ARTICLE

Guilty or Not Guilty?

Dr. Richard Sleeman, Scientific Director, Mass Spec Analytical Ltd

When evaluating evidence for use in a court of law, evidential data are presented together with two competing propositions which may account for this data: one from the prosecution and one from the defence. The aim of the court is to evaluate which of the two propositions is more likely. In other words, is the probability of the prosecution hypothesis given the evidence bigger than the probability of the defence hypothesis given the evidence?

We want to know whether the ratio of these probabilities is greater than one (overall support for the prosecution hypothesis), or less than one (support for the defence). This ratio can be thought of as the odds after taking the evidence into account. It can be written as the product of two terms: the odds before seeing any evidence and a ratio known as the Likelihood Ratio (LR). The likelihood ratio converts the odds before seeing any evidence to the odds taking the evidence into account. Thus, a likelihood ratio of greater than one means that the evidence has increased the odds in favour of the prosecution hypothesis.

A forensic scientist is concerned only with the evidence that he or she must analyse. Questions of what the odds were before involving the evidence are the responsibility of the judge or jury in the case. The likelihood ratio can therefore be used by a forensic scientist as a measure of the value of the evidence they present.

The presence of cocaine on banknotes is often presented in Courts of Law as evidence of involvement with criminal activities involving cocaine. However, it is well known that cocaine is present on banknotes from general circulation, in the form of environmental contamination. This leads to great difficulties in interpreting the evidence; difficulties which are further exacerbated by the fact that many banknotes known to be associated with crimes are not contaminated any more than those from general circulation.

There is a clear need to develop statistical methodology using the likelihood ratio framework to evaluate evidence in relation to cocaine traces on banknotes. Researchers from the University of Edinburgh and Mass Spec Analytical Ltd. in Bristol have tackled this problem by analysing two “populations” of banknotes; one from general circulation (the “Background dataset”), and one known to be associated with criminal activity (the “Crime dataset”). Complex mathematical modelling has been used to calculate Likelihood Ratios for seized exhibits, comparing them against the two populations. Two different mathematical models have been presented in peer reviewed publications; an “Autoregressive model”[1] and a “Hidden Markov Model”[2]. A computer programme has been written which performs millions of simulations and produces Likelihood Ratios for seized exhibits using both models.

Both models take into account the fact that cocaine transfers between surfaces. This means that contamination on one note has an effect on the readings from its neighbours with which it is in contact. This is known as “autocorrelation”. Failing to model this can result in misleading conclusions. Consider a simple example: an eye witness report that a criminal had blue eyes and blond hair. Assume that the proportion of people in the population is 0.4 (40%) with blue eyes, and 0.2 (20%) with blond hair. The suspect has blue eyes and blond hair, so what is the probability of the eye witness report if the suspect was not the perpetrator? If correlation is ignored, we simply multiply the two probabilities: i.e. 0:4 x 0:2 = 0:08. (an 8% chance). But perhaps we know that there is a correlation between having blond hair and blue eyes; for example assume that someone with blond hair has a 0.8 (80%) chance of having blue eyes. Taking this correlation into account, the probability is now: 0:8 x 0:2 = 0:16 (a 16% chance). The probability is twice as big if correlation is accounted for! This is a simple example, but the same issues apply with cocaine traces on banknotes.

Another problem which needed to be addressed by the model is that most samples and exhibits consist of multiple bundles of cash which often display different levels of contamination. Clearly, an exhibit with all banknotes having around average contamination should have a very different Likelihood Ratio to one where around half of the banknotes have very high contamination, and the other half have relatively low contamination, even though the average level may be the same.

So what are the prosecution and defence propositions? These have to match the data used to produce the models. The “Crime dataset” is composed of banknotes known to be associated with someone who was convicted of a crime involving cocaine.

The “Background dataset” comprises banknotes known to be taken from general circulation.
For the background dataset, the assumption is that banknotes taken from general circulation have the same distribution of cocaine contamination as those associated with a person who is not involved with criminal activity involving cocaine. The propositions are, therefore:

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

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

Note that with the Likelihood Ratio framework, we are not saying that one of these propositions is true or false. We are saying that, now the evidence has been analysed, the odds that the prosecution proposition is true over the defence proposition being true have changed by a factor equal to the Likelihood Ratio, compared to our position before the evidence was analysed.

As with any mathematical modelling, there are limitations. For example, a statement from the suspect stating where the banknotes are from changes the defence proposition, and hence (depending on the statement) may change the Likelihood Ratio. For example `I got the banknotes from the sale of a used car' would make the dataset used for this analysis void (you would need one from the sales of used cars rather than banks). Where there are multiple suspects, a likelihood ratio which is greater than one in this situation only provides support for the proposition that at least one but not all of the suspects from whom the banknotes have been seized is involved with a crime involving cocaine, not that the entire group is involved with a crime involving cocaine.

The models are not perfect, and never will be; the views and opinions of an experienced analyst will always be better than any algorithm. But they do represent a major step forward in the interpretation of complex data, which will greatly assist in placing such evidence in context, and assisting a Court to give the evidence the appropriate weight.

[1] Wilson, A., Aitken, C., Sleeman, R. and Carter, J. (2014), The evaluation of evidence relating to traces of cocaine on banknotes. Forensic Science International, Volume 236 , 67 - 76.
[2] Wilson, A., Aitken, C., Sleeman, R. and Carter, J. (2015), The evaluation of evidence for auto-correlated data in relation to traces of cocaine on banknotes. Applied Statistics, 64, 275-298.