J3   Handwriting Individuality: Probability Models, Subsampling Routines, and Implications

Christopher P. Saunders, PhD*, and Mark Lancaster, George Mason University, 4400 University Drive, MS 1G8, Fairfax, VA 22030; and JoAnn Buscaglia, PhD, Federal Bureau of Investigation Laboratory, Counterterrorism & Forensic Science Research Unit, Federal Bureau of Investigation Academy, Building 12, Quantico, VA 22135

After attending this presentation, attendees will understand the relationships between forensic handwriting individuality and random match probability, the use of subsampling to estimate the random match probability (RMP), and the interconnection between the strength of handwriting evidence and the size of the document.

This presentation will impact the forensic science community by providing bounds on the RMP that are based on the size of the questioned document, as well as quantifying the effect of the sample size of the document on the strength of handwriting evidence.

Forensic handwriting individuality refers to the proposition that each individual in a population has a unique writing profile. An empirical study cannot validate this proposition due to the impossibility of observing sample documents written by each person in a relevant population. However, the proposition that writing profiles are unique is one of the key premises underlying forensic handwriting comparisons.

In Saunders et al. (2008),[1] the relationship between forensic handwriting individuality, biometric individuality, and random match probabilities using a convenience sample of documents collected from 100 individuals was explored. The random match probability (RMP) of interest in handwriting analysis is the chance of randomly selecting two individuals from some relevant population and then randomly selecting a writing sample from each individual that are declared to "match" by a specific biometric matcher. A complementary probability, the random non-match probability (RNMP), is the chance of randomly selecting an individual and then randomly selecting two writing samples from the selected individual's body of handwriting that fail to "match" using the chosen biometric matcher. The RMP and the RNMP both depend upon the biometric matcher used and the sizes of the documents compared, as well as the relevant population from which individuals are selected.

Generally speaking, the RMP decreases as the writing sample size increases, with the theoretical minimum being the biometric individuality. The biometric individuality (of a population with respect to a comparison methodology) is defined as the probability that two (different) randomly selected writers from the population have indistinguishable writing profiles (with respect to the comparison methodology being used). Intuitively, two writing profiles being indistinguishable means that one concludes that the handwriting of two writers looks the "same" after observing their entire body of handwriting. The biometric individuality is bounded below by the handwriting individuality of the population. Therefore, a bound on the RMP will bound the biometric individuality and the handwriting individuality of a given population.

Using the subsampling routines presented in Saunders et al. (2008),[2] the behavior of the RMP is consistently estimated through an empirical subsampling routine. Research to investigate how subsampling from available writing samples in a data set can be used to investigate the dependency of the RMP and RNMP on the sizes of the writing samples being compared. The consistency of the subsampling estimates is dependent only on the number of writers. Based on an Federal Bureau of Investigation database of about 500 writers with approximately five writing samples per writer (the "Federal Bureau of Investigation500" data), results of the modeling will be presented and review the corresponding implications to a handwriting individuality study.

The subsampling routines provide a natural way to estimate the strength of handwriting evidence as a function of the subject's writing sample size. When evaluating the strength of evidence, there are two competing hypotheses of interest. The first hypothesis is that the subject is the writer of the questioned document, and the second hypothesis is that that random person (not the subject) is the writer of the questioned document. To evaluate the strength of the evidence, a likelihood ratio is used which compares two probabilities:

The probability of a match between the questioned document and a writing sample of a given size obtained from the subject. The probability of a match between the questioned document and a writing sample (of the same given sample size) obtained from a randomly selected individual from an appropriate population.

The results of a short empirical study from the Federal Bureau of Investigation500 data that models the strength of evidence as a function of the writing sample size is presented.

The proposed estimation methods and the associated conclusions concerning handwriting uniqueness will be illustrated and compared using the handwriting biometric identifiers.

References:

1. Saunders, et al., "A Comparison Between Biometric and Forensic Handwriting Individuality," *Journal of Forensic Science* (submitted 2008).

2**.** Saunders, et al., "Modeling the Relationship Between Random Match/No-Match Probabilities and the Sizes of Writing Samples," *Seventh International Conference on Forensic Inference and Statistics, Lausanne, Switzerland (2008).*
                 Handwriting Individuality, Random Match Probability, Subsampling