J Forensic Sci, March 2007, Vol. 52, No. 2
doi:10.1111/j.1556-4029.2006.00358.x
Available online at: www.blackwell-synergy.com

Colin G. G. Aitken,[1] *Ph.D.; Grzegorz Zadora,*[2] *Ph.D.; and David Lucy,*[1] *Ph.D.*

A Two-Level Model for Evidence Evaluation

ABSTRACT: A random effects model using two levels of hierarchical nesting has been applied to the calculation of a likelihood ratio as a solution to the problem of comparison between two sets of replicated multivariate continuous observations where it is unknown whether the sets of measurements shared a common origin. Replicate measurements from a population of such measurements allow the calculation of both within-group and between-group variances/covariances. The within-group distribution has been modelled assuming a Normal distribution, and the between-group distribution has been modelled using a kernel density estimation procedure. A graphical method of estimating the dependency structure among the variables has been used to reduce this highly multivariate problem to several problems of lower dimension. The approach was tested using a database comprising measurements of eight major elements from each of four fragments from each of 200 glass objects and found to perform well compared with previous approaches, achieving a 15.2% false-positive rate, and a 5.5% false-negative rate. The modelling was then applied to two examples of casework in which glass found at the scene of the criminal activity has been compared with that found in association with a suspect.

KEYWORDS: forensic science, evidence evaluation, likelihood ratio, graphical model, multivariate data, compositional data, glass

Consider a crime in which there was a breakage of glass. Fragments may remain at the location at which the offence took place. These fragments will be referred to as *control* fragments as their source is known. Other fragments from the same source may be transferred to the clothes and footwear of the offender. A suspect may be identified and, on subsequent examination, found to have glass fragments upon their person. These will be referred to as *recovered* fragments, as their source in not known. The purpose of the analysis of the fragments in this case is the evaluation of the evidence for comparison of the proposition that the glass associated with the suspect is from the same source as the fragments from the crime scene with the proposition that the glass associated with the suspect is not from the same source as the fragments from the crime scene.

The recovery of glass fragments from the suspect is the first stage of the examination of this kind of evidence. It is most frequently performed by shaking and/or brushing the garment(s). The debris collected is observed under an optical microscope and glass fragments separated manually. Most of these fragments have a linear dimension of <0.5 mm. Observation of morphological features, such as thickness and color, seldom contain enough information to deduce whether the fragments in question were derived from the same source as fragments found at the crime scene. Therefore, it is necessary to determine their physico-chemical properties. The Glass Refractive Index Measurements (GRIM) method is often used in the role, as are instrumental methods of elemental assay (e.g., μ-XRF, LA-ICP-MS, SEM-EDX). The comparison between recovered and control glass fragments is then made on the basis of the analytical results. The increasing ability to collect and store data relevant for identification in a forensic context has led to a corresponding increase in methods for the numerical evaluation of evidence associated with particular evidence types.

The comparison of two sets of glass fragments by numerical methods requires careful attention to the following considerations:

1. the similarity of recovered glass fragment(s) to a control sample;
2. the information about the rarity of the determined physico-chemical characteristics (e.g., elemental concentrations) for control and recovered samples in the relevant population;
3. the level of association between different characteristics where more than one characteristic has been measured; and
4. possible sources of variation that will include:

(a) variation of measurements of characteristics within the control items,
(b) variation of measurements of characteristics within the recovered items, and
(c) variation of measurements of characteristics between control and recovered items.

Commonly used significance tests, like the Student *t*-test for univariate data, and Hotelling's T^2 for multivariate data (1,2), take into account only information about within-source variation and the similarity of the compared items. Thus, the tests provide an answer to the question: *are the compared samples similar on the basis of their physico-chemical properties?* The answer to the question of interest from the forensic point of view, which is *what is the value of the evidence of these measurements in relation to the propositions that the two samples of glass fragments did, or did not, come from the same source?*, requires knowledge about the sources of variability and the rarity of the measured physico-chemical properties in the relevant population. For instance, one would expect refractive index (RI) values from different locations on the same glass object to be very similar. However, equally similar RI values could well be observed from different glass items. Without a wider context, it is not possible to ascribe meaning to the observed similarity. Therefore, inferences about the source of glass fragments made purely on the basis of similarity of measurements are incomplete. Information about the rarity of a

[1]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.
[2]Institute of Forensic Research, Westerplatte 9, PL-31-033, Krakow, Poland.

Received 6 July 2005; and in revised form 28 July 2006; accepted 1 Sept. 2006; published 30 Jan. 2007.

determined RI value has to be taken into account (3–5). Intuition suggests that the value of the evidence in support of the proposition that the recovered glass fragments and the control sample have a common origin is greater when the determined RI values are similar and rare in the relevant population, than when the RI values are equally similar but common in the same population. This kind of population information and information about the two sources of variability are taken into account in the two-level model described below (6,7).

Method

Glass Analysis

One large piece of glass from each of 200 glass objects was selected. Each of these 200 pieces was wrapped in a sheet of gray paper and further fragmented. The fragments from each piece were placed in a plastic Petri dish. Four glass fragments, of linear dimension <0.5 mm with surfaces as smooth and flat as possible, were selected for examination with the use of an SMXX Carl Zeiss (Jena, Germany) optical microscope ($\times$magnification 100).

The four selected glass fragments were placed on self-adhesive carbon tabs on an aluminum stub and then carbon coated using an SCD sputter (Bal-Tech, Balzers, Liechtenstein). The prepared stub was mounted in the sample chamber of a scanning electron microscope. Analysis of the elemental content of each glass fragment was carried out using a scanning electron microscope (JSM-5800 Jeol, Tokyo, Japan), with an energy-dispersive X-ray spectrometer (Link ISIS 300, Oxford Instruments Ltd., Witney, Oxfordshire, U.K.).

Three replicate measurements were taken from different areas on each of the four fragments, making 12 measurements from each glass object, but only four independent measurements. The four means of the measurements were used for the analysis. The measurement conditions were accelerating voltages 20 kV, life time 50 sec, magnification $\times 1000 - \times 2000$, 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 option applied a ZAF correction procedure, which takes into account corrections for the effects of difference in the atomic number (Z), absorption (A), and X-ray fluorescence (F).

The selected analytical conditions allowed the determination of all elements except Lithium (Li) and Boron (B). However, only the concentrations of oxygen (O), sodium (Na), magnesium (Mg), aluminum (Al), silicon (Si), potassium (K), calcium (Ca), and iron (Fe) are considered further in this paper as glass is essentially a silicon oxide with Na and/or Ca added to create a commonly produced glass, and potassium, magnesium, aluminum, and iron added to stabilize its structure and modify its physico-chemical properties (e.g., light transmission properties).

Data

From the eight elements measured, seven variables were derived by taking the $\log_{10}$ of each of the other elements normalized to oxygen. The data on the eight original elements are what is known as compositional data in that the sum is constrained to be 100%. Division of seven of the variables by the eighth removes this constraint. The logarithmic transformation of the ratio provides a better approximation to normality for the within-group distribution. The transformation also provides a statistic that is independent, subject to sign, as to whether oxygen is the numerator or the denominator. The normalization also effectively removes stochastic fluctuations in instrumental measurement, and is invariant to which element is used as the divisor. As all further usage of these variables in this paper refers to $\log_{10}$ of the ratio of the elemental concentration to O, then we shall denote $\log_{10}$(Na/O) as Na′, $\log_{10}$(Al/O) as Al′, $\log_{10}$(Si/O) as Si′, and so on.

TABLE 1—*Within-group variance–covariance matrix U ($\times 10^3$), for the data from the 200 glass objects.*

	Na′	Mg′	Al′	Si′	K′	Ca′	Fe′
Na′	0.182	0.167	−0.160	0.006	0.381	0.198	0.171
Mg′		43.94	−0.626	0.537	−0.573	−0.389	2.21
Al′			29.34	0.291	2.32	0.299	−0.104
Si′				1.048	2.82	1.68	0.588
K′					262.3	3.78	−31.41
Ca′						12.84	2.96
Fe′							94.35

Only the upper right triangle of the matrix is given; the lower left triangle is given by symmetry. The variances are the leading diagonal of this table, and the covariances are in the off-diagonal positions. For example, var(Na′) = 0.000182.

The two-level model requires the estimation from the background population of two variance/covariance matrices. The first is the matrix of variances and covariances for measurements on replicates within an object, and is denoted *U*. The second is the variance/covariance matrix for measurements between objects and is denoted *C*. The formulae for the estimation of these matrices are given in Appendix A. For the data from the 200 glass objects, the variance/covariance matrices are given in Tables 1 and 2.

Normal Assumptions

There is no evidence to indicate that the distribution of measurements within an object for the two-level model, following the above transformation, is anything other than multivariate normal, and this is what has been assumed from prior knowledge of the marginal distributions for these variables. It is not possible to test this assumption when there are only four measurements within a particular group. However, the same cannot be said of the between-object distributions. For this reason, a multivariate kernel density approach has been adopted for modelling the between-object distribution, and multivariate normal distributions for the within-object distribution.

Dependence Structures

A criticism of the modelling of multivariate databases, such as the glass database used in this study, is that there is a lack of

TABLE 2—*Between-group variance–covariance matrix C, for the data from the 200 glass objects.*

	Na′	Mg′	Al′	Si′	K′	Ca′	Fe′
Na′	0.0036	0.0354	−0.0048	0.0002	0.0335	0.0387	0.0111
Mg′		1.5243	−0.2258	−0.0060	−0.7537	0.6349	0.3780
Al′			0.9050	−0.0028	0.6987	−0.1517	−0.3435
Si′				0.0014	0.0023	−0.0024	0.0076
K′					2.2398	−0.5031	−0.2778
Ca′						0.8548	0.1692
Fe′							1.8682

Only the upper right triangle of the matrix is given; the lower left triangle is given by symmetry. The variances are the leading diagonal of this table, and the covariances are in the off-diagonal positions. For example, var(Na′) = 0.0036.

background data from which to estimate the parameters of the assumed distributions such as means, variances, and covariances. For example, when glass samples are described by seven variables then it is necessary to estimate, reliably, seven means, seven variances, and 21 covariances both within objects, and between objects. This requires far more analytical data than are accessible in many forensic databases, and observation of more variables, which, in applied forensic contexts, may be required, would necessitate the estimation of an exponentially larger number of means, variances, and covariances.

In the example given here, the between-object variance/covariance matrix is estimated from the means of the 200 cases for a maximum of seven variables. The means of the measurements within each object, and the within-object variance–covariance matrix is estimated from the mean of the three replicate measurements for each fragment. Thus, there are just four independent replicated sets of readings from each of the 200 glass objects.

Many approaches to the problem of multidimensionality have been proposed in forensic science. Often, it has been assumed that the variables are independent in order to reduce the number of parameters to be estimated (2). However, this assumption is seldom warranted. An alternative approach is to use some form of data reduction such as principal component analysis (3,4) but this involves problems of interpretability for a legal audience, and the loss of a certain amount of data that needs careful justification.

Calculation of a full model for the example under consideration requires the estimation of the probability density function

$$f(\mathrm{Na}', \mathrm{Mg}', \mathrm{Al}', \mathrm{Si}', \mathrm{K}', \mathrm{Ca}', \mathrm{Fe}')$$

under each of two propositions, H_p, the proposition that the glass measurements from the control and recovered material come from the same source, and H_d, the proposition that the measurements from the control and recovered material come from different sources. These density estimates are then used to calculate the likelihood ratio (LR) (see Appendix A)

$$\mathrm{LR} = \frac{f(\mathrm{Na}', \mathrm{Mg}', \mathrm{Al}', \mathrm{Si}', \mathrm{K}', \mathrm{Ca}', \mathrm{Fe}'|H_p)}{f(\mathrm{Na}', \mathrm{Mg}', \mathrm{Al}', \mathrm{Si}', \mathrm{K}', \mathrm{Ca}', \mathrm{Fe}'|H_d)} \qquad (1)$$

Unfortunately, as discussed above, the calculation of a full model is not practical as the database comprises only four independent measurements of seven variables from each of 200 glass objects. However, if the dependence structures of these data are taken into account, the problem can be reduced from one seven-dimensional problem to several problems in lower dimensions. Graph theory (8) is used to factorize the joint density function into the product of several density functions on lower dimensions. In a graph, each variable is represented as a node. In Fig. 1, the nodes are depicted as circles with the name of the variable within the circle. Nodes in a graph may, or may not be, connected by lines that are termed edges. The arrangement of nodes and edges represents the dependence structure within a graph. Variables (nodes) that are directly associated are joined by a line (edge). Variables that are conditionally independent, given the values of the other variables, are not directly connected. The dependence structure may be causal, in which case the causality may be represented by the addition of arrows to the edges. The graph is then known as a directed graph. Alternatively, the dependency model may not imply causality. The graph is then known as an undirected graph. As the relationships between the elemental concentrations in glass are not directly causal, the graphs used here will be undirected.

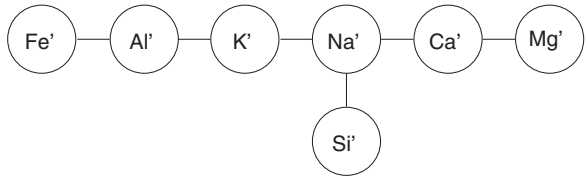

FIG. 1—*The decomposable undirected graphical model calculated from Table 3.*

The basis for the choice of edges is the partial correlation matrix, or its equivalent, the rescaled inverse of the variance–covariance matrix (8). Table 3 shows the partial correlation matrix for the seven variables from the 200 glass objects calculated from the between-object covariance matrix. Table 3 is obtained from Table 2 as follows: invert the matrix in Table 2. Scale the resultant matrix so that the diagonal terms are 1, and the off-diagonal terms are the correlation coefficients. Let a_{ij} be element from the ith row and jth column. The scaling is then implemented by setting the entry b_{ij} in the scaled matrix as

$$b_{ij} = a_{ij}/\sqrt{a_{ii}a_{jj}}$$

Thus, for $i = j$, $b_{ij} = 1$. It can be shown (8) that the elements of the scaled inverse correlation matrix are the negative partial correlation coefficients of the corresponding elements given the rest. The dependence structure between the variables can then be easily interpreted by inspection.

The graph is taken to represent a statistical model known as a graphical model. Subsets of variables, known as factors, are obtained from the graph by considering those parts of the graph in which each node is connected to each other node. For example, in Fig. 1, the nodes in the subset {Al′, Fe′} are connected to each other. The nodes in the subset {Fe′, K′} are not connected to each other. A subset in which all the nodes are connected to each other is known as a complete subgraph, and the corresponding subset of variables known as a clique. Any given graphical model can be characterized by the arrangement of its cliques.

The joint distribution f(Na′, Mg′, Al′, Si′, K′, Ca′, Fe′) was modelled using an undirected decomposable graphical model selected (9) using the covariance matrix C as a basis for the partial correlation matrix. A decomposable model is one that can be factorized in such a way that explicit formulae for the parameters may be derived.

Many automated methods for model selection rely upon a *goodness-of-fit* statistic, usually some deviance-based measure, to make a decision as to whether a model composed of a number of cliques adequately describes the full dataset. Deviance-based

TABLE 3—*Partial correlation matrix for the seven variables based on the variance–covariance matrix C for the two-level model.*

	Na′	Mg′	Al′	Si′	K′	Ca′	Fe′
Na′	1.000	0.139	0.193	0.262	−0.221	0.598	0.022
Mg′		1.000	0.008	−0.169	−0.201	0.288	0.183
Al′			1.000	−0.122	0.480	−0.113	−0.218
Si′				1.000	0.088	−0.155	0.154
K′					1.000	0.006	0.062
Ca′						1.000	−0.011
Fe′							1.000

Only the upper right triangle of the matrix is shown; the lower left triangle is given by symmetry.

statistics use an underlying assumption of normality. The method given in this paper specifically does not assume between-object normality; a deviance-based statistic is not relevant as a measure of fit, and thus cannot be used as a criterion for model selection.

Instead, the model was selected by the sequential addition of edges decided by inspection of the partial correlation matrix (Table 3). First, the partial correlation of the largest magnitude was selected. This was 0.598 between Ca′ and Na′, and an edge is added between these two nodes. Then, the partial correlation of the second largest magnitude, 0.480, was selected and an edge was added joining the corresponding nodes, K′ and Al′, to the graph. This process was repeated until all nodes were part of the model. After the addition of each edge, the model was checked to ensure that it was decomposable. Altogether, six edges were added, the final one linking Fe′ to Al′. Further additions of edges produced models that were not decomposable until so many edges were added that a model that was close to saturation, in which each node was connected to each other node, was obtained. Such models are not considered here.

The decomposable model is represented in Fig. 1, the cliques forming it being

$$(\mathrm{Al}', \mathrm{Fe}')(\mathrm{Al}', \mathrm{K}')(\mathrm{K}', \mathrm{Na}')(\mathrm{Ca}', \mathrm{Na}')(\mathrm{Ca}', \mathrm{Mg}')(\mathrm{Na}', \mathrm{Si}')$$

This model represents a minimal model that decomposes the seven variables of the full dataset into six sets of two variables. It is illustrated in Fig. 1.

Model Factorization

The factorization for an undirected decomposable graphical model is given (10) as

$$f(C_i|S_i) = \frac{f(C_i)}{f(S_i)} \tag{2}$$

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 that guarantees the factorization of the model. To find a set chain, the following algorithm may be followed:

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

The assigned ordering is a set chain. There may be many such set chains for a given graphical model; however, all will imply a single factorization for that graph (8).

For example, from Fig. 1 the node Na′ can be assigned to 1. If that is the case, then there is a choice of K′, Ca′ and Si′ for number 2, an arbitrary choice assigned 2 to Si′. There is no node linked to both Na′ and Si′, and no other nodes are linked to Si′, so there is a choice between K′and Ca′; K′ is arbitrarily assigned to 3. Both Al′ and Ca′ are connected to one already numbered node, so Al′ is arbitrarily numbered 4. Both Fe′ and Ca′ are linked to already numbered nodes, so 5 is arbitrarily assigned to Fe′. Ca′ is the only as yet unnumbered node connected to a numbered node, so is assigned to 6, leaving Mg′ to be assigned to 7.

TABLE 4—*Cliques, separators, and clique ordering for the graphical model.*

i	Clique (C_i)	Running Union (R_i)	Separator Set (S_i)
1	(Na′, Si′)	(Na′, Si′)	∅
2	(K′, Na′)	(Na′, Si′, K′)	(Na′)
3	(Al′, K′)	(Na′, Si′, K′, Al′)	(K′)
4	(Al′, Fe′)	(Na′, Si′, K′, Al′, Fe′)	(Al′)
5	(Ca′, Na′)	(Na′, Si′, K′, Al′, Fe′, Ca′)	(Na′)
6	(Ca′, Mg′)	(Na′, Si′, K′, Al′, Fe′, Ca′, Mg′)	(Ca′)

The clique ordering for those cliques suggested by the model is based upon the algorithm featured in the text. The separator sets S_i are composed of those elements of each clique that also appear in the running union along the set chain at position R_{i-1}. The running union R_i is composed of the set of all elements from each clique in the set up to and including clique C_i. ∅ denotes the empty set.

From the graphical model, the clique (Ca′, Mg′) has the highest numbered node Mg′, so the clique is given the same number as Mg′, which is 7. (Ca′, Na′) has the next highest numbered node Ca′, and so has the number 6. (Al′, Fe′) has 5 as its highest numbered node Fe′. (Al′, K′) has the highest numbered node Al′, and so has the number 4. The clique (K′, Na′) has the highest numbered node K′, and so is assigned to 3. The remaining clique (Na, Si) is numbered 2 as Si′ is the highest numbered node, and has number 2. Putting these into numerical order, the set chain in the first column of Table 4 is obtained.

Given the cliques for the model, and a suitable set chain, the sets of separators for each clique can be calculated. In the example in Table 4, the first clique in the set chain is (Na′, Si′); this is a complete subgraph, and at the moment there are no other cliques added to the graph, so there can be no separator sets. The next clique in the set chain is (K′, Na′), and so is added to the model. The intersection of elements between these two cliques is (Na′), and so this becomes the first separator set. The running union of the first two sets is now (Na′, Si′, K′). The third clique to be added to the model is (Al′, K′). The intersection between this clique and the running union is (K′), and this becomes the second separator set. Al′is then added to the running union to make it (Na′, Si′, K′, Al′). The fourth clique to be added is (Al′, Fe′). The intersection between it and the running union is (Al′), which becomes the third separator. Fe′ is now added to the running union, which becomes (Na′, Si′, K′, Al′, Fe′). The clique (Ca′, Na′) is the fifth clique to be added to the model. Its intersection with the running union is (Na′), which becomes the fourth separator. Ca′ is now added to the running union, which becomes (Na′, Si′, K′, Al′, Fe′, Ca′). Finally, the clique (Ca′, Mg′) is added to the model, the intersection between it and the running union is (Ca′), and becomes the fifth and final separator set for this model.

The above may be taken as terms for Eq. (2) and then the model can be factorized as

$$f(\mathrm{C}|\mathrm{S}) = \frac{f(\mathrm{Na}', \mathrm{Si}')f(\mathrm{K}', \mathrm{Na}')f(\mathrm{Al}', \mathrm{K}')f(\mathrm{Al}', \mathrm{Fe}')f(\mathrm{Ca}', \mathrm{Na}')f(\mathrm{Ca}', \mathrm{Mg}')}{f(\mathrm{Na}')f(\mathrm{K}')f(\mathrm{Al}')f(\mathrm{Na}')f(\mathrm{Ca}')}$$

Results

The formula for the LR for the comparison of control and recovered items is given in Appendix A. The generic notation is given there and values particular to the example discussed in the body of the text are given here.

TABLE 5—*Percentage distributions for the likelihood ratios (LR) calculated for all 200 items.*

LR	Different	Same
$\leq 10^{-6}$	72.58	2.0
10^{-6}–10^{-5}	2.50	0.0
10^{-5}–10^{-4}	2.25	1.0
10^{-4}–10^{-3}	2.16	0.5
10^{-3}–10^{-2}	2.15	0.5
10^{-2}–10^{-1}	1.49	0.5
10^{-1}–10^{0}	1.68	1.0
10^{0}–10^{1}	2.05	2.0
10^{1}–10^{2}	2.29	2.0
10^{2}–10^{3}	3.52	6.5
10^{3}–10^{4}	3.50	8.0
10^{4}–10^{5}	3.47	41.0
10^{5}–10^{6}	0.36	29.5
$\geq 10^{6}$	0.01	5.5
false +ve/ − ve	15.20	5.5

There are 200 within-source (same) comparisons and 19,900 between-source (different) comparisons. Any value in the denominator that equalled zero was set to 10^{-12}. False-positive (between-source comparison giving a value of LR >1) and false-negative (within-source comparison giving values of LR <1) rates are given.

There are 200 ($m = 200$) objects with seven ($p = 7$) variables. Each of four fragments from each object was measured three times, giving four ($n = 4$) independent replicate measurements (using the three means of the values of the measurements from each fragment) from each of the 200 glass objects. The variance–covariance matrices U and C were calculated using the formulae given in Appendix A.

The performance of the procedure was tested by evaluating the LR for comparisons of measurements from control and recovered fragments taken from the original dataset. LRs were calculated using the dependence structures described above for comparisons between each of the 200 glass objects.

The measurements chosen for the control and recovered measurements $y_{1,jk}$; $j = 1, \ldots, n_c$; $k = 1, \ldots, p$ and $y_{2,jk}$; $j = 1, \ldots, n_r$; $k = 1, \ldots, p$, respectively, were taken from the glass data. The first two replicated mean measurements, that is, the means of measurements from fragments one and two, were selected for the simulated control sample of measurements. The remainder, the means of measurements for fragments three and four, were assigned to the recovered sample. Thus, the control sample and the recovered sample each had two replicates, no measurement being common to both.

Each simulated control sample was compared with each simulated recovered sample of measurements. There were 200 instances where it was known that the two sets of measurements came from the same source, and 19,900 instances where it was known that the two sets of measurements came from different sources. Hence, 20,100 LRs were calculated. For instances where the measurements came from the same object, it is desirable for the LR to be >1. For instances where the measurements came from different objects, it is desirable for the LR to be <1. Summary information for these LRs is given in Table 5.

Application to Casework

Case 1

An elderly woman was found on the floor in a room at her home. It was determined that she had been killed by the use of a knife. Many glass fragments were found around the victim's body. The morphological features and locations of these fragments at the scene of the crime were such that it could safely be assumed that they had originated from a glass panel in an interior door. It was also assumed that this panel had been broken during the commission of the crime. Some of the fragments of glass found at the scene of crime were collected and sent for examination at the Institute of Forensic Research (IFR) for comparison purposes. Ten glass fragments (with linear dimensions <0.2 mm) were selected from these and analyzed by the SEM-EDX method. These 10 fragments will be referred to as the control material ($n_c = 10$).

Two suspects (identified as A and B) were arrested and articles of their clothing, listed in the columns of Table 6, were taken for examination. These articles included a jacket, a shirt, a sweater, a pair of trousers, and a T-shirt. Also recovered was a dishcloth lying under some glass fragments. The articles were sent to IFR for analysis. Debris from the clothing was collected in plastic Petri dishes by brushing. Each article was brushed separately. A number of glass fragments (with linear dimensions <0.2 mm) from each article were recovered during this procedure through the use of an optical microscope (with $\times 100$ magnification). Their elemental composition was determined by SEM-EDX methods using the procedure described earlier. The raw data were then transformed using a logarithmic transformation to base 10 as used for the LR approach described in this paper, and a factorization based

TABLE 6—*Values of the likelihood ratio for fragments of glass found on various pieces of clothing recovered from the home of two suspects, A and B, and a dishcloth from the crime scene.*

		Suspect A			Suspect A		
LR	Verbal Equivalent	Jacket	Shirt	Sweater	Trousers	T-Shirt	Dishcloth
$\leq 10^{0}$	Support for H_d	0	0	0	0	1*	0
10^{0}–10^{1}	Limited support H_p	0	0	0	0	0	0
10^{1} – 10^{2}	Moderate support H_p	0	0	0	0	0	0
10^{2}–10^{3}	Moderately strong support H_p	1	0	0	0	0	0
10^{3}–10^{4}	Strong support H_p	2	0	0	0	2	3
$\geq 10^{4}$	Very strong support H_p	10	3	1	3	4	6
Total fragments		13	3	1	3	7	9

*$\mathrm{LR} < 10^{-12}$.

These LRs were calculated using the model from Fig. 1. A verbal interpretation is associated with certain intervals, based on the ideas of Evett et al. (11). The prosecution (defense) proposition, H_p (H_d), is that the fragments of glass on the articles of clothing from the suspects, or dishcloth, as appropriate, were (were not) from the same source as the control, the glass panel on the victim's door.

LR, likelihood ratio.

on the model described above was used for these data. The value of the evidence in comparison with the control material was determined. The prosecution proposition, H_p, was taken to be that the fragments of glass on the articles of clothing from the suspects were from the same source as the control material: the glass panel on the victim's door. The defense proposition, H_d, was taken to be that the fragments of glass on the articles of clothing from the suspects were not from the same source as the control material: the glass panel on the victim's door. Interval values for the LRs, on a logarithmic scale, are given in Table 6. Note that each piece of glass is treated as an individual group. The value of n_r is taken as one in each case and the LR evaluated for each piece of glass. The overall value of the evidence of all the pieces of glass is then the product of the LRs evaluated for each piece of glass. There is an implicit assumption of independence among all the glass fragments. The formation of groups of fragments, for which n_r would be >1, is an exercise for the forensic scientist and is discussed elsewhere (5). The dependencies between measurements within such groups are accounted for in the expression for the LR given in Appendix A but cancel out between the numerator and denominator and so do not appear in the final expression.

In addition, nine glass fragments obtained from the brushing of a dishcloth found at the crime scene were analyzed by the SEM-EDX and LR approach. The results from these are in the rightmost column of Table 6. As the dishcloth was found lying under some glass fragments that had the morphological features of the broken glass panel from the interior door, it was thought that the fragments from the dishcloth were derived from the glass panel from the door. The values for the LR obtained for all of these fragments were larger than one, and can be assumed to be true positives.

The proposal of this paper is that the elemental content of fragments of glass can be analyzed by the SEM-EDX method and the evidential value of the results of the analysis can be analyzed by the LR method. The results obtained from the application of this approach to the fragments of glass analyzed from the dishcloth may be taken as confirmation that the approach is suitable for forensic purposes.

Case 2

A shirt and one piece of glass (control object) with a blood stain on it were delivered for examination. There were glass microtraces present on the shirt. Nine glass fragments were recovered from debris from the shirt, and the elemental concentrations were measured. Ten glass fragments were collected from the piece of glass ($n_c = 10$), and again the elemental concentrations were measured. The approach proposed in this paper was applied, calculating values of the LRs for the nine fragments individually (taking $n_r = 1$ in each case) for the propositions: H_P—the fragment came from the control object, and H_d—the fragment came from some object other than the control object. The LRs (all $\times 10^{-5}$) were

$$1.9, 2.2, 2.1, 3.4, 4.3, 1.1, 3.3, 1.5, 2.6$$

(For example, the expanded first result is 190,000.)

Thus, all provide very strong support (11), for the proposition that the fragments found on the shirt came from the control object.

Discussion

A procedure has been described for the evaluation of evidence of a multivariate nature at two levels. Multivariate data of this nature have been considered difficult to interpret and the error rates obtained by the procedure described provide encouragement that it has much to offer forensic scientists interested in an objective evaluation of their evidence where the data requirements can be met. Two examples have been given from casework to illustrate the usefulness of the approach for forensic purposes. The procedure described here has several advantages over existing methods.

1. Independence among variables is not assumed.
2. Loss of information is restricted to that unaccounted for by the graphical model.
3. It models distributions of between-group variability that are not normal, thus giving greater flexibility in the contexts in which it may be applicable.
4. It is able to model data with many more variables than can be modelled using a full dependence model, and is able to do so without making unrealistic assumptions, such as between-group normality and complete independence.

The procedure is able to model the two levels of variation inherent in many data structures considered in forensic science. These levels assess the variability between different items and the variability within items. The method can be adapted to situations where more levels of variability may be a necessary feature of the data (7).

A consideration not addressed earlier is the selection of an optimal graphical model. Here, only one decomposable model, which was not close to saturation, was found, so the question has not arisen. However, in a complex system of related variables it would be usual to find a limited set of such graphical models. The number of graphical models possible for any system of nodes is determined by the number of nodes. The minimal model is the model in which there are no edges; this corresponds to independence among all nodes. The maximal model is the model in which there exists an edge connecting each node to each other node; this corresponds to full dependence. All other possible models are intermediate between these two extremes. Were a set of models possible for these data, then considerations such as parsimony, and the availability of data from which to estimate the joint probabilities represented by the cliques may be factors in model choice. However, other considerations may be relevant to different datasets, and model selection may be made by subjective choice informed by background information drawn from the subject to which the data relate.

Another consideration is that the procedure requires the existence of population data from which the models may be constructed. These data exist for certain evidential types, such as glass, but not for all other data types of forensic interest. For those types for which the data do not exist, it is hoped that the models described here will show to forensic scientists the benefit of an objective evidence evaluation procedure that can be developed if such data were collected. Hence, it is hoped that some incentive has been provided for the collection of data to which these models can be applied.

Table 5 gives the distribution of LRs for the comparison of each of the set of 200 glass objects compared with each other; it also gives a summary of false positives and false negatives. The decision to call any individual a false negative or false positive has depended here on whether that LR is calculated from observations known to be from the same glass object, or from observations known to be from different glass objects, and whether that LR is >1, or <1, respectively. This characterization is to some extent missing the point of this analysis. The idea is not to say whether some fragment is likely to have been derived from some glass

object or not, but to make statements about to what extent, or how powerfully, the observations, in this case of elemental concentrations, lend support to the proposition that the fragment came from the glass object in question, or to the proposition that the fragment came from some other glass object from the relevant population of glass objects. The LR does not provide a direct attribution of source. It provides a measure of the strength of the evidence in support of one or other of the propositions in the case. For instance, one of the LRs from the T-shirt of suspect A in Case 1 was $<10^{-12}$. A value such as this would be regarded as very strong support that the particular fragment from which the observations were made had come from some source other than the broken window pane from the crime scene. However, this value, while providing very strong support, may not be persuasive enough, one way or the other, to affect the outcome of the case as a whole. The outcome is dependent upon other evidence and court requirements.

There also has to be an awareness of the possibility of false positives and false negatives. The existence of these is illustrated in Table 5. If the column labelled "different" (for truly different sources of the crime and recovered fragments) is examined, then 2.05% of LRs are in the 1–10 range and 2.29% are in the 10–100 range. LRs with these values given different sources offer support, erroneously, to the proposition of a common source. LRs offer support for a proposition but should not be used by an expert to ascribe a common, or different, source to the control and recovered evidence. These results of false positives and negatives emphasize the care that must be exercised with the analysis when interpreting any data of forensic significance.

Acknowledgments

The authors acknowledge the assistance of Richard Gill and Nuala Sheehan and two anonymous referees. The program MIM (mixed interaction modelling) aided in model determination and selection (http://www.hypergraph.dk/).

References

1. Koons RD, Buscaglia J. The forensic significance of glass composition and refractive index measurements. J Forensic Sci 1999;44(3):496–503.
2. Koons RD, Buscaglia J. Interpretation of glass composition measurements: the effects of match criteria on discrimination capability. J Forensic Sci 2002;47(3):505–12.
3. Curran JM, Triggs CM, Almirall JR, Buckleton JS, Walsh KAJ. The interpretation of elemental composition measurements from forensic glass evidence: I. Sci Justice 1997;37(4):241–4.
4. Curran JM, Triggs CM, Almirall JR, Buckleton JS, Walsh KAJ. The interpretation of elemental composition measurements from forensic glass evidence: II. Sci Justice 1997;37(4):245–9.
5. Curran JM, Hicks TN, Buckleton JS. Forensic interpretation of glass evidence. Boca Raton, FL: CRC Press, 2000.
6. Aitken CGG, Lucy D. Evaluation of trace evidence in the form of multivariate data. [published corrigendum appears Applied Statistics 2004; 53(4):665–666]. J Roy Stat Soc Ser C 2004;53(1):109–22.
7. 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.
8. Whittaker J. Graphical models in applied multivariate statistics. Chichester: John Wiley and Sons Inc., 1990.
9. Edwards D. Introduction to graphical modelling. Berlin: Springer, 2000.
10. Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. J R Stat Soc Ser B 1988;50(2):157–224.
11. Evett IW, Jackson G, Lambert JA, McCrossan S. The impact of the principles of evidence interpretation on the structure and content of statement. Sci Justice 2000;40:233–9.
12. Lindley DV. A problem in forensic science. Biometrika 1977;64:207–13.
13. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall, 1986.

Appendix: A Two-Level Model for Evidence Evaluation with a Kernel Density Estimate for the Between-Groups Distribution

Consider a class of m objects, or groups, which have p characteristics, or variables, $\mathbf{x}$, each of which is measured n times on a continuous scale within the groups. Denote the index of m as i so $i = \{1, 2, \ldots, m\}$, that of n as j, so $j = \{1, 2, \ldots, n\}$. There are $N = mn$ sets of measurements, and mnp measurements.

Suppose there are two sets: one of n_c, one of n_r measurements on the p characteristics and a comparison between the two sets are required. Let $\bar{\mathbf{y}}_1$ be a vector of means of the n_c measurements $\mathbf{y}_{1j}$; $j = 1, \ldots, n_c$ from the first object and $\bar{\mathbf{y}}_2$ be a vector of means of the n_r measurements $\mathbf{y}_{2j}$; $j = 1, \ldots, n_r$ from the second object. The first set is denoted as the *control* set, the set whose origin is known. This could be a set recovered at the scene of a crime. The second set is denoted the *recovered* set whose origin is unknown; this could be a set found with some association with a suspect. The two sets are indexed by l where $l = \{1, 2\}$. Note that it is a feature of the method that the variability within the two sets $\mathbf{y}_{1j}$; $j = 1, \ldots, n_c$ and $\mathbf{y}_{2j}$; $j = 1, \ldots, n_r$ does not form part of the final expression.

Each member of the class of objects $\mathbf{x}$ will have a vector of p means from the n measurements, taken from that member. Denote these as $\bar{\mathbf{x}}_i (i = 1, \ldots, m)$. Similarly, denote each vector of p measurements as $\mathbf{x}_{ij}$ so that $\bar{\mathbf{x}}_i = \sum_{j=1}^{n} \mathbf{x}_{ij}/n$.

If $\bar{Y}_1$ is the random variable from which observations $\bar{\mathbf{y}}_1$ are sampled, and $\bar{Y}_2$ is the random variable from which observations $\bar{\mathbf{y}}_2$ are sampled, then

$$\begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \Sigma \right)$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and

$$\Sigma_{11} = \frac{U}{n_c} + C,\ \Sigma_{12} = C = \Sigma_{21},\ \Sigma_{22} = \frac{U}{n_r} + C$$

U is the within-group variance, and is estimated as

$$U = \frac{S_w}{N - m} \quad \text{where} \quad S_w = \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T$$

It is assumed that this variance is constant from group to group. C is the between-groups variance, and is estimated as

$$C = \frac{S^*}{m - 1} - \frac{S_w}{n(N - m)} \quad \text{where} \quad S^* = \sum_{i=1}^{m} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T$$

Then

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

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.

Then

$$\int f(\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2|\mu) f(\mu)\, d\mu$$
$$= \int f(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2|\mu) f((n_c\bar{\mathbf{y}}_1 + n_r\bar{\mathbf{y}}_2)/(n_c + n_r)|\mu) f(\mu)\, d\mu$$

This is the multivariate analogue of the univariate example in Lindley (12).

The numerator is

$$(2\pi)^{-p/2}(2\pi)^{-p/2}\left|\frac{U}{n_c} + \frac{U}{n_r}\right|^{-1/2}(2\pi)^{-p/2}$$
$$\left|C + \frac{U}{n_c + n_r}\right|^{-1/2}\frac{1}{m}|h^2C|^{-1/2}$$
$$(2\pi)^{p/2}\left|\left(C + \frac{U}{n_c + n_r}\right)^{-1} + (h^2C)^{-1}\right|^{-1/2}$$
$$\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T\left(\frac{U}{n_c} + \frac{U}{n_r}\right)^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)\right\}$$
$$\sum_{i=1}^{m}\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_{12} - \bar{\mathbf{x}}_i)^T\left[\left(C + \frac{U}{n_c + n_r}\right)\right.\right.$$
$$\left.\left. + (h^2C)\right]^{-1}(\bar{\mathbf{y}}_{12} - \bar{\mathbf{x}}_i)\right\}$$

where $\boldsymbol{y}_{12}$, the overall mean of the control and recovered measurements, is

$$\boldsymbol{y}_{12} = \frac{n_c\bar{\mathbf{y}}_1 + n_r\bar{\mathbf{y}}_2}{n_c + n_r}$$

This can be simplified slightly so that the numerator of the LR is

$$\frac{1}{m}(2\pi)^{-p}\left|\frac{U}{n_c} + \frac{U}{n_r}\right|^{-1/2}\left|C + \frac{U}{n_c + n_r}\right|^{-1/2}|h^2C|^{-1/2}$$
$$\left|(C + \frac{U}{n_c + n_r})^{-1} + (h^2C)^{-1}\right|^{-1/2}$$
$$\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T\left(\frac{U}{n_c} + \frac{U}{n_r}\right)^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)\right\}$$
$$\sum_{i=1}^{m}\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_{12} - \bar{\mathbf{x}}_i)^T\left[\left(C + \frac{U}{n_c + n_r}\right)\right.\right.$$
$$\left.\left. + (h^2C)\right]^{-1}(\bar{\mathbf{y}}_{12} - \bar{\mathbf{x}}_i)\right\}$$

The denominator of the LR is

$$\left(\int f(\bar{\mathbf{y}}_1|\mu) f(\mu)\, d\mu\right) \times \left(\int f(\bar{\mathbf{y}}_2|\mu) f(\mu)\, d\mu\right)$$

where $\bar{Y}_1 \sim N(\mu, C + U/n_c)$and $\bar{Y}_2 \sim N(\mu, C + U/n_r)$.

The first term in the denominator is

$$\int f(\bar{\mathbf{y}}_1|\mu) f(\mu)\, d\mu$$
$$= \frac{1}{m}(2\pi)^{-p/2}\left|C + \frac{U}{n_c}\right|^{-1/2}|h^2C|^{-1/2}\left|\left(C + \frac{U}{n_c}\right)^{-1} + (h^2C)^{-1}\right|^{-1/2}$$
$$\sum_{i=1}^{m}\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{x}}_i)^T\left[\left(C + \frac{U}{n_c}\right) + (h^2C)\right]^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{x}}_i)\right\}$$

The second term in the denominator is

$$\int f(\bar{\mathbf{y}}_2|\mu) f(\mu)\, d\mu$$
$$= \frac{1}{m}(2\pi)^{-p/2}\left|C + \frac{U}{n_r}\right|^{-1/2}|h^2C|^{-1/2}\left|\left(C + \frac{U}{n_r}\right)^{-1} + (h^2C)^{-1}\right|^{-1/2}$$
$$\sum_{i=1}^{m}\exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_2 - \bar{\mathbf{x}}_i)^T\left[\left(C + \frac{U}{n_r}\right) + (h^2C)\right]^{-1}(\bar{\mathbf{y}}_2 - \bar{\mathbf{x}}_i)\right\}$$

An optimal value, h_{opt}, for the window smoothing parameter h (13) for the kernel distribution is estimated as

$$h = h_{opt} = \left(\frac{4}{2p+1}\right)^{\frac{1}{p+4}}\frac{1}{m^{\frac{1}{p+4}}}$$

For $m = 200$, $p = 2$, this equals 0.3984.

Additional information and reprint requests:
Colin G. G. Aitken, Ph.D.
School of Mathematics
The King's Buildings
The University of Edinburgh
Mayfield Road
Edinburgh EH9 3JZ
U.K.
E-mail: c.g.g.aitken@ed.ac.uk