Forensic Science International 236 (2014) 67–76

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

The evaluation of evidence relating to traces of cocaine on banknotes

Amy Wilson [a], Colin Aitken [a,*], Richard Sleeman [b], James Carter [c]

[a] School of Mathematics and Maxwell Institute, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom
[b] Mass Spec Analytical Ltd., Building 20F, Golf Course Lane, Bristol BS34 7RP, United Kingdom
[c] Clinical and Statewide Services Division, Queensland Health, PO Box 594, Archerfield, Queensland 4108, Australia

ARTICLE INFO

Article history:
Received 19 February 2013
Received in revised form 5 November 2013
Accepted 27 November 2013
Available online 17 December 2013

Keywords:
Autoregressive model
Banknotes
Cocaine
Evidence evaluation
Likelihood ratio
Nonparametric density estimation

ABSTRACT

Banknotes can be seized from crime scenes as evidence for suspected association with illicit drug dealing. Tandem mass spectrometry data are available from banknotes seized in criminal investigations, as well as from banknotes from general circulation. The aim of the research is to evaluate the support provided by the data gathered in a criminal investigation for the proposition that the banknotes from which the data were obtained are associated with a person who is associated with a criminal activity related to cocaine in contrast to the proposition that the banknotes are associated with a person who is not associated with a criminal activity involving cocaine. The data considered are the peak area for the ion count for cocaine product ion *m/z* 105. Previous methods for assessment of the relative support for these propositions were concerned with the percentage of banknotes contaminated or assume independence of measurements of quantities between adjacent banknotes. Methods which account for an association of the quantity of drug on a banknote with that on adjacent banknotes are described. The methods are based on an autoregressive model of order one and on two versions of a nonparametric approach. The results are compared with a standard model which assumes measurements on individual banknotes are independent; there is no autocorrelation. Performance is assessed using rates of misleading evidence and a recommendation made as to which method to use.

1. Introduction

A novel approach to the evaluation of evidence of the quantity of cocaine on banknotes is described. The methods are applicable more generally and can be applied to measurements of quantities of other drugs. The novelty is the consideration of autocorrelation which, in this context, is a measure of the association between the quantities of drugs on adjacent banknotes. A previous method, described in [1], uses only one banknote and states that one cannot assume independence in the measurements between adjacent banknotes.

The evidence is evaluated through use of the likelihood ratio (LR). In this context, the ratio is that of the probability density function of the data [1] under each of two propositions,

- H_C: the banknotes are associated with a person who is associated with a criminal activity involving cocaine, and
- H_B: the banknotes are associated with a person who is not associated with a criminal activity involving cocaine.

There has been some previous use of LRs in the area of drugs on banknotes. In [1], the likelihood ratio of the quantity of contamination of cocaine on a seized banknote was evaluated using a histogram. It is noted that calculating a LR for a set of multiple banknotes using this method is not possible without assuming independence (*i.e.*, that the quantity of cocaine on a particular banknote is unaffected by the quantity of cocaine on any other banknote, such as an immediate neighbour). In [2] the likelihood ratio for the quantity of cocaine contamination on a set of banknotes is calculated using a univariate kernel density estimate. An assumption of independence is made and it is noted that this assumption may not be warranted. This assumption is not made in the models introduced in this paper. The results obtained from these models are compared with a model which assumes independence.

The data used for the analysis, $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, are the logarithms of the peak areas of cocaine on a set of n banknotes. The strength of the evidence of $\mathbf{z}$ in support of H_C or H_B is to be assessed. The logarithmic transformation is made to reduce skewness in the

* Corresponding author. Tel. +44 0131 650 4877
E-mail address: c.g.g.aitken@ed.ac.uk (C. Aitken).

[1] The ratio is known as a likelihood ratio as this is a technical phrase in statistical theory in which a probability density function for data given parameter values may be thought of as a likelihood of the parameter values given the data. The phrase has been transferred in the forensic statistic literature to refer to the ratio of the probability density functions given propositions. Note that as the data are continuous it is not possible to refer to the probability of the data.

0379-0738/$ – see front matter © 2014 Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.forsciint.2013.11.011

data before fitting models. Training data are available from banknotes deemed to be associated with criminal activity involving cocaine and from banknotes deemed to be associated with general or background circulation. These training data are used to develop models associated with H_C and H_B, respectively.

The likelihood ratio LR associated with the propositions H_C and H_B is given by:

$$LR = \frac{f(\mathbf{z}|H_C)}{f(\mathbf{z}|H_B)},$$

where the function f is a probability density function for the measurements, conditional on H_C and on H_B in the numerator and the denominator, respectively. If this statistic is greater than one, then the evidence assigns more support to the proposition that the banknotes are associated with a person who is associated with drug crime involving cocaine. With an assumption of independence amongst the values in $\mathbf{z}$,

$$LR = \frac{\prod_{i=1}^{n} f(z_i|H_C)}{\prod_{i=1}^{n} f(z_i|H_B)}.$$

The interpretation of this LR is slightly different from the one used for comparison of possible sources of recovered and control evidence in [3]. Here there is only one set of evidence, the seized banknotes provided by the law enforcement agency. The LR provides a measure of support for one or other of the propositions as to whether the person with whom they are associated is himself associated or not with criminal activity involving cocaine.

The models described here give three approaches to the estimation of f in which there is no assumption of independence amongst the quantities on the n banknotes in the sample of unknown origin. One approach allows for an autocorrelation of lag one which models a dependency between measurements on adjacent banknotes. The other two approaches use a kernel density approach with multivariate conditional density functions instead of the univariate kernel density functions used in [2], again modelling a dependency between measurements from adjacent banknotes; one of these approaches has a fixed bandwidth, the other a bandwidth which varies with the density of the measurements.

2. Data

The method for acquisition of the data is described in [4]. Models were developed for the cocaine product ion m/z 105. For each sample of banknotes the data resemble a series of peaks, with each peak corresponding to a banknote. The height of the peak at any given scan number is given by the number of gas phase ion transitions giving rise to the cocaine product ion m/z 105 at that scan number. A peak detection algorithm was written in order to identify these peaks. The details are beyond the scope of this paper but are available from the corresponding author on request. Once identified, the area under each peak was measured and its logarithm used as a measure of the quantity of cocaine on each banknote.

Any sample for which the difference between the total number of recorded banknotes and the number of peaks detected by the peak detection algorithm was greater than 10% of the total number of recorded banknotes (either way) was removed as such a discrepancy meant there were difficulties with the peak detection algorithm. Any sample with fewer than twenty banknotes was removed as the statistical procedures were unreliable with such few banknotes. Any samples for which the information on the currency or the total number of banknotes were not available were also removed. Samples were often analysed in multiple runs and there were some missing runs within samples. For these, the longest contiguous section of the sample was included, with the rest discarded. The data from each sample were plotted and any outlying data points were further investigated and removed if they were found to be incorrectly identified peaks.

Two sets of training data were formed from the samples analysed. One set, C, containing data $\mathbf{y}$, was formed to develop the model for H_C, and another set B containing data $\mathbf{x}$, was formed to develop the model for H_B.

2.1. Banknotes that have been associated with criminal activity involving cocaine

The training data $\mathbf{y}$ for models developed for H_C are obtained from banknotes in criminal cases in which the defendant was convicted of a drug crime involving cocaine. Each case consisted of multiple exhibits, which may have been found in different locations. There were 29 cases containing at least one exhibit with greater than 20 banknotes. The 29 cases consist of between one and six exhibits, and there were a total of 70 exhibits which are known to have been associated with a person who has been involved in drug crime relating to cocaine. For future reference, any set of banknotes used in the analyses discussed that is said to be associated with crime will be known as an exhibit.

The training data $\mathbf{y}$ of banknotes used to develop models associated with H_C may include exhibits with two different types of cocaine contamination.

C(a) The banknotes have not been contaminated with cocaine any more than those banknotes in general circulation. The contamination detected on the banknotes is consistent with that typically detected on general circulation banknotes. This quantity of contamination could have arisen innocently, or because the banknotes were not contaminated in the course of a crime (perhaps no drug was present at an exchange of money) by the person with whom they are associated.

C(b) The banknotes were contaminated through their use in an illegal drug-related activity involving cocaine or in the course of other, legal, drug-related activity. This activity could have been carried out by some person other than the person eventually convicted.

It is expected that the quantities of cocaine on banknotes in C(a) will be lower than the quantities of cocaine on banknotes in C(b). See Fig. 1 for an illustration of the overlap in mean quantities of cocaine from samples from general circulation and from exhibits associated with crime (case). The left mode of the crime exhibits is formed of those exhibits in set C(a), and the right hand mode is formed of those exhibits in set C(b). Note that C(a) and C(b) are not propositions but descriptions of the possible sources of cocaine on the banknotes used as training data for the development of the models associated with the proposition H_C.

2.2. General circulation banknotes

The training data $\mathbf{x}$ of banknotes in set B are very unlikely to contain whole samples that are all associated with crime (although individual banknotes within a sample may well have been involved in a crime). Thus the banknotes defined by B and by $C(a)$ are likely to be similarly contaminated. The banknotes defined by $C(b)$ have been contaminated through their involvement in criminal activity involving cocaine, and so are likely to have higher levels of contamination.

There were 193 general circulation samples of English or Scottish currency obtained from a variety of locations around the UK. As shown in Table 1, a large number of the general circulation