Aitken, C.G.G. & Lucy, D. (2004) Evaluation of trace evidence in the form of multivariate data *Applied Statistics*, 53:**4**, p.109-122.

Abstract

The evaluation of measurements on characteristics of trace evidence found at a crime scene and on a suspect is an important part of forensic science. Five methods of assessments for the value of the evidence for multivariate data are described here. Two are based on significance tests and three on likelihood ratios. The likelihood ratio which compares the probability of the measurements on the evidence assuming a common source for the crime scene and suspect evidence with the probability of the measurements on the evidence assuming different sources for the crime scene and suspect evidence is a well-documented measure of the value of the evidence. One of the likelihood ratio approaches transforms the data to a univariate projection based on the first principal component. The other two versions of the likelihood ratio for multivariate data account for correlation amongst the variables and for two levels of variation, that between sources and that within sources. One version assumes between-source variability is modelled by a multivariate normal distribution; the other version models the variability with a multivariate kernel density estimate. An example is given with reference to the elemental composition of glass.

Keywords: Evaluation of evidence, forensic science, likelihood ratio, multivariate data.

1 Introduction

The increasing ability to collect and store data relevant for identification in a forensic context has led to a corresponding increase in methods for numerically evaluating evidence associated with particular crimes. Methods are described here which are applicable for multivariate random effects data and as such are applicable to a wide range of problems. Their application is illustrated here with an example of data of measurements on the elemental composition of glass. Another example of relevance to forensic science concerns the elemental analyses of consignments of illicit drugs where there may be variation in the elemental concentrations of consignments from different countries of manufacture.

Early approaches (*e.g.*, Lindley, 1977) to the evaluation of data using Bayesian ideas illustrated the theory on examples from the refractive index of glass fragments. More recently, Curran *et al* (1997a, b), and Koons and Buscaglia (1999a, 2002) have described methods for evaluating measurements of elemental concentrations. Koons and Buscaglia (1999a, 2002) describe methods for assessing the capability for discrimination of these concentrations. These methods answer a 'pre-data question: "what is the probability I would make a mistake if I carried out this matching procedure?"' (see the correspondence in Curran *et al*, 1999, and Koons and Buscaglia, 1999b). Curran *et al* (1997a, b) answer a so-called post-data question which evaluates the evidence in a particular case. Koons and Buscaglia (1999a, b) criticise the use of databases for the post-data analysis of glass measurements because the characteristics vary over location and time. They are also concerned that, as the data are multivariate, large numbers of samples are required to predict accurately probabilities for events with low frequencies of occurrence.

Koons and Buscaglia (2002) assess the capability of the elemental analysis of glass fragments to distinguish between glass fragments from different sources through the use of Type I and Type II errors. They use multiple t-tests, with modifications for unequal variances and Bonferroni corrections to allow for multiple comparisons. In their experiment there were 209 specimens. All pair-wise comparisons of the 209 specimens (21736 comparisons) for each of ten elements were considered using Student's t-tests. Every specimen represented a distinct source. A Type II error is one in which for a pair-wise comparison of specimens, known because of the experimental arrangement to come from different sources, a non-statistically significant result is obtained. If such a result were obtained in a case analysis, the specimens would be deemed, wrongly, to have come from the same source. Koons and Buscaglia (2002) report a Type II error rate of 0.009% when using a Type I error of 5% over ten variables. This provides a very good answer to the pre-data question and shows that elemental concentrations provide an excellent method for the distinction of glass from different sources. The approach could be criticised for the use of a definite cut-off point of 5% for the Type I error but this would be unduly critical for a method which is obviously so successful.

As is well-documented elsewhere (*e.g.*, Aitken, 1995) this source of criticism can be avoided through the use of the likelihood ratio. The likelihood ratio considers a particular case and answers the post-data question as to how the evidence in the particular case alters the odds in favour of a particular proposition (*e.g.*, that the defendant was in contact with the crime scene).

This paper describes multivariate methods for evaluating evidence through the likelihood ratio, using an example involving glass fragments. The data used are not from a real crime but were collected to provide appropriate data for the testing of analytical methods.

1.1 Control and recovered data

A number, n_1 (≥ 1), of replicate measurements are taken of a fragment of glass found at a crime scene which is assumed to come from a window P_1. These measurements are referred to as control data as the source, P_1, of the measurements is known. A number, n_2 $(\geq 1$ and not necessarily equal to $n_1)$, of replicate measurements are taken of a fragment of glass found on the clothing of a suspect assumed to come from a window P_2. These measurements are referred to as recovered data as they have been recovered from the suspect. The prosecution proposition, H_p, is that P_1 and P_2 are the same window. The defence proposition, H_d, is that they are not. Several elemental ratios are measured for each fragment.

1.2 Population database

Consider m groups with n members of each group, with $N = mn$. For the illustrative example, measurements were recorded on a fragment of glass from each of 62 (m) windows in one house. For each fragment, five (n) replicate measurements have been made of the concentration of several elements. Thus, there are 310 (N) fragments in total. The elemental ratios are deemed to be of relevance as the original compositional measurements were not standardised. Three (p) ratios (chosen from thirteen) which were thought, after consultation with the forensic scientist who provided the data, to be the most discriminatory, were used and these were Ca/K, Ca/Si and Ca/Fe, corresponding to variables $1, 2$ and 3 in the general notation, respectively. To ensure that the analysis is independent of which element is the numerator and which is the denominator, the natural logarithms of these ratios are used for the analysis. Also, a logarithmic transformation reduces positive skewness and the data are more likely to be normally distributed, a requirement of the method (at least for the within-group distribution). Thus, the data consist of three variables, which are measured five times on each of sixty-two panes of glass.

2 Models

Let Ω denote a population of p characteristics of items of a particular evidential type. For the illustrative example, the type is glass, the items are fragments and the characteristics are the measurements of the concentrations in three elemental ratios, log(Ca/K), log(Ca/Si) and log(Ca/Fe). The background data are measurements of these characteristics on a random sample of m members from Ω with $n(\geq 2)$ replicate measurements on each of the m members. Denote the background data as

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

with $\bar{\mathbf{x}}_\mathbf{i} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x_{ij}}$. Denote the control and recovered measurements by $\{\mathbf{y_l}\} = (\mathbf{y_{lj}}, \quad j = 1, \ldots, n_l; \quad l = 1, 2)$ where $\mathbf{y_{lj}} = (y_{lj1}, \ldots, y_{ljp})^T$. Let $\bar{\mathbf{y}}_\mathbf{l} = \frac{1}{n_l}\sum_{j=1}^{n_l} \mathbf{y_{lj}}$. Denote the individual variable means over n_l measurements as $\bar{\mathbf{y}}_\mathbf{l.k}$ for $k = (1, \ldots, p)$.

The model assumes two sources of variation, that between replicates within the same fragment (known as within-source variation) and that between fragments (known as between-source variation). It is assumed that the variation within-source is constant and normally distributed. The distribution of the variation between sources may be either normally distributed or estimated with a kernel density estimate.

Within-source: Denote the mean vector within source i by $\theta_\mathbf{i}$ and the matrix of within-source variances and covariances by U. Then, given $\theta_\mathbf{i}$ and U, the distribution of $\mathbf{X_{ij}}$ is taken to be normal:

$$(\mathbf{X_{ij}} \mid \theta_\mathbf{i}, U) \sim N(\theta_\mathbf{i}, U), \quad i = 1, \ldots, m; \quad j = 1, \ldots, n.$$

Between-source: Denote the mean vector between sources by μ and the matrix of between-source variances and covariances by C. The distribution of the $\theta_{\mathbf{i}}$, as measures of between-source variability, is initially taken to be normal:

$$(\theta_{\mathbf{i}} \mid \mu, C) \sim N(\mu, C), \quad i = 1, \ldots, m.$$

Later this distribution will be replaced by a kernel density estimate.

The distributions of the measurements $\mathbf{y_1}, \mathbf{y_2}$ on the control and recovered data, conditional on the source (crime or suspect), are also taken to be normal. The means, $\bar{\mathbf{Y}}_{\mathbf{l}}$, have normal distributions with mean $\theta_{\mathbf{l}}$ and variance-covariance matrix D_l where $D_1 = n_1^{-1}U$ and $D_2 = n_2^{-1}U$:

$$(\bar{\mathbf{Y}}_l \mid \theta_{\mathbf{l}}, D_l) \sim N(\theta_{\mathbf{l}}, D_l); \quad l = 1, 2.$$

Then, for the assumption of between-source normality,

$$(\bar{\mathbf{Y}}_l \mid \mu, C, D_l) \sim N(\mu, C + D_l); \quad l = 1, 2.$$

3 Methods

Five methods are considered for the evaluation of control and recovered data to compare propositions that the two sets of data have come from the same or from different sources. Three of these methods use likelihood ratios to compare H_p and H_d. The five methods are (a) multiple significance tests, (b)Hotelling's T^2, (c) likelihood ratio based on the ratio of a Hotelling's T^2 statistic and a univariate kernel density estimate, (d) likelihood ratio based on the ratio of the convolution of two multivariate normal densities and (e) likelihood ratio where within-group variability is assessed using a multivariate normal probability density function and between-group variability using a multivariate kernel density estimate.

The mean μ is estimated by $\bar{\mathbf{x}}$, the mean vector over all groups.

The within-group covariance matrix U is estimated from the background data $\{\mathbf{x_{ij}}\}$ by

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

where

$$S_w = \sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{x_{ij}} - \bar{\mathbf{x}}_\mathbf{i})(\mathbf{x_{ij}} - \bar{\mathbf{x}}_\mathbf{i})^\mathbf{T}.$$

The between-group covariance matrix C is estimated from the background data $\{\mathbf{x_{ij}}\}$ by

$$\hat{C} = \frac{S^*}{(m-1)} - \frac{S_w}{n(N-m)}, \tag{2}$$

where

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

The estimate s_k of the pooled within-group standard deviation for variable k is the square root of the $k-$th term on the leading diagonal of $\hat{U}$; $(s_1, \ldots, s_p)^T$ is denoted $\mathbf{s}$.

If normality is not supported by the data and the kernel procedure is used there may be a case for using robust estimators for these parameters.

3.1 Multiple significance tests

A significance test is conducted for each of the p variables, independently of the other variables, to determine if there is a significant difference between the mean of the control variable and the mean of the recovered variable. This approach is used by some forensic scientists (Koons and Buscaglia, 2002) and is colloquially known as the 'three-sigma' rule. This suggests rejection of H_p if any of the individual variable mean differences is greater than three standard errors. Account is taken of the multiple nature of the tests through the use of a Bonferroni correction, to reduce the significance level for an individual test.

Assume the within-source variability of the measurements is constant. The degrees of freedom for the two-sample $t-$test are $N - m - 2$. The means $\bar{\mathbf{y}}_{1.k}$ and $\bar{\mathbf{y}}_{2.k}$ are significantly different at the $100\alpha/p\%$ level (using the Bonferroni correction) in a two-sided test if

$$\mid \bar{\mathbf{y}}_{1.k} - \bar{\mathbf{y}}_{2.k} \mid > t_{(N-m-2)}(\frac{\alpha}{2p}) s_k \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \tag{3}$$

where $t_{(N-m-2)}(\frac{\alpha}{2p})$ is the $100(\frac{\alpha}{2p})\%$ point of the $t-$distribution with $(N - m - 2)$ degrees of freedom. Control and recovered samples are then said to be significantly different at the $100\alpha\%$ level if at least one of the individual tests is significant.

For the example of measurements on glass fragments, $N = 310$, $m = 62$, $p = 3$, $n_1 = n_2 = 3$ so for an overall 5% significance test, the critical value of t is $t_{248}(0.05/6) \simeq 2.4$. Thus, examination of absolute differences in this example leads to rejection of H_p if, for any one of the k variables, $k = 1, 2, 3$,

$$\mid \bar{\mathbf{y}}_{1.k} - \bar{\mathbf{y}}_{2.k} \mid > 2.4 s_k \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.4 s_k \sqrt{\frac{2}{3}}. \tag{4}$$

3.2 Hotelling's T^2

Hotelling's T^2 statistic is

$$T^2 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \hat{U} \right]^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \tag{5}$$

and

$$T^2 \sim \frac{(N - m - 2)p}{(N - m - p - 1)} F_{(p, N-m-p-1)}.$$

Control and recovered data are deemed to have come from a different source if the T^2 statistic is greater than some pre-determined significance level.

3.3 Ratio of Hotelling's T^2 and kernel density estimate

Curran *et al* (1997a, b) describe a likelihood ratio for the evaluation of multivariate continuous data. The numerator is the density value of the T^2 statistic given in (5). The denominator is a kernel density estimate of the distribution over a transformation to one dimension of the data from the population database.

The approach described here uses the within-group covariance matrix $\hat{U}$ estimated from the background data. Curran *et al* estimate this matrix from the control and recovered data. Estimation of U from the background data gives an increase in the degrees of freedom, replacing $n_1 + n_2$ when control and recovered data are used with $N - m$ when background data are used.

First, consider the numerator and Hotelling's T^2. Hotelling's T^2 statistic is given in (5). Following Curran *et al* (1997b), define $\hat{\mathbf{q}} = [\hat{U}]^{-1}(\bar{\mathbf{y}}_\mathbf{1} - \bar{\mathbf{y}}_\mathbf{2})$. Then

$$t_q^2 = \frac{(\bar{\mathbf{y}}_\mathbf{1} - \bar{\mathbf{y}}_\mathbf{2})^T(\bar{\mathbf{y}}_\mathbf{1} - \bar{\mathbf{y}}_\mathbf{2})}{(\frac{1}{n_1} + \frac{1}{n_2})\hat{\mathbf{q}}^T\hat{U}\hat{\mathbf{q}}}.$$

The statistic t_q^2/r, where $r = (N - m - 2)p/(N - m - p - 1)$, has an $F-$distribution with $(p, N - m - p - 1)$ degrees of freedom. The numerator of the likelihood ratio is given by

$$f_F\left(\frac{t_q^2}{r}\right) \times \frac{1}{G} \tag{6}$$

where $G = (\frac{1}{n_1}+\frac{1}{n_2})\hat{\mathbf{q}}^T\hat{U}\hat{\mathbf{q}}$ is a scaling factor and $f_F(t_q^2/r)$ is the ordinate of the probability density function of the $F-$distribution with $(p, N-m-p-1)$ degrees of freedom at the point (t_q^2/r).

The denominator is taken to be a kernel density estimate at the point $(\hat{\mathbf{q}}^T\bar{\mathbf{y}}_2)^2$.

The kernel density estimate is obtained from the background data. For the Curran kernel these are transformed to univariate scalars as $(\hat{\mathbf{q}}^T\bar{\mathbf{x}}_i; i = 1, \ldots, m)$ and then squared. Denote $(\hat{\mathbf{q}}^T\bar{\mathbf{y}}_2)^2$ as z and $(\hat{\mathbf{q}}^T\bar{\mathbf{x}}_i)^2$ as $(v_i; i = 1, \ldots, m)$.

The value of the denominator of the likelihood ratio is the value of the kernel density estimate of the probability density function at the point z given by

$$k(z) = \frac{1}{mhs_v} \sum_{i=1}^{m} K(\frac{z - v_i}{hs_v}), \tag{7}$$

where s_v is the standard deviation of the v_i, namely

$$s_v = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (v_i - \bar{v})^2}$$

and K is taken to be the standard normal probability density function

$$K(w) = \frac{1}{\sqrt{2\pi}} exp\ (-w^2/2).$$

The smoothing parameter h is estimated by the optimal value

$$\left(\frac{4}{2p+1}\right)^{\frac{1}{p+4}} m^{-\frac{1}{p+4}} \tag{8}$$

(Silverman, 1986, Scott, 1992). In the approach described here, $p = 1$ and $h = 1.06m^{-\frac{1}{5}}$.

The value of the evidence is then the ratio of (6) and (7) and the procedure is referred to as the UVK procedure.

3.4 Likelihood ratio using a multivariate random effects model and assumptions of normality

The previous approach used a univariate projection of the data. A multivariate approach is now considered. The value of the evidence $\mathbf{y_1}$ and $\mathbf{y_2}$ is the ratio of two probability density functions of the form $f(\mathbf{y_1}, \mathbf{y_2} \mid \mu, C, U)$, one for the numerator, where H_p is assumed true, and one for the denominator, where H_d is assumed true. In the numerator the source means $\theta_\mathbf{1}$ and $\theta_\mathbf{2}$ are assumed equal (to θ, say) but unknown. In the denominator it is assumed that the source means $\theta_\mathbf{1}$ and $\theta_\mathbf{2}$ are not equal.

In the numerator denote the probability density function by $f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \mu, U, C)$. It is given by

$$\int_\theta f(\mathbf{y}_1 \mid \theta, U) f(\mathbf{y}_2 \mid \theta, U) f(\theta \mid \mu, C) d\theta,$$

where the three probability density functions are multivariate normal. The integral can then be shown to be equal to

$$\begin{aligned}
&f_0(\mathbf{y}_1, \mathbf{y}_2 \mid \mu, U, C) = \\
&\quad \mid 2\pi U \mid^{-\frac{1}{2}(n_1+n_2)} \mid 2\pi C \mid^{-1/2} \mid 2\pi\{(n_1+n_2)U^{-1} + C^{-1}\}^{-1} \mid^{1/2} \\
&\quad exp\left\{-\frac{1}{2}(H_1 + H_2 + H_3)\right\} \qquad (9)
\end{aligned}$$

where

$$\begin{aligned}
H_1 &= \sum_{l=1}^{2} trace(S_l U^{-1}), \\
H_2 &= (\mathbf{y}^* - \mu)^T \left(\frac{U}{n_1+n_2} + C\right)^{-1} (\mathbf{y}^* - \mu), \\
H_3 &= (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \\
\mathbf{y}^* &= (n_1\bar{\mathbf{y}}_1 + n_2\bar{\mathbf{y}}_2)/(n_1+n_2), \\
S_l &= \sum_{j=1}^{n_l} (\mathbf{y}_{lj} - \bar{\mathbf{y}}_l)(\mathbf{y}_{lj} - \bar{\mathbf{y}}_l)^T.
\end{aligned}$$

The exponential term is a combination of three terms, H_3 which accounts for the difference $(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$ between the means of the measurements on the control and recovered items, H_2 which accounts for their rarity (as measured by the distance of the mean weighted by sample sizes from μ) and H_1 which accounts for internal variability.

In the denominator the probability density function, denoted $f_1(\mathbf{y}_1, \mathbf{y}_2 \mid \mu, U, C)$, is given by

$$\int_\theta \{f(\mathbf{y_1} \mid \theta, U) \times f(\theta \mid \mu, C)\} d\theta \times \int_\theta \{f(\mathbf{y_2} \mid \theta, U) \times f(\theta \mid \mu, C)\} d\theta, \qquad (10)$$

where $\mathbf{y_1}$ and $\mathbf{y_2}$ are taken to be independent as the data are assumed to be from different sources. The integral

$$\int_\theta \{f(\mathbf{y_1} \mid \theta, U) \times f(\theta \mid \mu, C)\} d\theta$$

can be shown to be equal to

$$
\begin{aligned}
& f(\mathbf{y}_1 \mid \mu, U, C) = \\
& \quad \mid 2\pi U \mid^{-\frac{1}{2}n_1} \mid 2\pi C \mid^{-1/2} \mid 2\pi (n_1 U^{-1} + C^{-1})^{-1} \mid^{1/2} \\
& \quad exp\left\{ -\frac{1}{2} trace(S_1 U^{-1}) - \frac{1}{2} (\bar{\mathbf{y}}_1 - \mu)^T \left(\frac{U}{n_1} + C \right)^{-1} (\bar{\mathbf{y}}_1 - \mu) \right\}
\end{aligned}
$$

with an analogous result for

$$\int_\theta \{ f(\mathbf{y_2} \mid \theta, U) \times f(\theta \mid \mu, C) \} d\theta.$$

The value of the evidence is the ratio of (9) to (10).

This is equal to the ratio of

$$\mid 2\pi \left[(n_1 + n_2) U^{-1} + C^{-1} \right]^{-1} \mid^{\frac{1}{2}} exp\{ -\frac{1}{2} (H_2 + H_3) \} \tag{11}$$

to

$$
\begin{aligned}
& \mid 2\pi C \mid^{-\frac{1}{2}} \mid 2\pi \left(n_1 U^{-1} + C^{-1} \right)^{-1} \mid^{\frac{1}{2}} \mid 2\pi \left(n_2 U^{-1} + C^{-1} \right)^{-1} \mid^{\frac{1}{2}} \\
& exp\{ -\frac{1}{2} (H_4 + H_5) \}
\end{aligned} \tag{12}
$$

where

$$
\begin{aligned}
H_4 &= (\mu - \mu^*)^T \left[(D_1 + C)^{-1} + (D_2 + C)^{-1} \right] (\mu - \mu^*), \\
H_5 &= (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2 + 2C)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2), \\
\mu^* &= \left\{ (D_1 + C)^{-1} + (D_2 + C)^{-1} \right\}^{-1} \left[(D_1 + C)^{-1} \bar{\mathbf{y}}_1 + (D_2 + C)^{-1} \bar{\mathbf{y}}_2 \right].
\end{aligned}
$$

This procedure will be referred to as the MVN procedure.

3.5 Likelihood ratio using a kernel distribution for the distribution of between-group variability

A multivariate normal distribution for θ may not always a reasonable assumption as can be seen by inspection of Figure 1. This is a set of three bivariate plots of the 62 means ($\bar{\mathbf{x}}_i$; $i = 1, \ldots, 62$) from the population database. The assumption of normality can be removed by considering a kernel density estimate for the between-group distribution. Earlier examples of the use of the kernel distribution in forensic science are Aitken (1986, 1995), Chan and Aitken (1989), Berry (1991) and Berry *et al* (1992). Given a data set, which in this case will be taken to be the group means $(\bar{\mathbf{x}}_{\mathbf{1}}, \ldots, \bar{\mathbf{x}}_{\mathbf{m}})$ of the database, the kernel density function is taken to be a multivariate normal density function, with a mean at $\bar{\mathbf{x}}_i$ and covariance matrix h^2C, and denoted by $K(\theta \mid \bar{\mathbf{x}}_i, C, h)$ where

$$K(\theta \mid \bar{\mathbf{x}}_i, C, h) = \frac{(2\pi)^{-p/2} \mid C \mid^{-1/2}}{h^p} exp\left\{-\frac{1}{2}h^{-2}(\theta - \bar{\mathbf{x}}_i)^T C^{-1}(\theta - \bar{\mathbf{x}}_i)\right\}.$$

The estimate $f(\theta \mid \bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m, C, h)$ of the overall probability density function is then

$$f(\theta \mid \bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_m, C, h) = \frac{1}{m}\sum_{i=1}^{m} K(\theta \mid \bar{\mathbf{x}}_i, C, h). \tag{13}$$

The smoothing parameter h is estimated, from (8) with $p = 3$, to be

$$\left(\frac{4}{7}\right)^{\frac{1}{7}} m^{-\frac{1}{7}} = 0.92m^{-\frac{1}{7}}.$$

The numerator of the likelihood ratio, for which H_p is assumed true, can be shown to be given by:

$$\begin{aligned} &f_0(\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2 \mid U, C) = \\ &\quad (2\pi)^{-p} \mid D_1 \mid^{-1/2} \mid D_2 \mid^{-1/2} \mid C \mid^{-1/2} \\ &\quad (mh^p)^{-1} \mid D_1^{-1} + D_2^{-1} + (h^2C)^{-1} \mid^{-1/2} \\ &\quad exp\left\{-\frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T (D_1 + D_2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)\right\} \\ &\quad \sum_{i=1}^{m} exp\left\{-\frac{1}{2}(\mathbf{y}^* - \bar{\mathbf{x}}_i)^T ((D_1^{-1} + D_2^{-1})^{-1} + (h^2C))^{-1} (\mathbf{y}^* - \bar{\mathbf{x}}_i)\right\} \end{aligned} \tag{14}$$

where

$$\mathbf{y}^* = \left(D_1^{-1} + D_2^{-1}\right)^{-1} \left(D_1^{-1}\bar{\mathbf{y}}_1 + D_2^{-1}\bar{\mathbf{y}}_2\right).$$

The denominator of the likelihood ratio, for which H_d is assumed true, can be shown to be given by:

$$\begin{aligned}
&f_1(\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2 \mid U, C) = \\
&\quad (2\pi)^{-p} \mid C \mid^{-1} (mh^p)^{-2} \\
&\quad \prod_{l=1}^{2} [\mid D_l \mid^{-1/2} \mid D_l^{-1} + (h^2 C)^{-1} \mid^{-1/2} \\
&\quad \sum_{i=1}^{m} exp\{-\frac{1}{2}(\bar{\mathbf{y}}_l - \bar{\mathbf{x}}_i)^T (D_l + h^2 C)^{-1} (\bar{\mathbf{y}}_l - \bar{\mathbf{x}}_i)\}]. \qquad (15)
\end{aligned}$$

The likelihood ratio is then the ratio of (14) to (15). This procedure will be denoted the MVK procedure.

4 Analysis

4.1 Summary statistics

The overall sample mean, $\bar{\mathbf{x}}$, used as an estimate of μ, is

$$\bar{\mathbf{x}} = (4.20, -0.75, 2.77)^T.$$

The within-group covariance matrix U is estimated by $\hat{U}$, from (1), as

$$\hat{U} = \begin{pmatrix} 1.68 \times 10^{-2} & 2.66 \times 10^{-5} & 2.21 \times 10^{-4} \\ 2.66 \times 10^{-5} & 6.53 \times 10^{-5} & 7.40 \times 10^{-6} \\ 2.21 \times 10^{-4} & 7.40 \times 10^{-6} & 1.33 \times 10^{-3} \end{pmatrix}. \qquad (16)$$

The between-group covariance matrix C is estimated by $\hat{C}$ from (2), as

$$\hat{C} = \begin{pmatrix} 7.06 \times 10^{-1} & 9.88 \times 10^{-2} & -4.63 \times 10^{-2} \\ 9.88 \times 10^{-2} & 6.21 \times 10^{-2} & 6.96 \times 10^{-3} \\ -4.63 \times 10^{-2} & 6.96 \times 10^{-3} & 1.01 \times 10^{-1} \end{pmatrix}. \qquad (17)$$

Figure 1 is a scatter diagrams of the 62 group means illustrating that an assumption of a normal distribution for the group means is not necessarily a good one. Figure 2 gives contour plots of the between-group distributions using a kernel density estimator with a normal kernel and the optimal smoothing parameter as given in (8) with $p = 3$ and $m = 62$.

Figures 1 and 2 about here.

4.2 Results

Measurements were selected from the replicate measurements for each of the 62 groups to act as control and recovered data in order to assess the performance of the methods. From each group the first three replicates were denoted as measurements on a control fragment ($n_1 = 3$), the final three replicates were denoted as measurements on a recovered fragment ($n_2 = 3$). The so-called control and recovered fragments from each group, therefore, consisted of three replicates each, with one replicate in common. All the measurements from all 62 groups were then used as a background population. Multiple significance tests and Hotelling's T^2 tests were conducted and likelihood ratios evaluated for 62 pairs of control and recovered measurements where it is known the members of each pair came from the same source, and 1891 ($\frac{1}{2} \times 62 \times 61$) sets of control and recovered measurements where it is known they did not come from the same source.

4.3 Support for propositions that fragments have the same source or different sources

For multiple $t-$tests (4) and Hotelling's T^2 test (5), H_p is not rejected (is supported) if the outcomes of the tests are not significant at the 5% level. For likelihood ratio statistics, described in Sections 3.3 to 3.5, H_p is supported if the likelihood ratio is greater than 1.

In all cases when the fragments came from the same source, the correct decision was reached. There were no false negatives. The method using multiple comparisons is very conservative with 1707 false positives. Considerable evidence is needed before H_p is rejected. The second testing procedure, based on Hotelling's T^2, has a much better performance (with 32 false positives) than the first, based on multiple comparisons, but also suffers from the general problem of the use of significance tests in the forensic context.

The approach of Curran *et al.*, the UVK approach, has 49 false positives. If the within-group covariance matrix is estimated from the control and recovered data as in Curran *et al.* (1997a, b) then there are 237 false positives. The UVK approach has the philosophical advantage of using a likelihood ratio but only considers a univariate projection of the data, and does not account for between-source variation. The performances using multivariate normal distributions (MVN with 62 false positives) or kernel density estimates (MVK with 53 false positives) are comparable to UVK, though the former is capable of providing very extreme estimates of the likelihood ratio as is shown in the next section.

4.4 Likelihood ratios

Tables 1 and 2 show the values of the likelihood ratio (within logarithmic intervals) when evidence comes from the same source (Table 1) and different sources (Table 2).

Tables 1 and 2 about here.

For the situation where control and recovered data come from the same source, the UVK and MVK procedures give likelihood ratio values in the range 1 to 10^6, values which are comparable with other approaches in the evaluation of glass evidence (Aitken, 1995). However, the MVN procedures give some values $(10^{17}, 10^{21})$ which are very large and must be treated with extreme scepticism. These values may have arisen because of the poor fit of the multivariate normal distribution to the between-group variability. Large false positive results for the MVN procedure are also given when the control and recovered data come from different sources.

5 Sensitivity analysis

Two experiments were conducted to investigate the sensitivity of the MVK procedure to changes in the position and separation of the control and recovered measurements.

The first experiment was designed to investigate how the likelihood ratio varied as the separation varied between the control and recovered sample. The separation between the control and recovered samples was taken as a proportion of the within-group standard deviation $\mathbf{s}$, for each of the 62 groups. The recovered measurements $\mathbf{y_2}$ were derived from the control measurements $\mathbf{y_1}$ using the equation

$$\mathbf{y_2} = \mathbf{y_1} + t\mathbf{s}$$

where $\mathbf{y_1}$ was taken to be equal to $\bar{\mathbf{x}}_\mathbf{i}$ for $i = 1, \ldots 62$ and t took values in the interval $(0, 2)$. Figure 3 shows how the likelihood ratio varies as t varies for each of the 62 group means.

Figure 3 about here.

As the separation between control and recovered measurements increases so the value of the likelihood ratio decreases. However, inspection of possible relationships between the distance of $\mathbf{y_1}$ from the overall mean $\bar{\mathbf{x}}$ and the likelihood ratio or between the kernel density estimate at $\mathbf{y_1}$ and the likelihood ratio showed no discernible relationship.

The second experiment was designed to investigate the relationship between the likelihood ratio, as a function of $\mathbf{y_1}$ and $\mathbf{y_2}$ and the underlying density, as a function of $\mathbf{y_1}$ where

$$\mathbf{y}_2 = \mathbf{y}_1 + t\mathbf{s}, \quad t = 0, 0.5, 1, 2. \tag{18}$$

The result was remarkable in that there was an almost perfect linear relationship between the logarithm of the likelihood ratio and the logarithm of the density (as estimated by the kernel density). This is shown to be a partial artefact of the kernel procedure and the design of the sensitivity test. However, it is still pleasing to note that the empirical results of this test agree with intuition that the value of the evidence for a given separation of the control and recovered measurements should be closely related to the underlying density of the background population.

The procedure was as follows. The point $\mathbf{y}_1$, and associated $\mathbf{y}_2$ for the four values of t, varied over the whole of the three-dimensional sample space of the three logarithmic elemental composition ratios. At each point, the density function was determined from the kernel density estimate given by (13) with θ replaced by $\mathbf{y}_1$ and the likelihood ratio was determined for the four pairs of values of $(\mathbf{y}_1, \mathbf{y}_2)$ given from the four values of t in (18). Plots of the logarithm of the likelihood ratio versus the logarithm of the density are given in Figure 4 and are, effectively, straight lines. The slopes of the lines are all -0.96 to two decimal places. The intercepts are 8.18 $(t = 0)$, 7.66 $(t = 0.5)$, 6.11 $(t = 1.0)$ and -0.12 $(t = 2.0)$.

Figure 4 about here.

From Figure 4, it appears that as the value of the underlying density increases, the value of the evidence decreases. As $\mathbf{y_1}$ and $\mathbf{y_2}$ move apart, as measured by increasing t, the value of the evidence decreases. If the slope of the line is taken to be -1 and the intercept varies with t then the relationship between the density d and the value V of the evidence can be represented as

$$log_e V \sim log_e\{g(t)d^{-1}\}$$

or

$$V \sim g(t)d^{-1},$$

where $g(t)$ is a decreasing function of t. An intuitive explanation for this result can be given by considering the relationship

$$\begin{aligned} V &= \frac{f(\mathbf{y}_1, \mathbf{y_2} \mid \mathbf{x}, H_p)}{f(\mathbf{y}_1 \mid \mathbf{x}, H_d) f(\mathbf{y}_2 \mid \mathbf{x}, H_d)} \\ &= \frac{f(\mathbf{y}_1 \mid \mathbf{y_2}, \mathbf{x}, H_p)}{f(\mathbf{y}_1 \mid \mathbf{x}, H_d)}, \end{aligned}$$

(Aitken, 1995).

Thus,

$$log(V) = log\{f(\mathbf{y}_1 \mid \mathbf{y_2}, \mathbf{x}, H_p)\} - log\{f(\mathbf{y}_1 \mid \mathbf{x}, H_d)\}. \tag{19}$$

The second term on the right-hand-side of (19) is the logarithm of the density function evaluated at $\mathbf{y_1}$. The first term on the right-hand-side of (19) can be shown to be a function of the ratio of

$$\sum_{i=1}^{m} exp\left\{-\frac{1}{2}(\mathbf{y}_1 + \frac{1}{2}t\mathbf{s} - \mathbf{x}_i)^T \left(\frac{\hat{U}}{2n} + h^2\hat{C}\right)^{-1} (\mathbf{y}_1 + \frac{1}{2}t\mathbf{s} - \mathbf{x}_i)\right\}$$

to

$$\sum_{i=1}^{m} exp\left\{-\frac{1}{2}(\mathbf{y}_1 + t\mathbf{s} - \mathbf{x}_i)^T \left(\frac{\hat{U}}{n} + h^2\hat{C}\right)^{-1} (\mathbf{y}_1 + t\mathbf{s} - \mathbf{x}_i)\right\},$$

For the example, $m = 62$, $n = 3$, $\hat{U}$ and $\hat{C}$ are as given in (16) and (17), and $h = 0.92 \times 62^{-\frac{1}{7}} = 0.51$. Thus, variation in the first term of the right-hand-side of (19) responds mainly to variation in t and explains the linear relationship illustrated in Figure 4.

6 Conclusion

In consideration of all five procedures, the MVK procedure is recommended because it models two levels of variation, it allows for a non-normal between-group distribution and the results are not extreme. However, if the distribution of the between-group means is well represented by a multivariate normal distribution then MVN may perform as well as the MVK procedure. The procedures based on significance tests suffer from the artificial nature of the chosen significance level and also, in the forensic context, from the implied shift of the burden of proof from the prosecution to the defence. The first sensitivity test shows a satisfactory relationship between the separation of the control and recovered samples and the likelihood ratio. The second sensitivity test provides a fascinating and intuitively pleasing result for the relationship between the logarithm of the likelihood ratio and the logarithm of the probability density, a result which warrants further investigation.

Computer programmes for the various probability distributions referred to in the paper are available from Dr. David Lucy, e-mail dlucy@maths.ed.ac.uk.

Table 1: Comparison of three methods of evaluation of continuous multivariate data where crime and suspect data come from the same source. The three values greater than 10^{10} are of the order 10^{17} (once) and 10^{21} (twice).

Likelihood ratio	Hotelling T^2 / Univariate kernel (6)/(7)	Normal (11) / (12)	MVK kernel (14)/15)
< 1	0	0	0
$1 - 10^2$	1	0	0
$10^2 - 10^3$	18	8	13
$10^3 - 10^4$	35	17	48
$10^4 - 10^5$	8	16	1
$10^5 - 10^6$	0	11	0
$10^6 - 10^7$	0	5	0
$10^7 - 10^8$	0	0	0
$10^8 - 10^9$	0	1	0
$10^9 - 10^{10}$	0	1	0
$> 10^{10}$	0	3	0
Total	62	62	62

Figure 1: Scatter diagrams of 62 group means of log elemental ratios.

Figure 2: Contour plots for the probability density function of group means assuming kernel density estimation.

Figure 3: Variation in likelihood ratio as control and recovered measurements move apart, using kernel density estimation for modelling between-group variation.

Figure 4: Variation in logarithm of likelihood ratio with logarithm of density using kernel density estimation for modelling between-group variation.

Table 2: Comparison of three methods of evaluation of continuous multivariate data where crime and suspect data come from different sources

Likelihood ratio	Hotelling T^2 / Univariate kernel (6)/(7)	Normal (11) / (12)	MVK kernel (14)/15)
$> 10^7$	0	1	0
$10^7 - 10^6$	0	3	0
$10^6 - 10^5$	0	8	0
$10^5 - 10^4$	0	9	0
$10^4 - 10^3$	22	11	16
$10^3 - 10^2$	10	17	19
$10^2 - 10^1$	12	5	13
$10^1 - 1$	5	8	5
$1 - 10^{-1}$	3	10	10
$10^{-1} - 10^{-2}$	5	7	9
$10^{-2} - 10^{-3}$	3	3	8
$10^{-3} - 10^{-4}$	6	6	6
$10^{-4} - 10^{-5}$	4	6	3
$< 10^{-5}$	1821	1797	1802
Total	1891	1891	1891

References

[1] Aitken,C.G.G. (1986) Statistical discriminant analysis in forensic science. *Journal of the Forensic Science Society*, **26**, 237-247.

[2] Aitken,C.G.G. (1995) *Statistics and the evaluation of evidence for forensic scientists.* John Wiley and Sons Ltd, Chichester.

[3] Berry,D.A. (1991) Inferences in forensic identification and paternity cases using hypervariable DNA sequences (with discussion). *Statist. Sci.*, **6**, 175-205.

[4] Berry,D.A., Evett,I.W. and Pinchin,R. (1992) Statistical inference in crime investigations using deoxyribonucleic profiling. *Appl. Statist.*, **41**, 499-531.

[5] Chan,K.P.S. and Aitken,C.G.G. (1989) Estimation of the Bayes' factor in a forensic science problem. *J. Statist. Computn. Simuln.*, **33**, 249-264.

[6] Curran,J.M., Triggs,C.M., Almirall,J.R., Buckleton,J.S. and Walsh,K.A.J. (1997a) The interpretation of elemental composition from forensic glass evidence: I. *Science and Justice*, **37**, 241-244.

[7] Curran,J.M., Triggs,C.M., Almirall,J.R., Buckleton,J.S. and Walsh,K.A.J. (1997b) The interpretation of elemental composition from forensic glass evidence: II. *Science and Justice*, **37**, 245-249.

[8] Curran,J.M., Buckleton,J.S. and Triggs,C.M. (1999) Commentary on Koons,RD, Buscaglia,J. The forensic significance of glass composition and refractive index measurements. J Forensic Sci 1999, 44, 496-503. *Journal of Forensic Sciences*, **44**, 1324-1325.

[9] Evett,I.W., Cage,P.E. and Aitken,C.G.G (1987) Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *Applied Statistics*, **36**, 174-180.

[10] Koons,R.D. and Buscaglia,J. (1999a) The forensic significance of glass composition and refractive index measurements. *Journal of Forensic Sciences*, **44**, 496-503.

[11] Koons,R.D. and Buscaglia,J. (1999b) Authors' response to Curran *et al* (1999) *Journal of Forensic Sciences*, **44**, 1326-1328.

[12] Koons,R.D. and Buscaglia,J. (2002) Interpretation of glass composition measurements: the effects of match criteria on discrimination capability *Journal of Forensic Sciences*, **47**, 505-512.

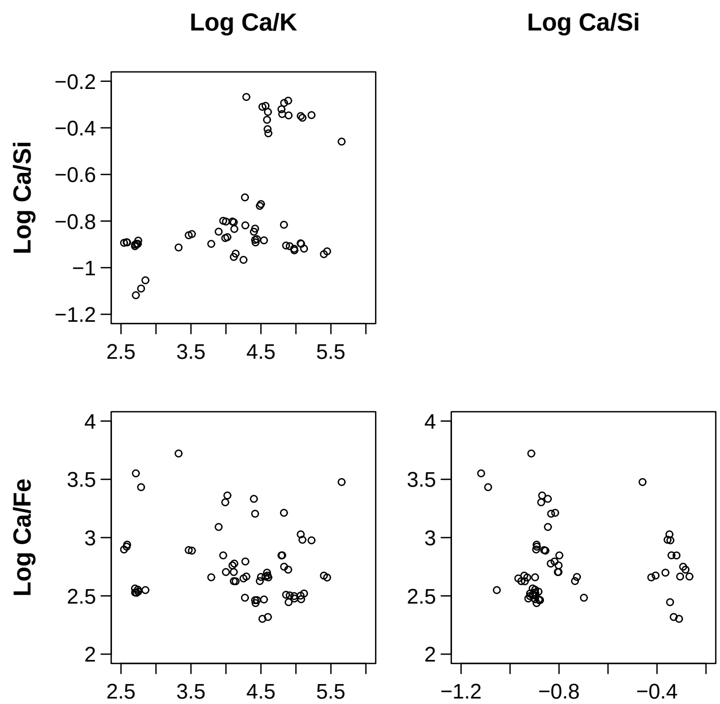
[13] Lindley,D.V. (1977) A problem in forensic science. *Biometrika*, **64**, 207-213.

[14] Scott, D.W. (1992) Multivariate density estimation. New York: John Wiley & Sons.

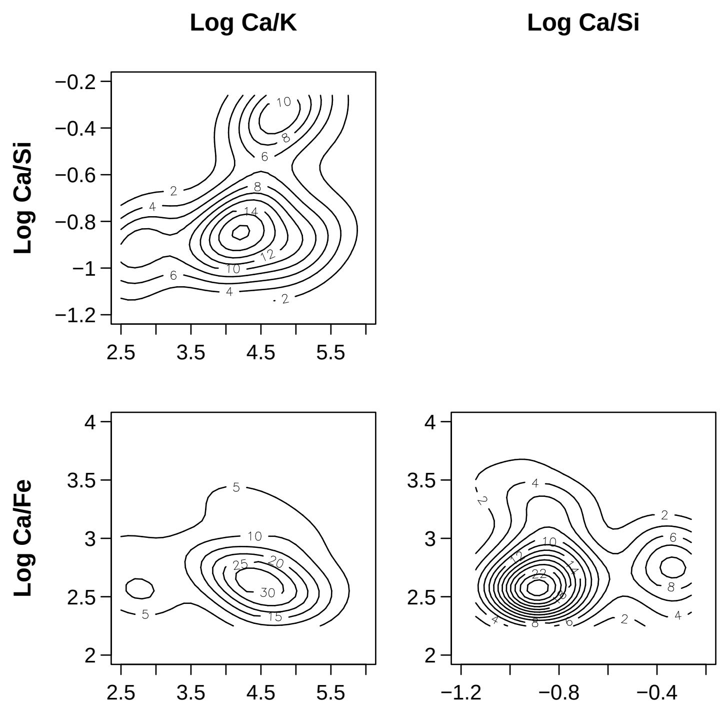
[15] Silverman, B.W. (1986) Density estimation for statistics and data analysis. London: Chapman & Hall.

Acknowledgements

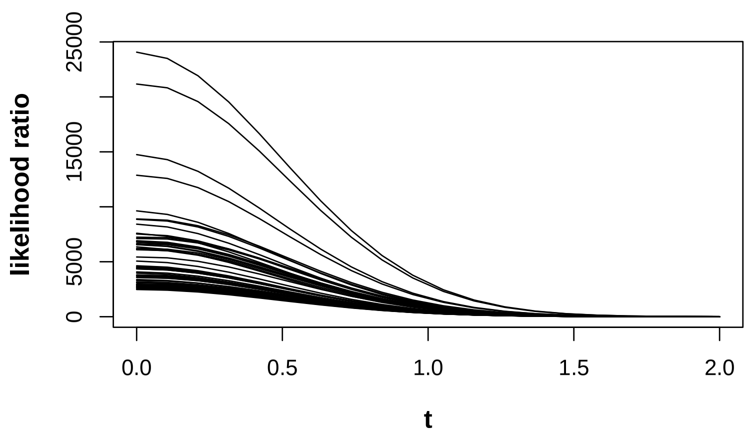
We gratefully acknowledge the assistance of Dr. JoAnn Buscaglia of the Counterterrorism & Forensic Science Research Unit, FBI Laboratory, FBI Academy, Quantico, Virginia, for the provision of the data and helpful discussions of our models and interpretation. This research was supported by a Research Development Grant from the Scottish Higher Education Funding Council. An Editor, Associate Editor and two referees made helpful comments on an earlier version of this paper.

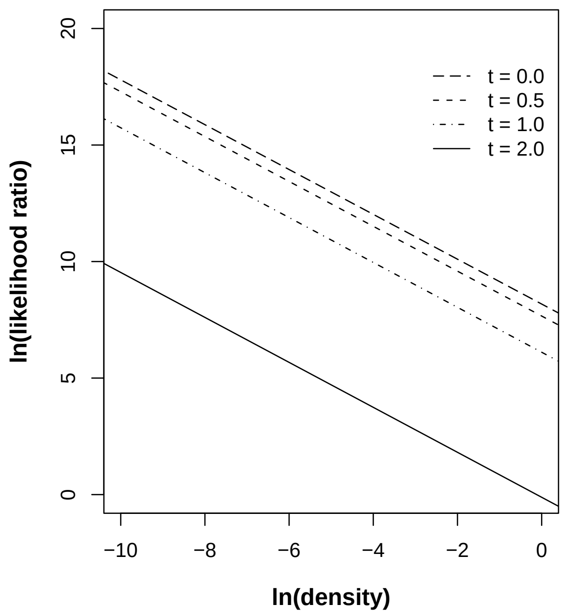


Log Ca/K
Log Ca/Si
Log Ca/Si
Log Ca/Fe
−0.2
−0.4
−0.6
−0.8
−1
−1.2
4
3.5
3
2.5
2
2.5
3.5
4.5
5.5
−1.2
−0.8
−0.4



Log Ca/K
Log Ca/Si
Log Ca/Si
Log Ca/Fe
−0.2
−0.4
−0.6
−0.8
−1
−1.2
2.5
3.5
4.5
5.5
4
3.5
3
2.5
2
−1.2
−0.8
−0.4


likelihood ratio
0
5000
15000
25000
0.0
0.5
1.0
1.5
2.0
t


t = 0.0
t = 0.5
t = 1.0
t = 2.0
ln(likelihood ratio)
ln(density)
20
15
10
5
0
−10
−8
−6
−4
−2
0