Effects of prescreening for likelihood ratio approaches in forensic source identification

November 27, 2023

Dylan Borchert, Semhar Michael, Andrew Simpson, Christopher P. Saunders, and Larry Tang

Dylan Borchert is a Ph.D. student in the Department of Mathematics Statistics at South Dakota State University

Semhar Michael is an Associate Professor in the Department of Mathematics and Statistics at South Dakota State University, Brookings SD 57007. email: semhar.michael@sdstate.edu.

Andrew Simpson is a Ph.D. student in the Department of Mathematics Statistics at South Dakota State University

Christopher P. Saunders is a Professor in the Department of Mathematics and Statistics at South Dakota State University, Brookings SD 57007.

Larry Tang is a Professor in the Department of Statistics and Data Science at the University of Central Florida, Orlando FL 32816, and Guest Researcher at the Rehabilitation Medicine Department, NIH Clinical Center, Bethesda, MD

Conflicts of Interest Statement:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Effects of prescreening for likelihood ratio approaches in the forensic identification of source problems

Abstract

Prescreening is a methodology where forensic examiners select samples similar to given trace evidence to represent the background population. This background evidence helps assign a value of evidence using a likelihood ratio or Bayes factor. A key advantage of prescreening is its ability to mitigate effects from subpopulation structures within the alternative source population by isolating the relevant subpopulation. This paper examines the impact of prescreening before assigning evidence value. Extensive simulations with synthetic and real data, including trace element and fingerprint score examples, were conducted. The findings indicate that prescreening can provide an accurate evidence value in cases of subpopulation structures but may also yield more extreme or dampened evidence values within specific subpopulations. The study suggests that prescreening is beneficial for presenting evidence relative to the subpopulation of interest, provided the prescreening method and level are transparently reported alongside the evidence value.

Keywords: value of evidence, likelihood ratio, relevant source population, latent population structures, forensic source identification

1. Introduction

In forensic source identification, the forensic examiner is often tasked with providing a value of evidence relative to two competing propositions about the source of the evidence [1, 2]. In the specific source problem, the first proposition is often referred to as the prosecution hypothesis, H_p, that the trace evidence is from the specified source of known origin. The second proposition is commonly called the defense hypothesis, H_d, that the trace evidence is from a randomly selected source in the alternative source population [3]. Often a sample from the background population is used to model the population of alternative sources [4, 5]. A slightly different problem that is encountered in forensic source identification is the common but unknown source problem. The hypotheses that the forensic examiner is evaluating now change to H_p : the two pieces of trace evidence are from the same randomly selected source in the alternative source population versus H_d : the two pieces of trace evidence are from two different randomly selected sources in the population of sources [3]. Like in the specific source problem, a sample from the background population is often introduced to model the alternative source population. The key difference between the two problems is that in the specific source problem, the source of one of the pieces of evidence is known [3].

The use of the likelihood ratio (LR) approach for presenting a value of evidence has been proposed for

a variety of evidence types such as glass [4, 5, 6], handwriting [7, 8], footwear marks [9], fingerprints [10], facial images [11], and speech [12]. The normal-based likelihood ratio methods such as Lindley's Bayes Factors (BFs) [4, 5, 13] and a multivariate kernel density-based LR [5] are commonly used methods for presenting a value of evidence. The normal-based Bayes factors are also presented as specific source BFs, but are equivalent to plug-in estimates of common but unknown source LRs due to prior specification [14]. They are often used in the specific source setting, as such the paper will discuss these LRs through the specific source lens. These methods also assume a hierarchical normal sampling structure. This normality assumption may not always be reliable, as there can be a subpopulation structure in the population of sources.

In likelihood ratio-based approaches, evidence about the relevant source population is needed to account for nuisance parameters. There has been discussion about what constitutes the relevant source population for use in speaker recognition [15, 16], glass, and textiles [17]. Discussions include whether information related to the trace evidence should be used in building the evidence about the background population, such as known categorical variables including speaker accent, fiber color, etc. There is also discussion of an important note, that by changing the relevant population considered, the hypotheses about the source of the trace evidence also change relative to that known subpopulation structure [16].

Prescreening is a methodology in which a forensic examiner uses a subset of the given evidence about background (or control) objects in comparison with the trace evidence. Using this subset changes the proposed population from which the alternative source can arise. For example, prior to producing a value of evidence for a fingerprint comparison, the forensic examiner may do a database search to find the most similar known fingerprints to the latent print to calculate the value of evidence. As different examiners may use a different number of similar known fingerprints for comparison, studying the changes in the value of evidence as a result of prescreening is of interest. The National Institute of Standards and Technology (NIST) has developed process maps indicating current practices in forensic science. The process maps for friction ridge [18], speaker recognition [19], and handwriting [20] discuss selecting data to use in providing a summary of evidence, in which a subset of the original database may be used in comparison with the trace. Though likelihood ratios may not be common practice for all of these evidence types, it suggests that examiners sometimes think in terms of a relevant source population or a subpopulation of the original population. This paper extends previous works such as Ref. [16] that considered LRs for known subpopulation structures to unknown subpopulation structures characterized by a latent variable that cannot be accounted for a priori.

This work discusses the implications of prescreening prior to obtaining an LR. One potential advantage of prescreening is that it may be useful in isolating samples from a subpopulation of interest. This subpopulation of interest may be characterized by a known variable, or a latent variable. In the case of the latent variable, when the subpopulation structure is well studied and the rate at which samples are en-

countered from the subpopulation are known, prescreening can be useful in isolating the samples from the subpopulation of interest in the database. This in turn can help assign a value of evidence relative to this subpopulation, which is the value of evidence assuming the true membership labels are known for objects belonging to the same subpopualtion as the trace evidence. In cases where there is not a subpopulation structure prescreening tends to give a dampened value of evidence relative to the LR in the entire population. In either case the propositions being evaluated change, as such the prescreening methodology should be transparently reported alongside the value of evidence, if it is used. The rest of the paper is organized as follows. Section 2 describes the proposed procedure to study prescreening prior to the evaluation of an LR or a SLR. Section 3 discusses the results from extensive simulation studies. Section 4 presents the application to various datasets in forensics. Section 5 concludes the paper.

2. Methods

2.1. Preliminaries

One method of comparing the sampling models corresponding to H_p and H_d is through a likelihood ratio (LR) or a Bayes factor (BF) [21], where sampling models M_p and M_d are specified corresponding to H_p and H_d, respectively and then the resulting ratio of (marginal) likelihoods is presented. An LR value greater than one suggests support for M_p, where an LR value less than one suggests support for M_d. It is important to note that M_p and M_d may differ under the specific source and common but unknown source problems. As such, an LR for the specific source problem will not necessarily be the same as an LR under the common but unknown source problem [3].

To begin discussing the sampling models consider $\boldsymbol{Y}_{sj}$ for $j = 1, \ldots, n_w$, which will denote the j^{th} sample from the specific source. We have the specific source evidence, E_s , which is alternatively called the control evidence, is the set of all n_w observation from the specific source. That is $E_s = \{\boldsymbol{Y}_{sj} : j = 1, \ldots, n_w\}$. Similarly, let $\boldsymbol{Y}_{uj}$ for $j = 1, \ldots, n_w$ denote the j^{th} sample of the recovered evidence of unknown source. We define E_u as the the collection of samples from the unknown source, which is also called the trace evidence. This gives $E_u = \{\boldsymbol{Y}_{uj} : j = 1, \ldots, n_w\}$. Finally, consider $\boldsymbol{Y}_{ij}$ for $i = 1, \ldots, n_a$ and $j = 1, \ldots, n_w$, which is the j^{th} sample from the i^{th} background object. We will denote the collection of samples from the i^{th} object or source as $O_i = \{\boldsymbol{Y}_{ij}, j = 1, \ldots, n_w\}$, and the collection of measurements from all of the n_a objects as $E_a = \{\boldsymbol{Y}_{ij} : i = 1, \ldots, n_a, j = 1, \ldots, n_w\}$. We will refer to E_a as the evidence about the alternative source population or evidence about the background population. Thus we have our complete set of evidence, $E = \{E_s, E_u, E_a\}$.

Following the developments in Ommen et al. [21], this model for generating background objects illustrates the case where there is not a latent subpopulation structure in the alternative source population. When there is not a subpopulation structure we will refer to the background population as homogeneous. In this case, we will assume that the observations are generated by a hierarchical normal sampling model,

M_1, written as

$$\begin{aligned} \boldsymbol{Y}_{ij} &= \boldsymbol{a}_i + \boldsymbol{\epsilon}_{ij}, \\ \boldsymbol{a}_i &\sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_a), \\ \boldsymbol{\epsilon}_{ij} &\sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon), \end{aligned} \tag{1}$$

for $i \in \{1, \ldots, n_a\}$ and $j \in \{1, \ldots, n_w\}$. In this model a_i represents the i^{th} randomly selected source's mean and ϵ_{ij} represents the deviation of the j^{th} sample from the i^{th} source's mean. Here $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_\epsilon$ represent the between- and within-source covariance matrices, respectively. For the case where the alternative source population has a subpopulation structure, we propose the following hierarchical sampling model with a Gaussian mixture model at the source level. This second model for generating background objects, M_2, for $i \in \{1, \ldots, n_a\}$ and $j \in \{1, \ldots, n_w\}$ is given as

$$\begin{aligned} \boldsymbol{Y}_{ij} &= \boldsymbol{a}_i + \boldsymbol{\epsilon}_{ij}, \\ Z_i &\sim Categorical(K, \pi_1, \ldots, \pi_K), \\ \boldsymbol{a}_i | Z_i = k &\sim MVN(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \\ \boldsymbol{\epsilon}_{ij} &\sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon). \end{aligned} \tag{2}$$

In this model, Z_i represents the result of randomly selecting one of the subpopulations in the alternative source population. Given the result of $Z_i = k$, the random selection of the source mean, a_i, is characterized by the source level distribution of the k^{th} subpopulation. Like before, ϵ_{ij} represents a deviation of the j^{th} sample from the i^{th} source mean. The distribution of these within source deviations is assumed to be the same across all of the subpopulations.

In addition, for the control evidence, we assume that E_s is generated according to the model M_s which is given by

$$\boldsymbol{Y}_{sj} \sim MVN(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{3}$$

where $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ are the mean and covariance matrix from the specific source from which the control evidence originated. Finally, to generate trace evidence, there are two competing models relative to the two competing propositions: the defense, H_d, and the prosecution, H_p. Under the prosecution proposition, M_p will be identical to M_s. Therefore, $\boldsymbol{Y}_{uj} \sim MVN(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$. On the other hand, under the defense proposition, E_u is treated as being randomly sampled from our alternative source population. This gives M_d is the same as M_1 or M_2 with $n_a = 1$, depending on the existence of subpopulations in the alternative source population. We will denote realizations of the evidence and random variables/vectors as lower case letters, where our observed set of evidence is $e = \{e_s, e_u, e_a\}$ and an observed sample is $\boldsymbol{y}_{i'j}$, for $i' \in \{s, u, 1, \ldots, n_a\}$ and $j = 1, \ldots, n_w$.

2.2. Prescreening for LR methods

The general prescreening procedure used by practitioners is defined as using a subset of the evidence about the background population for use in calculating the value of evidence. For ease, the general prescreening procedure for use with LR methods is written out in algorithmic form in Algorithm 1. Therefore, in our notation, given a database (e_a) the expert will select a subset e_a^* using some criterion for selecting background objects o_i for $i \in \{1, \ldots, n_a\}$ for use as the sample from the relevant population.

Algorithm 1 General Prescreening Procedure

1. Given evidence e_s, e_u, and e_a.
2. Construct e_a^*, a subset of e_a, using subsetting method of preference.
3. Calculate the LR using e_s, e_u, and e_a^*.

Here we develop an algorithm to study the effects of prescreening on trace-element data. We do this by using a distance or a dissimilarity score between the trace evidence e_u, the control evidence e_s, and o_i for $i \in \{1, \ldots, n_a\}$. A logical score to use is a test statistic for a difference of means. Due to the typical constraint that we have access to many classes with few samples within each class, we use a statistic representing Hotelling's T^2 statistic, but utilizing a within-source sample covariance matrix pooled across all background objects. The score used is $\delta(e_u, o_i) = \frac{2}{n_w}(\bar{\boldsymbol{y}}_{u\cdot} - \bar{\boldsymbol{y}}_{i\cdot})' \boldsymbol{S}_w^{-1} (\bar{\boldsymbol{y}}_{u\cdot} - \bar{\boldsymbol{y}}_{i\cdot})$, where $\boldsymbol{S}_w = \frac{1}{n_a(n_w-1)} \sum_{i=1}^{n_a} \sum_{j=1}^{n_w} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_{i\cdot})'$. This dissimilarity score will be used to build the sample from relevant background population, which will then be used to calculate two likelihood ratios. The likelihood ratios used are the normal-based (MVN-based) and density-based (KDE-based) LRs proposed by Aitken and Lucy [5] and summarized in Zadora et. al [22]. The general prescreening procedure for trace-element data is as follows. Given the control evidence, trace evidence, and evidence about the background population, the objects in the sample of the background population are ordered based on their similarity to the trace evidence. Then after specifying the proportion of background objects or a threshold for the precreening score, a subset of objects is selected for use as the sample from relevant source population to calculate the LR. It is important when constructing the relevant source population that we do not construct it only with objects that are more similar to the trace object than the control object. For a given trace e_u and control e_s elements and the number of subsets L, the details to study this procedure are outlined in Algorithm 2.

2.3. Presecreening for SLR methods

We also examine the effects of prescreening on score-based likelihood ratios in the context of fingerprints. Given a score function, $\rho(\cdot, \cdot)$, that produces a similarity score of full rolled fingerprints to slap fingerprints, we can prescreen our background samples using typicality of the same source scores [23]. This is done using the empirical CDF of the same source scores $\hat{G}(\cdot)$, which estimates the CDF $G(\rho(o, s)) = Pr(\rho(O, S) \leq$

Algorithm 2 Studying Prescreening for Trace Element Data

1. Given evidence e_s, e_u, and e_a.
2. Choose L, the number of subsets.
3. Let n_a be the number of sources and n_w be the number of samples in each source of e_a.
4. Calculate $\tau = \delta(e_s, e_u)$.
5. Calculate $\delta_i = \delta(e_u, o_i)$ for $i = 1, \ldots, n_a$.
6. Construct $Q = \{i : \delta_i > \tau\}$ and order Q based on decreasing order of $\{\delta_i : i = 1, \ldots, n_a\}$.
7. Construct a sequence $N_1 = 1, \ldots, N_L = |Q|$.

for $j = 1, \ldots, L$ **do**

8. Construct e_a^* by removing the first N_j sources in Q from e_a.
9. Calculate the LR using e_s, e_u, and e_a^* and store.

end for

Algorithm 3 Studying Prescreening for SLR

1. Given the score $\rho(e_s, e_u)$ for e_u and e_s
2. Choose L, the number of subsets
3. Define e_a as the remaining sources.
4. Compute $\hat{G}$, the Empirical CDF of the same source scores of objects in e_a.
5. Given $\tau = \hat{G}(\rho(e_s, e_u))$ and $\delta_i = \hat{G}(\rho(o_i, e_u))$ $i = 1, \ldots, n_a$
6. Construct $Q = \{i : \delta_i < \tau\}$ and order Q based on δ_i in increasing order.
7. Construct a sequence $N_1 = 1, \ldots, N_L = |Q|$.

for $j = 1, \ldots, L$ **do**

8. Construct e_a^* by removing the first N_j sources in Q from e_a.
9. Calculate the SLR using all pairwise comparisons of sources in e_a^*

end for

$\rho(o, s))$, where O and S are randomly generated full rolled and slap fingerprints from the same randomly selected source. This prescreening procedure involves keeping objects where $\hat{G}(\rho(o_j, e_u)) > \alpha$, *i.e.* for a given α we cannot exclude the source that generated o_j as giving rise to the trace slap fingerprint. This assumes that it is possible to get the score between the trace object and the background objects in the database. The score-based LR used is the plug-in estimate of the likelihood ratio using logistic regression. Given n_1 between-source scores and n_2 within-source scores of objects in the sample from the background population, we have $SLR = \frac{\tau(\rho(e_s,e_u))}{1-\tau(\rho(e_s,e_u))} / \frac{n_1}{n_2}$, where $\tau(\rho(e_s, e_u)) = \frac{1}{1+exp\{-\hat{\beta}_0 - \hat{\beta}_1 * \rho(e_s,e_u)\}}$ and $\hat{\beta}_0$ and $\hat{\beta}_1$ are the MLE's of the logistic regression fit on the known between-source and within-source scores in the background population [24]. Similarly to the trace-element data, the prescreening procedure for use with SLRs is as follows. Given the score comparing the trace object to the control object, the background

objects are ordered based on their scores with the trace object. Then after specifying a number of sources or a threshold for similarity, construct the sample from the relevant source population, and use the scores that compare two objects in the sample from the relevant source population to calculate the SLR. The details to study this procedure are shown in Algorithm 3.

3. Simulation study

To study the effects of prescreening on the value of LRs, we implement a Monte Carlo simulation to generate synthetic trace element data from different models for use as a sample from the background population. The background samples were generated from a population with and without subpopulation structures. We then control the mean of the control evidence e_s and the trace evidence e_u to see the effects of prescreening for these different scenarios. The simulations are implemented according to Algorithm 4.

Algorithm 4 Prescreening Simulation

1. Generate e_s and e_u from $MVN(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_\epsilon)$ and $MVN(\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_\epsilon)$, respectively with n_w samples each.
2. Calculate $\tau = \delta(e_s, e_u)$.
3. Choose L, the number of subsets

for $t = 1, \ldots, B$ **do**

4. Generate e_a with n_a objects and n_w samples from each object from the hierarchical model of choice (M_1 or M_2).
5. Calculate $\delta_i = \delta(o_i, e_u)$ for $i = 1, \ldots, n_a$.
6. Construct $Q = \{i : \delta_i > \tau\}$ and order Q based on decreasing order of $\{\delta_i : i = 1, \ldots, n_a\}$.
7. Construct a sequence $N_1 = 1, \ldots, N_L = |Q|$.

for $j = 1, \ldots, L$ **do**

8. Construct e_a^* by removing the first N_i sources in Q from e_a.
9. Calculate the MVN and KDE-based LRs using e_s, e_u, and e_a^* and store.

end for

end for

3.1. Simulation setup

Two models are used to study the effects of prescreening used in combination with the MVN-based and KDE-based LRs via Monte Carlo simulations. All the simulations are implemented according to Algorithm 4 with $n_a = 100$, $n_w = 10, B = 10$, and $L = 50$. The list of all parameters used for the generative models can be found in Appendix A.

The first simulation experiment (Sim 1) utilizes a bivariate model with two subpopulations generated with a small overlap between components derived from misclassification probabilities as given by Ref. [25].

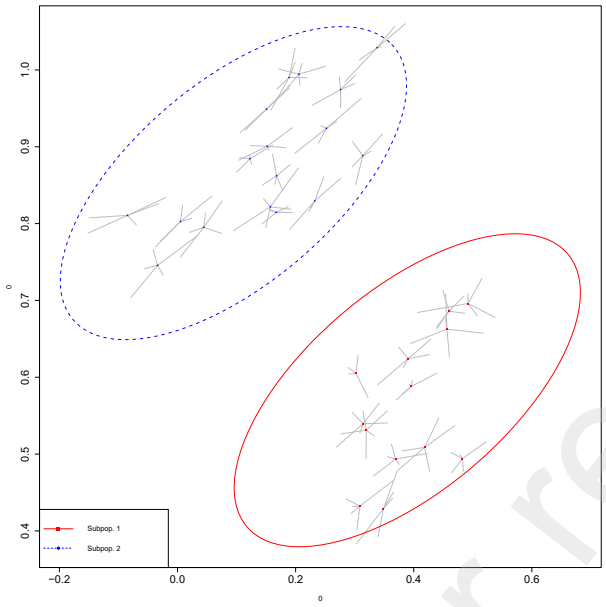

Figure 1: Sample generated from the bivariate model with two subpopulation structures (in Sim 1) represented by red and blue colors. The source means (red and blue points) and objects from each source (gray segments) for each subpopulation are given. The 95% probability contours for each component in the between-source distribution mixture model are shown in blue dashed line and red solid line.

The components of the source-level mixture have the same covariance matrix and the overlap is equal to 0.001 for both clusters. The covariance matrix for the within-source distribution is the covariance of the between-source distribution's components scaled by 0.1. The second simulation study (Sim 2) is based on data-suggested model. Sim 2 utilizes an 8-dimensional model with two subpopulations suggested by copper wire data presented in Dettman et al. [26] where the model parameters are given. Figures 1 and 2 show pairwise scatter plots of example data sets generated from the sampling models given in Sim 1 and Sim 2, respectively. For both cases, 30 sources with 5 observations from each source are generated and displayed along with the 95% probability contours [27] for the between source distribution. For the bivariate model, we can see that the subpopulations are well separated as shown by the elliptical contours in Figure 1. The Figure also visualizes the within source variation as shown by observations as line segments branching from each source mean. The 8-dimensional copper model is visualized in Figure 2. We can see that in certain dimensions there is more overlap and there is more separation in other dimensions. For these simulation studies the *MixSim* R package [28] is utilized to generate mixtures and/or data sets from the corresponding mixtures. Also, a modified version of the hotelling.stat() function from the *Hotelling* package is used to calculate the prescreening statistic [29].

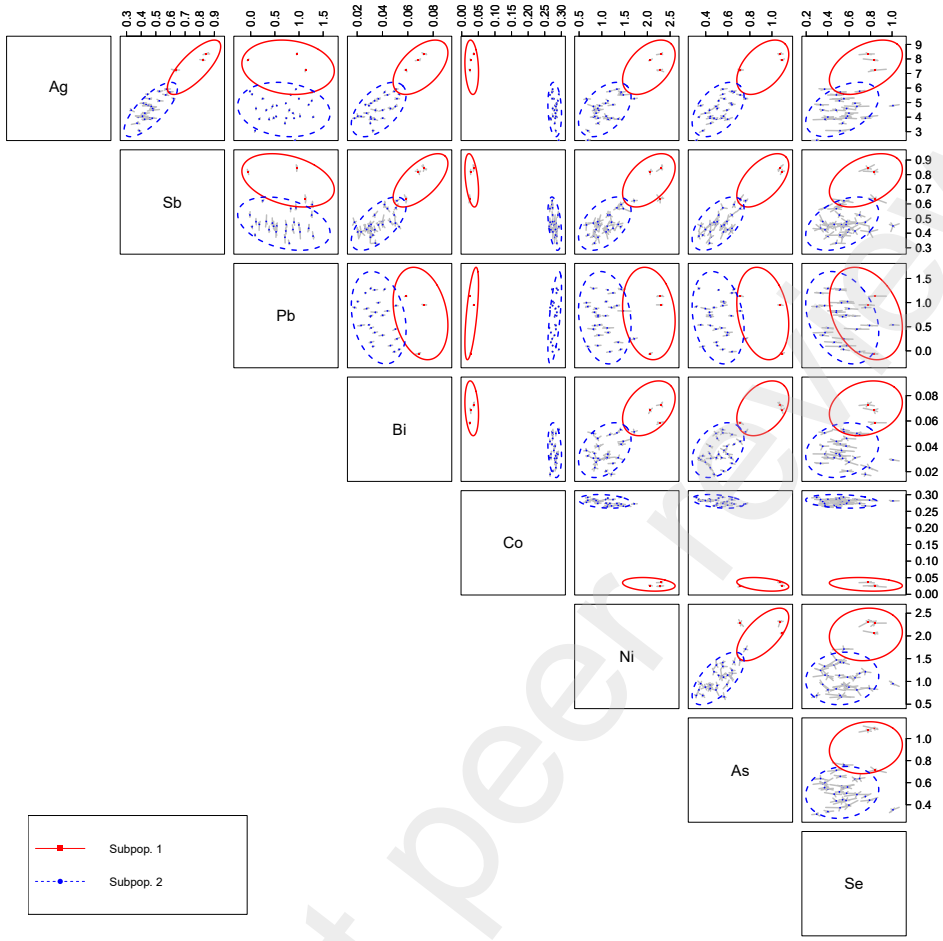

Figure 2: Sample generated from the copper data suggested model in Sim 2. The source means of the two subpopulations (red and blue points) and objects from each source (endpoints of the gray segments) are given. The 95% probability contours for each component in the between-source distribution mixture model are shown in blue dashed line and red solid.

3.2. Simulation results

The results of the Monte Carlo simulations implemented according to Algorithm 4 are displayed in Figures 3 - 6 with subplots in each figure representing different settings. In all the subplots within each figure, the vertical axis displays $\log_{10}(LR)$ (LLR) for both the MVN-based and KDE-based LRs, and the horizontal axis corresponds to the proportion of the candidate sources remaining in the sample from the relevant source population after prescreening. It is important to note that the horizontal axis has been reversed in order to visualize the value of the evidence as a function of prescreening, that is, subsetting the evidence about the background population further and further to consider a more refined relevant source population. In each subplot, the MVN-based LR, the KDE-based LR, and the true value of the LR calculated within the subpopulation containing e_u with the true parameters of the generative model are shown. If the number of sources remaining in the collection of background samples is less than the number

of sources needed to estimate the between-source covariance matrix before the stopping condition is met (all L subsets were considered), the last point in the plot is displayed as a star.

For the prosecution cases, we will consider three cases for the shared mean of the control and trace object. The first case we consider is when e_s and e_u have a mean at the center of the subpopulation of interest, that is when $\boldsymbol{\mu}_s = \boldsymbol{\mu}_u = \boldsymbol{\mu}_g$, where g is the index of the subpopulation of interest. In the simulations, $g = 1$ and $g = 2$, for Sim 1 and Sim 2, respectively. We will refer to this mean as $\boldsymbol{\mu}_{u_1}$. The second and third cases considered involve "rare" means. These utilize the points $\boldsymbol{w}_{\boldsymbol{v}} = \boldsymbol{\mu}_g + \alpha\frac{\boldsymbol{v}}{||\boldsymbol{v}||}$ such that $(\boldsymbol{w}_{\boldsymbol{v}} - \boldsymbol{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{w}_{\boldsymbol{v}} - \boldsymbol{\mu}_g) = \chi^2_{p,.99}$, where $\boldsymbol{v}$ is the direction vector and $\chi^2_{p,.99}$ is the 99^{th} percentile of the χ^2 distribution with p degrees of freedom. Solving for α we get $\alpha = \sqrt{\frac{\chi^2_{p,.99}||\boldsymbol{v}||}{\boldsymbol{v}'\boldsymbol{\Sigma}_g^{-1}\boldsymbol{v}}}$. The second case uses the mean $\boldsymbol{\mu}_{u_2}$, which is $\boldsymbol{\mu}_s = \boldsymbol{\mu}_u = \boldsymbol{w}_{\boldsymbol{v}_1}$, where $\boldsymbol{v}_1 = \boldsymbol{\mu}_g - \boldsymbol{\mu}_{g'}$ and g' is the index of the subpopulation that is not the subpopulation of interest. The third case uses the mean $\boldsymbol{\mu}_{u_3}$, which is $\boldsymbol{\mu}_s = \boldsymbol{\mu}_u = \boldsymbol{w}_{\boldsymbol{v}_2}$, where $\boldsymbol{v}_2 = \boldsymbol{\mu}_{g'} - \boldsymbol{\mu}_g$.

Figure 3 shows the results of the prescreening simulation under the prosecution cases when the alternative source population is homogeneous, that is when e_s and e_u share the same source and e_a was generated under M_1. The rows in the figure represent the two generative models used to generate data (Sim 1 and Sim 2) as set up in Section 3.1. The columns show the three locations for the mean, that is $\boldsymbol{\mu}_{u_1}, \boldsymbol{\mu}_{u_2}$, and $\boldsymbol{\mu}_{u_3}$. Looking at the first case $\boldsymbol{\mu}_{u_1}$, the KDE-based LR provides a larger value of evidence than the MVN-based LR. As we prescreen in this case, both LRs decrease gradually away from the true LR within the subpopulation of interest. When we consider the cases where $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_u$ are away from the mean of the subpopulation of interest ($\boldsymbol{\mu}_{u_2}$ and $\boldsymbol{\mu}_{u_3}$), both LLRs again start around the target value of the LLR and in some iterations remain relatively constant and in some cases increases as we prescreen e_a further and further. We see in Sim 2 that some of these increases are very dramatic where the LR increases in many orders of magnitude. Also, when the e_a contains a relatively small number of objects compared to the dimensionality of the data the value of the LLR behaves radically, likely due to parameter estimation.

Figure 4 shows the prosecution cases when the alternative source population is heterogeneous, that is when e_a was generated under M_2. The rows and columns have the same interpretation as in Figure 3. We see a similar behavior when there is a subpopulation structure in the alternative source population where the LR decreases when the means of e_u and e_s are centered in the subpopulation of interest and remains constant or increases when $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_u$ are moved away from the mean of the subpopulation of interest. When we have this subpopulation structure, we see sharp decreases or increases in the value of the LLR. The sharp changes happen when around 50% and 80% of the original objects remain in e_a in Sim 1 and Sim 2, respectively, which line up with the mixing proportions of the subpopulations of interest suggesting that they occur when the samples of background objects consists of only the objects from the subpopulation of interest. These sharp changes in the LLR slow around the true value of the LLR in the subpopulation of interest, suggesting that prescreening can be useful in returning a value of evidence relative to the

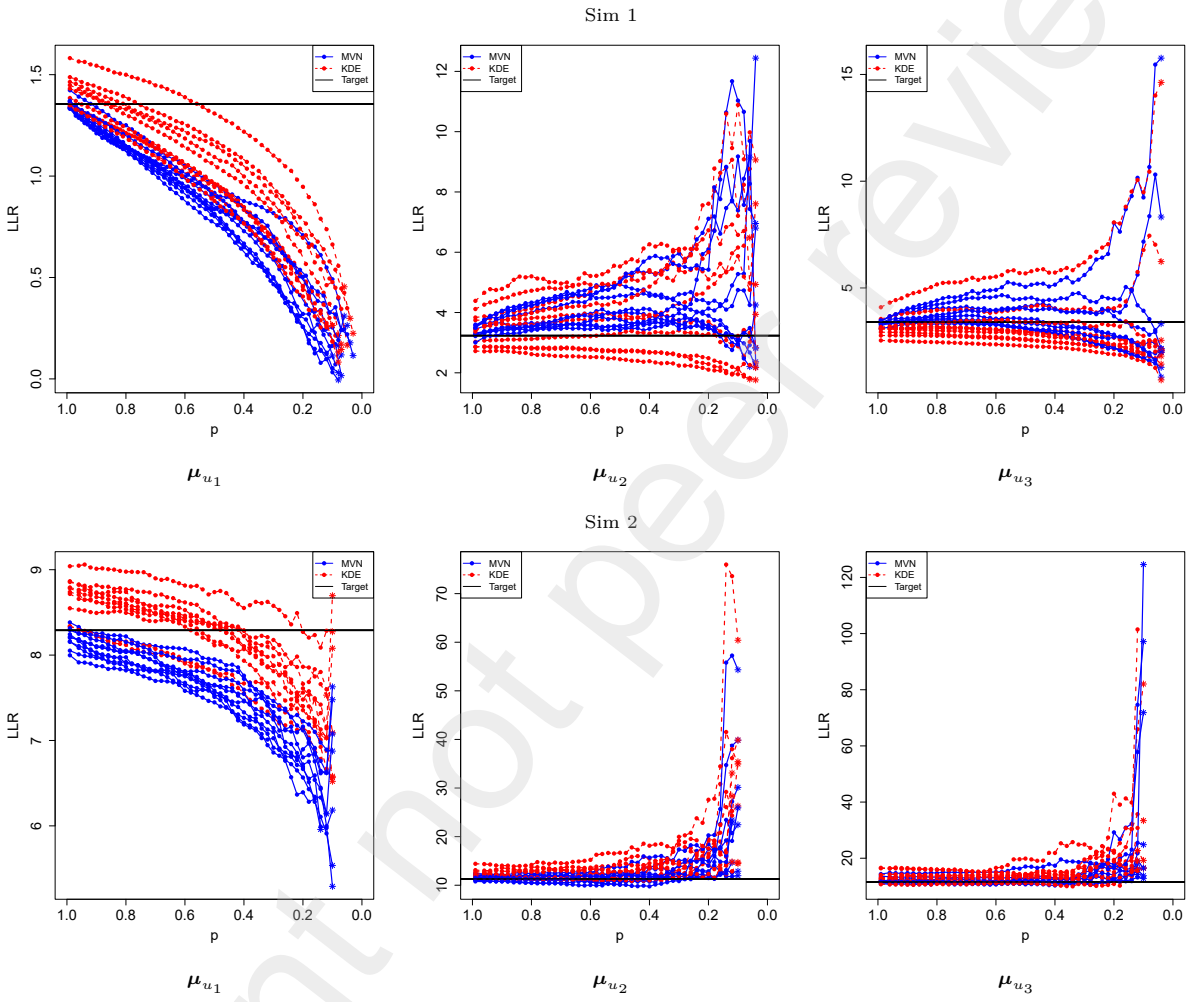

Figure 3: Prescreening simulation results under the prosecution cases when there is not a subpopulation structure in the alternative source population. The KDE-based LLR in (red dashed lines), the MVN-based LLR (solid blue lines), and the true LLR in the subpopulation that contains e_u (solid black line) are presented versus p the proportion of background objects remaining. The rows represent Sim 1 and 2 and the columns are for the different locations of e_u and e_s.

subpopulation of interest when the approximate value of mixing proportions is known. Again, when there are relatively few sources in the relevant source population the LLR value can drastically change.

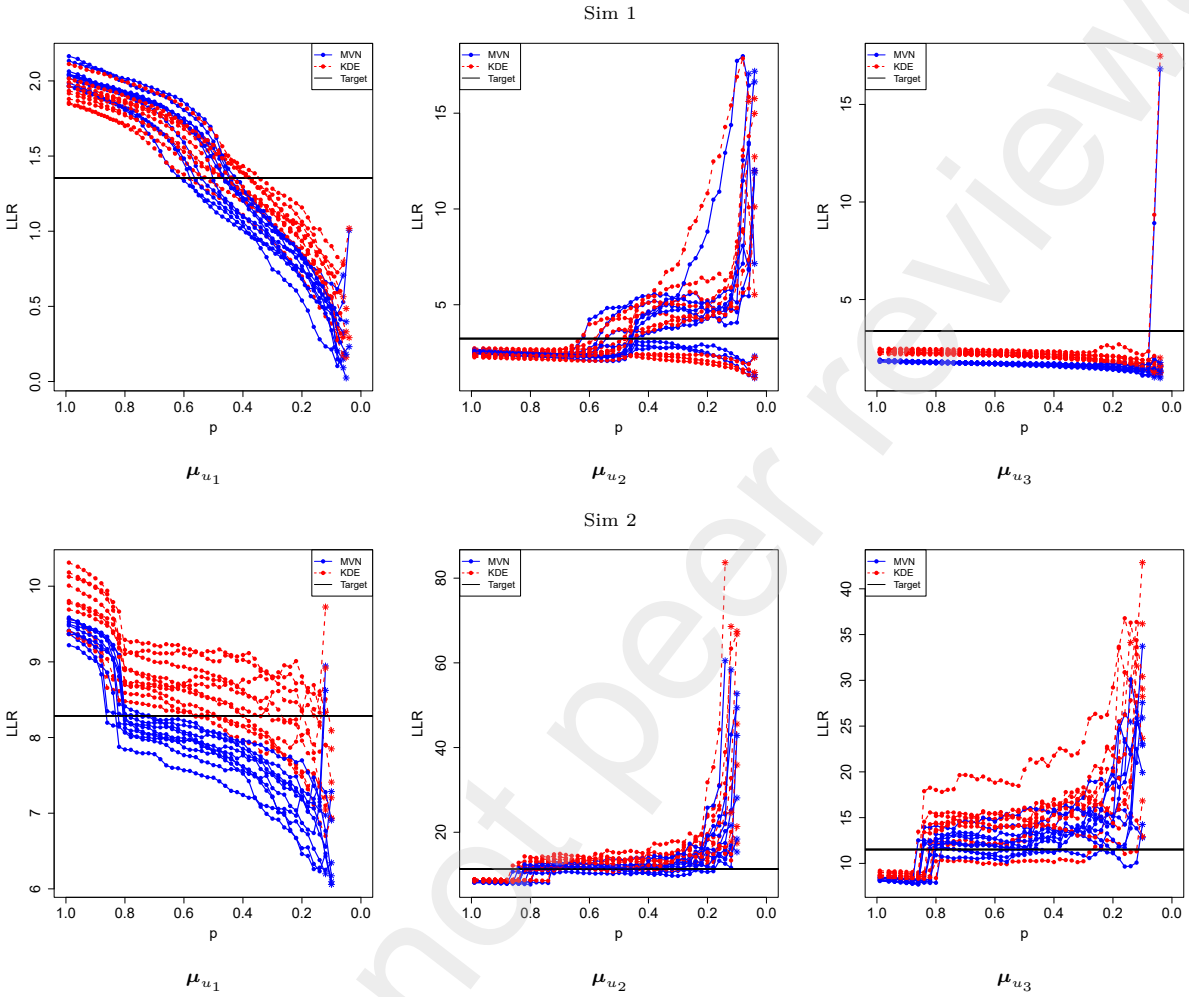

Figure 4: Prescreening simulation results under the prosecution cases when there is a subpopulation structure in the alternative source population. The KDE-based LLR in (red dashed lines), the MVN-based LLR (solid blue lines), and the true LLR in the subpopulation that contains e_u (solid black line) are presented versus p the proportion of background objects remaining. The rows represent Sim 1 and 2 and the columns are for the different locations of e_u and e_s.

For Sim 3, there are not any results displayed when e_a was generated under M_1, that is when there is a homogeneous background population. This is due to the covariance of the subpopulation of interest between-source component of the Gaussian mixture being small compared to the within-source covariance. This caused the estimate for the covariance used in the *comparison* R package [30] to have a negative determinant when we only have sources from the subpopulation of interest. For this reason, we were not comfortable using the LR values outputted for these cases. An illustrative example of this scenario can be found in Appendix B.

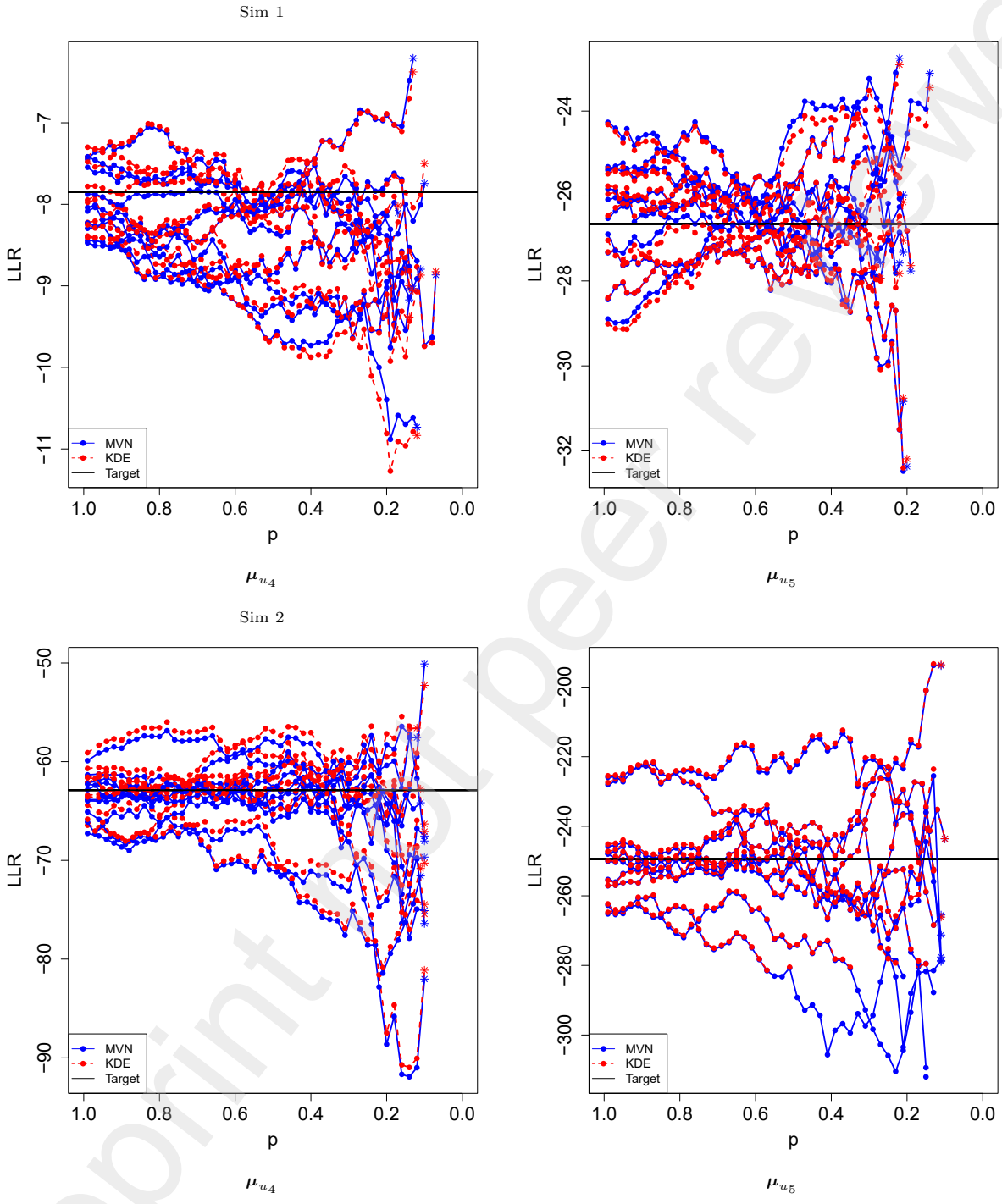

Figure 5: Prescreening simulation results under the defense cases when there is not a subpopulation structure in the alternative source population. The KDE-based LLR in (red dashed lines), the MVN-based LLR (solid blue lines), and the true LLR in the subpopulation that contains e_u (solid black line) are presented versus p the proportion of background objects remaining. The rows represent Sim 1 and 2 and the columns are for the different locations of e_u and e_s.

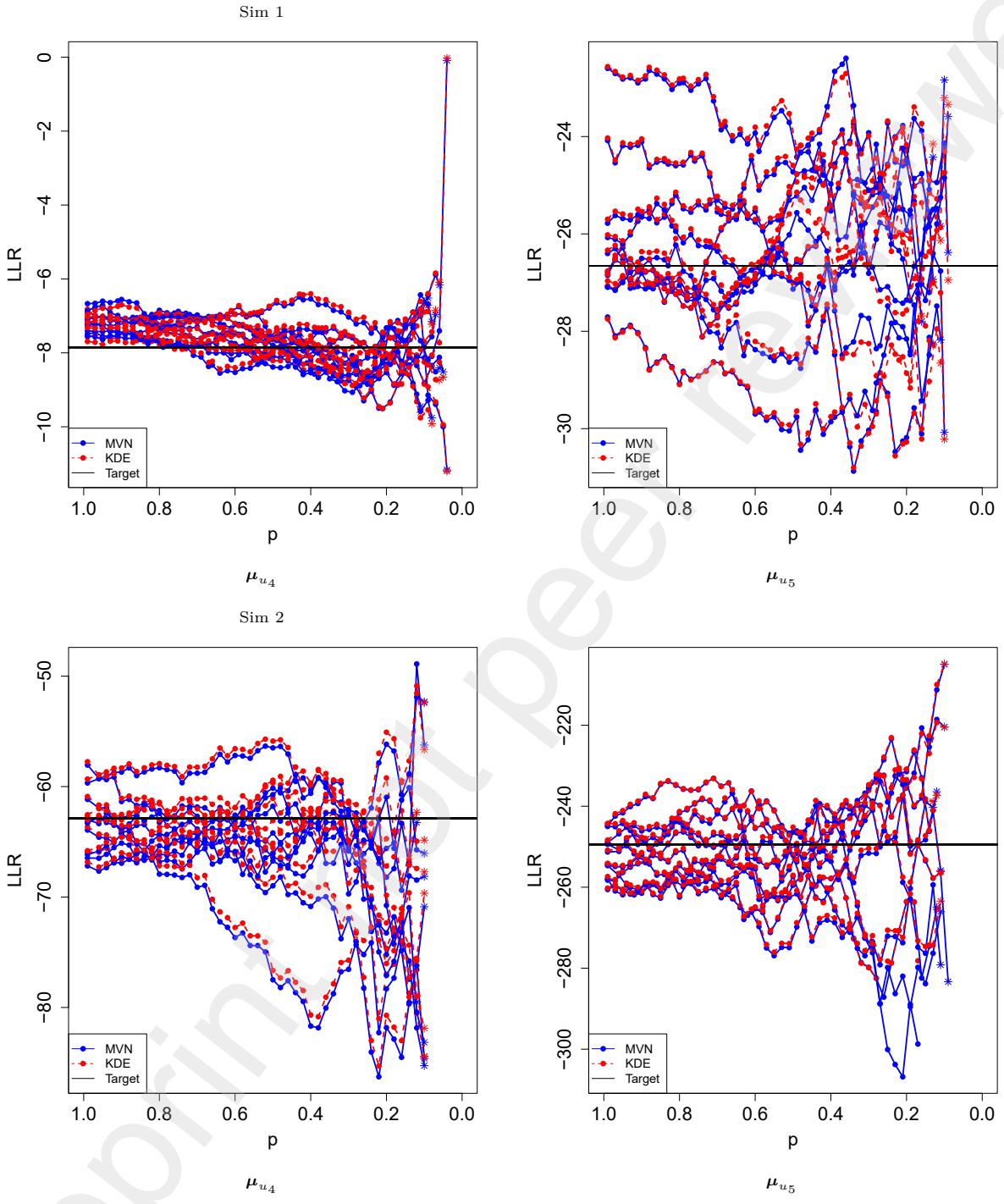

Figure 6: Prescreening simulation results under the defense cases when there is a subpopulation structure in the alternative source population. The KDE-based LLR in (red dashed lines), the MVN-based LLR (solid blue lines), and the true LLR in the subpopulation that contains e_u (solid black line) are presented versus p the proportion of background objects remaining. The rows represent Sim 1 and 2 and the columns are for the different locations of e_u and e_s.

For the defense cases when $\boldsymbol{\mu}_u \neq \boldsymbol{\mu}_s$, we will consider the case when $\boldsymbol{\mu}_s = \boldsymbol{\mu}_g$ and when $\boldsymbol{\mu}_u$ lies on the line between $\boldsymbol{\mu}_g$ and $\boldsymbol{\mu}_{g'}$. Two points will be considered for $\boldsymbol{\mu}_u$, which we will refer to as $\boldsymbol{\mu}_{u_4}$ and $\boldsymbol{\mu}_{u_5}$ where the latter is closer to $\boldsymbol{\mu}_{g'}$ than the former.

Figure 5 shows the results of the prescreening simulation under the defense cases when the alternative source population is homogeneous, that is, when e_s and e_u are generated from different sources and e_a is generated under M_1. Again rows indicate the model used to generate background samples (Sim 1 and Sim 2), and columns indicate how far the e_u is from e_s, that is when $\boldsymbol{\mu}_u$ equals $\boldsymbol{\mu}_{u_4}$ and $\boldsymbol{\mu}_{u_5}$. We see in these defense cases that there does not appear to be a clear trend in the value of the LLR as a results of prescreening as in the prosecution cases. There is again, however, the familiar unpredictable behavior of the LLR when there are few sources remaining in the relevant source population.

Figure 6 shows the results of the prescreening simulations under the defense cases when there is a subpopulation structure in the alternative source population. The rows and columns have the same meaning as in Figure 5. We again see that there is not a noticeable trend in the value of the LLR as a result of prescreening except a slight downward trend in the first subplot of the Sim 1 results. We again see that the value of the LLR is very unstable when there are only a few sources remaining the alternative source population.

4. Real data analysis

Three datasets are considered. The first two will illustrate the effects of prescreening on trace element data using LR methods and the third will illustrate the effects on fingerprint data using SLRs. All three examples have the prosecution model as ground truth, where the trace and control observations come from the same source.

4.1. Glass data sets

The first data set used to illustrate examples is a three-dimensional glass data set collected by JoAnn Buscaglia and distributed with Aiken and Lucy [5]. The data set has 62 windows with 3 labeled subpopulations. Each window has measurements on five fragments. The scatterplot of the data is shown in Figure 7 with colors denoting the three subpopulations in the data. We will focus on three windows for the examples. These are windows 46, 48, 58. The parameter estimates for the hierarchical models for this data set are given in Appendix C.

The second glass data set used is a 7-dimensional trace element glass data collected by Grzegorz Zadora and distributed with Aitken et al. [31]. The data set is also available through the *comparison* R package [30]. There are 200 windows with four fragments from each window and three technical replicates on each fragment. The mean was taken of the technical replicates to get an observation for each fragment. We will focus again on three windows from this data set, namely s127, s144, s188.

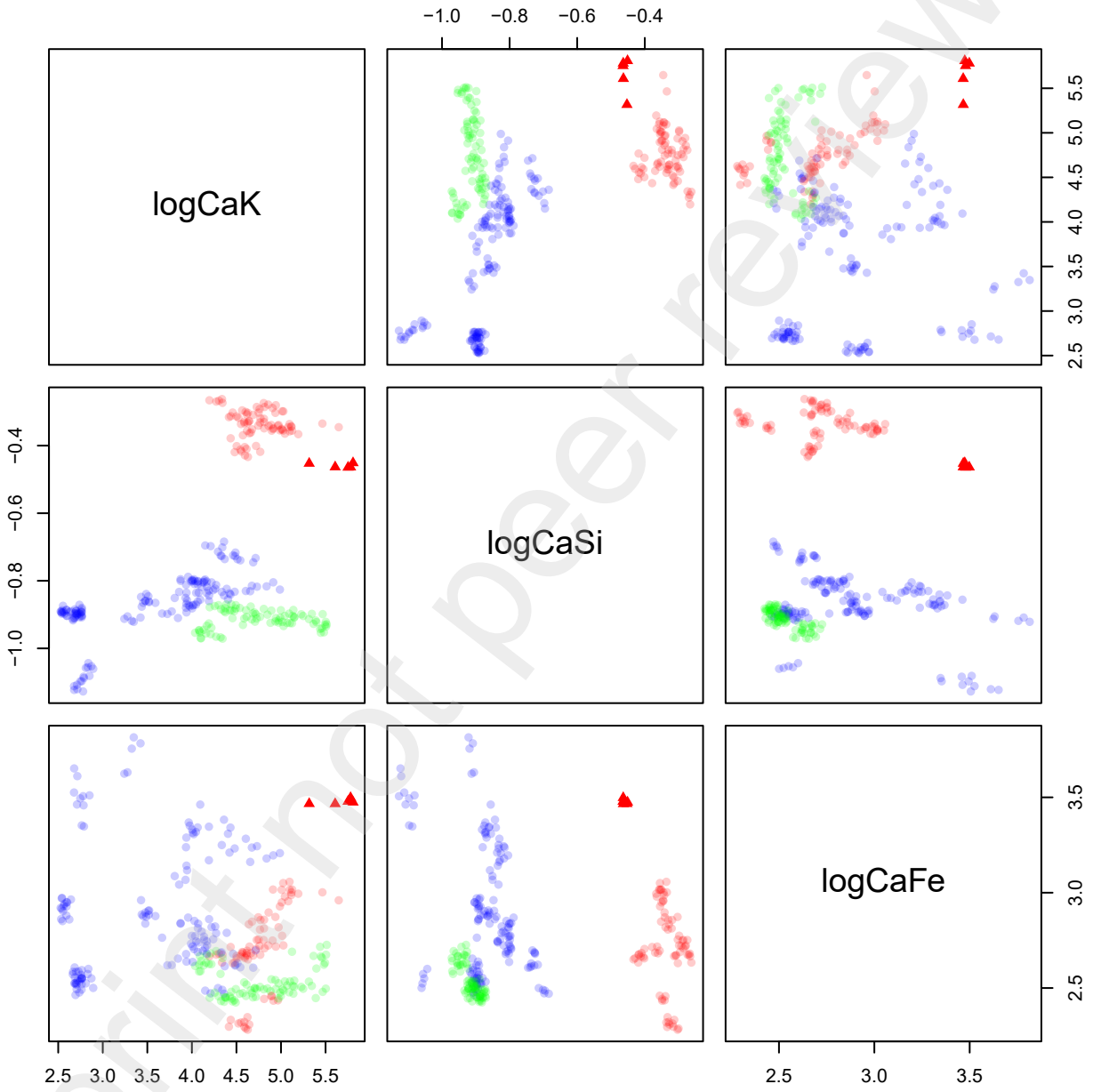

Figure 7: Scatter plot matrix visualizing the three-dimensional glass data. The colors denote the three subpopulations in the data set. The points corresponding to window 58 are displayed as triangles.

Figure 8 shows the results of prescreening on the two glass dataset examples. The rows denote the data set used, and the columns denote the object within the data set used as the trace and control. Row one shows the results of prescreening on these windows using the first two fragments within the window as the trace evidence, e_u, and the last three fragments as the control evidence, e_s for each of these windows. We see that windows 46 and 48, have a decrease in the value of the LR as we increase the prescreening level (decrease the number of sources in the background sample). There is also a sharp decrease around the value of 0.25, which agrees with the fact that the proportion of windows in the same subpopulation is approximately 0.258. Window 58 has an increase in the value of the LR as the prescreening level increases. Window 58 is "far away" from the other windows from the same subpopulation. This can be seen in Figure 7, where the red triangles show the points corresponding to window 58. This echos the results from the simulation, where when a same source pair is "rare" for its subpopulation, there can be an increase in the LR as a consequence of prescreening.

For the second glass data, the first two fragments for each window are used as the trace evidence, e_u and the last two fragments are used as the control evidence, e_s for that window. Figure 8-row 2 shows the results using this data set. We see the familiar trend in windows s127 and s188 where there is a decrease as we increase the prescreening level with a sharp decrease seen. Window s144 has an increase in the value of the LR as we prescreen, also with a sharp increase seen in the MVN-based LR when around 20-30% of the candidate background objects remain in the evidence about the relevant source population. It is interesting that there is also a sharp increase seen in the window at around a proportion of 0.9 to 0.8. This sharp increase is seen in both the MVN-based and KDE-based LRs, and the jumps is larger in the KDE-based LR. After this first sharp change the KDE-based LR levels off to around the same LR value as the MVN-based LR after both jumps. In this data set we do not have any labeled subpopulations, but after these changes suggest that we may be providing a value of evidence relative to an unknown subpopulation in the alternative source population.

4.2. Fingerprint scores

The third data set we use is a data set of scores on fingerprints [24]. We have all pairwise comparisons of full rolled and slap fingerprints. We focus on the scores that compare a full rolled fingerprint to a slap. We also have access to the pattern-level class of the fingerprint namely arch, left loop, right loop, tented arch, and whorl. In this data, we defined subpopulations based on the pattern-level class, which are all represented equally in the dataset. The prescreening procedure was implemented according to Algorithm 3.

Figure 9 shows the results of prescreening on three different sources namely, sources 797, 1013, and 1825. The value of the SLR decreased as we prescreen starting with the entire set of scores in the background population (blue solid lines). In this case, we see a steep decrease at some levels of prescreening. In the data set, there was an even split of each pattern-level fingerprints (*e.g.,* left or right loop). The plots

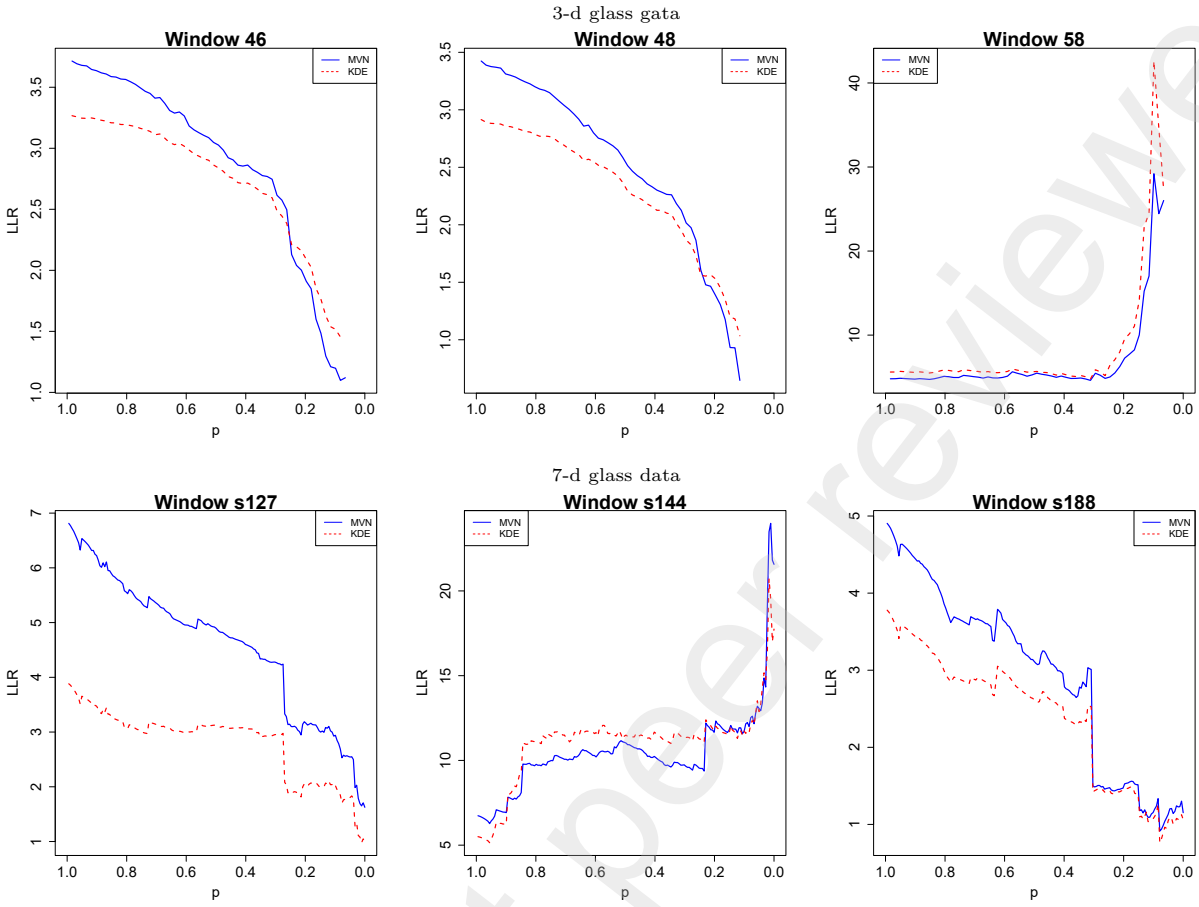

Figure 8: Prescreening results for select windows in two glass datasets under the prosecution cases. The KDE-based LLR in (red dashed lines) and the MVN-based LLR (solid blue lines) versus p the proportion of sources remaining in the relevant population are presented. The rows indicate the data set used where the first row is the 3-dimensional glass data and the second row is the 7-dimensional glass data.

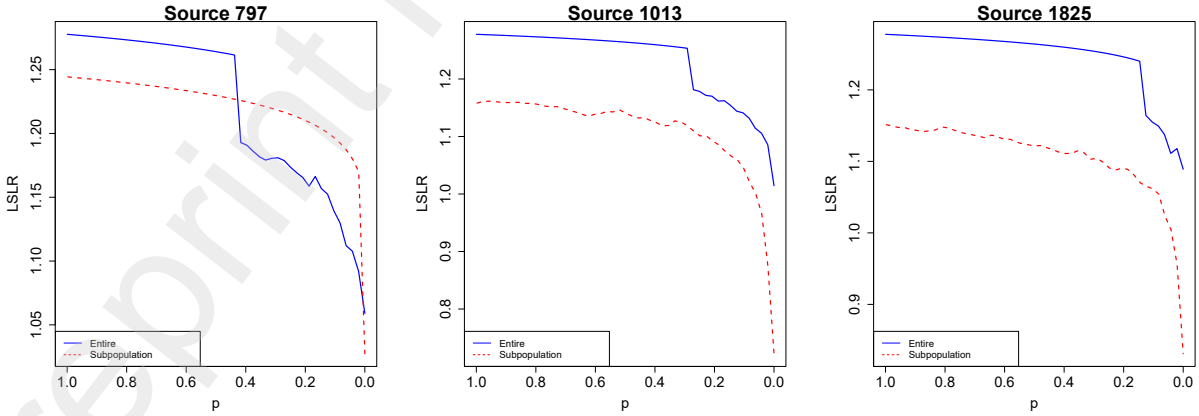

Figure 9: Prescreening results for select fingerprints in the fingerprint score data under the prosecution cases. The score-based SLR with the entire population (blue solid lines) and with pattern level subpopulation (red dashed lines) versus p the proportion of sources remaining in the relevant population are presented.

suggest that these sharp decreases may occur as pattern-level classes are removed from the alternative source populations, as the sharp changes occur around 0.4 for sources 797 and 1013, and around 0.2 for source 1825. When we only consider scores comparing objects within the same subpopulation (red dashed line), or those fingerprint of the same pattern-level class, the SLR value as a result of prescreening gradually decreases until a point of steep decrease.

5. Conclusion

This work demonstrates the empirical effect of population subsetting for some of the common methods for assigning values of evidence. The results of the simulation experiments show an interesting behavior of prescreening on multivariate normal and Kernel-based likelihood ratios. The simulation results suggest that prescreening the background samples when the control and trace are from the same source and common for their subpopulation causes a decrease in the LR. If there is a subpopulation structure in the alternative source population, prescreening the background samples can help give the value of the LR relative to the subpopulation of interest. This value of the LR seems to be indicated after a sharp decrease in the value of the LR as the prescreening level is increased. When there is no subpopulation structure in the alternative source population, prescreening dampens the value of the LR when the control and trace are common of their subpopulation. In either case, we must note that the propositions of interest will change with the prescreening levels, as the set of possible alternative sources at each level will be a subset of that which was originally considered; where this subset of possible alternative sources is selected to be the most similar to the evidence with an unknown source in some sense. Depending on the approach to defining "similar", this screening generally has the effect of increasing the likelihood of the evidence with an unknown source for each of the individual sources that make up the new background population.

When the control and trace are from the same source but rare for the subpopulation, the value of the likelihood ratio can increase, though we also saw cases where the LR decreased. Prescreening when the control and latent measurements come from different sources, yields no clear trend. When we prescreen too many sources, the value of the plug-in estimate of the LR becomes unstable, which is presumably due to a decrease in the quality of the parameter estimates.

Real glass data (two data sets) examples show similar results to the simulations. We see similar results with a decrease in the value of the likelihood ratio as we increase the prescreening level in certain cases, as well as cases with the increase. This behavior is also seen in a fingerprint score data set with a SLR.

These results suggest that prescreening the background sample to obtain a sample from the relevant source population for use with likelihood ratio can be useful, as it can provide a value of evidence relative to the subpopulation to which the trace evidence belongs. However, if we prescreen too far the value of the LR becomes unstable. Thus prescreening can be useful in giving a value of evidence relative to a given subpopulation that we do not know exists, but prescreening too far can also provide an unreliable

value of evidence. Though it may protect against overstating the value of evidence relative to the value of evidence in the latent subpopulation, prescreening fundamentally answers a different question than the non-prescreened LR. Thus, if prescreening is implemented prior to the evaluation of an LR, the prescreening methodology should be transparently presented along side the value of the LR. Our recommended wording is something similar to:

> "After prescreening $x\%$ of the candidate background objects to obtain a sample of the relevant source population according to the method outlined in Algorithm Y the plug-in estimate of the LR is z."

Appendix A. Parameters for simulation

Three models were used to generate synthetic data sets to study prescreening under simulation. The first model was a generated bivariate model, the second and third models were estimated from copper data and glass data, respectively.

Appendix A.1. Sim 1 parameters

Sim 1 utilized a bivariate model. The parameters for this model were generated using the *MixSim* R package [28]. The mixing proportions for the source level Gaussian mixture are equal $\boldsymbol{\pi} = (0.5, 0.5)'$. The means of the source level components are $\boldsymbol{\mu}_1 = (0.3891, 0.5831)'$, and $\boldsymbol{\mu}_2 = (0.0947, 0.8526)'$. The covariance matrices of the source level components are the same

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.0143 & 0.0062 \\ 0.0062 & 0.0069 \end{bmatrix},$$

and the within source covariance matrix is equal to the between source covariance matrix by a factor of 100, $\boldsymbol{\Sigma}_\epsilon = \frac{1}{100}\boldsymbol{\Sigma}_1$.

Appendix A.2. Sim 2 parameters

Sim 2 utilized a model suggested by copper wire data. For completeness we are including the estimates of the parameters for this model which is presented in Dettman et al. in the supporting material [26]. The mixing proportions for the source level model are $\boldsymbol{\pi} = (0.172, 0.828)'$, with means of the components of the source level Gaussian mixture given by

$$\boldsymbol{\mu}_1 = (7.429, 0.761, 0.774, 0.070, 0.030, 2.033, 0.917, 0.756)',$$

and

$$\boldsymbol{\mu}_2 = (4.524, 0.465, 0.687, 0.037, 0.28, 1.065, 0.512, 0.536)'.$$

The covariance matrices of the source level components are equal,

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.58300 & 0.043300 & -0.041700 & 3.61e-03 & -1.15e-03 & 0.083000 & 0.034700 & 0.048900 \\ 0.04330 & 0.005540 & -0.010100 & 4.17e-04 & -2.52e-04 & 0.008440 & 0.004040 & 0.004340 \\ -0.04170 & -0.010100 & 0.153000 & -8.82e-04 & 2.09e-03 & -0.014900 & -0.008780 & -0.016500 \\ 0.00361 & 0.000417 & -0.000882 & 7.80e-05 & -1.66e-05 & 0.000784 & 0.000309 & 0.000253 \\ -0.00115 & -0.000252 & 0.002090 & -1.66e-05 & 7.04e-05 & -0.000466 & -0.000285 & -0.000243 \\ 0.08300 & 0.008440 & -0.014900 & 7.84e-04 & -4.66e-04 & 0.055700 & 0.014700 & 0.002910 \\ 0.03470 & 0.004040 & -0.008780 & 3.09e-04 & -2.85e-04 & 0.014700 & 0.009250 & 0.001530 \\ 0.04890 & 0.004340 & -0.016500 & 2.53e-04 & -2.43e-04 & 0.002910 & 0.001530 & 0.019300 \end{bmatrix}.$$

The within-source covariance matrix is given as

$$\boldsymbol{\Sigma}_w = \begin{bmatrix} 4.33e-03 & 4.05e-04 & 3.89e-04 & 1.66e-05 & -2.45e-05 & 9.05e-04 & 6.79e-04 & 4.28e-04 \\ 4.05e-04 & 3.95e-04 & 1.34e-05 & -1.18e-06 & 2.39e-06 & 3.76e-04 & 2.36e-04 & 1.57e-04 \\ 3.89e-04 & 1.34e-05 & 1.71e-04 & 9.30e-06 & 4.89e-06 & 7.65e-05 & 4.46e-05 & -1.37e-04 \\ 1.66e-05 & -1.18e-06 & 9.30e-06 & 1.67e-06 & 7.29e-07 & 6.64e-06 & -5.34e-06 & -1.08e-05 \\ -2.45e-05 & 2.39e-06 & 4.89e-06 & 7.29e-07 & 1.69e-06 & 1.52e-05 & -8.06e-06 & -1.24e-05 \\ 9.05e-04 & 3.76e-04 & 7.65e-05 & 6.64e-06 & 1.52e-05 & 3.47e-03 & 1.49e-04 & 2.20e-04 \\ 6.79e-04 & 2.36e-04 & 4.46e-05 & -5.34e-06 & -8.06e-06 & 1.49e-04 & 3.61e-04 & 1.81e-04 \\ 4.28e-04 & 1.57e-04 & -1.37e-04 & -1.08e-05 & -1.24e-05 & 2.20e-04 & 1.81e-04 & 2.27e-03 \end{bmatrix}.$$

Appendix B. Covariance estimation

When working with the simulated glass data, an interesting scenario for the estimation of the covariance matrix was found. In particular when the between source covariance matrix, $\boldsymbol{\Sigma}_a$, was around the same "size" as or smaller than the within source covariance matrix, $\boldsymbol{\Sigma}_\epsilon$. The method used in the *comparison* R package gave an estimate of $\boldsymbol{\Sigma}_a$ that had a small but negative eigenvalue. Thus the resulting estimate of the between source covariance matrix does not lie within the parameter space of covariance matrices. We will walk through a simple bivariate example showing a case when this estimation can arise.

A sample of $n_a = 3$ sources with $n_w = 10$ observations from each source was generating according to M_1 with parameters $\boldsymbol{\mu} = (0, 0)'$,

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix},$$

and $\boldsymbol{\Sigma}_\epsilon = \frac{2}{3}\boldsymbol{\Sigma}_a$. Running this data through the two.level.components() function in the comparison package yields an estimate of the between source covariance matrix of

$$\hat{\boldsymbol{\Sigma}}_a = \begin{bmatrix} 2.493 & 3.588 \\ 3.588 & 5.045 \end{bmatrix},$$

which has eigenvalue of 7.577 and -0.0388. Thus this resulting matrix is not positive definite and is not a covariance matrix. When using e_a to calculate the KDE LR, an error stating "negative determinant - taking absolute value" is displayed, but still returns a value of the LR. In cases where this appeared, we were not comfortable reporting the value of the LR.

Appendix C. Glass data summaries

The parameter estimates for the for the 7-dimensional trace element glass data collected by Grzegorz Zadora and distributed with Aitken et al. [31] needed for the normal-based likelihood ratio are listed here. The data set is also available through the *comparison* R package [30]. After taking the mean of the technical replicates to get a measurement for each fragment, the overall mean estimate is

$$\hat{\boldsymbol{\mu}} = (-0.7101, -1.8473, -2.3097, -0.1539, -3.1066, -1.1383, -4.7975)'.$$

The estimate of the covariance matrix of the between source distribution is

$$\hat{\boldsymbol{\Sigma}}_a = \begin{bmatrix} 0.00355 & 0.03543 & -0.00476 & 0.00024 & -0.03345 & 0.03870 & 0.01106 \\ 0.03543 & 1.52429 & -0.22578 & -0.00598 & -0.75369 & 0.63492 & 0.37796 \\ -0.00476 & -0.22578 & 0.90573 & -0.00282 & 0.69872 & -0.15166 & -0.34354 \\ 0.00024 & -0.00598 & -0.00282 & 0.00136 & 0.00232 & -0.00237 & 0.00761 \\ -0.03345 & -0.75369 & 0.69872 & 0.00232 & 2.23979 & -0.50312 & -0.27782 \\ 0.03870 & 0.63492 & -0.15166 & -0.00237 & -0.50312 & 0.85480 & 0.16917 \\ 0.01106 & 0.37796 & -0.34354 & 0.00761 & -0.27782 & 0.16917 & 1.86819 \end{bmatrix}.$$

The estimate for the covariance matrix of the within source distribution is

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \begin{bmatrix} 0.00018 & 0.00017 & -0.00016 & 0.00001 & 0.00038 & 0.00020 & 0.00017 \\ 0.00017 & 0.04394 & -0.00063 & 0.00054 & -0.00057 & -0.00039 & 0.00221 \\ -0.00016 & -0.00063 & 0.02934 & 0.00029 & 0.00232 & 0.00030 & -0.00010 \\ 0.00001 & 0.00054 & 0.00029 & 0.00105 & 0.00282 & 0.00168 & 0.00059 \\ 0.00038 & -0.00057 & 0.00232 & 0.00282 & 0.26230 & 0.00378 & -0.03141 \\ 0.00020 & -0.00039 & 0.00030 & 0.00168 & 0.00378 & 0.01285 & 0.00295 \\ 0.00017 & 0.00221 & -0.00010 & 0.00059 & -0.03141 & 0.00295 & 0.09435 \end{bmatrix}.$$

The estimated parameter estimates for the three-dimensional glass data set collected by JoAnn Buscaglia and distributed with Aiken and Lucy [5] are listed as follows. The data set has 62 windows with 3 labeled subpopulations. The estimate of the mixing proportions for the three subpopulations are

$$\hat{\pi} = (0.258, 0.258, 0.484).$$

The overall mean estimate is

$$\hat{\boldsymbol{\mu}} = (4.1983, -0.7450, 2.7684)'.$$

The mean estimate for the first group is

$$\hat{\boldsymbol{\mu}}_1 = (4.8134, -0.3441, 2.7540)'.$$

The mean estimate for the second group is

$$\hat{\boldsymbol{\mu}}_2 = (4.7594, -0.9146, 2.5343)'.$$

Finally, the mean estimate for the third group is

$$\hat{\boldsymbol{\mu}}_3 = (3.5709, -0.8685, 2.9011)'.$$

The estimate of the covariance matrix of the within-source and between-source distributions are

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \begin{bmatrix} 1.679e-02 & 2.6619e-05 & 2.209e-04 \\ 2.661e-05 & 6.530e-05 & 7.399e-06 \\ 2.209e-04 & 7.399e-06 & 1.332e-03 \end{bmatrix} \text{ and } \hat{\boldsymbol{\Sigma}}_a = \begin{bmatrix} 0.7059 & 0.0988 & -0.0463 \\ 0.0988 & 0.0621 & -0.0070 \\ -0.0463 & -0.0070 & 0.1009 \end{bmatrix}.$$

The estimate of the between-source covariance matrix of the three subpopulations are

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 0.1060 & -0.0083 & 0.0762 \\ -0.0083 & 0.0026 & -0.0069 \\ 0.0762 & -0.0069 & 0.0825 \end{bmatrix}, \hat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 0.1771 & 0.0007 & 0.0019 \\ 0.0007 & 0.0007 & -0.0019 \\ 0.0019 & -0.0019 & 0.0065 \end{bmatrix},$$

$$\text{and } \hat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} 0.5490 & 0.0480 & 0.0257 \\ 0.0480 & 0.0085 & -0.0119 \\ 0.0257 & -0.0119 & 0.1178 \end{bmatrix}.$$

References

[1] C. Aitken, P. Roberts, G. Jackson, Practitioner guide no 1 : Fundamentals of probability and statistical evidence in criminal proceedings guidance for judges, lawyers, forensic scientists and expert witnesses, Prepared under the auspices of the Royal Statistical Society's Working Group on Statistics and the Law, London, United Kingdom (2010).

[2] European Network of Forensic Science Institutes, Stengthening the evaluation of forensic results across europe (STEOFRAE), ENFSI guideline for evaluative reporting in forensic science, approved version 3.0, European Network of Forensic Science Institutes, Wiesbaden, Germany (2015).

[3] D. M. Ommen, C. P. Saunders, Building a unified statistical framework for the forensic identification of source problems, Law, Probability and Risk 17 (2) (2018) 179–197. doi:10.1093/lpr/mgy008.

[4] D. V. Lindley, A problem in forensic science, Biometrika 64 (2) (1977) 207–213. doi:10.2307/2335686.

[5] C. G. G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Journal of the Royal Statistical Society. Series C (Applied Statistics) 53 (1) (2004) 109–122. doi:10.1046/j.0035-9254.2003.05271.x.

[6] J. M. Curran, The statistical interpretation of forensic glass evidence, International Statistical Review / Revue Internationale de Statistique 71 (3) (2003) 497–520.
URL http://www.jstor.org/stable/1403825

[7] S. Bozza, F. Taroni, R. Marquis, M. Schmittbuhl, Probabilistic evaluation of handwriting evidence: Likelihood ratio for authorship, Journal of the Royal Statistical Society. Series C (Applied Statistics) 57 (3) (2008) 329–341. doi:10.1111/j.1467-9876.2007.00616.x.

[8] A. B. Hepler, C. P. Saunders, L. J. Davis, J. Buscaglia, Score-based likelihood ratios for handwriting evidence, Forensic Science International 219 (1-3) (2012) 129–140. doi:10.1016/j.forsciint.2011.12.009.

[9] I. Evett, J. Lambert, J. Buckleton, A Bayesian approach to interpreting footwear marks in forensic casework, Science & Justice 38 (4) (1998) 241–247. doi:10.1016/S1355-0306(98)72118-5.

[10] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, A. Bromage-Griffiths, Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiæ, Journal of Forensic Sciences 52 (1) (2007) 54–64. doi:10.1111/j.1556-4029.2006.00327.x.

[11] A. Ruifrok, P. Vergeer, A. M. Rodrigues, From facial images of different quality to score based LR, Forensic Science International 332 (2022) 111201. doi:10.1016/j.forsciint.2022.111201.

[12] C. Champod, D. Meuwly, The inference of identity in forensic speaker recognition, Speech Communication 31 (2) (2000) 193–203. doi:10.1016/S0167-6393(99)00078-3.

[13] C. G. G. Aitken, D. Lucy, Corrigendum: Evaluation of Trace Evidence in the Form of Multivariate Data, Journal of the Royal Statistical Society Series C: Applied Statistics (2004) 665–666doi:10.1111/j.1467-9876.2004.02031.x.

[14] P. Vergeer, From specific-source feature-based to common-source score-based likelihood-ratio systems: ranking the stars, Law, Probability and Risk 22 (1) (2023) mgad005. doi:10.1093/lpr/mgad005.

[15] T. Hicks, A. Biedermann, J. de Koeijer, F. Taroni, C. Champod, I. Evett, The importance of distinguishing information from evidence/observations when formulating propositions, Science & Justice 55 (6) (2015) 520–525. doi:10.1016/j.scijus.2015.06.008.

[16] G. S. Morrison, E. Enzinger, C. Zhang, Refining the relevant population in forensic voice comparison – a response to hicks et alii (2015) the importance of distinguishing information from

evidence/observations when formulating propositions, Science & Justice 56 (6) (2016) 492–497. doi:10.1016/j.scijus.2016.07.002.

[17] D. de Zwart, J. van der Weerd, Extraction of the relevant population from a forensic database, Science & Justice 61 (4) (2021) 419–425. doi:10.1016/j.scijus.2021.03.008.

[18] National Institute of Standards and Technology, Friction ridge process map (current practice), https://www.nist.gov/system/files/documents/2019/12/10/Friction%20Ridge%20Process%20Map_December%202019.pdf, accessed July 10, 2023 (December 2019).

[19] National Institute of Standards and Technology, Process map of current practices in forensic speaker recognition, https://www.nist.gov/system/files/documents/2019/11/05/speaker_recognition_process_map_20190930.pdf, accessed July 10, 2023 (September 2019).

[20] Expert Working Group for Human Factors in Handwriting Examination, Forensic handwriting examination and human factors: Improving the practice through a systems approach, U.S. Department of Commerce, National Institute of Standards and Technology (May 2021). doi:10.6028/NIST.IR.8282r1.

[21] D. M. Ommen, C. P. Saunders, A Problem in Forensic Science Highlighting the Differences between the Bayes Factor and Likelihood Ratio, Statistical Science 36 (3) (2021) 344 – 359. doi:10.1214/20-STS805.

[22] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Statistical analysis in forensic science: evidential value of multivariate physicochemical data, John Wiley & Sons, 2014.

[23] A. O'Brien, A kernel based approach to determine atypicality, Ph.D. thesis, South Dakota State University (2017).
URL https://openprairie.sdstate.edu/etd/1711

[24] E. Tabassi, L. Tang, X. Zhu, Repeatability and reproducibility of forensic likelihood ratio methods when sample size ratio varies, The International Joint Conference on Biometrics (IJCB 2017), Denver, CO, 2018. doi:10.1109/BTAS.2017.8272737.

[25] R. Maitra, V. Melnykov, Simulating data to study performance of finite mixture modeling and clustering algorithms, Journal of Computational and Graphical Statistics 19 (2) (2010) 354–376. doi:10.1198/jcgs.2009.08054.

[26] J. R. Dettman, A. A. Cassabaum, C. P. Saunders, D. L. Snyder, J. Buscaglia, Forensic discrimination of copper wire using trace element concentrations, Analytical Chemistry 86 (16) (2014) 8176–8182. doi:10.1021/ac5013514.

[27] D. J. Murdoch, E. D. Chow, A graphical display of large correlation matrices, The American Statistician 50 (2) (1996) 178–180. doi:10.2307/2684435.

[28] V. Melnykov, W.-C. Chen, R. Maitra, Mixsim: An **R** package for simulating data to study performance of clustering algorithms, Journal of Statistical Software 51 (12) (2012) 1–25. doi:10.18637/jss.v051.i12.

[29] J. Curran, T. Hersh, Hotelling: Hotelling's T^2 Test and Variants, R package version 1.0-8 (2021). URL https://CRAN.R-project.org/package=Hotelling

[30] D. Lucy, J. Curran, A. Martyna, comparison: Multivariate Likelihood Ratio Calculation and Evaluation, R package version 1.0-5 (2020).
URL https://CRAN.R-project.org/package=comparison

[31] C. G. G. Aitken, G. Zadora, D. Lucy, A two-level model for evidence evaluation, Journal of Forensic Sciences 52 (2) (2007) 412–419. doi:10.1111/j.1556-4029.2006.00358.x.