Science and Justice 56 (2016) 397–401

Contents lists available at ScienceDirect

Science and Justice

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

Position paper

Posterior distributions for likelihood ratios in forensic science☆

Ardo van den Hout [a,*], Ivo Alberink [b]

[a] Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom
[b] Netherlands Forensic Institute, Laan van Ypenburg 6, The Hague, The Netherlands

ARTICLE INFO

Keywords:
Likelihood ratio
Bayes factor
Bayesian inference
Body height estimation
Posterior likelihood ratios
Precision
Reliability

ABSTRACT

Evaluation of evidence in forensic science is discussed using posterior distributions for likelihood ratios. Instead of eliminating the uncertainty by integrating (Bayes factor) or by conditioning on parameter values, uncertainty in the likelihood ratio is retained by parameter uncertainty derived from posterior distributions. A posterior distribution for a likelihood ratio can be summarised by the median and credible intervals. Using the posterior mean of the distribution is not recommended. An analysis of forensic data for body height estimation is undertaken. The posterior likelihood approach has been criticised both theoretically and with respect to applicability. This paper addresses the latter and illustrates an interesting application area.

1. Introduction and terminology

Evaluation of evidence in forensic science can be undertaken using the likelihood ratio framework. For a continuous random variable, the *likelihood ratio* (LR) is the ratio of two values of the probability function $p(x|\theta)$, given two values of model parameter θ, and data x. For values θ_1 and θ_2, we have $LR = p(x|\theta_1)/p(x|\theta_2)$, where function $p(\cdot)$ is a generic notation for a probability density function or a probability mass function.

Given two hypotheses H_1 and H_2 for assumptions for models M_1 and M_2, respectively, the *Bayes factor* (BF) in favour of H_1 is given by

$$BF = \frac{p(x|H_1)}{p(x|H_2)} = \frac{\int p(x\phi H)p(\phi|H_1)d\phi}{\int p(x|\psi,H_2)p(\psi|H_2)d\psi}. \qquad (1)$$

The BF is also called a *marginal likelihood ratio* as it is the ratio of two marginal likelihoods. It is not necessarily the case that $p(x|\phi,H_1)$ is the same function as $p(x|\psi,H_2)$. These probability functions are defined by M_1 and M_2, respectively. The same holds for $p(\phi|H_1)$ and $p(\psi|H_2)$. It is because of this that the BF can be used to compare non-nested models.

If, however, M_1 and M_2 are nested, i.e., one can be derived from the other by restricting a subset of the parameters, then the BF is still different from the LR, as the latter is defined for specific parameter values and the former is defined by integrating out the parameters. It is only in the specific case where the priors given by $p(\phi|H_1)$ and $p(\psi|H_2)$ identify parameter values with probability 1 (have a point mass 1 at those values), that the BF reduces to a LR.

The following example of a Bayes factor in forensic practice is taken from Lucy [1] (Section 12.5). An eyewitness height description of the male perpetrator is modelled as a normal distribution with mean 1.816 and standard deviation 0.054. The prosecution's hypothesis is H_p: perpetrator = suspect. The defence's hypothesis is H_d: perpetrator ≠ suspect. The assumed population distribution of men is normal with mean 1.775 and standard deviation 0.098. The evidence is the height $E = 1.855$ of the suspect.

The Bayes factor is in this case equal to the probability density of E under H_p divided by the probability density of E under H_d. That is, $BF = f(E|\mu_p,\sigma_p)/f(E|\mu_d,\sigma_d)$, where f is the density of a normal distribution with mean μ and standard deviation σ [1]. For $\mu_p = 1.816, \sigma_p = 0.054, \mu_d = 1.775, \sigma_d = 0.098$ this leads to a Bayes factor of 1.951.

We would like to add the following explanation in terms of the BF. The BF in this case is defined as

$$BF = \frac{p(E|H_p)}{p(E|H_d)} = \frac{\int p(E|\boldsymbol{\theta},H_p)p(\boldsymbol{\theta}|H_p)d\boldsymbol{\theta}}{\int p(E|\boldsymbol{\eta},H_d)p(\boldsymbol{\eta}|H_d)d\boldsymbol{\eta}}. \qquad (2)$$

There are no background data, i.e., there are no sample data from the relevant population. The models under both hypotheses are completely specified normal distributions. This means that $p(\boldsymbol{\theta}|H_p)$ specifies $\boldsymbol{\theta} = (\mu_p,\sigma_p)$ with probability one. Likewise $p(\boldsymbol{\eta}|H_b)$ specifies $\boldsymbol{\eta} = (\mu_d,\sigma_d)$ with probability one. As a result both integrals disappear in Eq. (2) and we end up with $p(E|\boldsymbol{\theta},H_p) = f(E|\mu_p,\sigma_p)$ and $p(E|\boldsymbol{\eta},H_d) = f(E|\mu_d,\sigma_d)$.

☆ 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.
E-mail address: ardo.vandenhout@ucl.ac.uk (A. van den Hout).

http://dx.doi.org/10.1016/j.scijus.2016.06.011

Note that there is no uncertainty associated with the *BF*. Consider the case where background data are used for the estimation of μ_d and σ_d. In that case, the denominator of Eq. (2) would have been

$$p(E|H_d, B) = \int p(E|\boldsymbol{\eta}, H_d, B)p(\boldsymbol{\eta}|H_d, B)d\boldsymbol{\eta} \quad (3)$$

$$= \int p(E|\boldsymbol{\eta}, H_d, B)\frac{p(B|\boldsymbol{\eta}, H_d)p(\boldsymbol{\eta}|H_d)}{p(B|H_d)}d\boldsymbol{\eta}, \quad (4)$$

where $p(B|\boldsymbol{\eta},H_d)$ is the likelihood and $p(\boldsymbol{\eta}|H_d)$ is the prior density. Because the *BF* is in this case defined conditional on background data *B*, there is still no uncertainty associated with the *BF*. The uncertainty with respect to $\boldsymbol{\eta}$ is integrated out. Nevertheless, if a new data set *B* were sampled, another *BF* would be the result. By conditioning on *B*, this sample uncertainty is not accounted for.

In Section 2, the posterior distribution of the likelihood ratio is explained within the context of forensic science. Section 3 presents an evaluation of evidence where the posterior distribution of the likelihood ratio is used for the measurement of body height. Background data in this case consist of measurements on test persons. A comparison is made with the Bayes factor approach. For the posterior sampling we use WinBUGS (Lunn et al. [2]). Section 4 concludes the paper.

2. Posterior likelihood ratio

As an alternative method for simple null hypothesis testing, Aitkin [3] advocates using a Bayesian framework and working with the posterior distribution of the *LR*. Instead of eliminating the uncertainty by maximising (*LR* test) or by integrating (*BF*), Aitkin proposes to retain uncertainty in the *LR* via parameter uncertainty derived from the posterior distributions.

Bayesian inference focusses on the posterior density of parameters. If θ is the parameter and x are the data, then the posterior is given by $p(\theta|x) = p(x|\theta)p(\theta)/p(x)$, where $p(x|\theta)$ is the likelihood of the data and $p(\theta)$ is the prior density of θ. Thus the posterior is proportional to the likelihood times the prior, and this is written as $p(\theta|x) \propto p(x|\theta)p(\theta)$.

The posterior likelihood ratio approach is readily explained in terms of sampling. The *LR* is considered a function of the parameters under both hypotheses. First, given H_1: $\theta = \theta_1$, the likelihood is a single value $L(\theta_1) = p(x|\theta_1)$. Second, given H_2: $\theta \neq \theta_1$, S parameter values θ^* are sampled from the posterior $p(\theta|x)$ and for each value the likelihood $L(\theta^*)$ is computed. Next, the S ratios $L(\theta_1)/L(\theta^*)$ provide a random sample from the posterior of the *LR*.

At first sight, the setting in Aitkin [3] is different from the forensic science setting. For the former, there is a data set and a model, and the hypotheses are about model parameters. For the latter, there is evidence *E* and background data *B*, and the hypotheses are about *E* - not about the model for *B*.

For the forensic science setting, we can define an *LR* given an estimate of model parameters for *B*. This only works if we assume that both the prosecution and the defence accept the same model for *B*. If the model parameter vector is denoted $\boldsymbol{\theta}$, then we can define a likelihood ratio as the ratio of two probability densities for the evidence. This conditional ratio is given by

$$LR = \frac{p(E|H_p, \boldsymbol{\theta})}{p(E|H_d, \boldsymbol{\theta})}. \quad (5)$$

For the forensic science setting, the *BF* is defined as

$$BF = \frac{p(E|H_p, B)}{p(E|H_d, B)} = \frac{\int p(E|H_p, \boldsymbol{\theta}_p)p(\boldsymbol{\theta}_p|B)d\boldsymbol{\theta}_p}{\int p(E|H_d, \boldsymbol{\theta}_d)p(\boldsymbol{\theta}_d|B)d\boldsymbol{\theta}_d}, \quad (6)$$

where $p(\boldsymbol{\theta}_p|B)$ and $p(\boldsymbol{\theta}_d|B)$ are posterior densities.

Given these definitions of *BF* and *LR*, we can apply the ideas of the posterior likelihood ratio and achieve a middle way between *BF* and *LR* such that the uncertainty in the *LR* is retained by parameter uncertainty derived from the posterior distribution of the model parameter vector for the background data. Thus we see *LR* as a function of sampled $\boldsymbol{\theta}$, and obtain its posterior by sampling from the posterior $p(\boldsymbol{\theta}|B)$.

The posterior *LR* distribution is very useful as it can be used to assess the strength of evidence by way of posterior probabilities such as $P(LR > c)$, for any $c > 0$. In this way it is possible to not only have knowledge about the central location of the LR, but also about the precision that is attached. For the end user of the LR (trier of fact) it may be important to know whether for a reported LR of 1000, a 5% lower bound is e.g. 20 or 990.

Care has to be taken not to summarise the posterior distribution of the likelihood ratio by its posterior mean. The posterior mean is not invariant under the switching of the order of the hypotheses in the sense that

$$E_{\boldsymbol{\theta}}\left[\frac{p(E|H_p, \boldsymbol{\theta})}{p(E|H_d, \boldsymbol{\theta})}\right] \neq \left(E_{\boldsymbol{\theta}}\left[\frac{p(E|H_d, \boldsymbol{\theta})}{p(E|H_p, \boldsymbol{\theta})}\right]\right)^{-1}. \quad (7)$$

This is important since the order of the hypotheses should not effect the statistical inference. Instead of assessing the posterior mean, the posterior median and credible intervals can be used for statistical inference.

3. Evaluation of evidence

In this section, the posterior of the likelihood ratio (5) is used for forensic data for height estimation of a perpetrator. A comparison with the Bayes factor (6) is made.

A perpetrator was well visible on a security camera and one image was chosen as the basis of height measuring. Background data *B* consist of additional measurements of six test persons who were positioned in the same stance as the perpetrator in front of the original camera (Edelman et al. [4]).

We use the following notation. Background data are measurements m_i, for test persons $i = 1, 2, \ldots, 6$, and known true heights h_i. The model for the height estimation is

$$m_i = \alpha + h_i + \varepsilon_i \quad with \quad \varepsilon_i \sim N(0, \sigma^2), \quad (8)$$

where α is the systematic measurement error, see Van den Hout and Alberink [5] for an extended model and details of the data. Let $\boldsymbol{\theta} = (\alpha, \log(\sigma))$.

The evidence is the measured height m_p of the perpetrator. The height of the suspect is h_s. The prosecution's hypothesis is H_p: perpetrator is suspect ($h_p = h_s$). The defence's hypothesis is H_d: perpetrator is not suspect ($h_p \neq h_s$). Assume that both the prosecution and the defence accept model (8). The *BF* is given by

$$BF = \frac{p(m_p|H_p, B)}{p(m_p|H_d, B)} = \frac{p(m_p|h_p = h_s, B)}{\int p(m_p|h_p = h, B)p(h)dh} \quad (9)$$

$$= \frac{\int p(m_p|\boldsymbol{\theta}, h_p = h_s)p(\boldsymbol{\theta}|B)d\boldsymbol{\theta}}{\int \left[\int p(m_p|\boldsymbol{\theta}, h_p = h)p(h)dh\right]p(\boldsymbol{\theta}|B)d\boldsymbol{\theta}}. \quad (10)$$

Let us assume that the height distribution of the population is given by $p(h|\mu_h, \sigma_h)$, a normal distribution with known mean μ_h and known standard deviation σ_h. The conditional *LR* is given by

$$LR = \frac{p(m_p|h_p = h_s, \boldsymbol{\theta})}{\int p(m_p|h_p = h, \boldsymbol{\theta})p(h|\mu_h, \sigma_h)dh}. \quad (11)$$