Lies, damned lies, expert witnesses and drugs on banknotes

Colin Aitken

School of Mathematics,
The University of Edinburgh,
Maxwell Institute

cgga@ed.ac.uk.

April 2016

CS Peirce (1878)

- Probability is the ratio of favourable cases to all cases.
- Chance is the ratio of favourable to unfavourable.
- Belief is the logarithm of chance and is proportional to the weight of chance; to multiply chances is to add beliefs.
- Balancing reasons: take the sum of all the feelings of belief which would be produced separately by all the arguments *pro* the proposition, subtract from that the similar sum for arguments *con*. The remainder is the feeling of belief which one ought to have on the whole.

As it is impossible to know the probability à priori, we will not be able to say such coincidence prove that the relation of the probability of the forgery to the inverse probability to such value. We would be only able to say, by the finding of this coincidence, this report becomes many times larger than before the finding.

Darboux, Appel, Poincaré
August 2nd, 1904

History

Good(1979)
Summary of statistical ideas of Alan Turing in 1940, 1941.

- Introduction of the expression '(Bayes) factor in favour of a hypothesis' without the qualification 'Bayes': the factor by which initial odds of H must be multiplied to obtain the final odds in favour of H provided by evidence E.
- Sequential analysis and log factors; log factor is the 'weight of evidence'; closely related to the amount of information concerning H provided by E.
- The ban and deciban: the unit by which weight of evidence is measured. A deciban is one-tenth of a ban.
- The weighted average of factors.
- Expected weight of evidence, variance of weight of evidence.

Likelihood ratios — 1

H_p	Prosecution proposition
H_d	Defence proposition
E	Evidence
I	Framework of circumstances

Likelihood ratios — 3

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

Posterior odds in favour of the prosecution proposition equals the likelihood ratio multiplied by the prior odds in favour of the prosecution proposition.

Likelihood ratios — 3

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

Posterior odds in favour of the prosecution proposition equals the likelihood ratio multiplied by the prior odds in favour of the prosecution proposition.

Innocent until proven guilty: $Pr(H_p \mid I)/Pr(H_d \mid I)$.

Likelihood ratios — 3

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

Posterior odds in favour of the prosecution proposition equals the likelihood ratio multiplied by the prior odds in favour of the prosecution proposition.

Innocent until proven guilty: $Pr(H_p \mid I)/Pr(H_d \mid I)$.

Guilty beyond reasonable doubt: $Pr(H_p \mid E, I)/Pr(H_d \mid E, I)$.

Notes on LR

Let

$$V = \frac{Pr(E \mid H_p)}{Pr(E \mid H_d)}.$$

Likelihood ratio V may be thought of as the value of the evidence.

Summarise evidence with phrase 'the evidence is V times more likely if H_p is true than if H_d is true'.

A value of $V > 1$ supports H_p.

A value of $V < 1$ supports H_d.

No statement is made about the probability of the truth of either proposition.

Mathematical justification of likelihood ratio - IJ Good

Assume that the value, $V(E)$ of the evidence E is a function, f, of $x = P(E \mid G)$ and $y = P(E \mid \bar{G})$ alone; $V = f(x, y)$.
Consider an event F that is entirely irrelevant to E and G. Let $P(F) = \lambda$. Then

$$\begin{aligned} P(E\&F \mid G) &= \lambda x \\ P(E\&F \mid \bar{G}) &= \lambda y \\ V(E\&F) &= f(\lambda x, \lambda y) \end{aligned}$$

However

$$V(E\&F) = V(E) \Rightarrow f(\lambda x, \lambda y) = f(x, y);$$

because F is irrelevant and therefore inadmissible as evidence. The equality is true for all $\lambda \in [0, 1]$.
Hence f is a function of x/y alone.

Lindley (1977)

$$BF = \frac{\tau}{2^{\frac{1}{2}}\sigma} \exp\left\{-\frac{(X-Y)^2}{4\sigma^2}\right\} \exp\left\{\frac{(Z-\mu)^2}{2\tau^2}\right\}.$$

Let $\lambda = \mid X - Y \mid /(2^{\frac{1}{2}}\sigma)$ and $\delta = \mid Z - \mu \mid /\tau$; $\tau/\sigma = 100$.

$$\begin{aligned} \lambda = \delta = 0 &\Rightarrow BF = 70.7; \\ \lambda = 0, \delta = 3.0 &\Rightarrow BF = 6370; \\ \lambda = 6.0, \delta = 0 &\Rightarrow BF = 1/(9.29 \times 10^5) \end{aligned}$$

Sally Clark - SIDS - Confidence intervals

CESDI study: case-control study:

Sally Clark - SIDS - Confidence intervals

CESDI study: case-control study:

Cases: 323 families in which there was a current death assigned to SIDS. Look backwards (retrospective study). Of those 323 families, five had had a previous infant death assigned to SIDS.

CESDI study: case-control study:

Cases: 323 families in which there was a current death assigned to SIDS. Look backwards (retrospective study). Of those 323 families, five had had a previous infant death assigned to SIDS.

Compare this with the statement in court that a double SIDS in a family can be expected once in 100 years.

Sally Clark - SIDS - Confidence intervals

CESDI study: case-control study:

Cases: 323 families in which there was a current death assigned to SIDS. Look backwards (retrospective study). Of those 323 families, five had had a previous infant death assigned to SIDS.

Compare this with the statement in court that a double SIDS in a family can be expected once in 100 years.

Controls: 1288 families alike in all other respects to the 323 case families except there had not been a current death assigned to SIDS.

Sally Clark - SIDS - Confidence intervals

CESDI study: case-control study:

Cases: 323 families in which there was a current death assigned to SIDS. Look backwards (retrospective study). Of those 323 families, five had had a previous infant death assigned to SIDS.

Compare this with the statement in court that a double SIDS in a family can be expected once in 100 years.

Controls: 1288 families alike in all other respects to the 323 case families except there had not been a current death assigned to SIDS.

Of those 1288 families, two had had a previous infant death assigned to SIDS.

Interpretation of SIDS - odds ratio: confidence interval

The odds ratio is 10.11; the log odds ratio is $\log_e(10.11) = 2.314$.

The odds ratio is 10.11; the log odds ratio is $\log_e(10.11) = 2.314$.

The standard error of the log odds ratio is 0.839

The odds ratio is 10.11; the log odds ratio is $\log_e(10.11) = 2.314$.

The standard error of the log odds ratio is 0.839 A 95% confidence interval for the log odds ratio is then

$$2.314 \pm 1.96 \times 0.839 = 2.314 \pm 1.644 = (0.670, 3.958).$$

The odds ratio is 10.11; the log odds ratio is $\log_e(10.11) = 2.314$.

The standard error of the log odds ratio is 0.839 A 95% confidence interval for the log odds ratio is then

$$2.314 \pm 1.96 \times 0.839 = 2.314 \pm 1.644 = (0.670, 3.958).$$

A 95% confidence interval for the odds ratio is

$$(\exp(0.670), \exp(3.958)) = (1.954, 52.353) \simeq (2, 52).$$

The odds in favour of a previous death assigned to SIDS is 10 times greater in a family with a current SIDS than in a family with no current death assigned to SIDS with associated 95% confidence interval of (2, 52).

The odds in favour of a previous death assigned to SIDS is 10 times greater in a family with a current SIDS than in a family with no current death assigned to SIDS with associated 95% confidence interval of (2, 52).

Thus there is evidence, significant at the 5% level, that the true odds ratio is greater than 1 and hence that the odds in favour of a previous SIDS in a family with a current SIDS is greater than the odds in favour of a previous SIDS in a family with no current SIDS. This is evidence of dependence between occurrences of SIDS in the same family.

Drugs on banknotes - 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 measured sequentially.
- It is known that cocaine can transfer between surfaces.
- Methods of evidence evaluation within the likelihood ratio framework have not been developed for autocorrelated data like this.

Likelihood ratio

The value of evidence E is its effect on the ‘odds’ in favour of a particular proposition H_C with respect to another mutually exclusive proposition H_B.

Likelihood ratio

The value of evidence E is its effect on the 'odds' in favour of a particular proposition H_C with respect to another mutually exclusive proposition H_B.

$$\frac{Pr(H_C \mid E)}{Pr(H_B \mid E)} = \frac{Pr(E \mid H_C)}{Pr(E \mid H_B)} \times \frac{Pr(H_C)}{Pr(H_B)}.$$

Propositions - discussion

Propositions referring to the notes:

- H_C: the banknotes have been associated with criminal activity involving cocaine.
- H_B: the banknotes are from general circulation.

Propositions - discussion

Propositions referring to a person. Two possibilities $C1$ and $C2$ are suggested for the prosecution proposition, one possibility B for the defence proposition.

Propositions - discussion

Propositions referring to a person. Two possibilities $C1$ and $C2$ are suggested for the prosecution proposition, one possibility B for the defence proposition.

- $C1$: all of the banknotes *have been found in the possession of* a person who has pled guilty or has been found guilty of a criminal activity involving cocaine.
- $C2$: all of the banknotes *are associated with* a person who has pled guilty or has been found guilty of a criminal activity involving cocaine.
- B : all of the banknotes are associated with general circulation.

A distinction is drawn between $C1$ and $C2$ to emphasise that the notes may have been found somewhere such as a property or car associated with the person ($C2$) rather than in their possession ($C1$).
Note that neither $(C1, B)$ nor $(C2, B)$ are mutually exclusive.

Propositions - discussion

Propositions referring to a person. Two possibilities $C1$ and $C2$ are suggested for the prosecution proposition, one possibility B for the defence proposition.

$C1$: all of the banknotes *have been found in the possession of* a person who has pled guilty or has been found guilty of a criminal activity involving cocaine.

$C2$: all of the banknotes *are associated with* a person who has pled guilty or has been found guilty of a criminal activity involving cocaine.

B : all of the banknotes are associated with general circulation.

A distinction is drawn between $C1$ and $C2$ to emphasise that the notes may have been found somewhere such as a property or car associated with the person ($C2$) rather than in their possession ($C1$).
Note that neither $(C1, B)$ nor $(C2, B)$ are mutually exclusive.
The person from whom the new banknotes were seized will not have pled guilty or been found guilty of a crime (yet).

Propositions - discussion

Propositions - discussion

$C1$: all of the banknotes *have been found in the possession of* a person who is involved in a criminal activity involving cocaine.

$C2$: all of the banknotes *are associated with* a person who is involved in a criminal activity involving cocaine.

B : all of the banknotes are from general circulation.

These are still not mutually exclusive.

Propositions - discussion

Notes found in the possession of a person who is involved in a criminal activity involving cocaine. The cocaine contamination may have come from two sources:

Propositions - discussion

Notes found in the possession of a person who is involved in a criminal activity involving cocaine. The cocaine contamination may have come from two sources:

(a) The notes are from general circulation. They have not been contaminated with cocaine any more than those notes in general circulation, so have obtained their contamination from being in general circulation. This could either be because they were not involved with a crime (involving cocaine) and were obtained innocently, or because they were not contaminated in the course of a crime (perhaps no drug was present at the money exchange).

Propositions - discussion

Notes found in the possession of a person who is involved in a criminal activity involving cocaine. The cocaine contamination may have come from two sources:

(a) The notes are from general circulation. They have not been contaminated with cocaine any more than those notes in general circulation, so have obtained their contamination from being in general circulation. This could either be because they were not involved with a crime (involving cocaine) and were obtained innocently, or because they were not contaminated in the course of a crime (perhaps no drug was present at the money exchange).

(b) The notes were contaminated through their use in a criminal activity involving cocaine.

Propositions - discussion

Propositions - discussion

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

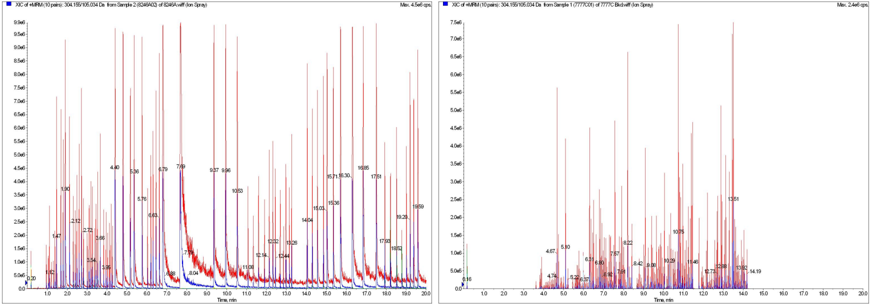
Drugs on banknotes: Propositions

Drugs on banknotes: Propositions

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

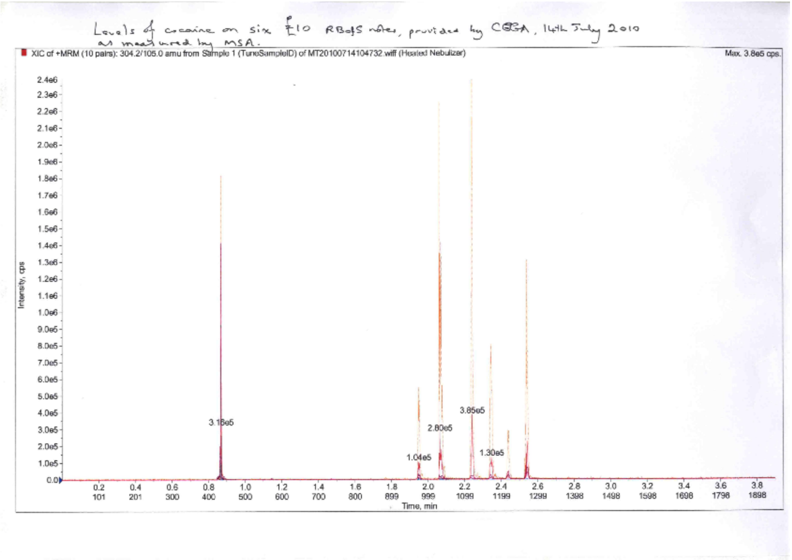
See Wilson et al. (2015)

Drugs on banknotes: The data



- Ion transition counts per second for one run from an exhibit in a criminal case (left) and from notes in general circulation (right). Ion transition 304 $\rightarrow$ 182 is shown in red and ion transition 305 $\rightarrow$ 105 is shown in blue.
- Log of cocaine peak areas used as data (ion 105).
- Two datasets available: banknotes associated with crime involving cocaine and banknotes from general circulation.
- Each sample consists of multiple banknotes (20 $-$ 1099). A dataset consists of a number of these samples.

Banknotes


Levels of cocaine on six £10 RBofS notes
MSA
XIC of +MRM (10 pairs): 304.2/105.0 amu from Sample 1 (TuneSampleID) of MT20100714104732.wiff (Heated Nebulizer)
Max. 3.8e5 cps.
Intensity, cps
Time, min
3.16e5
1.04e5
2.80e5
3.85e5
1.30e5

Banknotes

The training data

The training data

Banknotes associated with crime involving cocaine H_C:

Notes in criminal cases in which the defendant was convicted of a drug crime involving cocaine.

Each case consists 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 cases consisted 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 a drug crime relating to cocaine.

$\mathbf{y} = \{y_{ij}; i = 1, \ldots m_C, j = 1, \ldots n_{C_i}\}$: the logarithms of the peak areas of banknotes from criminal case exhibits for cocaine as defined in Section 2.1; there are m_C exhibits with n_{C_i} notes in exhibit i.

The training data

Banknotes associated with general circulation H_B:

193 general circulation samples of English and Scottish currency were obtained from a variety of locations around the UK. $\mathbf{x} = \{x_{ij}; i = 1, \ldots m_B, j = 1, \ldots n_{B_i}\}$: the logarithms of the peak areas for cocaine of banknotes from general circulation ; there are m_B samples with n_{B_i} notes in sample i.

The test data

Data z used for testing with the likelihood ratio will generally have been provided by the law enforcement agencies.

This may be thought to place z in C, by definition.

However, the definition of 'association' used here for the training set for C is that of conviction of a crime involving cocaine.

Data from other cases brought by the law enforcement agencies have not been included in the analysis. This definition of a case is different from definitions used in previous work, when all seized banknotes were used as cases.

Attributes of the data

- Cocaine is present on banknotes from general circulation.
- Some samples associated with crime are not contaminated.
- Over 80% of samples had significant autocorrelation at lag one.
- Samples consist of multiple bundles of cash. Often, these bundles have different levels of contamination.

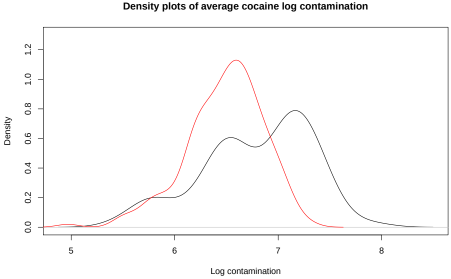


Figure: Density plots of mean contamination of samples/exhibits. Red - general circulation, black - positive case

Percentage of samples with significant autocorrelation at various lags

	Lag one	Lag two	Lag five
Positive case	80%	56%	39%
General Circulation	89%	62%	35%

The likelihood ratio

- H_C: the banknotes are associated with a person who is involved with criminal activity involving cocaine.
- H_B: the banknotes are associated with a person who is not involved with criminal activity involving cocaine.
- The measurements on a seized sample of n banknotes are given by $\mathbf{z} = (z_1, \ldots z_n)$.
- The likelihood ratio is given by:

$$V = \frac{f(\mathbf{z} \mid H_C)}{f(\mathbf{z} \mid H_B)}$$

- If $V > 1$ then the evidence $\mathbf{z}$ supports H_C, otherwise the evidence supports H_B.

Models fitted

- A standard AR(1) model - takes autocorrelation into account.
- A hidden Markov model - takes autocorrelation and bundles structure into account.
- A non-parametric model using conditional density functions - takes autocorrelation into account, no assumption of Normality of errors.
- A standard model assuming independence between banknotes

AR(1) model

The form of the models for B and for C is the same, only the parameters are different. The model is described with generic notation here, with w substituting for x and y as appropriate. The data of the logarithms of the peak areas of intensities of cocaine are denoted $\mathbf{w} = (w_1, \ldots, w_n)$. The corresponding random variable $\mathbf{W}$ is assumed Normally distributed with mean μ. An autoregressive model $AR(1)$ specifies the following relationship amongst the variables:

$$w_t - \mu = \alpha \left(w_{t-1} - \mu \right) + \epsilon_t \qquad (1)$$

where $t = 2, \ldots, n$; $\epsilon_t \sim N(0, \sigma^2)$ and $w_1 \sim N(\mu, \sigma^2)$, where $N(\mu, \sigma^2)$ is conventional notation denoting a Normal distribution with mean μ and variance σ^2.

Prior distributions

- $\mu \sim \mathrm{N}(\frac{1}{2}(\max(\mathbf{w}) + \min(\mathbf{w})), \mathrm{range}(\mathbf{w})^2)$;
- $\sigma^2 \sim \mathrm{IG}(2.5, \beta)$, where IG denotes the inverse gamma distribution;
- $\beta \sim \Gamma(0.5, 4/\mathrm{range}(\mathbf{w})^2)$;
- $\alpha \sim \mathrm{N}(0, 0.25)$, with the autocorrelation restricted to lie between -1 and 1.

The posterior distributions of the parameters μ, σ^2 and α were estimated using a Metropolis-Hastings sampler.

Hidden Markov model

In a hidden Markov model (HMM):

- Each observed data point is associated with a state
- States form a Markov chain
- States are unobserved
- States can determine the probability density function of the data point

Other examples: used in speech recognition (e.g. you may have multiple speakers on a recording), and economics (e.g. the economy could be in boom or bust)

State specification

We let the hidden states represent whether a banknote belongs to the criminal activity notes (c), or the background (general circulation) notes (b). A summary of the states is given below:

State (s)	Previous note	Current note
1	b	b
2	b	c
3	c	b
4	c	c

Transition Matrix

The transition matrix of the hidden states is given by:

$$P = \begin{array}{c c} \text{States} & \begin{array}{cccc} bb & bc & cb & cc \end{array} \\ \begin{array}{c} bb \\ bc \\ cb \\ cc \end{array} & \begin{pmatrix} 1-p_{01} & p_{01} & 0 & 0 \\ 0 & 0 & p_{10} & 1-p_{10} \\ 1-p_{01} & p_{01} & 0 & 0 \\ 0 & 0 & p_{10} & 1-p_{10} \end{pmatrix} \end{array}$$

This gives the probability of passing from one state to another

We assume that $(z_1, z_2, \ldots, z_n)$, the measurements on the seized banknotes, come from a hidden Markov model given by:

$$z_t - \mu_{s_t} = \alpha(z_{t-1} - \mu_{s_{t-1}}) + \epsilon_{s_t}$$

where

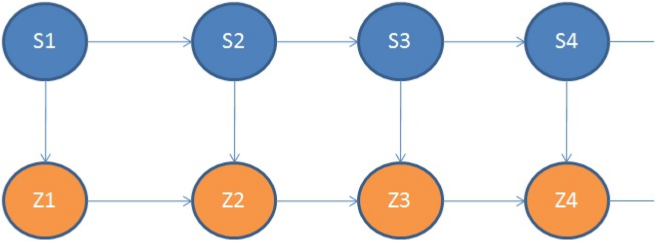
$$\epsilon_{s_t} \sim N(0, \sigma^2_{s_t}),\ \text{for } t \in (1, 2, \ldots n)$$

and,

- $(s_1, s_2, \ldots, s_n)$ are the hidden states, with $s_i \in [1, 2, 3, 4]$
- The subscript s_t denotes the parameter value in state s_t

Hidden Markov model

The Bayesian network of the hidden Markov model is given by:



- Different contamination levels on different bundles are taken into account via the hidden states. There is one hidden state for each banknote.

Parameter Estimation

- To calculate the likelihood ratio, need to estimate parameter θ_C (associated with proposition H_C) and parameter θ_B (associated with proposition H_B) for each of the parametric models.
- Used Bayesian approach with priors on all parameters and a Metropolis-Hastings sampler.
- Posterior distributions of parameters obtained for each individual sample in each of the training datasets.
- Likelihood for the hidden Markov model can be calculated using forward algorithm (Rabiner 1989). This sums out the states.

Nonparametric models

$$f_{D_i}(w_1, w_2, \dots w_n) \quad = \quad f_{D_i}(w_1) f_{D_i}(w_2|w_1) \dots f_{D_i}(w_n|w_{n-1})$$

The marginal density function $f_{D_i}(w_1)$ is estimated by a univariate kernel density estimate.

$$\hat{f}_{D_i}(w_t|w_{t-1}) = \frac{\hat{g}_{D_i}(w_t, w_{t-1})}{\hat{r}_{D_i}(w_{t-1})}.$$

$$\hat{g}_{D_i}(w_t, w_{t-1}) \quad = \quad \frac{1}{(n_{D_i} - 1)h_1 h_2} \sum_{j=2}^{j=n_{D_i}} K_1\left(\frac{w_t - w_{i,j}}{h_1}\right) K_2\left(\frac{w_{t-1} - w_{i,j-1}}{h_2}\right)$$

and

$$\hat{r}_{D_i}(w_{t-1}) \quad = \quad \frac{1}{(n_{D_i} - 1)h_3} \sum_{j=2}^{j=n_{D_i}} K_3\left(\frac{w_{t-1} - w_{i,j-1}}{h_3}\right),$$

Standard model

$$\text{LR} = \frac{\sum_{i=1}^{m_C} \left(m_C \sqrt{\tau_C^2 + n\lambda_C^2 s_C^2} \right)^{-1} \exp\left[-\frac{n(\bar{\mathbf{z}} - \bar{\mathbf{y}}_i)^2}{\tau_C^2 + n\lambda_C^2 s_C^2} \right]}{\sum_{i=1}^{m_B} \left(m_B \sqrt{\tau_B^2 + n\lambda_B^2 s_B^2} \right)^{-1} \exp\left[-\frac{n(\bar{\mathbf{z}} - \bar{\mathbf{x}}_i)^2}{\tau_B^2 + n\lambda_B^2 s_B^2} \right]}$$

Likelihood ratio for parametric models

- Posterior distributions were obtained for each individual sample or exhibit. Denote the two parameters for sample i by θ_{C_i} and θ_{B_i}.
- Need to combine these into overall estimates for θ_C and θ_B.

Write numerator of likelihood ratio, where $\mathbf{y}$ is training set associated with H_C, as

$$
\begin{aligned}
f(\mathbf{z} \mid H_C) &= \int_{\Theta_C} f(z_1 \mid \theta_C) \ldots f(z_n \mid z_{n-1}, \theta_C) f(\theta_C \mid \mathbf{y}) \, \mathrm{d}\theta_C \\
&\approx \sum_{i=1}^{i=70} v_i \int_{\Theta_{C_i}} f(z_1 \mid \theta_{C_i}) \ldots f(z_n \mid z_{n-1}, \theta_{C_i}) f(\theta_{C_i} \mid \mathbf{y_i}) \, \mathrm{d}\theta_{C_i}
\end{aligned}
$$

Integrals can be estimated using Monte Carlo integration.

Likelihood ratio for nonparametric models

The numerator of the likelihood ratio for the nonparametric models is estimated similarly by:

$$f(\mathbf{z} \mid H_C) \approx \sum_{i=1}^{i=70} v_i \hat{f}(z_1 \mid H_C) \ldots \hat{f}(z_n \mid z_{n-1}, H_C)$$

The functions $\hat{f}$ are nonparametric density estimates, based on the training data. Two different bandwidth selection methods were used.

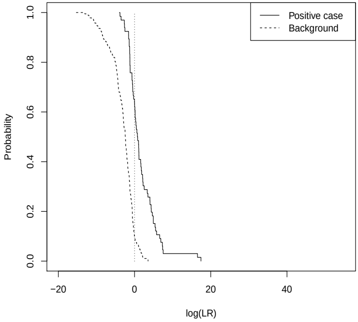
Rates of misleading evidence

	Crime exhibit	General circulation
Hidden Markov model	0.37 (25/67)	0.10 (18/188)
AR(1) model	0.37 (26/70)	0.15 (29/192)
Nonparametric fixed bw	0.27 (19/70)	0.32 (62/193)
Nonparametric adaptive nn	0.26 (18/70)	0.27 (52/193)
Standard model	0.50 (35/70)	0.14 (26/193)

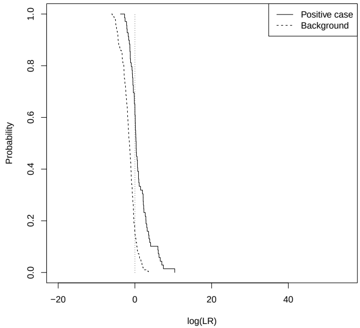
Table: Rates of misleading evidence out of (.) samples

Tippett plots - parametric

Hidden Markov model

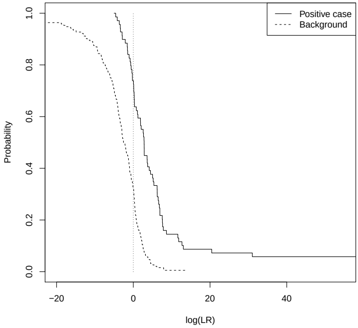


AR1 model

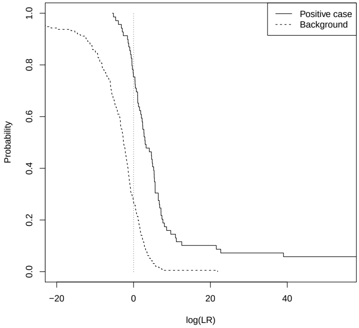


Tippett plots - nonparametric

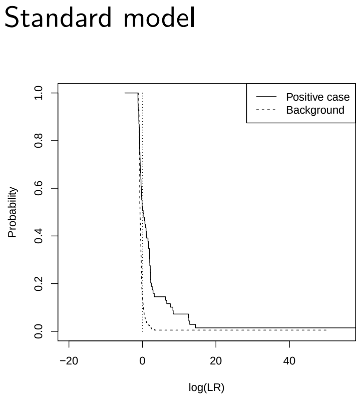
Nonparametric model - fixed bandwidth



Nonparametric model - adaptive bandwidth



Tippett plot - standard model



Comparison to a forensic expert

Exhibit number	HMM	AR(1)	Nonparametric fixed bw	Nonparametric adaptive nn	Standard model
1	7.37	6.05	31.02	39.02	32.61
3	3.51	3.67	5.19	6.43	4.68
16	6.61	7.51	6.92	7.14	2.89
23	7.51	6.32	8.64	7.64	7.72
38	5.38	6.64	11.61	12.55	7.39
39	7.31	10.39	20.43	22.69	8.51
40	4.91	2.24	0.05	21.53	0.60
42	4.35	4.09	6.23	8.03	2.47
43	6.89	7.06	6.80	8.61	2.06
57	4.67	3.58	6.24	11.13	5.45
67	16.52	0.57	244.80	262.25	7.51
69	17.42	0.48	128.69	169.64	5.44

Table: Log likelihood ratio values for 12 crime exhibits assessed by experts as being contaminated.

RSS Section on Statistics and the Law

- Started 2015, evolved from a working group of the same name.
- Working group produced four reports on 'Communicating and interpreting statistical evidence in the administration of criminal justice.' Topics:
  1. Fundamentals.
  2. DNA profiling.
  3. Inferential reasoning: Wigmore charts and Bayesian networks.
  4. Case assessment and interpretation.

  Available from www.rss.org.uk/statsandlaw

Legal Framework - Lord Thomas CJ; 2015

'In whatever system forensic evidence is given it is necessary to ensure that

- the expert evidence has a reliable scientific base;
- the scientists giving evidence are themselves reliable;
- the ambit of the expert's opinion is properly understood (issue to be addressed and the strength of the evaluative opinion);
- the system for collecting the evidence and safeguarding it during analysis provides clear continuity and
- the expert evidence is explained to the judge or jury in a way that they can properly assess it.'

Legal Framework - Lord Thomas CJ; 2015

Strength of an evaluative opinion

'A scientist is entitled and in most cases must express an evaluative opinion as to the conclusion to be drawn from the primary facts on which he gives evidence.'

'More difficult, however, is the question as to the extent to which such an evaluative opinion can be based on a numerical approach . . . It is an issue, however, that needs to be addressed.'

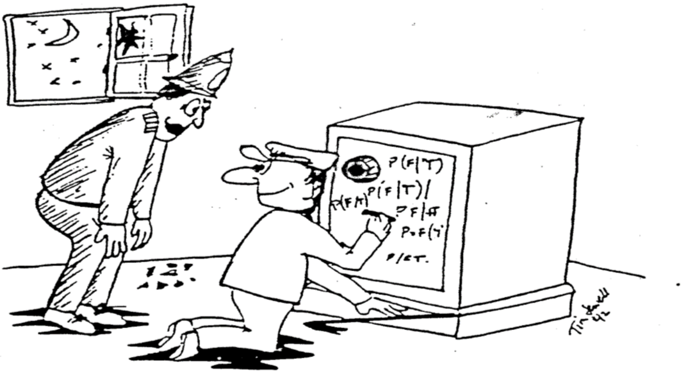
Bibliography 1

- Darboux,J.G., Appell,P.E. and Poincaré,J.H. (1908) Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau. In *L'affaire Drefus - La révision du procès de Rennes - enquête de la chambre criminelle de la Cour de Cassation.* Ligue française des droits de l'homme et du citoyen, Paris, France, 499-600.
- Good,I.J. (1979) Studies in the history of probability and statistics. XXXVII: A.M.Turing's statistical work in World War II. *Biometrika*, 66, 393-6.
- Lindley,D.V. (1977) A problem in forensic science. *Biometrika*, 64, 207-213.
- Lord Thomas CJ (2015) The legal framework for more robust forensic science evidence. *Phil. Trans. R. Soc. B*, 370: 2014258. http://dx.doi.org/10.1098/rstb.2014.0258.

Bibliography 2

- Peirce,C.S. (1878) The probability of induction. *Pop. Sci. Monthly* Reprinted (1956) in *The World of Mathematics*, volume 2, Ed. J.R. Newman, pp. 1341-54. New York: Simon and Schuster.
- Wilson,A., Aitken,C.G.G., Sleeman,R. and Carter,J. (2015) The evaluation of evidence for autocorrelated data with an example relating to traces of cocaine on banknotes. Applied Statistics. 64, 275-298.

The data



"Granted that there are thirty-five other safebreakers in the area that could have done the job, I don't think Bayes' Theorem is going to be any help to you in this case, Fingers."