Properties of likelihood ratios. Lecture I.3

Thore Egeland[(1),(2)] (thanks to Klaas Slooten)

(1) Norwegian University of Life Sciences, (2) Oslo University Hospital

Workshop. Monterrey, Mexico, Nov 11-13, 2017

Contents

- LR definition. Implications for the hypotheses under consideration
- Likelihood ratios as random variables: distributions
- Use and abuse of LR distributions

Likelihood Ratio (LR)

Definition of the LR

$$LR_{H_1,H_2}(E) = \frac{P(E \mid H_1)}{P(E \mid H_2)},$$

depending on

- The hypotheses H_1, H_2 under consideration
- The data E that we are considering

Meaning of the LR

- $P(E \mid H)$ is the probability to get E, if hypothesis H is true
- It is also called the likelihood of the hypothesis, given the evidence E
- The LR says how much better the explanation for E offered by H_1 is, compared to the explanation offered by H_2.
- The individual likelihoods $P(E \mid H_i)$ do not allow for any inference considered on their own: the issue is not to predict the evidence (as $P(E \mid H)$ does) but to see which mechanism explains it better

Application: Bayes rule

$$\frac{P(H_1 \mid E)}{P(H_2 \mid E)} = LR_{H_1,H_2}(E)\frac{P(H_1)}{P(H_2)},$$

$$\text{posterior odds} = \text{LR} \times \text{prior odds}.$$

Subjective probability

Let $P(H)$ be the maximum you want to pay for a bet that pays out 1 if H turns out be true; then these $P(\cdot)$ obey the usual probability rules, but they apply to non frequentist situations such as one-time events

Notion of probability in Bayes rule

Often the probabilities on the hypotheses can only sensibly be regarded as subjective; but for the LR calculation a frequentist interpretation is often possible in forensic DNA (kinship) analysis.

LR as support for a hypothesis ?

Interpretation

Often, one writes LR instead of $LR_{H_1,H_2}(E)$ and interprets

- $LR > 1$ as support for H_1
- $LR < 1$ as support for H_2

But..this is relative, not absolute

- This is only correct relative to the other hypothesis (Bayes rule)!
- $LR_{H_1,H_2}(E) > 1$ does not imply that $P(H_1 \mid E) > P(H_1)$
- $LR_{H_1,H_2}(E) > 1$ does not imply that $P(H_2 \mid E) < P(H_2)$.

Example 1: sibling test

Suppose two individuals may be siblings.

- H_1: sibs
- H_2: unrelated
- E: two identical profiles from individuals 1 and 2
- Obviously $LR_{H_1,H_2}(E) > 1$. Say, 10^6.

There is support in favour of these individuals being siblings, but only when compared to the alternative that they are *unrelated*. In this case, a third hypothesis comes to mind: the individuals are (genetically) identical:

- H_3 : monozygotic twins
- Obviously $LR_{H_3,H_2}(E) > 1$. Say, 10^{15}.

Uniform prior

Table 1: Prior probabilities, likelihoods and posterior probabilities under the three scenarios

	H_1	H_2	H_3
Prior probability	$1/3$	$1/3$	$1/3$
Scaled likelihood	10^6	1	10^{15}
Posterior probability	10^{-9}	10^{-15}	≈ 1

So indeed $LR_{H_1,H_2}(E) > 1$ but $P(H_1 \mid E) < P(H_1)$.

Dormant H_3

Table 2: Prior probabilities, likelihoods and posterior probabilities under the three scenarios

	H_1	H_2	H_3
Prior probability	0.4999995	0.4999995	10^{-6}
Scaled likelihood	10^6	1	10^{15}
Posterior probability	0.00049975	$4.9975 \cdot 10^{-10}$	0.9995

Again $P(H_1 \mid E) < P(H_1)$.
The hypothesis H_3, even though a priori extremely unlikely, is such a better explanation that it gets almost all posterior probability.

Conclusions based on this example

- $LR_{H_1,H_2}(E) >> 1$ does not mean that H_1 is a 'good' hypothesis; it may be that H_1 is a bad hypothesis and H_2 is an even worse one.
- The LR only allows for a statement relative to the two hypotheses it considers.
- It is therefore very important to choose *all* relevant hypotheses.
- If considered hypotheses are not exhaustive then only posterior odds on the considered hypotheses, but no posterior probabilities can be calculated.
- Often, the relevant hypotheses are clear from the context. But the evidence may be strongly in favour of a 'dormant' (=not considered hypothesis with very low prior) hypothesis and this can be important.

We have seen that

- $LR > 1$ does not imply evidence in favour of H_1, in an absolute sense, so we have no reason to believe more in H_1 than we did before learning E.
- Only our belief in H_1 relative to our belief in H_2 has increased, but both may have gone down or up; we can't know unless these two hypotheses are exhaustive.
- In the examples, both $P(H_1 \mid E)$ and $P(H_2 \mid E)$ were smaller than prior to E.
- Perhaps even more tricky: if $LR > 1$ and $P(H_1 \mid E) > P(H_1)$ this does not imply that we have any reason to believe less in H_2 than we did before seeing E. We may have $P(H_2 \mid E) = P(H_2)$ even if $LR_{H_1,H_2}(E) > 1$ and $P(H_1 \mid E) > P(H_1)$.

Example 2

- A deck of cards is face down on the table [2].
- One card is turned over: ace of spades.
- H_1: trick deck with only ace of spades.
- H_N: normal deck.
- Then $LR_{H_1,H_N} = 52$: there is evidence that the deck is tricked as containing only aces of spades, rather than being an ordinary deck
- But if we define all $H_1, H_2, \ldots, H_{52}$ as trick deck hypotheses then (with logical assumptions) $P(H_N \mid E) = P(H_N)$: there is no evidence that the deck is tricked.

Summary

- There is no evidence that the deck is tricked.
- We have only learned that *if* it is tricked as a deck with identical cards, then that must ace of spades.
- The ace of spades hypothesis H_1 is defined *in terms of the evidence*. Therefore by construction there is support for it, but in fact there is no evidence against its alternative.
- Testing hypotheses that are formulated in terms of the data is perfectly possible in this Bayesian framework, but results are easily misinterpreted.
- This is reminiscent of results from database searches, we may return to this later.

LR as random variable

- It is natural to consider that the evidence (i.e., genotypes) is the result of a stochastic process, and that the evidence that we have, could have been different and will typically be different in a subsequent case with identical hypotheses.
- Under each hypothesis, we can calculate with which probability we see all possible genotypes.
- Thus, under each hypothesis we can also calculate the probability to see a certain LR, or the probability to obtain at least a certain LR, etc.,: we can obtain the distribution of the likelihood ratio.
- Many properties of the resulting LR distributions are very general but some are at first glance surprising.

LR of LR is LR

Forensic hypotheses

We will assume that, for all possible evidence E,

$$P(E \mid H_p) > 0 \Rightarrow P(E \mid H_d) > 0. \tag{1}$$

This is a natural assumption: the defense hypothesis can explain anything that the prosecution hypothesis can explain.

Basic property

For all $x \geq 0$ we have

$$P(LR = x \mid H_p) = xP(LR = x \mid H_d), \tag{2}$$

i.e., for x that can be obtained as a LR, "the LR of the LR is the LR",

$$\frac{P(LR = x \mid H_p)}{P(LR = x \mid H_d)} = x. \tag{3}$$

as proved in [4].

Example: Sibling Index

Sample of 1,000,000 SI for sibling pairs (H_p) and for unrelated pairs (H_d)

Table 3: Observed probabilities for LR's in intervals

	$[10^{-5}, 10^{-4}]$	$[10^{-4}, 10^{-3}]$	$[10^{-3}, 10^{-2}]$	$[10^{-2}, 10^{-1}]$	$[10^{-1}, 1]$
H_p	$9 * 10^{-6}$	0.000087	0.000677	0.003334	0.011895
H_d	0.229912	0.25051	0.185839	0.096578	0.037755
Quotient	0.0000391454	0.000347292	0.00364294	0.0345213	0.315058

Table 4: Observed probabilities for LR's in intervals

	$[1, 10]$	$[10, 100]$	$[100, 10^3]$	$[10^3, 10^4]$	$[10^4, 10^5]$
H_p	0.032832	0.069986	0.118859	0.159493	0.175519
H_d	0.010826	0.002375	0.000401	0.000059	$5 * 10^{-6}$
Quotient	3.0327	29.4678	296.406	2703.27	35103.8

Average LR under H_d is one

If H_d is true we have, regardless of what H_p is,

$$E[LR \mid H_d] = 1,$$

since

$$\begin{aligned} E[LR \mid H_d] &= \sum_x xP(LR = x \mid H_d) \\ &= \sum_x P(LR = x \mid H_p) \\ &= 1 \end{aligned}$$

- This is an average. It may not even be possible to obtain $LR = 1$
- It reflects the fact that it is never possible to prove H_p, and that false positives ($LR > 1$ under H_d) exist
- As a consequence, if there are fewer false positives they must become stronger.

Bound on false positives

While false positives always exist, there is a bound to how many of them can exceed threshold x:

$$P(LR \geq x \mid H_d) \leq \frac{1}{x},$$

since

$$\begin{aligned} P(LR \geq x \mid H_d) &= \sum_y \mathbf{1}_{y \geq x} P(LR = y \mid H_d) \\ &= \sum_y \mathbf{1}_{y \geq x} \frac{1}{y} P(LR = y \mid H_p) \\ &\leq \sum_y \mathbf{1}_{y \geq x} \frac{1}{x} P(LR = y \mid H_p) \\ &\leq 1/x \end{aligned}$$

So, a LR of at least x can not happen more than one in x times if the second hypothesis is true

Importance sampling

- Because $PL(LR \geq x \mid H_d) \leq 1/x$, large sample size needed for big x
- But since $P(LR = x \mid H_p) = xP(LR = x \mid H_d)$, we can also select LR's from the H_p distribution and correct for the fact that we saw it too easily in that distribution
- Importance sample estimate: take $r_1, \ldots, r_N$ values from LR under H_p, [1]

$$P(LR \geq x \mid H_d) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{r_i \geq x} \frac{1}{x}.$$

Applications

LR distributions allow to estimate/compute

- Power: for a certain threshold t: what is the probability (true positive rate) $TPR(t) = P(LR \geq t \mid H_p)$?
- False positive rate: what is the probability $FPR(t) = P(LR \geq t \mid H_d)$?
- Bycatch: for a third hypothesis H_x, what is $P(LR \geq t \mid H_x)$?

Useful for

- Database uptake decisions
- Method assessment: does it generate strong enough evidence often enough?
- False positive and false negative rates

LR distribution not relevant within the case

- The odds on the hypotheses are updated with the LR from the obtained evidence via Bayes rule
- It is irrelevant what evidence could have been obtained, it only matters what the obtained evidence is!

Abuse of LR distributions

- Attach additional consequences for the two considered hypotheses to the LR depending on where it sits in the distribution or a proxy, e.g., based on the number of loci considered

Unusual LR can alert to third possibility

- Suppose $SI = 10$ is obtained from a pairwise comparison. What can we conclude?
- This increases the odds on H_p (sibs) versus H_d (unrelated) 10 times, irrespective of the number of loci that were typed.
- Is it important that this is an unusually small LR for sibs and unusually large for unrelated individuals?
- It is not for the considered hypotheses. If the considered individuals can *only be* either full siblings or unrelated, then it is irrelevant what LR values we could also have obtained, i.e., the distribution of the LR is irrelevant.
- However, if there could be other explanations possible, such as half-sibs, then the fact that the LR is 'unusual' may be an alert to consider other hypotheses.

Exceedance probabilities

Let

$$q_{10} = P(LR \geq 10 \mid H_d)$$

and

$$r_{10} = P(LR \leq 10 \mid H_p)$$

Number of loci	q_{10}	r_{10}
5	0.37	0.012
10	0.13	0.0066
15	0.050	0.0030
32	0.0022	0.00017

Table 5: Exceedance probabilities for $SI = 10$ based on various numbers of loci

On 5 loci, the result is less unexpected than on 32. Especially in the latter case, also consider other hypotheses.

Paternity testing: RMNE versus LR

Two approaches to paternity testing

- A man AF is possibly the father of a child C: genotypes consistent
- Method 1: compute the proportion of men in the general population that could genetically be the father of C (Probability called Random Man Not Excluded)
- Method 2: compute the LR in favour of paternity versus unrelated

Differences between these approaches

RMNE

- Does not use the genotype of the alleged father, only that of the child
- Can be computed prior to having any alleged father
- E.g., to judge whether to compare it to a database (how many possible fathers to expect)
- Can view it as $RMNE = P(LR > 0 \mid H_d)$

LR

- Uses all available genetic information on both individuals
- Is therefore better informed than the RMNE

Unified framework

H_1 AF is the father of C

H_2 AF and C are unrelated

E The genotypes of AF and C

E' The genotype of C and the fact that AF can genetically be the father of C

Then

$$1/RMNE = \frac{P(E' \mid H_p)}{P(E' \mid H_d)},$$

so the inverse RMNE is also a LR, corresponding only to the knowledge of a non-exclusion.

RMNE versus LR

If we do take into account the genotype of the alleged father,

$$E[LR \mid H_p] \geq 1/RMNE,$$

Thus,

- In expectation, the LR will be more than 1/RMNE, expressing that on average, taking more information into account yields stronger evidence in favour of the correct hypothesis
- The RMNE itself is also a likelihood ratio and an unbiased method
- Therefore, there has to be a non-zero probability that $LR < 1/RMNE$ (barring pathological cases such as $p_a = 1$)
- Heuristically, 1/RMNE being a LR, it summarizes all available information. If it could only increase when the genotypes are taken into account, it would not adequately summarize the information without.
- A mathematical proof: [3].


Maarten Kruijver.
Efficient computations with the likelihood ratio distribution.
Forensic Science International: Genetics, 14:116–124, 2015.


Richard Royall.
Statistical evidence: a likelihood paradigm.
CRC press, 1997.


Klaas Slooten and Thore Egeland.
Exclusion probabilities and likelihood ratios with applications to mixtures. (First of three papers).
International journal of legal medicine, 130(1):39–57, 2016.


Klaas Slooten and Ronald Meester.
Probabilistic strategies for familial DNA searching.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 63(3):361–384, 2014.