Paper

Probabilistic Issues in Biometric Template Design

Andrzej Pacut

Biometric Laboratories, Research and Academic Computer Network NASK, Warsaw, Poland
Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland

Abstract—Since the notion of biometric template is not well defined, various concepts are used in biometrics practice. In this paper we present a systematic view on a family of template concepts based on the L_1 or L_2 dissimilarities. In particular, for sample vectors of independent components we find out how likely it is for the median code to be a sample vector.

Keywords—biometrics, sample median, template.

1. Introduction

Biometric template is commonly understood as a certain *best representative* of a set of *enrolment data*. This description does not actually makes a definition, since the meaning of 'representative' is only intuitive and the meaning of 'best' is also not defined. In fact, various understanding of those terms lead researchers to quite different transformations of the enrolment data into the template.
In this paper we will sort out several meanings of the term "the best representative" and discuss the resulting methods of template construction.

2. Enrolment Measurements as the Sample

We assume that the biometric *enrolment sample* $\mathbf{X}$ for a given subject is a sample of size n in $\mathbb{R}^\ell$, i.e., it consist of a finite sequence of biometric measurements

$$\mathbf{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}), \tag{1}$$

where each measurement $\mathbf{x}^{(i)}$ is represented by a *vector*

$$\mathbf{x}^{(i)} = \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_\ell^{(i)} \end{bmatrix} \in \mathbb{R}^\ell. \tag{2}$$

To differentiate between the sample elements (the vectors) and vectors elements we call the latter the vector components. In the probabilistic context, we always assume that the sample vectors are independent and have identical distribution (the i.i.d. sample). It is often useful to reinterpret the measurements as finite *sequences* of real numbers $\mathbf{x}^{(i)} = \left(x_1^{(i)}, \ldots, x_\ell^i\right)$; we will use either interpretation. Note that we use upper indexes in parentheses to number the measurements, and reserve lower indexes for their components.

If $\ell = 1$, the sample is called scalar. In particular, one may consider scalar *component samples* that consist of selected components of all sample vectors, namely

$$X_j = (x_j^{(1)}, \ldots, x_j^{(n)}), \qquad j = 1, \ldots, \ell. \tag{3}$$

Elements of the scalar sample can be rearranged in a non-decreasing order

$$x^{[1]} \le x^{[2]} \le \cdots \le x^{[n]}, \tag{4}$$

so that $x^{[1]}$ is the smallest sample element, $x^{[r]}$ is the rth smallest, so that $x^{[n]}$ is the largest. The sample can thus be represented by the *ordered sample*

$$(x^{[1]}, \ldots, x^{[n]}) \tag{5}$$

if the original order of sample elements is irrelevant. Note that the ordered representation (5) is in general non unique due to possible repetitions in the sample. This happens in particular if a scalar sample is generated by a discrete random variable whose *support set* is finite $Y = \{y^{(j)}, j = 1, \ldots, M\}$, $y^{(1)} < \ldots < y^{(M)}$. The sample can be then characterized by the support values together with their multiplicities m^j, namely, by *the set*

$$\left\{(y^{(j)}, m^{(j)}),\ y^{(j)} \in Y,\ j = 1, \ldots, M\right\}. \tag{6}$$

Certainly $\sum_{j=1}^{M} m^{(j)} = n$. In particular, we will be interested in the *binary case* with $Y = \{0, 1\}$.
The enrolment sets used in biometrics can have more complex structure. For instance, the measurements can be of varying lengths, like in signature biometrics. In those cases, the concepts discussed in this paper must be appropriately modified.

3. Template Concepts

We now consider several concepts of the template for the enrolment sample Eq. (1). All concepts employ the notion of dissimilarity $\overline{D}_p$ between the sample and a vector, understood here as the average pth power of the L_p distance of the vector to all the enrollment vectors

$$\begin{aligned} \overline{D}_p(\mathbf{x}, \mathbf{X}) &= \frac{1}{n} \sum_{\mathbf{x}^{(i)} \in \mathbf{X}} d(\mathbf{x}, \mathbf{x}^{(i)})^p = \frac{1}{n} \sum_{\mathbf{x}^{(i)} \in \mathbf{X}} \sum_{j=1}^{\ell} \left| x_j - x_j^{(i)} \right|^p \\ &= \frac{1}{n} \sum_{j=1}^{\ell} \sum_{x_j^{(i)} \in X_j} \left| x_j - x_j^{(i)} \right|^p, \end{aligned} \tag{7}$$

where X_j is jth component sample. In particular, $\overline{D}_p$ comes down to the average distance ($p=1$) or the average squared distance ($p=2$) between a vector and the enrollment vectors. We often skip the index p when it is obvious from the context.

Within this approach, we distinguish four elementary concepts, each in either L_1 or L_2 versions, thus making together eight interpretations of the template. The templates obtained by a search for "the best" sample vector will be referred to as template-S (for **s**ample), and those obtained by looking for "the best" vector, not necessarily being a sample vector, will be called template-R (for **r**eal). The result of the approach that combines the two will be called template-RS. Finally, the template intended to minimize the average dissimilarity between the sample vectors and an unknown testing vector will be called template-T (for **t**esting). Certainly, these elementary concepts are far from being exhaustive, and many other, more sophisticated template concepts can be introduced.

The solutions to the underlying minimizations problems we discuss are typically not unique, and by $\mathrm{Arg}\min_{\mathbf{x}\in Z}(\overline{D}_p)$ we denote *the set* of vectors that minimize $\overline{D}_p$ over $x\in Z$.

In the first concept of the template, one of the enrolment vectors is chosen to represent the sample.

Definition of template-S. *The template is equal to any enrolment vector that minimizes the average dissimilarity Eq.* (7) *between this vector and the enrolment vectors of the same subject, namely*

$$\mathbf{x}^{*S}\in\mathbf{X}^{*S}\stackrel{\mathrm{def}}{=}\mathrm{Arg}\min_{\mathbf{x}^{(k)}\in\mathbf{X}}\overline{D}_p(\mathbf{x}^{(k)},\mathbf{X})\,. \tag{8}$$

Note that template-S is in general not defined uniquely; it is even possible that all the enrolment vectors fulfil the definition condition. Certainly, all $\mathbf{x}^{*S}\in\mathbf{X}^{*S}$ lead to the same minimal average dissimilarity

$$\overline{D}_p^{*S}\stackrel{\mathrm{def}}{=}\overline{D}_p(\mathbf{x}^{*S},\mathbf{X})\quad\text{for all}\quad\mathbf{x}^{*S}\in\mathbf{X}^{*S}. \tag{9}$$

Template-S definition restricts the search to the enrolment vectors. In the next definition the search is extended to the entire $\mathbb{R}^\ell$.

Definition of template-R. *The template is equal to any vector that minimizes the average dissimilarity Eq.* (7) *between this vector and all the enrolment vectors of the same subject, namely*

$$\mathbf{x}^{*R}\in\mathbf{X}^{*R}\stackrel{\mathrm{def}}{=}\mathrm{Arg}\min_{\mathbf{x}\in\mathbb{R}^\ell}\overline{D}_p(\mathbf{x},\mathbf{X})\,. \tag{10}$$

The template here *may not belong* to the enrolment sample. Again, the definition does not in general lead to a unique solution. Unlike Eq. (8), definition Eq. (10) can be substantially simplified: by Eq. (7), the minimization in Eq. (10) can be performed separately for the sample components, namely

$$\min_{\mathbf{x}\in\mathbb{R}^\ell}\overline{D}_p(\mathbf{x},\mathbf{X})=\frac{1}{n}\sum_{j=1}^{\ell}\min_{x_j\in\mathbb{R}}\sum_{x_j^{(i)}\in X_j}|x_j-x_j^{(i)}|^p\,. \tag{11}$$

In the result, the definition of template-R, can be expressed in an equivalent form:

The template is equal to any vector whose components minimize the average dissimilarities between these components and the corresponding components of the enrolment vectors, namely

$$x_j^{*R}\in X_j^{*R}\stackrel{\mathrm{def}}{=}\mathrm{Arg}\min_{x\in\mathbb{R}}\sum_{x_j^{(i)}\in X_j}|x-x_j^{(i)}|^p=\mathrm{Arg}\min_{x\in\mathbb{R}}\overline{D}_p(x,X_j),\qquad j=1,\ldots,\ell\,. \tag{12}$$

The minimization of template-R in $\mathbb{R}^\ell$ has been in the above formulation replaced by a series of minimizations in $\mathbb{R}$, which may computationally be much simpler. The minimal dissimilarity is the sum of the component dissimilarities, namely

$$\overline{D}_p^{*R}\stackrel{\mathrm{def}}{=}\overline{D}_p(\mathbf{x}^{*R},\mathbf{X})=\sum_{j=1}^{\ell}\overline{D}_p(x_j^{*R},X_j) \tag{13}$$

and certainly

$$\overline{D}_p^{*R}\le\overline{D}_p^{*S}\,, \tag{14}$$

hence template-R is "better" than template-S. Note that the minimization in Eq. (8) cannot be decomposed into component sample minimizations due to a dependence between the components of $\mathbf{x}$ induced by the restriction of $\mathbf{x}$ to enrolment vectors.

Simplification in template-R definition comes for the cost of the template being not an element of the enrollment sample. To overcome this, one may in a sense integrate a simplicity of template-R definition with an intuitive need of the template to be a sample element as realized by template-S. In the next template concept, we will be looking for the sample element closest to the reference vector calculated according to the definition of template-R.

Definition of template-RS. *The template is equal to any enrolment vector that minimizes the distance to template-R, namely*

$$\mathbf{x}^{*RS}\in\mathbf{X}^{*RS}=\arg\min_{\mathbf{x}^{(i)}\in\mathbf{X}}d(\mathbf{x}^{(i)},\mathbf{x}^{*R})\,. \tag{15}$$

Note that the dissimilarity comes down here to pth power of the distance in L_p, thus the minimization just calls for a minimization of the distance. The value of the $\overline{D}_p$ for template-RS

$$\overline{D}_p^{*RS}=\overline{D}_p(\mathbf{x}^{*RS},\mathbf{X}) \tag{16}$$

certainly fulfils the inequality

$$\overline{D}_p^{*R}\le\overline{D}_p^{*S}\le\overline{D}_p^{*RS}\,, \tag{17}$$

so it is the worst, *in the sense of the average dissimilarity*, of the three templates considered so far.

All the concepts outlined above define the template as a certain representation of the sample, with formally sound but arbitrary meanings of 'representation'. This raises a question whether 'representation' could not be defined less arbitrarily. We thus propose a template concept based on the

template use, introducing an unknown test vector $\mathbf{x^0}$ the template will be compared to. Let $\mathbf{g}$ be a (Borel) function that maps a sample (of a fixed size) in $\mathbb{R}^\ell$ into a vector in $\mathbb{R}^\ell$, i.e., $\mathbf{g}(\mathbf{X}) = \mathbf{g}(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}) \in \mathbb{R}^\ell$, and denote by $\mathcal{G}$ the family of all such functions.

Definition of template-T. *The template vector is equal to the value of any vector-valued (Borel) function of the enrolment data that minimizes the conditional expected distance to the (unknown) test vector given the enrollment data, namely*

$$\mathbf{x}^{*T} = \mathbf{g}^*(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}), \quad \text{where}$$
$$\mathbf{g}^* \in \operatorname{Arg}\min_{\mathbf{g}\in\mathcal{G}} \mathcal{E}_{\mathbf{X}} d\big(\mathbf{x^0}, \mathbf{g}(\mathbf{X})\big), \tag{18}$$

where $\mathbf{x^0}$ is a test vector, and $\mathcal{E}_{\mathbf{X}}$ denotes the conditional expectation given $\mathbf{X}$.

Here the minimization is performed over all (Borel) vector *functions* of the template. The result is obviously not necessarily one of the template elements. The minimization in Eq. (18) can be performed separately for each component of the vector function $\mathbf{g}$, similarly to what we did for template-R. Consequently, the definition of template-T, can be replaced by the following equivalent concept:
The template is equal to any (Borel) vector-valued function of the enrolment data whose each component minimizes the expected distance to the corresponding component of the test vector, namely

$$\mathbf{x}^{*T} = \begin{bmatrix} g_1^*(\mathbf{X}) \\ \vdots \\ g_\ell^*(\mathbf{X}) \end{bmatrix}, \quad \text{where}$$
$$g_j^* \in \mathcal{G}_j^* = \operatorname{Arg}\min_{g_j\in\mathcal{G}} \mathcal{E}_{\mathbf{X}} \big|x_j^0 - g_j(\mathbf{X})\big|^p, \tag{19}$$

where $\mathcal{G}$ is a family of (Borel) functions that map a scalar sample into a scalar.

Within Bayesian context, we assume here that the test vector $\mathbf{x^0}$, and the template vectors $\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(n)}$ are independent, identically distributed, parameterized by the same unknown parameter vector $\boldsymbol{\vartheta}$, and moreover, that $\mathbf{x^0}$ and $\mathbf{X}$ are conditionally independent given $\boldsymbol{\vartheta}$. We may further rewrite $\mathcal{G}_j^*$ of Eq. (19) in the form

$$\mathcal{G}_j^* = \operatorname{Arg}\min_{g_j\in\mathcal{G}} \mathcal{E}_{\mathbf{X}} |x_j^0(\boldsymbol{\vartheta}) - g_j(\mathbf{x}(\boldsymbol{\vartheta}))|^p$$
$$= \operatorname{Arg}\min_{g_j\in\mathcal{G}} \mathcal{E}_{\mathbf{X}} \mathcal{E}_{\mathbf{X},\boldsymbol{\vartheta}} |x_j^0(\boldsymbol{\vartheta}) - g_j(\mathbf{x})|^p, \tag{20}$$

so for each sample $\mathbf{X}$ and each (unknown) parameter vector $\boldsymbol{\vartheta}$

$$\mathcal{G}_j^* = \operatorname{Arg}\min_{g_j\in\mathcal{G}} \mathcal{E}_{\mathbf{X},\boldsymbol{\vartheta}} |x_j^0(\boldsymbol{\vartheta}) - g_j(\mathbf{x})|^p. \tag{21}$$

In what follows we analyze some properties of the four above template concepts in L_1 and L_2 spaces.

4. L_1 Version of Template-R

As we earlier noticed, calculation of template-R, comes down to a series of minimizations for scalar samples. We will thus remind a classical issue of finding a real number x^* closest on the average to all scalar sample elements, i.e., the one that minimizes the average dissimilarity Eq. (7) specified to L_1, namely the average absolute distance

$$\overline{D}_1(x^*;X) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} |x^{(i)} - x^*|. \tag{22}$$

Let us first notice that *for scalar samples*, the minimization of $\overline{D}_1$ over $\mathbb{R}$ leads to one of the sample elements, e.a. the minimizations over $\mathbb{R}$ and over X lead to the same result.

Proposition 1. *For one-dimensional samples, minimizations of the average L_1 distance to the sample elements over all real values (real domain), and over all sample values (sample domain) lead to the identical minimum*

$$\min_{x\in\mathbb{R}} \overline{D}_1(x,X) = \min_{x\in X} \overline{D}_1(x,X) \tag{23}$$

and the solution set in the sample domain is a subset of the one for the real domain

$$X_1^{*S} \subseteq X_1^{*R}. \tag{24}$$

Proof. The function to be minimized is piecewise linear and bounded from below. Hence the minimum always exists, and can be assumed either at a non-differentiability point, namely one of the sample points $x^{(1)},\ldots,x^{(n)}$, or at the points of the closed segment between two neighboring non-differentiability points (Fig. 1). In the finite support case, the non-differentiability points are just the supporting points $y^{(1)},\ldots,y^{(M)}$ (of non-zero multiplicities).

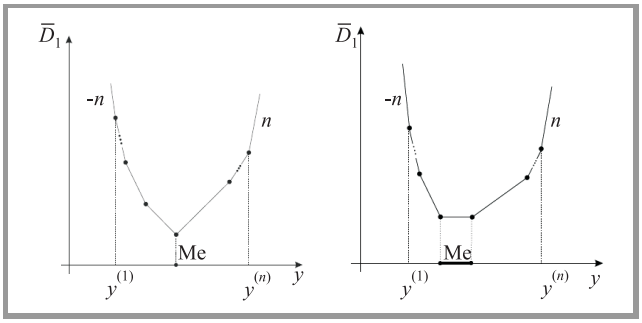

Fig. 1. Two cases of the minimum location of $\overline{D}_1$.

Consequently, for scalar samples $X \subset \mathbb{R}$, the points that minimize $\overline{D}_1$ in the sample domain, minimize also $\overline{D}_1$ in the real domain. In other words, in L_1, template-S fulfills the requirements of the definitions of template-R and template-RS.
Minimization of Eq. (22) has a well known solution, which for scalar samples is related to the *sample median*. To formulate it more precisely, we first recall the basic properties of the sample median.

4.1. Sample Median

For scalar samples, the sample median $\mathrm{me}(X)$ is understood as any number that "bisects the ordered sample". More precisely, it has the property

$$\begin{aligned} \mathrm{size}\{i : x^{(i)} \le \mathrm{me}(X)\} &\ge n/2\,, \\ \mathrm{size}\{i : x^{(i)} \ge \mathrm{me}(X)\} &\ge n/2\,. \end{aligned} \tag{25}$$

The set of all values that fulfills Ineq. (25) will be called the *median set* $\mathrm{Me}(x)$. If the sample size n is odd then simply

$$\mathrm{me}(X) = x^{[(n+1)/2]}.$$

Note that while in this case the sample median *is defined uniquely*, this value may be taken by more than one sample element: readily, when there are repetitions at the median value, more than one sample element may take the identical value equal to the median; we will call them the *median elements*. This is why the requirements of Ineq. (25) must also allow for sizes greater than the half of the sample size.

If the sample size n is even, any number in the median set

$$\mathrm{Me}(X) = \left[x^{[n/2]}, x^{[n/2+1]}\right], \tag{26}$$

called here the *median interval*, fulfills the requirements of Ineq. (25) hence the definition in this case *may not be unique*. Apart from the values *inside* the median interval, which are not sample values, Ineq. (25) is fulfilled also by the two end points of this interval, which *are* the sample values. If there are repetitions, $x^{[n/2]}$ can be equal to $x^{[n/2+1]}$ so then the median interval shrinks to a single value, and the median is again defined uniquely as the single element of the median interval. Note that both in odd and even sample sizes, *more than one sample element* can be equal to the median.

To make the definition of median unique for any sample size, one often chooses the middle of the median interval as the median in the even sample size case, so then $\mathrm{me}(X) \stackrel{\mathrm{def}}{=} \left(x^{[n/2]} + x^{[n/2+1]}\right)/2$. We are interested in the median as a – non necessarily unique – solution to a minimization problem, so we remain with the definition Eq. (26) for even sample sizes, and often deal with median intervals $\mathrm{Me}(X)$ rather then median values.

Summing up, the median, as an element of the median set, can be equal to one or more sample elements, or be equal to the values which are not sample elements at all (for even sizes).

In the special case of odd-sized binary samples

$$\mathrm{me}(X) = \begin{cases} 0 \text{ if } m^{(0)} > m^{(1)} \\ 1 \text{ if } m^{(0)} < m^{(1)} \end{cases} = \mathbf{1}(m^{(1)} - m^{(0)})\,, \tag{27}$$

where $\mathbf{1}$ denotes the step function, so the sample median is equal to the sample majority value.

In what follows, we will always focus on odd-value samples. It yet straightforward to include also the even-size samples.

4.2. L_1 Minimization

We are now prepared to minimize Eq. (22) for scalar samples using an elementary reasoning. First we characterize the function to be minimized. To avoid repetitions, sample points $x^{(i)}$ will be represented here by sample suport points $y^{(r)}$.

Proposition 2 (Average distance). *The following recursive formula applies in the finite support case*

$$\begin{aligned} \delta^{(r)} &= \delta^{(r-1)} + 2m^{(r-1)} \\ \overline{D}_1(y^{(r)}, X) &= \overline{D}_1(y^{(r-1)}, X) + |y^{(r)} - y^{(r-1)}|\,\delta^{(r)}\,, \end{aligned} \tag{28}$$

with

$$\delta^{(0)} = -n\,.$$

The proof of Eq. (28) is immediate and results directly from Proposition 1.

Consequently, the value

$$\min_{x\in\mathbb{R}} \overline{D}_1(x; X) \tag{29}$$

is for odd n uniquely attained by the sample median, and for even n is attained by any point of the median interval.

As seen from Proposition 2, for a given scalar sample, $\overline{D}_1$ is segmentwise linear, with the slopes increasing from some initial negative slope as x increases. Moreover, if n is odd then the function decreases to the left of $y^{((r-1)/2)}$ and increases to right of $y^{((r-1)/2)}$, where

$$\begin{aligned} \delta^{(r)} &< 0\,, \\ \delta^{(r+1)} &> 0\,. \end{aligned} \tag{30}$$

Consequently, the function attains its minimum at $y^{((r-1)/2)}$, which is the sample median. Similarly, if n is even, then $\overline{D}_1$ decreases to the left of $y^{(r/2)}$ and increases to right of $y^{(r/2+1)}$, hence it attains its minimum at all points of the segment $[y^{(r/2)}, y^{(r/2+1)})$ which is identical to the median set.

4.3. Template-R: The Explicit Formula

The above discussion enables to find the vector that fulfills the definition of the template-R in L_1:

Proposition 3 (L_1 minimization in $\mathbb{R}^n$). *The template-R is for odd n uniquely given by the vector of sample medians of the component samples*

$$\mathbf{x}^{*R} = \arg\min_{\mathbf{x}\in\mathbb{R}^\ell} \overline{D}_1(\mathbf{x}, \mathbf{X}) = \mathbf{me}(\mathbf{X}) = \begin{bmatrix} \mathrm{me}(X_1) \\ \vdots \\ \mathrm{me}(X_\ell) \end{bmatrix}. \tag{31}$$

For even n, the solution is not unique and is attained by any vector whose components belong to the median intervals of the corresponding component samples.

The solution Eq. (31) will be in short called the *median vector* or the *median sequence*, depending on the interpretation. Note that for binary vector samples, the median vector is identical to the *majority code*. The median vector (which fulfills the definition of template-R) *is not* in general a sample vector. In fact, it is easy to see that *for two-dimensional binary samples*, the median vector is always equal to some sample element and thus template-R and template-S are identical. However, there exist 3-dimensional binary samples for which the median vector *is not* an element of the sample. For instance, take $\mathbf{x}^{(1)} = \begin{bmatrix}1\\1\\0\end{bmatrix}$, $\mathbf{x}^{(2)} = \begin{bmatrix}1\\0\\1\end{bmatrix}$ and $\mathbf{x}^{(3)} = \begin{bmatrix}0\\1\\1\end{bmatrix}$. Then $\mathbf{me}(\mathbf{X}) = \begin{bmatrix}1\\1\\1\end{bmatrix}$ is not equal to any sample vector (Fig. 2).

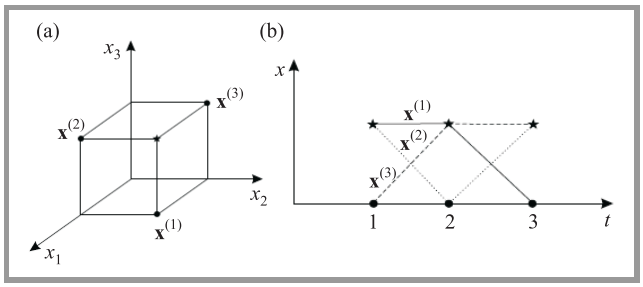

Fig. 2. Example of a 3-dimensional binary median sample for which the median vector is not a sample element. In the vector interpretation (a) the sample elements are marked with filled circles, and the median vector is marked with a star. Similarly, in the sequence interpretation of the vector sample (b) the sample elements are marked with filled circles and joined by different line types; the median sequence points are marked with stars.

5. L_1 Version of Template-S

In the definition of template-S, the minimization of $\overline{D}_1(\mathbf{X},\mathbf{x})$ is performed over the sample elements, instead of the entire $\mathbb{R}^\ell$. Consequently, this minimization *cannot be decomposed* into independent minimizations in $\mathbb{R}$, hence $\overline{D}_1$ is minimized by a different vector than the one solving (29), and the resulting minimal L_1 average distance $\overline{D}_1^*$ is certainly worse.

We will now analyze what is the chance that template-S is identical to template-R for finite support samples. We additionally assume that the finite support sample vectors have *independent components* (not necessarily binary), and derive the probability that the median vector *is* a sample vector. In this order, we first derive the median distribution for finite support samples, and then the distribution of the number of median vectors in the sample.

5.1. Sample Median Distribution

Consider a discrete scalar random variable ξ whose distribution has a finite support $Y = \{y^{(1)},\dots,y^{(M)}\}$. By P, F, and S, we denote its probability function, distribution function, and survival function, respectively, namely

$$\begin{aligned} P(y) &\stackrel{\text{def}}{=} \mathcal{P}(\xi = y)\,,\\ F(y) &\stackrel{\text{def}}{=} \mathcal{P}(\xi < y)\,,\\ S(y) &\stackrel{\text{def}}{=} \mathcal{P}(\xi > y)\ , \end{aligned} \tag{32}$$

for $y \in Y$. Note that some authors use $F(y) = \mathcal{P}(\xi \le y)$, $S(y) = \mathcal{P}(\xi \ge y)$ and then the formulas below would look differently.

To find the distribution of the ordered sample values for finite support i.i.d. sample, we use the result of [1]. The rth order statistic is equal to y if there are $u = 0,\dots,r-1$ values less than y and $w = 0,\dots,n-r$ values greater than y. The remaining $s = n-u-w$ values must be equal to y. Consequently, for n-element sample X, the probability that the rth order statistic is equal to some $y \in \{y^{(1)},\dots,y^{(M)}\}$ is given by

$$\mathcal{P}(x^{[r]} = y) = \begin{cases} \sum\limits_{w=0}^{n-r} \binom{n}{w} P(y)^{n-w} S(y)^w\,, \\ \qquad\qquad \text{for } y = y^{(1)} \\ \sum\limits_{u=0}^{r-1} \sum\limits_{w=0}^{n-r} \binom{n}{u} \binom{n-u}{w} F(y)^u P(y)^{n-u-w} S(y)^w\,, \\ \qquad\qquad \text{for } y = y^{(2)},\dots,y^{(M-1)} \\ \sum\limits_{u=0}^{r-1} \binom{n}{u} F(y)^u P(y)^{n-u}\,, \\ \qquad\qquad \text{for } y = y^{(M)}\,. \end{cases} \tag{33}$$

To simplify the notation, we assume from this moment on that *the sample size is odd*. Derivation for even-size samples must take into account the non-uniqueness of the median value, what makes the formulas a little more complex.

We now can easily find the median distribution for odd sample sizes. Setting r in Eq. (33) to $(n+1)/2$, which corresponds to the median, we obtain the distribution μ of the median

$$\mu(y) = \mathcal{P}(\mathrm{me}(X) = y) = \begin{cases} \sum\limits_{w=0}^{\overline{n}} \binom{n}{w} P(y)^{n-w} S(y)^w\,, \\ \qquad\qquad \text{for } y = y^{(1)} \\ \sum\limits_{u=0}^{\overline{n}} \binom{n}{u} \sum\limits_{w=0}^{\overline{n}} \binom{n-u}{w} \times \\ \quad \times P(y)^{n-u-w} F(y)^u S(y)^w\,, \\ \qquad\qquad \text{for } y = y^{(2)},\dots,y^{(M-1)} \\ \sum\limits_{u=0}^{\overline{n}} \binom{n}{u} P(y)^{n-u} F(y)^u\,, \\ \qquad\qquad \text{for } y = y^{(M)}\,, \end{cases} \tag{34}$$

where

$$\overline{n} \stackrel{\text{def}}{=} \frac{n-1}{2}\,. \tag{35}$$

For example, for the binary case ($Y = \{0,1\}$, $P(1) = p$, $P(0) = q = 1-p$) we have

$$\mu(y) = \begin{cases} \sum\limits_{w=0}^{\overline{n}} \binom{n}{w} q^{n-w} p^w\,, & \text{for } y = 0 \\ \sum\limits_{u=0}^{\overline{n}} \binom{n}{u} p^{n-u} q^u\,, & \text{for } y = 1\,. \end{cases} \tag{36}$$

While for the i.i.d. samples, the sample median converges with the sample size to the population median if such is uniquely defined, yet for samples sizes typically considered in biometrics, the two quantities may strongly differ (see Fig. 3).

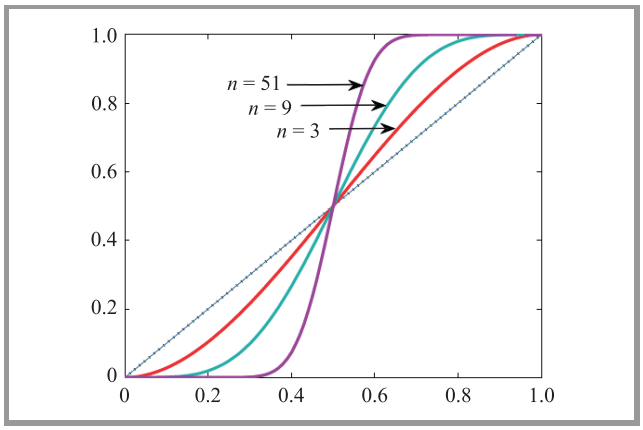

Fig. 3. Probability that the sample median is equal to one versus the probability of success p for the scalar binary i.i.d. sample, for several values of the sample size n. The population median is equal to 1 for $p > 0.5$.

5.2. Number of Median Elements

Note that in the scalar sample there is at least one element equal to the median, called here further the *median element*. Typically, for scalar finite-support samples, there are even more than one median element. We will derive the distribution of the number of median elements in a scalar finite support sample, irrespectively of the median value. Denote by $\mathcal{M}$ the number of median elements in X and by ν its distribution function, i.e.,

$$\nu(z) \stackrel{\text{def}}{=} \mathcal{P}\{\mathcal{M} = z\}, \quad z = 0, \ldots, n. \tag{37}$$

Proposition 4 (Distribution of the number of median elements in scalar finite-support samples). *The distribution of the number of median elements is given by*

$$\nu(s) = \begin{cases} 0, & \text{for } s = 0 \\ \binom{n}{s} \sum_{m=2}^{M-1} P(y^{(m)})^s \sum_{u=\overline{n}+1-s}^{\overline{n}} \binom{n-s}{u} F(y^{(m)})^u S(y^{(m)})^{n-s-u}, & \text{for } s = 1, \ldots, \overline{n} \\ \binom{n}{s} \sum_{n=1}^{M} P(y^{(m)})^s \left(1 - P(y^{(m)})\right)^{n-s}, & \text{for } s = \overline{n}+1, \ldots, n \end{cases} \tag{38}$$

Proof. The proof is given in Appendix A.

Corollary 1 (The binary case). For the binary sample we obtain

$$\nu(s) = \begin{cases} 0 & s = 0, \ldots, \overline{n} \\ \binom{n}{s} \left(p^s q^{n-s} + q^s p^{n-s}\right) & s = \overline{n}+1, \ldots, n. \end{cases} \tag{39}$$

As we stressed, for vector samples ($\ell > 1$) the median vector may not be equal to any sample vector. The question arises, how likely it is that the median vector *does* belong to the sample. For i.i.d. finite-support vector samples whose sample vectors *have independent components* we now derive the probability that there exists at least one median vector among n sample vectors. Since we will be dealing here with vector samples in $\mathbb{R}^\ell$ with various ℓ, we index the samples with their vectors lengths, i.e., $\mathbf{X}_\ell$ denotes a sample of ℓ-element vectors. Denote by $\mathfrak{M}_\ell$ the number of median vectors in $\mathbf{X}_\ell$ and by $\boldsymbol{\nu}_\ell$ its distribution function, i.e.,

$$\boldsymbol{\nu}_\ell(z) \stackrel{\text{def}}{=} \mathcal{P}\{\mathfrak{M}_\ell = z\}, \quad z = 0, \ldots, n. \tag{40}$$

Proposition 5 (Distribution of the number of median vectors for finite-support independent component vector samples). *The distribution of the number of median vectors in the sample is for $\ell = 2, 3, \ldots$ given recursively by*

$$\begin{aligned} \boldsymbol{\nu}_1(z) &= \nu(z), \quad z = 0, \ldots, n \\ \boldsymbol{\nu}_\ell(z) &= \sum_{z'=z}^{n} \binom{z'}{z} \boldsymbol{\nu}_{\ell-1}(z') \sum_{s=z}^{n-z'+z} \nu(s) \frac{\binom{n-z'}{s-z}}{\binom{n}{s}}, \\ & \qquad z = 0, \ldots, n, \quad \ell = 2, 3, \ldots \end{aligned} \tag{41}$$

Proof. The proof is presented in Appendix B.

The main problem to overcome is the dependence of random variables $\mathfrak{M}_\ell$ both for different n and for different ℓ. The former results from a possibility of changing the sample median by any sample vector, and the latter is caused by the dependence between the median vector components.

Corollary 2 (Binary case). For a binary sample with $Y = \{0, 1\}$, $P(0) = q$, $P(1) = p$, Eqs. (41) simplifies to

$$\begin{aligned} \boldsymbol{\nu}_\ell(z) = & \sum_{z'=z}^{\min(n, z+\overline{n})} \binom{z'}{z} \boldsymbol{\nu}_{\ell-1}(z') \\ & \sum_{s=\max(\overline{n}+1, z)}^{n-z'+z} \binom{n-z'}{s-z} \left(p^s q^{n-s} + q^s p^{n-s}\right). \end{aligned} \tag{42}$$

We now can easily calculate the probability of the presence of at least one median vector in the sample, which is equal to $1 - \boldsymbol{\nu}_\ell(0)$.

We illustrate the results for the vector binary case. For binary samples, the number of median vectors strongly depends on $P(1) = p$. Exemplary results are shown in Fig. 4 for a binary sample of a fixed size $n = 15$ and several vector lengths ℓ. For $0.3 \le p \le 0.7$, the median vectors cease to exist in samples in $\mathbb{R}^\ell$, $\ell > 30$, while for $p = 0.05$ or $p = 0.95$ they still exist with probability > 0.7 for $\ell = 50$.

Another view of the same results is shown in Fig. 5, where the probability $1 - \boldsymbol{\nu}_\ell(0)$ that the sample contains any median vectors is plotted versus the vector length ℓ, for two probabilities of the success: $p = 0.5$ (Fig. 5a) and

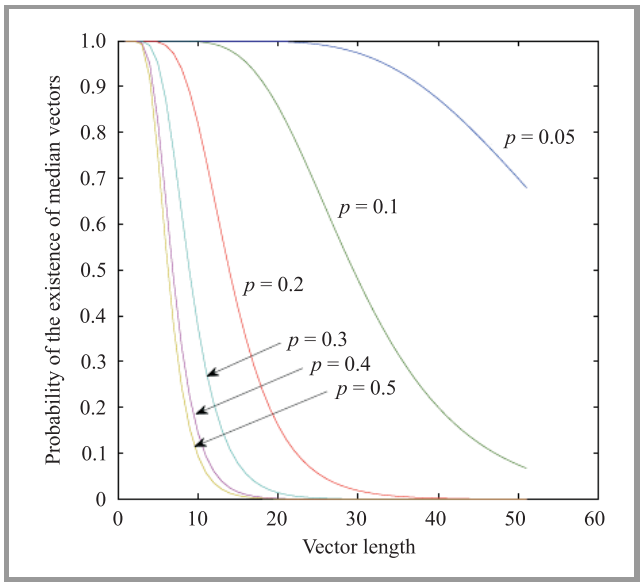

Fig. 4. Probability of the presence of the median vector in a $n = 15$-element binary sample versus the space dimension ℓ, with the probability of success p as a parameter, $p \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

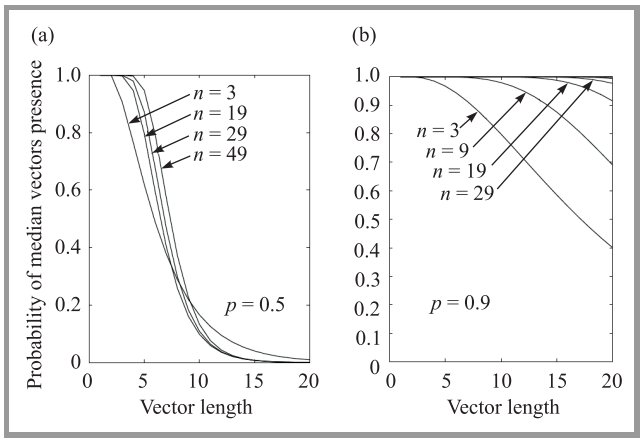

Fig. 5. Probability that there exist median vectors in the binary vector sample versus vector length ℓ, for the probability of success $p = 0.5$ and sample sizes $n \in \{3, 19, 29, 49\}$ (a), and for $p = 0.9$ and $n \in \{3, 9, 19, 29\}$ (b).

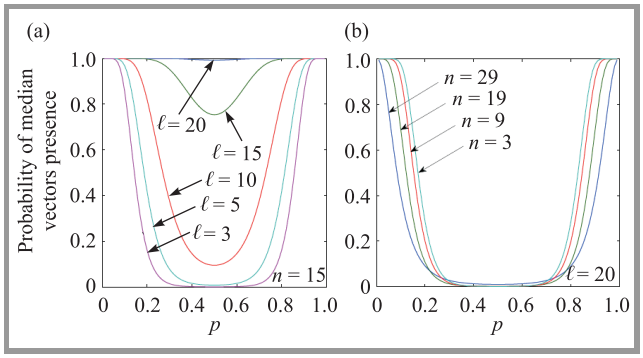

Fig. 6. Probability that there exist median vectors in a binary vector sample versus the probability of success p for the sample size $n = 15$ with vector lengths $\ell \in \{3, 5, 10, 15, 20\}$ (a) and for the vector length $\ell = 20$ and sample sizes $n \in \{3, 9, 19, 29\}$ (b).

$p = 0.9$ (Fig. 5b), each for several sample sizes n. The chance that the median vector exists quickly goes to zero with the increase of the vector length, and the velocity of the decrease grows as p get closer to $1/2$.

The last phenomenon is very well visible in yet another visualization of the results (Fig. 6) where the probability $1 - \boldsymbol{\nu}_\ell(0)$ of the median vectors presence in a binary vector sample is plotted versus the probability of success p, for a fixed sample size $n = 15$ and several vector lengths ℓ (Fig. 6a), and for a fixed vector length $\ell = 20$ and several sample sizes n (Fig. 6b). The probability quickly increases as $|p - 0.5|$ increases, and for each p it increases both with ℓ and n. Note that the characteristics (Fig. 6b and Fig. 5a) may intersect. This means that the dependence of the probability of n may not be monotonic for lower n.

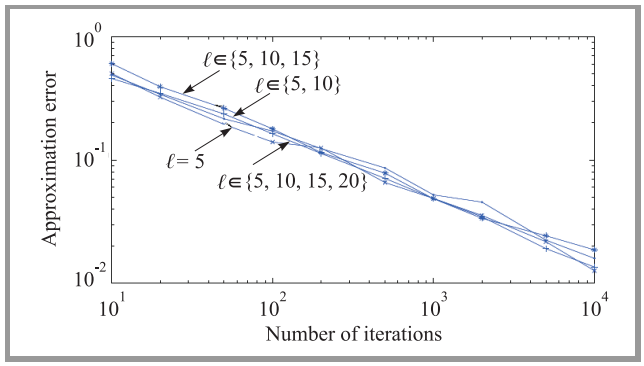

Fig. 7. The absolute difference between the simulated and theoretical distribution values versus number of iterations; logarithmic scales, averaged over ℓ and n.

An approximation of the discussed distribution can certainly be brought in by direct simulations. In our experiments, it was necessary to use about 10000 repetitions of the entire sample to obtain the simulation error of order of 0.01. The experiments show (Fig. 7) that in logarithmic scales the simulation error decreases almost linearly with the number of sample repetitions, which comes down to an exponential decrease of the simulation error with the number of repetitions. A decrease of the error by an order of one requires the increase in the number of repetitions by order of about one and a half.

6. L_1 Version of Template-RS

We now consider the definition of template-RS. In L_1 one can rewrite the defining formula (15) to the form

$$\mathbf{x}^{*RS} = \arg \min_{\mathbf{x}^{(i)} \in \mathbf{X}} \sum_{j=1}^{\ell} |x_j^{(i)} - \mathrm{me}_j| , \tag{43}$$

where $\mathrm{me}_j = \mathrm{me}(X_j)$ denotes jth component of $\mathbf{me}(\mathbf{X})$. We now consider odd-size *binary samples* in L_1 and compare

the definitions of template-RS and template-S. The latter is unique and by Eq. (8) equal to

$$\mathbf{x}^{*S} = \arg\min_{\mathbf{x}^{(k)}\in\mathbf{X}} \overline{D}_1(\mathbf{x}^{(k)},\mathbf{X}). \tag{44}$$

Since for binary vectors $\|\mathbf{a}-\mathbf{b}\|_1 = \|\mathbf{a}-\mathbf{b}\|_2^2$ hence $\overline{D}_1(\mathbf{x}^{(k)},\mathbf{X})$ can be rewritten as

$$\begin{aligned}\overline{D}_1(\mathbf{x}^{(k)},\mathbf{X}) &= \frac{1}{n}\sum_{\mathbf{x}^{(i)}\in\mathbf{X}} \left\|\mathbf{x}^{(k)}-\mathbf{x}^{(i)}\right\|_2^2 \\ &= \frac{1}{n}\sum_{\mathbf{x}^{(i)}\in\mathbf{X}} \left\|(\mathbf{x}^{(k)}-\mathbf{me}(\mathbf{X}))-(\mathbf{x}^{(i)}-\mathbf{me}(\mathbf{X}))\right\|_2^2 \\ &= \left\|\mathbf{x}^{(k)}-\mathbf{me}(\mathbf{X})\right\|_1 + \frac{1}{n}\sum_{\mathbf{x}^{(i)}\in\mathbf{X}} \left\|\mathbf{x}^{(i)}-\mathbf{me}(\mathbf{X})\right\|_1 \\ &\quad -\frac{2}{n}\sum_{j=1}^{\ell}(x_j^{(k)}-\mathrm{me}_j)\sum_{\mathbf{x}^{(i)}\in\mathbf{X}}(x_j^{(i)}-\mathrm{me}_j). \end{aligned}\tag{45}$$

Joining the first and the last terms we may write $\overline{D}_1$ in the form

$$\begin{aligned}\overline{D}_1(\mathbf{x}^{(k)},\mathbf{X}) = \sum_{j=1}^{\ell}\left|x_j^{(k)}-\mathrm{me}_j\right|\Big(1-\frac{2}{n}\mathrm{sign}(x_j^{(k)}-\mathrm{me}_j) \\ \sum_{x_j^{(i)}\in X_j}(x_j^{(i)}-\mathrm{me}_j)\Big) + \frac{1}{n}\sum_{\mathbf{x}^{(i)}\in\mathbf{X}}\left\|\mathbf{x}^{(i)}-\mathbf{me}(\mathbf{X})\right\|_1. \end{aligned}\tag{46}$$

Since the last term does not depend on $\mathbf{x}^{(k)}$ we finally may write

$$\mathbf{x}^{*S} = \arg\min_{\mathbf{x}^{(k)}\in\mathbf{X}} \sum_{j=1}^{\ell} w_j^k \left|x_j^{(k)}-\mathrm{me}_j\right|, \tag{47}$$

where

$$w_j^k = 1-\frac{2}{n}\mathrm{sign}(x_j^{(k)}-\mathrm{me}_j)\sum_{x_j^{(i)}\in X_j}(x_j^{(i)}-\mathrm{me}_j). \tag{48}$$

It is easy to show that $\mathrm{sign}(w_j^k)$ is always equal to 1. In fact, since the absolute value of the sum is not greater than the sum of absolute values, and for scalar binary samples there must be more elements equal to the median than those nonequal, we have

$$\begin{aligned}\left|\frac{2}{n}\mathrm{sign}(x_j^{(k)}-\mathrm{me}_j)\sum_{x_j^{(i)}\in X_j}(x_j^{(i)}-\mathrm{me}_j)\right| \\ \le \frac{2}{n}\sum_{x_j^{(i)}\in X_j}|x_j^{(i)}-\mathrm{me}_j| < 1 \end{aligned}\tag{49}$$

hence $\mathrm{sign}(w_j^k)=1$. Considering Eq. (43) and Eq. (47) as linear programming problems with respect to the variables $|x_j^{(k)}-\mathrm{me}_j|$, we see that the solutions of both problems are identical. In the other words, for vector binary samples, the definitions of template-R and template-RS lead to the same template. Note that we did not make any assumptions about independence of the components of sample elements.

7. L_1 Version of Template-T

In L_1, template-T can be by Eqs. (19) and (21) rewritten for each sample $\mathbf{X}$ and each (unknown) distribution parameter vector $\boldsymbol{\vartheta}$ in the form

$$\mathbf{x}^{*T} = \begin{bmatrix} g_1^*(\mathbf{X}) \\ \vdots \\ g_\ell^*(\mathbf{X}) \end{bmatrix}, \quad \text{where}$$

$$g_j^* \in \arg\min_{g_j\in\mathcal{G}} \mathcal{E}_{\mathbf{X},\boldsymbol{\vartheta}}\left|x_j^0(\boldsymbol{\vartheta}) - g_j(\mathbf{X})\right|. \tag{50}$$

The minimum is attained by the (non-random) median of the (conditional) distribution of $x_j^0(\boldsymbol{\vartheta})$, hence $g_j^*(\mathbf{X})$ should approximate this value. We will employ the component sample median $\mathrm{me}_j = \mathrm{me}(X_j)$ to estimate $g_j^*(\mathbf{X})$, and thus take the sample median to estimate $\mathbf{g}^*(\mathbf{X})$, namely

$$\mathbf{x}^{*T} \approx \begin{bmatrix} \mathrm{me}_1 \\ \vdots \\ \mathrm{me}_\ell \end{bmatrix}. \tag{51}$$

Note yet that for dependent components of sample elements, some information about component sample medians is contained also in other component samples, hence the solution (51) is suboptimal. In fact, for independent component vector samples, $\mathbf{x}^{*T}$ obtained here is identical to $\mathbf{x}^{*R}$.

8. Template Definitions in L_2

8.1. Template-R in L_2

As earlier noticed, definition of template-R in L_p comes down to a series of minimization subproblems for scalar samples. However, this feature is not needed to derive $\mathbf{x}^{*R}$ in L_2, since here we have just the classical issue of least squares: find a vector $\mathbf{x}^{*R}$ whose average squared distance to all other sample vectors is minimized

$$\mathbf{x}^{*R} = \arg\min_{\mathbf{x}\in\mathbb{R}} \overline{D}_2(\mathbf{x},\mathbf{X}), \tag{52}$$

where

$$\overline{D}_2(\mathbf{x},\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}^{(i)}-\mathbf{x}\|^2. \tag{53}$$

This is solved in a standard way by adding and subtracting the sample average

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}^{(i)} \tag{54}$$

to the terms inside the norm, so taking into account that the sum of the product term is equal to zero, we obtain

$$\overline{D}_2(\mathbf{x},\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}^{(i)}-\overline{\mathbf{x}})-(\mathbf{x}-\overline{\mathbf{x}})\|^2$$
$$= \|\mathbf{x}-\overline{\mathbf{x}}\|^2+\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}^{(i)}-\overline{\mathbf{x}}\|^2. \quad (55)$$

Readily, for any sample, irrespectively of any assumptions about independence of sample vector components, one obtains

$$\mathbf{x}^{*R}=\overline{\mathbf{x}}. \quad (56)$$

8.2. Template-S and Template-RS in L_2

Definition of template-S calls for minimization over the selected points of the vector space, namely

$$\mathbf{x}^{*S}=\arg\min_{\mathbf{x}\in\mathbf{X}}\overline{D}_2(\mathbf{x},\mathbf{X})=\arg\min_{\mathbf{x}\in\mathbf{X}}\|\mathbf{x}-\overline{\mathbf{x}}\|^2. \quad (57)$$

It is certainly unlikely that - even for finite support samples – the sample average is equal to any sample vectors. Note that in L_2, template-S Eq. (57) and template-RS: $\mathbf{x}_2^{*RS}\in\operatorname{Arg}\min_{\mathbf{x}\in\mathbf{X}}\|\mathbf{x}-\overline{\mathbf{x}}\|^2$ are equivalent, irrespectively of any independence conditions.

8.3. Template-T in L_2

In L_2, we rewrite template-T similarly as in L_1, namely

$$\mathbf{g}^*=\arg\min_{\mathbf{g}\in\mathcal{G}}\mathcal{E}_{\mathbf{X},\boldsymbol{\vartheta}}\|\mathbf{x}^{\mathbf{0}}(\boldsymbol{\vartheta})-\mathbf{g}(\mathbf{X})\|^2. \quad (58)$$

The minimum is attained by the mean value of the (non-random) mean value of the (conditional) distribution of $\mathbf{x}^{\mathbf{0}}(\boldsymbol{\vartheta})$, so $\mathbf{g}^*(\mathbf{X})$ should approximate this value. Employing the sample average $\overline{\mathbf{x}}$ to estimate the mean value of $\mathbf{x}^{\mathbf{0}}(\boldsymbol{\vartheta})$, we obtain

$$\mathbf{x}^{*T}\approx\overline{\mathbf{x}}. \quad (59)$$

For dependent components of sample vectors, the information about the conditional mean value is contained also in other components, hence the solution can be improved. In other words, the solution (59) is identical to $\mathbf{x}^{*R}$ for independent component samples.

9. Conclusions

Our analysis of the art of template creation only touches the problem of choosing "the best representative" of biometric samples. We discussed only the problems characterized by measurements that could be viewed as points of a metric space, and if so, the metric was assumed to be Euclidean. The problem in general touches the notion of information contents of biometric measurement systems. Even if one desires to assume that the biometric measurements lead to Euclidean spaces, there still are various possibilities of choosing the "best representative". Intuitively, such the representative must express some "stable" properties of the measurements for a single subject and as such, it may strongly depend on the biological quantities under scrutiny. Consequently, choosing the template calls for a thorough knowledge of the biological context. On the other hand, apart from this context, one may choose the template on the base of one of "black box" solutions and choose the solution that works best for the given biometric database(s). In the paper we in fact analyzed several "black box" solutions to show their properties and determine their mutual relations.

The concepts we analyzed were based on L_1 and L_2 distances between the measurements. The possibilities we examined included the template as an enrollment measurement that is on the average closest to all other enrollment measurements (template-S), and a vector (not necessarily any enrollment set vector) closest on the average to all enrollment measurements (template-R). Since the latter is not necessarily the enrollment vector, we may treat it as a reference measurement, and define the template as the enrollment vector closest to the reference (template-RS). Finally, we also introduced the template that aimed into minimization of the distance between the template and a test measurement (template-T). Each of those concepts was analyzed with the use of L_1 and L_2 distances, so eight versions of the template were investigated.

We investigated closer the L_1 concepts, since they are less known. We showed, using independent component binary samples, that template-S differs from template-R, and the difference grows with the dimension of the sample vectors. Also, the difference grows as p approaches 0.5. This suggests that in general, the difference between template-S and template-R is higher for the underlying (population) distributions of higher entropy. We also showed that for binary samples template-RS is identical to template-S.

One may notice that as the enrollment sample size grows, all the concepts considered here may lead to either the subject's theoretical median or the subject's theoretical expected value. The templates based on samples of finite size can be thus treated as various estimators of subject's theoretical characteristics. They may strongly differ from the theoretical characteristics because the enrolment sample size for a single subject can be very low (as low as three measurements). On the other hand, the dimension of the measurements can be very high, since it must have the information contents high enough to differentiate between many subjects of large biometric many-subject databases. Note that if the number of the subjects grow, as in attempts to build universal identity verifiers, the templates considered here must also be based on growing enrollment subject's sample sizes, to be as close as possible to the theoretics subjects' characteristics. The question remains open if these characteristics have sufficient information contents, and even what the underlying theoretical subject and the entire population distributions are. We only hope that the assumption about the very existence of these distributions, or – in other words – on the possibility of describing the biological variability in terms of probabilities, holds.

Appendix A

Proof of Proposition 4

To prove Proposition 4 for finite-support scalar samples X we first derive the joint probability function of the sample median $\mathrm{me}(X)$ and the number of median elements $\mathcal{M}$

$$\rho(y,s) \stackrel{\mathrm{def}}{=} \mathcal{P}\{\mathrm{me}(X)=y \wedge \mathcal{M}=s\}, \quad \text{where} \quad y=y^{(1)},\dots,y^{(M)}, \quad s=0,\dots,n. \tag{60}$$

Certainly, $\mu(y)=\sum_{s=0}^{n}\rho(y,s)$, for $y=y^{(1)},\dots,y^{(M)}$ and $\nu(s)=\sum_{m=1}^{M}\rho(y^{(m)},s)$, for $s=0,\dots,n$. Assuming that the sample size n is odd, we will find the joint probability function $\rho(y,s)$ for $s=0,\dots,n$, $y=y^{(1)},\dots,y^{(M)}$.

Proposition 6 (Joint distribution of the sample median and the number of median elements for scalar samples). *For n-element discrete support scalar sample, the joint probability function of the sample median and the number of median elements is given by*

$$\rho(y,s)=\begin{cases} 0, & \text{for } s=0,\dots,\overline{n}, \quad y=y^{(1)}\vee y=y^{(M)} \\ \binom{n}{s}P(y)^s\sum_{u=\overline{n}+1-s}^{\overline{n}}\binom{m-s}{u}F(y)^u S(y)^{n-s-u}, & \text{for } s=0,\dots,\overline{n}, \quad y=y^{(2)},\dots,y^{(M-1)} \\ \binom{n}{s}P(y)^s\big(1-P(y)\big)^{n-s}, & \text{for } s=\overline{n}+1,\dots,n, \end{cases} \tag{61}$$

where $\overline{n}=(n-1)/2$.

Proof. We first rearrange the summation in Eq. (34) to show the influence of the terms related to the number $s=n-u-w$ of the sample elements equal to the median, namely (Fig. 8)

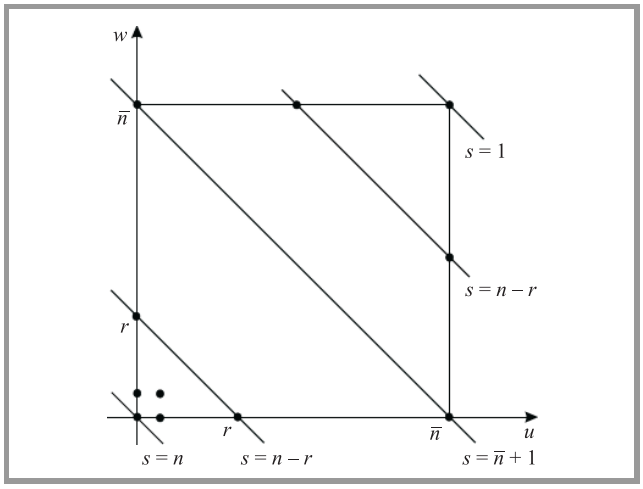

Fig. 8. Change of variables in Eq. (62).

$$\mu(y)=\begin{cases} \sum_{s=\overline{n}+1}^{n}\binom{n}{s}P(y)^s S(y)^{n-s}, & \text{for } y=y^{(1)} \\ \sum_{s=1}^{\overline{n}}\sum_{u=\overline{n}+1-s}^{\overline{n}}\binom{n}{s,u}P(y)^s F(y)^u S(y)^{n-s-u} + \sum_{s=\overline{n}+1}^{n}\sum_{u=0}^{n-s}\binom{n}{s,u}P(y)^s F(y)^u S(y)^{n-s-u}, & \text{for } y=y^{(2)},\dots,y^{(M-1)} \\ \sum_{s=\overline{n}+1}^{n}\binom{n}{s}P(y)^s F(y)^{n-s}, & \text{for } y=y^{(M)}. \end{cases} \tag{62}$$

We have

$$\rho(y,s)=\binom{n}{s}P(y)^s\begin{cases} S(y)^{n-s}, & \text{for } y=y^{(1)}, \quad s=\overline{n}+1,\dots,n \\ \sum_{u=\overline{n}+1-s}^{\overline{n}}\binom{n-s}{u}F(y)^u S(y)^{n-s-u}, & \text{for } y=y^{(2)},\dots,y^{(M-1)}, \quad s=1,\dots,\overline{n} \\ \sum_{u=0}^{n-s}\binom{n-s}{u}F(y)^u S(y)^{n-s-u}, & \text{for } y=y^{(2)},\dots,y^{(M-1)}, \quad s=\overline{n}+1,\dots,n \\ F(y)^{n-s}, & \text{for } y=y^{(M)}, \quad s=\overline{n}+1,\dots,n \\ 0 & \text{otherwise}, \end{cases} \tag{63}$$

hence

$$\rho(y,s)=\binom{n}{s}P(y)^s\begin{cases} \sum_{u=\overline{n}+1-s}^{\overline{n}}\binom{n-s}{u}F(y)^u S(y)^{n-s-u}, & \text{for } y=y^{(2)},\dots,y^{(M-1)}, \quad s=1,\dots,\overline{n} \\ \big(1-P(y)\big)^{n-s}, & \text{for } y=y^{(1)},\dots,y^{(M)}, \quad s=\overline{n}+1,\dots,n \\ 0 & \text{otherwise}, \end{cases} \tag{64}$$

so Eq. (61) follows.

Corollary 3 (The binary case). For a binary sample with $Y=\{0,1\}, P(0)=q$, $P(1)=p$ we have

$$\rho(y,s)=\begin{cases} 0, & s=0,\dots,\overline{n} \\ \binom{n}{s}q^s p^{n-s}, & y=0, \quad s=\overline{n}+1,\dots,n \\ \binom{n}{s}p^s q^{n-s}, & y=1, \quad s=\overline{n}+1,\dots,n. \end{cases} \tag{65}$$

Now, Proposition 4 can be easily proven.

Proof of Proposition 4. We can obtain the distribution of the number of median elements by summing up ρ in Eq. (61) over all y

$$\nu(s)=\sum_{y\in Y}\rho(y,s). \tag{66}$$

Readily,

$$\nu(s)=\binom{n}{s}\begin{cases}0, & \text{for } s=0\\ \sum\limits_{u=\overline{n}+1-s}^{\overline{n}}\binom{n-s}{u}\sum\limits_{m=2}^{M-1} P(y^{(m)})^s F(y^{(m)})^u S(y^{(m)})^{n-s-u}, & \text{for } s=1,\ldots,\overline{n}\\ \sum\limits_{m=1}^{M} P(y^{(m)})^s (1-P(y^{(m)}))^{n-s}, & \text{for } s=\overline{n}+1,\ldots,n\end{cases} \quad (67)$$

hence we obtain Eq. (38).

Appendix B

Proof of Proposition 5

Let $\mathfrak{M}_{1,\ldots,\ell}$ be the number of median vectors in ℓ-dimensional sample of size n, and denote by $\boldsymbol{\nu}_\ell(z)$ its probability function, namely

$$\boldsymbol{\nu}_\ell(z)\stackrel{\text{def}}{=}\mathcal{P}\{\mathfrak{M}_\ell=z\},\quad z=0,\ldots,n \quad (68)$$

for $\ell=1,\ldots$. We derive the probability function $\boldsymbol{\nu}_\ell$ recursively. Given an ℓ-dimensional sample $\mathbf{X}_\ell$ we form a $(\ell-1)$-dimensional sample $\mathbf{X}_{\ell-1}{}'$ by removing a single component (say the last one) of each vector in $\mathbf{X}_\ell$. We first calculate the conditional probability that there are exactly z median vectors in $\mathbf{X}_\ell$ given there are exactly z' median vectors in $\mathbf{X}_{\ell-1}{}'$ and s median elements in X, namely

$$p_\ell(z|z',s)\stackrel{\text{def}}{=}\mathcal{P}\big(\mathfrak{M}_\ell=z|\mathfrak{M}_{\ell-1}=z'\wedge\mathcal{M}=s\big). \quad (69)$$

It is easy to see that

$$p_\ell(z|z',s)=\begin{cases}\binom{z'}{z}\binom{n-z'}{s-z}\Big/\binom{n}{s}, & \text{for } z\le z' \text{ and } z'-z\le n-s \text{ and } z\le s\\ 0, & \text{for } z>z' \text{ or } z'-z>n-s \text{ or } z>s.\end{cases} \quad (70)$$

Note that $p_\ell(z|z',s)$ is null except for the points in the triangle $z'\ge z,\ s\ge z,\ s+z'\le s+z$ in the (s,z') plane. The distribution $\boldsymbol{\nu}_\ell(z)$ of $\mathfrak{M}_\ell$ can be thus be found by a summation of the conditional distribution Eq. (70) with respect to distributions: $\boldsymbol{\nu}_{\ell-1}$ of $\mathfrak{M}_{\ell-1}$ and ν of $\mathcal{M}$ of the two independent random variables. Therefore, $\boldsymbol{\nu}_\ell(z)$ can be determined recursively as

$$\begin{aligned}\boldsymbol{\nu}_1(z) &= \nu(z),\quad z=1,\ldots n\\ \boldsymbol{\nu}_\ell(z) &= \sum_{z'=0}^{n}\boldsymbol{\nu}_{\ell-1}(z')\sum_{s=0}^{n}p_\ell(z|z',s)\,\nu(s),\quad z=0,\ldots,n,\\ &\quad \ell=2,3,\ldots\end{aligned} \quad (71)$$

Plugging Eq. (70) into the above we obtain Eq. (41).

For the binary sample with $Y=\{0,1\}$, $P(1)=p$, $P(0)=q$, $\rho(s)$ is given by Eq. (39). Since this is equal to zero for $s\le\overline{n}$, the summation in Eq. (71) narrows down to the triangle $z'\ge z,\ s\ge\overline{n}+1,\ s+z'\le s+z$ for $z\le n$, and the triangle $z'\ge z,\ s\ge z,\ s+z'\le s+z$ otherwise. Consequently, Eq. (71) simplifies to

$$\boldsymbol{\nu}_\ell(z)=\begin{cases}\sum\limits_{z'=z}^{z+\overline{n}}\binom{z'}{z}\boldsymbol{\nu}_{\ell-1}(z')\sum\limits_{s=\overline{n}+1}^{n-z'+z}\binom{n-z'}{s-z}(p^sq^{n-s}+q^sp^{n-s}), & \text{for } z=0,\ldots,\overline{n}\\ \sum\limits_{z'=z}^{n}\binom{z'}{z}\boldsymbol{\nu}_{\ell-1}(z')\sum\limits_{s=z}^{n-z'+z}\binom{n-z'}{s-z}(p^sq^{n-s}+q^sp^{n-s}), & \text{for } z=\overline{n}+1,\ldots,n\end{cases} \quad (72)$$

and Eq. (42) follows.

Acknowledgements

This paper has been financed by the Ministry of Science and Higher Education grant OR00 0026 07 "A platform of secure biometrics implementations in personal verification and identification".

Reference

[1] D. L. Evans, L. M. Leemis, and J. H. Drew, "The distribution of order statistics for discrete random variables with applications to bootstrapping", *Informs J. Comput.*, vol. 18, no. 1, pp. 19–30, 2006.

Andrzej Pacut holds Ph.D. in electronics and D.Sc. in control and robotics. Since 1969 he is with Warsaw University of Technology, being presently a professor and the head of Biometrics and Machine Learning Group in the Institute of Control and Computation Engineering. Since 2001 he is also the head of Biometric Laboratory of Research and Academic Computer Network NASK. He was Visiting Prof. in the Lefschetz Center for Dynamic Systems at Brown University, Providence, Rhode Island 1980–1981, and Visiting Prof. in the Department of Electrical and Computer Engineering of Oregon State University, Corvallis, Oregon, 1986–1991. He is a senior member of the IEEE and served as the Vice-Chair and then Chair of Poland Section of the IEEE in 2002–2009. He is the head of Technical Committee 309 on Biometrics of Polish Committee for Standardization PKN. He is interested in learning systems, biometrics, identification, bio-inspired modeling and control, and related areas.
e-mail: A.Pacut@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Biometric Laboratories
Research and Academic Computer Network NASK
Wąwozowa st 18
02-796 Warsaw, Poland