal density functions and is equal to

$$= \exp \left\{ \frac{y - \mu}{2 \sigma^2} - \frac{y - x}{2 \sigma^2} \right\} \quad \{2\}$$

Table II [2] gives some illustrative values for this expression {2} with  $\sigma^2 / \sigma^2 = 100$ .

TABLE II. LIKELIHOOD RATIO VALUES FOR VARYING VALUES OF  $(y - x) / \sigma$  AND  $(y - \mu) / \sigma$

$(y - x) / \sigma$	$(y - \mu) / \sigma$		
	0	1	2
0	100	165	739
1	61	100	448
2	14	22	100
3	1	2	8

The rows provide a measure of how far apart the control mean and the recovered measurement are in units of within-window standard deviations. This is a measure of similarity. The columns provide a measure of how far apart the overall mean and the recovered measurement are in units of between-window standard deviations. This is a measure of rarity.

A perfect match of a common measurement gives a value for the evidence of  $(y - \mu) / \sigma = 0$ :  $y$  and  $x$  are equal and  $y$  is equal to the overall mean. In contrast, a perfect match of a rare measurement, say two standard deviations from the overall mean gives a very high likelihood ratio of 739. Two dissimilar measurements with  $y$  and  $x$  three standard deviations apart with  $y$  equal to the overall mean has a value of 1, which means that the evidence has no effect on the odds in favour of the suspect having been at the crime scene.

The above is a very simplistic illustration of the use of the likelihood ratio in that it makes several large distributional assumptions. More realistic illustrations are given in Aitken and Taroni [2].

### 3.3. Determination of the likelihood ratio for other evidential types of transfer evidence

A summary of the assessment of the values of various other evidential types is given in Aitken and Taroni [2]. The philosophy is that the evidence has to be assessed under each of two propositions, one proposed by the prosecution, the other proposed by the defence. The defence is under no obligation to propose a reason for the evidence; in the absence of a reason from the defence, the defence proposition is taken to be the opposite of that of the proposition. The evidence to be evaluated is transfer evidence and can be thought of as being in two pieces, known as control and recovered as described above. There will be that piece of ev-

idence found at the crime scene and that piece of evidence found in association with a suspect. There has to be a comparison of the two pieces. The likelihood ratio provides a comparison which takes account of similarity and rarity within the one statistic, as illustrated in the glass example above.

For certain evidential types, there are several layers of variation. These have to be modelled under each of the two propositions. An implication of the prosecution proposition is that the control and recovered evidence come from the same source. An implication of the defence proposition is that the control and recovered evidence come from different sources.

Examples of the propositions and evidence for different evidential types are given in Table III.

It is not always possible to provide a numerical value for the likelihood ratio. For example, in the example of earprints, there is no data set of earprints from which appropriate probability models may be derived. The example is given here to illustrate how the procedure may work and to aid development of data collection procedures to ensure that the appropriate data are collected.

Other evidential types mentioned in Aitken and Taroni [2] for which it is not yet possible to provide a likelihood ratio include fingerprints, hair, documents, envelopes and handwriting, though in this last example, some progress has been reported by Srihari et al. [7].

#### 4. Evidence interpretation

As stated in Aitken and Taroni [2], a “large part of the controversy over scientific evidence is due to the way in which the evidence is classically presented”. If the likelihood ratio is used then the evidence may be presented in the verbal form that the evidence is so many times more likely if the prosecution proposition is true than if the defence proposition is true. This is not a new idea and was used in a report in the Dreyfus case: “since it is absolutely impossible for us to know the *a priori* probability, we cannot say: this coincidence proves that the ratio of the forgery’s probability to the inverse probability is a real value. We can only say that, following the observation of this coincidence, this ratio becomes *X* times greater than before the observation” [5].

A more recent example of the use of the likelihood ratio was reported from New Zealand in relation to the interpretation of a DNA frequency of 1 in 12 400. “DNA profiling examination of the samples strongly supports a contention that the semen stain on the underpants of the complainants came from the accused. It is said that the likelihood of obtaining such DNA profiling results is at least 12 400 times greater if the semen stain originated from the accused than from another individual” *R. v. Montella*, 1 NZLR High Court, 1992, pp. 63–68.

TABLE III. PROSECUTION AND DEFENCE PROPOSITIONS AND EVIDENCE FOR DIFFERENT EVIDENTIAL TYPES

Type of evidence	Propositions		Control evidence	Recovered evidence
	Prosecution	Defence		
Earprint	Earmark has been left by the person who provided the earprint	Earmark has been left by the ear of some other unknown person	Earprint from person	Earmark impression from crime scene
Firearms and toolmarks	Bullet found at crime scene was fired from the suspect gun	Bullet found at crime scene was fired from a gun other than the suspect gun	Striations from a bullet fired from a known gun	Striations from a bullet found at the crime scene
Speaker recognition	Incriminating phone call, intercepted from a house, was made by the suspect, one of five men known to be in the house at the time of the call	Incriminating phone call, intercepted from a house, was made by one of the other four men known to be in the house at the time of the call.	Results of acoustic analyses of statements made by the suspect	Results of acoustic analyses of the phone call
Paint	Paint on the injured party’s vehicle originated from the suspect vehicle	Paint on the injured party’s vehicle originated from a random vehicle	Paint analysis of the suspect’s car	Paint analysis of the fragments found on the injured party’s vehicle.

Whilst the likelihood ratio is the ideal way to present the evidence it is difficult for a layman to understand [8] and so such a method of presentation has to be done with care. Careful use of the likelihood ratio reminds the user that the evidence has to be assessed under each of two propositions. Even if it is not possible to evaluate the probabilities under both of the propositions it is important to remember that knowledge of the probability of the evidence under one proposition only is not sufficient to provide a meaningful interpretation.

Consider the so-called prosecutor's fallacy and the defendant's fallacy [9]. For example, a crime is committed and a blood stain is found at the crime scene. There are no innocent explanations for the stain and, for the purposes of this example, the stain may be assumed to have been left by the criminal. A partial profile is obtained from the stain which shows that the profile has a frequency of 1% in the population to which the criminal is assumed to belong (the relevant population). A person is duly arrested and the profile is found to match that of the crime stain. At the subsequent trial the prosecutor states that the profile is found in only 1% of the population and, hence, the defendant has a 99% chance of being guilty. The defence counsel argues, conversely, that the defendant is one person in 1% of the relevant population. Assume the relevant population has 5000 members; 1% of 5000 is 50. Thus, the defendant has a 1 in 50 (or 2%) chance of being guilty. Hence the evidence of the match of the profiles of the crime stain and the suspect is not relevant.

One has to return to the odds form of Bayes' theorem {1} to understand the reasons for the differences in these interpretations. First, the prosecutor's argument. The evidence  $E$  is the profile of the crime stain, the prosecutor's proposition  $H_p$  is that the defendant left the crime stain, the defence proposition  $H_d$  is that some person unrelated to the defendant left the crime stain. Then  $Pr(E | H_d) = 0.01$ . The prosecutor argues that this result implies that  $Pr(H_p | E) = 0.99$  and hence that  $Pr(H_d | E) = 0.01$ . Assume that  $Pr(E | H_p) = 1$ , the profile of the crime stain and the suspect match if the defendant left the stain. Then the likelihood ratio is  $1/0.01 = 100$ . The posterior odds are  $0.99/0.01 = 99$ , or, approximately, 100. Thus, from {1}, the prior odds  $Pr(H_p)/Pr(H_d) = 1$ . Hence,  $Pr(H_p) = Pr(H_d) = 0.5$ . This suggests that the prior probability that the defendant left the crime stain, before any evidence of the profile match was led is 0.5. This is not an encouraging result from the perspective of the defendant who would like to believe in the maxim, "innocent until proven guilty".

Conversely, from the defence argument,  $Pr(H_p | E) = 1/50$ . Thus, the posterior odds are  $1/50$  divided by  $49/50$ , which is approximately  $1/50$ . The evidence has

a value of 100 thus, from {1}, the prior odds are 1 in 5000. This may be considered as 1 in the size of the relevant population. This is a perfectly reasonable interpretation. The dictum "innocent until proven guilty" may be taken to mean that the defendant is, *a priori*, just as likely as anyone else to be the source of the crime stain. The fallacy lies in suggesting that the evidence of the profiles is not relevant. Before the evidence of the profiles was presented, the defendant was 1 of 5000 possible suspects. After the evidence of the profiles was presented, the defendant was 1 of 50. Evidence which reduces the size of the pool of potential suspects by a factor of 100 is relevant.

Many errors arise because the expert witness makes a statement about the issue under discussion and not about the worth of the evidence. One example only is given here. The issue is the source of some blood on the clothing of a suspect: "with probability 1 in 7 million it can be said that the blood on the clothing of the suspect could be that of someone other than that of the victim" (*Wike v. State*, transcript, 596, S0 2<sup>nd</sup> 1020, Fla S.Ct., 1992, pp. 147–148.)

Consider the defence proposition ( $H_d$ ) in this case to be that the blood on the clothing came from someone unrelated to the victim. The evidence ( $E$ ) is a match of the profile of the blood on the clothing with that of the victim. Assume that the probability of a match if the prosecution proposition ( $H_p$ ) that the blood on the clothing came from the victim is 1. Then the likelihood ratio is 7 million. The evidence is 7 million times more likely if the blood on the clothing came from the victim than if it came from someone unrelated to the victim. However, the statement from the court is a statement for the value of  $Pr(H_d | E)$ . It is not possible to make a statement about this without knowledge of the prior probabilities as explained in the discussion above about the prosecutor's fallacy. When presenting evidence, care has to be taken not to make a comment on the issue under discussion.

Other errors of interpretation are discussed in Aitken and Taroni [2].

## 5. Future

The increasing role of the likelihood ratio highlights ways for the future development of statistics in forensic science.

- First, data collections are needed in various areas of evidential types where they do not exist at present in order to model the variability that is inherent in many of these types. For example, data collections

exist for DNA and for glass and these collections enable good statistical models to be developed.

- More sophisticated statistical models may be developed. For example, models which take account of several levels of variation and assess multivariate data, without assuming independence amongst the variables are being developed (see, for example, [4]).
- Bayesian networks to structure evidence. A network for profiling was shown in [4]. Networks may also be used for evidence evaluation. For example, Figure 1 [3] shows a network for the evaluation of cross-transfer evidence of bodily fluids where DNA profiles may be obtained. This network has 14 nodes. Factors considered include the transfer of material between the suspect ( $S$ ) and the victim ( $V$ ) (nodes 5 and 6; only one direction need be specified for the transfer explicitly), the backgrounds of the suspect (nodes 2 and 3) and of the victim (nodes 8 and 9), the choices of stains on the suspect (node 4) and the victim (node 7), matches in DNA profiles of what is found on the suspect with the profile of the victim (node 13) and of what is found on the victim with the profile of the suspect (node 14), reported matches of what is found on the suspect with the profile of the victim (node 1) and of what is found on the victim with the profile of the suspect (node 10), contact between the suspect and the victim (node 11) and the ultimate issue, that the suspect committed the crime (node 12). A like-

likelihood ratio for a piece of evidence may be determined from this network by first setting the ultimate issue node, 12, to “yes, the suspect committed the crime” and propagating this through the network. This propagation will assign a probability to the piece of evidence of interest, assuming the suspect committed the crime. The process may be repeated with the ultimate issue node set to “no, the suspect did not commit the crime”. Propagation of this will assign a probability to the piece of evidence of interest. The ratio of these two probabilities is then the likelihood ratio and is the value of that piece of evidence.

#### Acknowledgement

The author thanks the Institute of Forensic Research for their hospitality during the celebrations to mark the 75<sup>th</sup> anniversary of the founding of the Institute and for the invitation by the European Network of Forensic Science Institutes to chair the One Day One Issue Seminar on “The Role of Statistics in Forensic Science” in Krakow in November 2004 in association with the anniversary celebrations.

#### References

1. Aitken C. G. G., Statistics in forensic science I: An aid to investigation, *Problems of Forensic Sciences* 2006, 65, 53–67.

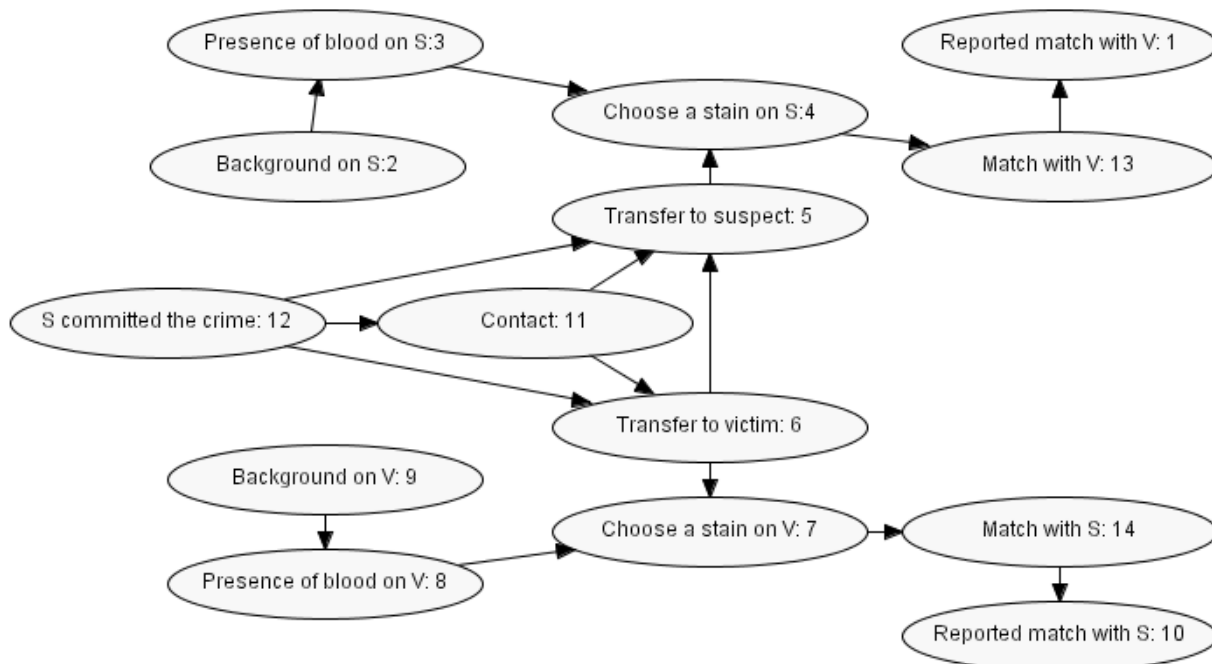


Fig. 1. Complete network for the evaluation of cross-transfer evidence in DNA profiles from [3] (reproduced by permission of Elsevier).

2. Aitken C. G. G., Taroni F., Statistics and the evaluation of evidence for forensic scientists, John Wiley and Sons, Chichester 2004.
3. Aitken C. G. G., Taroni F., Garbolino P., A graphical model for the evaluation of cross-transfer evidence in DNA profiles, *Theoretical Population Biology* 2003, 63, 179–190.
4. Aitken C. G. G., Lucy D., Zadora G. [et al.], Evaluation of transfer evidence for three-level multivariate data with the use of graphical models, *Computational Statistics and Data Analysis* 2006, 50, 2571–2588.
5. Darboux J. G., Appell P. E., Poincaré J. H., Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau, [in:] L'affaire Drefus – la révision du procès de Rennes enquête de la chambre criminelle de la Cour de Cassation, Ligue Française des Droits de l'Homme et du Citoyen, Paris 1908.
6. Evett I. W., A Bayesian approach to the problem of interpreting glass evidence in forensic science casework, *Journal of the Forensic Science Society* 1986, 26, 3–18.
7. Srihari S. N., Cha S. H., Arora H. [et al.], Individuality of handwriting, *Journal of Forensic Sciences* 2002, 47, 856–872.
8. Taroni F., Aitken C. G. G., Probabilistic reasoning in the law, part 1: assessment of probabilities and explanation of the value of DNA evidence, *Science & Justice* 1998, 38, 165–177.
9. Thompson W. C., Schumann E. L., Interpretation of statistical evidence in criminal trials. The prosecutor's fallacy and the defence attorney's fallacy, *Law and Human Behaviour* 1987, 11, 167–187.

---

**Corresponding author**

Colin G. G. Aitken  
The University of Edinburgh  
School of Mathematics  
Mayfield Road  
Edinburgh EH9 3JZ, United Kingdom  
e-mail: c.g.g.aitken@ed.ac.uk

---



## STATYSTYKA W NAUKACH SĄDOWYCH. CZĘŚĆ II. STATYSTYKA JAKO POMOC W OSZACOWANIU WARTOŚCI MATERIAŁU DOWODOWEGO

### 1. Wstęp

Metody statystyczne w naukach sądowych stosowane są zarówno w fazie dochodzeniowej śledztwa, czyli zanim podejrzany zostanie zidentyfikowany, jak też w fazie sądowej, kiedy podejrzany jest już znany. W pierwszym przypadku metody statystyczne znajdują zastosowanie jako narzędzie pomocnicze mające na celu dostarczenie informacji pozwalającej ustalić sprawcę przestępstwa. Natomiast w fazie procesowej metody statystyczne mogą być użyte w celu oceny wartości poszczególnych dowodów. Artykuł ten jest drugim z kolei, który omawia rolę statystyki w naukach sądowych. W poprzednim omówiono wykorzystanie statystyki jako narzędzia użytecznego na etapie dochodzeniowym śledztwa.

### 2. Testy istotności

Tak zwane podejście klasyczne oszacowania wartości dowodowej śladów kryminalistycznych jest metodą próbującą naśladować sposób, w jaki postępuje biegły sądowy w przypadku, gdy chce on porównać dowód zabezpieczony na miejscu zdarzenia z dowodem ujawnionym w otoczeniu podejrzanego. Podejście klasyczne opiera się na testach istotności stworzonych w pierwszej połowie dwudziestego wieku. Obecnie stosowane powszechnie w naukach sądowych metody statystyczne polegają na zastosowaniu teorii Bayesa. Teoria ta rozwijana była w osiemnastym stuleciu, czyli ponad sto lat wcześniej niż tzw. podejście klasyczne. Określenie „otoczenie podejrzanego” zostało użyte do opisania różnych możliwości pochodzenia materiału dowodowego. Może on bowiem pochodzić bezpośrednio od podejrzanego (np. profil DNA), może znajdować się na odzieży podejrzanego (np. okruchy szkła), może stanowić odcisk podeszwy jego obuwia (np. dowód w postaci odcisku znaku firmowego obecnego na podeszwie tego obuwia) lub może zostać zlokalizowany w miejscu zamieszkania podejrzanego (np. fragmenty obcego lakieru obecne na samochodzie zaparkowanym w garażu podejrzanego).

Klasyczne podejście służące oszacowaniu wartości dowodowej śladów kryminalistycznych składa się z dwóch faz, tzn. fazy porównywania dowodów oraz fazy późniejszej, związanej z oszacowaniem częstości występowania w populacji generalnej cech uwzględnianych w pierwszej fazie. Jak już wspomniano, w pierwszej fazie porównywany jest dowód znaleziony na miejscu zdarzenia

z dowodem znalezionym w otoczeniu oskarżonego. Wynik takiego porównania jest binarny, tj. dwa dowody mogą być uznane za podobne lub niepodobne. Jeżeli zostaną uznane za podobne, to wówczas przeprowadzana jest druga faza analizy polegająca na oszacowaniu wartości dowodowej analizowanego śladu. Podobieństwo dwóch dowodów, które opisane jest przez cechę rzadko występującą w populacji, jest znacznie silniejszym dowodem, niż gdy ta cecha jest często spotykana. W celach ilustracyjnych rozważmy dowód w postaci zeznań naoczego świadka na temat liczby rąk przestępcy. Na początku rozpatrzmy sytuację, gdy naoczny świadek zdarzenia stwierdził, że osoba popełniająca przestępstwo miała dwie ręce. Zatrzymany podejrzany również posiada dwie ręce. Liczba rąk podejrzanego jest więc identyczna z liczbą rąk przestępcy. Wniosek wynikający z analizy jest zatem taki, że dowody pasują do siebie w zakresie bardzo powszechnej cechy, jaką jest posiadanie przez człowieka dwóch rąk. Tym samym zeznanie naoczego świadka odnoszące się do liczby rąk nie jest nazbyt wartościową informacją i dowodem. Teraz rozważmy przypadek, gdy drugi świadek naoczny stwierdza, że przestępca posiadał tylko jedną rękę, a zatrzymany podejrzany również ma jedną rękę. Liczba rąk podejrzanego jest identyczna z liczbą rąk zaobserwowaną przez świadka, co oznacza zgodność dowodów w obrębie bardzo rzadkiej cechy, jaką jest posiadanie jednej ręki przez człowieka. W tym przypadku obserwacja poczyniona przez naoczego świadka jest bardzo wartościowym dowodem. Jeżeli dowody w pierwszej fazie zostaną uznane za różne, to wówczas analiza nie jest prowadzona dalej i uznaje się, że pochodzą one z dwóch różnych źródeł.

Jako inny przykład rozważmy przestępstwo, w trakcie którego zostało rozbite okno. Fragmenty szkła znalezione na miejscu zdarzenia są porównywane z fragmentami szkła znalezionymi na odzieży podejrzanego. W tym celu wykonywane zostają pomiary różnych parametrów opisujących te fragmenty, a posiadający charakter zmiennych ciągłych, takich jak współczynnik załamania światła lub skład chemiczny. Dane te dają wielowymiarowy opis dowodu i dlatego też w celu ich porównania powinien zostać zastosowany wielowymiarowy test istotności. Porównywane okruchy szkła określane są jako różne, jeżeli różnica wartości oznaczanych parametrów uzyskanych dla każdej z porównywanych próbek szkła jest większa niż dwa odchylenia standardowe, co odpowiada tzw. 5% poziomowi istotności.

Z zastosowaniem testów istotności związanych jest kilka problemów. Pierwszym z nich jest efekt, który można opisać jako „upadek z klifu”. W wybranym przez nas miejscu stoimy bezpiecznie na brzegu klifu, ale gdy wykonamy jeden mały krok w kierunku przepaści, wówczas spadniemy na jej dno. Tym samym mały krok posiada ogromne konsekwencje. Podobnie przedstawia się problem testu, w którym zakładamy 5% poziom istotności. Istnieje bowiem znaczna różnica, gdy stwierdzimy, że różnica wartości zmierzonych w trakcie analizy porównywanych próbek jest istotna na poziomie 4,9%, niż gdy różnica ta jest istotna na poziomie 5,1%. W pierwszym przypadku dowód jest poddawany analizie w drugiej fazie oszacowania jego wartości dowodowej i tym samym ciągle jest rozpatrywany jako dowód przeciwny podejrzanemu. W drugim przypadku dowód jest pomijany i tym samym nie będzie wzięty pod uwagę jako dowód przeciwko podejrzanemu. Dlatego też o tym, jak uznany zostanie dowód, decyduje wybór punktu granicznego, tzn. poziomu istotności, który jest wyborem subiektywnym.

Po drugie, przy zastosowaniu testów istotności nie jest łatwe uwzględnienie innych efektów, takich jak np. prawdopodobieństwo przeniesienia i pozostawiania okruszków szkła na odzieży. Trzeci problem związany jest z faktem, że testy te opierają się na założeniu, iż dowód z miejsca zdarzenia oraz z otoczenia podejrzanego pochodzą z identycznego źródła, o ile różnica zmierzonych wartości jest istotna na założonym poziomie 5%. Tym samym obowiązkiem obrońcy jest wykazanie, że różnica ta jest istotna na tym poziomie.

Niemniej jednak istnieje pozorna korzyść przy zastosowaniu testów istotności. Umożliwiają one bowiem porównanie dwóch dowodów wyłącznie przy wykorzystaniu danych zebranych w trakcie ich analizy bez konieczności dysponowania dużymi zbiorami danych, które z kolei są niezbędne w przypadku uwzględnienia informacji o rzadkości występowania danej cechy w populacji odniesienia.

### 3. Oszacowanie wartości dowodowej

Możliwe jest zastosowanie metody nieobciążonej wspomnianymi wyżej problemami, w której stanowiska prokuratora i obrońcy traktuje się równorzędnie. Nie zakłada się więc wyższości jednej hipotezy, jak to ma miejsce w opisanym powyżej podejściu, które opiera się na teście istotności.

#### 3.1. Możliwe hipotezy

W przedstawionym tutaj podejściu wartość dowodowa jest oszacowywana przez porównanie prawdopodobieństwa powstania dowodu, jeżeli wersja prokuratora

jest prawdziwa, z prawdopodobieństwem powstania dowodu, jeżeli wersja obrony jest prawdziwa.

Porównywane hipotezy niekoniecznie muszą sprawdzać się do stwierdzenia winny niewinny. W tabeli I wymieniono pięć par różnych hipotez [1].

Sąd zainteresowany jest tym, czy hipotezy te są prawdziwe, czy też nie. Biegłego sądowego interesuje natomiast uzyskanie informacji o tym, w jakim stopniu dowód wspiera hipotezę prokuratora lub hipotezę obrony.

Dla ułatwienia dalszego wywodu wprowadźmy następującą notację. Jako  $Pr$  zostanie oznaczone prawdopodobieństwo,  $H_p$  i  $H_d$  oznaczają odpowiednio hipotezy prokuratora i obrony,  $E$  dowód, a  $I$  inne dostępne informacje o przebiegu przestępstwa. Kreska pionowa | oznacza zależność zdarzeń. Zdarzenie po lewej stronie kreski jest tym, którego prawdopodobieństwem jesteśmy zainteresowani. Zdarzenie, bądź zdarzenia po prawej stronie kreski, są zdarzeniami, które w założeniu są znane. Na podstawie teorii Bayesa można napisać, że:

$$\frac{Pr(H_p|E,I)}{Pr(H_d|E,I)} = \frac{Pr(E|H_p,I)}{Pr(E|H_d,I)} \cdot \frac{Pr(H_p|I)}{Pr(H_d|I)} \quad \{1\}$$

Jako przykład interpretacji zależności zdarzeń rozważmy licznik stosunku prawdopodobieństw występujący po lewej stronie równania {1}. Po lewej stronie kreski znajduje się hipoteza  $H_p$ , której prawdopodobieństwo chcemy ustalić. Zdarzenia po prawej stronie kreski,  $E$  i  $I$ , są tymi, dla których zakładamy, że ich wynik jest znany. Tym samym wyrażenie  $Pr(H_p|E,I)$  jest prawdopodobieństwem zajścia zdarzenia, że hipoteza prokuratora jest prawdziwa, pod warunkiem, że ujawniono dowód  $E$  i uwzględniono informacje  $I$  opisujące przebieg zdarzenia.

Stosunek prawdopodobieństw po prawej stronie równania {1}, tj.  $\frac{Pr(H_p|I)}{Pr(H_d|I)}$ , jest tzw. stosunkiem szans *a priori*

na korzyść hipotezy  $H_p$ . Uwzględni on wartości prawdopodobieństw zajścia obu hipotez przed zdobyciem informacji o dowodzie (np. informacji uzyskanych w trakcie badań jego właściwości fizykochemicznych). Należy dodać, że określając prawdopodobieństwa *a priori*, możemy uwzględnić informacje o przebiegu przestępstwa ( $I$ ).

Wyrażenie po lewej stronie równania {1},  $\frac{Pr(H_p|E,I)}{Pr(H_d|E,I)}$ ,

jest stosunkiem prawdopodobieństw *a posteriori* na korzyść zdarzenia opisanego hipotezą  $H_p$ .

Współczynnik, który przekształca szanse *a priori* w szanse *a posteriori*, jest stosunkiem prawdopodobieństw, pierwszym wyrażeniem po prawej stronie równania. Zwany jest jako iloraz wiarygodności (ang. likelihood ratio). Mierzy on stosunek prawdopodobieństw powstania dowodu w przypadku, gdy każda z dwóch hipotez jest prawdziwa. Na przykład, gdy iloraz wiarygodności będzie posiadać wartość  $V$ , to można wówczas stwierdzić, że dowód jest  $V$  razy bardziej wiarygodny, jeżeli hipoteza

prokuratora jest prawdziwa, niż gdy hipoteza obrony jest prawdziwa.

### 3.2. Iloraz wiarygodności

Iloraz wiarygodności,  $\frac{Pr(E|H_p, I)}{Pr(E|H_d, I)}$ , można potraktować jako miarę wartości dowodowej analizowanego śladu kryminalistycznego. Jest to miara określająca, jak mocno dowód wspiera hipotezę prokuratora, a jak hipotezę obrony. Jako przykład rozważmy przypadek, w którym szanse *a priori* na korzyść, że podejrzany jest winny, wynoszą 1 do 1000. Załóżmy, że osobę należy uznać za niewinną aż do momentu, gdy szanse *a posteriori* wynoszą 1000 do 1 na korzyść hipotezy oskarżenia. Wówczas iloraz wiarygodności, który można określić jako stosunek szans *a posteriori* do szans *a priori*, w tym przypadku wyniesie 1 000 000. Tak więc przekształcenie szans *a priori* wynoszących 1:1000 do stosunku 1000:1 wymaga, aby dowód był 1 000 000 razy bardziej wiarygodny w przypadku, gdy hipoteza prokuratora jest prawdziwa, niż wtedy, gdy hipoteza obrony jest prawdziwa.

Wartych omówienia jest kilka właściwości relacji opisanych równaniem {1}. Szanse *a posteriori* dla jednego dowodu stają się szansami *a priori* dla następnego dowodu, przy czym należy uwzględnić fakt, że wcześniejszy dowód jest częścią warunku dla następnego dowodu. Iloraz wiarygodności, którego wartość jest większa (mniejsza) niż jeden, wskazuje, że dowód wspiera hipotezę prokuratora (obrony). Ponadto przyjmuje on wartości pomiędzy 0 a nieskończonością, a tym samym iloraz wiarygodności nie może być utożsamiany z prawdopodobieństwem. Iloraz wiarygodności nie może być interpretowany w ten sposób, że określa on prawdopodobieństwo zajścia którejkolwiek z hipotez. Należy traktować go jako metodę, która dostarcza logicznego wytłumaczenia dla rozważań nad wartością dowodu. Należy podkreślić, że nie jest wystarczające oszacowanie możliwości, że dowód powstał w przypadku, gdy jedynie hipoteza  $H_p$  jest prawdziwa. Inna cecha ilorazu wiarygodności związana jest z tym, że przy oszacowaniu wartości dowodowej śladów kryminalistycznych daje on możliwość uwzględnienia prawdopodobieństw zdarzeń odnoszących się do przeniesienia się materiału dowodowego z jednego obiektu na drugi, np. włókien z odzieży ofiary na odzież przestępcy.

Kolejną możliwość zastosowania ilorazu wiarygodności można pokazać na przykładzie przeniesienia materiału dowodowego gdy dowód został przeniesiony z miejsca przestępstwa na przestępcę lub *vice versa*. Przykładem pierwszego zdarzenia jest sytuacja, gdy fragmenty szkła z rozbitego okna zostaną przeniesione na odzież przestępcy. Jako przykład przeniesienia śladu z przestępcy na miejsce zdarzenia można uznać np. plamę krwi pozostawioną przez osobę, która skaleczyła się

w trakcie dokonywania włamania. Dowód, którego źródło jest znane, nazywany jest próbką porównawczą. Mogą to być fragmenty szkła znalezione na miejscu zdarzenia lub próbka krwi pobrana od podejrzanego, który niekoniecznie jest przestępcą. Dowód, którego źródło nie jest znane, nazywa się materiałem odzyskanym. Mogą to być fragmenty szkła znalezione na odzieży podejrzanego, który dokonał lub też nie dokonał zarzucanego mu czynu, bądź też plama krwi ujawniona na miejscu zdarzenia, na temat której istnieje podejrzenie, że pochodzi od przestępcy, choć nie został on jeszcze zidentyfikowany.

Iloraz wiarygodności łączy również ocenę podobieństwa porównywanych obiektów oraz częstości wystąpienia w populacji generalnej cech opisujących analizowane obiekty w jednym podejściu statystycznym, w przeciwieństwie do dwustopniowego procesu stosowanego w przypadku użycia testów istotności. Uproszczonego przykład wykorzystania ilorazu wiarygodności w praktyce można znaleźć w literaturze przedmiotu [2]. Autorzy zastosowali zależności rozwinięte przez Evetta [6]: odzyskano fragmenty szkła z odzieży osoby podejranej o popełnienie przestępstwa, w trakcie którego rozbite okno. Wyznaczona wartość współczynnika załamania światła dla tych fragmentów oznaczona została jako  $y$ . Próbkę kontrolną złożoną z  $m$  fragmentów szkła pobrano ze szkła zabezpieczonego na miejscu zdarzenia, a średnia wartość współczynnika załamania światła zmierzonego dla tych okruszków wynosiła  $x$ . Rozważamy następujące hipotezy na tzw. poziomie aktywności:

- $H_p$ : odzyskany fragment szkła pochodzi z okna rozbitego na miejscu zdarzenia;
- $H_d$ : odzyskany fragment szkła nie pochodzi z okna rozbitego na miejscu zdarzenia.

Jeżeli  $H_p$  jest prawdziwa, to wówczas należy założyć, że rozkład współczynnika załamania światła fragmentu szkła zabezpieczonego z okna rozbitego na miejscu zdarzenia jest rozkładem normalnym ze średnią  $x$  i wariancją  $\sigma^2$ .

Założono, że rozkład wartości współczynnika załamania światła uzyskanych dla populacji okien również jest rozkładem normalnym, ale ze średnią  $y$  i wariancją  $\sigma^2$ . Wariancja współczynnika załamania światła pomiędzy oknami  $\sigma^2$  jest znacznie większa niż wariancja wewnątrz okna  $\sigma^2$ . Założono, że rozkład wartości współczynnika załamania światła uzyskanych w obrębie pojedynczego okna jest rozkładem normalnym, ze średnią  $x$  i wariancją  $\sigma^2$  około 1000-krotnie, co odpowiada stosunkowi  $\sigma^2 = 100$ . Wartość dowodu w postaci rezultatu pomiaru współczynnika załamania światła fragmentu szkła znalezionego na odzieży podejrzanego oznaczona jako  $y$  jest wówczas stosunkiem dwóch funkcji gęstości rozkładu normalnego i może być wyrażona jako:

$$= \exp \left[ -\frac{y^2}{2\sigma^2} - \frac{y x^2}{2\sigma^2} \right] \quad \{2\}$$

Tabela II [2] ilustruje różne wartości wyrażenia  $\{2\}$  dla  $\frac{y}{x} = 100$ .

W wierszach określono, jak daleko od siebie, w jednostkach odchylenia standardowego  $\sigma$ , znajduje się średnia wartość współczynnika załamania światła wyznaczona dla próbki kontrolnej i średnia wartość współczynnika załamania światła wyznaczona dla próbki odzyskanej. Jest to miara podobieństwa tych obiektów. W kolumnach umieszczono dane określające, jak daleko od siebie, w jednostkach odchylenia standardowego  $\sigma$ , znajduje się średnia  $\bar{y}$  i średnia  $\bar{x}$  wyznaczone dla odzyskanej próbki. Jest to miara rzadkości zmierzonej wartości  $y$  w populacji okien.

W przypadku, gdy współczynniki załamania światła próbki odzyskanej i kontrolnej mają takie same wartości i ponadto, gdy zmierzona wartość współczynnika załamania światła jest często spotykana w populacji okien, to wówczas wartość ilorazu wiarygodności wyniesie 100. Natomiast w przypadku, gdy współczynniki załamania światła próbki odzyskanej i kontrolnej również są identyczne, ale zmierzona wartość jest rzadka w populacji okien, powiedzmy odległa o dwa odchylenia standardowe od średniej wartości współczynnika załamania światła w populacji okien  $\bar{x}$ , to wówczas uzyskamy wartość ilorazu wiarygodności równą 739. Gdy uzyskamy dwie różne wartości współczynnika załamania światła  $y$  i  $x$  odległe od siebie o np. trzy odchylenia standardowe  $\sigma$  oraz gdy  $y$  jest równe średniej wartości współczynnika załamania światła w populacji okien  $\bar{x}$ , to wówczas iloraz wiarygodności ma wartość 1, co oznacza, że dowód nie wspiera żadnej z rozpatrywanych hipotez  $H_p, H_d$ .

Zaprezentowany powyżej bardzo prosty przykład ilustruje sposób zastosowania ilorazu wiarygodności w zależności od różnych założeń związanych z uzyskanymi wartościami współczynnika załamania światła. Bardziej realne przykłady zawarte są w innej publikacji [2].

### 3.3. Oszacowanie ilorazu wiarygodności dla innych typów materiału dowodowego

Oszacowanie wartości dowodowej różnego rodzaju śladów kryminalistycznych opisane już zostało w literaturze przedmiotu [2]. Zasada ta jest następująca: dowód musi być oszacowywany z uwzględnieniem dwóch hipotez – jednej, zaproponowanej przez prokuratora, i drugiej, zaproponowanej przez obronę. Oczywiście obrona nie ma obowiązku proponowania hipotez odnoszących się do przyczyny powstawania dowodu. Dlatego też w przypadku braku hipotezy postawionej przez obronę hipoteza taka powinna zostać sformułowana jako przeciwna do hipotezy prokuratora. Jak już wspomniano, materiał dowodowy, którego źródło jest znane, zwany jest próbką kontrolną. Przypomnijmy również, że próbka odzyskana jest próbką znalezioną podczas śledztwa w otoczeniu podejrzanego. Te dwie próbki są ze sobą po-

równywane np. przez zastosowanie ilorazu wiarygodności, który określa podobieństwo i częstość występowania w populacji generalnej danego dowodu w ramach jednego podejścia statystycznego, podobnie jak to opisano powyżej w przypadku analizy szkła.

Należy dodać, że dla każdego typu materiału dowodowego można znaleźć kilka źródeł zmienności, jak np. zmienność analizowanych cech wewnątrz pojedynczego obiektu i pomiędzy różnymi obiektami, które muszą zostać uwzględnione przy oszacowaniu dowodu. Przykłady rozpatrywanych hipotez w przypadku różnego rodzaju dowodów zebrane zostały w tabeli III.

Uzyskanie wartości ilorazu wiarygodności w formie liczbowej nie zawsze jest możliwe. Na przykład brakuje danych opisujących odciski małżowin usznych, z których można by wyprowadzić odpowiednie modele statystyczne. Przedstawione w tej publikacji przykłady zostały przytoczone w celu zobrazowania, czym jest iloraz wiarygodności, a także w celu uświadomienia czytelnikowi, jak taki model powinien być opracowywany oraz w jaki sposób powinny być zbierane dane niezbędne do jego powstania.

Analiza statystyczna w oparciu o iloraz wiarygodności dla innego typu dowodów, dla których ciągle nie można podać liczbowej wartości ilorazu wiarygodności, opisano w oddzielnym opracowaniu [2]. Zaliczają się do nich odciski palców, włosy, analiza dokumentów lub analiza pisma ręcznego, jakkolwiek w tym ostatnim przypadku dokonał się pewien postęp [7].

## 4. Interpretacja dowodów

Jak wspomniano w literaturze przedmiotu [2], „duża część kontrowersji wokół dowodów naukowych związana jest ze sposobem ich prezentacji z zastosowaniem tzw. klasycznego podejścia do analizy dowodu”. Gdy używamy ilorazu wiarygodności, to wówczas dowód może zostać zinterpretowany tak, że w większym stopniu wspiera on hipotezę prokuratora niż obrony. Nie jest to nowe podejście i zostało ono już wykorzystane w sprawie Dreyfusa: „od kiedy nie możemy absolutnie poznać prawdopodobieństwa *a priori*, to nie możemy stwierdzić następująco: analizowane zdarzenie dowodzi, iż stosunek prawdopodobieństwa, że dokonano fałszerstwa do prawdopodobieństwa zdarzenia odwrotnego, jest rzeczywistą wartością. Możemy tylko powiedzieć, że obserwacja danego zdarzenia powoduje, że możliwość dokonania fałszerstwa jest  $X$  razy większa, niż była ona przed obserwacją tego zdarzenia” [5].

Bardziej współczesnym przykładem zastosowania ilorazu wiarygodności (pochodzącym Nowej Zelandii) jest interpretacja częstości wystąpienia profilu DNA w stosunku 1 do 12 400. „Analiza profilu DNA próbki dowodowej silnie wspierała twierdzenie, że plama spermy na majt-

kach pochodzi od podejrzanego. Stwierdzono, że wiarygodność uzyskanego profilu DNA jest co najmniej 12 400 większa, jeżeli sperma pochodzi od podejrzanego niż od innej osoby” (*R. v. Montella*, 1 NZLR High Court, 1992, 63–68).

Pomimo że iloraz wiarygodności jest idealną metodą oszacowania wartości dowodu, to jest ona jednak trudno zrozumiała dla prawników [8] i dlatego powinna być stosowana z dużą ostrożnością. Jak już wielokrotnie podkreślano, wykorzystanie ilorazu wiarygodności wymaga, by dowód był rozważany w związku z dwiema hipotezami.

Rozważmy na przykładzie tzw. błąd prokuratora i błąd obrońcy [9]. Popelniono przestępstwo i na miejscu zdarzenia została znaleziona plama krwi. Nie jest możliwe, że plama krwi powstała w sposób inny niż związany z przestępstwem. Dla potrzeb tego przykładu założono więc, że plama krwi pozostawiona została przez przestępcę. Profil DNA tej plamy krwi występuje w populacji odniesienia, do której jest przypisany podejrzany, z częstością 1%. Ustalono, że profil DNA osoby aresztowanej w związku z tym zdarzeniem jest zgodny z profilem DNA plamy krwi. Podczas procesu prokurator stwierdził, że skoro profil DNA istnieje w 1% populacji odniesienia, to oskarżony jest w 99% winny. Natomiast obrońca argumentował, że oskarżony jest tylko jedną z osób spośród jednoprocetowej populacji odniesienia. Jeśli przyjmie się, że populacja ta składa się z 5000 osób, to 1% z 5000 wynosi 50 osób. Tak więc oskarżony ma szansę 1 na 50 (2%), że jest winny. W związku z tym nie jest istotne, że dowodowa plama ma profil zgodny z profilem plamy krwi znalezionej na miejscu zdarzenia.

Powróćmy do równania określającego szanse w ramach teorii Bayesa  $\{1\}$  w celu zrozumienia różnic występujących w tych interpretacjach. Po pierwsze, prześledźmy argumenty prokuratora. Dowodem  $E$  jest profil DNA plamy krwi znalezionej na miejscu zdarzenia. Hipoteza prokuratora  $H_p$  mówi, że oskarżony pozostawił plamę krwi, zaś hipoteza obrony  $H_d$  zakłada, że tę plamę krwi pozostawiła osoba niespokrewniona z oskarżonym. Wówczas  $Pr(E | H_d) = 0,01$ . Prokurator argumentował, że rezultat ten wskazuje, iż  $Pr(H_p | E) = 0,99$  i w związku z tym  $Pr(H_d | E) = 0,01$ . Załóżmy, że  $Pr(E | H_p) = 1$ , tj. profil plamy krwi i oskarżonego pasują do siebie. Wówczas iloraz wiarygodności wynosi  $1/0,01 = 100$ . Szanse *a posteriori* wynoszą  $0,99/0,01 = 99$  lub w przybliżeniu 100. Tak więc z  $\{1\}$  wynika, że szanse *a priori*  $Pr(H_p) / Pr(H_d) = 1$ . Tak więc  $Pr(H_p) = Pr(H_d) = 0,5$ . Sugeruje to, że prawdopodobieństwo *a priori* (tj. ustalone zanim informacja, iż profil DNA plamy krwi i krwi pobranej od podejrzanego są zgodne), że oskarżony pozostawił plamę krwi na miejscu zdarzenia, jest równe 0,5. Z perspektywy oskarżonego nie jest to pocieszający rezultat, bowiem bardziej odpowiada mu maksimum, że „jest niewinny aż do udowodnienia winy”.

Konsekwentnie, obrona argumentuje, że  $Pr(H_p | E) = 1/50$ . Szanse *a posteriori* wynoszą więc  $1/50$  podzielne przez  $49/50$ , co w przybliżeniu daje  $1/50$ . Iloraz wiarygodności równa się 100, czyli z równania  $\{1\}$  wynika, że szanse *a priori* wynoszą 1 do 5 000. Wartość ta może to być interpretowana jako 1 w stosunku do rozmiaru populacji odniesienia i jest to rozsądna interpretacja. Powiedzenie „niewinny aż do udowodnienia winy” może być rozumiane w ten sposób, że oskarżony *a priori*, jak każda inna osoba, może być źródłem plamy pozostawionej na miejscu zdarzenia. Błąd zawiera się w sugestii, że dowód z profilowania DNA nie jest istotny. Zanim wzięto pod uwagę dowód z profilu DNA, to oskarżony był jednym z 5000 możliwych podejrzanych, natomiast gdy dowód ten został już uwzględniony, to był on jedną z 50 osób. Dowód, który stukrotnie redukuje liczbę potencjalnych podejrzanych, na pewno jest istotny.

Wiele błędów wynika stąd, że biegły zeznaje w kwestii będącej przedmiotem sprawy, a nie w kwestii wartości dowodu. W celu ilustracji rozważmy następujący przypadek. Sprawa związana była ze źródłem plamy krwi znalezionej na odzieży podejrzanego. „Z prawdopodobieństwem 1 do 7 milionów można uznać, że plama krwi na odzieży podejrzanego pochodzi od kogokolwiek innego niż ofiara” (*Wike v. State*, transcript, 596, S0 2<sup>nd</sup> 1020, Fla S.Ct., 1992, pp. 147–148.). Rozważmy hipotezę obrony ( $H_d$ ): krew na odzieży pochodzi od kogoś niezwiązanego z podejrzanym. Dowód ( $E$ ) jest zgodny z profilem DNA plamy krwi z odzieży ofiary. Załóżmy, że prawdopodobieństwo dopasowania profili DNA, jeżeli hipoteza prokuratora jest prawdziwa ( $H_p$ ), czyli że krew na odzieży pochodzi od ofiary, wynosi 1. Wówczas iloraz wiarygodności wynosi 7 milionów. Oznacza to, że dowód jest 7 milionów razy bardziej wiarygodny, jeżeli krew znalezionej na odzieży pochodzi od ofiary, niż gdy pochodzi od kogoś innego, niezwiązanego z przestępstwem. W tym przypadku werdykt sądu jest więc werdyktem związanym z wartością  $Pr(H_d | E)$ . Jednak nie jest możliwe wydawanie wyroków w sprawie bez wiedzy o prawdopodobieństwie *a priori*, jak to wyjaśniono powyżej, opisując błąd prokuratora. Ponadto, kiedy opinia prezentowana jest w sądzie, to wówczas biegły powinien się zatroszczyć o to, aby ukazać daną sprawę w różnym świetle. Inne rodzaje błędów dotyczących interpretacji dowodu w oparciu o iloraz wiarygodności omówione zostały w literaturze [2].

## 5. Perspektywy

Wzrastająca rola ilorazu wiarygodności wytycza kierunki przyszłych działań i badań w dziedzinie statystyki w naukach sądowych. Oznacz to, że:

- konieczne jest tworzenie baz danych dla różnego rodzaju dowodów, których wartość chcemy oszacować,

w celu opracowania odpowiednich modeli statystycznych, tzn. takich, które uwzględniać będą wpływ przyczyn różnego rodzaju zmienności właściwych dla konkretnego typu materiału dowodowego na wartość ilorazu wiarygodności. Na przykład istnieją odpowiednie bazy danych w przypadku profili DNA lub szkła, co umożliwiło opracowanie odpowiednich modeli statystycznych do interpretacji tych dowodów;

- powinny być tworzone bardziej precyzyjne metody statystyczne. Przykładem są modele, które uwzględniają jednocześnie wiele zmiennych i kilka różnych przyczyn zmienności cech fizykochemicznych [4];
- powinny być prowadzone badania nad zastosowaniem sieci bayesowskich do analizy struktury danych i wartości dowodowej różnego rodzaju śladów kryminalistycznych. Na rycinie 1 [3] pokazany został przykład sieci złożonej z 14 węzłów, służącej do oszacowania profilu DNA uzyskanego dla płynów ustrojowych w przypadku krzyżowego przeniesienia materiału. Zmienne rozważane w tej sieci uwzględniają przeniesienie materiału pomiędzy podejrzanym  $S$  a ofiarą  $V$  (węzły 5 i 6; istotny jest tylko jeden kierunek zależności w celu wytłumaczenia procesu przeniesienia), informacje o podejrzanym (węzły 2 i 3) i ofierze (węzły 8 i 9), wybór plam do analizy, które ujawniono na podejrzanym (węzeł 4) i na ofierze (węzeł 7), informacje o tym, czy profil DNA śladu znalezione na podejrzanym jest zgodny z profilem DNA śladu znalezione na ofierze (węzeł 13) i czy ślady znalezione na ofierze posiadają profil DNA zgodny z profilem DNA podejrzanego (węzeł 14); informacje o tym, czy w trakcie wykonanego badania uzyskano zgodność profilu DNA oznaczonego dla płynu ustrojowego znalezione na oskarżonym z profilem DNA ofiary (węzeł 1) oraz z profilu DNA śladu znalezione na ofierze z profilem DNA oskarżonego (węzeł 10), informacje o możliwości kontaktu fizycznego pomiędzy podejrzanym a ofiarą (węzeł 11) i o tym, czy podejrzany popełnił przestępstwo (węzeł 12). Iloraz wiarygodności może być określony za pomocą tej sieci następująco: po pierwsze, gdy założymy w węźle 12, że sytuacja „podejrzany popełnił przestępstwo” jest prawdziwa (tj. prawdopodobieństwo tego zdarzenia jest równe 1), to wówczas ta informacja jest przesyłana automatycznie przez całą sieć, co w efekcie końcowym umożliwia uzyskanie wartości prawdopodobieństwa zdarzenia, że rozpatrywany dowód w postaci profilu DNA powstał, gdy hipoteza, że podejrzany popełnił przestępstwo, jest prawdziwa, tj. gdy  $H_p$  jest prawdziwe. Procedura ta może być powtórzona w tym samym węźle, gdy założymy, że prawdziwe jest zdarzenie, iż „podejrzany nie popełnił przestępstwa”. Również wtedy następuje automatyczne przesłanie tej informacji przez sieć, co pozwala uzyskać wartość prawdopodobieństwa zda-

żenia, że rozważany dowód powstał w przypadku, gdy podejrzany nie popełnił przestępstwa, tj. gdy  $H_d$  jest prawdziwe. Stosunek tych dwóch prawdopodobieństw, jak już wspomniano wcześniej, jest ilorazem wiarygodności i określa wartość tego dowodu.

#### Podziękowania

Autor dziękuje dyrekcji i pracownikom Instytutu Ekspertyz Sądowych za gościnność podczas jubileuszu 75. rocznicy powstania tej placówki, zaś przedstawicielom Europejskiej Sieci Instytutów Nauk Sądowych za zaproszenie do przewodniczenia w jednodniowym seminarium ENFSI zatytułowanym „Rola statystyki w naukach sądowych”, które odbyło się w ramach obchodów rocznicowych w Krakowie w listopadzie 2004 roku.