Science and Justice 57 (2017) 73–75

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Forensic likelihood ratio: Statistical problems and pitfalls☆

A. Philip Dawid

Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK

ARTICLE INFO

Keywords:
Bayes factor
Estimation
Forensic
Likelihood ratio

ABSTRACT

This article is a response to the position papers published in the *Science & Justice* virtual special issue on measuring and reporting the precision of forensic likelihood ratios. I point out a number of serious statistical errors in some of these papers. These issues need to be properly addressed before the philosophical debate can be conducted in earnest.

1. Introduction

This collection of position papers focuses largely on philosophical issues. However, there are serious statistical difficulties that first need to be addressed.

To clarify the main points, I will focus on a very simple case. Suppose that a certain marker (not necessarily genetic) exists in just two variant types, coded 0 and 1. A trace found at the crime scene is of type T, the suspect's marker is of type S, and we have a database of N individuals, of which D are of type 1. We assume all markers from distinct individuals are independent, with the same probability θ of being of type 1. The "prosecution hypothesis", H_p, is the proposition that S and T come from the same individual; the defence hypothesis H_d is that they come from distinct individuals. We denote by H the hypothesis variable, with possible values H_p and H_d, and suppose that H is independent of (T, D, θ):

$$H \perp\!\!\!\perp (T, D, \theta). \tag{1}$$

The full evidence in the case at hand is $E : S = 1, T = 1, D = d$.[1]

2. Known parameter

Initially suppose θ is known. The likelihood function $\mathbf{L}(\theta)$ comprises two terms, $\mathbf{L}(\theta) = (L(H_p; \theta), L(H_d; \theta))$, where

$$\begin{aligned} L(H_p; \theta) &= P(E \mid \theta, H_p) \\ &= P(S = 1, T = 1, D = d \mid \theta, H_p) \\ &= \theta^{d+1}\,(1-\theta)^{N-d} \end{aligned} \tag{2}$$

$$\begin{aligned} L(H_d; \theta) &= P(E \mid \theta, H_d) \\ &= P(S = 1, T = 1, D = d \mid \theta, H_d) \\ &= \theta^{d+2}\,(1-\theta)^{N-d}. \end{aligned} \tag{3}$$

The likelihood ratio, in favour of H_p as against H_d, is $\mathrm{LR}(\theta) := L(H_p; \theta)/L(H_d; \theta) = 1/\theta$.

3. Unknown parameter: frequentist approach

When θ is unknown, we need to eliminate it somehow. The frequentist way of doing this is to replace θ by an estimate, $\widehat{\theta}$. The most obvious estimate is the relative frequency, $\widehat{\theta} = d/N$, of type 1 in the database. As noted by Curran [4, Section 4], when d is very small (type 1 is very rare), it is sometimes recommended to augment the database with the crime trace,[2] so yielding the larger value $\widehat{\theta} = (d+1)/(N+1)$; Dawid and Mortera [5] argue that this should be done in all cases. However it is computed, the estimate is then generally plugged in to $\mathrm{LR}(\theta)$ to yield the estimated likelihood ratio,

☆ This paper is part of the Virtual Special Issue entitled: Measuring and Reporting the Precision of Forensic Likelihood Ratios, [http://www.sciencedirect.com/science/journal/13550306/vsi], Guest Edited by G. S. Morrison.

E-mail address: apd@statslab.cam.ac.uk (A. Dawid).

[1] Do not confuse this instance of d with its use to denote the defence hypothesis.

[2] Or, the suspect's profile — it makes no difference, as these are identical.

http://dx.doi.org/10.1016/j.scijus.2016.09.002
1355-0306/

$\widehat{\mathrm{LR}} = 1/\widehat{\theta}$. The crux of much of the discussion is how to qualify this simple point estimate with a measure of the uncertainty resulting from the fact that θ is not known, but has been estimated from finite data. However, this discussion starts off on the wrong foot, since there is no justification for the above estimation procedure in the first place.

3.1. Estimation

There is a standard likelihood procedure for comparing two hypotheses in the presence of unknown parameters. Although the general theory supporting this applies in a somewhat different context, it is nevertheless informative to see where this approach leads.

The principal requirement is that, for each hypothesis, we compute the maximum likelihood estimate of θ *assuming that hypothesis is true*, and plug that in to the associated term in the likelihood function. Only then do we proceed to take a ratio. In particular, since we will typically be using different values for θ in the numerator and denominator of the likelihood ratio, we must not cancel common terms prematurely: we need to work with the full likelihood function $\mathrm{L}(\theta)$ represented by expressions (2) and (3). As we shall see, this will be particularly important in the case of a rare allele, and especially so when that has not been observed in the database ($d = 0$).

Under H_p we have seen a total of $d+1$ distinct individuals of type 1, out of $N+1$ distinct individuals in total. The value of θ maximising Eq. (2) is thus $\widehat{\theta}_p = (d+1)/(N+1)$, and the maximised value is

$$L(H_p; \widehat{\theta}_p) = \frac{(d+1)^{d+1}\,(N-d)^{N-d}}{(N+1)^{N+1}}.$$

Under H_d we have seen a total of $d+2$ distinct individuals of type 1, out of a total of $N+2$. The value of θ maximising Eq. (3) is $\widehat{\theta}_d = (d+2)/(N+2)$, and the maximised value is

$$L(H_d; \widehat{\theta}_d) = \frac{(d+2)^{d+2}\,(N-d)^{N-d}}{(N+2)^{N+2}}.$$

The appropriate likelihood ratio is thus

$$\widehat{\mathrm{LR}} = \frac{(d+1)^{d+1}\,(N+2)^{N+2}}{(d+2)^{d+2}\,(N+1)^{N+1}}.$$

Assuming N is reasonably large, the term $(N+2)^{N+2}/(N+1)^{N+1}$ approximates to Ne,[3] and so

$$\widehat{\mathrm{LR}} \approx Ne\,\frac{(d+1)^{d+1}}{(d+2)^{d+2}}. \tag{4}$$

This $\widehat{\mathrm{LR}}$ could be regarded as obtained from $\mathrm{LR}(\theta) = 1/\theta$ by plugging in the "compromise estimate" $\widehat{\theta}_c = (d+2)^{d+2}/(d+1)^{d+1}Ne$ — and for small d this can be very different from any estimate normally considered. When $d = 0$ we get $\widehat{\theta}_c = 4/Ne = 1.47/N$, and for $d = 1$ we get $\widehat{\theta}_c = 27/4Ne = 2.48/N$.

3.2. Uncertainty?

What uncertainty should we associate with the above likelihood ratio?[4] We might condition on the case evidence $S = T = 1$, and regard d in Eq. (4) as the outcome of the random variable D having the binomial distribution $\beta(N; \theta)$. But then we still need to decide what value of θ to use, and there does not seem to be a unique principled way of choosing this. One possibility is to use the defence's estimate $\widehat{\theta}_d = (d+2)/(N+2)$: this respects both the legal "presumption of innocence" and the standard statistical approach to testing the null hypothesis H_d. An alternative might be to use the compromise estimate $\widehat{\theta}_c$. These will typically give different assessments of uncertainty.

Moreover, it is not clear that variations in the outcome of D are the only relevant frequentist source of uncertainty. Both S and T might have turned out differently: we could have observed $S = T = 0$, which would generate a different form for the likelihood ratio. We should perhaps also consider the possibility that we might have observed an exclusion, $S \neq T$, which would yield a likelihood ratio of 0. I have no clever suggestions to offer here, just the warning that it's all much more complicated than any of the position papers recognise.

4. Unknown parameter: Bayesian approach

Whatever one's philosophical attitude to the Bayesian approach, it does offer detailed instructions as to exactly how to proceed: we must eliminate the uncertainty in θ by integrating it out over an appropriate distribution. And (similarly to the frequentist estimation procedure introduced above) this elimination is to be done, not directly on $\mathrm{LR}(\theta)$, but on each term of the likelihood function $\mathbf{L}(\theta)$, before their ratio is finally taken — resulting in the overall "Bayes Factor", BF.

By Eq. (1) $\theta \perp\!\!\!\perp H$, so that we have a common conditional prior density for θ:

$$\begin{aligned}\pi(\theta \mid H_p) &= \pi(\theta \mid H_d)\\ &= \pi(\theta),\end{aligned} \tag{5}$$

say.

Also by Eq. (1), $(T, D) \perp\!\!\!\perp H$, so that (marginally after eliminating θ)

$$P(T=1, D=d \mid H_p) = P(T=1, D=d \mid H_d) \tag{6}$$

and

$$P(D=d \mid H_p) = P(D=d \mid H_d). \tag{7}$$

4.1. Bayes factor

The following analysis is in agreement with the more general description in the position paper by Ommen et al. [6].

4.1.1. Using full data

The Bayes factor, based on the full data $E : S=1, T=1, D=d$, is $\mathrm{BF} = P(E \mid H_p)/P(E \mid H_d)$, where the numerator is

$$\begin{aligned}P(E \mid H_p) &= \int P(E \mid \theta, H_p)\,\pi(\theta \mid H_p)\,d\theta\\ &= \int P(E \mid \theta, H_p)\,\pi(\theta)\,d\theta\\ &= \int \theta^{d+1}\,(1-\theta)^{N-d}\,\pi(\theta)\,d\theta\end{aligned} \tag{8}$$

and similarly the denominator is

$$P(E \mid H_d) = \int \theta^{d+2}\,(1-\theta)^{N-d}\,\pi(\theta)\,d\theta. \tag{9}$$

Thus

$$\mathrm{BF} = \frac{\int \theta^{d+1}\,(1-\theta)^{N-d}\,\pi(\theta)\,d\theta}{\int \theta^{d+2}\,(1-\theta)^{N-d}\,\pi(\theta)\,d\theta}. \tag{10}$$

[3] Or, to higher accuracy, $Ne + 1.5$.

[4] Much of this subsection is equally relevant for other ways of estimating $\mathrm{LR}(\theta)$.

4.1.2. Conditional on D

Alternatively, we can write

$$P(E \mid H) = P(S = 1, T = 1 \mid D = d, H)\, P(D = d \mid H).$$

Since by Eq. (7) the second term does not depend on H, the Bayes factor can be expressed as

$$\mathrm{BF} = \frac{P(S = 1, T = 1 \mid D = d, H_p)}{P(S = 1, T = 1 \mid D = d, H_d)}$$

where

$$\begin{aligned} P(S = 1, T = 1 \mid D = d, H_p) &= \int P(S = 1, T = 1 \mid \theta, D = d, H_p) \\ &\quad \times \pi(\theta \mid D = d, H_p)\, d\theta \\ &= \int \theta\, \pi(\theta \mid D = d)\, d\theta, \end{aligned}$$

and similarly $P(S = 1, T = 1 \mid D = d, H_d) = \int \theta^2\, \pi(\theta \mid D = d)\, d\theta$. Thus

$$\mathrm{BF} = \frac{\mathrm{E}^*(\theta)}{\mathrm{E}^*(\theta^2)}, \tag{11}$$

with E^* denoting expectation with respect to the posterior density π^* of θ based on the database alone: $\pi^*(\theta) := \pi(\theta \mid D = d) \propto \theta^d(1-\theta)^{N-d}\pi(\theta)$. This agrees with the formula as given in the position paper by Berger et al.—and of course with Eq. (10).

4.1.3. Conditional on (T, D)

As yet another alternative, we can write

$$P(E \mid H) = P(S = 1 \mid T = 1, D = d, H)\, P(T = 1, D = d \mid H).$$

By Eq. (6) the second term does not depend on H, so the Bayes factor can be expressed as

$$\mathrm{BF} = \frac{P(S = 1 \mid T = 1, D = d, H_p)}{P(S = 1 \mid T = 1, D = d, H_d)},$$

where

$$P(S = 1 \mid T = 1, D = d, H_p) = 1$$

and

$$\begin{aligned} P(S = 1 \mid T = 1, D = d, H_d) &= \int P(S = 1 \mid \theta, T = 1, D = d, H_d) \\ &\quad \times \pi(\theta \mid T = 1, D = d, H_d)\, d\theta \\ &= \int \theta\, \pi(\theta \mid T = 1, D = d)\, d\theta. \end{aligned}$$

Thus we have the alternative expression

$$\mathrm{BF} = 1/\theta^\dagger, \tag{12}$$

with $\theta^\dagger = \mathrm{E}^\dagger(\theta)$, the expectation being with respect to the posterior density $\pi^\dagger$ of θ based on adding the crime trace to the database: $\pi^\dagger(\theta) := \pi(\theta \mid T = 1, D = d) \propto \theta^{d+1}(1-\theta)^{N-d}\pi(\theta)$. Again this is easily seen to agree with Eq. (10).

Since $\mathrm{LR}(\theta) = 1/\theta$, and $\mathrm{BF} = 1/\theta^\dagger$, the value $\theta^\dagger$ might be considered as the Bayesian's estimate of θ. For example, when the prior is $\beta(a, b)$, we get $\theta^\dagger = (d+1+a)/(N+1+b)$, which, in the "weak prior information" limit $a, b \downarrow 0$, will be close to the frequentist estimate $\widehat{\theta}_p = (d+1)/(N+1)$ computed under the prosecution hypothesis H_p, which involves adding the crime trace to the database. This might be interpreted as supplying a Bayesian justification for "estimating" $\mathrm{LR}(\theta)$ by plugging in $\widehat{\theta}_p$. For small d this will still be considerably more favourable to the defence than using the estimate $\widehat{\theta} = d/N$ based on the database alone.

4.2. Uncertainty

The position papers by Ommen et al. [6], Berger and Slooten [1], and Biedermann et al. [2] make good arguments for *not* qualifying a Bayesian "estimate" of the likelihood ratio with an assessment of its uncertainty. Others in contrast, including van den Hout and Alberink [7], consider that this should be done, regarding $\mathrm{LR}(\theta)$ as a function of the random parameter θ. Cereda [3] gives a careful and readable account of relevant issues.

Philosophically I take the former view; but if we were to take the latter view, it is still far from clear what distribution for θ should be used. Taking our cue from the variant distributions for θ appearing in the different but equivalent analyses of Sections 4.1.1, 4.1.2, and 4.1.3 above, we could make a case for using any of: $\pi(\theta)$, the prior; $\pi^*(\theta)$, the posterior given $D = d$; or $\pi^\dagger(\theta)$, the posterior given $D = d, T = 1$ (note that none of these depends on which hypothesis, H_p or H_d, is assumed.). Use of the prior, π, would seem to ignore obviously relevant information, but it is less obvious how one should decide between π^* and $\pi^\dagger$. van den Hout and Alberink [7] use π^*. I would incline towards $\pi^\dagger$, since this is based on the greatest amount of information acceptable to both sides, and relates to an analysis where the only uncertainty is in just one of the two components (the denominator) of the Bayes factor. Yet another possibility, not motivated by the above considerations but in line with the "presumption of innocence", would be to use $\pi^{**}(\theta) = P(\theta \mid E, H_d) \propto \theta^{d+2}(1-\theta)^{N-d}\pi(\theta)$, which would be appropriate under H_d (when we have two additional instances of the marker to add to the database).

5. Discussion

I have described some statistical ambiguities and difficulties inherent in frequentist approaches to estimating a likelihood ratio, as well as problems bedevilling any attempt, be it frequentist or Bayesian, at assessing the uncertainty in such an estimate. None of the position papers has taken serious account of these. Until these issues are clarified, much of the debate will remain pointless and fruitless.

Acknowledgements

This work was funded by a Leverhulme Trust Emeritus Fellowship. I am grateful to Julia Mortera for helpful comments.

References

[1] C.E.H. Berger, K. Slooten, The LR does not exist, Sci. Justice 56 (5) (2016) 388–391.
[2] A. Biedermann, S. Bozza, F. Taroni, C. Aitken, Reframing the debate: a question of probability, not of likelihood ratio, Sci. Justice 56 (5) (2016) 392–396.
[3] G. Cereda, Bayesian approach to LR assessment in case of rare type match: careful derivation and limits, (2016), http://arxiv.org/abs/1502.02406v9.
[4] J.M. Curran, Admitting to uncertainty in the LR, Sci. Justice 56 (5) (2016) 380–382.
[5] A.P. Dawid, J. Mortera, Coherent analysis of forensic identification evidence, J. R. Stat. Soc. Ser. B 58 (1996) 425–443.
[6] D.M. Ommen, C.P. Saunders, C. Neumann, An argument against presenting interval quantifications as a surrogate for the value of evidence, Sci. Justice 56 (5) (2016) 383–387.
[7] A. van den Hout, I. Alberink, Posterior distributions for likelihood ratios in forensic science, Sci. Justice 56 (5) (2016) 397–401.