 doi:10.20944/preprints201905.0295.v2

Optinalysis: A New Approach of Symmetry Detection and Similarity Measurement through a Looking-Glass

*Kabir Bindawa Abdullahi

Department of Biology, Faculty of Natural and Applied Sciences, Umaru Musa Yar'adua University, P.M.B., 2218 Katsina, Katsina State, Nigeria.
*Correspondence: kabir.abdullahi@umyu.edu.ng ; kabirnamallam@gmail.com
(+2348065995423)

Abstract

Optinalysis, as a method of symmetry detection, is a new algorithm that intrametrically (within elements or variables) or intermetrically (between elements or variables) computes and compares two or more univariate or multi-clustered or multivariate sequences as a mirror-like reflection of each other (optics-like manner), hence the name is driven. Optinalysis is based by the principles of reflection and moment about a symmetrical line which detects symmetry that reflects a similarity measurement. This proposed methodology was validated in comparison with Pearson method of skewness detection, and also with some algorithms for pairewise alignment and comparison of genomic sequences (Needle, Stretcher, Water, Matcher) on EMBL-EBI website. A results comparison shows that optinalysis is more advance, more sensitive, more inferential and simple alternative approach of skewness detection and pairewise sequence comparison.

Keywords: Sequence; Correspondence; Symmetry; Similarity: Kabirian Coefficient.

 doi:10.20944/preprints201905.0295.v2

Introduction

Natural and man-made structural entities and objects are everywhere, the information about these interesting structures and objects are routinely gathered and collected all around us, we appreciates the beauty of their nature, shapes, patterns and orientations. We recognize, identify, compare and distinguish amongst them by our innate senses. We often make rational decisions about these structural entities and objects base on their symmetry and structural orientations. Therefore, measure of symmetry is of great and global concern and for a wider interest in a variety of disciplines with theoretical concepts in mathematics and statistics; practical applications in biology, chemistry, medicine, image analysis, archaeology, bioinformatics, geology, particle science, genetics, geography, law, pharmacy and physiotherapy (Goodall, 1991; Bookstein, 1991; Dryden and Mardia, 1998; Cootes and Taylor 2010; Dryden, and Mardia, 2016; and Zheng *et al.*, 2017). Lines and points as components of a geometric concept were established and invented by Mathematicians. Symmetry, on the other hand, is everywhere around us. Almost all living creatures such as plants, animals, and even humans are symmetric to a certain degree of geometry (Dryden and Mardia, 2016).

In the literal texts, going from Weyl (1952); Darvas (2007), a widely accepted general definition of symmetry is not claimed coverable by a single mathematical definition and there is much to learn and to explore before stating whether or not a unique definition is possible. Even the practical definitions of symmetry are often based on strong assumptions and exemplified rather than defined (Petitjean, 2007). However, strong assumptions, such as the existence of the euclidean structure for geometric symmetries, Riemannian distance, Minkowski distance, Mahalanobis distance, simple matching and Jaccard coefficient are some measures of similarity.

Similar and symmetrical entities are invariance to transformational properties such as reflection, rotation, scaling, and translation. The decisions we made about this invariance under different transformations are based on strong assumption with no general formula to prove and explain. Petitjean (2007) associated the topic of symmetry with the classification of symmetries, which should be done on the basis of the symmetry group structure of the object and symmetry is considered as a quantity varying continuously.

In this paper, a new algorithm called Optinalysis, is proposed and explained. Optinalysis torches the most important aspects of statistical inferences on geometrical shapes and sequence comparisons. Optinalysis does not require assumption of normality, but it requires the existence or establishment of a clearly defined sequence order within and/or between the elements or variables of a sequence(s). Several examples were examined and analyzed, and in comparison with other standard methods, revealed that Optinalysis presented a uniquely new paradigm of sequence data analysis of univariate or multi-clustered or multivariate observations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

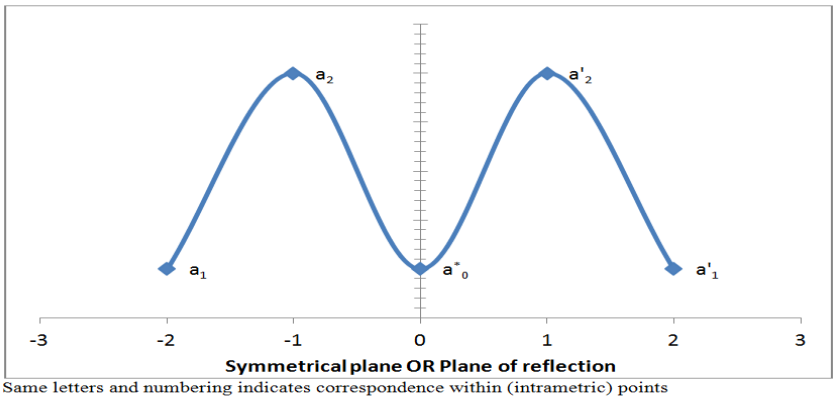
1.0 Theoretical Justification for the Algorithm of Optinalysis

The paradigm of the concept of symmetry is the mirror image. My mirror image and I are symmetric pair to each other by their corresponding points that matches within and between them. Other kinds of symmetry exist, but this is the one to start with. It is called *isosymmetry*.

To make this concept illustrative and precise, consider the case of one *M* letter (Figure 1) in a plane. Landmarks (elements) of letter *M* $(a_n \; and \; a'_n)$ may be related as follows: there is a straight line that separates these landmarks (*the line of reflection*) and each point (a_n) within *M* can be connected to a corresponding point (a'_n) within *M*. The correspondence connects all points $(a_n$ and $a'_n)$ in *M* such that corresponding points are equidistant from the line of reflection. Mardia *et al.* (2000) define this symmetry as object symmetry, and is also referred in this paper *intrametric symmetry or shape symmetry*.

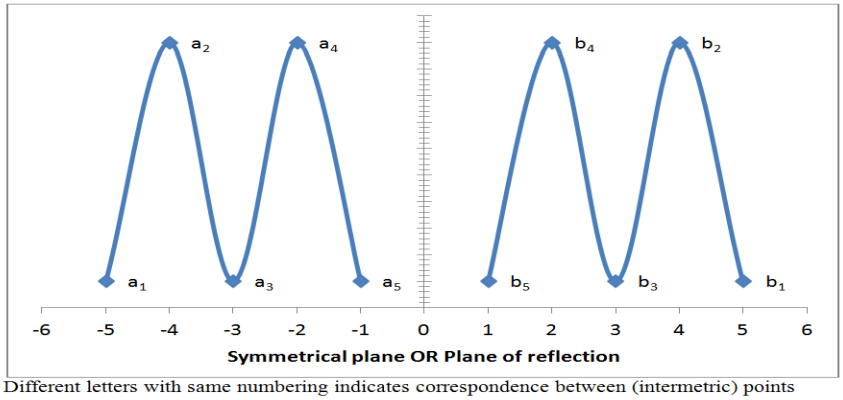
In another illustrative case of two *M* letters (Figure 2) in a plane. Letter M_1 and M_2 may be related as follows: there is a straight line that separates them (*the line of reflection*) and each point (a_n) in M_1 can be connected to a corresponding point (b_n) in M_2. The correspondence connects all points $(a_n$ and $b_n)$ in M_1 and M_2 and is such that corresponding points are equidistant from the line of reflection. Mardia *et al.* (2000) define this symmetry as matching symmetry, and is also referred in this paper *intermetric symmetry or comparative symmetry*.

In space, the definition is similar, but with a *plane of reflection.* Scholarly works such as Weyl (1952), Darvas (2007), Kendall (1984), Watson (1986), Bookstein 1986, 1991); Fraasen and Bsa (1989); Kent (1994), Lele and Richtsmeier (1991) and Dryden *et al.* (2008) follows same line with this principle.



Same letters and numbering indicates correspondence within (intrametric) points

Figure 1: Showing a symmetric correspondence within pair points of letter *M*.



Different letters with same numbering indicates correspondence between (intermetric) points

Figure 2: Showing a symmetric correspondence between pair points of two *M* letters.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

The Algorithm of Optinalysis is within this tradition and concept of symmetry. Optinalysis attempted to detects symmetry within and/or between the corresponding points of pair of structural elements or variables of sequences. Optinalysis is however designed to intrametrically or intermetrically compare two or more multi-clustered or multivariate sequences as a mirror-like reflection of each other (optics-like manner).

2.0 The principle of Optinalysis is Reflection and Moment

All symmetrical structures reflect momentarily (i.e, in same moment or in same total moments) about a symmetrical plane/point. Reflection can be: (1) Normal reflection, characterized by a plane mirror reflection (equidistance reflection from the central node), (2) Re-scaled reflection, characterized by the reduction in magnitude and increase in displacement or increase in magnitude and reduction in displacement. Therefore, reflection and moment are two companion mechanisms upon which the principle of Optinalysis operates.

If the query moment is equal to the reflector moment, then two comparing entities are geometrically and statistically symmetrical intrametrically.

$$a_n \times D_n = a'_n \times D_n$$

And/or if the total query moments is equal to the total reflector moments, then two comparing entities are geometrically and statistically symmetrical intrametrically.

$$\sum(a_n \times D_n) = \sum (a'_n \times D_n)$$

If the query moment is equal to the reflector moment, then two comparing entities are geometrically and statistically symmetrical intermetrically.

$$a_n \times D_n = b_n \times D_n$$

And/or the total query moments is equal to the total reflector moments, then two comparing entities are geometrically and statistically symmetrical intermetrically.

$$\sum(a_n \times D_n) = \sum(b_n \times D_n)$$

Suppose we refer to Figure 1-2, we find that intrametrically (within the sequence elements or variables),

a_1 is normally reflected momentarily about x-plane as a'_1
a_2 is normally reflected momentarily about x-plane as a'_2

And also intermetrically (between the sequence elements or variables),

a_1 is normally reflected momentarily about y-plane as b_1
a_2 is normally reflected momentarily about y-plane as b_2
a_3 is normally reflected momentarily about y-plane as b_3
a_4 is normally reflected momentarily about y-plane as b_4
a_5 is normally reflected momentarily about y-plane as b_5
a_6 is normally reflected momentarily about y-plane as b_6

In another case in Figure 3, we can find that intrametrically (within the sequence elements or variables),

a_1 is spherically reflected momentarily about y-plane as b_3
a_2 is spherically reflected momentarily about y-plane as b_4
a_3 is spherically reflected momentarily about y-plane as b_1
a_4 is spherically reflected momentarily about y-plane as b_2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

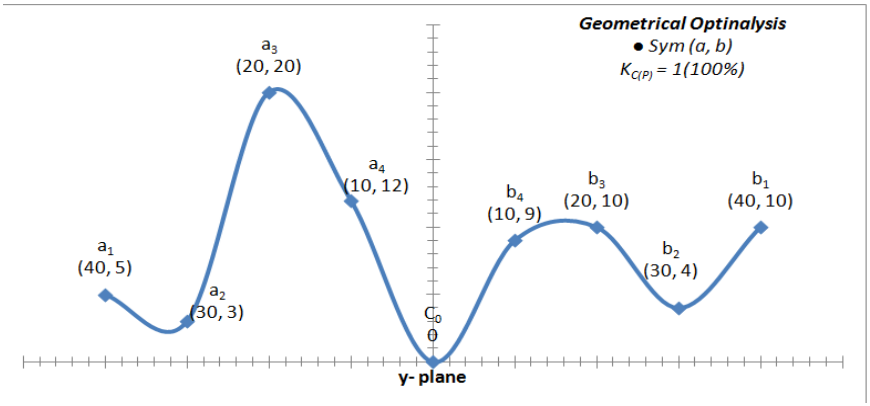


Figure 3: A pseudo-symmetrical distribution with a spherically reflected elements

3.0 Terms used and constructed

3.1 Quantitative scale: (denoted by $r_1, r_2, r_3 \ldots\ldots\ldots r_n$) are numbers arbitrarily assigned to rank every specific point, called the node, of a sequence, in a very logical manner, in such a way that every node has its own unique characteristic sensitivity to a changing magnitude. The symmetric status of a given shaped sequence remains invariant under any quantitative scaling provided that a uniform difference (common difference) is maintained between each scale point to its proceeding point (See Figure 4, and Table 1-2).

3.2 Elements or Variables: (denoted by $a_1, a_2, a_3, \ldots\ldots\ldots a_n$; $b_1, b_2, b_3 \ldots\ldots\ldots b_n$) refer to the main components of a sequence (See Figure 4, and Table 1-2).

3.3 Scalements: Denoted by *'S_m'* it is expressed as the product of element or variable and its bearing quantitative scale (See Table 1-2).

3.4 Node: Denoted by *'n'*. A node comprised of any specific quantitative scale's units, its bearing element or variable (See Table 1-2).

3.4.1 Left-sided and Right-sided Nodes: Left-sided and right-sided nodes describe respectively the nodes on which the elements or variables of left-sided and right-sided sequences are organized. The left-sided and right-sided sequences describe respectively the sequence on which the components of left-sided and right-sided nodes are organized (See Table 1-2).

3.4.2 Pericentral Node: Denoted by *'P_n'*. It describes one of the left-sided or right-sided node that divide each of the component sequence (i.e, the left-sided and the right-sided sequence) into two equal halves. Pericentral node exists only if and only two sequences are paired intermetrically (See Table 1-2).

4.3.3 The Central Node: Denoted by *'C_n'* or *'$a*_n$'*. It describes that point of the symmetrical plane or axis. It is the midpoint that divides a sequence or two paired sequences into two equal halves (See Table 1-2).

3.5 Nodality: Denoted by *'N'* is the total number of existing nodes in sets of sequences. Nodality directly correlates with the number of elements or variables (See Table 1-2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

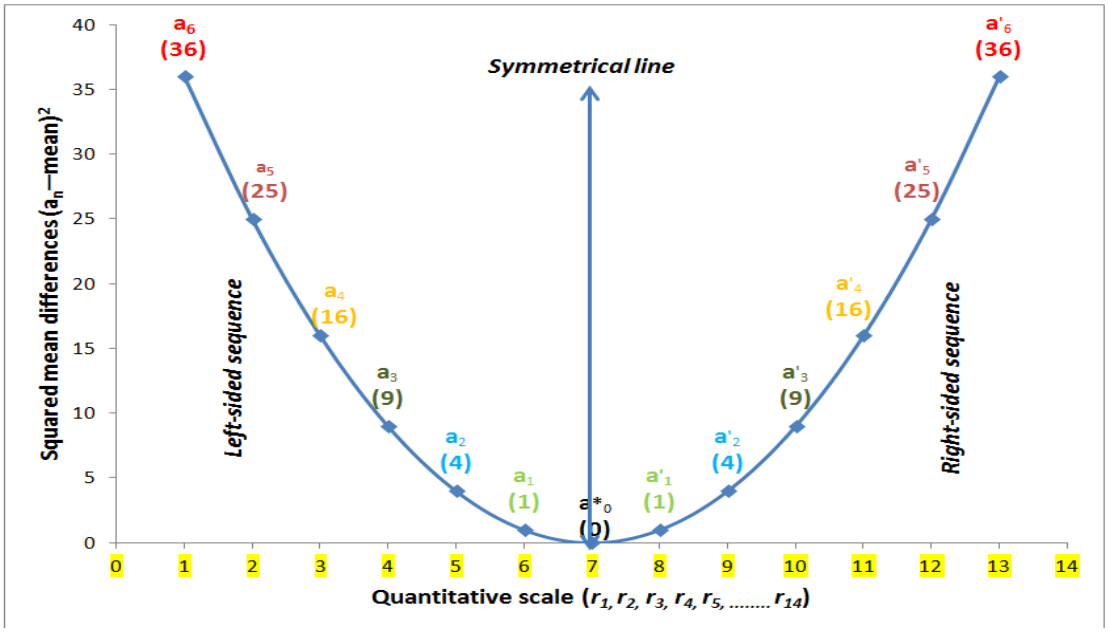


Figure 4: A symmetrical shape showing the spread of the data around a mean.
The same colored points show intrametric correspondence about a symmetrical line.

4.0 Optinalysis

Optinalysis is a new algorithm that intrametrically (within elements or variables) or intermetrically (between elements or variables) computes and compares two or more univariate or multi-clustered or multivariate sequences as a mirror-like reflection of each other (optics-like manner), hence the name is driven. Optinalysis is a useful tool for shape/pattern and comparative analysis. Intrametric optinalysis, also called shape optinalysis requires no pairing style to be chosen, because only one sequence is involved. But the intermetric optinalysis, also called computational optinalysis requires a suitable selection of a pairing style between the two sequences.

4.1 Step-by-step Guidelines to Optinalysis

Step-by-step guides to optinalysis are as follows:

Step 1: Identify the sequence data set(s) to be analyzed. Optinalysis welcomes all numerical data from any measurement scales. For nominal data, a suitable and appropriate transformation method need must be used to convert the nominal values to numerical values.

Step 2: Identify the elements or variables of the sequence(s) and establish or adopt any logical or empirical sequence order within the elements or variables. See further details in section item 5.1, 5.1.1, 5.1.2.

Step 3: Resolve the sequences using any suitable and appropriate resolution methods. See further details in section item 5.2.

Step 4: Assign symbolic annotations to the sequence(s) to show the head and tail of the sequence(s), and also the labeling of the sequence elements or variables. See further details in section item 3.2, 5.3.

Step 5: Select an appropriate pairing style if intermetric symmetry (symmetry/similarity detection between two independent sequences) is involved. For intrametric symmetry

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

detection (symmetry/similarity detection within a sequence or between two dependent parts of a sequence), no pairing style is required. See further details in section item 5.3.

Step 6: Select a controlled limit of normalization. Normalization can range from zero to any value. See further details in section item 5.4.

Step 7: Assign a quantitative scale to the sequence(s). See further details in section item 3.1.

Step 8: Using the suitable equations, compute the Kabirian coefficient of symmetry (similarity), and the probabilities or percentages. See further details in section item 6.0, 6.1.1, 6.1.2, 6.2, 6.3.

5.0 Further Details on Some Important Algorithmic Steps

5.1 Sequencing of the Data Set

Sequencing here refers to the adoption or establishing a logical and empirical order to a set of elements or variables.

5.1.1 Theoretical Sequence Order

This sequence order is based on the geometrical orientations, or theoretical explanations or natural phenomena. For instance, nucleotide base and amino acid sequences, systematic numbering of shape landmarks coordinates, chemical concentrations, rating and ranking responses of a questionnaire, and etc are some examples of a theoretical sequences. In this case, the position and pattern orientation of each element or variable of the attribute is preserve and kept in its natural order.

5.1.2 Ascending and Descending Sequence Order

In this case, the position and pattern orientation of all the random elements or variables of a given data set are reorganized in ascending or descending order. It disregards the inherent order of the random data set. This can be important for establishing an empirical sequence order to random univariate observations.

5.2 Resolution of univariate or multi-clustered or multivariate observations

Resolution of univariate and multi-clustered or multivariate observations are computed for the following reasons:

i. For constructing a shape or pattern to a shapeless sequence of univariate observations. To give a shape to a shapeless sequence of univariate observations, statistical functions such as mean differences of sequence, descaled mean differences of sequence, squared mean differences of sequence, and square root of squared mean differences of sequence can be used appropriately. Resolution by descaled mean differences is when all the scaling effect (positive and negative signs) is removed from the shaped sequence. Resolution by descaled mean difference is the same result as the square root of squared mean differences of sequence. Table A1-A2 of the appendix presented some worked examples.

ii. For simplification of repeated or replicated measurements of ordered sequence of multi-clustered or multivariate observations. To simplify repeated or replicated measurements of multi-clustered or multivariate observations, statistical functions such as variance, standard deviation, standard error of mean and etc can be used appropriately.

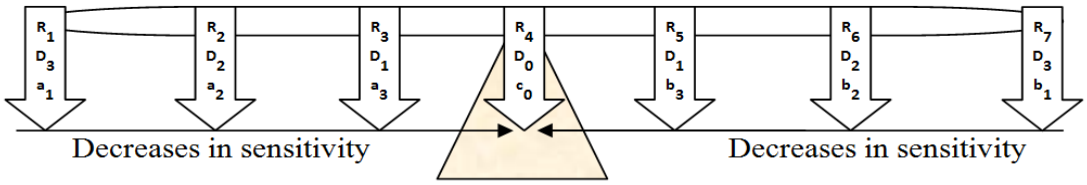
iii. For harmonizing the effect of co-factors of a structured or shaped distribution. Harmonization of co-factors' effect can be achieved appropriately by some functions such as differential moment resolution, differential surface area resolution, differential centroid size resolution, and etc.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

5.3 Existence of Sensitivity Points Necessitates for the Choice of Pairing Styles

Sensitivity point is any node that when considered a variable can exert a certain degree of imbalances in the distribution of elements (variables) about a dividing line or plane. Each node has its own unique characteristic sensitivity which increases away from the central node and decreases towards the central node(s). Sensitivity of a point generally decreased with increase in sequence elements. Figure 5 is an illustrative example.

The nodes with components R_1, D_3, a_1 and R_7, D_3, b_1 are the most sensitive points of the upper and lower stems respectively. The node with components $R_4D_0C_1$ is the central node.



Note: $\mathbf{R_n}$ = Quantitative scale; $\mathbf{D_n}$ = Displacements; $\mathbf{a_n}$ and $\mathbf{b_n}$ are paired variables/elements; $\mathbf{c_n}$ = Central variable

Figure 5: Sensitivity points of sequence elements.

Pairing style tells us how the sequences of two intermetric elements or variables pairewisely reflect. Sequences symmetry can be detected on different pairing style. The choice of appropriate pairing style depends on the consideration made on where (i.e, beginning or end of the sequence elements or variables) should be more sensitive to any imbalances/changes or otherwise.

5.3.1 Head-to-head Pairing (H-H): one ends of the two pairing sequences called the heads (the start point) are both allowed to be on the most sensitive node.

$$\overset{(\pm N)}{\underset{W\ or\ B}{\wedge}}:$$

5.3.2 Tail-to-tail Pairing (T-T): one ends of the two pairing sequences called the tails (the end point) are both allowed to be on the most sensitive node.

$$\overset{W\ or\ B}{\underset{(\pm N)}{\vee}}:$$

5.3.3 Head-to-tail Pairing (H-T) or Tail-to-head Pairing (T-H): one of the ends of the two pairing sequences called the head or tail (the start or end point) is allowed to be on the most sensitive node and other on the less sensitive node.

5.4 Normalization

Normalization refers to a deliberate positive or negative increase in magnitude of the central node of a given sequence distribution. A symmetrical distribution remains stable under any magnitude of central modulation. This explains that a symmetrical distribution is very flexible and stable to any limit of central modulation.

Asymmetrical distribution can be transformed symmetrical if the central node is positively or negatively modulated to a certain minimum magnitude called a normalization value ($\pm Nv$). Therefore, central modulation and normalization promotes unimodality and minimizes the skewness. See Figure 6-8 for visually illustrative examples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

OPTINALYSIS

$$\bigwedge_{w}^{(\pm Nv=0)} : \int_{\in(A\prime)}^{\in(A)} c(p) = 0.847(47\%)$$

$$\bigwedge_{w}^{(\pm Nv=3000)} : \int_{\in(A\prime)}^{\in(A)} c(p) = \sim 1(\sim 100\%)$$

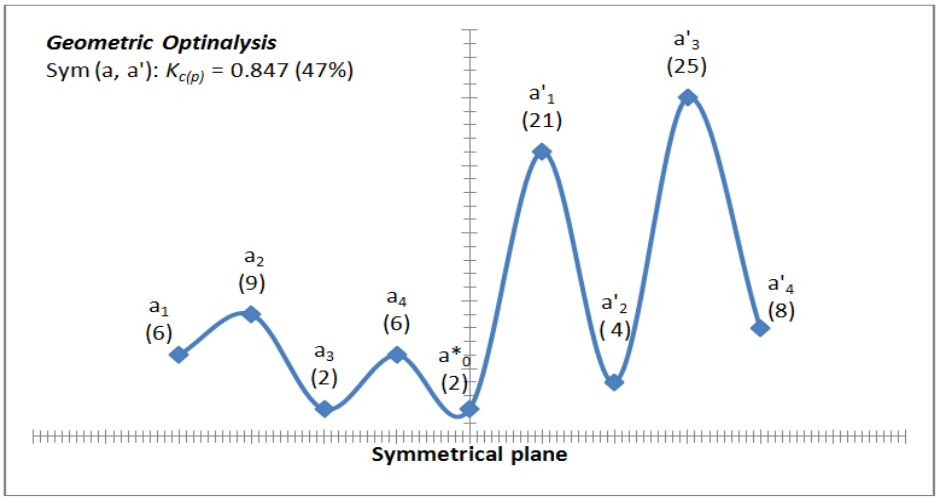


Figure 6: Asymmetrical distribution

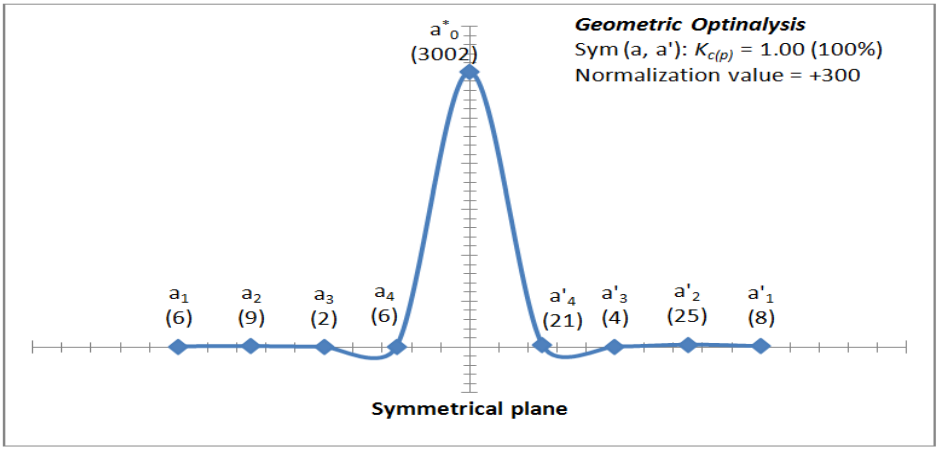


Figure 7: Transformed normalized distribution (By positive modulation)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

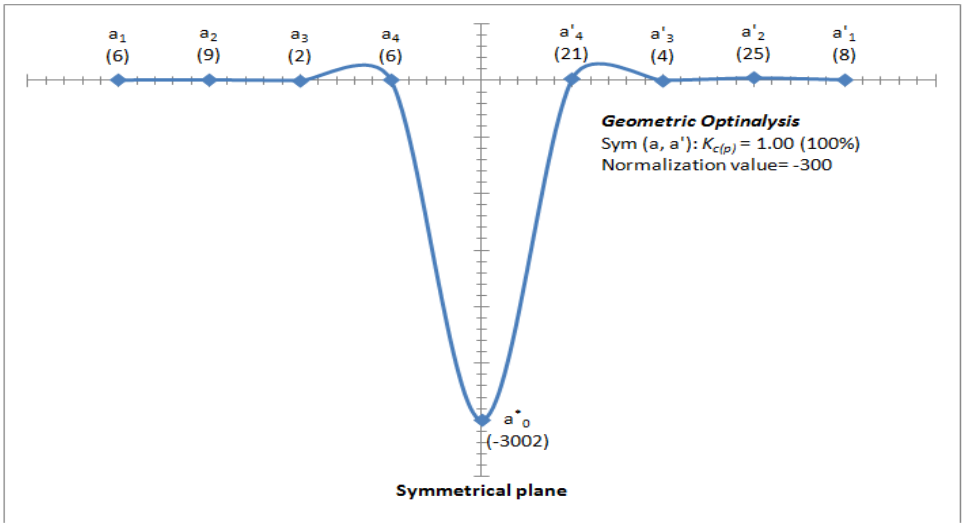


Figure 8: Transformed normalized distribution (By negative modulation)

6.0 Computations/Calculations

6.1 Kabirian Coefficient of Symmetry and Similarity

The Kabirian coefficients of symmetry and similarity (K_c) are values that quantify the magnitude and direction of balances or imbalances in the distribution of sequence elements or variables about a symmetric plane. It may exist in two value outcomes (from to central rotation) which translate the same significance level. It is calculated by intrametrically or intermetrically as described in Table 1-2 and equations 1-2.

6.1.1 Computations in intrametric symmetry detection (shape optinalysis)

As shown in Table 1, the Kabirian coefficient of symmetry that exists within the distribution of a_n elements or variables is given by the eq. (1) below. This is what quantifies intrametric symmetry and the approach is called shape optinalysis.

$$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + a_0^* + a'_n)}{1} \times \frac{1}{\sum r_n(a_n + a_0^* + a'_n)} \qquad (1.1)$$

$$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + a_0^* + a'_n)}{\sum r_n(a_n + a_0^* + a'_n)} \qquad (1.2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 1: Showing the computations in an intrametric symmetry detection (shape optinalysis)

QS-Unit (r_n)	Element (a_n)	Scalement Function $r_n\,(a_n + a_0^* + a'_n)$	Node (n_n)	Remarks
r_1	a_1	$(r_1 \times a_1)$	n_1	
r_2	a_2	$(r_2 \times a_2)$	n_2	
r_3	a_3	$(r_3 \times a_3)$	n_3	Pericentral node
r_4	a_4	$(r_4 \times a_4)$	n_4	
r_5	a_5	$(r_5 \times a_5)$	n_5	
r_6	a_0^*	$(r_6 \times a_0^*)$	n_6	Central node
r_7	a'_5	$(r_7 \times a'_5)$	n_7	
r_8	a'_4	$(r_8 \times a'_4)$	n_8	
r_9	a'_3	$(r_9 \times a'_3)$	n_9	Pericentral node
r_{10}	a'_2	$(r_{10} \times a'_2)$	n_{10}	
r_{11}	a'_1	$(r_{11} \times a'_1)$	n_{11}	
$\sum(r_n)$	$\sum(a_n + a_0^* + a'_n)$	$\sum r_n(a_n + a_0^* + a'_n)$		

6.1.2 Computations in intermetric symmetry detection (comparative optinalysis)

As sown in Table 2, the Kabirian coefficient of similarity that exists between the two paired sequences, a_n and b_n is given by eq. (2) below. This is what quantifies intermetric similarity.

$$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + c_0 + b_n)}{1} \times \frac{1}{\sum r_n(a_n + c_0 + b_n)} \qquad (2.1)$$

$$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + c_0 + b_n)}{\sum r_n(a_n + c_0 + b_n)} \qquad (2.2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 2: Showing the computations in intermetric symmetry detection (comparative optinalysis)

QS-Unit (r_n)	Elements (a_n)	Scalement Function $r_n\,(a_n + c_0 + b_n)$	Node (n_n)	Remarks
r_1	a_1	$(r_1 \times a_1)$	n_1	
r_2	a_2	$(r_2 \times a_2)$	n_2	
r_3	a_3	$(r_3 \times a_3)$	n_3	Pericentral node
r_4	a_4	$(r_4 \times a_4)$	n_4	
r_5	a_5	$(r_5 \times a_5)$	n_5	
r_6	c_0	$(r_6 \times c_0)$	n_6	Central node
r_7	b_5	$(r_7 \times b_5)$	n_7	
r_8	b_4	$(r_8 \times b_4)$	n_8	
r_9	b_3	$(r_9 \times b_3)$	n_9	Pericentral node
r_{10}	b_2	$(r_{10} \times b_2)$	n_{10}	
r_{11}	b_1	$(r_{11} \times b_1)$	n_{11}	
$\sum(r_n)$	$\sum(a_n + c_0 + b_n)$	$\sum r_n(a_n + c_0 + b_n)$		

6.2 Confidence level (probability value) of similarity and symmetry

The probability level of similarity or symmetry that exists within (intrametric) or between (intermetric) two comparing elements or variables can be calculated by the general formula below:

Suppose $r_n = 1, 2, 3;\ (b_1) = 1$ or $100;\ N_v(c_0) = 0$; and (a_1) = unknown (x).

Table 3: Bivariate Optinalysis under some constant parameters

QS-Unit (r_n)	Elements $(a_n + c_0 + b_n)$	Scalement Function $r_n(a_n + c_0 + b_n)$
1	x	$(x \times 1)$
2	0	(2×0)
3	1 or (100)	$(1 \times 1)\ or\ (1 \times 100)$
$\sum(r_n) = (6)$	$(x + 1)\ or\ (x + 100)$	$\sum r_n\,(a_n + c_0 + b_n) = (x + 0 + 1)\ or\ (x + 0 + 100)$

By substituting these variables into equation 2 (Table 3 gives further details), we have

$$K_c = \frac{2(x+1)}{x+3}\ or\ K_c = \frac{2(x+100)}{x+300}$$

Making x the subject of the formula, where x is represented as $P_{Sim.} - value$ or $\%\ Sim.$ (or $P_{Sym.} - value$ or $\%\ Sym.$), we now have:

$$P_{Sim.} - value = \frac{(2 \times 1) - (3 \times 1 \times S_c)}{K_c - 2} \qquad (3.1)$$

$$P_{Sym.} - value = \frac{(2 \times 1) - (3 \times 1 \times S_c)}{K_c - 2} \qquad (3.1)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

$$\%\ Sim. = \frac{(2 \times 100) - (3 \times 100 \times S_c)}{K_c - 2} \qquad (3.2)$$

$$\%\ Sym. = \frac{(2 \times 100) - (3 \times 100 \times S_c)}{K_c - 2} \qquad (3.2)$$

Equation (3) is appropriate to give positive outcomes if K_c is between 1 and tends to 0.66667

In the other turn, suppose $r_n = 1, 2, 3$; $(a_1) = 1$ or 100; $N_v(c_0) = 0$; and (b_1) = unknown (x).

Table 4: Bivariate Optinalysis under some constant parameters

QS-Unit (r_n)	**Elements** $(a_n + c_0 + b_n)$	**Scalement Function** $r_n(a_n + c_0 + b_n)$
1	1 or (100)	$(1 \times 1)\ or\ (1 \times 100)$
2	0	(2×0)
3	x	$(3 \times x)$
$\sum(r_n) = (6)$	$(1 + x)$ or $(100 + x)$	$\sum r_n\ (a_n + c_0 + b_n) = (1 + 0 + 3x)\ or\ (100 + 0 + 3x)$

By substituting these variables into equation 2 (Table 4, gives further details), we have

$$K_c = \frac{2(1 + x)}{3x}\ or\ K_c = \frac{2(100 + x)}{300x}$$

Making x the subject of the formula, where x is represented as $P_{Sim.} - value$ or $\%\ Sim.$ (or $P_{Sym.} - value$ or $\%\ Sym.$), we now have:

By probability as:

$$P_{Sim.} - value = \frac{(2 \times 1) - (1 \times S_c)}{(3 \times K_c) - 2} \qquad (4.1)$$

$$P_{Sym.} - value = \frac{(2 \times 1) - (1 \times S_c)}{(3 \times K_c) - 2} \qquad (4.1)$$

By percentages as:

$$\%\ Sim. - value = \frac{(2 \times 100) - (100 \times S_c)}{(3 \times K_c) - 2} \qquad (4.2)$$

$$\%\ Sym. - value = \frac{(2 \times 100) - (100 \times S_c)}{(3 \times K_c) - 2} \qquad (4.2)$$

Equation (3) is appropriate to give positive outcomes if K_c is between 1 and tends to 2, and all negative value results.

6.3 Confidence level (probability value) of dissimilarity or asymmetry

By probability as:

$$P_{Dsim.} - value = (\pm 1 - P_{Sim.} - value) \qquad (5.1)$$

$$P_{Asym.} - value = (\pm 1 - P_{Sim.} - value) \qquad (5.1)$$

By percentages as:

$$\%\ Dsim. - value = (\pm 100 - \%\ Sim. - value) \qquad (5.2)$$

$$\%\ Asym. - value = (\pm 100 - \%\ Sim. - value) \qquad (5.2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

6.4 Interpreting the Result of Optinalysis

Obtaining Kabirian coefficient equals to 1, >1, <1 indicates absolute symmetry or similarity (equal heaviness around the symmetrical line, at the left-sided sequence), asymmetry or dissimilarity (more heaviness below the symmetrical line), also asymmetry or dissimilarity (more heaviness above the symmetrical line, at the right-sided sequence) respectively. The probabilities or percentages obtained are the significance level at which the distribution of the elements/variables or the deviation of elements/variables is symmetrical about a mean.

7.0 Symbolic Notations in Optinalysis

The following symbols are used to express the algorithm of Optinalysis and the related arguments in consideration. Some symbolic demonstrations are given below and their full descriptions or meaning were followed.

Examples:

Let the left sided optinalytically reflects head-to-head (H-H) with the right sided by a normalization of 1000 units, such that elements of sequence (A) are intermetrically similar to the elements of sequence (B) with a resultant Kabirian coefficient of 1 and thus 100% similar/identical.

$$\bigwedge_{B}^{(\pm Nv=1000)} : \int_{\in(B)}^{\in(A)}{}_{c(p)} = 1(100\%)$$

Let the left sided optinalytically reflects head-to-head (H-H) with the right sided by zero normalization, such that elements of sequence (A) are intrametrically symmetrical to the elements of sequence ($A^{/}$) with a resultant Kabirian coefficient of 1 and thus 100% symmetrical.

$$\bigwedge_{w}^{(\pm Nv=0)} : \int_{\in(A')}^{\in(A)}{}_{c(p)} = 1(100\%)$$

Let the left sided optinalytically reflects tail-to-tail (T-T) with the right sided by normalization of a 50 units, such that elements of sequence (A) are intermetrically similar to the elements of sequence (B) with a resultant Kabirian coefficient of 1 and thus 100% similar/identical.

$$\bigvee_{(\pm Nv=50)}^{B} : \int_{\in(B)}^{\in(A)}{}_{c(p)} = 1(100\%)$$

Let the left sided optinalytically reflects head-to-head (H-H) with the right sided by zero normalization, such that elements of sequence (A) are intrametrically asymmetrical to the elements of sequence ($A^{/}$) by 65% and thus asymmetrical.

$$\bigwedge_{W}^{(\pm Nv=0)} : \oint_{(A')}^{(A)}{}_{(p)} = (65\%)$$

Let the left sided optinalytically reflects tail-to-tail (T-T) with the right sided by zero unit normalization, such that elements of sequence (A) are intermetrically similar to the elements of sequence (B) by 2% and thus dissimilar.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

$$\bigvee_{(\pm Nv=0)}^{B} : \oint_{(p)}^{(A)} \in (B) = (2\%)$$

It should be generally noted that, the upper and lower sequence denotation defines which sequence is on the left-sided and right-sided orientation in the pairing respectively.

8.0 Applications of Optinalysis and Method Validation

8.1 In Skewness Detection

Skewness measure is one of the very important aspects of statistics. In this subsection, new methods are presented for skewness detection using the algorithm of optinalysis. In this application, intrametric symmetry detection guidelines are used to measure how the sequence elements or variables spread around the mean or a symmetrical plane. Based on whether or not a sequence is resolved and the resolution approach used, four (4) types of skewness detection where identified as follows:

i. Raw skewness: this does not requires not any resolution, and as such the data has a meaningful shape or pattern. Raw skewness is suitably detected for multi-clustered or multivariate sequence. Table 8 presented an example.
ii. Absolute skewness: in this approach, the resolution approach for the construction of a shape to the sequence is the mean differences of the elements of the sequence. In this case, the positive and negative differences from the mean are taken into consideration. Absolute skewness is suitably detected for univariate sequence. Table 6-7 and Table A1-A2 of appendix A presented examples.
iii. Variance skewness: in this case also, the resolution approach for the construction of a shape to the sequence is the squared mean differences of the elements of the sequence. Variance skewness is suitably detected for univariate sequence. Table 6-7 and Table A1-A2 of appendix A presented examples.
iv. Standard skewness: in this case also, the resolution approach for the construction of a shape to the sequence is the square root of squared mean differences of the elements of the sequence. Standard skewness is suitably detected for univariate sequence. Table 6-7 and Table A1-A2 of appendix A presented examples.

8.1.1 Interpreting the Result of Skewness Measure

Obtaining Kabirian coefficient equals to 1, >1, <1 indicates zero skewness, negative skewness (more deviations below the mean), positive skewness (more deviations above the mean) respectively. The probabilities or percentages obtained are the significance level at which the distribution of the elements/variables or the deviation of elements/variables is symmetrical about a mean. Figure 4 is an illustration of a distribution of integers (1, 2, 3, ……, 13) with a zero skewness, the resultant shape (resolved by a squared mean differences) looks perfectly symmetrical about the mean value of 7. The optinalysis foe skewness detection described here gives a similar result with the standard method with zero skewness.

 doi:10.20944/preprints201905.0295.v2

Examples:

Table 6-7 presented an example of recorded random observations of a univarite character. The data skewness was calculated using a Graphad Prism software of 8.0.2 version, and then by the algorithms of shape optinalysis. The four (4) resolution methods as explained previosly (raw, absolute, variance and standard skewness detections) were used. The results in Table A1-A2 of appendix A shows that skewness detection by shape optinalysis is a more advance approach over the method used in the software (Pearson method), because it provide further details about the significance level of skewness, and also different approaches to symmetry detection. Both the three (3) methods considrered (absolute, variance and standard skewness detections) shows the same direction of skewness. Pearson skewness test is consitently compared with the skewness detection by optinalysis of seaquences that were sequenced in an ascending sequence order (Table 6) but not the descending sequence order (Table 7).

Table 8 presented an example of recorded frequencies (sequenced in multi-clusters of age groups) of age distribution of individuals in three (3) different populations A to C. The results of raw skewness detection by shape optinalysis in Table 8 shows that the frequencies of age distribution of individuals in each of the three (3) populations B and C are significantly (P_{Sym}>0.95) asymmetrical geometrically, while population A is significantly (P_{Sym}>0.95) symmetrical (similar) geometrically. Moreover, the histographic shape assessment of the age frequency distributions was compared to give same conclusion with the results (raw skewness) of optinalysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 6: Comparative results of skewness detection by Pearson and optinalysis methods

	Results and Methods skewness detection						
Ungrouped data	**Pearson Skewness**	***Standard Skewness by Optinalysis**			****Variance Skewness by Optinalysis**		
Ascending sequence order	**Value**	**Kc-value (H-H)**	**$P_{Sym.}$-value**	**$P_{Asym.}$-value**	**Kc-value (H-H)**	**$P_{Sym.}$-value**	**$P_{Asym.}$-value**
$^{(H)}$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	0.0000
$^{(H)}$3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3$^{(T)}$	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
$^{(H)}$1, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	-3.4641	1.625000	0.1304	0.8696	4.386700	-0.2139	-0.7861
$^{(H)}$1, 2, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	-1.9636	1.360465	0.3073	0.6927	2.423729	-0.0804	-0.9196
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 6, 7$^{(T)}$	1.9636	0.790541	0.3073	0.6927	0.629956	-0.0804	-0.9196
$^{(H)}$1, 2, 3, 4, 4, 4, 4, 4, 4, 5, 6, 7$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	-2.0000
$^{(H)}$3, 4, 5, 8, 8, 8, 8, 8, 8, 8, 8, 8$^{(T)}$	-1.5291	1.329545	0.3371	0.6629	2.093220	-0.0218	-0.9782
$^{(H)}$8, 8, 8, 8, 8, 8, 8, 8, 8, 11, 12, 13$^{(T)}$	1.5291	0.801370	0.3371	0.6629	0.656915	-0.0218	-0.9782
$^{(H)}$2, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 35$^{(T)}$	3.2630	0.672332	0.0128	0.9872	0.588446	-0.1662	-0.8338
$^{(H)}$2, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7$^{(T)}$	-3.4641	1.619718	0.1330	0.8670	4.283582	-0.2105	-0.7895
$^{(H)}$7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 35$^{(T)}$	3.4641	0.722035	0.1300	0.8700	0.565152	-0.2122	-0.7878
$^{(H)}$2, 2, 2, 2, 2, 2, 9, 9, 9, 9, 9, 9$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	0.0000

*The sequences were resolved by square root of squared mean differences to design it a shape.
**The sequences were resolved by squared mean differences to design it a shape.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 7: Comparative results of skewness detection by Pearson and optinalysis methods

	Results and Methods skewness detection						
Ungrouped data	**Pearson Skewness**	***Standard Skewness by Optinalysis**			****Variance Skewness by Optinalysis**		
Descending sequence order	**Value**	**Kc-value (H-H)**	**$P_{Sym.}$-value**	**$P_{Asym.}$-value**	**Kc-value (H-H)**	**$P_{Sym.}$-value**	**$P_{Asym.}$-value**
$^{(H)}$12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	0.0000
$^{(H)}$3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3$^{(T)}$	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 1$^{(T)}$	-3.4641	0.722222	0.1304	0.8696	0.564322	-0.2139	-0.7861
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 2, 1$^{(T)}$	-1.9636	0.790541	0.3073	0.6927	0.629956	-0.0804	-0.9196
$^{(H)}$7, 6, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	1.9636	1.360465	0.3073	0.6927	2.423729	-0.0804	-0.9196
$^{(H)}$7, 6, 5, 4, 4, 4, 4, 4, 4, 3, 2, 1$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	-2.0000
$^{(H)}$8, 8, 8, 8, 8, 8, 8, 8, 8, 5, 4, 3$^{(T)}$	-1.5291	0.801370	0.3371	0.6629	0.656915	-0.0218	-0.9782
$^{(H)}$13, 12, 11, 8, 8, 8, 8, 8, 8, 8, 8, 8$^{(T)}$	1.5291	1.329545	0.3371	0.6629	2.093220	-0.0218	-0.9782
$^{(H)}$35, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 2$^{(T)}$	3.2630	1.450496	0.2337	0.7663	3.326581	-0.1662	-0.8338
$^{(H)}$7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 2$^{(T)}$	-3.4641	0.723270	0.1330	0.8670	0.566075	-0.2105	-0.7895
$^{(H)}$35, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7$^{(T)}$	3.4641	1.625951	0.1300	0.8700	4.337196	-0.2122	-0.7878
$^{(H)}$9, 9, 9, 9, 9, 9, 2, 2, 2, 2, 2, 2$^{(T)}$	0.0000	1.000000	1.0000	0.0000	1.000000	1.0000	0.0000

*The sequences were resolved by square root of squared mean differences to design it a shape.
**The sequences were resolved by squared mean differences to design it a shape.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 8: Raw skewness detection of a grouped (multi-clustered) data by shape optinalysis

	y-axis (frequency of individuals)		
x-axis (age groups in years)	**Population A**	**Population B**	**Population C**
(H) 1-5	100	10	100
6-10	373	37	373
11-15	447	44	447
16-20	782	78	782
21-25	810	81	810
25-30	986	986	986
31-35	1537	1537	1537
36-40	1537	1537	1537
41-45	986	986	986
46-50	810	810	81
51-55	782	782	78
56-60	447	447	44
61-65	373	373	37
66-70 (T)	100	100	10
Raw skewness (Significance)			
Kc-value (H-H)	1.000000	0.872818	1.170568
$P_{Sym.}$-value	1.0000	0.5487	0.5487
$P_{Asym.}$-value	0.0000	0.4513	0.4513
Histographic assessment	Symmetrical	Negative skewed	Positive skewed

8.2 In Pairewise Genomic Sequence Comparison

An inferential sequence comparison is a very important aspect of applied mathematics and statistics such as comparative genomics. In this subsection, new method for genomic sequence comparison was presented following a sequence transformation approach here proposed.

Example:

Suppose we have a reference genomic sequence (S_0) and a set of mutant sequences ($S_{n=1-24}$) as shown in the below nucleotide sequences:

The shaded portions indicate a point of mismatch or mutation relative to the reference sequence. Since the algorithm of optinalysis works only with numerical values, an approach is proposed here to transform these nominal sequences to a numerical values based on their respective molecular mass (in g/mol) of each nucleotide base, as shown in Table 5 and appendix B.

Let the reference sequence (S_0) optinalytically reflects head-to-head (H-H or 5′—5′) with the mutant sequences (S_n) with a normalization of zero unit, such that elements of sequence S_0 are intermetrically similar to the elements of sequence S_n with a resultant Kabirain coefficient of x and thus y% similar/identical.

$$\bigwedge_{B}^{(\pm N=0)} : \int_{c(p)}^{\in(Sn)}{}_{\in(S0)} = x(y\%)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

OPTINALYSIS

Following the above argument, a pairewise comparison was made by comparative optinalysis. The results of the comparisons are presented in Table 9. To validate this method for suitability, other well known and adopted methods for pairewise genomic sequence comparison (*Needle, Stretcher, Water, Matcher*) were used on EMBL-EBI website, and the results of the analysis are presented in Table 9. The results in Table 9 show that optinalysis is more advanced over the all other algorithms of bioinformatics tools used here (i.e *Needle, Stretcher, Water, Matcher*) for biological sequence comparison. These well known existing bioinformatics tools are not absolutely geometric (position specific variations) computationally and little or no sensitivity to changes in magnitude and positions of the nucloetide bases of the examplified sequences. Therefore, optinalysis is a simple and suitable alternative approach for biological sequence compariosons.

Reference sequence

S_0	$^{(5')}$G	T	G	A	C	T	G	A	G	C	C	$T^{(3')}$

Mutant sequences

S_1	$^{(5')}$A	T	G	A	C	T	G	A	G	C	C	$T^{(3')}$
S_2	$^{(5')}$G	A	G	A	C	T	G	A	G	C	C	$T^{(3')}$
S_3	$^{(5')}$G	T	A	A	C	T	G	A	G	C	C	$T^{(3')}$
S_4	$^{(5')}$G	T	G	T	C	T	G	A	G	C	C	$T^{(3')}$
S_5	$^{(5')}$G	T	G	A	A	T	G	A	G	C	C	$T^{(3')}$
S_6	$^{(5')}$G	T	G	A	C	A	G	A	G	C	C	$T^{(3')}$
S_7	$^{(5')}$G	T	G	A	C	T	A	A	G	C	C	$T^{(3')}$
S_8	$^{(5')}$G	T	G	A	C	T	G	T	G	C	C	$T^{(3')}$
S_9	$^{(5')}$G	T	G	A	C	T	G	A	A	C	C	$T^{(3')}$
S_{10}	$^{(5')}$G	T	G	A	C	T	G	A	G	A	C	$T^{(3')}$
S_{11}	$^{(5')}$G	T	G	A	C	T	G	A	G	C	A	$T^{(3')}$
S_{12}	$^{(5')}$G	T	G	A	C	T	G	A	G	C	C	$A^{(3')}$
S_{13}	$^{(5')}$-	T	G	A	C	T	G	A	G	C	C	$T^{(3')}$
S_{14}	$^{(5')}$G	-	G	A	C	T	G	A	G	C	C	$T^{(3')}$
S_{15}	$^{(5')}$G	T	-	A	C	T	G	A	G	C	C	$T^{(3')}$
S_{16}	$^{(5')}$G	T	G	-	C	T	G	A	G	C	C	$T^{(3')}$
S_{17}	$^{(5')}$G	T	G	A	-	T	G	A	G	C	C	$T^{(3')}$
S_{18}	$^{(5')}$G	T	G	A	C	-	G	A	G	C	C	$T^{(3')}$
S_{19}	$^{(5')}$G	T	G	A	C	T	-	A	G	C	C	$T^{(3')}$
S_{20}	$^{(5')}$G	T	G	A	C	T	G	-	G	C	C	$T^{(3')}$
S_{21}	$^{(5')}$G	T	G	A	C	T	G	A	-	C	C	$T^{(3')}$
S_{22}	$^{(5')}$G	T	G	A	C	T	G	A	G	-	C	$T^{(3')}$
S_{23}	$^{(5')}$G	T	G	A	C	T	G	A	G	C	-	$T^{(3')}$
S_{24}	$^{(5')}$G	T	G	A	C	T	G	A	G	C	C	$-^{(3')}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

OPTINALYSIS

Table 5: Ordinal sequence transformation of nuclotide bases based on molecular mass

Nucleotide bases and gabs	**Molecular mass**
Adenine (A)	≈ 135 g/mol
Tymine (T)	≈ 126 g/mol
Cytosine (C)	≈ 111 g/mol
Guinine (G)	≈ 151 g/mol
Uracil (U)	≈ 112 g/mol
All other gabs	0

Table 9: Pairewise comparisons and percentage similarity and identity of nucleotide sequences

	Bioinformatics tools used/Reference sequence				
	Global Alignment		**Local Alignment**		
	Needle	***Stretcher***	***Water***	***Matcher***	****Optinalysis***
Mutant Sequences	**S_0**	**S_0**	**S_0**	**S_0**	**S_0**
S_0	100.00%	100.00%	100.00%	100.00%	100.00%
S_1	91.70%	91.70%	100.00%	100.00%	98.14%
S_2	91.70%	91.70%	91.70%	91.70%	99.05%
S_3	91.70%	91.70%	91.70%	91.70%	98.45%
S_4	91.70%	91.70%	91.70%	91.70%	99.21%
S_5	91.70%	91.70%	91.70%	91.70%	98.17%
S_6	91.70%	91.70%	91.70%	91.70%	99.39%
S_7	91.70%	91.70%	91.70%	91.70%	99.07%
S_8	91.70%	91.70%	91.70%	91.70%	99.56%
S_9	91.70%	91.70%	91.70%	91.70%	99.38%
S_{10}	91.70%	91.70%	91.70%	91.70%	99.31%
S_{11}	91.70%	91.70%	91.70%	91.70%	99.54%
S_{12}	91.70%	91.70%	100.00%	100.00%	99.91%
S_{13}	*91.70%*	*91.70%*	*100.00%*	*100.00%*	*83.09%*
S_{14}	*83.30%*	*91.70%*	*100.00%*	*100.00%*	*86.91%*
S_{15}	*91.70%*	*91.70%*	*91.70%*	*100.00%*	*85.71%*
S_{16}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*88.40%*
S_{17}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*91.45%*
S_{18}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*91.47%*
S_{19}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*91.17%*
S_{20}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*93.38%*
S_{21}	*91.70%*	*91.70%*	*91.70%*	*91.70%*	*94.03%*
S_{22}	*83.30%*	*91.70%*	*100.00%*	*100.00%*	*96.71%*
S_{23}	*83.30%*	*91.70%*	*100.00%*	*100.00%*	*97.79%*
S_{24}	*91.70%*	*91.70%*	*100.00%*	*100.00%*	*98.73%*

*Molecular mass approach of ordinal transformation was used.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Summary

Optinalysis can be summarized as:

- Optinalysis, as method of symmetry detection and similarity measurement, intrametrically (within elements or variables) or intermetrically (between elements or variables) computes and compares two or more univariate or multi-clustered or multivariate sequences as a mirror-like reflection of each other (optics-like manner).
- Elements or variables of symmetrical structures reflect in same moment (or in same total moments) about a symmetrical line.
- Lack of symmetry (asymmetry) exists when reflection is not in same moment (or toal moments) about a symmetrical line.
- Kabirian coefficient of symmetry or similarity is the fundamental value that gives further calculations of the statistical inferences about symmetry or similarity level.
- Symmetry detection reflects similarity measurement.
- Optinalysis is suitable alternative for skewness measure and also a pairewise sequence analysis and comparisons.
- The algorithm of shape optinalysis can be graphically summarized as illustrated in Table 10.
- The algorithm of comparative optinalysis can be graphically summarized as illustrated in Table 11.

Table 10: Summary of the algorithm of shape optinalysis

Instruction	Selection	Standard skewness detection by shape optinalysis within a sequence of elements: Steps and calculations with an example												Details in Sections:
Observations	Random repeated measurement	Sequenceless data:	A =	(6,	15,	4,	9,	6,	2,	7,	8,	5.	9,	3)	
Sequencing	Ascending order	Sequence:	A =	(2,	3,	4,	5,	6,	6,	7,	8,	9.	9,	15)	5.1, 5.1.2
Resolution	Square root of squared mean differences	Mean differences	A:A' =	(-4.73,	-3.73,	-2.73,	-1.73,	-0.73,	-0.73,	0.27,	1.27,	2.27,	2.27,	8.27)	5.2
		Sqr. of sqd. mean diff.	A:A' =	(4.73,	3.73,	2.73,	1.73,	0.73,	0.73,	0.27,	1.27,	2.27,	2.27,	8.27)	
Annotations	Symbolic representation	Sequence:	A:A' =	(4.73,	3.73,	2.73,	1.73,	0.73,	0.73,	0.27,	1.27,	2.27,	2.27,	8.27)	3.2, 5.3
		Annotation:		$^{(H)}(a_1$	a_2	a_3	a_4	a_5	a_0^*	a'_5	a'_4	a'_3	a'_2	$a'_1)^{(T)}$	
Normalization	Zero	Sequence:	A:A' =	(4.73,	3.73,	2.73,	1.73,	0.73,	0.73,	0.27,	1.27,	2.27,	2.27,	8.27)	5.4
		Annotation:		$^{(H)}(a_1$	a_2	a_3	a_4	a_5	a_0^*	a'_5	a'_4	a'_3	a'_2	$a'_1)^{(T)}$	
Q-scale Assignment and annotations	Scale of 1 unit to represent a specific Position	Sequence:	A:A' =	(4.73,	3.73,	2.73,	1.73,	0.73,	0.73,	0.27,	1.27,	2.27,	2.27,	8.27)	3.1
		Annotation:		$^{(H)}(a_1$	a_2	a_3	a_4	a_5	a_0^*	a'_5	a'_4	a'_3	a'_2	$a'_1)^{(T)}$	
		Q-scale:		1	2	3	4	5,	6	7	8	9	10	11	
		Q-S annotation:		$^{(H)}(r_1$	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	$r_{11})^{(T)}$	
Computes	Sum of elements	$\sum(a_n + a_0^* + a'_n) = 28.73$													6.1.1, 6.1.2
	Sum of scalements	$\sum r_n(a_n + a_0^* + a'_n) = 181.48$													6.1.1, 6.1.2
	Nodality	$N = 11$													3.5
Computes	Kabirian coefficient	$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + a_0^* + a'_n)}{\sum r_n(a_n + a_0^* + a'_n)} = 0.949857$													6.1, 6.1.1, 6.1.2
	Probabilities	Valid for $K_c \leq 1$	$P_{Sym.} - value = \frac{(2 \times 1) - (3 \times 1 \times S_c)}{K_c - 2} = 0.809004$												6.2
		Valid for $K_c \geq 1$ and $all - ve\ results$	$P_{Sym.} - value = \frac{(2 \times 1) - (1 \times S_c)}{(3 \times K_c) - 2} = ?$												6.2
			$P_{Asym.} - value = (\pm 1 - P_{Sim.} - value) = 0.190996$												6.2
	Percentages	Valid for $K_c \leq 1$	$\%\ Sym. = \frac{(2 \times 100) - (3 \times 100 \times S_c)}{K_c - 2} = 80.90$												6.3
		Valid for $K_c \geq 1$ and $all - ve\ results$	$\%\ Sym. - value = \frac{(2 \times 100) - (100 \times S_c)}{(3 \times K_c) - 2} = ?$												6.3
			$\%\ Asym. - value = (\pm 100 - \%\ Sim. - value) = 19.10$												6.3

Note: Same colored boxes indicate an intrametric correspondence within the attributes of the sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table 11: Summary of the algorithm of comparative optinalysis

Instruction	Selection	Pairewise similarity detection by comparative optinalysis between two nucleotide sequences: Steps and calculations with an example	Details in section:
Genomic (DNA) nucleotide base sequences	Identified sequences	Sequence: A = $^{(5')}$(G A G C C T)$^{(3')}$ Sequence: B = $^{(5')}$(G A - T A T)$^{(3')}$	
Pre-optinalytic stage (Ordinal transformation)	Molecular mass Approach	Sequence: A = $^{(5')}$(151, 135, 151, 111, 111, 126)$^{(3')}$ Sequence: B = $^{(5')}$(151, 135, 0, 126, 135, 126$^{(3')}$	
Sequencing	Theoretical order	Sequence: A = $^{(5')}$(151, 135, 151, 111, 111, 126)$^{(3')}$ Sequence: B = $^{(5')}$(151, 135, 0, 126, 135, 126)$^{(3')}$	5.1, 5.1.1
Resolution	Not required		5.2
Annotations	Symbolic representation	Sequence: A = $^{(5')}$(151, 135, 151, 111, 111, 126)$^{(3')}$ Annotation: $^{(H)}$(a_1 a_2 a_3 a_4 a_5 a_6)$^{(T)}$ Sequence: B = $^{(5')}$(151, 135, 0, 126, 135, 126)$^{(3')}$ Annotation: $^{(H)}$(b b_2 b_3 b_4 b_5 a_6)$^{(T)}$	3.2, 5.3
Pairing style	Head-to-head (H–H)/ (5'–5')	Sequence: A:B = $^{(5')}$(151, 135, 151, 111, 111, 126, N_v 126, 135, 126, 0, 135, 151)$^{(5')}$ Annotation: $^{(H)}$(a_1 a_2 a_3 a_4 a_5 a_6 c_0 b_6 b_5 b_4 b_3 b_2 b_1)$^{(T)}$	5.3
Normalization (N_v)	Zero	Sequence: A:B = $^{(5')}$(151, 135, 151, 111, 111, 126, 0, 126, 135, 126, 0, 135, 151)$^{(5')}$ Annotation: $^{(H)}$(a_1 a_2 a_3 a_4 a_5 a_6 c_0 b_6 b_5 b_4 b_3 b_2 b_1)$^{(T)}$	5.4
Q-scale assignment And annotations	Scale of 1 unit to represent a specific Position	Sequence: A:B = $^{(5')}$(151, 135, 151, 111, 111, 126, 0, 126, 135, 126, 0, 135, 151)$^{(5')}$ Annotation: $^{(H)}$(a_1 a_2 a_3 a_4 a_5 a_6 c_0 b_6 b_5 b_4 b_3 b_2 b_1)$^{(T)}$ Q-scale: 1 2 3 4 5, 6 7 8 9 10 11 12 13 Q-S annotat.: r_1 r_2 r_3 r_4 r_5 r_6 r_7 r_8 r_9 r_{10} r_{11} r_{12} r_{13}	3.1
Computes	Sum of variables	$\sum(a_n + c_0 + b_n) = 1458$	6.1.1, 6.1.2
	Sum of scalements	$\sum r_n(a_n + c_0 + b_n) = 9695$	6.1.1, 6.1.2
	Nodality	$N = 13$	3.5
Computes	Kabirian coefficient	$K_c = \frac{\sum(r_n)}{N} \times \frac{\sum(a_n + c_0 + b_n)}{\sum r_n(a_n + c_0 + b_n)} = 1.052708$	6.1, 6.1.2
	Probabilities	Valid for $K_c \leq 1$: $P_{Sim.} - value = \frac{(2 \times 1) - (3 \times 1 \times S_c)}{K_c - 2} = ?$	6.2
		Valid for $K_c \geq 1$ *and all – ve results*: $P_{Sim.} - value = \frac{(2 \times 1) - (1 \times S_c)}{(3 \times K_c) - 2} = 0.8180$	6.2
		$P_{Dsim.} - value = (\pm 1 - P_{Sim.} - value) = 0.1820$	6.2
	Percentages	Valid for $K_c \leq 1$: $\%\ Sim. = \frac{(2 \times 100) - (3 \times 100 \times S_c)}{K_c - 2} = ?$	6.3
		Valid for $K_c \geq 1$ *and all – ve results*: $\%\ Sim. - value = \frac{(2 \times 100) - (100 \times S_c)}{(3 \times K_c) - 2} = 81.80$	6.3
		$\%\ Dsim. - value = (\pm 100 - \%\ Sim. - value) = 18.20$	6.3

Note: Same colored boxes indicate an intermetric correspondence between the two attributes of the sequences.

Acknowledgement
I thank for the motivations and encouragement received from Abubakar Bello (PhD) of Biology Department, Umaru Musa Yar'adua University, Katsina. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest
The author declares no conflict of interest.

References

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1: 181–242.

Bookstein, F. L. (1991). Landmark methods for forms without landmarks: localizing group differences in outline shape. *Medical Image Analysis*, 1(3):225–244.

Cootes, T. F. and Taylor, C. J. (Oct 2001). *Statistical Models of Appearance for Computer Vision*. Tech. Report. University of Manchester,

Darvas G. (2007). *Symmetry*, Basel-Boston-Berlin: Birkhäuser, xi + 504 pp.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. John Wiley & Sons.

Dryden, I. L., and Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R*, Second Edition, John Wiley & Sons, Ltd.

Dryden, I. L., Kume, A., Le, H., and Wood, A. T. A. (2008). A multi-dimensional scaling approach to shape analysis. *Biometrika*, 95(4): 779–798.

Fraasen, V., and Bsa, C. (1989). *Laws and Symmetry*. Oxford University press Inc., New York

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Jour. Royal Statistical Society, Series B*, 53:285–339.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16: 81–121.

Kent, J. T. (1994). The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society, Series B*, 56: 285–299.

Lele, S. and Richtsmeier, J. T. (1991). Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86: 415–427.

Mardia, K.V.; Bookstein, F.L.; Moreton, I. J. (2000). Statistical assessment of bilateral symmetry of shapes. *Biometrika*, 87, 285–300.

Petitjean M. A. (2007). Definition of Symmetry. *Symmetry: Culture and Science*, 18 (2-3), pp.99 – 119.

 doi:10.20944/preprints201905.0295.v2

Watson, G. S. (1986). The shape of a random sequence of triangles. *Advances in Applied Probability*,18: 156–169.

Weyl H. (1952). *Symmetry*, Princeton, N.J: Princeton University Press.

Zheng, G. and S. Szekely, Li; G. (2017). *Statistical Shape and Deformation Analysis*. Academic Press. ISBN 9780128104941.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Appendix A

Table A1: Sequence order and resolution methods of univariate observations.

	Resolution methods to design a shape to the sequences.		
Ascending sequence order	**Mean differences**	**Square root of squared mean differences**	**Squared mean differences**
$^{(H)}$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$^{(T)}$	$^{(H)}$-5.5, -4.5, -3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5, 3.5, 4.5, 5.5$^{(T)}$	$^{(H)}$5.5, 4.5, 3.5, 2.5, 1.5, 0.5, 0.5, 1.5, 2.5, 3.5, 4.5, 5.5$^{(T)}$	$^{(H)}$30.25, 20.25, 12.25, 6.25, 2.25, 0.25, 0.25, 2.25, 6.25, 12.25, 20.25, 30.25$^{(T)}$
$^{(H)}$3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$
$^{(H)}$1, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	$^{(H)}$-2.75, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25$^{(T)}$	$^{(H)}$2.75, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25$^{(T)}$	$^{(H)}$7.56, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06$^{(T)}$
$^{(H)}$1, 2, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	$^{(H)}$-2.5, -1.5, -0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5$^{(T)}$	$^{(H)}$2.5, 1.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5$^{(T)}$	$^{(H)}$6.25, 2.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25$^{(T)}$
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 6, 7$^{(T)}$	$^{(H)}$-0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, 0.5, 1.5, 2.5$^{(T)}$	$^{(H)}$0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 2.5$^{(T)}$	$^{(H)}$0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 2.25, 6.25$^{(T)}$
$^{(H)}$1, 2, 3, 4, 4, 4, 4, 4, 4, 4, 5, 6, 7$^{(T)}$	$^{(H)}$-3, -2, -1, 0, 0, 0, 0, 0, 0, 1, 2, 3$^{(T)}$	$^{(H)}$3, 2, 1, 0, 0, 0, 0, 0, 0, 1, 2, 3$^{(T)}$	$^{(H)}$9, 4, 1, 0, 0, 0, 0, 0, 0, 1, 4, 9$^{(T)}$
$^{(H)}$3, 4, 5, 8, 8, 8, 8, 8, 8, 8, 8, 8$^{(T)}$	$^{(H)}$-4, -3, -2, 1, 1, 1, 1, 1, 1, 1, 1, 1$^{(T)}$	$^{(H)}$4, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1$^{(T)}$	$^{(H)}$16, 9, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1$^{(T)}$
$^{(H)}$8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 11, 12, 13$^{(T)}$	$^{(H)}$-1, -1, -1, -1, -1, -1, -1, -1, -1, 2, 3, 4$^{(T)}$	$^{(H)}$1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 4$^{(T)}$	$^{(H)}$1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 9, 16$^{(T)}$
$^{(H)}$2, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 35$^{(T)}$	$^{(H)}$-6.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, 26.08$^{(T)}$	$^{(H)}$6.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 26.08$^{(T)}$	$^{(H)}$47.89, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 680.17$^{(T)}$
$^{(H)}$2, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7$^{(T)}$	$^{(H)}$-4.58, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42$^{(T)}$	$^{(H)}$4.58, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42$^{(T)}$	$^{(H)}$20.98, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18$^{(T)}$
$^{(H)}$7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 35$^{(T)}$	$^{(H)}$-2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, 25.67$^{(T)}$	$^{(H)}$2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 25.67$^{(T)}$	$^{(H)}$5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 658.95$^{(T)}$
$^{(H)}$2, 2, 2, 2, 2, 2, 9, 9, 9, 9, 9, 9$^{(T)}$	$^{(H)}$-3.5, -3.5, -3.5, -3.5, -3.5, -3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5$^{(T)}$	$^{(H)}$3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5$^{(T)}$	$^{(H)}$12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25$^{(T)}$

*Descaled mean deviations refer to a form of sequence resolution that removes all the scaling effect (positive and negative signs) from a shaped sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Table A2: Sequence order and resolution methods of univariate observations.

	Resolution methods to design a shape to the sequences.		
Descending sequence order	**Mean differences**	**Square root of squared mean differences**	**Squared mean differences**
$^{(H)}$12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1$^{(T)}$	$^{(H)}$5.5, 4.5, 3.5, 2.5, 1.5, 0.5, -0.5, -1.5, -2.5, -3.5, -4.5, -5.5$^{(T)}$	$^{(H)}$5.5, 4.5, 3.5, 2.5, 1.5, 0.5, 0.5, 1.5, 2.5, 3.5, 4.5, 5.5$^{(T)}$	$^{(H)}$30.25, 20.25, 12.25, 6.25, 2.25, 0.25, 0.25, 2.25, 6.25, 12.25, 20.25, 30.25$^{(T)}$
$^{(H)}$3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$	$^{(H)}$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0$^{(T)}$
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 1$^{(T)}$	$^{(H)}$0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, -2.75$^{(T)}$	$^{(H)}$0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 2.75$^{(T)}$	$^{(H)}$0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 7.56$^{(T)}$
$^{(H)}$4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 2, 1$^{(T)}$	$^{(H)}$0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, -0.5, -1.5, -2.5$^{(T)}$	$^{(H)}$0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 2.5$^{(T)}$	$^{(H)}$0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 2.25, 6.25$^{(T)}$
$^{(H)}$7, 6, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4$^{(T)}$	$^{(H)}$2.5, 1.5, 0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5$^{(T)}$	$^{(H)}$2.5, 1.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5$^{(T)}$	$^{(H)}$6.25, 2.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25$^{(T)}$
$^{(H)}$7, 6, 5, 4, 4, 4, 4, 4, 4, 3, 2, 1$^{(T)}$	$^{(H)}$3, 2, 1, 0, 0, 0, 0, 0, 0, -1, -2, -3$^{(T)}$	$^{(H)}$3, 2, 1, 0, 0, 0, 0, 0, 0, 1, 2, 3$^{(T)}$	$^{(H)}$9, 4, 1, 0, 0, 0, 0, 0, 0, 1, 4, 9$^{(T)}$
$^{(H)}$8, 8, 8, 8, 8, 8, 8, 8, 8, 5, 4, 3$^{(T)}$	$^{(H)}$1, 1, 1, 1, 1, 1, 1, 1, 1, -2, -3, -4$^{(T)}$	$^{(H)}$1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 4$^{(T)}$	$^{(H)}$1, 1, 1, 1, 1, 1, 1, 1, 1, 4, 9, 16$^{(T)}$
$^{(H)}$13, 12, 11, 8, 8, 8, 8, 8, 8, 8, 8, 8$^{(T)}$	$^{(H)}$4, 3, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1$^{(T)}$	$^{(H)}$4, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1$^{(T)}$	$^{(H)}$16, 9, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1$^{(T)}$
$^{(H)}$35, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 2$^{(T)}$	$^{(H)}$26.08, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -1.92, -6.92$^{(T)}$	$^{(H)}$26.08, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 1.92, 6.92$^{(T)}$	$^{(H)}$680.17, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 3.69, 47.89$^{(T)}$
$^{(H)}$7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 2$^{(T)}$	$^{(H)}$0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, -4.58$^{(T)}$	$^{(H)}$0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 0.42, 4.58$^{(T)}$	$^{(H)}$0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 0.18, 20.98$^{(T)}$
$^{(H)}$35, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7$^{(T)}$	$^{(H)}$25.67, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33, -2.33$^{(T)}$	$^{(H)}$25.67, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33, 2.33$^{(T)}$	$^{(H)}$658.95, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43, 5.43$^{(T)}$
$^{(H)}$9, 9, 9, 9, 9, 9, 2, 2, 2, 2, 2, 2$^{(T)}$	$^{(H)}$3.5, 3.5, 3.5, 3.5, 3.5, 3.5, -3.5, -3.5, -3.5, -3.5, -3.5, -3.5$^{(T)}$	$^{(H)}$3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5$^{(T)}$	$^{(H)}$12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25$^{(T)}$

*Descaled mean deviations refer to a form of sequence resolution that removes all the scaling effect (positive and negative signs) from a shaped sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2019 doi:10.20944/preprints201905.0295.v2

Appendix B

The ordinal transformed sequences of nucleotide bases by their molecular masses (g/mol)

Reference sequence

S_0	$^{(H)}151$	126	151	135	111	126	151	135	151	111	111	$126^{(T)}$

Mutant sequences

S_1	$^{(H)}151$	126	151	135	111	126	151	135	151	111	111	$126^{(T)}$
S_2	$^{(H)}151$	135	151	135	111	126	151	135	151	111	111	$126^{(T)}$
S_3	$^{(H)}151$	126	135	135	111	126	151	135	151	111	111	$126^{(T)}$
S_4	$^{(H)}151$	126	151	126	111	126	151	135	151	111	111	$126^{(T)}$
S_5	$^{(H)}151$	126	151	135	135	126	151	135	151	111	111	$126^{(T)}$
S_6	$^{(H)}151$	126	151	135	111	135	151	135	151	111	111	$126^{(T)}$
S_7	$^{(H)}151$	126	151	135	111	126	135	135	151	111	111	$126^{(T)}$
S_8	$^{(H)}151$	126	151	135	111	126	151	126	151	111	111	$126^{(T)}$
S_9	$^{(H)}151$	126	151	135	111	126	151	135	135	111	111	$126^{(T)}$
S_{10}	$^{(H)}151$	126	151	135	111	126	151	135	151	135	111	$126^{(T)}$
S_{11}	$^{(H)}151$	126	151	135	111	126	151	135	151	111	135	$126^{(T)}$
S_{12}	$^{(H)}151$	126	151	135	111	126	151	135	151	111	111	$135^{(T)}$
S_{13}	$^{(H)}0$	126	151	135	111	126	151	135	151	111	111	$126^{(T)}$
S_{14}	$^{(H)}151$	0	151	135	111	126	151	135	151	111	111	$126^{(T)}$
S_{15}	$^{(H)}151$	126	0	135	111	126	151	135	151	111	111	$126^{(T)}$
S_{16}	$^{(H)}151$	126	151	0	111	126	151	135	151	111	111	$126^{(T)}$
S_{17}	$^{(H)}151$	126	151	135	0	126	151	135	151	111	111	$126^{(T)}$
S_{18}	$^{(H)}151$	126	151	135	111	0	151	135	151	111	111	$126^{(T)}$
S_{19}	$^{(H)}151$	126	151	135	111	126	0	135	151	111	111	$126^{(T)}$
S_{20}	$^{(H)}151$	126	151	135	111	126	151	0	151	111	111	$126^{(T)}$
S_{21}	$^{(H)}151$	126	151	135	111	126	151	135	0	111	111	$126^{(T)}$
S_{22}	$^{(H)}151$	126	151	135	111	126	151	135	151	0	111	$126^{(T)}$
S_{23}	$^{(H)}151$	126	151	135	111	126	151	135	151	111	0	$126^{(T)}$
S_{24}	$^{(H)}151$	126	151	135	111	126	151	135	151	111	111	$0^{(T)}$