The interpretation of elemental composition measurements from forensic glass evidence: I

JM CURRAN*, CM TRIGGS*

Department of Statistics, University of Auckland, Private Bag 90219, Auckland, New Zealand
e-mail: curran@stat.auckland.ac.nz or triggs@stat.auckland.ac.nz

JR ALMIRALL

Metro-Dade Crime Laboratory, Miami, Florida 33172, USA
e-mail: almirall@fiu.edu

and

JS BUCKLETON, KAJ WALSH

ESR, Private Bag 92-021, Auckland, New Zealand
e-mail: john.buckleton@esr.cri.nz or kevan.walsh@esr.cri.nz

Science & Justice 1997; 37: 241–244

Received 19 August 1996; accepted 15 January 1997

Glass fragments are known to transfer to the clothing of a person breaking a window. These fragments may be used as evidence, associating the breaker to the crime. Recent work in the characterization of glass evidence by its elemental composition has required a framework for the evaluation of this data. Traditional treatment of the data involves determining the mean concentration and the standard deviation for each element and then comparing the means using a '3 sigma rule' and testing the match criteria with a strict range overlap for each of the elements. This paper will demonstrate a statistical test that has advantages over the '3 sigma rule' approach.

Il est connu que des fragments de verre sont transférés sur les habits d'une personne qui casse une vitre. Ces fragments peuvent être utilisés comme indices associant le casseur au crime. Des travaux récents sur la caractérisation de l'indice verre par sa composition élémentaire a nécessité la mise en place d'une structure pour l'évaluation de ces données. Le traitement traditionnel veut que l'on détermine la concentration moyenne et l'écart-type de chaque élément et que l'on compare ensuite ces moyennes en utilisant une règle de '3 sigma' et que l'on teste la correspondance par une superposition stricte des domaines pour chacun des éléments. Cet article démontre un test statistique qui présente des avantages sur l'approche faisant appel à la règle des '3 sigma'.

Es ist bekannt, daß beim Einschlagen von Fensterscheiben Glassplitter auf die Kleidung übertragen werden. Diese Splitter können als Beweismittel benutzt werden, um die betreffende Person mit der Tat in Verbindung zu bringen. Bisher erforderte die Charakterisierung von Glasspuren anhand ihrer Elementzusammensetzung eine bestimmte Vorgehensweise bei der Datenauswertung. Üblicherweise beinhaltet die Bearbeitung der Daten die Bestimmung der mittleren Konzentration jedes Elements und deren Standardabweichung. Danach werden die Mittelwerte mittels der 3 Sigma-Regel verglichen und die Vergleichskriterien mit streng definierten Grenzen für jedes Element geprüft. Die Arbeit stellt einen statistischen Test vor, der gegenüber der Anwendung der 3 Sigma-Regel vorteilhafter ist.

Se sabe que los fragmentos de vidrio se transfieren a la ropa de una persona que rompe un cristal. Estos fragmentos se pueden usar como evidencia, asociando a la persona que lo rompe con el crimen. Los trabajos recientes en la caracterización de esta evidencia de vidrio por su composición elemental, requieren un marco para la evaluación de los datos. El tratamiento tradicional de los datos incluye la determinación de la concentración media y la desviación standard para cada elemento, la comparación de medias usando una regla de 3 sigma y chequear los criterios de correlación con un rango estricto de solapamiento para cada uno de los elementos. Este trabajo demostraría que un test estadístico tiene ventajas sobre el enfoque de la regla de 3 sigma.

Key Words: Forensic science; Interpretation of evidence; Statistics; Glass; Hotelling's T^2 test.

*Corresponding authors

Introduction

Glass evidence is considered to have a great potential for associating a person who has recently broken a glass window or container to the source of the glass. The transfer and persistence of glass fragments has been described by several workers [1–6] and the understanding of these mechanisms has led to a better evaluation of the glass evidence.

The analysis of glass evidence consists of comparing the physical and chemical properties of a fragment retrieved from a suspect to a possible source of the glass and then assessing the value of that association. In the case where the fragments are sufficiently large, coincidental edges may be found or density, colour and thickness comparisons can be attempted. The typical glass transfer case, however, produces very small recovered fragments and the only analyses usually performed are refractive index (RI) and elemental composition comparisons.

Since the introduction of the float process for the manufacture of flat glass by Pilkington in the 1950s, the quality of flat glass has improved. Along with the improved manufacturing methods and optimization of formulations, the range of variation for RI over the most common flat glass has become more narrow. An example of this improved manufacturing are observations from glass plants (float and container) where the RI of the glass product is found to be indistinguishable over a period of a year or more [7,8].

Workers have found that elemental composition comparisons add discrimination potential to distinguish between glass fragments when RI does not [9–11]. The elements of interest are the minor and trace elements: Aluminium (Al), Iron (Fe), Magnesium (Mg), Manganese (Mn), Strontium (Sr), Zirconium (Zr), Calcium (Ca), Barium (Ba) and Titanium (Ti). Preliminary work on small data sets shows that trace elements, such as Sr and Zr, are present in the low part per million range and have little or no observable correlation between the other elements [7,11]. These properties make them very good discriminating 'probes'.

Traditional treatment of the data involves determining the mean concentration and the standard deviation for each element and then comparing the means using a '3 sigma rule' and testing the match criteria to determine if the ranges overlap for all of the elements. If *any* of the elements fail this test, then the fragments are considered not to match.

This paper will demonstrate a statistical test that has advantages over the '3 sigma rule' approach.

The multiple comparison problem

The elemental composition of each fragment in the control and recovered samples is determined using either Scanning Electron Microscope Energy Dispersive X-ray Fluorescence (SEM/EDAX), Wavelength Dispersive X-ray Fluorescence (WD/XRF), Inductively Coupled Plasma Mass Spectrometry (ICP/MS) or Inductively Coupled Plasma Atomic Emission Spectrometry (ICP/AES). The range of concentrations measured are from the low percent to the low parts per million, depending on the element. The mean and the standard deviation are calculated for each elemental concentration in both samples. The standard treatment of the data has been to compare the intervals defined by adding and subtracting three times the standard error (or three times sigma) of an element concentration to the mean concentration for the control and recovered samples respectively. If the intervals overlap for every discriminating element then samples are said to match. However, if any one interval does not overlap then the samples are said to not match.

This approach has two problems. The first is the problem of multiple comparisons. The '3 sigma' rule has an approximate false rejection rate of p percent, i.e. on average in one hundred tests the scientist will say the mean concentrations are different when in fact they are the same p times. This false rejection rate is much higher when small numbers of fragments are being compared. Each comparison for each element has the same rate of false rejection, however the overall rate is much larger, even if the elemental concentrations are independent. Consider the following: I have an extremely biased coin, with the probability of getting a tail equal to $\Pr(T) = p = 0.01$. The outcome of each coin toss is independent of any previous toss. If the number of tosses, n, is fixed and X is the random variable that records the number of tails observed, then X is binomially distributed with parameters n, and p. This experiment is analogous to making pairwise comparisons on n element concentrations. If $n = 10$ and the probability of a false rejection on one element is 0.01, then the probability that *at least one* false rejection will be made is 0.096 or 9.6%, so the overall false rejection rate is nearly ten times higher than the desired rate. In general if the false rejection rate, or size, of a procedure is α for a single comparison, and n comparisons are performed in total, then the overall size is approximately $n \times \alpha$. A simple solution is to increase this width of the intervals so that the size of the individual comparison is $\frac{\alpha}{n}$. This is known as the Bonferroni correction and its immediate drawback is obvious – as n, the number of elements, increases it becomes almost impossible to detect any difference between the two means.

The second problem with the '3 sigma' approach is that it fails to take into account any estimated correlation between the elements. That is, the estimated concentration of one element will be associated with the estimated concentration of another element. Failure to include this information results in severe underestimation of any joint probability calculation.

The solution to both these problems is the multivariate analogue to Student's t-test.

Hotelling's T^2 – a method for comparing two multivariate mean vectors

Student's t-test, and Welch's modification have been used and discussed extensively in the treatment of refractive index measurements [12,13]. Comparison of glass samples with respect to refractive indices involves examining the standardised distance between the two sample means. Hotelling's T^2 (named after Harold Hotelling, the first statistician to obtain the distribution of the T^2 statistics) is a multivariate analogue of the t-test, that examines the standardised squared distance between two points in n-dimensional space [14]. These two points, of course, are given by the estimated mean concentration of the discriminating elements in both samples. The appendix gives the formulae for constructing the Hotelling's T^2 statistic. Parker and Holford [15] discussed the use of Hotelling's T^2 as a method of establishing whether or not two objects, each of which possesses a number of different attributes, could have come from a common source. The authors believe that this technology should be used with elemental composition data from forensic glass evidence.

Suppose that n_r fragments have been recovered from the suspect, and n_c control fragments have been selected from a crime scene sample, and that $n_c + n_r > p + 1$, where p is the number of elements considered, then T^2 has a scaled F-distribution

$$T^2 \sim \frac{(n_c + n_r - 2)p}{(n_c + n_r - p - 1)} F_{p,n_c+n_r-p-1}$$

Use of the F-distribution depends on two assumptions about the statistical distribution of the data. Firstly, both samples come from a multivariate normal distribution, and secondly, both populations have the same covariance structure, i.e. the spread of the elemental concentration is approximately the same in each [15]. Large values of T^2 are evidence against the hypothesis of no difference between the two populations, i.e., evidence against a match [14].

Examples

The data in the following examples (Tables 1 and 2) come from two distinct sources, one brown bottle and one colourless bottle removed from different process lines at the same plant at the same time. Ten fragments were taken from each bottle and the concentrations of aluminium, calcium, barium, iron and magnesium ($p = 5$) were determined by ICP/AES.

The first example uses five fragments from the brown bottle as a control sample ($n_c = 5$) and five fragments from the same bottle as a recovered sample ($n_r = 5$) so that the population means are truly equal. Hotelling's $T^2 = 11.69$ and $F_{5,4}(0.01) = 15.52$, so

$$T^2 = 11.69 << \frac{(5+5-2)5}{(5+5-5-1)} F_{5,4}(0.01) = 10F_{5,4}(0.01) = 155.21.$$

In this example T^2 is comparatively small in relation to F. This implies that there is no evidence to reject the null hypothesis, i.e. there is no evidence to suggest that the two samples come from different sources.

The second example takes the ten fragments from the brown bottle as the control sample ($n_c = 10$) and the ten fragments from the colourless bottle as the recovered sample ($n_r = 10$), so the null hypothesis is false, i.e. the population means are truly different. Hotelling's $T^2 = 708.86$ and $F_{5,14}(0.01) = 4.69$, so

$$T^2 = 708.86 >> \frac{(10+10-2)5}{(10+10-5-1)} F_{5,4}(0.01) = \frac{90}{14} F_{5,4}(0.01) = 70.42$$

In this example T^2 is comparatively large in relation to F. This implies there is very strong evidence to reject the null hypothesis, i.e. there is very strong evidence to suggest that the two samples come from different sources.

Conclusions

Hotelling's T^2 test for the difference in two mean vectors provides a valid statistical method for the discrimination

TABLE 1 Elemental concentration data for ten fragments from a brown bottle.

Fragment	*Al*	*Ca*	*Ba*	*Fe*	*Mg*
1	0.929	7.666	0.022	0.124	0.267
2	0.859	6.838	0.018	0.112	0.227
3	0.845	5.943	0.018	0.107	0.220
4	0.931	6.424	0.020	0.117	0.262
5	0.915	6.205	0.020	0.114	0.258
6	0.832	6.842	0.021	0.122	0.209
7	0.835	6.916	0.020	0.123	0.229
8	0.941	7.590	0.023	0.128	0.270
9	0.917	7.574	0.023	0.168	0.264
10	0.798	7.472	0.019	0.122	0.202

TABLE 2 Elemental concentration data for ten fragments from a colourless bottle.

Fragment	*Al*	*Ca*	*Ba*	*Fe*	*Mg*
1	0.751	6.378	0.011	0.032	0.117
2	0.659	5.511	0.009	0.029	0.090
3	0.746	6.864	0.012	0.031	0.113
4	0.772	7.061	0.012	0.032	0.117
5	0.722	6.417	0.011	0.030	0.109
6	0.752	7.684	0.013	0.028	0.115
7	0.739	7.543	0.013	0.028	0.113
8	0.695	7.780	0.013	0.031	0.115
9	0.741	7.548	0.013	0.028	0.113
10	0.715	7.755	0.012	0.027	0.108

between two samples of glass based on elemental data. The user may decide whether or not to include a refractive index measurement comparison, but the test remains the same. The properties of the test are closer to the desired properties of the '3 sigma' rule than the rule itself. Hotelling's T^2 must be advocated in place of any algorithms based on the '3 sigma' rule or modifications of it.

In general, however, it would be desirable to abandon tests altogether and move towards a direct calculation of a likelihood ratio from continuous multivariate data.

The authors realize that while calculation of the T^2 statistic is mathematically simple, it is computationally intensive. For that purpose a small validated software package that calculates T^2, and also returns the relevant probability from the F-distribution has been provided. Versions are available for MS-DOS, MS-Windows and UNIX by email (or post) from the corresponding authors.

Appendix [14]

Suppose that $x_i = [x_{i1},\ldots,x_{ip}]^T$ are the elemental concentrations of p elements on the ith control fragment, and $y_j = [y_{j1},\ldots,y_{jp}]^T$ are the elemental concentrations of p elements on the jth recovered fragment. If n_c control fragments are to be compared with n_r recovered fragments, the matrices

$$\begin{bmatrix} x_{11} & \cdots & x_{n_c p} \\ x_{12} & \ddots & x_{n_c p} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{n_c p} \end{bmatrix}, \quad \begin{bmatrix} y_{11} & \cdots & y_{n_r p} \\ y_{12} & \ddots & y_{n_r p} \\ \vdots & \ddots & \vdots \\ y_{1p} & \cdots & y_{n_r p} \end{bmatrix}$$

represent the measurements. The summary statistics are defined by the mean vectors

$$\bar{x} = \frac{1}{n_c}\sum_{i=1}^{n_c} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n_r}\sum_{j=1}^{n_r} y_j$$

and the sample covariance matrices

$$S_X = \frac{1}{n_c - 1}\sum_{j=1}^{n_c}(x_j - \bar{x})(x_j - \bar{x})^T \text{ and } S_Y = \frac{1}{n_r - 1}\sum_{j=1}^{n_r}(y_j - \bar{y})(y_j - \bar{y})^T$$

respectively. An estimate of the common covariance matrix, Σ, is given by

$$S_{pooled} = \frac{(n_c - 1)S_X + (n_r - 1)S_Y}{n_c + n_r - 2}$$

Hotelling's T^2 is then defined by

$$T^2 = (\bar{x} - \bar{y})^T \left[\left(\frac{1}{n_c} + \frac{1}{n_r}\right) S_{pooled}\right]^{-1} (\bar{x} - \bar{y})$$

References

1. Nelson DF and Revell BC. Backward fragmentation from breaking glass. Journal of the Forensic Science Society 1967; 7: 58–61.
2. Pounds CA and Smalldon KW. The distribution of glass fragments in front of a broken window and the transfer of fragments to individuals standing nearby. Journal of the Forensic Science Society 1978; 18: 197–203.
3. Luce RJW, Buckle JL and McInnis IA. Study on the backward fragmentation of window glass and the transfer of glass fragments to individual's clothing. Journal of the Canadian Society of Forensic Science 1991; 24: 79–89.
4. Locke J and Unikowski JA. Breaking of flat glass, Part 2: effect of pane parameters on particle distribution. Forensic Science International 1992; 56: 95–106.
5. Francis K. The backward fragmentation pattern of glass, as a result of impact from a firearm projectile. Internal Report, Australian Federal Police, 1993.
6. Hicks T, Vanina R and Margot P. Science & Justice 1996; 36: 101–107.
7. Almirall JR. Characterization of glass evidence by the statistical analysis of their inductively coupled plasma/atomic emission spectroscopy and refractive index data. Proceedings from the American Academy of Forensic Sciences meeting in Nashville, TN, February 1996: 50.
8. Stoecklein W. Determination of source and characterization of glass of international origin. Presentation to the International Symposium on the Forensic Examination of Trace Evidence, San Antonio, TX, June 1996.
9. Koons R, Fiedler C and Rawalt R. Classification and discrimination of sheet and container glasses by inductively coupled plasma-atomic emission spectrometry and pattern recognition. Journal of Forensic Sciences 1988; 33: 49–67.
10. Koons R, Peters C and Rebbert P. Comparison of refractive index, energy dispersive x-ray fluorescence and inductively coupled plasma atomic emission spectrometry for forensic characterization of sheet glass fragments. Journal of Analytical Atomic Spectrometry 1991; 6: 451–456.
11. Coleman R and Goode G. Comparison of glass fragments by neutron activation analysis. Journal of Radioanalytical Chemistry 1973; 15: 367–388.
12. Evett IW and Lambert JA. The interpretation of refractive index measurements III. Forensic Science International 1982; 20: 237–245.
13. Walsh KAJ, Triggs CM and Buckleton JS. A practical example of glass interpretation. Science & Justice 1996; 36: 213–218.
14. Johnson RA and Wichern DA. Applied Multivariate Analysis. New Jersey: Prentice-Hall, Inc., 1982: 238–239.
15. Seber GAF. Multivariate Observations. New York: John Wiley & Sons, 1984: 28.