Outline of writeup

Yawei Ge

July 1, 2020

Outline 1

Questions answered: (Is this a good question defined?)

The firearm examiners are focusing on the same source and different source problems of bullets and cartridge cases which serve as important forensic evidence in the court/to the jurors. The subjectivity in the traditional forensic process is called to be reduced or complemented by more objective methods. Some automatic matching algorithms are developed which usually return a similarity score (0-1) to quantify the similarity or the probability to be an actual match for a certain comparison. However, this raises questions about how to interpret the reported probabilities and how these probabilities are distributed. Thus, it is not all clear how to conduct inferences based on the similarity scores. [The paper] proposed binomial and beta-binomial for the number of matched cells of the CMC method. Thus, it provides a way to quantify the theoretical error rate of the algorithm. However, for the bullets LEA comparison, we haven't established similar distributional results. In this paper, we will evaluate the possible models/distributions for the LEA comparisons scores generated by the random forest proposed by Hare etc. And then, we will also evaluate the error rates based on the estimated distribution for the automatic matching algorithm. (threshold selection could also be mentioned if needed).

Simply speaking, the question is "how is the LAPD RF score distributed and how to evaluate the theoretical error rate for this particular combination of ammunition and firearm based on your (our) proposed distributional form, further, how will the conclusion be affected by the sample sizes"

Answer the question

To answer the question, basically, we will choose the distributional form, estimate the distribution, evaluate the fit, calculate the error rate; change sample sizes and evaluate the sensitivity to the sample sizes

1. We first propose the candidate distributional forms

2. then choose the one best describing the data while gain a proper explanation

3. evaluate the fit

4. analyze the result

5. evaluate the error rate

6. sample size effect

7. conclude

The main results

The Main results: which have been discussed for a while, need some modification, reproduce some results

Sections of the writeup

- introduction, literature review

- distribution selection (why select beta)

- data set (LAPD)

- the distribution estimated

- the error rate analysis associated

- with changing sample sizes, evaluate the performance (in terms of stability, variation, and error rates)

- conclusions