Nonparametric Statistical Data Analysis of Fingerprint Minutiae Exchange with Two-Finger Fusion

Jin Chu Wu and Michael D. Garris

Image Group, Information Access Division, Information Technology Laboratory

National Institute of Standards and Technology, Gaithersburg, MD 20899

Abstract

A nonparametric inferential statistical data analysis is presented. The utility of this method is demonstrated through analyzing results from minutiae exchange with two-finger fusion. The analysis focused on high-accuracy vendors and two modes of matching standard fingerprint templates: 1) Native Matching - where the same vendor generates the templates and the matcher, and 2) Scenario 1 Interoperability - where vendor A's enrollment template is matched to vendor B's authentication template using vendor B's matcher. The purpose of this analysis is to make inferences about the underlying population from sample data, which provide insights at an aggregate level. This is very different from the data analysis presented in the MINEX04 report in which vendors are individually ranked and compared. Using the nonparametric bootstrap bias-corrected and accelerated (BC_a) method, 95 % confidence intervals are computed for each mean error rate. Nonparametric significance tests are then applied to further determine if the difference between two underlying populations is real or by chance with a certain probability. Results from this method show that at a greater-than-95 % confidence level there is a significant degradation in accuracy of Scenario 1 Interoperability with respect to Native Matching. The difference of error rates can reach on average a two-fold increase in False Non-Match Rate. Additionally, it is proved why two-finger fusion using the sum rule is more accurate than single-finger matching under the same conditions. Results of a simulation are also presented to show the significance of the confidence intervals derived from the small size of samples, such as six error rates in some of our cases.

Keywords: bootstrap, fingerprint matching, inferential, interoperability, minutiae exchange, nonparametric, significance test, standard templates, statistical data analysis

1. Introduction

The purpose of this paper is to demonstrate the utility of applying nonparametric inferential statistics to biometric test results. There are significant advantages of this approach. Since there is no underlying distribution model for fingerprint data, the statistical data analysis must be model independent [1][2]. This nonparametric method is applicable on small sizes of samples where the Central Limit Theorem cannot be applied. This is particularly useful when the availability of samples is limited or the cost of generating more samples is prohibitively high. Additionally, the statistics invoked in this article are inferential rather than descriptive. In this way, properties of the population are inferred from the sample, which provides potentially deep insights rather than analyses that focus only on individuals in the sample.

This study was conducted on two-finger fusion results derived from the Minutiae Interoperability Exchange Test 2004 (MINEX04), which was organized and administered by the National Institute of Standards and Technology (NIST), and a complete description is found at [3] including an executive summary and report [4]. The two-finger fusion was conducted by using the sum rule. A proof is provided in Appendix A that shows why the two-finger sum rule is more accurate than single-finger matching results. The MINEX04 report conducted extensive analyses intended to rank participants and to determine a group of participants that demonstrate interoperability in different modes (or categories) above a minimum level of accuracy. The purpose of the nonparametric data analysis presented in this paper is different in that it is not primarily intended to rank participants, but rather, make statistically robust observations regarding the collective capabilities of the participants (the population) by analyzing their aggregate performance.

A brief overview of MINEX04 is provided in Section 2. Section 3 describes the steps taken to conduct the nonparametric data analysis and presents the statistical results. This includes selection of a representative sample of participants, comparison of mean performances, computing 95 % confidence intervals, and then conducting statistical significance tests. Conclusions are drawn in the final section.

2. MINEX04 Overview

The purpose of the MINEX04 test was to determine the feasibility of using minutiae data (rather than image data) as the interchange medium for fingerprint information between different fingerprint matching systems. The results of MINEX04 have implications that affect planning decisions for projects such as Personal Identity Verification (PIV). PIV was initiated by Homeland Security Presidential Directive 12 [5]. This mandated the establishment of a common identification standard for federal employees and contractors. It required interoperable use of identity credentials to control physical and logical access to federal government locations and systems.

MINEX04 was designed to evaluate whether various populations and combinations of encoding schemes, enrolled templates, probe templates, and fingerprint matchers will produce successful matches. There were two categories of encoding schemes; the first were proprietary minutiae templates generated by the participants (called vendors); the second were standard minutiae templates. These standard templates are based on INCITS (International Committee for Information Technology Standards) 378 Finger Minutiae Format for Data Interchange [6]. There were two standard template types evaluated in MINEX04, but for the purposes of this study, we focus on just the results of using the standard 'A' templates nicknamed "MIN:A", which contain only the minutiae attributes $\{x, y, \theta, type, quality\}$.

A total of 14 vendors participated in MINEX04. These vendors are identified in the MINEX04 report and subsequently in this report by assigned letters. The identities of these vendors are not germane to the purpose of this report, so identities are not revealed herein, however the vendor key is published in the full MINEX04 report. Each vendor had to supply NIST with a software development kit (SDK) that

- creates an INCITS 378 MIN:A template from an image
- produces a comparison score from two MIN:A templates

In this way, matching accuracy could be computed and compared in combinations of three dimensions: a gallery (enrolment) template, matched to a probe (authentication) template, matched with a specific vendor's fingerprint minutiae matcher. A simple nomenclature "XY_Z" has been adopted to represent the possible combinations, where X represents the vendor that generated the enrollment template, Y represents the vendor that generated the authentication template, and Z is the vendor that developed the template matcher.

Standard template matching within MINEX04 was tested in two modes. The first comprised of standard templates being generated and matched by the same vendor, referred to as *Native Matching* and nicknamed "MIN:A.XX_X" The second involved testing the interoperability of matching a standard template from one vendor with a standard template generated by a different vendor and then matched potentially by yet another different vendor.

While all the possible combinations of interoperability were studied in MINEX04, there is one combination that has greatest operational relevance. This is the interoperable scenario where a subject is enrolled in vendor-P's system, but then attempts to authenticate with a different vendor-Q's system. This is the case when a person enrolled by one agency's system visits and presents his credentials to a different agency. In this scenario, the subject presents his finger to vendor-Q's system and a standard template is generated; this template is then matched to vendor-P's enrolled template with the match being conducted by vendor-Q's matcher. This is referred to as *Scenario 1 Interoperability* and nicknamed "MIN:A.YX_X". For the purposes of this study, only MIN:A.XX_X and MIN:A.YX_X template-matcher combinations are analyzed.

MINEX04 used four different and distinct collections of fingerprints (called datasets) named: POEBVA, POE, DOS, & DHS2. A description of these datasets and their NIST Fingerprint Image Quality (NFIQ) distributions are documented in the MINEX04 report. All datasets used were comprised of left and right-index fingers using live-scan plain impressions. The subject sample sizes of each dataset were 60 thousand mates and 120 thousand non-mates. The testing was performed by using the second instance of the mates as the enrollment image and the first instance as the authentication image. So for each dataset there were 60 thousand mate (genuine) template comparison scores. The non-mate scores were generated by comparing the non-mate authentication samples to the same enrollment images used with the mates. This generated 120 thousand non-mate (impostor) template comparison scores.

One and two-finger authentication was evaluated in the MINEX04 test. The two-finger comparison scores were produced in a score-level fusion process by summing a subject's left and right-index finger comparison scores. Given a set of genuine and a set of impostor two-finger template comparison scores, performance measures of False Non-Match Rate (FNMR) and False Match Rate (FMR) were computed and Detection Error Tradeoff (DET) characteristic curves compared. The analysis in this paper focuses only on two-finger fusion results. Note the proof in Appendix A.

The scope of the MINEX04 test was large and varied as fully documented in Reference [3]. Only a subset of these results is used in the data analysis herein. This includes results from all four datasets, with standard MIN:A templates, on two-finger fusion scores, and with only two template-matchers combinations (Native Matching and Scenario 1 Interoperability) as described in the following section.

3. Nonparametric Inferential Statistical Data Analysis

The purpose of the nonparametric inferential statistical data analysis presented in this paper is to make statistical inferences regarding the collective capabilities of the fingerprint matching community (the population) based on analyzing the aggregate performance of fingerprint matching vendors (the sample). The utility of this statistical method is therefore greatest when addressing operational decisions requiring an assessment of aggregate performance.

In the case of the various combinations of template generators and matchers, one can think of an information technology manager who is responsible for procuring and deploying standard template fingerprint authentication systems within his enterprise. Furthermore, he/she may be required to procure standard template technology from a pool of competent providers and not just from a single source. The category MIN:A.XX_X represents this case. This manager must also consider visitors gaining access to his enterprise via credentials enrolled and issued by various vendors not within his control. What is the expected impact on the enterprise when dealing with standard templates interoperably? The category MIN:A.YX_X represents this case. What might this manager infer from the test results regarding the pool of such potential vendors?

These types of questions can be addressed by a nonparametric inferential statistical approach. The properties of the population of MIN:A.XX_X and the population of MIN:A.YX_X will be analyzed and compared.

3.1   The Method

The statistical approach to the data analysis in this paper is as follows. A group of high-accuracy vendors is selected from those who participated in the test. The mean of the sample is computed. To compute a confidence interval about the mean, the nonparametric bootstrap is used by making no assumption about the distribution of population. The sample, which can be relatively small in size, is replicated through a process of resampling with replacement and a confidence interval is calculated from the replicated set. Next, to determine if two samples are significantly different and to make the test results more convincible, nonparametric significance tests are applied. The details and results of each of these steps as applied to MIN:A.XX_X and MIN:A.YX_X are described in the following sections.

3.2   Sample Selection

In our analysis we desire to gain understanding and insight into the capabilities of vendors who demonstrate a relatively high level of matching accuracy. Not all the vendors participating in MINEX04 demonstrated such a desirable level. It was necessary therefore to determine a pool of sufficiently accurate vendors for us to conduct our analysis. The participation in the test is voluntary. Thus, such a sample selection focusing on high-accuracy vendors is random.

The Ongoing MINEX tests [7] were initiated while conducting the analyses in this paper. A primary purpose to the Ongoing MINEX tests is to evaluate and publish a certified list of vendors that exhibit a level of template matching interoperability above a minimum level of accuracy. This level of accuracy is set such that all vendors in the interoperable group achieve, within the context of Scenario 1 Interoperability, a FNMR less than or equal to 0.01 at a fixed FMR of 0.01. The details of this process and the current list of *compliant* vendors are posted on the Ongoing MINEX website. For the purposes of this report, we chose to analyze the six vendors {A,B,C,D,F,G} whose matcher performance in MINEX04 was determined compliant

using the Ongoing MINEX criteria. It is noted that an ad hoc sample selection of vendors was conducted independent of the Ongoing MINEX testing, and the results were complementary.

3.3   Comparison of Means

Given the selected set of six vendors {A,B,C,D,F,G}, error rates from each vendor performing Native Matching on each of the four datasets were calculated. This generated six error measures per dataset, and the mean Native Matching (MIN:A.XX_X) error rate was computed for each dataset. Error rates were also calculated from every combination of two vendors performing Scenario 1 Interoperability. This generated thirty error measures per dataset (each of six vendors used interoperably with the remaining five vendors), and the mean Scenario 1 Interoperability (MIN:A.YX_X) error rate was computed for each dataset.

Error rates were computed as the resulting FNMR at a specified level of FMR. Two levels of FNMR were computed and compared in this analysis. The first was FNMR at an FMR of 0.01, and results are reported in Table 1. The second level was FNMR at an FMR of 0.001, and these results are reported in Table 2. Comparing the means between MIN:A.XX_X and MIN:A.YX_X within each table, there is a consistent degradation in values when switching from Native Matching to Scenario 1 Interoperability. How accurate are these means?

	FMR = 0.01							
	POEBVA		POE		DOS		DHS2	
	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval
MIN:A.XX_X	0.0022	(0.0013, 0.0028)	0.0024	(0.0013, 0.0031)	0.0063	(0.0043, 0.0081)	0.0133	(0.0083, 0.0256)
MIN:A.YX_X	0.0048	(0.0041, 0.0057)	0.0049	(0.0041, 0.0057)	0.0117	(0.0102, 0.0134)	0.0183	(0.0146, 0.0237)

Table 1. Means and 95 % confidence intervals of FNMR with FMR at 0.01 by dataset.

	FMR = 0.001							
	POEBVA		POE		DOS		DHS2	
	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval	Mean FNMR	Conf. Interval
MIN:A.XX_X	0.0045	(0.0028, 0.0061)	0.0043	(0.0026, 0.0054)	0.0120	(0.0083, 0.0150)	0.0195	(0.0138, 0.0315)
MIN:A.YX_X	0.0099	(0.0081, 0.0120)	0.0095	(0.0080, 0.0113)	0.0204	(0.0176, 0.0236)	0.0295	(0.0242, 0.0372)

Table 2. Means and 95 % confidence intervals of FNMR with FMR at 0.001 by dataset.

3.4   95 % Confidence Intervals

To answer this question, the 95 % confidence intervals in different cases were computed using the nonparametric bootstrap bias-corrected and accelerated (BC_a) method [8].   The statistical computing program R (Version 2.0.1) was used with the package "boot."[1]   For the MIN:A.XX_X category from each dataset, the six error measurements were re-sampled with replacement up to 100000 bootstrap replications, and the limits of the nonsymmetrical 95 % confidence interval were computed.   The same process was followed for the MIN:A.YX_X category. Only in each case, thirty error measurements were re-sampled with replacement.   The resulting confidence intervals are reported along side their corresponding means in Table 1 & Table 2.   A simulation is provided in Appendix B that shows the significance of the confidence intervals computed from a sample of six error rates for the MIN:A.XX_X category.

The means and their associated confidence intervals for the MIN:A.XX_X and MIN:A.YX_X categories are plotted side by side for each dataset in Figure 1 & Figure 2 for two different values of FMR, respectively.   Looking at the results from the first three datasets in both figures, a clear pattern is observed.   The 95 % confidence intervals for MIN:A.XX_X are completely separated with no vertical overlap to the 95 % confidence intervals for the corresponding MIN:A.YX_X sample.   Comparing the results for the POEBVA dataset, the average error rate for Scenario 1 Interoperability is more than twice that of Native Matching with standard templates.   This is true for both FNMR measured at FMR=0.01 and FMR=0.001.

The results are quite different for the DHS2 dataset.   In this case, the mean error rates are considerably higher and the 95 % confidence intervals are much larger and the intervals for MIN:A.XX_X overlap largely with the corresponding intervals for MIN:A.YX_X.   This dataset is known to have unique image quality characteristics resulting in much poorer image quality as reflected in the MINEX04 Report which lists DHS2 as having the largest percentage of worst (NFIQ Quality 5) fingerprints.   The differences observed with DHS2 in our analysis are attributed to the difference in image quality of this dataset.   The confidence intervals are particularly large due the existence of an outlier from one vendor's sample distribution of FNMRs.

The results from the other three datasets are consistent.   These datasets are comprised of fingerprints captured with more modern sensors and quality control processes, while DHS2 contains legacy data captured with older sensor technology and with fewer quality controls. Therefore, the statistical results with DHS2 are considered not representative and dismissed.   The DHS2 results are intentionally left in this report as they point out how critical proper test design and sample selection are to achieving relevant conclusions using nonparametric inferential methods.

---

[1] Specific hardware and software products identified in this paper were used in order to perform the analyses described herein. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.
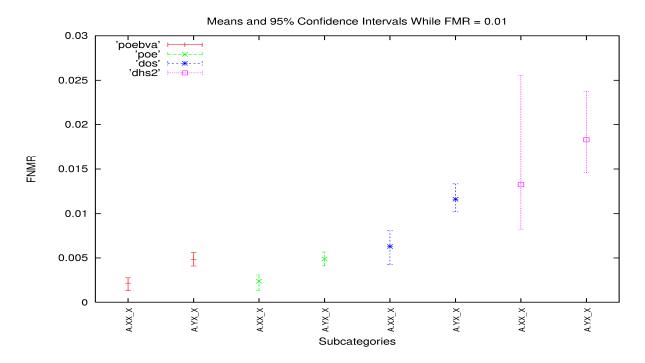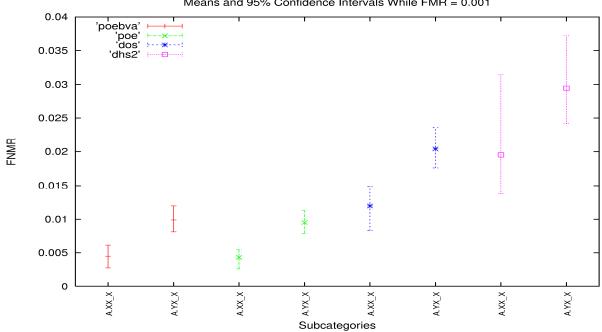
Figure 1. Means FNMR and 95 % confidence intervals with FMR at 0.01 by dataset

Figure 2. Means FNMR and 95 % confidence intervals with FMR at 0.001 by dataset

3.5   Statistical Significance Tests

Are all the above observed differences significant? To help answer this question, two nonparametric statistical significance tests, the permutation test [9] and the Wilcoxon rank sum test [10], are applied. Both test the null hypothesis (that there is no difference in distributions) between two populations. Two-sided p-values are generated that indicate the significance level whether the two distributions are the same. A p-value less than 5 % then represents that the two distributions are likely to be different with a confidence level of greater than 95 %.

The results of the permutation test are listed in Table 3. The first row of values are from samples of FNMR where FMR = 0.01, and the second row are from samples of FNMR where FMR = 0.001. The statistical computing program S-Plus (Version 7.0.0) with the package "resample" was used to compute p-values.   For each dataset, the MIN:A.XX_X sample and the MIN:A.YX_X sample were input to S-Plus and p-values were computed. Once again a pattern for the first three datasets is observed in that all p-values are well less than 0.05 indicating that the underlying populations of MIN:A.XX_X are significantly different from their corresponding populations of MIN:A.YX_X. The results on dataset DHS2 are strikingly different with p-values considerably larger than 0.05. This confirms the large and overlapping confidence intervals observed for this dataset in the previous figures for the same reasons.

Results of the Wilcoxon rank sum test are listed in Table 4 and were computed using R. The Wilcoxon rank sum test is used because samples between the MIN:A.XX_X and MIN:A.YX_X categories are unpaired.   The same pattern of p-values is observed confirming the earlier permutation test results and the mean error rates along with confidence intervals.

Category 1	Category 2	Permutation Test				FMR
		POEBVA	POE	DOS	DHS2	
MIN:A.XX_X	MIN:A.YX_X	0.0022	0.0053	0.0027	0.4214	0.01
MIN:A.XX_X	MIN:A.YX_X	0.0089	0.0034	0.0119	0.1773	0.001

Table 3 Two-sided p-values of the permutation test with FMR at 0.01 and 0.001, by dataset

Category 1	Category 2	Wilcoxon Rank Sum Test				FMR
		POEBVA	POE	DOS	DHS2	
MIN:A.XX_X	MIN:A.YX_X	0.0024	0.0063	0.0032	0.1463	0.01
MIN:A.XX_X	MIN:A.YX_X	0.0074	0.0063	0.0290	0.1721	0.001

Table 4 Two-sided p-values of Wilcoxon rank sum test with FMR at 0.01 and 0.001, by dataset

4. Conclusions

The application of nonparametric inferential statistics has been demonstrated on fingerprint minutiae exchange with two-finger fusion. Advantages include no assumptions of an underlying distribution model, the handling of confidence intervals and significance tests, small sample sizes can be used, and properties of the population are inferred providing important insights. The method begins with careful sample selection. If the sample is not representative, then results will be irrelevant and potentially misleading as demonstrated with the results reported on dataset DHS2. Given a sample, the 95 % confidence intervals are computed using the nonparametric bootstrap BC_a method. The underlying populations of two samples are then compared using nonparametric significance tests such as the permutation test and the Wilcoxon rank sum test. Using this method, MINEX04 results were studied. Six high-accuracy vendors were selected and their ability to match standard fingerprint templates natively (MIN:A.XX_X) and interoperably (MIN:A.YX_X) were compared using two-finger fusion. Analysis showed that at a greater-than-95 % confidence level there is a significant degradation in accuracy of Scenario 1 Interoperability with respect to Native matching. The difference of error rates can reach on average a two-fold increase in FNMR. A proof was also provided to show why two-finger fusion using the sum rule is better than single-finger matching results under the same conditions. Results of a simulation using the nonparametric bootstrap are also reported that show the significance of the confidence intervals derived from the small size of samples in our case.

APPENDIX A.   Proof of Two-Finger Fusion Sum Rule

There are several ways to deal with two-finger fusion from the output of single-finger matching. For instance, the sum rule adds up the two similarity scores that are generated from the right-index finger matching and the left-index finger matching. An alternative is the maximum rule where the maximum of the two similarity scores is selected as the fused score. In this paper, the sum rule is adopted.

A qualitative proof is as follows showing why two-finger fusion using the sum rule improves results of single-finger matching in terms of the operational criteria under the same conditions. Suppose that for right-index finger matching, a score G_r is selected from the distribution of genuine comparison scores, and a score I_r is selected from the distribution of impostor comparison scores. And for left-index finger matching, the corresponding genuine score and impostor score are G_l and I_l, respectively. Assume further that $I_r<G_r$ and $I_l<G_l$. Indeed, this assumption is valid in most cases. The distances between genuine scores and impostor scores in the single-finger case are $D_r = G_r - I_r$ and $D_l = G_l - I_l$, respectively. Using the sum rule for two-finger fusion, the fused genuine score is $(G_r + G_l)$ and the fused impostor score is $(I_r + I_l)$. The distance between these two fused scores is $D_f = (G_r + G_l) - (I_r + I_l) = D_r + D_l$. This indicates that D_f must be greater than D_r and D_l, respectively. In general, the greater is the distance between distributions of genuine and impostor scores, the more accurate the fingerprint matcher [1]. Therefore, the accuracy using the fused two-finger scores is increased.

The effect of the sum rule is to create greater separation between the genuine and impostor score distributions and thus reduce the overlapping area of the two distributions. This can also be evidenced by examining the discrete probability distribution functions for the genuine and impostor comparison scores. In this case, MIN:A.XX_X results from a single vendor on the POEBVA dataset are presented. Figure 3 & Figure 4 show the results from matching individually the right-index finger and the left-index finger, respectively. Figure 5 shows the results of two-finger fusion using the same vendor's technology. Note that the probabilities in these three figures have been cut at 0.001 in order to show clearly the relative positions of two distributions of the genuine and impostor scores. Comparing the two-finger fusion results with the single-finger results demonstrates the effect of the sum rule in that there is greater separation and less overlap of the genuine and impostor distributions in Figure 5.

As a result, the sum rule of two-finger fusion improves the shape of the receiver operating characteristic (ROC) curve as opposed to using one-finger matching under the same conditions [1]. An ROC curve is created by applying the operational criteria of a score threshold across the genuine and impostor distributions measuring the True Accept Rate (TAR)[2] and FMR at each threshold. An ROC analysis evaluates vendors' performance. It is shown in Figure 6 that the ROC curve of two-finger fusion invoking the sum rule is higher than the ROC curves using one-finger matching.
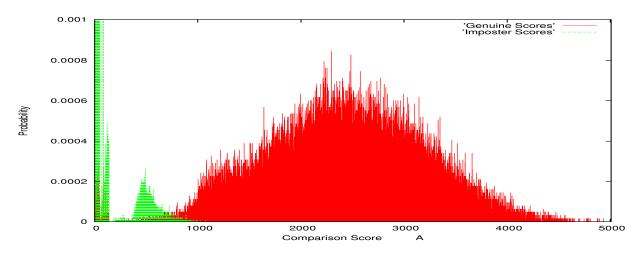
---

[2] Equivalently, (FNMR = 1 – TAR).

Figure 3. **The discrete probability distribution functions of the genuine and impostor scores for right-index finger matching.**
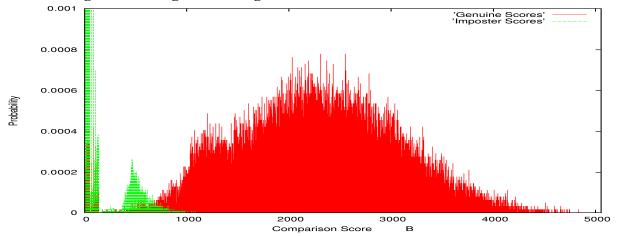

Figure 4. **The discrete probability distribution functions of the genuine and impostor scores for left-index finger matching.**

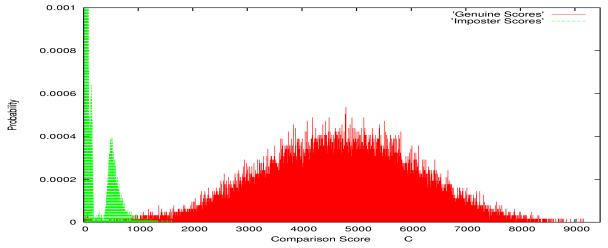**Figure 5.** **The discrete probability distribution functions of the genuine and impostor scores for two-finger fusion using the sum rule.**
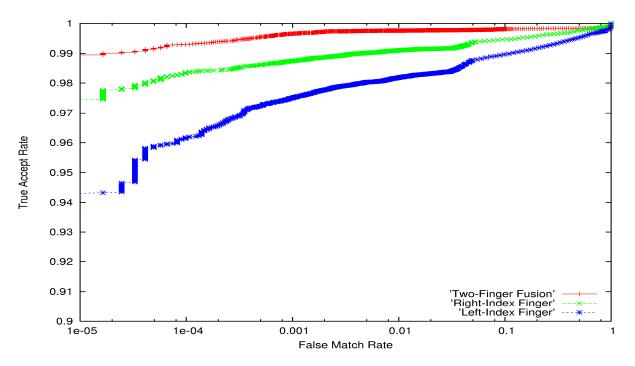
Figure 6 The three ROC curves of the right-index finger, the left-index finger, and the two-finger fusion.

APPENDIX B.   Significance of Confidence Intervals

The 95 % confidence intervals for the MIN:A.XX_X category were obtained using the bootstrap BC_a method on error measurements taken from a set of six vendors. With a sample size of six, one might question whether the results are significant. This section addresses this question, first in a theoretical manner [8][11], followed by supporting evidence from a simulation.

Suppose that a bootstrap is carried out on n distinct values. It implies that the bootstrap space is n^n in the sense that each possible bootstrap sample is selected with equal probability. Further, assume that a multinomial-distribution vector of the bootstrap is $(k_1, k_2, \ldots, k_n)$, where k_i stands for the number of times the ith value is selected in a bootstrap sample, subject to

$$0 \le k_i \le n, \, i = 1, \ldots, n; \text{ and } \sum_{i=1}^{n} k_i = n \,.$$

All bootstrap samples, corresponding to a bootstrap vector, can result in the same bootstrap replication of the considered statistic such as mean etc. In this sense, the number of distinct bootstrap samples is $\binom{2n-1}{n-1}$.

A distinct bootstrap sample has the following multinomial probability being selected,

$$\text{Prob}\,(k_1, k_2, \cdots, k_n) = \frac{n!}{k_1! k_2! \cdots k_n!} * \frac{1}{n^n} \,.$$

The original sample of n distinct values is obtained while all k_i's are equal to 1, and thus the corresponding probability is $n!/n^n$. It follows from the above probability formula that this probability is the largest one among all others.

Therefore, while the sample size is six for the MIN:A.XX_X category, the bootstrap space is 6^6 = 46,656, the number of distinct bootstrap samples is $\binom{11}{5}$ = 462, and the probability of selecting the observed statistic is $6!/6^6$ = 1.54 %. All these indicate that the confidence intervals of the mean derived from a sample of size six using the bootstrap are significant.

To support this position, a simulation was conducted on the MIN:A.XX_X results that were presented in Table 1 for the POEBVA dataset. The associated mean and 95 % confidence interval are shown in the top-right of Table 5 in the row labeled Trial 0. To the left of these statistics are the underlying FNMR measurements (at an FMR = 0.01) contributed by each of the six vendors.

For each of the subsequent trails (Trials 1-5) reported in Table 5, the underlying genuine and impostor scores from each vendor were re-sampled with replacement, respectively, and an FNMR was computed at FMR = 0.01 from two new distributions. These simulated FNMR's are reported under each vendor column in the table. Their mean is reported to the right, and the bootstrap BC_a method was applied to the simulated FNMR's to compute new 95 % confidence intervals, which are also reported.

Comparing the resulting simulated means and 95 % confidence intervals in the table with the original results of Trial 0, it is clear that the means as well as the upper bounds and the lower bounds of the confidence intervals fluctuate only by no more than ±0.0002. This indicates that the simulation results are very stable. It follows that the means and 95 % confidence intervals reported in Table 1 & Table 2 are significant, even though for the MIN:A.XX_X category only six error rates were used in the bootstrap.

Trial	MIN:A.XX_X						Mean	Conf. Interval
	AA_A	BB_B	CC_C	DD_D	FF_F	GG_G		
0	0.0024	0.0024	0.0032	0.0013	0.0031	0.0007	0.0022	(0.0013, 0.0028)
1	0.0026	0.0026	0.0033	0.0013	0.0035	0.0007	0.0023	(0.0014, 0.0030)
2	0.0023	0.0024	0.0032	0.0011	0.0033	0.0008	0.0022	(0.0013, 0.0029)
3	0.0024	0.0027	0.0031	0.0013	0.0029	0.0006	0.0021	(0.0012, 0.0028)
4	0.0024	0.0027	0.0030	0.0014	0.0035	0.0006	0.0023	(0.0013, 0.0029)
5	0.0019	0.0020	0.0035	0.0013	0.0028	0.0007	0.0020	(0.0013, 0.0028)

Table 5. Original results (Trial 0) of mean FNMR and 95 % confidence interval compared to bootstrap simulation results (Trails 1-5) computed by resampling the underlying genuine and impostor scores of Trial 0. (All results are from MIN:A.XX_X on the POEBVA dataset with FNMR's computed at an FMR = 0.01.)

REFERENCES

[1] J.C. Wu, C.L. Wilson; Nonparametric Analysis of Fingerprint Data on Large Datasets; Pattern Recognition, 2007; in press.

[2] J.C. Wu, C.L. Wilson; An Empirical Study of Sample Size in ROC-Curve Analysis of Fingerprint Data; Proceedings of SPIE Volume: 6202; Biometric Technology for Human Identification III, Editors: P.J. Flynn, S. Pankanti, April 2006.

[3] Minutiae Interoperability Exchange Test 2004 (MINEX04); http://fingerprint.nist.gov/minex04/index.html.

[4] P. Grother, et al; MINEX – Performance and Interoperability of the INCITS 378 Fingerprint Template; NISTIR 7296; National Institute of Standards and Technology; March 21, 2006.

[5] HSPD-12; http://www.whitehouse.gov/news/releases/2004/08/20040827-8.html.

[6] ANSI-INCITS 378-2004; Fingerprint Minutiae Format for Data Interchange 2004; American National Standard.

[7] On-going Minutiae Interoperability Exchange Test; http://fingerprint.nist.gov/MINEX/ .

[8] A.C. Davison, D.V. Hinkley; Bootstrap Methods and Their Application; Cambridge University Press, Cambridge; 2003.

[9] F. Pesarin; Multivariate Permutation Tests: With Applications in Biostatistics; John Wiley & Sons, Ltd; New York; 2001.

[10] B. Ostle, L.C. Malone; Statistics in Research: Basic Concepts and Techniques for Research Workers; fourth ed.; Iowa State University Press; Ames; 1988.

[11] B. Efron; The Jackknife, the Bootstrap and Other Resampling Plans; Society for Industrial and Applied Mathematics, J.W.Arrowsmith Ltd; Philadelphia, 1982.