New Approaches to the Quantification of Trace Evidence for Source Identification

Danica Ommen, Ph.D. Candidate
South Dakota State University

Christopher Saunders, Ph.D.
South Dakota State University

JoAnn Buscaglia, Ph.D.
FBI Laboratory

Outline

Acknowledgments and Collaborators

Cedric Neumann, South Dakota State University

Joshua R. Dettman, ORISE Visiting Scientist FBI Lab

Alyssa Cassabaum, ORISE Visiting Scientist FBI Lab

Deanna Synder, ORISE Visiting Scientist FBI Lab

Don Gantz, George Mason University

John Miller, George Mason University

Mark J. Lancaster, IC Post Doctoral Fellow (2006-2008)

Disclaimer

The research detailed in this presentation was supported in part by the National Institute of Justice, Office of Justice Programs, US Department of Justice under Award No. 2009-DN-BX-K234 and Award No. 2014-IJ-CX-K088.

This research was supported in part by an appointment to the Visiting Scientist Program at the Federal Bureau of Investigation Laboratory Division, administered by the Oak Ridge Institute for Science and Education, through an interagency agreement between the US Department of Energy and the Federal Bureau of Investigation.

The opinions and conclusions or recommendations expressed in this presentation are those of the author and do not necessarily represent those of the Department of Justice, the Federal Bureau of Investigation, or the US Department of Energy.

Outline

Example motivation

An improvised explosive device(s) (IED) is found.

A roll of copper wire is found at a suspect's house.

Is the copper wire used in the IED indistinguishable from the wire found at the suspect's house?

Can the copper wire in the two different IEDs be attributed to the same source?

Copper Mining and Wire Production

Mine

↓

Pre-treatment

↓

Initial refining

↓

Electrolytic refining

Cathode copper, >99.99% Cu

↓ ↓ ↓ ↓

Sheet/Strip Tubes Bars/Sections Rod

↓

Wire

Cathode copper
Average 6800 lbs/charge (bundle)
(picture: Freeport McMoRan Copper and Gold Inc.)

Copper Mining and Wire Production

Mine

↓

Pre-treatment

↓

Initial refining

↓

Electrolytic refining

Cathode copper, >99.99% Cu

Melted, two refining processes mixed

↓ ↓ ↓ ↓

Sheet/Strip | Tubes | Bars/Sections | Rod

↓

Wire

Furnace

Copper Mining and Wire Production

Mine

↓

Pre-treatment

↓

Initial refining

↓

Electrolytic refining

↓

Sheet/Strip Tubes Bars/Sections Rod

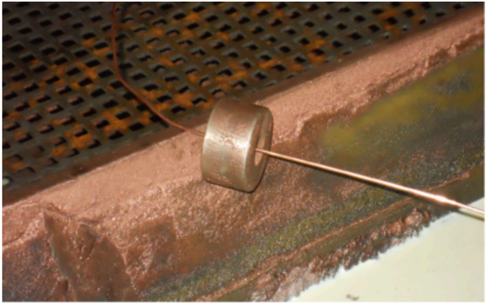
↓

Wire

Rod coils

Copper Mining and Wire Production

Mine
↓
Pre-treatment
↓
Initial refining
↓
Electrolytic refining
↓
Sheet/Strip Tubes Bars/Sections Rod
↓
Wire



Stretch rod using dies to form wire

Analytical Methodology

Preparation: dissolve copper into solution

Analysis: Inductively coupled plasma mass spectrometry (ICPMS)

- Standard method for trace element analysis
- Eight elements *(Ag, As, Bi, Co, Ni, Pb, Sb, Se)* were found at concentrations greater than the quantitation limit and consistently measured with good precision.
- Three separate pieces of each sample were carried through the preparation process and analyzed by ICPMS.

Validation: used NIST certified reference materials

Data Collection

Within-source Copper Samples

- Three wires, each 70 feet and sampled every 5 feet

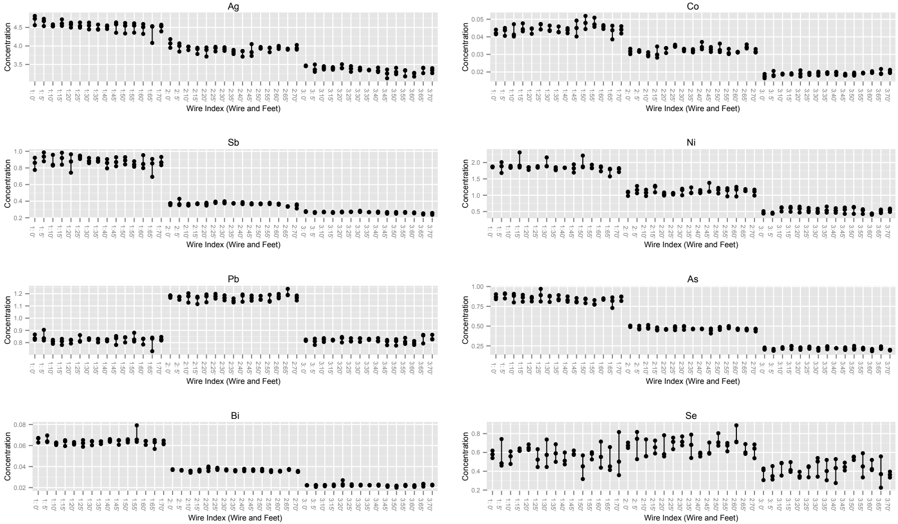
Between-source Copper samples

- One production sample every 90 minutes for 10 days

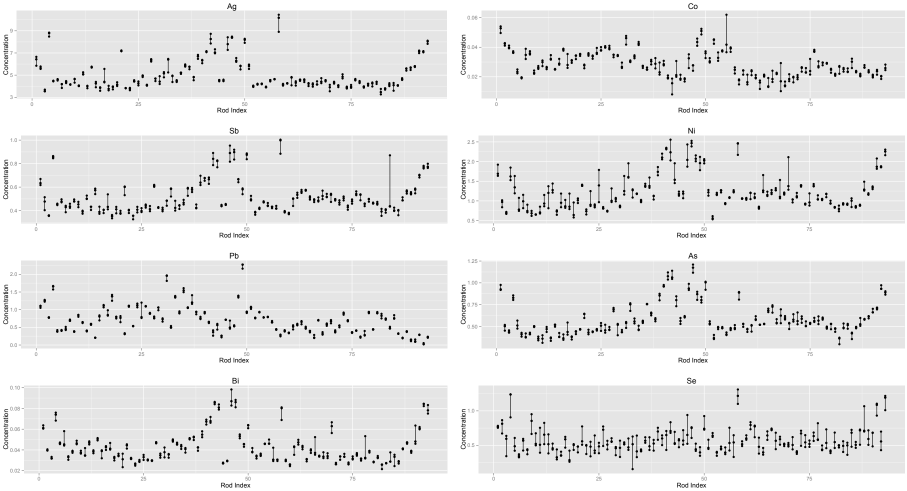
The datasets we have collected are exploratory in nature.

The purpose of their collection is to facilitate the study of the copper manufacturing process, with a focus on the development of new forensic analytical techniques.

Wire Sample Concentrations



Rod Sample Concentrations



The Problem

We have:

a collection of 50 feet of 12AWG (American Wire Gauge) copper wire for which we collect samples every 5 feet,

a second collection of 20 feet of 12AWG copper wire for which we collect samples every 5 feet,

and

the 93 samples collected from the copper Rods.

We would like to determine if the two collections of copper wire share the same source.

We will discuss the statistical and evidence interpretation aspects of how to address this question throughout the rest of this presentation.

Exact Nature of the Question

In general, when working on these problems, we are concerned with what is being asked by the practitioner.

Let us focus on these two questions-

- Q1: Are the copper wire samples found at the crime scene from the same wire coil found on the suspect?
- Q2: Are the copper wire samples found at these two different crime scenes from the same wire coil?

What is the difference between these two questions?

- Q1: The source (the wire coil related to the suspect) is fixed!
- Q2: There is not a specific source in mind- we are only concerned with whether or not the two samples share a common (but unknown) source.

Why is this distinction important?

Each of these questions can have radically different answers, even when the statistician is given the same information....

This issue is due to the following interrelated reasons:

- The evidence that is used to answer the question
- The probability models used to characterize the evidence
- Methods to solve the problem in an optimal manner
- The interpretation/presentation of the results of the identification process
- The definition of an error

Outline

The Value of Evidence Approach

Using the odds form of Bayes Theorem, we arrive at the following definition for the Value of Evidence

$$\underbrace{\frac{P\left(H_p|E,I\right)}{P\left(H_d|E,I\right)}}_{\text{Posterior Odds}} = \underbrace{\frac{P\left(E|H_p,I\right)}{P\left(E|H_d,I\right)}}_{\text{Value of Evidence}} \times \underbrace{\frac{P\left(H_p,I\right)}{P\left(H_d,I\right)}}_{\text{Prior Odds}}$$

where

P: Probability operator

E: Evidence

H_p: E has arisen according to the prosecution model

H_d: E has arisen according to the defense model

I: Background information

Value of Evidence Forms

The Bayes Factor

$$V_{BF}(e) = \frac{\pi(e|H_p, I)}{\pi(e|H_d, I)}$$

The Likelihood Ratio

$$V_{LR}(e) = \frac{f(e|\theta_{p_0})}{f(e|\theta_{d_0})}$$

The Neyman-Pearson Likelihood Ratio

$$V_{NP}(e) = \frac{\max\limits_{\theta_p \in \Theta_p} f\left(e \left| \theta_p \right.\right)}{\max\limits_{\theta_d \in \Theta_d} f\left(e \left| \theta_d \right.\right)}$$

Approximation Theorems

Let $f(\cdot|\theta_s)$ and $f(\cdot|\theta_a)$ be bounded continuous functions of θ_s and θ_a, respectively.

Theorem (1)

Let the assumptions of Doob's Consistency Theorem[a] *be satisfied. Then as* $n \to \infty$

$$V_{BF}(e) \xrightarrow{P} V_{LR}(e).$$

[a] Theorem 10.10 from van der Vaart *Asymptotic Statistics* p. 149

Theorem (2)

Let the assumptions of the Consistency of M-Estimators theorem[a] *and the Linearization of M-Estimators theorem*[b] *hold. Then as* $n \to \infty$

$$V_{NP}(e) \xrightarrow{P} V_{LR}(e).$$

[a] Corollary 3.2.3 on p. 287

[b] Theorem 3.3.1 on p. 310 from van der Vaart and Wellner *Weak Convergence and Emprirical Processes*

Outline

Specific Source Hypotheses

H_p: The wire samples from E_u came from the specific wire coil in question.

H_d: The wire samples from E_u came from a randomly selected wire coil in the alternative source population.

Since copper wires are manufactured from the copper rods, we will assume that the alternative source population of wire coils can be characterized by studying the distribution of samples taken from the copper rods at the manufacturing facility.

Specific Source Evidence

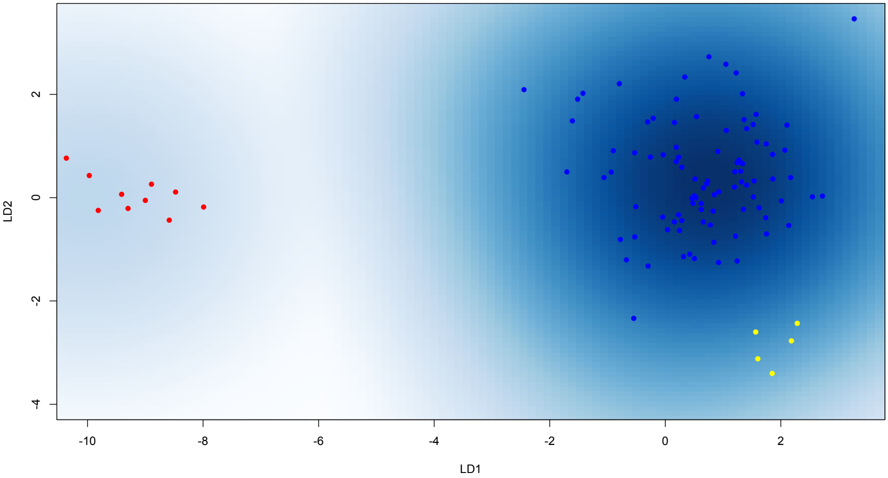
$E = \{E_s, E_a, E_u\}$ where:

E_s: The elemental compositions on 10 wire samples from the specific source (the 1^{st} wire coil).

E_a: Composed of the elemental compositions of 93 samples collected from a population of rods that are used for producing wire coils.

E_u: The elemental compositions of 5 wire samples from the 2^{nd} wire coil.

E: The Complete Evidence



E_s: Probability Model

Let y_{sj} denote the vector of measurements on the j^{th} wire sample, for $j = 1, 2, \ldots, 10$, from the specific wire coil. We assume these measurements follow a multivariate normal distribution;

$$y_{sj} \overset{iid}{\sim} MVN(\mu_s, \Sigma_s)$$

Then

$$\theta_s = \{\mu_s, \Sigma_s\}.$$

E_a: Probability Model

Let y_{ij} denote the vector of measurements on the j^{th} wire coil sample, for $j = 1,\, 2,\, \ldots\, m_i$, from the i^{th} randomly selected wire coil, for $i = 1,\, 2,\, \ldots\, n$. We assume these measurements follow a simple multivariate random effects model;

$$y_{ij} = \mu_a + a_i + w_{ij}.$$

$a_i \overset{iid}{\sim} MVN(0, \Sigma_a)$

$w_{ij} \overset{iid}{\sim} MVN(0, \Sigma_w)$

a_i and w_{ij} are independent of each other for all i and all j

Then

$$\theta_a = \{\mu_a, \Sigma_a, \Sigma_w\}.$$

E_a: Rods not Wires ...

We do not have samples from a large number of wires...

Let y_j denote the vector of measurements on the j^{th} rod sample for $j = 1, 2, \ldots 93$. We assume these measurements follow a simple multivariate random effects model;

$$y_j = \mu_r + r_j.$$

$$r_j \overset{iid}{\sim} MVN(0, \Sigma_r)$$

Compromises

So as long as we are willing to make the following assumptions about the alternative source population

$$\mu_a = \mu_r$$
$$\Sigma_a = \Sigma_r.$$

we can provide an answer to the specific source problem!

We are making one strong assumption, namely that the rod samples give us an idea of how the between wire coil samples behave.

Specific Source Neyman-Pearson LR

$$V_{NP}(e) = \frac{f(e_s|\hat{\theta}_s^*)f(e_u|\hat{\theta}_s^*)f(e_a|\hat{\theta}_a)}{f(e_s|\hat{\theta}_s)f(e_u|\hat{\theta}_a^*)f(e_a|\hat{\theta}_a^*)}$$

where

$\hat{\theta}_s$ is the MLE for θ_s under H_d

$\hat{\theta}_a$ is the MLE for θ_a under H_p

$\hat{\theta}_s^*$ is the MLE for θ_s under H_p

$\hat{\theta}_a^*$ is the MLE for θ_a under H_d

Then

$$V_{NP}(e) = 6.61 \times 10^{-12}.$$

Effectively zero!

Outline

Common Source Evidence

$E = \{E_{u_1}, E_{u_2}, E_a\}$ where:

E_{u_1}: The elemental compositions on 10 wire samples from a source (the 1^{st} wire coil).

E_{u_2}: The elemental compositions of 5 wire samples from a source (the 2^{nd} wire coil).

E_a: Composed of the elemental compositions of 93 samples collected from a population of rods that are used for constructing wire coils.

Restating the Problem

H_p: The wire samples in E_u and E_s came from the specific source wire coil.

H_d: The wire samples in E_u came from the randomly selected wire coil in the alternative source population, while the wire samples in E_s came from the specific source wire coil.

Versus:

H_p: The wire samples in E_{u_1} and E_{u_2} came from the same randomly selected wire coil in the alternative source population.

H_d: The wire samples in E_{u_1} and E_{u_2} came from two different randomly selected wire coils in the alternative source population.

E_a: Probability Models

Let y_{ij} denote the vector of measurements on the j^{th} wire sample for $j = 1,\, 2,\, \ldots\, m_i$ wire sample from the i^{th} wire coil, for $i = 1,\, 2,\, \ldots\, n$. simple multivariate random effects model;

$$y_{ij} = \mu_a + a_i + w_{ij}.$$

$a_i \overset{iid}{\sim} MVN(0, \Sigma_a)$

$w_{ij} \overset{iid}{\sim} MVN(0, \Sigma_w)$

a_i and w_{ij} are independent of each other for all i and all j

The only difference between the common source models H_p and H_d are whether or not e_{u_1} and e_{u_2} from one randomly selected coil or two different randomly selected coils.

We will make the same assumptions concerning the alternative source population of wire coils and the rods as we made under the specific source models.

Common Source Value of the Evidence

$$V_{NP}(e) = \frac{f(e_{u_1}|\hat{\theta}_a^p)f(e_{u_2}|\hat{\theta}_a^p)f(e_a|\hat{\theta}_a^p)}{f(e_{u_1}|\hat{\theta}_a^d)f(e_{u_2}|\hat{\theta}_a^d)f(e_a|\hat{\theta}_a^d)}$$

where

$\hat{\theta}_a^p$ is the MLE for θ_a under H_p

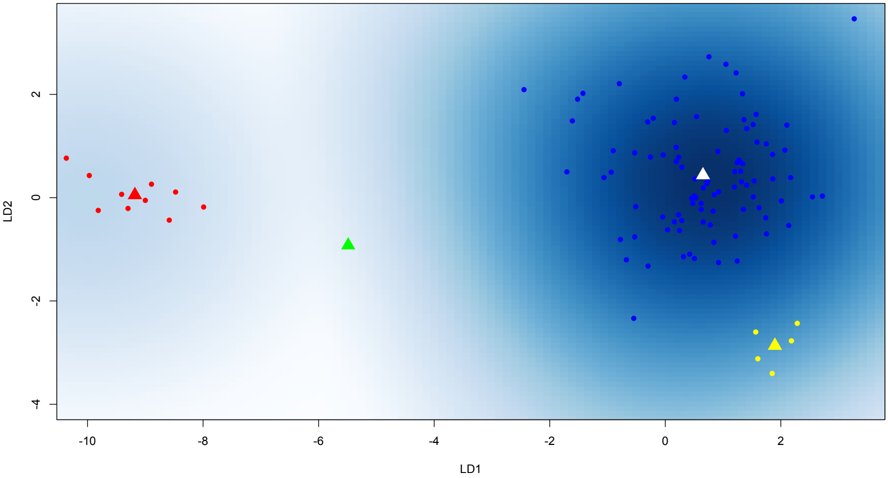
$\hat{\theta}_a^d$ is the MLE for θ_a under H_d

Then

$$V_{NP}(e) = 19,629.$$

Much larger than 1!

But E_{u_1} and E_{u_2} are far apart!



Recap

Specific Source

$$V_{NP}(e) = 6.61 \times 10^{-12}$$

Common Source

$$V_{NP}(e) = 19,629$$

The only difference is the question being asked.

Q1 Are the copper wire samples found at the crime scene from the same wire coil found on the suspect?

Q2 Are the copper wire samples found at these two different crime scenes from the same wire coil?

Conclusions

Lindley's Paradox

The copper wire example is an illustration of a phenomenon known as Lindley's Paradox.

It occurs in the common source attribution problem when e_{u_1} and e_{u_2} are far away from the center of the alternative source population.

This is the first time we have encountered it without methodologically creating the datasets.

Glen Shafer has studied this paradox in a number of settings.

In most situations, the common source and specific source values of evidence are approximately equal to each other.

Outline

So what question should we ask?

Depends ....

Who are you in the process?

What are you trying to do?

What evidential resources do you have available?