

Information-theoretical comparison of likelihood ratio methods of forensic evidence evaluation

Daniel Ramos and Joaquin Gonzalez-Rodriguez,

ATVS - Biometric Recognition Group, Universidad Autonoma de Madrid, Spain; daniel.ramos@uam.es

Grzegorz Zadora and Janina Zieba-Palus *

Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland; gzadora@ies.krakow.pl

Colin Aitken,

School of Mathematics and Joseph Bell Centre for Forensic Statistics and Legal Reasoning

University of Edinburgh, King's Buildings, Edinburgh, EH9 3JZ, UK; c.g.g.aitken@ed.ac.uk

Abstract

Forensic evidence in the form of two-level hierarchical multivariate continuous data is modelled using a likelihood ratio approach. Data are available from fragments of glass and of paint. Cross-entropy is used to compare the results with a neutral method and a method using the correct answers.

1 Introduction

Fragments of glass and of paint are common examples of transfer evidence and are frequently present at crime scenes. The recovery of paint and glass fragments from the environment of a suspect (evidence known as recovered evidence) is an important stage of the investigation of such transfer evidence. Recovered evidence is of an unknown origin; it may or may not have come from the crime scene. Fragments from paint and glass objects at a crime scene (evidence known as control evidence as its origin is known) are assumed also to be available for analysis. The value of transfer evidence is determined by comparing the probability of the evidence (E) if the control and recovered evidence have come from the same source (H_p : the evidence associated with the suspect has come from the crime scene) with the probability of the evidence if the control and recovered evidence have come from different sources (H_d : the evidence associated with the suspect has come from somewhere other than the crime scene). Most glass and paint fragments analysed by forensic experts are very small and evidence is eval-

uated by consideration of their physico-chemical features (e.g., chemical composition).

Evidence evaluation requires consideration of the variation of physico-chemical measurements within samples from the same source or group, the variation between samples from different sources or groups and the rarity of the physico-chemical measurements in the relevant population. Not all these factors are considered by significance tests. The most useful method of evaluation of evidence is that of the likelihood ratio [3]

$$LR = \frac{Pr(E | H_p)}{Pr(E | H_d)}.$$

This is the factor in the odds version of Bayes Theorem which converts prior odds in favour of H_p , prior to consideration of E , into posterior odds in favour of H_d , posterior to consideration of E . Values of LR above 1 support H_p and values of LR below 1 support H_d . A value of LR close to 1 provides little support for either proposition. Also the larger (the lower) the value of the LR , the stronger the support of E for H_p (H_d). The model which considers two levels of uncertainty, i.e., within and between group variability, in the case of multivariate data is of the form (1)

$$\begin{aligned} LR &= \frac{f(\mathbf{x}, \mathbf{y} | \mathbf{H}_p)}{f(\mathbf{x}, \mathbf{y} | \mathbf{H}_d)} \\ &= \frac{\int f(\mathbf{x}, \mathbf{y} | \mu, \mathbf{H}_p) f(\mu | \mathbf{H}_p) d\mu}{\int f(\mathbf{x} | \mu, \mathbf{H}_d) f(\mu | \mathbf{H}_d) d\mu \int f(\mathbf{y} | \mu, \mathbf{H}_d) f(\mu | \mathbf{H}_d) d\mu} \end{aligned} \quad (1)$$

In (1), E is represented by vectors of measurements (\mathbf{x}, \mathbf{y}) where \mathbf{x} are measurements on the control sample and \mathbf{y} are measurements on the recovered sample. The between-group variability is modelled by $f(\mu)$, the between-group distribution, where μ is the mean of a group and in these examples is independent of H_p and H_d . This distribution is modelled by a multivariate kernel density estimator, using data

*The physico-chemical data of paint samples were obtained within a grant supported by the State Committee for Scientific Research, Poland (Project 0T00C 013 26).

from a background database. The within-group distribution is modelled with a multivariate normal distribution. For the numerator, when the control and recovered data are from the same source, account is also taken of the correlation between \mathbf{x} and \mathbf{y} .

Many comparisons of data, taken from the background database, are made. Comparisons are from data known to come from the same (different) source(s), and hence are expected to have LR values greater (less) than 1. The mean information of a set of LR values obtained from these comparisons gives a measure of performance of the LR method. This is the main novel contribution of this work with respect to the work in [4], as the proposed information-theoretical evaluation technique not only accounts for the discrimination (false positive-false negative rate) of the computed LR values, but also for *calibration*, or deviation of the scientist's method from the probabilities obtained with knowledge of the correct answers. The performance of the LR based on (1) is compared with a neutral method ($LR = 1$ for all data) and a method which gives the correct answer (same source or different source) for all comparisons using a procedure based on *information theory* [6, 9], outlined in Section 2.4.

2 Methods

2.1 Data from glass fragments

Four glass fragments, with surfaces as smooth and flat as possible, collected from each of 270 glass objects (104 building windows, 63 car windows, 26 bulbs, 16 headlamps, 57 containers, 6 glasses) were placed on self-adhesive carbon tabs on an aluminium stub and then carbon coated using an SCD sputter (Bal-Tech, Switzerland). The means of three replicate measurements from each fragment were used for the analysis. These provided 270 groups, each with four members. The measurement conditions were accelerating voltages $20kV$, life time $50s$, magnification $1000 - 2000\times$, and the calibration element was cobalt. The SEMQuant option (part of the software LINK ISIS, Oxford Instruments Ltd, United Kingdom) was used in the process of determining the percentage of particular elements in a fragment. The selected analytical conditions allowed the determination of concentrations of oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). From the measurements on these elements, seven independent variables were derived by taking \log_{10} of the measurements on the seven other elements normalised to the oxygen measurements. There are 270 same-source comparisons and 36,315 different source comparisons.

2.2 Data from paint fragments

Py-GC/MS (TurboMass Gold, Perkin Elmer, USA) was used to analyse 20 samples, each with three members, of acrylic topcoat paints that were indistinguishable on the base of their infrared spectra and elemental composition. The GC program was: 40C held for 2 min; ramped $10 C \cdot min^{-1}$ to 300C, 300C maintained for 2 min; increase $30 C \cdot min^{-1}$ to 320C; 320 C maintained for 3 min. A RTx-35MS capillary column (30 m x 0.25 mm x 0.25 μm) was used. Pyrolysis was performed at 750C during 20 sec. without derivatisation. The \log_{10} of the ratio of the peak areas of the following seven organic compounds to the peak area of styrene were calculated: 2-hydroxyethylmethacrylate (M2E - abbreviation used in this paper); α -methylstyrene (MST); toluene (TOL); butylacrylate (BMA); 2-hydroxypropylmethacrylate (M2P); methylmethacrylate (MMA); 1,6-diisocyanatehexane (I16). There are 20 same-source comparisons and 190 different-source comparisons.

For both the glass and paint data, the normalisation effectively removes stochastic fluctuations in the instrumental measurement, and the subsequent analyses are invariant to the choice of compound as the divisor. A value of $LR > 1$ is associated with the same source of origin and a value of $LR < 1$ is associated with a different source of origin for the control and recovered data.

2.3 Two-level graphical models

A two-level model was used for the determination of the likelihood ratio (1). Values of partial correlations were used to aid the construction of a decomposable graphical model and hence to factorise the multivariate density function into the product of several density functions in lower dimensions. Such a factorisation avoids problems which may arise from an assumption that all the variables are mutually independent or from the fitting of a full model. The choice of edges is based on partial correlations obtained from the rescaled inverse of the between-source covariance matrix of the database.

2.4 Information-theoretical assessment

A measure of performance is derived based on information theory. The LR values determined by the forensic scientist (referred to hereafter as the 'scientist's' values) in the experiment described above are compared with the LR values determined from the true values of the propositions (same-source, H_p or different-source, H_d), which are referred to as the 'evaluator's' LR values. In an information-theoretical framework [6, 9], the uncertainty of a random variable (denoted H with two possible values $\{H_p, H_d\}$)

given the evidence E is measured by the conditional entropy¹, defined as:

$$U_{Pr}(H|E) = - \int \sum_{e \in \{p,d\}} f(e, H_i) \log_2 Pr(H_i|e) de. \quad (2)$$

where Pr denotes probability and f denotes probability density function (pdf). However, solution of (2) is not possible, since the integral cannot be evaluated. The problem is solved by comparing the scientist's LR values with the evaluator's LR values as follows. The evaluator knows the true states of H for each comparison. Thus, the corresponding LR values are ∞ for a same-source experiment and 0 for a different-source experiment. The corresponding posterior probabilities are 1 and 0 for H_p and H_d , respectively, using the odds form of Bayes Theorem. Also, the evaluator is *perfectly reliable*, because no error is introduced. A comparison of the scientist's LR values with the *reliable* evaluator's LR values is given by the *conditional cross-entropy*:

$$\begin{aligned} U_{\tilde{Pr}||Pr}(H|E) &= \mathcal{E}_{f(E)} \left\{ U_{\tilde{Pr}||Pr}(H|E=e) \right\} \\ &= - \sum_{i \in \{p,d\}} \tilde{Pr}(H_i) \int \tilde{f}(e|H_i) \log_2 Pr(H_i|e) de. \end{aligned} \quad (3)$$

where a tilde ($\tilde{\cdot}$) indicates the probabilities computed by the evaluator. It can be demonstrated that assigning $LR = \infty$ to all $e \in \mathbf{E}_p$ (same-source evidence values) and $LR = 0$ to all $e \in \mathbf{E}_d$ (different-source evidence values) gives

$$\begin{aligned} U_{\tilde{Pr}||Pr}(H|E) \\ \simeq - \sum_{i \in \{p,d\}} \tilde{Pr}(H_i) \frac{1}{N_i} \sum_{e_j \in \mathbf{E}_i} \log_2 Pr(H_i|e_j), \end{aligned} \quad (4)$$

where N_p is the number of same-source comparisons and N_d is the number of different-source comparisons. Thus, (4) measures the conditional cross-entropy (or simply cross-entropy) between the evaluator's LR values and the scientist's LR values. For this reason, it is proposed as the metric for evaluation of the performance of a LR estimation procedure. The cross-entropy values in (4) depend on the prior probabilities of H since

$$Pr(H_p|E) = \frac{LR \cdot \frac{Pr(H_p)}{Pr(H_d)}}{\left\{ 1 + LR \cdot \frac{Pr(H_p)}{Pr(H_d)} \right\}}. \quad (5)$$

2.4.1 Discrimination and calibration

The ability of a LR procedure to discriminate is assessed by its performance in distinguishing between (\mathbf{x}, \mathbf{y}) pairs from the same-source and from different-sources. For pairs

¹We use the notation U (uncertainty) for entropy for clarity with respect to the notation for propositions (H_p, H_d)

selected from a database, it is known whether they come from the same source or from a different source. The performance of a LR procedure may be assessed by counting the number of false assignments using LR for a given *threshold* θ . In Section 1, $\theta = 1$. In decision theory, θ is determined by $Pr(H_p)$ and the utilities involved in the decision process [10]. A pair for which the $LR > \theta$ is assigned to the same-source category; this is a false assignment if the pair have come from different sources, *i.e.*, a false positive for θ . Similarly, a pair for which the $LR < \theta$ is assigned to the different-source category; this is a false assignment if the pair have come from same source, *i.e.*, a false negative for θ . Thus, a *false positive* and *false negative* rate for every value of θ may be defined. A *discrimination performance* of a LR procedure, defined by a set of LR values, is itself defined as the relationship between false positives and false negatives for every value of the threshold θ . Typical representations of discrimination performances are ROC (Receiver Operating Characteristic Curves) or DET (Detection Error Tradeoff) curves, the former of which plots correct detection rate versus false negatives and the latter of which plots false positives versus false negatives, [8]. Two different sets of LR values have the same discrimination performance if, for every possible threshold θ in the first set of LR values, a threshold θ' can be found for the second set of LR values which presents the same false positive and false negative rates as the first set. A monotonic transformation does not change the discrimination performance of a LR procedure. On the other hand, two different LR procedures may have a very different performance in terms of cross-entropy even if they have the same discrimination performance. This is due to a different *calibration* of each LR procedure, as defined in [7]. Cross-entropy measures the mean deviation of $Pr(H|E)$, from (5), with respect to $\tilde{Pr}(H|E)$, which is the *perfectly reliable* evaluator's posterior distribution. Calibration is used to compare two or more LR procedures; it measures the deviation of $Pr(H|E)$ with respect to $\tilde{Pr}(H|E)$ for LR procedures for which the discrimination is the same.

The following is an example which illustrates the concepts of discrimination and calibration. Suppose there are two different (\mathbf{x}, \mathbf{y}) comparisons, e_s and e_d , which denote same-source and different-source comparisons respectively. Two different sets of LR values for e_s and e_d are computed. The first set, $\mathbf{\Lambda}_1$, has $LR = 5$ for e_s and $LR = 0.5$ for e_d . The second set, $\mathbf{\Lambda}_2$, has $LR = 5000$ for e_s and $LR = 500$ for e_d . For any threshold in $\mathbf{\Lambda}_1$ (*e.g.*, $\theta_1 = 2$), there is another threshold in $\mathbf{\Lambda}_2$ which will lead to the same false positive and false negative rate (*e.g.*, $\theta_2 = 2000$). Therefore, both sets have the same discrimination performance². However, using $\mathbf{\Lambda}_2$ the fact finder will infer posterior probabilities which will lead to erroneous conclusions. For in-

²It can be demonstrated that $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ will have the same ROC curve.

stance, for e_d the LR value in Λ_1 is 0.5, and the LR value in Λ_2 is 500. For a prior probability of $Pr(H_p) = 0.5$, Λ_1 will lead to a posterior probability of $Pr(H_p|E) = 0.333$, from (5), which correctly suggests the evidence supports H_d for a different-source comparison. However, Λ_2 will lead to a posterior probability $Pr(H_p|E) = 0.998$, from (5), which provides misleading evidence in favour of H_p for a different-source comparison. Therefore, even when Λ_1 and Λ_2 have the same discrimination performance, the lack of calibration in the LR values of Λ_2 will mislead the fact finder.

It is desirable to find a monotonic transformation which minimizes the value of the cross-entropy for a given LR procedure. This will provide a new set of transformed LR values which will be optimal in terms of cross-entropy constrained to not changing the discrimination performance of the original set of scientist's LR values. This new set of LR values will be *calibrated* in the sense that the cross-entropy cannot be reduced for the given discrimination performance. This new set of transformed LR values is called the *calibrated LR* set. Isotonic regression using the *Pool Adjacent Violators (PAV)* algorithm [5] is used to obtain the monotonic transformation which leads to the calibrated LR values. Details about the *PAV* algorithm may be found in [5, 1]. However, the forensic scientist cannot use *PAV* in order to obtain the calibrated LR set, because *PAV* requires the true values of H , which are unknown by the forensic scientist.

Calibration can be assessed with cross-entropy, because it measures the loss of information due to deviating from the evaluator's and calibrated states of knowledge. This is not possible with discrimination measures like ROC curves or average probability of error (average between false positives and false negatives).

2.4.2 Graphical representation of cross-entropy: the information plot

The measurements of discrimination and calibration described above allow cross-entropy to be represented as a function of $Pr(H_p)$ in a so-called *information plot*. Figure 1 shows several examples of information plots for the glass example. The cross-entropy is represented for three cases. First, the solid curve is the cross-entropy of the scientist's LR values. Second, the dashed curve is the cross-entropy of the calibrated LR values obtained with the *PAV* algorithm [5]. Finally, the dotted curve is the cross-entropy of a *reference LR* set, defined as the LR set which is neutral, *i.e.*, the LR is always 1. The reference curve does not change between LR procedures.

As can be seen in [6], the information shared by the propositions and the evidence is the difference between the dotted curve and the solid curve. Thus, as the calibration

of the system degrades, the cross-entropy (solid curve) increases, and the information about the propositions that is extracted from the evidence is lower. Moreover, the dashed curve shows the potential information that the system is able to deliver if it is properly calibrated and also gives an idea of the loss of information due to a lack of discrimination, because the calibration is optimal. Thus, the difference between the dashed and the solid curves at a given prior will be the information loss due to the lack of calibration.

3 Results

3.1 Glass and paint

The rescaled inverses of the variance-covariance matrices for the seven variables from the 270 glass objects and from the 20 paint samples are presented in Table 1. (Only the upper right triangles of the matrices are given, the lower left triangles are given by symmetry.)

Table 1. The partial correlation matrix for the seven variables of glass (upper table) and of paint (bottom table).

	Na'	Mg'	Al'	Si'	K'	Ca'	Fe'
Na'	1.00	0.10	0.17	0.27	-0.28	0.59	0.05
Mg'		1.00	0.05	0.09	-0.12	0.29	0.14
Al'			1.00	-0.19	0.43	-0.09	-0.20
Si'				1.00	0.14	-0.20	0.09
K'					1.00	0.03	0.05
Ca'						1.00	-0.02
Fe'							1.00
	MMA	TOL	BMA	MST	M2E	M2P	I16
MMA	1.00	-0.33	-0.49	0.15	0.13	0.69	-0.40
TOL		1.00	-0.37	0.20	0.09	0.05	-0.22
BMA			1.00	0.00	0.49	0.77	-0.19
MST				1.00	0.12	0.02	0.28
M2E					1.00	-0.72	0.33
M2P						1.00	0.53
I16							1.00

The models are obtained as follows. For glass, the partial correlation of largest magnitude is 0.59, between Ca' (*i.e.*, $(\log_{10}(Ca/O))$) and Na' (*i.e.*, $(\log_{10}(Na/O))$), and an edge is placed between these two nodes. The partial correlation of next largest magnitude is 0.43, and an edge is added to the graph joining the corresponding nodes, K' and Al'. This process is repeated until all nodes are part of the model. After the addition of each edge the model is checked to ensure it is decomposable in the sense that it can be factorised in such a way that explicit formulae for the parameters in the model may be derived. The further addition of edges to the graph produced a model that was not decomposable. The final graph is shown in the left-hand diagram of Figure 2. The corresponding LR is given in (6). A similar procedure was followed for paint and the result is shown in the

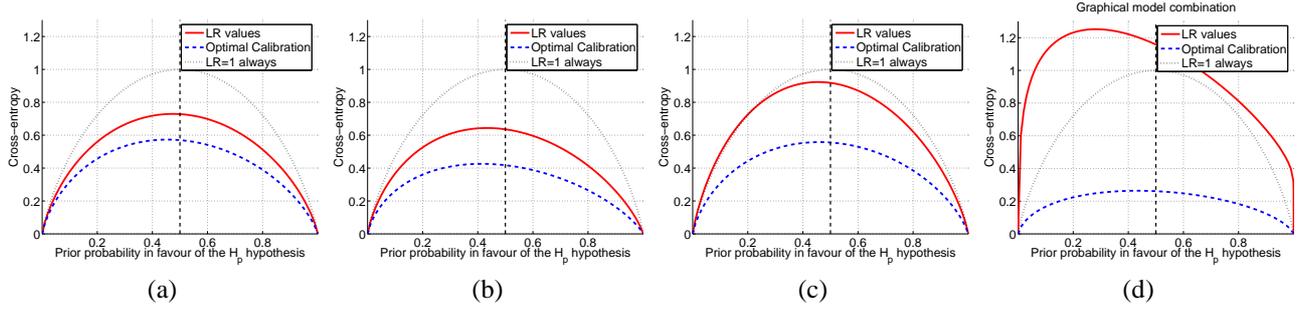
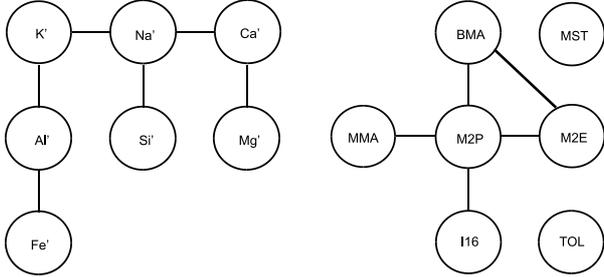


Figure 1. Information plots for glass evidence. Examples of three sets of bivariate LR values (a) (Na', Si'), (b) (K', Na') and (c) (Ca', Na') and of (d) the result for the LR for the graphical model.

right-hand diagram in Figure 2. The corresponding LR is given in (7). Notice that two nodes (those for MST and for TOL) are not connected to any other nodes and hence have univariate likelihood ratios in (7). General rules for the construction of such undirected graphs and models are given in [4].

Figure 2. The decomposable undirected graphical model for the glass fragments (left graph) and paint fragments (right graph) calculated from Table 1.



The following likelihood ratio (LR_{glass}) was obtained for the glass data:

$$\frac{LR_{Na', Si'} LR_{K', Na'} LR_{Al', K'} LR_{Al', Fe'} LR_{Ca', Na'} LR_{Ca', Mg'}}{(LR_{Na'})^2 LR_{K'} LR_{Al'} LR_{Ca'}} \quad (6)$$

where, for example, $LR_{Na', Si'}$ indicates the likelihood ratio in (1) for the bivariate data set $(\log_{10}(Na/O), \log_{10}(Si/O))$.

The following likelihood ratio (LR_{paint}) was obtained for the paint data.

$$\frac{LR_{M2E, M2P, BMA} LR_{MMA, M2P} LR_{I16, M2P} LR_{MST} LR_{TOL}}{(LR_{M2P})^2} \quad (7)$$

where, for example, $LR_{MMA, M2P}$ indicates the likelihood ratio in (1) for the bivariate data set of the \log_{10} of the peak area of M2P to the peak area of styrene and of the \log_{10} of the peak area of MMA to the peak area of styrene.

3.2 Information-theoretical analysis

In this section, information plots are presented in order to show the information loss of the LR values computed by the forensic scientist due to discrimination and calibration. In Figures 1a, 1b and 1c the performances of three of the bivariate LR values computed for the glass model are shown as an example. The cross-entropy of the *calibrated* LR values (dashed curve) is fairly similar (ranging from 0.4 to 0.6 for $P(H_p) = 0.5$, except for a set of variables (Ca', Na') where its value reaches 0.9). As was stated in Section 2.4.2, that means that the loss in information due to discrimination is similar for all univariate and bivariate LR procedures. On the other hand, the cross-entropy of the scientist's LR values (solid curve) varies significantly between univariate and bivariate models. Moreover, for some bivariate variables like (Ca', Na') (Figure 1c) the value of the cross-entropy is very high and comparable to the neutral LR set ($LR = 1$ always). Despite the discrimination capabilities of (Ca', Na'), the mean information of the scientist's LR values for (Ca', Na') will be almost the same as the information of a neutral expert ($LR = 1$ always).

Figure 1d shows the performance of the proposed LR computation technique based on the full graphical model (Figure 2 - left graph). It shows that the cross-entropy of the calibrated LR values (dashed curve) is better for the graphical model than for the univariate and bivariate models (cross-entropy of 0.23 for the graphical model for $Pr(H_p) = 0.5$). Thus, LR values computed with the proposed graphical model present a much lower discrimination loss than the univariate and bivariate LR values. However, the cross-entropy of the scientist's LR values is quite high because of miscalibration, which limits the information delivered by the scientist's LR values. This may be due to the effect of some of the bivariate LR values which have high cross-entropy, like (Ca', Na') (Figure 1c). The difference in cross-entropy between the scientist's LR values and the calibrated LR values indicates that while the scientist's LR

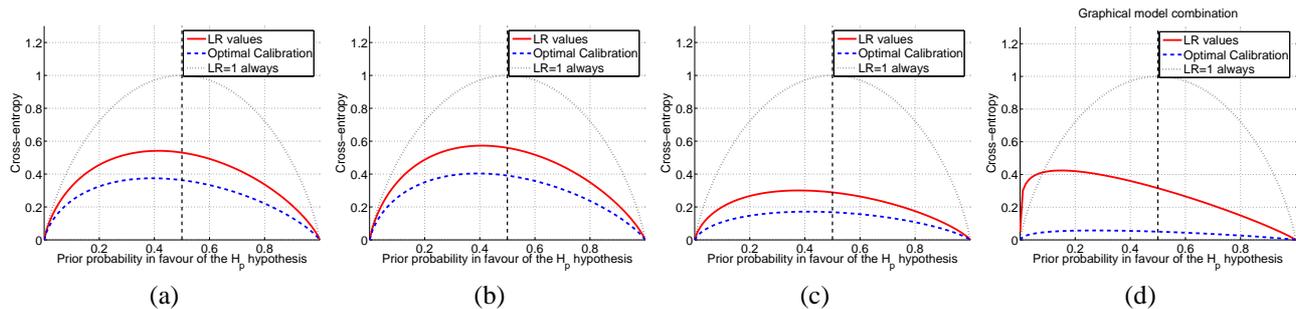


Figure 3. Information plots for paint evidence. Examples of two sets of univariate and one set of bivariate LR values (a) TOL, (b) MST and (c) (MMA, M2P) and of (d) the result for the LR for the graphical model.

values provide good discrimination [2] there is still a considerable room for improvement before they approach their calibrated values.

The results of a similar analysis for three univariate LR values for paint evidence are shown in Figures 3a, 3b and 3c. The performance of the calibrated LR values (dashed curve) in this case is excellent (ranging from 0.1 to 0.4 for $P(H_p) = 0.5$). Moreover, the difference between the cross-entropy of the calibrated LR values and the scientist's LR values is much lower than for the glass evidence case, which means an amount of information delivered by the scientist's LR values closer to the optimal calibration case. Also, the information plot for the graphical model for paint evidence is shown in Figure 3d. Again, the loss of information due to discrimination (dashed curve) is excellent (cross-entropy of 0.05 for $P(H_p) = 0.5$). Moreover the performance of the forensic scientist's LR set (solid curve) is fairly good, compared with the neutral LR value of 1, which means good delivery of information. As the dashed and solid curves are still separated, an improvement of the calibration would lead to an even better provision of information.

4 Conclusions

High-quality discrimination in the form of correct identification of evidence from the same source and of evidence from different sources is important. Calibration provides a measure of the quality of the discrimination process. Calibration measures the deviation of the results determined by the forensic scientist from the optimal values for given discrimination results. The deviation may increase the value of the cross-entropy dramatically. This corresponds to a reduction in the amount of information delivered to the fact-finder. The scientist can testify to the value of their evidence by quoting a likelihood ratio value obtained from a particular procedure. They can

testify to the quality of the procedure by giving the results of the assessment of tests on appropriate databases using the ideas of entropy. Future work includes the use of the proposed evaluation technique with LR values from different forensic disciplines and its comparison to other information theoretical measures of goodness.

References

- [1] R. K. Ahuja and J. B. Orlin. A fast scaling algorithm for minimizing separable convex functions subject to chain constraints. *Operations Research*, 49:784–789, 2001.
- [2] C. G. G. Aitken and D. Lucy. Evaluation of trace evidence in the form of multivariate data. *Applied Statistics*, 53, 109–122; with corrigendum 665–666, 2004.
- [3] C. G. G. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 2004.
- [4] C. G. G. Aitken, G. Zadora, and D. Lucy. A two-level model for evidence evaluation. *Journal of Forensic Sciences*, 52, 412–419, 2007.
- [5] N. Brümmer and J. du Preez. Application independent evaluation of speaker detection. *Computer Speech and Language*, 20(2-3):230–275, 2006.
- [6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd ed. Wiley Interscience, 2006.
- [7] M. H. deGroot and S. E. Fienberg. The comparison and evaluation of forecasters. *The Statistician*, 32:12–22, 1982.
- [8] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET curve in assessment of decision task performance. In *Proc. of ESCA Eur. Conf. on Speech Comm. and Tech., EuroSpeech*, pages 1895–1898, 1997.
- [9] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 623–656, 1948.
- [10] F. Taroni, S. Bozza, and C. G. G. Aitken. Decision analysis in forensic science. *Journal of Forensic Sciences*, 50(4):894–905, 2005.