A Problem in Forensic Science Highlighting the Differences between the Bayes Factor and Likelihood Ratio

Danica M. Ommen*† and Christopher P. Saunders*

Iowa State University and South Dakota State University

Abstract. This article is aimed at the growing number of statisticians interested in the important problem of interpreting evidence within the forensic identification of source problems. Our purpose is to formalize these forensic problems as statistical model selection problems. We use two different classes of statistics for quantifying the evidential value, the likelihood ratio and Bayes Factor. In forensics, both are commonly called the “likelihood ratio approach” and “the value of evidence” despite using different definitions of probability. In statistics, they are closely related to the traditional likelihood ratio from pattern recognition and the Bayes Factor used in model selection. For two different problem frameworks typical in forensic science, the common source and the specific source problems, we show the Bayes Factor and likelihood ratio are not equivalent, and highlight several interesting links between them. These contributions will help to elucidate the effects of choosing different definitions of probability when addressing the forensic identification of source problems. The broader population of statisticians may find this paper interesting as an introduction to forensic applications and for illuminating the connections between model selection methods from two different paradigms of statistics, particularly in view of the active recent discussions on the connections among Bayesian, Fiducial, Frequentist (BFF) approaches.

Key words and phrases: Bayesian, Frequentist, model selection, consistency, credible interval, forensics, common source, specific source.

*Iowa State University, Department of Statistics, Ames, IA (e-mail: dnommen@iastate.edu); South Dakota State University, Department of Mathematics & Statistics, Brookings, SD (e-mail: christopher.saunders@sdstate.edu).

*Research supported by the National Institute of Justice, Office of Justice Programs, US Department of Justice under Award No. 2014-IJ-CX-K088. The opinions and conclusions or recommendations expressed in this presentation are those of the author and do not necessarily represent those of the Department of Justice.

†Research partially supported by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.
1. INTRODUCTION

Recently, there has been a push to strengthen the statistical foundations for applications in forensic science [15, 21]. In general, Bayesian statistical methods of approaching the forensic identification of source problems have been advocated by a number of forensic science researchers, especially in Europe [9, 26, 4, 5] while alternative methods have been advocated by a number of forensic science researchers primarily in the United States [13, 25, 10]. This has resulted in an increased number of statisticians interested in analyzing and interpreting forensic evidence. This is a very important task because making mistakes means that some innocent people are convicted while others who are guilty go free. The likelihood ratio and the Bayes Factor play a critical role in quantifying the value of forensic evidence. Typically, the term likelihood ratio is used in reference to the Neyman-Pearson likelihood ratio test statistic for simple- and nested-model selection. However, this method is not directly applicable to non-nested models, like the models often encountered in forensic science applications. In this article, the phrase likelihood ratio function will be used to define the ratio of two different likelihood functions for data described under two competing models. The Bayes Factor is a well-studied statistic often used for performing model selection on a wide variety of models [11, 16], including non-nested models like those related to forensic evidence.

Both the likelihood ratio (associated with the classical paradigm of statistics) and the Bayes Factor (associated with the Bayesian paradigm) are appropriate statistics aimed at answering the same basic question: which model is better supported by the data. This has led to a great deal of confusion on precisely how the forensic value of evidence should be computed [26, 23]. Further complicating matters, the forensic science community uses ambiguous language to describe the methods. For example, Bayes Factors and likelihood ratios are both commonly called “likelihood ratio” or “value of evidence” when used in forensic science applications, regardless of the statistical paradigm used in the analysis. The goal of this article is to clearly define the Bayes Factor and likelihood ratio for forensic identification of source problems and to highlight an interesting relationship between them. This relationship can be used to further explore frequentist asymptotic properties of a Bayes Factor.

Another topic of debate in forensic statistics is whether or not an interval estimate is an appropriate quantification of the value of evidence [28, 4, 5, 7, 14, 19, 27]. Although no stance regarding this issue is taken here, the secondary goal of this article is to explore a Bayesian interval estimate for a likelihood ratio (a non-Bayesian statistic). The asymptotic relationship between the likelihood ratio and the Bayes Factor facilitates the derivation of a credible interval for the likelihood ratio centered on the Bayes Factor. Although the interval is not practically useful from an interpretation/inference perspective, it will be useful for deriving solutions to other practical issues in the forensic community.

Section 2 will summarize the two different model selections frameworks, called the common source and the specific source problems, developed in Ommen and Saunders [17]. In Section 3, the forms of both the Bayes Factor and the likelihood ratio for the common source and specific source problems will be given. Finally, the relationships between them are explored in Section 4. The interested reader is directed to the supplementary material [18] for further details that have been
omitted from this article for the sake of conciseness.

2. FORENSIC IDENTIFICATION OF SOURCE PROBLEMS

One of the most common, and increasingly the most difficult, areas of statistical application to forensic science is in the subject of forensic identification of source problems. The general idea of identification of source problems is that you have recovered physical evidence related to the perpetration of a crime, and you wish to determine where that evidence originated (the source). This is typically done by comparing the evidence with unknown source to some control samples collected from relevant known sources. For example, a fingerprint is left at the scene of a murder, and you want to determine if the print originates from a finger of the suspect. This type of problem is referred to as the specific source problem since one of the sources is fixed and known (i.e. a specific finger of the suspect) [17]. Similarly, suppose you have two different crime scenes with fingerprints recovered, and you want to know whether the prints were left by the same finger of an unknown perpetrator (this would suggest that the two crimes are related). This type of problem is referred to as the common source problem since the source(s) are unknown [17]. As it turns out, these two scenarios can be expressed statistically as two different non-nested model selection problems.

The main difference between the forensic identification of source problems discussed in this paper and traditional classification or pattern recognition problems is the order in which you observe the sets of data. In forensics applications, the unknown-source data is usually observed first, and then the other sets of data are subsequently collected. This is because the unknown source evidence is collected after a crime has occurred. Investigators don’t collect samples from a suspect unless a crime has been committed first, so the samples from the known sources are collected last. In traditional classification problems, the known-source data is typically collected first, and then the unknown-source data is analyzed as it is observed.

2.1 Specific Source Models

For the specific source problem, the evidence consists of recovered materials from an unknown source, denoted $e_u$, control materials from a known, fixed specific source, denoted $e_s$, and additional control materials from the population of alternative sources, denoted $e_a$, which is sometimes referred to as the background population [17]. The entire set of evidence will be denoted $E = \{e_u, e_s, e_a\}$. The alternative source population is defined using a hierarchical structure where $e_a$ is a sample generated by first randomly selecting $N_a$ sources from the background population, and then randomly selecting $N_w$ elements from within the $i^{th}$ source, for $i = 1, 2, \ldots, N_a$ [17]. Next, the subset $e_s$ is a sample generated by randomly selecting $N_s$ elements from the population associated with the specific source [17]. Finally, $e_u$ is a sample consisting of $N_u$ elements, and our goal is to select between two hypotheses for how it was generated.

There are typically two hypotheses in a criminal trial corresponding to the two competing “sides,” the prosecution (who works to show that the suspect is the guilty party) and the defense (who works to show that the suspect is innocent). If the unknown source evidence originates from the specific source, then that would support the prosecution’s argument of guilt, i.e. the fingerprint comes
from the suspect. However, if the unknown source evidence originates from some other source, then that would support the defense’s argument of innocence, i.e. the fingerprint is from someone other than the suspect. Therefore, the forensic hypotheses for the specific source problem are typically stated as follows, where \( H_p \) corresponds to the prosecution and \( H_d \) corresponds to the defense [17]:

\( H_p \): The unknown source evidence originates from the specific source.
\( H_d \): The unknown source evidence originates from some other source in the alternative source population.

As discussed in Ommen and Saunders [17], there is not enough information in the forensic hypotheses to perform proper statistical inference. So, we specify distributional models for the measurements (or the features) on the evidential objects. For notational clarity, we will denote the measurements made on the objects \( e_s \) by \( s \), measurements made on the objects \( e_a \) by \( A \), and the measurements made on the objects \( e_u \) by \( u \). (This distinction is not really necessary in most applications because there exists a one-to-one mapping from the space of evidential objects to the space of measurements made on those objects, but we choose to make the distinction here to clarify that the distributions on the measurements lead to the likelihood functions used in the statistical analysis that follows.) For example, the evidential objects could be fingerprints whereas the measurements could be the number, type, placement, and orientation of the minutiae. The entire collection of measurements on the evidence will be denoted by \( E = \{u, s, A\} \). The formal statistical models corresponding to the generation of the evidence in the specific source problem are as follows [20]:

\( M_s \): Let \( s_j \) denote the measurements of the \( j^{th} \) element of the sample \( s \) from the specific source and let \( F_s \) be a distribution indexed by the parameter \( \theta_s \) that describes their variation in the specific source population, then for \( j = 1, 2, \ldots, N_s \)

\[
(1) \quad s_j \overset{iid}{\sim} F_s(\cdot|\theta_s)
\]

\( M_a \): Let \( B_i \) denote a latent random variable that corresponds to a parameter characterizing the \( i^{th} \) source sampled from the alternative source population and let \( G \) be a distribution indexed by the parameter \( \theta_a \) that describes the between-source variation of these parameters from the alternative source population. Let \( A_{ij} \) denote the measurements of the \( j^{th} \) element from the \( i^{th} \) source of the sample from the alternative source population and let \( F_a \) be a distribution that describes their within-source variation in the alternative source population evidence, then for \( i = 1, 2, \ldots, N_a \) and \( j = 1, 2, \ldots, N_w \)

\[
(2) \quad B_i \overset{iid}{\sim} G(\cdot|\theta_a) \quad \text{and} \quad A_{ij}|B_i = b_i \overset{iid}{\sim} F_a(\cdot|b_i, \theta_a)
\]

\( M_p \): Let \( u_j \) denote the measurements of the \( j^{th} \) element of the sample \( u \), then for \( j = 1, 2, \ldots, N_u \)

\[
(3) \quad u_j \overset{iid}{\sim} F_s(\cdot|\theta_s)
\]

\( M_d \): Let \( u_j \) denote the measurements of the \( j^{th} \) element of the sample \( u \) and let \( B_u \) denote a latent random variable that corresponds to a parameter
characterizing one new source from the alternative source population, then
for \( j = 1, 2, \ldots, N_u \)

\[
B_u \sim G(\cdot|\theta_a) \quad \text{and} \quad u_j | B_u = b_u \overset{iid}{\sim} F_a(\cdot|b_u, \theta_a)
\]

For the specific source problem, both the prosecution and defense hypotheses imply that \( A \) and \( s \) have been generated according to models \( M_p \) and \( M_s \), respectively. Since the prosecution and defense agree on these models, the problem reduces to a selection between two different models for the unknown source evidence. Now, \( H_p \) implies that \( u \) has been generated according to model \( M_p \), and \( H_d \) implies that \( u \) has been generated according to model \( M_d \). In this scenario, both the distributional models which generated the unknown source evidence and the exchangeability models are different under \( M_p \) and \( M_d \) [17]. One further assumption made regarding the sampling models for the specific source problem is that each subset of the evidence is mutually independent of the other subsets, given the parameters \( \theta_s \) and \( \theta_a \), under both the prosecution and defense models, i.e.

\[
u \perp s \perp A.
\]

An additional reason to make the distinction between the evidential objects and their measurements is due to the improvement of scientific technology: the tools used to take the measurements can change. For example, the technology for measuring the trace elemental composition of glass fragments has changed from X-ray fluorescence (XRF) to inductively coupled plasma mass spectrometry (ICPMS). This will change the forms of the corresponding distributions on the measurements (defined by \( G, F_a, F_s \)) while the overall sampling structure (defined by \( M_s, M_a, M_p, M_d \)) and independence assumptions remain the same.

### 2.2 Common-Source Models

For the common-source problem, the evidence consists of recovered materials from the first unknown source, denoted \( e_{u_s} \), recovered materials from the second unknown source, denoted \( e_{u_y} \), and control materials from the population of alternative sources, denoted \( e_a \), and sometimes referred to as the background population [17]. The alternative source population is defined using the same hierarchical structure as the specific source problem. In contrast to the specific source problem, the common source problem does not have the additional set of control samples from a fixed, specific source. The entire set of evidence will be denoted \( E = \{ e_{u_s}, e_{u_y}, e_a \} \).

Similar to the specific source problem, the common source problem has two competing models corresponding to the two competing “sides” in a criminal trial. Therefore, the forensic hypotheses for the common source problem are typically stated as follows [17]:

- \( H_p \): The two sets of unknown source evidence originate from the same source.
- \( H_d \): The two sets of unknown source evidence originate from two different sources.

The prosecution hypothesis implies that \( e_{u_s} \) and \( e_{u_y} \) are two samples generated by first randomly selecting a single source from the population of alternative sources, and then \( e_{u_s} \) is generated by randomly selecting the first set of \( N_x \)
elements from within the common source, and finally \( e_{uy} \) is generated by randomly selecting the second set of \( N_y \) elements from within the common source [17]. The defense hypothesis implies that \( e_{ux} \) and \( e_{uy} \) are two samples generated by first randomly selecting two different sources from the population of alternative sources, and then \( e_{ux} \) is generated by randomly selecting \( N_x \) elements from within the first source, and finally \( e_{uy} \) is generated by randomly selecting \( N_y \) elements from within the second source [17].

Like the specific source problem, we will make the distinction between the evidential objects and the measurements on those objects following Ommen and Saunders [17] for notational clarity. The measurements made on the evidential objects \( E = \{ e_{ux}, e_{uy}, e_a \} \) will be denoted by \( E = \{ u_x, u_y, A \} \). The formal statistical models corresponding to the generation of the evidence in the common source problem are as follows [20]:

\( M_a: \) \( A_{ij}, B_i, F_a, G, \) and \( \theta_a \) are defined the same as for the specific source problem with corresponding models given by Equation 2:

\[
B_i \overset{iid}{\sim} G(\cdot|\theta_a) \quad \text{and} \quad A_{ij}|B_i = b_i \overset{iid}{\sim} F_a(\cdot|b_i, \theta_a)
\]

\( M_p: \) Let \( u_{xi} \) denote the measurements of the \( i^{th} \) element of the sample \( u_x \), \( u_{yj} \) denote the measurements of the \( j^{th} \) element of the sample \( u_y \), and let \( B_u \) denote a latent random variable that corresponds to a parameter characterizing one new source from the alternative source population, then for \( i = 1, 2, \ldots, N_x \) and \( j = 1, 2, \ldots, N_y \)

\[
B_u \sim G(\cdot|\theta_a) \quad \text{and} \quad u_{xi}|B_u = b_u \overset{iid}{\sim} F_a(\cdot|b_u, \theta_a)
\]

\[
u_{yj}|B_u = b_u \overset{iid}{\sim} F_a(\cdot|b_u, \theta_a)
\]

\( M_d: \) Let \( u_{xi} \) denote the measurements of the \( i^{th} \) element of the sample \( u_x \) and \( B_x \) denote a latent random variable that corresponds to a parameter characterizing the first new source from the alternative source population. Also, let \( u_{yj} \) denote the measurements of the \( j^{th} \) element of the sample \( u_y \) and \( B_y \) denote a latent random variable that corresponds to a parameter characterizing the second new source from the alternative source population, then for \( i = 1, 2, \ldots, N_x \) and \( j = 1, 2, \ldots, N_y \)

\[
B_x \sim G(\cdot|\theta_a) \quad \text{and} \quad u_{xi}|B_x = b_x \overset{iid}{\sim} F_a(\cdot|b_x, \theta_a)
\]

\[
B_y \sim G(\cdot|\theta_a) \quad \text{and} \quad u_{yj}|B_y = b_y \overset{iid}{\sim} F_a(\cdot|b_y, \theta_a)
\]

For the common source problem, both the prosecution and defense hypotheses imply that \( A \) has been generated according to model \( M_a \). Additionally, the prosecution hypothesis implies that the unknown source evidence has been generated according to model \( M_p \), whereas the defense hypothesis implies that the unknown source evidence has been generated according to model \( M_d \). The model selection problem is then a selection between \( M_p \) and \( M_d \) for the unknown source evidence. In contrast to the specific source problem, the distributional models which generated \( u_x \) and \( u_y \) are the same, but the exchangeability models for \( u_x \) and \( u_y \) are different under the two different sampling models [17]. Under the prosecution model, \( u_x \) and \( u_y \) are conditionally independent given the common
source, and \( \mathbf{u}_v \) and \( \mathbf{u}_x \) are unconditionally independent under the defense model [17]. Furthermore, all of the unknown source evidence is mutually independent of the alternative source population data. These independence assumptions are summarized below, and all are conditional on the parameter \( \theta_0 \).

\[
M_p : \quad (\mathbf{u}_x|\mathbf{b}_u) \perp (\mathbf{u}_y|\mathbf{b}_u) \perp \mathbf{A}
\]

\[
M_d : \quad \mathbf{u}_x \perp \mathbf{u}_y \perp \mathbf{A}
\]

Similar to the specific source problem, any change in measurement tool for the common source problem will change the forms of the distributional models (defined by \( G, F_0 \)), but the sampling models (defined by \( M_0, M_p, M_d \)) and independence assumptions will remain the same.

\subsection{2.3 Motivating Example}

For this example, we will only consider the specific source problem for the sake of conciseness, but the interested reader is directed to the supplemental material for details of the common source example [18]. Suppose that we invent a scenario in which a crime was committed by a perpetrator who intended to detonate an improvised explosive device (IED) at a public building. The IED failed to detonate, and the authorities found the unexploded device which contained several small lengths of copper wire (these will become the unknown source evidence). A paper by Dettman et al. [8] demonstrated that copper wire can be used as forensic evidence where the scientific analysis is based on the trace elemental compositions of the copper wire obtained through ICPMS. Based on other information, the authorities believe they have a suspect (let’s call her Mrs. Smith). Upon searching Mrs. Smith’s basement, the authorities find a roll of copper wire in her craft studio that shares the same gauge as the wire from the IED (this will become the specific source evidence). The question posed by the authorities is, “Does the copper wire found at the crime scene originate from the roll of copper wire obtained at Mrs. Smith’s residence?” In order to investigate this question, the authorities obtain several rolls of copper wire (sharing the same gauge) from hardware stores in the areas where Mrs. Smith lives and works (these will become the alternative source population).

Now, let’s suppose that all of the copper wire objects described above are sent to a forensic lab for measurements of their trace elemental compositions via ICPMS. According to Dettman et al. [8], copper wire consists mostly of the 29th element of the periodic table, copper (Cu), but also contains small amounts of these other periodic elements: silver (Ag), arsenic (As), bismuth (Bi), cobalt (Co), nickel (Ni), lead (Pb), antimony (Sb), and selenium (Se). The lengths of wire from the recovered IED each result in one measurement of each of the eight trace elements listed above. Therefore, \( \mathbf{u}_j \) is an 8-dimensional vector of measurements from the unknown source evidence for \( j = 1, 2, \ldots, N_u \) where \( N_u \) is the number of recovered lengths of wire. Also, suppose that the roll of copper wire from Mrs. Smith’s residence was prepared to obtain several small lengths of wire from various locations in the roll, each resulting in one measurement of the eight trace elements. Then, \( \mathbf{s}_j \) is an 8-dimensional vector of measurements from the specific source for \( j = 1, 2, \ldots, N_s \) where \( N_s \) is the number of lengths cut from the roll. Finally, suppose that each of the rolls from the hardware stores are similarly prepared to obtain several small lengths of wire from various locations in the roll,
each resulting in one measurement of the eight trace elements. Then $A_{ij}$ is an 8-dimensional vector of measurements from the $j^{th}$ length of wire cut from the $i^{th}$ roll in the alternative source population for $j = 1, 2, \ldots, N_w$ and $i = 1, 2, \ldots, N_a$ where $N_a$ is the number of rolls and $N_w$ is the number of lengths cut from each roll.

Following the assumptions of multivariate normality of the vectors of trace element measurements in Dettman et al. [8], reasonable statistical models for the specific source problem are given by:

$M_s$: Let $N_8$ be a multivariate normal distribution with an 8-dimensional mean vector $\mu_s$ and an $8 \times 8$ positive-definite, symmetric covariance matrix $\Sigma_s$.

$$s_j \overset{iid}{\sim} N_8(\mu_s, \Sigma_s)$$

$M_a$: Let $N_8$ be a multivariate normal distribution with an 8-dimensional mean vector $\mu_a$ and an $8 \times 8$ positive-definite, symmetric covariance matrix $\Sigma_a$, and let $N_8$ be a second multivariate normal distribution with an 8-dimensional vector of simulated means $b_i$ and an $8 \times 8$ positive-definite, symmetric covariance matrix $\Sigma_w$.

$$B_i \overset{iid}{\sim} N_8(\mu_a, \Sigma_b) \quad \text{and} \quad A_{ij}|B_i = b_i \overset{iid}{\sim} N_8(b_i, \Sigma_w)$$

$M_p$: Let $\mu_s$ and $\Sigma_s$ be defined as in $M_s$ above.

$$u_j \overset{iid}{\sim} N_8(\mu_s, \Sigma_s)$$

$M_d$: Let $\mu_a$, $\Sigma_b$, and $\Sigma_w$ be defined as in $M_a$ above.

$$B_u \sim N_8(\mu_a, \Sigma_b) \quad \text{and} \quad u_j|B_u = b_u \overset{iid}{\sim} N_8(b_u, \Sigma_w)$$

So, in this specific source scenario $\theta_s = \{\mu_s, \Sigma_s\}$ and $\theta_a = \{\mu_a, \Sigma_b, \Sigma_w\}$. We note that $\Sigma_s$ and $\Sigma_w$ both have the interpretation of characterizing a within-source variance, but we prefer to use separate notation because they characterize two different populations and it is not necessary for these two parameters to take the same values. Additionally, the separate notation allows for clearer explanations of the relationships between the prior distributions on these parameters in the Bayesian setting. Also, we should point out the fact that $B_i$ and $B_u$ are unobserved, so they are neither part of the evidence nor the parameter space. Note that the model for $M_a$ is slightly different than the one proposed in Dettman et al. [8]. We chose to use a more simplistic model than the Gaussian mixture model proposed for the alternative source population in Dettman et al. [8] to clearly demonstrate the approach.

### 3. METHODS

The main goal of the forensic identification of source problems described in Section 2 is to choose between two different models for how the unknown source evidence was generated. This process is often called “quantifying the value of evidence” within the forensic science community [1, 2]. While most researchers agree that the best way to quantify the value of evidence is to determine the
value of the “likelihood ratio,” many often disagree about the precise methods of
doing so.

\[
\frac{\Pr(E|H_p)}{\Pr(E|H_d)}
\]

“Likelihood Ratio”

In the expression above for the “likelihood ratio,” \(E\) is the evidence, and \(H_p\) and \(H_d\) are the prosecution and defense hypotheses, respectively. Additionally, \(H_p\) and \(H_d\) should be non-exhaustive, mutually exclusive propositions. Now, there are two popular ways of defining the probability \(\Pr\) within the forensic science
community. The first is to treat it as a measure of subjective belief following
the Bayesian paradigm and the second is to treat it as a likelihood following
the classical/Frequentist paradigm. Those who fall into the former category typically
compute a Bayes Factor following Lindley’s example [12], and those who fall into
the latter category typically compute a likelihood ratio.

The Bayesian method of quantifying the value of evidence relies on the \textit{odds form of Bayes’ Theorem} to convert prior odds (of the two competing hypothesess) to posterior odds via the “likelihood ratio” from Equation 11 (which is actually
called the “Bayes Factor” within the Bayesian paradigm of statistics) [2]. This
results in the following prescription

\[
\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}.
\]

In the case where the measurements on the evidence follow parametric models,
the Bayes Factor is defined by

\[
\beta(E) = \frac{\int f(E|\theta_p, H_p) \, d\Pi(\theta_p|H_p)}{\int f(E|\theta_d, H_d) \, d\Pi(\theta_d|H_d)}
\]

where the \(f\) denotes “the relevant probability densities,” \(\theta_p\) represents the parameters for \(f\) under \(H_p\), \(\theta_d\) represents the parameters for \(f\) under \(H_d\), and \(\Pi\) represents the prior distribution [12, 2]. Note that when \(E\) is an observed sample, then \(f\) becomes the likelihood function for the evidence under each of the competing propositions. A Bayes Factor greater than one indicates that the evidence supports \(H_p\) over \(H_d\), in contrast to a Bayes Factor less than one which indicates that the evidence supports \(H_d\) over \(H_p\). A Bayes Factor that is equal to
one means that the evidence cannot discriminate between the two competing hypothesess. The form of the Bayes Factor for the identification of source problems
will be summarized in Section 3.2.

The non-Bayesian way to compute the “likelihood ratio” given in Equation 12
is by conforming to the law of likelihood. The law of likelihood from Royall [22]
says,

“If Hypothesis A implies the probability that \(X = x\) is \(p_A(x)\), while
Hypothesis B implies the probability that \(X = x\) is \(p_B(x)\), then the obser-
vation \(X = x\) is evidence supporting Hypothesis A over Hypothesis B if and
only if \(p_A(x) > p_B(x)\) and the Likelihood Ratio \(p_A(x)/p_B(x)\) measures
the strength of that support.”
Following the notational conventions of this article, the likelihood ratio is defined by

\[ \lambda(E) = \frac{f(E|\theta_{p0}, H_p)}{f(E|\theta_{d0}, H_d)} \]  

where \( \theta_{p0} \) and \( \theta_{d0} \) represent the true values of the parameters \( \theta_p \) and \( \theta_d \), respectively [19]. Under the likelihood framework, the value of the likelihood ratio is interpreted in the same way as the Bayes Factor. In order to compute a single value for the likelihood ratio, the true values of the parameters need to be known or estimated with a high degree of certainty. If the values of the parameters are not known, then the likelihood ratio is a function of the unknown parameters. We call this the likelihood ratio function, and it is given by

\[ \lambda(\theta_p, \theta_d|E) = \frac{\mathcal{L}(\theta_p|E, H_p)}{\mathcal{L}(\theta_d|E, H_d)} \]  

where \( \mathcal{L} \) represents the likelihood function defined in the traditional way [19]. For given observations of the evidence, the likelihood ratio function is a function of the parameter vectors \( \theta_p \) and \( \theta_d \) which are constrained to take values in the space \( \Theta_p \) and \( \Theta_d \), respectively. The form of the likelihood ratio for the specific source and common source problems will be defined in Section 3.1.

Now, regardless of which method you choose for quantifying the value of evidence, the likelihood functions for the evidence under both of the competing hypotheses need to be defined. The likelihood functions for the evidence were originally given in the appendix of Ommen et al. [20], but they will be reproduced in Appendix A to match the notational conventions in this article. It should be noted that, for both the specific source and common source problems, \( B_i, B_u, B_x, \) and \( B_y \) are latent random variables in the models and they are not considered part of the parameter space for \( \theta_a \). This is made evident by the forms of the corresponding likelihood functions where the latent random variables are integrated out. Therefore, the parameter space under \( H_p \) is the same as the parameter space under \( H_d \), i.e. \( \Theta_p = \Theta_d \). For the specific source problem, this is the joint parameter space for \( \theta_s \) and \( \theta_a \). Similarly for the common source problem, this is the parameter space for \( \theta_a \). This reinforces the idea that the likelihood ratio and the Bayes Factor are designed to select between model \( M_p \) and model \( M_d \) for the unknown source evidence where the difference is in the exchangeability and/or the distributional assumptions. The two model selection methods are not designed in the traditional way to select between two different parameter spaces (which may or may not be nested) because there is no difference in the parameter space for the forensic identification of source problems.

### 3.1 The Likelihood Ratio

We’ll begin by giving the likelihood ratio function as defined by Equation 15 for the forensic identification of source problems using the likelihood functions defined in Appendix A. In particular, the likelihood ratio function for the specific source problem is given by

\[ \lambda_{ss}(\theta_s, \theta_a|u) = \frac{\mathcal{L}(\theta_s|u, M_p)}{\mathcal{L}(\theta_a|u, M_d)} = \frac{f(u|\theta_s)}{f(u|\theta_a)} \]
and is a function of the parameters $\theta_s$ and $\theta_a$ for a given observation of the unknown source evidence. You’ll notice that $s$ and $A$ are absent from the equation. This is because the corresponding likelihood functions are the same for the both the numerator (implied by $H_p$) and the denominator (implied by $H_d$) and so those terms cancel. Similarly, the likelihood ratio function for the common source problem is given by

$$
\lambda_{cs}(\theta_a|u_x, u_y) = \frac{\mathcal{L}(\theta_a|u_x, u_y, M_p)}{\mathcal{L}(\theta_a|u_x, M_d)\mathcal{L}(\theta_a|u_y, M_d)}
$$

$$
= \frac{f(u_x, u_y|\theta_a, M_p)}{f(u_x|\theta_a, M_d)f(u_y|\theta_a, M_d)}
$$

(17)

and is a function of the parameter $\theta_a$ for a given observation of the two sets of unknown source evidence.

Next, the true likelihood ratios are defined with respect to $\theta_{s0}$ and $\theta_{a0}$, the parameter values of the true sampling distributions implied by models $M_s$ and $M_a$, respectively. Since the values of these parameters are rarely known, then the value of the true likelihood ratio is fixed, but unknown. The forms of the true likelihood ratio for the specific source and common source problems are $\lambda_{ss}(\theta_{s0}, \theta_{a0}|u)$ and $\lambda_{cs}(\theta_{s0}|u_x, u_y)$, defined by plugging the true values of the parameters into Equation 16 and Equation 17, respectively. Note that the true likelihood ratio does not imply that either $\theta_{a0}$ or $\theta_{s0}$ is the true parameter for models generating the unknown source evidence. The parameter value $\theta_{a0}$ is used to indicate the true parameter for the alternative source population models while $\theta_{s0}$ is used to indicate the true parameter for the specific source population model. Therefore, the true likelihood ratio is not the limiting form of the likelihood ratio function as the sample size of the unknown source evidence grows ($N_u \to \infty$). Theorem 5.1 from Vuong [30] provides the limiting form of the likelihood ratio function for this scenario, which will not be considered further in this article. In fact, we will see in Section 4 that the true likelihood ratio is the limiting form of the Bayes Factor as the number of sources in the alternative source population ($N_a \to \infty$) and the number of observations from the specific source ($N_s \to \infty$) grow infinitely larger, but the number of observations from the unknown source evidence stays fixed ($N_u, N_x, N_y$).

3.2 The Bayes Factor

In the Bayesian paradigm, the method for dealing with parameters of distributions is to assign a prior distribution that describes the uncertainty about its value. So, the prior distributions for the parameters of the identification of source problems, $\theta_a \in \Theta_a$ and $\theta_s \in \Theta_s$, will be denoted by $\Pi(\theta_a)$ and $\Pi(\theta_s)$, respectively. The prior distributions $\Pi(\theta_a)$ and $\Pi(\theta_s)$ will be defined over the spaces $\Theta_a$ and $\Theta_s$, respectively. In addition, the prior densities corresponding to $\Pi(\theta_a)$ and $\Pi(\theta_s)$ will be denoted by $\pi(\theta_a)$ and $\pi(\theta_s)$, respectively. Also, the posterior distributions for the parameter given some data $x$ will be denoted in the usual way: $\Pi(\theta_a|x)$ and $\Pi(\theta_s|x)$.

Using the likelihood functions defined in Appendix A, we can derive forms for the common source and specific source Bayes Factors following the definition given in Equation 13. These expressions were originally derived in Ommen et al. [20]. The derivation of the Bayes Factor for the specific source problem
is reproduced below, where \( m \) denotes a marginal density and \( p \) denotes a posterior predictive density. One additional assumption is needed in order to obtain the result for the specific source problem: \( \theta_s \) is independent of \( \theta_a \) so that
\[
\Pi(\theta_s, \theta_a| M_p) = \Pi(\theta_s, \theta_a| M_d) = \Pi(\theta_s) \Pi(\theta_a) \tag{20}
\]

\[
\begin{align*}
\beta_{ss}(E) &= \frac{\int f(E|\theta_s, \theta_a, M_p) \, d\Pi(\theta_s, \theta_a| M_p)}{\int f(E|\theta_s, \theta_a, M_d) \, d\Pi(\theta_s, \theta_a| M_d)} \\
&= \frac{\int f(u|\theta_s)f(s|\theta_a)f(A|\theta_a) \, d\Pi(\theta_s)\Pi(\theta_a)}{\int f(u|\theta_a)f(s|\theta_a)f(A|\theta_a) \, d\Pi(\theta_a)\Pi(\theta_a)} \\
&= \frac{\int f(u|\theta_s)f(s|\theta_a) \, d\Pi(\theta_s)}{\int f(A|\theta_a)f(u|\theta_a) \, d\Pi(\theta_a)} \times \frac{\int f(A|\theta_a)f(u|\theta_a) \, d\Pi(\theta_a)}{\int f(s|\theta_a) \, d\Pi(\theta_a)} \\
&= \frac{\int f(u|\theta_s)f(s|\theta_a) \, d\Pi(\theta_s)}{\int f(A|\theta_a)f(u|\theta_a) \, d\Pi(\theta_a)} \tag{21}
\end{align*}
\]

Therefore, the Bayes Factor for the specific source problem is given by
\[
\beta_{ss}(E) = \frac{\int f(u|\theta_s) \, d\Pi(\theta_s| s)}{\int f(u|\theta_a) \, d\Pi(\theta_a| A)} = \frac{p(u|s, M_p)}{p(u|A, M_d)}. \tag{22}
\]

Similarly, the Bayes Factor for the common source problem is given by
\[
\begin{align*}
\beta_{cs}(E) &= \frac{\int f(u_x, u_y|\theta_a, M_p) \, d\Pi(\theta_a| A)}{\int f(u_x|\theta_a, M_d)f(u_y|\theta_a, M_d) \, d\Pi(\theta_a| A)} \\
&= \frac{\int f(u_x, u_y) \, d\Pi(\theta_a)}{\int f(u_x|\theta_a, M_d)f(u_y|\theta_a, M_d) \, d\Pi(\theta_a)} \times \frac{\int f(u_x|\theta_a, M_d)f(u_y|\theta_a, M_d) \, d\Pi(\theta_a)}{\int f(u_x, u_y) \, d\Pi(\theta_a)} \\
&= \frac{\int f(u_x, u_y) \, d\Pi(\theta_a)}{\int f(u_x|\theta_a, M_d)f(u_y|\theta_a, M_d) \, d\Pi(\theta_a)} \tag{23}
\end{align*}
\]

The derivation is omitted here because it is similar to the specific source problem, but it is reproduced in the supplemental material [18]. The Bayes Factors given by Equation 22 and Equation 23 are interesting because they are the ratio of the integrated likelihood functions given in Equation 16 and Equation 17, respectively. Additionally, these Bayes Factors integrate with respect to the posterior distribution for the parameters, as opposed to the prior distributions as given in Equation 13. Bayes Factors in these forms are often computationally intensive to compute since the integrals rarely have closed-form solutions [11, 16, 20] and because two approximations are needed, one for the integral in the numerator and one for the integral in the denominator. The expression derived in the following section allows us to eliminate one of these integrals so only one integral approximation is needed.

4. RELATIONSHIPS BETWEEN THE LR AND BF

To reiterate, the likelihood ratio is often used within the classical paradigm of statistics, while the Bayes Factor is often associated with the Bayesian paradigm. In this section, we provide some expressions which directly relate the statistics from the two different paradigms. These expressions show the Bayes Factor is
the expected value of the likelihood ratio function with respect to the posterior distribution of the parameters given the entire collection of evidence generated under the defense model. This expression is derived below for the specific source problem.

\[
\beta_{ss}(E) = \frac{\int f(u|\theta_s)f(s|\theta_s)f(A|\theta_a) \, d\Pi(\theta_s,\theta_a)}{m(u,s,A|M_d)}
\]

\[
= \int \frac{f(u|\theta_s)}{f(u|\theta_a)} \times \frac{f(s|\theta_s)f(A|\theta_a)}{m(u,s,A|M_d)} \, d\Pi(\theta_s,\theta_a)
\]

\[
= \int \frac{f(u|\theta_s)}{f(u|\theta_a)} \times \frac{f(E|\theta_s,\theta_a,M_d)}{m(E|M_d)} \, d\Pi(\theta_s,\theta_a)
\]

Equation 24 is obtained from independence relationship defined in Equation 5

and Equation 25 is obtained using the definition of the posterior distribution \(\Pi(\theta_s,\theta_a|E, M_d)\). The expression for the common source Bayes Factor can be derived using a similar method.

\[
\beta_{cs}(E) = \int \frac{f(u_x,u_y|\theta_a,M_p)}{f(u_x,u_y|\theta_a,M_d)} \, d\Pi(\theta_a|u_x,u_y,A,M_d)
\]

\[
= \int \frac{1}{\lambda_{cs}(\theta_a|u_x,u_y)} \, d\Pi(\theta_a|E,M_d)
\]

The derivation of Equation 26 for the common source Bayes Factor was originally given in the appendix of Ommen et al. [20], but a version using these notational conventions can be found in the supplemental material [18]. These expressions serve both theoretical and practical purposes. Theoretically, they make the proofs of the theorems in the following section easier by only having to work with one integral instead of the ratio of two integrals. Practically, these expressions provide some reduction in computational complexity when approximating the value of the Bayes Factor via Monte Carlo integration since only one integral needs to be estimated as opposed to the ratio of two integrals [20].

Imagine that an expert uses this expression to quantify the value of evidence and testifies to their methods in court. It may seem unfair to the prosecution that the defense’s model was used. So, a similar expression can be derived using the prosecution model for the evidence in the posterior distribution, with the introduction of two inverses. The derivations of Equation 27 and Equation 28 can be found in Appendix B.

\[
\beta_{ss}(E) = \left[\int \frac{1}{\lambda_{ss}(\theta_s,\theta_a|u)} \, d\Pi(\theta_s,\theta_a|E, M_p)\right]^{-1}
\]

\[
\beta_{cs}(E) = \left[\int \frac{1}{\lambda_{cs}(\theta_a|u_x,u_y)} \, d\Pi(\theta_a|E, M_p)\right]^{-1}
\]

In all of these expressions, notice that the unknown source evidence is included in the posterior distribution for the parameters, in contrast to the Bayes Factors
given in Equation 22 and Equation 23. Another interesting thing to note about these forms for the Bayes Factor is that one of them will be computed using a misspecified model for the evidence $E$ depending on the truth of which model, $M_p$ or $M_d$ actually generated the unknown source evidence. In unusual circumstances, it may be the case that both models are misspecified for the unknown source evidence since there are no assumptions with the identification of source framework that $M_p$ and $M_d$ partition the space of models which could have generated the unknown source evidence. This is due to the fact that there are no requirements that $H_p$ and $H_d$ be exhaustive propositions, they must only be mutually exclusive.

4.1 Frequentist Asymptotic Results for the Bayes Factor

In this section, we examine the consistency of the Bayes Factor by way of a well-known result, the Bernstein-von Mises Theorem. The Bernstein-von Mises Theorem is reproduced from Van der Vaart [29] for ease of reference in the supplementary material [18]. Generally, the asymptotic properties of the Bayes Factor have been examined with respect to data from a single population representing a single source of information. In this case, the Bayes Factor will diverge to positive infinity (in probability) when the model in the numerator is preferred and will converge to zero (in probability) when the model in the denominator is preferred [6]. However, the forensic identification of source problems have multiple sources of information. We will show that, under certain regularity conditions, the Bayes Factor for both the common source and specific source problems will converge to the true likelihood ratio.

For the specific source problem, there are three sources of information corresponding to the three subsets of the evidence, $E = \{u, s, A\}$, but there are only two populations which generate that evidence (one corresponding to the specific source population and the other to the alternative source population). From a theoretical perspective, it doesn’t make sense for the number of observations from the unknown source evidence, $N_u$, to be allowed to grow. This is due to the fact that the crime scene analysts cannot control how much evidence is left behind. Alternatively, there is control over how many samples can be collected from a suspect or from the background population (even if this is practically constrained by factors such as time and money). Therefore, the number of observations from the specific source, $N_s$, and the number of sources in the alternative source population, $N_a$ will be allowed to grow for the asymptotic results that follow.

In order to facilitate the understanding of these results, we will make the following notational additions. Let $A_{n_u}$ denote a sequence of random variables corresponding to the generation of hierarchical samples from the alternative source population $A_{n_u}$ where $n_u$ is the index that denotes the varying number of sources in the alternative source population with a fixed number of samples from within each source, $N_u$. Similarly, let $S_{n_s}$ denote a sequence random variables corresponding to the generation of the sample from the specific source $s_{n_s}$ where the index $n_s$ denotes the varying sample size. For simplicity, we will fix $n_u = n_s = n$ so that the number of alternative sources increases in the exact same way as the number of elements from the specific source. The proofs of the results can easily be modified to accommodate more flexible relationships between the two sample sizes. In addition, let $P_\theta^o$ denote the joint probability measure on $A_n$ and $S_n$ for
all \(\theta \in \Theta\) where \(\Theta\) is the joint parameter space, including \(\theta_0\) which denotes the true value of the joint parameter. Finally, let \(\beta_{ss}(u, S_n, A_n)\) denote the random function corresponding to the Bayes Factor given in Equation \(25\) before the values of \(A_n\) and \(S_n\) have been observed.

**Theorem 1** (Specific Source Bayes Factor Consistency):
Given a fixed observation of \(u\), suppose that the likelihood ratio function, \(\lambda_{ss}(\theta|u)\), is bounded in a neighborhood of \(\theta_0\) and that \(\hat{\theta}_n\) is a consistent estimator for \(\theta_0\). Furthermore, suppose the assumptions of the Bernstein-von Mises Theorem are satisfied. Then the Bayes Factor converges in \(P^m_{\theta_0}\)-probability to the true likelihood ratio as \(n \to \infty\),

\[
\beta_{ss}(u, S_n, A_n) \xrightarrow{P^m_{\theta_0}} \lambda_{ss}(\theta_0|u).
\]

The proof of this theorem is provided in Appendix C.

Similarly for the common source problem, there are three sources of information corresponding to the three subsets of evidence, \(E = \{\text{u}_x, \text{u}_y, A\}\), but in this case there is only one population that generates the evidence (corresponding to the alternative source population). Again, it doesn’t make sense for the numbers of observations from the two sets of unknown source evidence, \(N_x\) and \(N_y\), to grow because you can’t collect evidence that doesn’t exist. So, for the same reasons, only the number of sources in the alternative source population, \(N_o\) will be allowed to grow for the asymptotic results to follow while \(N_x\), \(N_y\), and \(N_w\) remain fixed. The same notational conventions \(A_{na}\) and \(P^n_{\theta_o}\) will be used as defined for the specific source problem. Finally, let \(\beta_{cs}(\text{u}_x, \text{u}_y, A_{na})\) denote the random function corresponding to the Bayes Factor given in Equation \(26\) before the value of \(A_{na}\) has been observed.

**Theorem 2** (Common Source Bayes Factor Consistency):
Given a fixed observation of \(\text{u}_x\) and \(\text{u}_y\), suppose that the likelihood ratio function \(\lambda_{cs}(\theta_0|\text{u}_x, \text{u}_y)\) is bounded in a neighborhood of \(\theta_{0o}\) and that \(\hat{\theta}_n\) is a consistent estimator for \(\theta_{0o}\). Furthermore, suppose the assumptions of the Bernstein-von Mises Theorem are satisfied. Then as \(n_o \to \infty\), the Bayes Factor converges in \(P^n_{\theta_o}\)-probability to the true likelihood ratio,

\[
\beta_{cs}(\text{u}_x, \text{u}_y, A_{na}) \xrightarrow{P^n_{\theta_o}} \lambda_{cs}(\theta_{0o}|\text{u}_x, \text{u}_y).
\]

The proof of this theorem is provided in Appendix C.

These two theorems imply an interesting relationship between the two statistical paradigms: the Bayes Factor can be used as an approximation to the value of the true likelihood ratio for forensic identification of source problems when there is uncertainty regarding the modeling parameters. Furthermore, Theorem 1 and Theorem 2 are an important contribution to the field of forensic statistics where the standard asymptotic results for the Bayes Factor are not expected to apply due to the nature of the evidence and sample sizes. Finally, we chose to derive asymptotic results for the Bayes Factors because it is difficult to derive asymptotic results for a “plug-in” likelihood ratio where the parameter values are
estimated because the size of the unknown source evidence is always fixed (and
the likelihood ratio is not a function of the other sets of evidence). Even if it were
possible to rationalize a growing sample size for the unknown source evidence,
the true likelihood ratio may not be the limiting form of the “plug-in” likelihood
ratio because the true parameter values \( \theta_{0} \) and \( \theta_{1} \) are the true parameter values
for the alternative source population and the specific source population evidence,
respectively, and not for the true data generating model for the unknown source
evidence.

### 4.2 Bayesian Credible Intervals for the Likelihood Ratio

We have first explored a frequentist asymptotic results for a Bayesian statistic
given by Theorem 1 and Theorem 2. Now, it will be interesting to explore
a relationship from the reverse perspective. The following theorem provides results
regarding a credible interval (a Bayesian method) for the value of the true
likelihood ratio (a classical statistic).

First, let \( E_{n} \) denote the random variable associated with generating the entire
collection of evidence \( E \). For the common source problem, this means that \( A_{n} \)
generates the hierarchical observations from the alternative source population and
the observations \( u_{x} \) and \( u_{y} \) are already fixed. For the specific source problem,
this means that \( A_{n} \) generates the hierarchical observations from the alternative
source population, \( S_{n} \) generates the observations from the specific source, and \( u \)
is already fixed. Furthermore, let \( P_{\theta}^{n} \) denote the joint probability measure on \( E_{n} \)
for all \( \theta \in \Theta \) including \( \theta_{0} \) where \( \Theta \) is the joint parameter space. Next, let \( \lambda(\theta) \)
denote the likelihood ratio function for either the common source or the specific
problem given by Equation 17 or Equation 16, respectively. Moreover, let
\( \Pi(\cdot|E_{n}, M_{d}) \) denote the sequence of posterior measures for \( \theta \) given the entire
collection of evidence generated under model \( M_{d} \). Similarly, let \( \Phi(\cdot|E_{n}, M_{d}) \) denote
the sequence of probability measures corresponding to the normal distribution
with a zero-vector mean and a covariance of \( \lambda'(\hat{\theta}_{n})'I_{\hat{\theta}_{n}}^{-1}\lambda'(\hat{\theta}_{n}) \) where \( \theta_{0} \) is a consistent estimator of \( \theta_{0} \) defined using the entire set of evidence under model \( M_{d} \),
\( \lambda' \) is the vector of first partial derivatives of the likelihood ratio function, and
\( I_{\hat{\theta}_{n}}^{-1} \) is the observed Fisher’s information matrix. Finally, let \( \Psi(\cdot|E_{n}, M_{d}) \) denote
the probability measure corresponding to the normal distribution with a mean
of \( \beta(E_{n}) \) and a variance of \( \sigma_{n}^{2} \) where \( \beta(E_{n}) \) represents the Bayes Factor given by
either Equation 25 or Equation 26 where the posterior distributions are defined
under model \( M_{d} \), and \( n\sigma_{n}^{2} \) converges in \( P_{\theta}^{n} \)-probability to a constant \( \sigma_{\theta}^{2} > 0 \) as
\( n \to \infty \).

**Theorem 3** (Approximate 1 - \( \alpha \) Credible Interval for the LR):

*Let the assumptions of Lemma 3.1 and Lemma 3.2 hold and let*

\[
I_{n} = \beta(E_{n}) \pm \Phi^{-1}(1 - \alpha/2)\sigma_{n}
\]

*where \( \beta(E_{n}) \) represents the sequence of either the common source or the
specific source Bayes Factors, 0 < \( \alpha \) < 1 is the desired significance level,
\( \sigma_{n} \) is the sequence of posterior standard errors for the likelihood ratio,
and \( \Phi^{-1} \) is the standard normal quantile function. Then as \( n \to \infty \)*

\[\Pi(\lambda(\theta) \in I_{n}|E_{n}, M_{d}) \to 1 - \alpha.\]
Please see Appendix D for a proof of this theorem which relies the the following lemmas.

**Lemma 3.1:**

For a given observation of the unknown source evidence, suppose that \( \lambda(\theta) \) is twice continuously differentiable, and that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_0 \) under \( P^{\theta}_n \)-probability. If the assumptions of the Bernstein-von Mises Theorem hold, then

\[
\left\| \Pi \left( \sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d \right) - \Phi \left( \sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d \right) \right\|_{TV}
\]

converges to zero in \( P^{\theta}_n \)-probability as \( n \to \infty \).

Please see Appendix D for a proof of this lemma.

**Lemma 3.2:**

For a given observation of the unknown source evidence, suppose that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_0 \) under \( P^{\theta}_n \)-probability and let \( n\sigma^2_n \) converge in \( P^{\theta}_n \)-probability to a constant \( \sigma^2 \) \( > 0 \) as \( n \to \infty \). If the assumptions of the Bernstein-von Mises Theorem hold for \( \beta(E_n) \), then

\[
\left\| \Phi(\lambda(\theta)|E_n, M_d) - \Psi(\lambda(\theta)|E_n, M_d) \right\|_{TV}
\]

converges to zero in \( P^{\theta}_n \)-probability as \( n \to \infty \).

Please see Appendix D for a proof of this lemma.

Because the Bernstein-von Mises Theorem results in posterior distributions for the likelihood ratio that are approximately normal, we can use a Wald-type credible interval. Theorem 3 is an interesting result because it centers the interval for the likelihood ratio on the value of the Bayes Factor. In Ommen et al. [19], one of the arguments against presenting intervals for the likelihood ratio is that the center of the commonly used interval (in the forensic discipline) tends to be larger than the corresponding Bayes Factor. This means that the usual estimate of the likelihood ratio from Ommen et al. [19] will be biased against the suspect. Creating intervals of the form given by Theorem 3 ensures that we do not find ourselves in a situation like the one described in Ommen et al. [19]. For ease of reference, the result from Ommen et al. [19] is summarized in the supplemental material [18].

While Theorem 3 provides some interesting insights into the relationships between Bayes Factors and likelihood ratios for forensic identification of source problems, it has some limitations regarding practicality. The first is that this interval estimate for a classical statistic relies on subjective Bayesian probabilities. It is unlikely that either a staunch Frequentist or a fervent Bayesian would be interested in using such an interval. Secondly, what should you do about the uncertainty associated with computational methods for the Bayes Factor? For example, if the Bayes Factor is computed using Monte Carlo integration, it is unclear how, or even if, the computational error should be incorporated into this credible interval for the corresponding likelihood ratio. Nevertheless, we believe
that credible intervals for the likelihood ratio in this form will be useful for exploring sample size determinations for forensic applications. For example, these credible intervals may help us answer the question, “How many control samples should be collected to ensure the value of the Bayes Factor is within some specified tolerance of the likelihood ratio?”

5. CONCLUSIONS

Because the Bayes Factor and likelihood ratio are so interconnected for the forensic identification of source problems, the two terms are used interchangeably in the discipline causing some serious issues and confusion. In this article, we clearly define both statistics, clarifying the distinction is largely due to the chosen definition of probability. Additionally, we have shown several interesting relationships between the two for forensic identification of source problems. The first relationship expresses the Bayes Factor as the expected value of the likelihood ratio function with respect to the posterior distribution for the parameters given the entire collection of evidence as generated under the defense model. This result is interesting because it reduces the complexity of using Bayes Factors both theoretically and practically, and because the resulting Bayes Factors are defined using posterior distributions for the parameters which use (possibly) misspecified models for the evidence. This expression led to a frequentist asymptotic result (using the Bernstein-von Mises theorem) for a Bayesian statistic, specifically that the likelihood ratio is the limiting form of the Bayes Factor when the number of samples from the specific source and/or the number of sources in the alternative source population increases. This result is an important contribution to the field of forensic statistics where the standard asymptotic results for the Bayes Factor do not apply. Finally, this asymptotic result was used to derive a Bayesian method of quantifying uncertainty for a classical statistic, i.e. a credible interval for the likelihood ratio. This is an interesting result because it avoids the issue commonly encountered in the discipline where the value of evidence resulting from the interval estimate is biased against the suspect.

Finally, there are many natural extensions of this work. We are interested in exploring whether the same relationships between the Bayes Factor and likelihood ratio hold when the parameters of the likelihood ratio must be estimated. Additionally, we are interested in exploring solutions to the same types of issues we encountered in the forensic identification of source problems within broader applications of statistical pattern recognition and machine learning.

ACKNOWLEDGEMENTS

We would like to thank our wonderful mentor Dr. JoAnn Buscaglia for the many helpful discussions throughout the years. Since the research presented in this article was performed as part of Dr. Ommen’s dissertation research at South Dakota State University, we would also like to thank the members of her departmental Ph.D. committee, Dr. Kurt Cogswell and Dr. Cedric Neumann, for their support and guidance. Finally, the authors are very grateful to Jakob Sohl and Jade Swanenburg from the Technical University of Delft for their careful technical review of our work and for noticing a mistake in the statement and proof of one of our theorems from an earlier version of this article. The interested reader
(who reads Dutch) can find more detailed versions of the proofs of our theorems in Ms. Swanenburg’s Bachelor’s thesis [24].

REFERENCES


**APPENDIX A: LIKELIHOOD FUNCTIONS FOR THE EVIDENCE**

Let’s begin with the likelihood function for the specific source evidence, s. Recall that the prosecution and defense both agree that the specific source evidence is generated according to model $M_s$. Let $f_s$ denote the density function corresponding to the distribution $F_s$ defined in Equation (1). Then, the likelihood function for the sample s from the specific source is given by

$$
L(\theta_s|s, M_s) = \prod_{j=1}^{N_s} f_s(s_j|\theta_s).
$$

(29)

It will be convenient to use the following notational equivalency, $f(s|\theta_s) = L(\theta_s|s, M_s)$ where the redundant dependence on the model $M_s$ is dropped.

Next, let’s consider the likelihood function for the alternative source population evidence (for both the specific source and common source problems). Again, the prosecution and defense both agree that the alternative source population evidence is generated according to model $M_a$. Let $g$ and $f_a$ denote the density functions corresponding to the distributions $G$ and $F_a$, respectively, from Equation (2) for the specific source and common source problems. Taking a closer look at model $M_a$, it is clear that the $A_{ij}$ elements are independent within a given source, $b_i$, i.e. for $i = 1, 2, \ldots, N_a$

$$(A_{i1}|b_i) \perp (A_{i2}|b_i) \perp \cdots \perp (A_{iN_a}|b_i).$$

So, it will be convenient to partition $A$ into samples consisting of elements from each source, $A = \{A_1, A_2, \ldots, A_{N_a}\}$. Then, model $M_a$ implies that these samples are mutually independent, i.e.

$$A_1 \perp A_2 \perp \cdots \perp A_{N_a}.$$
Therefore, define the likelihood function for the sample from the $i^{th}$ alternative source to be

$$\mathcal{L}_i(\theta_a|\mathbf{A}_i, M_a) = \int \prod_{j=1}^{N_w} f_a(\mathbf{A}_{ij}|b_i, \theta_a) \, g(b_i|\theta_a) \, db_i,$$

(30)

Then, the likelihood function for the hierarchical sample $\mathbf{A}$ from the alternative source population is given by

$$\mathcal{L}(\theta_a|\mathbf{A}, M_a) = \prod_{i=1}^{N_a} \mathcal{L}_i(\theta_a|\mathbf{A}_i, M_a).$$

(31)

It will be convenient to use the following notational equivalency, $f(\mathbf{A}|\theta_a) = \mathcal{L}(\theta_a|\mathbf{A}, M_a)$ where the redundant dependence on the model $M_a$ is dropped.

Now, the definition of the likelihood functions for the unknown source evidence is going to depend upon the identification of source problem and on the sampling model implied by the prosecution or the defense. First, consider the specific source problem. In this scenario, the prosecution hypothesis implies that the unknown source evidence is generated according to the same model as the specific source evidence. Therefore, the likelihood function for the sample $\mathbf{u}$ with unknown source is given by

$$\mathcal{L}(\theta_s|\mathbf{u}, M_p) = \prod_{j=1}^{N_u} f_s(\mathbf{u}_j|\theta_s).$$

(32)

Alternatively, the defense hypothesis implies that the unknown source evidence is generated from one new randomly selected source from the alternative source population, resulting in the following likelihood function

$$\mathcal{L}(\theta_a|\mathbf{u}, M_d) = \int \left[ \prod_{j=1}^{N_u} f_a(\mathbf{u}_j|b_u, \theta_a) \right] g(b_u|\theta_a) \, db_u.$$

(33)

Again, the following notational equivalencies are used for convenience, $f(\mathbf{u}|\theta_s) = \mathcal{L}(\theta_s|\mathbf{u}, M_p)$ and $f(\mathbf{u}|\theta_a) = \mathcal{L}(\theta_a|\mathbf{u}, M_d)$, where the redundant dependence on the models $M_p$ and $M_d$ are dropped.

Finally, consider the likelihood functions for the unknown source evidence in the common source problem. The prosecution hypothesis implies that the unknown source evidence is generated by one new randomly selected source from the alternative source population, where the likelihood function is

$$\mathcal{L}(\theta_a|\mathbf{u}_x, \mathbf{u}_y, M_p) = \int \left[ \prod_{i=1}^{N_x} f_a(\mathbf{u}_x_i|b_u, \theta_a) \right] \left[ \prod_{j=1}^{N_y} f_a(\mathbf{u}_y_j|b_u, \theta_a) \right] g(b_u|\theta_a) \, db_u.$$

(34)

In contrast, the defense hypothesis implies that the unknown source evidence is generated by two different randomly selected sources from the alternative source population. This means that we can find the likelihood function for $\mathbf{u}_x$ and $\mathbf{u}_y$
and then multiply them together because of the assumption that \( u_x \perp u_y \) under model \( M_d \).

\[
\mathcal{L}(\theta_a | u_x, M_d) = \int \left[ \prod_{i=1}^{N_x} f_a(u_{x_i} | b_x, \theta_a) \right] g(b_x | \theta_a) db_x
\]

\[
\mathcal{L}(\theta_a | u_y, M_d) = \int \left[ \prod_{j=1}^{N_y} f_a(u_{y_j} | b_y, \theta_a) \right] g(b_y | \theta_a) db_y
\]

\[
(35) \quad \mathcal{L}(\theta_a | u_x, u_y, M_d) = \mathcal{L}(\theta_a | u_x, M_d) \mathcal{L}(\theta_a | u_y, M_d)
\]

It will be convenient to use the following notational equivalencies, but in this case the dependence on the models are necessary and so it is retained:

\[ f(u_x, u_y | \theta_a, M_p) = \mathcal{L}(\theta_a | u_x, u_y, M_p) \quad \text{and} \quad \]

\[ f(u_x, u_y | \theta_a, M_d) = \mathcal{L}(\theta_a | u_x, u_y, M_d) = f(u_x | \theta_a, M_d) f(u_y | \theta_a, M_d). \]

**APPENDIX B: BAYES FACTOR DERIVATIONS**

**B.1 Equation 27**

First, consider the reciprocal of the specific source Bayes Factor. We will show that it is equal to the expected value of the reciprocal of the specific source likelihood ratio with respect to the posterior distribution given all the evidence has been generated according to the prosecution model.

\[
(36) \quad \frac{1}{\beta_{ss}(E)} = \frac{\int f(u | \theta_a) f(s | \theta_a) f(A | \theta_a) \ d\Pi(\theta_s, \theta_a)}{m(u, s, A | M_p)}
\]

\[
(37) \quad = \int \frac{f(u | \theta_a) \times f(s | \theta_a) \times f(A | \theta_a)}{m(u, s, A | M_p)} \ d\Pi(\theta_s, \theta_a)
\]

\[
(38) \quad = \int \frac{f(u | \theta_a)}{m(u, s, A | M_p)} \ d\Pi(\theta_s, \theta_a | u, s, A, M_p)
\]

\[
(39) \quad \frac{1}{\lambda_{ss}(\theta_s, \theta_a | u)} \ d\Pi(\theta_s, \theta_a | u, s, A, M_p)
\]

\[
(40) \quad = \int \frac{1}{\lambda_{ss}(\theta_s, \theta_a | u)} \ d\Pi(\theta_s, \theta_a | E, M_p)
\]

Therefore,

\[
(41) \quad \beta_{ss}(E) = \left[ \int \frac{1}{\lambda_{ss}(\theta_s, \theta_a | u)} \ d\Pi(\theta_s, \theta_a | E, M_p) \right]^{-1}
\]

**B.2 Equation 28**

First, consider the reciprocal of the common source Bayes Factor. We will show that it is equal to the expected value of the reciprocal of the common source likelihood ratio with respect to the posterior distribution given all the evidence has been generated according to the prosecution model.

\[
(42) \quad \frac{1}{\beta_{cs}(E)} = \frac{\int f(u_x | \theta_a, M_d) f(u_y | \theta_a, M_d) \ d\Pi(\theta_a | A)}{\int f(u_x, u_y | \theta_a, M_p) \ d\Pi(\theta_a | A)}
\]

\[
(43) \quad = \frac{\int f(u_x, u_y | \theta_a, M_d)}{p(u_x, u_y | A, M_p)} \ d\Pi(\theta_a | A)
\]
\[
\begin{align*}
(44) & \quad = \int \frac{f(u_x, y|\theta_a, M_d)}{f(u_x, y|\theta_a, M_p)} \times \frac{f(u_x, y|\theta_a, M_p)}{p(u_x, y|\mathbf{A}, M_p)} \ d\Pi(\theta_a|\mathbf{A}) \\
(45) & \quad = \int \frac{f(u_x, y|\theta_a, M_d)}{f(u_x, y|\theta_a, M_p)} \ d\Pi(\theta_a|u_x, y, \mathbf{A}, M_p) \\
(46) & \quad = \int \frac{1}{\lambda_{cs}(\theta_a|u_x, y)} \ d\Pi(\theta_a|E, M_p)
\end{align*}
\]

Therefore,

\[
(47) \quad \beta_{cs}(E) = \left[ \int \frac{1}{\lambda_{cs}(\theta_a|u_x, y)} \ d\Pi(\theta_a|E, M_p) \right]^{-1}
\]

**APPENDIX C: CONSISTENCY PROOFS**

**Theorem** (Common Source Bayes Factor Consistency):
Given a fixed observation of \(u_x\) and \(u_y\), suppose that the likelihood ratio function \(\lambda_{cs}(\theta_a|u_x, u_y)\) is bounded in a neighborhood of \(\theta_{a_0}\) and that \(\hat{\theta}_a\) is a consistent estimator for \(\theta_{a_0}\). Furthermore, suppose the assumptions of the Bernstein-von Mises Theorem are satisfied. Then as \(n_a \to \infty\), the Bayes Factor converges in \(P_{\theta_a}\)-probability to the true likelihood ratio,

\[
\beta_{cs}(u_x, u_y, A_{n_a}) \xrightarrow{P_{\theta_a}} \lambda_{cs}(\theta_{a_0}|u_x, u_y).
\]

**Proof.** First, let \(E_{n_a}\) denote the random variable associated with generating the entire dataset where \(A_{n_a}\) generates the hierarchical sample from the alternative source population and where the observations \(u_x\) and \(u_y\) are already fixed. Also, let \(\delta_{\theta_{a_0}}\) denote the probability measure that is degenerate at \(\theta_{a_0}\).

\[
\left| \beta_{cs}(u_x, u_y, A_{n_a}) - \lambda_{cs}(\theta_{a_0}|u_x, u_y) \right| = \left| \int \lambda_{cs}(\theta_a|u_x, u_y) \ d \left[ \Pi(\theta_a|E_{n_a}, M_d) - \delta_{\theta_{a_0}}(\theta_a) \right] \right|
\]

Note that \(\Pi(\theta_a|E_{n_a}, M_d) - \delta_{\theta_{a_0}}(\theta_a)\) is a sequence of signed measures. For any signed measure \(\mu\) it follows that \(|\mu(A)| \leq ||\mu||_{TV}\) where \(||\mu||_{TV}\) is the total variation norm [3]. Now, by the assumption that the likelihood ratio function is bounded, let \(\lambda_{cs}(\theta_a|u_x, u_y) \leq C\) for some real number \(C > 0\). Therefore,

\[
\left| \beta_{cs}(u_x, u_y, A_{n_a}) - \lambda_{cs}(\theta_a|u_x, u_y) \right| \leq \int \lambda_{cs}(\theta_a|u_x, u_y) \ d \left[ \Pi(\theta_a|E_{n_a}, M_d) - \delta_{\theta_{a_0}}(\theta_a) \right] \leq C \left\| \Pi(\theta_a|E_{n_a}, M_d) - \delta_{\theta_{a_0}}(\theta_a) \right\|_{TV}.
\]

Now, let \(\Phi(\theta_a|E_{n_a}, M_d)\) denote the probability measure corresponding to the normal distribution with mean \(\hat{\theta}_a\) and variance \(\frac{1}{n_a} I^{-1}_{\hat{\theta}_a}\) where \(I^{-1}_{\hat{\theta}_a}\) is the corresponding inverse of the observed Fisher’s information matrix. Then by the triangle inequality, we obtain

\[
\left\| \Pi(\theta_a|E_{n_a}, M_d) - \delta_{\theta_{a_0}}(\theta_a) \right\|_{TV}
\]
\[ \leq \left\| \Pi(\theta_0 | E_{n_a}, M_d) - \Phi(\theta_0 | E_{n_a}, M_d) \right\|_{TV} + \left\| \Phi(\theta_0 | E_{n_a}, M_d) - \delta_{\theta_0}(\theta_a) \right\|_{TV}. \]

By the Bernstein-von Mises Theorem, as \( n_a \to \infty \) then
\[ \left\| \Pi(\theta_0 | E_{n_a}, M_d) - \Phi(\theta_0 | E_{n_a}, M_d) \right\|_{TV}^{P^{n_a}_{\hat{\theta}_a}} \to 0. \]

By the assumption that \( \hat{\theta}_a \) is consistent and provided that \( I^{-1}_{\hat{\theta}_a} \) is bounded in \( P^{n_a}_{\theta_0} \)-probability in a neighborhood of \( \theta_0 \), then this implies that as \( n_a \to \infty \)
\[ \left\| \Phi(\theta_0 | E_{n_a}, M_d) - \delta_{\theta_0}(\theta_a) \right\|_{TV}^{P^{n_a}_{\hat{\theta}_a}} \to 0. \]

\[ \square \]

**Theorem (Specific Source Bayes Factor Consistency):**

Given a fixed observation of \( u \), suppose that the likelihood ratio function, \( \lambda_{ss}(\theta | u) \), is bounded in a neighborhood of \( \theta_0 \) and that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_0 \). Furthermore, suppose the assumptions of the Bernstein-von Mises Theorem are satisfied. Then the Bayes Factor converges in \( P^{n}_{\theta} \)-probability to the true likelihood ratio as \( n \to \infty \),

\[ \beta_{ss}(u, S_n, A_n) \xrightarrow{P^{n}_{\theta}} \lambda_{ss}(\theta_0 | u). \]

**Proof.** Similar to the proof of Theorem 2. See the supplemental material for more details [18]. \[ \square \]

**APPENDIX D: CREDIBLE INTERVAL PROOFS**

**Lemma (3.1):**

For a given observation of the unknown source evidence, suppose that \( \lambda(\theta) \) is twice continuously differentiable, and that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_0 \) under \( P^{n}_{\theta} \)-probability. If the assumptions of the Bernstein-von Mises Theorem hold, then
\[ \left\| \Pi \left( \sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)] | E_n, M_d \right) - \Phi \left( \sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)] | E_n, M_d \right) \right\|_{TV} \]

converges to zero in \( P^{n}_{\theta} \)-probability as \( n \to \infty \).

**Proof.** Consider the following expression for \( \lambda(\theta) \) resulting from its Taylor series expansion about the maximum likelihood estimate \( \hat{\theta}_n \) (defined using all of the evidence \( E_n \) where the unknown source evidence is generated according to model \( M_d \)):
\[ \sqrt{n} \left[ \lambda(\theta) - \lambda(\hat{\theta}_n) \right] = \sqrt{n}(\theta - \hat{\theta}_n)^T \lambda'(\hat{\theta}_n) + \frac{\sqrt{n}}{2}(\theta - \hat{\theta}_n)^T \lambda''(\hat{\theta}_n)(\theta - \hat{\theta}_n) \]

where \( \lambda' \) is the vector of first partial derivatives of \( \lambda \), \( \lambda'' \) is the matrix of second partial derivatives of \( \lambda \), and \( \hat{\theta}_n \) is a value on the line between \( \theta \) and \( \hat{\theta}_n \). First, consider the error term of the expansion given by
\[ \frac{\sqrt{n}}{2}(\theta - \hat{\theta}_n)^T \lambda''(\hat{\theta}_n)(\theta - \hat{\theta}_n) = \frac{1}{2} \left[ \sqrt{n}(\theta - \hat{\theta}_n) \right]^T \left[ \frac{1}{\sqrt{n}} \lambda''(\hat{\theta}_n) \right] \left[ \sqrt{n}(\theta - \hat{\theta}_n) \right] \]
The Bernstein-von Mises Theorem implies that the \( \left[ \sqrt{n}(\theta - \hat{\theta}_n) \right] \) terms are bounded in \( P_\theta^n \)-probability as \( n \to \infty \). Under the assumption that the maximum likelihood estimate \( \hat{\theta}_n \) is consistent, i.e. \( \hat{\theta}_n \overset{P_\theta^n}{\to} \theta_0 \) as \( n \to \infty \), then the Continuous Mapping Theorem and the Squeeze Theorem imply that the \( \frac{1}{\sqrt{n}} \lambda'(\hat{\theta}_n) \) term converges in \( P_\theta^n \)-probability to zero as \( n \to \infty \). Therefore, Slutsky’s Lemma provides the following result as \( n \to \infty \):

\[
\frac{\sqrt{n}}{2} (\theta - \hat{\theta}_n)^T \lambda'(\hat{\theta}_n)(\theta - \hat{\theta}_n) \overset{P_\theta^n}{\to} 0.
\]

Next, consider the remaining term in the expansion, \( \sqrt{n}(\theta - \hat{\theta}_n)^T \lambda'(\hat{\theta}_n) \). Let \( \Phi(\sqrt{n}(\theta - \hat{\theta}_n)|E_n, M_d) \) denote the sequence of probability measures corresponding to the normal distribution with a zero-vector mean and a covariance matrix of \( I^{-1}_{\hat{\theta}_n} \) where \( I^{-1}_{\hat{\theta}_n} \) is the inverse of the observed Fisher’s information matrix. By the Bernstein-von Mises Theorem, then as \( n \to \infty \)

\[
\left\| \Pi(\sqrt{n}(\theta - \hat{\theta}_n)|E_n, M_d) - \Phi(\sqrt{n}(\theta - \hat{\theta}_n)|E_n, M_d) \right\|_{TV} \overset{P_\theta^n}{\to} 0.
\]

Now, by properties of linear combinations of normal random variables, then this implies that as \( n \to \infty \)

\[
\left\| \Pi(\sqrt{n}(\theta - \hat{\theta}_n)^T \lambda'(\hat{\theta}_n)|E_n, M_d) - \Phi(\sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d) \right\|_{TV} \overset{P_\theta^n}{\to} 0
\]

where \( \Phi(\sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d) \) denotes the sequence of probability measures corresponding to the normal distribution with zero-vector mean and covariance of \( \lambda'(\hat{\theta}_n)^T I^{-1}_{\hat{\theta}_n} \lambda'(\hat{\theta}_n) \). Therefore, as \( n \to \infty \),

\[
\left\| \Pi(\sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d) - \Phi(\sqrt{n}[\lambda(\theta) - \lambda(\hat{\theta}_n)]|E_n, M_d) \right\|_{TV} \overset{P_\theta^n}{\to} 0.
\]

\( \square \)

**Lemma (3.2):**

For a given observation of the unknown source evidence, suppose that \( \hat{\theta}_n \) is a consistent estimator for \( \theta_0 \) under \( P_\theta^n \)-probability and let \( n\sigma^2_n \) converge in \( P_\theta^n \)-probability to a constant \( \sigma^2_n > 0 \) as \( n \to \infty \). If the assumptions of the Bernstein-von Mises Theorem hold for \( \beta(E_n) \), then

\[
\left\| \Phi(\lambda(\theta)|E_n, M_d) - \Psi(\lambda(\theta)|E_n, M_d) \right\|_{TV} \overset{P_\theta^n}{\to} 0.
\]

converges to zero in \( P_\theta^n \)-probability as \( n \to \infty \).

**Proof.** By Theorem 1 and Theorem 2, we have that \( \beta(E_n) \) converges in \( P_\theta^n \)-probability to the true likelihood ratio, \( \lambda(\theta_0) \), as \( n \to \infty \) for both the specific source and the common source problems. Now, consider the likelihood ratio function term, \( \lambda(\hat{\theta}_n) \). For the common source problem, this takes the form

\[
\lambda_{cs}(\hat{\theta}_n) = \frac{f(u_x, u_y|\hat{\theta}_n, M_p)}{f(u_x|\hat{\theta}_n, M_d)f(u_y|\hat{\theta}_n, M_d)}.
\]
where $\hat{\theta}_n$ is the maximum-likelihood estimator of $\theta_a$ that is defined using all the evidence, $E_n$, when the two set of unknown source evidence are generated under model $M_d$. Similarly for the specific-source problem, then $\hat{\theta}_n = \{\hat{\theta}_s, \hat{\theta}_a\}$ is the maximum likelihood estimator for the joint parameter $\theta$ where computed using the entire set of data, $E_n$, when the unknown source evidence is generated under model $M_d$. This means that $\hat{\theta}_a$ is defined using $A_n$, and $u$ under $M_d$, and that $\hat{\theta}_b$ is defined using $s$. Then, the likelihood ratio value is given by

$$\lambda_{ss}(\hat{\theta}_n) = \frac{f(u|\hat{\theta}_s, M_p)}{f(u|\hat{\theta}_a, M_d)}.$$  

It should be noted that there is no theoretical justification for using $\lambda(\hat{\theta}_n)$ as an estimate of $\lambda(\theta_0)$ for either of the forensic identification of source problems. This value is just a construct necessary for the proof. Next, by assumption, then $\hat{\theta}_n$ is a consistent estimator of $\theta_0$. By the Continuous Mapping Theorem and the fact that densities are always non-negative, then applying Slutsky’s Lemma gives the result

$$\lambda(\hat{\theta}_n) \xrightarrow{P_n} \lambda(\theta_0),$$

provided that the unknown source evidence is in the support of the density functions chosen under model $M_d$. Additionally, by the assumption that $n\sigma_n^2$ converges in $P_n$-probability to $\sigma_2^2 > 0$ and by the fact that $\gamma_n^2 \equiv \lambda'(\hat{\theta}_n)^T I_{\hat{\theta}_n}^{-1} \lambda'(\hat{\theta}_n)$ converges in $P_n$-probability to $\sigma_2^2$ due to the consistency of $\hat{\theta}_n$, then

$$n \left( \sigma_n^2 - \frac{1}{n} \gamma_n^2 \right) \xrightarrow{P_n} 0$$

as $n \to \infty$. Finally, because normal distributions are completely determined by their means and variances, and since the previous results give us that the means and variances of both distributions will converge to the same value (the means will converge to the true likelihood ratio value and the difference between the variances will converge to zero), then the Continuous Mapping Theorem gives the needed result:

$$\left\| \Phi(\lambda(\theta)|E_n, M_d) - \Psi(\lambda(\theta)|E_n, M_d) \right\| \rightarrow 0.$$  

\[\square\]

**Theorem** (Approximate $1 - \alpha$ Credible Interval for the LR):

*Let the assumptions of Lemma 3.1 and Lemma 3.2 hold and let*

$$\mathcal{I}_n = \beta(E_n) \pm \Phi^{-1}(1 - \alpha/2)\sigma_n$$  

*where $\beta(E_n)$ represents the sequence of either the common source or the specific source Bayes Factors, $0 < \alpha < 1$ is the desired significance level, $\sigma_n$ is the sequence of posterior standard errors for the likelihood ratio, and $\Phi^{-1}$ is the standard normal quantile function. Then as $n \to \infty$*

$$\Pi(\lambda(\theta) \in \mathcal{I}_n|E_n, M_d) \to 1 - \alpha.$$
PROOF. First, use the Triangle Inequality and Slutsky’s Lemma to combine the results of Lemma 3.1 and Lemma 3.2 to obtain the result, as $n \to \infty$

(48) \[ \left\| \Pi(\lambda(\theta)|E_n, M_d) - \Psi(\lambda(\theta)|E_n, M_d) \right\|_{TV} \xrightarrow{P_n} 0. \]

Since $\Psi(\lambda(\theta)|E_n, M_d)$ denotes the sequence of probability measures corresponding to the normal distribution with mean $\beta(E_n)$ and variance $\sigma_n^2$, then

\[ \mathcal{I}_n = \beta(E_n) \pm \Phi^{-1}(1 - \alpha/2)\sigma_n \]

is a sequence of credible intervals such that as $n \to \infty$

\[ \Psi(\lambda(\theta) \in \mathcal{I}_n|E_n, M_d) \to 1 - \alpha. \]

The previous result in Equation 48 then implies that as $n \to \infty$

\[ \Pi(\lambda(\theta) \in \mathcal{I}_n|E_n, M_d) \to 1 - \alpha. \]

\[ \square \]