An Introduction to Statistical Thinking for Forensic Practitioners

8:30am – 3:30pm, March 3, 2016
Palm Beach County Sheriff’s Office
Palm Beach, FL

PART I: Hal Stern, Alicia Carriquiry, Colin Lewis-Beck, CSAFE

Why Probability and Statistics?
• A forensics problem to motivate the discussion
• Where statistics meets forensics: match criteria, probative value
• A preview of coming attractions

Probability and Statistics Preliminaries
• Probability – the mathematical language of uncertainty
  o Basic rules
  o Conditional probability
• Bayes’ rule (a key to later discussion of Bayesian statistical inference)
• Bayes’ rule and the likelihood ratio
• Probability distributions
  o How probability can be used to model data values
  o Some examples
• Statistical Inference
  o Population / Sample
  o Role of probability distributions
  o Importance of data collection
  o Overview of inference procedures

Probability and Statistics for Forensic Science
• The forensic science problem in statistical terms
  o Evidence types (DNA, latent prints, firearms, ...)
  o A range of different types of variables
  o Critical issues in design of data collection and availability of population databases
• Use of basic statistical tools
  o Probability models for evidence measurements
  o Estimation, confidence intervals and testing in some settings
• Likelihood ratio approach
o The concept in simple cases (e.g. blood type or other discrete characters)
o Continuous data (e.g., trace evidence)
• Bayes vs frequentist inference
• What happens with more complex data (most pattern evidence?)

**PART II: Kevin McElfresh (FIU), Henry Swofford (DFSC)**

Communicating results of forensic analyses to jurors and law professionals

• 30 years of statistics in court, banding, binning, and calculating to a lay audience. What’s next? -- McElfresh
• Integrating Statistical Thinking and Methods into Practice - Latent Print Examination – Swofford

Open discussion, questions and feedback.
Evaluations for:
An Introduction to Statistical Thinking for Practitioners
Palm Beach County Sheriff’s Office
Palm Beach, FL – March 3, 2016

At the invitation of Dr. Cecelia Crouse, CSAFE staff Hal Stern, Alicia Carriquiry and Colin Lewis-Beck (PhD student) agreed to present a one-day course on statistical ideas relevant for forensic practitioners and applications of statistics in forensic examinations. Two additional presentations were included in the program: Henry Swofford (DFSC) discussed an algorithm to search and match latent prints, and Steven McElfresh (FIU) talked about challenges when presenting quantitative evidence to juries.

About 70 forensic practitioners from every crime lab in Florida and approximately 10 other individuals participated in the training. At the end of the day, an evaluation form was circulated so that participants could assess the content, the level of difficulty and the presentation of the training material. Fifty five participants completed an evaluation form. In what follows, we summarize the results of the evaluation.

The evaluation form asked about the four parts of the training:
• Probability and Statistics Preliminaries
• Probability and Statistics for Forensic Science
• Integrating Statistical Thinking and Methods into Practice – Latent Print Examination
• Communicating Results of Forensic Analysis to Jurors and Law Professionals.

The same three questions were posed in the four sections of the form:

Pace is:
1 = Too slow, 2 = Somewhat slow, 3 = Just right, 4 = Somewhat fast, 5 = Too fast

Mathematical level is:
1 = Too low, 2 = Somewhat low, 3 = Just right, 4 = Somewhat high, 5 = Too high

Presentation was:
1 = Not well organized, 2 = Somewhat dis-organized, 3 = Just right, 4 = Somewhat organized, 5 = Very well organized.

The histograms presented next summarize the responses to each of the three statements in each section of the evaluation form.
Probability and Statistics Preliminaries

Fig 1. Response frequencies for the three statements about Pace (left-most panel), Mathematical level (center panel) and Presentation (right-most panel).

Most participants (about 33) thought that the pace with which the probability and statistical preliminaries was presented was “just right”. There were, however, several participants (about 17) who thought that the pace was either “somewhat fast” or “too fast”. About 30 participants thought that the mathematical level was “just right”. About 20 participants, however, found that the mathematical level was too high. In terms of organization, the vast majority of participants stated that the presentation was “just right”, “somewhat well-organized” or “very well organized”.

Probability and Statistics for Forensic Science

Fig. 2: Response frequencies for the three statements about Pace (left-most panel), Mathematical level (center panel) and Presentation (right-most panel).
Results for this section are similar to those described above. In terms of mathematical difficulty, however, a larger proportion of respondents found the material to be too challenging.

**Integrating statistical thinking and methods into practice – Latent print examination**

![Graphs showing response frequencies for three statements about Pace, Mathematical level, and Presentation.](image)

Fig. 3: *Response frequencies for the three statements about Pace (left-most panel), Mathematical level (center panel) and Presentation (right-most panel).*

These responses refer to Henry Swofford’s presentation. Since this presentation was not made by CSAFE personnel, we show the summary just for information purposes.

**Communicating results of forensic analysis to jurors and law professionals**

![Graphs showing response frequencies for three statements about Pace, Mathematical level, and Presentation.](image)

Fig. 4: *Response frequencies for the three statements about Pace (left-most panel), Mathematical level (center panel) and Presentation (right-most panel).*

These responses refer to Steve McElfresh presentation. Since this presentation was not made by CSAFE personnel, we show the summary just for information purposes.
An Introduction to Statistical Thinking for Forensic Practitioners
Presented for the Palm Beach County Sheriff’s Office

Alicia Carriquiry
Hal Stern
Colin Lewis-Beck

3/5/2016
OUTLINE

- Part I - Statistical Preliminaries - pg. 3
  - review of probability concepts
  - review of statistical inference concepts
- Part II - Statistics for Forensic Science - pg. 62
  - significance test/coincidence probability
  - likelihood ratio
  - conclusions
Statistical Preliminaries
Outline and Overview

Outline
- probability
- statistical inference

Overview - “The Big Picture”
- Population = universe of objects of interest
  Sample = objects available for study
- Probability: population $\rightarrow$ sample (deductive)
- Statistics: sample $\rightarrow$ population (inductive)
- Often use both together
  1. build/assume model for population
  2. assess sample using model
  3. refine model; go back to step 2
Probability
Basic setup

- Experiment with uncertain outcomes
e.g., consecutive matching striae (CMS) for randomly selected bullets
- Sample space: list of all possible outcomes
e.g., for two randomly selected bullets $S = \{0, 1, 2, 3, 4, 5, 6, \ldots \}$
- Event - set of possible outcomes
  - simple events - obtain 6 matching striae in row
  - compound events - obtain greater than 15 matching striae
Probability
Interpretation

- Probability of an event is a number (between 0 and 1) assigned to event describing the likelihood it occurs
- Interpretations
  - long run frequency of occurrence of event (must be a repeatable experiment)
  - subjective belief of likelihood of an event (e.g., a suspect was present at a crime scene)
Probability
Axioms/basic laws of probability

- Consider an event $Y$
  - $0 \leq P(Y) \leq 1$
  - $P(Y) = 0$, event never happens
  - $P(Y) = 1$, event always happens

- Sum of probabilities of all possible outcomes is 1
- Define complement of an event $Y$ as the event that $Y$ does not happen, denoted as $\bar{Y}$
  - $P(Y) + P(\bar{Y}) = 1$
Probability
Adding probabilities

- Consider two events $Y$ and $S$
  - $P(Y \text{ or } S) = P(Y) + P(S) - P(Y \text{ and } S)$
  - if $Y$ and $S$ can’t happen together (they are mutually exclusive), then we get a simple addition formula

- Examples
  - $P(A \text{ or } B \text{ blood}) = P(A) + P(B)$
  - $P(RH^+ \text{ or } A \text{ blood}) = P(RH^+) + P(A) - P(RH^+ \text{ and } A)$
Probability
Conditional probability

- Consider two events $Y$ and $S$
- Suppose we know $S$ has occurred
- Does this tell us anything about $Y$?
- Define conditional probability
  - probability of event $Y$ given that another event $S$ has happened
  - $P(Y|S) = \frac{P(Y \text{ and } S)}{P(S)}$
- Note this gives a multiplication rule for probabilities, $P(Y \text{ and } S) = P(S) \times P(Y|S)$
- Tempting to think of conditional probability as being different than “ordinary” probability
- But all probabilities are conditional ... it is just a question of what information is assumed known
Probability
Conditional probability - example

- Consider the genetic markers Kell and Duffy
- Number of individuals in a population classified into four possible genetic marker groups

<table>
<thead>
<tr>
<th></th>
<th>Duffy+</th>
<th>Duffy−</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kell+</td>
<td>34</td>
<td>26</td>
<td>60</td>
</tr>
<tr>
<td>Kell−</td>
<td>36</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>30</td>
<td>100</td>
</tr>
</tbody>
</table>

- \( P(\text{have Kell}+ \text{ and Duffy}+) = 0.34 \)
- \( P(\text{have Duffy}+) = 0.70 \)
- Then

\[
P(\text{Kell}+ | \text{Duffy}+) = \frac{\Pr(\text{Kell}+ \text{ and Duffy}+)}{\Pr(\text{Duffy}+)} = \frac{0.34}{0.70} = 0.485
\]
Probability
Conditional probability - example

- What is the probability of seeing two known matching bullets (KM) and two known non-matching bullets (KNM) given the maximum number of consecutive matching striae (CMS)? Here we condition on the maximum number of CMS.

- As expected $P(KM)$ and $P(KNM)$ are not independent of the maximum number of CMS.
Probability
Conditional probability - example

- Add Glass Example
Probability

Independent vs. Dependent events

- Definition - if the likelihood of one event is not affected by knowing whether a second has occurred, then the two events are said to be independent (e.g., living in MW and driving a red car)

- Forensics example: if Y is event that glass matches and S is event that fiber matches, we may ask if these are independent

- Example of dependent events: two markers for breast cancer may be linked (if you have one, then you are more likely to have the other)

- Forensic example: connecting mitochondrial DNA results to physical characteristics of hair. Given a specific mtDNA sequence the probability an individual has certain hair traits increases.
Probability
Multiplying probabilities

- It follows from the definition of conditional probability that
  \[ P(A \text{ and } B) = P(B)P(A|B) = P(A)P(B|A) \]

- If \( A \) and \( B \) are independent, then
  \[ P(A \text{ and } B) = P(A)P(B) \]

- Example:
  - \( P(\text{left handed and from Florida}) \)
    \[ = P(\text{left handed}) P(\text{from Florida}) \]
  - If events are not independent, then
    \( P(\text{lung cancer and smoke}) \)
    \[ = P(\text{lung cancer} | \text{smoke}) P(\text{smoke}) \]
Probability
Independent events - example

- Consider blood type
  - triallelic genetic system
  - alleles denoted as A, B, O with $P(A) = p_1$, $P(B) = p_2$, $P(O) = 1 - p_1 - p_2$
  - individual gets one allele from mother and one from father
  - assume allele inherited from mother is independent of allele inherited from father
  - then can compute the probability of all six genotypes: AA, AB, BB, AO, BO, OO
  - $P(AA) = P(A)P(A)$, $P(AB) = P(A)P(B)$, ...
  - can also compute probability of a particular observed blood type (phenotype), e.g., AA and AO have phenotype A ... so
  - $P(A) = P(AA) + P(AO) = P(A)P(A) + P(A)P(O)$
Probability
Probabilities in court

- Let $S$ be event that suspect was present at the scene of a crime and $\bar{S}$ be event that the suspect was not present.
- Assume each juror has initial probability for each event.
- Witness says saw tall Caucasian male running from scene, defendant is tall Caucasian male:
  - jurors update probabilities.
- Window broken during the crime and fragments found on the defendant’s clothing match the broken window:
  - jurors update probabilities.
- How should jurors update their probabilities?
- Do jurors actually think this way?
Probability
Law of total probability

- Assume events $S_1, S_2, \ldots, S_n$ are mutually exclusive (two can’t happen together) and exhaustive (one of them must occur), then
  $P(S_1 \text{ or } \cdots \text{ or } S_n) = P(S_1) + \cdots + P(S_n) = 1$
- For example ($n = 2$), $P(S) + P(\overline{S}) = 1$
- Introduce a second event to see how updating might work, call it $R$
- Can write
  
  $P(R) = P(R \text{ and } S_1) + P(R \text{ and } S_2)$
  
  $= P(R|S_1)P(S_1) + P(R|S_2)P(S_2)$

  using law of total probability and definition of conditional probability
Probability
Law of total probability (cont’d)

- Example - DNA profiles from New Zealand
  - 83.47% of population is Caucasian \( Ca \), 12.19% of the population is Maori \( Ma \), and 4.34% of population is Pacific Islander \( Pa \)
  - let \( G \) be the event of finding the same YNH24 genotype as in a crime sample.
  - for a Caucasian, Maori, or Pacific Islander these probabilities are 0.012, 0.045, and 0.039, i.e. \( P(G|Ca) = 0.012 \)
  - can find \( P(G) \), the probability of finding the YNH24 genotype in a person sampled at random from the population
    \[
    P(G) = P(G|Ca)P(Ca) + P(G|Ma)P(Ma) + P(G|Pa)P(Pa)
    \]
    \[
    = 0.012 \times 0.8347 + 0.045 \times 0.1219 + 0.039 \times 0.0434 = 0.017
    \]
  - this type of calculation is key to updating probabilities
Probability
Bayes’ Rule (or Theorem)

- Bayes’ rule provides an updating formula for probabilities
- Formula

\[ P(S|R) = \frac{P(R \text{ and } S)}{P(R)} = \frac{P(R|S)P(S)}{P(R)} = \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|\bar{S})P(\bar{S})} \]

where last line is from law of total probability
Probability
Bayes’ Rule (or Theorem)

Example - diagnostic tests

- let $G$ denote the presence of gunshot primer residue and $T$ denote a positive test
- often know $P(T|G)$ (sensitivity) and $P(\bar{T}|\bar{G})$ (specificity) and $P(G)$ (prevalence)
- may want $P(G|T)$
- use Bayes rule $P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\bar{G})P(\bar{G})}$

key point is that, in general, $P(T|G) \neq P(G|T)$
Can sometimes get surprising results

Recall diagnostic test setup

- assume $P(T|G) = 0.98$ (sensitivity)
- assume $P(\bar{T}|\bar{G}) = 0.96$ (specificity)
- assume $P(G) = 0.9$ (prevalence)

  - then $P(G|T) = \frac{0.98 \times 0.9}{0.98 \times 0.9 + 0.04 \times 0.1} = 0.995$

- now assume $P(G) = 0.1$ (low prevalence)

  - $P(G|T) = \frac{0.98 \times 0.01}{0.98 \times 0.01 + 0.04 \times 0.99} = 0.20$

  - even if test is positive still have low probability of having gunshot residue
Probability
Bayes’ Rule to the likelihood ratio

- In forensic setting let \( S \) be same source and \( E \) be evidence and use Bayes’ rule to find

\[
P(S|E) = \frac{P(E|S)P(S)}{P(E)}
\]

- Slight detour here to introduce the concept of odds
  - odds in favor of event \( R \) are defined as
    \[
    O = \frac{P(R)}{P(\bar{R})} = \frac{P(R)}{1 - P(R)}
    \]
  - odds against event \( R \) are defined as
    \[
    O = \frac{P(\bar{R})}{P(R)} = \frac{1 - P(R)}{P(R)}
    \]
  - odds against and probabilities are related since \( P(R) = \frac{1}{(O + 1)} \)
    - if \( P(R) = .2 \), then odds against are \(.8/.2 = 4 \) (or “4 to 1”)
    - if \( P(R) = .8 \), then odds against are \(.2/.8 = .25 \) (or “1 to 4”)

(CSAFE) Statistical Thinking for Forensic Practitioners 3/5/2016 21 / 112
Probability
Bayes’ Rule to the likelihood ratio

Bayes rule can be rewritten in terms of odds

\[
P(S|E) = \frac{P(E|S) P(S)}{P(E|\bar{S}) P(\bar{S})}
\]

- left hand side: odds in favor of S given the evidence (E)
- last term on right hand side: odds in favor of S before evidence (E)
- first term on right hand side is known as the likelihood ratio
  - ratio of likelihood of evidence under S to likelihood of evidence under \( \bar{S} \)
- Likelihood ratio is the factor by which multiply prior odds of same source to get posterior odds of same source
- This is how probabilities get updated
Probability

Simple example of updating probabilities: Regina vs. Sally Clark

- Sally Clark's was the only person in the house when her first child died unexpectedly at 3 months. The cause of death was treated as sudden infant death syndrome (SIDS). A year later Sally had a second child, who also died at 2 months under similar circumstances.

- Sally was convicted of murder. During the trial a pediatrician testified that the probability of a single SIDS death was 1/8500, so the probability of two SIDS deaths was 1/73 million (1/8500^2).

- However, this simple analysis (which also assumes the deaths are independent) ignores the probability of murder.

- Consider the following two hypotheses:
  - S: The babies died of SIDS
  - M: Sally Clark murdered her children
Probability
Regina vs. Sally Clark (cont’d)

Using Bayes’s Theorem, we can calculate the posterior odds of the two hypotheses using the evidence (E) that the babies died and the prior odds of M and S

\[
\frac{P(M|E)}{P(S|E)} = \frac{P(E|M) \cdot P(M)}{P(E|S) \cdot P(S)}
\]

It is difficult to estimate the prior probability of two infants in the same household being murdered empirically. However, using crime data from the Office of National Statistics, a rough approximation is 1/8.4 billion. Note: the probability of E (evidence) under both hypotheses is 1

\[
\frac{P(M|E)}{P(S|E)} = \frac{(1/84 \text{ billion})}{(1/73 \text{ million})} = .009
\]

Expressing the odds in terms of probability, the posterior probability that the children died of SIDS rather than murder is .99
Probability
Collecting data

- Probability is of great interest to us in the way that it makes us think about data

- Where do data come from
  - sampling
  - experiment
  - convenient sample

- Types of data
  - qualitative
    - categorical (blood type: A,B,AB,0)
    - ordinal (eval of teacher: poor, avg, exc)
  - quantitative
    - discrete ( # of suspects in a case)
    - continuous (amount of silver in bullet frag.)
Probability

Probability distributions

- Suppose we are to collect data on some characteristic for a sample of individuals or objects (weight, trace element conc)
- Probability distribution is used to describe possible values and how likely each value is to occur
- Examples of distributions
  - Binomial: \# of success in n trials
  - Poisson: count \# of events
  - normal: bell-shaped curve
  - log normal: logarithm of observations follow a normal distribution
Probability

Probability distributions - normal

- Familiar bell-shaped curve
- Measurement error is often assumed to follow a normal distribution
- Described by two parameters, mean $\mu$ and standard deviation $\sigma$
- We write $X \sim N(\mu, \sigma)$
- 95% of values within 2 $\sigma$ of the mean and over 99% within 3 $\sigma$
- For $X \sim N(\mu, \sigma)$, $Z = (X - \mu)/\sigma \sim N(0, 1)$ is a standard normal and tables exist to compute probabilities for $Z$
- Example
  - consider a marker of DNA left at scene, $y$, and DNA taken from the suspect, $x$
  - let $d = y - x$: if DNA at scene came from the suspect, then $d$ is measurement error
  - so we may assume $D \sim N(0, \sigma)$
Probability
Probability distributions - normal

![Graph showing normal distributions](image_url)

- Standard normal distribution
- Normal distribution with mean 1 and standard deviation 1
- Normal distribution with mean 1 and standard deviation 2

(CSAFE) Statistical Thinking for Forensic Practitioners 3/5/2016 28 / 112
Probability

Probability distributions - normal

Glass Data Example Distribution
Probability
Probability distributions - lognormal

- Often act as if everything is normally distributed
- Of course this is not true
- Can't be exactly true for a quantity that is certain to be nonnegative (trace element concentration)
- In that case, may believe the logarithm of the quantity is normal, this gives a lognormal distribution for the quantity
- Specify lognormal distribution with two parameters, mean (on log scale) $\mu$ and standard deviation (on log scale) $\sigma$
Probability

Probability distributions - lognormal

Density

standard log normal
log normal (0, .5)
log normal (0, .25)
Probability
Probability distributions - discrete

- Some quantities take on very few values (discrete data)
- There are two popular discrete distributions
  - binomial
    - binary data (two categories)
    - $n$ independent 'trials'
    - $P(\text{success}) = p$ on each trial
    - expected number of successes = $np$
    - example: # correct answers in a cheating case
  - Poisson
    - count data (# events in a fixed time)
    - this distribution has mean = variance
    - variability increases with the mean
    - example: number of calls to 911
      between 10:00 and midnight on Friday nights
Probability distributions - discrete

- Distribution of the maximum number of CMS for a randomly selected bullet compared to 118 known lands follows a Poisson model.
A key point - mean of a sample and the mean of a population are different concepts

Definition - a parameter is a numerical characteristic of a population, e.g., a population mean

Statistical methods are usually concerned with learning about parameters

Idea: apply laws of probability to draw inferences from a sample
  - observe sample mean
  - if “good” sample this should be close to the population mean
  - probability and statistics tells us how close we can expect to be
Statistical Inference
Background

Recall - “The Big Picture”

- Population = universe of objects of interest
  Sample = objects available for study
- Probability: population $\rightarrow$ sample (deductive)
- Statistics: sample $\rightarrow$ population (inductive)
- Often use both together
  1. build/assume model for population
  2. assess sample using model
  3. refine model; go back to step 2
Statistical Inference
Example

What if we are interested in the average height of all Floridians?

- Population = all Florida residents
  Sample = everyone in this room
- We take the average height of everyone here and use this statistic to make inference about the population mean for all Floridians
- This assumes of course that our sample is a random sample from the population. This assumption may be questionable
Statistical Inference

Background

- Goal: inference about a parameter
- Possible parameters
  - mean
  - variance
  - proportion
- Different kinds of inferential statements
  - single estimate of parameter (point estimate)
  - range of plausible values for parameter (interval estimate)
  - perform test of hypothesis about parameter
Statistical Inference
Background

- Summarize qualitative data by recording frequency of various outcomes
- Summaries of quantitative data
  - mean or average
  - median = middle value
  - standard deviation (s.d.) = measure of spread
  - percentiles
- Example: if data = (19, 20, 21, 22, 23), then mean = median = 21, s.d. = 1.58
- Example: if data = (19, 20, 21, 22, 93), then mean = 35, median = 21, s.d. = 32.4
Statistical Inference
Point estimation

- Estimator is a rule for estimating a population parameter from a sample
- Evaluate estimator by considering certain properties
  - bias - how close on average to population value
  - variability - how variable is the estimate
- For population mean, might use sample mean as an estimator
  - no bias
  - low variability if sample is large
- Point estimates provide no measure of accuracy so are of limited use
Statistical Inference
Different estimators for $\theta$
Statistical Inference
Interval estimation

- Before talking about interval estimation need to introduce concept of a standard error.
- Recall that standard deviation is a measure of the spread (variability) in a sample or in a population.
- When we look at a summary statistic (mean, median, extreme value) it is also a random quantity (would give different value in different samples).
- Standard error of an estimate is how we measure its variability.
Statistical Inference

Interval estimation

- Example: population with mean 100 and s.d. 15
  - review meaning of s.d.: expect 95% of observations to be between 70 and 130
  - now suppose we compute mean of 25 responses
  - standard error is $\frac{15}{\sqrt{25}} = 3$
  - sample mean should be around 100
  - 95% of the time it will be between 94 and 106
How do we interpret this interval?

The Frequentist approach assumes there is a true fixed value of the mean. If we grab 100 samples and construct 100 confidence intervals, 95 of those will contain the population mean. Each interval either contains the population mean or it doesn’t. We hope our constructed interval is one of the 95 “‘good’” intervals. What is random is the confidence interval, not the true population mean.
Statistical Inference
Bayesian vs. Frequentist

Bayesians assume the population mean has a probability distribution. We have a prior belief about the mean, and given observed data, we can update our prior knowledge to obtain a posterior distribution.

Rather than make a confidence interval, Bayesians construct a credible 95% interval. The credible interval captures the uncertainty of the location of the mean. Moreover, it can be interpreted as a probabilistic statement about the location of the true mean rather than in terms of resampling, as in the frequentist approach.
Statistical Inference
Normal distributions and standard errors

- Plots of the normal distribution of an observation that has mean 100 and standard deviation 15
- On the same graph: the distributions of the means computed from samples of size 10, 25, and 50
- Note that distributions of means are tighter around the expected value as the sample size increases
Statistical Inference
Normal distributions and standard errors

mean = 100, sd = 15

mean = 100.04, sd = 3.09

mean = 99.85, sd = 2.98

mean = 100.02, sd = 2.01
Statistical Inference
Interval estimation

- A confidence interval is an interval based on sample data that contains a population parameter with some specified confidence level.
- Essentially a confidence interval takes a point estimate and then adds some information about uncertainty.
- Typically we get a 95% confidence interval for a quantity by taking point estimate ± 2 std errors.
Statistical Inference
Hypothesis testing

- Sometimes we wish to formally test a hypothesis about a population parameter.
- The hypothesis to be evaluated is known as the null hypothesis, usually means the status quo. We look for evidence against the null.
- There is an alternative (or research) hypothesis.
- Test procedure is described on the next slide.
- If we reject the null hypothesis then we say we have a statistically significant result.
Statistical Inference
Hypothesis testing

- Two types of errors
  - type I: reject the null hypothesis when it is true
  - type II: fail to reject the null when it is false
- Type I error often considered more serious: only reject null if strong evidence against it
- For juries: null hypoth=innocent, alternative=guilty
  - type I error is to say guilty when innocent
  - type II error is to say innocent when guilty
Statistical Inference
Hypothesis testing

- Basic idea of hypothesis testing is to compute a test statistic that measures ‘distance’ between the data we have collected and what we would expect under the null hypothesis.
- Typically use a statistic of the form 
  \[(\text{point estimate} - \text{null hypothesis value})/\text{SE(estimate)}\]
  where \(SE\) is a standard error.
- Can be interpreted as the number of standard errors the sample estimate is from the hypothesized value under the null hypothesis.
Statistical Inference
Hypothesis testing

- Summarize test by attaching a probability to the test statistic
- Definition: a \( p \)-value gives the probability that we would get data like that we have in the sample (or something even more extreme) given that the null hypothesis is true
- Small \( p \)-values mean unusual data that lead us to question the null hypothesis (since sample data are unlikely to happen by chance)
- However, the \( p \)-value only addresses the null hypothesis. It does not speak to the likelihood of the alternative hypothesis being true
Statistical Inference
Procedures for normal data

- Most well established procedures are those related to drawing conclusions about the mean of a normal population.
- Suppose that we have acquired a random sample of $n$ observations from the population.
- Natural point estimate of population mean is the sample mean $\bar{X} = \sum X_i / n$.
- 95% confidence interval is: $\bar{x} \pm 1.96 \times SE(\bar{x})$.
  - $SE(\bar{X})$ is the standard error of the sample mean.
  - $SE(\bar{X}) = SD(\text{population})/\sqrt{n}$.
- Can test a hypothesis about $\mu$ say $H_0 : \mu = \mu_0$ using the test statistic,
  $$t = \frac{\bar{x} - \mu_0}{se(\bar{x})}$$
  where $p$-value is obtained from a $t$ table.
- Key result is that these procedures work well even if population is not normally distributed as long as the sample size is large.
Statistical Inference
Procedures for normal data
Example: Suppose nationally, 35% of all athletic shoes are Nike (\( p = .35 \)). We want to test if this population proportion holds in East LA. We take a random sample of 100 individuals, and 42 are wearing Nike’s (\( \hat{p} = .42 \)). Can we still use the national proportion as a reference or is East LA different?

By the CLT, \( \hat{p} \) has a normal distribution with mean \( p \) and standard deviation \( \sqrt{\frac{p(1-p)}{n}} \)

\[
z = \frac{(\hat{p} - p)}{\sqrt{\frac{p(1-p)}{n}}} = \frac{(0.42 - 0.35)}{\sqrt{\frac{0.35(1-0.35)}{100}}} = 1.47
\]

p-value = .0705. Thus, using an alpha level of .05, we fail to reject the null hypothesis that the reference proportion holds in East LA.
**Statistical Inference**

Procedures for normal data - example

- Want to estimate the mean amount of a trace element for population = all bullets in Southern Iowa
- Get a random sample of 500 bullets from the area
  - sample mean is 55, standard deviation is 22
  - standard error of the mean is \( \frac{22}{\sqrt{500}} = 0.98 \)
- 95% confidence interval: \( 55 \pm 1.96 \times 0.98 = (53.1, 56.9) \)
- Suppose we have reason to believe that Remington (mean = 58) is main producer in this area, can check with hypothesis test
  - hypothesis test: \( H_0 : \mu = 58 \)
  - test statistic, \( t = \frac{(55 - 58)}{0.98} = 3.06 \)
  - estimate is 3 standard errors from the mean under the null hypothesis
  - \( p \)-value = 0.00135: if null hypothesis is true, then observe a value 3 standard errors or more .1% of the time, hence reject null hypothesis
Statistical Inference
Comparing two means

- Previous discussion addresses methods for one sample
- In practice most often interested in comparing two samples (or more precisely two populations)
- Assume random samples from each of the two populations are available
- Test for relationship between parameters of the two populations
- Forensic example
  - suppose we have broken glass at the scene and glass fragments on the suspect
  - define $\mu_{scene}$ to be mean trace element level for the “population” of glass on the scene
  - define $\mu_{suspect}$ to be the mean element level for “population” of glass on the suspect (what population?)
  - compare means to address if glass fragments could plausibly have come from scene
Statistical Inference
Comparing two means

- Continue discussion using terminology from forensic example
- Take null hypothesis to be that evidence matches, $\mu_{\text{scene}} = \mu_{\text{suspect}}$
- Take alternative hypothesis to be that the evidence does not match, $\mu_{\text{scene}} \neq \mu_{\text{suspect}}$
- Note: this hypothesis test runs counter to the earlier null of assuming the evidence is not same source.
Equivalence testing assumes the population means are different, and uses data to test if the means are actually the same

\[ H_0 : |\mu_{\text{scene}} - \mu_{\text{suspect}}| > \Delta \]
\[ H_A : |\mu_{\text{scene}} - \mu_{\text{suspect}}| < \Delta \]

\( \Delta \) is set to capture "practical" significance, which varies depending on the data.

Once again, we reject the null hypothesis that the two means are different if we get a \( p \)-value less that a specified \( \alpha \) level for both hypothesis tests.
Suppose 10 glass fragments are taken from glass at the scene and 9 fragments are found on the suspect. Based on previous research a difference of 1 is considered significant

\[ \bar{X} = 5.3, \text{s.d.} = 0.9, \ SE(\bar{X}) = 0.9/\sqrt{10} = .28 \]

\[ \bar{Y} = 5.9, \text{s.d.} = 0.85, \ SE(\bar{X}) = 0.85/\sqrt{9} = .28 \]

Then \( SE(\text{diff } \bar{X} - \bar{Y}) = \sqrt{.28^2 + .28^2} = .4 \)

Test statistic: \[ t_1 = \frac{(\bar{X} - \bar{Y}) - 1}{SE} = -4, \ p\text{-value} < .001 \]

Test statistic: \[ t_2 = \frac{(\bar{X} - \bar{Y}) - (-1)}{SE} = 1, \ p\text{-value} = .317 \]

Therefore, we fail to reject the null hypothesis that the mean trace levels are different between the scene and suspect.
A recent study used 2 dimensional profile plots to compare bullet striation marks. Aligning the striation marks from two different bullets, the number of consecutive matching peaks and valleys were used to determine if the bullets came from the same gun. This measure is defined as consecutive matching striae (CMS).
The distributions of CMS for known matching and non-matching bullets are not normal, but we can compare the two distributions empirically to determine how well the variable CMS correctly identifies matching bullets.
Statistics for Forensic Science
Part II - Outline

- Brief review of probability/statistics
- The forensic examination
- Significance test/coincidence probability approach
- Likelihood ratio approach
- Conclusions

Brief Review of Probability and Statistics

- **Probability**
  - language for describing uncertainty
  - assigns number between 0 and 1 to events
  - depends on information available
    (information conditioned upon)
  - used to deduce likely values for individuals or samples from given (or hypothesized) information about the population

- **Probability distributions**
  - suppose we have a random quantity
    (e.g., trace element concentration in a glass fragment)
  - probability distribution gives possible values and relative likelihood of each value
Brief Review of Probability and Statistics

Statistics

- drawing inferences about a population (i.e., learning about some characteristic of the population) based on sample data
- need to carefully define “population”
- method used for data collection is very important
- there are a variety of inference procedures
  - point estimates
  - confidence intervals
  - hypothesis tests
The Forensic Examination

- 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., race of criminal, evidence substrate)

- Two hypotheses
  - $S$ - items from crime scene and suspect have common source
    (or suspect is source of crime scene item)
  - $\bar{S}$ - no common source

- Goal: assessment of evidence
  - do items have a common origin
  - how unusual is it to find items of a common origin like these
The Forensic Examination

- Evidence types
  - biological evidence (blood type, DNA)
  - glass fragments
  - fibers
  - latent prints
  - shoe prints / tire tracks
  - and others

- A key point is that different issues arise for different evidence types
  - discrete/continuous variables
  - information about the probability distribution of measurements
  - existence of reference database
  - role of manufacturing process
The Forensic Examination

- Daubert factors relevant for establishing validity of scientific testimony
  - theory/method should be testable
  - subject to peer review / publication
  - error rates
  - existence of standards and controls
  - generally accepted by a relevant scientific community
The Forensic Examination

- National Research Council (2009) findings
  - heterogeneous provider community (federal, state, local)
  - heterogeneity across disciplines
  - lack of standardization in practices
  - insufficient resources
  - questions underlying scientific basis for some conclusions

- Impact of quantification/reliability of DNA evidence on other disciplines
The Forensic Examination

- A community in transition
  - National Commission on Forensic Science
    (advisory to Attorney General)
  - Organization of Scientific Area Committees (NIST)
  - Research
    - NIJ-funded projects
    - NIST Forensic Sciences
    - NIST Center of Excellence
      (CSAFE = Center for Statistical Applications in Forensic Evidence)
Reliability and Validity

- Focus today is on statistical methods, a basic understanding and how they can be applied in a forensic context.
- In many cases there are no commonly accepted approaches to quantifying the evidence.
- A common question is what should forensic practice look like now, while novel methods are being explored.
- Daubert factors can present some insight:
  - studies of reliability and repeatability (same examiner, different examiners).
  - studies of validity (require gold standard).
  - examples include “black box” and “white box” studies of latent print examiners.
- Statistical methods are relevant here as well:
  - design of experiments (importance of randomization, blinding, etc.).
Common Approaches to Assessing Forensic Evidence

- Expert assessment
- Statistical approaches
  - Statistical-test based approach
  - Likelihood ratio
- Ultimately perhaps a combination of the two?
Signif. Test/Coincidence Prob. Approach

- One common statistical approach solves the forensic problem in two stages
- Step 1: determine if the crime scene and suspect objects “match” (typically using a significance test)
- Step 2: assess the significance of the match by finding the likelihood of a match occurring merely by coincidence
- Also known as the comparison/significance approach
- Note: DNA analysis can be categorized in this way but is usually thought of as a likelihood ratio approach
Signif. Test/Coincidence Prob. Approach

- Determining a match is straightforward for discrete data like blood type, gender
  - note that there are still limitations - laboratory or measurement error
  - usually more straightforward to think about discrete data in terms of the likelihood ratio
- Focus on continuous data for remainder of match discussion
- Statistical significance tests can be used for continuous data like trace element concentrations (e.g., in glass fragments)
Signif. Test/Coincidence Prob. Approach

Testing procedure
- characterize each object(s) by mean value (e.g., mean trace element concentrations in set of glass fragments)
- this is the “population mean” in statistics terminology (one for crime scene, one for glass on suspect)
- obtain sample values from crime scene object
- obtain sample values from suspect’s object
- use sample values to test hypothesis that two objects have the same mean
- common tool is Student’s \( t \)-test
- summary is \( p \)-value, probability of data like the observed data, assuming population means are the same
- small \( p \) (less than .05 or .01) rejects “match”
- otherwise accept
  (.... but is this evidence of a match?)
Example 1: Two glass samples (from Curran et al. 1997)

Five measurements of aluminum concentration in control

\[ 0.751, 0.659, 0.746, 0.772, 0.722 \]

Five measurements of aluminum concentration in recovered sample

\[ 0.752, 0.739, 0.695, 0.741, 0.715 \]

Control: mean = 0.730, std.err. = \( \frac{0.0435}{\sqrt{5}} = 0.019 \)

Sample: mean = 0.728, std.err. = \( \frac{0.0230}{\sqrt{5}} = 0.010 \)

Test statistic = \( \frac{0.730 - 0.728}{\sqrt{0.019^2 + 0.010^2}} = \frac{0.002}{0.0215} \approx 0.1 \)

\( p \)-value = 0.70 ..... no reason to reject hypothesis of equal means

In fact, these are 10 measurements from same bottle
Example 2: Two glass samples (from Curran et al. 1997)

Five measurements of aluminum concentration in control

\[ .751, .659, .746, .772, .722 \]

Five measurements of aluminum concentration in recovered sample

\[ .929, .859, .845, .931, .915 \]

Control: mean = .730, std.err. = \[ \frac{.0435}{\sqrt{5}} \] = .019

Sample: mean = .896, std.err. = \[ \frac{.0408}{\sqrt{5}} \] = .018

Test statistic = \[ \frac{.896 - .730}{\sqrt{.019^2 + .018^2}} \] = \[ \frac{.166}{.026} \] = 6.38

\( p \)-value = .00015 ..... reject hypothesis of equal means

In fact, these are from two different bottles
Signif. Test/Coincidence Prob. Approach

- Alternative methods exist
  - 3-sigma methods create interval for each element in each sample (mean conc. +/- 3 standard errors) and check for overlap
  - range overlap uses "control" sample to obtain an expected range and check with "test" samples are in/out of range
  - Hotelling’s $T^2$ test compares all elements simultaneously (take account of dependence)
Signif. Test/Coincidence Prob. Approach

- Before moving to assessing the probability of a coincidental match it is important to understand some objections to testing
  - significance tests do not treat the two hypotheses (equal mean, unequal mean) symmetrically
    - null hypothesis (equal mean) is assumed true unless the data rejects
    - acceptance of null is taken as evidence against suspect which seems to run counter to usual assumption in our justice system
  - binary decision of match/no-match requires an arbitrary cutoff (e.g., why 3 sigma rather than 2.5 sigma)
  - separation of match decision from assessment of the significance of the match
Signif. Test/Coincidence Prob. Approach

- Some more technical objections
  - many test procedures (t-test, Hotelling’s test) require assumptions about the probability distribution of the data
  - univariate procedures are repeated on multiple elements which must be accounted for
  - univariate procedures ignore information in the correlation of elements
  - multivariate procedures require large samples
Signif. Test/Coincidence Prob. Approach

- Alternatives to significance tests
  - Equivalence testing instead of significance testing (changes the null hypothesis)
  - Bayesian methods and the likelihood ratio (more on this later)
Signif. Test/Coincidence Prob. Approach

- Second stage of analysis is assessing the “significance” of a match
- Because significance has a formal statistical meaning we try not to use that term here
- Other terms - strength of evidence, quality of evidence, usefulness of evidence, probative value
- Examples: we know that suspect and criminal . . . are both male - limited usefulness
  . . . are both one-armed males - more useful
- Key idea is to determine probability of a match by coincidence
- This step is crucial for courtroom setting
Signif. Test/Coincidence Prob. Approach

- Discrete data (e.g., blood type, DNA)
  - want to find probability of a match by chance
  - several important considerations
    - usually crime-scene centered: material from scene is considered fixed and want likelihood that individual would have similar object
    - depends on relevant “information” (suspect is male, suspect is Chinese, etc.)
    - where do data come from (population records, convenience sample)
    - discussed further in the context of likelihood ratios later
Signif. Test/Coincidence Prob. Approach

- Continuous data
  - typically a bit harder to do
  - need likelihood that objects (e.g., glass fragments) selected at random would match “control” sample
  - basic idea (described in terms of $t$-test)
    - suppose we know the ”population” mean of randomly chosen object
    - can find probability that $t$-test based on a sample from this random object will result in match to given “control” sample
    - so if we can specify a probability distribution for mean value of randomly chosen object, then total coincidence probability is an average

\[
\text{C} = \sum_{\text{means}} \Pr(\text{mean}) \Pr(\text{match} \mid \text{mean})
\]

- this is challenging but it can be done
Signif. Test/Coincidence Prob. Approach

- Take $\bar{X} = .730, \sigma = .04, n = 5$ as in Curran et al. glass example
- Use $c = 1.96$ as cutoff ($p$-value roughly .05)
- Suppose that different glass samples in population can be described as a normal distn
- Results
  - if mean is .73 and s.d. .20 then .... .20
  - if mean is .73 and s.d. .10 then .... .37
  - if mean is .73 and s.d. .05 then .... .65
  - if mean is .83 and s.d. .20 then .... .37
  - if mean is .83 and s.d. .10 then .... .24
  - if mean is .83 and s.d. .05 then .... .17
  - if mean is .93 and s.d. .20 then .... .12
  - if mean is .93 and s.d. .10 then .... .06
  - if mean is .93 and s.d. .05 then .... .002
Signif. Test/Coincidence Prob. Approach

- Probability of a coincidental match (i.e., related to false match probability) is high when:
  - small difference between control sample and population of samples we are likely to find
  - large amount of heterogeneity among samples in the population
  - large amount of variability in the measurement process
Likelihood Ratio Approach

Introduction

- Real goal in courtroom setting is statement about relative likelihood of two hypotheses (prosecution, defense) given data
- In statistical terms this is a Bayesian formulation (asks for probabilities about the state of the world given observed data)
- Review of Bayes’ rule (or Bayes’ Theorem):
  Given two events $A$ and $B$ we have

$$
\Pr(A|B) = \Pr(A) \cdot \frac{\Pr(B|A)}{\Pr(B)}
$$

- This is a way of reversing direction of conditional probabilities ... we go from statements about likelihood of evidence given hypothesis to statements about likelihood of hypothesis given evidence
Likelihood Ratio Approach

Introduction

- Formally using $E$ (evidence) and $S$ (same source)

$$
Pr(S|E) = \frac{Pr(S)Pr(E|S)}{Pr(E)} = \frac{Pr(S)Pr(E|S)}{Pr(S)Pr(E|S) + Pr(\bar{S})Pr(E|\bar{S})}
$$

- Bayes’ rule can be rewritten in terms of the odds in favor of the same source hypothesis

$$
\frac{Pr(S|E)}{Pr(\bar{S}|E)} = \frac{Pr(E|S)}{Pr(E|\bar{S})} \times \frac{Pr(S)}{Pr(\bar{S})}
$$

- In words

  Posterior odds = Likelihood ratio $\times$ Prior odds

- Thus likelihood ratio is a (“the”?) measure of the value of the evidence (does not depend on prior beliefs)
Likelihood Ratio Approach

Introduction

- Reminder: \[ LR = \frac{\Pr(E|S)}{\Pr(E|\overline{S})} \]

- Some observations
  - 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 (more on this later)
  - Other available information (e.g., background) can be incorporated into the LR (more on this later)

- Interpretation
  - Derivation shows that LR is factor we should use to change our odds
  - There are some proposals for scales (e.g., ENFSI) that map LRs to words:
    - 2-10: weak support; 10-100: moderate support; etc.
Likelihood Ratio Approach

Introduction

Reminder: \( LR = \frac{\Pr(E|S)}{\Pr(E|\bar{S})} \)

Some observations

- numerator assumes common source and asks about the likelihood of the evidence in that case
  - somewhat related to finding a \( p \)-value for testing the hypothesis of equal means
  - but ... no binary decision regarding match
  - instead a quantitative measure of likelihood of evidence under \( S \)
- denominator assumes no common origin and asks about the likelihood of the evidence in that case
  - analogous to finding coincidence probability
  - here too, doesn’t require a binary decision regarding match
  - a quantitative measure of likelihood of evidence under \( \bar{S} \)
Likelihood Ratio Approach
Introduction

- There is some subtlety here ...
  - Prosecutor’s fallacy: interpreting $Pr(E|\bar{S})$ as $Pr(\bar{S}|E)$
    - Evidence is unlikely under $\bar{S}$ is interpreted as saying that $\bar{S}$ is unlikely
  - Defense attorney’s fallacy: other misinterpretations of $Pr(E|\bar{S})$
    - If $Pr(E|\bar{S}) = 1/1000000$, then there are 300 other people who could have done it
Likelihood Ratio Approach
Some notes on implementation

- Reminder: $LR = \frac{Pr(E|S)}{Pr(E|\bar{S})}$
- Assume $E = (x, y)$ where $x$ is measurement of evidence from crime scene and $y$ is measurement of evidence from suspect
- Replace $Pr()$ by $p()$ to cover discrete and continuous cases
- Then we have from laws of probability

$$LR = \frac{p(x, y|S)}{p(x, y|\bar{S})} = \frac{p(y|x, S) p(x|S)}{p(y|x, \bar{S}) p(x|\bar{S})}$$

- Often likelihood of $x$ is same for $S$ and $\bar{S}$, (i.e., $p(x|S) = p(x|\bar{S})$) e.g., trace element measurements for glass at crime scene don’t depend on who committed the crime
- If so ...... $LR = \frac{p(y|x,S)}{p(y|x,\bar{S})}$
Likelihood Ratio Approach
Some notes on implementation

Assume we can start with

\[ LR = \frac{p(y|x, S)}{p(y|x, \bar{S})} \]

- **Discrete case**
  - Numerator is typically one or zero
    (Does zero really mean zero? Should consider lab error or other explanations)
  - Denominator is probability of a coincidental match
- **Continuous case**
  - Numerator is measure of how likely it is to observe the numbers \( x, y \) if they represent multiple measures from the same source
  - Denominator is a measure of how likely it is to observe the numbers \( x, y \) if measures from different sources
Likelihood Ratio Approach
A simple example

- Suppose evidence is blood types for a crime scene sample and suspect sample.
- We have information about the distribution of blood types in the population:

<table>
<thead>
<tr>
<th>Type</th>
<th>A</th>
<th>B</th>
<th>AB</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Freq</td>
<td>.42</td>
<td>.10</td>
<td>.04</td>
<td>.44</td>
</tr>
</tbody>
</table>

- Suppose both samples are observed to be of blood type O.
- Pr(E|S) ≈ 1 (we’d expect to see the same blood type if S is true).
- Pr(E|¬S) = 0.44 (type O blood is relatively common in the U.S.).
- LR ≈ 1/0.44 ≈ 2.3
- Evidence provides weak support for “same source.”
- Blood type AB is rare in the U.S.. If the two samples were both AB, then LR would indicate stronger evidence.
Likelihood Ratio Approach
Where it works ..... DNA

- A DNA profile identifies alleles at a number of different locations along the genome (e.g., TH01 is 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

<table>
<thead>
<tr>
<th>TH01</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>9.3</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>.001</td>
<td>.001</td>
<td>.266</td>
<td>.160</td>
<td>.135</td>
<td>.199</td>
<td>.200</td>
<td>.038</td>
<td>.001</td>
</tr>
</tbody>
</table>

and then combine the results
- Often end up with very large numbers
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 from elemental analyses of bullet lead
Likelihood Ratio Approach
Where it can work .... Trace evidence

One-dimensional plots
Likelihood Ratio Approach
Where it can work .... Trace evidence

Three projections of 5-dimensional trace element data
Likelihood Ratio Approach
Continuous measures for trace data - example

- Take $x$ and $y$ to be trace element concentrations for a single element from glass fragments at the scene ($x$) and on the subject ($y$)
- Can generalize to multiple $x, y$ measurements
- Assume normal distn for trace element concentrations (may be more reasonable for logarithms)
- Within a single source (e.g., sheet of glass) assume that $x \sim N(\mu_s, \sigma^2)$
  where $\mu_s$ is the mean concentration in the sheet and $\sigma^2$ is the (presumably small) variance
- Two different sheets will have different $\mu_s$’s
- Assume sheet-to-sheet variation is described by normal distribution, $N(\mu, \tau^2)$, where $\mu$ is a “manufacturer” mean and $\tau^2$ is sheet-to-sheet variance
- Implication of the two normal assumptions: distn of a random fragment is $N(\mu, \tau^2 + \sigma^2)$
Likelihood Ratio Approach
Continuous measures for trace data - example (cont’d)

The following is not precise (ignores transfer, makes some simplifying approximations)

Numerator $p(y|x, S)$
- assume $\sigma^2$ is known
- use $x$ (or avg of $x$’s) to estimate $\mu_s$
- $p(y|x, S) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left\{-\frac{(y - x)^2}{2\sigma^2}\right\}$
- not exact (treats $x$ as measured precisely)

Denominator $p(y|x, \bar{S})$
- assume $\sigma^2$ and $\tau^2$ are known
- assume $\mu$ (company mean) is known, otherwise use $x$ to estimate
- $p(y|x, \bar{S}) = \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\left\{-\frac{(y - \mu)^2}{2(\sigma^2 + \tau^2)}\right\}$
Likelihood Ratio Approach
Continuous measures for trace data - example (cont’d)

\[
LR = \sqrt{1 + \frac{\tau^2}{\sigma^2}} \exp \left\{ \frac{(y - \mu)^2}{2(\sigma^2 + \tau^2)} - \frac{(y - x)^2}{2\sigma^2} \right\}
\]

- LR is small if \(y - x\) is large (no match)
- LR is big if \(y - \mu\) is large and \(y - x\) is small (a match on an unusual value)
- Evett’s work supplies values of parameters for refractive index, typical LR are of order of magnitude 100
- LR idea works in same basic way for other non-normal distributions
Likelihood Ratio Approach
Where it can might 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 ...
  - Likelihood ratios can be very sensitive to assumptions
  - Assessing the ”population” is hard (and may vary from case to case)
Likelihood Ratio Approach
Where it might work .... Pattern evidence

- Many forensic disciplines are focused on comparing a sample (mark) at the crime scene ("unknown” or "questioned") and a potential source ("known”)
- Need to assess whether two samples have same source or different source
- Many examples
  - Latent print examinations
  - Shoe prints and tire tracks
  - Questioned documents
  - 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
  - Data are very high dimensional (often images)
  - Flexibility in defining the numbers/types of features to look at
  - Lack of probability models for multivariate features / patterns
  - Need to study variation across a relevant population

- Very hard, but there is work under way
  - e.g., Neumann et al. (2015) on fingerprints
Score-based likelihood ratios

- Currently there is some interest in an empirical approach
- Define a score measuring the "difference" between the questioned and known sample
- Obtain distribution of scores for a sample of known matches
- Obtain distribution of scores for a sample of known non-matches
- Use these score distributions to obtain a likelihood ratio
Likelihood Ratio Approach

Complications

- It is not hard to find issues that complicate the complications
  - accounting for transfer evidence with glass or fibers
  - bullets in a box are heterogeneous (containing representatives from a number of different manufacturing lots)
  - usage/lifetime of products (e.g., sneakers)
- Understanding likelihood ratios can be a challenge for non-statisticians, including juries
Likelihood Ratio Approach

Contextual information

- There is some discussion in the community about the role of contextual bias, task-relevant/task-irrelevant information, etc.
- Likelihood ratio framework can accommodate this discussion
- \( LR = \frac{\Pr(E|S, I)}{\Pr(E|\bar{S}, I)} \)
  where \( I \) refers to other information that is being considered
- \( I \) should include task-relevant information (e.g., substrate)
- \( I \) should not include results of other forensic examinations, other case information
Likelihood Ratio Approach

Multiple sources of evidence

- Likelihood ratio can accommodate multiple types of evidence
- \[ LR = \frac{Pr(E_1, E_2|S)}{Pr(E_1, E_2|\bar{S})} \]
  where \( E_1, E_2 \) are two evidence types
- If evidence types are independent, then this simplifies considerably
  \[ LR = \frac{Pr(E_1|S)}{Pr(E_1|\bar{S})} \times \frac{Pr(E_2|S)}{Pr(E_2|\bar{S})} \]
- If dependent, then the joint analysis can be tricky
Likelihood Ratio Approach
ENFSI Guideline for Evaluative Reporting

- Reporting requirements
  - balance - should consider two propositions
  - logic - should focus on likelihood of evidence given hypotheses
  - robustness - should withstand scrutiny
  - transparency - clear case file and report

- Propositions (different kinds, etc.)

- Assignment of likelihood ratio
  - data and/or expert knowledge used to assign probabilities required for likelihood ratio
  - subjective elements can be used
  - avoid undefined qualifiers (e.g., rare)
  - account for uncertainty

- LR forms basis for evaluation (verbal equivalents)
Likelihood Ratio Approach

Summary

- **Advantages**
  - explicitly compares the two relevant hypotheses/propositions
  - provides a quantitative summary of the evidence
  - 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

- **Disadvantages**
  - requires assumptions about distributions
  - 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
Conclusions

- Quantitative analysis of forensic evidence requires use of statistical methods
- Reviewed testing-based approaches and likelihood ratios
- Any approach must account for the two (or more) competing hypotheses for how the data was generated
- Need to be explicit about reasoning and data on which reasoning is based
- None of today’s statistical presentation addresses how to testify
  - Two presentations to follow
  - A great deal of discussion about reporting conclusions in the OSAC