Theoretical Statistical Correlation for Biometric Identification Performance

Michael E. Schuckers

schuckers@stlawu.edu

Overview

Problem Statement: There is a need for appropriate statistical methods for estimation of the performance measures for biometric identification devices.

Contribution: This paper lays out **general** correlation structure that allows for estimation, specifically standard errors, for FTE, FTA, FMR, FNMR.

Biometric Process and Notation

- Enrollment, (E_i)

Biometric Process and Notation

- Enrollment, (E_i)
- Acquisition, (A_{ij})

Biometric Process and Notation

- Enrollment, (E_i)
- Acquisition, (A_{ij})
- Feature Extraction, $(S_{ik}, \mathbf{X}_{ik})$

Biometric Process and Notation

- Enrollment, (E_i)
- Acquisition, (A_{ij})
- Feature Extraction, $(S_{ik}, \mathbf{x}_{ik})$
- Matching, $(Y_{ik,i'k'})$

Biometric Process and Notation

- Enrollment, (E_i)
- Acquisition, (A_{ij})
- Feature Extraction, $(S_{ik}, \mathbf{X}_{ik})$
- Matching, $(Y_{ik,i'k'})$
- Decision, $(D_{ik,i'k'})$

Failure to Enrol (FTE)

Notation

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

Failure to enroll (FTE) rate is

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

where N_E is the total number of individuals who attempted to enrol in the system.

Failure to Acquire (FTA)

Notation

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

where

- a_i is the total number of attempts for the i^{th} individual
- $\mathcal{A}$ represents all individuals who attempt to have their biometric image collected.

Failure to acquire (FTA) rate is

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

False Match Rate (FMR)

Notation

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

False match rate (FMR) is then

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

where $n_{ii'}$ is the number of decisions on the ordered pair of individuals i and i'.

False Non-Match Rate (FNMR)

Notation

(7) $$D_{ii'\ell} = \begin{cases} 1 & \text{if} \quad i = i', Y_{ik,ik'} > \tau \\ 0 & \text{if} \quad i = i', Y_{ik,ik'} \leq \tau \end{cases}$$

False non-match rate (FNMR) is then

(8) $$FNMR = \frac{\sum_i \sum_\ell D_{ii\ell}}{\sum_i n_{ii}},$$

where n_{ii} is the number of decisions on individual i.

Statistical Methods

Estimated FTE, FTA, FMR, FNMR are all linear combinations of binary RV's

- Linear Combination $R^\star = \sum_v b_v R_v$
- Expectation $E[R^\star] = \sum_{v=1}^V a_v E[R_v]$
- Variance $V[R^\star] = \sum_{v=1}^V \sum_{w=1}^V a_v a_w Cov(R_v, R_w)$
- If variance is constant then $V[R^\star] = \sigma_R^2 [\sum_{v=1}^V a_v^2 + 2 \sum_{v=1}^V \sum_{w>v}^V a_v a_w Corr(R_v, R_w)]$.

Assumptions

For the rest of this paper, we assume that FTE, FTA, FMR, and FNMR

- Constant mean
- Constant variance/covariance/correlation

Get simple right move to more complicated

- E's, A's, D's are binary (variance $\propto$ mean(1-mean))
- Work from correlation

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1
- **B** 0 0 0 0 0 0 0 0 1 0

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1
- **B** 0 0 0 0 0 0 0 0 1 0
- **C** 1 1 1 1 0 1 1 1 1 1

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1
- **B** 0 0 0 0 0 0 0 0 1 0
- **C** 1 1 1 1 0 1 1 1 1 1
- **D** 1 0 1 1 0 0 1 0 0 0

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1
- **B** 0 0 0 0 0 0 0 0 1 0
- **C** 1 1 1 1 0 1 1 1 1 1
- **D** 1 0 1 1 0 0 1 0 0 0

Answer: None, all are binomial $\rightarrow Corr = 0$

Binary Correlations

Which of the following rows of binary sequences is correlated?

- **A** 1 0 1 1 1 0 0 0 0 1
- **B** 0 0 0 0 0 0 0 0 1 0
- **C** 1 1 1 1 0 1 1 1 1 1
- **D** 1 0 1 1 0 0 1 0 0 0

Answer: None, all are binomial $\rightarrow Corr = 0$

Correlation Structure: FTE

Uncorrelated Between Individuals Structure

(9)
$$Corr(E_i, E_{i'}) = \begin{cases} 1 & \text{if} \quad i = i' \\ 0 & \quad\quad \text{otherwise.} \end{cases}$$

Then estimated variance is

(10)
$$V[F\hat{T}E] = \frac{F\hat{T}E(1 - F\hat{T}E)}{N_E}$$

Correlation Structure: FTA

Intra-individual Correlation

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

Then estimated variance is

(12) $$V[F\hat{T}A] = \frac{F\hat{T}A(1 - F\hat{T}A)}{N_A^2} \left[N_A + \psi \sum_{i=1}^{n} a_i(a_i - 1) \right]$$

where $N_A = \sum a_i$.

Correlation Structure: FNMR

Intra-individual Correlation

(13)
$$Corr(D_{ii\ell}, D_{i'i'\ell'}) = \begin{cases} 1 & \text{if} \quad i = i', \ell = \ell' \\ \rho & \text{if} \quad i = i', \ell \neq \ell' \\ 0 & \quad otherwise \end{cases}$$

$$V[FN\hat{M}R] = \frac{FN\hat{M}R(1 - FN\hat{M}R)}{N_G^2}\left[N_G + \rho \sum_{i=1}^{n} n_{ii}(n_{ii} - 1)\right]$$

(14)

where $N_G = \sum n_{ii}$.

Correlation Structure: FMR

Shared elements of a pair correlation OR
ALL HELL BREAKS LOOSE

$$Corr(D_{ik\ell}, D_{i'k'\ell'}) = \begin{cases} 1 & \text{if} \quad i = i', k = k', \ell = \ell' \\ \eta & \text{if} \quad i = i', k = k', \ell \neq \ell' \\ \omega_1 & \text{if} \quad i = i', k \neq k', i \neq k, i \neq k' \\ \omega_2 & \text{if} \quad i \neq i', k = k', i \neq k, i \neq k' \\ \omega_3 & \text{if} \quad i = k', i' \neq k, i \neq i', i \neq k \\ \omega_3 & \text{if} \quad i' = k, i \neq k', i' \neq i, k \neq k' \\ \xi_1 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell = \ell' \\ \xi_2 & \text{if} \quad i = k', k = i', i \neq i', k \neq k', \ell \neq \ell' \\ 0 & \quad\quad \text{otherwise} \end{cases} \tag{15}$$

Correlation Structure: FMR (Variance)

$$
\begin{aligned}
V[F\hat{N}MR] \quad = \quad & \frac{F\hat{N}MR_I(1 - F\hat{N}MR)}{N_I^2}\Bigg[N_I + \eta \sum_{i=1} \sum_{\substack{k=1\\k\neq i}} n_{ik}(n_{ik} - 1) \\
+ \quad & \omega_1 \sum_{i=1} \sum_{\substack{k=1\\k\neq i}} n_{ik} \left(\sum_{\substack{k'=1\\k'\neq i,k'\neq k}} n_{ik'} \right) + \omega_2 \sum_{i=1} \sum_{\substack{k=1\\k\neq i}} n_{ik} \left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}} n_{i'k} \right) \\
+ \quad & \omega_3 \sum_{i=1} \sum_{\substack{k=1\\k\neq i}} n_{ik} \left(\sum_{\substack{i'=1\\i'\neq i,i'\neq k}} n_{i'i} + \sum_{\substack{k'=1\\k'\neq i,k'\neq k}} n_{kk'} \right) \\
+ \quad & \xi_1 \sum_{i=1} \sum_{\substack{k=1\\k\neq i}} n_{ki} + \xi_2 \sum_{i=1}^{n} \sum_{\substack{k=1\\k\neq i}} n_{ki}(n_{ki} - 1)\Bigg] \qquad (16)
\end{aligned}
$$

where $N_I = \sum_i \sum_{k\neq i} n_{ik}$.

Application:

FMR from Ross and Jain (2003)

Modality	τ	$F\hat{M}R$	$\hat{\eta}$	$\hat{\omega_1}$	$\hat{\omega_2}$	$\hat{\omega_3}$	$\hat{\xi_1}$	$\hat{\xi_2}$	$\sqrt{\hat{V}[\hat{\pi_I}]}$
Hand	100	0.203	0.021	0.021	0.004	*0.000	*0.000	0.006	0.010
FP	20	0.060	0.005	0.0003	0.002	*0.000	0.019	0.014	0.003
Face	50	0.044	*0.000	0.001	0.004	*0.000	0.004	*0.000	0.003

τ here is the threshold

* indicates truncated correlation at zero

$N_I = 12250$

Overdispersion here: $2.64, 1.25, 1.45$ respectively relative to binomial

Application:

FNMR from Ross and Jain (2003)

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

τ here is the threshold

* indicates truncated correlation at zero

$N_G = 500$

Overdispersion here: $1.16, 1.00, 1.04$ respectively relative to binomial

Structure and Methods

Metric	Parametric model	Non-parametric model
FTE	Binomial	IID
FTA	Beta-binomial	User-specific Bootstrap
FNMR	Beta-binomial	User-specific Bootstrap
FMR	N/A	N/A

User-specific Bootstrap due to Poh *et al* 2007

FMR model generalization of implicit structure by Bickel

Central Limit Theorems apply to all cases.

Discussion

- Theoretical Formulation of Correlation

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent
- Basis for More Complicated Correlation Structures

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent
- Basis for More Complicated Correlation Structures
- Appropriate Statistical CI's and Hypo Tests

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent
- Basis for More Complicated Correlation Structures
- Appropriate Statistical CI's and Hypo Tests
- Parametric approach used for Sample Size Calc's

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent
- Basis for More Complicated Correlation Structures
- Appropriate Statistical CI's and Hypo Tests
- Parametric approach used for Sample Size Calc's
- Needed for GLM's, Covariate approaches

Discussion

- Theoretical Formulation of Correlation
- All correlations are threshold dependent
- Basis for More Complicated Correlation Structures
- Appropriate Statistical CI's and Hypo Tests
- Parametric approach used for Sample Size Calc's
- Needed for GLM's, Covariate approaches
- Mean Transaction Time??

Thank You

Questions?

schuckers@stlawu.edu