

Evaluation of evidence in forensic science: a multivariate Bayesian approach ^(*)

La valutazione dell'evidenza in ambito forense: un approccio bayesiano multivariato

Silvia Bozza¹, Franco Taroni², Raymond Marquis², Matthieu Schmittbuhl³

¹ Dipartimento di Statistica, Università di Venezia
e-mail: silvia.bozza@unive.it

² School of Criminal Justice, University of Lausanne

³ Team Research EA 3428, Dental Faculty, Louis Pasteur University, Strasbourg

Riassunto: Le perizie calligrafiche su manoscritti provenienti da fonte anonima e da un sospetto costituiscono un problema aperto in ambito forense. Le valutazioni dipendono fortemente dalle analisi condotte dagli esperti, i quali valutano le caratteristiche calligrafiche in modo qualitativo e soggettivo. Una individualizzazione precisa della forma dei caratteri manoscritti è possibile attraverso una tecnica basata sull'analisi di Fourier: la variabilità dei caratteri può essere studiata così in modo quantitativo. La valutazione dell'evidenza è condotta attraverso il calcolo del rapporto di verosimiglianza per dati multivariati. La metodologia individuata mette l'esperto forense nelle condizioni di considerare la dipendenza tra variabili e due distinti livelli di variabilità, all'interno di ciascun scrittore e tra gruppi di scrittori.

Keywords: evaluation of evidence, forensic science, likelihood ratio, multivariate data, Bayes' theorem, handwriting evidence.

1. Introduction

The individualization of handwriting is largely dependent on analysis by examiners, who evaluate the characteristics of the writing in a qualitative or subjective way. Various studies have already been undertaken to reduce or eliminate the subjective part of the handwriting analysis process. Letter shape can be studied in a global and precise way following the methodology based on Fourier analysis proposed by Marquis *et al.* (2005a). In this work, Fourier descriptors are used to study the variability of handwritten characters. In particular, each contour can be characterized by a series of harmonics; each harmonic is defined by a frequency, an amplitude and a phase. The amplitude and the phase are called the Fourier descriptors. The amplitude of a harmonic represents the relative importance of its contribution to the original shape of the contour; the phase represents the orientation of the harmonic contribution.

The assessment of the value of a scientific evidence is performed through the derivation of a likelihood ratio; its use in forensic science is widely discussed in Aitken and Taroni (2004). A method for the evaluation of evidence through the derivation of an appropriate likelihood ratio for multivariate data has been developed by Aitken and Lucy (2004) in the context of elemental composition of glass fragments.

(*) This work was supported by the Swiss National Science Foundation (grant no. 100011-105291)

The paper is organized as follows. The population database and the models adopted are introduced in Section 2. Section 3 illustrates the evaluation of the evidence through the derivation of a likelihood ratio for multivariate data and finally Section 4 presents the results and concludes the paper.

2. Population database and models

Extensive handwriting of several writers showing closed loops for their characters a , d , o and q , were collected and analyzed (Marquis *et al.*, 2005b). Global shape of the retained letters can be described by Fourier descriptors obtained through the image analysis procedure presented in Marquis *et al.* (2005a). In particular, the size and the shape of each letter can be described through p variables representing respectively the surface and the amplitude and phase of the first four harmonics.

The background data contains the measurements of these characteristics on a sample of $m = 13$ writers, with n_i measurements on each writer, $i = 1, \dots, m$. The background data is denoted as $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})'$, $i = 1, \dots, m$, $j = 1, \dots, n_i$, with $\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$.

Two sources of variation are considered: the within-source variation (between replicates within the same writer) and the between-source variation (between writers). In a first instance, it is assumed that the variation within writers is constant and normally distributed.

Let us denote the mean vector within source i by $\boldsymbol{\theta}_i$ and the matrix of within-source variances and covariances by W . Then, the distribution of $\mathbf{X}_{ij}$ is taken to be normal with $(\mathbf{X}_{ij} | \boldsymbol{\theta}_i, W) \sim N(\boldsymbol{\theta}_i, W)$. For the between-source variation, let us denote the mean vector between sources by $\boldsymbol{\mu}$ and the matrix of between-source variances and covariances by B . The distribution of the $\boldsymbol{\theta}_i$ is taken to be normal with $(\boldsymbol{\theta}_i | \boldsymbol{\mu}, B) \sim N(\boldsymbol{\mu}, B)$, $i = 1, \dots, m$. The overall mean $\boldsymbol{\mu}$, the within- and between-groups covariance matrices W and B are estimated from the background data.

If the variation within writers is assumed to be not constant, an Inverted Wishart distribution is introduced to model the within-writers covariance matrix, say $W_i \sim IW(U, n_w)$.

3. Likelihood ratio for authorship

The likelihood ratio considers a particular case and answers the post-data question about how the evidence in the particular case alters the odds in favor of a particular proposition. For the sake of illustration, let us consider the following two propositions of interest:

H_p : The suspect is the author of the manuscript.

H_d : The suspect is not the author of the manuscript: a random person wrote the manuscript.

An anonymous letter is available for comparative analysis. A suspect is apprehended; written material from this suspect is selected and will be analyzed for comparative purposes. A number $n_1 (\geq 1)$ of measurements, along the manuscript, (on a specific letter, let us say a), are performed from the anonymous manuscript. These measurements are referred to as recovered data. A number $n_2 (\geq 1$ and not necessarily equal to $n_1)$ of measurements are obtained from manuscripts written by the suspect. These measurements

are referred to as control data. Let us denote the control and the recovered replicate measurements by $\{\mathbf{y}_l\} = (\mathbf{y}_{lj}, j = 1, \dots, n_l, l = 1, 2)$, where $\mathbf{y}_{lj} = (\mathbf{y}_{lj1}, \dots, \mathbf{y}_{ljp})'$.

If the variability within writers is assumed to be constant, then the likelihood ratio can be computed straightforwardly. The value of the evidence $\mathbf{y}_1$ and $\mathbf{y}_2$ is the ratio of two probability density functions of the form $f(\mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\mu}, W, B)$: one for the numerator, where H_p is assumed true, and one for the denominator, where H_d is assumed to be true. In the numerator, the source means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are assumed to be equal (to $\boldsymbol{\theta}$, say) but unknown. In the denominator, it is assumed that the source means $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are not equal. In the numerator the probability density function is given by

$$f_p(\mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\mu}, W, B) = \int_{\boldsymbol{\theta}} f(\mathbf{y}_1 | \boldsymbol{\theta}, W) f(\mathbf{y}_2 | \boldsymbol{\theta}, W) f(\boldsymbol{\theta} | \boldsymbol{\mu}, B) d\boldsymbol{\theta}. \quad (1)$$

In the denominator, the probability density function is given by

$$\begin{aligned} f_d(\mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\mu}, W, B) \\ = \int_{\boldsymbol{\theta}} \{f(\mathbf{y}_1 | \boldsymbol{\theta}, W) f(\boldsymbol{\theta} | \boldsymbol{\mu}, B)\} d\boldsymbol{\theta} \int_{\boldsymbol{\theta}} \{f(\mathbf{y}_2 | \boldsymbol{\theta}, W) f(\boldsymbol{\theta} | \boldsymbol{\mu}, B)\} d\boldsymbol{\theta}. \end{aligned} \quad (2)$$

The value of the evidence is the ratio of expressions (1) to (2). See Aitken and Lucy (2004) for details.

Vice versa, if the variability within sources is assumed to be not constant and an Inverted Wishart distribution is introduced, then the foregoing integration has no analytical form. In this case, a numerical procedure (a Gibbs sampling) is implemented to handle the complexity and compute the marginal likelihood under the competitive propositions.

4. Results and conclusions

To assess the performance of the method, from each writer the n_i measurements were randomly divided into two groups to act as control and recovered data: values coming from the first group were denoted as measurements of a control sample, while values coming from the second group were denoted as measurements of a recovered manuscript. In order to take into account the within-writers variability, one thousand draws were considered for each writer, and the likelihood ratio was computed for each draw.

Table 1 shows the apportionments of likelihood ratio considered (in logarithmic intervals), as proposed in forensic literature. Table 2 shows the frequency of false negatives and false positives obtained for each separate letter. These values represent the frequency of the states showing a likelihood ratio less than 1 (states 7 to 12) obtained under proposition H_p and greater than 1 (states 1 to 6) obtained under H_d , respectively.

Figure 1 shows as an example the values of the evidence obtained considering letter *a* of writer 1 and 10 under proposition H_p , notably the frequency of the apportionments of likelihood ratio according to states described in Table 1. It can be observed that variability changes amongst writers.

Results such as those presented in Table 2 are very encouraging, though from Figure 1 it can be observed that for writer 10 the frequency of state 1 (a likelihood ratio greater than 10^5) is rather high: these values are very high and must be treated with scepticism. Results under the second model (the one that allows for not constant variability) still presented

States	1	2	3	4	5	6
LR	$> 10^5$	$10^5 - 10^4$	$10^4 - 10^3$	$10^3 - 10^2$	$10^2 - 10^1$	$10^1 - 1$
States	7	8	9	10	11	12
LR	$1 - 10^{-1}$	$10^{-1} - 10^{-2}$	$10^{-2} - 10^{-3}$	$10^{-3} - 10^{-4}$	$10^{-4} - 10^{-5}$	$< 10^{-5}$

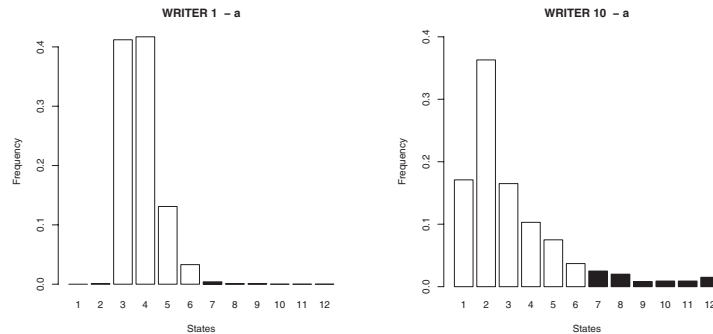
Table 1: Apportionments of Likelihood Ratio

Letter	False negatives	False positives
Letter a	0.02	0.0264
Letter d	0.015	0.0195
Letter o	0.019	0.009
Letter q	0.02	0.003

Table 2: Frequency of false-negatives and false-positives

some extreme values. This represents a critical point which will be addressed with utmost attention in the future developments.

Figure 1: Frequency of the apportionments of likelihood ratio, writers 1 and 10, letter a.



References

- Aitken, C.G.G. and Taroni, F. (2004) *Statistics and the evaluation of evidence for forensic scientists*, 2nd ed. John Wiley and Sons, Inc., Chichester.
- Aitken, C.G.G., Lucy, D., (2004) Evaluation of trace evidence in the form of multivariate data, *Applied Statistics*, 53, 109–122.
- Marquis, R., Schmittbuhl, M., Mazzella, W.D., Taroni, F. (2005a) Quantification of the shape of handwritten characters: a step to objective discrimination between writers based on the study of the capital character O, *Forensic Science International*, 150, 23–32.
- Marquis, R., Taroni, F., Bozza, S., Schmittbuhl, M. (2005b) Quantitative characterization of morphological polymorphism of handwritten character loops, *Forensic Science International*, in press.