Communicating Forensic Findings: Framing the Issues

NIST June 25-26, 2024

Hal Stern
Department of Statistics
University of California, Irvine
sternh@uci.edu

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

csafe
Center for Statistics and Applications in Forensic Evidence

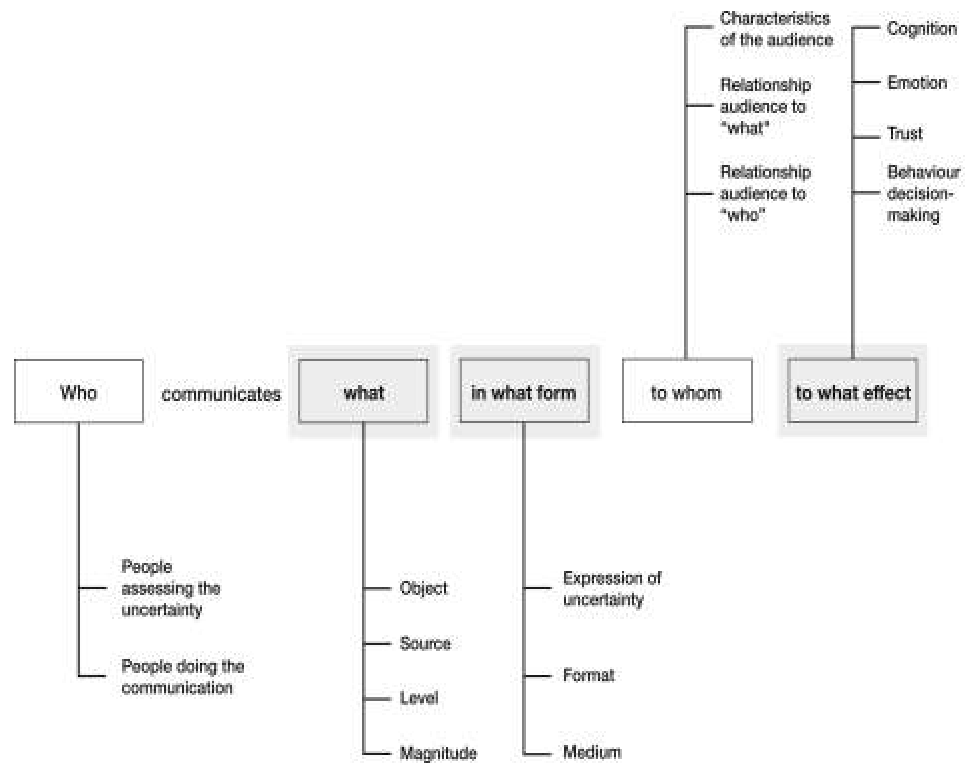
Communicating Uncertainty

- August 2014 NIST Federal Funding Opportunity :

  "A critical need in the forensic science research community is a **more thorough understanding and contextualizing of the uncertainty associated with scientific measurements** and/or analytical techniques. **Reporting uncertainty in forensic science measurements is currently an uncommon practice**, largely because the forensic science community demands an unequivocal conclusion of a binary analysis .."

Communicating Uncertainty

- van der Bles et al., 2019, Royal Society Open Science "Communicating Uncertainty About Facts, Numbers and Science"
- Framework:
  - Who is communicating?
  - **What are they communicating?**
  - **In what form is the uncertainty communicated?**
  - Communicated to whom?
  - Communicated to what effect?



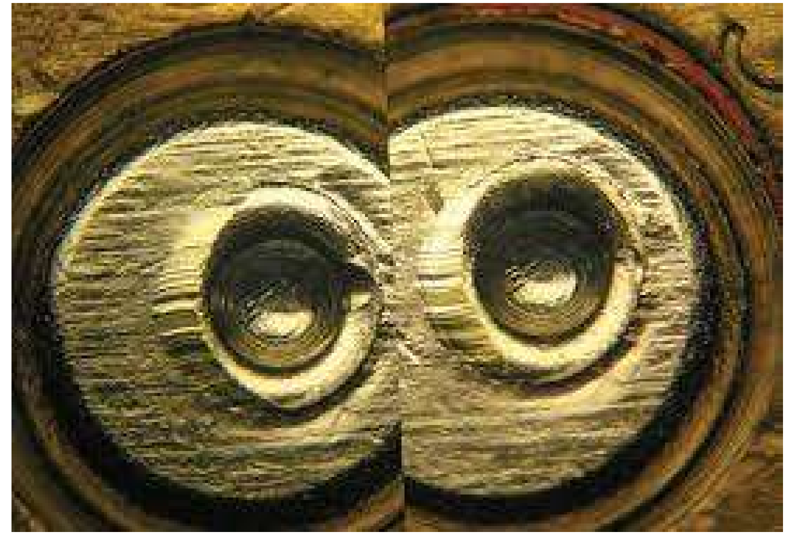
Communicating Uncertainty

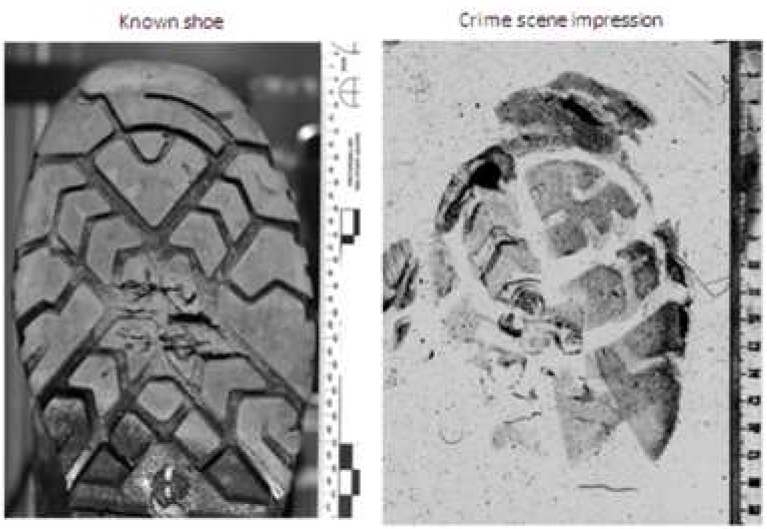
- Who is communicating?
  - Forensic examiner
  - Attorney (prosecutor / defense)
- Communicated to whom?
  - Trier of fact (jury or judge)
    - 2018 – Only 2% of federal criminal cases went to jury trial
    - 2013-14 – Only 2% of felony cases in CA went to jury trial
  - Investigators
  - Attorneys
- Communicated to what effect?
  - Primarily about decision-making
  - Decision making by jury, investigators, attorneys

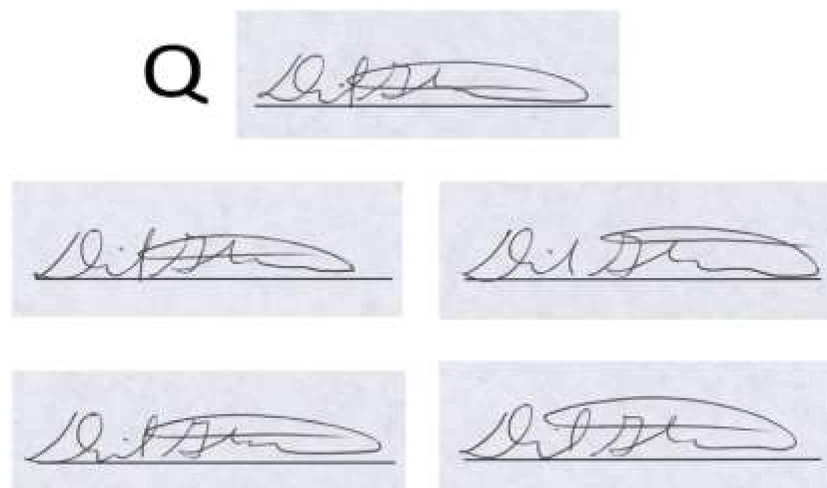
What is being communicated?

- The task of interest for purposes of this presentation:
  assess two items of evidence,
  one from a known source and one from an unknown source,
  to assess the proposition that the two samples originate from the same source

- Clearly, there are other scenarios
  - Digital evidence (collecting evidence)
  - Bloodstain pattern analysis (causal mechanism)
  - DNA mixture analysis (inclusion of suspect)





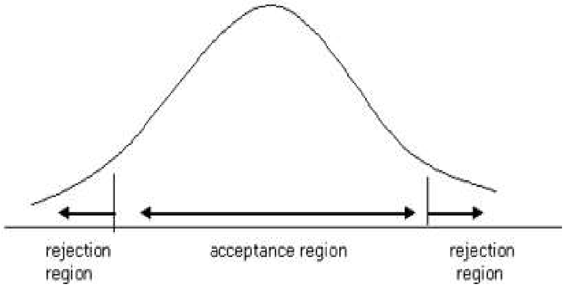


UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

csafe
Center for Statistics and Applications in Forensic Evidence

In what form is the uncertainty communicated?

- Approaches
  - Expert assessment based on experience, training, use of accepted methods. Typically summarized by a categorical conclusion (e.g., identification / exclusion / inconclusive)
  - Two-stage procedure (see, e.g., Parker and Holford in the 1960s)
    - similarity (can the Q and K be distinguished)
    - discrimination (is the observed agreement a coincidence)



  - Likelihood ratio (or the closely related Bayes factor)

Forensic Evidence as Expert Opinion

- Status quo in pattern disciplines (fingerprints, shoe prints, firearms, toolmarks, questioned documents, etc.)
- Examiner analyzes evidence based on
  - Experience
  - Training
  - Use of accepted methods in the field
- Assessment of the evidence reflects examiner's expert opinion
- Conclusions typically reported as categorical conclusions
  - Identification, Exclusion, Inconclusive
  - Multi-category scales (e.g., questioned documents)
  - Potentially via OSAC-developed interpretation scales

Forensic Evidence as Expert Opinion

- Strengths and Weaknesses:
  - Conclusions can be easily understood
  - Black-box studies can be used to provide discipline-level performance data
    - Measure reliability (reproducibility/repeatability) and accuracy
    - For example: Ulery et al. (2011) latent print study found:

Nonmated pairs:	0.15% ID	11.14% Inconcl	88.71% Excl
Mated pairs:	61.37% ID	31.09% Inconcl	7.54% Excl

    - But these studies have limitations
      - Does not address individual case/expert
      - Imc; with "inconclusive" results
      - Studies vs real casework
  - Existing scales don't address uncertainty (other than through "inconclusive")
  - It is proving challenging to develop scales that integrate uncertainty assessment
    - Some support ? Strong support?

The Two-Stage Approach

- Stage 1 - Similarity
  - Statistical test or procedure to determine if the two samples "are indistinguishable", "can't be distinguished", "match", etc.
- Stage 2 - Discrimination
  - Assessment of the probability that two samples from different sources would be found indistinguishable
- Used in assessment of trace evidence (like glass)
- Conceptually many other disciplines appear to act in this way (e.g., a footwear examination)

The Two-Stage Approach

- Strengths and Weaknesses
  - Stage 1 is a natural thing to do for discrete / categorical variables (blood type, DNA alleles)
  - Stage 1 is more challenging when the evidence are summarized by quantitative measurements (e.g., element concentrations for glass)
    - Requires a statistical procedure of some sort (e.g., ASTM E2927 for glass)
    - The usual null hypothesis (samples can't be distinguished) seems to be the wrong starting point
    - A binary decision here (distinguished / not) can involve a loss of information
  - Stage 2 is difficult (what is the relevant population?)
  - Stage 2 is not usually provided in a quantitative way

The likelihood ratio (LR)

- A current focus of much attention in forensic science research is the likelihood ratio
- The LR is a statistical concept seen as a potential unifying logic for evaluation and interpretation of forensic evidence
- The LR already plays a role outside forensics in …
  - Statistical inference (hypothesis tests)
  - Evaluating evidence provided by medical diagnostic tests
- Europe has moved in this direction (ENFSI Guidelines and work of NFI)

The likelihood ratio (LR)

- E = evidence
- H_s = "same source" proposition (two samples have the same source)
  H_d = "different source" proposition (two samples have different sources)
- Bayes' Theorem

$$\frac{Pr(H_s \mid E)}{Pr(H_d \mid E)} = \frac{Pr(E \mid H_s)}{Pr(E \mid H_d)} \; \frac{Pr(H_s)}{Pr(H_d)}$$

"a posteriori" odds in favor of same source hypothesis

Likelihood ratio or Bayes factor

"a priori" odds in favor of same source hypothesis

- Details: role of task-relevant contextual information, terminology (LR vs Bayes factor)

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences
csafe
Center for Statistics and Applications in Forensic Evidence

Likelihood ratio (LR)

- Current state
  - Successfully used for "single source" DNA
    - Underlying biology is understood
    - Biological theory provides a probability model
    - Data is available
    - Note that DNA mixtures remain challenging
  - Examples in other disciplines
    - Glass (Aitken and Lucy)
    - Bullet lead (Carriquiry, Daniels, Stern)
  - Pattern evidence has proven challenging
    - How to represent the evidence as quantitative data
    - Score-based approaches are often used
      (replace evidence E by score S)

Likelihood ratio (LR)

- Strengths and Weaknesses
  - Explicitly compares two (or more) relevant hypotheses/propositions
  - Provides a mapping from a specified set of assumptions to a quantitative summary of the evidence
    - Assumptions regarding probability distributions, manufacturing, transfer of evidence, etc.
    - Making such assumptions explicit has the potential to enhance the transparency of the evidence assessment process
    - But LR can be quite sensitive to the assumptions (Lund and Iyer, 2017)
  - Avoids arbitrary match/non-match decisions when faced with continuous data
  - Can potentially accommodate a wide range of factors (e.g., manufacturing, distribution, wear)
  - Very difficult to develop models for pattern evidence; score-based models have promise but also limitations
  - Challenging for people (especially non-quantitative people) to understand and interpret

Putting ideas together – LR & Expert Opinion

- Black box studies provide field-level data about error rates
- Can think about evidence E as being the expert opinion
  (not the prints, but the expert's opinion about the prints)
- LR would then tell us to find Pr(E | known match) and Pr(E | known non-match)
- From Ulery et al.
  - If E = "ident", then *LR* = (3663/5969) / (6/4083) = 418 in favor of same source
  - If E = "exclude", then LR = .085 in favor of same source
    or LR = 1/.085 = 12 in favor of different source
  - If E = "inconclusive", then LR = 2.8 in favor of same source
- From the recent Monson et al. firearms (bullet) data
  - If E = "ident", then LR = 109 in favor of same source
  - If E = "elimination", then LR = .086 in favor of same source
    or LR = 1/.086 = 12 in favor of different source
  - If E = "inconclusive-A", then LR = 1 (not informative)
  - If E = "inconclusive-B", then LR = 3 in favor of different source
  - If E = "inconclusive-C", then LR = 10 in favor of different source

Putting ideas together – LR & Two-Stage

- Stage 1 of two-stage approach determines whether two evidence samples (e.g., glass) are ”indistinguishable”
- Can think about evidence E being ”observation that samples are indistinguishable”
- LR would then tell us to evaluate Pr(E | same source) and Pr(E | different source)
- Pr(E | same source) is usually very high (depends on statistical procedure used to determine whether we can distinguish), typically .95 or higher
- Stage 2 is our attempt to calculate Pr(E | different source)
- Stage 2 is key to understanding the value of the evidence

Conclusions

- Any approach to assessing the probative value of forensic evidence should:
  - Account for the two (or more) competing hypotheses about how the evidence (data) were generated
  - Be explicit about the reasoning and assumptions on which the assessment is based
  - Have relevant empirical support for the reasoning and assumptions
  - **Include an assessment of the level of uncertainty associated with the assessment**
- The language used in reports, testimony, opening/closing statements are critical.
- Contact: sternh@uci.edu

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

csafe
Center for Statistics and Applications in Forensic Evidence