The evaluation of evidence relating to traces of cocaine on banknotes

Amy Wilson

September 2015
Table of Contents

1. Likelihood ratio framework
2. Cocaine on banknotes
3. Data and propositions
4. Models
5. Results and conclusion
Contents

1 Likelihood ratio framework
2 Cocaine on banknotes
3 Data and propositions
4 Models
5 Results and conclusion
Evidence evaluation

- When evaluating evidence for use in a court of law, typically have:
  - some evidential data \((E)\),
  - two competing propositions relating to the data: one from the prosecution \((H_p)\) and one from the defence \((H_d)\).
- Aim of court is to evaluate which of two propositions is more likely, given the evidence.
- In other words, is the probability of \(H_p\) given the evidence bigger than the probability of \(H_d\) given the evidence?
Evidence evaluation

We want to know whether the ratio

\[
\frac{P(H_p \mid E)}{P(H_d \mid E)}
\]

is greater than one, or less than one.
Bayes’ theorem

Can write as:

\[
\frac{P(H_p \mid E)}{P(H_d \mid E)} = \frac{P(E \mid H_p)}{P(E \mid H_d)} \times \frac{P(H_p)}{P(H_d)}
\]

Odds after evidence = Likelihood ratio $\times$ Odds before evidence

- Likelihood ratio greater than one: the evidence has increased the odds in favour of $H_p$ (in comparison to prior odds),
- Forensic scientists can use likelihood ratio to measure strength of evidence.
Contents

1 Likelihood ratio framework

2 Cocaine on banknotes

3 Data and propositions

4 Models

5 Results and conclusion
Motivation

- Banknotes can be seized from a crime scene as evidence,
- Methods exist to measure the amount of cocaine on each banknote within a sample of notes,
- Banknotes are generally stored in bundles and cocaine measurements are taken sequentially,
- It is known that cocaine can transfer between surfaces.
Previous approaches

- Ad-hoc, expert view.
- Hypothesis testing
  - No consideration of contamination on notes associated with crime,
  - Can’t be used for multiple banknotes - contamination is not independent.
- Likelihood ratios based on kernel density estimates
  - Considers contamination on notes associated with crime,
  - Still cannot be used for multiple banknotes - requirement of an assumption of independence.
Aims

- Develop statistical methodology using the likelihood ratio framework to evaluate autocorrelated evidence.
- Apply this to the evaluation of evidence relating to traces of cocaine on banknotes.
The evaluation of evidence relating to traces of cocaine on banknotes

Contents

1. Likelihood ratio framework

2. Cocaine on banknotes

3. Data and propositions

4. Models

5. Results and conclusion
The evaluation of evidence relating to traces of cocaine on banknotes

Data and propositions

General circulation

- Training dataset: 193 samples of banknotes from general circulation.
- Each sample had between 20 and 257 banknotes.
- Measurements taken using mass spectrometer. Each measurement is the logarithm of the peak area for the cocaine $m/z$ 105 ion.
Crime

- 70 samples of banknotes that were seized from a suspect by law enforcement agencies, where the suspect was later convicted of a crime involving cocaine.
- Known as ‘exhibits’.
- Each exhibit had between 20 and 1099 banknotes.
What should the propositions be?

- Propositions need to match data used to produce the models.
- Crime dataset: banknotes known to be associated with someone who was convicted of a crime involving cocaine.
- Background dataset: banknotes known to be taken from general circulation.
- Assume that banknotes from general circulation have same distribution of cocaine contamination as those associated with a person who is not involved with criminal activity involving cocaine?
Propositions chosen

- $H_C$: the banknotes have been seized by law enforcement agencies as evidence in a criminal case against a group of one or more people, and that at least one of these people is guilty (in the eyes of the law) of a crime involving cocaine.

- $H_B$: the banknotes have been seized by law enforcement agencies as evidence in a criminal case against a group of one or more people, and that none of these people is guilty (in the eyes of the law) of a crime involving cocaine.
Limitations

- Multiple suspects- ‘at least one of the suspects is involved with a crime involving cocaine’.
- Suspect may state where the banknotes are from - notes from general circulation may then not have same distribution as notes from this source.
- General circulation samples mainly from banks. May not be representative of situation.
- Propositions may not match what court/forensic scientists want.
Data - statistical issues

- Cocaine is present on banknotes from general circulation.
- Many crime exhibits are not contaminated any more than general circulation (58 of 70 were not declared as contaminated by experts).
- Over 80% of samples and exhibits had significant autocorrelation at lag one.

Density plots of average cocaine log contamination

Figure: Density plots of mean contamination of samples/exhibits. Red - general circulation, black - crime exhibit.
Contamination levels

Samples and exhibits consist of multiple bundles of cash. Often, these bundles have different levels of contamination.
The evaluation of evidence relating to traces of cocaine on banknotes

Contents

1 Likelihood ratio framework
2 Cocaine on banknotes
3 Data and propositions
4 Models
5 Results and conclusion
The evaluation of evidence relating to traces of cocaine on banknotes

Some notation

Define:

- \( C \) as the training dataset of crime exhibits.
- \( B \) as the training dataset of general circulation samples.
- \( z = (z_1, z_2, \ldots, z_n) \) as the logarithms of the peak areas of a sample of banknotes found on a suspect (i.e. the evidence).

The likelihood ratio associated with \( H_B \) and \( H_C \) is

\[
V = \frac{f(z \mid H_C)}{f(z \mid H_B)} = \frac{\int f(z \mid \theta_C)f(\theta_C)d\theta_C}{\int f(z \mid \theta_B)f(\theta_B)d\theta_B}
\]
Which models were fitted?

- AR(1) model - takes autocorrelation into account.
- Hidden Markov model - takes autocorrelation and ‘bundles’ structure into account.
- Non-parametric model using conditional density functions - takes autocorrelation into account, no assumption of Normality of errors.
- A model which assumes independence, for comparison.
The hidden Markov model

The Bayesian network of the hidden Markov model used is:

- Bundles modelled using hidden states. There is one hidden state for each banknote.
- Independence of observations, conditional on the hidden states, is not assumed.
Parameter Estimation

- Crime and background datasets used to estimate parameters $\theta_C$ and $\theta_B$,
- Bayesian approach - priors on all parameters and Metropolis-Hastings sampler,
- Likelihood ratio estimated using Monte Carlo integration (for hidden Markov model can use forward algorithm (Rabiner 1989) to sum out hidden states).
Contents

1. Likelihood ratio framework
2. Cocaine on banknotes
3. Data and propositions
4. Models
5. Results and conclusion
## Rates of misleading evidence

<table>
<thead>
<tr>
<th>Model</th>
<th>Crime exhibit</th>
<th>General circulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden Markov model</td>
<td>0.36 (25/70)</td>
<td>0.10 (20/193)</td>
</tr>
<tr>
<td>AR(1) model</td>
<td>0.37 (26/70)</td>
<td>0.16 (30/193)</td>
</tr>
<tr>
<td>Nonparametric fixed bw</td>
<td>0.27 (19/70)</td>
<td>0.32 (62/193)</td>
</tr>
<tr>
<td>Nonparametric adaptive nn</td>
<td>0.26 (18/70)</td>
<td>0.27 (52/193)</td>
</tr>
<tr>
<td>Model assuming independence</td>
<td>0.50 (35/70)</td>
<td>0.14 (26/193)</td>
</tr>
</tbody>
</table>

**Table**: Rates of misleading evidence, estimated as $(r/n)$ where $r$ is the number of samples or exhibits out of $n$ analysed which gave misleading support in each context.
**Results and conclusion**

**Tippett plots - parametric**

- **HM model**
- **AR1 model**
- **Standard model**
The evaluation of evidence relating to traces of cocaine on banknotes

Results and conclusion

Tippett plots - nonparametric

Nonparametric model - fixed bandwidth

Nonparametric model - adaptive bandwidth
Comparison to a forensic expert

<table>
<thead>
<tr>
<th>Exhibit number</th>
<th>HMM</th>
<th>AR(1)</th>
<th>nonparametric fixed bandwidth</th>
<th>nonparametric variable bandwidth</th>
<th>Standard model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.37</td>
<td>6.05</td>
<td>31.02</td>
<td>39.02</td>
<td>32.61</td>
</tr>
<tr>
<td>3</td>
<td>3.51</td>
<td>3.67</td>
<td>5.19</td>
<td>6.43</td>
<td>4.68</td>
</tr>
<tr>
<td>16</td>
<td>6.61</td>
<td>7.51</td>
<td>6.92</td>
<td>7.14</td>
<td>2.89</td>
</tr>
<tr>
<td>23</td>
<td>7.51</td>
<td>6.32</td>
<td>8.64</td>
<td>7.64</td>
<td>7.72</td>
</tr>
<tr>
<td>38</td>
<td>5.38</td>
<td>6.64</td>
<td>11.61</td>
<td>12.55</td>
<td>7.39</td>
</tr>
<tr>
<td>39</td>
<td>7.31</td>
<td>10.39</td>
<td>20.43</td>
<td>22.69</td>
<td>8.51</td>
</tr>
<tr>
<td>40</td>
<td>4.91</td>
<td>2.24</td>
<td>0.05</td>
<td>21.53</td>
<td>0.60</td>
</tr>
<tr>
<td>42</td>
<td>4.35</td>
<td>4.09</td>
<td>6.23</td>
<td>8.03</td>
<td>2.47</td>
</tr>
<tr>
<td>43</td>
<td>6.89</td>
<td>7.06</td>
<td>6.80</td>
<td>8.61</td>
<td>2.06</td>
</tr>
<tr>
<td>57</td>
<td>4.66</td>
<td>3.58</td>
<td>6.24</td>
<td>11.13</td>
<td>5.45</td>
</tr>
<tr>
<td>67</td>
<td>16.52</td>
<td>0.57</td>
<td>244.80</td>
<td>262.25</td>
<td>7.51</td>
</tr>
<tr>
<td>69</td>
<td>17.42</td>
<td>0.48</td>
<td>128.69</td>
<td>169.64</td>
<td>5.44</td>
</tr>
</tbody>
</table>

**Table:** Log likelihood ratios of exhibits declared as contaminated by an expert.
Conclusions

- Models developed to calculate likelihood ratios for autocorrelated evidential data and data that are driven by a latent variable (the hidden Markov model).
- Applied to data which consist of cocaine traces on banknotes.
- Lowest rate of misleading evidence for the banknotes data achieved with the hidden Markov model.
- Problems with outliers when using non-parametric models.
- Not modelling autocorrelation when it is present seems to result in overstatement of likelihood ratios for small exhibits.