The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Score-based Likelihood Ratio Properties for the Specific-Source Identification Problem**
Author(s):	**Janean Hanka**
Document Number:	**308570**
Date Received:	**January 2024**
Award Number:	**Travel Support Report**

Department of Justice
National Institute of Justice

Travel Support Report

Score-based Likelihood Ratio Properties for the Specific-Source Identification Problem

Travel Support Period: June 11 2023 – June 17 2023
Report for the DOJ/NIJ Travel Support Project Code 1.1.4.2.13

Submitted: July 15, 2023

Submitted by:
Janean Hanka
janean.hanka@jacks.sdstate.edu
(605) 415-5531

To:
Mr. Tom Wilson and Dr. Greg Dutton
U.S. Department of Justice
National Institute of Justice

Table of Contents

Coauthors

The coauthors for this work were Dr. Christopher Saunders[1], Dr. Danica Ommen[2], and Dr. JoAnn Buscaglia[3].

Abstract

In the identification of source problem, a likelihood ratio (LR) is used to quantify the value of evidence under two competing models for how the evidence has arisen. When the feature space of this evidence is very complex, a score-based likelihood ratio (SLR) can be used as a surrogate for the value of evidence. Using a SLR results in the use of simpler underlying densities due to the score function mapping the complex evidence to a univariate score; however, it is expected that some information is lost when using a score. Hence, the SLR can perform slightly differently than the LR. In this poster, we discuss four reasonable properties that should be expected of a SLR when used for the specific source identification problem: first, that the SLR can be constructed when the background population consists of one alternative source. Second, when the background population consists of a single alternative source, and we invert the role of the specific source and the alternative source, the full SLR is also inverted. Third, when the alternative source population is composed of multiple sources, the inverse of the omnibus SLR can be written in terms of the average of the inverse of the simple SLR, where the simple SLR is the SLR of the specific source vs one alternative source. Finally, that the SLR does not provide stronger support for either model than a LR. These properties will be formally written and demonstrated on trace element concentrations in aluminum foil sources.

Keywords: Score-based likelihood ratio; trace evidence; explosives

[1]South Dakota State University, Department of Mathematics and Statistics, Brookings, SD 57007
[2]Iowa State University, Department of Statistics, Ames, IA 50011
[3]FBI Laboratory, Research and Support Unit, Quantico, VA 22135

Introduction

This is the report of the poster presented at the International Conference on Forensic Inferences and Statistics (ICFIS) held in Lund, Sweden from June 12-15, 2023. The title of the poster is Score-based Likelihood Ratio Properties for the Specific-Source Identification Problem.

Background and Motivation

Aluminum (Al) is an inexpensive and easy-to-obtain component in improvised explosive devices (IEDs). Al is found in cans, spray paints, foil, binary exploding targets, fireworks, etc. Thus, an amateur bombmaker could go to a local store, purchase Al foil, grind it up in a coffee grinder, blender, or rock tumbler (ball-mill) and use the resulting Al powder in an IED, such as a pipe bomb.

This research determined trace element profiles of recovered Al materials from known sources with sufficient sampling to understand the sources of variability in the measurements and materials, and to assess the potential for using this method for further characterizing sources of Al used in IEDs.

The value of evidence can be presented in a form of a likelihood ratio (LR) when possible; when there is not sufficient information to present an LR, a Bayes Factor is a typically a reasonable alternative. In certain scenarios, it may be preferable to use a score-based likelihood ratio (SLR) instead. This report will discuss what properties a SLR can have as a surrogate for the LR.

Instrumentation and Samples

The trace elemental analysis of samples of 169 rolls of Al foil was performed with Inductively Coupled Plasma Mass Spectrometry (ICP-MS) using ThermoFisher Scientific iCAP RQ with STD and KED mode. The mode used depended on the element being measured. A total of 29 different trace elements were measured per analytical sample. The quantitative analysis used external calibration, internal standardization across the entire mass range, and external standard replicates for quality control [11].

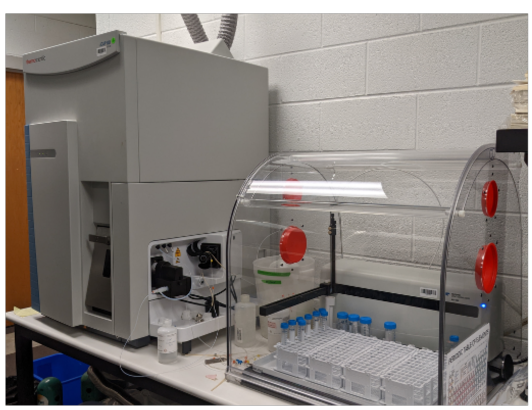

Figure 1: ICP-MS and autosampler used for analysis of digested Al samples.

Sampling Scheme

A specific sampling scheme was used to characterize within- and between-day variability, as well as instrument stability. On the 169 rolls of Al foil, nine measurements were taken: three each from the left (L), center (C), and right (R) sides of the roll. On each subsample, three replicates were taken, for a total of 27 measurements of 29 features each per foil source [11].

The sampling scheme was performed in day pairs. On the first day of a day pair, L, R subsamples (6 subsamples per Al foil source) were analyzed for half of the sources, and C subsamples (3 subsamples per source) were analyzed for the remaining half of the sources. On the second day, the method is reversed. For example, if the L and R subsamples were run on the first day for the 1^{st} and 2^{nd} sources of Al foil, then the C subsamples from those two sources would be run on the second day. In this scenario, the C subsamples from the 3^{rd} and 4^{th} sources of Al foil would run on the first day (with the L and R subsamples from the 1^{st} and 2^{nd} sources), and then the L and R subsamples from the 3^{rd} and 4^{th} sources would be run on the second day.

The full length of foil was sampled for the first 46 rolls, and then the remaining 123 rolls were only sampled from the first 12-18 inches. The different position on the lengths of the foil was found to be irrelevant [11].

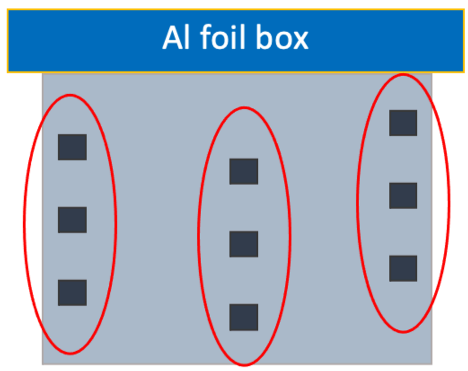

Figure 2: Schematic of subsample locations from a roll of Al foil.

A further summary was taken on each subsample. Recall each subsample consists of three replicate samples. The median of these replicates was taken to result in nine measurements of 29 features each per source. The median was chosen as opposed to the mean, or to no summary, based on the Receiver Operating Characteristic (ROC) curve in Figure 3 (note all statistical work was performed in R [5]). The median line had the highest/flattest ends at the far left and right ends of the plot. This implies that there will be fewer inconclusive values for the SLRs [11].

Score Function

The score function we will use for the SLRs is a modification [3] of the ASTM method used in forensic glass comparisons (hereafter the "ASTM score") [1, 2]. Denote $\Delta(k, q)$, as the ASTM score between a specific/known source observation k and an unknown source/question observation q. We allow k to have $i = 1, \ldots, m$ features and $j = 1, \ldots, n$ observations per feature. The observation q consists of $i = 1, \ldots, m$ features, with only one observation per feature. Thus, let k_{ij} be the i^{th} feature on the j^{th} observation with a known source. We define the following terms:

The known source mean of the i^{th} feature: $\bar{k}_i = \frac{1}{n}\sum_{j=1}^{n} k_{ij}$.

The known source variance of the i^{th} feature: $sd_i = \frac{1}{n-1}\sum_{j=1}^{n}(k_{ij} - \bar{k}_i)^2$.

The ASTM score for the i^{th} feature: $ASTM_i = \frac{|\bar{k}_i - q|}{\max\{sd_i, 0.03\bar{k}_i\}}$.

The overall ASTM score of k and q is then $\Delta(k, q) = \max\{ASTM_i\}$.

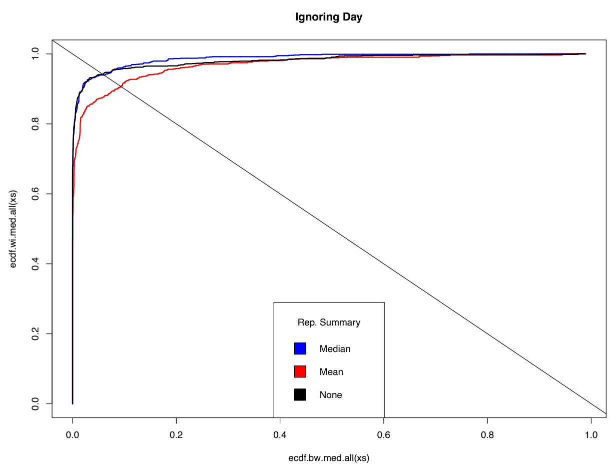

Figure 3: ROC curves of within- and between- scores using median, mean, and no summary statistic across all data collection days.

Specific Source Problem and Likelihood Ratios

Before exploring SLRs, we will first explain the specific source problem setup, and LRs. For more detailed explanations on this topic, see [4, 7, 6, 8, 9, 10].

Define our evidence as $E = \{E_a, E_u, E_s\}$, where E_a are random samples from alternative source(s), comprising a database from the relevant background population; E_u are random samples from the unknown source; E_s are random samples from the specific/known source. Lowercase e_a, e_u, e_s denote observed samples drawn from the random variables E_a, E_u, E_s respectively.

We define a *profile* as the data-generating process of random variables and observations. A *template* is a series of observed or yet-to-be-observed (the random variables) samples from the profile. A *trace* is a single observed or yet-to-be-observed sample from the profile. For example, E_a are random variables of the templates and traces in the background population mapping back to their respective profiles, and e_u is a trace drawn from E_u, which arose from the profile of the specific source samples E_s and e_s, or from a profile in the background population.

The models for the specific source identification problem are

M_0 : The unknown source evidence e_u arose from the same profile as e_s;

M_1 : The unknown source evidence e_u arose from a randomly selected profile in the background population.

Note that these hypotheses imply a simple random sample.

Now, we can define our LR as

$$LR(e_u, e_s, M_0, M_1) = \frac{f(e_u, e_s|M_0, I)}{f(e_u e_s|M_1, I)}$$

where I is any additional information, f is the probability distribution associated with the random variables E_u, E_s, and we are evaluating f at the evidence e_u, e_s. In other words, the LR is defined as the likelihood of observing e_u, e_s if M_0 is true vs the likelihood of observing e_u, e_s if M_1 is true.

Score-based Likelihood Ratios

Recall that we have 29 features. This puts the likelihood function in a 29-dimensional space. This is a complex space to work in, and as the number of features increases, a likelihood function may not exist in that space, or- as is the case with the Al foil- it is too difficult to model. When this occurs, we can use a SLR as a surrogate for the LR. The SLR maps the feature space to the real line using a score function. In this report, we will be using the ASTM score function described previously. While the mapping to the real line provides a simpler underlying density than a LR, the reduction in dimensionality results in a loss of information. For more detailed explanations on this topic, see [4, 7, 6, 8, 9, 10].

Define $\delta = \Delta(e_u, e_s)$ with the ASTM score defined previously. We will refer to this as the "evidence score." Now, we will redefine our models from the LR to models for the SLR:

M_0^* : The evidence score δ arose from pairing a template and a trace object that arose from the same profile;

M_1^* : The evidence score δ arose from the profile of scores obtained by pairing a randomly selected observation from the relevant population with observations from the profile of the specific source.

These models rely on two main assumptions. First, we are anchoring on the profile that generates E_s, and not the observed template e_s; in other words, we are conditioning on the specific source distribution for M_1^*. We can see from the Tippett plot in Figure 4 that this anchoring method results in good separation of same- vs between-source scores. The second assumption is that all within-source comparisons have the same distribution. In other words, the distribution of scores of templates and traces both from the same profile is the same as the distribution of scores from templates and traces both from another, different profile. Under this assumption, we can better approximate the distribution of scores under M_0^* using all of e_a, not just e_s.

Now, we can define our SLR as

$$\lambda(\delta, M_0^*, M_1^*) = \frac{g(\delta|M_0, I)}{g(\delta|M_1, I)}$$

where I is any additional information, g is the probability distribution associated with the scores of the random variables E_u, E_s, and we are evaluating g at the evidence score $\delta = \Delta(e_u, e_s)$. In other words, the SLR is defined as the likelihood of observing the evidence score if M_0^* is true vs the likelihood of observing the evidence score if M_1^* is true.

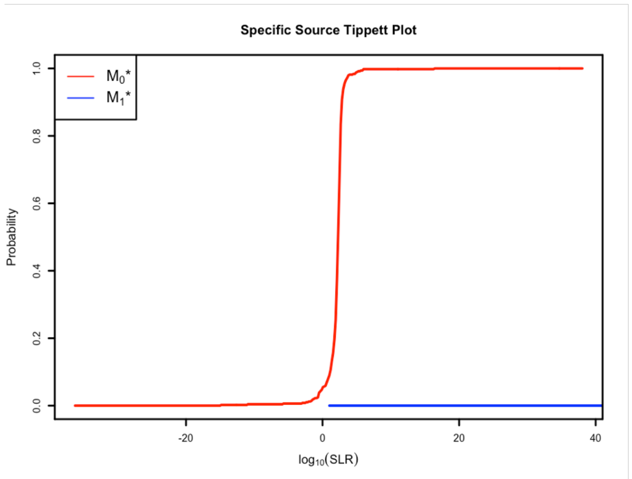

Figure 4: Tippett plot of the probability of the SLRs. This shows that the rate of misleading evidence is almost zero.

Simulation Set Up

Briefly, we will describe how the actual SLR value was computed in the simulation. First, select a source for e_u, and then the same or different source for e_s, depending on which model is true. The background population e_a is created from the N remaining sources. For each source in the background, $i = 1, \ldots, N$, compute the same-source scores $\Delta(e_{a_i}, e_{a_i})$ and denote these scores as $Type = 1$. Next, compute the between-source scores $\Delta(e_s, e_{a_i})$ and denote these scores as $Type = 0$. Then, fit a logistic regression model on all scores to distinguish between $Type$. Finally, use the posteriors on the model to find the SLR, and divide out the base rates. Dividing out the base rates protects against making same- or different-source decisions based on sample size alone. There will be more between-source scores than within-source scores, and thus without dividing out the base rates of same-or different-source comparisons, the model would select the different source option, $Type = 0$, simply because it occurs most often and there are so few $Type = 1$ that selecting $Type = 0$ for all scores minimizes the overall error rate.

Desired Properties

The LR has several desired properties, and since the SLR is a surrogate to the LR, we would like the SLR to have as many of the LR properties as possible. Below is an exploration of four properties of the LR, and a discussion on which of those properties a SLR can have.

Property One

The SLR can be constructed when the background population consists of a single alternative source. Let M_{11}^* be the model that the single alternative source is the actual source of e_u. Then, $\lambda(\delta, M_0^*, M_{11}^*)$ exists, where M_{11}^* is the model for when the background population consists of one source [6].

From our simulation, the SLR exists, thus the property holds.

For M_{11}^* true, we have a value of $\lambda(\delta, M_0^*, M_{11^*}) = 5.525 \times 10^{-12}$.
For M_0^* true, we have a value of $\lambda(\delta, M_0^*, M_{11^*}) = 4.168$.

Property Two

When the background population consists of a single alternative source, and we invert the role of the specific source and the alternative source, the full SLR is also inverted. In other words, $\lambda(\delta, M_0^*, M_{11}^*) = \lambda(\delta, M_{11}^*, M_0^*)^{-1}$ [6].

This property does not hold because the ASTM score function is not a monotonic transform of the LR, and is also an asymmetric function. If the score function is a monotonic transform of the LR, Property Two will hold.

For M_{11}^* true, we have values of $\lambda(\delta, M_0^*, M_{11}^*) = 5.525 \times 10^{-12}$,
$\lambda(\delta, M_{11}^*, M_0^*)^{-1} = 1.987 \times 10^{10}$.
Under M_0^* true, we have values of $\lambda(\delta, M_0^*, M_{11}^*) = 54.168$,
$\lambda(\delta, M_{11}^*, M_0^*)^{-1} = 6.029 \times 10^{15}$.

Property Three

Let M_{1i}^* be the model that the i^{th} source in the alternative population composed of N alternative sources is the actual source of e_u. Then, the inverse of the full SLR can be written in terms of the average of the inverse of the simple SLR which is the SLR of the specific source vs one alternative source. In other words, $\lambda(\delta, M_0^*, M_1^*)^{-1} = \frac{1}{N}\sum_{i=1}^{N} \lambda(\delta, M_0^*, M_{1i}^*)^{-1}$ [6].

See the Appendix for a brief proof of why Property 3 holds for the LR.

This property does not quite hold for the same reasons as those in Property Two, and because we are using a naive suspect profile-anchored approach.

For M_1^* true, we have values of $\lambda(\delta, M_0^*, M_1^*)^{-1} = 7.506 \times 10^{14}$,
$\frac{1}{N}\sum_{i=1}^{N} \lambda(M_0^*, M_{1i}^*)^{-1} = 4.404 \times 10^{16}$.
For M_0^* true, we have values of $\lambda(\delta, M_0^*, M_1^*)^{-1} = 3.212 \times 10^{-3}$,
$\frac{1}{N}\sum_{i=1}^{N} \lambda(M_0^*, M_{1i}^*)^{-1} = 4.055 \times 10^{-2}$.

Property Four

> *The SLR does not provide stronger support for either model than the LR for every possible value of e_u.*
> *If* $LR(e_u, e_s, M_0, M_1) \geq 1,$ *then* $LR(e_u, e_s, M_0, M_1) \geq \lambda(\delta, M_0^*, M_1^*) \geq 1.$
> *If* $LR(e_u, e_s, M_0, M_1) \leq 1,$ *then* $LR(e_u, e_s, M_0, M_1) \leq \lambda(\delta, M_0^*, M_1^*) \leq 1$ [6].

This property holds by the sufficiency principle. Recall we started with a 29-dimensional space for the likelihood function. By reducing the space to one dimension via the score function instead of using all the features, we are losing information. Thus the SLR is not as informative (i.e., cannot provide as strong support) as the LR. We cannot provide stronger support for either model with less information.

Conclusions and Future Work

We can see that the ASTM SLR is not behaving exactly like a LR. This means we cannot claim the SLR is a perfect surrogate for the LR. Note that the score function used is very important. If a score function is chosen such that it is a monotonic transform of the LR, the resulting SLR will posses properties two and three. However, it is not always clear what this monotonic transform is, or even if the LR framework exists in the feature space where the evidence naturally occurs. But if that transform can be found, clearly it would be the score function to use.

Another method to consider for evaluating these properties is using only a cluster of Al sources instead of the full data set. The full data set has high between-source variability relative to the within-source variability, resulting in extreme SLRs (either very large or very small values, for example Property Three under M_1^*). Using just a cluster of Al sources reduces the between-source variability, and so the small values will result in less rounding and numerical imprecision which will help when studying these properties.

A final method is using multiple scores ($N+1$ total), consisting of our evidence score with e_u and e_s, as well as scores between e_u and each of the N observations in e_a. Using more scores will contain more information, thus providing fewer inconclusive values and reducing the rate of misleading evidence. However, it puts the SLR in $N+1$ dimensions instead of just one.

Acknowledgements

The authors would like to thank former FBI Laboratory ORISE postdoctoral research fellows, Drs. Anjuli Bhandari and Michelle Jordan for the ICP-MS method development and sample analysis that provided the data used in this statistical research.

This is an ongoing research effort that was supported in part by an appointment to the Visiting Scientist Program at the Federal Bureau of Investigation Laboratory Division, administered by the Oak Ridge Institute for Science and Education, through an interagency agreement between the US Department of Energy and the FBI. Names of commercial manufacturers are provided for information only and inclusion does not imply endorsement by the FBI or the U.S. Government. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the U.S. Government.

This work was also made possible in part by the National Science Foundation through the National Research Traineeship titled "Cyber-Physical-Social System for Understanding and Thwarting the Illicit Economy under NSF award numbers: DGE 1828492,1828302, 1828462, 1828288.

The research detailed in this presentation was begun with support in part by Award No. 2009-DN-BX-K234 awarded by the National Institute of Justice, of Justice Programs, US Department of Justice.

The presentation of this work at ICFIS in Lund, Sweden was made possible by travel support provided by the National Institute of Justice, Project Code 1.1.4.2.13.

References

1. E2927-16e1 Standard Test Method for the Determination of Trace Elements in Soda-Lime Glass Samples Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry for Forensic Comparisons, ASTM International, 2022.
2. E2330-19 Standard Test Method for Determination of Concentrations of Elements in Glass Samples Using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) for Forensic Comparisons, ASTM International, 2022.
3. O. Loutain, H. Brown, K. Redman, S. Wilczek, M.A. Humphries, A multivariate extension to the standard 4 criterion for comparison of forensic glass evidence. Forensic Sci Int. 338 (111386) (2022).
4. A. Hepler, C. P. Saunders, L. Davis, J. Buscaglia. Score-based likelihood ratios for handwriting evidence. Forensic Sci Int. 219 (1-3) (2012) 129-140.
5. R: A Language and Environment for Statistical Computing. Vienna, Austria (2011). http://www.R-project.org.
6. C. P. Saunders, J. Miller. On Desiderata for Score-based Likelihood Ratios For Forensic Evidence. 2013 Joint Statistical Meetings Conference Contributed Talk 337. Montreal, Canada.
7. D. M. Ommen. Approximate Statistical Solutions to the Forensic Identification of Source Problem (2017).Electronic Theses and Dissertations. 1710. https://openprairie.sdstate.edu/etd/1710
8. C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Applied Statistics 53 (1) (2004) 109-122.
9. C.G.G. Aitken, F. Taroni, S. Bozza, Statistics and the Evaluation of Evidence for Forensic Scientists, 3rd ed., John Wiley and Sons, Chichester, UK, 2021.
10. D. M. Ommen and C. P. Saunders. Building a unified statistical framework for the forensic identification of source problems. Law, Probability and Risk. 17 (2) (2018) 179-197.
11. A. Bhandari, M. Jordan, D. M. Ommen, C. P. Saunders, J. Buscaglia. Statistical Characterization of the Discriminating Power of Trace Elemental Concentrations of Aluminum Foils used in Improvised Explosive Devices. American Academy of Forensic Sciences Conference, Criminalistics Session II. Orlando, FL.

Appendix

Derivation of Property 3 for the LR

We can write the LR as

$$\begin{aligned} LR(e_u, e_s, M_0, M_1) &= \frac{f(e_u, e_s|M_0, I)}{f(e_u, e_s|M_1, I)} \\ &= \frac{f(e_u, e_s|M_0, I)}{N^{-1}\sum_{i=1}^{N} f(e_u, e_s|M_{1i}, I)}. \end{aligned} \tag{1}$$

Then, considering the inverse (and thus avoiding Jensen's Inequality)

$$\begin{aligned} \frac{1}{LR(e_u, e_s, M_0, M_1)} &= \frac{N^{-1}\sum_{i=1}^{N} f(e_u, e_s|M_{1i}, I)}{f(e_u, e_s|M_0, I)} \\ &= N^{-1}\sum_{i=1}^{N} \frac{f(e_u, e_s|M_{1i}, I)}{f(e_u, e_s|M_0, I)} \\ &= N^{-1}\sum_{i=1}^{N} \frac{1}{LR(e_u, e_s, M_0, M_{1i})}. \end{aligned} \tag{2}$$

Where we get Equation 1 by a strong laws version of the denominator, and Equation 2 by moving a constant inside the sum.