Test Sample Size Determination for Biometric Systems based on Confidence Elasticity

Rongfeng Li  
Key Laboratory of Machine Perception (MOE)  
Peking University  
Beijing, China  
rongfeng@pku.edu.cn

Beining Huang  
Key Laboratory of Machine Perception (MOE)  
Peking University  
Beijing, China  
hbn@pku.edu.cn

Wenxin Li  
Key Laboratory of Machine Perception (MOE)  
Peking University  
Beijing, China  
lwx@pku.edu.cn

Many researchers estimated the confidence interval of the evaluation index such as FMR and studied the relationship between the confidence interval width of FMR and sample size. In our research, we firstly indicated the relationship between confidence interval width \(w\) and sample size \(n\) by the equation: \(w^2 = b_1/n + b_2/n\). Apparently, the more test samples, the narrower of the confidence interval and the more convincible of the evaluation will come out. Most biometric test such as Fingerprint Verification Competitions (FVC) and NIST Fingerprint Evaluation collect as many samples as possible to get a convinced evaluation. However, based on the relationship between confidence interval and sample size, a big expansion of sample size only brings a little effect in reducing the confidence interval width when the sample size is very large. It has not been discussed till now whether it is worth achieving a narrow confidence interval by collecting a very big database. In this paper, we propose the concept of confidence elasticity which is defined by the ratio of the percentage change in confidence interval width to the percentage change in sample size to indicate the cost-effectiveness of the collecting data. Then we determine the test sample size of a deployed finger-vein biometric system according to empirical confidence elasticity. Experimental result shows that if we enlarge the sample size 150 which used in FVC2006 to 632 (about 4 times), the confidence interval width will reduce from 2.4% to 1.2% (about 1/2) based on the confidence elasticity of 0.5.

**Keywords**: performance evaluation; sample size; elasticity

### I. INTRODUCTION

One of the most important tasks of the biometric systems is to evaluate how well it makes correct decision about accepting or denying an individual’s access on a particular group in a specified environment. We measure the result by some error rates such as false match rate (FMR) or false non-match rate (FNMR). The number of the biometric systems users is usually expected to exceed millions. It is costly to collect all the data to evaluate the system, and ordinary evaluation only incompletely tests a part of the entire data set, namely the testing data set. It is impossible for the incomplete testing to evaluate the systems 100% accurately. Therefore, once we evaluate a system, the convincement of the evaluation should be taken into consideration. We can describe the convincement of the evaluation by its confidence intervals.

Many researchers worked at estimating the relationship between the confidence interval width and the sample size. Shen et al.[1], Wayman et al. [2], Bolle et al. [3] calculated the sample size by distribution estimation or resampling. When estimating the confidence interval of the evaluation index (such as FMR) which is calculated with multiple comparisons of a same individual, the methodology above will face a dependent problem: If A, B, C represent three different people, the matching of (A, B) is depend on the matching of (A, C) because they share the same person A. Incorrect assumption of the independent distribution leads to too small or too large variance. Introducing a correlation coefficient, Dass et al. [4] set up the correlation model based on Copula functions. Schuckers[5] developed a parametric methodology to set up a correlation framework. (The parametric here is not in the sense of Bolle[3] who assumes the shape of the distribution. Instead, Schuckers[5] used parameters to specify the first two moments of the distributions of interest and the parameters.) Meanwhile, Li[6] introduced a nonparametric methodology based on second-level partition for dependent resampling. This paper will discuss and deduce that the inference of the nonparametric methodology is the same as the parametric one: The relation between the confidence interval width: \(w\) and sample size: \(n\) can be formulated as the same equation: \(w^2 = \beta_1/n^2 + \beta_2/n\); \(\beta_1\) and \(\beta_2\) can be predicted by a regression model using a small database.

Apparently, the more test samples, the narrower of the confidence interval and the more convincible of the evaluation will come out. Most biometric test such as Fingerprint Verification Competitions 2006 (FVC2006[7]) and NIST Fingerprint Evaluation [8] collect as much samples as possible to get a convinced evaluation. However, based on the relationship between confidence interval as sample size, a big

---

This work is supported by the NSFC(No. 60933004).
expansion of sample size only brings a little effect in reducing the confidence interval width. Whether it is worth achieving a narrow confidence interval by collecting a very big database has not been discussed. According to the equation \( w^2 = \frac{\beta_1}{n^2} + \frac{\beta_2}{n} \), when \( n \) is very large, \( w^2 \approx \frac{1}{n} \). In NIST Fingerprint Evaluation [8], the sample size of the largest database is in the order of \( 10^8 \). Compared to a \( 10^8 \) samples database, the confidence interval width is only 10 times narrower based on the relation of \( w^2 \approx \frac{1}{n} \). In this case, we should consider if it is worth extending the sample size by 100 times to reduce the confidence interval width to 1/10. In this paper, we propose the concept of confidence elasticity which is defined by the ratio of the percentage change in confidence interval width to the percentage change in sample size. In economics, elasticity measures the degree to which the demand or supply reacts to a change in price. We use confidence elasticity to indicate the cost-effectiveness of the collecting data. High elasticity indicates that change in sample size result in big change of cost-effectiveness of the collecting data. High elasticity change in price. We use confidence elasticity to indicate the measures the degree to which the demand or supply reacts to a percentage change in sample size. In economics, elasticity is always decreasing when enlarging the sample size. In our research, we specify an empirical value of elasticity to determine a test sample size.

This paper takes FMR as an example of evaluation index to calculate the test sample size. We do not discuss FNMR because the correlation structure for false non-match decision data is much simpler than the correlation structure for false match decision data which is mentioned by Schuckers[5]. Also, the sample size decided by the confidence interval of FNMR is usually much smaller than FMR.

The rest of the paper is organized as follows: Section 2 introduces the calculation of confidence interval by the resampling model introduced by Li[6]. We conclude that the relationship between confidence interval width \( w \) and sample size \( n \) can be formulated by the equation \( w^2 = \frac{\beta_1}{n^2} + \frac{\beta_2}{n} \) where the coefficient \( \beta_1 \) and \( \beta_2 \) can be predicted by a regression model. In Section 3, we put forward our concept of confidence elasticity and propose a strategy to determine the test sample size. Section 4 provides experimental results of the test sample size determination on a finger vein database. Finally, the conclusion and future work discussion is shown in section 5.

II. RELATION BETWEEN CONFIDENCE INTERVAL WIDTH AND SAMPLE SIZE

In this section, we will calculate the relationship between confidence interval width and sample size based on dependent resampling introduced by Li[6], and we will conclude that the relation is consistent with the inference of Schuckers[5]. We start from the denotations of false rate and confidence and then describe the confidence interval estimation of FMR based on Li[6]'s dependent resampling. After that, we will provide the equation \( w^2 = \frac{\beta_1}{n^2} + \frac{\beta_2}{n} \) which indicates the relationship between confidence interval width and sample size. Finally, a regression model is introduced to predict the coefficient by a small sample size.

A. False Rates and Confidence Intervals

The process in which a biometric system accepts or rejects a person’s claim of identity begins with the system’s storing sample template of each person \( I \). We indicate \( D(I) \) as the template set of \( I \). Most biometric recognition algorithms define a similarity function \( s(x,y) \) that is used to show how similar two templates: \( x \) and \( y \) are. A threshold \( t \) is set to decide whether \( x \) and \( y \) belong to the same individual by \( s(x,y) \geq t \) or to different individuals by \( s(x,y) < t \). Two types of recognition error rate are led by the setting of the threshold: False Nonmatch Rate (FNMR) and False Match Rate (FMR). (Here, we use FNMR and FMR rather than the false reject rate and false accept rate, to be consistent with Wayman [2]).

They are defined as follows:

\[
FNMR = P(s(x,y) < t \mid x \in D(I), y \in D(I'), I = I')
\]

(1)

\[
FMR = P(s(x,y) \geq t \mid x \in D(I), y \in D(I'), I \neq I')
\]

(2)

\( P \) is the probability function.

Pay attention that the test data is randomly sampled from the entire data set, and the FMR are random variables which enable us to measure to what extent the evaluation is convincible, by the following \((1-\alpha)100\%\) confident interval \([FMR_{\alpha}, FMR_{1-\alpha}]\):

\[
P(FMR_{\alpha} < FMR < FMR_{1-\alpha}) > 1-\alpha
\]

(3)

We note the variance of FMR by \( V(FMR) \), and the confidence interval can be written as:

\[
[FMR - z_{1-\alpha/2} \sqrt{V(FMR)}, FMR + z_{1-\alpha/2} \sqrt{V(FMR)}]
\]

(4)

Here, \( z_{1-\alpha/2} \) is the \((1-\alpha/2)\)th percentile of a standard Gaussian distribution. Thus, the confidence interval \( w = z_{1-\alpha/2} \sqrt{V(FMR)} \).

Schuckers [5] has noted two kinds of sample size which affect the confidence interval \( w \), one is \( n \): the number of individuals to be tested and the other is \( m \): the number of times that each individual should be tested. He then demonstrated the relation between \( w \) and \( n,m \):

\[
w^2 \propto \frac{1+\xi_1 + (\eta + \xi_2)(m-1) + 4\omega(n-2)}{n(n-1)m}
\]

(5)

\( \xi_1, \xi_2, \eta, \omega \) are parameters which are used to set up the correlation structure and are independent from \( n, m \). This paper only pays attention to the number of individuals \( n \) (for the number of times is usually fixed in testing), and (5) can be written as:

\[
\frac{1 + \xi_1 + (\eta + \xi_2)(m-1) + 4\omega(n-2)}{n(n-1)m} \approx \frac{1}{n^2} + \frac{m}{n(n-1)} + \frac{n-2}{n(n-1)}
\]

Where \( \frac{1}{n(n-1)} \approx \frac{1}{n^2} \) and \( \frac{n-2}{n(n-1)} \approx \frac{1}{n} \) (Consider that \( n \) is a very large number). We rewrite (5) as:

\[
w^2 = \frac{\beta_1}{n^2} + \frac{\beta_2}{n}
\]

(6)

The following paragraphs will demonstrate a consistent inference by the nonparametric methodology.
B. Dependent Resampling

Resampling is the most widely-used nonparametric methodology. Bolle et al. [3] has introduced the independent resampling: subset Bootstrap for the estimation of variance of FMR, and Li[6] has developed the subset Bootstrap to a dependent resampling with second-level partitions. Noticing that the confidence interval width \( w = z_{1-\alpha/2} \sqrt{V(FMR)} \) Thus we only need to estimate the variance of FMR: \( V(FMR) \).

1) Subset Bootstrap Resampling

Subset bootstrap can be formulated as follows.

Suppose that we have \( n \) different individuals and each one has been tested for \( N \) times. Comparing each individual to the rest of others, we can get \( N=n(n-1)/2 \) individual pairs and \( Nd(d-1)/2 \) comparisons. We denote \( S \) as the collection of comparisons and \( S_i \), \( i=1,2,\ldots,N \) as comparison set of the \( i \)-th pair, and the subset bootstrap estimation of variance of FMR can be described with the following steps:

Step 1) For \( k=1,2,\ldots,B \),
   i) generate random integer array \( r_1,r_2,\ldots,r_N \) with replacement from \( \{1,2,\ldots,N\} \); 
   ii) generate bootstrap resample set \( S(r_1) \cup S(r_2) \cup \ldots \cup S(r_N) \);
   iii) calculate \( FMR_k \) of the bootstrap samples.

Step 2) Calculate the variance of the \( FMR_1,FMR_2,\ldots,FMR_k \) as the variance estimation:

\[
V(FMR) = \frac{1}{B} \sum_{k=1}^{B} \left( FMR_k - \frac{1}{B} \sum_{k=1}^{B} FMR_k \right)^2 
\]  

2) Second-Level Partition Resampling

The subset bootstrap partitions individual pair set to \( N \) subsets and we assume they are independent from others. However, if we denote \( S_{ij} \) as the false match set of two different individuals \( i,j \), for three different individuals \( i,j,k \), we cannot assume \( S_{ij} \) and \( S_{ik} \) are independent because they share the common individual \( i \). Therefore, we introduce a second-level partition to \( S \) so that we can separately process the second-level subsets to avoid the dependence problem.

Suppose that \( M \) is an integer and can divide \( N \) exactly, i.e. \( N=MK \).

The second-level partition is as follows:

\[
S^{(k)} = S^{(1)} \cup S^{(2)} \cup \ldots \cup S^{(K)} 
\]

\[
S = \bigcup_{k=1}^{K} S^{(k)} \tag{9}
\]

For each \( k=1,2,\ldots,K \), if the indices \( i_1^{(k)}, j_1^{(k)} \), \( i_2^{(k)}, j_2^{(k)} \), \ldots, \( i_M^{(k)}, j_M^{(k)} \) are exactly \( 2M \) different integers, we name the partition (9) as an independent partition. We then indicate Subset False Match Rate(SFMR) for each subset \( S^{(k)} \):

\[
SFMR^{(k)} = P(s > t \mid s \in S^{(k)}) \tag{10}
\]

Note that

\[
FMR = P(S > t \mid s \in S) = \sum_{k=1}^{K} P(s > t \mid s \in S^{(k)}) P(s \in S^{(k)}) \tag{11}
\]

Thus,

\[
FMR = \frac{1}{K} \sum_{k=1}^{K} SFMR^{(k)} \tag{12}
\]

The equation (12) allows us to resample the second-level subset and calculate SFMR separately to get the FMR resample.

When implementing the second-level partition (9), we will soon face the question that how to partition it independently such that: \( i_1^{(b)}, j_1^{(b)}, i_2^{(b)}, j_2^{(b)}, \ldots, i_M^{(b)}, j_M^{(b)} \) are \( 2M \) different integers. One way to answer the question is by using the following partition:

\[
S^{(b)} = \bigcup_{i+j=b (mod b)} S_{i,j}, \quad M=(n-1)/2
\]

\[
S^{(k)} = \bigcup_{i+j=k (mod n-1)} S_{i,j} \cup \bigcup_{2i+k(n-1) (mod n-1)} S_{i,n} \tag{14}
\]

The above resampling procedure by second-level partition can be illustrated by Figure 1.

![Figure 1. Procedure of resampling for estimating \( V(FMR) \)](image)

3) Relation Between Confidence Interval and Sample Size

Although nonparametric methodology cannot provide an equation like (5) which calculates the sample size with
parametric statistics, we can still demonstrate the relation between sample size \( n \) and variance of FMR \( V(FMR) \) estimated by dependent resampling.

According to (12), assume that SFAR\((k)\) obeys identically distribution, and we have:

\[
V(FMR) = \frac{1}{K^2} \left( KV(SFMR) + \sum_{u=v} \text{Cov}(SFMR(u), SFMR(v)) \right)
\]

(15)

Where

\[
\text{Cov}(SFMR(u), SFMR(v)) = \frac{1}{M} \sum_{r=1}^{M} \text{Cov}(FMR(S_{r, j}(u)), FMR(S_{r, j}(v)))
\]

\[
= \frac{1}{M} \sum_{r=1}^{M} \text{Cov}(FMR(S_{r, j}(u)), FMR(S_{r, j}(v)))
\]

3

\[
= \frac{1}{M} \text{Cov}(FMR(S_{r, j}), FMR(S_{r, j}))
\]

Thus, (21) is a linear regression model with weights

\[
y_k = w^2(n_k) + \epsilon_k = \frac{\beta_1}{n_k} + \frac{\beta_2}{n_k} + \epsilon_k, \quad k=1,2,\ldots,R
\]

(23)

Here, we assume that \( \epsilon_k \) obeys to a Gaussian distribution, the standard error of which is proportional to \( y_k \). Approximately, we assume that:

\[
\text{var}(\epsilon_k) = \left( \frac{\sigma}{n_k} \right)^2, \quad k=1,2,\ldots,R
\]

(24)

Thus, (21) is a linear regression model with weights \( f_k \):

\[
f_k = \frac{1}{n_k^2}, \quad k=1,2,\ldots,R
\]

(25)

1) Weighted linear regression

We fit the following linear model:

\[
y_k = w^2(n_k) + \epsilon_k = \frac{\beta_1}{n_k} + \frac{\beta_2}{n_k} + \epsilon_k, \quad k=1,2,\ldots,R
\]

2) Weighted least square and confidence interval of prediction

We denote that:

\[
Y = \begin{bmatrix} y_1 \\ y_R \end{bmatrix}, \quad X = \begin{bmatrix} \frac{1}{n_1^2} & \frac{1}{n_1} \\ \vdots & \vdots \\ \frac{1}{n_k^2} & \frac{1}{n_k} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \cdots & 0 \\ 0 & \ddots & \vdots \\ \vdots & \cdots & 1 \end{bmatrix}
\]

We denote that:

\[
\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}
\]

The coefficient \( \beta \) and \( \sigma \) are estimated by minimizing the following weighted residuals:

\[
\text{RSS} = (Y - X\beta)^\top \Sigma^{-1} (Y - X\beta)
\]

(26)

Results from Weisberg[10] are as follows:

\[
\hat{\beta} = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} Y
\]

(27)

Figure 2 show the procedure of coefficient prediction by the regression model.
Elasticity is used in the economy to measure to which degree the demand or supply reacts to the change in the price. In this paper, the confidence elasticity indicates the cost-effectiveness of reducing the confidence interval width by enlarging the sample size. Denoting that \( k = \frac{n}{n_0} \), and replacing \( w \) and \( w_0 \) by (6), we have:

\[
E_c(n_0, n) = \frac{w_0}{w/n} = \frac{n_0w_0}{n_w} \tag{29}
\]

From (29) we can see the \( E_c \) is always decreasing by the increasing of \( n_0 \) and \( k \). The equation (3) indicates that the cost-effectiveness is decreasing when enlarging the sample size.

When \( n_0 \) is very large, \( E_c \approx \frac{1}{k} \). Thus, enlarging the sample size by 100 times only results in reducing the confidence interval width to 1/10. Considering the low elasticity \( E_c=0.1 \), we are reluctant to enlarge the sample size to achieve a narrower confidence interval width.

### B. Strategy of Determining the Sample size

Our strategy of determination is based on the above confidence elasticity. The determination consists of the following two steps:

- **Setp 1:** Setting a base line sample size \( n_0 \) and calculate the corresponding confidence interval width \( w_0 \) by the equation (6):
  \[
w_0^2 = \frac{\beta_1}{n_0^2} + \frac{\beta_2}{n_0} \tag{30}
\]

- **Step 2:** Enlarge the sample size to \( n \), where the confidence elasticity \( E_c \) decrease to a base line elasticity \( E_0 \).

According to (30), we have:

\[
E_0 = E_c(n_0, n) = \frac{\beta_1 + \beta_2n_0}{\beta_1 + \beta_2n} \tag{32}
\]

Therefore,

\[
n = \frac{(1-E_c^2)\beta_1 + \beta_2n_0}{\beta_1 \beta_2 \beta_0^2} \tag{33}
\]

In this paper, the base line value of \( n_0 \) and \( E_0 \) are determined empirically.

### IV. EXPERIMENTAL RESULT

Our experiment is based on the data of PKU Finger Vein Database which is collected by the finger vein biometric system based on [11]. This system serves as a terminal checking attendance of exercise by students who take P.E lessons, with 10 terminals, 6,212 users in total. Some 4,000 people in average use the system every day. The system has worked well since deployed on March 9th, 2009. The database contains 50,700 different infrared finger-vein images from 10,140 fingers (5 images per finger) and 5,208 persons. Samples of the database are shown in Figure 3. Two finger vein recognition algorithms: A832 developed by Miura [12] and A835 developed by Huang[13] are implemented for the evaluation. Parts of the results of the evaluations are shown on an online testing website [http://rate.pku.edu.cn](http://rate.pku.edu.cn).
Our experiment consists of the following four steps:

Step 1: Randomly sample the dataset $T_1, T_2, ..., T_R$ from the database $S$ such that:

$$T_1 \subset T_2 \subset \cdots \subset T_R \subset S$$

Where the sample size of $T_1, T_2, ..., T_R$ are 10, 20, 30, ..., 10000.

Step 2: Estimating the confidence interval width $w_1, w_2, ..., w_R$ of $T_1, T_2, ..., T_R$ by the independent resampling model introduced in section 2.

Step 3: Predict the coefficient $\beta_1$ and $\beta_2$ of equation (6) using $w_1, w_2, ..., w_{10}$ by the regression model introduced in section 3. We validate the prediction by the $t$-test using $w_{100}, w_{5000}, w_{10000}$ and $w_{100000}$.

Step 4: Taking a base line value of sample size $n_0=150$ which is the sample size of FVC 2006. Confidence elasticity $E_0$ is set to 0.5 empirically, and we determine the sample size $n$ by the equation (31) and (33).

The result of coefficient prediction of $\beta_1$ and $\beta_2$ using the database of algorithms A832 and A835 are shown in Table 1.

Table I. Coefficient Prediction

<table>
<thead>
<tr>
<th></th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A832</td>
<td>8753</td>
<td>831</td>
</tr>
<tr>
<td>A835</td>
<td>33831</td>
<td>636</td>
</tr>
</tbody>
</table>

The result of $t$-test is shown in Table 2.

Table II. T-test of Regression Prediction

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>p value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A832</td>
<td>96.8%</td>
<td>98.4%</td>
<td>98.8%</td>
<td>99.2%</td>
<td>99.7%</td>
</tr>
<tr>
<td>A835</td>
<td>95.8%</td>
<td>97.4%</td>
<td>98.3%</td>
<td>99.1%</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

Table 2 shows that the prediction of $\beta_1$ and $\beta_2$ is convincing.

The result of sample size determination step is shown by Table 3.

Table III. Result of Sample Size Determination

<table>
<thead>
<tr>
<th></th>
<th>$n_0$</th>
<th>$w_0$</th>
<th>$n$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A832</td>
<td>150</td>
<td>2.43%</td>
<td>632</td>
<td>1.2%</td>
</tr>
<tr>
<td>A835</td>
<td>150</td>
<td>2.41%</td>
<td>760</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Take A832 as an example, the result of Table 3 indicates that by the baseline sample size 150 which is used in FVC2006, we would enlarge the sample size from 150 to 632 (about 4 times) to reduce the confidence interval from 2.4% to 1.2% (about 1/2) based on the base line value of confidence elasticity $E_0=0.5$. If we continue to enlarge the sample size, the confidence elasticity will be smaller than the $E_0=0.5$ which shows that the cost-effectiveness is lower than the base line value. For example, if we enlarge the sample size to the order of 1500000 (about $10^7$ times) which achieve the same order of magnitude of the largest database of NIST Fingerprint Evaluation, the confidence interval width only reduce to 0.02% (about 1/100 of 2.4%). The confidence elasticity is 0.01. In this case, the cost-effectiveness is low to collect so many samples.

V. Conclusions and Future Discussions

As much other work in performance evaluation of Biometric systems this research concentrates on determining the test sample size for convincible evaluation. In this paper, we applied the independent resampling model by Li[6] to calculate the relation between the confidence interval width $w$ and sample size $n$. The relationship is formulated by the equation: $w^2 = \beta_1^2 + \beta_2^2$, where the coefficient $\beta_1$ and $\beta_2$ are predicted by a regression model. Then we propose the concept of confidence elasticity which is defined by the ratio of the percentage change in confidence interval width to the percentage change in sample size. The elasticity can indicate the cost-effectiveness of the collecting data. In the experiment based on a finger-vein database, we enlarge the sample size 150 which is used in FVC2006 to 653 (about 4 times expansion) to reduce the confidence interval width from 2.4% to 1.2% (about 1/2).

The base line value of the sample size $n_0$ and the confidence elasticity $E_0$ play important roles in sample size determination. However, in this paper, $n_0$ and $E_0$ are set to an empirical value. The method to set $n_0$ and $E_0$ still needs to be discussed and figured out in the future.

Although different recognition algorithms are used, we only experiment on a finger-vein system. Thus, the future work shall also include experiments on different Biometric Systems.

REFERENCES


