Distribution-Free Statistical Methods for Biometric Performance Evaluation

Michael E. Schuckers

St. Lawrence University

Center for Identification Technology Research (CITeR) [a]

schuckers@stlawu.edu

[a]This work is supported by NSF grants CNS-0325640 and CNS-0520990 which is cooperatively funded by the National Science Foundation and the United States Department of Homeland Security

Overview

Goal: Develop inferential statistical methods for FTA, FTE, FMR, FNMR without worrying about *sampling* distribution of those statistics.

Note: We *can* use a central limit theorem for all of these .... but that is another talk (SPIE 2010).

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

where N_E is the total number of individuals who attempted to enrol in the system and $\mathcal{E}$ is the collection of individuals that tried to enrol.

Failure to Acquire (FTA)

Notation

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

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

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

False Non-Match Rate (FNMR)

Let $Y_{iw,kw'}$ be the match score for comparing the w^{th} presentation from the i^{th} individual to the w'^{th} presentation from the k^{th} individual.
Notation

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

False non-match rate (FNMR) is then

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

where n_{ii} is the number of decisions on individual i.
We assume some ordering of decisions $\ell = 1, \ldots, n_{ii}$.

False Match Rate (FMR)

Let $Y_{iw,kw'}$ be the match score for comparing the w^{th} presentation from the i^{th} individual to the w'^{th} presentation from the k^{th} individual.
Notation

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

False match rate (FMR) is then

(8)
$$\widehat{FMR} = \frac{\sum_i \sum_{k \neq i} \sum_\ell D_{ik\ell}}{\sum_i \sum_{k \neq i} n_{ik}}$$

where n_{ik} is the number of decisions on the ordered pair of individuals i and k.
We assume some ordering of decisions $\ell = 1, \ldots, n_{ik}$.

IID Bootstrap Review

Suppose we have n decisions $X_1, X_2, \ldots, X_n$

- Calculate statistic, say $\hat{\Phi} = \hat{\Phi}(X_1, X_2, \ldots, X_n)$.
- Sample n values from $\{1, 2, \ldots, n\}$ with replacement and call those values $b_1, b_2, \ldots, b_n$.
- Calculate and store $\hat{\Phi}^* = \hat{\Phi}(X_{b_1}, X_{b_2}, \ldots, X_{b_n})$
- Repeat the previous two steps some large number of times
- Use distribution of the $\hat{\Phi}^*$ to approximate the sampling distribution of $\hat{\Phi}$.

Only good if data are uncorrelated (FTE) but extensions and improvements exist.... as we shall see.

Correlation Structure: FTE

Uncorrelated Between Individuals Structure

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

Then estimated variance is

(10) $$V[\widehat{FTE}] = \frac{\widehat{FTE}(1 - \widehat{FTE})}{N_E}$$

where N_E is the total number of enrolment attempts.

Correlation Structure: FTA

Intra-individual Correlation

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

Then estimated variance is

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

where $N_A = \sum a_i$ and a_i is the number of decisions from the i^{th} individual.

Correlation Structure: FNMR

Intra-individual Correlation

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

$$V[\widehat{FNMR}] = \frac{\widehat{FNMR}(1 - \widehat{FNMR})}{N_G^2} \left[N_G + \rho \sum_{i=1}^{n} n_{ii}(n_{ii} - 1) \right]$$
(14)

where $N_G = \sum n_{ii}$ and $n_i i$ is the number of decisions from the i^{th} individual.

Subsets Bootstrap: FNMR and FTA

- Calculate statistic, say $\widehat{FTA} = \frac{\sum_{i\in\mathcal{A}}\sum_{j=1}^{a_i} A_{ij}}{\sum_{i\in\mathcal{A}} a_i}$ where $n = \mid \mathcal{A} \mid$ is the number of individuals and a_i is the number of decisions on i^{th} individual.
- Sample n values from $\{1, 2, \ldots, n\}$ with replacement and call those values $b_1, b_2, \ldots, b_n$.
- Calculate and store $\widehat{FTA}^* = \frac{\sum_{i\in\mathcal{A}}\sum_{j=1}^{a_{b_i}} A_{b_i j}}{\sum_{i\in\mathcal{A}} a_{b_i}}$
- Repeat the previous two steps some large number of times
- Use distribution of the $\widehat{FTA}^*$ to approximate the sampling distribution of $\widehat{FTA}$.

Correlation Structure: FMR

Two-instance 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)

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

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

Two-instance Bootstrap

- Resample $n_{\mathcal{P}}$ individuals from the list of all $n_{\mathcal{P}}$ individuals in the probe and call those individuals $b_1, b_2, \ldots, b_{n_{\mathcal{P}}}$.

Two-instance Bootstrap

- Resample $n_{\mathcal{P}}$ individuals from the list of all $n_{\mathcal{P}}$ individuals in the probe and call those individuals $b_1, b_2, \ldots, b_{n_{\mathcal{P}}}$.
- For the b_i^{th} individual selected in the previous step we then resample from the $n_{\mathcal{G}}$ individuals in the gallery and call those individuals $h_{b_i,k}$'s for $k = 1, \ldots, n_{\mathcal{G}}$.

Two-instance Bootstrap

- Resample $n_{\mathcal{P}}$ individuals from the list of all $n_{\mathcal{P}}$ individuals in the probe and call those individuals $b_1, b_2, \ldots, b_{n_{\mathcal{P}}}$.
- For the b_i^{th} individual selected in the previous step we then resample from the $n_{\mathcal{G}}$ individuals in the gallery and call those individuals $h_{b_i,k}$'s for $k = 1, \ldots, n_{\mathcal{G}}$.
- Take all of the decisions $D_{b_i h_{b_i,k} 1}, \ldots, D_{b_i h_{b_i,k} m_{b_i h_{b_i,k}}}$ from the resampled pairs of individuals, $(b_i, h_{b_i,k})$ found in the two previous steps.

Two-instance Bootstrap

- Resample $n_{\mathcal{P}}$ individuals from the list of all $n_{\mathcal{P}}$ individuals in the probe and call those individuals $b_1, b_2, \ldots, b_{n_{\mathcal{P}}}$.
- For the b_i^{th} individual selected in the previous step we then resample from the $n_{\mathcal{G}}$ individuals in the gallery and call those individuals $h_{b_i,k}$'s for $k = 1, \ldots, n_{\mathcal{G}}$.
- Take all of the decisions $D_{b_i h_{b_i,k} 1}, \ldots, D_{b_i h_{b_i,k} m_{b_i h_{b_i,k}}}$ from the resampled pairs of individuals, $(b_i, h_{b_i,k})$ found in the two previous steps.
- The replicated value for $\widehat{FMR}$ is then given by

(20) $$\widehat{FMR}^* = \frac{\sum_{i=1}^{n_{\mathcal{P}}} \sum_{k=1}^{n_{\mathcal{G}}} \sum_{\ell=1}^{m_{b_i h_{b_i,k}}} D_{b_i h_{b_i,k} \ell}}{\sum_{i=1}^{n_{\mathcal{P}}} \sum_{k=1}^{n_{\mathcal{G}}} m_{b_i h_{b_i,k}}}.$$

Structure and Methods

Metric	Distribution model	Distribution-Free model
FTE	Binomial	*IID*
FTA	Beta-binomial	Subset Bootstrap
FNMR	Beta-binomial	Subset Bootstrap
FMR	??	Two-instance Bootstrap

Subset Bootstrap due to Bolle *et al* (2003)

FMR model generalization of implicit structure by Bickel

Two-instance Bootstrap Schuckers(2010)

Central Limit Theorems apply to all cases.

Variance estimation: FMR

XM2VTS Database

Matcher	Threshold	$\hat{\nu}$	$s_{\hat{\nu}}$	$s_{\hat{\nu},boot}$
(FH,MLP)	0.0	0.0038	0.0007	0.0009
(DCTs,GMM)	0.0	0.0582	0.0067	0.0066
(DCTb,GMM)	0.2	0.0041	0.0007	0.0008
(DCTs,MLP)	-0.8	0.1057	0.0086	0.0091
(DCTb,MLP)	-0.5	0.0580	0.0072	0.0074
(LFCC,GMM)	3.0	0.0142	0.0039	0.0038
(PAC,GMM)	2.0	0.0570	0.0090	0.0090
(SSC,GMM)	1.0	0.0692	0.0105	0.0099

$s_{\hat{\nu}}$ is the standard error from Schuckers (2009) and $s_{\hat{\nu},boot}$ is the bootstrapped standard error.

Great! So What?

Great! So What?

- There's that significance thing.

Great! So What?

- There's that significance thing.
- And that sampling variability thing.

Great! So What?

- There's that significance thing.
- And that sampling variability thing.
- Just cause one rate/curve is 'better' than another doesn't make it *significantly* better.

Great! So What?

- There's that significance thing.
- And that sampling variability thing.
- Just cause one rate/curve is 'better' than another doesn't make it *significantly* better.
- Ask about significance, confidence intervals, p-values.

Great! So What?

- There's that significance thing.
- And that sampling variability thing.
- Just cause one rate/curve is ‘better’ than another doesn't make it *significantly* better.
- Ask about significance, confidence intervals, p-values.
- Tools now in place.

Beyond a Single Metric

Move to methods for:

- Comparing two metrics

Beyond a Single Metric

Move to methods for:

- Comparing two metrics
- Comparing three or more metrics

Beyond a Single Metric

Move to methods for:

- Comparing two metrics
- Comparing three or more metrics
- Considering paired data collections

Beyond a Single Metric

Move to methods for:

- Comparing two metrics
- Comparing three or more metrics
- Considering paired data collections
- Considering independent data collections

Example:FNMR

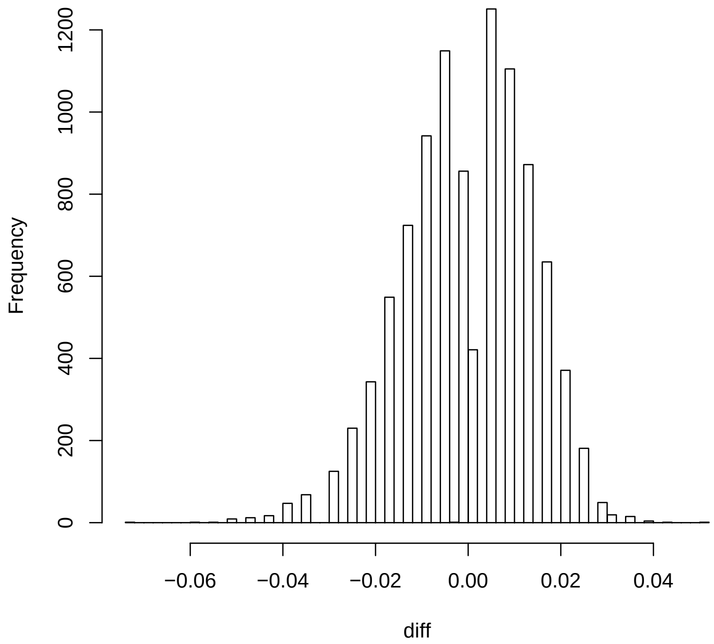
We are considering difference between two FNMR's for the same matcher but on different groups. Each group 312 decisions from 26 individuals.

$\widehat{FNMR}_1 = \mathbf{0.0086}$ and
$\widehat{FNMR}_2 = \mathbf{0.0214}$.

Make 95% confidence interval for the difference in the FNMR's

Example:FNMR

Bootstrap each FNMR and take difference at each bootstrap



The 95% CI for $FNMR_1 - FNMR_2$ is $(\mathbf{-0.0384, 0.0171})$ from 10000 bootstrap replicates.

Example:FMR

Comparing equality of FMR for 5 matchers from BANCA database based upon 312 decisions.

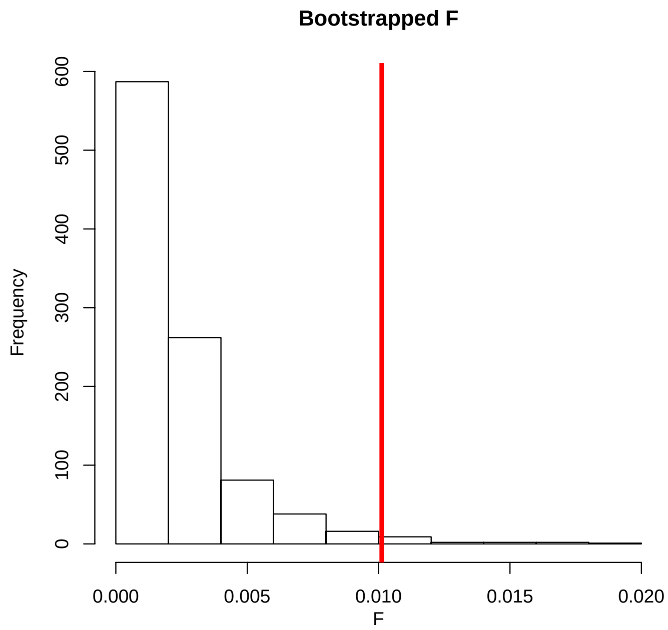
$\widehat{FMR}$'s are $\mathbf{0.0769}$, $\mathbf{0.0449}$, $\mathbf{0.0513}$, $\mathbf{0.0737}$, $\mathbf{0.1699}$.

Are they significantly different?

Formally, $H_0 : \nu_1 = \ldots - \nu_5$, H_1 : at least one different.

Example:FMR

Calculate F test statistic (red line) for raw data. Bootstrap assuming H_0 is true and recalculate $F*$.

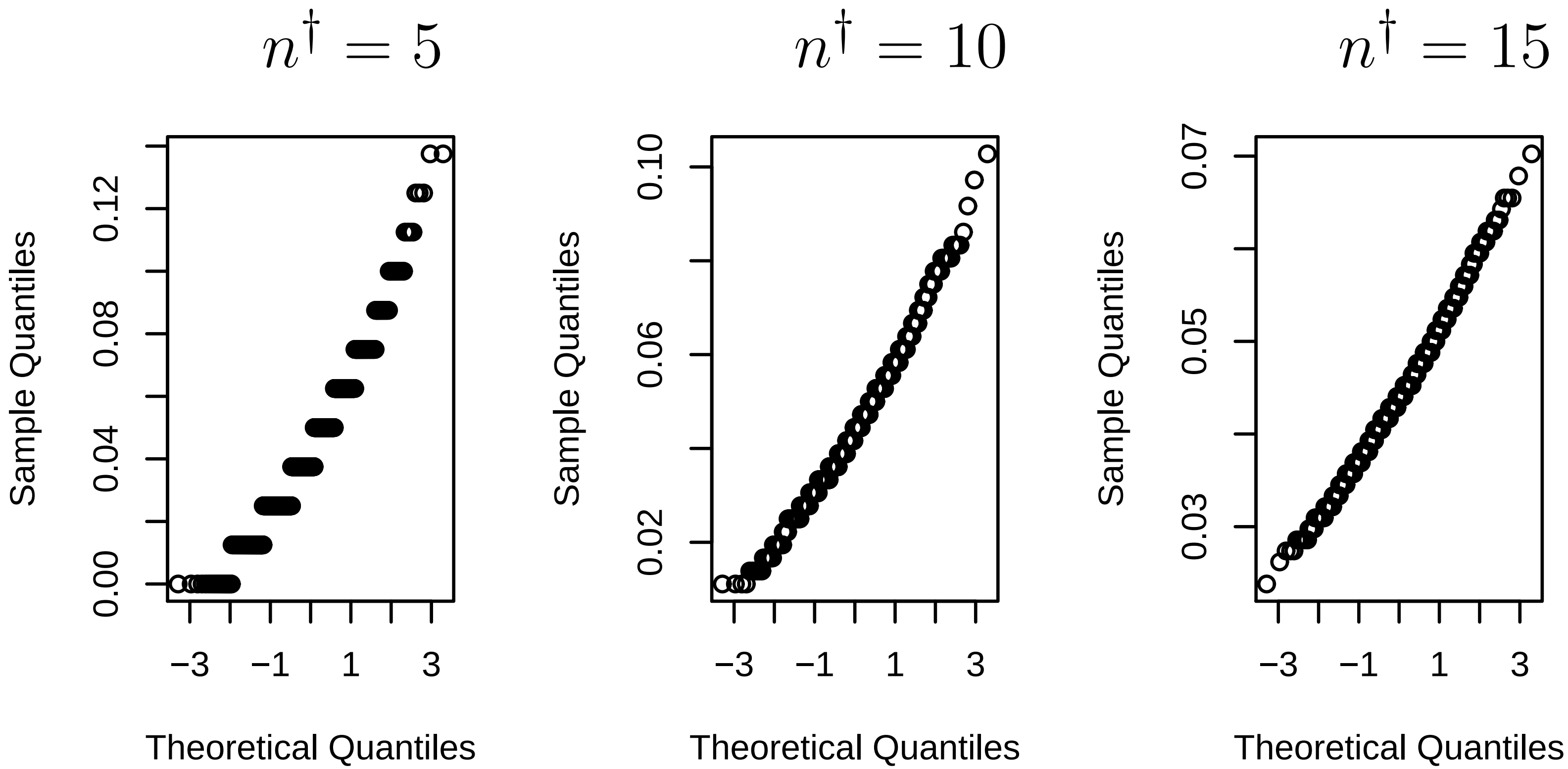


P-value for this test is 0.016.

CLT

Distribution of sampling distributions converge to Gaussian for large sample sizes

Example: Facial FMR (Data from Ross and Jain(2003))



Discussion

- All correlations are threshold dependent but model is not

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures
- Parametric approach used for sample size calc's, Schuckers (2009,2010)

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures
- Parametric approach used for sample size calc's, Schuckers (2009,2010)
- Needed for GLM's, covariate approaches (Colo. State et al. for FNMR)

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures
- Parametric approach used for sample size calc's, Schuckers (2009,2010)
- Needed for GLM's, covariate approaches (Colo. State et al. for FNMR)
- *Computational Methods in Biometrics* (June/July 2010) Springer

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures
- Parametric approach used for sample size calc's, Schuckers (2009,2010)
- Needed for GLM's, covariate approaches (Colo. State et al. for FNMR)
- *Computational Methods in Biometrics* (June/July 2010) Springer
- CITeR funded software PRESS v2.0

Discussion

- All correlations are threshold dependent but model is not
- Basis for more complicated correlation structures
- Parametric approach used for sample size calc's, Schuckers (2009,2010)
- Needed for GLM's, covariate approaches (Colo. State et al. for FNMR)
- *Computational Methods in Biometrics* (June/July 2010) Springer
- CITeR funded software PRESS v2.0

Schuckers (2010)

Computation Methods in Biometric Authentication :
Statistical Methods for Performance Evaluation
Text by *Springer* to appear Summer 2010

- Methods for FTE, FTA, FNMR, FMR, ROC, EER
- Methods for one sample, two sample, three or more samples
- Distribution-free and Large sample methods
- Independent and Paired data collections

Thank You

Questions?

`schuckers@stlawu.edu`