Validating a Biometric Authentication System:[1] Sample Size Requirements

Sarat C. Dass, Yongfang Zhu

Dept. of Statistics & Probability
Michigan State University
E. Lansing, MI 48823

Anil K. Jain

Dept. of Computer Science & Engineering
Michigan State University
E. Lansing, MI 48823

ABSTRACT

Authentication systems based on biometric features (e.g., fingerprint impressions, iris scans, human face images, etc.) are increasingly gaining widespread use and popularity. Often, vendors and owners of these commercial biometric systems claim impressive performance based on some propriety data. In such situations, there is a need to independently validate the claimed performance levels. System performance is typically evaluated by collecting biometric templates from n different subjects, and for convenience, acquiring multiple instances of the biometric for each of the n subjects. Very little work has been done in constructing confidence regions for the claimed performance levels and also in determining the required number of subjects needed to establish pre-specified confidence regions. To simplify the analysis, several previous studies have assumed that multiple acquisitions of the biometric entity are statistically independent. This assumption is too restrictive and is generally not valid. In this paper, we present a semi-parametric model based on multivariate copula models for correlated biometric acquisitions. Based on this approach, confidence bands for the claimed performance, expressed in terms of the ROC curve, are constructed. We also give the minimum number of subjects needed to achieve a width of 1% for the ROC confidence bands. Using fingerprints as our biometric, experimental results show that we need approximately 11,000 subjects to achieve a width of 1% for the 95% ROC confidence band. The required number of subjects increases when either one or both of the inter-finger and intra-finger correlations increase. For achieving the desired confidence level, the required number of subjects based on our method is much smaller compared to other methods of selecting sample sizes, such as the subset bootstrap.

Keywords: Biometric authentication, Error estimation, Gaussian copula models, bootstrap, ROC confidence bands.

I. INTRODUCTION

The purpose of a biometric authentication system is to validate the claimed identity of a user based on his/her physiological characteristics. In such a system operating in the verification mode, we are interested in accepting queries which are "close" or "similar" to the template of the claimed identity, and rejecting those that are "far" or "dissimilar". Suppose a user with true identity I_t supplies a biometric query Q and a claimed identity I_c. We are interested in testing the hypothesis

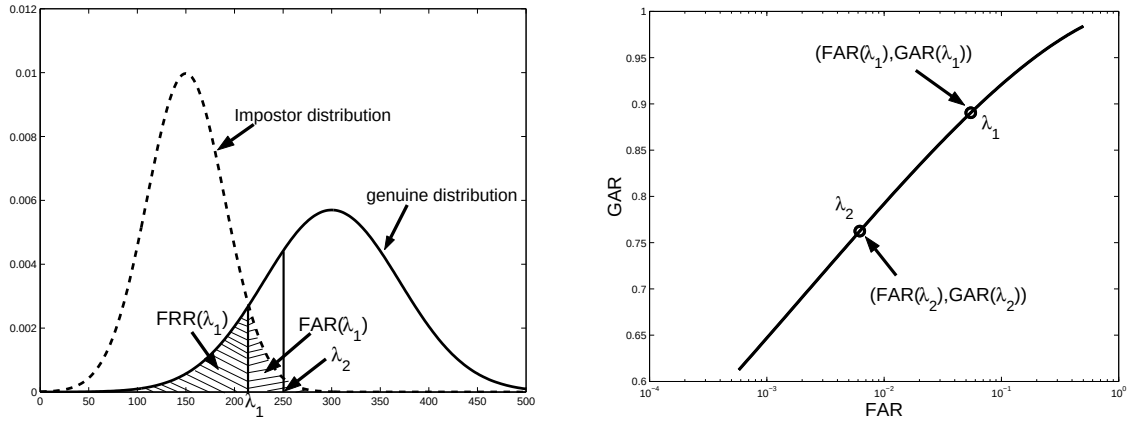
$$H_0 : I_c = I_t \quad \text{vs.} \quad H_1 : I_c \neq I_t \tag{1}$$

based on the query Q and the template T of the claimed identity in the database; in Equation (1), H_0 (respectively, H_1) is the null (alternative) hypothesis that the user is genuine (impostor). The testing in (1) is carried out by computing a similarity measure, $S(Q, T)$ where large (respectively, small) values of S indicate that T and Q are close to (far from) each other. A threshold, λ, is specified so that all similarity values lower (respectively, greater) than λ lead to the rejection (acceptance) of H_0. Thus, when a decision is made whether to accept or reject H_0, the testing procedure (1) is prone to two types of errors: the false reject rate (FRR) is the probability of rejecting H_0 when in fact the user is genuine, and the false accept rate (FAR) is the probability of accepting H_0 when in fact the user is an impostor. The genuine accept rate (GAR) is $1-$FRR, which is the probability that the user is accepted given that he/she is genuine. Both the FRR and the FAR (and hence GAR) are functions of the threshold value λ (see Figure 1 (a)). The Receiver Operating Curve (ROC) is a graph that expresses the relationship between the FAR versus GAR when λ varies, that is,

$$ROC(\lambda) = (FAR(\lambda), GAR(\lambda)), \tag{2}$$

and is commonly used to report the performance of a biometric authentication system (see Figures 1 (a) and (b)).

In marketing commercial biometric systems, it is often the case that error rates are either not reported or poorly reported (i.e., reported without giving details on how it was determined). In a controlled environment such as in laboratory experiments, one may achieve very high accuracies when the underlying biometric templates are of very good quality. However, these accuracies may not reflect the true performance of the biometric system in real field applications where uncontrolled factors such as noise and distortions can significantly degrade the system's performance. It is common to report the performance of a biometric system in terms of its ROC curve. Thus, the problem we address in this paper is the validation of a claimed ROC curve, $\text{ROC}_c(\lambda)$, by a biometric vendor. Of course, reporting just $\text{ROC}_c(\lambda)$ does not give the complete picture. One should also report as much information as one



(a) (b)

Fig. 1. Obtaining the ROC curve by varying the threshold λ. Panel (a) shows the FRR and FAR corresponding to a threshold λ_1. λ_2 is another threshold different from λ_1. Panel (b) shows the ROC curve obtained when λ varies. The values of (FAR, GAR) on the ROC curve corresponding to the thresholds λ_1 and λ_2 are shown.

can about the underlying biometric samples, such as the quality, the sample acquisition process, sample size as well as a brief description of the subjects themselves. If the subjects used in the experiments for reporting $\text{ROC}_c(\lambda)$ are not representative of the target population, then $\text{ROC}_c(\lambda)$ is not very useful. But assuming that the underlying samples are representative and can be replicated by other experimenters under similar conditions, one can then proceed to give margins of errors for validating $\text{ROC}_c(\lambda)$.

The process of obtaining biometric samples usually involves selecting n individuals (or, subjects) and using c different biometrics from each individual (such as different fingers of each subject, or iris images of the left and right eyes, etc.). Additional biometric samples can be obtained by sampling each biometric multiple times, d, over a period of time. It is well known that multiple acquisitions corresponding to each biometric exhibit a certain degree of dependence (or, correlation); see, for example, [1], [3], [9], [15]–[18]). There have been several earlier efforts to validate the performance of a biometric system based on multiple biometric acquisitions. Bolle et al. [4] first obtained confidence intervals for the FRR (respectively, FAR) using a binomial model for the total number of false rejects (false accepts) made by the system. In order to avoid errors due to an inappropriate statistical model, they also presented a non-model based approach for obtaining confidence intervals based on the bootstrap technique. However, the bootstrap methodology assumed that the multiple biometric acquisitions were independent of each other. To account for correlation, Bolle et al. [2], [3] introduced the subsets bootstrap approach to construct confidence intervals for the FAR, FRR and the ROC curve. Schuckers [15] proposed the beta-binomial family to model the correlation between the multiple biometric acquisitions as well as to account for

varying FRR and FAR values for different subjects. He showed that the beta-binomial model gives rise to extra variability in the FRR and FAR estimates when correlation is present. However, a limitation of this approach is that it models correlation for a single threshold value. Thus, this method cannot be used to obtain a confidence region for the entire ROC curve. Further, Schucker's approach is strictly model-based; inference drawn from this model may be inappropriate when the true underlying model does not belong to the beta-binomial family.

To construct confidence bands for the ROC curve, Bolle et al. [3] select T threshold values, $\lambda_1, \lambda_2, \ldots, \lambda_T$ and compute the 90% confidence intervals for the associated FARs and GARs. At each threshold value λ_i, *combining* these 90% confidence intervals results in a confidence rectangle for $ROC(\lambda_i)$ (see (2)). Repeating this procedure for each $i = 1, 2, \ldots, T$ and *combining* the confidence rectangles obtained gives rise to a confidence region for ROC(λ). A major limitation of this approach is that the 90% confidence intervals for the FARs and GARs will neither automatically guarantee a 90% confidence rectangle at each λ_i nor a 90% confidence region for the ROC curve. In other words, ensuring a confidence level of 90% for each of the individual intervals cannot, in general, ensure a specific confidence level for the *combined* approach. This is the well-known problem of combining evidence from simultaneous hypothesis testing scenarios [8], [10], [11]: In essence, for each i, we are performing the tests

$$H_{0,i} \,:\, FAR(\lambda_i) = FAR_c(\lambda_i) \qquad \text{versus} \qquad H_{1,i} \,:\, FAR(\lambda_i) \neq FAR_c(\lambda_i), \tag{3}$$

and

$$H^*_{0,i} \,:\, GAR(\lambda_i) = GAR_c(\lambda_i) \qquad \text{versus} \qquad H^*_{1,i} \,:\, GAR(\lambda_i) \neq GAR_c(\lambda_i), \tag{4}$$

where $FAR(\lambda_i)$ (respectively, $FAR_c(\lambda_i)$) are the true but unknown (respectively, claimed) FAR at λ_i, and $GAR(\lambda_i)$ (respectively, $GAR_c(\lambda_i)$) are the true but unknown (respectively, claimed) GAR at λ_i. To test each $H_{0,i}$ (and $H^*_{0,i}$) individually, the 90% confidence interval for FAR (and GAR) can be used, and the resulting decision has a FRR of at most $100 - 90 = 10\%$. The confidence region for the ROC curve combines the $2T$ confidence intervals above and is used to test the hypothesis

$$H_0 \,:\, \cap_{i=1}^{T} \{\, H_{0,i} \cap H^*_{0,i} \,\} \qquad \text{versus} \qquad H_1 \,:\, \text{not } H_0. \tag{5}$$

However, the combined confidence region is not guaranteed to have a confidence level of 90%. In other words, the decision of whether to accept or reject H_0 does not have an associated FRR of 10% as in the case of the individual hypotheses. In fact, for a number α where $0 < \alpha < 1$, combining $2T$ $100(1-\alpha)\%$ level confidence intervals based on a-priori selected thresholds can only guarantee a lower bound of $100(1-2T\alpha)\%$ on the confidence level. This fact is based on Bonferroni's inequality, and is well-known in the statistics literature. Instead of trying to derive this inequality, we point the reader to the relevant

literature in statistics on simultaneous hypotheses testing procedures; see, for example, the following references [8], [10], [11]. The lower bound $100(1-2T\alpha)\%$ on the confidence level is not useful when T is large; in this case, $100(1-2T\alpha)\%$ is negative, and we know that any confidence level should range between 0% and 100%. In Bolle et al.'s procedure, the value of T is large since the confidence rectangles are reported at various locations of the *entire* ROC curve.

In this paper, we present a new approach for constructing confidence regions for the ROC curve with a guaranteed pre-specified confidence level. In fact, we are able to construct confidence regions for a *continuum* of threshold values, and not only for finite pre-selected threshold values. In contrast to the non-parametric bootstrap approach of [3], we develop a semi-parametric approach for constructing confidence regions for ROC(λ). This is done by estimating the genuine and impostor distributions of similarity scores obtained from multiple biometric acquisitions of the n subjects where the marginals are first estimated non-parametrically (without any model assumptions), and then coupled together to form a multivariate joint distribution via a parametric family of Gaussian copula models [12]. The parametric form of the copula models enables us to investigate how correlation between the multiple biometric acquisitions affects the confidence regions. Confidence regions for the ROC are constructed using bootstrap re-samples from our estimated semi-parametric model. The main steps of our procedure are shown in Figure 2. Note that our approach based on modeling the distribution of similarity scores is fundamentally different from that of [15], where binary (0 and 1) observations are used to construct confidence intervals for the FRRs and FARs.

Our approach also varies from that of [1], [3], [9], [15] in several respects. First, we explicitly model the correlation via a parametric copula model, and thus, are able to demonstrate the effects of varying the correlation on the width of the ROC confidence regions. We also obtain a confidence *band*, rather than confidence rectangles as in [3], consisting of an upper and lower bound for the ROC curve. Further, the confidence bands come with a guaranteed confidence level for the *entire* ROC in the region of interest. Thus, we are able to perform tests of significance for the ROC curve and report error rates corresponding to our decision of whether to accept or reject the claimed ROC curve.

Another important issue that we address is that of the test sample size: How many subjects and how many biometric acquisitions per subject should be considered in order to obtain a confidence band for the ROC with a pre-specified width? Based on the multivariate Gaussian copula model for correlated biometric acquisitions, we give the minimum number of subjects required to achieve the desired width. In presence of non-zero correlation, increasing the number of subjects is more effective in reducing the width of the confidence band compared to increasing the number of biometric acquisitions per subject.

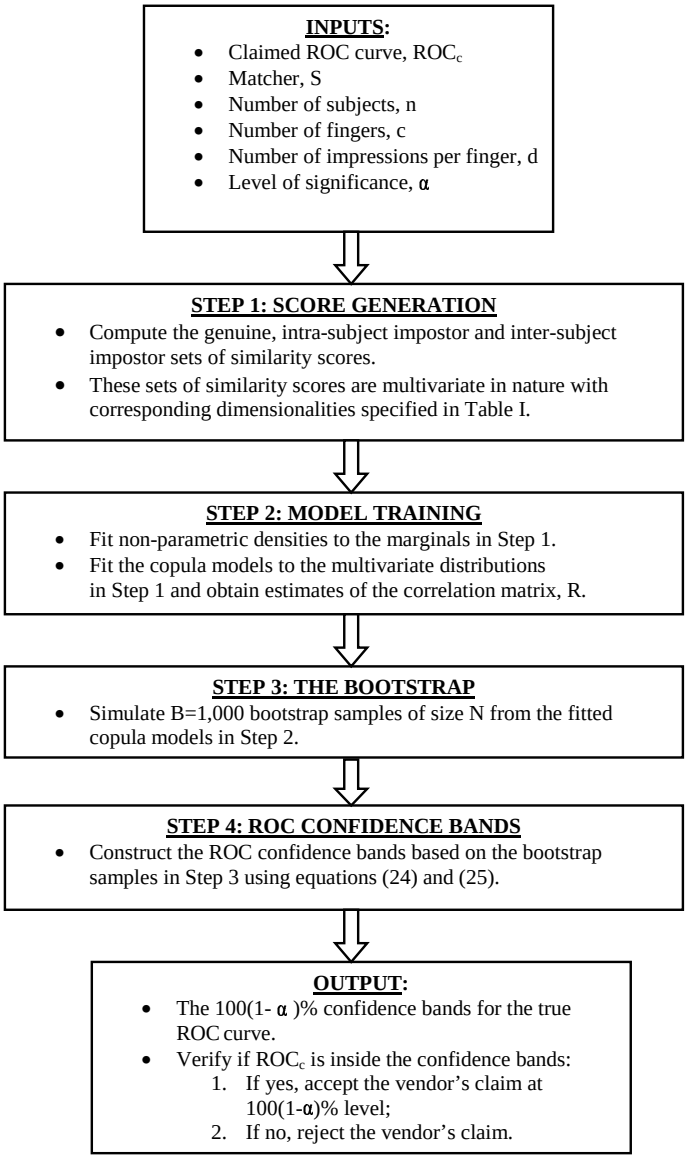


Fig. 2. The main steps involved in constructing the ROC confidence bands for validating the claim of a fingerprint vendor.

For achieving the desired confidence level, the required number of subjects based on our method is much smaller compared to the subset bootstrap. Rules of thumb such as the Rule of 3 [19] and the Rule of 30 [13] grossly underestimate the number of users required to obtain a specific width. The underestimation becomes more severe as the correlation between any two acquisitions of a subject increases.

The paper is organized as follows: Section II presents an introduction to the problem. Section III discusses the use of multivariate copula functions to model the correlation between multiple queries per subject for the genuine and impostor similarity score distributions. Section IV presents the construction of confidence bands for the ROC curve as well as confidence intervals for the FRR and FAR. Section V discusses minimum sample size required for obtaining confidence bands for the ROC of a pre-specified width.

II. Preliminaries

Suppose we have n subjects available for validating a biometric authentication system. Often, during the data collection stage, multiple biometrics (e.g., different fingers) from the same subject are used. We denote the number of biometrics used per subject by c. To obtain additional data, each biometric of a subject is usually sampled a multiple number of times, d, over a period of time. Thus, at the end of the data collection stage, we acquire a total of ncd biometric samples from the n subjects. This collection of ncd biometric samples will be denoted by $\mathcal{B}$. To obtain similarity scores, a pair of biometric samples, B and $B^{'}$ with $B \neq B^{'}$, are taken from $\mathcal{B}$ and a matcher S is applied to them, resulting in the similarity score $S(B, B^{'})$. We will consider asymmetric matchers for S in this paper: The matcher S is asymmetric if $S(B, B^{'}) \neq S(B^{'}, B)$ for the pair of biometric samples $(B, B^{'})$ (a symmetric matcher implies that $S(B, B^{'}) = S(B, B^{'})$).

In the subsequent text, we will use a fingerprint authentication system as the generic biometric system that needs to be validated. Thus, the c different biometrics will be represented as c different fingers from each subject, and the d acquisitions will be represented by d impressions of each finger. When B and $B^{'}$ are multiple impressions of the same finger from the same user, the similarity score $S(B, B^{'})$ is termed as a genuine similarity score, whereas when B and $B^{'}$ are impressions from either (i) different fingers from the same subject, or (ii) different subjects, the similarity score $S(B, B^{'})$ is termed as an impostor score. The impostor scores arising from (i) (respectively, (ii)) are termed as the intra-subject (respectively, inter-subject) impostor scores.

We give some intuitive understanding of why similarity scores arising from certain pairs of fingerprint impressions in $\mathcal{B}$ are correlated (or, dependent). During the fingerprint acquisition process, multiple impressions of a finger are obtained by successive placement of the finger onto the sensor. Therefore,

given the first impression, B, and two subsequent impressions B_1 and B_2, the similarity scores $S(B, B_1)$ and $S(B, B_2)$ are most likely going to be correlated. Further, the fingerprint acquisition process is prone to many different types of uncontrollable factors such as fingertip pressure, fingertip moisture and skin elasticity factor. These factors cause some level of dependence between fingerprint impressions of two different fingers of the same user. If this is the case, then we expect to see some level of correlation between the similarity scores $S(B_1, B_2)$ where B_1 and B_2 are impressions from different fingers. Also, as noted in [3], even the scores $S(B_1, B_2)$ from different fingers of different subjects could be dependent. All these facts lead us to statistically model the correlation for similarity scores in the three major categories, namely the genuine, intra-user impostor and inter-user impostor similarity scores.

In order to develop the framework that incorporates correlation, we need to introduce some more notations. We denote the set consisting of the d impressions of finger f, $f = 1, 2, \ldots, c$, from subject i by $\mathcal{M}_{i,f}$. The notation

$$\mathcal{S}(i, j, f, f') = \{\, S(B_u, B_v);\, B_u \in \mathcal{M}_{i,f},\, B_v \in \mathcal{M}_{j,f'},\, B_u \neq B_v \,\} \tag{6}$$

represents the set of all similarity scores available from matching the fingerprint impressions of finger f from subject i and those of finger f' from subject j. Three disjoint sets of (6) are of importance, namely, the set of genuine similarity scores obtained by taking $i = j$ and $f = f'$ in (6) (that is, the similarity scores are obtained from multiple impressions of the same finger corresponding to the same subject), the set of intra-subject impostor scores obtained by taking $i = j$ and $f \neq f'$ in (6) (that is, the similarity scores are obtained from different fingers of the same subject), and the set of inter-subject impostor scores obtained by taking $i \neq j$ in (6) (that is, the similarity scores are obtained from different subjects). We denote the genuine, intra-subject impostor and inter-subject impostor score sets for subject i by

$$\mathcal{G}_i \equiv \bigcup_{f=1}^{c} \mathcal{S}(i, i, f, f), \qquad \mathcal{I}_i \equiv \bigcup_{\substack{f, f'=1 \\ f \neq f'}}^{c} \mathcal{S}(i, i, f, f'), \qquad \text{and} \qquad \mathcal{I}_{ij} \equiv \bigcup_{f, f'=1}^{c} \mathcal{S}(i, j, f, f'), \tag{7}$$

respectively.

The above sets can be specialized also to the case when the biometric entities are chosen to be the different fingers, instead of the subjects. In that case, we have $c = 1$ since there is only one finger per "subject", and the sets $\mathcal{M}_{i,1}$ denote the multiple impressions corresponding to each finger. The set of intra-subject impostor scores, $\mathcal{I}_i$, corresponding to the i-th finger will consequently be empty, and the set of impostor similarity scores will now comprise of scores from different pairs of fingers only. The sets of genuine and impostor scores when the biometric entities are fingers will be denoted by $\mathcal{G}_i$ and

Entities	$\mathcal{N}$	$\lvert\mathcal{G}_i\rvert$	$\lvert\mathcal{I}_i\rvert$	$\lvert\mathcal{I}_{ij}\rvert$
Subjects	n	$cd(d-1)$	$c(c-1)d^2$	c^2d^2
Fingers	nc	$d(d-1)$	0	d^2

TABLE I

VALUES OF K ACCORDING TO WHETHER THE BIOMETRIC ENTITIES ARE SUBJECTS OR FINGERS, AND $|\cdot|$ DENOTES THE CARDINALITY OR DIMENSION OF A SET; n IS THE NUMBER OF SUBJECTS, c IS THE NUMBER OF FINGERS AND d IS THE NUMBER OF IMPRESSIONS PER FINGER.

$\mathcal{I}_{ij}$, respectively, corresponding to the i-th finger and the (i,j)-th finger pair. The number of biometric entities, whether they are different subjects or different fingers, will be denoted by $\mathcal{N}$.

We give the cardinality or dimension (the number of possibly distinct similarity scores) of each of the sets discussed above. When the biometric entities are the different subjects, we have $\mathcal{N} = n$. Further, the dimensions of $\mathcal{G}_i$, $\mathcal{I}_i$ and $\mathcal{I}_{ij}$ are $cd(d-1)$, $c(c-1)d^2$ and c^2d^2, respectively, when the matcher S is asymmetric. When the biometric entities are the different fingers, we have $\mathcal{N} = nc$, and the dimensions of $\mathcal{G}_i$ and $\mathcal{I}_{ij}$, respectively, are $d(d-1)$ and d^2. In all of these scenarios, we will denote the dimension corresponding to each set by K. Table I summarizes the different cases of choosing biometric entities and the corresponding value of K.

The total number of sets of similarity scores arising from $\mathcal{N}$ entities for the genuine, intra- and inter-impostor cases will be denoted by N. When the entities are subjects, we have that $N = \mathcal{N}$, $N = \mathcal{N}$ and $N = \mathcal{N}(\mathcal{N}-1)$, respectively, for the total number of sets of genuine, intra-subject impostor and inter-subject scores with $\mathcal{N} = n$. When the entities are the different fingers, $N = \mathcal{N}$ and $N = \mathcal{N}(\mathcal{N}-1)$ for the number of genuine and impostor sets of similarity scores with $\mathcal{N} = nc$.

When the matcher S is symmetric, the dimension associated with each of the genuine, intra-subject impostor and inter-subject impostor sets of similarity scores gets reduced since many of the similarity scores in each of the three sets will be identical to each other. In the subsequent text, we outline the methodology for validating a vendor's claim for an asymmetric matcher. Our methodology for constructing the ROC confidence bands for a symmetric matcher can be handled in a similar fashion, keeping in mind the reduction in dimensions of each of the three sets of similarity scores discussed above (two sets when the entities are the different fingers).

Subsequently, N will denote the total number of independent sets of similarity scores, and K will denote the dimension of each of these N sets. For $i = 1, 2, \ldots, N$, the i-th set of similarity scores will

be denoted by the K-dimensional vector

$$\mathcal{S}_i = (s(i,1), s(i,2), \ldots, s(i,K))^T, \tag{8}$$

where $s(i,k)$ is the generic score corresponding to the k-th component of $\mathcal{S}_i$, for $k = 1, 2, \ldots, K$. If the scores $s(i,k)$ are bounded between two numbers a and b, the order preserving transformation

$$\mathcal{T}(s(i,k)) = \log\left(\frac{s(i,k)-a}{b-s(i,k)}\right) \tag{9}$$

converts each score onto the entire real line. The transformed scores will be represented by the same notation $s(i,k)$. The transformation in (9) yields better non-parametric density estimates for the marginal densities compared to using the original scores (see Section III-B). The distribution function for each $\mathcal{S}_i$ will be denoted by F, that is,

$$P\{\, s(i,1) \le s_1,\, s(i,2) \le s_2,\, \ldots, s(i,K) \le s_K \,\} = F(s_1, s_2, \ldots, s_K), \tag{10}$$

for real numbers $s_1, s_2, \ldots, s_K$. Note that (i) F is a multivariate joint distribution function on R^K, and (ii) we assume that F is the common distribution function for every $i = 1, 2, \ldots, N$. The distribution function F has K associated marginals; we denote the marginals by F_k, $k = 1, 2, \ldots, K$, where

$$P\{\, s(i,k) \le s_k \,\} = F_k(s_k). \tag{11}$$

In the next few sections, a semi-parametric family of Gaussian copula models are presented as candidates for F, and the procedure for obtaining parameter estimates based on the observed scores $\mathcal{S}_i$, $i = 1, 2, \ldots, N$ is described.

III. Copula Models for F

Let $H_1, H_2, \ldots, H_K$ be K continuous distribution functions on the real line. Suppose that H is a K-dimensional distribution function with the k-th marginal given by H_k for $k = 1, 2, \ldots, K$. According to Sklar's Theorem [12], there exists a unique function $C(u_1, u_2, \ldots, u_K)$ from $[0,1]^K$ to $[0,1]$ satisfying

$$H(s_1, s_2, \ldots, s_K) = C(H_1(s_1), H_2(s_2), \ldots, H_K(s_K)), \tag{12}$$

where $s_1, s_2, \ldots, s_K$ are K real numbers. The function C is known as a K-copula function that "couples" the one-dimensional distribution functions $H_k\,, k = 1, 2, \ldots, K$ to obtain H. Basically, K-copula functions are K-dimensional distribution functions on $[0,1]^K$ whose marginals are uniform. Equation (12) can also be used to construct K-dimensional distribution functions H whose marginals are the pre-specified distributions $H_k\,, k = 1, 2, \ldots, K$: choose a copula function C and define the function H as in (12). It follows that H is a K-dimensional distribution function with marginals $H_k\,,\, k = 1, 2, \ldots, K$.

The choice of C we consider in this paper is the K-dimensional Gaussian copulas [5] given by

$$C_R(u_1, u_2, \ldots, u_K) = \Phi_R^K(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_K)) \tag{13}$$

where each $u_k \in [0,1]$ for $k = 1, 2, \ldots, K$, $\Phi(\cdot)$ is the distribution function of the standard normal, $\Phi^{-1}(\cdot)$ is its inverse, and Φ_R^K is the K-dimensional distribution function of a normal random vector $\mathcal{Z} = (Z_1, Z_2, \ldots, Z_K)^T$ with component means and variances given by 0 and 1, respectively, and with correlation matrix R. We define the $(k, k^{'})$-th entry of R as $\rho_{kk'}$, where $k, k^{'} = 1, 2, \ldots, K$, and write $R = ((\rho_{kk'}))$. Note that R is a positive definite matrix with diagonal entries ρ_{kk}, $k = 1, 2, \ldots, K$ equal to unity. The distribution function F will be assumed to be of the form (12) with $H_k = F_k$ for $k = 1, 2, \ldots, K$, and $C = C_R$; thus, we have

$$F(s_1, s_2, \ldots, s_K) = C_R(F_1(s_1), F_2(s_2), \ldots, F_K(s_K)). \tag{14}$$

A. Simulation from F

We now describe how to simulate samples from F assuming that F is of the form (14). The following steps outline a simulation procedure to generate N samples from F: (1) Generate a vector $Z = (Z_1, Z_2, \ldots, Z_K)^T$ from Φ_R^K, the K-dimensional multivariate normal with mean 0, variance 1, and correlation matrix R, (2) Obtain the vector $U = (U_1, U_2, \ldots, U_K)^T$ by letting $U_k = \Phi(Z_k)$ for $k = 1, 2, \ldots, K$, and (3) Obtain the vector $\mathcal{S}^* = (s_1^*, s_2^*, \ldots, s_K^*)^T$ using $s_k^* = F_k^{-1}(U_k)$ for $k = 1, 2, \ldots, K$, where F_k^{-1} is the inverse of F_k. It follows that $\mathcal{S}^*$ is a sample from F. In order to obtain a sample of size N, steps (1-3) are repeated N times resulting in the simulated samples $\{ s^*(i,k)\,,\, k = 1, 2, \ldots, K \}$ for $i = 1, 2, \ldots, N$. In practice, one difficulty is that the marginal distributions and the correlation matrices for the genuine and impostor similarity scores will generally be unknown, and will have to be estimated from the observed scores (this is discussed in the subsequent section). Once they have been estimated, we can follow steps (1-3) to obtain samples from the fitted copula models.

B. Estimation of F_k and R

The marginal distribution functions, F_k, and the correlation matrix R are generally unknown and have to be estimated from the observed vector of similarity scores, $\{\mathcal{S}_i, :\ i = 1, 2, \ldots, N\}$. The empirical distribution function for the k-th marginal is given by

$$E_k(s) = \frac{1}{N} \sum_{i=1}^{N} I\{\, s(i,k) \leq s \,\}, \tag{15}$$

where $I(A)$ is the indicator function of the set A; $I(A) = 1$ if A is true, and 0 otherwise. Note that $E_k(s) = 0$ for all $s < s_{min}$ and $E_k(s) = 1$ for all $s \geq s_{max}$, where s_{min} and s_{max}, respectively, are

the minimum and maximum of the observations $\{s(i,k) : i = 1, 2, \ldots, N\}$. Next, we define $\mathcal{H}(s) \overset{12}{\equiv} -\log(1 - E_k(s))$, and note that discontinuity points of $E_k(s)$ will also be points of discontinuity of $\mathcal{H}(s)$. In order to obtain a continuous estimate of $\mathcal{H}(s)$, the following procedure is adopted: For an M-partition $s_{min} \equiv s_0 < s_1 < \ldots < s_M \equiv s_{max}$ of $[s_{min}, s_{max}]$, the value of $\mathcal{H}(s)$ at a point $s \in [s_{min}, s_{max}]$ is redefined via the linear interpolation formula

$$\hat{\mathcal{H}}(s) = \mathcal{H}(s_m) + (\mathcal{H}(s_{m+1} - \mathcal{H}(s_m)) \cdot \frac{s - s_m}{s_{m+1} - s_m} \tag{16}$$

whenever $s_m \leq s \leq s_{m+1}$ and subsequently, the estimated distribution function, $\hat{F}_k(s)$, of $F_k(s)$ is obtained as

$$\hat{F}_k(s) = 1 - \exp\{-\hat{\mathcal{H}}(s)\}. \tag{17}$$

It follows that each $\hat{F}_k(s)$ is a continuous distribution function. Next we generate $B^* = 1,000$ samples from $\hat{F}_k$: (1) Generate a uniform random variable U in $[0, 1]$, (2) Define $V = -\log(1 - U)$, and (3) Find the value V^* such that $\hat{\mathcal{H}}(V^*) = V$. It follows that V^* is a random variable with distribution function $\hat{F}_k$. To generate $1,000$ independent realizations from $\hat{F}_k$, we repeat the steps (1-3) $1,000$ times. Finally, a non-parametric density estimate of F_k is obtained based on the simulated samples using a Gaussian kernel.

The estimate of R based on the observed similarity score vectors $\{\mathcal{S}_i : i = 1, 2, \ldots, N\}$ is obtained in the following way: Define a new vector $\mathcal{Z}_i = (Z(i,1), Z(i,2), \ldots, Z(i,K))^T$ where

$$Z(i,k) = \Phi^{-1}(E_k(s(i,k)), \tag{18}$$

for $k = 1, 2, \ldots, K$. The mean vector $\bar{\mathcal{Z}}$ is then obtained by averaging over the vectors $\mathcal{Z}_i$, that is,

$$\bar{\mathcal{Z}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{Z}_i \tag{19}$$

and the covariance matrix is defined as

$$J = \frac{1}{N} \sum_{i=1}^{N} (\mathcal{Z}_i - \bar{\mathcal{Z}}) \cdot (\mathcal{Z}_i - \bar{\mathcal{Z}})^T. \tag{20}$$

The estimate of $\rho_{kk'}$ is given by

$$\hat{\rho}_{kk'} = \frac{\sigma_{kk'}}{\sqrt{\sigma_{kk}\sigma_{k'k'}}}, \tag{21}$$

where $\sigma_{kk'}$ is the (k, k')-th entry of J in (20), and the estimated correlation matrix is given by $\hat{R} = ((\hat{\rho}_{kk'}))$. The total number of correlation parameters that need to be estimated is $K(K-1)/2$; thus, it is necessary to have $K(K-1)/2$ much smaller than N to avoid over-fitting.

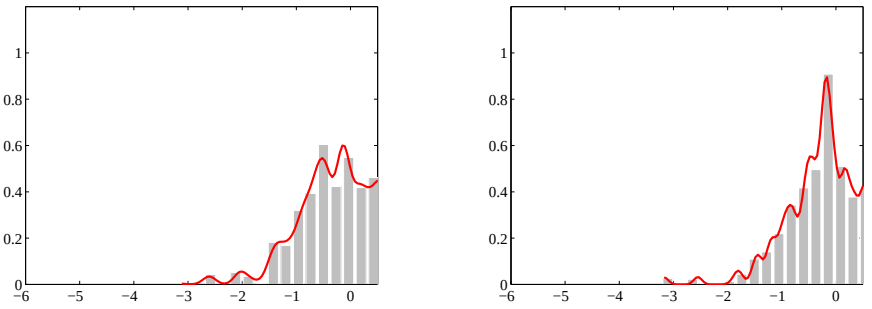
We demonstrate the estimation of F_k and R using the fingerprint database described in [7]. This fingerprint database consists of impressions acquired from 4 different fingers (the right index, right

middle, left index and left middle fingers) from 160 different users. A total of 4 impressions per finger were obtained; 2 impressions were obtained on the first day and the remaining two after a period of a week. See Section V for details of this database. Genuine and impostor raw similarity scores were generated using the asymmetric matcher described in [6]. All raw scores ranged between 0 and 1000, and thus, the transformation (9) with $a = 0$ and $b = 1000$ was used to convert the scores onto the real line. All subsequent analysis was performed on the transformed similarity scores.

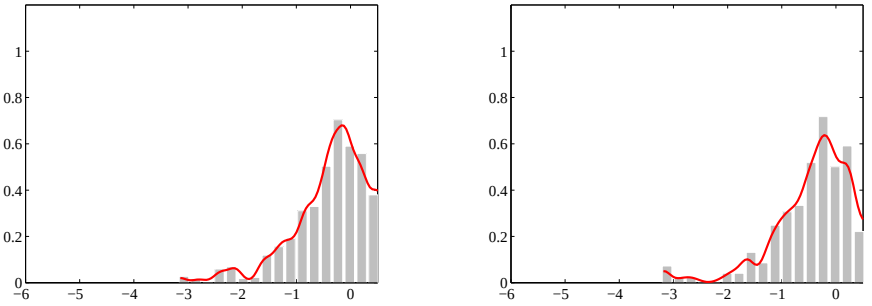
We will illustrate our methodology in the subsequent text by considering subjects as the different biometric entities. The choice of fingers for the entities can be handled as a special case by viewing each finger as a different "subject", and setting $c = 1$. Thus, we postpone the results when fingers are chosen as the entities to Section V. In the present case, we have the following values for N and K (with $c = d = 4$): $N = 160$ and dimensionality $K = 4 \times 4 \times 3 = 48$ for the set of genuine scores, $N = 160$ and $K = 4 \times 3 \times 4^2 = 192$ for the set of intra-subject impostor scores, and $N = 160 \times 159 = 25,440$ and $K = 4^2 \times 4^2 = 256$ for the set of inter-subject impostor scores. The number of parameters in the correlation matrices that need to be estimated for the genuine, intra-subject impostor and inter-subject impostor scores are, respectively, $(48 \times 47)/2 = 1128$, $(192 \times 191)/2 = 18,336$ and $(256 \times 255)/2 = 32,640$. Thus, the number of parameters far exceeds the total number of observations in each of the three sets of scores. In order to avoid over-fitting, we reduce the value of K in each case. Instead of selecting all 4 fingers, we choose only 2, namely, the right index and right middle fingers, and use the $d = 2$ impressions per finger obtained on the first day. These values of c and d will be used subsequently for our experimental results. With these choices for c and d, the K values (with $c = d = 2$) are 4, 8 and 16, respectively, for the set of genuine, intra-subject impostor, and inter-subject impostor scores.

We show the results of the non-parametric estimation procedure for the first 4 marginal distributions corresponding to each of the genuine, intra-subject impostor and inter-subject impostor scores when the biometric entities are the different subjects. These are shown, respectively, in Figures 3, 4, and 5, respectively. The red line in each figure gives the graph obtained using a density estimator with a Gaussian kernel based on $B^* = 1,000$ simulated values. Note the very good agreement between the observed density histogram and the fitted density curve for each figure, especially at the tails of the distributions. A good fit at the tails is essential for the construction of a valid ROC curve that accurately reflects the authentication performance based on the observed data of similarity scores.

Correlation estimates of the genuine (4×4), inter-subject impostor (8×8), and intra-subject impostor



(a) $k = 1$ (b) $k = 2$



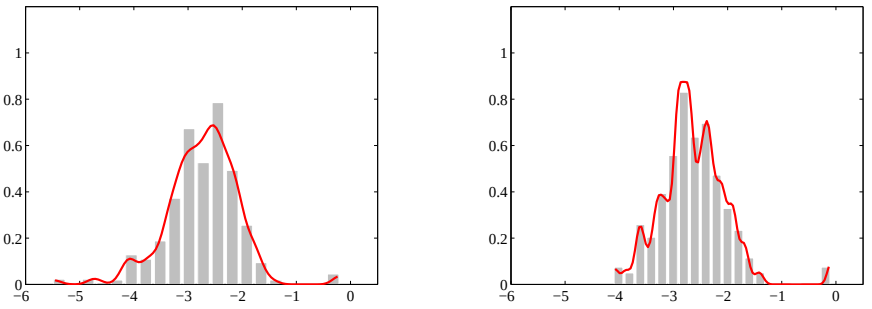
(c) $k = 3$ (d) $k = 4$

Fig. 3. Fitted density functions for the genuine marginal distributions based on 1000 simulated samples.

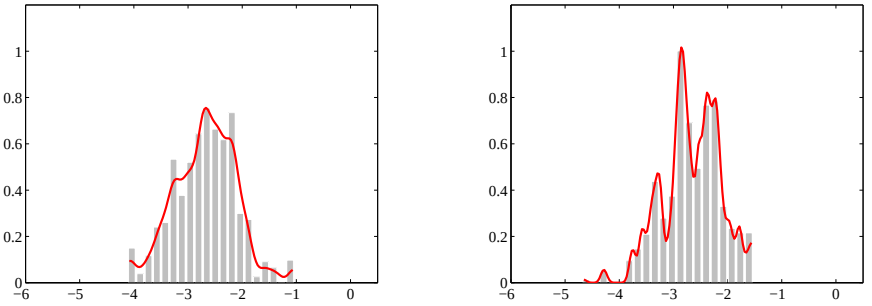
sets (16×16) are, respectively, given by

$$\hat{R}_0 = \begin{pmatrix} 1.00 & 0.99 & 0.15 & 0.16 \\ 0.99 & 1.00 & 0.15 & 0.16 \\ 0.15 & 0.15 & 1.00 & 0.99 \\ 0.16 & 0.16 & 0.99 & 1.00 \end{pmatrix}, \tag{22}$$

$$\hat{R}_{11} = \begin{pmatrix} 1.00 & 0.58 & 0.52 & 0.42 & 0.90 & 0.53 & 0.54 & 0.41 \\ 0.58 & 1.00 & 0.44 & 0.47 & 0.58 & 0.46 & 0.88 & 0.46 \\ 0.52 & 0.44 & 1.00 & 0.45 & 0.50 & 0.86 & 0.37 & 0.42 \\ 0.42 & 0.47 & 0.45 & 1.00 & 0.41 & 0.41 & 0.43 & 0.87 \\ 0.90 & 0.58 & 0.50 & 0.41 & 1.00 & 0.53 & 0.55 & 0.41 \\ 0.53 & 0.46 & 0.86 & 0.41 & 0.53 & 1.00 & 0.40 & 0.42 \\ 0.54 & 0.88 & 0.37 & 0.43 & 0.55 & 0.40 & 1.00 & 0.44 \\ 0.41 & 0.46 & 0.42 & 0.87 & 0.41 & 0.42 & 0.44 & 1.00 \end{pmatrix}, \tag{23}$$



(a) $k = 1$ (b) $k = 2$

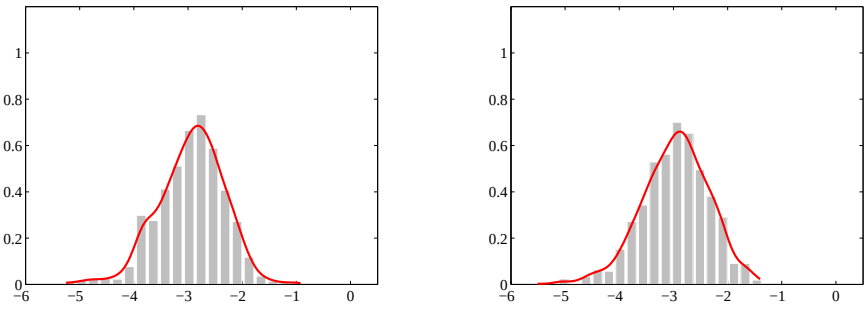


(c) $k = 3$ (d) $k = 4$

Fig. 4. Fitted density functions for the intra-impostor marginal distributions based on 1000 simulated samples.

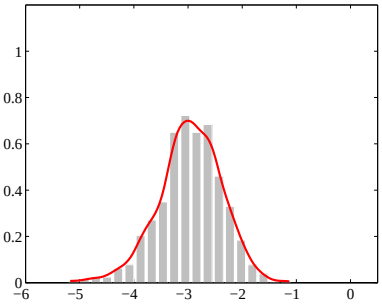
and

$$\hat{R}_{12} = \begin{pmatrix} 1.00 & 0.42 & 0.17 & 0.16 & 0.42 & 0.31 & 0.15 & 0.13 & 0.17 & 0.14 & 0.08 & 0.07 & 0.14 & 0.12 & 0.07 & 0.07 \\ 0.42 & 1.00 & 0.17 & 0.17 & 0.32 & 0.42 & 0.14 & 0.13 & 0.13 & 0.17 & 0.08 & 0.09 & 0.12 & 0.16 & 0.07 & 0.08 \\ 0.17 & 0.17 & 1.00 & 0.37 & 0.14 & 0.14 & 0.38 & 0.26 & 0.05 & 0.05 & 0.15 & 0.12 & 0.03 & 0.03 & 0.13 & 0.11 \\ 0.16 & 0.17 & 0.37 & 1.00 & 0.13 & 0.13 & 0.28 & 0.38 & 0.04 & 0.05 & 0.11 & 0.15 & 0.04 & 0.04 & 0.09 & 0.15 \\ 0.42 & 0.32 & 0.14 & 0.13 & 1.00 & 0.43 & 0.18 & 0.16 & 0.18 & 0.14 & 0.08 & 0.07 & 0.16 & 0.13 & 0.08 & 0.09 \\ 0.31 & 0.42 & 0.14 & 0.13 & 0.43 & 1.00 & 0.17 & 0.17 & 0.13 & 0.17 & 0.08 & 0.07 & 0.12 & 0.16 & 0.08 & 0.08 \\ 0.15 & 0.14 & 0.38 & 0.28 & 0.18 & 0.17 & 1.00 & 0.37 & 0.05 & 0.06 & 0.14 & 0.11 & 0.04 & 0.04 & 0.13 & 0.12 \\ 0.13 & 0.13 & 0.26 & 0.38 & 0.16 & 0.17 & 0.37 & 1.00 & 0.03 & 0.05 & 0.11 & 0.15 & 0.03 & 0.04 & 0.11 & 0.16 \\ 0.17 & 0.13 & 0.05 & 0.04 & 0.18 & 0.13 & 0.05 & 0.03 & 1.00 & 0.38 & 0.15 & 0.14 & 0.36 & 0.26 & 0.12 & 0.11 \\ 0.14 & 0.17 & 0.05 & 0.05 & 0.14 & 0.17 & 0.06 & 0.05 & 0.38 & 1.00 & 0.14 & 0.14 & 0.26 & 0.37 & 0.11 & 0.11 \\ 0.08 & 0.08 & 0.15 & 0.11 & 0.08 & 0.08 & 0.14 & 0.11 & 0.15 & 0.14 & 1.00 & 0.36 & 0.12 & 0.11 & 0.37 & 0.28 \\ 0.07 & 0.09 & 0.12 & 0.15 & 0.07 & 0.07 & 0.11 & 0.15 & 0.14 & 0.14 & 0.36 & 1.00 & 0.10 & 0.11 & 0.26 & 0.37 \\ 0.14 & 0.12 & 0.03 & 0.04 & 0.16 & 0.12 & 0.04 & 0.03 & 0.36 & 0.26 & 0.12 & 0.10 & 1.00 & 0.37 & 0.16 & 0.15 \\ 0.12 & 0.16 & 0.03 & 0.04 & 0.13 & 0.16 & 0.04 & 0.04 & 0.26 & 0.37 & 0.11 & 0.11 & 0.37 & 1.00 & 0.15 & 0.16 \\ 0.07 & 0.07 & 0.13 & 0.09 & 0.08 & 0.08 & 0.13 & 0.11 & 0.12 & 0.11 & 0.37 & 0.26 & 0.16 & 0.15 & 1.00 & 0.38 \\ 0.07 & 0.08 & 0.11 & 0.15 & 0.09 & 0.08 & 0.12 & 0.16 & 0.11 & 0.11 & 0.28 & 0.37 & 0.15 & 0.16 & 0.38 & 1.00 \end{pmatrix}, \quad (24)$$

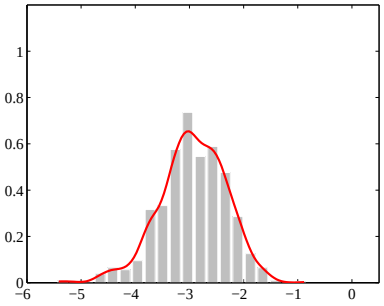


(a) $k = 1$

(b) $k = 2$



(c) $k = 3$



(d) $k = 4$

Fig. 5. Fitted density functions for the inter-impostor marginal distributions based on 1000 simulated samples.

respectively.

The off-diagonal entries of the estimated correlation matrices indicate the degree of correlation between the corresponding row and column dimensions. For example, the entry 0.15 in the 2-nd row and 3-rd column of matrix $\hat{R}_0$ is the correlation between the 2-nd and 3-rd component of $\mathcal{S}_i$ in the genuine case (that is, the correlation between $s(B_2, B_1)$ for finger 1 and $s(B_1, B_2)$ for finger 2; recall that $c = 2$ and $d = 2$). The off-diagonal entries of $\hat{R}_0$, $\hat{R}_{11}$ and $\hat{R}_{12}$ indicate that there are significant amounts of correlations in both the genuine and impostor sets of similarity scores.

C. Assessing the Goodness of Fit

We present a method here for graphically assessing the goodness of fit of the estimated multivariate Gaussian K-copula model to the observed data. We first give the general methodology, and then apply it to the observed genuine and impostor similarity scores. Lower dimensional marginals of a K-copula function $C(u_1, u_2, \ldots, u_K)$ can be obtained by fixing the irrelevant u_ks to be equal to one: For example, if we require the 2-dimensional copula function in the dimensions of k and k', where $k \neq k'$, $k, k' =$

$1, 2, \ldots, K$, this can be obtained by setting the other u_js ($j \neq k$, $j \neq k^{'}$) to 1, that is,

$$C_{k,k'}(u_k, u_{k'}) \equiv C(1, 1, \ldots, u_k, 1, \ldots, 1, u_{k'}, 1, \ldots, 1). \tag{25}$$

It follows that all lower k-dimensional ($k < K$) marginals of the multivariate Gaussian K-copula are Gaussian k-copulas. In particular, for $k = 2$, we obtain $\binom{K}{2}$ bivariate Gaussian copulas from a single Gaussian K-copula as in (13). Each bivariate Gaussian copula is characterized by a single correlation parameter; for dimensions p and q, this parameter is $\rho_{kk'}$ of matrix R.

The bivariate empirical copula based on N independent bivariate observations (X_i, Y_i) , $i = 1, 2, \ldots, N$ is defined as follows: For each $0 \leq x \leq 1$ and $0 \leq y \leq 1$,

$$C_{emp}(x, y) = \frac{1}{N} \sum_{i=1}^{N} I\{X_i \leq X_{([Nx])} \, , \, Y_i \leq Y_{([Ny])}\}, \tag{26}$$

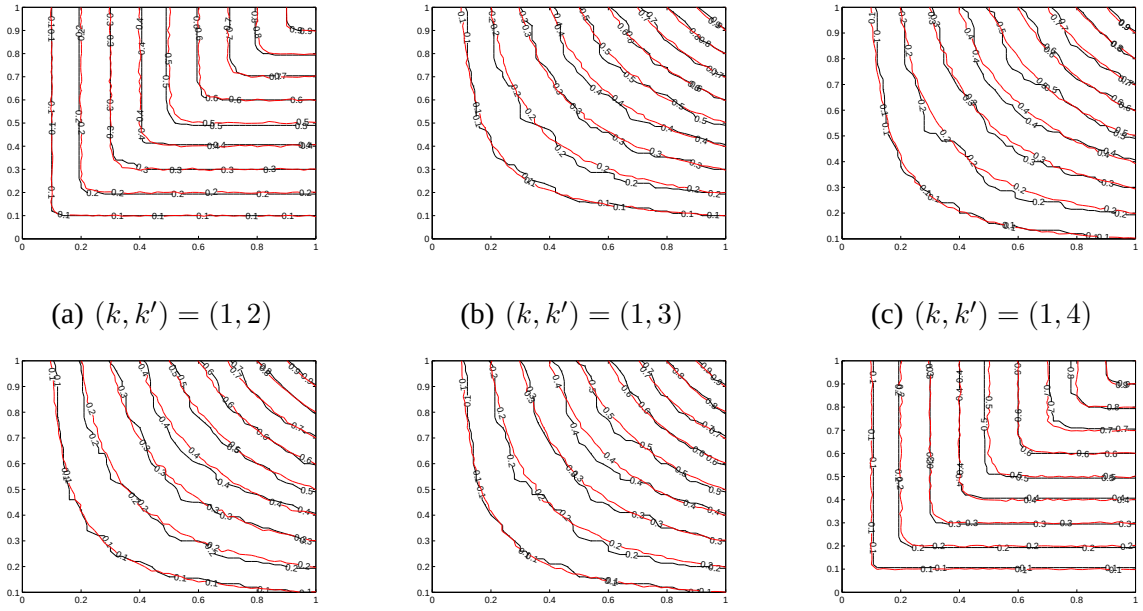
where $X_{([Nx])}$ (respectively, $Y_{([Ny])}$) is the $[Nx]$-th ($[Ny]$-th) element in the ordered list of X (Y) samples, and the notation $[u]$ represents the greatest integer less than or equal to u. The empirical copula function gives the best approximation to the true but unknown copula function that generated the observed data (X_i, Y_i) , $i = 1, 2, \ldots, N$.

Our graphical test for checking goodness of fit consists of the following steps: (i) Obtain the $\binom{K}{2}$ 2-dimensional marginal copulas based on $\hat{R}$. For the dimension pair $(k, k^{'})$, we obtain the contour plot of $C_{k,k'}(u_k, u_{k'})$ given by

$$C_{k,k'}(u_k, u_{k'}) = \Phi^2_{\hat{\rho}_{kk'}}(\Phi^{-1}(u_k), \Phi^{-1}(u_{k'})). \tag{27}$$

(ii) Obtain the empirical copula based on the score vectors $s(i, k), s(i, k^{'}))^T$ for $i = 1, 2, \ldots, N$ using equation (26); here $s(i, k)$ are the X samples and $s(i, k^{'})$ are the Y samples.

The total number of pairs of components for the sets of genuine, intra-subject and inter-subject scores are, respectively, $\binom{4}{2} = 6$, $\binom{8}{2} = 28$, and $\binom{16}{2} = 120$. Figures 6, 7 and 8 respectively give the plots of 6 component pairs for the genuine, intra-subject impostor and inter-subject impostor sets in this case. Note that the figures indicate that there is a good agreement between the empirical and the proposed Gaussian copula functions. We checked all of the pair-wise copula plots and found that there were no major discrepancies between the empirical contours and the fitted Gaussian copula contours. Thus, we conclude that the proposed Gaussian copula functions are good models for representing the correlation structures in all of the genuine, intra-subject and inter-subject sets of scores. There is always a problem of quantitatively assessing the quality of a model fit to the observed data when the sample size is very large (as in the case of the genuine and impostor sets of similarity scores here). A small discrepancy between the observed data and model fit will magnify due to the large sample size and cause a quantitative



(a) $(k, k') = (1, 2)$ (b) $(k, k') = (1, 3)$ (c) $(k, k') = (1, 4)$

(d) $(k, k') = (2, 3)$ (e) $(k, k') = (2, 4)$ (f) $(k, k') = (3, 4)$

Fig. 6. Nine level curves (at levels $0.1, 0.2, \ldots, 0.9$) indicating a good match between the empirical copula (black lines) and the estimated bivariate Gaussian copula (red lines) along dimensions k and k' for the genuine scores.

goodness of fit test to be statistically significant. The point to note here is that the test can potentially be statistically significant even if the models are a good fit to the observed data set.

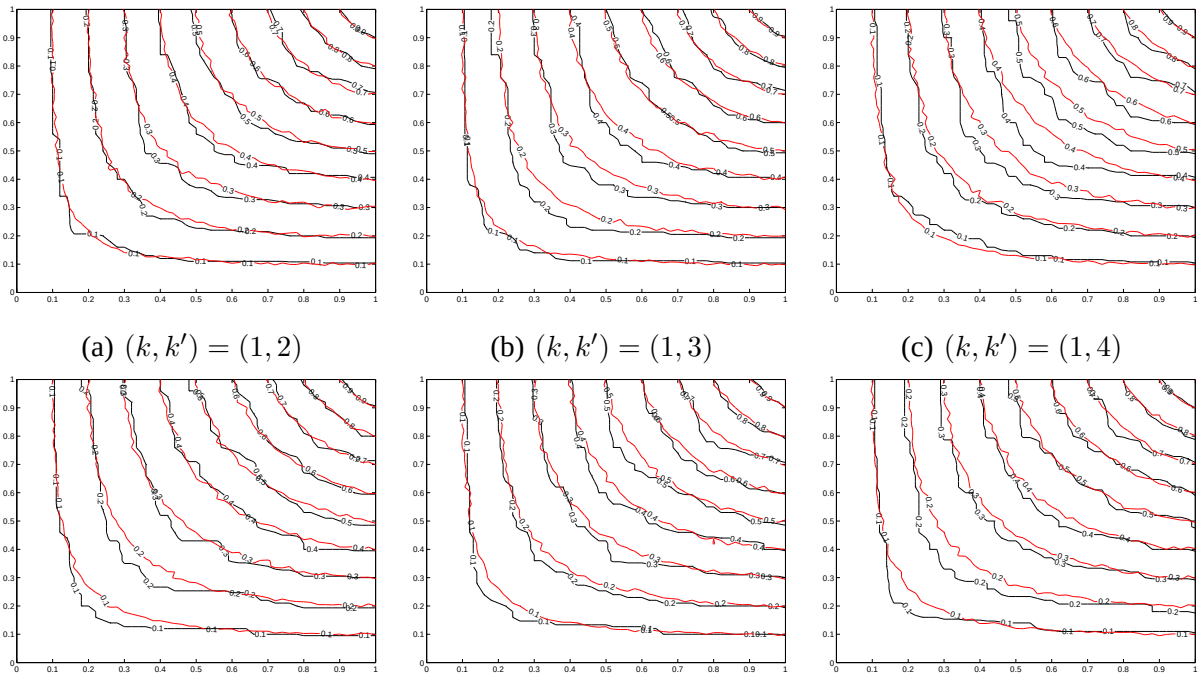
IV. Confidence bands for the ROC curve

The Receiver Operating Curve (ROC) is a graph that expresses the relationship between the Genuine Accept Rate (GAR) and the False Accept Rate (FAR), and is used to report the performance of a biometric authentication system. We denote the set of genuine, intra-impostor and inter-impostor similarity scores by $\mathcal{S}_0 \equiv \{ s_0(i,k),\, k = 1, 2, \ldots, K_0,\, i = 1, 2, \ldots, N_0\}$, $\mathcal{S}_{11} \equiv \{ s_{11}(i,k),\, k = 1, 2, \ldots, K_{11},\, i = 1, 2, \ldots, N_{11}\}$ and $\mathcal{S}_{12} \equiv \{ s_{12}(i,k),\, k = 1, 2, \ldots, K_{12},\, i = 1, 2, \ldots, N_{12}\}$, respectively. Thus, the total number of impostor similarity scores, denoted by N_1, is $N_1 = N_{11}K_{11} + N_{12}K_{12}$. For the threshold λ, the empirical GAR and FAR can be computed using the formulas

$$GAR(\lambda) = \frac{1}{N_0 K_0} \sum_{i=1}^{N_0} \sum_{k=1}^{K_0} I\{ s_0(i,k) > \lambda\}, \tag{28}$$

and

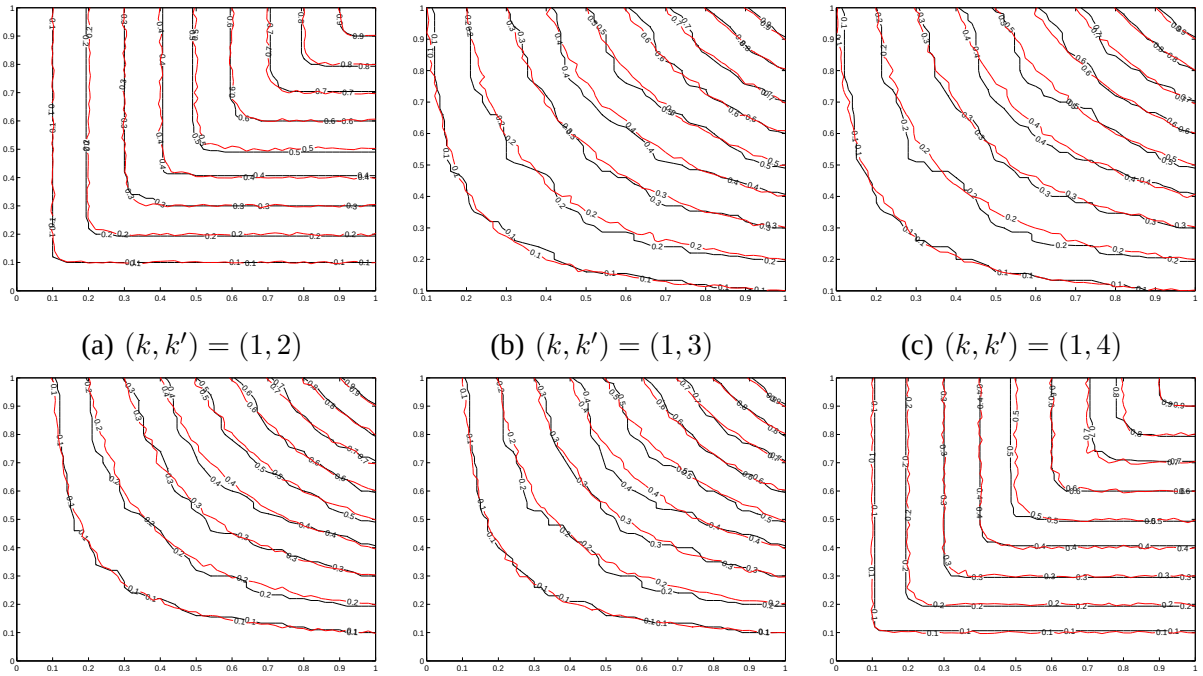
$$FAR(\lambda) = \frac{1}{N_1} \left\{ \sum_{i=1}^{N_{11}} \sum_{k=1}^{K_{11}} I\{ s_{11}(i,k) > \lambda\} + \sum_{i=1}^{N_{12}} \sum_{k=1}^{K_{12}} I\{ s_{12}(i,k) > \lambda\} \right\}. \tag{29}$$



(a) $(k, k') = (1, 2)$ (b) $(k, k') = (1, 3)$ (c) $(k, k') = (1, 4)$

(d) $(k, k') = (2, 3)$ (e) $(k, k') = (2, 4)$ (f) $(k, k') = (3, 4)$

Fig. 7. Nine level curves (at levels $0.1, 0.2, \ldots, 0.9$) indicating a good match between the empirical copula (black lines) and the estimated bivariate Gaussian copula (red lines) along dimensions k and k' for the intra-subject impostor scores.



(a) $(k, k') = (1, 2)$ (b) $(k, k') = (1, 3)$ (c) $(k, k') = (1, 4)$

(d) $(k, k') = (2, 3)$ (e) $(k, k') = (2, 4)$ (f) $(k, k') = (3, 4)$

Fig. 8. Nine level curves (at levels $0.1, 0.2, \ldots, 0.9$) indicating a good match between the empirical copula (black lines) and the estimated bivariate Gaussian copula (red lines) along dimensions k and k' for the inter-subject impostor scores.

The true but unknown values of $GAR(\lambda)$ and $FAR(\lambda)$ are the population versions of (28) and (29). The expression for the population $GAR(\lambda)$ is given by

$$E(GAR(\lambda)) = \frac{1}{N_0 K_0} \sum_{i=1}^{N_0} \sum_{k=1}^{K_0} P\{\, s_0(i,k) > \lambda\} = \frac{1}{K_0} \sum_{k=1}^{K_0} P\{\, s_0(1,k) > \lambda\} \equiv G_0(\lambda), \qquad (30)$$

where each set $\{\, s_0(i,k),\, k = 1, 2, \ldots, K_0\}$ for $i = 1, 2, \ldots, N_0$ are independent and identically distributed according to the copula model (14). The probabilities in (30) are functions of the unknown genuine marginal distributions, $F_{0,k}$, $k = 1, 2, \ldots, K_0$, and the genuine correlation matrix, R_0. Also, the second equality in (30) is a consequence of the identically distributed assumption. In a similar fashion, the population $FAR(\lambda)$ is given by

$$\begin{aligned} E(FAR(\lambda)) &= \frac{1}{N_1} \left\{ \sum_{i=1}^{N_{11}} \sum_{k=1}^{K_{11}} P\{\, s_{11}(i,k) > \lambda\} + \sum_{i=1}^{N_{12}} \sum_{k=1}^{K_{12}} P\{\, s_{12}(i,k) > \lambda\} \right\} \\ &= \frac{N_{11}}{N_1} \sum_{k=1}^{K_{11}} P\{\, s_{11}(i,k) > \lambda\} + \frac{N_{12}}{N_1} \sum_{k=1}^{K_{12}} P\{\, s_{12}(i,k) > \lambda\} \\ &\equiv G_1(\lambda), \end{aligned} \qquad (31)$$

where now, elements within each of the sets $\{\, s_{11}(i,k),\, k = 1, 2, \ldots, K_{11}\}$ for $i = 1, 2, \ldots, N_{11}$, and $\{\, s_{12}(i,k),\, k = 1, 2, \ldots, K_{12}\}$ for $i = 1, 2, \ldots, N_{12}$ are independent and identically distributed according to different forms of the copula model (14). The probabilities in (31) are functions of the unknown marginal distributions, $F_{11,k}$ for $k = 1, 2, \ldots, K_{11}$ and $F_{12,k}$ for $k = 1, 2, \ldots, K_{12}$, and the correlation matrices, R_{11} and R_{12}, for the intra-subject and inter-subject impostor scores, respectively. In the special case when the biometric entities are the different fingers, we have $K_{11} = N_{11} = 0$; thus, $N_1 = N_{12} K_{12}$, and

$$G_1(\lambda) = E(FAR(\lambda)) = \frac{1}{N_1} \sum_{i=1}^{N_{12}} \sum_{k=1}^{K_{12}} P\{\, s_{12}(i,k) > \lambda\} = \frac{1}{K_{12}} \sum_{k=1}^{K_{12}} P\{\, s_{12}(i,k) > \lambda\}. \qquad (32)$$

In light of the notations used for the population versions of FAR and GAR, equations (28) and (29) are sample versions of $G_0(\lambda)$ and $G_1(\lambda)$. Thus, we define

$$\hat{G}_0(\lambda) \equiv GAR(\lambda) \quad \text{and} \quad \hat{G}_1(\lambda) \equiv FAR(\lambda). \qquad (33)$$

The empirical ROC curve can be obtained by evaluating the expressions for GAR and FAR in (28) and (29) at various values λ based on the observed similarity scores, and plotting the resulting curve $(\hat{G}_1(\lambda), \hat{G}_0(\lambda))$. However, there is an alternative way in which an ROC curve can be constructed. Note that the ROC expresses the relationship between the FAR and GAR, and the threshold values are necessary only at the intermediate step for linking the FAR and GAR values. Thus, another representation of the ROC curve can be obtained by the following re-parameterization: we fix p as a value of FAR and obtain

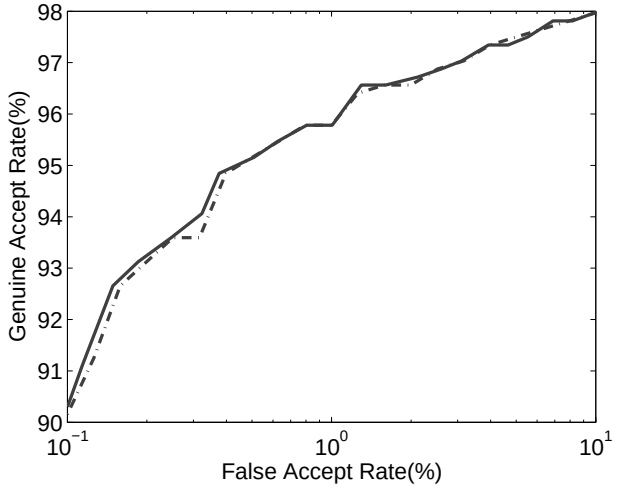


Fig. 9. ROC curves obtained using two different methods: The curves $(\hat{G}_1(\lambda), \hat{G}_0(\lambda))$ and $(p, \hat{W}(p))$ are represented by the dot-dashed and solid lines, respectively.

the threshold λ_* such that $\hat{G}_1(\lambda_*) = p$ or, $\lambda_* \equiv \hat{G}_1^{-1}(p)$. Substituting λ_* in (28) gives the ROC curve in the form $(p, \hat{W}(p))$, where

$$\hat{W}(p) = \hat{G}_0(\lambda_*) \equiv \hat{G}_0(\hat{G}_1^{-1}(p)). \tag{34}$$

Note that in the case when there is no λ_* such that $\hat{G}_1(\lambda_*) = p$, one can re-define the inverse, $\hat{G}_1^{-1}(p) \equiv \lambda_*$, where λ_* is the smallest λ satisfying $\hat{G}_1(\lambda) \leq p$. This definition of the inverse of $\hat{G}_1$ is more general and always yields a unique λ_*. The true but unknown ROC curve can be obtained in the same way as above by replacing the empirical versions with the corresponding population version; thus, we have

$$W(p) = G_0(G_1^{-1}(p)), \tag{35}$$

where $G_1^{-1}(p) \equiv \lambda_*$, where λ_* is the smallest λ satisfying $G_1(\lambda) \leq p$.

Although the two methods, $(\hat{G}_1(\lambda), \hat{G}_0(\lambda))$ and $(p, \hat{W}(p))$, of deriving the ROC curves are slightly different, they, nevertheless, are close approximations of one another for large N; Figure 9 present the two ROC curves for the fingerprint database described in [7]. We will use the parametrization given in (34) and (35) subsequently to develop confidence bands for the true ROC curve.

For fixed numbers C_0 and C_1 satisfying $0 \leq C_0 < C_1 \leq 1$, let us consider all $p = FAR$ values that fall in $[C_0, C_1]$. A confidence band for the true (claimed) ROC curve of a biometric system at confidence level $100(1-\alpha)\%$ gives two envelope functions, $e_L(p)$ and $e_U(p)$, so that for *all* p in $[C_0, C_1]$, the true ROC curve lies inside the interval $(\, e_L(p),\, e_U(p)\,)$ with probability of *at least* $100(1-\alpha)\%$. The numbers C_0 and C_1 form the lower and upper bounds of the range of FAR, and will be chosen to cover typical

reported values of FAR in biometric applications. If $C_0 = 0$ and $C_1 = 1$, the resulting ROC confidence band is constructed for the true ROC curve for all p in $(0, 1)$.

To obtain the envelopes, we first form the quantity

$$z \equiv \max_{C_0 \le p \le C_1} \sqrt{N_0}\, |\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)}|. \tag{36}$$

The function $f(x) = \sin^{-1}\sqrt{x}$ serves as a variance stabilizing transformation for quantities taking values in $(0, 1)$, such as $W(p)$ and $\hat{W}(p)$ (see, for example, [14]). Assume for the moment that the distribution of z is known. If $z_{1-\alpha}$ denotes the $100(1-\alpha)\%$ percentile of z, the envelopes are given by

$$e_L(p) = (\sin(\sin^{-1}\sqrt{\hat{W}(p)} - z_{1-\alpha}/\sqrt{N_0}))^2 \quad \text{and} \quad e_U(p) = (\sin(\sin^{-1}\sqrt{\hat{W}(p)} + z_{1-\alpha}/\sqrt{N_0}))^2. \tag{37}$$

However, the distribution of z is difficult to obtain analytically, and thus, we present two approaches to approximate the distribution of z in (36) based on (i) the bootstrap methodology, and (ii) an asymptotic representation of the distribution of z for large N_0.

A. The semi-parametric bootstrap approach

The value $z_{1-\alpha}$ will be found based on bootstrap samples from the fitted semi-parametric Gaussian copula models described in Section III. Let $\mathcal{S}_0^* \equiv \{\, s_0^*(i,k),\, k = 1, 2, \ldots, K_0,\, i = 1, 2, \ldots, N_0\}$, $\mathcal{S}_{11}^* \equiv \{\, s_{11}^*(i,k),\, k = 1, 2, \ldots, K_{11},\, i = 1, 2, \ldots, N_{11}\}$ and $\mathcal{S}_{12}^* \equiv \{\, s_{12}^*(i,k),\, k = 1, 2, \ldots, K_{12},\, i = 1, 2, \ldots, N_{12}\}$ denote the sets of genuine, intra-impostor and inter-impostor similarity scores obtained by one simulation from the fitted copula models. Also let

$$W^*(p) = G_0^*(G_1^{*^{-1}}(p)), \tag{38}$$

where $G_0^*(\lambda)$ (respectively, $G_1^*(\lambda)$) is obtained from equation (28) (respectively, (29)) with the bootstrap samples $s^*(i,k)$ used in place of the $s(i,k)$s. We form the quantity

$$z^* \equiv \max_{C_0 \le p \le C_1} \sqrt{N_0}\, |\sin^{-1}\sqrt{W^*(p)} - \sin^{-1}\sqrt{\hat{W}(p)}|, \tag{39}$$

with $\hat{W}(p)$ and $W^*(p)$ defined as in equations (34) and (38), respectively. By repeating the above procedure a large number of times, $B^* = 1,000$, we obtain $1,000$ values of z^*, $z_1^*, z_2^*, \ldots, z_{1,000}^*$. The $100(1-\alpha)\%$ percentile of the distribution of z^* can be approximated by $z^*_{[1000(1-\alpha)]}$, which is the $[B(1-\alpha)]$-th element in the ordered list of $z_1^*, z_2^*, \ldots, z_{1000}^*$. Thus, we approximate $z_{1-\alpha}$ by $z^*_{[1000(1-\alpha)]}$.

B. An asymptotic representation of z

We approximate the distribution of z asymptotically when N, the number of biometric entities, is large. Let $C_0 \equiv p_1 < p_2 < \ldots < p_m < p_{m+1} < \ldots < p_M \equiv C_1$ be a partition of the interval $[C_0, C_1]$. In the Appendix, we show that

$$z \equiv \max_{C_0<p<C_1} \sqrt{N_0}\,|\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)}| \approx \max_{1\le m\le M} |D\cdot\hat{G}_{0,M} + D\cdot\hat{G}_{1,M}|, \quad (40)$$

where D is a diagonal matrix with the (m,m)-th entry given by $1/\sqrt{4W(p)(1-W(p))}$, $D\cdot\hat{G}_{0,M}$ and $D\cdot\hat{G}_{1,M}$ are independent of each other, the distribution of $D\cdot\hat{G}_{0,M}$ (respectively, $D\cdot\hat{G}_{1,M}$) is approximately a M-dimensional multivariate normal with mean 0 (respectively, 0) and covariance matrix given by Γ_0 (respectively, Γ_1); see Appendix for the forms of Γ_0 and Γ_1. The maximum in $[C_0, C_1]$ is approximated by the component of the multivariate normal that takes on the maximum absolute value. We define

$$\max_{1\le m\le M} |D\cdot\hat{G}_{0,M} + D\cdot\hat{G}_{1,M}| \equiv z_M. \quad (41)$$

The distribution of z is approximated by the distribution of z_M for large M. Denoting the $100(1-\alpha)\%$ percentile of z_M by $z_{1-\alpha,M}$, the $100(1-\alpha)\%$ confidence interval for $W(p)$ is given by $(e_L(p), e_U(p))$ where

$$e_L(p) = (\sin(\sin^{-1}\sqrt{\hat{W}(p)} - z_{1-\alpha,M}/\sqrt{N_0}))^2 \quad \text{and} \quad e_U(p) = (\sin(\sin^{-1}\sqrt{\hat{W}(p)} + z_{1-\alpha,M}/\sqrt{N_0}))^2. \quad (42)$$

We compared the results of the semi-parametric and the asymptotic approaches with the non-parametric bootstrap, and found that the resulting upper and lower bounds of all the three approaches closely match with each other (see Figure 10) for our fingerprint database [7] with $N = 160$ subjects. This shows that the semi-parametric bootstrap approach and the asymptotic approach give good approximations to the true upper and lower confidence bands even for moderate sample sizes.

C. Confidence intervals for the correlation

The aim in this section is to set up confidence intervals for the correlation parameters, $\rho_{kk'}$, which are entries of the correlation matrix R of a multivariate Gaussian K-copula function. The confidence intervals can be obtained based on a bootstrap procedure similar to the one discussed in Section IV-A: If $B^* = 1,000$ samples, each of size N, are drawn from the estimated distribution $\hat{F}$, the estimation of the correlation matrix presented in Section III-B can be used to obtain $1,000$ estimated correlation matrices, $R^*_b, b = 1, 2, \ldots, 1,000$ with the (k, k')-th entry given by $\rho^*_{kk',b}$. The $100(1-\alpha)\%$ confidence interval

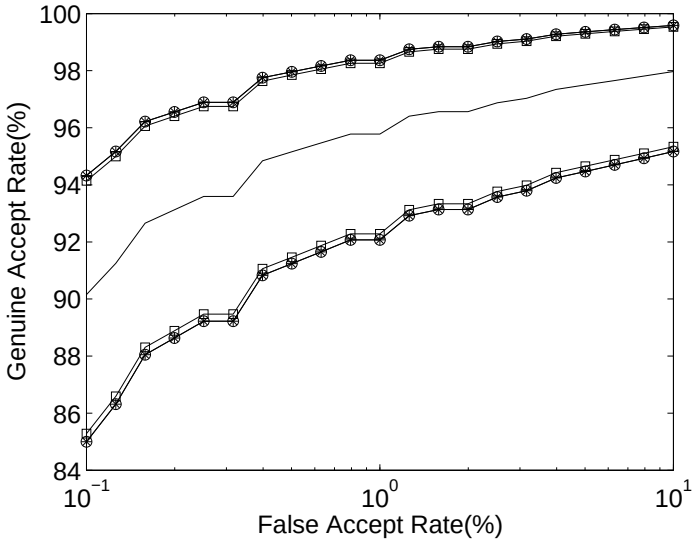


Fig. 10. Upper and lower ROC envelopes obtained using the three different methods: The non-parametric, semi-parametric bootstrap, and asymptotic envelopes are represented by the symbols $\circ$, $\square$, and $*$, respectively. The middle solid line is the non-parametric ROC curve.

for $\rho_{kk'}$ is given by (ρ_L^*, ρ_U^*), where ρ_L^* and $\rho_U^*)$ are, respectively, the $100\alpha/2$-th and $100(1-\alpha/2)$-th percentiles of the distribution of $\rho_{kk',b}^*, b = 1, 2, \ldots, 1,000$.

The 95% bootstrap confidence intervals for the genuine correlations are given in Table II. Note that the genuine correlations $\rho_{0,12}$ and $\rho_{0,34}$ are very high indicating significant amount of dependence between the similarity scores $s(B_1, B_2)$ and $s(B_2, B_1)$ for each of fingers 1 and 2, as expected. More importantly, we can also infer that there is no significant inter-finger correlation in the genuine case, that is, the different fingers are independent in the genuine case. This is because all of the confidence intervals for $\rho_{0,kk'}$ for $(k, k') = (1,3), (1,4), (2,3)$ and $(2,4)$ contain 0. We have also obtained the 95% confidence intervals for all the intra-subject and inter-subject impostor correlations using the bootstrap approach but do not present them here due to the larger dimensionality of the correlation matrices. We found that none of the confidence intervals in this case contain 0 indicating significant intra-finger and inter-finger correlations.

D. Testing the claim of a biometric vendor

Suppose that a vendor of a biometric authentication system claims that his/her biometric authentication system has a ROC curve given by $ROC_c = (p, W_c(p))$, for p in some interval $[C_0, C_1]$. Based on acquisitions from N biometric entities, we can test the validity of this claim by generating our own genuine and impostor similarity scores, and obtaining the $100(1-\alpha)\%$ confidence band for the true

k/k'	1	2	3	4
1	1	(0.976, 0.988)	(-0.008, 0.296)	(-0.003, 0.303)
2	*	1	(-0.006, 0.297)	(-0.002, 0.302)
3	*	*	1	(0.978, 0.989)
4	*	*	*	1

TABLE II

95% CONFIDENCE INTERVALS FOR THE GENUINE CORRELATION PARAMETERS $\rho_{0,kk'}$ USING 1000 BOOTSTRAP RE-SAMPLES. DUE TO SYMMETRY OF THE CORRELATION MATRIX, ONLY THE UPPER DIAGONAL ENTRIES ARE SHOWN. THE DIAGONAL ENTRIES OF THE CORRELATION MATRIX ARE ALWAYS UNITY.

ROC curve, $(p, W(p))$, for $p \in [C_0, C_1]$. We assume that the scores generated from users in the vendor's database are samples from an underlying population of genuine and impostor scores, and our samples based on N biometric entities are also samples from the same underlying population. If this assumption is true, then the confidence bands constructed from the previous section can be used for validating the vendor's claim. We perform the test

$$H_0 : W(p) = W_c(p) \quad \text{vs.} \quad H_1 : W(p) \neq W_c(p) \quad \text{for some } p, \tag{43}$$

and will accept H_0 (the claimed ROC curve) if

$$e_L(p) \leq W_c(p) \leq e_U(p) \tag{44}$$

for all $p \in (C_0, C_1)$; otherwise, we will reject it. We can also perform a test for claims of specific values of FRR and FAR, FRR_c and FAR_c. At $p_c = FAR_c$, we obtain the upper and lower limits of $GAR(p_c)$, $GAR_L(p_c)$ and $GAR_U(p)$. We will accept the claimed error rates if

$$GAR_L(p_c) \leq GAR_c \leq GAR_U(p_c) \tag{45}$$

where $GAR_c = 1 - FRR_c$, and reject it otherwise.

V. EXPERIMENTAL RESULTS

A. Construction of the ROC

Experiments were performed on the database [7] consisting of fingerprint impressions acquired in our laboratory from 4 different fingers (the right index, right middle, left index and left middle fingers) of 160 users. A total of 4 impressions per finger were obtained; 2 impressions were obtained on the first day and the remaining two after a period of a week. The fingerprint images were acquired using a solid state sensor manufactured by Veridicom, Inc, with image sizes 300×300 and resolution 500 dpi. A

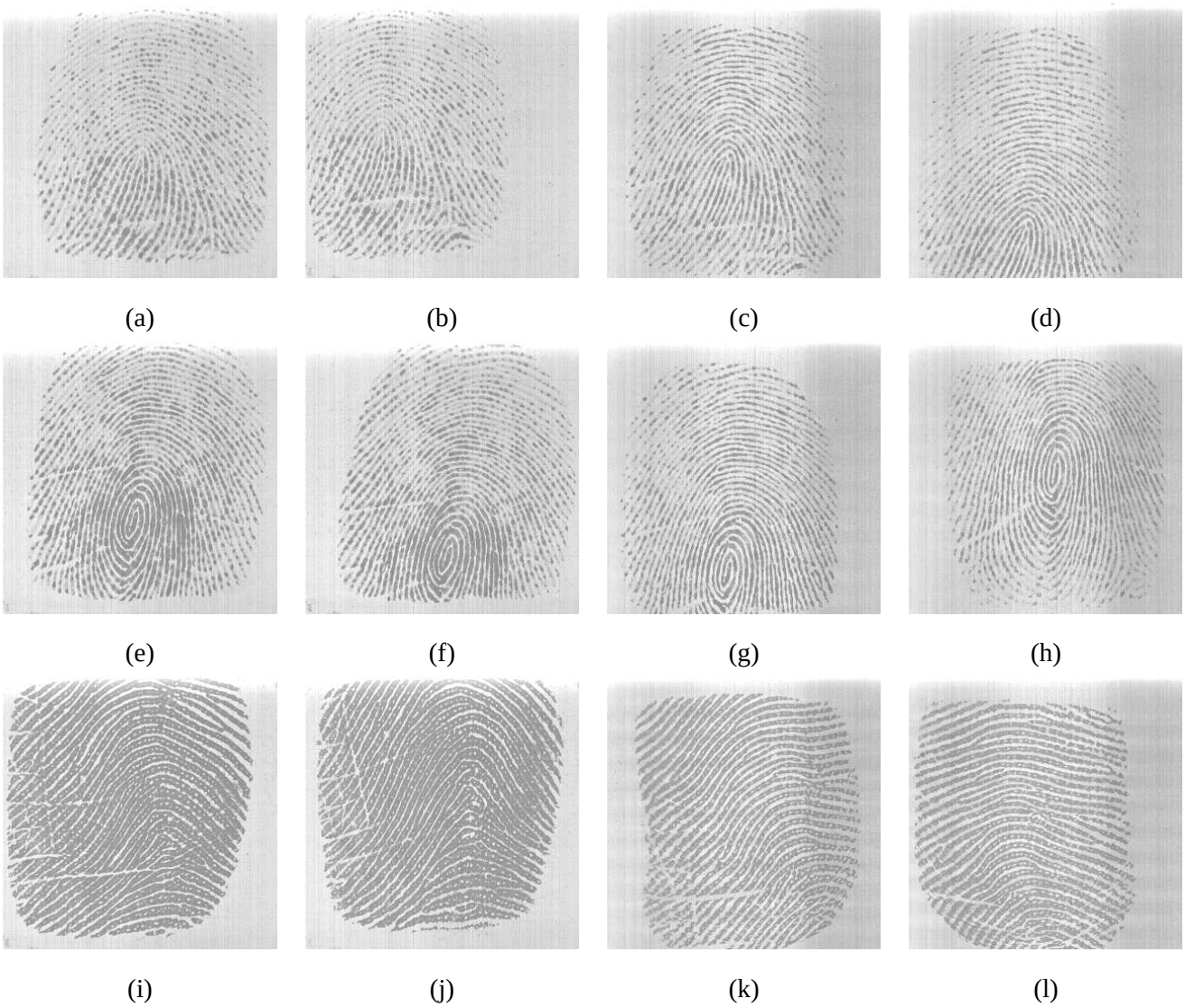


Fig. 11. Examples of fingerprint impressions from [7]: Each row gives the 4 impressions per finger collected. The first two rows are different fingers from the same subject, whereas the last row are fingerprint impressions from a different subject.

fingerprint similarity score was generated using an asymmetric matcher, described in [6]. Figure 11 show all 4 impressions of 3 fingers in this database. The first two fingers (first two rows) are from the same subject whereas the images in the last row are from a different subject.

The semi-parametric models based on the Gaussian copula functions were fit to all the genuine, intra-impostor and inter-impostor sets of similarity scores. Figures 3, 4 and 5 give the fitted density histograms (estimates of the marginals) based on 1000 samples simulated. We have also demonstrated that the elicited models are a good representation of the observed similarity scores (see Figures 3, 4 and 5 as well as Figures 6, 7 and 8). We also constructed the 95% confidence bands for the ROC based on the semi-

parametric bootstrap and the asymptotic approximation derived using the methodology in Section IV (see Figure 10) and have shown that the two methods give bands that are close to each other.

B. Effects of correlation on the ROC confidence bands

Our second set of experiments consisted of studying the effect of correlation among the multiple impressions of a user on the width of the ROC confidence band. As we require to vary the correlation, this experiment is not possible using real data since real data would give only one estimate of correlation for each of the sets of genuine, intra-subject and inter-subject impostor similarity scores. Our experiment was based on simulated sets of genuine, inter-subject impostor and intra-subject impostor similarity scores from the multivariate Gaussian K-copula models with Toeplitz forms for the correlation matrix. Let

$$R(\rho) = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} \tag{46}$$

denote the correlation matrix with all off-diagonal entries equal to ρ. The dimension of $R(\rho)$ will be different according to whether the sets of scores are genuine, intra-subject or inter-subject impostor scores. One advantage of selecting correlation matrices to be of the form (46) is that the matrix can be determined from specifying only a single real number, ρ, and is therefore, easy to use for illustrative purposes. Let ρ_c and ρ_d denote the intra- and inter-finger correlations (that is, the correlation between multiple impressions of a finger, and the correlation between impressions of two different fingers), respectively. For a given estimated correlation matrix $\hat{R}$, we try to find the most likely values for the intra-finger and inter-finger correlations. This is done by selecting those values of ρ_c and ρ_d that minimize the sum of Euclidean distances between $\hat{R}$ and $R(\rho_c) \otimes R(\rho_d)$,

$$||\hat{R} - R(\rho_c) \otimes R(\rho_d)||^2, \tag{47}$$

where $R(\rho_c)$ and $R(\rho_d)$ are as in (46) with ρ_c and ρ_d plugged in for ρ, respectively, and $\otimes$ is the Kronecker delta product. The minimizers of ρ_c and ρ_d, $\hat{\rho}_c$ and $\hat{\rho}_d$, for each of the genuine, intra-subject impostor and inter-subject impostor sets of scores, as well as the dimensions of each of $R(\rho_c)$ and $R(\rho_d)$ are given in Table III. The entries in Table III for the dimension assumes that the matcher is asymmetric.

In order to show the effects of increasing correlation on the ROC confidence bands, four combinations of (ρ_c,ρ_d) were selected. The first three combinations are (i) $(\rho_c = 0, \rho_d = 0)$, (ii) $(\rho_c = 0, \rho_d = \hat{\rho}_d)$, and (iii) $(\rho_c = \hat{\rho}_c, \rho_d = \hat{\rho}_d)$, where $\hat{\rho}_c$ and $\hat{\rho}_d$ are selected according to the entries of Table III for each set of

Sets/Estimates	$\hat{\rho}_c$	$\hat{\rho}_d$	dim $R(\rho_c)$	dim $R(\rho_d)$
Genuine	0.15	0.99	c	$d(d-1)$
Intra-Impostor	0.80	0.27	$c(c-1)$	d^2
Inter-Impostor	0.26	0.55	c^2	d^2

TABLE III

DIFFERENT VALUES OF $\hat{\rho}_c$ AND $\hat{\rho}_d$ FOR THE GENUINE, INTRA-SUBJECT IMPOSTOR AND INTER-SUBJECT IMPOSTOR SIMILARITY SCORES, AS WELL AS THE DIFFERENT DIMENSIONS OF $R(\rho_c)$ AND $R(\rho_d)$ FOR AN ASYMMETRIC MATCHER.

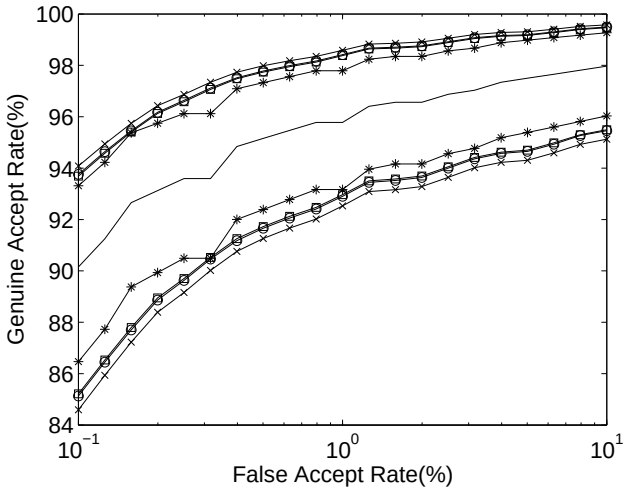


Fig. 12. Effects of correlation on the ROC confidence bands. The lines with '*', □, ○ and ×, respectively, denote the four different combinations of intra-finger and inter-finger correlations (i), (ii), (iii) and (iv).

genuine, intra-subject impostor and inter-subject impostor similarity scores. The fourth combination (iv) is obtained by setting the genuine ρ_c to 0.6 and the remaining ρ_cs and ρ_ds selected according to the entries in Table III. The 95% ($\alpha = 0.05$) level confidence bands for the ROC curve were constructed based on $B^* = 1,000$ bootstrap resamples. Figure 12 gives the ROC confidence bands based on the semi-parametric bootstrap. Note that the width of the confidence bands generally increases as we move from combination (i) to (iv). Another important point to be noted is that the confidence bands for combinations (ii) and (iii) are similar. This is because the width of the confidence bands relies more on the genuine, rather than the impostor, inter-finger and intra-finger correlations. Further, from the 95% confidence intervals obtained for the genuine inter-finger correlations, we found that these correlations were not significantly different from 0. The median widths of the confidence bands for the four combinations are 4.62, 5.41, 5.51, 6.06 respectively. The effects of correlation on the confidence bands using the asymptotic approach were similar to the bootstrap approach, and therefore, are not presented here.

Correlations	Values of c and d		
(ρ_c, ρ_d)	$c = 1, d = 2$	$c = 2, d = 2$	$c = 2, d = 3$
(0,0)	11,443 (246)	5,809 (148)	1,967 (31)
	22,885 (492)	23,235 (590)	11,801 (190)
$(0, \hat{\rho}_d)$	20,439 (790)	10,476 (279)	9,505 (263)
	40,877 (1,581)	41,905 (1,115)	57,028 (1,580)
$(\hat{\rho}_c, \hat{\rho}_d)$	21,403 (1,004)	11,056 (346)	9,749 (163)
	42,806 (2,008)	44,223 (1,382)	58,492 (977)
$(0.6, \hat{\rho}_d)$	19,015 (503)	13,321 (506)	11,558 (423)
	38,029 (1,006)	53,285 (2,026)	69,346 (2,540)

TABLE IV

VALUES OF n^* FOR ACHIEVING A WIDTH OF 1% FOR THE 95% CONFIDENCE BAND. THE TOTAL NUMBER OF OBSERVATIONS, n^*cd, ARE GIVEN BELOW THE n^* ENTRIES. ENTRIES ARE CALCULATED AS THE MEANS OF 10 SIMULATION RUNS. THE CORRESPONDING STANDARD DEVIATIONS ARE GIVEN IN PARENTHESIS.

C. Validation of the ROC confidence bands

The next experiment we conducted was to validate the ROC confidence bands at a specified confidence level. Recall that the $100(1-\alpha)\%$ ROC confidence bands, by definition, cover the true ROC curve with a probability of *at least* $100(1-\alpha)\%$ on repeated sampling from the underlying population of similarity scores. Treating the database [7] with $n = 160$ subjects as the underlying population, we selected a subset of 120 subjects from this population and constructed the semi-parametric bootstrap ROC confidence bands. We then determined if the population ROC curve (the empirical ROC curve for the 160 subjects) was within the constructed confidence bands. This procedure was repeated 200 times (with different subsets of 120 subjects from the population of 160), and each time, we determined if the population ROC curve was within the constructed ROC confidence bands. The percentage of coverage based on this validation procedure should be at least $100(1-\alpha)\%$. In our experiments we selected $\alpha = 0.05$ for the 95% ROC confidence bands, and obtained a coverage proportion of 99.5%.

D. Sample size requirements

As correlated multiple biometric observations affect the width of the ROC confidence bands, we now proceed to determine the number of users, n^*, required by a system to report a $100(1-\alpha)\%$ ROC confidence band with a width of at most w. We take $w = 1\%$. Our results are based on simulation with correlations selected according to combinations (i-iv) in Section V-B. Thus, the results reported here can be generalized to real data which exhibit different degrees of intra-finger and inter-finger correlations.

The values of n^* are given for different combinations of c and d, and therefore, varying dimensionality of the genuine, intra-subject and inter-subject sets of similarity scores. Consequently, we assume a common marginal for each of the three sets given by the mixture over component scores. We selected $C_0 = 0.1\%$, $C_1 = 10\%$ and $M = 21$ here, and $p_m = 10^{-1:0.1:1}$. For each $m = 1, 2, \ldots, M$, the width of the ROC confidence band at each $FAR = p_m$ is given by

$$w(p_m) = e_U(p_m) - e_L(p_m) = \frac{4z_{1-\alpha,M}\sqrt{W(p_m)(1-W(p_m))}}{\sqrt{n}} \tag{48}$$

for large n, where $z_{1-\alpha,M}$ is the $100(1-\alpha)\%$ percentile of the distribution of z_M defined in (41). In order to determine $z_{1-\alpha,M}$, we first estimate the covariance matrices Γ_0 and Γ_1 (see Section VI) as accurately as possible. This estimation is performed based on 1000 simulated samples from each of the correlation combinations (i-iv) for $n = 1000$ subjects. To achieve a width of w for the confidence band implies that $w(p_m) \leq w$ for all p_m, $m = 1, 2, \ldots, M$. Thus, the minimum number of users required is given by the formula

$$n^* = \left\lceil \max_{1\leq m\leq M} \left(\frac{4z_{1-\alpha,M}\sqrt{W(p_m)(1-W(p_m))}}{w(p_m)} \right)^2 \right\rceil + 1. \tag{49}$$

Table IV reports the average n^* over 10 runs of the estimation procedure. The numbers below n^* in Table IV give the average total number of observations n^*cd. The numbers in the parenthesis are the corresponding standard deviations over the 10 runs. In other words, if a biometric authentication system was tested based on n users, where n is chosen according to the entries in Table IV, we will be 95% certain that the true ROC *curve* will lie in the interval $[\hat{W} - 0.5\,,\, \hat{W} + 0.5]$. Table IV indicates that as the correlation among the multiple impressions of a finger increases for each fixed c and d, the total number of observations needed to achieve the width w for the confidence band increases. The same holds true when c and d values are increased for each correlation combination. Thus, in the presence of non-zero correlation, we are better off selecting a larger number of users rather than increasing the number of acquisitions per user. We also obtained the minimum sample sizes determined by the "Rule of 3" [19] and the "Rule of 30" [13]. Recall that the Rule of 3 and the Rule of 30 are rules of thumb to select the sample size, n, for the reliable estimation of an error probability, p, based on n independent binary observations, $x_1, x_2, \ldots, x_n$, with $P(x_i = 1) = 1 - P(x_i = 0) = p$ (see [19] and [13] for details). Since both the rules were derived for setting up confidence intervals for specific values of FAR and GAR (and not confidence bands for a range of FAR and GAR values), we were required to modify them slightly to suit the present case. For the Rule of 3, we computed the quantity $FRR_m = 1 - GAR(p_m)$ for $m = 1, 2, \ldots, M$ and derived the minimum sample size as

$$n_3 = \max_{1\leq m\leq M} \frac{3}{FRR_m}. \tag{50}$$

The smallest sample size based on the Rule of 30 was obtained using the formula

$$n_{30} = \max_{1 \leq m \leq M} \frac{(2 * 1.96)^2}{FRR_m}. \tag{51}$$

For the fingerprint database [7], n_3 was approximately 150 for all pairs of correlation combination, c and d, while n_{30} was approximately 770. Comparing the values of n_3 and n_{30} with n^*cd, we see that both n_3 and n_{30} grossly underestimate the total number of biometric acquisitions required to achieve a desired width. The underestimation becomes more prominent when significant correlation is present between multiple acquisitions of the biometric templates from a subject.

We also compared the minimum sample size requirements given by our method to that of the subset bootstrap approach [3]. One important point is that [3] obtains confidence rectangles, and not confidence bands, at each threshold value on the ROC curve. In order to perform a valid band to band comparison of the two methods, we applied the subset bootstrap procedure to the alternative parametrization of the ROC curve given in (34). As mentioned earlier, the subset bootstrap is not able to give an overall confidence level of $100(1-\alpha)\%$ using M individual $100(1-\alpha)\%$ confidence intervals. To guarantee a $100(1-\alpha)\%$ confidence level, the level of each individual confidence interval would have to be $100(1-\alpha/M)\%$ using Bonferroni's inequality. For $m = 1, 2, \ldots, M$, the minimum sample size requirement, $n_{sb}(m)$, for the m-th confidence interval can be obtained using similar asymptotic arguments as in Section IV-B with $C_0 = C_1 = p_m$. It follows that the minimum sample size required to achieve the pre-specified width for all M confidence intervals is given by

$$n^*_{sb} = \max_{1 \leq m \leq M} n_{sb}(m). \tag{52}$$

Table V presents the minimum sample size requirements for the subset bootstrap approach based on the $M = 21$ pre-selected FAR values. Note that the sample sizes required by our method is smaller compared to n^*_{sb} for achieving the same overall confidence level.

We performed similar experiments as mentioned above on a different database, namely, the West Virginia University fingerprint database consisting of fingerprint impressions from 263 different users. We used the first 2 impressions of the right index finger to obtain similarity scores with the same matcher used in the experiments above. Thus, in this case, we have $c = 1$ and $d = 2$. We found that the fit of the Gaussian copula functions to both the genuine and impostor sets of similarity scores were satisfactory. For validating the ROC confidence bands, the procedure described in Section V-C was carried out with sub-samples of 198 users. The procedure of constructing the ROC confidence bands was repeated 500 times. The empirical ROC curve (ROC curve based on the 263 users) was found to be inside the 95% confidence bands in 497 (out of the 500) trials, resulting in a coverage probability of 99.4%. We also obtained the

Correlations	Values of c and d		
(ρ_c, ρ_d)	$c = 1, d = 2$	$c = 2, d = 2$	$c = 2, d = 3$
(0,0)	48,674 (600)	24,201 (373)	8,143 (136)
	97,350 (1,200)	96,810 (1,493)	48,860 (814)
$(0, \hat{\rho}_d)$	90,725 (315)	46,209 (837)	43,500 (455)
	181,450 (630)	184,840 (3,346)	261,000 (2,729)
$(\hat{\rho}_c, \hat{\rho}_d)$	90,477 (407)	47,855 (430)	46,269 (968)
	180,950 (813)	191,420 (1,720)	277,620 (5,811)
$(0.6, \hat{\rho}_d)$	89,993 (429)	61,394 (884)	56,723 (826)
	179,990 (858)	245,570 (3,536)	340,340 (4,956)

TABLE V

VALUES OF n^*_{sb} AND $n^*_{sb}cd$. ENTRIES ARE CALCULATED AS THE MEANS OF 10 SIMULATION RUNS. THE CORRESPONDING STANDARD DEVIATIONS ARE GIVEN IN PARENTHESIS.

minimum sample size requirements for this database based on the methodology represented in Section V-D and compared it with the subset bootstrap approach. The estimated values of ρ_d (within finger correlation) was found to be 0.99 and 0.39, respectively, for the genuine and impostor sets of similarity scores. To illustrate the effects of correlation on the sample size requirement, we choose three combinations of the genuine and impostor within finger correlations, namely, $(\rho_d^{gen}, \rho_d^{imp}) = (0, 0), (0.49, 0.19)$ and $(0.99, 0.39)$ to reflect the no correlation (or, independence), intermediate and high correlation states. Table VI reports the average n^* over 10 runs for the width $w = 1\%$, with the average total number of observations, n^*d, given by the entries directly below the n^*s. The numbers in the parenthesis are the corresponding standard deviations over the 10 runs. Also, Table VII presents the minimum sample size requirements for the subset bootstrap approach based on the $M = 21$ pre-selected FAR values. Note here, again, that the sample sizes required by our method is smaller compared to the subset bootstrap approach for achieving the same overall confidence level.

VI. SUMMARY & CONCLUSION

With the growing popularity of biometric systems, it is often necessary to validate the performance levels of a system claimed by a vendor. One challenge in verifying the vendor's claim is that it is difficult to elicit models for the distribution of observed similarity scores. We present a flexible semi-parametric approach for estimating both the genuine and impostor distributions of similarity scores using multivariate Gaussian copula functions with non-parametric marginals. We emphasize that the Gaussian copula functions are used only to model the correlation and not the marginal density functions.

Correlations	Values of c and d		
$(\rho_d^{gen}, \rho_d^{imp})$	$c=1, d=2$	$c=1, d=3$	$c=1, d=4$
(0,0)	12,875 (283)	4,251 (77)	2,103 (37)
	25,749 (477)	12,754 (231)	8,412 (148)
$(0.49, 0.19)$	15,215 (513)	7,719 (215)	6,200 (299)
	30,430 (1,025)	23,158 (645)	24,799 (1,197)
$(0.99, 0.39)$	23,802 (886)	20,898 (414)	18,748 (698)
	47,604 (1,772)	62,693 (1,244)	74,991 (2,793)

TABLE VI

VALUES OF n^* FOR ACHIEVING A WIDTH OF 1% FOR THE 95% CONFIDENCE BAND BASED ON THE WEST VIRGINIA UNIVERSITY DATABASE. THE TOTAL NUMBER OF OBSERVATIONS, n^*d, ARE GIVEN BELOW THE n^* ENTRIES. ENTRIES ARE CALCULATED AS THE MEANS OF 10 SIMULATION RUNS. THE CORRESPONDING STANDARD DEVIATIONS ARE GIVEN IN PARENTHESIS.

Correlations	Values of c and d		
$(\rho_d^{gen}, \rho_d^{imp})$	$c=1, d=2$	$c=1, d=3$	$c=1, d=4$
(0,0)	47,526 (655)	16,170 (280)	8,144 (169)
	95,050 (1,310)	48,510 (841)	32,580 (676)
$(0.49, 0.19)$	61,195 (1,074)	35,053 (697)	29,149 (940)
	122,390 (2,148)	105,160 (2,091)	116,600 (3,761)
$(0.99, 0.39)$	90,334 (170)	86,357 (400)	84,478 (766)
	180,670 (304)	259,070 (1,200)	337,910 (3,064)

TABLE VII

VALUES OF n_{sb}^* AND $n_{sb}^* d$ BASED ON THE WEST VIRGINIA UNIVERSITY DATABASE. ENTRIES ARE CALCULATED AS THE MEANS OF 10 SIMULATION RUNS. THE CORRESPONDING STANDARD DEVIATIONS ARE GIVEN IN PARENTHESIS.

This approach enables modeling of correlation between pairs of similarity scores. Confidence bands for the ROC curve are constructed using (i) semi-parametric bootstrap re-samples, and (ii) asymptotic approximations derived from the estimated models. We also determine the minimum required number of subjects needed to achieve a desired width for the confidence band of the ROC curve. A loss of efficiency is observed (in terms of the total number of acquisitions required) when there is considerable correlation present between multiple acquisitions of the biometric per subject. In this case, we are better off selecting a larger number of users rather than increasing the number of acquisitions per user. The required number of subjects for achieving the desired confidence level for the ROC confidence band is much smaller

compared to other methods of selecting sample sizes, such as the subset bootstrap. Our future work will focus on extending the current framework to validate reported performance for multimodal systems.

ACKNOWLEDGMENTS

The authors wish to thank Karthik Nandakumar, Anoop Namboodiri, Arun Ross, Umut Uludag and Yi Chen for their help in conducting experiments. This research is partially supported by the NSF ITR grant 0312646.

APPENDIX

We derive several results below to validate the asymptotic representation of z in equation (40). In proving these results, we assume that the biometric entities considered are the different subjects, and the matcher S is asymmetric. These results can be extended in a straightforward manner to the case when the biometric entities are the different fingers (see Table I). Recall that the total number of subjects was denoted by n, and d impressions of c fingers for each subject were acquired for validating a vendor's claim. In this case, $N_0 = n$, $K_0 = cd(d-1)$, $N_{11} = n$, $K_{11} = c(c-1)d^2$, $N_{12} = n(n-1)$ and $K_{12} = c^2d^2$. The asymptotic results presented here will be for $n \to \infty$ with c and d fixed.

We will first derive the asymptotic theory for $\sqrt{N_0}(\hat{W}(p) - W(p))$, and then extend it to the quantity $\sqrt{N_0}(\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)})$. We denote the densities of G_0 and G_1, assuming they exist, by g_0 and g_1, respectively. The quantity $\sqrt{N_0}(\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)})$ is a continuous function of $p \in [C_0, C_1]$ since the component marginals and their estimates for the genuine, intra-subject impostor and inter-subject impostor joint distributions are continuous. In order to find the asymptotic distribution of $\sqrt{N_0}\max_{C_0 \le p \le C_1} |\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)}|$, we first define a partition of $[C_0, C_1]$: $C_0 \equiv p_1 < p_2 < \ldots < p_M \equiv C_1$. For large M, we have

$$\sqrt{N_0}\max_{C_0 \le p \le C_1} |\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)}| \approx \sqrt{N_0}\max_{1 \le m \le M} |\sin^{-1}\sqrt{\hat{W}(p_m)} - \sin^{-1}\sqrt{W(p_m)}|. \tag{53}$$

Thus, we first derive the joint asymptotic distribution of the M-dimensional vector $\sqrt{N_0}(\sin^{-1}\sqrt{\hat{W}(p_m)} - \sin^{-1}\sqrt{W(p_m)})$, $m = 1, 2, \ldots, M$, and then obtain the distribution of the maximum of the absolute values of these m components. Note that by Taylor's expansion, we have

$$\sqrt{N_0}(\sin^{-1}\sqrt{\hat{W}(p)} - \sin^{-1}\sqrt{W(p)}) \approx \frac{1}{\sqrt{4W(p)(1-W(p))}}\sqrt{N_0}(\hat{W}(p) - W(p)) = D(p)\sqrt{N_0}(\hat{W}(p) - W(p)) \tag{54}$$

for large N_0, where $D(p) = \frac{1}{\sqrt{4W(p)(1-W(p))}}$. In other words, we require to find the distribution of $D_M \cdot \hat{W}_M$ where

$$\hat{W}_M \equiv \sqrt{N_0}\,(\hat{W}(p_1) - W(p_1), \hat{W}(p_2) - W(p_2), \ldots, \hat{W}(p_M) - W(p_M))^T \tag{55}$$

is an M-dimensional vector and D_M is the diagonal matrix with the (m,m)-th entry given by $D(p_m)$. We introduce some notation before stating the main results. For $m = 1, 2, \ldots, M$, define ξ_m and $\hat{\xi}_m$ to be the p_m-th upper quantiles of G_1 and $\hat{G}_1$, respectively, that is

$$\xi_m \equiv G_1^{-1}(p_m) \qquad \text{and} \qquad \hat{\xi}_m \equiv \hat{G}_1^{-1}(p_m). \tag{56}$$

Since $\hat{G}_1 - G_1$ converges almost surely to 0, we have $\hat{\xi}_m - \xi_m \to 0$ as $N_0 \to \infty$. Also, denoting $\hat{G}_{0,M} \equiv \sqrt{N_0}\,(\,\hat{G}_0(\hat{\xi}_1) - G_0(\hat{\xi}_1),\, \hat{G}_0(\hat{\xi}_2) - G_0(\hat{\xi}_2),\, \ldots,\, \hat{G}_0(\hat{\xi}_M) - G_0(\hat{\xi}_M)\,)^T$ and $\hat{G}_{1,M} \equiv \sqrt{N_0}\,(\,G_0(\hat{\xi}_1) - G_0(\xi_1),\, G_0(\hat{\xi}_2) - G_0(\xi_2),\, \ldots,\, G_0(\hat{\xi}_M) - G_0(\xi_M)\,)^T$, we have

$$\hat{W}_M = \hat{G}_{0,M} + \hat{G}_{1,M}. \tag{57}$$

Results 1 - 4 below enable us to conclude that the components $\hat{G}_{0,M}$ and $\hat{G}_{1,M}$ are asymptotically independent, and the limiting distributions of $\hat{G}_{0,M}$ and $\hat{G}_{1,M}$ are multivariate normals with means 0 and covariance matrices given by Θ_0 and $\frac{N_0}{N_1}\Theta_1$, respectively; see Results 2 and 3 for the forms of Θ_0 and Θ_1, respectively. Thus, it follows that for the M-partition $C_0 \equiv p_1 < p_2 < \ldots p_M \equiv C_1$, the distribution of $\sqrt{N_0}(\sin^{-1}\sqrt{\hat{W}(p_m)} - \sin^{-1}\sqrt{W(p_m)})$ for $m = 1, 2, \ldots, M$ is given by

$$D \cdot \hat{W}_M = D \cdot \hat{G}_{0,M} + D \cdot \hat{G}_{1,M}. \tag{58}$$

Since $\hat{G}_{0,M}$ and $\hat{G}_{1,M}$ are asymptotically independent, it follows that $D \cdot \hat{G}_{0,M}$ and $D \cdot \hat{G}_{1,M}$ are also asymptotically independent, and the limiting distributions of $D \cdot \hat{G}_{0,M}$ and $D \cdot \hat{G}_{1,M}$ are multivariate normals with means 0 and covariance matrices given by $\Gamma_0 = D\Theta_0 D^T$ and $\Gamma_1 = D\Theta_1 D^T$, respectively. Since the covariance matrices above depend on unknown parameters, they will, in practice, be determined by plugging in parameter estimates in place of the unknown parameters; for example, the (m,m)-th entry of D, $D(p_m) = \frac{1}{\sqrt{4W(p_m)(1-W(p_m))}}$, will be estimated by plugging in $\hat{W}(p_m)$ in place of $W(p_m)$.

We now state the required results. Define $G_{11}(\lambda) = \frac{1}{K_{11}} \sum_{k=1}^{K_{11}} P\{s_{11}(1,k) > \lambda\}$ and $G_{12}(\lambda) = \frac{1}{K_{12}} \sum_{k=1}^{K_{12}} P\{s_{12}(1,k) > \lambda\}$. It follows then, that $G_1(\lambda) = \frac{N_{11}K_{11}}{N_1} G_{11}(\lambda) + \frac{N_{12}K_{12}}{N_1} G_{12}(\lambda)$. For $m = 1, 2, \ldots, M$, define $\xi_{12,m} = G_{12}^{-1}(p_m)$. The first result is

Result 1: The M-dimensional vector

$$\sqrt{N_{12}} \left(\frac{g_1(\xi_1)(\,\hat{\xi}_1 - \xi_1)}{\sqrt{p_1(1-p_1)}}, \frac{g_1(\xi_2)(\hat{\xi}_2 - \xi_2)}{\sqrt{p_2(1-p_2)}}, \ldots, \frac{g_1(\xi_M)(\hat{\xi}_M - \xi_M)}{\sqrt{p_M(1-p_M)}} \right)^T \to Z_M \tag{59}$$

where Z_M is a multivariate normal random variable with zero means, unit variances and correlation matrix given by

$$\Theta_{12}(m,m') = \frac{1}{K_{12}^2} \sum_{k=1}^{K_{12}} \sum_{k'=1}^{K_{12}} \theta_{12}(k,k',m,m') \tag{60}$$

where

$$\theta_{12}(k,k',m,m') = \frac{P\{s_{12}(1,k) > \xi_{12,m},\, s_{12}(1,k') > \xi_{12,m'}\} - P\{s_{12}(1,k) > \xi_{12,m}\}P\{s_{12}(1,k') > \xi_{12,m'}\}}{\sqrt{p_m(1-p_m)}\cdot\sqrt{p_{m'}(1-p_{m'})}}. \tag{61}$$

Proof: We consider the expression

$$\begin{aligned}
& P\left\{\sqrt{N_{12}}\frac{(g_1(\xi_m)(\hat{\xi}_m - \xi_m))}{\sqrt{p_m(1-p_m)}} \le x_m,\, m = 1,2,\ldots,M\right\} \\
= \;& P\left\{\hat{\xi}_m \le \xi_m + \frac{x_m}{g_1(\xi_m)}\sqrt{\frac{p_m(1-p_m)}{N_{12}}},\, m = 1,2,\ldots,M\right\} \\
= \;& P\left\{\hat{G}_1\left(\xi_m + \frac{x_m}{g_1(\xi_m)}\sqrt{\frac{p_m(1-p_m)}{N_{12}}}\right) > p_m,\, m = 1,2,\ldots,M\right\} \\
= \;& P\left\{K_{11}X_{11} + K_{12}X_{12} > N_{12}p_m,\, m = 1,2,\ldots,M\right\},
\end{aligned}$$

where X_u is a Binomial random variable with parameters N_u for the total number of trials and $p_u^m \equiv G_u(\xi_m + \frac{x_m}{g_1(\xi_m)}\sqrt{\frac{p_m(1-p_m)}{N_{12}}})$ as the probability of success in each trial, for $u = \{11\}$ and $\{12\}$. It follows that the last expression above can be re-written as $P\{K_{12}\, Z_{12}^m > Q^m,\, m = 1,2,\ldots,M\}$ where

$$Q^m = \frac{N_1 p_m - N_1 G_1\left(\xi_m + \frac{x_m}{g_1(\xi_m)}\sqrt{\frac{p_m(1-p_m)}{N_{12}}}\right) - K_{11}Z_{11}^m\sqrt{N_{11}p_{11}^m(1-p_{11}^m)}}{\sqrt{N_{12}p_{12}^m(1-p_{12}^m)}},$$

$Z_{11}^m = (X_{11} - N_{11}p_{11}^m)/\sqrt{N_{11}p_{11}^m(1-p_{11}^m)}$, and $Z_{12}^m = (X_{12} - N_{12}p_{12}^m)/\sqrt{N_{12}p_{12}^m(1-p_{12}^m)}$. As $n \to \infty$, using the Taylor's expansion for $G_1\left(\xi_m + \frac{x_m}{g_1(\xi_m)}\sqrt{\frac{p_m(1-p_m)}{N_{12}}}\right)$ and the facts that $N_{11}/N_{12} \to 0$, $N_1/N_{12} \to K_{12}$ and $p_{12}^m \to p_m$, we get $Q^m \to -K_{12}\, x_m$. The limiting distributions of each Z_u^m is normal with mean 0 and variance 1, for $u = \{11\}$ and $\{12\}$. Further, a computation of the covariance gives the expression (60) for the covariance between Z_{12}^m and $Z_{12}^{m'}$. QED.

For the next result, we need a few more notations. Define $\theta_0(k,k',m,m')$ by

$$\theta_0(k,k',m,m') = P\{s_0(1,k) > \xi_m\,,\, s_0(1,k') > \xi_{m'}\} - P\{s_0(1,k) > \xi_{12,m}\}\, P\{\, s_0(1,k') > \xi_{12,m'}\}, \tag{62}$$

and let Θ_0 be the $M \times M$ matrix whose (m,m')-th entry is given by

$$\Theta_0(m,m') = \frac{1}{K_0^2}\sum_{k=1}^{K_0}\sum_{k'=1}^{K_0}\theta_0(k,k',m,m'). \tag{63}$$

We state the following result:

Result 2: Let $\underline{t} = (t_1, t_2, \ldots, t_M)^T$. If $\hat{\varphi}_0(\underline{t})$ denotes the characteristic function of $\hat{G}_{0,M}$, and $\varphi_0(\underline{t}) \equiv \exp\left\{-\frac{1}{2}\underline{t}^T\,\Theta_0\,\underline{t}\right\}$ is the characteristic function of an M-dimensional normal with mean 0 and covariance matrix Θ_0, then

$$|\hat{\varphi}_0(\underline{t}) - \varphi_0(\underline{t})| \to 0 \tag{64}$$

as $n \rightarrow \infty$.

Proof: The proof of Result 2 will first involve conditioning on $\hat{\xi}_m$ for $m = 1, 2, \ldots, M$. Using the multivariate Central Limit Theorem [14], it follows that $\sqrt{N_0}(\hat{G}_0(\hat{\xi}_m) - G_0(\hat{\xi}_m))$ converges to an M-variate normal distribution with zero means and covariance matrix given by $\hat{\Theta}_0$, where $\hat{\Theta}_0$ is the matrix Θ_0 in (63) with $\hat{\xi}_m$ used in place of $\xi_{12,m}$. But, note that, $\hat{\xi}_m \rightarrow \xi_{12,m}$ so that $\hat{\Theta}_0 \rightarrow \Theta_0$. Result 2 follows. QED.

Next, we define the matrix Θ_1 with the (m, m')-th entry given by

$$\sigma_{12}(m, m') = \sqrt{p_m(1-p_m)} \cdot \frac{g_0(\xi_m)}{g_1(\xi_m)} \cdot \left(\frac{1}{K_{12}^2} \sum_{k=1}^{K_{12}} \sum_{k'=1}^{K_{12}} \theta_{12}(k, k', m, m') \right) \cdot \frac{g_0(\xi_{m'})}{g_1(\xi_{m'})} \cdot \sqrt{p_{m'}(1-p_{m'})}, \quad (65)$$

where $\theta_{12}(k, k', m, m')$ is as given in (61). We state

Result 3: Let $\underline{u} = (u_1, u_2, \ldots, u_M)^T$. If $\hat{\varphi}_1(\underline{u})$ denotes the characteristic function of $\hat{G}_{1,M}$ and $\varphi_1(\underline{u}) \equiv \exp\left\{ -\frac{1}{2} \frac{N_0}{N_{12}} \underline{u}^T \Theta_1 \underline{u} \right\}$, then

$$|\hat{\varphi}_1(\underline{u}) - \varphi_1(\underline{u})| \rightarrow 0 \quad (66)$$

as $n \rightarrow \infty$.

Proof: The m-th component of $\hat{G}_{1,M}$, $\sqrt{N_0}(G_0(\hat{\xi}_m) - G_0(\xi_m))$, can be written as $\sqrt{N_0} g_0(\xi_m)(\hat{\xi}_m - \xi_m)$ using Taylor's expansion for large n since $\hat{\xi}_m - \xi_m \rightarrow 0$. We can re-write this as

$$\sqrt{\frac{N_0}{N_{12}}} \frac{g_0(\xi_m)}{g_1(\xi_m)} \cdot \sqrt{p_m(1-p_m)} \cdot \left(\sqrt{N_{12}} \frac{(\hat{\xi}_m - \xi_m)}{\sqrt{p_m(1-p_m)}} \right). \quad (67)$$

Result (3) follows from applying Result 1 to (67). QED.

The next result is

Result 4: Let $\varphi_{0,1}(\underline{t}, \underline{u}) \equiv E(\mathrm{e}^{i\underline{t}^T \hat{G}_{0,M} + i\underline{u}^T \hat{G}_{1,M}})$ to be the characteristic function of $(\hat{G}_{0,M}, \hat{G}_{1,M})$. Then,

$$|\varphi_{0,1}(\underline{t}, \underline{u}) - \varphi_0(\underline{t}) \cdot \varphi_1(\underline{u})| \rightarrow 0 \quad (68)$$

as $n \rightarrow \infty$, where $\varphi_0(\underline{t})$ and $\varphi_1(\underline{u})$ are as defined in Results 2 and 3, respectively.

Proof: We first condition on all the impostor similarity scores. Thus, we have

$$\varphi_{0,1}(\underline{t}, \underline{u}) = E(\mathrm{e}^{i\underline{t}^T \hat{G}_{0,M} + i\underline{u}^T \hat{G}_{1,M}}) = E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,M}} E(\mathrm{e}^{i\underline{t}^T \hat{G}_{0,M}} \mid \mathcal{S}_{11} \cup \mathcal{S}_{12})) = E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,M}} \varphi_0^*(\underline{t})),$$

where $\varphi_0^*(\underline{t})$ is $\varphi_0(\underline{t})$ with Θ_0 replaced by $\hat{\Theta}_0$. Next, we have

$$|\varphi_{0,1}(\underline{t}, \underline{u}) - \varphi_0(\underline{t})\varphi_1(\underline{u})| = |E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,M}}(\varphi_0^*(\underline{t}) - \varphi_0(\underline{t})) + E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,M}} \varphi_0(\underline{t})) - \varphi_0(\underline{t})\varphi_1(\underline{u})| = |M_1 + M_2|,$$

where $|M_1| = |E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,L}}(\varphi_0{}^*(\underline{t}) - \varphi_0(\underline{t}))| \leq E|\varphi_0^*(\underline{t}) - \varphi_0(\underline{t}))| \rightarrow 0$ as $n \rightarrow \infty$ (since $\varphi_0^*(\underline{t})$ and $\varphi_0(\underline{t})$ are bounded functions by Result 2, and pointwise convergence implies convergence in expectation), and $|M_2| = |E(\mathrm{e}^{i\underline{u}^T \hat{G}_{1,L}} \varphi_0(\underline{t})) - \varphi_0(\underline{t})\varphi_1(\underline{u})| \leq |\hat{\varphi}_1(\underline{u}) - \varphi_1(\underline{u})| \rightarrow 0$. Result 4 follows. QED.

REFERENCES

[1] J. R. Beveridge, K. She, and B. A. Draper. A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), Hawaii*, December 2001.

[2] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. *Guide to Biometrics*. Springer, 2004.

[3] R. Bolle, N. Ratha, and S. Pankanti. Error analysis of pattern recognition systems: The subsets bootstrap. *Computer Vision and Image Understanding*, 93(1):1–33, January 2004.

[4] R. M. Bolle, S. Pankanti, and N. K. Ratha. Evaluation techniques for biometrics-based authentication systems (FRR). In *Proceedings of the 14th International Conference on Pattern Recognition, ICPR*, pages 2831–2837, August 2000.

[5] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Wiley, 2004.

[6] A. K. Jain, L. Hong, and R. Bolle. On-line fingerprint verification. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 19(4):302–314, 1997.

[7] A. K. Jain, S. Prabhakar, and A. Ross. Fingerprint matching: Data acquisition and performance evaluation. Technical Report TR99-14, Michigan State University, 1999.

[8] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[9] R. J. Micheals and T. E. Boult. Efficient evaluation of classification and recognition systems. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), Hawaii*, December 2001.

[10] R. G. Miller. *Simultaneous Statistical Inference*. Springer-Verlag, NY, 1981.

[11] D. F. Morrison. *Multivariate Statistical Methods*. McGraw-Hill, NY, 1990.

[12] R. B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

[13] J. Porter. On the “30 error criterion”. In J. Wayman, editor, *National Biometric Center Collected Works*, pages 51–56, 2000.

[14] C. R. Rao. *Linear Statistical Inference And Its Applications*. Wiley, 1991.

[15] M. E. Schuckers. Using the beta-binomial distribution to assess performance of a biometric identification device. *International Journal of Image and Graphics (Special Issue on Biometrics)*, 3(3):523–529, July 2003.

[16] U. K. Biometrics Working Group. Best practices in testing and reporting performance of biometric devices, 2000. Online: www.cesg.gov.uk/technology/biometrics.

[17] J. Wayman. Technical testing and evaluation of biometric identification devices. In A. K. Jain, R. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, 1999.

[18] J. Wayman. Confidence interval and test size estimation for biometric data. In J. Wayman, editor, *National Biometric Center Collected Works*, pages 91–95, 2000.

[19] J. Wayman. Technical testing and evaluation of biometric identification devices. In J. Wayman, editor, *National Biometric Center Collected Works*, pages 67–89, 2000.