JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2010, Vol. 55, No. 2
doi: 10.1111/j.1556-4029.2009.01316.x
Available online at: interscience.wiley.com

PAPER

GENERAL

Grzegorz Zadora,[1] *Ph.D.; Tereza Neocleous,*[2] *Ph.D.; and Colin Aitken,*[3] *Ph.D.*

A Two-Level Model for Evidence Evaluation in the Presence of Zeros*

ABSTRACT: Likelihood ratios (LRs) provide a natural way of computing the value of evidence under competing propositions. We propose LR models for classification and comparison that extend the ideas of Aitken, Zadora, and Lucy and Aitken and Lucy to include consideration of zeros. Instead of substituting zeros by a small value, we view the presence of zeros as informative and model it using Bernoulli distributions. The proposed models are used for evaluation of forensic glass (comparison and classification problem) and paint data (comparison problem). Two hundred and sixty-four glass samples were analyzed by scanning electron microscopy, coupled with an energy dispersive X-ray spectrometer method and 36 acrylic topcoat paint samples by pyrolysis gas chromatography hyphened with mass spectrometer method. The proposed LR model gave very satisfactory results for the glass comparison problem and for most of the classification tasks for glass. Results of comparison of paints were also highly satisfactory, with only 3.0% false positive answers and 2.8% false negative answers.

KEYWORDS: forensic science, evidence evaluation, likelihood ratio, missing data, multivariate data, graphical models

Various types of materials such as glass and paint fragments are routinely subjected to physico-chemical examination by forensic scientists. For example, glass and paint fragments could be obtained from hit-and-run accident debris comprising a mixture of pieces of sand, soil, dust, and other transfer evidence such as glass and paint. These fragments are called the *recovered sample*. If it is also possible, glass from the glass object broken, e.g., during the car accident, and/or paint fragments from the car paint surface destroyed during the accident are collected for analysis. Such samples are called the *control sample*.

One of the problems of analysis of materials found in debris for forensic purposes is their classification into use-type category (e.g., whether a glass object originates from the window category or the container category). In the case of transfer evidence such as glass this process is especially important in the absence of a control sample as it could help investigators (policemen, prosecutors) focus their search for appropriate control materials. Another common task is the comparison problem. This task relates to the question—could two samples (e.g., a paint fragment recovered from a smear found on the clothes of the victim of a hit-and-run accident and a paint fragment collected from the suspected car) have originated from the same object?

The size of recovered fragments of glass and paint is very small. Therefore, addressing the above-mentioned problems requires data obtained during physico-chemical analysis that are quantitative and semi-quantitative such as the concentration of elements in a glass fragment (1,2) or peak areas of compounds present in a chromatogram of pyrolysis gas chromatography (Py-GC) (3,4).

The importance of glass as evidence was recognized many years ago (5,6) as very small glass fragments (of linear dimension 0.1–0.5 mm) that arise during car accidents, burglaries, fights, etc. could be carried on the clothes, shoes, and hair of participants. The glass refractive index measurement method and scanning electron microscopy, coupled with an energy dispersive X-ray spectrometer (SEM–EDX), are routinely used in many forensic institutes for the investigation of glass and other forensic problems (1,7). Other methods of elemental analysis of glass fragments are μ-X-ray fluorescence (8) and laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) (9). However, these methods require relatively large fragments of glass, for example LA-ICP-MS gives good results with pieces of glass larger than 0.5 mm. SEM–EDX has the drawback that it can only provide information about major and minor elements, such as O, Na, Al, Mg, Si, K, Ca, Fe, from any glass fragment. Trace elements exist in concentrations below the detection limits of this method. It is commonly believed that trace element concentrations are essential to enable the glass investigator to compare and individualize glass evidence effectively. However, it has been shown that some headway can be made on the basis of the major and minor element concentrations (1,7). These data could be used for classification of recovered glass fragments and/or in the comparison problem if a control glass sample is available.

The comparison of car paints is an important problem in road accident investigations (5). Paint fragments are collected from the road or the victim's clothes and compared with paint that originated from the suspected car. It is necessary to obtain information on both the morphology and the chemical composition of the analyzed samples in order to solve the comparison problem mentioned

[1]Institute of Forensic Research, Westerplatte 9, PL-31-033 Krakow, Poland.
[2]Department of Statistics, University of Glasgow, 15 University Gardens, Glasgow G12 8QW, U.K.
[3]School of Mathematics and The Joseph Bell Centre for Forensic Statistics and Legal Reasoning, The King's Buildings, The University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, U.K.
*The research was financially supported by the State Committee for Scientific Research, Poland, within project no. 0T00C02829 and by U.K. Engineering and Physical Sciences Research Council grant *EP/C*532627/1.

Received 29 Oct. 2008; and in revised form 26 Jan. 2009; accepted 4 Mar. 2009.

earlier. However, if the paint samples belong to the same class, i.e., contain similar polymer binder and pigments, then further individualization requires the application of a more sensitive and discriminating analytical method, Py-GC hyphened with MS as a detector (Py-GC/MS) (3,4). Moreover, Py-GC/MS is able to find subtle structural or trace compositional variations within a sample that appears overwhelmingly similar even when compared with similar samples of the same category. The disadvantage of application of Py-GC/MS for forensic purposes is that samples are destroyed during the analysis. It should be kept in mind that the analyzed paint sample is the evidence, and it should be available for various analyses at any time.

Evaluation of evidence is based on analytical data obtained during the physico-chemical analysis. The comparison of control and recovered materials and the classification of the analyzed materials require careful attention to the following considerations:

- possible sources of uncertainty (sources of error), which will include, at least:

(a) variation of measurements of characteristics within the recovered and/or control items,

(b) variation of measurements of characteristics between various objects in the relevant population (e.g., glass object population);

- information about the rarity of the determined physico-chemical characteristics (e.g., elemental and/or chemical composition of compared samples) for recovered and/or control samples in the relevant population;
- the level of association between different characteristics when more than one characteristic has been measured; and
- in the case of the comparison problem, the similarity of the recovered material to the control sample.

The best way of obtaining the value of the evidence, especially in the case when the observed differences of the analyzed features are small and/or the number of observed features is larger than one, is using statistical methods, especially those related to the likelihood ratio (LR), which is a well-documented measure of evidence value in the forensic field (1,2,10):

$$\mathrm{LR} = \frac{\Pr(E|H_1, I)}{\Pr(E|H_2, I)} \quad (1)$$

where H_1 and H_2 are the considered propositions; I is the background information of the case not related to the evidence E. Values of LR above 1 support H_1 and values of LR below 1 support H_2. A value of LR close to 1 provides little support for either proposition. Also the larger (the lower) the value of the LR, the stronger (the weaker) the support of E for H_1 (H_2).

Consider a case where the fact finder (prosecutor, judge, etc.) asks a forensic scientist to evaluate evidence in the form of a recovered material, of unknown origin, and a control material, whose origin is known. The result of such a comparison will be referred to as E. The relevant propositions for the fact finder arise from the circumstances of the case, and often because of the adversarial nature of the system, but for evidence evaluation they typically are as follows:

- H_1: the control and recovered samples come from the same source (*prosecutor proposition*),
- H_2: the control and recovered samples come from different sources both belonging to a relevant population (*defense proposition*).

Also, there are several methods which could be used for the classification of forensic evidence (11). For example, the category for type of use of glass fragments described by the content of the main elements (E) can be determined by application of different classifiers such as Support Vector Machines and Naïve Bayes Classifiers (7). These methods allocate an observation on an object to one of the considered glass categories (class 1 or class 2) according to posterior probabilities $P(H_1|E)$ and $P(H_2|E)$ where H_1 is the hypothesis that the object came from class 1 and H_2 is the hypothesis that the object came from class 2. In forensic science, it is more common to evaluate evidence (E) in the context of two propositions (H_1 and H_2), i.e., to determine the conditional probabilities $P(E|H_1)$ and $P(E|H_2)$. This information can be expressed in terms of the LR mentioned earlier, with the relevant propositions in the form:

- H_1: the recovered samples come from category 1 (*prosecutor proposition*),
- H_2: the recovered samples come from category 2 (*defense proposition*).

Several models exist (e.g., [1,12]), which allow calculation of the LR for multivariate data, but these models assume that there are no missing data (i.e., no zeros in any of the considered variables). Based on the experience of the authors, it is almost impossible to have such multidimensional data, especially in the field of the physico-chemical analysis of microtraces like glass or paint where the presence or absence of a particular component is related to the nature of the analyzed objects. One novelty in the proposed model is that it allows for the presence of zero concentrations. The zeros are assumed to be structural. Of course, absolute zero does not exist in analytical chemistry as measurements depend on the detection level of the method used. Elements and components present in trace levels (below the detection limit) will always be undetectable for the specific method, and hence the measurements obtained can be considered as structural zeros for the purposes of the statistical analysis.

The performance of the model is discussed as applied to glass and paint data, for classification of glass fragments into one of two known categories, and second, for evidence evaluation by comparison of recovered and control samples of glass and paint. Related work with an application to food compositions using latent Gaussian models, i.e., ones in which the data are assumed to arise from a Euclidean projection of a multivariate Gaussian random model to the unit simplex, may be found in (13).

Physico-Chemical Data

Glass Database

Four glass fragments, with surfaces as smooth and flat as possible, collected from each of 264 glass objects (79 building windows [*w*], 86 car windows [*c*], 26 bulbs [*b*], 16 headlamps [*h*], and 57 containers [*p*]) were placed on self-adhesive carbon tabs on an aluminum stub and then carbon coated using an SCD sputter (Bal-Tech, Liechtenstein, Switzerland). Three replicate measurements on each fragment were made. Analysis of the elemental content of each glass fragment was carried out using a SEM (JSM-5800; Jeol, Tokyo, Japan), with an EDX detector (Link ISIS 300; Oxford Instruments Ltd., Witney, Oxfordshire, U.K.). The measurement conditions were accelerating voltages 20 kV, life time 50 sec, magnification 1000–2000×, and the calibration element was cobalt. The SEMQuant option (part of the software LINK ISIS; Oxford Instruments Ltd.) was used in the process of determining the percentage of particular elements in a

fragment. The selected analytical conditions allowed the determination of concentrations of oxygen (O), sodium (Na), magnesium (Mg), aluminum (Al), silicon (Si), potassium (K), calcium (Ca), and iron (Fe).

Each of the four glass fragments selected for analysis was measured three times and the mean of the three measurements was taken for each fragment in the comparison problem. In addition, the mean of the four fragments was taken to obtain just one value for each object in the classification problem. Descriptive statistics are presented in the form of boxplots (Fig. 1) depicting the median, upper and lower quartiles (box), outliers (dots), and smallest and largest nonoutlier values (fences). Values lying more than 1.5 interquartile ranges (IQRs) below the lower quartile or more than 1.5 IQRs above the upper quartile are considered to be outliers.

The data consist of seven variables obtained as the $\log_{10}$ of the ratio with respect to the oxygen (O) concentration, which is always nonzero for glass data. The seven variables thus obtained will be denoted as (Na′, Mg′, Al′, Si′, K′, Ca′, Fe′). When raw data take the value 0 the logratio is not available as $\log_{10}$ of zero cannot be obtained. The proportions of nonzero values for each variable are given in Table 1.

Paint Database

These data were originally published in Table 2 of (4) and obtained in the following way (4): a pyrolyser (CSD2000; Analytix, Peterlee, U.K.) coupled with GC/MS (TurboMass Gold; Perkin Elmer, Wellesley, MA) was used to analyze 36 samples, each with

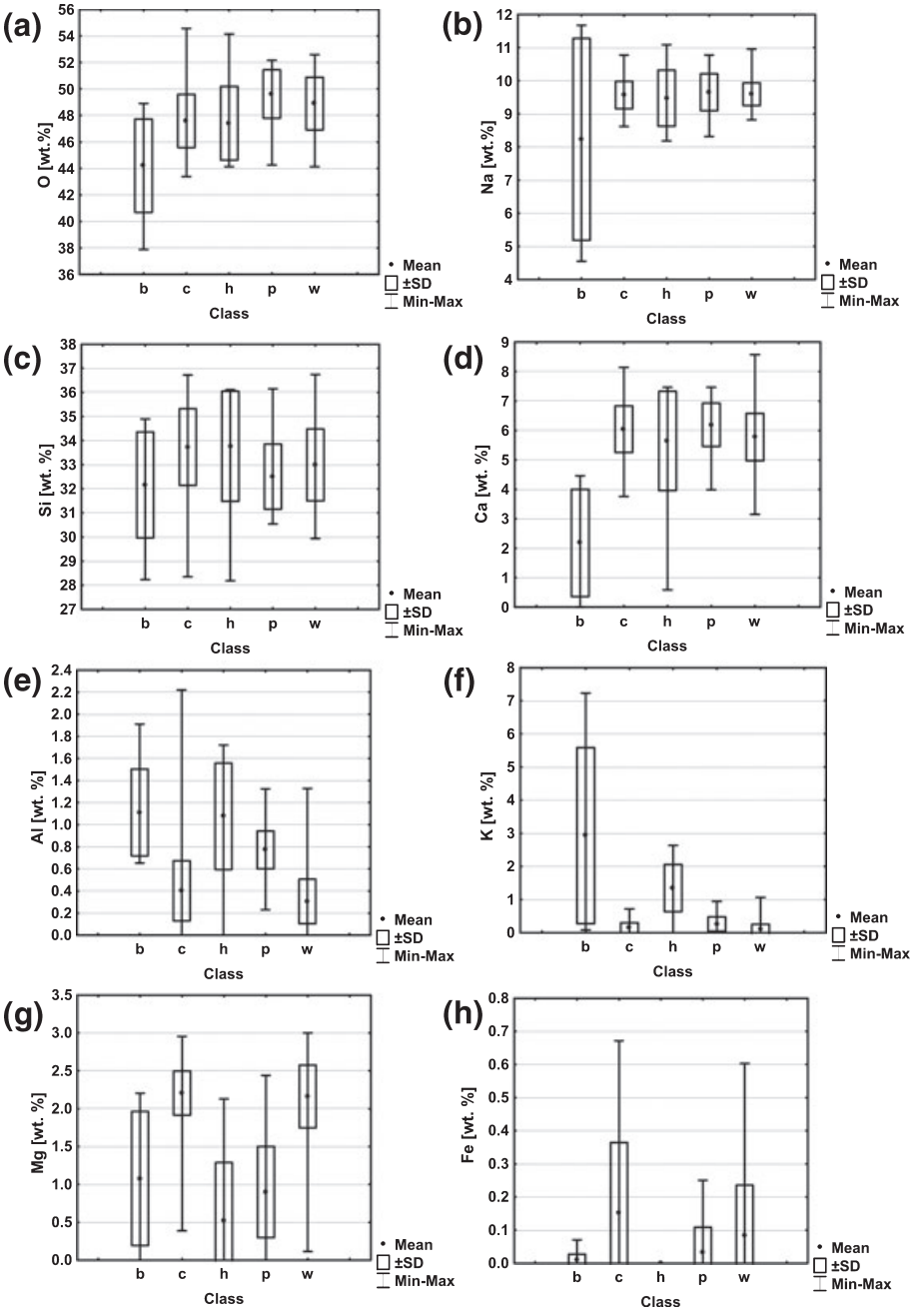

FIG. 1—*Univariate descriptive statistics of variables in each of the five glass categories (b, bulbs, c, car windows, h, headlamps, p, containers, w, building windows); (a) oxygen, (b) sodium, (c) silica, (d) calcium, (e), aluminum, (f) potassium, (g) magnesium, (h) iron.*

TABLE 1—*Number (out of 264) and proportion not zero for the glass data by category (b, c, h, p, w) and overall, where b represents bulbs, c represents car windows, h represents headlamps, p represents containers, and w represents building windows.*

Variable	O	Na	Mg	Al	Si	K	Ca	Fe
Proportion *b*	1.000	1.000	0.923	1.000	1.000	1.000	0.769	0.346
Proportion *c*	1.000	1.000	1.000	0.975	1.000	0.747	1.000	0.456
Proportion *h*	1.000	1.000	0.812	0.938	1.000	0.875	1.000	0.000
Proportion *p*	1.000	1.000	0.982	1.000	1.000	0.807	1.000	0.211
Proportion *w*	1.000	1.000	1.000	0.907	1.000	0.744	1.000	0.256
Overall proportion	1.000	1.000	0.977	0.958	1.000	0.792	0.977	0.299
Overall number	264	264	258	253	264	209	258	73

TABLE 2—*Estimated probability vectors for the glass comparison problem. For the numerator of the likelihood ratio, these are obtained from same-object pairs, and for the denominator from different-object pairs. Thus, $p(u_1 = 0, u_2 = 0)$ is an estimate of the probability that both variables will be zero, [$p(u_1 = 0, u_2 = 1)$ or $p(u_1 = 1, u_2 = 0)$] is an estimate of the probability that one variable will be zero and the other nonzero, $p(u_1 = 1, u_2 = 1)$ is an estimate of the probability that both variables will be nonzero.*

Variable	O	Na	Mg	Al	Si	K	Ca	Fe
Numerator								
$p(u_1 = 0, u_2 = 0)$	0.000	0.000	0.060	0.056	0.000	0.199	0.036	0.718
$p(u_1 = 0, u_2 = 1$ or $u_1 = 1, u_2 = 0)$	0.000	0.000	0.037	0.011	0.000	0.084	0.008	0.043
$p(u_1 = 1, u_2 = 1)$	1.000	1.000	0.902	0.933	1.000	0.717	0.956	0.238
Denominator								
$p(u_1 = 0, u_2 = 0)$	0.000	0.000	0.006	0.003	0.000	0.057	0.001	0.547
$p(u_1 = 0, u_2 = 1$ or $u_1 = 1, u_2 = 0)$	0.000	0.000	0.146	0.116	0.000	0.368	0.077	0.387
$p(u_1 = 1, u_2 = 1)$	1.000	1.000	0.848	0.881	1.000	0.575	0.921	0.067

three replications, of acrylic topcoat paints—called *clear coat*—that were indistinguishable in terms of their infrared spectra and elemental composition. The clear coat was isolated from paints under an optical microscope (40× magnification; SMXX Carl Zeiss, Jena, Germany) using a scalpel, and this was placed in a quartz tube (Analytix Ltd.) in the platinum coil of a pyrolyser. The sample size was 50–100 μg. The GC program was 40°C held for 2 min; ramped 10°C/min to 300°C, 300°C held for 2 min; increase 30°C/min to 320°C; 320°C held for 3 min. A RTx-35MS (Restek, Bellefonte, PA) capillary column (30 m × 0.25 mm × 0.25 μm) was used. Py was performed at 750°C during 20 sec. Peak areas were collected by application of the software TurboMass 4.4.0.014; Perkin Elmer.

The $\log_{10}$ of the ratio of the peak areas for each of the following seven organic compounds compared to the peak area of styrene was calculated: methylmethacrylate (MMA – abbreviation used in this paper for $\log_{10}$ ratios); toluene (TOL); butylmethacrylate (BMA); α−methylstyrene (MST); 2-hydroxyethylmethacrylate (M2E); 2-hydroxypropylmethacrylate (M2P); and 1,6-diisocyanatehexane. These are the most frequently occurring compounds in the analyzed samples. The choice of variables was subjective as the number of samples in the database was relatively small. Peak values below 10^{-4} in (4) were practically zero peaks and their logratio was thus not available. The two-level LR model published in (7) was used in the calculations of (4). This LR model did not allow for the presence of zeros. Instead, in (4), missing chromatographical data (logratios corresponding to zeros) were replaced by the relatively small value of 1000, when other values of peak areas had order of magnitude 10^7–10^8. Descriptive statistics are shown in Table 3 and Fig. 2.

TABLE 3—*Proportion and number (out of 36) not zero for the paint data.*

Variable	MMA	TOL	BMA	MST	M2E	M2P	I16
Proportion	0.694	0.694	0.778	0.806	0.500	0.667	0.833
Number	25	25	28	29	18	24	30

Statistical Methods

Likelihood Models in the Presence of Zeros

A common way to deal with data obtained from physico-chemical analysis is to take the logratio of a composition $\mathbf{x} = (x_1,\ldots,x_D)$, $x_D \neq 0$ given by

$$y_1 = \log\left(\frac{x_1}{x_D}\right), \ldots, y_{D-1} = \log\left(\frac{x_{D-1}}{x_D}\right)$$

and then apply the methodology for continuous data to $\mathbf{y} = (y_1, \ldots,y_{D-1})$. This effectively removes stochastic fluctuations in instrumental measurement (especially those that occur in the use of various chromatography techniques) and also allows the elimination of the problem of large differences in the level of magnitude of the various values of the variables. However, it is often the case that some of the $\{x_i;\ i \neq D\}$ are zero thus making it impossible to compute the logratio.

Methods for replacing the zeros by some small number have been proposed, e.g., (14), and these are useful when the proportion of zeros is small and the zeros are not structural. If the zeros are structural, this should be taken into account when building a model. One possible way of doing this is to look at the logratios for nonzero subcompositions (15). These can be assumed to have multivariate distributions with varying dimensions as each observation has up to $D-1$ variables for nonzero concentrations. The following independent binary model is assumed for the zeros. Let u_i be an indicator of whether the ith variable in the composition is nonzero, for $i = 1,\ldots,D$. For example, the eight-component composition $\mathbf{x} = (0.4,0.2,0.1,0.0,0.1,0.1,0.0,0.1)$ has corresponding $\mathbf{u}$-vector (1,1,1,0,1,1,0,1). The probability of the vector $\mathbf{u}$ is taken to be

$$b(\mathbf{u}|\mathbf{p}) = \prod_{i=1}^{D} p_i^{u_i}(1-p_i)^{(1-u_i)}$$

which is the product of D independent Bernoulli variables with different probabilities, p_i, of success. In the example above, $D = 8$ and the p_i; $i = 1,\ldots,D$ are estimated from background data as the

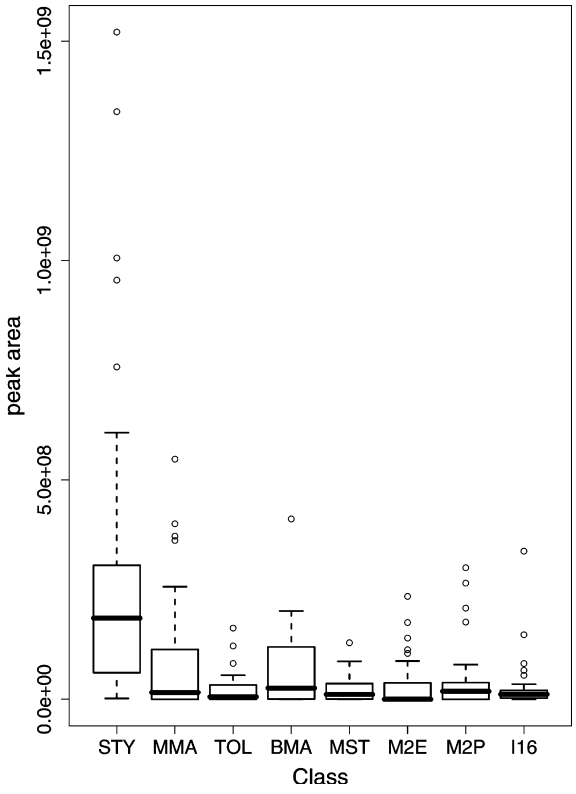

FIG. 2—*Boxplots of the univariate distributions for paint item means. STY, styrenel; MMA, methylmethacrylate; TOL, toluene; BMA, butylmethacrylate; MST, α-methylstyrene; M2E, 2-hydroxyethylmethacrylate; M2P, 2-hydroxypropylmethacrylate; I16, 1,6-diisocyanatehexane*.

proportion of observations that are not zero for each of the eight components separately. Suppose that the estimated **p**-vector is (0.3, 0.6, 0.9, 0.2, 0.7, 0.5, 0.1, 0.8). Thus, in this example, 30% of observations of component 1 are not zero, whereas 90% of observations of component 3 are not zero. Since the estimated value of p_1 is 0.3, the term for the first component for **u**-vector (1, 1, 1, 0, 1, 1, 0, 1) will be $0.3^1(1-0.3)^0 = 0.3$. Similarly, the term for the fourth component will be $0.2^0(1-0.2)^1 = 0.8$ and so on. All eight factors ($i = 1,\ldots,8$) can be obtained in this way and multiplication of them together will give $b(\mathbf{u}|\mathbf{p})$. For an observation without zeros, the incidence vector **u** would equal (1, 1, 1, 1, 1, 1, 1, 1) and the corresponding probability $b(\mathbf{u}|\mathbf{p})$ with p as given earlier would be 0.3×0.6×0.9×0.2×0.7×0.5×0.1×0.8.

Let the logratios Y be taken with respect to the last column of X, which is assumed to contain only nonzero data. For a logratio vector **y**, let Q be the matrix picking out only the elements of **y** coming from nonzero elements of **x**. The contribution of each observation to the likelihood will then be $b(\mathbf{u}|\mathbf{p})\ f(Q\mathbf{y})$, where $f(Q\mathbf{y})$ is the multivariate density for the nonzero subcomposition. This multivariate density could be assumed to be normal or, if necessary, obtained by kernel density estimation (KDE) (e.g., by application of a Gaussian kernel). More details on the LR calculation can be found in the Appendix.

The Classification Problem

Consider the problem in which an observation $\mathbf{x_0} = (x_1,\ldots,x_D)$ is to be classified into one of two groups. Background data X_1,X_2 of dimension N_1×D and N_2×D (N_1 observations from group 1 and N_2 observations from group 2), respectively, exist for each group, thus allowing parameter estimation under the propositions

H_1: $\mathbf{x_0}$ comes from group 1, H_2: $\mathbf{x_0}$ comes from group 2.

A LR can thus be constructed to assess whether the observation is more likely if H_1 were true or if H_2 were true. More details on parameter estimation and the derivation of the LR can be found in the Appendix.

The Comparison Problem

Another problem, perhaps more common in forensic science, is to compare two pieces of evidence (the control and recovered samples, x_1 and x_2, respectively) in order to determine whether they could originate from the same source. For this comparison problem, the propositions of interest are

H_1: $\mathbf{x_1}$ and $\mathbf{x_2}$ come from the same group, H_2: $\mathbf{x_1}$ and $\mathbf{x_2}$ come from different groups,

and a two-part model similar to that for classification is used. The binary part is assumed to be independent of the continuous part of the model. Under H_1, $\mathbf{u_1}$ and $\mathbf{u_2}$, the incidence vectors from $\mathbf{x_1}$ and $\mathbf{x_2}$, respectively, are not independent and their joint distribution must be considered. This can be estimated from background data from the pairs coming from the same group for the numerator and from the pairs coming from different groups for the denominator. Thus, with background data with m groups of size n each, there are $\frac{1}{2}mn(n-1)$ same-group pairs and $\frac{1}{2}mn^2(m-1)$ different-group pairs for estimation of the probabilities $p(u_1 = 1,\ u_2 = 1)$, $p(u_1 = 1,\ u_2 = 0)$, $p(u_1 = 0,\ u_2 = 1)$, and $p(u_1 = 0,\ u_2 = 0)$. Same-object pairs give estimates for the joint distribution of (u_1, u_2) in the numerator of the LR (i.e., under H_1), and different-object pairs for the distribution in the denominator (i.e., under H_2). The estimates thus obtained are shown in Table 2.

For the continuous part of the model, i.e., the nonzero subcompositions, two levels of variation are assumed for the logratios, $\mathbf{y_1}$ (control) and $\mathbf{y_2}$ (recovered), of the two observations. The propositions H_1 and H_2 can be expressed in terms of the logratios and a LR can be constructed as shown in the Appendix by multiplying together the binary and continuous parts for the numerator and denominator. These likelihood calculations take into account the differing sample sizes for $\mathbf{y_1}$ and $\mathbf{y_2}$.

Variance Estimation

Formulae for the estimated within- and between-group variances in the comparison problem are given in the Appendix and follow (12). For both comparison and classification, these estimation procedures were adapted to deal with missing data corresponding to logratios of zero values. Two methods for missing data were considered: available cases and multiple imputation. The "available cases" method computes covariances from pairs of observations with complete data only, while imputation "fills in" missing data with plausible values. Multiple imputation (16) is a particular form of imputation in which the missing values are replaced by M>1 simulated versions, where M is typically small (e.g., 3–10). The multiple imputation method used here involves a combination of the Expectation–Maximization algorithm and bootstrap as described in (17,18) and is implemented in the Amelia package in R (19). From each simulated dataset with imputed values, a variance matrix is obtained. These datasets result in M variances and their average is the estimate used in the LR calculations.

Dimension Reduction for High-Dimensional Data

A limitation of multivariate modeling is the lack of background data from which to estimate the parameters of the assumed distributions such as means, variances, and covariances. This requires far more analytical data than are usually accessible in many forensic databases. This effect has been dubbed the "curse of dimensionality." For the examples described in the Physico-Chemical Data section, 264 glass samples and 36 paint samples with seven variables each are used as background data. Therefore, it is necessary to reliably estimate up to seven means, seven variances, and 21 covariances if all variables (V_i; $i = 1,\ldots,d$) where $d = D-1$ are to be taken into account in calculating the LR in

$$\mathrm{LR} = \frac{f(V_1, V_2, \ldots, V_d|H_1)}{f(V_1, V_2, \ldots, V_d|H_2)} \quad (2)$$

(LR-full-model). The estimates from a small dataset could be very imprecise. The simplest way around this problem is to assume independence of the variables considered. However, this is a naive assumption as by nature the measured features are correlated, especially those pertaining to the main components of the analyzed materials. Where feasible, models should take into account this correlation, possibly in combination with some form of dimension reduction, e.g., graphical models (20), and the independence assumption should be used only when the number of observations is extremely limited.

Graphical models are a dimension-reduction technique based on graph theory. They are obtained from the matrix of negative partial correlation coefficients, that is the rescaled inverse of the covariance matrix calculated from the between-object covariance matrix. These partial correlations are used to aid the construction of a decomposable graphical model, that is a model with complete irreducible components (20), and hence to factorize the multivariate density function into the product of several density functions in lower dimensions, allowing more precise estimates of parameters from small background sample sizes, much smaller than those for a full model.

The factorization is done according to

$$f(C_i|S_i) = \frac{f(C_i)}{f(S_i)} \quad (3)$$

where C_i is the ith clique in the model and S_i is the set of all separators for the ith clique calculated from a set chain of the cliques for the model.

A set chain is a particular ordering of the cliques in the model and may be constructed using the following factorization algorithm.

- Select a node arbitrarily from the model graph and denote this as the lowest numbered node, number each remaining node in turn ordered by the number of edges linking it to any other already numbered node.
- Break ties arbitrarily; assign a rank to each clique based upon the highest numbered node in the clique; if two cliques share the highest numbered node then rank arbitrarily between the two nodes.

Given the cliques for the model, and a suitable set chain, the set of separators (S_i) for each clique is found. The first clique in the set chain is always a complete subgraph (i.e., all nodes or elements in the clique are connected by edges) and there are no separator sets. After that, the next clique present in the set chain is added to the model. The intersection of elements between these two cliques becomes the first separator set. The process is continued until all cliques are joined to the model. This process is repeated until all nodes are part of the graphical model. If this does not happen, the model cannot be factorized (20).

In this article, the graphical model shown in Fig. 3 is used for the analysis of the glass data. The clique (Si′, Na′, Ca′) involves elements that form the basis of soda-lime-silica glass. A high correlation between Al′ and K′ was also observed in previous analyses (e.g., [1]). These variables represent compounds used in glass as additives whose role is to improve the physical properties of glass samples (optical—K_2O—or to avoid recrystallization process—Al_2O_3). The variables Mg′ and Fe′ are treated as independent variables. Variables Mg′ and Ca′ are correlated, most probably because compounds of Mg are substitutes of Ca compounds in glass since they have the same oxidation number. However, if this correlation is taken into account, the resulting model is nondecomposable (20). The proposed decomposable graphical model (Fig. 3) thus ignores the Mg′–Ca′ correlation. This model was used in all the problems studied, i.e., classification and comparison of glass samples. This is a simplified version of a graphical model previously obtained for glass data (1), and it accounts for the main relationships between the most important chemical compounds in glass.

This model was chosen because the graphical models obtained from the matrices of negative partial correlation coeficients were heavily influenced by the few observations that did contain all elements, and as such they did not represent the chemical relationships between the variables. For example, the graphical model obtained from the matrix shown in Table 4 cannot be interpreted based on chemical knowledge as the Fe′ variable has a relatively high correlation with Ca′ and Si′ (|0.464| and |0.442|, respectively). Therefore, the graphical model (Fig. 4) has an unusual construction from the point of view of the chemistry. At the same time Na′ is very weakly correlated to Ca′ (|0.033|) and to Si′ (|0.050|). However, these three variables involve elements that form the basis of soda-lime-silica glass so they should be relatively highly correlated, as

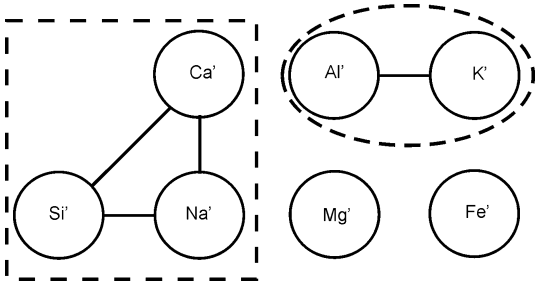

FIG. 3—*The graphical model used in this research.*

TABLE 4—*Partial correlation matrix for the seven variables based on the between-object covariance matrix C for the glass comparison problem. Only the upper right triangle of the matrix is shown, the lower left triangle is given by symmetry.*

	Na′	Mg′	Al′	Si′	K′	Ca′	Fe′
Na′	1.000	0.128	0.166	−0.050	0.162	−0.033	−0.366
Mg′		1.000	0.121	−0.157	−0.211	−0.035	−0.065
Al′			1.000	−0.143	−0.250	0.179	−0.034
Si′				1.000	−0.102	−0.517	−0.442
K′					1.000	−0.009	0.007
Ca′						1.000	0.464
Fe′							1.000

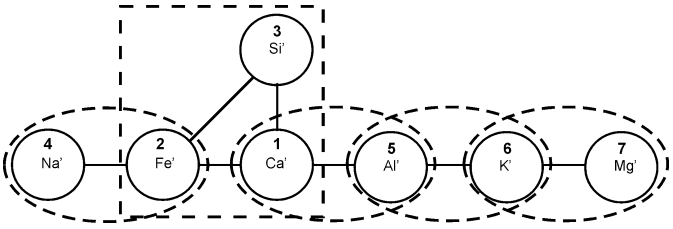

FIG. 4—*The graphical model obtained for the comparison problem of glass samples from the matrix of partial correlation coefficients in Table 4.*

observed in previous research, e.g., the partial correlation coefficients of |0.598| for Na′–Ca′ and |0.267| for Na′–Si′ in Table 3 of (1). These values allowed these three nodes to be localized near each other in the graph in Fig. 3 of (1). Similar values of partial correlation coefficients were obtained for classification tasks described in (21).

Experiments—Determination of Incorrect Answers

Experiments on Glass Data

The classification problems examined in this article arose from the practical experience of the authors. The first problem was the classification of a glass object into either the car windows, building windows, and containers category (*cwp*) or the bulbs and headlamps category (*bh*). Glass objects from category *c*, *w*, and *p* represent glass commonly occurring in forensic practice. It is expected (7) that bulbs and car headlamps (*bh*) would have systematically different elemental compositions to container glass (*p*—jars and bottles) because of their optical properties. If a glass sample was originally classified to the *cwp* category, an attempt was made to classify it further into the *cw* or *p* category. Glass objects from the *c* (car windows) and *w* (building windows) categories have very similar elemental content because they are manufactured in a similar way (by a float glass manufacturing method). Therefore, glass objects from categories *c* and *w* are treated as one category—*cw*. Two additional classification problems were studied, i.e., *b* versus *h* and *c* versus *w*. The latter problem, whether glass objects from the *c* and *w* categories could be distinguished despite the similarity of their chemical composition, was especially interesting. It was found that this was possible (7) but that the accuracy of the process was lower in comparison to other classification tasks.

The factorization of the LR

$$\mathrm{LR} = \mathrm{LR}(\mathrm{Na}', \mathrm{Ca}', \mathrm{Si}') \cdot \mathrm{LR}(\mathrm{Al}', \mathrm{K}') \cdot \mathrm{LR}(\mathrm{Mg}') \cdot \mathrm{LR}(\mathrm{Fe}') \quad (4)$$

derived from the graphical model shown in Fig. 3 was selected based on the authors' experience from previous analyses of a glass database (e.g. [1]). The notation $\mathrm{LR}(V_1,\ldots,V_k)$ refers to the LR calculated using Eq. (2) with *k*-variate densities $f(V_1,\ldots,V_k|H_1)$ and $f(V_1,\ldots,V_k|H_2)$. Note that there are no separators S_i (see Eq. [3]) in Eq. (4). The statistical approach described in the previous section was used for the evaluation of evidence obtained by glass SEM–EDX analysis in the classification and comparison problems. Two types of LR models were considered; first, those that assumed a multivariate normal (MVN) distribution for the continuous part of the LR, and, second, those in which the multivariate distribution of the continuous part of the LR was obtained by a KDE method. In addition, the outcomes of the use of two methods of calculation of the variance–covariance matrices (imputation and available cases) were studied.

The performance of the models for the classification problems was evaluated by determination of the number and percentage of samples incorrectly classified, the latter referred to as *accuracy*. The expected LR value was LR>1 if the true category of the glass object was that assumed in the numerator of Eq. (1) and LR<1 if the true category was that assumed in the denominator. An incorrect answer would yield LR>1 when LR<1 is expected or LR<1 when LR>1 is expected. Results of the classification experiments are shown in Tables 7–10.

The performance of the models for the comparison problem was evaluated in terms of the number and percentage of false negative and positive answers. A false negative answer (type I error) is an answer where the compared glass samples originated from the same glass object but were evaluated as having originated from different glass objects (LR<1). A false positive answer (type II error) is an answer where the compared glass samples originated from different glass objects but were evaluated as having originated from the same glass object (LR>1). Control of the level of false positive answers is especially important from the forensic point of view as such a result lends support, falsely, to the proposition that the defendant was associated with the crime scene and hence, perhaps, the crime. The following experiments were performed in order to study the level of false positive and false negative answers, obtained by the proposed model for the comparison problem:

- Experiment 1 (estimation of the percentage of false negative answers). The results of the analysis of the first two glass fragments of a total of four analyzed from a particular glass sample were selected for the simulated measurements of the A sample (recovered). The results of the analysis of the other two glass fragments were assigned to the B sample (control). Each simulated A sample was compared with the simulated B sample from the same total sample of four fragments. Two hundred and sixty-four such sets were created. The desirable answer was LR>1, and each answer with LR<1 was a false negative answer.
- Experiment 2 (estimation of percentage of false positive answers). The results of all four analyses performed on two glass samples of size two from different total samples of size four were selected for creation of a pair of samples to compare, i.e., samples A and B. Two hundred and sixty-four glass samples were available in the database, and thus $2^{264} = 34{,}716$ such pairs were created. The desirable answer was LR<1, and each answer with LR>1 was a false positive answer.

A comparison between the performance of the graphical model approach (Fig. 3) and two extreme assumptions, i.e., the assumption that all the variables are independent (called the *LR-independent model*) – LR = LR(Na′)LR(Mg′)LR(Al′)LR(Si′)LR(K′)LR(Ca′)LR(Fe′)—and that all variables are correlated—LR = LR(Na′,Mg′,Al′,Si′,K′,Ca′,Fe′)—called the *LR-full model*, was made. Results in the form of Tippett plots and receiver operating characteristic (ROC) curves are shown in Figs. 5 and 6.

Experiments on Paint Data

Because of the small size of the database (36 paint samples), only the LR-independent model LR = LR(MMA)LR(TOL)LR(BMA)LR(MST)LR(M2E)LR(M2P)LR(I16) was considered. The following experiments were performed in order to check the level of false positive and false negative answers obtained by using the proposed model for the comparison of paints:

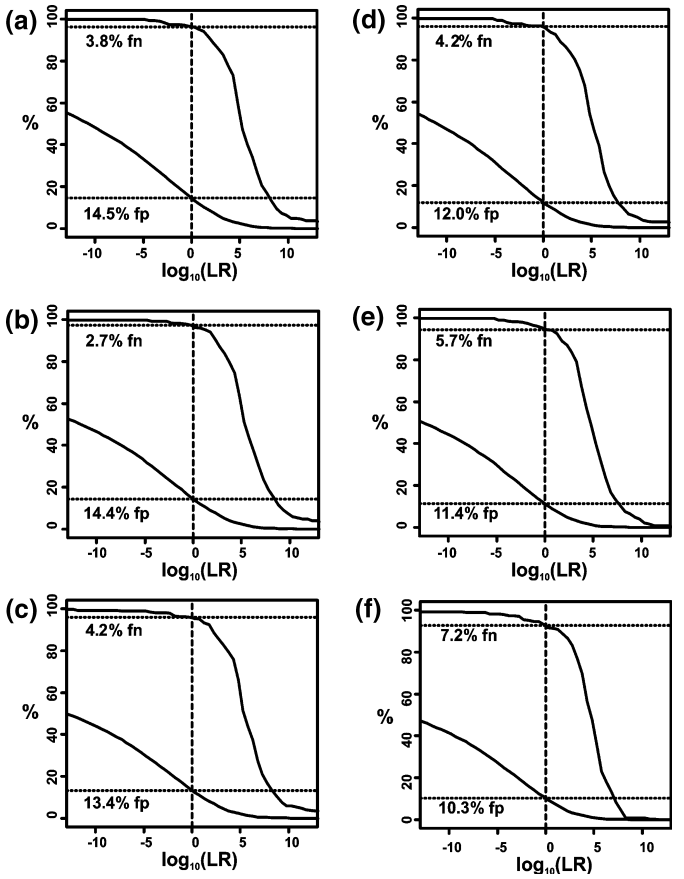

FIG. 5—*Tippett plots—comparison problem of glass samples when the variance was estimated using multiple imputation: (a) independent model with multivariate normality (MVN) assumption, (b) graphical model with MVN assumption, (c) full model with MVN assumption, (d) independent model with kernel density estimation (KDE) assumption, (e) graphical model with KDE assumption, (f) full model with KDE assumption. fn, false negative answers; fp, false positive answers.*

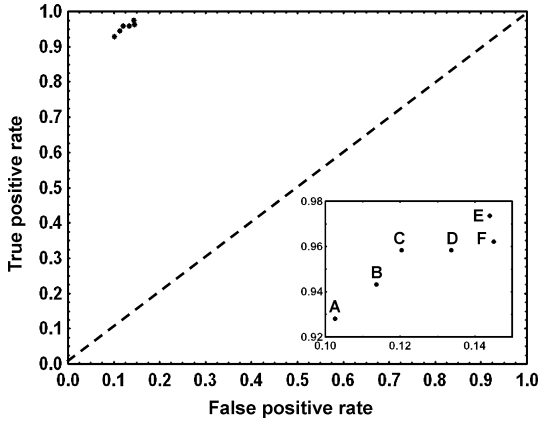

FIG. 6—*Plot of true positive rate against false positive rate for the performance of models applied to the comparison problem of glass samples: (a) full model with kernel density estimation (KDE) assumption (A), (b) graphical model with KDE assumption (B), (c) independence model with KDE assumption (C), (d) full model with multivariate normality (MVN) assumption (D), (e) graphical model with MVN assumption (E), (f) independence model with MVN assumption (F).*

- Experiment 1 (estimation of the percentage of false negative answers). The results of the first two analyses of a total of three performed on a particular paint sample were selected for the simulated measurements of the A sample. The result of the third analysis of the paint sample was assigned to the B sample. Each simulated A sample was compared with a simulated B sample from the same source. Thirty-six such sets were created. The desirable answer was LR>1 and each answer with LR<1 was a false negative answer.
- Experiment 2 (estimation of percentage of false positive answers). The results of all three analyses performed on two different paint samples were selected for creation of a set of compared samples, i.e., samples A and B. Thirty-six paint samples were available in the database, and thus $2^{36} = 630$ such sets were created. The desirable answer was LR<1 and each answer with LR>1 was a false positive answer.

Results are presented in the form of Tippett plots in Fig. 7.

Discussion

Glass—Classification Results

Because of the varying number of observations for each of the classification tasks, multiple imputation for variance estimation was not always feasible; hence, the available-cases method was used for this part of the statistical analysis.

Examination of Table 5 shows that classification into *cwp* versus *bh* gave in general very satisfactory results, i.e., low rates of false answers. Slightly better results were obtained when normality (MVN) was assumed—minimum cumulative percent of false answers equal to 1.5% (a total of four incorrectly classified samples). This seems to be in contrast to the expectation that KDE would give better results. A more detailed inspection shows that most of the incorrect classifications in the KDE model came from the *bh* category, where the full (F), graphical (G), and independent (I) models gave 4, 11, and 25 incorrectly classified samples, respectively, out of 42 (Table 5). At the same time, normal models gave just two incorrectly classified samples. This is not surprising because the *bh* category contains only 42 samples, which could be too small for reliable nonparametric multivariate density estimation. Moreover, the elemental composition for the set of bulbs (*b*) varied much more between objects than in other categories (Fig. 1) implying that it affects the variability of the *bh* category too. This may also have contributed to problems with estimation for the *bh* category. On the other hand, the number of samples in the *cwp* dataset (222) was sufficiently large for reliable nonparametric density estimation. The full, graphical, and independent models gave two (1.0%), five (4.8%), and four (2.3%) incorrectly classified *cwp* samples, respectively, under the MVN assumption and one (0.5%), one, and zero incorrectly classified *cwp* samples when KDE was used (Table 5).

The graphical models in combination with KDE gave the best cumulative results for the problem *cw* versus *p*. The full, graphical, and independent models yielded eight (4.9%), four (2.4%), and one (0.6%) incorrectly classified *cw* samples and nine (15.8%), seven (12.3%), and 31 (54.4%) incorrectly classified *p* samples, respectively (Table 6). At the same time, normal models gave 13 (7.9%), eight (4.9%), and 12 (7.3%) incorrectly classified *cw* samples and 12 (21.1%), five (8.8%), and five incorrectly classified *p* samples, respectively.

Very similar patterns were observed in Table 7, which shows results obtained for the *b* versus *h* task. However, because of the small number of samples and limited availability of Fe′

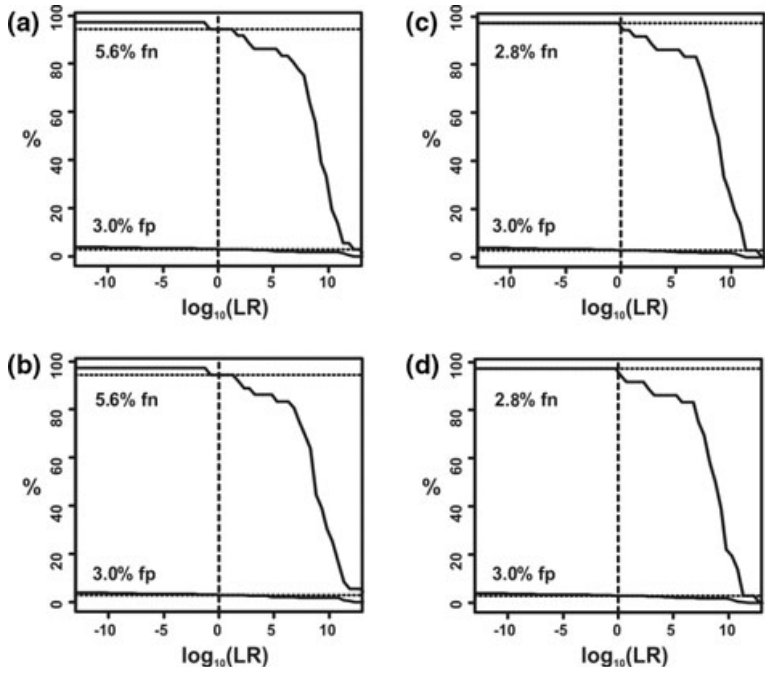

FIG. 7—*Tippett plots—comparison problem of paint samples: (a) variance estimation using imputation with multivariate normality (MVN) assumption, (b) available case variance estimation with MVN assumption, (c) variance estimation using imputation with kernel density estimation (KDE) assumption, (d) available case variance estimation with KDE assumption. fn, false negative answers; fp, false positive answers.*

TABLE 5—*Likelihood ratio (LR) distribution for glass object classification into car windows, building windows, and containers (cwp) or bulbs and headlamps (bh). MVN, multivariate normal distribution; KDE, kernel density estimation; F, full model; G, graphical model; I, independent model. There are 264 samples in the full dataset of which 42 are bh and 222 are cwp; percentages in the table are percentages of 222 for columns headed cwp and of 42 for columns headed bh and of 264 for the last row, labeled "cumulative %."*

	MVN						KDE					
	F		G		I		F		G		I	
LR	*cwp*	*bh*	*cwp*	*bh*	*cwp*	*bh*	*cwp*	*bh*	*cwp*	*bh*	*cwp*	*bh*
$>10^7$	53.2	52.4	35.6	33.3	35.1	33.3	53.2	54.8	51.8	33.3	71.6	33.3
10^6–10^7	9.0	2.4	9.9	4.8	10.4	0.0	7.2	2.4	8.6	0.0	14.9	0.0
10^5–10^6	6.3	0.0	8.1	0.0	9.9	14.3	7.2	4.8	16.2	4.8	9.5	0.0
10^4–10^5	9.9	14.3	10.8	23.8	9.5	16.7	14.0	7.1	12.2	4.7	2.7	0.0
10^3–10^4	7.2	16.7	9.0	19.1	9.0	4.8	10.8	9.5	8.1	7.1	0.9	0.0
10^2–10^3	7.7	4.8	9.0	7.1	8.6	14.3	5.9	2.4	1.4	4.8	0.0	0.0
10^1–10^2	3.6	4.8	9.9	4.8	9.0	11.9	0.5	4.8	1.4	9.5	0.5	2.4
1–10	2.3	0.0	5.4	2.4	6.8	0.0	0.9	4.8	0.0	9.5	0.0	4.8
10^{-1}–1	0.5	2.4	0.9	2.4	0.9	2.4	0.0	0.0	0.0	2.4	0.0	4.8
10^{-2}–10^{-1}	0.0	2.4	0.9	2.4	0.9	2.4	0.0	7.1	0.0	11.9	0.0	14.3
10^{-3}–10^{-2}	0.0	0.0	0.5	0.0	0.0	0.0	0.0	2.4	0.0	4.8	0.0	16.7
10^{-4}–10^{-3}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.8	0.0	16.7
10^{-5}–10^{-4}	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	4.8
10^{-6}–10^{-5}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	2.4
$<10^{-6}$	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.5	2.4	0.0	0.0
False answers %	1.0	4.8	2.3	4.8	1.8	4.8	0.5	9.5	0.5	26.3	0.0	59.7
Cumulative %	1.5		2.7		2.3		1.9		4.6		9.5	

observations, this variable was not used in the analysis. KDE results were poor because of the small number of data points (26 *b* and 16 *h*), i.e., five (11.9%), three (7.1%), and six (14.3%) incorrectly classified samples for the full, graphical, and independent models, respectively. The LR model with the normality assumption and the graphical approach for dimension reduction gave, respectively, 2.4%, 4.8%, and 9.5% false answers (one, two, and four incorrect classifications). The LR-full model gave slightly better results (one incorrectly classified sample; Table 7) but it also gave, as in other classification problems, many more LR values above 10^7 than the LR-graphical model.

Results obtained for the *c* versus *w* problem were the worst among all classification problems considered (Table 8). The lowest cumulative number of incorrectly classified samples (18.2%) was observed for the normality assumption and the graphical approach for dimension reduction. This could be expected taking into

TABLE 6—*Likelihood ratio (LR) distribution for glass object classification into car windows and building windows (cw) or containers (p). MVN, multivariate normal distribution; KDE, kernel density estimation; F, full model; G, graphical model; I, independent model. There are 264 samples in the full dataset of which 165 are cw and 57 are p; percentages in the table are percentages of 165 for columns headed cw and of 57 for columns headed p and of 222 for the last row, labeled "cumulative %."*

	MVN						KDE					
	F		G		I		F		G		I	
LR	*cw*	*p*	*cw*	*p*	*cw*	*p*	*cw*	*p*	*cw*	*p*	*cw*	*p*
$>10^7$	13.9	29.8	12.7	33.3	12.1	33.3	43.0	35.1	24.2	1.8	23.6	1.8
10^6–10^7	5.5	3.5	6.7	0.0	4.2	0.0	9.7	12.3	6.1	0.0	17.0	0.0
10^5–10^6	4.2	1.8	4.2	1.8	5.5	1.8	6.1	5.3	20.6	5.3	19.4	0.0
10^4–10^5	19.4	3.5	15.8	5.3	12.7	5.3	9.1	5.3	20.6	1.8	20.6	1.8
10^3–10^4	20.0	10.5	12.7	14.0	14.6	8.8	6.1	5.3	14.6	24.6	10.9	1.8
10^2–10^3	15.2	5.3	16.4	5.3	20.6	14.0	13.3	8.8	5.5	8.8	5.5	8.8
10^1–10^2	7.3	14.0	19.4	21.1	13.3	15.8	6.7	8.8	4.9	26.3	2.4	15.8
1–10	7.3	10.5	7.3	10.5	9.7	12.3	1.2	3.5	1.2	19.3	0.0	15.8
10^{-1}–1	4.2	10.5	3.0	7.0	5.5	5.3	1.8	3.5	0.6	1.8	0.0	33.3
10^{-2}–10^{-1}	1.2	5.3	0.6	0.0	0.6	1.8	0.6	5.3	0.0	7.0	0.0	8.8
10^{-3}–10^{-2}	0.6	5.3	0.6	0.0	0.0	0.0	0.6	3.5	0.6	1.8	0.0	7.0
10^{-4}–10^{-3}	0.0	0.0	0.0	1.8	0.0	0.0	0.6	1.8	0.0	0.0	0.0	3.5
10^{-5}–10^{-4}	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.6	0.0	0.0	0.0
10^{-6}–10^{-5}	0.0	0.0	0.0	0.0	0.0	1.8	0.6	0.0	0.0	0.0	0.0	0.0
$<10^{-6}$	1.2	0.0	0.6	0.0	0.6	0.0	0.6	1.8	0.6	1.8	0.6	1.8
False answers %	7.9	21.1	4.9	8.8	7.3	8.8	4.9	15.8	2.4	12.3	0.6	54.4
Cumulative %	11.3		5.9		7.7		7.7		5.0		14.4	

TABLE 7—*Likelihood ratio (LR) distribution for glass object classification into bulbs (b) or headlamps (h). MVN, multivariate normal distribution; KDE, kernel density estimation; F, full model; G, graphical model; I, independent model. There are 264 samples in the full dataset of which 26 are b and 16 are h; percentages in the table are percentages of 26 for columns headed b and of 16 for columns headed h and of 42 for the last row, labeled "cumulative %."*

	MVN						KDE					
	F		G		I		F		G		I	
LR	*b*	*h*	*b*	*h*	*b*	*h*	*b*	*h*	*b*	*h*	*b*	*h*
$>10^7$	84.6	56.3	61.5	12.5	57.7	6.3	50.0	12.5	50.0	18.8	69.2	12.5
10^6–10^7	7.7	0.0	3.9	0.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0
10^5–10^6	0.0	0.0	3.9	0.0	0.0	0.0	30.8	0.0	0.0	6.3	0.0	0.0
10^4–10^5	3.9	12.5	0.0	0.0	0.0	0.0	3.8	12.5	0.0	6.3	0.0	6.3
10^3–10^4	0.0	6.3	7.7	18.8	11.5	6.3	7.7	12.5	0.0	6.3	0.0	0.0
10^2–10^3	3.9	6.3	0.0	25.0	0.0	18.8	0.0	12.5	0.0	25.0	7.7	6.3
10^1–10^2	0.0	6.3	11.5	12.5	7.7	31.3	0.0	18.8	38.5	12.5	19.2	18.8
1–10	0.0	6.3	11.5	18.8	15.4	18.8	0.0	12.5	11.5	6.3	3.9	18.8
10^{-1}–1	0.0	6.3	0.0	6.3	7.7	12.5	7.7	6.3	0.0	12.5	0.0	6.3
10^{-2}–10^{-1}	0.0	0.0	0.0	6.3	0.0	0.0	0.0	12.5	0.0	0.0	0.0	25.0
10^{-3}–10^{-2}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	6.3
10^{-4}–10^{-3}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10^{-5}–10^{-4}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10^{-6}–10^{-5}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$<10^{-6}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
False answers %	0.00	6.3	0.0	12.5	7.7	12.5	7.7	18.8	0.0	18.8	0.0	37.5
Cumulative %	2.4		4.8		9.5		11.9		7.1		14.3	

account that samples from both categories have very similar elemental composition because of their similar manufacturing process (float glass production). Moreover, LR values are relatively small, i.e., in the range 0.01–100, and hence lend limited support to either of the considered propositions (H_1 and H_2).

In general, the independent models performed poorly because they do not take into account the correlation between variables. With enough background data, this correlation can be estimated relatively accurately, and use of a multivariate model can improve performance in both classification and comparison tasks.

Glass—Comparison Problem Results

Both methods of variance estimation (available cases and multiple imputation) gave very similar results for the glass comparison problem, with the imputation method yielding slightly lower percentages of false positive answers. As it is more important that the percentage of false positive answers is controlled, in what follows results are presented that are obtained using the multiple imputation method for variance estimation. Application of the KDE approach returned fewer false positive answers than the model assuming

TABLE 8—*Likelihood ratio (LR) distribution for glass objects classification into car windows (c) or building windows (w). MVN, multivariate normal distribution; KDE, kernel density estimation; F, full model; G, graphical model; I, independent model. There are 264 samples in the full dataset of which 86 are c and 79 are w; percentages in the table are percentages of 86 for columns headed c and of 79 for columns headed w and of 165 for the last row, labeled "cumulative %."*

	MVN						KDE					
	F		G		I		F		G		I	
LR	*c*	*w*	*c*	*w*	*c*	*w*	*c*	*w*	*c*	*w*	*c*	*w*
$>10^7$	0.0	3.5	0.0	1.2	0.0	1.2	1.3	2.3	0.0	1.2	0.0	1.2
10^6–10^7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0
10^5–10^6	0.0	0.0	0.0	2.3	0.0	1.2	0.0	1.2	0.0	0.0	0.0	0.0
10^4–10^5	0.0	1.2	0.0	0.0	0.0	0.0	3.8	1.2	0.0	1.2	0.0	1.2
10^3–10^4	2.5	0.0	1.3	0.0	0.0	0.0	6.3	8.1	2.5	1.2	0.0	0.0
10^2–10^3	2.5	0.0	21.5	2.3	1.3	1.2	8.9	17.4	2.5	4.7	3.8	5.8
10^1–10^2	27.9	12.8	68.4	1.2	30.4	12.8	24.1	23.3	22.8	29.1	25.3	30.2
1–10	58.2	46.5	7.6	55.8	50.6	50.0	38.0	24.4	51.9	31.4	44.3	29.1
10^{-1}–1	5.1	33.7	0.0	34.9	15.2	29.1	12.7	10.5	16.5	20.9	21.5	24.4
10^{-2}–10^{-1}	2.5	2.3	0.0	2.3	1.3	4.7	1.3	7.0	2.5	9.3	3.8	5.8
10^{-3}–10^{-2}	0.0	0.0	0.0	0.0	0.0	0.0	1.3	2.3	0.0	0.0	0.0	1.2
10^{-4}–10^{-3}	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	1.2
10^{-5}–10^{-4}	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10^{-6}–10^{-5}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$<10^{-6}$	0.0	0.0	1.3	0.0	1.3	0.0	1.3	1.2	1.3	1.2	1.3	0.0
False answers %	8.9	36.1	1.3	37.2	17.7	33.7	17.7	31.0	20.3	31.4	26.6	32.6
Cumulative %	21.2		18.2		24.9		18.8		25.5		28.5	

MVN, as can be seen in Figs. 5 and 6. For example, when the assumption of independence of all variables was made (Figs. 5(*a*) and (*d*)), the level of false positive answers was equal to 14.5% for the MVN model and 12.0% for the KDE model. The improvement in the rates of false positive answers is followed by deterioration of the rates of false negative answers—from 3.8% (10 incorrect results; Fig. 5(*a*)) to 4.2% (11 incorrect results; Fig. 5(*d*)), respectively. The differences in levels of false positive answers seem to be very small but in fact 0.1% means that the wrong result was obtained in 35 comparisons. At the same time, one incorrect comparison corresponds to 0.4% in the experiment for estimation of false negative answers. Therefore, the kernel estimation procedure is recommended for the glass comparison problem.

Moreover, the KDE procedure and the LR-full model shown in Fig. 5(*f*) gave the lowest percent false positives (10.3%) but at the same time the highest rates of false negative answers (7.2%, Fig. 5(*f*)). The opposite was observed for the LR-independent model, and the graphical model gave results somewhere in the middle. Similar patterns were observed in the MVN models, except that use of the LR-graphical model yielded the lowest level of false negative answers (Fig. 5(*b*)).

Because a direct comparison of the above six models is not straightforward, ROC analysis was also performed in order to compare the performance of the six models (Fig. 5(*a*)–(*f*)). Results are shown in Fig. 6, where it can be seen that all models gave very satisfactory results as all points are located near the upper-left corner where the point characterizing the perfect model (false positive rates = 0, true positive rates = 1) is located.

In the Tippett plots in Fig. 5, the lower solid line (H_2; obtained from different-source comparisons) does not reach 100% on the left side in the range up to −12 (negative values of log [LR]). This means that around 50% of the obtained results gave very small values of LR, i.e., much lower than 10^{-12}. At the same time, all LR values obtained from same-source comparisons (rates of false negative answers) were in the range from 10^{-12} to 10^{12}.

Paint—Comparison Problem

The LR-independent model was the only model considered because of the limited number of paint samples (36). The level of false positive answers was 3.0% for both the available case and imputation methods of variance estimation, and also for both MVN and KDE. The level of false negative answers was the same for both variance estimation methods but KDE gave slightly better rates of false negative answers, i.e., 2.8% (one incorrect comparison) than MVN (5.6% false negatives or two incorrect comparisons).

LR values in the range from 10^{-12} to 10^{12} (upper line, Fig. 7) were obtained in the experiment for estimation of the percentage of false negative answers. The lower line in Fig. 7, which shows log(LR) results from different-source comparisons, suggests that

TABLE 9—*Estimated probability vectors for the paint comparison problem. For the numerator of the likelihood ratio, these are obtained from same-object pairs, and for the denominator from different-object pairs.*

Variable	MMA	TOL	BMA	MST	M2E	M2P	I16
Numerator							
$p(u_1 = 0, u_2 = 0)$	0.306	0.306	0.222	0.194	0.500	0.333	0.167
$p(u_1 = 0, u_2 = 1$ or $u_1 = 1, u_2 = 0)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$p(u_1 = 1, u_2 = 1)$	0.694	0.694	0.778	0.806	0.500	0.667	0.833
Denominator							
$p(u_1 = 0, u_2 = 0)$	0.087	0.087	0.044	0.033	0.243	0.105	0.024
$p(u_1 = 0, u_2 = 1$ or $u_1 = 1, u_2 = 0)$	0.437	0.437	0.356	0.322	0.514	0.457	0.286
$p(u_1 = 1, u_2 = 1)$	0.476	0.476	0.600	0.644	0.243	0.438	0.690

TABLE 10—*Values of the optimal kernel bandwidth h_{opt} given by Eq. (5) when d variables are considered. The sample size n is shown in brackets for each of the five categories of glass objects (b, bulbs; c, car windows; h, headlamps; p, containers; w, building windows).*

	Glass Type				
d	*b* (26)	*h* (16)	*c* (86)	*w* (79)	*p* (57)
1	0.55	0.61	0.43	0.44	0.47
2	0.56	0.61	0.46	0.47	0.49
3	0.58	0.62	0.49	0.49	0.52
7	0.66	0.69	0.59	0.60	0.61

most of the LR values are lower than 10^{-12}, i.e., they are very large values in support of the proposition in the denominator. This is due to a feature of the small paint background database, namely that no pair of different-source observations had a zero and a nonzero value for the same variable (Table 9). Thus, the likelihood of such a pair coming from the same source was estimated as zero (and hence the log-likelihood was $-\infty$).

These results are slightly better than those obtained in (4), which were 5.6% of false negative answers and 3.3% of false positive answers, but as they were obtained from a rather small database (36 paint samples), more Py-GC/MS data are needed before this approach can be used in forensic practice. For a larger database, a multivariate model is anticipated to perform better than the model used here.

Conclusions

The proposed procedure considers a two-part, two-level multivariate model for evidence evaluation in the presence of zeros. The main difference between this model and the model in (12) is that it is no longer necessary to substitute the zeros by a small value. In addition, the resulting nonzero subcompositions can be closer to normality when the influence of many small values caused by zero substitution is removed. Data with zeros occur routinely in forensic science and modeling the zeros is desirable in many cases. In addition, the procedure offers the flexibility of multivariate modeling even when the amount of background data is not very large, e.g., by means of density factorization using decomposable graphical models. For small datasets, univariate densities can be of some use, but it is desirable to obtain more data for fitting multivariate models. The proposed LR models gave very satisfactory results for both problems considered (classification and comparison) of glass and car paint data. For relatively large datasets (such as the glass data in the comparison problem), use of KDE gives better results than normal models. Simpler models were used for the paint problem because of the small amount of background data and some promising results were obtained. This suggests that more effort should be made in forensic laboratories toward data collection, especially in the case of routine forensic problems (e.g., glass, paint, or fire debris analysis) and calculation of their evidential value.

Acknowledgment

The authors thank Prof. Janina Zieba-Palus, Institute of Forensic Research, Krakow, Poland for providing some of the paint data.

References

1. Aitken CGG, Zadora G, Lucy D. A two level model for evidence evaluation. J Forensic Sci 2007;52:412–9.
2. Aitken CGG, Lucy D, Zadora G, Curran JM. Evaluation of trace evidence for three-level multivariate data with the use of graphical models. Comput Stat Data Anal 2006;50:2571–88.
3. Zieba-Palus J, Zadora G, Milczarek JM, Koscielniak P. Pyrolysis-gas chromatography/mass spectrometry analysis as a useful tool in forensic examination of automotive paint traces. J Chromatogr A 2008;1179: 41–6.
4. Zieba-Palus J, Zadora G, Milczarek JM, Koscielniak P. Differentiation and evaluation of evidence value of styrene acrylic urethane topcoat car paints analyzed by pyrolysis-gas chromatography. J Chromatogr A 2008;1179:47–58.
5. Caddy B. Forensic examination of glass and paint. Boca Raton, FL: CRC Press, 2001.
6. Curran JM, Hicks TN, Buckleton JS. Forensic interpretation of glass evidence. Boca Raton, FL: CRC Press, 2000.
7. Zadora G. Glass analysis for forensic purposes—a comparison of classification methods. J Chemom 2007;21:174–86.
8. Hicks TC, Monard Sermier F, Goldmann T, Brunelle A, Champod C, Margot P. The classification and discrimination of glass fragments using non destructive energy dispersive x-ray fluorescence. Forensic Sci Int 2003;137:107–18.
9. Trejos T, Almirall JR. Sampling strategies for the analysis of glass fragments LA-ICP-MS part 1: micro-homogeneity study of glass and its application to the interpretation of forensic evidence. Talanta 2005;67:388–95.
10. Aitken CGG, Taroni F. Statistics and the evaluation of evidence for forensic scientists. Chichester, U.K.: John Wiley & Sons, 2004.
11. Bishop C.M. Pattern recognition and machine learning. New York: Springer, 2006.
12. Aitken CGG, and Lucy D. Evaluation of trace evidence in the form of multivariate data. J R Stat Soc Ser C Appl Stat 2004;53:109–22.
13. Butler A, Glasbey CA. A latent Gaussian model for compositional data with zeros. J R Stat Soc Ser C Appl Stat 2008;57:505–20.
14. Fry J, Fry T, McLaren K. Compositional data analysis and zeros in micro data. Appl Econ 2000;32:953–9.
15. Aitchison J, Kay J. Possible solutions of some essential zero problems in compositional data analysis. Girona, Spain: In Compositional Data Analysis Workshop, 2004.
16. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons, 1987.
17. Honaker J, King G. What to do about missing values in time-series cross-section data? 2007. http://gking.harvard.edu/files/pr.pdf (accessed January 23, 2009).
18. Honaker J, King G, Blackwell M. Amelia II: a program for missing data. 2007. http://gking.harvard.edu/amelia/ (accessed January 23, 2009).
19. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0. http://www.R-project.org (accessed January 23, 2009).
20. Whittaker J. Graphical models in applied multivariate statistics. Chichester: John Wiley & Sons, 1990.
21. Zadora G. Classification of glass fragments based on elemental composition and refractive index. J Forensic Sci 2009;54:49–59.
22. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall, 1986.
23. Lindley DV. A problem in forensic science. Biometrika 1977;64:207–13.

Additional information and reprint requests:
Grzegorz Zadora, Ph.D.
Institute of Forensic Research
Westerplatte 9
31-033 Krakow
Poland
E-mail: gzadora@ies.krakow.pl

Appendix

Likelihood Ratio (LR) Model for Classification

The LR for a composition $\mathbf{x_0}$ with an incidence vector $\mathbf{u_0}$ and a logratio vector $\mathbf{y_0}$ is assumed to be

$$\frac{b(\mathbf{u_0}|\mathbf{p_1})f(Q_1\mathbf{y_0}|\theta_1)}{b(\mathbf{u_0}|\mathbf{p_2})f(Q_2\mathbf{y_0}|\theta_2)}$$

where $f(\cdot)$ is the multivariate density function for the logratio-transformed nonzero subcomposition vector $Q_g\mathbf{y_0}$, $g = 1,2$. The parameters $(\mathbf{p_g},\theta_\mathbf{g})$ are replaced by their estimates for $g = 1,2$ obtained from the background data. The probability vectors $\mathbf{p}$ are simply the proportions of each component that are not zero, as shown in Table 1.

If $f(\cdot)$ is assumed to be MVN, the parameters $\theta_\mathbf{g}$ are the mean and covariance matrix of the MVN distribution, taking distinct values for the two distinct populations. If the logratios are not normally distributed, the continuous parts of the likelihood are estimated using Gaussian kernels. Given background data (transformed logratios) $\{\mathbf{v_{gi}}, i = 1,\ldots,m_g, g = 1,2\}$ from the population database and smoothing parameter (kernel bandwidth) h, the multivariate Gaussian kernel for the gth population is

$$K_g(\mathbf{y_0}|\mathbf{v_{g1}},\ldots,\mathbf{v_{gm_g}},\Sigma_g,h) = \frac{1}{m_g}\sum_{i=1}^{m_g} f_{gi}$$

where $f_{gi} = (2\pi)^{-d_{gi}/2}|Q_{gi}(h^2\Sigma_g)Q_{gi}^\top|^{-1/2}\exp\{-\frac{1}{2}(Q_{gi}\mathbf{y_0} - Q_{gi}\mathbf{v_{gi}})^\top\ [Q_{gi}(h^2\Sigma_g)Q_{gi}^\top]^{-1}(Q_{gi}\mathbf{y_0} - Q_{gi}\mathbf{v_{gi}})\}$ and d_{gi} equals the number of nonzero variables in data point gi.

The Q_{gi} select the nonzero variables from both $\mathbf{y_0}$ and $\mathbf{v_{gi}}$. The optimal value for the kernel bandwidth is

$$h_{\text{opt}} = \left(\frac{4}{n(2d+1)}\right)^{\frac{1}{d+4}}, \tag{5}$$

where d is the number of variables and n the number of observations (22). Numerical values of h_{opt} for the various numbers of parameters, d, used in the models for classification, are shown in Table 10. In practice, sensitivity analysis shows that values of h between 0.4 and 0.5 perform well; and thus, for simplicity, the value $h = 0.45$ was used for all glass classification problems.

LR Model for Comparison Problem

Suppose that the controlled and recovered vectors (means of n_1 and n_2 observations, respectively) are $\mathbf{x_1}$ and $\mathbf{x_2}$ with corresponding incident vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ and logratios $\mathbf{y_1}$ and $\mathbf{y_2}$, respectively. The background data comprise logratios $\mathbf{v}_{ij}$ $i = 1,\ldots,m$, $j = 1,\ldots,n_i$ from m groups. The group means $\bar{\mathbf{v}}_i(=(\sum_{j=1}^{n_i}\mathbf{v}_{ij})/n_i)$, as well as the within-group covariance matrix U, and the between-group covariance matrix, C, can be estimated using multiple imputation or available cases to deal with missing data due to zeros.

The LR is assumed to be of the form

$$\frac{b(\mathbf{u_1},\mathbf{u_2}|H_1)f(\mathbf{y_1},\mathbf{y_2}|H_1)}{b(\mathbf{u_1},\mathbf{u_2}|H_2)f(\mathbf{y_1},\mathbf{y_2}|H_2)}$$

where the binary part, $b(\cdot)$, considers two separate sets of joint probabilities for the incidence vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ under the prosecution proposition H_1 and the defense proposition H_2, and the continuous part is obtained as follows.

The denominator assumes independence of $\mathbf{y_1}$ and $\mathbf{y_2}$, with normal within-group distributions and Gaussian KDE for between-group distributions, with kernel bandwidth h. Thus

$$f(\mathbf{y_1},\mathbf{y_2}|H_2) = f(\mathbf{y_1}|U,C,h)f(\mathbf{y_2}|U,C,h)$$

with $f(\mathbf{y_l}|U,C) = \frac{1}{m}\sum_{i=1}^{m}\{(2\pi)^{-d_{li}/2}|Q_{li}^*(\frac{U}{n_l} + h^2C)Q_{li}^{*\top})|^{-1/2}\times\ \exp\{-\frac{1}{2}(Q_{li}^*\mathbf{y_l} - Q_{li}^*\bar{\mathbf{v}}_i)^\top(Q_{li}^*(\frac{U}{n_l} + h^2C)Q_{li}^{*\top})^{-1}(Q_{li}^*\mathbf{y_l} - Q_{li}^*\bar{\mathbf{v}}_i)\}\}$ for Q_{li}^* defined as the operator selecting only variables that are nonzero for both $\mathbf{y_l}$ and $\bar{\mathbf{v}}_i$ for $l = 1,2$ and $i = 1,\ldots,m$.

In the numerator, the joint probability density function of $\mathbf{y_1}$ and $\mathbf{y_2}$ is the "average" of products of two independent MVN terms, one for $(\mathbf{y_1}-\mathbf{y_2})$ and another for $(\mathbf{y}^* - \bar{\mathbf{v}}_i)$, where

$$\mathbf{y}^* = \frac{n_1\mathbf{y_1} + n_2\mathbf{y_2}}{n_1 + n_2}$$

Let A be the matrix that selects the nonzero subcomposition of $(\mathbf{y_1}-\mathbf{y_2})$ and B_i the corresponding matrices for $(\mathbf{y}^* - \bar{\mathbf{v}}_i)$, $i = 1,\ldots,m$. This results in

$$A(\mathbf{y_1}-\mathbf{y_2}) \sim \mathrm{MVN}\left(0, A\left(\frac{U}{n_1}+\frac{U}{n_2}\right)A^{\top}\right)$$

and, independently,

$$B_i(\mathbf{y}^*-\bar{\mathbf{v}_i}) \sim \mathrm{MVN}\left(0, B_i\left[\left(\left(\frac{U}{n_1}\right)^{-1}+\left(\frac{U}{n_2}\right)^{-1}\right)^{-1}+h^2C\right]B_i^{\top}\right)$$

which yield

$$f(\mathbf{y_1},\mathbf{y_2}|H_1) = f(\mathbf{y_1}-\mathbf{y_2},\mathbf{y}^*|U,C,h) = (2\pi)^{-d_A/2}|A\left(\frac{U}{n_1}+\frac{U}{n_2}\right)A^{\top}|^{-1/2}\exp\left\{-\frac{1}{2}(A\mathbf{y_1}-A\mathbf{y_2})^{\top}\left[A\left(\frac{U}{n_1}+\frac{U}{n_2}\right)A^{\top}\right]^{-1}(A\mathbf{y_1}-A\mathbf{y_2})\right\}$$
$$\times\frac{1}{m}\sum_{i=1}^{m}\left\{(2\pi)^{-d_i/2}|B_i\left[\left(\left(\frac{U}{n_1}\right)^{-1}+\left(\frac{U}{n_2}\right)^{-1}\right)^{-1}+h^2C\right]B_i^{\top}|^{-1/2}\right.$$
$$\left.\times\exp\left[-\frac{1}{2}(B_i\mathbf{y}^*-B_i\bar{\mathbf{v}_i})^{\top}\left(B_i\left[\left(\left(\frac{U}{n_1}\right)^{-1}+\left(\frac{U}{n_2}\right)^{-1}\right)^{-1}+h^2C\right]B_i^{\top}\right)^{-1}(B_i\mathbf{y}^*-B_i\bar{\mathbf{v}_i})\right]\right\}$$

The kernel bandwidth, h, is taken to equal its optimal value given by Eq. (5), assuming the maximal dimension $d = 7$ for glass, and $d = 1$, i.e., univariate, for paint.

Note that if Q, A, and B are taken to be identity matrices, then the expression for $f(\mathbf{y_1},\mathbf{y_2}|H_1) = f(\mathbf{y_1}-\mathbf{y_2},\mathbf{y}^*|U,C,h)$ reduces to Eq. (14) of (12). Also, the expression $f(\mathbf{y_1}|U,C,h)f(\mathbf{y_2}|U,C,h)$ reduces to Eq. (15) of (12).

Some clarification is needed of the differences between the models in (12) and (1). The expression for the numerator in the Appendix in (1) includes a third component of variation for measurement error using the between-group covariance matrix. It is hard to envisage a situation where this may arise and the authors recommend the use of the results in (12). Note also that the comment on page 419 of (1) about the "multivariate analog of the univariate example in (23)" refers to the expressions for

$$\bar{Y}_1-\bar{Y}_2 \sim N(0,\frac{U}{n_c}+\frac{U}{n_r})$$

and

$$(n_c\bar{Y}_1+n_r\bar{Y}_2)/(n_c+n_r) \sim N\left(\mu, C+\frac{U}{n_c+n_r}\right)$$

and $(\bar{Y}_1-\bar{Y}_2)$ and $(n_c\bar{Y}_1+n_r\bar{Y}_2)/(n_c+n_r)$ are independent with unit Jacobian at the bottom of page 418 of (1).

Variance Estimation

Assuming a background database of m groups of n objects, denoted as

$$\mathbf{x}_{ij} = (x_{ij1},\ldots,x_{ijp})^{\top}; i = 1,\ldots,m, j = 1,\ldots,n,$$

with $\bar{\mathbf{x}}_i = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_{ij}$, $\bar{\mathbf{x}} = \sum_{i=1}^{m}\bar{\mathbf{x}}_i/m$, the within-group variance estimate is

$$\hat{U} = \frac{S_w}{m(n-1)}$$

where

$$S_w = \sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)(\mathbf{x}_{ij}-\bar{\mathbf{x}}_i)^{\top},$$

and the between-group variance estimate is

$$\hat{C} = \frac{S^*}{m-1}-\frac{S_w}{mn(n-1)}$$

where

$$S^* = \sum_{i=1}^{m}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\top}.$$