

Forensic Literature on higher order LR

August 2023

1 Purpose of this document

- Summarize existing forensic science literature on higher order LR
- Outline the structure of the paper/chapter on higher-order legal probabilism.

2 Relevant literature summaries

2.1 What should a forensic practitioner's likelihood ratio be? (Geoffrey Stewart Morrison, Ewald Enzinger)

Authors believe that LR has a true but unknown value. More specifically, once the following items are specified, the LR should be fixed and have a true value:

- hypotheses to compare
- definition of relevant population
- type of measurements
- sample of relevant population

Quotation:

Once the forensic practitioner has stated what they understand to be the relevant circumstances of the case and the prosecution and defence hypotheses they have adopted, including the definition of the relevant population, and they have stated the type of measurements they will make on the known-origin sample, the questioned-origin specimen, and a sample of the relevant population, their task is to calculate an estimate of a likelihood ratio which has a true (but unknown) value. Without these specifications there is no true likelihood ratio value to be estimated. (p. 374)

So forensic scientists need to calculate the precision of their estimate of the LR:

we believe that the task of the practitioner is to calculate an estimate of the true but unknown value of the likelihood ratio given a specified relevant population and one or more specified types of measurement. (p. 375)

Now, the question is, how to model and report to the fact-finders the precision of the LR estimate? Using intervals would be too conservative, so distributions of LRs are better:

A trier of fact who always used the bound of a credible interval closest to the neutral value of 1 would end up being very conservative, probably more conservative than they anticipated.

A better way for a trier of fact to handle imprecision would be for them to consider the distributions of the likelihood ratios in the forensic practitioner's assessment of the reliability of the system. (p. 377)

So the trier of fact would combine (log) prior distribution with (log) LR distribution and arrive at (log) posterior distribution.

How would the triers of fact make decisions? The proposal is cumbersome and not completely clear. Authors rely on the proportion of the distribution that is above the neutral value of zero (using log distributions). So if more than 95% of the (log) posterior distribution is above 0, this would meet decision threshold of 95% (say needed in criminal cases).

In the example shown in Fig. 1, 96% of the log posterior odds distribution is greater than 0. If, before the beginning of the presentation of evidence the trier of fact had decided that their decision threshold would be 95%, then they would now conclude in favour of the prosecution hypothesis, but if they had decided on a threshold of 99% they would conclude in favour of the defence hypothesis. (p. 378)

Question: how do these “area based thresholds” relate to the more standard “probability thresholds”? Is a 95% area based threshold the same as a 95% probability threshold? Probably not. So what is the meaning of these area based thresholds?

2.2 Using sensitivity analyses in Bayesian Networks to highlight the impact of data paucity and direct future analyses (Duncan Taylor, Tacha Hick, Christophe Champod)

There is no true value of the LR since LRs are not parameters to be estimated. But, it is still useful for forensic scientists to use *distributions of LRs*. How?

- First, the distribution of LRs convey information about the *robustness* of the LR assessment, that is, whether or not it is going to change (or would have been different) in light of other data.
- Second, the distribution guides decisions whether more data is needed.

They use Bayesian networks with different nodes to carry out the sensitivity analysis. Relevant quotations:

Because LR_s depend on the data used to inform probabilities, it goes without saying that using different data, will lead to different LR_s. (p. 403)

This means that our evaluation will be considered robust, if the system, informed by a different set of experiments, leads to similar LR values. Thus, exploring the sensitivity of the LR to the underlying data allows us to explore whether our knowledge is sufficient to ensure robust conclusions. Sensitivity analysis is achieved by simulating cases under a range of datasets and exploring the impact of this on the output. (p. 404)

[sensitivity analysis] provides the distribution of LR_s shown in Fig. 2 ... for which the 50%, 5% and 1% quantiles LR_s are 160, 40 and 30 respectively. This distribution shows the impact of the data on our evaluation. We wish to stress here that there is no “true” value for a LR so this distribution does not show the uncertainty of the LR, but how robust (or sensitive) it is depending on the data used. Each data point on the distribution only represents a LR computed under the conditioning of different data. (p. 404)

The benefits of sensitivity analyses are two-fold. Firstly, it demonstrates if the basis of our opinion is robust and what the impact would have been if another practitioner had chosen different casefiles from which to collect data. The second benefit is that the BN can be probed in order to determine which nodes have the most impact on evaluation. This can direct us to which experiments would provide the most benefit if we were to collect more data to inform the conditional probability tables of the BN. (p. 404)

They make a distinction between scientific assessment of the evidence (say DNA match) and reporting. They do not think the full distribution of LR_s should be reported in court, even though they think the distribution is useful for forensic scientists.

We are of the opinion that sensitivity analysis can be useful for scientists to decide whether or not they should perform more experiments or if they have sufficient knowledge to report. But, we believe that it is their duty to decide the value of their LR as we will explain below.

When it comes to reporting, we need to strive to inform in a fair but useful manner. Fig. 2 (or at least some summary of it) should no doubt find its place in the case file and be amenable to review by any forensic scientist. We have doubt that the results of sensitivity analyses will help in a statement that can be used in court. (p. 407)

They compare two cases, i.e., LR of 1 with few data versus LR of 1 with lot of data:

the first the LR of 1 reflects upon a “state of ignorance” where we chose to assign equal opportunity to each state whereas the second case leads to an uninformative system but based on substantial knowledge that the states are equally likely. Pragmatically speaking, the first case tells us that it would be wise to invest on data acquisition (p. 407)

But they also seem somewhat undecided on what is best options in the end:

The authors of the present paper are still exploring and debating various options on how to convey the sensitivity that the scientist is prepared to associate with his/her LR (p. 407-408)

2.3 The meaning of justified subjectivism and its role in the reconciliation of recent disagreements over forensic probabilism (A. Biedermann, S. Bozza, F. Taronia, C. Aitken)

The authors simply say that probability is a single number and exclude the need for intervals (or distributions of LRs):

The main theme in this article collection is that of the construction of intervals for probabilities and likelihood ratios. Consideration of this theme requires one to ask what probability theory says on this topic. Briefly, the answer to this is that ... in their most fundamental form (axiomatic definition), the three rules of probability, say nothing about Intervals, Precision, Application of the theory, Assignment of probabilities. (p. 478)

2.4 The LR does not exist (Charles E.H. Berger, Klaas Slooten)

Authors state that probabilities (and LRs) reflect information that we do have. They consider the following two cases probabilistic equivalent (50% cure success in both):

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 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. (p. 388-389)

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 (p. 389)

They do admit that different data will yield different LRs, so they seem open to sensitivity analysis but they say this is a different matter than LR assessment. And they do not think that intervals are the right modeling choice.

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 LR, 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. (p. 389)

Authors argue that all (most?) uncertainty – including variability based on little data – can be encapsulated in the LR. If we know the population proportion, then LR is simply $\frac{p}{p^2} = \frac{1}{p}$. But what if the variance is greater than 1? Then the LR is as follows (derivation on p. 390):

$$LR = \frac{1}{E[X] + \frac{VAR[X]}{E[X]}}$$

So it is clear that the greater the variance, with fixed $E[X]$, the lower the LR. The LR is greatest, when variance is zero, $LR=1/p$. So they say:

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. (p. 391)

Authors issue a challenge to defenders of the view intervals-for-LR (how should updating go?):

We invite 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.

Authors are however open to the fact that model uncertainty cannot be captured by one single LR. So they seem open to reporting multiple LRs in some cases:

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. (p. 391)

Here is a specific example when two LR might be needed, but this seems different from issue of meta-uncertainty:

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. . . . 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. (p. 391)

2.5 Posterior distributions for likelihood ratios in forensic science (Ardo van den Hout, Ivo Alberink)

Authors provide a way to calculate posterior distributions of LR (highly technical, also code).

Using the posterior likelihood ratio is not frequentist as sampling from a posterior is required, but it is also not fully Bayesian since it does not use the Bayes factor for hypothesis testing. (p. 400)

Interestingly, they only consider one item of evidence and leave open what to do with two items or more:

In this paper the situation is considered in which there is only one piece of evidence. If there is more than one piece of evidence, a posterior distribution may be determined of the LR of the combination of the evidence. This topic may be explored elsewhere. (p. 400)

2.6 An argument against presenting interval quantifications as a surrogate for the value of evidence (Ommen, Saunders, Neumann)

Authors present a compelling case against using intervals for LRs for presenting the values of the evidence. They distinguish between:

- uncertainty in LR/BF due to numerical error (e.g. using Monte Carlo simulation)
- uncertainty in LR/BF due to sampling variability

They agree that interval estimation is necessary when numerical error is involved. When it comes to sampling variability, though, they argue that:

- using the end point (upper or lower) of the intervals will give rise to an incoherent decision process
- using the point in the interval closer to one (neutral point) will be biased in favor of one side or another (prosecution or defense)
- using the midpoint of the interval will overvalue the value of the evidence

They think sensitivity analysis is more appropriate or simply just a single value:

To conduct a rigorous formal Bayesian sensitivity analysis, the expert would require some basic knowledge about the priors for the parameters that other experts tend to use (which is likely outside the scope of the relevant expertise of a typical forensic expert). Additionally, we do not suggest presenting intervals to a decision-maker for capturing this type of variability since the results of sensitivity analyses are typically more complex than can be expressed through an interval. (p. 386)

Our foremost recommendation is that when presenting forensic evidence to a decision-maker, a single value of evidence needs to be presented from which a logical and coherent decision can be made regarding the two competing hypotheses/models (p. 386)

2.7 Admitting to uncertainty in the LR (Curran)

Nice paper illustrating, in simple terms, why there is uncertainty in the random match probabilities or likelihood ratios. Strangely, though, the paper argues that, in the best case scenario, we would use Bayesian factors, to encapsulate all uncertainty.

In late 1997 I began to consider the problem of addressing sampling uncertainty when assessing random match probabilities, or likelihood ratios for DNA evidence, especially in complex mixtures. The issue of concern ... was that the allele frequencies used in calculating genotype probabilities typically arose from very small samples of individuals (p. 380)

DNA experts testifying in the 1990s and 2000s were routinely asked to justify how they could give a random match probability in the order of one in a billion ... from such a small sample of people. An astute lawyer would also ask "If I took another sample of size 200, would this figure change?" The single most effective response to this question is "Yes, and my method for assessing this probability has already taken this into account." (p. 380-381)

He makes a point that others (e.g. Berger/Slooten) disagree with:

If we regard the likelihood ratio as function of parameters which have a distribution, then this means that the LR will also have a distribution

But even admitting that LRs will have a distribution, still he thinks that, ideally, the solution is not distributions of LRs, but rather, the Bayes factor:

Bayes factors are what we ultimately want (p. 381)

3 Structure of paper/chapter

As I see it, the argument of the paper should be structured around three key points:

- **First:** The debate among forensic scientists and proposals on the table
- **Second:** The higher order approach as a novel proposal, how it works, etc.
- **Third:** Why the higher order approach is better than existing approaches

This is standard structure for any scholarly paper. Below each point is developed more precisely.

3.1 Charitable reconstruction of the debate in the literature

- The paper should begin with a **charitable reconstruction** of the debate in the forensic science literature. (See, e.g., the 2016 special issue of *Science and Justice*, “Special issue on measuring and reporting the precision of forensic likelihood ratios”, edited by G.S. Morrison. All the papers are in a special folder called 2016ScienceJustruce-SpecIssue-LR etc.) The Lund slides try to do that, but – in retrospect – they do not do succeed.
- Here is a charitable reconstruction:
 - (i) Everybody agrees that LR_s (as averages, expectations) do not convey all the underlying uncertainty. There is uncertainty originating with the sample size (call it, *data uncertainty*) and there is uncertainty about the model assumptions (call it, *model uncertainty*) and there is uncertainty about the amount of information considered (call it, *information uncertainty*). Call these additional forms of uncertainty *meta-uncertainty*.¹ This tripartite distinction is not clearly stated in the literature, but data uncertainty and model certainty are certainly recognized. Some think that data uncertainty – data sets can be larger or smaller – is not captured by the LR itself, while others (Berger/Slooten) believe that this uncertainty can be captured as a function of Variance VAR in the denominator of the LR. Still, even Berger/Slooten admit that model uncertainty cannot be captured by the LR and that sometimes multiple LR_s will need to be considered.
 - (ii) The question, then, is to what extent the meta-uncertainty that is not captured by the LR should be communicated by the experts to the fact-finders besides the first-order uncertainty captured by LR_s. This is subject to debate. Some think that fact-finders should be made aware of this extra level of uncertainty, while others (Taylor/Hick/Champod) think that considerations about meta-uncertainty should only be entertained by

¹An interesting philosophy paper that might be good to reference on meta-uncertainty is “Meta-uncertainty and the proof paradoxes” by Katie Steele and Mark Colyvan, *Philosophical Studies*, in particular sections 4 and 5.

experts. This is partly a psychological question: to what extent can fact-finders understand issues of meta-uncertainty? Isn't it worth at all to present questions of meta-uncertainty to them? But, suppose the fact-finders are ideally rational and tasked with making decisions about the ultimate issues. In that case, the question is whether meta-uncertainty should influence decisions about the ultimate issue.

- (iii) If we think that meta-uncertainty should be conveyed to the fact-finders – as many, though not all, in the literature think – the next question is, how should this meta-uncertainty be conveyed? What is the best way to convey it? This is where we encounter significant disagreement.
- (iv) Many proposals are on the table: describe data collection and model assumptions (Taroni/Bozza); place confidence intervals over LRs; model possible differences in terms of robustness (Biederman); precision estimates using distribution of LRs; model robustness using distribution of LRs.
- (v) At the mere technical level, the difference is between confidence/interval of LR versus distribution of LRs. Besides these two options, we have the informal approach of just describing data collection and model assumptions (Taroni/Bozza). But there is a disagreement about how to interpret distributions of LRs: some think that they are a measure of robustness (Taylor/Hick/Champod) while others that they are a measure of precision (GS Morrison, Enzinger).
- (vi) The interval approach is criticized because LRs are not parameters to be estimated.

3.2 The higher order approach

The next section of the paper should outline the higher approach as a novel answer to the question of how to model (at least part of) the meta-uncertainty that cannot be modeled by LRs alone. Some key points:

- The **mathematics** should be clear. Are higher order probabilities well-behaved probabilities? Is the mathematics all working properly? How do we show that? (The paper by Philip Dawid “Forensic likelihood ratio: Statistical problems and pitfalls” in the 2016 special issue might be a good starting point)
- How much of the meta-uncertainty is **captured** by distribution over probabilities and how much of the meta-uncertainty is still **left out** of the formal model?
- Parallels with the debate between precise/imprecise probabilism in philosophy. (Not sure if this is really necessary though.)
- Good to show that everything that can be done with LRs (e.g. combine them) can be done using higher order LRs and higher order Bayesian networks. So the higher order approach combines reasoning about first-order uncertainty with reasoning about meta-uncertainty.

3.3 Comparing higher order approach to existing approaches

In what way is the higher order approach different and better than other existing approaches for conveying meta-uncertainty to fact-finders?

- One key argument in favor of the higher-order approach concerns how different pieces of evidence (with different meta-uncertainties) can be combined. No one of the existing approaches has anything to say about how to combine different pieces and keep track of the meta-uncertainty. For example, it is not clear how the resilience approach or the interval approach could do that correctly (explain why).
- Respond to objections against the higher order approach, e.g. higher order probabilities do not exist or make no sense mathematically.