JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2011, Vol. 56, No. 2
doi: 10.1111/j.1556-4029.2010.01665.x
Available online at: onlinelibrary.wiley.com

PAPER

DIGITAL & MULTIMEDIA SCIENCES

Anders Nordgaard,[1,2] *Ph.D. and Tobias Höglund,*[1] *M.Sc.*

Assessment of Approximate Likelihood Ratios from Continuous Distributions: A Case Study of Digital Camera Identification*

ABSTRACT: A reported likelihood ratio for the value of evidence is very often a point estimate based on various types of reference data. When presented in court, such frequentist likelihood ratio gets a higher scientific value if it is accompanied by an error bound. This becomes particularly important when the magnitude of the likelihood ratio is modest and thus is giving less support for the forwarded proposition. Here, we investigate methods for error bound estimation for the specific case of digital camera identification. The underlying probability distributions are continuous and previously proposed models for those are used, but the derived methodology is otherwise general. Both asymptotic and resampling distributions are applied in combination with different types of point estimators. The results show that resampling is preferable for assessment based on asymptotic distributions. Further, assessment of parametric estimators is superior to evaluation of kernel estimators when background data are limited.

KEYWORDS: forensic science, likelihood ratio, digital cameras, generalized Gaussian distribution, confidence intervals, bootstrap

Today, the scientific perspective of forensic interpretation of evidence is based largely on statistical theory, particularly Bayesian hypothesis testing. Aitken and Taroni (1) use the term *likelihood ratio*, which in their context is equal to the Bayes factor (2). What is often called the logical approach to evidence evaluation (3) stipulates that the likelihood ratio is the only aspect of an interpretation of a forensic analysis that should be reported. Although this ratio is not an independent measure, it is part of a completely logical framework, and in that context it represents the logical evidence. However, it would be unlucky to denote this strategy "Bayesian," because the forensic analyst should not apply the full Bayes' theorem to come to conclusions (3). The likelihood ratio can be explained, understood, and interpreted without the inclusion of prior odds, but it cannot be used as independent support for decisions about conviction or acquittal.

A likelihood ratio is often connected with estimating probabilities or expectations related to the findings of a forensic analysis, unless such probabilities or expectations are already well known. There is always a degree of uncertainty when probabilities must be estimated from reference data (compiled databases or historical case records), and thus an obtained likelihood ratio must almost always be treated as a point estimate of the underlying true likelihood ratio. Assessment of this estimate with respect to sampling variation is therefore important and is particularly necessary if the magnitude of the obtained estimate is small. Assessment of likelihood ratios for DNA evidence in terms of uncertainty limits has previously been investigated (4–6) including Bayesian credible intervals and highest posterior density regions as well as frequentistic confidence intervals obtained from asymptotic distribution or by bootstrapping. It might be argued that for several cases of DNA identification, the evidence value is so high that uncertainty limits become superfluous. As an example, if the obtained likelihood ratio says that a match is 20 million times more probable that the suspect was the donor of the stain than that he was not, and the lower bound of that number is 18 million, it would most probably not affect the conclusions in court. However, as pointed out (6), a calculated likelihood ratio with an error bound included gives a more trustworthy impression. Moreover, in cases in which the reported likelihood ratio is very close to one, it would be of no practical use to provide the value with a measure of uncertainty, because we would already have reached the point where the findings are inconclusive. Reported error bounds would however strengthen the scientific value of the expert testimony.

Nevertheless, forensic casework includes the reporting of likelihood ratios in the whole range from zero to infinity, and scrutinization of those quantities may sometimes lead to unpleasant surprises, for instance when it cannot be stated that there is high probability that the true counterpart of a reported likelihood ratio of 1000 is actually above 100. The results of evidence evaluation cannot be more certain for information derived from an entire logical framework than for that acquired from each of its components. When the evidence value is a ratio of conditional probabilities (usually denoted $P(E|H_P)/P(E|H_D)$), its interpretation is that of a Bayes factor and becomes straightforward in cases where the probabilities can be estimated from relative frequencies obtained from existing databases (typical for DNA evidence, but also for a number of other types of evidence such as shoe prints and glass fragments). When such databases are very large, the discrepancies between the true probabilities and the relative frequencies are negligible and calculated posterior odds can be considered to be the true

[1]The Swedish National Laboratory of Forensic Science, SE-581 94 Linköping, Sweden.
[2]Department of Computer and Information Science, Linköping University, SE-581 83 Linköping, Sweden.

*Presented in part at the Seventh International Conference on Forensic Inference and Statistics, August 20–23, 2008, in Lausanne, Switzerland.

Received 20 June 2009; and in revised form 22 Feb. 2010; accepted 6 Mar. 2010.

ones. For other types of evidence, and in particular the one that is studied here, the value of evidence is more complicated and proposed likelihood ratios will not be equal to true Bayes factors. Additional models that can be estimated from available reference data are needed, and the resulting likelihood ratios can be considered as parameters in the frequentistic sense. The estimated likelihood ratios can be used as approximate Bayes factors in Bayesian hypothesis testing, but the uncertainty induced by the estimation must be considered. Confidence intervals for likelihood ratios in heroin seizures have been proposed (7) by the use of bootstrapping including bias correction. If the likelihood ratio can be expressed in terms of parameters of the sampling distributions, like for height measurements (8), sampling distributions of estimators of those parameters can be used to construct confidence intervals for the likelihood ratio.

Common to the works of Curran et al. (5,6) is the use of continuous distributions in the numerator and denominator of the likelihood ratio, and this paper describes another study within this area. The intention is not to present a general framework, but rather to obtain possible approaches for a certain type of application, i.e., digital camera identification. Use of the derived methodology is however not restricted to this application, which we have chosen simply to facilitate understanding of the present results. It should also be mentioned that we use only continuous univariate data in our approaches, to agree with the application of interest. In the following section, we describe the background of digital camera identification and derive likelihood ratios. Thereafter, we develop the methodology needed for assessment and subsequently present a small simulation study to validate the devised techniques.

Camera Identification and Likelihood Ratios

One of the key issues in the forensic analysis of an image concerns the source of that particular graphic material. For a digital image, the most important thing is to determine whether it came from a certain camera, and that issue can be rephrased into two competing hypotheses:

$$\begin{aligned} &H_P: \text{The image was taken with the camera in question.} \\ &H_D: \text{The image was taken with a different camera than} \\ &\quad \text{the one in question.} \end{aligned} \tag{1}$$

It can be noted that it is essential to know how hypothesis H_D is specified. In a standard manner, it might be expressed as the complement of H_P, in other words, it could simply be stated that the camera in question did *not* capture the image. However, that would complicate the situation, because it would mean that the alternative explanation for the source of the image could even include devices other than cameras (e.g., some means of artificial image production). To comply with routine intelligence procedures, here we restrict the second hypothesis to the involvement of a camera (or cameras) other than the one that is specified as source of the image of interest in hypothesis H_P, and we leave cases involving sources other than cameras to future work. It should also be said that in many cases, hypothesis H_D comprises one or more specific cameras (that are available for analysis), while in other cases H_D would comprise a randomly selected camera, but different from the questioned one.

Measuring the Distance Between an Image and a Camera

Once we have established the hypotheses that should be compared when evaluating the evidence, the next step is to decide what kind of measurements should be used to calculate (estimate) the likelihoods of the hypotheses. Following suggestions made by other investigators (9,10), we choose what is known as *pixel nonuniformity* (*PNU*) *noise* for that purpose. Generally, the noise can be extracted as the rest or *residual* after application of a noise filter. If $\boldsymbol{Y}_P$ is the pixel brightness (two-dimensional array) of an image, and G is the noise filter operator, the *noise residual* is

$$\Delta_P = \boldsymbol{Y}_P - G(\boldsymbol{Y}_P) \tag{2}$$

We prefer to use the term *array* instead of matrix as it will be used as a vector in the calculations below, but still appear as a two-dimensional entity. For the most part, PNU is considered to be the only type of pixel/camera-specific noise (9), and thus we can reduce the other noise components by averaging over images taken with the same camera. The *reference pattern* for a specific camera is estimated by shooting a satisfactorily large number (N) of images and then calculating

$$\boldsymbol{R} = N^{-1} \sum_{i=1}^{N} \Delta_{P_i} \tag{3}$$

where Δ_{P_i} is the noise residual obtained from shot image i. Note that $\boldsymbol{R}$ is a two-dimensional array of the same size as $\boldsymbol{Y}$. Now, it is possible to judge whether an image in question (Q) was taken with a particular camera (C) by comparing the noise residual of that image ($\boldsymbol{\Delta}_Q$) with a reference pattern ($\boldsymbol{R}_C$) obtained from that camera. In particular, we measure the "distance" between the image and the camera by determining the Pearson correlation coefficient between $\boldsymbol{\Delta}_Q$ and $\boldsymbol{R}_C$ as follows:

$$r(\Delta_Q, \boldsymbol{R}_C) = \frac{(\Delta_Q - \overline{\Delta_Q}) \circ (\boldsymbol{R}_C - \overline{\boldsymbol{R}_C})}{\|\Delta_Q - \overline{\Delta_Q}\| \cdot \|\boldsymbol{R}_C - \overline{\boldsymbol{R}_C}\|} \tag{4}$$

where $\overline{\Delta_Q}$ and $\overline{\boldsymbol{R}_C}$ are averages over the whole array, "$\circ$" is the inner product, and $\| \cdot \|$ represents the Euclidean norm.[1] It could be expected that this correlation coefficient would generally be low, because PNU noise does not dominate the noise residual. However, the higher that value, the more it seems that the image in question was indeed taken using camera C. Frequently hereafter, we refer to this measure as "correlation" or "correlation coefficient" to avoid the cumbersome expressions including "noise residual" and "reference pattern." Let us consider a case involving two cameras, either of which could have produced an image in question. If we use $Q \leftarrow C$ to represent the event of the image Q being taken by camera C, then the pair of "competing" hypotheses can be written

$$\begin{aligned} &H_P: Q \leftarrow C \\ &H_D: Q \leftarrow \overline{C} \end{aligned} \tag{5}$$

and $\overline{C}$ in this case denotes the alternative camera. If we shoot a satisfactory number of images with each camera, we can obtain empirical probability distributions of the correlation coefficient (4) for each of the two cameras involved. Note that such sets of images are taken independently of the set used to estimate the reference pattern $\boldsymbol{R}_C$. The current state-of-the art strategy is to compare those two distributions with the correlation coefficient obtained for the image of interest. This is illustrated in Fig. 1,

[1]Equation (2.4) writes the correlation coefficient on condensed form. Letting $\delta_1, \ldots, \delta_m$ denote the individual values of $\boldsymbol{\Delta}_Q$ and $R_1, \ldots, R_m$, the individual values of $\boldsymbol{R}_C$ the correlation coefficient can be written using standard notation as $\sum_1^m (\delta_i - \overline{\delta}) \cdot (R_i - \overline{R}) \Big/ \sqrt{\sum_1^m (\delta_i - \overline{\delta})^2 \cdot \sum_1^m (R_i - \overline{R})^2}$.

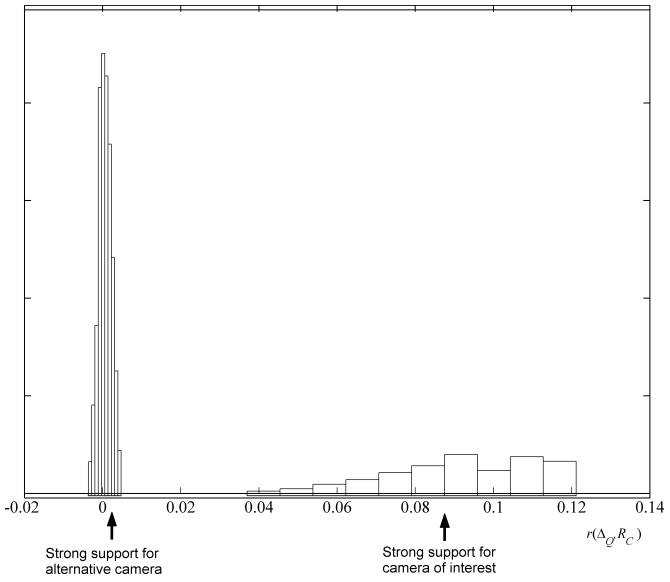

FIG. 1—*Empirical probability distributions for a camera of interest (*C*; right histogram) and an alternative camera (*$\overline{\text{C}}$*; left histogram). The arrows show the positions of a calculated correlation coefficient (2.4) that can provide strong support for the conclusion that either of the respective cameras is the source of an image in question (*Q*).*

in which the two arrows indicate the positions of a calculated correlation coefficient that are necessary to conclude strong support for each of the hypotheses H_P and H_D. It can be seen that, in the area between the arrows, the conclusion about support is weaker, but the camera whose empirical distribution is ''closest'' to the observed correlation coefficient is the one that can be concluded to be the source of the image *Q*.

It is apparent that the empirical probability distributions shown in Fig. 1 are very different in shape. However, in general, dispersions of the distributions of correlation coefficients will be larger for images captured with the camera for which the reference pattern has been calculated than for images from other cameras. The reason for this is that other cameras will usually produce much lower correlation coefficients, but the dispersion will diminish with the level.

Likelihood Ratios Based on Correlation Coefficients

An image used as a piece of evidence is often referred to as recovered data, and a likelihood ratio should be used in the forensic evaluation of such data. Bayes' theorem in odds form for hypothesis testing using continuous measurements is:

$$\frac{\Pr(\mathrm{H}_P|\boldsymbol{x},I)}{\Pr(\mathrm{H}_D|\boldsymbol{x},I)}=\frac{f(\boldsymbol{x}\,|\mathrm{H}_P,I)}{f(\boldsymbol{x}\,|\,\mathrm{H}_D,I)}\cdot\frac{\Pr(\mathrm{H}_P|I)}{\Pr(\mathrm{H}_D|I)} \tag{6}$$

As before, H_P and H_D represent, respectively, the hypothesis that the image was taken with a particular camera and the hypothesis that it was taken with some other camera; and $\boldsymbol{x}$ denotes the continuous measurements (recovered data) made on the (piece of) evidence, which in our case will be the correlation coefficient obtained. The term *I* represents the background information that is relevant for all components involved. The ratio

$$\frac{f(\boldsymbol{x}\,|\mathrm{H}_P,I)}{f(\boldsymbol{x}\,|\,\mathrm{H}_D,I)}=V \tag{7}$$

is referred to in the forensic literature as the likelihood ratio that stands for the evidence value, where $f(\cdot\,|\mathrm{H}_P,I)$ and $f(\cdot\,|\,\mathrm{H}_D,I)$ are the probability density functions valid under the respective hypothesis (1). The generic symbol *f* is used for both functions, but the density of the numerator can be very different from the density in the denominator depending on how each hypothesis and the background information relate to the recovered data $\boldsymbol{x}$. We will hereafter omit the background information (*I*) from the formulas to simplify the expressions. With true densities (implying that the sampling models for the evidence under the respective hypotheses are known), the ratio in (7), the *frequentist* likelihood ratio is equal to the Bayes factor. This is easy to see because the Bayes factor is defined as the ratio of the posterior odds to the prior odds of the hypotheses. In the following, we shall however use an *approximate* Bayes factor as we will use estimates of the probability densities. When referring to likelihood ratios below, it shall therefore be understood that we are using frequentist likelihood ratios and not true Bayes factors.

We now turn to establishing a likelihood ratio from the correlation coefficient (4) obtained for an image of interest. Figure 1 shows empirical probability distributions of correlation coefficients obtained from sets of images taken with each of two ''competing'' cameras. Indeed, the comparison of these two distributions with the correlation coefficient $r(\Delta_Q, \boldsymbol{R}_C)$ for the image in question is a likelihood ratio evaluation, although not numerically specified. To achieve such specification, we need estimates of the probability densities governing the empirical data.

As before, we let *C* stand for the camera that is supposed to be the source of an image *Q*, $\boldsymbol{R}_C$ is the reference pattern obtained for *C*, and $\overline{C}$ denotes all other cameras that are potential sources of *Q*. This is an extension of the procedure described in the previous section, which assumed only one alternative camera. Now, independent of the $\boldsymbol{R}_C$, a set of n_C images are captured with camera *C*, and the correlations of each image's noise residual with $\boldsymbol{R}_C$ are obtained (i.e., a collection of correlations):

$$\left\{r\left(\Delta_{C,j},\boldsymbol{R}_C\right)\right\}_{j=1}^{n_C} \tag{8}$$

The set $\overline{C}$ can be written $\{B_1,\ldots, B_m\}$, where each B_i represents a camera other than *C*. A set of images is produced with each of those cameras, and the correlations between each image's noise residual with $\boldsymbol{R}_C$ are obtained. This gives *m* sets of correlations:

$$\left\{r\left(\Delta_{B_1,j},\boldsymbol{R}_C\right)\right\}_{j=1}^{n_{B_1}},\ldots,\left\{r\left(\Delta_{B_m,j},\boldsymbol{R}_C\right)\right\}_{j=1}^{n_{B_m}} \tag{9}$$

($n_{B_1},\ldots,n_{B_m}$ are the number of images shot with cameras B_1, …, B_m, respectively). Let Ψ_C be the population of correlations for images taken with camera *C* and $\Psi_{\overline{C}}$ the population of correlations for images obtained with the alternative cameras. Each population is thus built on the set of infinite number of images that can theoretically be taken with that camera (those cameras). Then, (8) constitutes a random sample from Ψ_C, and (9) a random sample from $\Psi_{\overline{C}}$. Further, let f_C and $f_{\overline{C}}$ be the probability density functions for populations Ψ_C and $\Psi_{\overline{C}}$, respectively. The former density can be assumed to be unimodal, while the latter is either unimodal (when $m = 1$) or multimodal (when $m > 1$). When the recovered data $\boldsymbol{x}$ constitute the correlation coefficient (4) for the image *Q*, then the *true* likelihood ratio for the pair of hypotheses (2.1) becomes

$$V=\frac{f_C(r(\Delta_Q,\boldsymbol{R}_C))}{f_{\overline{C}}(r(\Delta_Q,\boldsymbol{R}_C))} \tag{10}$$

For the sake of brevity, hereafter we write r_0 for $r(\boldsymbol{\Delta}_Q, \boldsymbol{R}_C)$, $r_{C,j}$ for $r(\Delta_{C,j}, \boldsymbol{R}_C)$, and $r_{B_i,j}$ for $r\left(\Delta_{B_i,j},\boldsymbol{R}_C\right)$. We need to

estimate the densities f_C and $f_{\overline{C}}$ from the empirical data (8) and (9), or, more specifically, we need to use these data to estimate the numerator and denominator of (10). There are two main ways to accomplish that task: parametric and nonparametric estimation. The former requires functional forms of f_C and $f_{\overline{C}}$, whereas the latter can be used with kernel estimation (11,12).

Parametric Density Estimation

The generalized Gaussian distribution function has been suggested (13) as a suitable model of correlations obtained from a single camera. The density (GGD) has the form

$$f(x\,|\mu,\sigma,\beta) = \frac{\beta}{2\sigma\cdot\Gamma(1/\beta)}\sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}\exp\left\{-\left(\frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}\cdot|x-\mu|\right)^{\beta}\right\} \tag{11}$$

where μ and σ are the population mean and standard deviation, respectively, β is a shape parameter, and Γ is the gamma function. With $\beta = 2$, (2.11) becomes an ordinary Gaussian density. Assuming this functional form of f_C, we can estimate it by replacing the parameters μ, σ, and β with point estimates obtained from the empirical data (2.8). More precisely, we set

$$f_C(x) = f(x\,|\mu_0,\sigma_0,\beta_0) \tag{12}$$

where μ_0, σ_0, and β_0 are the mean, standard deviation, and shape of Ψ_C, respectively. For $f_{\overline{C}}$, we use a mix of GGD functions:

$$f_{\overline{C}}(x) = \sum_{i=1}^{m} w_i\cdot f(x\,|\mu_i,\sigma_i,\beta_i) \tag{13}$$

where μ_i, σ_i, and β_i are the mean, standard deviation, and shape of the subpopulation of $\Psi_{\overline{C}}$ defined by the correlations of images taken with camera B_i, $i = 1, \ldots, m$. The weights $w_1, \ldots, w_m$ are chosen to be the relative sizes of the sets (9), i.e., $w_i = n_{B_i}/\sum_{j=1}^{m} n_{B_j}$.

The parameters $\mu_0, \mu_1, \ldots, \mu_m$; $\sigma_0, \sigma_1, \ldots, \sigma_m$; $\beta_0, \beta_1, \ldots, \beta_m$ can be, respectively, estimated by the maximum-likelihood (ML) method or by the method of moments (MM). However, the former technique can be problematic, because an ML estimate of β may not exist, or there may be convergence difficulties associated with numerical determination of that estimate. Conditions for the existence of such an estimate and algorithms for investigating its existence and computation have previously been proposed (11). Application of those conditions to the present case is described in Appendix 1. We do not recommend the use of ML estimation when computing likelihood ratios as part of daily casework at a forensic laboratory. By comparison, although MM estimation is not free of problems, that method is easier to handle. In short, it is not difficult to estimate μ, and that estimate can be used to transform the data so that they satisfy the case $\mu = 0$, in which there are conditions that allow the other estimates to exist. Empirical studies (13) have shown that MM estimates of σ and β can be found, if the sample size is fairly large (above 16) and the population value of β is below 4. Explicit formulas for the MM estimates are given in Appendix 1. We can now estimate the true likelihood ratio (10) by

$$\hat{V} = \frac{f\left(r_0|\hat{\mu}_0,\hat{\sigma}_0,\hat{\beta}_0\right)}{\sum_{i=1}^{m} w_i\cdot f\left(r_0|\hat{\mu}_i,\hat{\sigma}_i,\hat{\beta}_i\right)} \tag{14}$$

where the point estimates $\hat{\mu}_0, \hat{\sigma}_0, \hat{\beta}_0$ and $\hat{\mu}_i, \hat{\sigma}_i, \hat{\beta}_i$, $i = 1, \ldots, m$ are either ML or MM estimates.

Nonparametric Kernel Density Estimation

A kernel is a symmetric non-negative function $K(y)$ satisfying $\int_{-\infty}^{\infty} K(y)dy = 1$. From a probabilistic point of view, a kernel can be thought of as a probability density function for a continuous random variable with mean and median equal to zero. Very often, the standard normal density function $\varphi(y) = (2\pi)^{-1/2}\exp(-y^2/2)$ is used as what is referred to as a Gaussian kernel. A kernel estimate of a probability density function f evaluated at x_0 and based on a random sample $(x_1, \ldots, x_n)$ from a population with this probability density is

$$\hat{f}_h(x_0) = \frac{1}{n\cdot h}\sum_{j=1}^{n} K\left(\frac{x_0 - x_j}{h}\right) \tag{15}$$

where h is what is known as bandwidth for the estimation. The choice of K is usually not very important for the properties of this estimate, whereas the choice of bandwidth is crucial. Therefore, we use a Gaussian kernel for all sets of correlations, but allow for different bandwidths. A kernel density estimate of f_C (r_0) is now suggested as

$$\hat{f}_{C,h_C}(r_0) = n_C^{-1}h_C^{-1}\sum_{j=1}^{n_C}\varphi\left(h_C^{-1}\cdot\left(r_0 - r_{C,j}\right)\right) \tag{16}$$

and a kernel density estimate of $f_{\overline{C}}(r_0)$ is proposed as

$$\hat{f}_{\overline{C},h_{\overline{C}}}(r_0) = n_{\overline{C}}^{-1}h_{\overline{C}}^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_{B_i}}\varphi\left(h_{\overline{C}}^{-1}\cdot\left(r_0 - r_{B_i,j}\right)\right) \tag{17}$$

Note the difference between (13) and (17). In the latter, we do not weigh individual density estimates, because this is taken care of "automatically" by the kernel technique. Potential multimodalities in $f_{\overline{C}}$ will also occur in $\hat{f}_{\overline{C},h_{\overline{C}}}$, provided that the bandwidth $h_{\overline{C}}$ is properly chosen. The estimates (16) and (17) are combined to give the *kernel likelihood ratio estimate*

$$\hat{V}_{h_C,h_{\overline{C}}} = \frac{n_C^{-1}h_C^{-1}\sum_{j=1}^{n_C}\varphi\left(h_C^{-1}\cdot\left(r_0 - r_{C,j}\right)\right)}{n_{\overline{C}}^{-1}h_{\overline{C}}^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_{B_i}}\varphi\left(h_{\overline{C}}^{-1}\cdot\left(r_0 - r_{B_i,j}\right)\right)} \tag{18}$$

As previously mentioned, the bandwidth h is essential for obtaining a good kernel density estimate. According to other investigators (14), the bandwidth that minimizes the mean square error of the estimate of $f_\bullet(r_0)$ (f_C or $f_{\overline{C}}$) using a Gaussian kernel is

$$h_l(r_0) = \left(\frac{f_\bullet(r_0)}{n\cdot 2\pi\cdot\left(f_\bullet''(r_0)\right)^2}\right)^{1/5} \tag{19}$$

where we have used the notation h_l to indicate that this is a locally optimized bandwidth. A *global* optimal bandwidth can be derived by integration over the space of r-values in (19).

In our analysis, the density function and its second derivative are approximated by fitting generalized Gaussian densities, but in other applications that may instead be achieved by ordinary Gaussian density fitting. It is possible that this step is not crucial for the success of the estimation, but that is not being investigated here. We

point out that there are several ways of choosing the bandwidth that are more or less data driven, although it seems that (19) and its global counterpart appear most frequently in the literature and free sites and software.

Example 1—A system camera C is used as the imaging device in question, and a reference pattern $\boldsymbol{R}_C$ is calculated from 50 images taken with that camera. Comparison is made with two other system cameras of the same kind, which are designated B_1 and B_2. With camera C, $n_C = 204$ images are captured to form a set of correlations (cf. [8]) for density estimation. For the same purpose, $n_{B_1} = 109$ and $n_{B_2} = 201$ images are shot using cameras B_1 and B_2, respectively (cf. [9]). Figure 2 shows the estimated density functions (denoted f_C and $f_{(B_1,B_2)}$ in the graphs) obtained using parametric GGD estimation and kernel density estimation with locally optimized bandwidths, in both together with plotted log-likelihood ratios ($\ln \hat{V}$) for different values of r_0. It is obvious from Fig. 2 that the parametric estimation in this case is superior to the kernel density estimation, in the sense that likelihood ratios increase monotonically with increasing r_0. This can be explained by the fairly small sample sizes used for the density estimation. In regions where sample values are sparse, the precisions of the kernel density estimates are low.

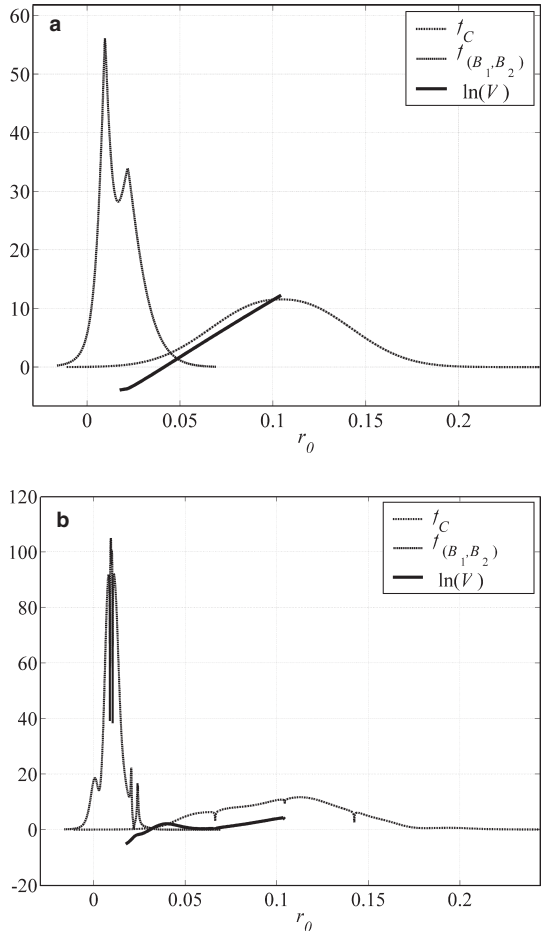

FIG. 2—*Log-likelihood ratios versus correlation (r_0) of an image of interest (solid lines), with parametric GGD estimation (a) and kernel density estimation (b) applied to correlations of images obtained with one particular camera (dashed dotted lines) and with two alternative cameras (dashed lines).*

It should also be kept in mind that the estimated likelihood ratio $\hat{V}$ will increase dramatically with r_0 in the case of parametric estimation. This is not shown in Fig. 2, but can be easily understood.

Assessment Methodology

Let us now consider inferential assessment of the likelihood ratio estimates (14) and (18).

Such a procedure usually involves evaluation of bias and variance, performing the latter either to establish error bounds or to test hypotheses. From a jurisdictional perspective, it is of particular interest to provide lower error bounds for the estimated likelihood ratios (to give the benefit of the doubt). There are both frequentistic and Bayesian approaches to do this (5,6), where the former is about confidence intervals and the latter is about credible intervals or highest posterior density regions. In a full Bayesian approach, the true likelihood ratio (the Bayes factor) is considered as a function of the parameters, each with a prior distribution and the prior density for that function needs to be derived and integrated with the sampling distribution of the estimated likelihood ratio to obtain the posterior density. The integration part would be carried out with application of a Markov chain Monte Carlo method, but the derivation of the prior and the sampling distributions is more complex. Frequentistic confidence intervals can be obtained either analytically by use of exact sampling distributions (if possible) or asymptotic sampling distributions, or by application of resampling techniques.

In this work, we propose and evaluate two frequentistic approaches: use of asymptotic sampling distributions and resampling (bootstrap). The former strategy is more straightforward but requires sufficiently large sample sizes, whereas the latter is more flexible with regard to sample size but needs efficient computer programming.

Three important things about the assessment of the likelihood ratio estimates should be mentioned before we proceed to the different approaches. The first of these is that the evaluation here is carried out with respect to the uncertainty induced by the samples of correlations used to estimate the density functions. The correlation coefficient of the image of interest is treated as a fact and not as a random observation. Accordingly, the procedure might be regarded as something between Bayesian and frequentistic, where the former governs the hypothesis testing, and the latter is a tool for likelihood estimation. There is no conflict in this, but merely a simplification compared with an otherwise very complex full Bayesian approach.

The second important aspect of the present assessment is that inferential properties are derived for the natural logarithm of the estimated likelihood ratio, i.e., $\ln \hat{V}$, hereafter referred to as the log-likelihood ratio. The reason for this is partly that calculations are simpler, but also that the results will be more consistent with the mathematical conditions of the likelihood ratio. A confidence interval for V obtained from direct estimation may cover negative values that are outside the range of V. A confidence interval for $\ln V$ transformed to the original scale covers only positive values, but can, of course, still be less informative if the lower limit is far below 1. A drawback of such confidence intervals is that they will always be right-skewed when transformed to original scale. However, as the upper limit is seldom of interest when examining forensic evidence, this is of minor importance for the conclusions in a real case. In the following sections, we derive formulas for the likelihood ratio in original scale, but, for the sake of clarity, the examples given illustrate the log-likelihood ratio.

The third aspect of interest is that we restrict our assessment to cases involving only one alternative camera; in other words, the task is to determine which of the two cameras was used to produce a particular image. This restriction does not result in loss of the generality of the assessment methodology, but it does make the mathematical calculation more transparent, and it also facilitates the validation of the methodology discussed in a later section. This validation is not intended to be adequate, but merely to serve as verification that the assessment can work in practice.

Assessment Using Asymptotic Sampling Distributions

When considering only one alternative camera, the natural logarithm of the likelihood ratio (14) becomes

$$\hat{v} = \ln f(r_0 \,|\hat{\mu}_0, \hat{\sigma}_0, \hat{\beta}_0) - \ln f(r_0 \,|\hat{\mu}_1, \hat{\sigma}_1, \hat{\beta}_1) \tag{20}$$

To avoid the inclusion of too many parameter symbols in the expressions below, we set $\boldsymbol{\Theta} = (\mu_0, \sigma_0, \beta_0, \mu_1, \sigma_1, \beta_1)$, i.e., the vector of all parameters and $\hat{\boldsymbol{\Theta}} = (\hat{\mu}_0, \hat{\sigma}_0, \hat{\beta}_0, \hat{\mu}_1, \hat{\sigma}_1, \hat{\beta}_1)$, the vector of parameter estimates, and write

$$v = \ln f(r_0 \,|\mu_0, \sigma_0, \beta_0) - \ln f(r_0 \,|\mu_1, \sigma_1, \beta_1) = v(\boldsymbol{\Theta}) \tag{21}$$

$$\hat{v} = v(\hat{\boldsymbol{\Theta}}) \tag{22}$$

assuming $\hat{\Theta}$ is either a maximum-likelihood estimate $\hat{\boldsymbol{\Theta}}_{ML}$ or a method-of-moments estimate $\hat{\boldsymbol{\Theta}}_{MM}$. Some of the results presented below are not related to the method of estimation, and thus we mostly use the generic notation $\hat{\boldsymbol{\Theta}}$, except when discussing the choice of estimation method. Bias-adjusted estimates of v and V can be obtained as

$$\hat{v}_b = \hat{v} - \mathrm{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}})) \Big/ 2 \tag{23}$$

$$\hat{V}_b = \hat{V} \Big/ \exp(\mathrm{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}})) \Big/ 2) \tag{24}$$

where $\mathbf{H}_v(\hat{\boldsymbol{\Theta}})$ stands for the Hessian of the log likelihood and $\mathbf{J}(\hat{\boldsymbol{\Theta}})$ for the observed Fisher information. Details about the derivation can be found in Appendix 2. Lower $100(1-\alpha)\%$ confidence limits v_{LO} and V_{LO} for v and V, respectively, can be estimated as follows:

$$v_{\mathrm{LO}} = \hat{v} - bias(\hat{v}) - z_\alpha \sqrt{\nabla_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \nabla_v(\hat{\boldsymbol{\Theta}})^{\mathrm{T}} + 2 \cdot \mathrm{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}))} \tag{25}$$

$$V_{\mathrm{LO}} = \exp\{v_{\mathrm{LO}}\} \tag{26}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$, i.e., z_α is the $100(1-\alpha)$th percentile of the standard Gaussian distribution. Again, details can be found in Appendix 2.

Example 2—Using the sets of correlations obtained for images taken with cameras C and B_1 in Example 1, we estimate the densities f_C and $f_{\overline{C}} = f_{B_1}$, and calculate $\hat{v}$ from (20), $\hat{v}_b$ from (23), and v_{LO} with $\alpha = 0.05$ from (25) for a range of potential values of r_0. Densities and graphs of calculated statistics are shown in Fig. 3. The bias is negligible in size, although it does increase with increasing r_0. However, the lower confidence limit is problematic, because, with increasing r_0, the lower limit for v should be >0, if it is to provide statistical evidence that the likelihood ratio V is above 1 and hence supports the hypothesis that camera C is the source of the image in question. Notably, the graph in Fig. 3 indicates the opposite, and even if the confidence interval could be expected to be wider when r_0 deviates from the more data-dense parts on the axis, it is not acceptable that the lower limit degenerates in this way. The reason for this behavior lies in the sample sizes, which, from an asymptotic standpoint, are quite modest. Notwithstanding, in a real case we want to achieve satisfactory outcomes without having to take too many images with the cameras of interest, and therefore the use of asymptotic results cannot be recommend when the sample sizes are in the order of magnitude seen in the current example.

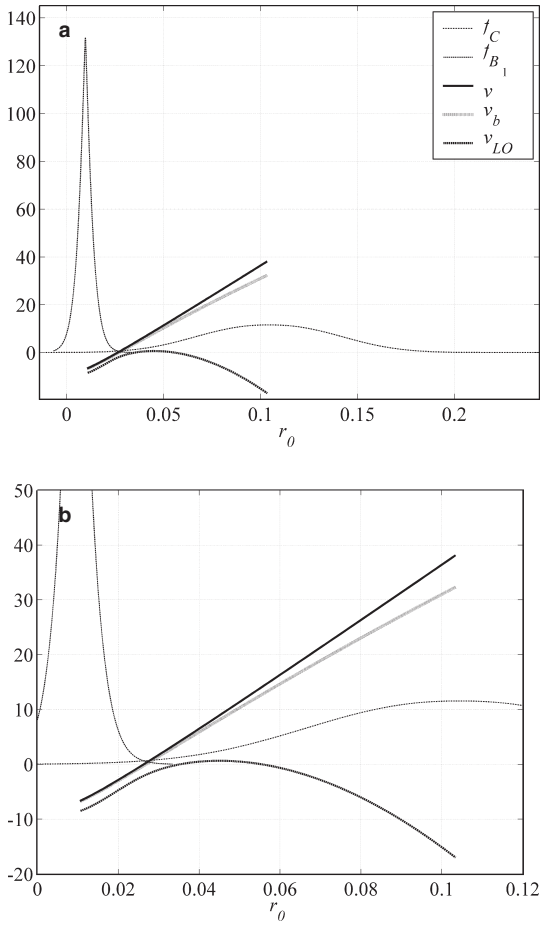

FIG. 3—*Results of an assessment using asymptotic distributions. Estimates (solid line), bias-adjusted estimates (dotted line), and lower 95% confidence limits (dashed line) for the log-likelihood ratios were calculated for a range of potential values of* r_0 *as evidence that one camera, but not another, was the source of a particular image. ([b] shows the central parts augmented.)*

Bootstrap Assessment

The bootstrap method is now widely used and has had ramifications in several areas of statistics. In many cases, it is necessary to modify the original idea (15), for example, to account for differences in distributional moments (such as means and variances) between sampling points or correlations between obtained values. Nevertheless, here we use the original idea, including both

parametric and nonparametric resampling, but we are especially cautious when determining the variance of the statistics in the assessment step. For a full description of the methodology, see papers published by other investigators (16,17).

Parametric Resampling

Inferential assessment of parametric models is commonly carried out in a "classical" way, either by direct derivation of small-sample properties of the estimators or, as in a previous section, by asymptotic inference. Less information has been published about the possibility of using resampling methods. However, if we apply the model (11), in which obtained correlation coefficients have a generalized Gaussian distribution, and we have estimates $\hat{f}_C$ and $\hat{f}_{\overline{C}}$ of the density functions based on this model, we can resample from the population defined by these density estimates. A resampling inferential assessment would then entail estimation of the bias and variance of $\hat{V}$ as

$$\text{bias}_*\left(\hat{V}\right) = E\left(\hat{V}\middle|f_C = \hat{f}_C, f_{\overline{C}} = \hat{f}_{\overline{C}}\right) - \hat{V} \tag{27}$$

$$\text{var}_*\left(\hat{V}\right) = \text{Var}\left(\hat{V}\middle|f_C = \hat{f}_C, f_{\overline{C}} = \hat{f}_{\overline{C}}\right) \tag{28}$$

The estimates (27) and (28) are commonly referred to as bootstrap estimates of bias and variance. The notation with an asterisk emphasizes that conclusions drawn within the empirical distributions defined by $\hat{f}_C$ and $\hat{f}_{\overline{C}}$ are transferred to the original population according to the bootstrap principle (16). Although (27) and (28) are theoretically correct, they are seldom easy to derive analytically, and it is for that reason that Monte Carlo simulation is often used. However, we would like to point out that performing bootstrapping or resampling does *not* necessarily imply simulation, which represents a common misunderstanding that can interfere with the conclusions that are drawn from a bootstrap procedure. Furthermore, (27) is not generally used for estimating bias, although it is valid when the parameter of interest in the original distribution (population) corresponds to the point estimate in the empirical distribution. This can be seen here, because our likelihood ratio estimate has the same form as the true likelihood ratio. Nonetheless, there are cases when the point estimate is not derived in that manner and, as a result, does not correspond to the population parameter.

For Monte Carlo simulation, let $\hat{F}_C$ and $\hat{F}_{\overline{C}}$ be the corresponding cumulative distribution functions (i.e., $\hat{F}_\bullet(r) = \int_{-\infty}^{r} \hat{f}_\bullet(y)dy$). Then, a random observation from the empirical population of potential correlations between an image of interest and camera C is

$$r^*(\Delta_C, \boldsymbol{R}_C) = \hat{F}_C^{-1}(U) \tag{29}$$

where U is a random quantity uniformly distributed over the interval (0,1) and $\hat{F}_C^{-1}$ is the inverse of $\hat{F}_C$. Replacing $\hat{F}_C^{-1}$ with $\hat{F}_{\overline{C}}^{-1}$ in (29) gives a random observation from the empirical population of potential correlations between the target image and the (set of) alternative camera(s). As before, we restrict our interest to the existence of only one alternative camera. By repeated use of (29) and its counterpart for $\overline{C}$, we can produce samples of values that are usually of the same sizes as the original samples. Such samples are commonly referred to as *pseudo-samples* of the original population. We define a *sample set* as comprising one sample using (29) and one sample using its counterpart for $\overline{C}$. From this set, we obtain parameter estimates $\hat{\mu}_0^*, \hat{\sigma}_0^*, \hat{\beta}_0^*, \hat{\mu}_1^*, \hat{\sigma}_1^*$, and $\hat{\beta}_1^*$ (ML or MM), and, analogous to (20), we can estimate the log-likelihood ratio

$$\hat{v}^* = \ln f\left(r_0 \middle| \hat{\mu}_0^*, \hat{\sigma}_0^*, \hat{\beta}_0^*\right) - \ln f\left(r_0 \middle| \hat{\mu}_1^*, \hat{\sigma}_1^*, \hat{\beta}_1^*\right) \tag{30}$$

Repeating this procedure (i.e., producing a sample set and calculating [3.7]), a suitable number of times will give an empirical distribution for $\hat{v}^*$ as a random variable. From this empirical distribution, it is possible to obtain estimates of bias and variance of $\hat{v}$ as an estimator of v. However, for many estimators, such measures can be improved by studentizing, provided that an estimate of its standard error can be obtained from a sample. Applying this procedure to (3.7), we obtain repeated versions of

$$T_{\hat{v}^*} = \frac{\hat{v}^* - \hat{v}}{\text{SE}_*(\hat{v}^*)} \tag{31}$$

where $\text{SE}_*(\hat{v}^*)$ is the estimate of standard error from a sample set. As a standard error estimator, we can use any statistic that is computed in closed form and is monotonically related to the standard deviation of $\hat{v}$; it does not have to be an unbiased estimator. We suggest the asymptotic standard deviation used in (3.4a), i.e.,

$$s_{\hat{v}} = \sqrt{\nabla_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \nabla_v(\hat{\boldsymbol{\Theta}})^{\mathrm{T}} + 2 \cdot \text{tr}(\mathrm{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \mathrm{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}))} \tag{32}$$

which is computable from any sample set. The point is that (32) can be calculated both with the original samples and with the resampled sets, and the properties obtained from the empirical distribution of $\hat{v}^*$ can be transferred to $\hat{v}$ using the same kind of standard error estimate. By applying (32) evaluated at $\hat{\boldsymbol{\Theta}}^* = (\hat{\mu}_0^*, \hat{\sigma}_0^*, \hat{\beta}_0^*, \hat{\mu}_1^*, \hat{\sigma}_1^*, \hat{\beta}_1^*)$ as the denominator of (31), we obtain the following bias-adjusted estimate and lower $100(1-\alpha)\%$ confidence limit for v:

$$\hat{v}_{*b} = \hat{v} - s_{\hat{v}} \cdot B^{-1} \sum_{b=1}^{B} T_{\hat{v}^*}^{(b)} \tag{33}$$

$$v_{*\text{LO}} = \hat{v} - s_{\hat{v}} \cdot F_T^{-1}(1-\alpha) \tag{34}$$

where B is the number of resampled sets, $T_{\hat{v}^*}^{(b)}$ is the statistic (31) of sample set b, and F_T^{-1} is the inverse of the empirical cumulative distribution function induced by the set of statistics $\{T_{\hat{v}^*}^{(1)}, \ldots, T_{\hat{v}^*}^{(B)}\}$.

Example 3—Using the same setup as in Example 2, Fig. 4 shows densities and graphs of calculated statistics based on $B = 200$ resampled sets. Compared to the assessment with asymptotic distributions (Fig. 3), the estimated bias is still negligible, but negative, and the lower confidence limit is much more stable, although it decreases with increasing r_0.

Nonparametric Resampling

Much of what has been discussed thus far with regard to parametric resampling can also be applied in nonparametric resampling. The main difference between those two approaches lies in the actual resampling of data. Instead of sampling from an estimated generalized Gaussian distribution, we resample directly from the original samples. Let $F_{n_C,C}$ and $F_{n_{\overline{C}},\overline{C}}$ represent the empirical cumulative distribution functions obtained from the sets of correlations (8) and (9) (here assuming the latter came from only one camera). The counterparts of (27) and (28) become

$$\text{bias}_*\left(\hat{V}\right) = E\left(\hat{V}\middle|F_C = F_{n_C,C}, F_{\overline{C}} = F_{n_C,\overline{C}}\right) - \hat{V} \tag{35}$$

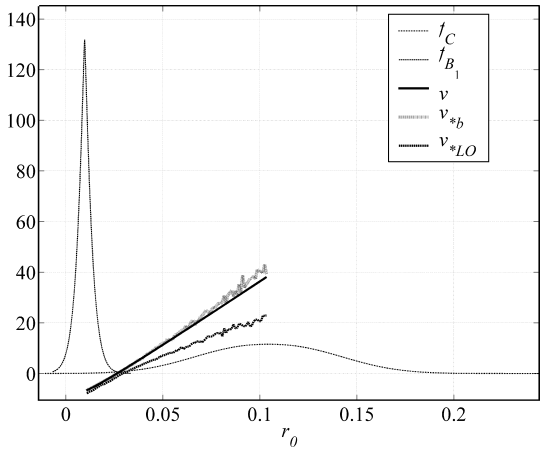

FIG. 4—*Results of a bootstrap assessment. Estimates (solid line), bias-adjusted estimates (dotted line), and lower 95% confidence limits (dashed line) for the log-likelihood ratios were calculated for a range of potential values of* r_0 *as evidence that one particular camera, but not another, was the source of an image in question.*

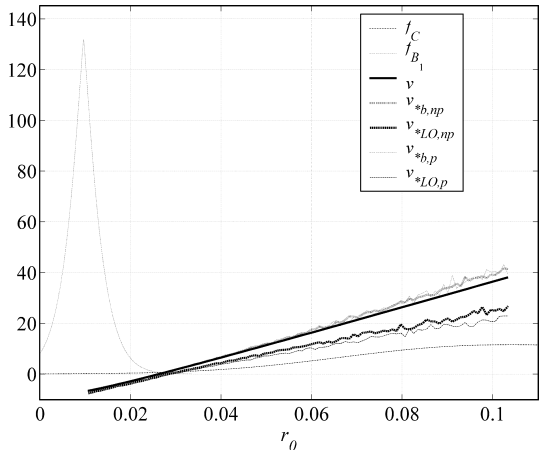

FIG. 5—*Results of a bootstrap assessment using nonparametric and parametric resampling. Estimates (solid line), bias-adjusted estimates (dotted lines), and lower 95% confidence limits (dashed lines) were calculated for the log-likelihood ratios of a range of potential values of* r_0 *as evidence that one camera, but not another, was the source of a particular image. The bold dotted and dashed lines indicate nonparametric resampling.*

$$\mathrm{var}_*\left(\hat{V}\right) = \mathrm{Var}\left(\hat{V}\middle| F_C = F_{n_C,C}, F_{\overline{C}} = F_{n_{\overline{C}},\overline{C}}\right) \tag{36}$$

where F_C and $F_{\overline{C}}$ denote the true cumulative distribution functions of the two sets of correlations. The resampling counterpart of (29) is

$$r^*(\Delta_C, \boldsymbol{R}_C) = F^{-1}_{n_C,C}(U) \tag{37}$$

with a corresponding expression for the resampling from $F_{n_{\overline{C}},\overline{C}}$. In practice, this means that we draw correlations independently of each other and with replacement from each of the two original samples to form a sample set.

For parametric estimation of v (and V), we can use the formulas (30) to (33, 34) given above without modifications. It might be argued that a parametric model should be accompanied by parametric resampling. However, even though such resampling can succeed if the point estimates of the parameters (μ, σ, and β) successfully reflect the cumulative distribution function (or the probability density function), the empirical cumulative distribution function (also an ML estimate) might still be a better alternative.

Example 4—Again, we choose the setup from Example 3, and, as in Example 3.2, we use 200 resampled sample sets. Figure 5 presents graphs of statistics calculated with nonparametric and with parametric resampling. The ranges of the axes are reduced compared to what is shown in Fig. 4 to make it easier to see the differences between the methods. The subscripts "np" and "p" are used to distinguish between statistics from the two resampling procedures. The bias-adjusted estimates are quite similar, whereas the lower confidence limits are somewhat higher for the nonparametric resampling.

Kernel Density Estimation

In a parametric resampling setup, the use of kernel density estimates would not seem consistent, because such estimation is performed when knowledge is lacking about the background population or there are difficulties associated with the estimation of some parameters. In contrast, it is natural to use nonparametric resampling in combination with kernel density estimation. However, with that approach, there is no longer a parametric standard error in closed form, and it is computationally expensive to calculate an estimate by using either the kernel density estimation procedure or double bootstrap (16,17). In this paper, we consistently use the estimate (3.9) motivated by the discussion preceding that equation, which emphasized that it is important that the estimate of standard error is monotonically related to the true standard deviation of the point estimate. We do not attempt to validate that that is actually the case, but it seems that there are no arguments supporting the opposite.

Example 5—Nonparametric resampling is used in the same setup as in Example 3.3. As in the previous examples, Fig. 6 shows graphs of calculated statistics, but in this case obtained by parametric and kernel density estimation, using a Gaussian kernel for the latter. The bandwidth is locally optimized for the original samples and retained in the resampled sets. For clarity, the kernel density-based statistics in Fig. 6 are denoted h_C and h_B. The curves in the graph may give a noisy impression, but all statistics for the kernel density estimation are far below the statistics for the parametric estimation. The estimated log-likelihood ratios, the bias-adjusted estimates, and the curve of lower confidence limits nearly coincide in the lower part of the graph. The illustrated results clearly demonstrate that kernel density estimation does not work satisfactorily for this data set.

Validation of the Methodology

The methods of assessment used in the previous section have not been justified by any theoretical calculations, and it may be argued that such an investigation is required to prove the reliability of the techniques. However, there is no other way to theoretically demonstrate reliability than to establish asymptotic validity. Furthermore, it cannot honestly be said that such results will actually provide any benefits, knowing that the methods in question will very seldom be applied to samples that are of sizes that can be assumed to exhibit asymptotic behavior. Nevertheless, we cannot accept a method that might fail even if a fairly large sample size is used, and hence we proceed here by performing two simulation studies to validate the methodology.

The assessment methodology to be validated is restricted to the comparison of two cameras. Our objective of the first simulation

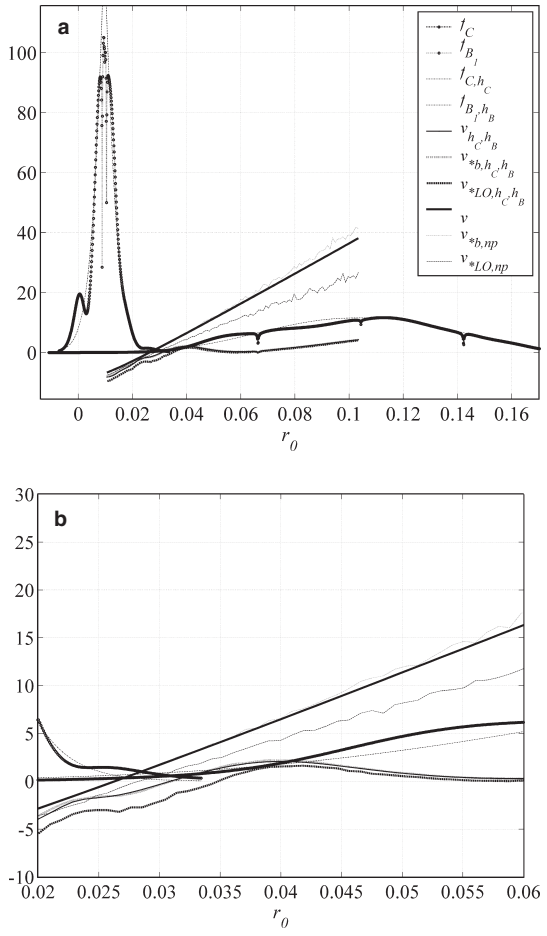

FIG. 6—*Results of a bootstrap assessment involving nonparametric resampling using parametric and kernel density estimation. Estimates (solid lines), bias-adjusted estimates (dotted lines), and lower 95% confidence limits (dashed lines) were calculated for the log-likelihood ratios of a range of potential values of* r_0 *as evidence that one camera, but not another, was the source of a particular image. The three lowest statistics curves all represent kernel density estimation. (The figure to the right shows the central parts augmented.)*

study is to show how distances between the means of the populations of correlations for the two cameras affect the estimated log-likelihood ratio, the bias of this estimate, and the lower confidence limit for the true log-likelihood ratio. Data are simulated from generalized Gaussian distributions, where the mean of the distribution for a particular camera is fixed and the other mean varies. To account for differences in shape, a prior distribution is used for this parameter. In the calculations, the standard deviation will vary with the mean, but the coefficient of variation will be constant. There are several ways of choosing the parameters, and it is clear that a simulation study of this type cannot be fully adequate. However, our aim is not to prove general validity, but to give indications of how well the methods suggested earlier may work. Details about the distributional setup are given in Table 1.

The sample sizes used are $n_C = 2000$ and $n_{\overline{C}} = 2000$, respectively, because they lie between what could be expected to be used in daily casework and what can be assumed to be sufficient for approaching asymptotic properties. Moreover, it would not be infeasible, albeit very tedious, to take 2000 images with each of the cameras involved.

For each pair of samples (sample set), we apply five different methods of assessment:

TABLE 1—*Distributional setup of the first simulation study.*

	"Camera"	
Simulation Parameter	Questioned	Alternative
Mean (μ)	0.15	Varying*
Shape (β)	$\sim 2 \cdot Beta(5,5)^{\dagger} + 1$	$\sim 2 \cdot Beta(5,5)^{\dagger} + 1$
Coefficient of variation (σ/μ)	0.5	0.5
r_0	0.0869[§]	

*The mean for the alternative camera is varied in five equidistant descending steps, starting with the value for which the true likelihood ratio V is equal to 1 when β is equal to 2, and stopping with the value for which V is equal to 10,000 when β is equal to 2.

†The correlation in question (r_0) is chosen as the 20th percentile of a generalized Gaussian distribution with $\mu = 0.15$, $\beta = 2$, and $\sigma = 0.5$, $\mu = 0.075$ (i.e., a Gaussian distribution). This corresponds to the "mean" simulated density for the camera of interest.

§*Beta*(5,5) denotes a beta distribution with parameters 5 and 5. The linear transformation sets the range from 1 to 3 with mean 2.

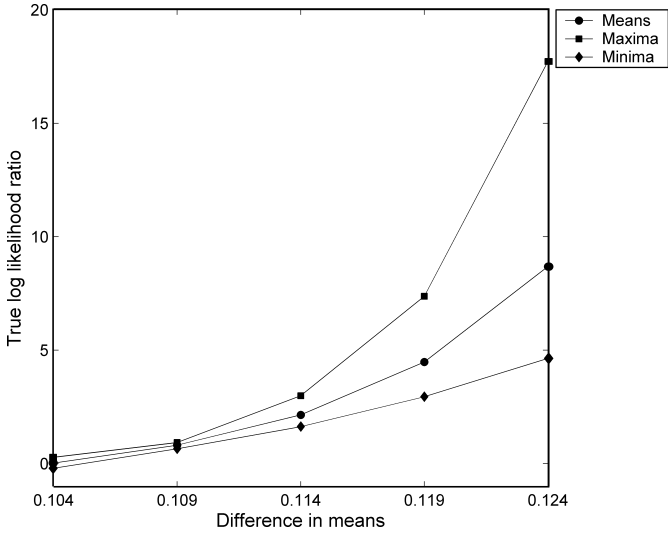

FIG. 7—*Means, maxima, and minima of true likelihood ratios for all pairs of distributions used in the simulation study.*

Method 1: Parametric estimation and assessment with asymptotic distributions.

Method 2: Parametric estimation and assessment with parametric resampling.

Method 3: Parametric estimation and assessment with nonparametric resampling.

Method 4: Kernel density estimation with nonparametric resampling, and bandwidth locally optimized for the original sample set and retained in resampled sets.

Method 5: Kernel density estimation with nonparametric resampling, and bandwidth locally optimized for each sample set (original and resampled).

For each method and each combination of means, 300 original sample sets are generated. Because the true likelihood is known for each combination of means, we can compare each set of 300 assessments with the truth. In Fig. 7, the ranges and means of the true likelihood ratio are plotted against the difference in distribution means for all pairs of distributions used in the simulation. It might seem strange that a single value is not used for each difference in means, but the reason for this is that the shape varies randomly

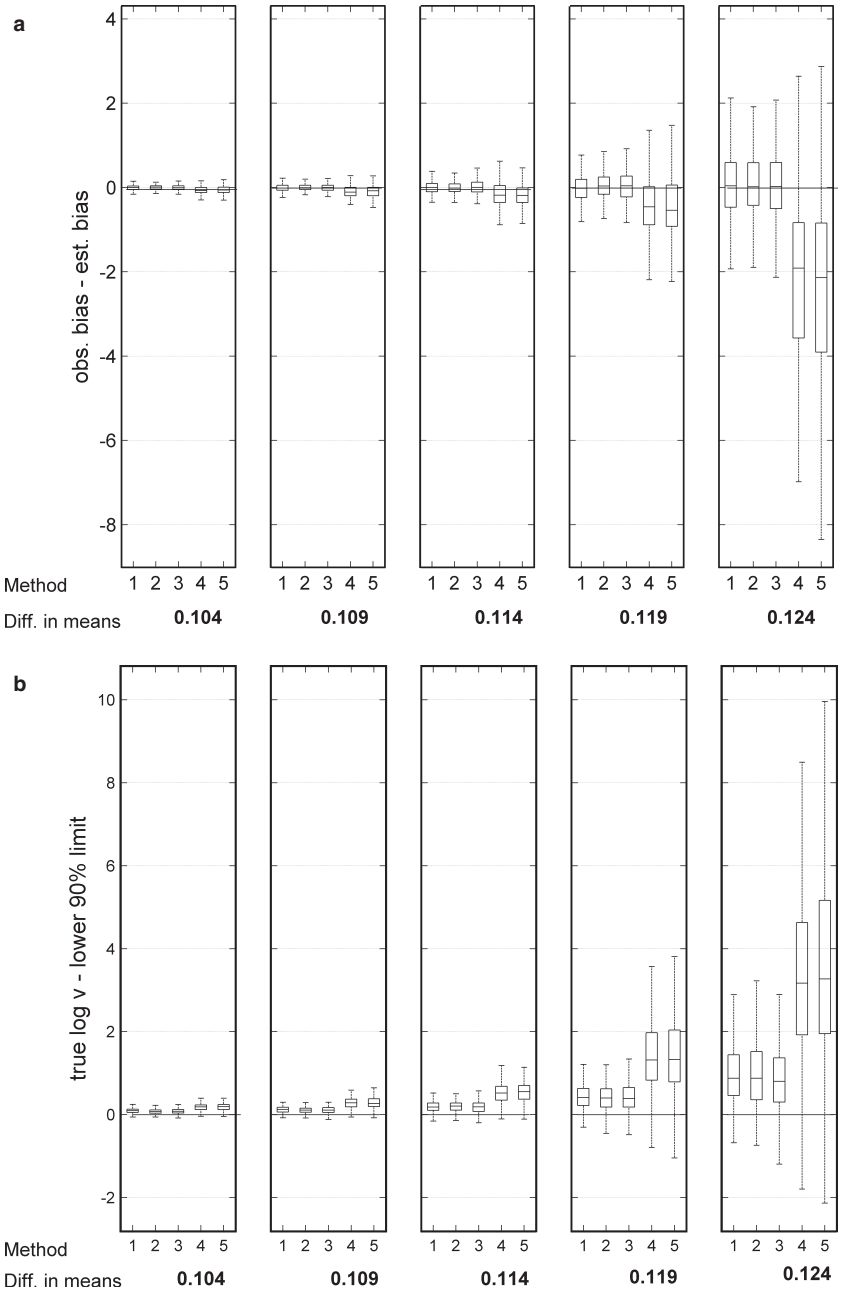

FIG. 8—*Differences between observed bias and estimated bias (a) and between the true log-likelihood ratio and the lower 90% confidence limit (b) calculated using methods 1–5 for each difference in means in a pair of distributions.*

between the distributions. Notably, the range increases with the difference in means.

To show how the bias estimation works, we compare the biases that can actually be observed (i.e., $\hat{v} - v$) with the estimated biases. Differences between the observed and the estimated bias are shown in Fig. 8*a* as box plots for each of the methods and for each difference in the means. It is obvious that the estimation is satisfactory when using methods 1, 2, and 3, whereas it collapses with methods 4 and 5. It should also be noted that the variation increases with the difference in means but is still symmetric around zero for the first three methods.

In a similar manner, we compare calculated 90% lower confidence limits for the log-likelihood ratio (v_{LO}) with the true log-likelihood ratio (v), and in Fig. 8*b* this is shown as box plots for each of the methods and each difference in the means. For methods 1, 2, and 3, the variation is small (although it increases with the difference in means), and the box plots show that the confidence limits are satisfactorily close to the true log-likelihood ratio. However, for methods 4 and 5, intervals are wider, because the boxes are more extensively separated from zero. Table 2 gives the attained coverage levels for the one-sided 90% confidence intervals defined by the lower limits for each method. Not surprisingly,

TABLE 2—*Attained coverage levels for one-sided 90% confidence intervals obtained in the first simulation study.*

Method	1	2	3	4	5
Coverage level (%)	92.3	90.1	90.5	95.1	95.7

TABLE 3—*Simulation setup and attained coverage levels for one-sided 90% confidence intervals obtained in the second simulation study. A total of 500 sample sets were simulated.*

	Camera in Question			Alternative Camera					
Method	Mean μ_C	Shape β_C	CV $\frac{\sigma_C}{\mu_C}$	Mean $\mu_{\overline{C}}$	Shape $\beta_{\overline{C}}$	CV $\frac{\sigma_{\overline{C}}}{\mu_{\overline{C}}}$	Sample Size*	Likelihood Ratio (V) for $r_0 = 0.0869$	Coverage Level 90% CI†
1	0.15	2.10	0.5	0.0257	1.90	0.5	2000	4655	0.926 (0.012)
2									0.878 (0.015)
3									0.906 (0.013)

*The sample size is used for both cameras.
†Lower-bounded confidence intervals. Standard error is given within parentheses.

methods 4 and 5 give higher coverage levels than the designed 90% intervals, and it is also apparent that method 1 differs from methods 2 and 3 in that the latter two techniques result in coverage levels that are substantially closer to the designed 90% intervals. These observations agree with our finding that the assessment with asymptotic distributions described earlier proved to be less satisfactory than the parametric and nonparametric bootstrap. Furthermore, it might be said that all the disappointing results obtained using methods 4 and 5 were indeed expected, considering the poor behavior of those techniques in the examples given in the previous section.

In the second simulation study, we investigate the coverage levels of the proposed confidence intervals with methods 1, 2, and 3 above. As before, we limit ourselves to two cameras. The populations of correlations obtained with the two cameras are modeled with generalized Gaussian distributions with different means and shapes and the value of r_0 is set to 0.0869, which corresponds with a true likelihood ratio of 4655 under the models used. The sample sizes are $n_C = 2000$ and $n_{\overline{C}} = 2000$, respectively, and a total of 500 original sample sets are generated. The difference between the two means equals the largest difference used in the first simulation study. However, in this second simulation study, the shapes are fixed, which means that we are studying two fixed cameras. In Table 3 are given details of the distributions used and attained coverage levels with standard errors for each of the three methods investigated. Comparing with the results in Table 2, we note that the attained coverage level with method 1 still exceeds the nominal level and does so even within 95% error bounds ($0.926 \pm 2 \times 0.012$). The attained coverage level for method 2 does not reach the nominal 90%, but the latter is within the 95% error bounds ($0.878 \pm 2 \times 0.015$). For method 3, the attained coverage level is satisfactory.

Conclusions

In this paper, we have demonstrated how a number of different methods can be employed to assess a likelihood ratio (or actually the logarithm of a likelihood ratio) that is calculated from density functions estimated from available data. Use of these techniques specifically to assess digital images as forensic evidence is in itself an interesting future application of likelihood ratios, but the methodology also offers more general possibilities for evaluating evidence in almost any area of forensics in which continuous distributions govern the outcome of the findings. It is important to emphasize that this type of assessment is carried out in relation to the background data. In the application in focus in our study, such data can be produced for each particular case, whereas in other applications background data can exist in the form of compiled databases. An alternative approach would be to work with the true Bayes factor and prior distributions for the parameters of the numerator and denominator densities. The uncertainty of the Bayes factor would then lie in the choice of prior densities, but this has not been investigated in the current study.

Considering the examples and the validation study described in this paper, it seems that resampling represents the most successful approach, regardless of whether it is carried out in a parametric or a nonparametric fashion. Furthermore, it appears that densities should be estimated from parametric models. Also, assessment using asymptotic distributions does have potential, but in general it requires larger sample sizes. This means that, in cases when extensive databases are available, it can be expected that this type of assessment will be successful, provided we can obtain reasonable parameterizations of the data. Of course, it can be suitable to employ resampling in cases involving enormous amounts of background data, although that may be very time-consuming and require extensive computer efforts.

Estimating density functions with kernel smoothers was not successful in our examples. However, we have no reason to believe that such methods will collapse if larger amounts of background data are used. Even so, it is important to realize that we cannot expect to obtain good results from either the calculation of a likelihood ratio or the assessment of that ratio, if data from the forensic findings are far out in the tails of any of the distributions involved. Inasmuch as the kernel density estimates are too uncertain in such cases, the results are not reliable.

Acknowledgments

The authors direct special thanks to CSE Patricia Ödman for the language revision and to the two reviewers of this manuscript for valuable comments.

References

1. Aitken CGG, Taroni F. Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Chichester, U.K.: John Wiley and Sons Ltd, 2004.
2. Garthwaite P, Jolliffe I, Jones B. Statistical inference, 2nd edn. New York, NY: Oxford University Press, 2002.
3. Buckleton J. A framework for interpreting evidence. In: Buckleton J, Triggs CM, Walsh SJ, editors. Forensic DNA evidence interpretation. Boca Raton, FL: CRC Press, 2005;27–63.
4. Balding DJ. Estimating products in forensic identification using DNA profiles. J Am Stat Assoc 1995;90:839–44.
5. Curran JM, Buckleton JS, Triggs CM, Weir BS. Assessing uncertainty in DNA evidence caused by sampling effects. Sci Justice 2002;42(1): 29–37.
6. Curran JM. An introduction to Bayesian credible intervals for sampling error in DNA profiles. Law Probab Risk 2005;4:115–26.
7. Dujourdy L, Barbati G, Taroni F, Guéniat O, Esseiva P, Anglada F, et al. Evaluation of links in heroin seizures. Forensic Sci Int 2003;131: 171–83.

8. Alberink I, Bolck A. Obtaining confidence intervals and likelihood ratios for body height estimations in images. Forensic Sci Int 2008;177:228–37.
9. Höglund T. Digital camera identification—a brief test of the possibility to identify a camera based on the sensor noise. J Forensic Identif 2009; 59(5):475–501.
10. Lukáš J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensic Secur 2006;1(2):205–14.
11. Parzen E. On estimation of a probability density function and mode. Ann Math Stat 1962;33(3):1065–76.
12. Silverman BW. Density estimation for statistics and data analysis. New York, NY: Chapman and Hall, 1986.
13. Meignen S, Meignen H. On the modeling of small sample distributions with generalized Gaussian density in a maximum likelihood framework. IEEE Trans Image Process 2006;15(6):1647–52.
14. Katkovnik V, Shmulevich I. Kernel density estimation with adaptive varying window size. Pattern Recognit Lett 2002;23:1641–8.
15. Efron B. Bootstrap methods—another look at the jackknife. Ann Stat 1979;7:1.
16. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York, NY: Chapman & Hall, 1993.
17. Hjorth JSU. Computer intensive statistical methods: validation, model selection and bootstrap. London, U.K.: Chapman & Hall, 1993.

Additional information and reprint requests:
Anders Nordgaard, Ph.D.
The Swedish National Laboratory of Forensic Science
SE-581 94 Linköping
Sweden
E-mail: Anders.Nordgaard@liu.se

Appendix 1—Conditions for the existence of method-of-moments estimators

MM estimates of σ and β for a GGD with zero mean exist, provided that

$$\left(n^{-1}\sum|x_j|^2\right)^2\Big/\left(n^{-1}\sum|x_j|\right)<3/4,$$

where $(x_1, \ldots, x_n)$ is the random sample to be used for the estimation (13). Translated to the setup of the present study (the samples defined by [8] and [9]), we require

$$\begin{aligned}&\frac{\left(n_C^{-1}\sum_{j=1}^{n_C}\left|r\left(\Delta_{C,j},\boldsymbol{R}_C\right)\right|^2\right)^2}{n_C^{-1}\sum_{j=1}^{n_C}\left|r\left(\Delta_{C,j},\boldsymbol{R}_C\right)\right|}<\frac{3}{4}\ \text{and}\\&\frac{\left(n_{B_i}^{-1}\sum_{j=1}^{n_{B_i}}\left|r\left(\Delta_{B_i,j},\boldsymbol{R}_C\right)\right|^2\right)^2}{n_{B_i}^{-1}\sum_{j=1}^{n_{B_i}}\left|r\left(\Delta_{B_i,j},\boldsymbol{R}_C\right)\right|}<\frac{3}{4};\ i=1,\ldots,m\end{aligned}\qquad(A1.1)$$

The conditions in (A1.1) are fairly easy to check by use of a routine for finding the estimates. In addition, these conditions are sufficient, but not necessary, for the MM estimates to exist.

The MM estimates of μ, σ, and β for any set of n correlations $\left\{r\left(\Delta_{\bullet,j},\boldsymbol{R}_C\right)\right\}_{j=1}^{n}$ are found by solving this system of equations:

$$\hat{\mu}=n^{-1}\sum_{j=1}^{n}r\left(\Delta_{\bullet,j},\boldsymbol{R}_C\right);\quad G(y)=\frac{[\Gamma(2/y)]^2}{\Gamma(3/y)\cdot\Gamma(1/y)}$$

$$M_1=n^{-1}\sum_{j=1}^{n}\left|r\left(\Delta_{\bullet,j},\boldsymbol{R}_C\right)-\hat{\mu}\right|;\quad\hat{\beta}=G^{-1}\left(M_1^2/M_2\right)$$

$$M_2=n^{-1}\sum_{j=1}^{n}\left|r\left(\Delta_{\bullet,j},\boldsymbol{R}_C\right)-\hat{\mu}\right|^2;\quad\hat{\sigma}=M_1\cdot\frac{\Gamma\left(1/\hat{\beta}\right)}{\Gamma\left(2/\hat{\beta}\right)}\cdot\sqrt{\frac{\Gamma\left(3/\hat{\beta}\right)}{\Gamma\left(1/\hat{\beta}\right)}}\qquad(A1.2)$$

Appendix 2—Mathematical derivation of bias-adjusted estimates and confidence limits in assessment using asymptotic distributions

Applying the second-order Taylor approximation[2] of (3.2*b*), we have

$$E\left\{v(\hat{\boldsymbol{\Theta}})\right\}\approx v(\boldsymbol{\Theta})+\operatorname{tr}(\mathrm{H}_v(\hat{\boldsymbol{\Theta}})\cdot\Sigma)\Big/2\qquad(A2.1a)$$

$$\begin{aligned}Var\left\{v(\hat{\boldsymbol{\Theta}})\right\}\approx{}&\nabla_v(\hat{\boldsymbol{\Theta}})\cdot\Sigma\cdot\nabla_v(\hat{\boldsymbol{\Theta}})^{\mathrm{T}}\\&+\phi\cdot\operatorname{tr}(\mathrm{H}_v(\hat{\boldsymbol{\Theta}})\cdot\Sigma\cdot\mathrm{H}_v(\hat{\boldsymbol{\Theta}})\cdot\Sigma)+\psi\end{aligned}\qquad(A2.1b)$$

where $\nabla_v(\hat{\boldsymbol{\Theta}})$ and $\mathrm{H}_v(\hat{\boldsymbol{\Theta}})$ are, respectively, the gradient and Hessian of v evaluated at $\hat{\boldsymbol{\Theta}}$; $\boldsymbol{\Sigma}$ is the variance–covariance matrix of $\hat{\boldsymbol{\Theta}}$; ϕ and ψ are factors depending on the third and fourth moments of $\mathbf{H}_v(\hat{\boldsymbol{\Theta}})$and on the covariance between $(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})\cdot\nabla_v(\hat{\boldsymbol{\Theta}})$ and $(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})\cdot\mathrm{H}_v(\hat{\boldsymbol{\Theta}})\cdot(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})^{\mathrm{T}}$. Choosing ϕ and ψ is a complicated task, which we simplify here by using $\phi=2$ (an overestimate of Gaussian-distributed $\hat{\boldsymbol{\Theta}}$) and ignoring asymptotic covariance through the use of $\psi\approx 0$. The gradient and Hessian in (A2.1*a*) and (A2.1*b*) can both be derived analytically, but we simplify the calculation of the Hessian by employing numerical differentiation. For the variance–covariance matrix $\boldsymbol{\Sigma}$, we use the observed Fisher information

$$\mathbf{J}\left(\hat{\boldsymbol{\Theta}}\right)=\left(-\frac{\partial^2 l(\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}^2}\Big|_{\boldsymbol{\Theta}=\hat{\boldsymbol{\Theta}}}\right)\qquad(A2.2)$$

where $l(\Theta)$ represents the log-likelihood function. It is valid to use (A2.2) as the approximate asymptotic variance–covariance matrix of a maximum-likelihood estimator $\hat{\boldsymbol{\Theta}}_{ML}$, whereas the validity of this approximation for the method-of-moments estimator $\hat{\boldsymbol{\Theta}}_{MM}$ has not been thoroughly investigated. Some empirical results show that when a maximum-likelihood estimate can be obtained, it deviates only slightly from the method-of-moments estimate. Thus, it is plausible that if we obtain a methods-of-moments estimate, the results below will hold for a likelihood ratio based on that estimate.

From (A2.1*a*) and (A2.2), we obtain approximate estimates of bias:

$$\operatorname{bias}\{\hat{v}\}=\operatorname{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}})\cdot\mathbf{J}(\hat{\boldsymbol{\Theta}}))/2\qquad(A2.3a)$$

$$\operatorname{relbias}\left(\hat{V}\right)=\exp(\operatorname{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}})\cdot\mathbf{J}(\hat{\boldsymbol{\Theta}}))/2),\qquad(A2.3b)$$

and we note that the estimated bias of $\hat{V}$ is relative. Bias-adjusted estimates of v and V are now obtained as

$$\hat{v}_b=\hat{v}-\operatorname{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}})\cdot\mathbf{J}(\hat{\boldsymbol{\Theta}}))/2\qquad(A2.4a)$$

$$\hat{V}_b=\hat{V}/\exp(\operatorname{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}})\cdot\mathbf{J}(\hat{\boldsymbol{\Theta}}))/2)\qquad(A2.4b)$$

For a maximum-likelihood estimate $\hat{\boldsymbol{\Theta}}_{ML}$, we can make use of well-known asymptotic distributional properties. By combining

[2] $v(\hat{\boldsymbol{\Theta}})\approx v(\boldsymbol{\Theta})+(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})\cdot\nabla_v(\hat{\boldsymbol{\Theta}})+(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})\cdot\mathrm{H}_v(\hat{\boldsymbol{\Theta}})\cdot(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta})^{\mathrm{T}}/2$

(A2.3*a*) with (A2.1*b*) and (A2.2), the lower $100(1-\alpha)\%$ confidence limits v_{LO} and V_{LO} for v and V, respectively, can be estimated as follows:

$$v_{\mathrm{LO}} = \hat{v} - bias(\hat{v}) - z_\alpha \sqrt{\nabla_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \nabla_v(\hat{\boldsymbol{\Theta}})^{\mathrm{T}} + 2 \cdot \mathrm{tr}(\mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{H}_v(\hat{\boldsymbol{\Theta}}) \cdot \mathbf{J}(\hat{\boldsymbol{\Theta}}))} \quad \text{(A2.5}a\text{)}$$

$$V_{\mathrm{LO}} = \exp\{v_{\mathrm{LO}}\} \quad \text{(A2.5}b\text{)}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$, i.e., z_α is the $100(1-\alpha)$th percentile of the standard Gaussian distribution. We also use these limits for a method-of-moments estimator $\hat{\boldsymbol{\Theta}}_{MM}$, taking into account that the validity of this may be questionable, especially in the tails of the true distributions of $\hat{v}$ and $\hat{V}$.