



**9th International
Conference on**

**Forensic
Inference
and
Statistics**

Abstracts

Session themes

Theme [1] Evaluating evidence of emerging analytical technologies

New kinds of forensic analysis give results which require novel methods of interpretation.

Theme [2] Epidemiological evidence and the law

Consideration of the use of epidemiologic methods and evidence to resolve problematic questions of causation in law, most notably within the context of personal injury litigation. Epidemiological evidence helps courts deal with alleged causal connections between plaintiffs' diseases or other harm and exposure to specific noxious agents (such as asbestos, toxic waste, radiation, and pharmaceuticals).

Theme [3] Education of methods for evidence interpretation

Efforts to educate legal and forensic practitioners, aimed at improving understanding where it concerns forensic evidence and inference.

Theme [4] Communication between lawyers and experts

How can we improve the way legal and forensic practitioners communicate through reporting and in court? How can the legal practitioner deal with the knowledge paradox: judging the expertise of the forensic practitioner if by definition there is a big difference in knowledge about the field of expertise?

Theme [5] Legal versus scientific reasoning

Where do the ways of reasoning by legal and forensic practitioners differ and what does science have to say about it?

Theme [6] Quantifying evidential value: [6a] DNA; [6b] Chemistry, Fingerprints; [6c] Other

Quantitative ways of assigning evidential value, based on models and empirical data. This includes methods for model improvement and assessment such as calibration, and validation studies of their performance. It involves both general theory and practical applications to a specific field. Views on current debates - such as dealing with various sources of uncertainty when reporting likelihood ratios - are also welcome.

Theme [7] DNA mixture interpretation

How can we assign a quantitative evidential value to a profile of a DNA mixture? This theme includes software to assign evidential value, their performance for different types of mixture and profile quality, and experiences in casework.

Theme [8] Interpretation at activity level

Since legal practitioners will inevitably need to interpret evidence in relation to possible activities, forensic practitioners may increasingly be asked to do so. This theme is concerned with when and how to do that based on hypothesis formation, case circumstances, and available knowledge and data. Views on dealing with a lack of information that is required for properly addressing activity level hypotheses are also welcome.

Theme [9] Combining evidence

Combining evidence is an inferential challenge for legal and forensic practitioners. This theme discusses legal and scientific ways of doing that, e.g., using Bayesian networks.

Theme [10] Deciding under uncertainty

When inference is extended towards making decisions, this will involve the costs and benefits related to the decisions made. This theme is concerned with application and theory of such decision making in the scientific and legal domain.

Theme [11] Statistics as evidence

This theme is concerned with those cases where the evidence is statistical information. For instance, statistical evidence in human rights tribunals, or civil cases where discrimination is an issue. There are counting problems associated with human rights in conflicts such as estimation of the number of civilian casualties.

Tuesday Morning

Workshop Evidence Interpretation Basics

Charles Berger^{1,2}, Colin Aitken³

¹Netherlands Forensic Institute, ²Leiden University, ³University of Edinburgh

This workshop will treat the history, science, and practice of the interpretation, evaluation, and reporting of forensic evidence. It will not focus on any single forensic area of expertise, but on the important things that all of the areas have in common, with many examples

The content will include topics such as:

- Classical principles of forensic science
- Interpreting evidence, evidential value
- Working with hypotheses
- Obtaining a more objective value of evidence
- Modern forensic methodology
- Forensic interpretation
- Common reasoning mistakes: fallacies
- Taking it to the next level: the hierarchy of propositions
- Cases Assessment and Interpretation (CAI)

The workshop will consist of lectures, as well as exercises and discussions. Target audience: Forensic practitioners and other stakeholders such as police and lawyers. The subject will be made accessible to anybody with an interest in truth-finding, whether with the police, in the forensic lab, or in the courtroom.

Tuesday Afternoon

Workshop A New Software Approach for the Exploration of Forensic Reference Data and Calculation of Score-based Likelihood Ratios

Martin Lopatka^{1,2}, Jacob de Zoete^{1,2}

¹University of Amsterdam, ²Netherlands Forensic Institute

Modern technology affords us vast capacities for large-scale data storage, high speed computation of complex models, and near instantaneous data transmission. These tools have facilitated numerous advances in forensic science and created the ability to amass sizeable reference data collections. We demonstrate a fit-for-purpose software application that is designed for the exploration of continuous multivariate data intended for forensic use. The application allows users to quickly evaluate different possible workflows for likelihood ratio estimation. We examine the many steps involved in LR calculations including: feature selection, data pre-processing, choice of similarity metric, and fitting of various distributions (parametric and non-parametric). Performance is appraised using log likelihood ratio cost, quality of fit metrics, class separability, and cross validation performance. This tool may assist users in determining whether their data is suitable as a reference population from which score-based likelihood assertions may be drawn.

The performance of a score-based likelihood ratio, in the context of a batch membership paradigm, is heavily dependent on accurate and complete representation of a reference population for which batch labels are known. It is also notoriously sensitive to numerous parameters that can be involved in data pre-processing, choice of metric, and the fitting of parametric and semi-parametric distributions. Reference distributions relating to pairwise similarity of samples known to originate from a single origin must be compared to a distribution of pairwise similarities for different origin comparisons. Throughout the course of the workshop a brief overview of the development and validation procedure for such software will be summarized. We will present a multitude of real examples from a variety of forensic domains including: illicit drugs, forensic ballistics, digital image forensics, fibre analysis, soil analysis (microbial and elemental), and arson investigation. We will discuss the role of this software for use by

forensic practitioners, for educational purposes, and as a legal aid to demonstrate the sensitivity of LR methods to parameter selection.

The use of score-based likelihood ratios in forensic intelligence tasks will be discussed in the context of a full walkthrough of an investigative workflow from data collection to evidential reporting.

Keynote [2] Causal Inference from Epidemiological Data

Philip Dawid¹, Monica Musio², Stephen Fienberg³

¹University of Cambridge, ²University of Cagliari, ³Carnegie-Mellon University

While Science seeks to understand the effects of causes (EoC), Law is more concerned with identifying the causes of effects (CoE). However, incorporating statistical or epidemiological evidence into CoE reasoning is highly problematic. We give a careful account of the many strong conditions and assumptions required for this, and show that, even with the best possible and most extensive data, in general only interval bounds can be set for the probability of causation. With less than perfect data these bounds will themselves be uncertain, leading to a novel compounding of two kinds of uncertainty. We illustrate these points with a novel Bayesian analysis of a case study in child protection (Best et al., 2013).

- Dawid AP, Musio M, Fienberg SE (2013), From statistical evidence to evidence of causality. arXiv:1311.7513
- Best N, Ashby D, Dunstan F, Foreman D, McIntosh, N (2013), A Bayesian approach to complex clinical diagnoses: A case-study in child abuse (with Discussion). Journal of the Royal Statistical Society, Series A, 176:53-96.

Oral [2] Debunking Some Judicial Myths about Epidemiology and Its Relevance to UK Tort Law

Claire Mclvor

University of Birmingham

The UK courts are highly skeptical about the value of using epidemiological evidence to help determine questions of causation in complicated disease litigation. This paper seeks to address the sources of this skepticism and to highlight the potential benefits of epidemiology to tort law. It identifies and corrects two common judicial misconceptions about epidemiology as a scientific discipline: (i) that it is concerned with bare statistics; and (ii) that epidemiologists treat relative risk results of two or more as definitive proof of causation. The paper then proceeds to demonstrate that the current judicial approach to determining questions of probabilistic causation in tort law (where the standard of proof is the balance of probabilities) is fundamentally flawed and suggests that it could be significantly improved through greater use of specialist epidemiologic expert witnesses. On the issue of expertise, the paper concludes by highlighting some of the problems associated with the current tendency of UK tort lawyers to turn to clinicians as experts on all matters of biomedical science.

Oral [2] Using Epidemiology to Prove Causation

Maurice Zeegers

the Maastricht Forensic Institute

The field of personal injury and tort law is developing rapidly and complex. During the last 5-10 years, there is a rapidly increasing reliance on evidence from scientific medical research in human populations to find proof for causality. In these circumstances the expertise of the Epidemiologist could be helpful. Forensic Epidemiology uses epidemiological and statistical methods for the evaluation and quantification of cause and effect relationships in legal settings. The result of an epidemiological investigation is an evidence-based probability that a suspected relationship between a cause and an effect is true on a more probable than not basis. This calculated probability can support legal decision-making regarding the causation element of proof of liability in civil actions. This lecture will outline how the medical scientific literature can be appraised from an epidemiological point of view, how this evidence can be applied to the individual case, how the chance of injury can be quantified under different scenarios and how to use this information to build an evidence-based argumentation.

The speaker, Professor Maurice Zeegers, is active in teaching, research and practice in the field of Forensic Epidemiology. He is the co-chair of the International Association for Law and Epidemiology and the president of the Netherlands Epidemiology Society.

Oral [2] How Epidemiology Can Inform Legal Decisions: Compensation Claims for Asbestos-related Lung Cancer in the Netherlands

Alex Burdorf
Erasmus MC

Exposure to asbestos is a well-established occupational risk factor for lung cancer. In many countries asbestos-related lung cancer has been acknowledged as an occupational disease. Three classical mistakes may arise in the interpretation of epidemiological knowledge on the association between asbestos and lung cancer in litigation at individual level in order to determine whether asbestos exposure has played a role in developing lung cancer.

The first issue is the interpretation of attribution. In several countries it was estimated that the proportion of lung cancer cases in the population that is attributable to asbestos varies between 2% and 10%. This estimate represents the population attributable fraction (PAF), which is a summary measure of population health. This measure is often confused with the attributable fraction (AF), which represents the proportion of lung cancer cases among those exposed to asbestos, that is attributed to that asbestos exposure. The latter can be informative at individual level, the former is by definition a population measure among exposed and non-exposed and, thus, is useless for asbestos litigation.

The second issue is the relative contribution of different risk factors to a single disease, e.g. smoking and asbestos in lung cancer. Some countries require for notification of an occupational disease that the AF of asbestos exposure is over 50%, whereas The Netherlands has accepted proportional liability whereby the magnitude of AF determines the compensation. The requirement of an AF > 50% in litigation is heavily influenced by the choice of the definition of non-exposed workers, i.e. did they smoke or not? The third issue is the difference between relative risk and risk difference. A high increased lung cancer risk due to asbestos exposure (relative risk) will increase the likelihood to die of lung cancer during lifetime only modestly (risk difference). Again, the risk difference is a measure of population health, and not a suitable measure to evaluate among those with lung cancer that asbestos has played a role.

Oral [2] Forensic Applications of Epidemiology in Civil and Criminal Litigation

Michael Freeman^{1,2}, Maurice Zeegers³

¹Oregon Health & Science University, ²Aarhus University,

³University of Maastricht

Causation is a key feature of both criminal and civil litigation in which the alleged harm is injury or disease in a person or persons. Once it is established that an action (either commission or omission) has occurred and that an adverse health outcome is temporally or spatially associated with the action, there are two questions that must be answered in order for the claim to advance legally. First, the action (alleged "hazard") must be plausibly related to the adverse outcome.¹ Next, it must be demonstrated (to some degree of probability, typically >50%), that absent the exposure to the hazard, the outcome would not have occurred in the individual [1]. The process of answering this question is referred to as specific or individual causation. In non-legal settings the evaluation of specific causation is invariably performed by clinicians, and as such it is rare that a causal determination is ever revisited or challenged. In a legal setting, however, causation is routinely disputed, often leaving judge or jury fact finders with conflicting opinions and no objective means of assessing them. The practice of forensic epidemiology, also referred to as legal epidemiology, is generally described as concerning the

intersection of epidemiology and law. More specifically, forensic epidemiology (FE) provides a systematic approach to the investigation of general and specific causation in civil and criminal matters [2,3,4]. In an FE assessment general causation is first assessed via application of the Hill Criteria,¹ and specific causation is quantified by a Comparative Risk assessment in which the risk of the injury or disease associated with the hazard is compared to the risk of the injury or disease occurring at the same point in time or space absent the hazard, given the relevant predictive characteristics of the individual [5,6]. The result of the analysis may be presented as a Comparative Risk ratio or a Probability of Causation (attributable risk percent). In this presentation we describe several real world examples of applied FE in criminal and civil settings as a demonstration of the practicality and accessibility of the methodology

[1] Reference Guide on Epidemiology, in Reference Manual on Scientific Evidence. 3rd Edition. 2011 National Academies Press, Washington DC.

[2] Freeman MD, Rossignol AC, Hand M. Forensic Epidemiology: A systematic approach to probabilistic determinations in disputed matters. *J Forensic Legal Med* 2008;15(5):281-90.

[3] Freeman MD, Kohles SS. An examination of the threshold criteria for the evaluation of specific causation of mesothelioma following a history of significant exposure to chrysotile asbestos-containing brake dust, *Int J Occ Env Hlth* 2012;18(4):329-36.

[4] Loue S. Forensic Epidemiology: Integrating Public Health and Law Enforcement, Jones and Bartlett Publishers, Boston, 2010.

[5] Koehler S, Freeman MD. Forensic epidemiology: a methodology for investigating and quantifying specific causation. *Forens Sci Med Path* (in press)

[6] Freeman MD, Cahn PJ, Franklin FA. Applied forensic epidemiology. Part 1: medical negligence. *OA Epidemiology* (in press)

Wednesday Morning

Plenary Uncertainty in Law and Forensic Science, Strategies for Coping and Communicating

David Kaye
Penn State University

Plenary English T-time
Paul Roberts
University of Nottingham

Keynote [3] Supporting Collaborative and Reflective Learning Activities on Pre-assessment in Forensic Science

Alex Biedermann, Romain Voisard, Emanuele Sironi, Matteo Gallidabino, Julien Furrer
University of Lausanne

In the late 1990s, forensic scientists in the UK developed an important framework for reasoning to help forensic practitioners and other participants in legal proceedings evaluate both the requirements in a case at hand, and assign a value to potential outcomes of subsequent examinations (Cook et al. 1998). This framework, based on sound probabilistic underpinnings, became known as 'Case Assessment and Interpretation (CAI)'. Its contributions to balance, transparency and logic have recently been reaffirmed in a report issued by the Royal Statistical Society's Working Group on Statistics and the Law (Jackson et al. 2013). Despite the widely recognised potential of CAI to add value to the overall process of forensic evaluation, experience shows that practitioners might not find it straightforward to apply the approach in practice. Similarly, instructors of workshops and training sessions may find it difficult to conceive learning activities that appropriately support participants in acquiring knowledge and competence in pre-assessment.

To help overcome these practical challenges, the authors' project intends to (i) design learning scenarios that emphasise key aspects of pre-assessment (for various domains of forensic practice), and (ii) develop a web-interface to support both individual and group learning activities. This application provides participants with a means to clarify their reasoning in the various steps of pre-assessment, to structure their argument and to document their analyses. In turn, instructors are supported in their task of supervising the participants' work (e.g., compare results from participants both within and between different study groups),

recognise difficulties and collect other relevant information for giving constructive feedback.

To ensure a wide share of this approach, the web interface is designed in an open format. It thus is well suited for collaboration with external partners who may wish to engage in joint learning and training activities. This supports the overall aim of contributing towards common understanding and proficiency in CAI.

- Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA, A Model for Case Assessment and Interpretation. *Science & Justice* 38 (1998) 151-156.
- Jackson G, Aitken C, Roberts P, Case Assessment and Interpretation of Expert Evidence. *Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. Royal Statistical Society's Working Group on Statistics and the Law, 2013.

Oral [3] A Practical Solution to Training U.S. Forensic DNA Practitioners on Implementing Probabilistic Approaches to Weighting Forensic DNA Evidence

Keith Inman^{1,2}, Norah Rudin²

¹*California State University,* ²*Forensic DNA Consulting*

Although leading authorities have made clear that forensic DNA evidence interpretation must transition to probabilistic methods, very little practical direction exists. In the U.S. in particular, which has lagged behind in implementing appropriate statistical approaches, forensic DNA practitioners have finally come to accept that binary methods for interpreting complex profiles are not only insufficient, but dangerous. However a void still exists with regard to practical, formal instruction for implementation of probabilistic methods in DNA interpretation. While a selection of software tools are now available for this specific task, most are produced by commercial companies that do not, and cannot, offer the appropriate education to competently use the tools in a forensic casework setting. Required prerequisites to competent use of a software tool include understanding the basics of a holistic casework approach, which incorporates probabilistic approaches to weighting evidence and formulation of relevant competing hypotheses. Absence of this competency leaves a large void in the expertise required to intelligently choose an appropriate software solution. Certainly high-level academic instruction can be found, but many laboratories and practitioners are not in a position to undertake extended academic study. Over the past few years, we have developed a practical solution to this dilemma. In various venues around the U.S. we have been offering an intensive training program that can vary from 1 to 3 days, depending on the wishes of the agency or group. We have developed a modular curriculum that can be adapted to various venues. While we do provide training on the use of our free-of-charge open source software tool, Lab Retriever, the education is independent of the software and could easily be offered separately if requested.

We will present the specifics of the curriculum we have developed, and discuss the implementation and response from several groups to whom we have presented the training.

Oral [3] A Review of Recent Forensic Document Examiner Training in North America Relating to the Logical Approach to Evidence Evaluation

Brent Ostrum

Canada Border Services Agency

A critical component in the adoption of a logical approach for evidence evaluation by forensic practitioners is expert-level knowledge and understanding of precisely how the approach can work within a given discipline of interest. Practitioners must develop an intimate and thorough understanding of why this approach is preferable, how it differs from whatever approach is traditionally used, and how they can explain the process in a court of law. At the same time, they must also be aware of limitations and concerns that exist when using this approach. The key to such understanding is education which, regrettably, appears to be an area where little effort has been made to date. The author believes that this is one of the main, though not the sole, reason why there has been only limited adoption of the approach. The discipline of forensic document examination (FDE) is arguably one of the more challenging to address given that this

type of evidence is usually not quantified and requires a largely subjective evaluation process. The author has presented a number of workshops on this topic aimed at forensic document examiners, primarily from North American laboratories [1-5]. These were 1 or 2-day workshops tailored to the FDE audience and focused entirely on problems and issues pertaining to that domain.

This presentation will provide the author's perspective on those workshops, both as a FDE practitioner and as an educator. It will outline the approach taken and discuss some of the challenges faced when working with already-trained examiners who do not generally understand this approach or see the need to make any change to their normal evaluation process. While the future of the logical approach for evidence evaluation is rather unclear at this time, at least in the North American context, a prognosis with recommendations for improving that future will be given.

[1] "Questioned Documents: Conclusion Scales and Logical Inference (Part 1)", presented August 2013 at the A.S.Q.D.E. Conference, Indianapolis, Indiana

[2] "Practical Applications of Logical Inference and Reasoning for QDE", presented November 2012, CSFS QD Workshops, Ottawa, Ontario

[3] "Logical Inference and Evidence Evaluation for QDE", presented October 2012, FBI Laboratory, QD Section Members

[4] "Logical Inference and Evidence Evaluation for QDE", presented October 2012, Scientific Working Group for Forensic Document Examination (SWGDOC)

[5] "Conclusion Scales and Logical Inference - Part 1", presented May 2012 at the C.S.F.S. Conference, Burnaby, British Columbia

Keynote [7] TrueAllele® Interpretation of DNA Mixture Evidence

Mark Perlin

Cybergenetics, Duquesne University

A DNA mixture arises when two or more individuals contribute their DNA to biological evidence. STR data derived from this evidence contain peaks whose heights are in rough proportion to the genotype contributions [1,2]. The peak height data patterns at each genetic locus can be described by a hierarchical Bayesian model, which accounts for the genotypes, their relative quantities, PCR amplification and detection artifacts, and uncertainty in the data and variables.

The interpretation task is to objectively infer genotypes from the data, representing uncertainty as posterior probability.

Comparison can be made afterwards between inferred genotypes to calculate a likelihood ratio (LR) that assesses evidential value [3].

Cybergenetics TrueAllele® Casework system frames the STR data generation process in a hierarchical model [4]. First developed in the late 1990s, the TrueAllele model evolved through 25 versions as new explanatory variables were included or refined, and more hierarchical layers were added for robustness. Markov chain Monte Carlo (MCMC) was introduced early on to statistically solve probability equations of increasing dimensionality.

The system is designed around a fast and efficient forensic analyst workflow. A visual user interface (VUIer™) client program lets a user examine data, ask mixture questions, review genotype answers, and calculate LRs [5,6]. MCMC genotyping is done on parallel server computers connected to a coordinating TrueAllele database. All genotypes are uploaded to the database, and can be automatically compared with LR assessment for investigative applications. A small lab system can process hundreds of mixture items every day.

TrueAllele has been tested with diverse STR kits and instruments in over twenty validation studies, on both laboratory and casework samples [4,7,9,10]. The mixture samples contain up to five contributors, have high or low template amounts, may exhibit differential degradation, and can include relatives. log(LR) match comparisons show that TrueAllele is highly sensitive, specific and reproducible. Identification information varies predictably with DNA contributor amount, regardless of contributor number. With a sufficient number of assumed contributors, LR values remain constant.

TrueAllele has processed mixture items in hundreds of criminal cases [8], including the World Trade Center disaster. Giving evidence in court entails educating lawyers and jurors about

genotype inference and LR's before stating statistical results[5]. TrueAllele has been admitted into evidence after challenge in the United States and the United Kingdom.

The talk will describe the system's operation in the context of a criminal case, as it proceeds from DNA mixture interpretation through trial testimony.

- [1] Perlin MW, Lancia G, Ng S-K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *Am J Hum Genet.* 1995;57(5):1199-210.
- [2] Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J Forensic Sci.* 2001;46(6):1372-7.
- [3] Perlin MW, Kadane JB, Cotton RW. Match likelihood ratio for uncertain genotypes. *Law, Probability and Risk.* 2009;8(3):289-302.
- [4] Perlin MW, Sineelnikov A. An information gap in DNA evidence interpretation. *PLoS ONE.* 2009;4(12):e8327.
- [5] Perlin MW. Explaining the likelihood ratio in DNA mixture interpretation. Twenty First International Symposium Human Ident, 2010; San Antonio, TX.
- [6] Perlin MW. Sherlock Holmes and the DNA likelihood ratio (A142). AAFS 63rd Annual Scientific Meeting; Chicago, IL. *Amer Acad Forensic Sci*; 2011. p. 95.
- [7] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. *J Forensic Sci.* 2011;56(6):1430-47.
- [8] Perlin MW. When good DNA goes bad. *J Forensic Res.* 2013; S11:003.
- [9] Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations. *Sci Justice.* 2013;53(2):103-14.
- [10] Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *J Forensic Sci.* 2013;58(6):1458-66.

Oral [7] Challenging the Interpretation of a DNA Mixture Therese Graversen

University of Oxford

Cowell et al. (2013) proposes a full statistical model for DNA mixtures in the potential presence of stutter and dropout. The model describes the contributors to the mixture and the peak heights observed in the crime scene profile. The model may be used both for the evaluation of the weight of evidence and for mixture deconvolution, also in the case of a joint analysis of multiple mixtures or replicates.

In this talk, I will demonstrate how the model can be challenged, exploiting the set of diagnostic tools introduced in Graversen and Lauritzen (2014). It is important to investigate whether the model adequately describes the specific case at hand. In particular, one should not only compare the prosecution and defence hypotheses, but an effort should be made to justify that each of the hypotheses represents a plausible explanation of the mixture under analysis. The diagnostic tools can reveal problematic aspects of a particular interpretation of a DNA sample, and assumptions about the parameters can be formally tested. All reasoning is entirely within the framework of the adopted model, ensuring a consistent analysis.

An implementation of the statistical model of Cowell et al. (2013) and diagnostic tools is freely available in the R package DNAmixtures (Graversen, 2014).

- Analysis of Forensic DNA mixtures with Artefacts. Cowell R, Graversen T, Lauritzen S, Mortera J (2013). arXiv:1302.4404
- Computational Aspects of DNA Mixture Analysis. Graversen T, Lauritzen S (2014). *Statistics and Computing*, DOI: 10.1007/s11222-014-9451-7
- DNAmixtures: Statistical analysis of mixed traces of DNA with artefacts. Graversen T (2014). R package version 0.1-2. <http://dnamixtures.r-forge.r-project.org>

Oral [7] Evaluation of a Gamma Model Using the Boston University DNA Mixtures Dataset

Robert Cowell

City University London

In 2012, a dataset of the STR PCR amplifications of DNA samples was released to the public by Boston University, in the form of binary ".fsa" files, by Cotton et. al. The number of amplifications released is quite large, and consists of both single source profiles and mixed profiles of up to four individuals, all experimentally controlled for the amount of input DNA and, in the case of the mixtures, of the proportions contributed by each individual. Amplifications from several different kits were given.

In 2013 Cowell et al. proposed a gamma model for the peak heights arising from STR DNA mixtures. The model takes into account quantitative peak height information, as well as dropout, dropin, and stutter artifacts and a peak threshold detection level. By likelihood maximization it is possible to carry out both evidential and separation analyses.

This presentation looks at evaluating the performance of the gamma model using the subset of DNA samples amplified with the Identifiler kit. This subset consists of 1100 sample files, plus additional files of amplifications of positive and negative controls and allelic ladders. It thus provides a significant and large dataset suitable for examining the assumptions and performance of the gamma model.

One of the crucial assumptions of the gamma model is that the distributions of peaks heights of distinct alleles are independent both within and between markers, given the amounts of DNA from the contributors and their genotypes. This assumption appears to be violated in the Boston dataset samples, with significant correlations between the peak heights of distinct alleles. This finding is in line with that reported by Tvedebrink et. al. (2010). The implications of this are discussed, and a possible resolution presented.

Modifications of the gamma model suggested by the nature of the Boston University DNA Mixtures Dataset are also presented. The enhanced and simple gamma models are compared with respect to their predictions on the dataset. Suggestions for future work are made.

- Cotton RW, Grgicak CM, Word CJ, Terrill M, Boston University Biomedical Forensic Sciences: DNA Mixtures. Available from <http://www.bu.edu/dnamixtures/>
- Cowell RG, Graversen T, Lauritzen S, Mortera J (2013), Analysis of DNA mixtures with artefacts. ArXiv:1302.4404.
- Tvedebrink T, Eriksen PS, Mogensen HS, Morling N, Evaluating the weight of evidence using quantitative STR data in DNA mixtures, *Appl. Stat.* 59 (2010) 855–874.

Keynote [6b] Bayes Factor Evidence of a Ricin Preparation Method

Crister Åstot, Susanne Wiklund Lindström, Anders Nordgaard

Swedish Defence Research Agency (FOI)

Ricin is a protein toxin, produced in the seeds of the castor plant *Ricinus communis*, which grows wild in tropical and subtropical climates. Castor seeds are an unregulated agricultural product but the intentional purification of the toxin, present in the seed's pulp, is highly regulated and classified as a Schedule 1 controlled substance under the Chemical Weapons Convention and the implementation of the Convention in the national laws of the 189 signatory States makes undeclared ricin purification a crime, globally.

The presented study includes the extraction of ricin using 4 different preparation methods. The four methods to prepare a toxic ricin sample were ranging from simple protocols found in various publications and internet sites, to more sophisticated purification procedures from the scientific literature. Samples from four different *Ricinus communis* cultivars were prepared for each of the methods and their chemical profiles were determined by mass spectrometric methods. Peptides and small proteins were analysed by capillary electrophoresis and LC-MS. Ricin and the closely related *Ricinus communis* agglutinin were mapped by LC-MS and LC-MSMS after isolation by affinity separation and enzymatic digestion. Carbohydrates, ricin fatty acids and solvent residues were analysed by GC-MS. The resulting data matrix included in total 38 samples and 58 variables.

The main goal of the study is to evaluate the possibility to calculate Bayes Factor (BF) for evidence of a criminal action of the extraction of ricin. The second goal is to find out what preparation method were used in the extraction of ricin and the corresponding BF evidence. A strategy to handle the complicated situation of multiple variables and few samples will be presented.

- Kass RE, Raftery AE (1995). "Bayes Factors". *J. Am. Statist. Assoc.* 90, 791.
- Wunschel DS, Melville AM, Ehrhardt CJ, Colburn HA, Victry KD, Antolick KC, Wahl JH, Wahl, KL (2012) Integration of gas chromatography mass spectrometry methods for differentiation ricin preparation methods. *Analyst* 137, 2077.

Oral [6b] Evaluation of Evidential Value of Spectra Agnieszka Martyna¹, Grzegorz Zadora^{2,3}, Aleksandra Michalska²

¹Jagiellonian University Krakow, ²Institute of Forensic Research, ³University of Silesia

Nowadays there is a growing interest in searching for new analytical techniques suitable for microtraces analysis. As a consequence of the ongoing development in the field of advanced methods of analytical chemistry, there is growing amount of information describing the analysed materials, e.g. information about absorption of infrared light corresponding to many wavenumbers in the form of spectra (containing more than thousand of variables). On the one hand this implies that the analytical methods may become suitable for analysing smaller samples called microtraces as despite their sizes, they can be described by many relevant parameters, but on the other hand multidimensional data are still difficult to interpret using statistical methods, e.g. likelihood ratio. The reason for this is that the analysis of multidimensional data requires the population parameters to be reliably estimated. This can only be achieved when working with a huge databases with number of objects (m) far greater than the number of parameters (p) they are described by [1]. Till now, all the LR models are created for situations in which the number of objects in the database (m) extends the number of the variables (p) that describe them ($m \gg p$) [1]. Since most analytical methods deliver many more parameters measured for samples than their number (i.e. the situation is that $m \ll p$), this condition is difficult to meet.

In this presentation the procedure of applying the wavelet transform [2] as a data compression method will be proposed as it was found that it allows for effective reduction of FTIR and Raman data dimensionality (so that $m \gg p$). Its advantage is that it does not waste the chemical information by focusing on the local spectra features as e.g. bands associated with particular bond vibrations, especially important from the chemical perspective. Wavelets are useful tool for generating sparse representation of the signals by preserving its features, which may be especially crucial from chemical perspective and beneficial for compressing the data. They proved to be capable of extracting the relevant information from spectra by creating fewer new variables, which appeared useful for solving the comparison problem of samples described by infrared or Raman spectra with application of LR approach. The effectiveness of the proposed LR models based on the variables derived from wavelet transform was proved with the use of the empirical cross entropy (ECE) approach [3].

[1] Zadora G, Martyna A, Ramos D, Aitken C, Statistical Analysis in Forensic Science. Evidential value of multivariate physicochemical data. Wiley 2014.

[2] Walczak B (eds), Wavelets in Chemistry. Elsevier 2000.

[3] Ramos D, Gonzalez-Rodriguez J, Zadora G, Ramos D, Information-theoretical assessment of the performance of likelihood ratio computation methods. Journal of Forensic Sciences 58 (2013) 1503.

Oral [6b] The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

Amy Wilson^{1,2}, Colin Aitken², Richard Sleeman, James Carter³
¹University of Durham, ²University of Edinburgh, ³Queensland Health Forensic and Scientific Services

Much research in recent years for evidence evaluation in forensic science has focussed on methods for determining the likelihood ratio where the data have been generated by various random phenomena. The likelihood of the evidence is calculated under each of two propositions, that proposed by the prosecution and that proposed by the defence. The value of the evidence is given by the ratio of the likelihoods associated with these two propositions. One form of evidence evaluation is related to discrimination in which the problem is one of source identity. The prosecution proposition is that the banknotes are associated with a person who is associated with criminal activity, the defence proposition is that the banknotes are not associated with a person who is associated with criminal activity.

The aim of this research is to evaluate this likelihood ratio under two explanations, one an extension of the other, for the random phenomena by which the data have been generated. The first is when the evidence consists of continuous autocorrelated data. The second is when the observed data are also believed to be driven by an underlying latent Markov chain. Four models have been developed to take these attributes into account: an autoregressive model of order one, a hidden Markov model with

autocorrelation of lag one and a nonparametric model with two different bandwidth selection methods. Application of these methods is illustrated with an example where the data relate to traces of cocaine on banknotes. The likelihood ratios using these four models are calculated for these data, and the results compared.

- Wilson A, Aitken CGG, Sleeman R, Carter JF (2014) The evaluation of evidence relating to traces of cocaine on banknotes, Forensic Science International, 236, 67-76.

Wednesday Afternoon

Keynote [9] Analysing Convergent and Conflicting Evidence in Forensic Scenarios

Patrick Juchli, Alex Biedermann, Franco Taroni
 University of Lausanne

Forensic and judicial practices are routinely faced with multiple (if not masses of) items of evidence. The coherent assessment of items of evidence in combination is therefore of vital interest. Numerous legal cases, including historically well-known examples such as *People vs. Collins*, suggest that such an assessment is not a trivial task and that inferential issues ensued are either not well understood, or even unknown. This illustrates that a systematic study on the combination of items of evidence and its inferential issues plays a fundamental role in evidence interpretation.

To approach this topic, the following two questions can provide guidance: (i) what is the relationship among a set of propositions of judicial interest and a set of scientific findings? and (ii) What is the joint inferential force of a set of findings and how is it affected by each finding?

These questions guided our analysis – using graphical probabilistic models – of two forensic scenarios. The first scenario involves issues in combination within a single item of trace evidence: that is, the combination of two separate aspects of a single trace, as given by the general pattern and the size of a footwear mark. The second scenario comprises two distinct items of trace evidence, namely a finger- and a footwear-mark. These two instances of evidential combination differ substantially from a physical viewpoint. From a probabilistic perspective, however, they turn out to be closely related: they share the same inference structure, and are, thus, ruled by the same inferential mechanisms. In fact, they lead to a generic reasoning pattern that embraces what David A. Schum called “convergent evidence” and “conflicting evidence”. This reasoning pattern clarifies the mechanisms involved in inferences upon hypotheses based on two or more distinct items of information (such as separate aspects of a single trace or distinct items of trace evidence). The analysis of such generic patterns of reasoning can help forensic scientists gain better insight in their thought process, and raise their awareness of inferential subtleties that – without close inspection – may go unnoticed.

- Juchli P, Biedermann A, Taroni F, Graphical probabilistic analysis of the combination of items of evidence, Law, Probability and Risk 11 (2012) 51.
- Schum DA, Evidential Foundations of Probabilistic Reasoning, Northwestern University Press, Evanston, IL, 2001.

Oral [9] Crime Scenarios in a Bayesian Network: Modeling Forensic Evidence with Narrative

Charlotte Vlek¹, Henry Prakken^{1,2}, Silja Renooij², Bart Verheij^{1,3}

¹University of Groningen, ²Utrecht University, ³Stanford University

When a criminal trial requires reasoning with a combination of evidence, Bayesian networks are considered a good tool to work with combined probabilities. However, a judge or jury is more inclined to think in terms of arguments or scenarios rather than probabilities. An approach that combines narrative and probability can form the basis for a better communication between a judge or jury and a forensic expert.

In this talk we present a method for building Bayesian networks on the basis of narrative. We apply our method to the Dutch case

of the Anjum murders, with a resulting model showing that an alternative scenario was perhaps more probable than the conclusion of the court case. Our method aims to combine the best of two worlds: while Bayesian networks enable a solid formalization of the details of the case, scenarios serve as a coherent account of what may have happened, thereby providing the context needed for finding relevant variables for the network. Inspired upon work by Fenton, Neil and Lagnado [Fenton et al., 2013], who developed legal idioms as building blocks for the construction of a legal Bayesian network, we propose four narrative idioms. These narrative idioms capture the narrative concepts of a scenario, a subscenario, small variations within a scenario and the combination of multiple scenarios in a case. For the construction of a Bayesian network we use these narrative idioms as building blocks, and employ the property of narrative that it can be told at various levels of detail. By slowly unfolding a scenario into more detail, a Bayesian network structure is built incrementally. This results in a modular structure in which the various scenarios and subscenarios are clearly visible and in which the probabilities of scenarios can be compared. By modeling the Anjum case, we illustrate some strengths and weaknesses of our method. Finally, our model of the Anjum case allows us to draw the conclusion that an alternative scenario was perhaps more probable. Rather than the scenario that sounds most obvious and for which Marjan van der E. was convicted in court, an accomplice cooperating with the investigation to keep suspicion from himself is, in this model, the more probable explanation of the combination of all evidence.

- Fenton N, Neil M, Lagnado D, A general structure for legal arguments using Bayesian networks, *Cognitive Science* 37 (2013), 61.

Oral [9] Modelling Crime Linkage with Bayesian Networks
Jacob de Zoete¹, Marjan Sjerps¹, David Lagnado², Norman Fenton³, Koen Vriend¹, Menno Dolman¹, Ronald Meester⁴
¹University of Amsterdam, ²University College London, ³Queen Mary University of London, ⁴VU University Amsterdam

When two or more crimes show specific similarities, such as a very distinct modus operandi, the probability that they were committed by the same offender becomes relevant. This probability depends on the degree of similarity and distinctiveness. We show how Bayesian networks can be used to model different evidential structures that can occur when linking crimes, and how they assist in understanding the complex underlying dependencies. That is, how evidence that is obtained in one case can be used in another and vice versa. The flip side of this is that the intuitive decision to "unlink" a case in which exculpatory evidence is obtained leads to serious overestimation of the strength of the remaining cases.

The examples that will be discussed include Bayesian networks for different numbers of cases, situations where there are multiple offenders and a situation where it is not clear whether it is a crime (for example arson vs. accident).

Although the examples discussed are about crime linkage, they generalise to other types of problems where multiple items are combined, e.g. multiple stain problems in forensic practice. The same questions are relevant and similar caution is needed. Lastly, the differences between the use of crime linkage in legal practice and what is possible according to a logical evaluation of the evidence are discussed.

Oral [9] The Interpretation at Activity Level of Traces Found on Adhesive Tapes

Remi Wieten^{1,2}, Bart Blankers², Bas Kokshoorn², Jacob de Zoete^{1,2}

¹University of Amsterdam, ²Netherlands Forensic Institute

In violent crimes (e.g. homicides and terrorist attacks) adhesive tapes are often used by perpetrators. Duct tapes are used for example to tie up a victim or to bind together parts of an explosive. In the forensic examination of such tapes many different types of traces, such as finger marks, DNA and fibers can be found [1]. These traces can be used to link a suspect to

the crime. The evidential value of these traces (at source level) is determined by assessing the rarity of relevant features.

To determine the combined evidential value of these traces at activity level (i.e. how probable are the findings if the suspect used the tape to tie up the victim compared to when he only used the tape for normal purposes) the evidential value of the individual traces has to be combined. To do so, many other factors have to be taken into consideration. One of the most important factors is the position of the trace on a piece of tape in relation to other pieces of tape and the original roll of tape. The position of the trace strongly determines its evidential value since it is for instance unlikely to find traces matching the suspect's deep into a roll of tape when he claimed that he previously only touched the outside of the roll. By determining the order in which the tape pieces found at the crime scene came from the roll (by physical end matching) the location of the traces can be deduced. Whereas evidence in tape cases is currently combined on a case by case basis, a more uniform and objective approach making use of Bayesian Networks may be preferred [2]. In a Bayesian Network all the relevant variables and their conditional dependencies can be charted and the evidential value at activity level in a case can be determined. The conditional probabilities necessary to form a conclusion are determined by consulting literature and the (subjective) opinions of forensic experts. By performing sensitivity analysis it can be evaluated which probabilities most strongly influence the likelihood ratio. Further studies will be set up to collect the data necessary to better estimate these probabilities.

To demonstrate the versatility of the developed Bayesian Network in casework we will look at some case examples. This study demonstrates how a Bayesian Network can assist forensic experts in providing the court with a robust and transparent evaluation of their findings.

[1] Smith JM (2007). *Forensic Examination of Pressure Sensitive Tape*. In: ed. R.D. Blackledge. *Forensic Analysis on the Cutting Edge: New Methods for Trace Evidence Analysis*. John Wiley & Sons, New Jersey; p291-332.

[2] Taroni F, Aitken CGG, Garbolino G, Biedermann A (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*. Wiley, Chichester.

Oral [9] Combining Uni-directional Statistics to Evaluate Circumstantial Evidence of Anomalies

Dennis Maynes

Caveon Test Security

Propositional inference using Bayesian Statistics has been shown to be very effective for arguing and presenting circumstantial evidence. However, some situations involving anomalous data do not easily lend themselves to this class of propositional inferences. As a case in point, the detection of potential fraud on certification exams or licensure tests usually involves anomalous data that are not readily suited for evaluating the proposition that a person committed exam fraud given the data. The primary difficulty in these cases is the lack of a database of known instances of exam fraud. In medicine, biopsies and autopsies can be performed to determine whether the individual had X or not. In crime, physical evidence such as lost property, signs of unlawful entry or personal harm can be used to reasonably infer that a crime was committed. Exam fraud constitutes a class of activities which do not allow verification that fraud was actually committed or the manner in which it might have been committed. As a result, the analyst must present the circumstantial evidence using anomalous data, usually based upon frequentist statistics. This presentation will demonstrate two fundamental approaches for combining probability statements concerning anomalies that can be used as a basis for presenting the degree of extremity of all the circumstantial evidence. The two approaches are based on order statistics and chi-square statistics. Furthermore, each approach shows how to probabilistically evaluate compound statements, written as OR and AND propositions, concerning circumstantial evidence. The presentation will show that OR propositions, when using order statistics, correspond to the popular Bonferroni correction or control of simultaneous error when several propositions (or hypothesis tests) are evaluated jointly.

Oral [9] Evaluation of Findings in Hit and Run Investigations**Amanda Lennon***Forensic Science Laboratory, Ireland*

Hit and run investigations involving a motorcar and another vehicle, person, structure etc are regularly encountered in our laboratory. Where possible we try to assist the investigation by evaluating the evidence relative to two propositions, i.e. we assess the probability of the evidence given that the suspect motorcar is the vehicle involved in the collision and the probability of the evidence given that it was a vehicle other than the suspect motorcar. We make use of a National Vehicle Database along with knowledge of how paint varies between vehicles to assign frequencies to the paint evidence recovered.

An important aspect of the evaluation is the background information. The identification of the suspect car on the basis of colour, make, model or partial registration number affects the strength of the evidence. The presentation will outline firstly how this background information influences the selection of the appropriate propositions for consideration. It will also address the assignment of the frequency of a paint type based on the relevant population for the particular case circumstances. This allows for fairer evaluation of the evidence and guards against "double counting" of evidence.

Oral [7] Building blocks for continuous DNA interpretation models**James Curran¹, Jo-Anne Bright^{1,2}, John Buckleton², Duncan Taylor³**¹University of Auckland, ²Environmental Science & Research,³Forensic Science South Australia

Continuous likelihood ratio models are the Holy Grail for researchers interested in the modern interpretation of DNA evidence. Continuous models aim to take full advantage of the additional information contained in electropherograms by incorporating it using a statistically sound framework. If these models are correctly implemented, then they can remove the need for the thresholds, guidelines, and subjective decision making that is involved in current interpretation practice. Such models depend on substantial laboratory work and statistical data analysis to have any hope of functioning correctly. In this talk I will describe some of the work our team has been doing to model stochastic PCR phenomena such as stutter and heterozygous balance.

Oral [7] STRmix: The New Australasian Standard for Forensic DNA Profile Interpretation**Jo-Anne Bright^{1,2}, Duncan Taylor³, James Curran², John Buckleton¹***Environmental Science & Research, University of Auckland, Forensic Science South Australia*

STRmix is an expert system that applies a fully continuous approach to the interpretation of forensic DNA profiles. STRmix was created for implementation across Australia and New Zealand and has been adopted as the new Australasian standard. It has attracted international interest.

The Institute of Environmental Science and Research Limited, New Zealand, implemented STRmix into routine forensic casework in August 2012. This presentation gives a brief overview of the continuous statistical model which underpins the software and describes our laboratory's experience and the benefits realised through its introduction.

The improved interpretative capability and advanced database search functionality are both illustrated with examples of challenging profiles.

With appropriate care an increased number of mixed DNA profiling results which would have previously been deemed 'unresolvable', can be effectively and objectively assessed. Where appropriate, comparison to reference samples from persons of interest can be made and a likelihood ratio calculated.

- Taylor D, Bright J-A, Buckleton JS, The interpretation of single source and mixed DNA profiles, *Forensic Sci Int Genet* 7 (2013) 516.
- Bright J-A, Taylor D, Curran J, Buckleton J, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, *Forensic Sci Int Genet* 7 (2013) 296.
- Bright J-A, Taylor D, Curran J, Buckleton J, Searching mixed DNA profiles directly against profile databases, *Forensic Sci Int Genet* 9 (2014) 102.
- Bright J-A, Taylor D, Curran J, Buckleton J, Degradation of forensic DNA profiles, *Australian Journal of Forensic Sciences* 45 (2013) 445.

Oral [7] An Investigation of Software Programs Using "semi-continuous" and "continuous" Methods for Complex Mixture Interpretation**Michael Coble¹, Ate Kloosterman²**¹National Institute of Standards and Technology, ²Netherlands Forensic Institute

In 2006, the ISFG commissioned an expert committee to develop guidelines for mixture interpretation (Gill et al. 2006) and these recommendations have been widely accepted. Among the recommendations was a need for the laboratory to establish a stochastic threshold to determine the risk associated with the loss (dropout) of an allele in the mixture.

Recent improvements in both STR chemistry and CE instrumentation have exacerbated interpretation as laboratories try to analyze highly complex mixtures such as "touch" items with (a) more than two contributors and/or (b) low-level contributors with possible dropout. Current strategies to evaluate low-level mixtures with dropout using the binary Likelihood Ratio (LR) are insufficient and may overstate the weight of the evidence.

As a continuation to the 2006 guidelines, the ISFG recently published recommendations for the interpretation of low-level mixtures when dropout is possible (Gill et al. 2012). Recently, a number of software programs are now available that utilize a "semi-continuous" approach, incorporating a probability of dropout in the LR (e.g. LRMix: Gill and Haned (2013), Lab Retriever: Lohmueller and Rudin (2013), LikelTD: Balding (2013)). Other software programs utilize a "continuous" model to incorporate variation within the data (e.g. STRmix: Taylor et al. (2013); True Allele: Perlin and Sinelnikov (2009)).

We have examined a set of complex DNA mixtures containing different quantities of DNA from known reference samples. These mixtures include low-level contributors and three- and four-person contributors with differing ratios and allele sharing. Using the data from the mixed DNA-profiles we have investigated the performance of the existing platforms to calculate LR's.

We examined the accuracy of probabilistic models in terms of (1) the ability to discriminate between donors and non-donors and (2) the stability of output data across reasonable settings of, for example, probabilities of drop-out and drop-in (calibration of the probabilistic model). We also considered other parameters such as reproducibility when replicates of the same mixture are analyzed. Finally practical aspects such as time of analysis and ease of use are compared.

- Gill P, et al. (2006) *Forensic Sci Int* 160, 90-101.
- Gill P, et al. (2012) *Forensic Sci Int Genet* 6, 679-688.
- Gill P, Haned H (2013) *Forensic Sci Int Genet* 7, 251-263.
- Lohmueller KE, Rudin N (2013) *J Forensic Sci* 58 Suppl 1: S243-249.
- Balding DJ (2013) *Proc Natl Acad Sci U S A* 110, 12241-12246.
- Taylor D, Bright J-A, Buckleton J (2013) *Forensic Sci Int Genet* 7, 516-528.
- Perlin MW, Sinelnikov A (2009) *PLoS One*. 4, e8327.

Oral [7] The DNA•VIEW Mixture Solution**Charles Brenner***University of California, Berkeley*

There is an emerging consensus and general awareness in the forensic world that there is no very simple yet adequate method, no "royal road" to DNA mixture analysis. At a minimum the computational method must model allelic signal intensities (peak height) and stochastic variation – a so-called "continuous" model. However there are always advantages to choosing a model that is as simple as possible. For example, there are physical reasons to expect that stutter intensity is correlated with the intensity of its mother allele so it's less accurate to model the two as varying independently and tempting to complicate the model with a more subtle treatment of stutter. But each added complication impedes

the nimbleness of the model and soon it can be implemented only through simulations. That means compromises in other directions including a very large penalty of execution speed. Speed and simplicity come at some cost in accuracy, but that need not mean a bias against the suspect. One strategy to avoid that bias is to build in a contrary bias by choosing intentionally generous parameter values for the defence hypothesis. The DNA•VIEW Mixture Solution computes and compares likelihoods for prosecution and defence hypotheses under the model given a mixture. The model has several parameters representing realities such as stochastic variation and stutter proportions. Individually the two computed likelihoods are sensitive to the choices of parameter values, but fortunately and comfortingly the model is robust in that the likelihood ratio is not very sensitive to moderate changes in the parameter values. The DNA•VIEW Mixture Solution embodies a simple coherent model that is evaluated without Monte Carlo simulations and is by comparison blindingly fast. Orders of magnitude advantage in speed is of course convenient for the user of the final product, but even more importantly facilitates understanding, testing, experimenting, comparing approaches, validating, extending, avoiding distractions, and keeping focus on the goal.

Oral [7] How Can We Assign a Quantitative Evidential Value to a Profile of a DNA Mixture?

Keith Inman^{1,3}, Kirk Lohmueller², Norah Rudin³

¹California State University East Bay, ²University of California Los Angeles, ³Forensic DNA Consulting

We have developed a software tool to estimate the quantitative evidential value of complex DNA profiles. *Lab Retriever* uses likelihood ratios that incorporate an empirically-derived estimate of the probability of allelic drop-out. The software is an implementation of the approach suggested by Balding and Buckleton in 2009. To meet the needs of the forensic DNA community, we have committed to an open source software platform that is provided free of charge. We will discuss the capabilities, advantages and limitations of our implementation of the semi-continuous model as compared to both the binary and fully continuous models. We will offer casework examples and report our experiences testifying about our calculations and conclusions in court.

- Balding DJ, Buckleton J. Interpreting low template DNA profile, *Forensic Sci Int Genet* 4 (2009) 1–10.

Oral [7] An Investigation of the Potential of DIP-STR Markers for DNA Mixture Analyses

Giulia Cereda, Alex Biedermann, Franco Taroni

University of Lausanne

The genetic characterization of unbalanced mixed stains remains an important area where improvement is imperative. In fact, in cases of mixed DNA stains, the use of the standard tools of forensic DNA profiling (i.e., Short Tandem Repeat (STR) markers) does not allow one to detect the profile of the minor contributor if its quantitative share of DNA is less than 10% of the mixed trace. This is due to the fact that the major contributor's profile "masks" that of the minor contributor.

Besides known remedies to this problem, such as Y-STR analysis, a new compound genetic marker has recently been developed: it consists of a Deletion/Insertion Polymorphism (DIP) linked to a STR polymorphism. These novel markers are also known as DIP-STR.

The first part of this presentation will develop and describe a probabilistic framework for the assessment of profiling results obtained with this novel typing technique, using the particular context of unbalanced DNA mixtures of two contributors as an example. The approach relies on probabilistic graphical models, in particular object-oriented Bayesian networks (OOBNs), and uses the likelihood ratio as an expression of the probative value of DIP-STR profiling results.

In a second part, the presentation will deal with the investigation of the robustness of these novel markers. In particular, the performance of DIP-STR markers will be compared (i) to that of

results obtained with traditional STR markers in cases of moderately unbalanced mixtures, and (ii) to that of results obtained with Y-STR markers in cases of female-male mixtures.

Keynote [6c] A Composite Model for Compositional Data with Large Concentrations of Zeros

Gary Napier¹, Tereza Neocleous¹, Agostino Nobile²

¹University of Glasgow, ²University of York

We present an effective approach for modelling compositional data with large concentrations of zeros and several levels of variation, applied to an experimental database of elemental compositions of forensic glass of various use types. The procedure consists of (i) partitioning the dataset in subsets characterized by the same pattern of presence/absence of chemical elements; (ii) fitting a Bayesian hierarchical model to the transformed compositions in each data subset; and (iii) using a composite model which combines all the models in (ii) for classification and evidence evaluation of glass fragments. The composite model is implemented in a web app which returns the posterior predictive probability that newly observed fragments of glass are of a certain use type (classification task), or a numerical measure of the evidential value of two sets of glass fragments under two competing propositions about their source (evidence evaluation task). The model performs well in both tasks in a validation study. We present the results and discuss some issues relating to calibration performance and how this can be improved by a small adjustment in the way the measure of evidential value is reported.

Oral [6c] Probabilistic Graphical Models to Deal with Age Estimation of Living Persons

Emanuele Sironi, Matteo Gallidabino, Franco Taroni

University of Lausanne

The rise of criminal, civil and administrative judicial situations involving young people lacking of valid identity documents has generated scientific needs to deal with age estimation of living persons. Age estimation, therefore, represents an important operational activity for numerous forensic and medico-legal services worldwide. Methods used to estimate the age of given persons are generally based on the examination of several physical attributes, such as the mineralization of the dentition or the skeletal maturity of hand or collarbones. The results of each exam are assessed in order to estimate a chronological age by reference to atlas or tables, regression analysis or descriptive statistics. All these methods are generally specific for one particular physical attribute. Therefore, a global conclusion could be difficult to produce and its coherence hard to justify.

In this presentation, the Bayesian approach is adopted to deal with the problem at hand, aimed at combining and interpreting data issued from different physical attributes. A graphical probabilistic model (i.e., Bayesian network) is built in order to allow the scientist or jurist an easy communication of the probabilities for exclusive hypotheses of judicial interest such as "the given person is above (or below) the 18-years-old threshold" or for estimating the chronological age.

Results will show the benefits of the suggested model. For example, (a) the facility to incorporate a large number of influential variables, such as sex or observational errors, and, (b) the extension of the model for the combination of items of evidence, i.e., the possibility to simultaneously evaluate results issued from different physical attributes in order to improve the estimation of the chronological age estimation variable.

Oral [6c] A Bayesian Multilevel Model for the Evaluation of Multivariate Scientific Findings

Silvia Bozza¹, Raymond Marquis², Franco Taroni²

¹Ca' Foscari University of Venice, ²University of Lausanne

Multivariate continuous data are becoming more prevalent in forensic science. Scientific findings, say glass fragments, illicit substances or handwritten characters can be described by several variables. The evaluation of multivariate measurements

on characteristics associated to questioned material can be performed through the computation of a Bayes factor, a rigorous concept that provides a balanced measure of the degree to which the available observations are capable of discriminating among propositions forwarded by opposing parties at trial. Data often present a complex dependence structure with several levels of variation, a large number of variables and a non-Normal distribution. A Bayesian multilevel model is proposed to deal with such constraints, and it is implemented in a case study involving handwriting examination in presence of questioned documents.

- Aitken CGG, Lucy D (2004) Evaluation of trace evidence in the form of multivariate data, *Applied Statistics*, 53, 109-122.
- Gelman A, Carlin JB, Stern H, Denson DB, Vehtari A, Rubin DB (2014) *Bayesian Data Analysis*, 3rd ed, CRC Press.
- Bozza S, Taroni F, Marquis R, Schmittbul M (2008) Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship, *Applied Statistics*, 57, 329-341.

Oral [6c] Computational and Statistical Aspects of the Forensic Identification Source Problem

Danica Ommen, Cedric Neumann, Christopher Saunders
South Dakota State University

In 1977, Lindley and Evett introduced modern Bayesian methods for forensic evidence interpretation to the forensic science community. This and related approaches have dominated the academic research related to the interpretation and presentation of forensic evidence. However, in recent years there have been number of debates, in both academic circles and forensic communities, related to the applicability of these methods in the U.S. judicial system.

These methods require the explicit statement of two mutually exclusive, but non-exhaustive, models about how the evidence in a given situation has arisen; one usually corresponding to a defense model and one corresponding to a prosecution model. Once these models have been defined and the evidence is collected, the forensic scientist is then required to present the evidence in a concise and transparent manner so that a decision maker can ultimately decide between the two proposed models of how the evidence has arisen. This presentation usually takes the form of a ratio of two marginal likelihoods known as a Bayes Factor (BF) or a likelihood ratio.

The information that a forensic scientist has available to evaluate between the two models includes the following forms: (1) samples from a recovered trace of unknown origin; (2) samples from a control source and (3) a collection of samples from an alternative source population. In certain applications, the choice of the samples used to characterize the defense and prosecution probabilistic models will be mandated by available databases or, in extreme situations, there will be no such samples available.

The BF in these situations is traditionally calculated using Monte Carlo integration. Well-known research has illustrated the inefficiency of these methods for calculating the marginal likelihoods necessary to compute the Bayes factor. This inefficiency has a significant negative impact on the implementations of these methods in real world applications. We will review how the common sets of propositions and statistical approaches used in forensic science result in different definitions of the evidence and of its support for deciding between the prosecution and defense models. Our focus will be on introducing the different types of problems that may arise while highlighting the computational difficulties in calculating the Bayes factors for the different model selection problems.

- Lindley DV (1977), A Problem in Forensic Science, *Biometrika* 64, 207-213.
- Kass, Raftery (1995), Bayes Factors, *J. Amer. Statist. Assoc.* 90, p773-795.
- Neal (2011), MCMC using Hamiltonian dynamics, in the *Handbook of Markov Chain Monte Carlo*, Brooks, Gelman, Jones, Meng (eds), Chapman & Hall / CRC Press, p113-162.

Oral [6c] Forensic Likelihood Ratios Should Not Be Based on Similarity Scores or Difference Scores

Geoffrey Morrison^{1,2}, Ewald Enzinger^{1,3}

¹University of New South Wales, ²University of Alberta, ³Austrian Academy of Sciences

A forensic likelihood ratio quantifies the probability of the evidence if the prosecution hypothesis were true versus if the defence

hypothesis were true. For source-level comparison of objects which when measured produce continuously-valued data, the forensic likelihood ratio can often be characterised as the similarity of the offender (questioned) sample to the suspect (known) sample versus the typicality of the offender sample with respect to the relevant population. An approach for calculating likelihood ratios proposed on multiple occasions in the literature consists of first measuring the difference between pairs of objects in a relevant database, some same-origin pairs and some different-origin pairs, and then modelling the distributions of the same-origin scores and of the different-origin scores in order to calculate the relative likelihoods for a score based on the difference between the actual suspect and offender samples. An alternative to a distance metric is a similarity metric (one is conceptually the inverse of the other). Such an approach, however, does not take account of the typicality of the offender sample with respect to the relevant population, either in the first step of calculating scores or in the second step of calculating likelihood ratios on the basis of those scores. We have previously argued (Morrison, 2013) that in order to calculate forensically interpretable likelihood ratios on the basis of scores, those scores must take account of both similarity and typicality. In this presentation we use Monte Carlo simulation to illustrate the problem with the use of similarity-only or difference-only scores. Monte Carlo simulation allows us to specify the distribution for a population and then draw samples from that population. We have invented the population, it is not a real population, but because we have specified the distribution for the population we have the advantage over a real population that we can calculate the true value of the likelihood ratio for any specified test samples. We can then draw a sample from the population, apply a method to calculate a likelihood ratio and compare the result with what we know to be the true likelihood ratio. We show that the values of likelihood ratios calculated using similarity-only or difference-only scores deviate substantially from the true likelihood ratios. We also compare the performance of several score-based methods which take account of both similarity and typicality, and find that some of these produce values which are relatively close to the true likelihood-ratio values.

- Morrison GS (2013), Tutorial on logistic-regression calibration and fusion: Converting a score to a likelihood ratio. *Aust J Forensic Sci* 45, 173-197.

Oral [6c] Quantifying Uncertainty in Estimations of the Total Weight of Drugs in Groups of Complex Matrices

Ivo Alberink¹, Annette Sprong¹, Annabel Bolck¹, James Curran²

¹Netherlands Forensic Institute, ²University of Auckland

At the drugs department of the Netherlands Forensic Institute (NFI) frequently materials are received from the national airport (Schiphol) such as clothing, towels, suitcases, lemonade powders etcetera which are impregnated with illicit drugs, mainly cocaine. Establishing the presence of cocaine is not problematic: this can be done by analysing a sample of the material with GC-MSD. Determining the total amount of cocaine if the matrix is not a powder or a liquid is more complex. For homogeneous powders and liquids containing cocaine the corresponding (combined) measurement uncertainty is known on the basis of control chart information. When concentrations of cocaine are determined, confidence intervals may be determined for the total amount of cocaine in the powder or liquid. For cocaine impregnated in for example clothing this calculation is more complex because the distribution of the drug may be much more inhomogeneous. Further the weight of the items in the seizure might differ, which will result in a different measurement uncertainty on the estimation of the drug content per item. Thirdly, as indicated above, determination of drug concentration can sometimes only be done on (smaller) subsamples for practical reasons, producing an extra complicating factor. Literature gives guidelines on the expression of uncertainty in analytical measurements, cf. e.g. [1,2]. ISO 17025 [3] requires forensic laboratories to determine uncertainty on their measurements. The NAS-NRC report [4] stresses the need for a scientific approach in the forensic field,

which typically includes determination and reporting of uncertainty on measurements. There are ample publications on the issue how to sample from consignments of drugs and how large samples should be to obtain required results. However, these studies do not address the question of how to obtain estimations of a total combined amount of drugs in items of different weight, containing drugs in concentrations which have considerable measurement uncertainty attached. A guideline is proposed on how to estimate the total amount of drugs in cases like these, and how to quantify uncertainty of the above using confidence intervals.

[1] Guide to the Expression of Uncertainty in Measurement, 2nd edition, ISO, Geneva, 2010.

[2] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, 2nd edition, 2000.

[3] ISO/IEC Standard 17025, General Requirements for the Competence of Testing and Calibration Laboratories, ISO, Geneva, 1999.

[4] Strengthening Forensic Science in the United States: A Path Forward, NAS/NRC Committee Report, 2009.

Thursday Morning

Plenary Evaluating DNA Findings at Activity Level, the Contribution of Bayesian Networks and Simulations of Cases
Christophe Champod

Keynote [11] Statistical Evidence in International Criminal Cases

Wendy Betts

Human Rights Data Analysis Group

International criminal law cases, specifically crimes against humanity and genocide, are complex and require evidence of patterns of conduct or the magnitude of the abuses that generally is not required when prosecuting domestic crimes. In addition to demonstrating that specified prohibited conduct has occurred, the prosecution must also prove the overarching contextual or “chapeau” elements. In the case of crimes against humanity, the prosecution must show that the abuses were part of a widespread or systematic campaign. In the case of genocide, the prosecution must prove the abuses were committed with the intent to destroy a protected group as part of a manifest pattern of similar conduct. Statistical analysis of human rights abuses can provide important circumstantial evidence of the existence of such a pattern or plan. Statistics can quantify victims, those reported as well as unreported, to paint a picture of the extent as well as temporal and geographic distribution of abuses. Perhaps more importantly, statistical analysis of human rights data can discern patterns of abuses in what otherwise may appear to be a chaotic fog of war. However, despite this potential, the use of statistical methods in international criminal cases to date has highlighted a series of challenges to the understanding and acceptance of statistical analysis as evidence. Some challenges result from weaknesses inherent in human rights data, particularly imperfect or potentially biased data. Other challenges stem from key differences between scientific caution and the courtroom. These challenges include appropriately evaluating expertise, the difference between the notions of proof in science and the law, the limitations of statistics in establishing causal links, and the difficulty of adapting the broad statistical picture to the narrow focus of a prosecution narrative.

This paper will discuss the important role statistical analysis can play in international criminal cases, the challenges to the effective introduction of statistics as evidence, and lessons for legal professionals and social scientists for future cases. The paper draws on case studies from international tribunals as well as domestic cases.

Oral [11] Social Science Evidence and the Special Court for Sierra Leone

Catrien Bijleveld^{1,2}, Barbora Hola¹, Aurelie Poelhekke³

¹VU University Amsterdam, ²Netherlands Institute for the Study of Crime and Law Enforcement, ³Human Rights Watch

International crimes, such as genocide, crimes against humanity and war crimes, are widespread, systematic violations of international human rights often committed within a period of armed conflict. In the last decade, many international criminal courts and tribunals have been established to prosecute the most responsible perpetrators of these crimes. Large-scale victimization, settings of upheaval, the collapse of state bureaucracies, and attempts by officials to cover up the crimes and often their own involvement usually create particular problems with respect to evidence. As a result, getting information about the extent of these crimes and number of victims is extremely challenging and often can be done only by social science research using statistical modelling and estimation. In this paper we examine the use of social science documentary evidence by the Special Court for Sierra Leone (“SCSL”). We investigate how often such social science was used by the courts, whether it was contested at trial, whether there are trends over time in the employment of such evidence, and for what elements of the crimes such evidence was admitted.

Oral [11] Statistical Issues Arising in Class Action Cases: Application to the Analyses Presented to the Court in *Dukes v. Wal-Mart II*

Weiwen Miao¹, Joseph Gastwirth²

¹Haverford College, ²George Washington University

After the Supreme Court rejected plaintiffs’ request to certify a nation-wide class action alleging that women were discriminated against in pay and promotion in *Dukes v. Wal-Mart*, the lawyers filed a new suit focusing on all three regions of the company in California. For each of the managerial positions at issue, the plaintiffs’ analysis considered data for each of the six years in each of the districts in a region. Then the yearly data were combined to carry out a test for each district. For each region, the analyses were summarized in a table reporting the number of districts in the region where women received fewer than their expected number of promotions and the number of those which were statistically significant. The court noted that for Managerial Trainees, who were selected at the district level, in only one of the three regions was there a statistically significant shortfall of promotions given to women in a majority of its districts. By using a power analysis, which provides an estimate of the number of significant results would be found if the odds a woman received a promotion were only a fixed fraction (e.g. 0.70) of those of a man, it will be shown that the court’s criteria is far too stringent. The plaintiffs did not provide a power analysis, and actually used the normal approximation (Z-statistic) to find the p-values of statistical tests to data at the district level. As the numbers of promotions were very small in some districts, it will be seen that the normal approximation to binomial under-estimated the “tail probability”. Statistical checks to ensure that the patterns of data from smaller units are sufficiently similar that they can be combined into one summary analysis will be described and used in conjunction with appropriate combination methods to reanalyze several data sets from the case. It will be seen that some the results might have strengthened the plaintiffs’ presentation.

- Gastwirth JL, Greenhouse SW (1987) Estimating a common relative risk: application in equal employment, *J. Amer. Statist. Assoc.* 82, 38-45.
- Gail M, Simon R (1985) Testing for qualitative interactions between treatment effects and patient subsets, *Biometrics* 41, 361-371.

Oral [11] Whose Statistical Analysis Was Correct?

Joseph Gastwirth¹, Wenjing Xu², Qing Pan¹

¹George Washington University, ²KPMG

An error in the computer program that selected eligible residents of Kent County for jury service led to a noticeable shortfall of African-Americans on jury venires from April 2001 to August 2002. Several African-Americans who were found guilty in trials that time appealed their convictions because their right to a jury composed of a “fair cross-section” of the community was violated. The same statistical analyses were submitted in virtually all these appeals. On June 28, 2012, the Sixth Circuit found that the statistics supported the defendant’s claim in *Ambrose v. Booker*

while the Michigan Supreme Court decided that they did not in *People v. Bryant*. The major reason for this discrepancy is that the Michigan Court adopted the “disparity of the risk” measure for evaluating the difference between proportion (p) of African-Americans on the jury venires and their proportion (π) of the jury eligible population. In contrast, the 6th Circuit accepted the results of a statistical hypothesis test demonstrating that the shortfall was statistically significant in conjunction with meaningful differences in both the absolute and comparative disparity measures of minority shortfall. The “disparity of the risk” measure is shown to be the Kolmogorov-Smirnov (KS) distance between two binomial distributions, with success probabilities p and π , respectively. The criteria that its value is at least .50, adopted the Michigan Supreme Court, will be shown to be quite stringent; corresponding to an “effect size” of at least 1.3 standard deviation units in order for the KS distance to reach this value. Such large “effect sizes” are rarely seen in practice, where a shift of .8 (.5) is considered large (moderate). Rather than adopt a “new” measure, courts should rely on the selection ratio or relative risk or odds ratio, which are statistically sound measures that have been used in both discrimination and epidemiologic studies.

- This paper continues the research in Gastwirth, J.L. and Pan, Q. Statistical measures and methods for assessing the representativeness of juries: a reanalysis of the data in *Berghuis v. Smith*. *Law, Probability & Risk* 2011, 10(1): 17-57. Many references to the statistical and legal literature on the topic are given there. The statistical studies used in the cases discussed were carried out by Professors Stephenson and Rothman but they have not been published.

Oral [6b] The Multivariate Kernel Likelihood Ratio Method Applied on Comparison of Amphetamine Seizures
Kjell Andersson, Carina Högborg, Anders Nordgaard
Swedish National Laboratory of Forensic Science (SKL)

Comparison of seizures of amphetamine with respect to their origins of illicit manufacturing can be done by investigating the amphetamine impurity pattern. Such an impurity pattern is a result of an incomplete cleaning-up process – typical for illicit manufacturing – when producing the drug. The manufacturing process can be divided into three steps: (1) choosing a recipe for how to produce; (2) producing amphetamine oil; and finally (3) precipitating the amphetamine from the oil.

The impurity pattern of the amphetamine will depend on the recipe itself, the conditions used for the synthesis, the precipitation process and the method of cleaning-up. The impurity profile is a chromatogram of around 150 different contaminants, of these contaminants 26 have been used by several European countries in police intelligence work to link manufacturers of illicit drugs [1]. However, the linkage methods used are investigative and not evaluative.

The issue addressed when two specific seizures are to be compared, and the results are going to be used in the court of law, is whether they originate from the same precipitation batch. When this is true the impurity patterns of the two seizures are in general expected to be similar, at least for stable contaminants. This is a less expected result if the seizures originate from different batches.

Interpretation of observed similarities and differences between the impurity patterns of two seizures is still to a large extent based on subjective judgements where in Sweden the experiences of two forensic experts are used. In this presentation we show how the so-called multivariate kernel likelihood ratio approach [2] can be used for this interpretation. From a designed experiment comprising several recipes, the variance components for a subset or for a lower-dimensional projection of all contaminants are estimated and likelihood ratios can then be easily calculated. A cross-validated study shows high sensitivity as well as high specificity of the likelihood ratios.

[1] Final forensic report JLS/2006/AGIS/128 Collaborative Harmonised Amphetamine Initiative (CHAIN) AGIS Programme, European Commission, Directorate – General Justice, Freedom and Security.

[2] Aitken CGG, Lucy D (2004), Evaluation of trace evidence in the form of multivariate data, *Appl Statist* 53, 109-122.

Oral [6b] Evaluation and Interpretation of Handlebar Grip Residue Evidence in Hit and Run Cases

Yuanfeng Wang, Zeyu Lin, Jianwei Liu, Baoguo Fang, Fumin Chu, Tianshui Yu
Key Laboratory of Evidence Science (CUPL), Ministry of Education

In this report, the authors presented a study on the analysis and evaluation of handlebar grip residues from suspect vehicles in hit and run traffic cases in China. As we all know, China is a developing country where the bicycle, the electric bike and the motor play important roles in the modern transport. The three common means of transport in China are practical and economical. However, more and more bicycle/electric bike/motor have been involved in traffic cases. Meanwhile, serious injuries have frequently happened to the rider, because most of them did not wear a helmet. Sometimes, there was indeed a collision between the perpetrator and accident victim or their vehicles. In this case, the handlebar grip residue from the bicycle/electric bike/motor always transferred to the surface of the suspect vehicle. Whereas, there was sometimes no collision between the two parts and the accident victim fell down because he/she changed the direction suddenly during the accident in order to avoid collision. In this case, the grip material contamination from daily life might confuse the fact finder. Different facts will lead to different judgements in the courtroom. Thus, it is necessary to set up a scientific method for tracing the facts in hit and run cases. Especially, we need to be qualified enough to distinguish the situation with slight collision from the situation without collision. 50 samples of bicycle handlebar grip, 50 samples of electric bike handlebar grip and 100 samples of motor handlebar grip have been collected randomly by the local police. Scanning electron microscope / Energy dispersive X-ray spectroscopy (SEM/EDX) and Fourier transform infra-red micro-spectroscopy were utilized as analytical technology. First, samples were grouped by the typical elements shown in their EDX spectrum. Second, samples from the same group were differentiated further by the statistical data of their FTIR spectrum. Meanwhile, the contamination of grip material on the vehicle surface has been investigated as well. Likelihood ratios and Bayesian networks have been utilized to achieve the evaluation and interpretation of handlebar grip residue evidence.

It is possible to explore the evidential value of handlebar grip residue evidence through instrumental analysis and statistical methods.

Oral [6b] Progress toward Development of Error Rates for the Presence of Ignitable Liquid Residue in Fire Debris
Michael Sigman, Erin Waddell, Mary Williams
University of Central Florida

Fire debris analysis in arson cases is complicated by heat-induced changes to the composition of the ignitable liquid used to set the fire and the production of interfering products through pyrolysis mechanisms. Current practices for determining if fire debris samples are positive or negative for ignitable liquid residue are subjective and dependent on the experience of the analyst. This presentation will give an overview of research aimed at developing a numerical method for classifying samples as positive or negative for ignitable liquid residue and the determination of true positive and false positive classification rates. Numerical methods used in the research include linear and quadratic discriminant analysis, and soft independent modelling of class analogy (LDA, QDA and SIMCA) [1,2].

The data used in these analyses were the average electron ionization mass spectra across the chromatographic profile, which has been shown to have sufficient information content to classify ignitable liquid complex mixtures with fewer than approximately 80 indistinguishable pairs in one million comparisons [3,4]. The sets of average mass spectra were analysed by principal components analysis and the number of principal components retained was determined by the DRMAD method. Models for LDA, QDA and SIMCA were developed by withholding 20% of the training sets for cross validation. One training set was composed of data from 460 ignitable liquids and 88 burned substrates. The

second training set included the initial data set along with 4600 electronically generated samples containing 20% burned substrate contribution and 4400 electronically generated mixtures of burned substrates. The models were tested on 89 fire debris samples that were deemed positive for ignitable liquid (samples taken on the ignitable liquid pour trail) and 40 that were negative for ignitable liquid (collected off the pour trail). True positive rates from cross validation ranged from 81.7%–98.3%, with false positive rates of 1.1%–25.4%. Testing on fire debris samples resulted in a true positive rate of 82.9%–96.8% and false positive rates of 1.3%–21.6%.

Results for the discriminant methods will be discussed in the context of the database used to develop the discriminant models and the properties of an optimal database for model development.

[1] Waddell EE, Song ET, Rinke CN, Williams MR, Sigman ME, (2013) *J. Forensic Sci.* 58, 887-896.

[2] Waddell EE, Williams MR, Sigman ME (2014) *J. Forensic Sci.* 59, 927-935.

[3] Sigman ME, Williams MR, Castelbuono JA, Colca JG, Clark CD (2008) *Instrumentation Sci Technol* 36, 375-393.

[4] Frisch-Daiello JL, Williams MR, Waddell EE, Sigman ME (2014) *Forensic Sci Int* 236, 84-89.

Oral [6b] Lead Isotopes as Evidence of Bullet Source

David Lucy¹, Knut-Endre Sjøstad²

¹Lancaster University, ²University of Oslo

Conventional forensic bullet interpretation involves the microscopic examination of acquired striation features on a fired projectile, and comparison to a firearm supposed to have fired said bullet. However, in the absence of a firearm with which to compare to any fired bullet, or extreme deformation of the bullet due to impact, forensic scientists have, in the past, used a comparison of the chemical properties of the bullet, and bullets in the possession of a suspect to establish a link between bullet and suspect. This practice has now been all but abandoned by forensic laboratories worldwide due to some disquiet over the safety of comparisons made by such methods. This paper considers the problems of within-box heterogeneity, and a possible solution using a hierarchical mixture model, and subsequent likelihood ratio calculation. This is applied to a sample of lead isotopic observations and seeks to address many of the criticisms made of these comparisons.

Keynote [8] Expressing Activity Level Propositions Using Chain Event Graphs: How to Address Asymmetric Evidence

Anjali Mazumder, James Smith

University of Warwick

The case assessment and interpretation framework (Cook et al. 1998a, 1998b) commonly used by forensic scientists in the UK involves a hierarchy of propositions to be addressed in casework. Forensic scientists are often focused on addressing source or sub-source level propositions, e.g. what is the source of the DNA sample or glass fragment. However, there is increased interest in addressing activity level propositions (e.g. did Mr Y break the window) to better assist the courts (without straying outside the bounds of scientific knowledge). The pairing, framing and development of at least two competing propositions must be done with careful consideration of the case circumstances and evidence (Evetts et al., 2000).

There is a plethora of literature on the use of Bayesian Networks (BNs) for forensic science (Taroni et al, 2006) which expresses the graphical and probabilistic relationship between measured variables. BNs have been particularly useful in providing a graphical representation of the problem, calculating marginal and conditional probabilities of interest, and making inferences particularly addressing lower level propositions. To address activity level propositions, there is a need to account for different plausible explanations of suspect/perpetrator's actions and events as it relates to the evidence. In this talk, we propose the use of another class of graphical models, chain event graphs (CEGs), exploiting event tree structures to depict the unfolding of events (Smith and Anderson, 2007) as postulated by each side (defence and prosecution) and differing explanations/scenarios. Different scenarios can introduce different sets of relevant information

affecting the dependence relationship between variables and symmetry of the structure. CEGs are a flexible class of graphical models which can model the asymmetric story structure directly in its topology. Yet because of its graph modular structure it also inherits many benefits of the BN. A BN can be represented as a symmetric CEG but the BN is not always a rich enough structure to incorporate all obtainable information (Barclay et al. 2013). We demonstrate how CEGs can be very useful in addressing activity level propositions and assist the courts by directly supporting the barrister's argument within the topology of a graph, particularly in complex cases. With the use of case examples involving transfer and persistence and different evidence types, we further show how CEGs can assist in the careful pairing and development of propositions and analysis of the evidence by addressing the uncertainty and asymmetric unfolding of the events to better assist the courts.

- Barclay LM, Hutton JL, Smith JQ (2013) Refining a Bayesian Network using a Chain Event Graph, *International Journal of Approximate Reasoning* 54, 1300-1309.
- Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998a) A model for case assessment and interpretation, *Science & Justice* 38, 151-156.
- Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998b) A hierarchy of propositions: deciding which level to address in casework, *Science & Justice* 38, 231-239.
- Evett IW, Jackson G, Lambert JA (2000) More on the hierarchy of propositions: exploring the distinction between explanations and propositions, *Science & Justice* 40, 3-10.
- Smith JQ, Anderson PE (2008) Conditional independence and chain event graphs, *Artificial Intelligence* 172, 42-68.
- Taroni F, Aitken C, Garbolino P, Biedermann A (2006) *Bayesian Networks and Probabilistic Inference for Forensic Science*, Wiley.

Oral [8] Evaluating New and Missing Evidence in Criminal Cases Using Bayesian Networks

Nadine Smit¹, Ruth Morgan¹, David Lagnado¹, Norman Fenton²

¹University College London, ²Queen Mary University of London

Interpreting forensic evidence in the context of a case requires logical powers of reasoning, which can possibly be achieved by using multivariate analytical methods such as Bayesian networks. Currently, in forensic domains with a less apparent probabilistic nature, there is a lack of studies that focus on an integrated framework of collection, analysis and interpretation. The Bayesian approach can address this by offering a general but context-sensitive approach to guide the evaluation of evidence and to support decision making in such investigations.

This presentation aims to show how a Bayesian network can be used to represent the reasoning of the prosecution and defence regarding new evidence in a criminal case that is considered for appeal. A transparent model is provided showing the relation between source and activity level hypotheses and the underlying assumptions and probabilities. Additionally, the results of a sensitivity analysis show that if incorrect assumptions are made, they can have a major impact on the significance of the evidence. It is argued and shown that without specific experimental studies (even though an exact replication is not possible due to the complex nature of most forensic evidence), the items of evidence may not provide the desired impact. In this case, the evidence presented by the defence could support the hypothesis of the prosecution depending on the results of yet unexecuted experimental studies. The presentation will discuss the findings that, in contrast to the current practice of many experts, in this case it is more valuable to determine the probability of observing (dis)similarities when items of evidence are different than to confirm similarities between items of evidence that are thought to be the same. Lastly, the problem of missing information at a source level is addressed using a case involving geo-forensic evidence. It is shown how a terrible misinterpretation at the activity level could have been avoided by following a Bayesian framework.

Oral [8] Human Biological Trace Investigations: A Probabilistic Approach to Interpreting Test Results

Tialda de Wolff, Arnoud Kal, Bas Kokshoorn

Netherlands Forensic Institute

The interpretation of forensic investigations follows a hierarchy of propositions. From subsorce ('who is the donor of the DNA?') to source ('what kind of biological trace is present and who is the donor?'), from source to activity ('how did the defendant's semen get on the victim's clothing?'), and from activity to offense ('did the defendant rape the victim?').

It is the task of the forensic scientist to provide the judge or jury with robust and transparent evaluations of forensic results that may help them to decide on the crucial issues in the case.

The forensic scientific community has invested strongly in the interpretation and evaluation of results at subsorce [1] and activity level [2]. Interpretation at source level has not yet received as much attention. Most efforts at this level have addressed the association of body fluid testing and DNA results [3]. However, a robust and transparent interpretation of the body fluid testing results has -to our knowledge- not yet been attempted.

Interpretation of test results of biological trace investigations is exceedingly complex, given that the tests that are commonly used all display false positive and false negative rates, that degradation of the material is a factor, and that tests show cross-reactivity in other human body fluids and substances.

One such a forensically relevant body fluid is saliva. The presence or absence of saliva (of a person of interest) may have considerable evidential value depending on the case context. Reporting officers will most often report either plain test results ('test for saliva positive'), provide a suggestive statement ('indication for the presence of saliva in the sample') or use a 'fall-of-the-cliff' decision model ('given that the sample was taken from a cigarette butt, and that the test was positive, the sample contains saliva').

Here we present a graphical model (Bayesian Network) for the interpretation of test results for human saliva, based on the presence of human α -amylase.

The model can be used by reporting officers as an exploratory tool to formulate their expert opinion on the presence or absence of saliva in a sample. Use of the model will promote correct evaluation, increase uniformity and reduce bias in reporting. The model will be extended to other body fluids and tests. It may also be integrated in networks addressing activity level questions. The practical use of the model is demonstrated using case examples.

[1] Kelly et al. (2014) A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science and Justice* 54, 66-70.

[2] McKenna L (2013) Understanding DNA results within the case context: importance of the alternative proposition. *Frontiers in Genetics* 4, article 242.

[3] Hartevelde et al. (2013) RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated? *Science and Justice* 53, 261-269.

Oral [8] Bayesian Networks as a Tool for Interpreting DNA Results at Activity Level

Bart Blankers, Marjan Sjerps

Netherlands Forensic Institute

A forensic expert has a duty to the court to provide the judge or jury with robust and objective information that may help them decide on crucial issues in the case. Often this information is provided by means of a likelihood ratio. This likelihood ratio can be used by the court as a measure for the evidential value of the findings.

A likelihood ratio follows from an evaluation of the results under competing propositions. The propositions can address issues on several levels. These levels are known as the hierarchy of propositions. This hierarchy distinguishes source ('what is the source of the material?'), activity ('what activity resulted in the findings?') and offence ('what offence has been committed and by who?').

Over the last decade much emphasis has been on evaluations on source level. The significance of interpreting results from DNA investigations on activity level has been acknowledged by the forensic and judicial communities. However, a relatively small amount of resources have been invested in addressing activity level issues. In part this is due to the large number of relevant variables (such as transfer, persistence and recovery) and their conditional dependencies. In combination with a need for relevant

data this can make interpreting DNA results on activity level very complex.

Bayesian Networks can graphically model a large number of variables and their conditional dependencies. They have been recognized as a means for interpreting forensic findings [1] and the application for evaluating DNA results on activity level has been described [2,3]. A Bayesian Network can model all the relevant variables and their conditional dependencies, provide insight into the necessary data and shows the effect a variable has on the conclusion.

We show an example of a Bayesian Network that can be used for interpreting DNA results on activity level. The model incorporates all relevant variables that can influence the findings. Using a case example we will also demonstrate how Bayesian Networks can be implemented in casework and how they can assist in forming a conclusion.

[1] Taroni F, Aitken CGG, Garbolino G, Biedermann A, Bayesian Networks and Probabilistic Inference in Forensic Science, John Wiley & Sons, Chichester, 2006.

[2] Biedermann A, Taroni F (2012) Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature, *Forensic Sci Int Genet* 6, 147-157.

[3] Evett IW, Gill PD, Jackson G, Whitaker J, Champod C (2002) Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, *J Forensic Sci* 47, 520-530.

Thursday Afternoon

Keynote [6] Bayesian Calibration for Forensic Evidence Reporting

Niko Brummer, Albert Swart

AGNITiO Research

State-of-the-art algorithms for automatic, biometric speaker comparison are based on probabilistic models that formally produce output in the form of the ratio of likelihoods, for the same-source hypothesis, versus the different-source hypothesis [1]. This suggests these likelihood-ratios could be used directly for evidence reporting [2,3]. Unfortunately, these models do not represent the speech with sufficient accuracy to give well-calibrated likelihood-ratios. That is, when Bayes decision theory is applied in practice to these likelihood-ratios, on average, sub-optimal decisions result [4].

This can be improved by applying a second layer of probabilistic modelling, known as calibration, where the output (known as the score) of the automatic speaker recognizer is regarded as the evidence. The score can then be modelled in turn via a much simpler, and generally more accurate, probabilistic score model, which gives demonstrably well-calibrated likelihood-ratios [5-12].

In this paper, we model scores via hypothesis-dependent univariate normal distributions, the parameters of which can be estimated from the scores that result from running the same algorithm on a collection of background speaker comparison experiments. If these background speech samples are hand-selected by a human practitioner to match the circumstances of the trial, typically very few samples may be available, leading to non-negligible uncertainty in the parameter estimates. We demonstrate experimentally that simply plugging in such uncertain parameter estimates into the calibration model, leads on average, to overconfident belief in one or the other hypothesis, as well as suboptimal decisions.

This problem can be addressed by extending probabilistic modelling to the model parameters, giving a Bayesian joint posterior distribution over both unknowns: the parameters and the hypothesis. Marginalizing over the parameters gives a Bayesian likelihood ratio [11], which we experimentally demonstrate gives better decisions, on average, than the above plugin recipe.

We further show that the Bayesian likelihood-ratio has the intuitively pleasing behaviour of becoming less confident (giving smaller log-likelihood-ratio magnitude), as the number of background samples used for calibration becomes smaller. In contrast, the plugin method behaves counter-intuitively, becoming more confident as the number of calibration samples decreases. For full details, see [12].

- [1] Kenny P, Bayesian speaker verification with heavy-tailed priors, in *Odyssey 2010*, 249.
- [2] Balding D, *Weight-of-Evidence for Forensic DNA Profiles*, Wiley, 2005.
- [3] Jaynes ET, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
- [4] Brummer N, du Preez JA (2006) Application-independent evaluation of speaker detection, *Computer Speech and Language* 20, 230.
- [5] Brummer N, Measuring, refining and calibrating speaker and language information extracted from speech, Ph.D. dissertation, Stellenbosch University, 2010.
- [6] Gonzalez-Rodriguez et al. (2007) Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition, *IEEE/ACM Trans. Audio, Speech, Language Process* 15, 2104.
- [7] Ramos D, Forensic evaluation of the evidence using automatic speaker recognition systems, Ph.D. dissertation, Universidad Autónoma de Madrid, 2007.
- [8] Mandasari et al. (2013) Calibration based on duration quality measure function in noise robust speaker recognition for NIST SRE'12, in *Biometric Technologies in Forensic Science*, Nijmegen, 2013.
- [9] van Leeuwen D, Brummer N (2013) The distribution of calibrated likelihood ratios, *Interspeech*.
- [10] Brummer N, Doddington G (2013) Likelihood-ratio calibration using prior-weighted proper scoring rules, *Interspeech*.
- [11] Brummer et al. (2014) A comparison of linear and non-linear calibrations for speaker recognition, in *Odyssey 2014*, Joensuu, Finland.
- [12] McLaren et al. (2014) Trial-based calibration for speaker recognition in unseen conditions, in *Odyssey 2014*, Joensuu, Finland.
- [13] Brummer N (2011) Fully Bayesian score calibration assuming Gaussian distributions, AGNITIO Research, South Africa, Tech. Rep., <http://niko.brummer.googlepages.com>
- [14] Brummer N, Swart A, Bayesian calibration for forensic evidence reporting, *Interspeech 2014*, accepted.

Oral [7] Familial Searching on Degraded Mixture Samples *Klaas Slooten*

Netherlands Forensic Institute

Traditionally, familial searching aims to identify relatives of the unknown donor of a single source crime stain by computing likelihood ratios for relatedness between the profile obtained from the stain and that of a known person.

Recently, Chung, Fung and Hu have extended this approach to a two-person mixture composed of a known victim and an unknown perpetrator, in case all alleles of both contributors have been obtained in the DNA profile.

However, many mixtures, e.g. when low template DNA profiling has been carried out, are susceptible to allelic dropout (meaning that not all alleles of all donors are visible in the profile) and drop-in (detected alleles that do not come from the contributors).

We have therefore further extended the method to be able to deal with mixtures where both drop-out and drop-in are possible.

Our results show that also for such degraded mixtures, a familial search targeted at finding parents, children or siblings of the unknown offender is often feasible. We use likelihood ratio computations based on a semi-continuous model that allows different drop-out rates per contributor.

If a database search is performed, we will show what the probability is of finding relatives in a top-k in a database search for various mixture classes (two or three contributors, with various ranges of drop-out and with or without known contributors) and what the false positive and false negative rates are when a LR-threshold is chosen.

We will show how these values can be predicted prior to carrying out a familial search, so that investigators can choose a search strategy depending on the outcome of such a pre-assessment.

- Haned et al. (2012) Exploratory data analysis for the interpretation of low template DNA mixtures, *Forensic Sci Int Genet* 6, 762-774.

Oral [7] Implementation of Probabilistic Models for the Evaluation of DNA-Mixtures in Casework

Jerien Klaver, Klaas Slooten

Netherlands Forensic Institute

Since 2013 the Netherlands Forensic Institute (NFI) routinely uses a binary and a semi-continuous model for the evaluation of DNA-mixtures. A binary model only takes the called alleles into account, whereas the semi-continuous model assumes different dropout rates per donor and is therefore to some extent able to deconvolute the mixture. In many cases these models are able to extract enough information from the profiles in order to discriminate well between donors and non-donors. They are used by reporting officers as a tool to assess the strength of DNA

evidence. The results of the analyses are used to support an expert opinion which is what is ultimately reported to the court. We describe how these analysis have changed DNA-mixture interpretation over the last year. We show cases in which we are able to calculate the strength of the evidence, where it was not possible to do so without using these models. We describe the methodological framework that is currently used at the NFI and how we incorporated the use of these models in this framework and in our case-work management to decide when to perform (sometimes laborious analysis) and when not. We will focus on types of DNA-mixtures that are suitable for an analysis, how this analysis is performed and what the strength of the evidence is for some typical examples.

- Haned et al. (2012) Exploratory data analysis for the interpretation of low template DNA mixtures, *Forensic Sci Int Genet* 6, 762-774.

Oral [7] Evidential Evaluation of DNA Profiles Using Discrete and Continuous Models

Roberto Puch-Solis, Tim Clayton
LGC Forensics

The ever increasing sensitivity in the technology for producing DNA profiles poses interpretation challenges where stain profiles are affected by spurious alleles (allelic dropout), failure to detect alleles (allelic dropout) and degradation. In addition, for low template DNA (LTDNA), two or three stain replicates may be produced. Statistical methods for assessing the evidential strength of profiles are required to model these phenomena whilst take into account multiple replicates and to factor them simultaneously into the likelihood ratio (LR).

Two types of model have been proposed in the literature: discrete (or semi-continuous) and continuous models. The former utilises the absence/presence of allelic and artefactual peaks, whilst continuous models incorporate additional quantitative information embodied by the peak height or area. Continuous models utilise more of the information from the evidential profiles and for this reason are construed as a fuller treatment of the problem. However, the models are more complex as a result and so are more difficult to implement. A number of groups worldwide have developed different solutions – some continuous and some discrete. The relative performance of these models is a matter of genuine interest given the different approaches being taken within the same jurisdiction and internationally.

In this presentation we report the findings of a comparative study in which DNA profiles, including profiles from more than one person, are evaluated using both a discrete model, (LiRa) (Puch-Solis & Clayton, 2014), and also a continuous model, (LiRaHt) (Puch-Solis et al., 2013; Puch-Solis, 2014). The discrete model takes account of uncertain allelic designation, degradation, dropout, dropin and multiple replicates, whilst the continuous model also considers allelic, stutter and dropin peak heights. We compare outputs from both models and attempt to explore the attributes and drawbacks of both. We believe that this work will contribute to the body of knowledge that underpins the use of statistical models in casework.

- Puch-Solis R, Clayton T (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software, LGC Research Report LGC/P/2013/165, submitted to *Forensic Sci Int Genet*.
- Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, Balding D (2013) Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters, *Forensic Sci Int Genet* 7, 555-563.
- Puch-Solis R (2014) A dropin peak height model, accepted for *Forensic Sci Int Genet*.

Oral [7] Verifying Likelihoods for Low Template DNA Profiles Using Multiple Replicates

Christopher Steele¹, Matthew Greenhalgh², David Balding¹
¹UCL Genetics Institute, ²Orchid Cellmark Ltd

To date there is no generally-accepted method to test the validity of algorithms used to compute likelihood ratios (LR) evaluating forensic DNA profiles from low-template and/or degraded samples. An upper bound on the LR is provided by the inverse of the match probability, which is the usual measure of weight of evidence for standard DNA profiles not subject to the stochastic effects that are the hallmark of low-template profiles. However,

even for low-template profiles the LR in favour of a true prosecution hypothesis should approach this bound as the number of profiling replicates increases, provided that the queried contributor is the major contributor. Moreover, for sufficiently many replicates the standard LR for mixtures is often surpassed by the low-template LR. It follows that multiple LTDNA replicates can provide stronger evidence for a contributor to a mixture than a standard analysis of a good-quality profile. Here, we examine the performance of the likeLTD software for up to eight replicate profiling runs. We consider simulated and laboratory-generated replicates as well as resampling replicates from a real crime case. We show that LRs generated by likeLTD usually do exceed the mixture LR given sufficient replicates, are bounded above by the inverse match probability and do approach this bound closely when this is expected. We also show good performance of likeLTD even when a large majority of alleles are designated as uncertain, and that there can be advantages to using different profiling sensitivities for different replicates. Overall, our results support both the validity of the underlying mathematical model and its correct implementation in the likeLTD software.

Oral [7] Ramifications of the Dlugosz Judgment in the UK
Ian Evelt, Sue Pope

Principal Forensic Services Ltd

The judgment in *R v Dlugosz* relates to three cases in each of which the issue was the interpretation of a complex mixed DNA profile from a crime sample. The judgment ruled that if a scientist cannot provide a numerical assessment of the weight of the evidence in such a case, the jury should form their own assessment based on the number of components in the mixed profile which match the profile of the defendant. It is also ruled admissible for the scientist to assist the jury with a qualitative opinion based on experience.

We argue that: (a) an assessment of weight based on the number of matching bands fosters a prejudicial view of the evidence; (b) scientists are not trained to provide qualitative assessments of weight of evidence in such cases and no system exists for assessing the robustness of such assessments.

- *R v Dlugosz* 2013 EWCA Crim 2

Keynote [1] Assessment of Noise in next-Generation DNA Sequencing Applications for Forensic Purposes

Mark Wilson, Brittanina Bintz, Erin Burnside
Western Carolina University

Using newly emerging methods of DNA sequencing, often referred to as next-generation sequencing (NGS), we have developed robust protocols designed to generate whole human mitochondrial DNA genome sequencing data from DNA samples, such as buccal swabs, blood samples, and hair shafts. Our analyses have revealed that there are many potential sources of variation within mtDNA sequences obtained from a questioned sample or a reference sample using NGS. These sources generally fall into four categories: fixed changes resulting from mutational events (polymorphisms); background noise obtained from the instrument or procedure; low-level short-lived mutational variants subject to loss via genetic drift; the co-amplification of nuclear pseudogenes.

An analysis of the instrument or procedure-based noise is essential to developing an understanding of the physical limitations of the procedure and thereby subsequently formulating a proper and defensible interpretation of comparisons for forensic purposes. We have identified a number of parameters pertaining to our particular NGS procedure that affect the quantitative assessment of noise. These include such things as sequence quality, alignment, bi-directionality (balance), depth, and directional sequence motif.

Each of these parameters must be estimated from empirical data. Once sufficient data has accumulated and been subject to statistical analysis, then it will be possible to utilize NGS for forensic comparison purposes.

Oral [1] Tracing the Source of a Collected Microbial Pathogen to a Specific Laboratory Derived Culture

Petter Lindgren, Mats Forsman, Jon Ahlinder
Swedish Defence Research Agency (FOI)

Amerithrax, the investigation of the anthrax letter attacks in USA 2001, resulted in the emergence of a new scientific area in biology: the field of microbial forensics. Since then, microbial forensics has become more standardized and has played an increasingly important role in crime and epidemiological investigations. A unified model-based statistical framework is of key importance for classification of pathogenic bacterial culture (Budowle et al. 2005).

Massively parallel sequencing techniques allow characterization of a sub-population, such as a growing culture of bacteria, at a much finer scale than previously possible. Microevolution within a growing culture will induce mutations at low frequencies that were not detectable with previous sequencing techniques. With the next-generation-sequencing, however, detection of low frequent single nucleotide polymorphisms (SNP) in a culture is now feasible, which enables inferences of identity matching based subpopulation profiles occurring in a culture.

Source identification in microbial forensic is a supervised learning or classification problem where an unknown sample needs to be probabilistically assigned to known classes consisted of training or reference data. The classifier builds a predictive model from the training data to trace the correct class. In the microbial forensic context here, such classes could, for example, correspond to different laboratories which have access to the pathogen of interest. Thus, samples needs to be retrieved from each laboratory including the suspect one.

A statistical framework is presented, with the objective of calculating evidence values of interest in legal trials, where source attribution is the key concern. Evidence values in favor of the hypothesis, that a suspect laboratory is the source of the pathogenic culture, are based on the Bayes factor method (Kass and Raftery 1995) for comparison of alternative hypothesis.

A method to simulate large scale bacterial growth is developed, which can be used to obtain data typical to microbial forensic investigations. The step underpinning the method is based on a population genetic model. Simulated mutational patterns, specific to each laboratory, help to trace the source of the pathogenic culture in a number of scenarios.

Result suggests that a pathogenic culture could be allocated to its correct laboratory source with high precision. We believe that this approach will have important impact on future microbial forensic investigations.

- Budowle et al. (2005) Genetic analysis and attribution of microbial forensic evidence, *Crit Rev Microbiol* 31, 233-254.
- Kass RE, Raftery AE (1995) Bayes Factors, *J Am Statist Assoc* 90, 791.

Oral [1] On the Measurement and Interpretation of Bullet Lead Isotope Ratios, a Not so Stable Approach

Knut-Endre Sjøstad^{1,2}, David Lucy³, Tom Andersen¹

¹University of Oslo, ²National Criminal Investigation Service Norway, ³Lancaster University

An increasing focus of interest in the forensic sciences has been the analysis of stable isotope ratios of both light, and heavy elements. The analysis of radiogenic isotopes, such as lead, has been given less attention, despite the fact, geologically, these isotopes are inherently more variable than their stable isotopes counterparts.

In this presentation, the authors will present the lead isotope ratios as a useful method for forensic scientists in the comparison of any lead bearing artefacts, or naturally generated traces. Measurements are made by state of the art instrumentation, with extremely low detection limits, high sensitivity and high resolution. Due to the simplicity of the data, an illustrative method of "LR-capes" will be introduced, a concept inherited from isoscapes, a (geo)graphical presentation of the variation of stable isotopes. The LR-capes will be used as an example of how to communicate scientific findings in an easy accessible way for the layman as well as between scientist dealing with forensic interpretation.

- Stupian GW (1975) Lead isotope ratio measurements: A potential method for bullet Identification, *J Forensic Sci Soc* 15, 161-164.
- Andraszko et al. (1993) Lead-isotope ratios in lead smears and bullet fragments and application in firearm investigations, *J Forensic Sci* 38.
- McFadden et al. (2004) *Forensic Analysis: Weighing Bullet Lead Evidence*, National Academies Press, Washington, D.C.

Oral [1] Challenges in the Evaluation of Modern Chromatographic Data as Forensic Evidence

Martin Lopatka^{1,2}, Gabriel Vivó-Truyols¹, Marjan Sjerps^{1,2}

¹University of Amsterdam, ²Netherlands Forensic Institute

The abundance of data produced by modern analytical chemical analysis far exceeds the capacity of computational tools normally employed in the calculation of likelihood ratios for forensic evidence. For forensic profiling applications, an assertion of similarity between two chromatograms is often the basis of evidence. This similarity is frequently based on feature selection such that only a fraction of the original data is compared. We present a new method for full signal comparison of chromatographic evidence that does not discard data a priori. Initial analysis of chromatographic data is performed over the first-order signal. We employ a new probabilistic peak detection algorithm that assesses the probability that a compound related peak is exerting an effect for each point.

A comparative analysis between two samples is then performed point-by-point under the assumption that a pair of chromatograms are measurements of the same object. This step is comparable to a sequential Bayesian updating procedure where a priori state estimates are improved by the inclusion of new information to achieve superior a posteriori estimates. The compounding deviance of the error covariance from a fixed model of system noise acts as a quantitative similarity metric between the chromatograms. The probabilistic peak detection serves to determine regions of relevance where comparison of the two signals is most informative.

This resulting global similarity may be used as a metric indicating the closeness of relevant portions of the pair of signals. In this way we derive a similarity metric for comparing full chromatographic signals. A collection of pairwise similarity comparisons between samples of known origins is then used as a reference population for modelling probability densities for same-batch comparisons and different-batch comparisons. These reference distributions may then be used for the estimation of score-based likelihood ratios pertaining to a source-identification paradigm employing full chromatographic similarity as the evidential component. We discuss how this likelihood ratio may be reported in casework.

The presentation will focus on outlining the problem and basic ideas of our approach and will not dwell on the technical details.

Oral [1] Modelling the Sensitivity of Detection Methods in Microbial Forensics

Ronny Hedell^{1,3}, Gunnar Andersson², Petter Mostad³

¹Swedish National Laboratory of Forensic Science (SKL),

²Swedish National Veterinary Institute (SVA), ³Chalmers University of Technology, University of Gothenburg

The bacterium *Bacillus anthracis* is the cause of anthrax, a potentially lethal disease for both humans and animals. As the negative consequences of the presence of *Bacillus anthracis* can be large, it is important to screen objects and areas for it whenever the prior probability of its presence is non-negligible. Such cases may be related to feed and food quality and to bioterrorism. If any of the samples taken is positive for the bacterium, the sampling results can be used as argument for taking actions to hinder or minimize the further spread of the bacteria and as evidence in a juridical process. If the bacterium is not detected, it is harder to draw a conclusion – can one be reasonably sure it is not present? If the sensitivity of the detection method is not perfect, or if not all parts of the object or area are investigated, there is a risk of a false negative conclusion. To evaluate the sampling results with respect to hypotheses about the concentration or distribution of the bacteria the performance and the limitations of the detection method has to be

known. To assess the sensitivity of the detection method an experiment is usually performed where the target material is artificially contaminated with different concentrations of the bacterium, followed by analyses of the samples using the pre-enrichment method and analytical detection method of interest. For safety reasons, economical limits or time limits it is not always possible to perform exactly such an experiment, with the desired number of samples.

We investigate how one can learn about the false negative rate by using data from experiments where some of the experimental parameters are somewhat different compared to the current situation: the bacterium type, material type, the pre-enrichment method or the analytical detection method. Part of our investigation concerns how the modelling can be split up in two parts; into the pre-enrichment step, where the bacteria are growing, and into the analytical detection step with polymerase chain reaction (PCR) technology or with plate counts. We mainly use hierarchical Bayesian methods to assess the uncertainty of the model parameters and to estimate the false negative rates for new combinations of bacterium type and material type.

Oral [1] Food Authenticity as a Forensic Problem

Grzegorz Zadora^{1,2}, Agnieszka Martyna³, Ivana Stanimirova², Daniel Ramos⁴, Patryk Własiuk³

Polish Institute of Forensic Research, University of Silesia, Jagiellonian University in Krakow, Universidad Autónoma de Madrid

The authenticity of food products may be an issue of forensic interest, specifically when it involves economic or health consequences and upholds the consumer rights. Then, the forensic expert can face the classification problem, whether the analysed sample may come from the declared brand A, or another brand named B.

The problem of authenticity is in fact the classification/discriminant problem that can be solved using the physicochemical data obtained from various analytical methods. In order to give an objective answer to the question of the products authenticity, chemometric and statistical methods should be employed for the interpretation of the analytical results. There are many classification methods, but not all of them can express the results in the form which is required if counterfeiting of food products become a legal case. We propose a likelihood ratio methodology to address this problem, aimed to determine the ratio of conditional probabilities $\Pr(E|H_1)$ and $\Pr(E|H_2)$ (known as likelihood ratio, LR), which enables the assessment of data (E) in the context of two contrasting hypotheses: H_1 - the analysed sample comes from brand A, and H_2 - the analysed sample comes from brand B of the wines samples [1, 2].

The aim of the presented research work was to investigate if the LR models used in forensic science could be successfully adapted to verify the authenticity of wine products. The examinations involved 178 wine samples labeled as the wine brands Barolo (BAR), Grignolino (GRI) and Barbera (BRB). Each sample was characterised by 27 parameters e.g. amount of alcohol, hue, and elemental composition. The LR calculations [3] were performed for the raw analytical data and logarithmically transformed. There were proposed 27 univariate models concerning single variables. F-test was applied for selecting the most informative wine parameters for the authenticity determination of selected wines and PCA was used to decorrelate the variables. The selected variables created naïve multivariate LR models. All LR models were checked for their correctness by estimating their misclassification rates and using the Empirical Cross Entropy approach (ECE) [1,2].

For the flavonoids content, hue, colour intensity, alcohol and proline content and the transmittance parameter, the levels of correctly classified wine samples exceeded 90% [3]. The multivariate models built with the reduced number of variables (using F-test and PCA for transformation) proved to deliver satisfying results. The results of our work showed that the LR models could be applied successfully for verification of a food

products authenticity, which was also confirmed by the ECE plots [3].

[1] Ramos D, Zadora G (2011) Information-theoretical feature selection using data obtained by SEM coupled with an EDX spectrometer for the classification of glass traces, *Analytica Chimica Acta* 705, 207-217.

[2] Zadora G, Martyna A, Ramos D, Aitken C (2014) *Statistical Analysis in Forensic Science*, Evidential value of multivariate physicochemical data, Wiley.

[3] Martyna A, Zadora G, Stanimirova I, Ramos D (2013) Wine authenticity verification as a forensic problem, *Food Chemistry* 150, 287-295.

Keynote [5] To Catch a Thief With and Without Numbers

Bart Verheij

Stanford University, University of Groningen

Evidential reasoning is a notoriously complex task, and mistakes can have severe consequences, for instance, when errors in statistical reasoning lead to miscarriages of justice. To help prevent mistakes, three kinds of normative frameworks have been proposed focusing on arguments, scenarios and probabilities, respectively (Kaptein et al., 2009). In an arguments framework, the structure and evaluation of the available arguments for and against claims are analysed. In a scenario framework, guidelines are proposed for the collection and comparison of different scenarios of what may have happened in a criminal case. In a probabilities framework, it is specified how the probabilities of hypothetical events change in the light of new evidence. These three kinds of normative frameworks emphasise very different ingredients of evidential reasoning, and in many treatments, differences prevail over commonalities.

The aim of this presentation is to discuss an integrating perspective in which arguments, scenarios and probabilities have their place. It is concluded that there is no need to choose a normative framework focusing only on arguments, scenarios or probabilities. Argumentative, scenario and probabilistic methods are all valuable in the normative regulation of evidential reasoning, aimed at the prevention of errors.

Specific attention will be paid to the following issues that are relevant for the use of statistical reasoning in criminal trials:

Specificity. The probability of a hypothesis about what has happened depends on how specific it is. So how specific should hypotheses be?

Unavailability. A central criticism of probabilistic methods is that more numbers are required than feasibly available. Are all numbers always necessary?

Relevance and rationality: Even when meaningful numbers are available, final decisions may not depend on their exact value.

Some have even argued against the relevance of exact probabilities. Still the rationality constraints provided by probability theory stand strong. How is that paradoxical situation possible?

Burden of proof: High probability of a hypothesis given the evidence is among the constraints of good evidential reasoning with probabilities. But what is the burden of proof? When is a high probability sufficiently high?

Reasonable doubt: Every decision leaves some room for doubt, sometimes reasonable, sometimes not reasonable. But when is doubt reasonable?

The presented integrated perspective on arguments, scenarios and probabilities in reasoning with evidence will be used to provide an answer to each of these issues.

- Kaptein H, Prakken H, Verheij B (eds.) (2009). *Legal Evidence and Proof: Statistics, Stories, Logic* (Applied Legal Philosophy Series). Farnham: Ashgate.
- NWO Forensic Science project: Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios.

Oral [5] The Judge's Choice

Lonneke Stevens, Linda Kesteloo

VU University Amsterdam

In certain criminal cases, only little incriminating evidence is available to the judge. Additionally, this evidence is often challenged on its reliability by the defense. If part of the evidence is probative for the defendant's guilt, whereas other information is relevant to his innocence, a judge cannot – unlike (legal) scholars – suffice with the consideration that the evidence is inconclusive. The judge has to decide between conviction or acquittal. An important question is how judges should choose between these

two verdicts. How can he assess the reliability of (contradicting) pieces of evidence and how can he give a reliable verdict? These are the central questions of this paper, and we will specifically discuss them in relation to two types of evidence.

The first evidentiary issue concerns the identification of the perpetrator through identification procedures with witnesses. The reliability of identifications can be safeguarded when the used method is in conformity with procedural rules that are based upon psychological research. Interestingly, when the identification procedure does not meet scientific standards and the reliability cannot be (easily) determined anymore, judges assess the probative value of identification results differently. Some judges base their assessment on the relation between the result of the confrontation procedure and other evidence (holistic approach), whereas other judges assess the identification in isolation (atomistic approach). The atomistic approach raises the question which methodological mistakes lead to an unacceptable high risk of a mistaken identification. Another point of discussion concerns the validity of the holistic approach, since procedural mistakes are, strictly speaking, not corrected by the availability of other evidence.

The second evidentiary issue concerns the witness statement that is disputed by the defendant. According to Dutch evidence law, a judge can convict the defendant on a single witness statement, as long as this statement is sufficiently supported by other evidence (of which it is not required that it is directly linked to the criminal conduct itself). Some judges presume that the witness' reliability already support the statement sufficiently. But how does the alleged reliability of only one witness (statement) result in reliable fact-finding? Other judges explain 'sufficient support' with scenario-reasonings. They ask themselves whether certain pieces of information discriminate between the scenario of guilt and the scenario of innocence. However, pieces of information usually provide the judge only with an indication as to which scenario is plausible. But when is the plausibility of a scenario of guilt sufficiently demonstrated?

- Bex FJ, Verheij B (2012) Solving a Murder Case by Asking Critical Questions: An Approach to Fact-Finding in Terms of Argumentation and Story Schemes, *Argumentation* 26, 325-353.
- van Koppen PJ (2011) *Overtuigend bewijs. Indammen van rechterlijke dwalingen*, Amsterdam: Nieuw Amsterdam Uitgevers.
- Loftus EF (1996) *Eyewitness testimony*, Cambridge, MA: Harvard University Press.
- Dubelaar MJ (2014) *Betrouwbaar getuigenbewijs, Totstandkoming en waarderings van strafrechtelijke getuigenverklaringen in perspectief* (diss. Leiden), Deventer: Kluwer.

Oral [5] Limitations and Opportunities of the Likelihood Ratio Approach for Evidence Evaluation

Norman Fenton

Queen Mary University of London

By Bayes' Theorem, the likelihood ratio (LR) is, in principle, an ideal measure of the probative value of a single piece of evidence against a hypothesis and its negation.

However, due to the choice and context of hypotheses, there are common situations where the LR of a piece of evidence may be quite unrelated to its 'probative value'. Hence, the use of the LR may be misleading (I will demonstrate the serious implications of this with examples of probabilistic arguments that were used in both the Barry George and Sally Clark cases). In particular, the most common scenarios are:

1) Where the LR is applied (as is common) to a piece of evidence E with two hypotheses H1 and H2 that are not mutually exclusive and exhaustive. Contrary to most expectations, in this case the LR may tell us nothing about the probative value of H1 against 'not H1'.

2) Where the LR is applied to a piece of evidence with respect to a source-level hypothesis and its negation (such as defendant was/was not at the crime scene). Contrary to most expectations, in this case the LR may tell us nothing about the probative value of the evidence on the relevant offence-level hypothesis (defendant is/is not guilty) of the case.

It is also important to note that the computation of the LR – even for a single piece of evidence against a source level hypothesis –

is often erroneous due to failure to incorporate different types of potential process and testing errors.

All of the above limitations on the use of the LR for evidence evaluation can be addressed by using Bayesian networks that incorporate the necessary offence level hypotheses as well as other (normally unstated) hypotheses needed for correct computation of the LR. Of course, there is much resistance to such an approach (by both forensic practitioners and lawyers) and so the talk will also address why there is resistance and how it can be tackled effectively.

Oral [5] An Integrated Theory of Causal Stories and Evidential Arguments

Floris Bex

Utrecht University

It has been argued that both *arguments* and *stories* are needed in order to do justice to all the relevant reasoning mechanisms as they are recognised and used by legal decision makers and forensic investigators. Stories – in the sense of coherent sequences of events – are needed to organise the complex mass of facts in a case into one or more hypotheses about “what happened” in the case. Arguments consisting of (a chain of) defeasible inferences based on evidence can then be used to support or attack the individual facts in these hypothetical stories. In this talk, I will discuss the *integrated argumentative-narrative approach* to reasoning in the process of proof, in which arguments and narratives can be used in conjunction as well as interchangeably. Thus, the decision about which facts to accept is based on the acceptability of the stories in light of the evidential arguments in the case.

The talk will first discuss the integrated approach from an informal perspective, briefly touching on the strengths and weaknesses of stories and arguments. After this a more formal approach is exposed, in which story-based reasoning with causal rules (i.e. *fire causes smoke*) is combined with argument-based reasoning with evidential rules (i.e. *smoke is evidence for fire*) using well-known defeasible logics. I will finish by showing how the work on the integrated theory can inform probabilistic Bayesian approaches to reasoning with evidence, as they have to deal with the same knowledge representation issues as the logical integrated theory.

Oral [5] Measurement Uncertainty and Inference in the Courtroom

Ted Vosk

Of Counsel, Cowan Kirk Gaston Wolff

Measurement is a quantitative activity performed to determine the quantity values attributable to a measurand. Regardless of how good a measurement is, though, it doesn't permit us to know a quantity's unique true value. Hence, there is a degree of doubt associated with all measurement results. A measurement's uncertainty characterizes our state of knowledge about the measurand's value by providing a range of values that can be reasonably attributed to a quantity with a specified level of confidence based upon the results obtained.

Forensic measurements are commonly relied upon as evidence in criminal prosecutions. The determination of an amount of drugs seized, the concentration of drugs or alcohol in blood or breath and the speed at which a motor vehicle is traveling are just a few. Although Courts are increasingly considering the uncertainty of qualitative results (DNA, fingerprint, etc.) in determining their admissibility, and sometimes even requiring that this uncertainty be reported to fact finders along with those results, the uncertainty of forensic measurements is rarely considered.

Like many lay people, courts routinely accept measured results as expressing a quantity's value with near absolute certitude. This conviction even leads some judges to preclude the parties themselves from presenting evidence of a measured result's uncertainty to triers of fact. On those occasions when measurement uncertainty is considered, it is typically so that *the court* can make a determination of whether the results are reliable

enough to be admitted. Seldom do courts require measured results presented as evidence be accompanied by their uncertainty. Rather, the traditional practice is for such results to be presented as “accurate and reliable” without any quantitative measure of the inferences they actually support. When presented in this manner, measured results cannot be rationally weighed and the inferences they support are speculative at best. Without more, the description of results as “accurate and reliable” is misleading, beckoning jurors to make inferences that are not supported by the science underlying a measurement. This undermines the search for truth and is directly responsible for convictions of the innocent and exonerations of the guilty. If systems of justice are to yield results consistent with the science relied upon as evidence, then fact finders must be supplied with the uncertainty associated with measured results. Only by so doing can the public have confidence in verdicts reached in reliance upon forensic measurements.

[1] International Organization for Standardization, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, ISO/TS 21748 (2004).

[2] Joint Committee for Guides in Metrology, Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement (GUM) (2008).

[3] Vosk T, Trial by Numbers: Uncertainty in the Quest for Truth and Justice, The NACDL Champion, Nov. 2010.

[4] Vosk T, Measurement Uncertainty, Encyclopedia of Forensic Sciences, p.322-331 (2nd ed. Elsevier – 2013).

[5] Vosk T, Emery A (2014) Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists, CRC/Taylor Francis Group, in press.

[6] Metrology: The New Honesty about the Uncertainty of Measurements in Scientific Analysis Research Paper 317, UC Davis Legal Studies Research Paper Series, 2012.

[7] Kacker et al. (2007) Evolution of modern approaches to express uncertainty in measurement, Metrologia 44, 513.

[8] Estler WT, Measurement as Inference: Fundamental Ideas, 48 Annal. CIRP 611 (1999).

Friday Morning

Plenary On the Statistical Analysis of DNA Mixtures

Steffen Lauritzen

University of Oxford

Commentary David Balding

University College London

Keynote [4] Recommendations from the NIST Working Group on Reporting Uncertainty in Forensic Conclusions

Cedric Neumann¹, David Kaye², Valerie Reyna³, Anjali Ranadive⁴

¹South Dakota State University, ²Pennsylvania State University,

³Cornell University, ⁴SciLawForensics Inc Brookings

Recent recommendations in Europe and the United States advocate a movement away from categorical opinions on the source of a particular piece of evidence. This movement proposes to frame the uncertainty associated with forensic conclusions into a logically coherent inference process supported by quantitative information [1-5]. The adoption of such a framework by the forensic science community is slowed in part by the difficulty of communicating logically sound and scientifically justifiable conclusions to police investigators, officers of the court, and jurors.

The communication of the results of forensic examinations needs to balance two critical elements: (a) forensic scientists need to represent evidence fairly and logically; (b) they need to present the information so that the audience can understand and use it appropriately.

Unfortunately, several jury studies have shown that research participants find categorical opinions more persuasive than conclusions containing some measurement of uncertainty; that different framings of given pieces of information influenced their behaviour; and finally, that they were not necessarily using the forensic information in a logical and coherent manner when compared to a normative framework (see [6-9] for some examples). Other studies have shown that individuals are not only affected by the framing of the information, but that the same

information, whether expressed quantitatively or qualitatively, has different meanings for different persons (see [10] for an introduction).

The U.S. National Institute of Standards and Technology is supporting a working group on the communication of uncertainty in conclusions resulting from forensic examinations. This working group has considered the nature of forensic inference and aspects of the psychology of effective communication of uncertainty and risk by individuals (see [10-11]). It plans to issue recommendations on how to best present conclusions from forensic examinations.

This talk will briefly contrast the context of presenting uncertainty in the forensic sciences with well-studied contexts such as the medical field, or policy/military-decision making. We will then review various aspects of the psychology of effective communication. Finally, we will present “best practices” recommendations of the working group for presenting conclusions from forensic examinations, as well as open questions for further research.

[1] Ass. Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion *Science and Justice* 49, 161–164.

[2] Aitken CGG, Roberts P, Jackson G, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings, Practitioner Guide No. 1, Statistics and the Law Working Group, Royal Statistical Society, 2011.*

[3] Saks MJ, Koehler JJ (2008) The Individualization Fallacy in Forensic Science Evidence, *Vand L Rev* 61, 199-219.

[4] Saks MJ (2003) The Legal and Scientific Evaluation of Forensic Science, *Seton Hall Law Review* 33, 1167-1187.

[5] NRC Committee on Identifying the Needs of the forensic Science Community (2009), *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, D.C.

[6] McQuiston-Surrett D, Saks M (2009) The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear, *Law Hum Behav.* 33, 436-453.

[7] Thompson W, Kaasa SO, Peterson T (2013) Do Jurors Give Appropriate Weight to Forensic Identification Evidence?, *Journal of Empirical Legal Studies*, Forthcoming; UC Irvine School of Law Research Paper No. 2013-124.

[8] Howes et al. (2013) Forensic scientists' conclusions: How readable are they for non-scientist report-users?, *Forensic Sci Int* 231, 102–112.

[9] Martire et al. (2013) The Expression and Interpretation of Uncertain Forensic Science Evidence, *Law and Human Behavior* 37, 197–207.

[10] Fischhoff B, Brewer NT, Downs JS, Eds (2011) *Communicating Risks and Benefits*, Food and Drug Administration (FDA), 242.

[11] Reyna VF, Brainerd CJ (2008) Numeracy, ratio bias and denominator neglect in judgments of risk and probability, *Learning and Individual Differences* 18, 89–107.

Oral [4] Discussion - A Forensic Practitioner's Perspective in Bridging the Gap Between the Forensic Statistician and the Courts

Tina Lovelock

Cellmark Forensic Services

In the UK the forensic scientist is in what could be considered a precarious position regarding written testimony for the courts. Nowadays, in order to expedite cases to the police and to some extent the CJS, short reports, abbreviated statements and streamlined forensic reports are permeating their way into the courtroom. They have their place but how should these be underpinned? The practitioner will deliver her viewpoint and experience as an expert.

Consequently, how should the final conclusion ‘be delivered’, in full or in abbreviated reports? In light of the practitioner’s experience for written and oral testimony she will discuss the use of the Bayesian scale of evidence in communicating the evaluation of the findings and the issues being encountered frequently speaking to judges, lawyers, magistrates, jury and the police. She aims to present the results of a small survey of experienced practitioners regarding the use of the scale in a range of evidence types. This may help address the questions of ‘how can the expert and statistician work with the Criminal Justice System to use jargon that is easily understood by the courts’, and ‘whether or not there should be greater level of standardisation of phraseology and language within the courtroom’ – which could lead to an accreditation of the expert’s ability, in evaluation/interpretation of evidence and formulating an opinion, by UKAS and the Forensic regulator.

Can the forensic practitioner bridge the gap between the statistician and the courts? This highlights a possible need for a forensic statistics practitioner (something akin to a nurse

practitioner between a nurse and doctor) bridging the void between the forensic statistician and the casework practitioner in the laboratory? Often the ideas presented by the statistician are academically sound but sometimes practically difficult to achieve in a busy laboratory and to convey in the courtroom. Some ideas of the pitfalls and experiences will be given to help bridge the gap – if indeed these are perceived to exist.

Oral [4] Constructing Arguments from Bayesian Networks about Forensic Evidence

Sjoerd Timmer¹, John-Jules Meyer¹, Henry Prakken^{1,2}, Silja Renooij¹, Bart Verheij^{2,3}

¹Utrecht University, ²University of Groningen, ³Stanford University

With the growing popularity of scientific evidence, such as DNA and fingerprint matching, the communication and interpretation of legal evidence has become increasingly challenging [1, 2, 3]. Recent miscarriages of justice, like the notorious cases of Sally Clark and Lucia de Berk, illustrate this problem. Bayesian networks (BNs) are models that allow for reasoning with uncertain information in a mathematically correct way [6]. Argumentation models [4, 5], however, are closer to the natural way of human reasoning. Forensic experts may be familiar with the type of reasoning involved in a BN, while lawyers and judges are used to argumentative styles of reasoning. To facilitate the communication between these two groups, we propose a method that automatically extracts rules, arguments and counter-arguments from BNs. Using the numerical information from the BN, we compute strengths for potential rules and extract those rules that are sufficiently strong. We use the ASPIC+ framework [5] for argumentation which can resolve conflicting arguments based on an ordering of the rules. An easy and straightforward rule ordering is provided by the strengths.

Our method of identifying rules inspects every two variables A and B connected by an arc in the Bayesian network’s directed graph and returns a rule from A=a to B=b if the ratio of the probability of B=b given A=a, and the prior probability of B=b is larger than one. This measure of strength equals the normalized likelihood [7], and corresponds with the intuition of a positive influence. With the extracted rules and their ordering, the ASPIC+ argumentation framework defines the applicable arguments and their attack relation. Since in BNs, inferences can be made both in the direction and against the direction of arcs, rules can also exist for both directions simultaneously. Within the ASPIC+ framework, however, we will only allow the construction of arguments that coincide with valid reasoning chains in the BN. As such, the method described serves for explaining the probabilistic results of a forensic BN in terms of natural arguments.

[1] Fenton et al. (2013) A general structure for legal arguments about evidence using Bayesian Networks, *Cognitive Science* 37, 61–102.

[2] Keppens J (2012) Argument diagram extraction from evidential Bayesian Networks, *Artificial Intelligence and Law* 20, 109–143.

[3] Lacave et al. (2007) Explanation of Bayesian Networks and influence diagrams in elvira. *Systems, Man, and Cybernetics, Part B*, 37, 952–965.

[4] Dung PM (2005) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77, 321–357.

[5] Modgil S, Prakken H (2013) A general account of argumentation with preferences. *Artificial Intelligence* 195, 361–397.

[6] Pearl J (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[7] Crupi et al. (2007) On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science* 74, 229–252.

[8] Pearl J (1988) Embracing causality in default reasoning. *Artificial Intelligence* 35, 259–271.

Oral [6a] Considering the Implementation of the Balding-Nichols Model for Database Applications

Rori Rohlfis¹, Vitor Aguiar², Kirk Lohmueller³, Amanda Castro⁴, Alessandro Ferreira⁴, Frederico Malta⁴, Iuri Louro², Rasmus Nielsen¹

¹University of California Berkeley, ²Universidade Federal do Espírito Santo Brazil, ³University of California Los Angeles, ⁴Laboratório Hermes Pardini Brazil

The Balding-Nichols model undergirds forensic genetic calculations, most notably, the estimated frequency of genotypes.

The Balding-Nichols model takes into account the increased probability that unrelated individuals share some alleles by distant shared ancestry using the parameter θ . The standard implementation has been used to appropriately estimate genotype frequency to test hypotheses of a particular individual leaving a particular genetic sample. The same implementation of the Balding-Nichols model has been applied to analyse allelic sharing observed in large forensic databases. Yet, it is not clear that this implementation applies directly to pairwise analyses. When examining a large dataset of over 100,000 individuals, we observe an excess of individuals sharing few alleles as compared to the best estimate under the Balding-Nichols model. However, this observation may be explained by further examining the specific implementation of the Balding-Nichols model in the context of complex human populations where pairs of individuals have highly varying times until their most recent common ancestor. These results may inform further analyses of genetic variation and allelic sharing observed in increasingly large forensic databases.

Keynote [6a] Analysis of Y-Chromosomal STR Population Data Using the Discrete Laplace Model

Mikkel Meyer Andersen¹, Poul Svante Eriksen¹, Niels Morling²
¹Aalborg University, ²University of Copenhagen

We present a new method based on the discrete Laplace probability distribution that approximates the distribution of short tandem repeat (STR) alleles when assuming a haploid Fisher-Wright model of evolution with a single-step mutation model. Creating a consistent statistical model of the haplotypes enables us to perform a wide range of analyses. Both simulated data and real Y-chromosomal STR haplotype databases were analysed using the discrete Laplace method. The analyses could be performed on a laptop computer.

The simulation study consisted of 9,000 data sets with 500, 1,000 or 5,000 Y-STR-profiles sampled from 60 different populations of size 20,000,000. The average deviation of the estimated probabilities of the Y-STR-profiles from the true population frequencies using the discrete Laplace method was smaller than those calculated with the naïve count estimate method (like $1/n$ or $1/(n+1)$ for data set size n) and Brenner's kappa method. When analysing real Y-STR databases, sound results were obtained, e.g. similar pairwise distances (between geographically separated sampling locations) to those obtained using the AMOVA method for a 7-loci European Y-STR database with approximately 12,700 males from 91 different locations and a 10-loci African Y-STR database with approximately 2,700 males from 26 different locations. Further analyses that are impossible with AMOVA were made using the discrete Laplace method: analysis of the homogeneity in two different ways and calculating marginal STR distributions.

- Andersen et al. (2013) The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies, *Journal of Theoretical Biology* 329, 39-51.
- Brenner C (2010) Fundamental problem of forensic mathematics - The evidential value of a rare haplotype, *Forensic Sci. Int. Genet.* 4, 281-291.
- Excoffier et al. (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes, *Genetics* 131, 479-491.
- Roewer et al. (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution, *Human Genetics* 116, 279-291.

Oral [6a] The Fundamental Problem of Forensic Statistics: Taking Account of Three Levels of Uncertainty

Richard Gill¹, Giulia Cereda²

¹Leiden University, ²University of Lausanne

When using Y-chromosome DNA profiles, it often happens that a DNA profile found on a crime scene and matching the suspect's profile does not appear in the relevant data-base. This creates a big challenge to the analyst who is required to supply a likelihood ratio (LR) or match-probability in order to quantify the evidential value of the match. Sensible estimation of the LR seems to rely on sensible estimation of the population frequency of this previously unseen haplotype.

There are three existing proposals of quite different nature: Roewer et al. (2000), based on Bayesian estimation of the

haplotype frequency with a Beta prior; Brenner (2010), based on the number of singletons observed in the database; and Andersen et al. (2013) using a mixture of independent discrete Laplace distributions as a parametric approximation of the distribution of allelic frequencies.

We add two new methods. One is similar to Brenner's, and like Brenner's is strongly related to the Good-Turing estimator. A second method is based on Anevski, Gill and Zohren's study of a non-parametric maximum-likelihood estimator. It is somehow intermediate between the parametric approach of Andersen and non-parametric methods based on Good-Turing estimators. We believe that it avoids the disadvantages of those while moreover providing a supplementary means of evaluating their accuracy. For all methods it is imperative to assess two more levels of uncertainty, beyond the uncertainty about which hypothesis is true given the evidence which would hold if we knew everything about the population probability distribution. LR is a ratio of probabilities which are usually based on a model, and that model is at best only a good approximation to the truth. Moreover, we only estimate parameters of that model by fitting it to the data in our database. We will see that sometimes "less can be more". It can pay to ignore some of the information at our disposal, since trying to incorporate it, in other words trying to reduce some of the uncertainties by using more information, only magnifies the impact of the other uncertainties. For statisticians, this is the familiar pay-off between bias and variance; which we also meet when discussing over-fitting versus under-fitting; the pay-off between good prediction of new observations versus good description of observations done earlier.

- Roewer et al. (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males, *Forensic Sci Int* 114, 31- 43.
- Brenner CH (2010) Fundamental problem of forensic mathematics — The evidential value of a rare haplotype, *Forensic Sci Int Genet* 4, 281-291.
- Andersen MM, Eriksen PS, Morling N (2013) The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies, *J Theor Biol* 329, 39-51.
- Anevski D, Gill RD, Zohren S (2013) Estimating a probability mass function with unknown labels, submitted to *Ann. Statist.*

Oral [6a] The Effect of R-Allele and Wild Card Designations in Forensic DNA Database Searches

Torben Tvedebrink

Aalborg University, University of Copenhagen

Forensic DNA databases are powerful tools used for the identification of persons of interest in criminal investigations. Typically they consist of two parts; a database containing profiles of known individuals and a database of profiles associated with a crime. The risk of adventitious or chance matches between crimes and innocent people increases as the number of profiles within a database grows, and more data is shared between different jurisdictions.

The DNA profiles obtained from crime scenes are often partial because crime samples may be compromised in quantity or quality. Where an individual's profile cannot be resolved from a mixture ambiguity is introduced. A wild card, F, may be used in place of an allele that has dropped out or where an ambiguous profile is resolved from a mixture.

Variant alleles that do not correspond to a marker in the allelic ladder or that appear above or below the extent of the allelic ladder range are assigned the allele designation R, for rare allele. R alleles are position specific with respect to the surviving/unambiguous allele. The F and R designations are made where the exact genotype has not been determined. The R and F designation are treated as wild cards for searching. They can match any other allele without penalty. Such profiles generally have a higher chance of an adventitious match. In this presentation we investigate the probability of an adventitious match given wild cards under different minimum load criteria and matching strategies, and generate a logical construct in which to assess the value of the profiles typed with older multiplexes. Furthermore, we discuss what effect this has on the utility of the forensic DNA databases.

- Butler JM (2012) *Advanced topics in forensic DNA typing: methodology*, Elsevier.

- Walsh SJ, Moss DS, Kleim C, Vintiner GM (2002) The collation of forensic DNA case data into a multi-dimensional intelligence database, *Science & Justice* 42, 205–214.
- Walsh SJ, Moss DS (2001) Forensic DNA databasing - solving crime in New Zealand, *Australasian Science*, 34–36.
- Walsh SJ, Roux C, Ross A, Ribaux O, Buckleton JS, Examining the role and impact of forensic DNA profiling on key areas of the criminal justice system, Presented at the 16th ANZFSS International Symposium for the Forensic Sciences.
- Weir BS (2004) Matching and partially-matching DNA profiles, *J Forensic Sci* 49, 1–6.
- Weir BS (2007) The rarity of DNA profiles, *Ann Appl Stat* 1, 358–370.
- Tvedebrink et al. (2012) Analysis of matches and partial-matches in a Danish STR data set, *Forensic Sci Int Genet* 6, 387–392.
- Tvedebrink T, Curran J, DNAtools: Statistical functions for analysing forensic DNA databases (R-package).

Oral [6a] Using Match Proportions to Characterize the Effects of Population Structure on the Strengths of DNA Evidence

Bruce Weir

University of Washington

One of the early advances in interpreting DNA evidence was the introduction of match probabilities that included a population structure

parameter [1]. This has become known as the “theta correction” following the NRC report in the US, although that report did not clarify the distinction between profile and match probabilities.

Estimation of the parameter theta, when suitable data are available, has followed standard population genetic approaches, such as that in [2]. Prompted by a recent paper [3] I have reformulated the estimates as un-weighted averages of the proportions of matching profiles, within and between population pairs. The emphasis on matching has a natural forensic interpretation, and this new approach clarifies the use of whole-population matching for estimating sub-population matching. The predicted match probability M_W , for single-allele profiles, within a subpopulation is $\theta_W + (1-\theta_W)M_T$ where M_T is the match probability in the whole population and θ_W is a within sub-population quantity. To estimate M_W with a sample value of M_T it is necessary to replace θ_W by $\beta_W = (\theta_W - \theta_B)/(1-\theta_B)$ where θ_B is a between-subpopulation-pair quantity. There is a natural extension to autosomal genotypes and lineage marker haplotypes.

Whereas in the past I have used mean squares of allele frequencies to estimate quantities like β_W I now suggest phrasing them in terms of sample heterozygosities, which themselves can be expressed as functions of sample matching proportions. The numerical differences in the two approaches are generally small, but using sample matching proportions is appealing from a forensic perspective. This approach also offers a direct demonstration of the effects of the number of loci on matching probabilities. If there are a large number of sub-populations, an approximate estimate of β_W is $(m_W - m_B)/(1-m_B)$ where m_B and m_W are the averages of the sample matching proportions within and between subpopulations.

I present more detailed estimating equations and numerical results from forensic and other databases of STR and SNP data.

[1] Balding DJ, Nichols RA (1997) *Heredity* 78, 583-589.

[2] Weir BS, Cockerham CC (1984) *Evolution* 38, 1358-1370.

[3] Bhatia et al. (2013) *Genome Research* 23, 1514-1521.

Keynote [10] How to Make Rational Decisions in the Analysis of Low-Template DNA Traces

Simone Gittelson, Carolyn Hill, Michael Coble

National Institute of Standards and Technology

The analysis of forensic science casework evidence involves making a series of decisions. Given the serious consequences that these decisions can lead to in the administration of justice, they should be based on coherent foundations and consider the particular circumstances and objectives of the case at hand. A decision-theoretic approach to these decision problems provides a transparent, normative framework for accomplishing this. In this presentation, we illustrate such a decision-theoretic approach to two decision points in the analysis of low-template DNA traces: (i) choosing how many replicate PCR amplifications to perform, and (ii) choosing how to designate the genotype of the trace's donor. We model the first of these as a two-stage decision problem solved by maximizing the expected net gain, and the second as a

one-stage decision problem solved by selecting the Bayes action [1]. Sensitivity analyses show how the most rational choice for each depends on the scientist's probability distributions and loss function for the particular case at hand. Hence the major advantage of this approach (with regard to fixed laboratory standards applied to all cases alike) is that it leads to case-specific choices. On the one hand, the scientist's probability distributions reflect the particular case circumstances, such as the genetic traits of the population of potential donors, as well as the locus-specific characteristics, such as the possibility of an allelic peak missing in the profile, known as allele drop-out. Alternatively, the loss function reflects the scientist's objectives and preferences in the particular case with regard to the trade-off between narrowing down the population of potential donors and the possibility of falsely excluding a potential donor. This framework therefore provides a tool for the forensic scientist to coherently think about these decision problems, allowing him or her to make better informed choices. From a broader perspective on forensic science, the decision problems treated here are not restricted to the domain of low-template DNA traces. They represent the two fundamental types of decision problems: one-stage decision problems and two-stage decision problems. Hence this theoretical framework also lends itself to applications addressing decisions in other domains of forensic science.

[1] Gittelson S, Biedermann A, Bozza S, Taroni F (2014) Decision analysis for the genotype designation in low-template-DNA profiles, *Forensic Sci Int Genet* 9, 118-133.

Oral [10] Beyond the LR Approach: The Development of a Decision Support Tool for Real Time Identification

Peter Vergeer, Reinoud Stoel

Netherlands Forensic Institute

In the context of terrorist attacks or organized crime, police forces and secret services keep suspects under surveillance. When a threat becomes acute, a fast confirmation of the identity of a suspect needs to be performed, before intervention takes place. Consequences of wrong interventions can be severe, see e.g. the Stockwell case. On the other hand, refraining to intervene may result in a serious crime or terrorist attack being committed. Apart from a need for fast and reliable means to confirm the identity of a suspect, the decision to intervene requires an analyses of the consequences of the combined decision and identity space. We have developed a decision support tool based on a Bayesian network combined with an expected utility analyses. Our decision support tool can support decisions made in the practice of real time identification.

The tool combines several evidence modalities including finger print, speech and facial photo's, with information about the relevance of reference and trace material, and a means to elicit utilities, prior odds, means of intervention, and hypotheses. A graphical user interface was developed that combines this information and represents the advised means of action in a user friendly way.

In the current presentation we will discuss a number of lessons learned while modelling this decision process. First, the strength of using a Bayesian network in a case depends also on its simplicity. We have modelled parameters deemed relevant by police forces involved in identity confirmation. Second, a large portion of the uncertainty in the decision comes from elicitation of the consequences, which has received little attention in the forensic science literature. Third, the value of LR's relatively close to 1 is generally underestimated. Elicited prior odds and decision thresholds often result in situations where LR's in the order of 10 to 100 change the decision advice.

• Lindley DV (1990) *Making decisions*, 2nd ed., Wiley, Chichester.

• O'Hagan et al. (2006) *Uncertain Judgements, Eliciting experts' probabilities*, 1st ed., Wiley.

Oral [10] Game Theoretical Structures and Forensic Decisions

Rainhard Bengesz

Technische Universität München, National University Taipei

Game theory is a collection of mathematical concepts to investigate the nature of interdependent actions based on certain assumptions of strategies, preferences (orders), and methods of evaluation and anticipation. Classical and Bayesian, evolutionary, combinatorial and algorithmic game theory may serve as examples of some conceptual frameworks under the umbrella of game theory. Besides the structure of specific games and its relation to socio-technical dynamics the most crucial part is to identify methods of evaluation of certain outcomes of strategies because the important concept of steady state (a classical concept of the 17th to 19th century), e.g. the well-known Nash equilibrium, is crucially based on it.

In forensic science & digital forensic science or jurisprudence it is often necessary to make decisions under uncertainty and lack of data (or information). A common way to handle this asymmetric situation is to model the relevant issues in terms of costs and benefits, i.e. in concepts of (linear) optimization.

Our aim is to investigate in the game theoretical structures in forensic decisions, i.e. in the connections between forensic decisions under uncertainty and its game theoretical counterparts. This connection will be exemplified, firstly, by so called P2P networks, and, secondly, by evolutionary games. The first concept connects game theory and forensic decisions via Bayesian networks which are used extensively in forensic science. The second concept connects both theories via so called extended replicator equations. In these replicator equations we will express the costs & benefits for forensic strategies, i.e. decisions under uncertainty.

What we finally gain is a way to connect forensics with game theory in a very natural way. This will be helpful in using the entire methodologies and results game theory provides. From a scientific communication viewpoint it is easier to communicate problems and concepts in terms of game theory and its elaborated schemes.

- Bengesz R (2014) Note on a Second Order Game in Legal Practice, in: Armgardt M, Rachman S (eds), Jurisprudence and Logic, Springer.

Oral [6b] Performance Study of a Score-based Likelihood Ratio System for Forensic Fingerprint Comparison

Jeannette Leegwater, Didier Meuwly, Marjan Sjerps, Peter Vergeer, Ivo Alberink

Netherlands Forensic Institute

Traditionally, fingerprint evidence is evaluated by a fingerprint examiner leading to a categorical conclusion of identification or exclusion. These conclusions, however, refer to a decision. Decisions are based, amongst others, on posterior probabilities, i.e. the probabilities of the hypotheses themselves, which also depend on factors other than the forensic findings. The only statement that may allow for any uncertainty in the discrimination between hypotheses is the inconclusive statement, but even if this statement expresses the fact that there is uncertainty, it does not give a degree of certainty.

Following the seminal approach of Lindley [1], an approach to address these issues is to report evidence according to the Bayesian framework, i.e. in terms of likelihood ratios (LRs). This contribution describes a score-based likelihood ratio system for fingerprint to fingerprint comparisons containing 6 up to 11 minutiae. The LR system is based on an AFIS (Automated Fingerprint Identification System) comparison algorithm and is designed in accordance with the rules of probability; the same evidence is evaluated in the numerator and denominator of the LR. Furthermore, the hypotheses are on the level of the suspect, not the finger. The resulting LRs are presented with 95%-bootstrap intervals, indicating the precision of the results. Currently there is a fundamental discussion in the scientific community about the best way of dealing with uncertainty when reporting LRs. In our contribution we will briefly explain what the discussion is about.

The databases used consist of 58 fingerprints containing at least 12 minutiae with corresponding fingerprints from forensic casework, and 246,745 ten-print cards from an anonymized copy of the Dutch fingerprint database. Subsamples of the marks were

taken to obtain marks with 6-11 minutiae. Parameterization of the scores resulted in a same and different source score distribution, used to obtain the score-to-LR transformation function. The performance of the system is tested at several levels. Leave-one-out cross-validation was applied to calculate LRs given a same and a different source scenario. Rates of misleading evidence were studied at two levels. At a logarithmic level, Empirical Cross-Entropy [2] indicates that the system is well calibrated. At the level of the LRs, the expected value of the LR given a different source and that of 1/LR given a same source comparison were close to 1, as they should be. A simulation study showed a coverage of 90-98% of the bootstrap intervals.

"Essentially, all models are wrong, but some are useful" is a famous quote from Box [3]. This performance study sheds light on the question: "How useful?"

[1] Lindley DV (1977) A Problem in Forensic Science, *Biometrika* 64, 207-213.

[2] Ramos D (2007) Forensic Evaluation of the Evidence Using Automatic Speaker Recognition Systems, PhD dissertation, Universidad Autonoma de Madrid.

[3] Box GEP, Draper NR (1987) *Empirical Model Building and Response Surfaces*, Wiley, New York.

Poster Abstracts

Poster [1] Innovative Analytical Strategies in Forensic Toxicology – Combining Chemometrics with the Bayesian Approach

Eugenio Alladio^{1,2}, Valentina Pirro¹, Enrico Gerace², Alberto Salomone², Marco Vincenti^{1,2}

¹Università degli Studi di Torino, ²Centro Regionale Antidoping "A. Bertinaria"

Recent innovations in the fields of liquid/gas chromatography and mass spectrometry have turned around the traditional strategy of chemical analysis of biological samples. Nowadays, multi-residue high performance liquid chromatography and gas chromatography combined with tandem mass spectrometry (HPLC-MS/MS and GC-MS/MS) methods have increasingly been developed to screen large sets of analytes, that may be simultaneously contained in single biological sample and detected within a single run for each sample [1].

Multivariate Data Analysis is gradually playing a key role in laboratory practice. Since the developments of mass spectrometry and chromatography allow operators to deal with large amounts of data, containing several information to be extracted from, multivariate approaches provide great advantages in several fields, i.e. forensic toxicology, clinical chemistry, etc. Particularly, the combination of Bayesian inference approach – peculiarly exploited in forensic sciences – with Multivariate Data Analysis techniques proved to increase the sensitivity of the decision power. When the biomarkers present smaller intra-individual than inter-individual variability, longitudinal trends become very useful. Aim of our study was to evaluate the possibility of increasing the sensitivity of several biomarkers evaluation by adding Multivariate Data Analysis information in well-defined Bayesian frameworks. Therefore, as Bayesian longitudinal approach looks really encouraging, with self-adapting limits in function of prior test results performed on the individual [2], we moved from population based- to individual-based cut-off values using Bayesian approach combined with several chemometric techniques, e.g. Partial Least Squares–Discriminant Analysis (PLS-DA), Unequal Dispersed Classes (UNEQ), Soft Independent Models of Class Analogy (SIMCA), etc. Specifically, our goal was to develop HPLC-MS/MS and GC-MS methods, combined with Bayesian inference and Multivariate Data Analysis, in order to detect alternative markers of endogenous anabolic androgenic steroids (EAAS) misuse in urine samples, thus developing comprehensive analytical strategies to easily recognize effective anti-doping rule violations [3]. In fact, since the distinction between the endogenous or exogenous origin from the substance is still challenging for anti-doping laboratories, endogenous steroids increased their popularity as anabolic and androgenic doping agents, and thereby became the

most misused substances in elite sports. The developed analytical methods were capable to detect 18 EAAS, 7 glucuronide- and 5 sulphate-derivates and their relative ratios. The advantages of Multivariate Data Analysis techniques ease the decision process of recognizing anti-doping rule violations. Therefore, the combined use of Multivariate Data Analysis and Bayesian adaptive model provide significant approaches in doping control, particularly in the complex field of EAAS misuse.

[1] Vincenti M, Cavanna D, Gerace E, Pirro V, Petrarulo M, Di Corcia D, Salomone A (2013) *Anal. Bioanal. Chem.* 405, 863–887.

[2] Berry DA (2006) *Nature Reviews Drug Discovery* 15, 27–36.

[3] Schulze JJ, Lundmark J, Garle M, Ekström L, Sottas PE, Rane A (2009) *Steroids* 74, 365–368.

Poster [1] Bayesian Contribution for the Analysis of Y-STRs in Paternity Testing and Pedigree Analysis

Cláudia Vieira da Silva^{1,4}, António Amorim^{1,2,4}, Heloísa Afonso Costa^{1,4}, Rosa Espinheira^{1,4}, Teresa Ribeiro^{1,4}, Jorge Costa Santos^{1,2}, Antónia Turkmann³

¹Instituto Nacional de Medicina Legal e Ciências Forenses,

²Universidade de Lisboa, ³Departamento de Estatística e Investigação Operacional, ⁴Centro de Ciências Forenses (CENCIFOR)

The analysis of Y-chromosomal STRs (Y-STR) has become a widely accepted tool for criminal casework, paternity testing and human identification. The establishment of a worldwide Y-STR database and the supply of commercial Y-STR typing kits contributed for its use in most forensic genetic laboratories. These markers contribute to identify the male lineage, valuable information to solve sexual assault cases and also complex kinship tests, as well as anthropological and population studies. Knowledge about mutation rates and the mutational process of Y-STR or microsatellite loci used in paternity testing is crucial for the correct interpretation of resulting genetic profiles. It seems to be convenient that mutation counts and father/son pair counts are used for simple frequency estimation of mutations rates.

In order to estimate Y-STR germline mutation rates in south Portugal population a 17 Y-STR kit, was applied in 197 father/son pairs samples selected from our laboratory, but a limited number of mutation events were detected, suggesting an alternative way of estimating mutation rates.

The locus specific mutation rates (μ) were estimated with Bayesian formulation, considering that X , the random variable corresponding to the number of mutations per locus in n meioses is well estimated by a binomial distribution. Also, according to Bayesian formulation the probability of a mutation event, μ , is also a variable with a *prior* Beta (α , β) distribution. This problem was approached with a beta-binomial model. Expected values from Bayesian estimation of locus-specific mutation rates ranged from 0.0004123 for DYS392 to 0.006444 for DYS458. Our findings were important to conclude future applications of Y-STRs in paternity testing and pedigree analysis.

- Kayser M, Sajantila A (2001) Mutations at Y-STR loci: implications for paternity testing and forensic analysis, *Forensic Sci Int* 118, 116–121.

Poster [1] Likelihood Ratio Model as a Tool for Forensic Expert in Olive Oil Geographical Origin Determination

Patryk Własiuk¹, Agnieszka Martyna¹, Grzegorz Zadora^{2,3}

¹Jagiellonian University in Krakow, ²Institute of Forensic Research, ³University of Silesia

Food fraud or food adulteration may be of forensic interest for instance in a case of a suspected deliberate mislabelling. On account of its potential health benefits and nutritional qualities, an olive oil geographical origin determination might be of special interest. With the use of chemometrics in olive oil chemical compositional data analysis an assessment of food traceability can be made [1]. However, such methods do not take into account possible sources of uncertainty or the information about the rarity in relevant population, which are of particular interest to forensic expert. These criteria are met by likelihood ratio (LR) model, which has been used in this study to examine an issue of olive oil classification. The aim of this work was to inspect different LR models and their pertinence under selected data

pre-processing methods (logarithm based data transformations) and feature selection technique. This was carried out on a data describing 572 Italian olive oil samples characterised by the content of 8 fatty acids in lipid fraction of olive oils [2]. Six classification problems related to three regions of Italy (South, North and Sardinia) have been considered with the use of LR models. The correct classification rate and empirical cross entropy were taken into account as a measure of performance of each model. The application of LR models in the olive oil geographical origin determination has proven to be satisfactorily useful for the considered issues analysed in terms of many variants of data pre-processing.

[1] Ben-Ayed R, Kamoun-Grati N, Rebai A (2013) An Overview of the Authentication of Olive Tree and Oil, *Comprehensive Reviews in Food Science and Food Safety* 12, 218–227.

[2] Forina M, Armano C, Lanteri S, Tiscornia E (1983) Classification of olive oils from their fatty acid composition, [in:] H. Martens, H. Russwurm Jr., eds, *Food Research and Data Analysis*, Applied Science Publishers, London, 189–214.

Poster [2] Improved Bounds for the Probability of Causation

Philip Dawid¹, Rossella Murtas², Monica Musio²

¹University of Cambridge, ²University of Cagliari

Given even the best evidence for the dependence of a disease Y on an exposure X , we can typically only provide bounds for the “probability of causation” in the case of an individual I who has developed the disease after being exposed. However these bounds can be improved or adapted if further information becomes available. We consider such improved bounds in a variety of such cases, including:

- A covariate that can be observed in experimental data and in I [1]
- A covariate that can be observed in observational data and in I [2]
- A covariate that can be observed only in experimental data [3]
- A mediator that can be observed in experimental data and in I

We also pay attention to the additional assumptions that need to be made

in order to justify the validity of these bounds.

[1] Dawid AP, Musio M, Fienberg SE (2014) From statistical evidence to evidence of causality, arXiv:1311.7513.

[2] Tian J, Pearl J (2000) Probabilities of causation: Bounds and identification, *Annals of Mathematics and Artificial Intelligence* 28, 287–313.

[3] Dawid AP (2011) The role of scientific and statistical evidence in assessing causality. In *Perspectives on Causation* (R. Goldberg, Ed.). Oxford: Hart Publishing, 133–147.

Poster [9] Indirect Evaluation by Simulation of a Bayesian Network

Anna Emanuelson, Susanne Mannerskog, Ulf Svensson, Kajsa Ekberg, Lennart Jonasson, Märtha Lövbj, Anders Nordgaard

Swedish National Laboratory of Forensic Science (SKL)

Evidence evaluation when addressing source level propositions is usually done by comparing a piece of recovered material to (specimens of) control material. When the control material source is not available for taking specimens or for investigating it in its entirety, we must stick to photographs or video take-ups for making comparisons. An example is the comparison of class characteristics between a recovered footwear print and a picture of a seized shoe, where the evaluation is occasionally made that way. However, this way of pursuing the investigation is due to needs of quick answers, when there is no or little time to send in the entire footwear for the comparison. Moreover, the pictures taken of the sole of the seized footwear are taken by the police under controlled conditions and with high quality equipment. When the suspected source is captured on a lower quality video take-up and the recovered material consists of fragments from the original body of material – for instance fire debris – the comparison with the control material source is naturally more difficult. In this paper we present a case where the question is whether recovered fire debris originate from a piece of garment captured on a CCTV take-up. We show how a likelihood ratio for the two propositions can be indirectly obtained from a classification of the source of the fire debris, by using a Bayesian network model. Results from fire debris analysis as well as the

results from image comparisons can be evaluated against propositions of class and the updating of the class node for fire debris propagates back to the propositions for source. Feeding the network with uniform priors for the class nodes we show how simulation can be used to obtain the correct level of the likelihood ratio for further reporting.

Poster [9] Social and Cognitive Factors Affecting Asymmetric Social Hypothesis Testing

Roberta Capellini, Simona Sacchi, Patrice Rusconi, Paolo Cherubini

University of Milano-Bicocca

The search and evaluation of pieces of evidence for testing alternative hypotheses is a cognitive activity of critical importance in many human domains, including criminal investigations and judgments. The subset of gathered evidences is determined by the information search strategies that are used by the decision maker. However, some highly spontaneous and frequent information search strategies can thwart the hypothesis testing process, causing (over)confirmation biases. The yes-no format of the question is not always a guarantee of effective discrimination between alternative hypotheses. For instance, a dichotomous question can be strongly asymmetric, when affirmative replies are more diagnostic than negative ones (i.e., asymmetric confirming questions, $LR(D_h) \gg LR(D-h)$). Although the modern forensic methodology are particularly careful about procedural accuracy and leading questions, the use of asymmetric-confirming questions could elude the cognitive control and consequently lead to an unconscious confirmation of the focal hypothesis. The present contribution explores cognitive and motivational factors affecting the use of asymmetric questions. Study 1 ($N = 253$) aimed at exploring if people may have cognitive control over the use of asymmetric strategy. To this goal, specific instructions in order to avoid confirmation tendencies were experimentally manipulated and their efficacy was tested. Study 2 ($N = 112$) focused on the effect of power on the use of asymmetric strategies. Participants will be placed in more or less powerful position than the social target and their information search process was analyzed. Overall, results showed that, under certain conditions, people may have cognitive control on questions asymmetry and that powerful people are more prone to use asymmetric questions when searching information than weaker persons. Implications for forensic psychology are discussed.

Poster [10] On the Level of Detail of the Alternative Hypothesis

Jan Peter van Zandwijk

Netherlands Forensic Institute

In forensic research, the evidential value of a trace E in the light of two hypotheses can be expressed by means of the likelihood ratio $LR = P(E|H_1)/P(E|H_2)$. Here, $P(E|H_i)$ is the probability of seeing the evidence E under hypothesis H_i . Usually, the only requirement made for the two hypotheses H_1 and H_2 is, that they are mutually exclusive, e.g. they cannot both be true at the same time (see e.g. [1]).

In case of individualization research, the question may be asked whether some reference object is the source of a trace. In such cases the hypothesis H_1 is obvious. The alternative hypothesis H_2 would be that the trace is produced by a somewhat arbitrary 'other source'.

An object, e.g. a screwdriver x , can be viewed as being a member of a hierarchy of nested sets of increasing detail. At each level of nesting, a hypothesis H_1 concerning the object can be formulated based on the set it belongs to at that level. At the same time, there would be a natural candidate for the alternative hypothesis H_2 in the form of one of the other sets present at the same nesting level. In this picture, the level of detail specified by the hypothesis H_1 automatically defines the level of detail of the alternative hypothesis H_2 .

Within such a framework, valid pairs of hypotheses would be: (1) the mark is produced by a screwdriver vs. the mark is produced

by another tool; (2) the mark produced by this particular screwdriver vs. the mark is produced by another screwdriver of the same brand. However, an invalid pair would be: The mark is produced by this particular screwdriver vs. the mark is produced by another tool.

Conversely, such a hierarchy of nested sets can also be used to select the level of detail of H_1 worthwhile considering given the amount of information present in the evidence E . At a certain level of detail, the evidence E can be equally likely under H_1 as under the alternative H_2 , yielding a LR equal to one. In those cases, one can consider going up in the hierarchy and choose hypothesis H_1 from a more general set and compare it to an alternative H_2 at the same level of nesting.

A verbal statement of the LR in this case would be: the marks are equally likely under the hypothesis that they are produced by this particular screwdriver as under the hypothesis that they are produced by another screwdriver of the same brand. But the marks are more likely under the hypothesis that they are produced by a screwdriver than under the hypothesis that they are produced by another tool.

[1] Robertson B, Vignaux GA (1995) *Interpreting scientific evidence: Evaluating Forensic Science in the Courtroom*, John Wiley and Sons, Chichester.

Poster [6a] Interpretation of DNA Evidence: Implications of Thresholds Used in the Forensic Laboratory

Ricky Ansell^{1,2}, Ronny Hedell^{1,3}, Anders Nordgaard^{1,2}

¹Swedish National Laboratory of Forensic Science (SKL),

²Linköping University, ³University of Gothenburg

Evaluation of forensic evidence is a process lined with decisions and balancing, not infrequently with a substantial deal of subjectivity. Already at the crime scene a lot of decisions have to be made about search strategies, the amount of evidence and traces recovered, later prioritised and sent further to the forensic laboratory etc. Within the laboratory there must be several criteria (often in terms of numbers) on how much and what parts of the material should be analysed. In addition there is often a restricted timeframe for delivery of a statement to the commissioner, which in reality might influence on the work done.

The path of DNA evidence from the recovery of a trace at the crime scene to the interpretation and evaluation made in court involves several decisions based on cut-offs of different kinds. These include quality assurance thresholds like limits of detection and quantitation, but also less strictly defined thresholds like upper limits on prevalence of alleles not observed in DNA databases.

In a verbal scale of conclusions there are lower limits on likelihood ratios for DNA evidence above which the evidence can be said to strongly support, very strongly support, etc. a proposition about the source of the evidence. Such thresholds may be arbitrarily chosen or based on logical reasoning with probabilities. However, likelihood ratios for DNA evidence depend strongly on the population of potential donors, and this may not be understood among the end-users of such a verbal scale. Even apparently strong DNA evidence against a suspect may be reported on each side of a threshold in the scale depending on whether a close relative is part of the donor population or not.

In this presentation we review the use of thresholds and cut-offs in DNA analysis and interpretation and investigate the sensitivity of the final evaluation to how such rules are defined. In particular we show what are the effects of cut-offs when multiple propositions about alternative sources of a trace cannot be avoided, e.g. when there are close relatives to the suspect with high propensities to have left the trace. Moreover, we discuss the possibility of including costs (in terms of time or money) for a decision-theoretic approach in which expected values of information could be analysed.

Poster [6a] Assessment of Homozygous Allele Calls from Low Level DNA Samples

Vanessa Duchamp^{1,2}, Ingo Bastisch², Peter Schneider¹

¹University of Cologne, ²Federal Criminal Police Office (BKA)

Using optimized assays for amplification of DNA samples using the polymerase chain reaction, the methods of forensic DNA analysis have become increasingly sensitive and are able to detect very low DNA quantities even in routine casework. DNA analysis is based on the examination of distinct polymorphic regions of different chromosomes, called Short Tandem Repeat Systems (STRs). As the human genome contains two copies of each chromosome, these STRs could have either the same or a different number of tetrameric repeat motifs. Therefore, in the first case, only one peak will appear in the electropherogram; these are homozygous alleles. In the second case, two peaks of different fragment lengths will be present on the electropherogram; these are heterozygous alleles.

Some known effects are associated with these sensitive analyses. The first one is the peak height imbalance; when the DNA quantity is high, the two heterozygous alleles of a given STR should have approximately the same peak height. That is not the case anymore when the DNA quantity decreases so that stochastic effects may influence the amplification efficiency during PCR. The second effect is the occurrence of allele drop-out; i.e. one allele is not amplified due to a very low copy number of the target sequence. It is an extreme case of imbalance and the result could be wrongly interpreted as a homozygous allele. The third effect possibly encountered is an allele drop-in; it is a spurious allele which is typically not related to the crime case, most likely originating from a low-level contamination.

Although these effects are well characterized for heterozygous alleles [1], we think that this information is still incomplete for apparently homozygous alleles. Some approaches are proposed in the literature to handle these alleles [2,3,4] one of these is to consider homozygous peak height as twice a heterozygous peak height [3] and to calculate the homozygous drop-out rate based on heterozygous drop-out rate squared [4].

The aim of this project is to study, in a pragmatic way, the statistical relationship between homozygous and heterozygous alleles, based on experimental data. We hope to be able to propose a simple mathematical model to explain the behavior of peak height and drop-out of homozygous alleles. This experiment is carried out initially with mock casework samples. These are obtained from buccal cell swabs from different unrelated persons. One part of these samples are generated from serial DNA dilutions and the other part from degraded DNA. For degradation, the DNA is exposed to irradiation with UV light. Mixtures are not considered at this point in our experiment. Data resulting from DNA analysis, principally information about the allele peak height and the absence of allele, are statistically analysed using the open-source software R.

[1] Gill P et al. (2012) DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods, *Forensic Science International: Genetics* 6 (2012) 679-688.

[2] Balding D, Buckleton J (2009) Interpreting low template DNA profiles, *Forensic Sci Int Genet* 4, 1-10.

[3] Benschop C et al. (2011) Low template STR typing: Effect of replicate number and consensus method on genotyping reliability and DNA database search results, *Forensic Sci Int Genet* 5, 316-328.

[4] Haned H et al. (2012) Exploratory data analysis for the interpretation of low template DNA mixtures, *Forensic Sci Int Genet* 6, 762-774.

Poster [6a] Toward Practical Implementation of an Empirical Drop-out Model for Casework: Comparison of Benchmark vs. Estimated Probabilities of Allelic Drop-out

Keith Inman^{2,3}, Kirk Lohmueller¹, Norah Rudin³

¹University of California Los Angeles, ²California State University East Bay, ³Forensic DNA Consulting

Complex DNA profiles (cDp), characterized by some combination of low template, mixtures, degraded, or inhibited DNA, continue to present interpretational challenges to the forensic community. Whether the LT contribution comprises the main profile, or whether it is present as the minor component of a mixture, ambiguity arises from the possibility that alleles present in the biological sample may not be detected in the resulting DNA profile. This phenomenon is known as allelic drop-out. This ambiguity complicates both the assessment of the potential

number of contributors and estimation of the weight of the DNA evidence for or against specific propositions. One solution to estimating the weight of the evidence is to use a likelihood ratio (LR) that incorporates the probability of allelic drop-out ($P(D_o)$) estimated for the specific evidence sample under consideration. However, although a vast repository of data exists, few empirical studies to determine allelic drop-out probabilities have been performed to date. Here we characterized patterns of allelic drop-out in single-source samples using both universal and run-specific analytical thresholds. Not surprisingly, we found fewer instances of apparent drop-out when using a lower (run-specific) detection threshold. Also, unsurprisingly, a positive correlation exists between allele drop-out and allele length, even in good quality samples. We used logistic regression to model the fraction of alleles that dropped out of a profile as a function of the average height of the detected peaks. The equation derived from the logistic regression model allowed us to estimate the expected drop-out probability for an evidentiary sample based on the average peak height of the profile. We show that the LRs calculated using the estimated drop-out probabilities were similar to those calculated using the benchmark drop-out probabilities, suggesting that the estimates of the drop-out probability are accurate and useful. This trend holds even when using the data from the PowerPlex[®] 16 typing system to estimate the drop-out probability for an Identifiler[®] profile, and vice versa. Thus we demonstrate that use of a LR that incorporates empirically estimated allelic drop-out probabilities provides a reliable means for extracting additional information from LT forensic DNA profiles.

Poster [6a] Statistical Evaluation of Composite DNA Profiles Using a Discrete Model

Roberto Puch-Solis¹, Miguel Juárez²
LGC Forensics¹, University of Sheffield²

The technology for producing DNA profiles has greatly improved in recent years and will continue to evolve. These developments make DNA profiling systems very sensitive and pose a number of interpretation challenges. Phenomena such as drop-in, allelic dropout and degradation in stain profiles make the evaluation more complex. In addition, for low template DNA (LTDNA), two or three stain profile replicates are produced and need to be evaluated simultaneously whilst taking into account the phenomena described above.

In recent years a number of discrete and continuous statistical models have been proposed. Discrete models use absence and presence of artefactual, spurious and allelic peaks in the stain profiles, whilst continuous models also use peak height of area. Some of the models assume that replicates are independent given putative donors at a locus. In this poster we assess this assumption in the context of discrete models, recognising that the same analysis is key for continuous models and will be addressed in further work. We specifically use the discrete model LiRa (Puch-Solis & Clayton, 2014).

In Bright et al. (2012), the composite method is defined as the production of a composite profile from the replicates (the set union of the alleles in the replicates), followed by the calculation of a likelihood ratio (LR) using the Binary model. In this poster we use a discrete model, LiRa, which is capable of accommodating the phenomena described above in the calculation of the LR from the composite.

Profile replicates with the characteristics described above will be produced by computer simulation. LRs will be calculated by entering the replicates separately and by entering the composite profile and adjusting drop-in probabilities accordingly. LRs will be compared and ascertain the effect of the assumption. We envisage that this work will contribute to the body of knowledge underpinning the use of discrete models in casework.

- Puch-Solis R, Clayton R (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. LGC Research Report LGC/P/2013/165. Submitted to *Forensic Sci Int Genet*.
- Bright J, Gill P, Buckleton J (2012) Composite profiles in DNA analysis. *Forensic Sci Int Genet* 6, 317-321.

Poster [6b] Classification of Volatiles in Fire Debris

Anna Emanuelson, Klas Brorsson-Läthen, Anders Nordgaard
Swedish National Laboratory of Forensic Science (SKL)

Much of the discussion in the area of evaluation of forensic evidence according to the Bayesian approach concerns comparisons at source level and activity level. However, besides such comparisons there is in daily case work also a need to report conclusions of classification by the use of a likelihood ratio (or a general Bayes factor).

The aim of this presentation is to illustrate how the Bayesian approach can be applied in a common arson case. The issue is to classify whether a piece of burnt wood contains ignitable liquid or not. When evaluating the results, the properties of the matrix must be taken into account. The piece of wood may contain volatiles although the volatiles do not originate from ignitable liquid. The propositions are as follows:

Prosecutor's proposition: The item contains ignitable liquid.

Defence proposition: The volatiles come from the matrix.

From case work there is knowledge about ignitable liquids used in arsons and from matrix surveys there is knowledge about volatiles in fire debris. In the case described, a high amount of aromatics was found in the item. Aromatics can be found in ignitable liquid, preferably in gasoline, but also in diesel oil and kerosene. However, in diesel oil and kerosene the amount is normally lower than what was found in this case. Aromatics are normally not found in wood. However, they can be found in some lacquers, impregnating agents, oils and glues, which in turn can be found on wood, but normally not in such high amounts as in the case described.

The probabilities of obtaining the results if the prosecutor's proposition and the defence proposition respectively are true are calculated as weighted sums of probabilities assigned under subsets of each of the propositions. The ratio between these weighted sums is then a Bayes factor that is presented according to the scale of conclusion adopted by The Swedish National Laboratory of Forensic Science (SKL) [1]:

The results support that the item contains ignitable liquid rather than that the volatiles come from the matrix (Level +2).

The effect of changing the prosecutor's proposition according to the results is further discussed. A change of the prosecutor's proposition to "The item contains gasoline" will result in a higher value of the Bayes factor.

[1] Nordgaard A, Ansell R, Drotz W, Jaeger L (2012) Scale of conclusions for the value of evidence, *Law, Probability and Risk* 11, 1-24.

Poster [6b] Quantifying Fingerprint Evidence Using Bayesian Alignment

Peter Forbes

University of Oxford

Fingerprint evidence has long been considered infallible by courtrooms worldwide. However, subjective human judgement plays a large role in determining whether or not two fingerprints match, especially when dealing with the blurry prints typical of crime scenes. Despite this uncertainty, courtroom fingerprint evidence is always presented categorically as a match or non-match. This leads to inflated confidence in the forensic evidence. There has been a recent push within the forensics community to present courtroom evidence as a probability of matching rather than a categorical match. Before this is possible, a standardized method for quantifying the strength of fingerprint evidence needs to be developed. We are developing one such model using the mathematical theory of spatial point processes. The minutiae of an observed fingerprint or fingermark are viewed as distorted copies of some true, unobserved fingerprint. Given an observed fingerprint and fingermark, we compute the likelihood ratio of the probability that they originated from the same finger over the probability that they originated from independent fingers. The model has been tested on a small database of 258 forensic fingerprints provided by the American Forensic Bureau of Investigation. Joint work with Steffen Lauritzen and Jesper Møller.

- Garris M, McCabe R (2000) NIST special database 27: Fingerprint minutiae from latent and matching tenprint images. Technical report, NIST, Gaithersburg, MD, USA.
- Green PJ, Mardia KV (2006) Bayesian alignment using hierarchical models, with applications in protein bioinformatics, *Biometrika* 93, 235-254.

- Neumann C, Evett IW, Skerrett JE (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, *J. R. Stat. Soc A* 175, 371-415.

Poster [6b] Developing Appropriate Score-Based LR's: The Example of Fingerprints

Douglas Armstrong, Cedric Neumann, Christopher Saunders, Danica Ommen, Austin O'Brien
South Dakota State University

The evaluation of the weight of complex evidential forms, such as traditional impression and pattern evidence (fingerprint, shoeprint, handwriting), and data resulting from the chemical analysis of trace material, need to involve sophisticated quantification procedures due to the usually extremely large number of dimensions of the samples. In these situations traditional statistical methods for evidence interpretation tend to lead to ambiguous conclusions within the context of the formal Bayesian paradigm.

Very few attempts have been made to quantify the weight of forensic evidence characterised by multidimensional variables [1,2] with even fewer considering truly high dimensional data. Most attempts to develop formal evidence interpretation strategies for complex evidence are based on using an ad hoc proximity measure between pairs of (trace, control or reference) objects. This facilitates a reduction of the dimensionality of the problem, allowing for approximating likelihood ratios (LRs) based on tractable distributions. Unfortunately, the use of proximity measures in LR's is not straightforward and, to the best of our knowledge, there are no score-based LR methods that have achieved the statistical rigor normally associated with formal Bayesian evidence interpretation. Most published methods happen to contain logical errors, or result in score based LR's not quantifying the weight of given evidence as the authors intended. The complexity of the problem is significantly magnified when multiple samples are used to represent the trace, control and reference objects since variance/co-variance effects between samples from same/different sources need to be taken into account.

In this presentation, we will review some of the on-going research programs relating to these topics. Focusing first on various attempts to use proximity measures in score-based LR's [citations omitted], we will examine the various logical fallacies and shortcomings that are inherent to these approaches. Secondly, our presentation will concentrate on novel developments that build on the score-based LR published in [3] and address two of the weaknesses highlighted by discussants: (1) the use of a *weighting function*; (2) the fact that the proposed model was not formally a LR.

[1] Aitken CGG, Zadora G, Lucy D (2007) A Two-Level Model for Evidence Evaluation. *J Forensic Sci* 52, 412-419.

[2] Neumann C, Champod C, Yoo M, Genessay T, Langenburg G (2013) Improving the Understanding and the Reliability of the Concept of Sufficiency in Friction Ridge Examination, *NJ Research Report 2010-DN-BX-K267*.

[3] Neumann C, Evett IW, Skerrett JE (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm, *J. R. Stat. Soc A* 175, 371-415.

Poster [6b] Does a Change of SEM-EDX Equipment Require the Creation of New Glass Database?

Grzegorz Zadora^{1,2}, Agnieszka Martyna³, Aleksandra Michalska¹

¹*Institute of Forensic Research Krakow*, ²*University of Silesia*,

³*Jagiellonian University Krakow*

It is well known that database creation, e.g. database of elemental composition of glass fragments determined by application of Scanning Electron Microscopy coupled with an Energy Dispersive X-ray spectrometer (SEM-EDX) is a time and money consuming process [1]. Nevertheless, it happens from time to time that a particular model of analytical equipment is replaced by a new one. Then, the following question arises - does a change of SEM-EDX equipment in laboratory require a creation of a new glass database? It is especially important when information from database is used for evaluation of evidential values of determined physicochemical data when so-called comparison and

classification problems are solved, e.g. for forensic purposes, by application of likelihood ratio approach [2].

The aim of the research was to check whether it is necessary or not to create a new database when an old SEM-EDX equipment (Jeol JSM-5800) was replaced by the new one (JSM-6610LV). In the aim to solve this problem elemental composition for 100 glass objects (50 float glass and 50 containers) was determined. Next, it was checked what is a difference between obtained LR values [2,3] when comparison or classification of samples from the new-SEM database is performed on a basis of the old-SEM database or on a basis of the new-SEM database.

Within a comparison problem, levels of false answers were low and similar when samples from the new-SEM database were compared on the basis of data collected in the new-SEM or the old-SEM database (ca. 1.5% false positive and 15% false negative). Also, levels of correct classifications were very high and similar (ca. 95%) when samples from the new-SEM database were classified on the basis of data collected in the new-SEM or the old-SEM database. Moreover, differences between LR values obtained for the same classified or compared samples on the basis of information gathered in an old-SEM database and new-SEM database were calculated. In theory, such LR values obtained for respective classifications or comparisons should be the same, i.e. the difference between LR values should be zero. Such results were obtained in most of the cases.

It all suggests that the database created by application of SEM-EDX equipment, which is replaced by the new one, could be still used for evaluation of evidential value of glass fragments and extended by data obtained by application of the new equipment.

[1] Zadora G, Neocleous T (2010) Evidential Value of Physicochemical Data – Comparison of Methods of Glass Database Creation, *Journal of Chemometrics* 21, 174-186.

[2] Zadora G, Martyna A, Ramos D, Aitken C (2014) *Statistical Analysis in Forensic Science, Evidential value of multivariate physicochemical data*, Wiley.

[3] Aitken C, Zadora G, Lucy D (2007) A Two-Level Model for Evidence Evaluation. *J Forensic Sci* 52, 412-419.

Poster [6c] Evidence Evaluation for Forensic Voice Comparison

Tereza Neocleous¹, Colin Aitken², Paul Foulkes³, Peter French³, Erica Gold³, Vincent Hughes³

¹University of Glasgow, ²University of Edinburgh, ³University of York

In forensic voice comparison, experts are presented with recordings of a criminal (e.g. threatening phone calls) and recordings of a suspect (e.g. police interview). They are asked to assess the possibility that the recordings contain the voice of the same person. A 'paradigm shift' is underway in forensic science, with moves to express expert conclusions in the form of likelihood ratios (LRs). The LR is increasingly accepted by forensic scientists as the logically and legally correct framework for assessing forensic evidence, and is used for many evidence types, including DNA and chemometrics.

Where several features are analysed (as with speech: vowels, consonants, fundamental frequency(f_0), etc) LR results for each individual feature must be combined into an overall LR. Speech is a particularly complex biometric, as vocal features are highly variable within any individual (they are affected by illness, style, emotions, etc), and they affect each other in diverse ways (e.g. fundamental frequency (f_0) may affect vowel acoustics). Current applications of LRs to voice evidence, however, generally fail to account for the complexity and inter-relatedness of features.

Voice evidence in the form of an LR therefore tends either to focus on a small subset of continuous acoustic features (potentially overlooking other discriminatory features), or to ignore the inter-relatedness of the features and thus present a potentially misleading overall LR. See [1] for an example with discrete data. Current developments arising from a collaboration between statisticians in the Universities of Glasgow and Edinburgh and forensic phoneticians in the University of York will be described.

[1] Aitken CGG, Gold E (2013) Evidence evaluation for discrete data, *Forensic Sci Int* 230, 147-155.

Poster [8] Critical Questions for Fact-Finding: Heuristics and Logical Models

Floris Bex

Utrecht University, University of Groningen

The reasoning performed by forensic scientists and practitioners takes place in the larger context of the process of proof, in which activity-level hypotheses are constructed, modified and dismissed on the basis of the evidence. In order to do justice to all the relevant reasoning mechanisms in the process of proof, both arguments and narratives (or stories) are needed. Stories – coherent sequences of events – can be used to organise the complex mass of facts in a case into one or more hypotheses at the activity level, overviews of "what happened" in a case.

Arguments – single conclusions drawn from, e.g., forensic tests or witness testimonies – are then used to support or attack these hypothetical stories.

Arguments will often be based on specific background knowledge about types of evidence and forensic tests. For example, we know that analyses of a DNA sample have a certain probative value, or that when a witness says something we can, at least in the first instance, believe the witness. Given this knowledge, we can define argumentation schemes, generic types of inference that underlie arguments, as well as critical questions, typical sources of doubt for each of the inference types.

Stories are also based on background knowledge about different types of crime, what we call story schemes. For example, a story scheme for "murder" specifies the typical elements of a murder scenario (e.g. the motive, time of the killing, place of the killing, weapon). Like argumentation schemes, story schemes also have associated critical questions that point to possible weaknesses in a story such as missing elements or weak causal connections between events.

I will present three types of questions for critically analysing a complex criminal case in a dialectical way: (i) critical questions with which individual (forensic) inferences based on evidence can be analysed; (ii) critical questions with which the activity-level stories about the facts can be analysed; and (iii) critical questions that can be used to analyse the case as-a-whole, that is, the combination of stories and arguments. I will show how not only how these critical questions provide informal heuristics for the process of proof, but also how they can be encapsulated in a formal logical framework that provides a mathematical approach to the process of proof.

- Bex FJ, Verheij B (2012) Solving a Murder Case by Asking Critical Questions: An Approach to Fact-Finding in Terms of Argumentation and Story Schemes, *Argumentation* 26, 325-353.
- Bex FJ, van Koppen PJ, Prakken H, Verheij B (2010) A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence, *Artificial Intelligence and Law* 18, 123-152.

Poster [11] Lessons Learned from an Overturned Exam Fraud Case

Dennis Maynes

Caveon Test Security

A difficulty with enforcement actions (such as revocation of scores or dismissal from a position) after exam fraud has been detected is the lack of direct evidence that fraud actually occurred. There are no broken windows, missing valuables, or dead bodies.

However, there are tantalizing clues. A school district received a phone call from a concerned guardian that a student did not have mastery of basic skills, even though the test results indicated otherwise. As a result, the school district monitored the test administration at the student's school, which resulted in a precipitous pass rate drop for reading and math in elementary school grades 3, 4, and 5. Visual examination of the answer sheet documents from the prior year revealed a large number of wrong-to-right answer changes or erasures. Statistical analysis confirmed that the presence of erasures were significantly higher in the prior year than in the monitored year. The school district decided to dismiss the school testing coordinator. The case was heard three times. In the first two hearings, each hearing examiner found the school district was justified in dismissing the school testing coordinator. But, on appeal the hearing examiner

overturned the previous rulings. This session will discuss how the erasure data was collected, analyzed, and presented in a compelling manner that confirmed unauthorized modification of the answer sheet documents. The school district used different expert witnesses in each hearing, each of whom presented their own interpretation of the data, which added complexity to the case. Even though the statistical evidence that test fraud occurred was compelling, the third hearing examiner overturned the previous rulings. The session will focus on the lessons learned in presenting data, in presenting evidence that supports the enforcement action taken by the school district, and in arguments for and against the statistical evidence and its relevance to the case.

Poster [11] A Time Series Analysis of Crime Rates

Wafaa Abdullah Alkhareef

Public Authority for Applied Education and Training (PAAET)

Kuwait

This study aims to illustrate predictability performance among different competitive models of time series to forecast the crime rates in Kuwait.

Exponential smoothing is the widely used class of procedures for smoothing discrete time series in order to forecast the immediate future. This popularity is attributed to its simplicity and computational efficiency; ease of adjusting responsiveness to changes in the process being forecast; and its reasonable accuracy. The idea of exponential smoothing is to smooth the original series the way the moving average does and to use the smoothed series in forecasting future values of the variable of interest.

Exponential smoothing is a simple and pragmatic approach to forecasting, whereby the forecast is constructed from an exponentially weighted average of past observations. The largest weight is given to the present observation, less weight to the immediately preceding observation, even less weight to the observation before that, and so on exponential decay of influence of past data.

The current study was designed to compare the performance of different exponential smoothing methods in forecasting the crime rates and also to examine the trend movement of it. Performance of the models is assessed with the performance criteria of mean absolute percentage error (MAPE), root mean square errors (RMSE) and mean absolute error (MAE). Results from the study revealed that the Winter's method is the most appropriate method to forecast exchange rates for the given time interval in Kuwait.

Index

Ahlinder	17	Leegwater	24
Aitken	3, 7, 29	Lennon	9
Alberink	11	Lindström	6
Alladio	24	Lopatka	3, 18
Armstrong	28	Lovelock	21
Andersen	22	Lucy	14
Ansell	26	Martyna	6, 25
Balding	20	Maynes	8, 29
Bengez	23	Mazumder	14
Berger	3	Mclvor	3
Betts	12	Miao	12
Bex	20, 29	Morrison	11
Biedermann	4	Neocleous	10
Bijleveld	12	Neumann	20
Blankers	15	Nordgaard	13, 25
Bozza	10	Ostrum	5
Brenner	9	Perlin	5
Bright	9	Puch-Solis	16, 27
Brorsson-Läthen	28	Roberts	4
Brummer	15	Rohlf's	21
Burdorf	4	Rudin	10
Capellini	26	Saunders	11
Cereda	10	Sigman	13
Cowell	6	Sironi	10
Curran	9	Sjåstad	17
Dawid	3, 25	Slooten	16
Duchamp	26	Smit	14
Evelt	17	Steele	16
Fenton	19	Stevens	19
Forbes	28	Timmer	21
Freeman	4	Tvedebrink	22
Gastwirth	12	Vergeer	23
Gill	22	Verheij	19
Gittelsohn	23	Vieira da Silva	25
Graversen	6	Vlek	7
Hedell	18	Vosk	20
Hola	12	Wang	13
Inman	5, 27	Weir	23
Juchli	7	Wieten	8
Kaye	4	Wilson	17
Kesteloo	19	Zadora	18, 28
Klaver	16	Zandwijk	26
Kloosterman	9	Zeegers	3
Kokshoorn	14	Zoete	3, 8
Lauritzen	20		

Silver sponsor

Co van Ledden Hulsebosch Centrum



Qiagen



Exhibitor

STRMIX, Institute of Environmental Science and Research



Supporter

Netherlands Forensic Institute



Netherlands Forensic Institute
Ministry of Security and Justice

HUGIN Expert

HUGINEXPERT

The leading decision support tool



Stochastics - Theoretical and Applied Research (NWO-STAR)



Leiden University



**Universiteit
Leiden**

University of Amsterdam



UNIVERSITY OF AMSTERDAM

Stichting Leerstoel Criminalistiek



STICHTING
LEERSTOEL
Criminalistiek

Royal Netherlands Academy of Arts and Sciences (KNAW)



KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN