THEORETICAL STATISTICAL CORRELATION FOR BIOMETRIC IDENTIFICATION PERFORMANCE

Michael E. Schuckers

schuckers@stlawu.edu
St. Lawrence University, Canton, NY, USA
and Center for Identification Technology Research (CITeR)

ABSTRACT

Measurement and evaluation of biometric device performance is critical to end users and consumers of these devices. In this paper we present explicit theoretical correlation models when the biometric matching process is stationary that can be used to derive variance estimates of biometric performance metrics. We focus in this paper on the failure to enroll, the failure to acquire, the false match rate and the false non-match rate. We further present unified notation that makes our correlation frameworks possible.

Index Terms— Statistics, Covariance matrices, Image Classification, Moment methods, Random Variables

1. INTRODUCTION AND NOTATION

An understanding of how a biometric device performs is essential for decisions regarding Many methods for carrying out an evaluation of a biometric identification have been proposed. All of these evaluations utilize general statistical methods and, in particular, statistical methods that estimate variances. The evaluation of the performance of a biometric device is a critical tool for users and decision makers. This work builds on previous attempts to describe the biometric process mathematically by Mansfield and Wayman [1] and Ma *et al* [2]. We extend those works to give explicit (and appropriate) correlation structures that allow for the calculation of variance estimates of the typical biometrics performance metrics.

The biometric matching process can be thought of in the following steps with accompanying notation:

1. Enrollment, (E_i)

2. Acquisition, (A_{ij})

3. Feature Extraction, $(S_{ik}, \mathbf{X}_{ik})$

4. Matching, $(Y_{ik,i'k'})$

5. Decision, $(D_{ik,i'k'})$.

(We will assume here a $1 - 1$ matching process but it is straightforward to generalize the work here to $1 - n$ matching.) The first step in this process is the enrollment of individuals. As part of the enrollment we measure whether or not the ith individual is capable of enrolling. We call this binary random variable E_i for $i \in \mathcal{E}$ where $\mathcal{E}$ is the set of all individuals who attempt to enroll and n_E is the number of individuals who attempted to enroll.

$$E_i = \begin{cases} 1 & \text{if individual } i \text{ is unable to enroll} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Then the failure to enroll (FTE) rate is

$$FTE = \frac{\sum_{i \in \mathcal{E}} E_i}{n_E}. \quad (2)$$

It is certainly the case that additional measurements are collected, e.g. biometric samples and templates, as part of the enrollment process. We will discuss these below as part of the matching process.

The second step in the process is the acquisition of images in non-enrollment attempts. Since it is possible for multiple acquisition attempts to occur per individual, we use A_{ij} to denote the jth binary outcome of the acquisition attempt by the ith individual where $i \in \mathcal{A}$ and $j = 1, \ldots, a_i$. We will assume A_{ij} is an indicator of failure to acquire. Then

$$A_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ acquisition attempt by} \\ & \text{individual } i \text{ is not acquired} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Thus, a_i is the total number of attempts for the ith individual and $\mathcal{A}$ represents all individuals who attempt to have their biometric image collected. Then the failure to acquire (FTA) rate is

$$FTA = \frac{\sum_{i \in \mathcal{A}} \sum_{j=1}^{a_i} A_{ij}}{\sum_{i \in \mathcal{A}} a_i} \quad (4)$$

We are not assuming that $\mathcal{A} = \mathcal{E}$.

We next move to the biometric sample that is collected. We will denote biometric samples by S_{ik} where we are only interested in samples taken when $A_{ik} = 0$. From each sample, S_{ik}, image processing is done and features are extracted. Since it is nearly always the case that multiple features are extracted we denote the features that are extracted on the kth successful acquisition for the ith individual by $\mathbf{X}_{ik}$. Having extracted these features, biometric systems then compare features extracted to those taken during the enrollment process. This matching process yields a match score for each pair of images. Thus, we need to denote these scores by $Y_{ik,i'k'}$ which is the match score for a comparison of the kth image collected on the ith individual to the k'^{th} collected image for the i'^{th} individual. This notation allows for consideration of both genuine ($i = i'$) and imposter ($i \neq i'$) match scores with a single notation. The imposter and genuine score distributions are build by combining these comparison scores when $i \neq i'$ and when $i = i'$ respectively.

The final step of the matching process is the decision whether to accept or reject an individual based upon their presented biometric. Without loss of generality we can say that we will reject a given decision if $Y_{ik,i'k'}$ is above some threshold, τ. We let $D_{ii'\ell}$ be a binary decision based upon the match score $Y_{ik,i'k'}$. $D_{ii'\ell}$ will be a binary decision for the ℓth decision from comparing images from individual i and i'. We use this notation since it is more general and simpler than the notation for the match scores. Note that the $D_{ii'\ell}$ will be 1 if a mistake is made in matching and 0 otherwise. More formally let

$$
D_{ii'\ell} = \begin{cases} 1 & \text{if} \quad i = i', Y_{ik,ik'} > \tau \\ 0 & \text{if} \quad i = i', Y_{ik,ik'} \leq \tau \\ 0 & \text{if} \quad i \neq i', Y_{ik,i'k'} > \tau \\ 1 & \text{if} \quad i \neq i', Y_{ik,i'k'} \leq \tau \end{cases} \tag{5}
$$

The false match rate (FMR) and the false non-match rate (FNMR) are then

$$
FMR = \frac{\sum_i \sum_{i' \neq i} \sum_\ell D_{ii'\ell}}{\sum_i \sum_{i' \neq i} n_{ii'}} \tag{6}
$$

and

$$
FNMR = \frac{\sum_i \sum_\ell D_{ii\ell}}{\sum_i n_{ii}}, \tag{7}
$$

respectively. In the above equations, $n_{ii'}$ represents the number of comparisons made between individuals i and i' and n_{ii} is the number of comparisons made between individual i and themselves. If $i = i'$ then the comparisons are genuine and if $i \neq i'$ then the comparisons are imposter.

2. STATISTICAL BACKGROUND

All of the statistical methods used to evaluate classification and matching performance - FTE, FTA, FMR, FNMR - are all averages and, therefore, they are linear combinations. We

simply note here that a proportion is an average of 0's and 1's. Clearly then FTE, FTA, FMR and FNMR are averages. Any discussion or estimation of these quantities utilizes statistical methods for linear combinations. Letting R_v, $v = 1, \ldots, V$ be a random variable and let a_v's be known constants. Then $R^\star = \sum_{v=1}^V a_v R_V$ is our linear combination of interest. Note that if $a_v = V^{-1}$ then $R^\star$ is an average. We note that the first two moments of $R^\star$ are

$$
E[R^\star] = \sum_{v=1}^V a_v E[R_v] \tag{8}
$$

and

$$
\begin{aligned}
V[R^\star] &= \sum_{v=1}^V \sum_{w=1}^V a_v a_w Cov(R_v, R_w) \\
&= \sum_{v=1}^V a_v^2 V[R_v] \\
&\quad + 2\sum_{v=1}^V \sum_{w>v}^V a_v a_w Cov(R_v, R_w) \tag{9}
\end{aligned}
$$

Here we have used the standard notation that $E[\ ]$, $V[\ ]$, $Cov(\ ,\ )$ represent the expectation, variance and covariance, respectively.

If we can assume that the process mean is constant, i.e. $E[R_v] = \mu_R$, then Equation 8 becomes

$$
E[R^\star] = \mu_R \left(\sum_{v=1}^V a_v \right). \tag{10}
$$

If we can further assume that the variances for the process described by the random variables are equal - $V[R_v] = \sigma_R^2$ then we can rewrite Equation 9 as

$$
\sum_{v=1}^V a_v^2 \sigma_R^2 + 2\sum_{v=1}^V \sum_{w>v}^V a_v a_w \sigma_R^2 Corr(R_v, R_w). \tag{11}
$$

Here we denote $Corr(\ ,\ )$ as the correlation of two random variables. Note that the assumption of constant mean and variance is often referred to as wide-sense stationarity (WSS) or covariance stationarity. One concern that is raised about the assumption of WSS for biometric classification is the issue of the biometric menagerie or biometric zoo first proposed by Doddington *et al* [3]. There has been much work in recent years on the issue of a biometric menagerie and whether or not it truly exists. For example, see [4], [5], [6], or [7] for recent works in this area. Setting aside the issue of the existence of the menagerie, it is possible to maintain a biometric process whose mean and variance are constant but whose *correlation* structure can accommodate the menagerie. We will present such structures below.

We motivated the above derivations by pointing out the need for understanding variances of linear combinations. There are two other important reasons for focusing on the

correlation and the covariance structure. The first of these is the need to consider asymptotic behavior of random variables and their linear combinations. We note that central limit theorems are most often associated with averages; however, central limit theory is more generally applicable to linear combinations. See, for example, Jacod and Shiryaev [8], for additional results in this area. The second reason for the importance of the correlation structure is the need for appropriate resampling methods. Several resampling methods have been proposed in the biometrics literature including Bolle *et al* [9] and Poh *et al* [10].

3. BIOMETRIC CORRELATION STRUCTURES

In this section we present correlations for the performance metrics of FTE, FTA, FMR and FNMR. As mentioned in the previous section, understanding of correlation is crucial to understanding of variability estimates and, hence, statistical inference. Many of the commonly used statistical tools - sample size calculations, confidence intervals, generalized linear models - that are potentially available to biometrics require the ability to articulate the variance structure of a linear combinations. It is important to remember here that these correlation structures depend on WSS being constant for the process under consideration. In addition, we note that the correlations here yield variability beyond that already due to independent events. We begin here with the correlation for FTE. This correlation structure is straightforward since only individuals are involved.

3.1. FTE Correlation

Above we introduced the notation for FTE's. Since a single individual is involved in each enrollment attempt and for each individual there is only a single enrollment decision, we can assume that there is no correlation between enrollment decisions. Formally this is

$$Corr(E_i, E_{i'}) = \begin{cases} 1 & \text{if} \quad i = i' \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

The implications from this correlation structure is that the observed enrollment decisions are uncorrelated and, thus, we can use methods based upon independence for appropriate inference.

3.2. FTA Correlation

The correlation structure for FTA is necessarily more complicated since multiple attempts are possible for each individual. We will assume here that acquisition attempts will be correlated if they involve the same individual. If different individuals are involved then we assume that acquisition attempts are

uncorrelated. Mathematically, we write.

$$Corr(A_{ij}, A_{i'j'}) = \begin{cases} 1 & \text{if} \quad i = i', j = j' \\ \psi & \text{if} \quad i = i', j \neq j' \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

In this case ψ measures the degree of similarity that exists between acquisition decisions of the same individual. It is often referred to as the intra-individual or intra-class correlation.

3.3. FNMR Correlation

FNMR has a structure that is similar to FTA because only a single individual is involved though there are potentially numerous decisions based upon that individual. The correlation structure we propose explicitly here has been formulated implicitly by at least two authors Poh *et al* [10] – the user-specific bootstrap – and Schuckers [11] – the beta-binomial approach. The correlation structure for the false non-match decisions, the $D_{ii\ell}$'s, is

$$Corr(D_{ii\ell}, D_{i'i'\ell'}) = \begin{cases} 1 & \text{if} \quad i = i', \ell = \ell' \\ \rho & \text{if} \quad i = i', \ell \neq \ell' \\ 0 & \quad otherwise \end{cases} \quad (14)$$

For this correlation, ρ is a measure of how similar each individual is to themselves.

3.4. FMR Correlation

False match rate correlation is necessarily the most complicated because each decision involves two individuals and thus the correlation depends on four individuals. Notationally we try to group possible correlations. Below η represents the correlation between two comparisons made on the same pair of individuals. The ω's represent the case where one individual from each pair is shared and the ξ's represent the correlation between individuals when the order is asymmetric. If the matcher is symmetric then this model will simplify to simply η and the ω's. The degree to which someone is a goat or sheep is measured by the ω's. The the correlation structure for the imposter decisions, the $D_{ii'\ell}$'s where $i \neq i'$ is given by

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

$$(15)$$

The correlation described above is a complex and sophisticated one. A similar sort of correlation structure was first

proposed by Bickel [12]. Our correlation structure generalizes Bickel's approach. It also important to note that resampling or nonparametric methods currently in use are not sufficient to accurately assess the variability in the FMR based upon this .

4. DISCUSSION

An understanding of correlation structure is necessary for appropriate use of statistical methods. The evaluation of biometric identification performance depends on these statistical methods. This paper has proposed explicit correlation structures to further statistical research on biometric device performance. The metrics for which we have proposed a correlation structure ar e the failure to enroll, the failure to acquire, the false match rate and the false non-match rate. The correlation structures here are dependent upon each process having constant mean and variance. Changes in the process can be modelled by a generalization of the methods considered here. It is hoped that the correlation frameworks described here will lead to the development of improved statistical methods now that these correlations have been implicitly described. In particular, the use of statistical theory for linear combinations should lead to the development of appropriate variance estimates of biometric performance metrics. Finally we note that these correlations allow for the possibility of the biometrics menagerie.

5. REFERENCES

[1] Tony Mansfield and James 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.

[2] Yan Ma, Michael Schuckers, and Bojan Cukic, "Guidelines for appropriate use of simulated data for bioauthentication research," in *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, Vijayakumar Bhagavatula and Venu Govindaraju, Eds., 2005, pp. 251–256.

[3] George R. Doddington, Walter Liggett, Alvin F. Martin, Mark A. Przybocki, and Douglas A. Reynolds, "Sheep, Goats, Lambs and Wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *ICSLP*, 1998, pp. 608–611.

[4] Austin Hicklin, Craig Watson, and Brad Ulery, "The myth of goats: How many people have fingerprints that are hard to match," Tech. Rep. NISTIR 7271, National Institutes for Standards and Technology, 2005.

[5] M. Wittman, P. Davis, and P.J. Flynn, "Empirical studies of the existence of the biometric menagerie in the frgc 2.0 color image corpus," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006.

[6] Norman Poh, Samy Bengio, and Arun Ross, "Revisiting doddingtons zoo: A systematic method to assess user-dependent variabilities," IDIAPRR 06-04, IDIAP Research Institute, , 2006, .

[7] Neil Yager and Ted Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *2007 IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'07)*, 2007.

[8] Jean Jacod and Albert N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer, 2002.

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

[10] Norman Poh, Alvin Martin, and Sammy Bengio, "Performance generalization in biometric authentication using joint user-specific and sample bootstraps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear.

[11] Michael E. Schuckers, Anne Hawley, Katie Livingstone, Nona Mramba, and Collen J. Knickerbocker, "A comparison of statistical methods for evaluating matching performance of a biometric identification device- a preliminary report," in *Biometric Technology for Human Identification*, Anil K. Jain and Nalini K. Ratha, Eds., 2004, vol. 5404 of *SPIE Proceedings Series*, pp. 144–155.

[12] James L. Wayman, "Confidence interval and test size estimation for biometric data," in *National Biometrics Test Center, Collected Works 1997-2000*, pp. 89–99. www.engr.sjsu.edu/biometrics/nbtccw.pdf, 2000.