Toward a Likelihood Ratio for Bullet Lead Evidence

Hal S. Stern, Alicia L. Carriquiry,and Michael Daniels
University of California, Irvine, Iowa State University and University of Florida

August 21, 2003

1 Introduction

Forensic evidence plays an important role in the final courtroom decision concerning the guilt or innocence of a suspected criminal. The question that forensic examiners try to answer is whether two items, one found at the crime scene and one found on the suspect, have a common origin. They may also try to quantify the quality of the evidence, e.g., by specifying the probability that two items might appear to match by coincidence. Examples of the kinds of evidence that may be considered include traces of DNA, traces of blood, glass fragments, and the subject of the work reported here, trace element concentrations in bullet lead.

The application of modern scientific methods to determine whether traces of DNA found at the crime scene in the form of hair or bodily fluids match the DNA of a suspect is now commonplace, as is the use of probability and statistics to quantify the strength of such evidence. There is now a push to extend such approaches to other types of evidence such as synthetic fibers, glass, and bullets. These classes of evidence introduce complications not present with biological specimens; the population distribution of biological specimens, such as DNA or blood type, are generally accepted as being adequate and assumed invariant over time whereas for other types of evidence, such as bullets, the population distribution is largely unknown. To determine such population (or reference) distributions for trace element concentrations in bullet lead or other non-biological evidence types, one must try to exploit features specific to the manufacturing process of the product.

Our approach to assessing the quality of bullet lead trace evidence uses the likelihood ratio to provide a quantitative measure of the quality of the evidence. The likelihood ratio approach can be motivated by Bayes' Theorem. The odds form of Bayes' Theorem holds that the posterior odds for an event (in this case, that the evidence resulted from contact between the suspect and the crime scene) are the product of the prior odds and the likelihood ratio. The likelihood ratio compares the chances of obtaining measurements like those obtained for the bullet fragment at the crime scene and for the suspect's bullets under the two competing hypotheses (the fragment comes from the same box or boxes as the suspect's bullets or not). From this perspective the likelihood ratio supplies a one-number summary of the degree of match and the significance thereof. The main disadvantage of this approach is that a large number of assumptions is often required to estimate a likelihood ratio.

This report emphasizes two related findings. First, our results suggest some difficulty in reliably measuring the quality of bullet lead evidence. The likelihood ratio approach is developed, but only for a special case that is likely to be rare. There are some computational difficulties in extending the likelihood ratio approach to more realistic scenarios, especially with respect to computing the probability of a match by coincidence. The second important finding is that, despite the difficulty, it is crucial to consider the possibility that the crime scene bullet fragment and the bullet (or bullets) found on the suspect could match by coincidence. It does not suffice to just establish that two samples are a match since the match does not imply a common origin for the two samples. Our primarily negative results are noteworthy because they highlight the importance of the manufacturing process in assessing bullet evidence. The data made available to us have been collected after the manufacturing process is complete, from bullets purchased at stores or found during the course of investigations. Our results clearly demonstrate that there would be a benefit to data collected from the manufacturer prior to the packaging of bullets into boxes.

The next section provides some preliminary information about the bullet manufacturing process and the data available to us. An informal preliminary analysis of the bullet data is presented in Section 3. Some of the things learned there helped guide subsequent work on the likelihood ratio approach (Section 4). A numerical example is provided in Section 5. A summary and some final comments are given in Section 6.

2 Preliminaries

2.1 Bullet Manufacturing Process

At the time of this analysis (1999 and 2000) there were four major U.S. manufacturers of bullets: Cascade Cartridge, Federal, Remington, and Winchester. Though there is some variation in the manufacturing process across companies, the basic process is fairly consistent. It was described for us by Charles Peters, Scientist at the FBI laboratory in Washington, DC.

Bullets are produced from lead alloy obtained from local smelters. Manufacturers set specifications or guidelines for the alloy, e.g., a target range for the concentration of a particular element. One potential difficulty is that these need not be fixed over time. At the factory, the lead alloy from the smelter is melted down and mixed in large vats. Bullets are produced from the raw material in these vats and then subsequently packaged in boxes with each box containing fifty bullets. Depending on caliber, approximately 300,000 to 800,000 bullets may be produced from the raw material in a single vat. That is, the bullets produced from a single vat of raw material can be packaged into 6,000 to 16,000 boxes. The storage of bullets within the manufacturing plant and the packaging process are such that bullets from several different vats can end up in the same box, though the degree to which this happens varies from manufacturer to manufacturer.

It is natural to expect some variability among bullets that are produced from the same vat due to imperfect mixing of the molten lead. We expect however that bullets produced from the same vat of raw material are more alike than bullets produced from different vats. These are the key suppositions that guide the attempt to determine whether two bullets were manufactured at the same time. One difficulty in the analyses that follow is the absence of controlled data collection from the manufacturers that would allow this assumption to be examined more carefully. If samples of bullets known to have been produced from the same raw material could be obtained, it would be possible to quantify the variation in trace element concentrations among bullets from the same vat, and for bullets from different vats.

2.2 Data

Building sensible methods for assessing bullet evidence requires data. The primary data source on which we have relied is an FBI laboratory study (Peele et al., 1991) carried out using four full boxes (50 bullets each) of .38 caliber cartridges loaded with 158 grain, round nose bullets from each of the four major U. S. manufacturers. The boxes were generally packaged within three years of each other, except for the manufacturer Remington where there is a gap of more than ten years. Two of the boxes in each group of four boxes were packaged on the same date. The lead tip of each bullet was quartered and concentrations of five trace elements were measured on three of the quarters: copper, arsenic, bismuth, silver, and antimony. We relied primarily on the 800-bullet FBI lab data set to develop our methods. After developing methods we examine their performance in a large (13,000 bullet) FBI investigation database.

3 Exploratory Data Analysis

The first step is to examine the data from the 800 bullet FBI laboratory study to explore the nature of trace element data. The exploratory analysis also focuses on the possibility of identifying groups of bullets that originated from the same vat of raw material during the manufacturing process.

3.1 Summary statistics

Table 1 gives the mean concentration (measured in parts per million (ppm)) of each of five trace elements in the 200 bullets from each manufacturer. This table provides support for the notion that trace element concentrations can be used to assess bullet evidence. Bullets manufactured by Cascade and Federal have extremely high average concentrations of antimony whereas Remington and Winchester bullets are quite low on that element. Figure 1 contains a number of scatterplots showing the average measurement (over the three

Figure 1: Scatterplots showing mean of the three measurements (on logarithmic scale) for each bullet on two elements. Bullets from different manufacturers are plotted with different symbols as indicated in the legend.

measurements) for each bullet on two of the elements; bullets from different manufacturers are identified by different symbols. Figure 1 suggests that having measurements on just two elements is often enough to identify the manufacturer of a bullet, at least within the context of this study.

Manufacturer	Sb	Cu	As	Bi	Ag
Cascade	26836	262	233	128	37.6
Federal	27437	278	1381	16	65.5
Remington	7289	400	105	169	37.0
Winchester	4605	238	36	115	40.3

Table 1: Mean concentration (ppm) of each of five trace elements for 200 bullets from each of four manufacturers.

The three measurements of each single bullet provide data for estimating the standard deviation of the measurement process on a single bullet. Table 2 presents these standard deviations by manufacturer for each trace element. There is considerable variability in the standard deviations. Table 3 repeats the standard deviation calculations after taking the logarithm of each measurement. Note that taking the logarithm of the measurements makes the measurement variability more consistent across manufacturers and elements. There are still some large differences in variability but not as extreme as for the untransformed data. For the analyses that follow we use the logarithms of the measurements in our analyses.

Manufacturer	Sb	Cu	As	Bi	Ag
Cascade	427	6.5	5.4	4.8	0.7
Federal	317	5.2	21.5	0.4	4.2
Remington	60	5.3	2.4	5.2	0.6
Winchester	46	4.2	1.5	4.8	0.7

Table 2: Estimated standard deviation of measurement errors (ppm) averaged over the 200 bullets from each of four manufacturers.

Manufacturer	Sb	Cu	As	Bi	Ag
Cascade	0.016	0.025	0.028	0.038	0.019
Federal	0.012	0.019	0.016	0.023	0.067
Remington	0.008	0.016	0.038	0.034	0.018
Winchester	0.010	0.021	0.057	0.044	0.019

Table 3: Estimated standard deviation of measurement errors after taking logarithms of measurements averaged over the 200 bullets from each of four manufacturers.

3.2 Differentiating among manufacturers

For this data set the manufacturer of each bullet is known. The problem of building a rule from a training sample to identify the manufacturer of a bullet from a vector of trace element concentration measurements is an example of a classification problem. There are a number of statistical approaches to such problems including both formal methods (discriminant analysis, classification trees) and informal methods (graphical techniques). In the present case any of these approaches can easily discriminate among the four manufacturers for the 800 bullets in the laboratory study.

For example, the following simple rules provide nearly perfect classification on the training sample. If a bullet has antimony concentration less than 5713.34 then it is a Winchester bullet, if the antimony concentration is greater than 5713.34 and less than 16965.8 then it is a Remington bullet, if the antimony concentration is greater than 16965.8 then we must consult a second element, with high levels of arsenic (greater than 856.835) indicating a Federal bullet and low arsenic indicating a Cascade bullet. This can be represented graphically as in Figure 2. Using these simple rules all but one of the 800 bullets is classified correctly. Note that these classification rules serve to codify the differences between manufacturers evident in Figure 1.

Figure 2: Classification tree for identifying the manufacturer of a bullet.

As this part of the problem is relatively straightforward for the laboratory data, we put it aside for the moment and assume that bullet manufacturer can be easily identified. It should be noted however that automatically applying the classification tree from above to the large FBI bullet database was not successful - for one thing the measurements appear to be recorded in different units or on different scales over time. This is discussed further at the close of this Section.

3.3 Groups within manufacturers

Recall that the goal of using quantitative measures to assess bullet evidence is based on the supposition that bullets manufactured from the same raw material (i.e., in the same vat) are likely to have more similar trace element concentration measurements than bullets manufactured from different raw material at the same manufacturer. Identifying groups of bullets manufactured from the same raw material, these are called compositional groups in the FBI study, is an example of what is known as a clustering problem because the number (or even existence) of such groups is not known for sure. We explored a large number of approaches to this problem including traditional clustering methods (e.g., hierarchical clustering), graphical methods based on viewing representations of the data in one-, two-, or three-dimensions, and a method designed to mimic the manual clustering algorithm that was used in the FBI's laboratory study. The method that seemed to work best is known as model-based clustering.

Model-based clustering (Banfield and Raftery, 1993) hypothesizes that measurements in each cluster (group) can be modeled using a Gaussian distribution with a specified covariance structure; the distribution of the trace element concentrations is better approximated by a normal distribution if the logarithm of the measurements is used. For cluster k the distribution of the vector of measurements z is Gaussian with mean vector μ_k, and variance matrix Σ_k. The probability density for such a cluster, which specifies the relative likelihood of observing a value z is

$$\phi_k(z|\mu_k, \Sigma_k) = (2\pi)^{-\frac{p}{2}}|\Sigma_k|^{-\frac{1}{2}}\exp\left\{ -\frac{1}{2}(z-\mu_k)^T\Sigma_k^{-1}(z-\mu_k) \right\} .$$

These clusters are ellipsoidal, centered at the mean μ_k. The variance matrix Σ_k determines the other geometric features of the cluster (e.g., shape, orientation, volume/size). Before fitting the cluster model a variance structure must be specified indicating whether clusters are assumed to agree or vary on the features mentioned above.

Given indicator variables that identify the cluster to which each bullet belongs, the data can be viewed as independent Gaussian samples. As the indicator variables are not known, the data can be viewed as coming from a finite mixture model. The EM algorithm can be used to simultaneously estimate the parameters for each cluster and determine (probabilistically) the cluster membership of each observation. We obtained the clusters (or compositional groups in the bullet context) using MCLUST (Fraley and Raftery, 1998), a software package for model-based cluster analysis. A formal model selection measure is proposed by the developers of model-based clustering for estimating the number of groups present in a data set.

Model-based clustering was applied to the 200 bullets for each manufacturer to identify groups of bullets that appear to have come from a single batch of raw material. According to the results, there are 8 compositional groups represented in the 200 bullets for the manufacturer Cascade, 4 compositional groups for Federal, 9 compositional groups for Remington, and 36 compositional groups for Winchester. Tables 4-7 give the clusters found for each manufacturer, the size of the clusters, and identifies the boxes from which the bullets in each cluster came. It is worth noting that these groups are similar to those reported by the FBI study (Peele et al. 1991) in their analysis of the same data, but that model-based clustering typically yields a smaller number of clusters. The algorithm used in the FBI study which analyzed the bullets one at a time, comparing each to all existing groups, is more conservative in that it forms new groups more often than the model-based approach.

GROUP	# OF BULLETS	SOURCE OF BULLETS
1	22	22 from box 1
2	9	9 from box 1
3	19	19 from box 1
4	87	44 from box 2, 43 from box 3
5	9	5 from box 2, 4 from box 3
6	5	1 from box 2, 3 from box 3, 1 from box 4
7	40	40 from box 4
8	9	9 from box 4

Table 4: Model-based clustering results for Cascade.

GROUP	# OF BULLETS	SOURCE OF BULLETS
1	49	18 from box 1, 13 from box 2, 18 from box 4
2	95	32 from box 1, 37 from box 2, 26 from box 4
3	6	6 from box 4
4	50	50 from box 3

Table 5: Model-based clustering results for Federal.

GROUP	# OF BULLETS	SOURCE OF BULLETS
1	63	30 from box 1, 33 from box 2
2	19	10 from box 1, 9 from box 2
3	10	4 from box 1, 1 from box 2, 5 from box 3
4	13	6 from box 1, 7 from box 2
5	31	31 from box 3
6	13	13 from box 3
7	11	1 from box 3, 10 from box 4
8	22	22 from box 4
9	18	18 from box 4

Table 6: Model-based clustering results for Remington.

GROUP	# OF BULLETS	SOURCE OF BULLETS
1	1	1 from box 1
2	25	25 from box 1
3	1	1 from box 1
4	1	1 from box 1
5	7	7 from box 1
6	1	1 from box 1
7	1	1 from box 1
8	2	2 from box 1
9	2	2 from box 1
10	4	4 from box 1
11	1	1 from box 1
12	1	1 from box 1
13	2	2 from box 1
14	1	1 from box 1
15	13	9 from box 2, 4 from box 4
16	8	4 from box 2, 4 from box 4
17	14	8 from box 2, 6 from box 4
18	13	10 from box 2, 3 from box 4
19	8	4 from box 2, 4 from box 4
20	1	1 from box 2
21	7	5 from box 2, 2 from box 4
22	2	2 from box 2
23	1	1 from box 2
24	3	1 from box 2, 2 from box 4
25	1	1 from box 2
26	9	3 from box 2, 6 from box 4
27	1	1 from box 2
28	37	13 from box 4, 24 from box 3
29	8	2 from box 4, 6 from box 3
30	1	1 from box 4
31	1	1 from box 4
32	1	1 from box 4
33	8	8 from box 3
34	6	6 from box 3
35	3	3 from box 3
36	4	1 from box 4, 3 from box 3

Table 7: Model-based clustering results for Winchester.

3.4 Performance on the larger data set

Several things made it difficult to apply the classification and clustering findings of this section to the larger FBI bullet database. First, the FBI database includes data from many manufacturers including those from outside the U.S.; for some only a limited amount of data are available. Second, the measurements of the elements were much less consistent over time than was expected based on the lab study. In the lab study the four boxes with varying packaging dates yielded good evidence that a manufacturer's specifications would be fairly consistent. The larger database sheds doubt on whether this is true. Bullets were entered into the database over a long time period and may represent bullets manufactured over an even longer time period. There are two consequences of our difficulty working with the larger database. First, we have not used these data in the remainder of our work. Second, if there are not in fact consistent manufacturer patterns over long periods of time, then periodic samples from the manufacturers would be required to make assessing evidence practical.

4 The Likelihood Ratio Approach

The likelihood ratio approach to assessing the evidence from a crime scene and a suspect provides a single quantitative measure of the probative value of the evidence. References to the likelihood ratio approach in the context of other evidence types include Evett et al. (1987) and Wakefield et al. (1991) for fibers, and Curran et al. (1997) for glass.

4.1 Introduction

To begin, let G denote the hypothesis that the bullet fragment (or fragments) from the crime scene and the bullets found with the suspect have a common source (i.e., come from a single source box (or boxes)). We abbreviate this as G for "Guilt" recognizing of course that a common source does not necessarily imply that the suspect is guilty. It is common to take $\bar{G}$ to be the complementary or opposite event, that the bullet fragments and the bullets do not have a common source. Take E to represent all of the bullet evidence. Formally E could contain evidence of many types but we restrict attention to trace element measures from the bullets. One reasonable goal of a forensic examination is a measure of the probability of the hypothesis G. Using Bayes' Theorem we find an expression for the probability of guilt based on the evidence E,

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

This expression can be written more simply in terms of the odds in favor of the hypothesis G given the evidence E,

$$\frac{\Pr(G|E)}{\Pr(\bar{G}|E)} = \frac{\Pr(E|G)}{\Pr(E|\bar{G})}\frac{\Pr(G)}{\Pr(\bar{G})}.$$

This last expression gives the odds of the hypothesis G given the evidence E (sometimes known as the posterior odds) as the product of two terms, the odds in favor of the hypothesis G before looking at the evidence, $\Pr(G)/\Pr(\bar{G})$ (sometimes known as the prior odds), and the *likelihood ratio*, $\Pr(E|G)/\Pr(E|\bar{G})$. The numerator of the likelihood ratio measures the probability (or likelihood) of the evidence under the hypothesis G while the denominator measures the probability (or likelihood) of the evidence under the hypothesis $\bar{G}$.

The two quantities that define the likelihood ratio are related to the two steps of the traditional significance test/coincidence probability approach. The numerator is related to determining whether the two objects match; as with a significance test, the evidence is evaluated under the null hypothesis that the two objects do in fact have a common source (hypothesis G). The denominator is related to the step in which the significance of the match is assessed by determining the probability of a coincidental match, i.e., a match under the hypothesis $\bar{G}$. One nice feature of the likelihood ratio is that it provides a single quantitative measure without requiring a binary decision as to whether the two objects match or not. Uncertainty about the match status is factored in to the single measure along with the probability of alternative explanations of the evidence.

It turns out however that calculation of the likelihood ratio requires that a number of assumptions be made. Moreover, we show below that even under these assumptions computing the likelihood ratio in the context of trace element data for bullet evidence becomes increasingly difficult as the amount of evidence increases.

4.2 Application to bullet trace element data

We now consider the likelihood ratio in the context of trace element data for bullet lead. Assume that a crime has been committed and that the evidence includes k bullet fragments found at the crime scene and m bullets found with a suspect. Often k is small; we initially attempt to find a likelihood ratio for the case with $k = 1$. The fragments and bullets are analyzed yielding a set of trace element concentrations. The number of trace elements can vary depending on the bullet manufacturer and perhaps other factors. Formally then, the evidence E includes all of the available trace element measurements for the $k + m$ bullets and fragments.

The hypothesis G is that the fragments originate from the same box (or even set of boxes) as the bullets found on the suspect. The hypothesis $\bar{G}$ is that the fragments do not originate from the same box (or set of boxes) as the bullets found on the suspect. One complication with bullet data is a consequence of the manufacturing process. It is assumed that bullets manufactured from the same vat of raw material share relatively similar trace element concentrations. It is known that bullets manufactured from the same raw material can end up in different boxes (obvious given the number of bullets manufactured from a single vat of raw material) and that each box is likely to contain bullets manufactured from different vats of raw material (less obvious but still true). Thus it is possible that under the hypothesis G, which specifies that the fragments come from the same box as the suspect's bullets, the fragments will not have trace element concentrations that match those of the bullets in the box. This differs from other forms of evidence (DNA, blood type) where under G we can be reasonably certain (barring contamination) of a close or even identical match. It is also possible that the fragments' trace element concentrations will match those of the suspect's bullets even though the true hypothesis is $\bar{G}$ (this is the coincidental match we worry about). Given that 6,000 or more boxes can be produced from a single vat of lead, the probability of finding two identical bullets is likely to be non-negligible. Thus, the relative chance of these two events must be determined as part of the likelihood ratio calculation.

Based on the preliminary analysis reported in Section 3, we take it as given that the manufacturer of the bullet fragment and the manufacturer of the suspect's bullets can be determined without error. If so, then it is reasonable to only work on the case where the fragments and suspect's bullets have the same manufacturer. If they do not, then the likelihood ratio is near zero (nonzero because of the possibility that manufacturers buy bullet lead from one another).

4.3 The one fragment/one bullet case

Notation. Consider the case with $k = 1$ fragment and $m = 1$ bullet on the suspect. This scenario may not be realistic in that the suspect is likely to have more than a single bullet, most likely the unused portion of a box. The scenario is however very useful in explaining how the likelihood ratio is constructed and in pointing out some of the difficulties that occur when there is more evidence. Let x denote the vector of trace element measurements (recall that these are logarithms of the actual measurements) on a bullet fragment found at a crime scene and y denote the vector of trace element measurements on a bullet found with a suspect. The evidence E in this case is the combined data x, y. The likelihood ratio is then defined as

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

where $p(\cdot|H)$ is notation describing the probability or likelihood of the observed value conditional on the hypothesis H being true. If the evidence in question can only take a limited number of values, as in the case of blood type, then p will represent actual probabilities, but when the evidence is a continuous measurement it is more natural to think of p as a probability density function. It is possible to simplify the form of the likelihood ratio in this case because the likelihood or probability associated with just a single measurement, say y, ignoring the other is the same under either hypothesis (in the sense that the likelihood is governed completely by the manufacturing process since there is no other information to be taken into account),

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

A difficulty with the form of the likelihood ratio specified above in terms of G and $\bar{G}$ is that, as mentioned earlier, boxes consist of bullets from a number of compositional groups (vats of raw material). It is reasonably straightforward to think about the likelihood of x given y if the two lead pieces are hypothesized to come from the same compositional group or from different compositional groups, but harder to think about the likelihood given only the hypothesis G or $\bar{G}$. Fortunately this additional structure can be taken advantage of by introducing the event S that two bullets come from the same compositional group and rewriting the likelihood ratio as

$$\begin{aligned} LR &= \frac{p(x|y,G)}{p(x|y,\bar{G})} \\ &= \frac{p(x|y,S,G)p(S|G) + p(x|y,\bar{S},G)p(\bar{S}|G)}{p(x|y,S,\bar{G})p(S|\bar{G}) + p(x|y,\bar{S},\bar{G})p(\bar{S}|\bar{G})}. \end{aligned}$$

This last expression involves two different types of terms: terms describing the variation in two bullets' measurements given that they come from the same (or different) compositional groups, and terms describing the makeup of bullet boxes in terms of the compositional groups represented. The former type, terms of the form $p(x|y,S,G)$, can be simplified because given S (or $\bar{S}$) we need no longer concern ourselves with the hypothesis G. In other words $p(x|y,S,G) = p(x|y,S,\bar{G}) = p(x|y,S)$ and similarly for $\bar{S}$.

Bullet measurements within (between) compositional groups. For each manufacturer, there are measurements of p (typically 5 or 6) trace element concentrations for each bullet. We assume that the vectors of trace element measurements for bullets from the kth compositional group have a multivariate normal distribution with mean μ_k and variance matrix Σ_{within}. Note that this is more restrictive then the most general clustering model in which each cluster has its own variance matrix. In practice we found it useful to combine the information from all of the clusters to estimate a common variance matrix for all of the compositional groups for a single manufacturer (though different manufacturers have different variance matrices). It is difficult to reliably estimate separate variance matrices for each different compositional group. The normality assumption is made more tenable by the fact that the trace element measurements are typically an average of several readings per bullet. We have considered alternatives to the multivariate normal assumption. If we think of the different compositional groups as perhaps having different variance matrices, then we are led to a multivariate t-distribution in place of the normal distribution. This yields additional variation which should yield more conservative likelihood ratios in practice.

To describe the variation among bullets from different compositional groups additional assumptions are required. If we further assume that the means for each compositional group (the μ_k's) have a normal distribution with mean μ_{manuf} and variance $\Sigma_{between}$, then the distribution of a randomly chosen bullet from that manufacturer can be viewed as a normal distribution with mean μ_{manuf} and variance $\Sigma_{between} + \Sigma_{within}$. This is the relevant distribution when we are looking at a bullet with no knowledge about its compositional group. It is natural to assume that the manufacturer mean differs for different manufacturers. Once again alternatives to the normal distribution can be considered.

Each likelihood term concerns the distribution of x given y and information regarding whether the fragment and bullet are from the same compositional group. Under this model, if the fragment and bullet are from different compositional groups, then x and y are independent and so $p(x|y, \bar{S})$ is just the usual multivariate normal density for a randomly drawn bullet, i.e., with mean μ_{manuf} and variance $\Sigma_{within} + \Sigma_{between}$ evaluated at x. If the bullet and fragment are from the same compositional group, then the two vectors of measurements can be viewed as coming from a joint multivariate normal distribution,

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{manuf} \\ \mu_{manuf} \end{pmatrix}, \begin{pmatrix} \Sigma_{within} + \Sigma_{between} & \Sigma_{between} \\ \Sigma_{between} & \Sigma_{within} + \Sigma_{between} \end{pmatrix}\right).$$

It follows that the conditional distribution of x given y for the same compositional group case is multivariate normal with mean

$$m_{same} = \mu_{manuf} + \Sigma_{between}(\Sigma_{within} + \Sigma_{between})^{-1}(y - \mu_{manuf})$$

and variance matrix

$$V_{same} = \Sigma_{within} + \Sigma_{within}(\Sigma_{within} + \Sigma_{between})^{-1}\Sigma_{between}.$$

The likelihood ratio can then be calculated from (1) once the parameters of the normal model and the relevant combinatorial probabilities have been estimated. We conclude this discussion of the one bullet, one fragment case by addressing these final two issues.

Estimation of model parameters. The conditional distribution for x given y is normal under either the assumption of a common compositional group or the assumption of different compositional groups. To compute the two normal likelihoods requires that the parameters $\mu_{manuf}, \Sigma_{within}, \Sigma_{between}$ be estimated. Given the results of our clustering we estimate Σ_{within} using the pooled within group variance matrix. If z_{ij} is the vector of trace element concentrations for the jth bullet in compositional group i of the training data, then

$$S_{within} = \frac{1}{N-g}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(z_{ij} - \bar{z}_i)(z_{ij} - \bar{z}_i)',$$

where $\bar{z}_i$ is the mean of the trace element measurements for the bullets in group i, N is the total number of bullets and g is the number of groups.

When the number of bullets in each group/cluster is the same there are standard procedures for estimating $\Sigma_{between}$. That is not the case here so we use the following approach (described for example in Section 13.7 of the text by Snedecor and Cochran (1989)). The variation among cluster means can be estimated directly as

$$S_{between} = \frac{1}{g-1}\sum_{i=1}^{g}(\bar{z}_i - \bar{z})(\bar{z}_i - \bar{z})',$$

where $\bar{z}$ is the mean of all of the trace element measurements. If the sample sizes in the clusters were equal (say n observations per cluster), then $S_{between}$ would be a good estimate of $\Sigma_{between} + \Sigma_{within}/n$. In the present case with unequal sample sizes one can show that $S_{between}$ is actually providing an estimate of $\Sigma_{between} + \Sigma_{within}/n_h$ where n_h is the harmonic mean ($n_h = g/(\sum_{i=1}^{g} \frac{1}{n_i})$). Then a natural estimate of $\Sigma_{within} + \Sigma_{between}$ can be constructed from S_{within} and $S_{between}$ as $\hat{\Sigma}_{within} + \hat{\Sigma}_{between} = \frac{n_h - 1}{n_h} S_{within} + S_{between}$.

Throughout this study we have taken a fairly crude approach to the estimation of μ_{manuf} and assumed that the bullet measurement y is the best estimate for the manufacturer (or compositional group) mean μ_{manuf}. There are advantages and disadvantages to this approach. One advantage is that with this assumption the conditional mean of x is the same (and equal to y) under the assumption of a common compositional group or of different compositional groups. This can be verified from the conditional distributions discussed above. This mathemtical results means that the normal likelihood is a kind of standardized distance between x and y which facilitates its interpretation. A second advantage is that we insure that the estimate of μ_{manuf} is reasonably close to the data relevant to the case, even if the manufacturer has in facted changed their guidelines over time. Of course, a disadvantage of this approach is that ignores a great deal of information known to be relevant, namely the other trace element measurements from this manufacturer's products (including the fragment x. Using additional information to estimate μ_{manuf} requires only that we use the conditional distributions of the previous discussion to compute the likelihood ratio.

The distribution of compositional groups within boxes. The terms in the likelihood ratio concerning the events S and G (and their complements) do not depend on the measurement values at all, they need only be estimated once (and updated periodically as the industry changes). Ideally their estimation would be done with a large, carefully designed study. For now we have used the convenience sample available, the data from the 800 bullet FBI laboratory study, to illustrate our approach to estimating these quantities. For the conditional probabilities $p(S|G)$ and $p(S|\bar{G})$, we use a combinatorial argument. For example, suppose that there are b boxes of bullets from a given manufacturer, g_i homogeneous groups represented in the ith box, and n_{ij} bullets from the jth compositional group within the ith box for a manufacturer. The number of distinct compositional groups is at most $\sum_{i=1}^{b} g_i$ but could be less if the same groups appear in more than one box. Then one can calculate the proportion of two bullet samples from a single box that come from the same compositional group,

$$p(S|G) \quad = \quad \frac{p(S,G)}{p(G)} = \frac{\sum_{i=1}^{b} \sum_{j=1}^{g_i} \binom{n_{ij}}{2} / \binom{N}{2}}{\sum_{i=1}^{b} \binom{n_i}{2} / \binom{N}{2}},$$

where $n_i = \sum_{j=1}^{g_i} n_{ij}$ for $i = 1, \ldots, b$ and $N = \sum_i n_i$ is the total number of bullets. Note that $n_i = 50$ for a complete box, however in our example we chose to ignore one compositional group (it was small) that contained bullets from some of the boxes so that n_i will be less then 50 in those cases. Naturally, the probability that two bullets from the same box come from different compositional groups is just one minus the previous quantity.

The probability that two bullets come from the same group though they are in different boxes can be computed in a similar way. It is a bit more difficult to write down because more detailed notation is required. We explain the approach but do not give detailed equations.

$$p(S|\bar{G}) \quad = \quad \frac{p(S,\bar{G})}{p(\bar{G})} = \frac{p(S,\bar{G})}{1-p(S)}.$$

The denominator is easy to compute; we haThe likelihood ratio terms describing measurement variation for bullets from the same (different) compositional groups have been approximated by normal distributions.ve already computed $p(S)$. The calculation of $p(S,\bar{G})$ requires finding the total number of ways to select bullets from two different boxes such that the two bullets are from the same compositional group. We illustrate in the numerical example below. Occasionally it is possible that additional information (like packaging date) may exist to simplify this calculation.

4.4 Difficulty in applying the likelihood ratio

Having explained the basic principle of the likelihood ratio in the context of a simple example, it remains only to elaborate on how the calculation changes for more complex scenarios. It is here, however, that difficulties arise. The basic difficulty is that this approach to computing the likelihood ratio in the one-bullet, one-fragment case considers each of the possible scenarios for the two bullets: same compositional group and same box, same group and different box, different group and same box, different group and different box. Each possibility contributes to the likelihood ratio calculation. What happens when there are more bullets or fragments?

Extending the likelihood ratio approach to address the case when there is more than one bullet found in the suspect's partial box is difficult because it requires considering all possible configurations of the $m+1$ bullets (m from the suspect and the fragment). The fragment may come from the same compositional group as all m of the suspect's bullets or just a subset of the suspect's bullets. To illustrate, consider the scenario with $k=1$ and $m=2$. Let y_1 and y_2 denote the trace element measurements from the two bullets found with the suspect and, as usual let x denote the measurements from the fragment. There are five possible scenarios: (i) x, y_1, and y_2 all belong to different compositional groups (different manufacturing batches); (ii) y_1 and y_2 belong to the same homogeneous group which is different than x; (iii) x and y_1 belong to the same homogeneous group but y_2 is from a different group; (iv) x and y_2 belong to the same homogeneous group but y_1 is from a different homogeneous group; and (v) x, y_1, and y_2 belong to the same homogeneous group. Then the numerator and denominator of the likelihood ratio each need to consider the five scenarios along with their probability of occurring under the two hypotheses. Clearly the task is more difficult than it was in the simple case. In addition, the data collection required to reliably estimate the various probabilities increases as well.

Working with additional fragments is also difficult but in that case it is possible to avoid the problem by carrying out a separate analysis for each fragment. This ignores potentially useful information (e.g., whether the two fragments appear to have come from a single compositional group) but should provide reasonable information for courtroom use.

Note that there are other difficulties with the likelihood ratio approach as well. Our likelihood ratio derivation does not account for bullet distribution patterns or bullet usage patterns in assessing the likelihood of two bullets coming from the same vat of raw material. Thus if all of the bullets in a given local area are from the same manufacturer, and were manufactured at approximately the same time, then it will not be surprising to find matching bullets with most any partial box found by investigators in this locality. One can contrast that with a situation in which the fragment and suspect's box found by investigators are unusual for a given town. Such considerations are not insignificant and could vary considerably from case to case.

5 An example

We now apply the likelihood ratio approach for the assessment of evidence for the case with $k = 1$ and $m = 1$ to actual data. As we have assumed that manufacturer can be easily identified, we restrict attention to the 200 bullets produced by Cascade Cartridge Industries (CCI) analyzed in the FBI study. The data are the measurements of five trace element concentrations (antimony(Sb), copper(Cu), arsenic(As), bismuth(Bi), and silver(Ag)) for each of the 200 bullets obtained from four boxes, including two boxes (box 2 and box 4) with the same packaging date. We have taken the logarithms of the three measurements for each bullet (recall this makes the measurement variation more consistent) and then averaged these to obtain a single measure for each element for each bullet. The cluster analysis yields eight compositional groups using the model-based clustering method. The results are summarized in Table 4 of Section 4. Cluster (group) 6 was quite small and did not appear too homogeneous. As a result we did not use this in the remaining analysis.

To illustrate calculations consider a fragment with trace element concentrations $x = (10.219, 5.195, 4.536,\ 5.157,\ 4.065)$ and a bullet with trace element concentrations $y = (10.221, 5.202, 4.571, 5.157, 4.032)$. The estimated within compositional group variance matrix for Cascade is

$$\hat{\Sigma}_{within} = S_{within} = \begin{pmatrix} 0.000333 & 0.000167 & 0.000285 & 0.000071 & 0.000026 \\ 0.000167 & 0.000642 & 0.000375 & -0.000071 & 0.000080 \\ 0.000285 & 0.000375 & 0.001097 & 0.000216 & -0.000116 \\ 0.000071 & -0.000071 & 0.000216 & 0.003487 & 0.000424 \\ 0.000026 & 0.000080 & -0.000116 & 0.000424 & 0.000786 \end{pmatrix}.$$

The estimate for the sum $\Sigma_{between} + \Sigma_{within}$ based on the approach described in the previous section is

$$\hat{\Sigma}_{within} + \hat{\Sigma}_{between} = \begin{pmatrix} 0.001903 & 0.009086 & -0.026462 & 0.009287 & 0.014005 \\ 0.009086 & 0.122653 & -0.085549 & 0.036607 & 0.016389 \\ -0.026462 & -0.085549 & 0.524556 & -0.172858 & -0.296696 \\ 0.009287 & 0.036607 & -0.172858 & 0.068801 & 0.089066 \\ 0.014005 & 0.016389 & -0.296696 & 0.089066 & 0.192721 \end{pmatrix}.$$

Then $p(x|y, S)$ (recall S means same compositional group) is a multivariate normal distribution evaluated at the given x with y used as the estimated mean of the multivariate normal and $\hat{V}_{same}$ used as an estimate of the relevant variance matrix (formula given in Section 4),

$$p(x|y, S) = (2\pi)^{-5/2}|\hat{V}_{same}|^{-1/2}e^{-0.5(x-y)^T\hat{V}_{same}^{-1}(x-y)} = 90996.$$

The other relevant likelihood $p(x|y, \bar{S})$ is the same multivariate likelihood with $\hat{\Sigma}_{within} + \hat{\Sigma}_{between}$ used in place of $\hat{V}_{same}$. The resulting likelihood value is 196.4.

The remaining pieces of the likelihood ratio are probabilities that a bullet and fragment randomly chosen from the same (or different) box would come from the same group. Again we rely on the clustering data that found 8 compositional groups for Cascade (see Table 4 in Section 4) represented among the four boxes and we omit the five bullet group. Given the makeup of the four boxes, we calculated the probability that two randomly chosen bullets from the same box would come from the same (or different) compositional groups and the probability that two randomly chosen bullets from different boxes would come from the same (or different) compositional groups. Using the combinatoric arguments from the previous section we find that,

$$\begin{aligned} p(S|G) &= \frac{p(S,G)}{p(G)} \\ &= \frac{\binom{22}{2} + \binom{9}{2} + \binom{19}{2} + \binom{44}{2} + \binom{5}{2} + \binom{40}{2} + \binom{9}{2} + \binom{43}{2} + \binom{4}{2}}{\binom{50}{2} + \binom{49}{2} + \binom{49}{2} + \binom{47}{2}} = 0.67 \end{aligned}$$

and

$$\begin{aligned} p(S|\bar{G}) &= \frac{p(S,\bar{G})}{p(\bar{G})} \\ &= \frac{2\left[\ \binom{44}{1}\binom{43}{1} + \binom{5}{1}\binom{4}{1}\ \right]}{\binom{195}{2} - \left[\ \binom{50}{2} + \binom{49}{2} + \binom{49}{2} + \binom{47}{2}\ \right]} = 0.27. \end{aligned}$$

For these data, therefore, the estimated probability that two bullets match (i.e., come from the same compositional group) given that they come from the same box is .67, while the probability that two bullets match given that they come from different boxes is .27. The former Of course, these results are sensitive to the 200 bullets used – in particular since 100 of the 200 bullets were packed on the same day our estimate that the probability of a coincidental match of compositional groups is .27 is likely an overestimate. A larger study to refine these estimates would be needed before computing likelihood ratios for courtroom use.

Then the likelihood ratio for this pair of bullets is

$$LR = \frac{.67(90996) + .33(196.4)}{.27(90996) + .73(196.4)} = 2.47. \tag{1}$$

Note that the likelihood of obtaining measurements like these is much higher under hypothesis G than under hypothesis $\bar{G}$ though both are possible explanations. In fact x and y in this case were chosen randomly from the same box so that the LR should be large, favoring that hypothesis. The likelihood ratio is greater than one but not very large because we have estimated that there is a probability .27 that two bullets will have come from the same vat even if they are in different boxes. As argued above this number would probably be lower if based on a more representative data set which would tend to increase the size of the likelihood ratio.

To illustrate the variation that one can expect in computing likelihood ratios we have carried out two small experiments. First, we computed the likelihood ratio for 50 randomly chosen pairs of Cascade bullets, where each pair did in fact come from a common box. The values of the likelihood ratio ranged from 0.45 to 2.48. Four of the observed LR's are less than one suggesting that the bullet lead evidence favors $\bar{G}$ despite the fact that the bullets actually come from the same box. The presence of such values indicates again that variation in the manufacturing process is such that bullets from the same box may look quite different. As a second experiment we computed the likelihood ratio for 50 randomly chosen pairs of bullets, where the pairs do not come from a common group. Here we'd expect small likelihood ratios. All fifty values were approximately 0.45 (which is the smallest possible value given our estimated probabilities). This suggests that bullets from different clusters are very different and can be easily detected as not coming from the same cluster.

6 Summary and Discussion

Modern analytical techniques permit measuring the concentration of various trace elements in bullet lead with a high degree of accuracy. Naturally, these accurage measurements can then be used to decide whether two lead samples (perhaps a fragment and a bullet) match. If the samples match, they are said to belong to the same compositional group. Quantitative analyses such as these have been used recently in the courts of law to try to determine whether a bullet fired in the course of criminal activity can be matched to unspent bullets in the possession of a suspect. While establishing a match between the two sets of samples would seem to be a necessary step to tie the suspect to the crime scene, it is by no means sufficient.

The goal of this project was to develop a means for assessing bullet evidence, especially to be able to quantify the significance of matching bullet lead. A likelihood ratio approach is considered. The likelihood ratio approach is feasible at the current time only for small evidence sets, e.g., a single fragment and a single bullet. Moreover there are a number of issues associated with bullet distribution and usage patterns that were not factored into the current discussion. These also would tend to complicate the application of the likelihood ratio approach to bullets. The natural conclusion is that an automated quantitative approach to assessing bullet lead evidence is not possible at this time.

Despite this negative conclusion there are two constructive points that come from this study. One key point is that it is the bullet manufacturing process itself that makes the analysis of forensic bullet evidence difficult. Bullets manufactured from different batches of raw material may end up in the same box of bullets; similarly the many bullets manufactured from a single batch will definitely end up in different boxes. It has been difficult from the limited data available to estimate the relative frequency of these events. This appears to be crucial information for developing quantitative methods for the assessment of bullet evidence. A more focused data collection effort, collecting data from manufacturers at various points in the manufacturing process, is likely to provide reliable information for improving upon this work. As a start it would be worth a dedicated effort to measure bullet lead from bullets known to come from the same vat of raw material and then to evaluate batch-to-batch variability over time with additional samples.

Another key point concerns the practical issue of assessing bullet evidence in the absence of definitive quantitative measurements. The results here suggest that it is not possible to assume that the probability of a coincidental match is negligible and can therefore be ignored in court. While the estimate of the probability of a coincidental match that was obtained from the Cascade bullets in this study (.27) is likely a significant overestimate, it seems equally incorrect to treat this number as if it were zero. Again data collection might be a key to addressing this crucial question.

References

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **49**, pp. 803-821.

Curran, J. M., Triggs, C. M., Almirall, J. R., Buckleton, J. S., and Walsh, K. A. J. (1997). The interpretation of elemental composition measurements from forensic glass evidence: II. *Science and Justice*, **37**, pp. 245-249.

Evett, I. W., Cage, P. E., and Aitken, C. G. G. (1987). Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *Applied Statistics*, **36**, pp. 174-180.

Fraley, C., and Raftery, A. E. (1998). MCLUST: Software for model-based cluster analysis. Technical Report, Department of Statistics, University of Washington. (software available from www.stat.cmu.edu/S/mclust or
www.stat.washington.edu/fraley/mclust/soft.shtml).

Peele, E. R., Havekost, D. G., Peters, C. A., Riley, J. P., Halberstam, R. C., and Koons, R. D. (1991). Comparison of bullets using the elemental composition of the lead component. Proceedings of the 1991 International Symposium on the Forensic Aspects of

Trace Evidence, pp. 57-68.

Snedecor, G. W., and Cochran, W. G., (1989). *Statistical Methods*, 8th edition. Iowa State University Press: Ames, IA.

Wakefield, J. C., Skene, A. M., Smith, A. F. M., and Evett, I. W. (1991). *Applied Statistics*,**40**, pp 461-476.

Acknowledgements

Ying Huang, Ji-Yeon Kim, and Jennifer Schumi, research assistants on the project, provided support for the research and preparation of this report. We are also appreciative of the help provided by Robert Koons and Jo Ann Buscaglia of the Federal Bureau of Investigation's Forensics Unit.