Research Information

Available online at www.researchinformation.co.uk
The Journal of Grey System
Volume 29 No.1, 2017

Correspondence Analysis With Grey Data: The Grey Eigenvalue Problem

Philippos D. Alevizos[1*], Yiannis Theodorou[2], Michael N. Vrahatis[1]

1. *Department of Mathematics, University of Patras, GR-26110 Patras, Greece*
2. *Department of Electronics, Technological Educational Institute of Lamia, GR-35100 Lamia, Greece*

Abstract

Correspondence analysis with fuzzy data has been proposed a few years ago and the mathematical foundation for fuzzy contingency tables as well as for the related fuzzy eigenvalue problem has been investigated. These approaches have been named correspondence analysis with fuzzy data. In the paper at hand, we enrich these theoretical results as well as the mathematical foundation on grey contingency table in the case where its entries-data are grey numbers. We name this approach correspondence analysis with grey data. The aim of the paper at hand is to show that the correspondence analysis can be enriched to tackle uncertainties caused by grey data. We investigate here mainly the grey eigenvalue problem. The proposed approach is validated by using data from a real-life application and the corresponding computational processes are explained in detail through a simple representative numerical example. Our experience is that the proposed procedures can be easily implemented computationally. Furthermore, although our results are comparable to those obtained using fuzzy numbers, the corresponding required computational burden is significantly reduced.

Keywords: Grey Systems; Correspondence Analysis; Grey Numbers; Grey Matrices; Grey Eigenvalue Problem; Whitenization Methods; Grey Statistics and Data Analysis

1. Introduction

Correspondence Analysis (CA) proposed by Hirschfeld[8] and later developed by Benzécri[3, 4] is a traditional multivariate geometrical statistical method that converts a two-way (and in general multi-way) contingency table into a particular type of graphical display, in which the matrix rows and matrix columns are exhibited as points simultaneously. CA is similar to *Principal Component Analysis* (PCA) and is applied to categorical (nonnegative on the same scale) data rather than continuous data. CA is a well-known and widely used statistical technique aiming to analyze a contingency table by losing relevant information as much as necessary. This can be achieved by the factorization of a certain matrix which is

[*] ***Corresponding Author:*** *Philippos D. Alevizos, Department of Mathematics, University of Patras, GR-26110 Patras, Greece; Email: philipos@math.upatras.gr*

based on the *eigendecomposition*. Such a matrix can be decomposed by the product of four matrices. By using this decomposition technique we are in a position to project the observed data onto a lower dimensional subspace.

In many issues that arise in practice as well as in many applications of science and engineering, including among others, engineering control systems, psychology sentimental situations, economy price indices, opinion polls, medicine diagnostics, palaeontology classification of discoveries as well as linguistics structure of expressions, the usage of uncertain information becomes a "conditio sine qua non". Thus, data of such issues can not be easily determined precisely and, in general, an estimate of the data is used. Therefore, inevitably approximate quantities enter into statistical analysis that usually can be handled by using classical real intervals.

An interval of classical Mathematics, can be used in technological applications, as in the case of grey systems[2, 6, 7, 11, 12, 13, 15, 16] to determine an uncertain value. More specifically, assume that a variable with uncertain value, (that is, when the available information is insufficient and not precisely determinable), is estimated to lie between a minimum value a_1 and a maximum value a_2 then this variable can be handled by considering that it belongs in a real interval $[a_1, a_2]$. In this case, "grey quantities" are inevitably entered into the applications and the statistical treatments, as grey numbers, grey matrices, grey eigenvalue, etc., which, in general, can not be handled by the classical statistical methods.

The paper at hand constitutes an enrichment to tackle uncertainties caused by grey data of a previous work[25], where the mathematical foundation and the algebraic treatment of the fuzzy eigenvalue problem of *Correspondence Analysis with Fuzzy Data* (CAFD) is investigated. Here, we implement these theoretical results to the application of *Correspondence Analysis with Grey Data* (CAGD). Furthermore, an objective of the paper at hand is to show how CAGD can work in practice. To this end, a numerical approach is provided and the used processes are explained through a simple representative example taken from real life applications. This example has been studied in detail in[25] and we use it in the study at hand for comparison purposes.

Our experience is that the proposed procedures can be easily implemented computationally. Furthermore, although the obtained results using the proposed approach are comparable to those obtained using fuzzy numbers the corresponding required computational burden is significantly reduced.

The paper is organized as follows. In Section 2, the basic formulation of the standard CA as well as the grey arithmetic are provided. The mathematical framework of CAGD is synopsized in Section 3. The substance of this work is in Section 4, where a representative numerical example is presented in details, in order to illustrate in practice the computational processes developed mathematically earlier. Finally, a synopsis and some concluding remarks are presented in Section 5.

2. Background Material and Mathematical Formulation

2.1 Correspondence Analysis–CA

The basic mathematical framework of the standard two-way CA (which must be distinguished from its multi-way extension called Multiple Correspondence Analysis (MCA)) has briefly the formulation presented below. For a basic bibliography on the standard CA we refer the interested reader to[3, 4, 5, 9, 22].

Let us assume that $K(I \times J)$ is a two-way contingency table between two finite sets I (rows) and J (columns), which is also denoted by the $(n \times p)$ matrix $K_{n,p} = (k_i^j)$ with nonnegative elements $k_i^j \geq 0,\ i \in I = \{1,\dots,n\},\ j \in J = \{1,\dots,p\}$. Then we consider the following matrices, which are generated by K:

$$X_{n,p} = (x_i^j), \qquad \text{(row stochastic-profiles matrix)}, \qquad (1)$$

$$D_{n,n} = \operatorname{diag}(m_i) \qquad \text{and}$$

$$Q_{p,p} = \operatorname{diag}\left(\frac{1}{m^j}\right).$$

where:

$$x_i^j = \frac{k_i^j}{k_i}, \qquad m_i = \frac{k_i}{k}, \qquad m^j = \frac{k^j}{k},$$

and $\quad k_i = \sum_{j=1}^{p} k_i^j, \qquad k^j = \sum_{i=1}^{n} k_i^j, \qquad k = \sum_{i=1}^{n}\sum_{j=1}^{p} k_i^j$.

We also consider the following sets-clouds:

$$\begin{aligned} N_I &= \left\{(x_i, m_i) \ \mid\ i \in I,\quad x_i \in \mathbb{R}^p,\ \ m_i > 0,\quad \sum_{i=1}^{n} m_i = 1\right\}, \\ N_J &= \left\{(y^j, m^j) \ \mid\ j \in J,\ \ y^j \in \mathbb{R}^n,\ \ m^j > 0,\ \ \sum_{j=1}^{p} m^j = 1\right\}, \end{aligned} \qquad (2)$$

where:

$$x_i = (x_i^j) = (x_i^1,\dots,x_i^j,\dots,x_i^p) \in \mathbb{R}^p, \ \text{(row-profile)} \ \text{ and } \ x_i^j = \frac{k_i^j}{k_i},\ \ j \in J,$$

$$y^j = (y_i^j) = (y_1^j,\dots,y_i^j,\dots,y_n^j) \in \mathbb{R}^n, \ \text{(column-profile)} \ \text{ and } \ y_i^j = \frac{k_i^j}{k^j},\ \ i \in I .$$

The distance d between two row-profiles $x_i, x_{i'} \in N_I$, is a weighted Euclidean distance, and is given by the following relation:

$$d^2(x_i, x_{i'}) = \sum_{j=1}^{p} \frac{1}{m^j}(x_i^j - x_{i'}^j)^2 = (x_i - x_{i'})^{\mathrm{T}}\, Q\, (x_i - x_{i'}) . \qquad (3)$$

The center of gravity of the set-cloud N_I is the vector:

$$g_I = \sum_{i=1}^{n} m_i x_i = (m^1,\dots,m^j,\dots,m^p) . \qquad (4)$$

The formal algebraic treatment of CA, leads finally to the eigenvalue problem of the column-stochastic matrix:

$$S_{p,p} = V_{p,p}\ Q_{p,p} = \left(X_{p,n}^{\mathrm{T}}\ D_{n,n}\ X_{n,p}\right)\ Q_{p,p}, \qquad (5)$$

where p = rank(S), or, equivalently by diagonalization of the symmetric matrix:

$$A_{p,p} = Q_{p,p}^{1/2}\ X_{p,n}^{\mathrm{T}}\ D_{n,n}\ X_{n,p}\ Q_{p,p}^{1/2}, \qquad (6)$$

where S and A have the same eigenvalues λ, whereas the eigenvectors u of S and w of A are related as follows:

$$u = Q^{1/2}\ w .$$

The eigenvalues λ_s of the matrix S or A satisfy, in general, the following conditions:

$$\lambda_s \in [0,1],\ (s = 1,2,\ldots,r), \quad \text{with}$$

$$\lambda_1 = 1 \geq \lambda_2 \geq \cdots \geq \lambda_s \geq \cdots \geq \lambda_r \geq 0, \quad r = \operatorname{rank}(S) \leq \min\{n, p\},$$

and
$$\lambda_s = \sum_{i=1}^{n} m_i\, \psi_s^2(x_i), \tag{7}$$

where:

$$\psi_s(x_i) = \operatorname{proj}_{\Delta u_s}(x_i) = (x_i)^{\mathrm{T}} Q\, u_s \in \mathbb{R}, \quad \psi_{u_s}(X) = \operatorname{proj}_{u_s}(X) = X\, Q\, u_s, \tag{8}$$

or
$$\psi_{w_s}(X) = \operatorname{proj}_{w_s}(X) = X\, Q^{1/2} w_s \quad \text{and} \quad \sum_{i=1}^{n} m_i \psi_s(x_i) = 0, \quad \sum_{i=1}^{n} m_i = 1.$$

For each point-vector or row-profile

$$x_i = (x_i^1, \ldots, x_i^j, \ldots, x_i^p) \in N_I \subseteq \mathbb{R}^p,$$

the new coordinates with respect to the system of factorial (or principal) axes Δu_s are as follows:

$$x_i = (\psi_1(x_i), \psi_2(x_i), \ldots, \psi_r(x_i)) = (\operatorname{proj}_{\Delta u_1}(x_i), \operatorname{proj}_{\Delta u_2}(x_i), \ldots, \operatorname{proj}_{\Delta u_r}(x_i)) \in \mathbb{R}^r \tag{9}$$

Thus, according to Eq. (7), we can consider the eigenvalue λ_s as the weighted variance of the factor-function ψ_s.

2.2 Grey Arithmetic

A grey number is an indeterminate number that takes its possible value within an interval[10, 18]. In other words, a number whose probable range is known with clear upper and lower boundaries but which has an unknown position within the boundaries is called a grey number.

A grey number A is expressed mathematically as follows:

$$A \in [a_1, a_2] = \{x \in \mathbb{R} \mid a_1 \leq x \leq a_2\}.$$

Clearly, a grey number represents the range of the possible variance of the underlying number. In this sense, a grey number is the same as an interval value with the same limits[1]. That is, the algebra of grey numbers is very similar to interval arithmetic.

However, compared with interval values, a grey number enriches its uncertainty representation with the whitenization function and the degree of greyness[13, 14, 19]. Thus, the arithmetic of grey numbers can be defined by the well-known and widely used arithmetic of the ordinary real intervals, (see e.g.[20, 21]).

Therefore, if $A \in [a_1, a_2]$ and $B \in [b_1, b_2]$ are two grey numbers with corresponding interval values

$$[a_1, a_2] = \{x \in \mathbb{R} \mid a_1 \leq x \leq a_2\} \quad \text{and} \quad [b_1, b_2] = \{x \in \mathbb{R} \mid b_1 \leq x \leq b_2\},$$

then generally we have,

$$A \circledast B \in [a_1, a_2] \circledast [b_1, b_2] = \{x \circledast y \mid a_1 \leq x \leq a_2,\ b_1 \leq y \leq b_2\}, \tag{10}$$

where the symbol "$\circledast$" denotes any one of the four basic grey arithmetic operations.

More specifically for the basic grey arithmetic operations we have the following:

Addition:

$A \oplus B \in [a_1, a_2] \oplus [b_1, b_2] = [a_1 + b_1, a_2 + b_2]$,

Subtraction:

$A \ominus B \in [a_1, a_2] \ominus [b_1, b_2] = [a_1 - b_2, a_2 - b_1]$,

Multiplication:

$A \otimes B \in [a_1, a_2] \otimes [b_1, b_2] = [\min\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\},\ \max\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}]$,

Division:

$A \oslash B \in [a_1, a_2] \oslash [b_1, b_2] = [a_1, a_2] \otimes [1/b_1, 1/b_2] =$

$= [\min\{a_1/b_1,\ a_1/b_2,\ a_2/b_1,\ a_2/b_2\},\ \max\{a_1/b_1,\ a_1/b_2,\ a_2/b_1,\ a_2/b_2\}]$,

where $0 \notin [b_1, b_2]$.

Furthermore, the additive inverse of $A \in [a_1, a_2]$ is given by:

$$-A \in [-a_2,\ -a_1],$$

while its multiplicative inverse is given by:

$$A^{-1} \in [1/a_2,\ 1/a_1].$$

For scalar multiplication by a positive real number *k*, we obtain:

$$k \otimes A \in [ka_1,\ ka_2],$$

while for exponent*s* we have:

$$A^k \in [a_1{}^k,\ a_2{}^k].$$

Algebraically, the set of all grey numbers forms a field, and the totality of all grey numbers constitutes a grey linear space. Any matrix containing grey entries will be referred to as a grey matrix.

A grey number $A \in [a_1, a_2]$, with $a_1 = a_2$ is called a white number. When $A \in (-\infty, \infty)$ then it is called a black number. Also, the technique to transfer a grey number $A \in [a_1, a_2]$ into a white number $w(A)$ is called whitenization. That is, $w(A)$ is the white number after the whitenization of the grey number *A*, which has the highest probability to be the representative real value of *A*.

Whitening is a weight function which usually is defined as follows:

$$w(A) = (1-\gamma)a_1 + \gamma a_2, \quad \gamma \in [0,1], \tag{11}$$

which is known as the equal weight whitenization. If $\gamma = 1/2$, the resultant whitenization is called equal-weight mean whitenization (often in practice, when the distribution of a grey number is unknown, the equal-weight mean whitenization is employed)[18, p.24].

3. Mathematical Framework of CAGD

Let us assume that *K* is a two-way $(n \times p)$ contingency table, where some or all of its elements (k_i^j) are grey nonnegative numbers (note that a real number is a special case of a grey number $A \in [a_1, a_2]$, if $a_1 = a_2$).

In accordance to a previous work[25], we transform here the initial grey matrix K to the grey matrix of row-profiles *X* and we calculate the associated grey matrices Q and D. Thus, by using the grey arithmetic operations, we find the grey $(p \times p)$ nonnegative matrix:

$$S_{p,p} = X^{\mathrm{T}} \otimes D \otimes X \otimes Q,$$

where for simplicity we consider the following case:

$p = \operatorname{rank}(S) \leq \min\{n, p\}.$

The entries (s_i^j), $\left(i, j = 1,2,...,p\right)$ of S are clearly nonnegative grey numbers, because all the grey matrices X, Q and D are in the nonnegative real line $\mathbb{R}^+$. Thus, this grey extension of CA (Correspondence Analysis for Grey Data CAGD) leads finally to the grey eigenvalue problem of the grey matrix S. Therefore, the grey eigenvalues of S are the grey numbers Λ, which are the solutions of the grey characteristic equation:

$$S \otimes u = \Lambda \otimes u\ , \tag{12}$$

where:

$$S_{p,p} = X^{\mathrm{T}} \otimes D \otimes X \otimes Q = (s_i^j) = [(s_i^j)_\ell, (s_i^j)_r] \geq 0\,,$$

$$u^{\mathrm{T}} = (u_j) = (u_1, \ldots, u_j, \ldots, u_p)\,,$$

and Λ, u_j are grey numbers, or, equivalently, according to the grey arithmetic we obtain:

$$\begin{bmatrix} s_1^1 & \cdots & s_1^j & \cdots & s_1^p \\ \vdots & & \vdots & & \vdots \\ s_i^1 & \cdots & s_i^j & \cdots & s_i^p \\ \vdots & & \vdots & & \vdots \\ s_p^1 & \cdots & s_p^j & \cdots & s_p^p \end{bmatrix} \otimes \begin{bmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_p \end{bmatrix} = \Lambda \otimes \begin{bmatrix} u_1 \\ \vdots \\ u_j \\ \vdots \\ u_p \end{bmatrix},$$

which is equivalent to the following system of equations:

$$(s_i^1 \otimes u_1) \oplus \cdots \oplus (s_i^j \otimes u_j) \oplus \cdots \oplus (s_i^p \otimes u_p) = \Lambda \otimes (u_j),\ \ (i, j = 1,2,\ldots,p).$$

Thus, the algebraic foundations of CAGD lead to the grey eigenvalue problem, i.e., to the solution of the above grey equations (12).

Similarly to the two-step method introduced in the paper[25], for $a = 0$, in the case of CAGD we can determine the grey eigenvalues Λ of the grey matrix S, as follows:

1) At the first step, we write the CA-grey matrix S as an interval matrix, which its entries are closed, bounded and nonnegative real intervals,

$$S = (s_i^j) = [(s_i^j)_\ell, (s_i^j)_r] = [S_\ell, S_r].$$

2) At the second step the interval matrix S is represented by a line-segment $[S_\ell, S_r]$ of the matrix vector space $\mathrm{M}_p(\mathbb{R})$. In order to find the ordinary eigenvalues λ_ℓ of $S_\ell = (s_i^j)_\ell$ and λ_r of $S_r = (s_i^j)_r$ the line-segment $[S_\ell, S_r]$ of the vector space $\mathrm{M}_p(\mathbb{R})$ is transformed into the line-segment $\Lambda = [\lambda_\ell, \lambda_r]$ of the vector space $\mathbb{R}$.

Therefore, the grey eigenvalue problem is equivalently reduced to equations (12), as follows:

$$[S_\ell, S_r] \otimes [u_\ell, u_r] = [\lambda_\ell, \lambda_r] \otimes [u_\ell, u_r]\,, \tag{13}$$

Each entry-interval $[(s_i^j)_\ell, (s_i^j)_r] = (s_i^j)$ of the matrix S can be expressed as a convex combination of its boundaries $(s_i^j)_\ell$ and $(s_i^j)_r$. More specifically we write

$$[(s_i^j)_\ell, (s_i^j)_r] = (1-\theta)(s_i^j)_\ell + \theta(s_i^j)_\ell\ ,\ \ \theta \in \left[0,1\right],$$

or $$[S_\ell, S_r] = \left\{ S(\theta) \,/\, (1-\theta)S_\ell + \theta S_r,\ \ \theta \in [0,1] \right\}, \tag{14}$$

where S_ℓ and S_r are the crisp matrices of left-end and right-end points, respectively. Relation (14) is just a line-segment in $\mathrm{M}_p(\mathbb{R}^+)$ that expresses in matrix form the entries of S.

Thus, by using the two-step method for the eigenvalue problem of a $(p \times p)$ CAGD-grey matrix S, similarly as it has been proved in the paper [25], and for the case $a = 0$ especially, the proofs of the following theoretical results are evident:

Proposition 1 The process of obtaining any ordinary eigenvalue $\lambda(\theta)$ of the crisp matrix $S(\theta)$ can be expressed by the eigenvalue function-transformation E_λ,

$$E_\lambda : \Theta \subseteq \mathrm{M}_p(\mathbb{R}^+) \to \mathbb{R}^+ \,//\, S(\theta) \mapsto E_\lambda(S(\theta)) = \lambda(\theta), \tag{15}$$

where:

$$\Theta = \{ S(\theta) \,/\, S(\theta) = (1-\theta)S_\ell + \theta S_r,\ \theta \in [0,1] \}, \tag{16}$$

is a family of crisp matrices with nonnegative real elements constituting a convex polytope.

Remark 1 The eigenvalue function transformation E_λ can also be considered as a linear functional on a convex polyhedron.

Theorem 1 (Linearity of the eigenvalue function-transformation). Let S be a $(p \times p)$ CA-grey matrix. Then the eigenvalue function-transformation E_λ is linear; that is, by setting

$E_\lambda(S_\ell) = \lambda_\ell$ and $E_\lambda(S_r) = \lambda_r$,

we have:

$$E_\lambda((1-\theta)S_\ell + \theta S_r) = (1-\theta)E_\lambda(S_\ell) + \theta E_\lambda(S_r) = (1-\theta)\lambda_\ell + \theta\lambda_r . \tag{17}$$

Corollary 1 The eigenvalue function transformation

$$E_\lambda : \Theta \to \mathbb{R}^+ \,//\, S(\theta) \mapsto \lambda(\theta),$$

is uniformly continuous and increasing, for each eigenvalue of the $(p \times p)$ matrix $S(\theta),\ \theta \in [0,1]$

Theorem 2 Every point

$$\lambda(\theta) = (1-\theta)\lambda_\ell + \theta\lambda_r,\quad \theta \in [0,1],$$

in the interval $\Lambda = [\lambda_\ell, \lambda_r]$ provided by the two-step method, satisfies the corresponding characteristic equation; that is, for every $\theta \in [0,1]$ we have,

$$S(\theta)u(\theta) = \lambda(\theta)u(\theta),$$

since $S(\theta), u(\theta)$ and $\lambda(\theta)$ are crisp data.

Theorem 3 Any $(p \times p)$ CAGD-grey matrix S has grey eigenvalues Λ_1, $\Lambda_2, \ldots, \Lambda_p$, that are nonnegative grey (interval) numbers, solution of the grey characteristic equation (12).

4. Practical Application and Defuzzified – Geometrical Display of CAGD

4.1 A representative numerical example

The main ideas of our approach are illustrated in the following representative example taken from real life applications. This example has been studied in detail in[25] and we use it here for comparison purposes.

Consider three groups of candidate-students who are preparing to enter university, where each group represents a different geographical community, indicated as, *Stud-1*, *Stud-2* and *Stud-3*. In order to succeed in university, the students should choose to participate at exactly one from a choice of three directions-departments of the university, indicated as *Mathematic*s *(MATH)*, *Physic*s *(PHYS*) and *Literatur*e *(LITR)*. Education experts make an approximate estimation (using grey numbers) for how many candidates from each "student group-community" will succeed at each "department", which is exhibited in Table 1.

Table 1 Grey contingency table that provides "the success of candidates" with respect to "student groups-communities" and to "departments"

Name	*MATH*	*PHYS*	*LITR*
Stud-1	[6, 10]	[20, 26]	[12, 24]
Stud-2	[15, 25]	[30, 50]	[17, 17]
Stud-3	[1, 3]	[5, 5]	[7, 9]

Therefore, in this case, we have a 3 by 3 grey nonnegative matrix (grey contingency table) that provides "the success of candidates" with respect to "student groups–communities" and to "departments", i.e., the occurrences of two qualitative variables *I*, *J*, where the rows (*I*) are the "student groups– communities", *Stud-1*, *Stud-2* and *Stud-3*, and the columns (*J*) are the "departments" of the university, *MATH*, *PHYS* and *LITR*.

For this simple representative problem, we can obtain the initial grey contingency table *K* according to CAGD. More specifically we obtain the following grey matrix:

$$K_{3,3} = \begin{bmatrix} k_1^1 \in [6,10] & k_1^2 \in [20,26] & k_1^3 \in [12,24] \\ k_2^1 \in [15,25] & k_2^2 \in [30,50] & k_2^3 \in [17,17] \\ k_3^1 \in [1,3] & k_3^2 \in [5,5] & k_3^3 \in [7,9] \end{bmatrix},$$

where the entries of the 3×3 grey initial matrix *K* are considered as nonnegative grey numbers.

4.2 The two-step method for computing the grey eigenvalues of CAGD

Using INTLAB[23], and according to the process of the standard CA[24, 25], we get the associated to *K* grey (interval) matrices:

$$X_{3,3} = \begin{bmatrix} x_1^1 \in [0.1, 0.26] & x_1^2 \in [0.33, 0.68] & x_1^3 \in [0.2, 0.63] \\ x_2^1 \in [0.16, 0.4] & x_2^2 \in [0.33, 0.81] & x_2^3 \in [0.18, 0.27] \\ x_3^1 \in [0.06, 0.23] & x_3^2 \in [0.29, 0.38] & x_3^3 \in [0.41, 0.69] \end{bmatrix},$$

$$D_{3,3} = \begin{bmatrix} [0.23,0.53] & 0 & 0 \\ 0 & [0.37,0.82] & 0 \\ 0 & 0 & [0.08,0.15] \end{bmatrix},$$

and $$Q_{3,3} = \begin{bmatrix} [2.97,7.69] & 0 & 0 \\ 0 & [1.39,2.75] & 0 \\ 0 & 0 & [2.26,4.7] \end{bmatrix}.$$

Then, we get the nonnegative interval matrix,

$$S_{p,p} = X_{p,n}^T \otimes D_{n,n} \otimes X_{n,p} \otimes Q_{p,p},$$

which is associated to *K*:

$$S_{3,3} = X^{\mathrm{T}} \otimes D \otimes X \otimes Q = (s_i^j) = [(s_i^j)_\ell,\ (s_i^j)_r] = [S_\ell,\ \mathrm{S}_r] =$$

$$= \begin{bmatrix} s_1^1 \in [0.04,1.36] & s_1^2 \in [0.04,1.15] & s_1^3 \in [0.04,0.95] \\ s_2^1 \in [0.08,2.88] & s_2^2 \in [0.10,2.46] & s_2^3 \in [0.11,2.11] \\ s_3^1 \in [0.05,1.56] & s_3^2 \in [0.06,1.38] & s_3^3 \in [0.08,1.62] \end{bmatrix}.$$

From where we get the ordinary square matrices:

$$S_\ell = (s_i^j)_\ell = \begin{bmatrix} 0.04 & 0.04 & 0.04 \\ 0.08 & 0.10 & 0.11 \\ 0.05 & 0.06 & 0.08 \end{bmatrix}$$

and $S_r = (s_i^j)_r = \begin{bmatrix} 1.36 & 1.15 & 0.95 \\ 2.88 & 2.46 & 2.11 \\ 1.56 & 1.38 & 1.62 \end{bmatrix}$.

Thus, according to the two-step method, we have,

$$S = (s_i^j) = [(s_i^j)_\ell,\ (s_i^j)_r] = [S_\ell,\ \mathrm{S}_r] = \left\{ S(\theta) \mid S(\theta) = (1-\theta)S_\ell + \theta S_r,\ \theta \in [0,1] \right\},$$

from where we can get the corresponding interval (or grey) eigenvalues

$$\Lambda = [\lambda_\ell, \lambda_r] = \left\{ \lambda(\theta) \mid \lambda(\theta) = (1-\theta)\lambda_\ell + \theta\lambda_r,\ \theta \in [0,1] \right\}.$$

Equivalently, we can take the same eigenvalues by diagonalization of the corresponding to S grey *symmetric* matrix (cf. Eq. (6)),

$$A_{p,p} = Q_{p,p}^{1/2} \otimes X_{p,n}^{\mathrm{T}} \otimes D_{n,n} \otimes X_{n,p} \otimes Q_{p,p}^{1/2},$$

which, for this example, using the grey (or interval) arithmetic is calculated as follows:

$$A_{3,3} = Q_{3,3}^{1/2} \otimes X_{3,3}^{\mathrm{T}} \otimes D_{3,3} \otimes X_{3,3} \otimes Q_{3,3}^{1/2} =$$

$$= \begin{bmatrix} [1.72,2.77] & 0 & 0 \\ 0 & [1.18,1.66] & 0 \\ 0 & 0 & [1.5,2.17] \end{bmatrix} \otimes$$

$$\otimes \begin{bmatrix} [0.1,0.26] & [0.33,0.68] & [0.2,0.63] \\ [0.16,0.4] & [0.33,0.81] & [0.18,0.27] \\ [0.06,0.23] & [0.29,0.38] & [0.41,0.69] \end{bmatrix}^{\mathrm{T}} \otimes$$

$$\otimes \begin{bmatrix} [0.23,0.53] & 0 & 0 \\ 0 & [0.37,0.82] & 0 \\ 0 & 0 & [0.08,0.15] \end{bmatrix} \otimes$$

$$\otimes \begin{bmatrix} [0.1,0.26] & [0.33,0.68] & [0.2,0.63] \\ [0.16,0.4] & [0.33,0.81] & [0.18,0.27] \\ [0.06,0.23] & [0.29,0.38] & [0.41,0.69] \end{bmatrix} \otimes$$

$$\otimes \begin{bmatrix} [1.72,2.77] & 0 & 0 \\ 0 & [1.18,1.66] & 0 \\ 0 & 0 & [1.5,2.17] \end{bmatrix}.$$

Thus, we obtain

$$A_{3,3}=\begin{bmatrix} a_1^1\in[0.04,1.34] & a_1^2\in[0.06,1.71] & a_1^3\in[0.04,1.2] \\ a_2^1\in[0.06,1.71] & a_2^2\in[0.1,2.22] & a_2^3\in[0.08,1.6] \\ a_3^1\in[0.04,1.2] & a_3^2\in[0.08,1.6] & a_3^3\in[0.08,1.6] \end{bmatrix}.$$

Therefore, we have

$$A_\ell=(a_i^j)_\ell=\begin{bmatrix} 0.04 & 0.06 & 0.04 \\ 0.06 & 0.1 & 0.08 \\ 0.04 & 0.08 & 0.08 \end{bmatrix},$$

and

$$A_r=(a_i^j)_r=\begin{bmatrix} 1.34 & 1.71 & 1.2 \\ 1.71 & 2.22 & 1.6 \\ 1.2 & 1.6 & 1.6 \end{bmatrix}.$$

Consequently, according to the two-step method, we obtain,

$$A=(a_i^j)=[(a_i^j)_\ell,\ (a_i^j)_r]=[A_\ell,\ \mathrm{A}_r]=$$
$$=\left\{A(\theta)\mid A(\theta)=(1-\theta)A_\ell+\theta A_r,\ \theta\in[0,1]\right\},$$

from where we can get the corresponding interval (or grey) eigenvalues:

$$\Lambda=[\lambda_\ell,\lambda_r]=\left\{\lambda(\theta)\mid\lambda(\theta)=(1-\theta)\lambda_\ell+\theta\lambda_r,\ \theta\in[0,1]\right\}.$$

The respective ordinary arithmetical matrices $A(\theta)$ from the family

$$A=\left\{A(\theta)\mid A(\theta)=(1-\theta)A_\ell+\theta A_r,\ \theta\in[0,1]\right\},$$

for various values of $[\in \theta 0,1]$ and with the corresponding eigenvalues-eigenvectors are as follows:

$$A(\theta=0)=(1-\theta)A_\ell+\theta A_r=A_\ell=(a_i^j)_\ell=\begin{bmatrix} 0.04 & 0.06 & 0.04 \\ 0.06 & 0.1 & 0.08 \\ 0.04 & 0.08 & 0.08 \end{bmatrix},$$

with the corresponding eigenvalues-eigenvectors:

$$\lambda_{1(\theta=0)}=(\lambda_1)_\ell=0.2022 \leftrightarrow w_{1(\theta=0)}=\begin{pmatrix} 0.4041 \\ 0.6991 \\ 0.5899 \end{pmatrix},$$

$$\lambda_{2(\theta=0)}=(\lambda_2)_\ell=0.0178 \leftrightarrow w_{2(\theta=0)}=\begin{pmatrix} 0.6263 \\ 0.2586 \\ -0.7354 \end{pmatrix},$$

$$\lambda_{3(\theta=0)}=(\lambda_3)_\ell=-4.9009\times10^{-30}\simeq 0 \leftrightarrow w_{3(\theta=0)}=\begin{pmatrix} 0.6667 \\ -0.6667 \\ 0.3333 \end{pmatrix},$$

and $$A(\theta=1)=(1-\theta)A_\ell+\theta A_r=A_r=(a_i^j)_r=\begin{bmatrix} 1.34 & 1.71 & 1.2 \\ 1.71 & 2.22 & 1.6 \\ 1.2 & 1.6 & 1.6 \end{bmatrix},$$

with the corresponding eigenvalues-eigenvectors:

$$\lambda_{1(\theta=1)} = (\lambda_1)_r = 4.7967 \leftrightarrow \mathrm{w}_{1(\theta=1)} = \begin{pmatrix} 0.5166 \\ 0.6721 \\ 0.5303 \end{pmatrix},$$

$$\lambda_{2(\theta=1)} = (\lambda_2)_r = 0.3512 \leftrightarrow w_{2(\theta=1)} = \begin{pmatrix} 0.3873 \\ 0.3689 \\ -0.8448 \end{pmatrix},$$

$$\lambda_{3(\theta=1)} = (\lambda_3)_r = 0.0121 \leftrightarrow w_{3(\theta=1)} = \begin{pmatrix} 0.7635 \\ -0.6419 \\ 0.0698 \end{pmatrix},$$

as well as

$$A\left(\theta = \frac{1}{2}\right) = (1-\theta)A_\ell + \theta A_r = \frac{1}{2}A_\ell + \frac{1}{2}A_r =$$

$$= \frac{1}{2}\begin{bmatrix} 0.04 & 0.06 & 0.04 \\ 0.06 & 0.1 & 0.08 \\ 0.04 & 0.08 & 0.08 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1.34 & 1.71 & 1.2 \\ 1.71 & 2.22 & 1.6 \\ 1.2 & 1.6 & 1.6 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.69 & 0.88 & 0.62 \\ 0.88 & 1.16 & 0.84 \\ 0.62 & 0.84 & 0.84 \end{bmatrix},$$

with the corresponding eigenvalues-eigenvectors:

$$\lambda_{1(\theta=1/2)} = 2.4945 \leftrightarrow w_{1(\theta=1/2)} = \begin{pmatrix} 0.5117 \\ 0.6733 \\ 0.5336 \end{pmatrix},$$

$$\lambda_{2(\theta=1/2)} = 0.1831 \leftrightarrow w_{2(\theta=1/2)} = \begin{pmatrix} 0.4023 \\ 0.3609 \\ -0.8413 \end{pmatrix},$$

$$\lambda_{3(\theta=1/2)} = 0.0124 \leftrightarrow w_{3(\theta=1/2)} = \begin{pmatrix} 0.7591 \\ -0.6452 \\ 0.0861 \end{pmatrix}.$$

Also, the ordinary eigenvalues-eigenvectors that have been provided from initial matrix K for the mean of each grey entry, (that is for the corresponding classical CA), are as follows:

$$\lambda_1 = 1 \;\leftrightarrow\; w_1 = \begin{pmatrix} 0.46 \\ 0.69 \\ 0.55 \end{pmatrix},$$

$$\lambda_2 = 0.05 \;\leftrightarrow\; w_2 = \begin{pmatrix} 0.48 \\ 0.33 \\ -0.81 \end{pmatrix},$$

$$\lambda_3 = 0.006 \;\leftrightarrow\; w_3 = \begin{pmatrix} 0.75 \\ -0.64 \\ 0.18 \end{pmatrix},$$

where w_1, w_2 and w_3 are the classical eingenvectors of the ordinary symmetric matrix A_{mean} which are pairwise orthogonal.

Therefore, the grey eigenvalues Λ_1, Λ_2 and Λ_3 of the CAGD-matrix A (or equivalently of S) are grey numbers, as follows:

$$\Lambda_1 = [(\lambda_1)_\ell,\ (\lambda_1)_r)] = [\lambda_1(\theta = 0),\ \lambda_1(\theta = 1)] = [0.2022,\ 4.7967]\ ,$$
$$\Lambda_2 = [(\lambda_2)_\ell,\ (\lambda_2)_r)] = [\lambda_2(\theta = 0),\ \lambda_2(\theta = 1)] = [0.0178,\ 0.3512]\ ,$$
$$\Lambda_3 = [(\lambda_3)_\ell,\ (\lambda_3)_r)] = [\lambda_3(\theta = 0),\ \lambda_3(\theta = 1)] = [0.0,\ \ 0.0121]\ .$$

In comparison to the standard CA, the following inequalities-relations hold:
Left limits of grey-interval eigenvalues Λ of CAGD:

$$(\lambda_1)_\ell = 0.2022 > (\lambda_2)_\ell = 0.0178 > (\lambda_3)_\ell = 0.0.$$

The crisp eigenvalues of standard CA:

$$\lambda_1 = 1 > \lambda_2 = 0.05 > \lambda_3 = 0.006.$$

Right limits of grey-interval eigenvalues Λ of CAGD:

$$(\lambda_1)_r = 4.7967 > (\lambda_2)_r = 0.3512 > (\lambda_3)_r = 0.0121.$$

5. Synopsis and Concluding Remarks

In the paper at hand, we give an enrichment to tackle uncertainties caused by grey data of a previous work[25], where the mathematical foundation and the algebraic treatment of the fuzzy eigenvalue problem of correspondence analysis with fuzzy data have been investigated.

More specifically, we enrich here these theoretical results as well as the mathematical foundation on grey contingency table in the case where its entries-data are grey numbers. We name this approach correspondence analysis with grey data.

In the paper at hand, we mainly focus on the grey eigenvalue problem and we show that the correspondence analysis can be enriched to tackle uncertainties caused by grey data.

Furthermore, the proposed approach is validated by using data from a real-life application and the corresponding computational processes are explained in detail through a simple representative numerical example.

Our experience is that the proposed procedures can be easily implemented computationally. Moreover, although the obtained results using the proposed approach are comparable to those obtained using fuzzy numbers the corresponding required computational burden is significantly reduced.

Acknowledgements

The authors wish to thank the anonymous reviewers for their careful reading and helpful comments.

References

[1] Alefeld, G. and Herzberger, J. Introduction to interval computation, Academic Press, New York, 1983.
[2] Andrew, A.M. Why the world is grey, Grey Systems: Theory and Application, 2011, 1(2):112–116.
[3] Benzécri, J.P. (Ed.) . L'analyse des données, Tome 1: La taxinomie, Dunod, Paris, 1973.
[4] Benzécri, J.P. and Benzécri, F. L'analyse des données, Tome 2: L'Analyse des correspondances, 4e éd, Dunod, Paris, 1984.
[5] Bertier, P. and Bouroche, J.M. Analyse des données multidimensionnelles, PUF, Paris, 1977.
[6] Deng, J.L. Control problems of grey systems, Systems & Control Letters, 1982, 1(5): 288–294.
[7] Deng, J.L. Introduction to grey system theory, Journal of Grey System, 1989, 1(1):1–24.
[8] Hirschfeld, H.O. A connection between correlation and contingency, Proceedings of the Cambridge Philosophical Society, 1935, 31:520–524.
[9] Lebart, L., Morineau, A. and Warwick, K. Multivariate descriptive statistical analysis, John Wiley and Sons, Chichester, 1984, (4th French edition, Dunod, Paris, 2006).
[10] Li Q.X. and Liu, S.F. Some results about grey mathematics, Kybernetes, 2009, 38(3/4): 297–305.
[11] Lin Y. and Liu, S.F. A systemic analysis with data, International Journal of General Systems, 2000, 29(6):989–999.
[12] Lin Y. and Liu, S.F. Solving problems with incomplete information: A grey systems approach, In "Advances in imaging and electron physics", Vol. 141 (Ed. P. Hawkes), Elsevier, Oxford, pp.77–174, 2006.
[13] Lin Y., Chen M.Y. and Liu, S.F. Theory of grey systems: Capturing uncertainties of grey information, Kybernetes, 2004, 33(2):196–218.
[14] Liu, S.F., Fang, Z.G., and Yang, Y.J. and Forrest, J. General grey numbers and their operations, Grey Systems: Theory and Application, 2012, 2(3):341–349.
[15] Liu, S.F., Forrest, J. and Yang, Y.J. A brief introduction to grey systems theory, Grey Systems: Theory and Application, 2012, 2(2):89–104.
[16] Liu, S.F., Forrest, J. and Yang, Y.J. Grey system: Thinking, methods, and models with applications, In "Contemporary issues in systems science and engineering" (Eds. M.C. Zhou, H.X. Li and M. Weijnen), IEEE Press, John Wiley and Sons, Inc., Haboken, New Jersey, pp.153–224, 2015.
[17] Liu, S.F. and Lin, Y. Grey information: Theory and practical applications, Springer, London, 2006.
[18] Liu, S.F. and Lin, Y. Grey systems: Theory and applications, Springer, Berlin/Heidelberg, 2010.
[19] Liu, S.F. and Lin, Y. (Eds.). Advances in grey systems research, Springer, Berlin/Heidelberg, 2010.
[20] Moore, R.E. (1995): Methods and applications of interval analysis, 2nd Printing, SIAM, Philadelphia.
[21] Moore, R.E., Kearfott, R.B. and Cloud M.J. Introduction to interval analysis, SIAM, Philadelphia, 2009.
[22] Roux, Br. and Rouanet, H. Geometric data analysis, Kluwer Academic Publishers, Dordrecht, 2004.
[23] Rump, S.M. INTLAB–INTerval LABoratory. In "Developments in reliable computing" (Ed. T. Csendes), Kluwer Academic Publishers, Dordrecht, pp.77–104, 1999.
[24] Theodorou, Y., Alevizos, Ph. and Kechriniotis, A. Correspondence Analysis for Fuzzy Data (CAFD): The practical application; Representative example with defuzzified-geometrical display, International Journal of Applied Mathematics and Statistics, 2012, 25(1):1–19.
[25] Theodorou, Y., Drossos, C. and Alevizos, Ph. Correspondence analysis with fuzzy data: The fuzzy eigenvalue problem, Fuzzy Sets and Systems, 2007, 158:704–721.