A parametric correlation framework for the statistical evaluation and estimation of biometric-based classification performance in a single environment

Michael E. Schuckers

***Abstract*— In this paper, we propose a parametric correlation models for assessment of biometric classification error rates. Correctly specified correlations are integral to variance estimation and the corresponding inferential quantities which depend upon these estimates. We present methodology here for both false match and false non-match error rates for a single environment. This paper generalizes other work that has previously appeared in the bioauthentication literature. Since both symmetric and asymmetric matching algorithms are used in practice, we present a general correlation structure for both types of algorithms. Along with the correlation structure, we describe estimators for the parameters in these models. The correlation structure described here for binary decision data is then used to derive explicit confidence intervals and sample size calculations for estimation of false match and false non-match error rates. We then apply the correlation structure described herein to two match scores databases to illustrate our approach. A discussion of the utility and consequences of this correlation structure is also provided.**

***Index Terms*— false accept rate , false reject rate , false match rate, false non-match rate, biometric authentication, variance structure, effective sample size, confidence intervals, sample size calculations**

I. BACKGROUND

Bio-authentication devices have become increasingly prevalent and increasingly important for securing a wide range of locations, applications and data. One important facet of these devices is their matching performance — how well they make accurate decisions about whom to allow and whom to deny access. To estimate the matching performance error rates, some statistical methodologies have been proposed, e.g. Mansfield and Wayman [10]. Most of this previous work focuses on assessing matching performance through the false match and/or false non-match rate. We note here that false match and false non-match rates are averages of zero's and one's and, consequently, they follow statistical methodology for averages and linear combinations. Our goal here is methodology – confidence intervals and sample size calculations – for the statistical evaluation and assessment of matching performance in a single environment and not methodology for improved matching performance. To accomplish this, we propose a new parametric correlation structure for binary classification decisions that is useful in estimation of the variability in false match and false non-match rates. We use parametric here, not in the sense of Bolle et al. [2] who stated that "[p]arametric ... methods impose assumptions about the shape of the distribution...." Instead we use parametric in the sense that we are specifying the first two moments of the distributions of interest and the parameters involved in those moments; however, we are *not*, as will become clear, assuming a particular shape of the distribution. It is important to properly specify the correlation structure, and, hence, the variance structure, for statistical inference. Incorrect assumptions or other misspecifications regarding the correlations can lead to inaccurate standard errors and to confidence intervals for estimation of these error rates that are drastically too large or too small. Estimation of the parameters in the correlation structure using a method of moments approach is also presented.

The correlation structure is an integral aspect of assessing the variability in an estimate. This is especially important since biometric decision data, though binary, is thought to not be binomial, Mansfield and Wayman [10], and, hence, not uncorrelated. Methodologies such as those proposed by Givens et al [7], Mitra [11] and Schuckers [20] depend heavily on a correctly specified correlation structure for their approaches. Misspecification of such models can have significant consequences for the outcomes of any study. Here we will propose the first general parametric correlation structure that accounts for the source of the correlations, the same individual being used for multiple comparisons in biometric decision data. It is possible to estimate the variance (and, hence, the standard errors) of the false match and false non-match rates without assigning various components to a particular parameter; however, be-

---

Michael E. Schuckers is with St. Lawrence University and the Center for Identification Technology Research.

cause we are also interested in sample size calculations we use parameters for the various components of the variance. Wayman [23] proposed an estimator for the variance of a FNMR assuming only a single image per individual. That estimator seems to be based on simplified version of the correlation structure we will propose here. Bolle *et al* [2], Poh *et al* [14] and Schuckers [19] previously assumed an implicit correlation structure as part of their work. We generalize all of these approaches here. Our approach is designed specifically for decision data, i.e. the binary decision by an algorithm to claim a match or non-match since that is required for use when one reports error rates. See Ma *et al.* [9] for a taxonomy of biometric data types. The parameter estimates are dependent on the chosen threshold; however, the general structure itself is not. For notational simplicity we will not explicitly denote this dependence. That the estimates differ by threshold is made clear upon application of the model to data using varying thresholds. Additionally, we will deal with correlation for a single environment over which the process has a stationary covariance, that is for which the process has a fixed mean and variance.

We begin the description of our approach by presenting the theoretical basis for each parametric correlation structure. Following this, we provide moments-based estimators for the correlation parameters. Confidence interval approaches and sample size calculations are then derived based upon these methods. Our confidence interval approach is one that can be carried out solely on observed data; while the sample size approach depends - like all statistical sample size calculations - on some estimates of the process parameters. This is followed by illustrations of our approach by applying this methodology to several datasets. The structure of this paper is as follows. We begin by introducing the correlation structure for false non match rates in Section II. Section III discusses correlation structure for estimation of false match rates. Both symmetric and asymmetric classification algorithms are discussed. We illustrate these methods by applying them to biometric data from multiple modalities in Section IV. Finally, we discuss the implications for these correlation structures in Section V.

II. False Non-match Rate

In this section, we present a parametric correlation structure for decisions made when comparing two biometric captures from the same individual. This correlation structure is necessary for estimation of the the false non-match rate (FNMR). (Here we will use FNMR and FMR rather than False Reject Rate and False Accept Rate, respectively, to be consistent with Mansfield and Wayman [10].) Since we are interested in the FNMR, our decisions will be based on comparisons from two biometric captures from the same individual. These intra-individual decisions are derived from dichotomizing the genuine distribution, where the genuine distribution is defined as the match scores from intra-individual comparisons. Conditional on the error rate, we utilize a model that ignores — treats as zero — correlations between different individuals. We begin by introducing our correlation structure, then deriving the variance for the estimated error rate. That is followed by confidence intervals and sample size calculations for estimation of the FNMR.

A. Correlation Structure

Let Y_{ij} represent the decision for the j^{th} pair of captures collected on the i^{th} individual, where n is the number of individuals, $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Thus, the number of capture pairs that are compared for the i^{th} individual is m_i, and n is the number of different individuals being compared. We then define

$$Y_{ij} = \begin{cases} 1 & \text{if } j^{th} \text{ pair of captures from} \\ & \text{individual } i \text{ is declared a non-match} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We assume for the Y_{ij}'s that $E[Y_{ij}] = \pi_G$ and $V[Y_{ij}] = \pi_G(1-\pi_G)$ where $E[X]$ and $V[X]$ represent the mean and variance of X, respectively. Thus, π_G represents the FNMR and G here stands for genuine. The correlation structure for the $Y'_{ij}s$ is

$$Corr(Y_{ij}, Y_{i'j'}) = \begin{cases} 1 & \text{if} \quad i = i', j = j' \\ \rho & \text{if} \quad i = i', j \neq j' \\ 0 & \quad otherwise \end{cases} \tag{2}$$

This correlation structure for the FNMR is based upon the idea that the there will only be correlations between decisions made on the same individual but not between decisions made on different individuals. Thus, conditional upon the error rate, there is no correlation between decisions on the i^{th} individual and decisions on the $i'\ ^{th}$ individual, when $i \neq i'$. The degree of correlation is summarized by ρ.

Then

$$\begin{aligned} V[\hat{\pi}_G] &= V[N_G^{-1}\mathbf{1}^T\mathbf{Y}_G] = N_G^{-2}V[\mathbf{1}^T\mathbf{Y}_G] \\ &= N_G^{-2}\mathbf{1}^T\mathbf{\Sigma}_G\mathbf{1} = N_G^{-2}\pi_G(1-\pi_G)\mathbf{1}^T\mathbf{\Phi}_G\mathbf{1} \\ &= N_G^{-2}\pi_G(1-\pi_G) \\ &\times \left[N_G + \rho\sum_{i=1}^{n} m_i(m_i-1)\right] \end{aligned} \tag{3}$$

where $\mathbf{\Sigma}_G = Var(\mathbf{Y}_G)$, $N_G = \sum_{i=1}^{n} m_i$, $\mathbf{\Phi}_G = Corr(\mathbf{Y}_G)$, $\mathbf{1} = (1, 1, \ldots, 1)^T$, $\hat{\pi}_G = (N_G)^{-1}\mathbf{1}^T\mathbf{Y}_G$ and $\mathbf{Y}_G = (Y_{11}, \ldots, Y_{1m_1}, Y_{21}, \ldots, Y_{2m_2}, \ldots, Y_{n1}, \ldots, Y_{nm_n})^T$. Fleiss et al.

[6] has suggested the following estimator for ρ

$$\begin{aligned}\hat{\rho} &= \left(\hat{\pi}_G(1-\hat{\pi}_G)\sum_{i=1}^{n} m_i(m_i-1)\right)^{-1} \\ &\times \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{j'=1 \\ j'\neq j}}^{m_i}(Y_{ij}-\hat{\pi}_G)(Y_{ij'}-\hat{\pi}_G). \qquad (4)\end{aligned}$$

Note that this is equivalent to averaging all of the terms in the estimated correlation matrix that correspond to ρ to get a single estimate. Models like that found in Equation (2) are known as intra-individual or intra-class models and have been studied extensively in the statistics literature, e.g. Williams [24] or Ridout *et al* [15]. The parameter ρ in the models above represents the intra-class correlation. That is, it measures the degree of similarity between the binary decisions made on each individual. If all of the decisions on every individual are the same (all ones or all zeros), then ρ will be one. If the decisions on each individual are varying in a way that suggests that the decisions are not dependent upon the individual then ρ is zero, meaning that the observations are uncorrelated. Note that because we are considering correlations of binary observations that the usual intuition based upon, for example, Pearson's correlation coefficient is not applicable. Negative values of ρ are possible but such values suggest that decisions on captures from the same individual are less similar to each other than they are to all of the other decisions. This seems unlikely to be the case in the context of biometric authentication. Schuckers [18] implicitly used the model in Equation (2) to create confidence intervals and to derive sample size calculations. Several authors, including Fleiss et al. [6], have suggested using the following approximation to Equation (3)

$$V[\hat{\pi_G}] = N_G^{-1}\pi_G(1-\pi_G)(1+(m_0-1)\rho) \qquad (5)$$

where $m_0 = n^{-1}\sum_{i=1}^{n} m_i - (n\bar{m})^{-1}\sum\limits_{i=1}^{n}(m_i-\bar{m})^2$. If $m_i = m$ for all i, then $N_G = nm$ and the variance of $\hat{\pi}_G$ from Equation (5) becomes

$$\begin{aligned}V[\hat{\pi}] &= \frac{\pi_G(1-\pi_G)}{(nm)^2} * (nm + \rho nm(m-1)) \\ &= (nm)^{-1}(1+(m-1)\rho). \qquad (6)\end{aligned}$$

As mentioned above ρ represents the intra-class correlation. This quantity has a direct relationship with the variance of $\hat{\pi}$ found in Equation (3) and the simplified version found in Equation (6). As ρ increases, the variance in both cases increases. This is a consequence of the lack of independent information from each individual. If ρ is large, then each additional decision on a previously observed individual is providing little new information.

B. Statistical Estimation Methods

Being able to make statistical inference about an error rate based upon observed data is often of interest for evaluators of biometric authentication devices. Being able to determine the margin of error or the width of confidence intervals is often of interest to testers for determining the number of individuals and the number of attempts per individual. One advantage that a parametric approach to variance estimation enjoys is the ability to invert confidence intervals to determine the sample sizes necessary to achieve a particular margin of error for a confidence interval. Schuckers [18] has noted that there are effectively two quantities involved in any sample size calculation for biometric matching testing. The first is the number of individuals to be tested and the second is the number of times that each individual is to be tested. Below we describe confidence interval (CI) and sample size calculations for estimation of the FNMR. These methods are dependent on the correlation structure given above in Equation (2). Our estimand here is π_G. A $100(1-\alpha)\%$ CI for π_G is then

$$\hat{\pi}_G \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}_G(1-\hat{\pi}_G)[1+(m_0-1)\hat{\rho}]}{N_G}}, \qquad (7)$$

where $z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}^{th}$ percentile of a standard Gaussian distribution. We assume a Gaussian distribution following the generalized Central Limit Theorem in Moore [12]. Recall that N_G is the total number of decisions used to estimate π_G.

We now turn to sample size calculations. In order to create sample size calculations for a confidence interval, it is necessary to determine, among other things, the desired margin of error, B, for the interval. As mentioned above there are effectively two sample sizes when dealing with performance evaluation for biometric authentication devices. Here we will focus on the number of individuals, n, that need to be tested and assume that the number of decisions per individual is fixed and known. This is equivalent to assuming that $m_i = m$ for all i and that m is known. In practice it will be possible to determine different values for n by varying m before proceeding with a evaluation. We can obtain the following sample size calculation for making a $100(1-\alpha)\%$ CI with a margin of error of B.

$$n = \left\lceil \frac{z^2_{1-\frac{\alpha}{2}}\pi_G(1-\pi_G)(1+(m-1)\rho)}{mB^2} \right\rceil \qquad (8)$$

having replaced m_0 in Equation (7) with m. As with all sample size calculations it is important to note that we must specify *a priori* values for the parameters in the model. In this case we must estimate values for π_G and ρ before we determine the number of individuals, n. Traditionally, there are several ways to obtain estimates for these quantities. The two primary choices are to use previously collected similar data or to conduct a small pilot study. Having obtained those estimates, it is straightforward to then determine n.

III. FALSE MATCH RATE

We next focus on correlation models for false match rate (FMR) estimation in verification mode. As with the FMNR correlation sturctures, these models are based upon binary decisions because all biometric authentication decisions result in either an acceptance or a rejection. We are concerned here with the correlation between decisions made between two individuals and another decision made on two other individuals with the possibility of overlap among the individuals involved in each decision. The amount and type of overlap in individuals will be crucial to this structure. Thus, we aim to model the correlation between inter-individual decisions. These inter-individual decisions are based on dichotomizing match scores from the imposter distribution. We follow Mansfield and Wayman [10] in using the term imposter distribution to refer to match scores from inter-individual comparisons. Below we differentiate between classification algorithms that are symmetric and those that are asymmetric. Symmetric matching algorithms are those for whom the order of the individual captures does not matter and asymmetric ones are those for whom the order does matter. Thus for asymmetric matchers comparing capture A against capture B may result in a different decision than comparing capture B against capture A. The decisions resulting from a symmetric matching algorithm would always be the same regardless of the order of the captures. Below we propose a general correlation structure for the asymmetric case and show that the correlation structure for a symmetric matcher is a special case of this. As with the FNMR, we provide confidence intervals and sample size calculations based upon this structure.

A. General Correlation structure

In this section we derive a general correlation structure for use in estimation of the FMR. This general structure allows for the matching algorithm to be asymmetric. See Bistarelli et al. [1] for an example of such a matching algorithm. Let $Y_{ik\ell}$ represent the ℓ^{th} decision from the comparison pair $(i,k), \ell = 1,\ldots,m_{ik}, i = 1,\ldots,n, 1 \leq k \leq n, k \neq i$. Also let n be the total number of individuals from whom biometric data has been taken and let $m_{ik} \geq 0$ represent the number of comparisons on the pair of individuals (i,k) where the order of the pair matters, i.e. that m_{ik} is not necessary equal to m_{ki}. It is possible for $m_{ik} = 0$ when decisions for a particular comparison pair are not observed. In the general case, we consider both comparison pairs (i,k) and (k,i). This allows for possible asymmetry in the decisions, $Y_{ik\ell}$ not guaranteed to equal $Y_{ki\ell}$, is because we assume that when we compare the ℓ^{th} decision from the comparison pair, (i,k) to the ℓ^{th} decision from the pair (k,i) based on the same captures that the resulting decision may be different. We assume that the data is well ordered in the sense that it is possible to assume that the order of the captures for a given pair of individuals (i,k) is the same for the pair (k,i). Define

$$Y_{ik\ell} = \begin{cases} 1 & \text{if } \ell^{th} \text{ pair of captures from} \\ & \text{individual } i \text{ and individual } k \text{ is} \\ & \text{declared a match} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Let $E[Y_{ik\ell}] = \pi_I$ and $V[Y_{ik\ell}] = \pi_I(1-\pi_I)$. Thus the mean error rate or FMR is assumed to be π_I. We estimated this quantity through $\hat{\pi}_I = N_I^{-1}\mathbf{1}^T\mathbf{Y}_I$, where $\mathbf{1} = (1,1,\ldots,1)^T$. Further, let $\mathbf{Y}_{ik} = (Y_{ik1}, \ldots, Y_{ikm_{ik}})^T$ and $\mathbf{Y}_I = (\mathbf{Y}_{12}^T, \mathbf{Y}_{13}^T, \ldots, \mathbf{Y}_{1n}^T, \mathbf{Y}_{21}^T, \mathbf{Y}_{23}^T, \mathbf{Y}_{24}^T, \ldots, \mathbf{Y}_{2n}^T, \ldots, \mathbf{Y}_{n1}^T, \mathbf{Y}_{n2}^T, \ldots, \mathbf{Y}_{nn-1}^T)^T$. Here I stands for imposter.

As before all structure is dependent on the threshold from the matching process, but we will suppress that dependence for notational simplicity. Then,

$$Corr(Y_{ik\ell}, Y_{i'k'\ell'}) = \begin{cases} 1 & \text{if} \quad i=i', k=k', \ell=\ell' \\ \eta & \text{if} \quad i=i', k=k', \ell \neq \ell' \\ \omega_1 & \text{if} \quad i=i', k \neq k', i \neq k, i \neq k' \\ \omega_2 & \text{if} \quad i \neq i', k=k', i \neq k, i \neq k' \\ \omega_3 & \text{if} \quad i=k', i' \neq k, i \neq i', i \neq k \\ \omega_3 & \text{if} \quad i'=k, i \neq k', i' \neq i, k \neq k' \\ \xi_1 & \text{if} \quad i=k', k=i', i \neq i', k \neq k', \ell=\ell' \\ \xi_2 & \text{if} \quad i=k', k=i', i \neq i', k \neq k', \ell \neq \ell' \\ 0 & \quad\quad \text{otherwise} \end{cases} \tag{10}$$

This correlation structure is necessarily more complicated than the FNMR equivalent presented in (15) because of the correlation between decisions on captures from different individuals. In this correlation structure, η represents the correlation between decisions when the same individuals appear in the same order but the captures considered are different. In other words if the first decision is the 2nd capture from the 1st and 3rd individuals, Y_{132}, and the second decision is the 4th capture from the 1st and 3rd individuals, Y_{134}, then the correlation between these two decisions will be denoted by $\eta = Corr(Y_{132}, Y_{134})$. This is then an intra-comparison pair correlation similar to ρ in Equation (2). The ω's here represent the correlation between two decisions when those decisions contain one and only one individual in common.

ω_1 represents the case when the first individual in each decision is the same. For example, the first decision is based on individuals 1 and 2, $Y_{12\ell}$, and the second decision is based on individuals 1 and 3, $Y_{13\ell'}$. $\omega_1 = Corr(Y_{12\ell}, Y_{13\ell'})$. When the second individual is the same in each decision, we will assume that correlation will be represented by ω_2. An example of this would be when the first decision involves individuals the 3rd and 4th individuals, $Y_{34\ell}$ and the second decision involves the 2nd and 4th individuals, $Y_{24\ell'}$. Then, $\omega_2 = Corr(Y_{34\ell}, Y_{24\ell'})$. When an individual is shared between the decisions in either the 'inside' or the 'outside' positions, we will assume the correlation is ω_3. For example, the first decision might involve the 2nd and 4th individuals, $Y_{24\ell}$, and the second decision might involve the 4th and 1st individuals, $Y_{41\ell'}$. In that case we would assume the correlation between those decisions is $\omega_3 = Corr(Y_{24\ell}, Y_{41\ell'}) = Corr(Y_{41\ell}, Y_{24\ell'})$. The correlations denoted by the ξ's are those where the individuals are the same in each decision but the order is reversed. That is, decisions involved must include the same individuals, but the order of those individuals is different in one of the decisions. ξ_1 represents the case where the capture is the same but the order is different; while xi_2 represents the case where the captures are not the same and the order is not the same. To illustrate this, consider the 3rd and 4th capture on the individuals 5 and 7. So $\xi_1 = Corr(Y_{573}, Y_{754})$ and $\xi_2 = Corr(Y_{573}, Y_{754})$. Further, we assume that if all individuals involved in the decisions are all distinct then the correlation, conditional on the FMR, is zero. For example $Corr(Y_{24\ell}, Y_{37\ell'}$. In that case we will assume, conditional on the error rate, that the decisions are uncorrelated or have correlation zero. In general, we will assume that all of the correlation parameters here are non-negative. A negative correlation for the binary decisions described above would imply that having individuals in common would make decisions less likely to be same than those involving four distinct individuals. In a biometrics context this seems counterintuitive. Thus we assume a non-negative correlation when the same individual is part of both comparisons, but assume no correlation when four distinct individuals are part of these comparisons.

Using the structure in Equation (10), we can calculate the variance of our estimated error rate

$$
\begin{aligned}
& V[\hat{\pi}_I] \\
&= V[N_I^{-1}\mathbf{1}^T\mathbf{Y}_I] \\
&= N_I^{-2}V[\mathbf{1}^T\mathbf{Y}_I] \\
&= N_I^{-2}\mathbf{1}^T\mathbf{\Sigma}_I\mathbf{1} \\
&= N_I^{-2}\pi_I(1-\pi_I)\mathbf{1}^T\mathbf{\Phi}_I\mathbf{1}.
\end{aligned} \tag{11}
$$

This can be expanded to

$$
\begin{aligned}
& V[\hat{\pi}_I] \\
&= \frac{\pi_I(1-\pi_I)}{N_I^2}\left[N_I + \eta\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}(m_{ik}-1)\right. \\
&+ \omega_1\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{k'=1\\k'\neq i,k'\neq k}}^{n} m_{ik'}\right) \\
&+ \omega_2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}}^{n} m_{i'k}\right) \\
&+ \omega_3\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}}^{n} m_{i'i}+\right. \\
&\left.\sum_{\substack{k'=1\\k'\neq i,k'\neq k}}^{n} m_{kk'}\right) \\
&+ \xi_1\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki} \\
&+ \left.\xi_2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki}(m_{ki}-1)\right]
\end{aligned} \tag{12}
$$

where $\mathbf{\Sigma}_I = V[\mathbf{Y}_I]$ and $\mathbf{\Phi}_I = Corr(\mathbf{Y}_I)$. We derive parameter estimators using a method of moments approach by setting the sample variance of $\mathbf{Y}$, $(\mathbf{Y}_I - \hat{\pi}_I\mathbf{1})(\mathbf{Y}_I - \hat{\pi}_I\mathbf{1})^T$, equal to the model variance, Σ_I, whose correlation structure is defined in Equation (10) and solving for the correlation parameters. As in the previous section, this is equivalent to averaging the individual components of the sample correlation matrix that correspond to each correlation parameter. The moment-based estimator for η is then

$$
\begin{aligned}
\hat{\eta} &= \left(\hat{\pi}_I(1-\hat{\pi}_I)\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}(m_{ik}-1)\right)^{-1} \\
&\times \sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\ell=1}^{m_{ik}}\sum_{\substack{\ell'=1\\\ell'\neq\ell}}^{m_{ik}} (Y_{ik\ell}-\hat{\pi}_I)(Y_{ik\ell'}-\hat{\pi}_I).
\end{aligned} \tag{13}
$$

Note the similarities in this estimator to the estimator for ρ in the previous section. The other moment-based estimators for the correlation parameters are

$$
\begin{aligned}
&\hat{\omega_1} = \\
&\left(\hat{\pi}_I(1-\hat{\pi}_I)\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{k'=1\\k'\neq i,k'\neq k}}^{n} m_{ik'}\right)\right)^{-1} \times \\
&\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{k'=1\\k'\neq i\\k'\neq k}}^{n}\sum_{\ell=1}^{m_{ik}}\sum_{\ell'=1}^{m_{ik'}}(Y_{ik\ell}-\hat{\pi}_I)(Y_{ik'\ell'}-\hat{\pi}_I)\right), \\
&\hat{\omega_2} = \\
&\left(\hat{\pi}_I(1-\hat{\pi}_I)\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}}^{n} m_{i'k}\right)\right)^{-1} \times \\
&\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{i'=1\\i'\neq i\\i'\neq k}}^{n}\sum_{\ell=1}^{m_{ik}}\sum_{\ell'=1}^{m_{i'k}}(Y_{ik\ell}-\hat{\pi}_I)(Y_{i'k\ell'}-\hat{\pi}_I)\right), \\
&\hat{\omega}_3 = \\
&\left(\hat{\pi}_I(1-\hat{\pi}_I)\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\sum_{\substack{i'=1\\i'\neq i\\i'\neq k}}^{n} m_{i'i}+\right.\right. \\
&\left.\left.\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\sum_{\substack{k=1\\k\neq i}}^{n} m_{kk'}\right)\right)^{-1} \times \\
&\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{i'=1\\i'\neq i\\i'\neq k}}^{n}\sum_{\ell=1}^{m_{ik}}\sum_{\ell'=1}^{m_{i'i}}(Y_{ik\ell}-\hat{\pi}_I)(Y_{i'i\ell'}-\hat{\pi}_I)\right. \\
&+ \\
&\left.\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{k'=1\\k'\neq i\\k'\neq k}}^{n}\sum_{\ell=1}^{m_{ik}}\sum_{\ell'=1}^{m_{kk'}}(Y_{ik\ell}-\hat{\pi}_I)(Y_{kk'\ell'}-\hat{\pi}_I)\right), \\
&\hat{\xi}_1 = \\
&\left(\hat{\pi}_I(1-\hat{\pi}_I)\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki}\right)^{-1} \times \\
&\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\ell=1}^{m_{ki}}(Y_{ki\ell}-\hat{\pi}_I)(Y_{ki\ell}-\hat{\pi}_I)\right), \\
&\hat{\xi}_2 = \\
&\left(\hat{\pi}_I(1-\hat{\pi}_I)\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki}m_{ki}\right)^{-1} \times \\
&\left(\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\ell=1}^{m_{ki}}\sum_{\substack{\ell'=1\\\ell'\neq\ell}}^{m_{ki}}(Y_{ki\ell}-\hat{\pi_I})(Y_{ki\ell'}-\hat{\pi_I})\right).
\end{aligned}
\tag{14}
$$

Below we describe some possible simplifications to the general structure in Equation (10). We then provide confidence interval and sample size methodology.

B. Some simplified models of the general FMR correlation

Under certain conditions, it is reasonable to simplify the model in Equation (10). One such simplification could occur when the matcher is symmetric. As a consequence of this the correlation parameters ξ_1 and ξ_2, become 1 and η respectively. This is because the symmetric case results in identical results regardless of the order of the captures. Consequently with a symmetric matcher the correlation $\xi_1 = 1$ and the correlation $\xi_2 = \eta$. This can be written as

$$
\begin{aligned}
&Corr(Y_{ik\ell}, Y_{i'k'\ell'}) = \\
&\begin{cases}
1 & \text{if} \quad i = i', k = k', \ell = \ell', \\
\eta & \text{if} \quad i = i', k = k', \ell \neq \ell', \\
\omega_1 & \text{if} \quad i = i', k \neq k', i \neq k', \\
\omega_2 & \text{if} \quad k = k', i \neq i', k \neq k', \\
\omega_3 & \text{if} \quad i' = k, i \neq k', i \neq i', k \neq k' \\
\omega_3 & \text{if} \quad i = k', i' \neq k, i \neq i', k \neq k' \\
0 & \text{otherwise.}
\end{cases}
\end{aligned}
\tag{15}
$$

Estimation for this symmetric version of Equation (10) follows from the method of moments approach described previously.

For some statistical methods, including sample size calculations, a simplified version of the general model is helpful. Several such constrained models are reasonable. We discuss one of these below. Here we let $\omega = \omega_1 = \omega_2 = \omega_3$. This is equivalent to stating that the correlation between decisions is the same when two and only two of the four individuals involved in the decisions is the same. This yields a correlation structure of

$$
\begin{aligned}
&Corr(Y_{ik\ell}, Y_{i'k'\ell'}) = \\
&\begin{cases}
1 & \text{if} \quad i = i', k = k', \ell = \ell' \\
\eta & \text{if} \quad i = i', k = k', \ell \neq \ell' \\
\omega & \text{if} \quad i = i', k \neq k', i \neq k, i \neq k' \\
\omega & \text{if} \quad i \neq i', k = k', k \neq i, k \neq i' \\
\omega & \text{if} \quad i = k', i' \neq k, i \neq i', i \neq k \\
\omega & \text{if} \quad i' = k, i \neq k', i' \neq i, i' \neq k' \\
\xi_1 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell = \ell' \\
\xi_2 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell \neq \ell' \\
0 & \text{otherwise.}
\end{cases}
\end{aligned}
\tag{16}
$$

Other simplified models – e.g. $\omega_1 = \omega_2 = \omega_3$, $\xi_1 = \xi_2$ – may be reasonable for a given data collection or with some additional knowledge about a particular matching algorithm. For a given simplified model, deriving the method of moments based estimation for the correlation parameters should be straightforward.

C. Statistical Estimation Methods

We begin by deriving a confidence interval for the error rate of interest, the FNMR, based upon the estimated variance of our estimator. Here that error rate is the FMR, π_I. Thus a $100(1-\alpha)\%$ CI for π_I is

$$\hat{\pi}_I \pm z_{1-\alpha/2}\sqrt{\hat{V}[\hat{\pi}_I]} \tag{17}$$

where

$$\begin{aligned}
\hat{V}[\hat{\pi}_I] &= N_I^{-2}\hat{\pi}_I(1-\hat{\pi}_I) \\
&\times \left[\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik} + \hat{\eta}\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}(m_{ik}-1)\right. \\
&+ \hat{\omega}_1\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{k'=1\\k'\neq i,k'\neq k}}^{n} m_{ik'}\right) \\
&+ \hat{\omega}_2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}}^{n} m_{i'k}\right) \\
&+ \hat{\omega}_3\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ik}\left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}}^{n} m_{i'i}\right. \\
&+ \left.\sum_{\substack{k'=1\\k'\neq i,k'\neq k}}^{n} m_{kk'}\right) \\
&+ \hat{\xi}_1\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki} \\
&+ \left.\hat{\xi}_2\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} m_{ki}(m_{ki}-1)\right].
\end{aligned} \tag{18}$$

The above interval is based upon the general correlation structure given in Equation (10). We get the estimated variance, $\hat{V}[\hat{\pi}_I]$, by substituting our parameter estimates into Equation (11). It is possible to derive confidence intervals by assuming a simplified correlation structure and following this same process. Our assumption of Gaussianity is asymptotically valid since the estimate error rate $\hat{\pi}_I$ is a linear combination, Christofides and Mavrikiou [3] If we assume that $m_{ik} = m$ for all i and that $\omega_1 = \omega_2 = \omega_3 = \omega$, then the variance given in Equation (18) becomes

$$\begin{aligned}
V[\hat{\pi}_I] &= \hat{\pi}_I(1-\hat{\pi}_I)[n(n-1)m]^{-1}[(1+\hat{\xi}_1) \\
&+ (\hat{\eta}+\hat{\xi}_2)(m-1)+4\hat{\omega}(n-2)].
\end{aligned} \tag{19}$$

We can then invert Equation (19) and solve for n. As we did above, we will assume that the number of decisions per comparison pair is constant and known, $m_{jk} = m$, and then determine the value for n based upon this. Doing this and further assuming that $n \approx n-1 \approx n-2$, we get

$$\begin{aligned}
n &= \left\lceil \frac{2z_{1-\alpha/2}^2\omega\pi_I(1-\pi_I)}{B^2m}\right. \\
&+ \left(\left(\frac{2z_{1-\alpha/2}^2\omega\pi_I(1-\pi_I)}{B^2m}\right)^2\right. \\
&+ \frac{z_{1-\alpha/2}^2\pi_I(1-\pi_I)}{B^2m} \\
&\times \left.\left.[(1+\xi_1)+(\eta+\xi_2)(m-1)]\right)^{1/2}\right\rceil .
\end{aligned} \tag{20}$$

Here, as before, B is the margin of error desired and $z_{1-\alpha/2}$ is the $1-\alpha/2^{th}$ percentile of a standard Normal distribution. (That is, the desired confidence interval is $\hat{\pi}_I \pm B$.) *A priori* estimates of the parameters π, ω, ξ_1, ξ_2 and η need to be determined before these confidence intervals can be made. We also note that for a given n the total number of decisions will be $N_I = n(n-1)m$ assuming that all cross-comparisons are run. For Equation (20) we use only the 'plus' part of the solution to the quadratic equation. We do this since the only way for the 'minus' part of the quadratic equation to yield a positive result is for the $\eta+\xi_2$ to be negative. Negative values for these parameters in the context of biometric authentication decisions indicate negative correlations. This seems counterintuitive in the biometrics context and, thus, we truncate these correlations at zero. Other sample size calculations can be derived from the simplified models above. Similar calculations can also be found based on the symmetric correlation model found in the appendix.

IV. ILLUSTRATION

In this section we apply the methods from the two previous sections to actual biometric authentication data. The data we use here comes from two sources. The first set of data was collected at Michigan State University and will be referred to as the MSU database, [16]. The second database to which we apply our methods is the XM2VTS database as processed under the Lausanne Protocol I [13]. Estimates for the parameters of the correlation structures, both symmetric and asymmetric, are given for a range of error rates and thresholds. In addition, we give the standard errors, $\sqrt{\hat{V}[\hat{\pi}]}$, for the estimated error rates. These quantities are essential for making confidence intervals for π. Note that the standard errors would be $(N^{-1}\hat{\pi}(1-\hat{\pi}))^{1/2}$ if all of the correlation parameters are zero. The larger the correlation parameters the larger the variance. The impact of each particular correlation parameter is governed by the number of times that correlation appears within the correlation matrix, $\mathbf{\Phi}$. Again these estimates are assumed to be non-negative and so are truncated at zero when estimates are negative.

It is worth clarifying here that the correlations we are studying are for decisions made on the same device not the correlation between two devices or between two classifiers. We will discuss these issues further in Section V below The results below are not meant to be representative of possible values for all classifiers but rather they are meant to serve as illustrations of the methodology proposed above.

A. MSU database

The MSU database is a collection of similarity scores from three different biometric modalities - face, fingerprint and hand geometry. To convert these similarity scores to decisions, we dichotomize by a threshold, τ. Thresholds were chosen as reasonable choices for possible applications or to facilitate understanding of the consequences of different thresholds on the correlation structure. Tables I to IV have estimated parameters from applying the methods described in Sections II and III. The algorithms that produced the imposter match scores in this data set are all asymmetric. For each modality 50 individuals were involved in the data collection and each individual was compared against all of the others. The inter-individual data (or imposter data) contained 50×49 different comparison pairs each of which were observed 5 times. Thus there were $N_I = 12250$ total imposter match scores for the MSU database. This number is the same across all three modalities for this database. For the intra-individual or genuine comparisons, each individual was compared against themselves 10 times. Thus, there are $N_G = 500$ genuine match scores for each modality in the MSU database. For both the genuine and imposter match scores, we dichotomize each of these match scores by reasonable thresholds below to obtain decision data.

Table I contains parameter estimates based upon the correlation model given in Equation (10) for each of the three modalities in the MSU database. We denote by $*$ estimates that are treated as zero since they take negative values. From this table, we can see the impact of the correlations on the standard errors through the relationship between correlation and effective sample size is also dependent on the number of decisions that comprise each parameter estimate. As mentioned above, the higher the correlation the higher the standard error. Clearly, the standard error for the hand geometry FMR is much more than the equivalent for the facial or fingerprint FMR. This is expected based on two factors: the estimated error rate, $\hat{\pi}_I$, and the estimated correlation parameters. Both are directly related to the size of the standard error when the error rate is less than 0.5. The standard errors for the hand geometry, fingerprint and facial FMR's would be 0.0036, 0.0022, and 0.0019, respectively, if the decisions where uncorrelated. Thus we can see that the hand geometry's standard error is most inflated by the correlations in this model. The larger magnitude of the ω_1 plays a large role here since it is so much larger than the other values of the ω's in this table. ω's because of the number of permutations that they encompass tend to be present in the correlation matrix in larger numbers than η, ξ_1 or ξ_2. As a consequence, small differences in the ω's propagate to larger differences in the standard errors. Lastly regarding this Table I, we note that $\hat{\psi}$ is truncated at zero for all three classifiers.

Table II focuses on the FNMR's for the facial classifier in the MSU database. Each row of this table corresponds to a different threshold, τ. We calculate the parameter estimates at each threshold to illustrate that the estimates are different for each threshold. For this face classifier, we also observe several trends in this table. First, the parameter estimates differ from one threshold to the next. For example, $\hat{\omega}_1$ takes values that fluctuate from positive to truncated at zero. Second, we see that the standard errors vary significantly but they decrease as the error rate decreases. There is one exception to this when $\tau = 70$. This aberration is due to the estimated correlations - $\hat{\omega}_1$, $\hat{\omega}_2$ and $\hat{\omega}_3$ - between decisions when they share a single individual. It is possible for a very small correlation, e.g. $\hat{\omega}_1 = 0.0033$ when $\tau = 80$, to have a large impact on the standard error because of the number of pairs of decisions to which $\hat{\omega}_3$ is multiplied. For this case, the standard error for the correlated model is 50% larger, 0.0048 than the standard error would be in the uncorrelated case, $(0.1994(1-0.1994)/12250)^{1/2}$. Note that uncorrelated standard error is the same estimate that we should get by applying either the user-specific bootstrap of Poh *et al* [14] or the Beta-binomial approach of Schuckers [18] Recall Equation (10) for more details. Third, $\hat{\eta}$ and $\hat{\xi}_2$ are consistently near zero or truncated at zero indicating that the correlation between decisions on the same comparison pair is zero if different captures are being considered regardless of order. As we will see this seems to be peculiar to this particular matcher. Lastly, we see that $\hat{\omega}_3$ take positive values when the threshold, τ, is 40, 60 or 70. This is in contrast to the previous table where estimates of ω_3 were zero. Thus, we reiterate that the correlation parameters are dependent upon the threshold.

In Table III we move to presenting parameter estimates for the decisions based upon genuine match score data. The estimates here are based upon the correlation model found in Equation (2). Recall that for the MSU database $m_{ij} = 10$ intra-individual comparisons were observed for each of 50 individuals. Thus the total number of decisions here is $N_G = 500$. Table III calculates the estimated error rate,

TABLE I
Estimates for correlation structure parameters from MSU database,
False Match Rate, Asymmetric matchers, $N_I = 12250$

Modality	τ	$\hat{\pi}_I$	$\hat{\eta}$	$\hat{\omega_1}$	$\hat{\omega_2}$	$\hat{\omega_3}$	$\hat{\xi_1}$	$\hat{\xi_2}$	$\sqrt{\hat{V}[\hat{\pi}_I]}$
Hand	100	0.2031	0.0205	0.0209	0.0036	*0.0000	*0.0000	0.0061	0.0096
Finger	20	0.0603	0.0049	0.0003	0.0016	*0.0000	0.0193	0.0143	0.0027
Face	50	0.0443	*0.0000	0.0005	0.0042	*0.0000	0.0037	*0.0000	0.0027

* indicates truncated at zero

TABLE II
Estimates for correlation structure parameters from MSU database,
False Match Rate, Face, Asymmetric matcher, $N_I = 12250$

τ	$\hat{\pi}_I$	$\hat{\eta}$	$\hat{\omega_1}$	$\hat{\omega_2}$	$\hat{\omega_3}$	$\hat{\xi_1}$	$\hat{\xi_2}$	$\sqrt{\hat{V}[\hat{\pi}_I]}$
100	0.3773	*0.0000	0.0048	0.0023	*0.0000	0.0035	0.0000	0.0072
90	0.2811	*0.0000	0.0006	0.0020	*0.0000	0.0120	*0.0000	0.0052
80	0.1994	*0.0000	*0.0000	0.0033	*0.0000	*0.0000	*0.0000	0.0048
70	0.1357	*0.0000	0.0008	0.0058	0.0008	0.0031	*0.0000	0.0053
60	0.0863	*0.0000	0.0010	0.0052	0.0009	0.0029	*0.0000	0.0044
50	0.0443	*0.0000	0.0005	0.0042	*0.0000	0.0037	*0.0000	0.0027
40	0.0167	*0.0000	0.0008	0.0087	0.0001	0.0229	*0.0000	0.0021
30	0.0040	*0.0000	*0.0000	0.0042	*0.0000	*0.0000	*0.0000	0.0008

* indicates truncated at zero

TABLE III
Estimates for correlation structure parameters
from MSU database,
False Non-Match Rate, Asymmetric matchers,
$N_G = 500$

Modality	τ	$\hat{\pi}_G$	$\hat{\rho}$	$\sqrt{\hat{V}[\hat{\pi}_G]}$
Hand	100	0.1120	0.0392	0.0164
Finger	20	0.0760	*0.0000	0.0119
Face	50	0.0660	0.0086	0.0115

* indicates truncated at zero

TABLE IV
Estimates for correlation structure parameters
from MSU database,
False Non-Match Rate, Face, Asymmetric matcher,
$N_G = 500$

τ	$\hat{\pi}_G$	$\hat{\rho}$	$\sqrt{\hat{V}[\hat{\pi}_G]}$
100	0.0040	*0.0000	0.0028
90	0.0040	*0.0000	0.0028
80	0.0060	*0.0000	0.0035
70	0.0180	*0.0000	0.0059
60	0.0500	0.0129	0.0103
50	0.0660	0.0086	0.0115
40	0.1320	*0.0000	0.0151
30	0.2140	*0.0000	0.0183

* indicates truncated at zero

$\hat{\pi}_G$, the estimated intra-class correlation $\hat{\rho}$ for each threshold, τ, as well as the standard error for each estimated error rate. For the FNMR, the relationship between the observed correlation and the standard errors is clearer than for the FMR since there is only a single correlation term. Here we also note that the decisions for the false non-match rate of the fingerprint classifier effectively have a correlation of zero which results in the standard errors being what we would expect them to be for uncorrelated data. To see that the parameter estimates change for a particular classifier as the threshold changes for the correlation structure of the FNMR, we present Table IV. We note that ρ takes positive values only when $\tau = 50$ or 60. Thus, it is necessary to estimate all of the parameters each time we are interested in a different threshold. This table uses the same thresholds and the same classifiers that appeared in Table II. For the FNMR the correlation estimates change little as the thresholds change.

B. XM2VTS database

This data was processed by Poh and Bengio under the Lausanne Protocol I [13]. Included in this data are both speaker and facial recognition match scores. With the results of estimating the correlation models above, we include the matching algorithm as well as the modality for data from the XM2VTS database. More details on these classifiers and the data processing can be found in [13]. The methodology for this data collection ensured that not all cross-comparisons were considered. Individuals were partitioned into a gallery and a probe set for the inter-individual or imposter comparisons. Individuals in each of those groups were only compared against members of the other group, not against members of their own group. Because of this structure to the data collection, we force $\omega_3 = 0$ since no data is available for estimation of this quantity. Additionally, the order of the comparisons was never reversed, so that in cases where a comparison of a capture from individual A versus an capture from individual B was recorded there is no comparison reversing the order of the individuals. As a consequence of this, we will treat all of the classifiers in this database as symmetric. There were 200 individuals in the gallery and 25 in the probe. 8 match scores were recorded for each inter-individuals comparison pair. Thus, the total number of imposter decisions is $N_S = 200 \times 25 \times 8 = 40000$. For the intra-individual or genuine comparisons, the 200 gallery individuals were each compared to themselves 3 times. This yields $N_G = 600$ total genuine decisions.

Table V has the parameter estimates for the symmetric correlation model found in Equation (15) applied to the XM2VTS database. We selected a range of thresholds to achieve different error rates for the different classifiers. The estimated correlations for this data are appreciably larger than those for the MSU database, cf. Table I. One consequence of these larger correlations is standard errors that would be much larger than if the decisions were uncorrelated. For example, consider the Facial matcher DCTb, MLP. The observed standard error using our correlation model is 0.0072 while the standard error assuming uncorrelated decisions is $(0.0580(1-0.0580)/40000)^{1/2} = 0.0012$. Thus the former is six times larger than the naïve uncorrelated approach. The estimated correlations between decisions on the same comparison pair, η, are much higher for the XM2VTS database relative to the equivalent quantities, η, ξ_1 and ξ_2, in the MSU database. This is clearly one reason for the large standard errors that we observe here. Similarly, the other correlation parameters estimates are consistently larger in magnitude than the equivalent estimates for the MSU database. The results in Table V also illustrate that the same modality may have different correlation parameter estimates depending upon the classifier. This can easily be seen by looking at the columns corresponding to $\hat{\omega}_1$ and $\hat{\omega}_2$.

The last table of output we consider contains estimates of the correlation structure for the FNMR found in Equation (2) for the XM2VTS database. These results can be seen in Table VI. Here we utilize the same thresholds and the same matchers as found in Table V. As we noted in the previous paragraph, the estimates for the XM2VTS database are generally much larger than those for the MSU database. This is again true for the results in estimates of the correlation parameters for the FNMR. However, the impact on the effective sample size are not as drastic since the number of comparisons that go into estimation of ρ is smaller relative to the number of comparisons that go into estimation of the parameters of the symmetric false match correlation model. Similarly the effective sample sizes are larger relative to the sample size for the FNMR than for the FMR in the XM2VTS database.

In this section we have applied the correlation structure proposed in Sections III and II to data from the MSU and XM2VTS databases. The results of this application, found in Tables I to VI, have illustrated the utility of the method of moments estimators we have proposed here. To further demonstrate the consequences of the correlation structures we propose, we introduced the effective sample size. From these results here it is clear that the estimated correlations vary from threshold to threshold even for the same classifier.

V. DISCUSSION

This paper presents parametric correlation structures for the evaluation of the matching performance of a single biometric authentication device. This correlation structure is based upon binary decisions of match or non-match. In addition to these correlation structures we have introduced confidence interval and sample size calculations based upon this structure. For decision data involving inter-individual comparisons (imposter decisions) we have proposed correlation structures for both symmetric and asymmetric matching algorithms. The correlation structure for false match decision data is much simpler than the correlation structure for false non-match decision data. This is due to the need to address correlations between two decisions that involve one or more individual that is part of both decisions. False non-match data only involves intra-individual comparisons and, hence, the correlation structure is more straightforward. We have also provided method of moments estimators for the parameters in these models. The models we have proposed are specific to verification or one-to-one mode. However, the model for false non-match decisions might reasonably be applied to identification mode decisions for false match data because identification mode only specifies a single individual and so we might reasonably assume that the correlation between repeated decisions concerning that same individual were correlated but there would be no correlation between identification decisions for different individuals.

Along with the general correlation models presented above, we have presented constrained models which simplify these structures. The correlation models here, both the more general and the simplified, are the first to be explicitly provided for biometric data of any kind. The models presented here are a generalization of the implicit correlation structures found in work by Schuckers [18], Poh *et al* [14] and in Bolle *et al.* [2]. The first two works implicitly assume the model we have explicitly described here.

TABLE V
Estimates for correlation structure parameters from XM2VTS database,
False Match Rate, Symmetric matchers, $N_I = 40000$

Modality	Matcher	τ	$\hat{\pi}_S$	$\hat{\eta}$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\sqrt{\hat{V}[\hat{\pi}_S]}$
Face	(FH, MLP)	0.0	0.0038	0.2271	0.0014	*0.0000	^{t}0.0000	0.0007
Face	(DCTs, GMM)	0.0	0.0582	0.2935	0.0163	0.0193	^{t}0.0000	0.0067
Face	(DCTb, GMM)	0.2	0.0041	0.2599	0.0009	0.0009	^{t}0.0000	0.0007
Face	(DCTs, MLP)	-0.8	0.1057	0.3484	0.0165	0.0075	^{t}0.0000	0.0086
Face	(DCTb, MLP)	-0.5	0.0580	0.2565	0.0215	0.0067	^{t}0.0000	0.0072
Speaker	(LFCC, GMM)	3.0	0.0142	0.3329	0.0226	0.0184	^{t}0.0000	0.0039
Speaker	(PAC, GMM)	2.0	0.0570	0.3500	0.0324	0.0247	^{t}0.0000	0.0090
Speaker	(SSC, GMM)	1.0	0.0692	0.4435	0.0339	0.0505	^{t}0.0000	0.0105

* indicates truncated at zero

^{t} indicates that it is a theoretical zero in the case of this data since the combinations of (i,k) that lead to the estimation of these parameters did not occur in this data.

TABLE VI
Estimates for correlation structure parameters from XM2VTS database,
False Non-Match Rate, Symmetric matchers,$N_G = 600$

Modality	Matcher	τ	$\hat{\pi}_G$	$\hat{\rho}$	$\sqrt{\hat{V}[\hat{\pi}_G]}$
Face	(FH, MLP)	0.0	0.0350	0.3091	0.0095
Face	(DCTs, GMM)	0.0	0.0400	0.1319	0.0090
Face	(DCTb, GMM)	0.2	0.0283	0.2130	0.0081
Face	(DCTs, MLP)	-0.8	0.0033	*0.0000	0.0024
Face	(DCTb, MLP)	-0.5	0.0400	0.2188	0.0096
Speaker	(LFCC, GMM)	3.0	0.0050	0.3300	0.0037
Speaker	(PAC, GMM)	2.0	0.0583	0.3628	0.0126
Speaker	(SSC, GMM)	1.0	0.0267	0.3579	0.0086

* indicates truncated at zero

In the case of the latter, Bolle *et al.* aim to make the subsets of observations ”as independent as possible” (p.15). This works well for the FNMR decision data; however, if the ω’s in the FMR correlation model in Equation (10) are non-zero then this approach underestimates the variances of the estimated error rates. There are some analogies here to the difficulties of bootstrapping spatial data. See Solow [21], Cressie [4] and Tang [22] for more details. The work of Lahiri [8] holds some potential for non-parametric inference for the full correlation structure described here. Our approach is quite flexible in the type of correlation structure among decisions that it addresses and thus our approach generalizes these previous methodologies. Our resulting structure provides a more thorough understanding of the variability in error rate estimation as well as a more accurate approach to inference that comes with such knowledge. With a misspecified correlation structure it is possible to overestimate the precision of the FMR or to claim that a FNMR is significantly less than a given value when that is not the case.

We have also presented here methods for confidence intervals , Equations (7) and (17), and sample size calculations, Equations (8) and (20). These methods allow for appropriate inference to be made about a single error rate and for the determination of the number of individuals that need to be tested. We recommend that these confidence interval methods only be used when N – either N_G or N_I is large, i.e. when $N\left[\frac{\hat{\pi}(1-\hat{\pi})/N}{\hat{V}[\hat{\pi}]}\right]\hat{\pi} \geq 10$. This corresponds to an effective sample size times our estimated error rate being greater than or equal to 10. See Eng[5] for discussions of the effective sample size. These correlation structures also have application to other statistical methods already in the literature which depend upon specification of the correlation structures, e.g. Givens *et al.* [7], Mitra [11] and Schuckers [20]. While our focus has been on decision data, it is likely that a similar correlation structure would apply to match score/similarity score data. While the correlation structure may be appropriate, the individual variances will clearly be different. It will be necessary to replace the $\pi(1-\pi)$ in each variance equation with a general term, perhaps σ^2 for the variance of the match scores. The models and structures specified here will also find application in the estimation of ROC curves. In particular, it may be possible to create confidence region estimates for the entire ROC curve. This approach would utilize the improved variance estimation techniques given here to derive these regions.

As part of this work, we have also offered methodology for estimating the correlation parameters that are part of each model. These estimators are based on a method-of-moments methodology.

It should be clear from the examples that we have given that the larger the correlation the the larger the variance of the estimated error rates. Further, we have illustrated how these methods can be applied to various decision data from a range of modalities. While these estimating equations are lengthy there application using a software package such as Matlab^{TM} or the statistical software R is straightforward. It is clear from these examples that it is important to understand this correlation structure in order to properly estimate and infer about biometric matching error rates. Additionally, we again note that the correlations for our proposed models may vary from threshold to threshold. Thus it is important to re-estimate the parameters of the correlation model at each threshold.

With the growth of multimodal biometrics work, Ross *et al* [17,], there is growing interest in the correlation between different classifiers. Because the correlation between any two matchers is dependent on the variance of each, it is important to first treat and assess the performance each matcher individually as we have done here. It will be important to an accurate understanding of multibiometric performance to derive models similar to those given here. Likewise it is also important to consider inferential methods for looking at the relationship and possibly significant differences between the performance of two or more biometric authentication systems. This will no doubt be dependent on the correlations in the decisions for each system. This serves to underscore the important of getting the correlation structure for biometric authentication decisions correct.

ACKNOWLEDGMENTS

Funding for this work comes from NSF grants CNS-0325640 and CNS-0520990. CNS-0325640 is cooperatively funded by the National Science Foundation and the United States Department of Homeland Security. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] S. Bistarelli, F. Santini, and A. Vaccarelli, "An asymmetric fingerprint matching algorithm for java card^{TM}," in *Proceedings of Audio- and Video-based Biometric Person Authentication 2005 (AVBPA2005)*, ser. Lecture Notes in Computer Science, T. Kanade, A. Jain, and N. K. Ratha, Eds., vol. 3546. Springer, 2005, pp. 279–288.

[2] R. M. Bolle, N. K. Ratha, and S. Pankanti, "Error analysis of pattern recognition systems – the subsets bootstrap," *Computer Vision and Image Understanding*, vol. 93, pp. 1–33, 2004.

[3] T. C. Christofides and P. M. Mavrikiou, "Central limit theorem for dependent multidimensionally indexed random variables," *Statistics and Probability Letters*, vol. 63, no. 1, pp. 67–78, 2003.

[4] N. Cressie, *Statistics for Spatial Data*. New York: John Wiley and Sons, 1993.

[5] J. Eng, "Sample size estimation: A glimpse beyond simple formulas," *Radiology*, vol. 230, pp. 606–612, 2004.

[6] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc., 2003.

[7] G. Givens, J. R. Beveridge, B. A. Draper, and D. Bolme, "Statistical assessment of subject factors in the pca recognition of human faces," in *IEEE CVPR 2003 Workshop on Statistical Analysis in Computer Vision Workshop*, 2003.

[8] S. N. Lahiri, *Resampling Methods for Dependent Data*. New York: Springer, 2003.

[9] Y. Ma, M. Schuckers, and B. Cukic, "Guidelines for appropriate use of simulated data for bio-authentication research," in *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, V. Bhagavatula and V. Govindaraju, Eds., 2005, pp. 251–256.

[10] T. Mansfield and J. L. Wayman, "Best practices in testing and reporting performance of biometric devices," on the web at www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf, 2002.

[11] S. Mitra, M. Savvides, and B. V. K. V. Kumar, "Biometric authentication based on novel frequency domain facial asymmetry measures," in *Proceedings of Defense and Security Symposium on Biometric Technology for Human Identification II*, A. K. Jain and N. Ratha, Eds., 2005, pp. 409–417.

[12] D. F. Moore, "Asymptotic properties of moment estimators for overdispersed counts and proportions," *Biometrika*, vol. 73, no. 3, pp. 583–588, December 1986.

[13] N. Poh and S. Bengio, "Database, protocol and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognition Journal*, 2005.

[14] N. Poh, A. Martin, and S. Bengio, "Performance generalization in biometric authentication using joint user-specific and sample bootstraps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007.

[15] M. S. Ridout, C. G. B. Demétrio, and D. Firth, "Estimating intraclass correlation for binary data," *Biometrics*, vol. 55, pp. 137–148, 1999.

[16] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, September 2003.

[17] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer Publishers, 2006.

[18] M. E. Schuckers, "Estimation and sample size calculations for correlated binary error rates of biometric identification rates," in *Proceedings of the American Statistical Association: Biometrics Section [CD-ROM]*. Alexandria, VA: American Statistical Association, 2003.

[19] ——, "Using the beta-binomial distribution to assess performance of a biometric identification device," *International Journal of Image and Graphics*, vol. 3, no. 3, pp. 523–529, 2003.

[20] M. E. Schuckers and N. L. Lopez, "Template aging: A study of the NIST biometric score set release 1," in *Proceedings of the Biometric Symposium*, 2005, pp. 66–67.

[21] A. R. Solow, "Bootstrapping correlated data," *Mathematical Geology*, vol. 17, no. 7, pp. 769–775, 1985.

[22] L. Tang, "Undercoverage of wavelet-based resampling confidence intervals and a parameteric spatial bootstrap," Ph.D. in Statistics, Southern Methodist University, Dallas, TX, 2005.

[23] J. L. Wayman, "National biometrics test center, collected works 1997-2000," www.engr.sjsu.edu/biometrics/nbtccw.pdf, 2000.

[24] D. A. Williams, "The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity," *Biometrics*, vol. 31, pp. 949–952, 1975.