Statistics and the Fair Administration of Justice

Alicia Carriquiry and Heike Hofmann
Iowa State University and CSAFE, Ames

csafe
Center for Statistics and Applications in Forensic Evidence

Outline

- Justice as a human right.
- Current practice
  - DNA evidence
  - Everything else
- A defensible paradigm based on Bayes' theorem
- Unique challenges associated with pattern evidence.
- Final comments.

Justice has failed an unknown number of individuals

- In the last two decades, the cases of hundreds of persons who spent years in jail for crimes they did not commit have come to light.
- The Innocence Project estimates that lack of scientific validity of many forensic tools, and exaggerated claims by forensic examiners is the second cause for this miscarriage of justice. (First cause is mistaken eye witness identifications.)
- **Lack of scientific validity**: Many forensic tools have no scientific underpinning or have not been evaluated in controlled, well-designed experiments.
- **Insufficient data**: Datasets for evaluation have been small, non-representative, or both.
- **Over-statement of significance of a "match"**: Forensic experts have routinely over-stated the significance of their findings, using terms such as *scientific reliability*, *practical certainty*, etc., that convey to juries a degree of precision that is not justified by the available science.

Basic ideas

- *Lockhart Principle*: every contact leaves a trace.
- When a crime is committed, evidence of many types left at the crime scene can help determine whether the suspect was the source of the evidence.
- Advances in the physical and biological sciences have encouraged the development of many new forensic tools, including:
  - Touch DNA analysis
  - 3D imaging of topographic surfaces of fired bullets
  - Mass spectrometry of glass fragments.
- Even though metrology has advanced tremendously, the correct interpretation of forensic evidence is lagging far behind.

Is there science in forensic *science*?

- With the exception of DNA, a long list of National Academy of Sciences reports have found that science is missing in forensics.
- A 2009 report entitled *Strengthening Forensic Sciences in the United States: A Path Forward* found that most forensic tools lack scientific foundation, are plagued with subjectivity and are subject to unknown error rates.
- Junk science is routinely admitted in Court as “expert witnesses” are unchallenged by lawyers and even judges who do not have the background to decide what is good science.
- Much of this unsupported “science” is offered by the forensic practitioners themselves who often view their role as one of support for the prosecution.
- Defendants without the means to hire their own experts are at a clear disadvantage in the US criminal justice system.

No access to information

- Even in the best case, experts for the defense do not have access to much of the data that the State may have used to build their case.
- In the United States, forensic scientists and prosecutors can access databases to which the defense (and consequently, the person accused of a crime) has no access.
- These databases and the search and matching algorithms that come with them are used to establish, for example, that a bullet was fired by a specific gun, or that the latent print at a crime scene matches the suspect's prints.

Access to information (cont'd)

- In the current standard of practice, defendants are denied the means to challenge:
    - The completeness and representativeness of databases used by the prosecution
    - The algorithms that are used to search those databases.
    - The criteria that are implemented to establish a match.
    - The probative value of the match.
- These issues have begun to be addressed, but progress is slow and there is a lot of work to be done.

Current practice in forensics

- In a typical Court testimony, a competent forensic examiner will:
  - Describe the analytical techniques used to process the evidence.
  - Provide an assessment of the error rate associated with the instrument or test.
  - Interpret the results of the analyses for the jury.
- The interpretation step roughly consists in declaring that the crime scene and suspect's samples **match** or **do not match**, or that the tests were **inconclusive**.
- There is a push in the forensics community to talk about *degrees* of the strength of the match.

Current practice (cont'd)

- Most lay jurors equate a match with *same source*.
- **Except in the case of single donor DNA among non-relatives, a match does not imply same source.**
- This is a critically important concept, and one that has escaped the attention of the forensic and legal professions until recently.
- The concept of a **coincidental match** in evidence other than DNA has emerged as important only in the last 20 years or so.
- For the vast majority of evidence types, we do not know the probability of a coincidental match.

Probability of a coincidental match

- Loosely, defined as the probability of observing that two samples are indistinguishable even though they have a different source.
- Consider blood types (A,B,O) in the US. Blood found at the crime scene is type A+ and so is the suspect. We consider two scenarios:
  - The suspect left the sample at the crime scene and if so, we would expect a match.
  - The suspect was not the donor. What is the probability some some other random person might have left an A+ sample at the crime scene?
- The latter is the probability of a coincidental match and for this example, we know that this probability is about 0.35.
- The 0.35 value is derived from the known frequency of A+ persons in the US, which is $\sim 35.7\%$.

The "weight of evidence" paradigm

- To move away from the binary "match/non-match" framework, statisticians have proposed a one-step approach, that consists in evaluating the evidence under two competing hypotheses:

$$\begin{aligned} H_p &: \quad \text{The suspect is the source of the evidence} \\ H_d &: \quad \text{Someone else is the source of the evidence} \end{aligned}$$

- A likelihood ratio statistic

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

can be computed to decide whether the evidence E supports the prosecutor's or the defense's hypothesis.

Computing a LR

- In the absence of contamination or lab error, the numerator in the LR will be close to 1.
- Computation of the numerator only requires comparison of known and questioned samples.
- The denominator is much more challenging:
  - We need to define H_d.
  - For each potential H_d we need a statistical model to compute $P(E|H_d)$.
  - We have to have reference databases that are relevant for each plausible H_d.
- At this time, we can compute $P(E|H_d)$ only for single-donor DNA samples or for simple DNA mixtures.

The challenging area of pattern evidence

- Pattern evidence includes fingerprints, ballistics and other toolmarks, handwriting, shoe prints....
- The evidence typically consists of a 2D or a 3D image of the sample.

- No statistical models, no (or questionable) reference databases.

The case of bullet striae

- When bullets travel down a barrel after they are fired, the “rifling” of the barrel and manufacturing imperfections leave marks or striation on the bullet surface.

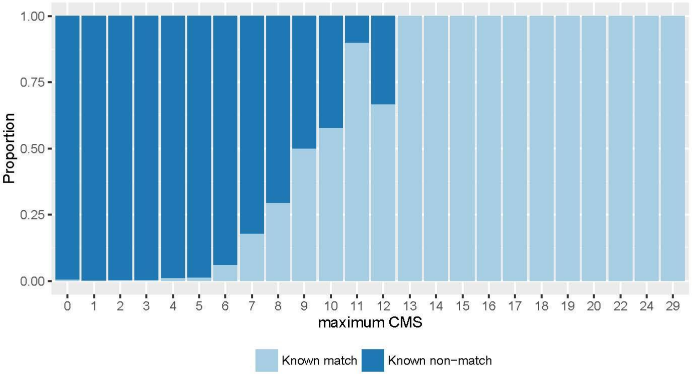
The art of ballistics

- By comparing striation in two bullets, firearms examiners attempt to determine whether a specific gun fired both bullets (identification) or at least whether the two bullets were fired by the same gun (same source).
- Current practice: "I know a match when I see one".
- The AFTE Theory of Identification:

...opinions of common origin to be made when the unique surface contours of two toolmarks are in *sufficient agreement*....Agreement is significant when it exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool................

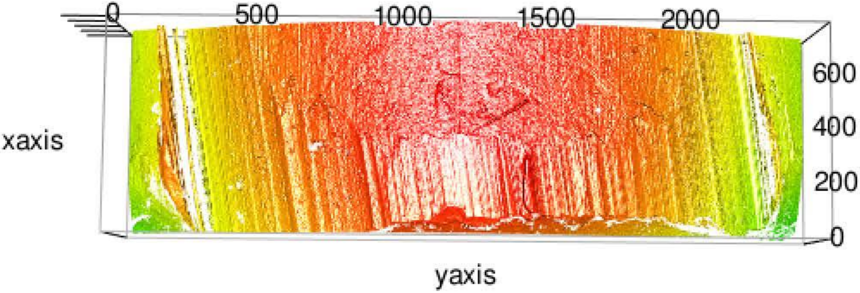
An attempt at objectivity - CMS

- CMS: consecutively matching striae.
- Idea: many consecutively matching striae are indication of a "match"; a match suggests that the two bullets were fired by the same gun.
- But what is the threshold?

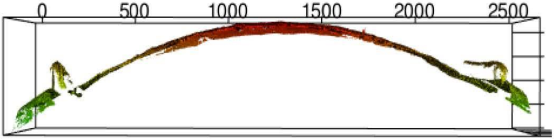


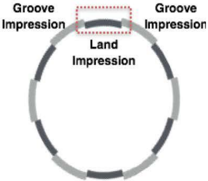
A more objective match criterion

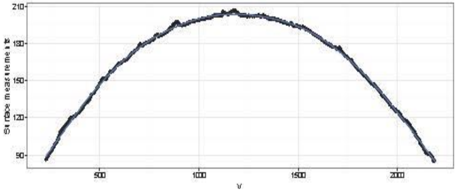
- Cutting edge: confocal 3D microscopy to capture surface topography of bullets.

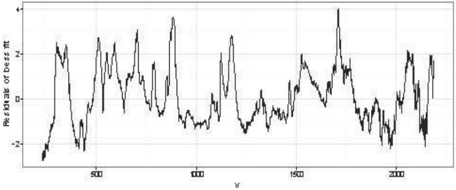


Extracting a signature



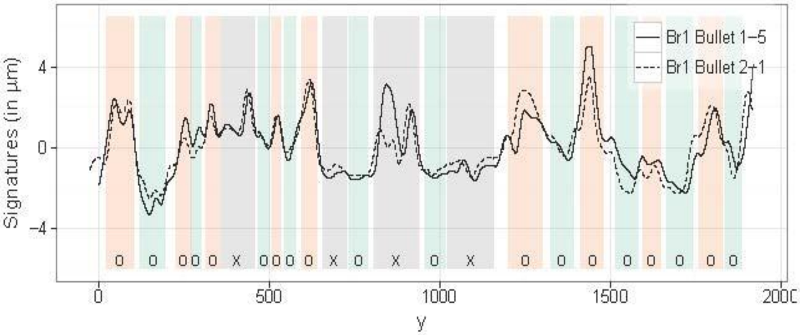






Comparing signatures

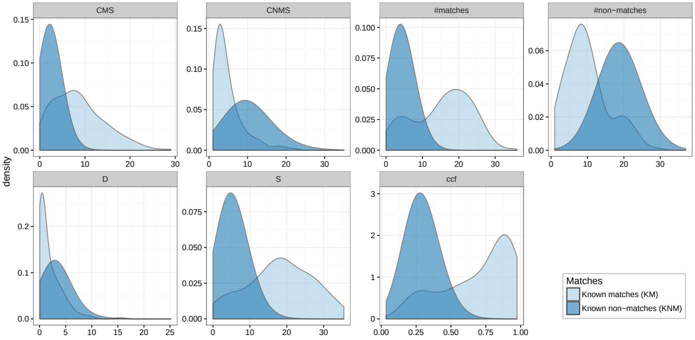
- Given two signatures, we can construct a “score” for the difference using machine learning methods.



Properties of a match criterion

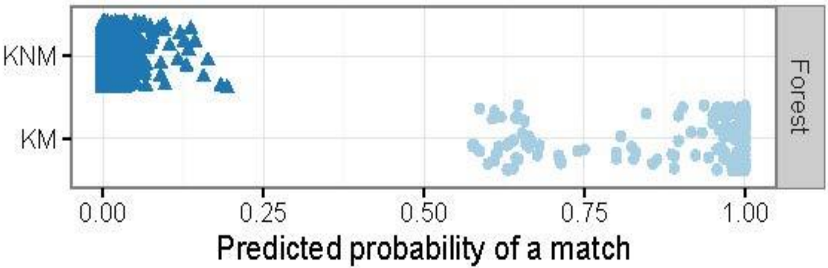
- Ideally:
  - High sensitivity or low probability of false positives
  - High specificity or low probability of false negatives.
- Need an experimental dataset with known matching pairs and known non-matching pairs to determine behavior of score.

Extracted features



Performance of score

- Features combined into score using random forests.
- Applied method to small (but challenging) dataset:



- Tested in several other datasets: no classification errors.
- **We still know nothing about the probative value of a match!**

Score-based LRs

- In principle, we could:
  - Fit densities to the empirical distributions of scores for matching pairs and for non-matching pairs.
  - Construct a likelihood ratio using the fitted densities $L(f(y)|H_p, \theta_p)$ and $L(f(y)|H_d, \theta_d)$, for y the vector of features, $f(.)$ the function that maps features into scores, and θ_p, θ_d the vectors indexing the likelihoods under the two competing hypothesis.
- **Problem:**

$$LR = \frac{L(y)|H_p)}{L(y|H_d)} \neq LR_f = \frac{L(f(y)|H_p)}{L(f(y)|H_d)},$$

  except for trivial $f(.)$. In fact, LR and LR_f are typically not even proportional to each other.
- Still and open problem.

The work ahead of us

- The Obama administration was aware that the scientific and statistical bases of most forensic disciplines must be shored up.
- The new administration has radically different priorities, so we do not know whether federal support will be continued.
- Statisticians can play a critical role – many interesting problems with tremendous social impact.
- Until now, research funding for forensic statistics was on the rise at various federal agencies: DoJ, NSF, DARPA, DoD, even NIH.

csafe

Center for Statistics and Applications in Forensic Evidence

CSAFE is a consortium of 4 universities: ISU, CMU, UCI, UVA. Please contact me if you are interested in this area

alicia@iastate.edu