

Probabilistic Decision Support Systems in the criminal justice system: draft report

Department of Statistics,
University of Warwick

3rd and 4th January 2024

Contents

Introduction

This workshop brought together statisticians, social researchers, forensic scientists and lawyers from government departments, police forces and universities. Colleagues from the Netherlands, Denmark and Switzerland provided a wider view. A range of statistical graphical models and probability trees, and integrated decision support systems in forensic work were presented. Applications included police procedures and predictive policing, finger print evaluation, and crime investigations. The approaches of the UK Forensic Science Regulator, and the European Network of Forensic Science Institutes to managing and reporting forensic evidence were considered.

The support of the Royal Statistical Society through the Mardia Prize 2023 is gratefully acknowledged.

1 Overview

The workshop began with participants briefly introducing themselves.

Roberto Puch-Solis (Principal Investigator, Leverhulme Research Centre for forensic science) then discussed how to integrate existing probabilistic decision support systems into forensic practice. The core requirement is full collaboration between forensic scientists, statisticians and software engineers in defining particular questions, developing statistical models and producing software which can be part of the work flow of forensic science. Developing relevant statistical methods is possible only if there is substantial involvement of forensic scientists so that methods are fit for purpose, with standard operating procedures and appropriate training and verification. Data quality is one aspect of standard operating procedures. Statisticians should create clear Statistical Specification Documents, which help

to ensure that methods are robust. Some aspects of forensic work might still require specialist statistical and forensic input, so operating documents should include advice on when further expert knowledge should be sought. Two major challenges are hiring a software engineer, and acquiring funding for development. Other issues include maintaining continuity of responsibility for methods and software, and assuring software for which proprietary claims are made.

Danyela Kellett (Head of Forensic Services for Lancashire Constabulary) described how in-house police work makes use of external expertise in case assessment and interpretation. Forensic scientists need to be transparent about the data and the accuracy of findings, following standard operating procedures. Likelihood ratios (LRs) are widely used to interpret forensic evidence. Despite not knowing the alternative explanations of evidence which the defence will propose, scientists have to consider which explanations are supported by the data, and the prior odds ratios to get LRs. The difficulties encountered include lack of suitable reference data, differing methods of calculating LRs, validating methods and verifying results. Evidence from various sources, some complex has to be combined and presented in reports. These reports have to be accessible to several audiences: police, solicitors, barristers, judges, and jury. Statistical methods contribute to clarity and reliability at most stages of the process.

Lisa Hall (Metropolitan Police Force Fingerprint Consultant) considered how probabilistic models for Friction ridge details (FRD) might help with volume of prints processed (80,000 marks per year), if solutions were scaleable. The two main questions are the source of the FRD in a print, and the activity which led to the print. The evidence from FRD for identification of people has been based on a categorical scale since 1901, but there are no standard measures of uncertainty or agreement on the accuracy of identifications or eliminations. Various statistical models are being developed, and decisions on reliability and validity, practical utility, comparisons between models and FRD experts will affect which models can be adopted. Theory and practice in medical diagnostic testing could inform approaches to identification. To answer these questions, suitable data, statistical tools and sufficient interest are necessary. The lack of UK data and lack of detailed statistical investigation of the MPS database hamper efforts to improve the speed with which initial putative identification is made.

Melissa Hamilton (Professor of Law & Criminal Justice, University of Surrey School of Law) addressed probabilistic risk assessment in criminal justice, where the focus is on the future. Decisions about the risks which people pose to others can use statistical models to predict recidivism from data on past offenders. A person journey through the criminal justice process involves several decisions: arrests, releasing suspects, sentences given, security classification, parole and final release. Algorithms or statistical models are beneficial if their use results in reduced detention rates, reliance on bail, and human biases while increasing accountability and effective use of resources. However, risks can be expressed in several ways, with equivalent information interpreted differently. Data science, law and psychology do not agree on how to measure risk, and define high risk. The balance between false positives and false negatives is only one aspect of risk. Decisions will ideally take severity, time scales and frequency of adverse acts into account.

Julia Mortera (retired Professor of Statistics, University Roma Tre, Italy) illustrated statistical decision support systems for simple and complex problems in police investigations. It

is rare to have evidence which leads to indisputable conclusions, and probability provides the science for dealing with uncertainty. Forensic identification might assess the presence of a single individual, or of several, whether in terms of traces left at a scene, or disputed family relationships. Graphical representation of the plausible relationships between a set of variables, such as items of evidence and hypotheses of interest through Bayesian Networks allows the dependence between factors to be illustrated and the uncertainty assessed using probability. A simple example of eyewitness reports of the colour of a taxicab introduced the idea, which was extended to include paint flakes and different fleets. The murder of Meredith Kercher in Italy provide an interesting example of complex evidence.

Jim Smith (Professor of Statistics, University of Warwick) has developed graphical frameworks to support police decision-making to counter crime, to pursuit of possible suspects, and to consider sequences of events relevant to deciding on forensic analysis of existing crimes. The plausible stories considered by police analysts to explain the series of events in a crime, or effective ways to present evidence in court can be expressed as a series of events. Chain event graphs (CEGs), a class of models which generalise discrete Bayes Nets, can express such stories in an event tree with probabilities associated with the possible events. As evidence accumulates, the staged tree can be updated and revised likelihood ratios calculated. The events which represent prosecution and defence representations can be embellished with alternative explanations and Bayesian inference used for effective comparisons. The relevant likelihood ratio is the ratio of the product of edge probabilities along the prosecution narrative to that for the defence narrative. Complex elements of evidence can be explored using subgraphs, and can be used with other models. The main challenge is that the graphs get big very quickly. The approach was illustrated using the Meredith Kercher murder.

Gill Tully (Professor of Practice for Forensic Science Policy and Regulation, King's College London) gave an overview of the development of regulatory guidance for forensic science in England and Wales. Five categories of information are considered: Factual, Investigative, analytical, categorical and evaluative, though evidence can fall into several categories. Evaluative opinions should follow a specific examination strategy, identify the issues to be addressed and consider the competence of experts, the assumptions made and how methods are validated as well as a clear statement on the sources of data. Various problems remain. Determining the boundaries between facts, inference and opinion is not always straightforward. If there is no defence proposition, it is not clear how best to assess the weight of evidence for prosecution and defence. If the framework within which evidence is evaluated is incomplete, a forensic scientist can only provide a limited evaluation. Logical reasoning when there is uncertainty should use probability. Lack of transparency, of access to peer review and misunderstanding of subtle issues and logical fallacies are further challenges. A consensus has developed in favour of using likelihood ratios as the appropriate framework which can work across a range of evidence types. Guidance can encourage robust science approaches and procedures, as well as transparency and clarity of conclusions. It might be possible to suggest approaches to use in the absence of a defence proposition, but preliminary work shows there are no simple solutions. Culture change and access to relevant high quality datasets is difficult.

Alex Biederman (Associate Professor, School of Criminal Justice, University of Lausanne (UNIL), Switzerland) provided a view from continental Europe. The theoretical background of the European Network of Forensic Science Institutes (ENFSI) Guideline for Evaluative

Reporting in Forensic Science comes from several monographs and academic journals, published from 1989. The practical context is the need for an audit template to help bridge the gap between mathematical results and a substantive application. An evaluative report (not investigative, intelligence or technical reports) requires an instruction to examine or compare material in order to evaluate the support for competing propositions, and assign a likelihood ratio in the context of the key issues on which a judgement is sought. The guidance assists forensic scientists in the preparation of defensible reports, which are balanced, transparent, logical and robust. Assignment of probabilities must be based on an available body of knowledge, and understanding of the type of data used. The ENFSI guideline has a unique form and content, and resistance is being addressed by showing how change benefits forensic scientists. A summary of the current status in Switzerland concluded the presentation.

2 Small group discussions

2.1 Bayesian Networks

Julia illustrated some simple and slightly more complex examples of Bayesian Networks to Danyela Kellett and Davin Parrot.

They both found them very useful for their case work and research. We discussed the use of decision support systems to aid in policework to simplify the investigation and exclude pathways that are redundant. One issue I raised was to explore how they can be implemented so as to combine different elements of evidence (for example footwear pattern evidence, fingertips, DNA etc.) A problem that arises in this case is that we cannot simply multiply the resulting likelihood ratios obtained for each piece of evidence, as the propositions in each case are usually different. We thus would need some general overall proposition that could lead into each specific proposition.

2.2 Knowledge transfer partnerships

Knowledge transfer partnerships are grant-funded projects which aim to transfer and embed academic knowledge in external entities. It is prudent to start with an Impact accelerator piece of work and funding. This provides about £30,000 to £60,000 so that collaborators can identify the potential and scope of a full project.

Three areas with potential for projects were discussed: friction ridge details, use of AI in reviewing images, and analyses of extensive records held by the Crown Prosecution Service and Ministry of Justice.

Developing models for friction ridge details and match probabilities is timely, as most fingerprint bureaus, including the Metropolitan Police Service (MPS), are moving to a digital workflow in the next two years. The MPS and others have invested in *the Xchange*. This is a fingerprint specific software application that enables FRD to be captured from a scene or laboratory and forwarded to the bureau within seconds. The examiners have the digital tools to undertake an onscreen comparison. The Xchange is hosted by Police Digital Services

in the Home office. Most bureaux are expected to implement the technology within the next few years. There is then scope to develop tools that could be accessed from this platform.

Decision support systems require data. The National fingerprint collection, FINDS, is managed by the Home Office. This consists of 8.6 million sets of tenprint forms or FRD from arrestees. The bureaux search the database from terminals housed within their own officers. This database is a rich vein of data. An essential task in the impact accelerator phase would be agreements for data access. A project which was able to provide even basic frequency data from FINDS, such as numbers of men and women, and the common pattern types on each digit would assist in estimating the frequency of particularly characteristics. The search algorithm managed by FINDS generates a match score. As not all FRD on the system are searched, work on converting a match score into a form more useful for evidential purposes is a further task. Assessment of models already in use in other countries, such as Xena freeware in Switzerland would inform developments. A further opportunity is developing and delivering basic training on evaluative models, using verbal scales or using probabilistic decision support models.

The use of AI to review images is linked to FRD. Two aspects to explore are:

1. Can we apply technology to consider which developed areas of FRD on items should be submitted for examination? The aim would be to develop a mechanism that triages or filters useful ridge detail. Only useful FRD would be submitted from the laboratories or from scenes for inclusion in evidence.
2. Can we apply models to identify images from video footage or from libraries that depict the hands or feet of individuals. We have the capability to identify victims and perpetrators of crime (on line child exploitation) from images where we can observe the underside of an individual's hand or foot. Instead of humans viewing and selecting suitable images can we get AI or data models to sift this? This is a new and emerging service and is deployed when police are unable to identify a person in the image. This service is also offered for images on the dark web where criminals are selling drugs or weapons online and they proudly show it off placed on their hands. But the big market here is for child exploitation and abuse cases.

Exploring and summarising Ministry of Justice and Crown Prosecution Service data is relevant to improving the efficiency and effectiveness of services, and evaluating the performance. This was not the focus of this workshop, but was considered, and links to the MoJ Data First Project and a Virtual Study Group on Mathematics for Justice made.

2.3 Validation of methodology for evaluating evidence

The issue of validation was addressed rather broadly. Many aspects can be validated: the end-to-end process (often referred to as a method), specific software, or some of the scientific technical components or methods underpinning the software.

A key element in validations is ground truth tests. Those tests may be designed in many ways, but the general aim is to explore the space of cases for which it is of interest to validate the methodology. As that space of cases is enormous, we might want to direct the attention to particularly difficult cases. Note that, evidently, results depend heavily on the batch of cases included in the test. In particular, this means that we need to be very careful that the

associated measures of performance are not bluntly generalized to a measure of the overall performance of the reliability.

Another key element is to set out acceptance criteria of some sort; they do not have to be quantitative statements. Further, it is important to address limitations and risks and how to handle them. Approaches from medical diagnostic tests, with published guidelines, might inform discussions.

Any validation should rely on some kind of independent expert review. It is unclear how this should be achieved – should it be “free labour” as for academic journals, or would an expert be instructed and given remuneration? There are pros and cons for both solutions.

Related to this discussion is where validation documents should be made available. Validation studies are not necessarily scientifically novel nor broadly applicable (so not appropriate for scientific journals), and many will be incremental of nature as modifications to systems become available. Some kind of stable reference (url, doi) seems desirable and for academics it is generally important that the huge effort can be recognized in a scientific publication or similar. Possibly a new publication venue should be established, but it is worth noting that Forensic Science International has already established a Reports series that may be suitable for validation studies.

Finally, we discussed also that some of the published guidelines should be updated to match state of the art in software development – this feedback will be given to the appropriate instance.

2.4 Guidance for judges and jury members in relation to evidence containing statistics

The group considered a proposal to provide judges with a simple decision tree based models that would assist them in evaluating forensic evidence. Although only some elements of such a decision tree concern statistics, one which encompassed all steps was discussed. This would extend from the question as to whether the evidence is relevant to the case through to whether an alternative proposition has been considered.

We also discussed whether the presence or absence of an alternative proposition should be considered by the judge prior to evidence being presented to the jury, or during the evidence being presented, and at what point and to what degree would it be expected that a judge becomes involved with a point that concerns statistics. Once this has been determined, the scope of training that could be provided to judges could be evaluated. In terms of implementation, it was made clear earlier in the day by a speaker that introducing statistics to the bar exam was a method already attempted but had not been adopted by the majority of organisations. The only other method would be for the training to be optional, potentially through the Royal Statistics Society, and using Continuing Professional Development points as an incentive.

For juries, we discussed the creation of a short and simple video that presents broad statistical concepts a jury is likely to hear in the form of analogies that relate to their time as a jury member. This could then be made available through a repository and the judge make it

available to the jury – either by replaying it in court or providing it to them for review when they are deliberating. The challenge in doing this would be two fold – firstly finding the resources to create such a video, including the time of experienced judges, and secondly, making judges aware of its presence and the benefits it would afford the court (informing juries, remove the need for multiple experts to explain statistics, known provenance, can be watched by the jury outside of the courtroom).

In short, the discussion led to the proposal of:

1. A decision tree based model to assist judges in accepting and questioning evidence;
2. A video presentation that presents key concepts to juries in a simple manner to assist their understanding and thus interpretation of statistics.

3 Feedback from participants

All participants made positive comments. The fascinating workshop was well run and structured well, with a nice balance between group work, talks, and general conversation. The flexibility of schedule was appreciated, with topics planned for small groups discussed. A small group of participants is very useful for interaction among people. It was interesting to include different countries and legal systems. Connections were made with a diverse group of practitioners.

Suggested improvements were to have a group session where people work on an actual case and report solutions; more time for discussions; and having a judge or barrister speaking about what they perceive to be the strengths & weaknesses of how forensic evidence is presented & communicated.

Julia Mortera We need to consider how to present a case study so that more participants get involved. I learned a lot about the various methods used in the UK for analyzing and reporting fingerprint and footwear pattern evidence, as well as the guidelines and rules in forensic science regulation in the UK. I think the workshop would have benefited by having more law scholars, barristers and judges both for presenting and for the discussions.

Danyela Kellett I thought the perspectives of how it works (or doesn't) in the UK versus what is in place elsewhere was really useful, although I think there is always the caveat of different legal systems across Europe. I think it would have been useful to have had a judge or barrister speaking about what they perceive to be the strengths & weaknesses of how forensic evidence is presented & communicated. I was reassured that there was consistency amongst all attendees as to what the issues were. I would be interested to explore further what academia can do to support policing and forensic scientists in better presenting forensic findings & working with the legal profession to understand their needs & provide training.

Michael Fletcher is an Ofqual expert on probability theory, whose role is to review syllabus submissions and give critical advice on specimen examination papers. He also

writes statistical puzzles for *Significance*, to entertain and educate. He and his wife attended to learn about the use of probability theory in forensic science.

In 2014 when statistics, and consequently Bayes' Theorem, became part of the A level mathematics syllabus, Ofqual with the help of the Winton centre advised the examination boards to suggest, in their programmes of study, teaching Bayes' Theorem using natural frequencies.

Michael achieved all his aims, including creating some material for NRICH, and writing a puzzle for the RSS and a teaser for the Sunday Times using the theme of forensics.

Ruoyun Hui I have a better idea about the regulatory space now, especially around what “validation” entails. Most people seem on board with wider adoption of probabilistic models but constrained by the opportunities for engaged collaboration. Other smaller points that stayed with me: the importance of software engineers; exceptionalism among digital forensics; challenges around a lack of defence proposition.

Rowland Seymour noted that the workshop was quite forensic heavy. As his work is more in the area of crime intelligence and not forensics, he will be invited to contribute to the workshop on “Understanding of legal systems”. This will provide a forum to discuss intelligence data such as crime reporting, or secondary data from finance or internet companies.

Peter Elston attended from general interest. The workshop helped him to know where to focus his attention in relation to statistics and the law.

4 Conclusion and future projects

The workshop successfully brought together professionals from different disciplines and institutions.

The group discussions outlined several research or knowledge transfer projects, with potential for follow on impact funding or collaborations.

It would be useful to identify funding for a further series of meetings, as most people wished to have more time for discussions than was possible in a short workshop.

The major challenges include developing and maintaining good quality data sets, hiring statisticians and software engineers, and acquiring funding for development. A Forensic Information Common Service was mooted. The work of NICE in evaluating which devices or medicines are economically worthwhile might provide a starting point for an approach to enhancing the quality of forensic processes. The UK Data Archive is world-leading with regard to acquisition of data, management, storage and accessibility.

5 Participants

Twenty-nine participants attended the workshop. A list of attendees is presented in the Appendix.

This report will be sent as a draft to the funders, the Royal Statistical Society, and to participants for comment.

Professor J L Hutton,
Department of Statistics,
University of Warwick
April 17, 2024

Table 1: Appendix: List of attendees

Name	Role	Institution
Dr Martine Barons	Director of Applied Statistics and Risk Unit	University of Warwick
Dr Alex Biedermann	Associate Professor	University of Lausanne
Ovidiu Brudan	Senior Social Researcher	Crown Prosecution Service
Kai Budrikas	PhD Student	IT University of Copenhagen
Dr Tim Clayton	Forensic Scientist	Eurofins Forensic Services
Jan de Koeijer	Senior Forensic Scientist	Netherlands Forensic Institute
Mr Peter Elston		Royal Statistical Society
Dr Gemma Escott	Senior Forensic Scientist	Eurofins Forensic Services
Linda Fletcher	Retired teacher	
Michael Fletcher	Retired lecturer	Royal Statistical Society
Dr Therese Graversen	Associate Professor of Statistics	IT University of Copenhagen
Lisa Hall	Fingerprint Consultant	Metropolitan Police Service
Professor Melissa Hamilton	Professor of Law & Criminal Justice	University of Surrey
Dr Ruoyun Hui	Statistician	Alan Turing Institute
Professor Jane Hutton	Professor of Statistics	University of Warwick
Danyela Kellett	Head of Forensic Services	Lancashire Constabulary
Dr Rupert Macey-Dare	Barrister and Economist	Minerva Chambers and St Cross College Oxford
Professor Julia Mortera	University Professor (retired)	Università Roma Tre
Dr Linda Nichols	Assistant Professor	University of Warwick
Davin Parrott	Data Scientist	West Midlands Police
Katie Pottage	Junior Forensic Imagery Examiner	Verden Forensics
Dr Roberto Puch-Solis	Statistician	Leverhulme Research Centre for Forensic Science
Perle Russel	Professional Doctorate Candidate	Hogeschool van Amsterdam
Dr Rowland Seymour	Assistant Professor	University of Birmingham
Professor Jim Smith	Professor of Statistics	University of Warwick
Dr Lizzie Tiarks	Lecturer; Barrister (currently non-practising)	Aberdeen University
Professor Gillian Tully	Forensic Scientist	King's College London
Dr Philip Wilson	Statistician	Eurofins Forensic Services
Mr James Zjalic	Multimedia Forensics	Verden Forensics