Statistical Thinking for Forensic Practitioners

Hal Stern
University of California, Irvine

Center for Statistics and Applications in Forensic Evidence

October / November 2022

Outline

- Part 1 - Probability Concepts and Their Relevance to Forensic Science
  - review of probability concepts
  - conditional probability and independence
  - Bayes' Theorem and likelihood ratio
- Part 2 - Data, Measurement, Reliability and Expert Opinion
  - collecting data
  - measurement, variability, reliability and accuracy
  - forensic evidence evaluation as expert opinion / black box studies
- Part 3 - Statistical Inference and the Two-Stage Approach for Assessing Forensic Evidence
  - estimation, confidence intervals, significance tests
  - two-stage approach (significance test/coincidence probability)
- **Part 4 - The Likelihood Ratio Approach - Strengths and Weaknesses**
  - **introducing the likelihood ratio**
  - **examples – the good, the bad, and the ugly**

Learning Objectives for Part 4

- Understand the underlying logic of the likelihood ratio
- Develop familiarity with the likelihood ratio approach for categorical evidence
- Understand the challenges associated with likelihood ratios for pattern evidence
- Understand the strength and weaknesses of the likelihood ratio approach

Probability
A short review

- Probability is the mathematical language of uncertainty
- Provides a common scale (0 to 1) for describing the chance that an event will occur
- Need to think about where probabilities come from – data, theory, subjective opinion
- Conditional probability is a key concept ...
  the probability of an event depends on what information is considered
- Independent events can be powerful (allows us to multiply probabilities as is common in DNA analysis) ... but the assumption needs to be confirmed
- Important to carefully interpret conditional probability $P(A \mid B)$
  - what events are we assigning probabilities to (the event A)
  - what information are we assuming to be true (the event B)

Data, Measurement, Reliability and Expert Opinion

A short review

- Random samples allow for generalization to the population
- Controlled experiments are best for drawing cause/effect conclusions
- Understanding uncertainty of measurements / decisions is crucial (e.g., ISO standard)
  - reliability refers to the consistency of measurements / decisions
  - validity refers to the accuracy of measurements / decisions
- Black box studies provide useful "discipline"-wide metrics regarding the use of expert opinion to summarize evidence

Data, Measurement, Reliability and Expert Opinion

A short review

- Statisticians distinguish between different types of data
- The different types require different measurement and analysis methods
  - qualitative data
    - categorical (blood type: A,B,AB,O)
    - ordinal (grades: A, B, C, D, F)
  - quantitative data
    - discrete (consecutive matching striae)
    - continuous (refractive index of a glass fragment)
- Probability distributions (e.g., normal) are used to describe how likely it is to see different measurement values

Statistical Inference and The Two-Stage Approach
A short review

- Statistical inference uses sample data to draw conclusions about a population
- Point estimation, interval estimation, hypothesis tests are main tools
- Statistical hypothesis tests can be useful but difficult to interpret at times
- Two-stage approach to forensic inference
  - First stage determines if the known and unknown samples appear to "match" or "be indistinguishable"
    - relies on statistical tests (or intervals)
    - important to recognize the asymmetry in testing a null hypothesis
  - Second stage attempts to quantify the probability of a coincidental match
    - requires careful consideration of the relevant population
    - can be challenging to compute (no standard procedure)
    - this step is important but unfortunately sometimes omitted

The Forensic Examination

- There are a range of questions that arise in forensic examinations - source conclusions, timing of events, substance ID, cause/effect
- Focus today on source conclusions
  - Topics addressed (e.g., need to address uncertainty, logic of the likelihood ratio) apply more broadly
  - Evidence E are items/objects found at crime scene and on suspect (or measurements of items)
    - occasionally write E_c(crime scene), E_s(suspect)
    - may be other information available, I (e.g., evidence substrate)
  - Two hypotheses
    - H_s - items from crime scene and suspect have the same (common) source or the suspect is (specific) source of crime scene item
    - H_d - different source / no common source
  - Goal: assessment of evidence
    - do items appear to have a common source
    - how unusual is it to find observed evidence / observed agreement by chance

Logic of the Forensic Examination

- Examine the evidence (E_c, E_s) to identify similarities and differences
- Assess the observed similarities and differences to see if they are expected (or likely) under the **same source** hypothesis
- Assess the observed similarities and differences to see if they are expected (or likely) under the **different source** hypothesis
  - Note that this includes assessing how unusual or rare the matching features are among the population

Approaches to Assessing Forensic Evidence

- There are multiple approaches to carrying out an examination of this type to assess the evidence
- Many different categorizations of the approaches
- We focus on three common approaches
  - Expert assessment based on experience, training, accepted methods
    - "black box" studies are relevant here
  - Two-stage approach
    - determination of similarity (often based on a statistical procedure)
    - identification (assessing likelihood of a coincidental match)
  - **Likelihood ratio / Bayes factor**

Recall the State (CT) vs Skipper Case

- Defendant charged with sexual assault
- State's expert witness reported on results of a genetic paternity test
- Expert reported a paternity index (likelihood ratio) of 3496 (probability that defendant would produce a child with the given genotype is 3496 times as large as the probability that a random male would produce such a child)
- Expert indicated the paternity index could be converted into a statistic that gave the defendant's probability of paternity
- He did so and reported the probability of paternity $= 3496/3497 = 0.9997$

Recall the State (CT) vs Skipper Case

The details

- $E =$ genetic evidence
- $H_d =$ "defendant is the father" hypothesis
  $H_r =$ "random man is the father" hypothesis
- Bayes' Theorem

$$\frac{P(H_d \mid E)}{P(H_r \mid E)} = \frac{P(E \mid H_d)}{P(E \mid H_r)} \frac{P(H_d)}{P(H_r)}$$

- We will discuss this in more detail shortly
- Expert testified that $\Pr(E \mid H_d)/\Pr(E \mid H_r) = LR = 3496$
  (evidence is much more likely if defendant is father than if a random man is the father)
- Expert assumed prior odds of 1-to-1
  (50% probability for H_d and 50% probability for H_r)
- Expert computed posterior odds are 3496-to-1 which gives
  $\Pr(H_d \mid E) = 3496/3497 = .9997$

State (CT) vs Skipper - the role of prior information

- Skipper was convicted
- He filed appeal claiming the statistical evidence was improperly admitted
- State Supreme Court found the expert's application of Bayes' Theorem was inconsistent with the presumption of innocence and remanded for new trial
  - Court determined that the conversion done by the expert to go from LR to posterior odds assumed prior probability of paternity was 0.50
  - Found this to violate presumption of innocence

Review of Bayes' Theorem - Gunshot residue test

- Consider a diagnostic test for gunshot residue on an individual
  - Let G denote the event that gunshot residue is present (we will say "not G" to denote the opposite event)
  - Let T denote the event that our diagnostic test is positive (indicates gunshot residue is present) and "not T" to indicate a negative test

True Status	Test Result T	Not T
G	True Positive	False Negative
Not G	False Positive	True Negative

  - We frequently have information about the performance of the test
  - $P(T \mid G)$ = true positive rate, sensitivity
  - $P(\text{not } T \mid \text{ not } G)$ = true negative rate, specificity
  - $p(T \mid \text{ not } G)$ = false positive rate
  - $p(\text{not } T \mid G)$ = false negative rate

Review of Bayes' Theorem - Gunshot residue test

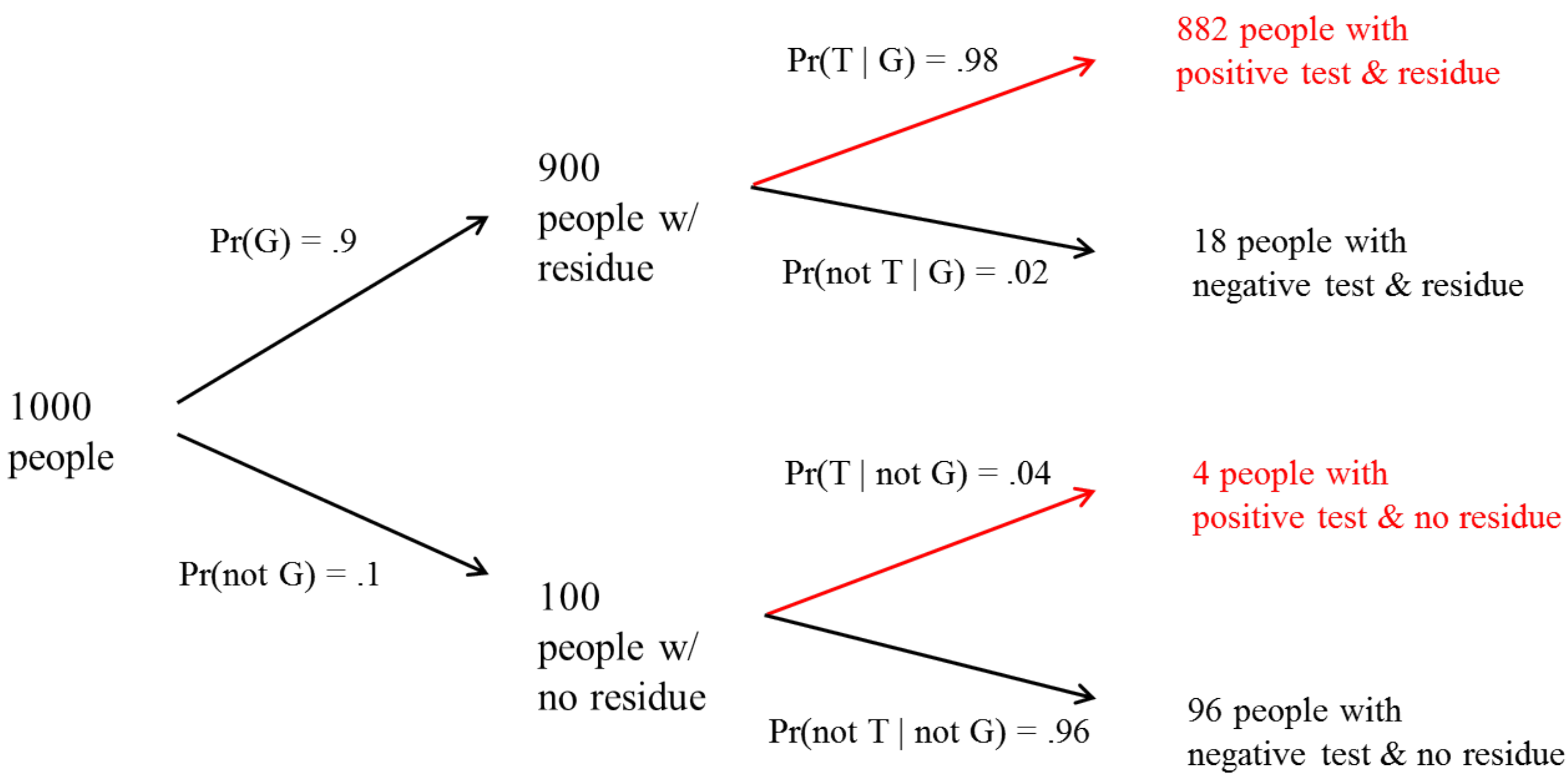
- Bayes' Theorem provides a means of taking information we have about the test (how the test performs on people whose status is known) to infer the status of an individual who has been tested
  - We know $\Pr(T \mid G)$ and $\Pr(T \mid \text{ not } G)$
  - Bayes' Theorem allows us to calculate $\Pr(G \mid T)$
- To do this **Bayes' Theorem also requires some "prior" information** about the prevalence of gunshot residue in the population of interest (i.e., $\Pr(G)$)
- A couple of key points
  - The "prior" information is important and it is not clear where it should come from
  - $P(T \mid G) \neq P(G \mid T)$ (sensitivity of the test is not the same as our certainty given the test result)

Review of Bayes' Theorem - Gunshot residue test

- Example 1
  - Assume $\Pr(T \mid G) = 0.98$
    (the test is very sensitive - many true positives)
  - Assume $\Pr(\text{not } T \mid \text{not } G) = 0.96$
    (the test is pretty specific - low rate of false positives)
  - For now assume $\Pr(G) = 0.90$ in the population of interest
  - We test an individual and get a positive test result
  - What can we say about the probability that the individual actually has gunshot residue on them?
    - It turns out that $\Pr(G \mid T) = 0.995$

Review of Bayes' Theorem - Gunshot residue test

- Suppose we have a population of 1000 individuals



- Conclusion: Note that there are 886 positive tests and 882 are "true" $Pr(G|T) = Pr(\text{residue} \mid \text{positive}) = 882/886 = .995$

Review of Bayes' Theorem - Gunshot residue test

- Example 1 - repeated from earlier slide
  - Assume $\Pr(T \mid G) = 0.98$ (high sensitivity)
  - Assume $\Pr(\text{not } T \mid \text{not } G) = 0.96$ (high specificity)
  - Assume $\Pr(G) = 0.90$ in the population of interest
  - We test an individual and get a positive test result
  - It turns out that $\Pr(G \mid T) = 0.995$
- Example 2
  - Assume $P(T \mid G) = .98$
  - Assume $\Pr(\text{not } T \mid \text{not } G) = 0.96$
  - **Now assume** $P(G) = .05$ **(a low prevalence scenario)** (i.e., testing in a population where gun usage is rare)
  - If we get a positive test result, then Bayes' theorem tells us $\Pr(G \mid T) = 0.56$
- Prior information about the prevalence has a big impact on the interpretation of the test result

Conditional Probability / Bayes' Theorem in the Courtroom

- E = evidence
  - DNA markers from the crime scene sample and suspect sample
  - Measurements on glass fragments from crime scene / suspect's clothing
  - Image of bullet cartridges found at crime scene / test fire from suspect's weapon
- H_s = "same source" proposition (two samples have same source)
  H_d = "different source" proposition (two samples w/ different sources)
- Then
  $\Pr(E \mid H_s)$ = probability of seeing evidence if suspect is the source
  $\Pr(E \mid H_d)$ = probability of seeing evidence if suspect is not the source
- And
  $\Pr(H_s \mid E)$ = probability suspect is the source given the evidence
  $\Pr(H_d \mid E)$ = probability suspect is not the source given the evidence

Conditional Probability / Bayes' Theorem in the Courtroom

- It is critical to be clear about the event to which we are assigning probability
- $\Pr(E \mid H)$ is about the probability of obtaining evidence E (assuming hypothesis H is true)
- $\Pr(H \mid E)$ is about the probability of hypothesis H being true (given we have evidence E)
- Confusion about these different probabilities is common
- The prosecutor's fallacy is one famous example
  - Involves interpreting $\Pr(E|H_d)$ as $\Pr(H_d|E)$
  - Finding that the evidence is unlikely under H_d is interpreted by prosecutor as saying that H_d is unlikely
  - Prosecutor in the Collins case did this!

Likelihood Ratio Approach

Introduction

- The goal for the trier of fact in courtroom setting is a decision about the relative likelihood of two hypotheses (e.g., same or different source) given data
- In statistical terms this suggests a Bayesian formulation (asks for probabilities about the state of the world given observed data)
- Recall that Bayes' rule is a way of reversing direction of conditional probabilities
  - We can go from statements about the **likelihood of the evidence given the hypotheses** to statements about **the likelihood of the hypotheses given the evidence**
- Bayes' rule points us towards the likelihood ratio approach

Likelihood Ratio Approach
Introduction

- Notation: E (evidence), H_s (same source), H_d (different source)
- Recall Bayes' Theorem written in terms of the odds in favor of the same source hypothesis

$$\frac{\Pr(H_s|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_s)}{\Pr(E|H_d)} \times \frac{\Pr(H_s)}{\Pr(H_d)}$$

- In words: Posterior odds = Likelihood ratio $\times$ Prior odds
- The likelihood ratio is the summary of the evidence that is relevant to applying Bayes' Theorem
- The likelihood ratio already plays a role outside of forensics (e.g., in medical diagnostic tests)
- Europe has moved in this direction (ENFSI Guideline)

Likelihood Ratio Approach

Introduction

- Reminder: $LR = \frac{\Pr(E|H_s)}{\Pr(E|H_d)}$
- Some observations
  - The LR speaks to the relative likelihood of the evidence under the two hypotheses
  - The LR **does not** make any direct statement about the probability of the hypotheses
  - If we wish to talk about the probability of the hypothesis, then we are interested in the posterior probabilities of the hypotheses
    - But to talk about posterior probabilities ....
      we must have had prior probabilities to start with
      and we should be willing to say what they are
    - Does not seem that this is the role of the forensic examiner

Likelihood Ratio Approach
Introduction

- Reminder: $LR = \frac{\Pr(E|H_s)}{\Pr(E|H_d)}$
- Some observations about the numerator
  - numerator assumes same source and asks about the likelihood of the evidence in that situation
  - if E contains many similarities and no major dissimilarities, then $\Pr(E|H_s)$ is high
  - if E contains major unexplainable differences, then $\Pr(E|H_s)$ is low
  - somewhat related to finding a p-value for testing the hypothesis of equal means in the two stage approach
  - but ... no binary decision regarding match!
  - instead a quantitative measure of likelihood of evidence under H_s

Likelihood Ratio Approach

Introduction

- Reminder: $LR = \frac{\Pr(E|H_s)}{\Pr(E|H_d)}$
- Some observations about the denominator
  - denominator assumes different sources and asks about the likelihood of the evidence in that case
  - if E contains many similarities and no major dissimilarities, then $\Pr(E|H_d)$ is low
  - if E contains major unexplainable differences, then $\Pr(E|H_d)$ is probably not small
  - analogous to finding probability of coincidental match in the two stage approach
  - here too, doesn't require a binary decision regarding match
  - a quantitative measure of likelihood of evidence under H_d

Likelihood Ratio Approach

Introduction

- Reminder: $LR = \frac{\Pr(E|H_s)}{\Pr(E|H_d)}$
- Some technicalities
  - the term likelihood is used because if E includes continuous measurements then can't talk about probability
  - could in principle be used with E equal to "all" evidence of all types but this would be very challenging
  - other available information (e.g., background) can be incorporated into the LR
  - confusion about the terms likelihood ratio / Bayes factor / Bayesian approach
    - Likelihood ratio (LR) is sometimes called the Bayes factor (BF)
    - LR and BF are both relevant in a Bayesian approach to evidence
    - distinction between LR and BF is quite technical and depends on how various parameters are treated
    - for our purposes ... the LR and BF are the same

Likelihood Ratio Approach

Introduction

- Makes explicit the need to consider the evidence under two different hypotheses
- Keeps us focused on reasoning about the evidence rather than reasoning about the hypotheses directly
- Interpretation
    - $LR > 1$ means the evidence is more likely to be obtained if H_s is true
    - $LR < 1$ means the evidence is more likely to be obtained if H_d is true
    - $LR = 1$ means the evidence is equally likely under the two hypotheses (so not informative)
    - Does not matter which hypothesis appears in numerator or denominator; we just have to make sure we interpret correctly
    - Sample LR statement: "The evidence (e.g., level of agreement) is LR times more likely if the objects have the same source than if the objects have different sources"
    - No hard and fast rules for what makes a "big" LR (or what makes a "small" LR in the case when $LR < 1$)
    - There are proposals (e.g., ENFSI) that map LRs to verbal scales (2-10: weak support; 10-100: moderate support; ....)

Test yourself

Likelihood ratios - the numerator

- The numerator of the likelihood ratio $(\Pr(E \mid H_s))$ measures
  - The probability that the suspect committed the crime
  - The probability that the expert witness is trying to confuse the jury
  - The probability of observing evidence like the evidence in this case if the two samples came from the same source
  - The probability that the two samples came from the same source given the observed evidence

Test yourself

Likelihood ratios - the numerator - answer

- The numerator of the likelihood ratio ($\Pr(E \mid H_s)$) measures
  - The probability that the suspect committed the crime - **INCORRECT**
  - The probability that the expert witness is trying to confuse the jury - **INCORRECT**
  - The probability of observing evidence like the evidence in this case if the two samples came from the same source - **CORRECT**
  - The probability that the two samples came from the same source given the observed evidence - **INCORRECT, BUT A COMMON MISUNDERSTANDING**

Test yourself

Likelihood ratios - denominator

- The denominator of the likelihood ratio ($\Pr(E \mid H_d)$) measures
  - The probability that the suspect did not commit the crime
  - The probability that the expert witness is trying to confuse the jury
  - The probability of observing evidence like the evidence in this case if the two samples came from different sources
  - The probability that the two samples came from different sources given the observed evidence

Test yourself

Likelihood ratios - denominator - answer

- The denominator of the likelihood ratio ($\Pr(E \mid H_d)$) measures
    - The probability that the suspect did not commit the crime - **INCORRECT**
    - The probability that the expert witness is trying to confuse the jury - **INCORRECT**
    - The probability of observing evidence like the evidence in this case if the two samples came from different sources - **CORRECT**
    - The probability that the two samples came from different sources given the observed evidence - **INCORRECT, BUT A COMMON MISUNDERSTANDING**

Likelihood Ratio Approach
A simple example

- Suppose evidence is blood types for a crime scene sample (y) and suspect sample (x)
- The source of the suspect sample is known
- The source of crime scene sample is unknown (random); we want to assign probabilities to y given observed data and different hypotheses
- Information about the distribution of blood types in the U.S.

Type	A	B	AB	O
U.S. Freq	.42	.10	.04	.44

- Suppose both samples are observed to be of blood type O
- To assess the evidence we should assess the likelihood of observing the random y having blood type O under the two hypotheses

Likelihood Ratio Approach

A simple example

- Information about the distribution of blood types in the U.S.

Type	A	B	AB	O
U.S. Freq	.42	.10	.04	.44

- To assess the evidence we should assess the likelihood of observing the random y having blood type O under the two hypotheses
- $\Pr(y = O|x = O, H_s) \approx 1$
  (expect crime scene sample to match suspect's type O if H_s is true)
- $\Pr(y = O|x = O, H_d) = \Pr(y = O|H_d) = 0.44$
  (type O blood is relatively common in the U.S.)
- $LR \approx \frac{1}{0.44} \approx 2.3$
- Evidence provides weak support for the "same source" hypothesis
- Note: If y and x don't match, then the numerator will be very small (evidence favors H_d)
- Note: If y and x match on a rare blood type (AB), then the denominator is small and the LR is big

Likelihood Ratio Approach

Where it works ..... DNA

- A DNA profile identifies alleles at a number of different locations along the genome (e.g., alleles at location TH01 are 7,9)
- As with blood type, we may see matching profiles (crime scene and suspect)
- Numerator is approximately one (as in blood type example)
- Can determine probability of a coincidental match for each marker or location

TH01	4	5	6	7	8	9	9.3	10	11
Freq.	.001	.001	.266	.160	.135	.199	.200	.038	.001

- For TH01 agreeing on alleles 7, 9, the probability of a random agreement is 2*.16*.199 $=$ .064 so LR $\approx \frac{1}{.064} \approx 15$

Likelihood Ratio Approach

Where it works ..... DNA

- DNA evidence consists of data for a number of locations (CODIS used 13 locations pre-2017 and more now)
- Locations on different chromosomes are independent
- Recall that if events are independent, then we can multiply probabilities
  (which basically means multiplying likelihood ratios)
- A match at all locations can lead to likelihood ratios in the billions (or even larger)

Likelihood Ratio Approach

Where it works ..... DNA

- Underlying biology is well understood
- Probability model for the evidence follows from genetic theory
- Population databases are available
- Peer–reviewed and well accepted by scientific community
- Note: Even with the above information, there are still issues in the DNA world
    - Allele calling still has some subjective elements
    - Samples containing multiple sources (i.e., mixtures)

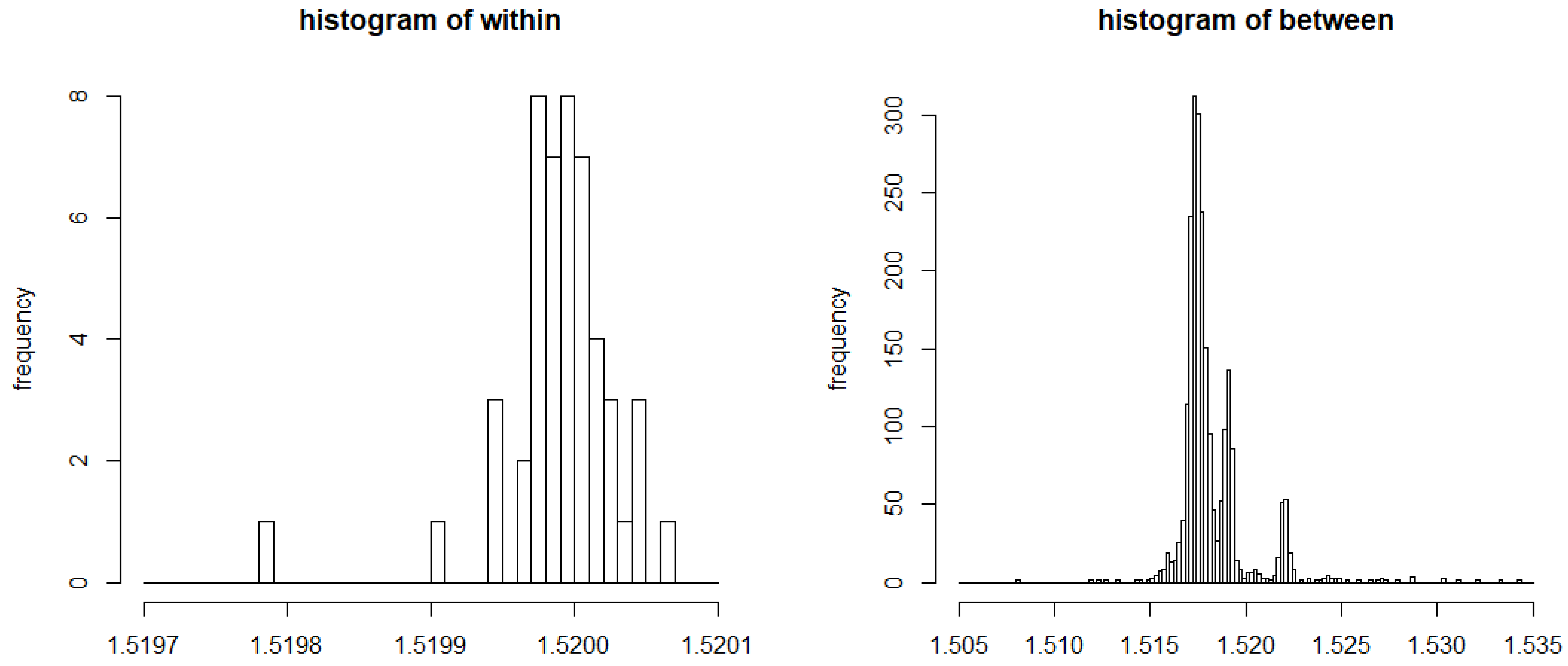
Likelihood Ratio Approach
Where it can work .... Trace evidence

- Glass and bullet lead are examples
- Can measure chemical concentrations of elements in glass (or bullet lead)
- May have broken glass at crime scene and glass fragments on suspect
- Can we construct a likelihood ratio for evidence of this type?
    - Perhaps .... motivate with some pictures of distributions of refractive indices of glass

Likelihood Ratio Approach

Where it can work .... Trace evidence

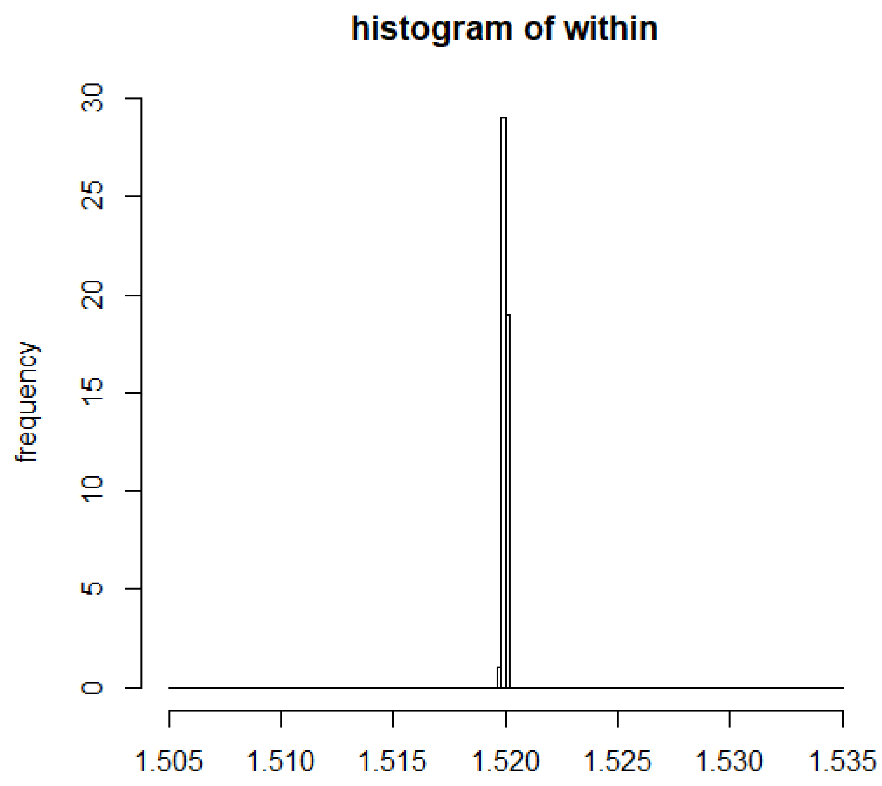
- Left plot shows distribution of 49 measurements from a single glass source
- Right plot shows distribution of (mean) measurements for samples from 2269 glass sources

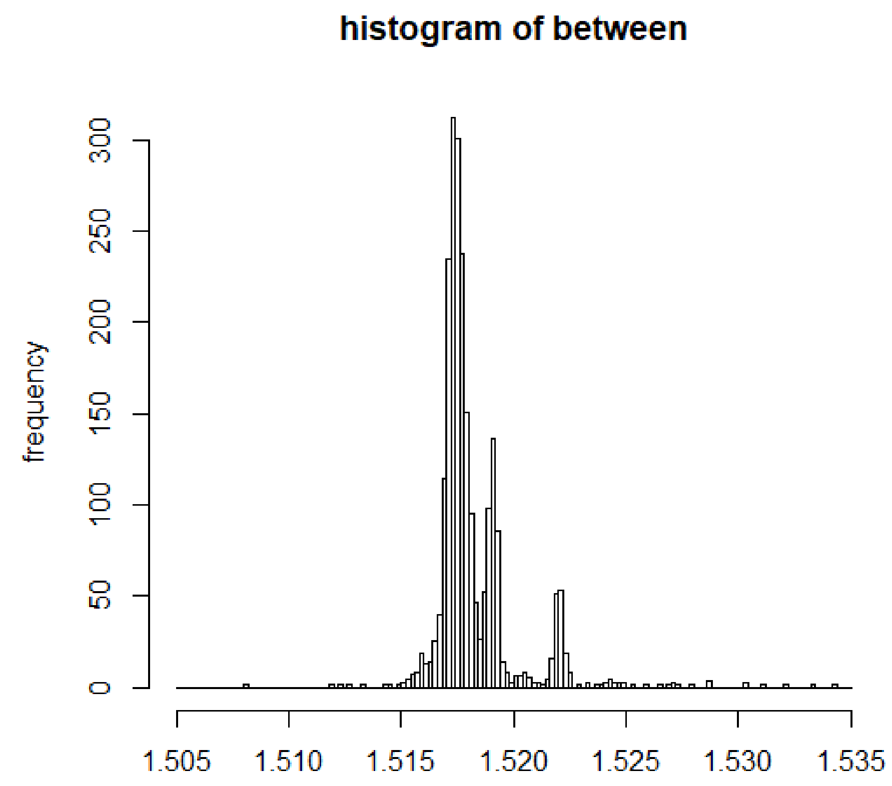


Likelihood Ratio Approach

Where it can work .... Trace evidence

- Now show the same images using a common horizontal axis
- Little variation within a single source; More variation between sources





Likelihood Ratio Approach

Continuous measures for trace data - example

- A short conceptual discussion of the approach described by Aitken and Lucy (Applied Statistics, 2004)
- Take y and x to be measurements (element concentration, refractive index) from several glass fragments at the crime scene (the control x) and the subject (the questioned y)
- Note that here the source of crime scene sample is known and the randomness is about whether y is from the same source
- Assume normal distribution for trace element concentrations (may be more reasonable for logarithms)
- Under the same source hypothesis H_s
  - x and y are two sets of measurements from a single source (i.e., from a single "within source" normal distribution)
- Under the different source hypothesis H_d
  - x and y are sets of measurements from two different sources (i.e., from two different normal distributions with means drawn from the relevant "between source" distribution of possible sources)

Likelihood Ratio Approach

Continuous measures for trace data - example (cont'd)

- It is possible compute a likelihood ratio in this scenario if we have information about
  - variability of repeated measurements from "within" a single source
  - variability among the mean measurements of sources in the population of interest (i.e., the "between" source variability)
- Key findings:
  - LR is small if y and x are very different (i.e., no match)
  - LR is big if y and x are similar and y is unusual for the population of interest (i.e., a match on an unusual value)
- Aitken and Lucy examples find typical LRs for glass evidence in the 100s or 1000s

Likelihood Ratio Approach

Where it can work .... Trace evidence

- Well-defined set of measurements (e.g., chemical concentrations)
- Plausible probability models to describe variation within a sample (e.g., normal distribution or less restrictive models)
- Possible to sample from a population (e.g., other windows) to assess variation across different sources
- Can and has been done
  - Aitken and Lucy (2004) - glass
  - Carriquiry, Daniels and Stern (2000 technical report) - bullet lead
- But ...
  - Assessing the relevant "population" is hard (and may vary from case to case)
  - Likelihood ratios can be very sensitive to assumptions that are made (Lund and Iyer, NIST 2017)

Test yourself

Likelihood ratios - bullets

- Recall the data below from Part I of the course
- Li, 2012 thesis, U Cent. Okla - maximum consecutive matching striae (CMS) in comparing 9mm bullet groove impressions from known matches and known non-matches

	2	3	4	5	6	7	8	Total
Known Matches	55	54	23	11	2	0	1	146
Known Non-Matches	48	11	1	0	0	0	0	60

- A forensic examiner comparing 9mm bullet groove impressions from a questioned and known sample in a case identifies the maximum CMS as 4. Based on the table above, the LR ratio for comparing the same source hypothesis to the different source hypothesis is .....
  - $23 \ / \ 1 = 23$
  - $1 \ / \ 23 = 0.043$
  - $(23/146)/(1/60) = .158 \ / \ .0167 = 9.5$
  - $(23/24)/(146/60) = .958 \ / \ 2.43 = 0.39$
  - You must be kidding!

Test yourself

Likelihood ratios - bullets - answer

	2	3	4	5	6	7	8	Total
Known Matches	55	54	23	11	2	0	1	146
Known Non-Matches	48	11	1	0	0	0	0	60

- Suppose we observe the maximum CMS as 4.
  - The evidence E is the observation that the maximum CMS is 4
  - The known match row of the table allows us to estimate that $\Pr(CMS = 4 \mid \text{same source}) = 23/146 = .158$
  - The known non-match row of the table allows us to estimate that $\Pr(CMS = 4 \mid \text{different source}) = 1/60 = .0167$
  - The LR based just on the table is thus $.158/.0167 = 9.5$. The data supports the same source hypothesis but is not overwhelming
  - Note that **for these data**, CMS=2 would yield an LR of .47 and thus provide limited support for the different source hypothesis
  - Note that **for these data**, CMS=5 would yield an infinite LR in favor of same source since we never saw that in the non-match population. But of course we would need more data to get a reliable LR
  - We need more data in any case! This just serves as an illustration.

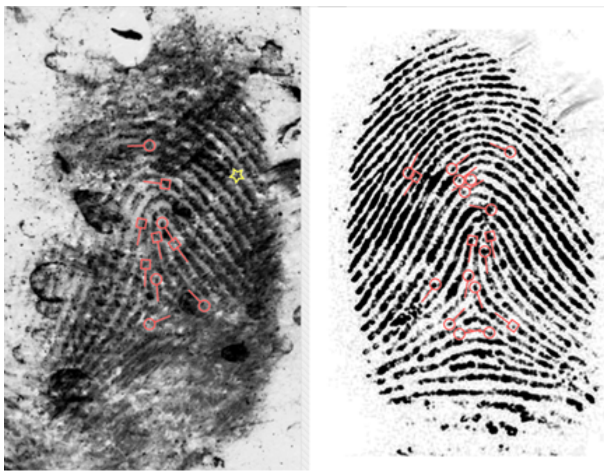
Likelihood Ratio Approach

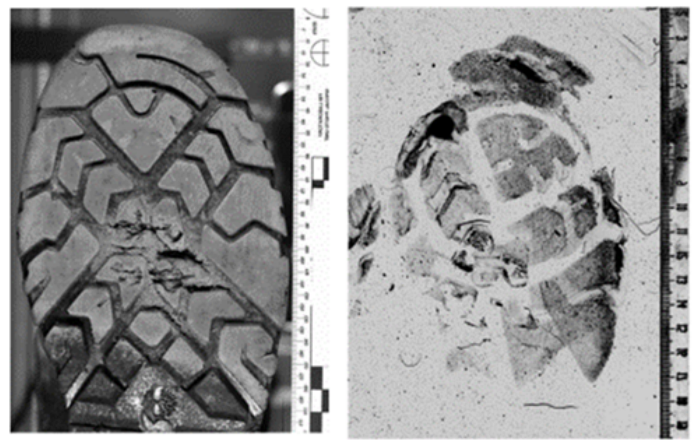
Where it might (?) work .... Pattern evidence

- Many forensic disciplines are focused on comparing a sample (mark) at the crime scene (the "unknown" or "questioned") and a potential source (the "known")
- Need to assess whether two samples have same source or different source
- Many examples
  - Latent prints
  - Shoe prints and tire tracks
  - Questioned documents / handwriting
  - Firearms
  - Tool marks

Likelihood Ratio Approach

Where it might work .... Pattern evidence





Likelihood Ratio Approach

Where it might work .... Pattern evidence

- A number of challenges in constructing likelihood ratios
- Even defining what we mean by the evidence E is challenging
  - Data are very high dimensional (often images)
  - Flexibility in defining the types of features and the number of features to examine
  - Typically E is taken to include
    - observed features from the crime scene and suspect samples
    - observations regarding their similarities and differences

Likelihood Ratio Approach
Where it might work .... Pattern evidence

- As with trace evidence, formal evaluation here requires that we study two different types of variation
  - Require information about the variation expected in repeated impressions from the same source (e.g., distortion of fingerprints) to talk about $\Pr(E \mid H_s)$
  - Require information about the variation expected in impressions from different items in the population (i.e., the "coincidental match" probability) to talk about $\Pr(E \mid H_d)$
  - May also need information about manufacturing, distribution, wear patterns (e.g., for shoes)

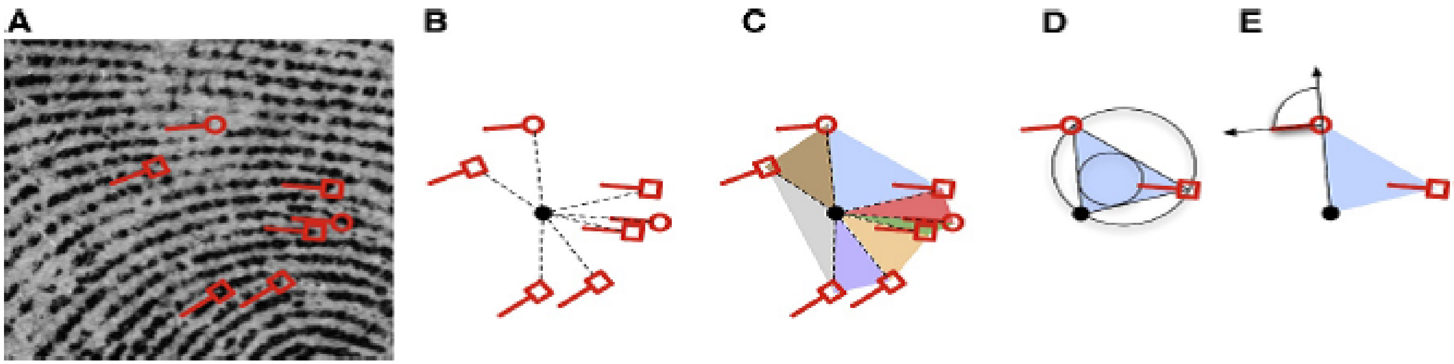
Likelihood Ratio Approach

Where it might work .... Pattern evidence

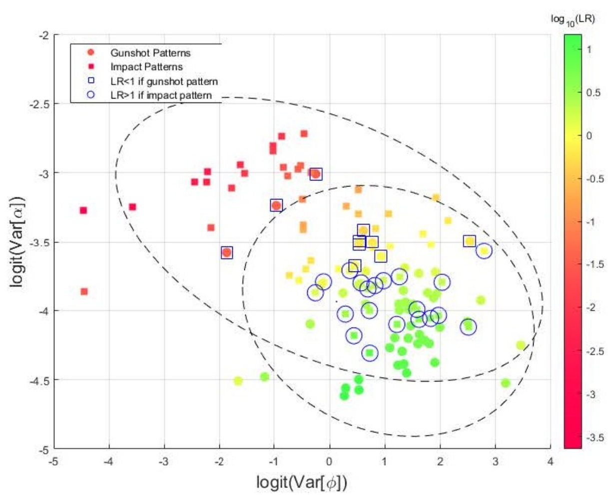
- How do we measure $\Pr(E \mid H_s)$ and $\Pr(E \mid H_d)$
- This is a very hard problem!!
  - Need to assign probabilities to all possible observations E
- Some approaches:
  - Probability models for features
  - Subjective likelihood ratios (permitted by ENFSI Guideline)
  - Score-based likelihood ratios

Likelihood Ratio Approach
Probability Models for Features

- Example 1: latent prints (from Neumann et al., 2015)



- Example 2: bloodstain pattern analysis (from Zou and Stern, submitted)



Likelihood Ratio Approach
ENFSI Guideline for Evaluative Reporting

- ENFSI has officially endorsed likelihood ratios (ENFSI Guideline)
- Guideline cites four requirements for evaluative reporting: balance, logic, robustness, transparency
- Some key statements from the Guideline:
  - Evaluate findings (evidence) with respect to competing hypotheses
  - Evaluation should use probability as a measure of uncertainty
  - Evaluation should be based on the assignment of a likelihood ratio
- According to the Guideline, probabilities in the likelihood ratio are ideally based on published data but experience, subjective assessments, case-specific surveys can be used as long as justified
  - The use of experience-based or subjective probabilities has been viewed a bit more skeptically in the U.S.

Likelihood Ratio Approach
ENFSI Guideline for Evaluative Reporting

- LRs reported as numbers or as verbal equipments
- Verbal equivalents are less precise, but may be easier to understand

Value of LR	Verbal equivalent: "The forensic findings ...
1	do not support one proposition over the other
2 – 10	provide **weak support** for the same source proposition relative to the different source proposition
10 - 100	provide **moderate support** for the same source proposition relative …
100 - 1000	provide **moderately strong support** for the same source proposition relative …
1000 - 10000	provide **strong support** for the same source proposition relative …
10000 – 1 mill.	provide **very strong support** for the same source proposition relative …
1 million +	provide **extremely strong support** for the same source proposition relative …

Likelihood Ratio Approach

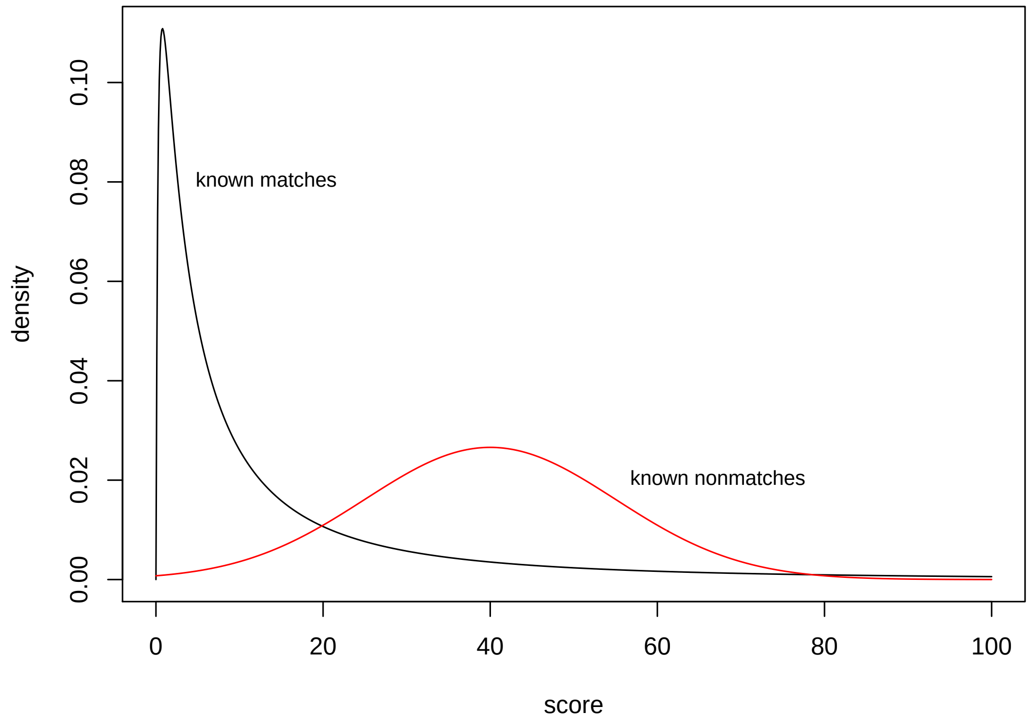
Score-based likelihood ratios

- Given the challenge in developing LRs for pattern evidence there has been recent work developing score-based approaches
- Define a score measuring the "difference" between the questioned and known samples (let's call the score D)
- Essentially we are replacing the evidence "E" by the score "D"
- As an example, we earlier saw an analysis of bullet groove impressions in which the evidence were replaced by the maximum CMS

Likelihood Ratio Approach

Score-based likelihood ratios

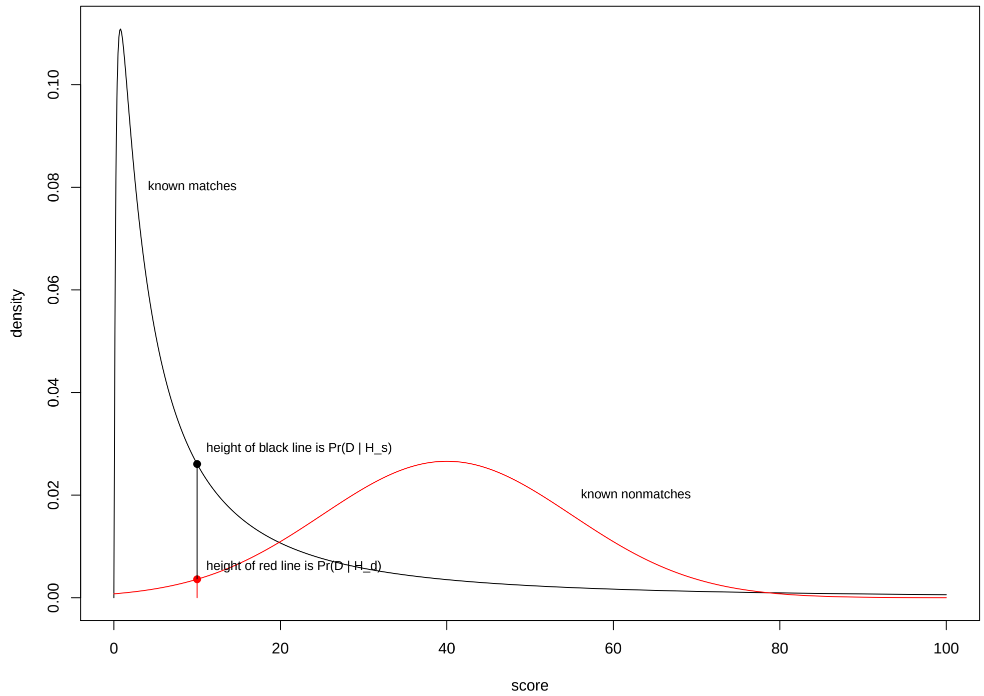
- Apply the likelihood ratio idea to the scores
- Obtain the distribution of scores for a sample of known matches (i.e., under same source hypothesis H_s)
- Obtain the distribution of scores for a sample of known non-matches (i.e., under different source hypothesis H_d)



Likelihood Ratio Approach

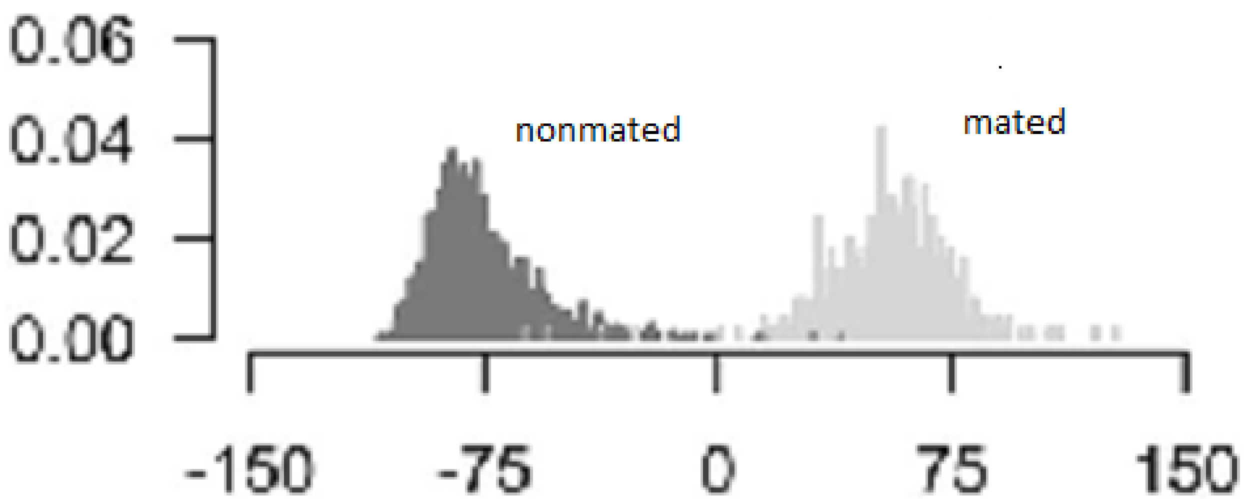
Score-based likelihood ratios

- The score-based likelihood ratio idea
  - Fit a probability distribution to the scores of known matches ($Pr(D \mid H_s)$)
  - Fit a probability distribution to the scores of known nonmatches ($Pr(D \mid H_d)$)
  - Score-based likelihood ratio if we observe score D is $SLR = Pr(D \mid H_s)/Pr(D \mid H_d)$



Likelihood Ratio Approach
Score-based likelihood ratios

- Example: FRSTATS for latent prints (Swofford et al, 2018)



Likelihood Ratio Approach

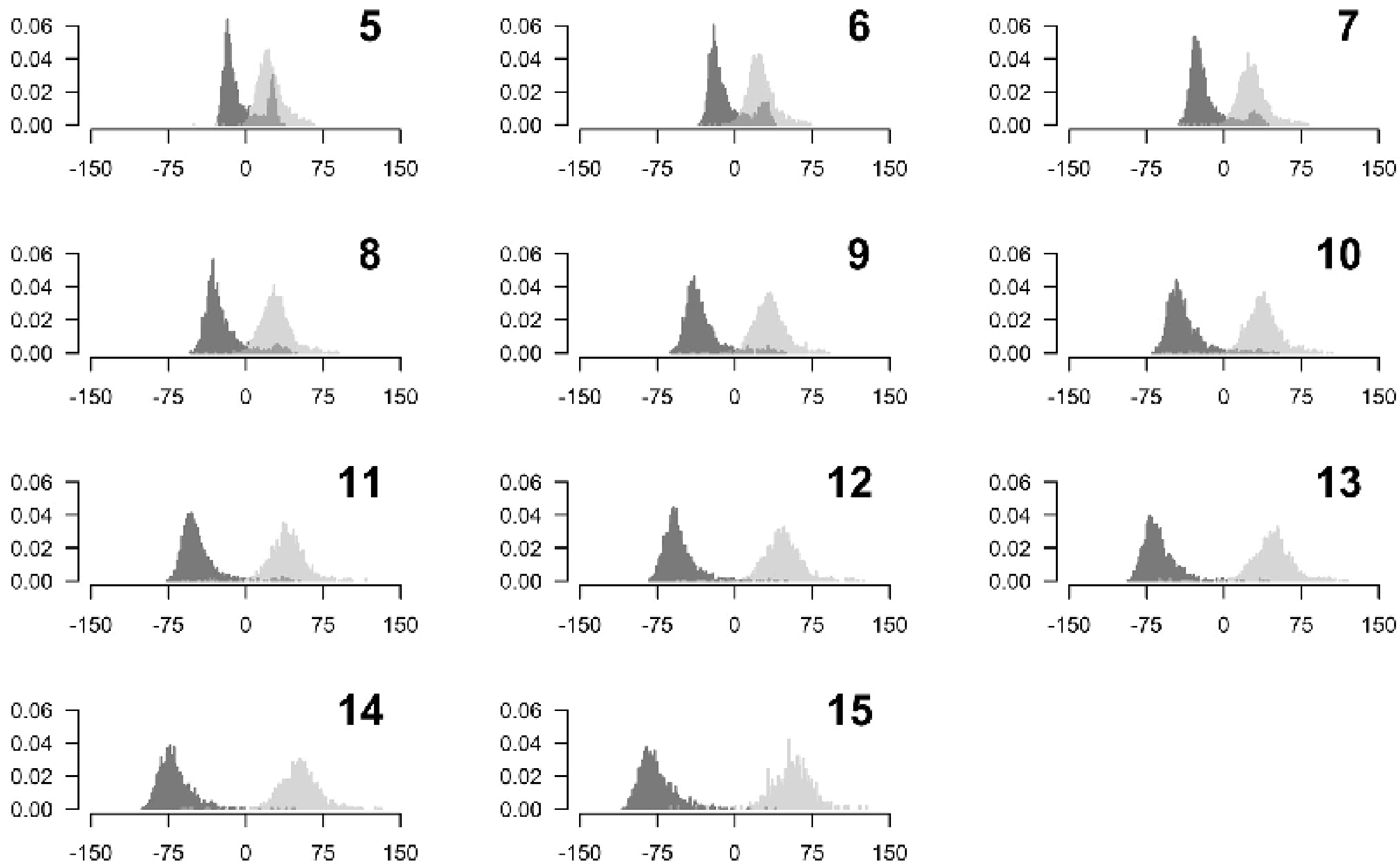
Score-based likelihood ratios - challenges

- Across a number of existing examples the score distribution for known matches seems relatively straightforward to characterize
- Careful thought is required to define the relevant non-match population
  - Is there a single non-match score distribution?
  - Should the non-match score distribution depend on characteristics of the crime scene sample?
  - If so, which characteristics?

Likelihood Ratio Approach

Score-based likelihood ratios - challenges

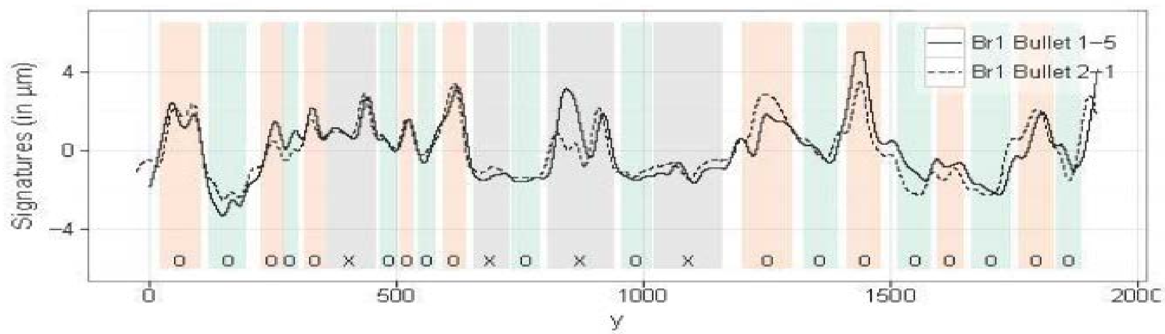
- Example: FRSTATS (Swofford et al., Forensic Science International, 2018)

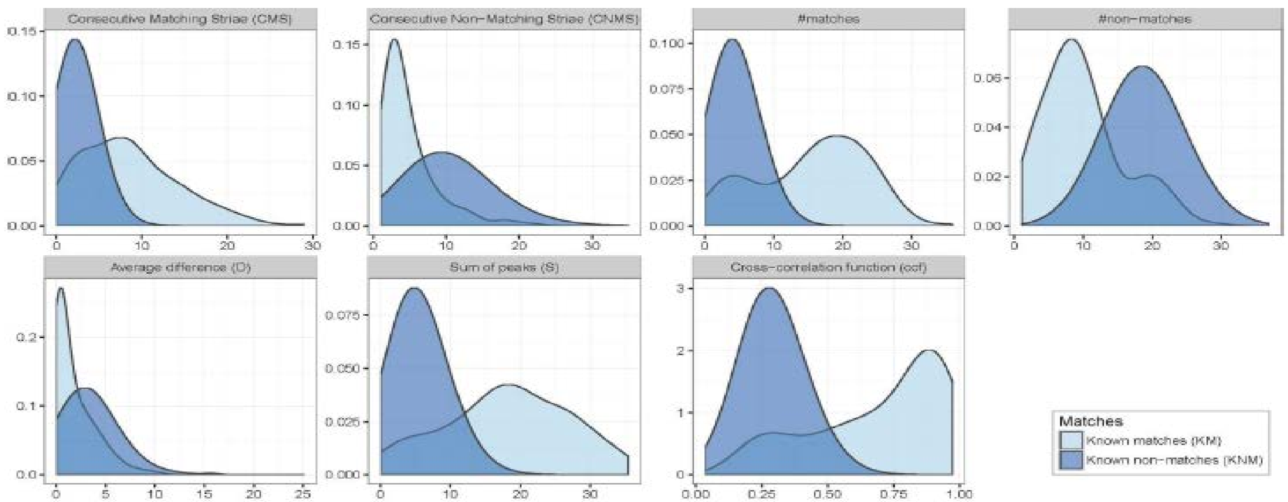


Likelihood Ratio Approach

Score-based likelihood ratios - challenges

- Another challenge is that sometimes no single score is found that distinguishes well between same-source and different-source pairs
- Example - bullet land signatures

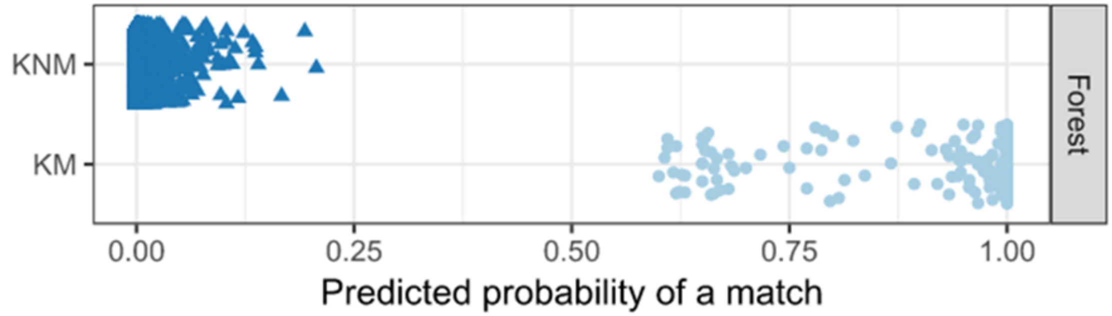




Likelihood Ratio Approach

Score-based analyses - challenges

- One approach in this case is to combine multiple scores (the D's) into a single summary score
- A possible summary score is an estimate of the probability (based on training data) that a given set of scores indicates a match
- Example - bullet lands (Hare et al., 2017, Annals of Applied Statistics)
    - Use features from previous slide
    - Fit a statistical model (random forest) to training examples (matches / nonmatches) and use the model and scores to estimate the probability of a match in the given case
    - Call the summary score S
    - Can show distribution of S for same source and different source pairs
    - Example from the paper for bullet lands - pairs are well separated



Likelihood Ratio Approach

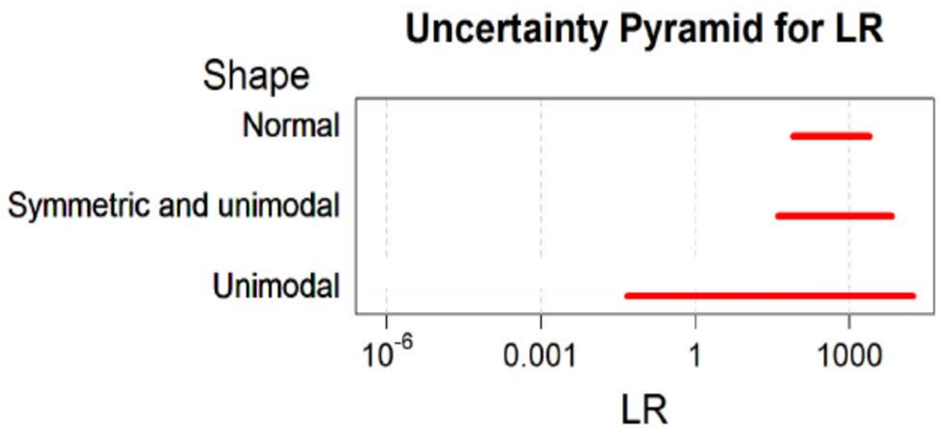
Score-based analyses - another approach

- The summary scores from the previous slide can also be used in a way that is related to the two-stage approach
- Choose a threshold for the summary score
- If S (estimated probability of a match) is above the threshold, then declare same source
- If S is below the threshold, then declare different source
- Can compute error rates for this type of procedure (and can repeat this for different thresholds)

Likelihood Ratio Approach

Sensitivity of results to assumptions

- Lund and Iyer (NIST, 2017) noted that a range of different statistical models can be used in deriving the likelihood ratios
- For a given set of observed data they considered a range of "plausible" models and explored the range of LRs observed
- Figure below is adapted from Lund and Iyer
    - Red lines show range of LRs obtained using distributions in the identified class (e.g., normal) that are consistent with the observed data
    - Lund/Iyer make a critical point
    - Lund/Iyer are (in my view) extremely generous in defining what is "plausible"



Likelihood Ratio Approach
Complications

- Many issues can complicate the calculation of LRs in practice. Examples include ...
  - accounting for transfer process with glass or fibers
  - accounting for heterogeneity due to packaging of bullets into boxes
  - accounting for usage/lifetime of products (e.g., sneakers)
- Though good work is being done, it seems likely that it will be some time before LRs are available for pattern evidence
- Important to remember that there is not one LR for a given item of evidence
  - The LR calculation depends on assumptions/models for the measured data
  - The LR calculation depends on assumptions/models for the relevant population

Test yourself

Likelihood ratios

- Which of the following statements about likelihood ratios are true
  - Likelihood ratios provide a continuous measure of the strength of the evidence
  - Likelihood ratios are the greatest invention since sliced bread
  - The likelihood ratio is intriguing but seems very difficult to apply in practice
  - Only statisticians are interested in the likelihood ratio
  - Likelihood ratios can be very sensitive to the assumptions made

Test yourself

Likelihood ratios - answer

- Which of the following statements about likelihood ratios are true
  - Likelihood ratios provide a continuous measure of the strength of the evidence - **TRUE**
  - Likelihood ratios are the greatest invention since sliced bread - **SOME PEOPLE THINK SO**
  - The likelihood ratio is intriguing but seems very difficult to apply for pattern evidence - **TRUE**
  - Only statisticians are interested in the likelihood ratio - **MAYBE TRUE**
  - Likelihood ratios values can be sensitive to the assumptions made - **TRUE**

Likelihood Ratio Approach
Summary

- Advantages
  - explicitly compares two relevant hypotheses/propositions
  - provides a quantitative summary of the evidence
  - assumptions being used are (or should be) made explicit and open to question
  - no need for arbitrary match/non-match decisions when faced with continuous data
  - can accommodate a wide range of factors
  - flexible enough to accommodate multiple pieces and multiple types of evidence

Likelihood Ratio Approach

Summary

- Disadvantages
  - requires assumptions about distributions
  - calculated LR can be sensitive to these assumptions
  - in the US there is no requirement for defense to provide a specific alternative hypothesis
  - need for reference distributions to define denominator (although this needs to be done implicitly in any examination)
  - can be difficult to account for all relevant factors
  - how should this information be conveyed to the trier of fact

Putting Some Ideas Together

Expert Opinion and the Likelihood Ratio

- 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 $LR = \frac{3663/5969}{6/4083} = 418$ 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
- 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 Some Ideas Together

Two-Stage Approach and the Likelihood Ratio

- 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 | S) 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

Short Course Summary / Conclusions

- Analysis of forensic evidence requires some familiarity with concepts from probability and statistics
- Course reviewed basics of probability and statistics
- Discussed statistical considerations relevant to current practice (expert opinion)
- Reviewed statistical approaches to forensic evidence (two-stage approach and likelihood ratio/Bayes factor)
- Addressed statistical considerations needed for each of these approaches to forensic evidence
- Key points
    - Any approach should account for the two (or more) competing hypotheses about how the data was generated
    - Need to be explicit about reasoning and data on which reasoning is based
    - Need to describe the level of certainty associated with a conclusion