Posterior distributions for likelihood ratios in forensic science*

Ardo van den Hout

Department of Statistical Science, University College London, UK

Ivo Alberink

Netherlands Forensic Institute,

Department of Digital Technology and Biometry, The Netherlands

Abstract

Hypothesis testing in forensic science is discussed and using posterior distributions for likelihood ratios is illustrated. 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.

Key Words: Bayes factor, Bayesian inference, body height estimation, hypothesis testing, posterior likelihood ratios.

---

* *Unpublished manuscript. Online in 2015*

1 Introduction

Evaluation of evidence in forensic science starts with formulating the prosecutor's hypothesis (H_p) and the defender's hypothesis (H_d). A typical example is H_p: the perpetrator is the suspect, and H_d: the perpetrator is not the suspect (but a random person). Let us assume that there are background data. For example, if the evidence is with regard to the height of the suspect, then background data may consist of height measurements in the relevant population.

The evaluation of evidence is the investigation of how the evidence changes the relation between the probability that H_p is true and the probability that H_d is true. Formally, it is the investigation of the multiplicative factor that turns the prior odds in favour of guilt $P(H_p|B)/P(H_d|B)$ into a posterior odds $P(H_p|E,B)/P(H_d|E,B)$, where E is the evidence and B are the background data. This factor is the Bayes factor, and is given by

$$\frac{P(H_p|E,B)}{P(H_d|E,B)} = \underbrace{\frac{p(E|H_p,B)}{p(E|H_d,B)}}_{\text{Bayes factor}} \times \frac{P(H_p|B)}{P(H_d|B)},$$

where $p(\cdot)$ is a generic notation for a probability density function (for continuous evidence data) or a probability mass function (for discrete evidence data). If the Bayes factor is 1, then the evidence does not help to choose between H_p and H_d. If the factor is between 0 and 1, then the evidence makes H_d more likely. If the factor is larger than 1, then the evidence makes H_p more likely. For applications of the Bayes factor in forensic practice, see Lindley (1977), Evett *et al.* (1987), Wakefield *et al.* (1991), Sjerps and Kloosterman (2003), Aitken and Taroni (2004), and Bozza *et al.* (2008).

The definition of the Bayes factor above is in line with the formulation of the Bayes factor within the statistical framework of model comparison. Given two hypotheses H_1 and H_2 corresponding to assumptions of alternative models, M_1 and M_2, for data $\boldsymbol{x}$, the Bayes factor in favour of H_1 is $p(\boldsymbol{x}|M_1)/p(\boldsymbol{x}|M_2)$ (Bernardo and Smith, 2000). In the rest of this paper, the conditioning on a model M will be indicated by the conditioning on the corresponding hypothesis H.

The Bayes factor has often been used in the statistical literature, but it usage is not without problems. First, using the Bayes factor can result in paradoxical inference (*Lindley's paradox*, see Bernardo and Smith, 2000). When a vague prior density is used, non-zero prior weights are assigned to values that have negligible likelihood. When integrating the likelihood with respect to the prior, the vague prior may result in a marginal probability of the data close to zero (Aitkin 1999, p. 115). Second, the factor can be hard to compute in some situations, see the recent discussion in Carlin and Louis (2009). Third, using the factor for a comparison between two models only works when there is no obvious middle ground that can be described by a third model (Gelman *et al.,* 2004, p. 185). Fourth, the Bayes factor is a value without an associated uncertainty. The factor quantifies an aspect of the data and as such it is a statistic without uncertainty. Whether the factor is close to 1 or not can only be decided by following general guidelines about what is considered to be a big or small value (Kass and Raftery, 1995).

The first point is not an issue in the evaluation of evidence when the prior density for the parameters of the model for the background data is the same in the numerator and in denominator of the Bayes factor. Also the third point is not a problem in forensic practice, as proper formulation of H_p and H_d is exclusive. With respect to the fourth point, a subjective assessment of a numerical outcome will never be eliminated in statistics. Nevertheless, we think that the method proposed in the current paper yields a more intuitive way of interpreting the outcome of hypothesis testing in forensic science.

Forensic evidence is sometimes evaluated by conditioning on model parameters for the background data B. This leads to the evaluation of a conditional likelihood ratio given by $p(E|H_p, \boldsymbol{\theta})/p(E|H_d, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector for the model for B. Typically, the likelihood ratio is evaluated conditional on the maximum likelihood estimate of $\boldsymbol{\theta}$, see, e.g., Aitken and Lucy (2004). The advantage of this is that the problems with the Bayes factor are avoided and there is no need for the specification of prior densities. Using the likelihood ratio in this way is a two-stage frequentist approach, where the second stage is statistical inference conditional on the estimate of $\boldsymbol{\theta}$. The disadvantage is that the uncertainty associated with the estimation of $\boldsymbol{\theta}$ is ignored.

As a method for simple null hypothesis testing, Dempster (1974) and Aitkin (1991, 1997, 2010) suggest to use a Bayesian framework but instead of working with the Bayes factor, they propose to consider the posterior distribution of the likelihood ratio. The current paper will illustrate using posterior distributions for likelihood ratios in forensic practice. Careful terminology and formulation are necessary to explain the posterior distribution of a likelihood ratio in a forensic setting and the way the method can be applied. Curran (2005) used the posterior distribution of the likelihood ratio in a specific application with DNA profiles. We will provide a more extensive discussion showing the general applicability of the method and the differences with a fully Bayesian approach.

The idea of using the posterior distribution of a likelihood ratio has received criticism, see the discussion of Aitkin's 1991 paper (in the same journal) and the discussion of Aitkin's 2010 book by Gelman *et al.* (2010). The two main points made by Gelman *et al.* are that the Aitkin's approach is incompatible with a Bayesian perspective, and that the approach does not seem to be useful for common applications in statistics. The aim of our work here is with respect to the second point: we hope to illustrate that forensic science provides interesting applications for Aitkin's method. With regard to the first point, Gelman *et al.* make a strong case. But - as stated by Gelman *et al.* themselves - it does not imply that Aitkin's approach is wrong. It just means that the approach is not purely Bayesian. Nevertheless, for reasons of consistency, we will use Aitkin's terminology such as *posterior distribution of the likelihood ratio* and *posterior probability* throughout this paper. Further discussion of the merits and the disadvantages of the approach are presented in the conclusion.

Section 2 introduces terminology. In Section 3, the posterior distribution of the likelihood ratio is explained within the context of forensic science. Section 4 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.,* 2000). Section 5 concludes the paper.

2 Terminology

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, as before, 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

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

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 LR can be used for simple null hypothesis testing. For example, let the null hypothesis be given by H_1: $\theta = \theta_1$, and the alternative by H_2: $\theta \neq \theta_1$. If $\widehat{\theta}$ is the maximum likelihood estimate, then the probability distribution of the test statistic $-2 \log[p(x|\theta_1)/p(y|\widehat{\theta})]$ can be approximated by a chi-square distribution with 1 degree of freedom (conditional on some assumptions). This is the well-established likelihood ratio test.

The BF can also be used for null hypothesis testing. For H_1 and H_2, BF is given by $p(x|\theta_1)/\int p(x|\theta)p(\theta)d\theta$, where $p(\theta)$ is the prior density under the alternative hypothesis. In this case, H_1 is rejected if $BF < 1$ and close to zero, and H_2 is rejected if $BF > 1$ and large.

The following example of a Bayes factor in forensic practice is taken from Lucy (2005, Section 12.5). An eyewitness height description of the male perpetrator is modelled as a normal distribution with mean 1.816 meter and standard deviation 0.054. The prosecutor's hypothesis is H_p: perpetrator = suspect. The defender's hypothesis is H_d: perpetrator $\neq$ 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, $BF = f(E|\mu_p = 1.816, \sigma_p = 0.054)/f(E|\mu_d = 1.775, \sigma_d = 0.098) = 1.951$, where f is the density of a normal distribution with mean μ and standard deviation σ (Lucy 2005).

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

$$BF = \frac{p(E|H_p)}{p(E|H_d)} = \frac{\int p(E|\boldsymbol{\theta}, H_p)p(\boldsymbol{\theta}|H_p)d\boldsymbol{\theta}}{\int p(E|\boldsymbol{\eta}, H_d)p(\boldsymbol{\eta}|H_d)d\boldsymbol{\eta}}. \tag{1}$$

There are no background data. 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 (1) and we end up with $p(E|\boldsymbol{\theta}, H_p) = f(E|\mu_p, \sigma_p)$ and $p(E|\boldsymbol{\eta}, H_d) = f(E|\mu_d, \sigma_d)$.

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 (1) would have been

$$\begin{aligned} p(E|H_d, B) &= \int p(E|\boldsymbol{\eta}, H_d, B)p(\boldsymbol{\eta}|H_d, B)d\boldsymbol{\eta} \\ &= \int p(E|\boldsymbol{\eta}, H_d, B)\frac{p(B|\boldsymbol{\eta}, H_d)p(\boldsymbol{\eta}|H_d)}{p(B|H_d)}d\boldsymbol{\eta}, \end{aligned}$$

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 would be sampled, another BF would be the result. By conditioning on B, this sample uncertainty is not accounted for.

3 Posterior likelihood ratio

As an alternative method for simple null hypothesis testing, Aitkin (2010) advocates to use a Bayesian framework but instead of working with the BF, and proposes to consider the posterior distribution of the LR. Instead of eliminating the uncertainty by maximising (LR test) or by integrating (BF), uncertainty in the LR is retained by 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 as 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 (2010) 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 prosecutor and the defender accept the same model for B. If the model parameter vector is denoted $\boldsymbol{\theta}$, then we can define a likelihood ratio by the ratio of two probability densities for the evidence. This conditional ratio is given by

$$LR = \frac{p(E|H_p, \boldsymbol{\theta})}{p(E|H_d, \boldsymbol{\theta})}. \tag{2}$$

For the forensic science setting, the BF is defined as

$$BF \;=\; \frac{p(E|H_p, B)}{p(E|H_d, B)} = \frac{\int p(E|H_p, \boldsymbol{\theta}_p)p(\boldsymbol{\theta}_p|B)d\boldsymbol{\theta}_p}{\int p(E|H_d, \boldsymbol{\theta}_d)p(\boldsymbol{\theta}_d|B)d\boldsymbol{\theta}_d}, \tag{3}$$

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 of LR 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 $c > 0$.

Care has to be taken not the 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

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

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.

4 Evaluation of evidence

In this section, the posterior of the likelihood ratio (2) is used for forensic data for height estimation of a perpetrator. A comparison with the Bayes factor (3) 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 perpetrators in front of the original camera (Edelman *et al.,* 2010).

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

$$m_i = \alpha + h_i + \epsilon_i \qquad \text{with} \qquad \epsilon_i \sim N(0, \sigma^2), \tag{4}$$

where α is the systematic measurement error, see Van den Hout and Alberink (2010) 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 prosecutor's hypothesis is H_p: perpetrator is suspect ($h_p = h_s$). The defender's hypothesis is H_d: perpetrator is not suspect ($h_p \neq h_s$). Assume that both the prosecutor and the defender accept model (4). The BF is given by

$$\begin{aligned} BF &= \frac{p(m_p|H_p, B)}{p(m_p|H_d, B)} = \frac{p(m_p|h_p = h_s, B)}{\int p(m_p|h_p = h, B)p(h)dh} \\ &= \frac{\int p(m_p|\boldsymbol{\theta}, h_p = h_s)p(\boldsymbol{\theta}|B)d\boldsymbol{\theta}}{\int \left[\int p(m_p|\boldsymbol{\theta}, h_p = h)p(h)dh\right] p(\boldsymbol{\theta}|B)d\boldsymbol{\theta}}. \end{aligned} \tag{5}$$

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

$$LR = \frac{p(m_p|h_p = h_s, \boldsymbol{\theta})}{\int p(m_p|h_p = h, \boldsymbol{\theta})p(h|\mu_h, \sigma_h)dh}. \tag{6}$$

The numerator of (6) is a normal density and is given by

$$p(m_p|h_p = h_s, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(m_p - \alpha - h_s)^2}{\sigma^2}\right]$$

Since $p(h|\mu_h, \sigma_h)$ is a normal distribution, there is a closed-form solution for the integral in the denominator of (6). The integrand is a convolution of two normal distributions and the denominator is given by

$$\int p(m_p|h_p = h, \boldsymbol{\theta})p(h|\mu_h, \sigma_h)dh = \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_h^2)}} \exp\left(-\frac{1}{2}\frac{(m_p - \alpha - \mu_h)^2}{\sigma^2 + \sigma_h^2}\right),$$

see, e.g., Gelman *et al.* (2004, Section 2.6) for a similar computation. If $\boldsymbol{\theta}$ is treated as a fixed value, then there is no uncertainty associated with LR.

For the posterior of LR, firstly, we sample $\boldsymbol{\theta}^*$ from the posterior $p(\boldsymbol{\theta}|B)$. Secondly we compute LR for each sampled $\boldsymbol{\theta}^*$.

To obtain the posterior $p(\boldsymbol{\theta}|B)$, we have to specify the prior of the model parameter vector $\boldsymbol{\theta}$. Gelman *et al.* (2004) discuss the definition of the prior density in the context of the normal distribution, and also the sampling from the resulting posterior. Various levels of informativeness and conjugacy are presented by Gelman *et al.*

For the evaluation of evidence in the present setting, we specify an informative proper prior $p(\boldsymbol{\theta})$ without worrying about conjugacy as we will rely on the automatic MCMC procedures in WinBUGS to do the sampling.

To compare the posterior likelihood ratio approach with the Bayes Factor (5), we approximate the integrals in the latter by using the trapezoidal rule (with 500 nodes). This computation includes the estimation of the marginal density $p(B)$ since the posterior for $\boldsymbol{\theta}$ is given by $p(\boldsymbol{\theta}|B) = p(B|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(B)$. In general, the estimation of marginal density can be complex (Carlin and Louis, 2009). Since $\boldsymbol{\theta}$ consists of only two parameters, numerical approximation of the integrals works fine. Sampling from the posterior of LR is undertaken in WinBUGS (Lunn *et al.,* 2000). WinBUGS is freely available software for the Bayesian analysis of statistical models using Markov chain Monte Carlo (MCMC) methods, see also `www.mrc-bsu.cam.ac.uk/bugs`. Code is provided in the Appendix. For the inference in this application, the MCMC consisted of two chains, each with a

Table 1: Background data on measured heights and true heights of test persons, and measured height of perpetrator.

	Test persons						Perpetrator
Measured height	1.964	1.832	1.900	1.780	1.937	1.865	1.885
True height	1.950	1.795	1.865	1.755	1.910	1.825	-

burn-in of 10000, and a further 10000 updates for inference. Convergence of the MCMC was checked by using the diagnostic tools provided within WinBUGS.

Evidence m_p and background data for the height estimation are presented in Table 1. The population distribution of Dutch Caucasian men is assumed to be normal with mean $\mu_h = 1.806$ and standard deviation $\sigma_h = 0.1$ (Statistics Netherlands, `www.cbs.nl`, 2006). This specifies $p(h|\mu_h, \sigma_h)$. For the prior of $\boldsymbol{\theta}$ we assume $p(\boldsymbol{\theta}) = p(\alpha, \log(\sigma)) = p(\alpha)p(\log(\sigma))$, and furthermore $\alpha \sim N(0, 0.1)$ and $\log(\sigma) \sim U(-10, 0)$. These priors are informative and take into account that the measurements are in meters.

Bayesian inference using WinBUGS yields a posterior mean 0.029 for α with 95% credible interval (CI) $(0.017, 0.042)$. So there is a systematic overestimation of the height of about 3cm. For σ the figures are 0.012 $(0.006, 0.024)$. The posterior density $p(\boldsymbol{\theta}|B)$ has a regular shape and is depicted in Figure 1.

We will illustrate the evaluation of the evidence $m_p = 1.885$ for various values of the height of the suspect h_s. Say that the suspect has the same height as the perpetrator. In that case $m_p - \alpha \approx 1.885 - 0.029 = 1.856 = h_s$. If this is indeed the case we would expect the value 1 to be located in the left tail of the density of LR because it is likely that the suspect is the perpetrator and hence the mean of LR should be larger than 1. In other words, $P(LR < 1)$ should be small. For the same reason, we would expect BF to be larger than 1. This is indeed

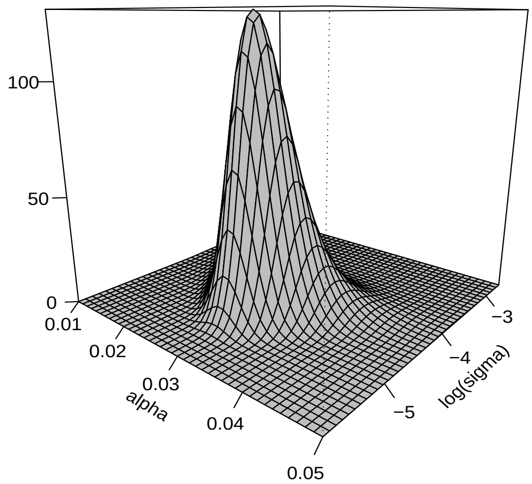

Figure 1: Posterior density $p(\boldsymbol{\theta}|B) = p(\alpha, \log(\sigma)|B)$.

the case, the posterior median of LR is 10.03 with 95% CI $(4.25, 17.69)$, and the BF is estimated at 9.27. The big advantage of the posterior of LR is clear: we have estimated its distribution. The 95% CI for instance shows immediately that $LR = 1$ is not very likely. This is information that the computation of BF does not provide. Because we have sampled values of LR, probabilities such as $P(LR < 1)$ are easy to estimate. In this case all sampled values of LR were larger than one, so we may safely conclude that $P(LR < 1) < 0.01$. The evidence is clearly in favour of the prosecutor's hypothesis.

Next we consider two values of h_s that are clearly in favour of the defender's hypothesis. Both values $h_s = 1.7$ and $h_s = 2.0$ yield a posterior median of LR smaller than 0.01 and $P(LR < 1) > 0.99$. The corresponding BF are both smaller than 0.001. For both theses values of the suspect's height, the evidence is clearly in favour of the defender's hypothesis.

The value $h_s = 1.825$ illustrates a situation where the extra information of the posterior of the LR is of particular use. The BF is estimated at 0.53. This is dissimilar to the posterior median 0.15 of the LR, whereas the posterior mean 0.521 of the LR is close to the BF. Where the BF gives no uncertainty information, the sampled values of the LR allow many possible quantities to be estimated to assess whether the evidence is in favour of the defender's hypothesis. The latter is not the case. The 95% CI for the LR is $(< 0.01, 2.86)$ which includes the value 1. Probability $P(LR < 1)$ is estimated at 0.82.

For the value $h_s = 1.825$, we investigate the sensitivity of the results with regard to the specification of the prior $p(\boldsymbol{\theta}) = p(\alpha, \log(\sigma))$. First, we use priors which are less informative. We specify $\alpha \sim N(0, 1)$ and $\log(\sigma) \sim U(-10, 5)$. Given that measurements are in meters, these priors do not contain much information. For the LR, we obtain median 0.150 and 95% CI $(< 0.01, 3.03)$, the BF is estimated at 0.54. Next we specify $\alpha \sim N(0, 0.05)$ and $\log(\sigma) \sim U(-10, -3)$. The prior for α implies that about 95% of the systematic error falls with the interval (-10cm, 10cm), the prior for σ implies that σ is less than 10cm. These priors are informative, but are still reasonable for this case. For the LR, we obtain median 0.145 and CI $(< 0.01, 2.80)$, the BF is estimated at 0.51. Given these alternative specifications of the priors, results are very similar to the previous results.

5 Conclusion

A fully Bayesian evaluation of evidence requires the computation of a Bayes factor. For complex models, this factor may be hard to compute. Using the ideas in Dempster (1974) and Aitkin (2010), the posterior distribution of the likelihood ratio is used in a forensic science setting as an alternative to the Bayes factor. 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.

The application discussed forensic data where heights were estimated on the bases of images from a security camera. The posterior mean of the likelihood ratio was similar to the Bayes factor. With samples available from the posterior of the likelihood ratio, an all-round inference was possible by investigating posterior percentiles and credible intervals.

As stated in the introduction, Gelman *et al.* (2010) criticise the posterior likelihood ratio approach by arguing that it is incompatible with a Bayesian perspective, and that it does not seem to be useful for common applications in statistics. We hope to have shown in this paper that forensic science is an area where the approach seems useful. The points raised by Gelman *et al.* (2010) with respect to using vague priors, comparing discrete hypotheses, and the problem with product of posteriors, are not applicable in our setting: In forensic science, it make sense to use vague prior densities for the parameters in the model for the background data, researchers are interested in comparing discrete hypotheses, and - at least in the current application - there is no assessment of a product of posteriors.

Nevertheless, we acknowledge that there are still important issues in the posterior likelihood ratio approach that need further attention. Using the posterior distribution of LR for hypothesis testing can be seen as a hybrid of Bayesian and frequentist methods. It is not fully Bayesian, but it is also not a frequentist analysis. This ambiguity causes interpretation problems. For example, in a fully Bayesian framework, a 95% credible interval of a parameter means that the posterior probability that the parameter lies in that interval is 0.95. A frequentist 95% confidence interval means that given a large number of repeated samples, 95% of the estimated confidence intervals includes the true value of the parameter. What are the properties of the credible intervals for LR that we computed in the current application?

Using the posterior likelihood ratio has a wide range of possible applications in forensic practice. Computationally it is a feasible method to evaluate evidence. It takes into account the uncertainty with regard to inference from background and at the same time allows to model prior knowledge.

Appendix

WinBUGS code used in the evaluation of evidence. For more information on the software and MCMC sampling see `www.mrc-bsu.cam.ac.uk/bugs`.

```
Data:
list(h = c(1.950, 1.795, 1.865, 1.755 ,1.910, 1.825) ,
     m = c(1.964, 1.832, 1.900, 1.780, 1.937, 1.865))

Inits:
list(alpha=0, logsigma= -4)
list(alpha=0.02, logsigma= -5)

Model:
model{
# Model for measurement:
for(i in 1:6){ mu[i]<-h[i]+alpha;  m[i]~dnorm(mu[i], tau) }

# Evaluation of evidence:
h_s<- 1.825; m.p <- 1.885
# Under H_p:
pi<-3.141593; p_Hp<-1/(sqrt(2*pi)*sigma)*exp(-1/2*tau*pow(m.p-(h_s+alpha),2))
# Under H_d:
mu_h.pop<-1.806; var_h.pop<-0.01
tau_h.pop<-1/var_h.pop
p_Hd<-1/sqrt(2*pi*(var+var_h.pop) )*exp(-1/(2*(var+var_h.pop))*pow(m.p-alpha-mu_h.pop,2))
# LR:
LR<-p_Hp/p_Hd
# Strength of evidence:
c<-1; pprob<-step(c-LR)
# Converting precision to sd and var:
tau<-pow(sigma,-2); var<-pow(tau,-1)
# Priors:
alpha~dnorm(0,0.1); logsigma~dunif(-10,0); sigma<-exp(logsigma)}
```

References

Aitken, C.G.G and F. Taroni (2004). *Statistics and the Evaluation of Evidence for Forensics Scientists, Second Edition*, Wiley, Chichester, UK.

Aitken, C.G.G. and D. Lucy (2004). Evaluation of trace evidence in the form of multivariate data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *53*, 109–122.

Aitkin, M. (1991). Posterior Bayes factors (with discussion) *Journal of the Royal Statistical Society, Series B*, 53, 111–142.

Aitkin, M. (1997). The calibration of P-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood, *Statistics and Computing*, *7*, 253–261.

Aitkin, M. (2010). *Statistical Inference. An Integrated Bayesian/Likelihood Approach*, Chapman & Hall/CRC, Boca Raton, FL.

Aitkin, M., R.J. Boys, and T. Chadwick (2005). Bayesian point null hypothesis testing via the posterior likelihood ratio, *Statistics and Computing*, *15*, 217–230.

Bernardo, J.M. and A.F.M. Smith (2000). *Bayesian Theory.* Wiley, Chichester, UK.

Bozza, S., F. Taroni, R. Marquis and M. Schmittbuhl (2008). Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *57*, 329–341.

Carlin, B.P. and T.A. Louis (2009). *Bayesian Methods and Data Analysis. Third Edition*, Chapman & Hall, London.

Curran, J.M. (2005). An introduction to Bayesian credible intervals for sampling error in DNA profiles, *Law, Probability and Risk*, *4*, 115–126.

Dempster, A.P. (1974). *The direct use of likelihood for significance testing*, in O. Barndorff-Nielsen, P. Bleasild and G. Sihon (eds.), Proc. Conf.

Foundational Questions in Statistical Inference, University of Aarhus, 335–352.

Edelman G., I. Alberink and B. Hoogeboom (2010). Comparison of the performance of two methods for height estimation, *Journal of Forensic Science*, *55*, 358–365.

Evett, I.W., P.E. Cage, and C.G.G. Aitken (1987). Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *36*, 174–180.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin (2004), *Bayesian data analysis*, Chapman & Hall, London.

Gelman, A., C.P. Robert, and J. Rousseau (2010), Do we need an integrated Bayesian/likelihood inference? Unpublished. `arxiv.org`: arXiv:1012.2184v1

Kass, R.E. and A.E. Raftery (1995). Bayes factors, *Journal of the American Statistical Association*, *90*, 773–795.

Lindley, D.V. (1977). A problem in forensic science. *Biometrika* **64**, 207–213.

Lucy, D. (2005). *Introduction to Statistics for Forensic Scientists*, Wiley, Chichester, UK.

Lunn, D.J., A. Thomas, N.G. Best and D.J. Spiegelhalter (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility, *Statistics and Computing*, *10*, 325–337.

Sjerps, M. and A.D. Kloosterman (2003). Statistical aspects of interpreting DNA profiling in legal cases, *Statistica Neerlandica*, *57*, 368–389.

Van den Hout, A. and I. Alberink (2010). A hierarchical model for body height estimation in images, *Forensic Science International*, *197*, 48-53.

Wakefield, J.C., A.M. Skene, A.F.M. Smith and I.W. Evett (1991). The Evaluation of Fibre Transfer Evidence in Forensic Science: A Case Study in Statistical Modelling, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *40*, 461–476.