Science and Justice 56 (2016) 388–391

Contents lists available at ScienceDirect

Science and Justice

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

Position paper

The LR does not exist☆

Charles E.H. Berger [a,b,*], Klaas Slooten [a,c]

[a] Netherlands Forensic Institute, PO Box 24044, 2490 AA, The Hague, The Netherlands
[b] Leiden University, Institute for Criminal Law and Criminology, PO Box 9520, 2300 RA Leiden, The Netherlands
[c] VU University Amsterdam, Department of Mathematics, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Evidence interpretation
Likelihood ratio
Subjective probability

ABSTRACT

More than 40 years ago, De Finetti warned that probability is a misleading misconception when regarded as objectively existing exterior to the mind. According to De Finetti, probabilities are necessarily subjective, and quantify our belief in the truth of events in the real world. Given evidence of a shared feature of a trace and an accused, we apply this framework to assign an evidential value to this correspondence. Dividing 1 by the objectively existing proportion of the population sharing that feature would give that evidential value — expressed as a likelihood ratio (LR) — only if that proportion were known. As in practice the proportion can only be estimated, this leads some to project their sampling uncertainty — or precision — associated with the estimated proportion onto the likelihood ratio, and to report an interval. Limited data should limit our LR however, because as we will demonstrate the LR is given by what we *know* about the proportion rather than by the unknown proportion itself. Encapsulating all uncertainty — including sampling uncertainty of the proportion — our LR reflects how much information we have retrieved from the feature regarding the trace's origin, based on our present knowledge. Not an interval but a number represents this amount of information, equal to the logarithm of the LR. As long as we know how to interpret the evidence with a well-defined probabilistic model, we know what our evidence is worth.

1. Introduction

The title of our position paper paraphrases De Finetti's 'probability does not exist', which he stated in 1974 at the beginning of his two volume book "Theory of Probability", explaining [1]:

> *Probability …. if regarded as something endowed with some kind of objective existence, is …. a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.*

De Finetti meant that probabilities do not exist objectively outside the human mind, but only subjectively as a very useful concept to help us reason rationally in the face of uncertainty:

> [*I*]*n the conception we follow and sustain here, only subjective probabilities exist – i.e., the degree of belief in the occurrence of an event attributed by a given person at a given instant and with a given set of information.*

Much like probability, the likelihood ratio (LR) is not something real that exists in the external world, but rather a construct of the human mind, developed for reasoning under uncertainty. It is not intrinsic to the evidential material, but it is our assignment of a value to our observation in order to answer a certain question. Note that this observation is not all that we see, but rather those observed aspects that we chose to take into account in our interpretation.

1.1. Probability reflects information

A subjective probability $P(E|I)$ represents our rational degree of belief in E, given information I. It is therefore not a quantity determined only by E. Probability is conditional on our knowledge,[1] and therefore, if we obtain additional information to update our knowledge to I' we get $P(E|I')$ which need not be the same as $P(E|I)$.

For example, suppose that a patient is diagnosed with a disease, for which it is known that 50% of patients are ultimately cured (information I). What value would we assign for the probability that the patient will be cured? Obviously, this is 0.5, or 50%. Next, suppose that the disease can actually be further diagnosed into one of two variants: a severe

☆ 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.

* Corresponding author.

[1] And since we all may have different knowledge, probability is also personal.

http://dx.doi.org/10.1016/j.scijus.2016.06.005
1355-0306/

one where 20% of the patients are cured, and a milder one where 80% of the patients are cured. We assume both variants occur equally often. What value would we assign now for the probability that the patient can be cured, still based only on the information that he has the disease, without knowing whether it is the severe or the milder variant?

This would still be 50%. There are several ways to see this, perhaps the simplest one being that we have not learned anything new regarding our patient. The probability that we assign if we do not know which variant our patient has, is the same as our probability if we do not even know about the existence of these variants. So, based only on our information that the patient has the disease, we assign the probability of being cured as 50%, even if we know that the variants exist but not which one the patient has.

If we find out about the variant that the patient has (new information I'), we will revise our probability to either 20% or 80%. This does not mean that we should have told our patient that he has a probability of being cured of 50% plus or minus some interval. This would not have been of any use to the patient with the knowledge at the time (I). But if obtaining the new information is feasible we should additionally inform the patient in a way that is similar to pre-assessment: there are two variants and based on the future knowledge of which variant he has, the probability will be updated to either 20% or 80%. Note that the 50% still addresses the question as to whether he will be cured or not, while the pre-assessment informs a different question: whether it will be useful to obtain the information on the variant or not.

1.2. Likelihood ratios

The above also applies to the evidential value expressed as a likelihood ratio,[2] which is a quotient of probabilities. The evidential value is the value we can rationally assign to our chosen observations E of evidential material for the purpose of discriminating between two hypotheses H_p and H_d under consideration, on the basis of our information I and a probability model which is used to evaluate $P(E|H_p,I)$ and $P(E|H_d,I)$.

Thus, we have

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

Let us consider each of the necessary ingredients for this LR.

1) We need to choose a probabilistic model to assign a probability P. For example, we use Mendelian inheritance to model how DNA is passed on from one generation to the next, statistical independence of the loci for which we have data, etc.
2) We need to choose exactly which observations constitute E. For example, for DNA profiles we may consider E to be the set of assigned peak positions of the DNA profile of a trace, or alternatively, the peak positions and the peak heights. Including more information into E allows for a more informative interpretation but requires a more sophisticated model.
3) We need to define the issue at stake in the case in hypotheses H_p and H_d for the positions taken by the prosecution and defence, respectively. The hypotheses might be statements about e.g. the source of a trace, or the activity that led to a trace.
4) Our chosen background information I may, for example, consist of a population sample for which we have measured characteristics such as those in E (e.g., a database of DNA profiles).

Summarising, the evidential value of E depends of course on E, but also on the issues at stake in the case (H_p versus H_d), on our knowledge (I), and on our understanding of what data found under H_p and H_d look like, i.e., our probabilistic model to assign a probability P. The corresponding weight of evidence, defined as $\log_{10}(LR)$, can be viewed as an amount of information, which places LR and weight of evidence calculations in an information theoretic context [2].

1.3. Precision?

An LR assigns a value to our data — however much or little data there are — for the purpose of updating our relative belief in the truth of one proposition versus another, as a step towards decision making. Limited data and resulting sampling variation relates to the precision of e.g. our estimate of the real and existing proportion of a feature in the population, but not to the precision of an LR. Our LR is not determined by a proportion in the population that we could only determine if we had all the population data that we *do not* have, our LR is determined by the data that we *do* have. This fundamental difference becomes obvious when we take it to the extreme: if we have no data or the data is otherwise worthless, there would not be an infinite imprecision of the LR, but no imprecision at all. The LR would be 1, as the word 'worthless' already implied ($\log(LR) = 0$), and there is no reason to change our odds.

The concepts of *accuracy* and *precision*[3] are closely linked to that of measuring real and existing quantities. The introductory paper to the special issue however, applies these concepts to LRs [3]. It does not include accuracy in the present debate – which would in our opinion require a 'true value' to exist – and limits itself to precision. More 'precisely', the issue is the *variability* in a *set* of output LRs when varying the input parameters. Different data *should* lead to different LRs, but we would like to know how sensitive the LRs coming out of our system are to variability in the input. Note that such a sensitivity analysis is meant to characterise the *system* that generates the LRs, and does not characterise the *evidence* in a particular case. Whether or not to gather more data to narrow down the input parameters further is a different and separate issue from the issue defined by the propositions in the case.

For the measurement of real and existing quantities we routinely use terms like 'accuracy' and 'precision', and we use intervals to describe our uncertainty about the true value. We might measure the length of a desk to be:

$$L = 156 \pm 0.5 \text{ cm}. \quad (2)$$

Proponents of the reporting of LRs with intervals might look at the 156 value as being analogous to an LR in that sense that its 'precision' could be reported with an interval. But an LR is not like the length of a desk which is a real and existing quantity in the external world and can be estimated. It is the value of 0.5 that, analogous to an LR, attempts to convey the quality of the information we have, encapsulating all uncertainty.[4]

Similar to the situation for an LR, it would not make sense to further characterise the margin of error as 0.5 ± 0.1, which would lead to $L = 156 \pm 0.5 \pm 0.1$, and by extension of the argument to an infinite recursion. So in this analogy the role of the 0.5 number is analogous with that of the LR, which can be properly expressed as a single number without an interval.

Example 1. A shared characteristic; general equation.

In this example, to address the question of the source of a trace, we look at the evidential value of a characteristic that is shared between an accused and the trace. We treat the proportion of the shared

[2] We adhere to the terminology likelihood ratio rather than Bayes factor, as this is customary within the forensic community.

[3] We object strongly against the use of the much more general terms 'reliability' and 'validity' for the concepts referred to here.

[4] It makes sense to convey this measure with a single significant digit, and not report e.g. $L = 156 \pm 0.483$ cm.

characteristic in the population as a random variable X. The probability distribution for X is defined with prior parameters, and our expected value[5] for the proportion of the characteristic is $E[X]$. When the variance $\mathrm{Var}(X) > 0$, we are uncertain about the true value of X. If we obtain data, for example a population sample registering the characteristic, then we have to combine these data with the prior information to arrive at a better informed X.

Suppose — for the sake of argument — that we know the population proportion of the relevant characteristic, and we denote it by p. Let H_p state that the accused was the source of the trace whereas H_d states that some unknown person was the source, who we model to have or not to have the characteristic independently from the accused. Since under H_p we have seen the characteristic once and under H_d we have seen it twice, our $LR = p/p^2$ which simplifies to $1/p$.

Now we can move to the more general case where $\mathrm{Var}(X) > 0$. Under H_p we have seen the characteristic once, and our probability for that event is $E[X]$. Under H_d we have seen it twice, and we obtain $E[X^2]$ as our probability for that to happen. Thus, our LR is given by [4]:

$$LR = \frac{E[X]}{E\left[X^2\right]}. \tag{3}$$

We can work out this equation[6] further as:

$$LR = \frac{E[X]}{E\left[X^2\right]} = \frac{E[X]}{E[X]^2 + \mathrm{Var}(X)} = \frac{1}{E[X] + \frac{\mathrm{Var}(X)}{E[X]}}. \tag{4}$$

It is now immediately clear that a larger variance, keeping $E[X]$ fixed, reduces our LR. The largest LR for the same $E[X]$ is obtained when we are certain about the true value of X, in which case we obtain $\mathrm{LR} = 1/E[X] = 1/p$. The LR therefore is not a reflection of the rarity of the characteristic, but of *our knowledge* of that rarity.

Example 2. A shared characteristic; numerical example with Beta prior.

Suppose we model the probability of having a matching characteristic with a Beta distribution with parameters (a,b): $X \sim \mathrm{Beta}(a,b)$. This means that we expect the population proportion with the characteristic to be $E[X] = a/(a+b)$, and our uncertainty is reflected in the variance [5]:

$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}. \tag{5}$$

Our LR for the matching characteristic then simplifies to

$$LR = \frac{1+a+b}{1+a}. \tag{6}$$

This is the inverse of the expectation of a beta distribution $\mathrm{Beta}(a+1, b)$ and reflects the fact that the LR is determined by the probability that a randomly selected individual has the characteristic, given that the trace has the characteristic and based on our knowledge of the characteristic prior to the observation in the trace. If the parameters a and b are prior parameters not corresponding to a population sample, this is a subjective result depending on the scientist's opinion on the proportion of the characteristic in the relevant population and his or her uncertainty regarding that proportion. If we accumulate data on the prevalence in the relevant population, we will for example obtain m matching and n no-matches. It can be shown that our Beta distribution, if we take this into account, becomes $\mathrm{Beta}(a+m, b+n)$, and

$$LR = \frac{1+a+m+b+n}{1+a+m}. \tag{7}$$

Suppose again that a scientist observes a certain corresponding characteristic in trace and reference material. To assign an evidential value to this observation, the scientist might wonder what probability to assign to observing the same correspondence for the trace and some other reference material. Suppose that prior to the case the scientist had never seen this characteristic in 1000 cases ($m = 0$, $n = 1000$), but would also not be surprised if the actual proportion turned out to be just 1 in 100,000. The scientist may wonder if assigning an evidential value of 1000 implies that the proportion is believed to be 1 in 1000, and thus not 1 in 100,000?

From the above discussion we see that the LR reflects the proportion expected given that the trace has the characteristic. This is not the same as the proportion expected to have the characteristic prior to the case, because the variance plays a role. The value of the evidence (1000) assigned for the characteristic of uncertain relative frequency is the same as for a characteristic that is known with certainty to occur with relative frequency 1/1000. But the assigned evidential value is based on the limited information you have, and not on the hypothetical outcome of an experiment that has not been carried out. If the limited data can only justify the assignment of a 1000 to the evidential value, this cannot be translated into a claim that the proportion was believed to be 1 in 1000 prior to the case. It does mean, however, that we expect this proportion of the population to have the characteristic, based on the fact that the trace (i.e., the unknown offender) has it and on our knowledge prior to the case.

The forensic scientist has seen this characteristic for the first time in about 1000 cases, but does not rule out that it occurs in only one in 100,000 cases. How can we model this? The fact that we have not seen the characteristic in 1000 cases translates to $m = 0$, $n = 1000$. Suppose, for example, that we had used a Beta distribution with prior parameters $a = 0.1$ and $b = 0.1$. With these parameters, the expected value prior to the sample data for the proportion is $E[X] = a/(a+b) = 1/2$, but the probability that the true proportion is smaller than one in 100,000 (10^{-5}) can be shown to be about 16%. Incorporating the sample ($m = 0, n = 1000$) gives an updated $\mathrm{Beta}(0.1, 1000.1)$-distribution describing our probability distribution of the population frequency of the characteristic, which has an expected value of 1/10,002. The probability that the frequency is now less than one in 100,000 has increased to 66%. However, when we learn that the trace has the characteristic, the LR associated with a match is easily seen to be 910 (using Eq. (7)). This is because even though we believed it is very rare to see the characteristic based on our prior belief and the sample, we did see it in the trace. This observation raises our probability to see it again in some unknown person from 1 in 10,002 to 1 in 910.

The evidential value of 910 does not mean that we expected one in 910 people to have the characteristic prior to the case. In fact, prior to our observation in the trace we were very uncertain about the frequency. We set the expectation at 0.5 but also allowed 16% for the probability that less than one in 100,000 people have it. The change is because we have incorporated our experience ($m = 0$, $n = 1000$) and the one observation of the characteristic in the trace. Now that the trace has this characteristic, the evidence has the same value as a characteristic that one in 910 people are known to have with certainty.

The evidential value (not its 'certainty' or 'precision') can only be increased at the cost of getting more external information. For example, if we had taken 10,000 measurements resulting in $m = 0$, $n = 10.000$, we would assign $LR = 9092$ (using Eq. (7)).

If we had started with $a = 1$, $b = 1$ then we would have obtained $\mathrm{LR} = 501.5$ ($m = 0$, $n = 1000$) and $\mathrm{LR} = 5001.5$ ($m = 0$, $n = 10.000$).

[5] If the probability distribution of X is given by a probability density function $f(x)$, then the expected value is $\int_{-\infty}^{\infty} xf(x)dx$.

[6] We use $\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - E[2\mu X] + E[\mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - E^2[X]$, substituting $E[X] = \mu$.

Thus, inevitably the LR partially depends on the scientist's opinion on the prevalence as reflected by the parameters a and b. This may be viewed as a drawback by some, as conclusions are not purely data driven. It may also be seen as an unavoidable reflection of the fact that data processing takes places in a context. However, as more data is gathered the resulting LRs starting with different prior parameters will end up closer to each other.

Summarising, we see that if we can formulate a probability model and an appropriate prior, we obtain the LR by integration of the likelihoods under H_p and H_d. Less certainty in the prior, or fewer data at our disposal, automatically leads to a weaker LR. This is the consequence of the weaker information position that we have in these cases. The fact that the LR may change if we obtain more information does not mean that we should adapt our LR, but only that we should consider whether it is worthwhile and feasible to obtain this extra information.

Conclusion

The title of this paper refers to the fact that — like probabilities — likelihood ratios are conditional on information, and do not exist in the external world. Therefore, LRs are personal and subjective; qualifications which most scientists are not very comfortable with. However, as more data is obtained, the use of subjective prior information becomes less relevant and different scientists will end up in close agreement. Especially in situations where there is little data, it is not surprising that there can be substantially different subjective assessments (see also the discussion of subjective probability in [6]). Since the LR depends on the evidence, the probability model (including choices of priors), the hypotheses, and the information that has been considered, the forensic scientist needs to grant access to all these ingredients in order to be transparent.

We have argued conceptually that all uncertainty should be encapsulated in the evidential value assigned itself, and not in an interval which would be another measure of uncertainty on top of the LR. One can be uncertain about a proportion of a population in which a certain characteristic occurs, and words like ‘precision’ and ‘accuracy’ do apply for such real and existing quantities, that have a true value. This uncertainty however, does not translate into an uncertainty of ‘the LR’. We have demonstrated mathematically how a paucity of data does not lead to an uncertain LR or an interval, but will increase the variance, and limit the value of our LR.

We have thus shown that there is no conceptual need for reporting an interval for an LR, and we have not seen a rational criterion for choosing such an interval around an LR. As a separate argument we think there is no rational way to use such an interval, if presented with one. We invite[7] those that propose to report an LR with an interval to demonstrate how one should update one's prior odds into posterior odds, based on that interval, for the purpose of decision making. Taking into account the benefit of the doubt for the accused should be done when the trier of fact has assigned probabilities to the hypotheses of guilt and innocence. It should not be done before, by choosing the LR in the interval that is closest to 1, and on separate items of evidence.

Further thoughts

In this position statement we have not addressed model uncertainty: we have assumed that we have an appropriate model for the evaluation of the evidence. When modelling error or uncertainty needs to be dealt with, there may be circumstances where the required integration over nuisance parameters may be hard or impossible in practice. In such a case we might not feel able to report a single LR. Reporting a single value does not mean that there is no uncertainty about the model parameters, but it does mean that there is no uncertainty about how to deal with the evidence. It remains the duty of the forensic scientist to judge whether this is the case.

If it is not the case, we might choose to report a verbal statement, or even ask the requester whether the various values could make a difference for the decision on the hypotheses. We should remember that even though the forensic community is mostly concerned with the assignment of the LR, it is not the central quantity of interest in the fact finding process. The trier of fact is really interested in the posterior odds on the hypotheses. If our evidence is the only evidence where no single value can be reported, the trier of fact might be satisfied with any of the values, if they lead to the same decision.

Another situation in which we might report more than one LR is where not all members of the alternative population have the same probability of having the characteristic that a trace and an accused share. In the DNA context, brothers of an accused will have a higher probability to have the accused's DNA profile than unrelated individuals. If such family members are among the alternative trace donors, then an LR (for the source-level question) can only be computed when a prior probability for them to have left the trace can be assigned, which is rarely the case. It is possible to compute separate LRs for the various kinds of alternative donors. The uncertainty that we deal with in this way is a lack of knowledge regarding the more specific alternative scenario H_d. This amounts effectively to setting up several alternative hypotheses and the computation of an LR for each, and not in the computation of an interval on the LR.

References

[1] B. De Finetti, Theory of Probability, Vol. 1 (John Wiley & Sons, New York, 1974), p. x and pp. 3–4.
[2] I.J. Good, Studies in the history of probability and statistics. XXXVII A.M. Turing's statistical work in World War II, Biometrika 66 (1979) 393–396.
[3] G.S. Morrison, Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate, Sci. Justice 56 (2016) 371–373.
[4] D.J. Balding, 0Weight-of-Evidence for Forensic DNA, Profiles, (John Wiley & Sons), Chichester 2005, p. 38.
[5] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, second ed. Vol. 2 Wiley, 1995 (Chapter 21: “Beta Distributions”).
[6] A.P. Dawid, Probability and Proof, on-Line Appendix to Analysis of Evidence (2nd Ed.), by T.J. Anderson, D.A. Schum, W.L. Twining. (Cambridge University Press, Cambridge, 2005).

[7] Charles Brenner has put forward a similar challenge in 2000 on http://dna-view.com/noconfid.htm#challenge and reports that no scientist has come forward to explain how to rationally use an interval (for profile frequency) in decision making.