Forensic Statistics - Evidence Evaluation
Colin Aitken

c.g.g.aitken@ed.ac.uk.

October 2014

Basic formula

$$\frac{Pr(H_p \mid E, I)}{Pr(H_d \mid E, I)} = \frac{Pr(E \mid H_p, I)}{Pr(E \mid H_d, I)} \times \frac{Pr(H_p \mid I)}{Pr(H_d \mid I)}.$$

- H_p: prosecution proposition;
- H_d: defence proposition;
- E : evidence to be evaluated;
- I : background information.

Continuous data

Continuous data

$E = \mathbf{x}, \mathbf{y};\ I = \mathbf{z}$

- $\mathbf{x}$: control evidence measurements, source known;
- $\mathbf{y}$: recovered evidence measurements, source unknown;
- $\mathbf{z}$: background data: measurements from some relevant source.

Continuous data

$E = \mathbf{x}, \mathbf{y};\ I = \mathbf{z}$

- $\mathbf{x}$: control evidence measurements, source known;
- $\mathbf{y}$: recovered evidence measurements, source unknown;
- $\mathbf{z}$: background data: measurements from some relevant source.

Hierarchical models:
$\mathbf{X}, \mathbf{Y}, \mathbf{Z} \sim f_1(\cdot \mid \theta); \Theta \sim f_2(\theta \mid \omega)$.

Continuous data

$E = \mathbf{x}, \mathbf{y};\ I = \mathbf{z}$

- $\mathbf{x}$: control evidence measurements, source known;
- $\mathbf{y}$: recovered evidence measurements, source unknown;
- $\mathbf{z}$: background data: measurements from some relevant source.

Hierarchical models:
$\mathbf{X}, \mathbf{Y}, \mathbf{Z} \sim f_1(\cdot \mid \theta); \Theta \sim f_2(\theta \mid \omega)$.
Possible distributions:
f_1: multivariate Normal;
f_2: multivariate Normal / Wishart or nonparametric.

Continuous data - likelihood ratio

$$\frac{Pr(\mathbf{x},\mathbf{y} \mid H_p, I)}{Pr(\mathbf{x},\mathbf{y} \mid H_d, I)} = \frac{\int f_1(\mathbf{x} \mid \theta) f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}{\int f_1(\mathbf{x} \mid \theta) f_2(\theta) d\theta \int f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}.$$

Continuous data - likelihood ratio

$$\frac{Pr(\mathbf{x}, \mathbf{y} \mid H_p, I)}{Pr(\mathbf{x}, \mathbf{y} \mid H_d, I)} = \frac{\int f_1(\mathbf{x} \mid \theta) f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}{\int f_1(\mathbf{x} \mid \theta) f_2(\theta) d\theta \int f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}.$$

Examples:

- Chemical composition of drugs; multivariate independent;
- Elemental composition of glass; multivariate independent;
- Refraction units of ions in cocaine on banknotes; univariate autocorrelated.

Continuous data - likelihood ratio

$$\frac{Pr(\mathbf{x}, \mathbf{y} \mid H_p, I)}{Pr(\mathbf{x}, \mathbf{y} \mid H_d, I)} = \frac{\int f_1(\mathbf{x} \mid \theta) f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}{\int f_1(\mathbf{x} \mid \theta) f_2(\theta) d\theta \int f_1(\mathbf{y} \mid \theta) f_2(\theta) d\theta}.$$

Examples:

- Chemical composition of drugs; multivariate independent;
- Elemental composition of glass; multivariate independent;
- Refraction units of ions in cocaine on banknotes; univariate autocorrelated.

Keywords: multivariate Bayesian hierarchical modelling, Hidden Markov models, graphical models, autocorrelation, nonparametric density estimation.

Comparison of mass spectra for inks

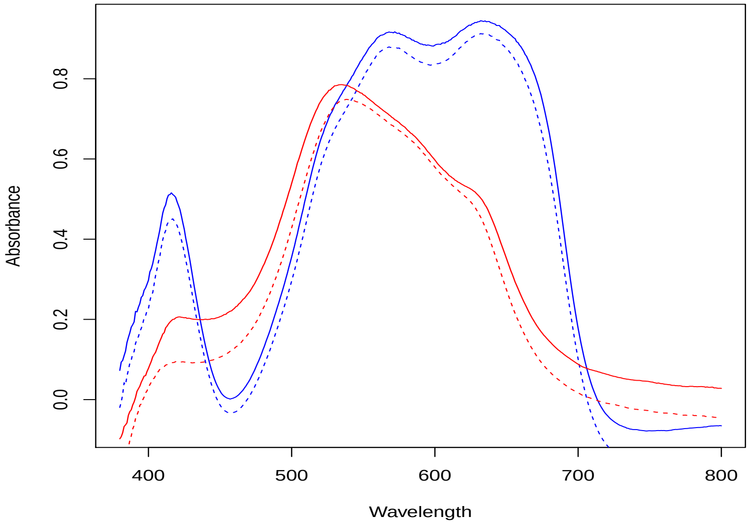
Ink - collected in criminal cases such as suicides, documents and will forgeries, and blackmail. Does the ink from such a document share a common origin with the ink sample from a document constituting a control material?

Comparison of mass spectra for inks

Ink - collected in criminal cases such as suicides, documents and will forgeries, and blackmail. Does the ink from such a document share a common origin with the ink sample from a document constituting a control material?

Given mass spectrometry data $f(t_i, y_i), i = 1, ..., n$, for n inks of the form of curves for absorbance y vs. wavelength t, the purpose of the research is to develop a model for the likelihood ratio for evidence in the form of data from two sets of writing in ink under the propositions that the two sets have a common type of ink and the proposition that the two sets are different types of ink.

Comparison of mass spectra for inks



Tool	Mark(1) / No Mark (0)					
1	0	1	1	1	1	1
1	1	1	1	1	0	0
1	1	1	1	1	1	1
2	1	0	0	0	1	0
2	1	1	0	0	0	1
2	1	1	0	0	0	0
3	0	1	0	0	0	0
3	0	0	0	0	0	0
3	0	1	0	0	0	0

Consider distance between one set of marks and another as the number of differences between sets.

Comparison of distances between marks