The Sixth International Symposium on Operations Research and Its Applications (ISORA'06) Xinjiang, China, August 8–12, 2006 Copyright © 2006 ORSC & APORC, pp. 95–105

Dual Scaling Using Mathematical Programming and Its Application

Tohru Ueda*

Seikei University, 3-3-1 Kichijoji-Kitamachi, Musashino, Tokyo 180-8633, Japan

Abstract In subjective performance measurement, paired comparison data or successive categories data are often utilized. The AHP or conjoint analysis is not very appropriate for aggregated evaluation of these data, but Dual Scaling aims preferably aggregated evaluation. Application and easy formulation of Dual Scaling for these data are proposed.

Keywords dual scaling; AHP; mathematical programming; paired comparison

1 Introduction

In subjective performance measurement, paired comparison data or successive categories data are often utilized. Individual evaluation can be obtained by the AHP (Analytic Hierarchy Process), conjoint analysis and so on. However, these methods are not appropriate for aggregated evaluation, that is, overall evaluation.

Dual scaling aims preferably aggregated evaluation. Formulation of dual scaling for successive categories data proposed by Nishisato is troublesome. However, in mathematical programming system successiveness is presented easily as constraints. Merits of mathematical programming system are easiness of addition and modification on objective functions and constraints. Thus mathematical programming models with various objective functions are proposed. Also a model which treats fuzzy numbers is proposed for the purpose of presenting lack of assurance or vagueness in answers.

EspeciallyI I propose methods of analyzing paired comparison data or successive categories data

2 Dual scaling for successive data

In this section a method proposed by Nishisato (1980) is shown. Let us consider the situation where each of N subjects evaluates M objects. Evaluation is done by selection of a category, where K categories are put in order, the category K shows the best one and the category 1 shows the worst one. Moreover, suppose that there

^{*}E-mail: ueda@st.seikei.ac.jp

is a boundary value, τ_k , between category *k* and category (*k*+1), where a relation $(\tau_k \le \tau_{k+1})$ must be hold. If a value, μ_n , given for object O_n satisfies a relation $(\tau_k < \mu_n \le \tau_{k+1})$, object O_n belongs to category *k*.

A component, f_{ik} , of a data matrix, F, with N rows and $\{M(K-1)\}$ columns is given by

$$f_{i,j(K-1)+h} = \begin{cases} 1 & \text{subject } i \text{ evaluates as } \mu_j < \tau_h \\ -1 & \text{subject } i \text{ evaluates as } \mu_j > \tau_h \end{cases}$$

Suppose that there are three objects, O_1 , O_2 and O_3 . If subject *i* evaluates objects, O_1 , O_2 and O_3 as category 1, 2 and 3 respectively, the *i*-th row of matrix *F* is

 $1 \quad 1 \quad | \quad -1 \quad 1 \quad | \quad -1 \quad -1$

The design matrix, A, with $\{M(K-1)\}$ rows and $\{M+(K-1)\}$ columns is given by

$$A = \begin{bmatrix} I_m & -1_m & 0_m & \cdots & 0_m \\ I_m & 0_m & -1_m & \cdots & 0_m \\ I_m & 0_m & 0_m & \cdots & 0_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_m & 0_m & 0_m & \cdots & -1_m \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ M \end{bmatrix}$$
(1)

where m = K - 1, $I_m = m \times m$ unit matrix, -1_m and 0_m : column vectors with all of *m* elements equal to (-1) and 0, respectively. Let a parameter vector be

$$x = (\tau_1, \tau_2, \cdots, \tau_{K-1}, \mu_1, \mu_2, \cdots, \mu_M)^t.$$

The first block of m elements of Ax is

$$(\tau_1-\mu_1,\tau_2-\mu_1,\cdots,\tau_{K-1}-\mu_1)^t.$$

The *h*-th block of *m* elements of *Ax* is

$$(\tau_1-\mu_h,\tau_2-\mu_h,\cdots,\tau_{K-1}-\mu_h)^t.$$

This means that Ax can be used at evaluation between τ_k and μ_h . In order to use at evaluation among τ_k or among μ_h matrices F_a and A_a are also introduced. A parameter vector, x, is obtained as an eigenvector corresponding to the maximum eigenvalue of $E^t E$ where

$$E = \begin{bmatrix} F, F_a \end{bmatrix} \begin{bmatrix} A \\ A_a \end{bmatrix}$$
(2)

This formulation means derivation of the maximum between-group variance under constant total variance of x. A vector presenting subjects

$$y = (y_1, y_2, ..., y_N)^t$$

. .

is obtained as an eigenvector corresponding to the maximum eigenvalue of EE^t . Between x and y the relations

$$x = aE^{t}y; \quad y = bEx \qquad (a,b:scalar)$$
 (3)

hold.

3 Formulation as mathematical programming for successive data

Table 1 is used as an example for explanation. Suppose that subject *i* is given a value, y_i , object *j* is given a value, x_j and category *k* is given a value, t_k . Let e_{ij} be a value as which subject *i* evaluates object *j*. For example, since subject 1 evaluated object A(=1) as category 2 and object B(=2) as category 3, $e_{11} = t_2$ and $e_{12} = t_3$. Then, the following formulations can be considered.

Formulation 1

• •

$$\max \sum_{i=1}^{N} (y_i - \mu)^2$$
 (4)

s.t.
$$\sum_{i=1}^{N} y_i / N = \mu = 0; \quad y_i = \sum_{j=1}^{M} e_{ij} / M$$
 (5)

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (e_{ij} - \mu)^2 / (NM) = 1$$
(6)

$$t_2 - t_1 \ge C, \ t_3 - t_2 \ge C, \ \cdots, t_K - t_{K-1} \ge C \tag{7}$$

 μ : grand mean; *C*: a nonnegative constant

Here, let h_{ik} be a number of objects which subject *i* evaluates as category *k*. Then

$$y_i = \sum_{k=1}^{K} h_{ik} t_k / M.$$
(8)

Objective, max $\sum_{i=1}^{N} (y_i - \mu)^2 + kC^2$, may be used, instead of max $\sum_{i=1}^{N} (y_i - \mu)^2$. Formulation 2

$$\max \sum_{j=1}^{M} (x_j - \mu)^2$$
 (9)

s.t.
$$\sum_{j=1}^{M} x_j / M = \mu = 0; \quad x_j = \sum_{i=1}^{N} e_{ij} / N$$
 (10)

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (e_{ij} - \mu)^2 / (NM) = 1$$
(11)

$$t_2 - t_1 \ge C, \ t_3 - t_2 \ge C, \ \cdots, t_K - t_{K-1} \ge C$$
 (12)

 μ : grand mean; C: a nonnegative constant

If values t_k ($k = 1, \dots, K$) which are obtained under Formulation 1 are substituted into Eq.10, *x* different from Formulation 2 is obtained. Inversely, if values t_k ($k = 1, \dots, K$) which are obtained under Formulation 2 are substituted into Eq.5, *y* different from Formulation 1 is obtained. This means that there may be various solutions according objectives and constraints. Compounded objectives for *x* and *y* may be desired. The following objectives can be considered.

(i)
$$\max \sum_{j=1}^{M} (x_j - \mu)^2 / M + \sum_{i=1}^{N} (y_i - \mu)^2 / N$$
 (13)

(ii) max
$$w_1 \sum_{j=1}^{M} (x_j - \mu)^2 / M + w_2 \sum_{i=1}^{N} (y_i - \mu)^2 / N$$
 (14)

(iii)
$$\max \sum_{j=1}^{M} (x_j - \mu)^2 / M; \quad \sum_{i=1}^{N} (y_i - \mu)^2 / N \ge C_1$$
 (15)

(iv)
$$\max \sum_{i=1}^{N} (y_i - \mu)^2 / N; \quad \sum_{j=1}^{M} (x_j - \mu)^2 / M \ge C_2$$
 (16)

(v) min
$$\sum_{i=1}^{N} (y_i - \mu)^2 / N;$$
 $\sum_{j=1}^{M} (x_j - \mu)^2 / M \ge C_2$ (17)

(vi)
$$\max \sum_{k} (t_{k+1} - t_k)^2$$
 (18)

Here, objective (v) is different from others, because it aims at minimum difference among subjects, while others aim at maximum difference among subjects.

4 A model which treats fuzzy numbers

Also a model which treats fuzzy numbers is proposed for the purpose of presenting lack of assurance or vagueness in answers. Let t_k be triangular fuzzy numbers with lower bound $(t_k - c_k)$, mode t_k and upper bound $(t_k + d_k)$. The following formulation corresponding to Formulation 1 is proposed, where

$$T_k = \{(t_k - c_k) + 2t_k + (t_k + d_k)\}/4$$
(19)

Formulation F1

$$\max \sum_{i=1}^{N} (y_i - \mu)^2 - \sum_{k=1}^{K} (c_k^2 + d_k^2)$$
(20)

s.t.
$$\sum_{i=1}^{N} y_i / N = \mu = 0; \quad y_i = \sum_{k=1}^{K} h_{ik} T_k / M$$
 (21)

$$\sum_{k=1}^{K} H_k T_k^2 / (NM) = 1; \quad H_k = \sum_{i=1}^{N} h_{ik}$$
(22)

$$T_1 \le T_2 \le \dots \le T_K \tag{23}$$

 μ : grand mean

Table 1: Example 1							
A	В	C	D	Е			
2	3	3	3	3			
1	2	3	1	1			
2	3	2	2	1			
1	1	1	1	1			
3	1	1	2	3			
3	3	3	3	3			
2	2	2	2	1			
1	3	3	2	1			
3	1	1	3	3			
1	2	2	1	1			
19	21	21	20	18			
	A 2 1 2 1 3 3 2 1 3 1	A B 2 3 1 2 2 3 1 1 3 1 3 3 2 2 1 3 3 1 1 2 1 3 1 2 19 21	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$			

.

1: bad, 2: medium, 3: good

Instead of T_k^2 the following quantity can be used.

$$\left[\int_{0}^{1} \{t_{k} - c_{k}(1-\alpha)\}^{2} d\alpha + \int_{0}^{1} \{t_{k} + d_{k}(1-\alpha)\}^{2} d\alpha\right] / 2$$

$$= t_{k}^{2} - t_{k}(c_{k} - d_{k}) / 2 + (c_{k}^{2} + d_{k}^{2}) / 6$$
(24)

Under all formulations the same value for many t_k may be given. Therefore, a condition

$$t_k - t_{k-1} \ge C$$
 C : a positive constant (25)

may be necessary.

Example 1. Table 2 and Table 3 show values of x and y under Formulation 1 (max V(y)). Table 4 and Table 5 show values of x and y under Formulation 2 (maxV(x)). As shown in Table 3 and Table 5,

$$\max V(y) >> \max V(x)$$

Therefore, results of $\max\{V(y) + V(x)\}$ coincided with results of $\max V(y)$. Under Formulation F1, $c_k = d_k = 0$, that is, the same results as Formulation 1 were obtained.

Dual scaling for paired comparison data 5

In this section paired comparison data are treated on the line of Nishisato (1980, Sec.6.2). Suppose that N subjects evaluate pairs among M objects. If object h > 1object j, $g_{hj}=1$ is given and if object h < object j, $g_{hj}=0$. The data matrix F has N rows and M(M-1)/2 columns. Consider a situation where N=8 and M=4 (objects: A, B, C, D).

		А	В	С	D	E	Objective
Nishisato		-0.402(4)	0.440(1)	0.426(2)	-0.210(3)	-0.546(5)	
			· · /	· · ·	-0.125(4)		4.792
					-0.101(4)		4.774
	C = 0.4	-0.117(5)	0.105(1)	0.105(1)	-0.077(4)	-0.015(3)	4.734

Table 2: Values of x in Formulation 1

(): order

Table 3: Values of *y* in Formulation 1

	Nishisato	C=0	C=0.2	<i>C</i> =0.4	Category Sum
1	-0.070(5)	0.917(2)	0.938(2)	0.953(2)	14(2)
2	-0.413(10)	-0.333(6)	-0.357(7)	-0.379(7)	8(8)
3	-0.329(8)	-0.333(6)	-0.277(6)	-0.219(6)	10(4)
4	-0.106(6)	-0.750(8)	-0.829(10)	-0.903(10)	5(10)
5	0.362(3)	0.083(4)	0.074(4)	0.065(4)	10(4)
6	0.071(4)	1.333(1)	1.329(1)	1.317(1)	15(1)
7	-0.255(7)	-0.750(8)	-0.669(8)	-0.583(8)	9(7)
8	0.433(1)	0.083(4)	0.074(4)	0.065(4)	10(4)
9	-0.387(9)	0.500(3)	0.466(3)	0.429(3)	11(3)
10	0.407(2)	-0.750(8)	-0.749(9)	-0.743(9)	7(9)
V(y)		0.479	0.477	0.473	
V(x))	0.010	0.009	0.008	
V(y) + V(x)		0.490	0.486	0.482	

(): order

Table 4: Values of *x* in Formulation 2

		А	В	С	С	E	Objective
Nishisato					-0.210(3)		
					0.165(1)		
$\max V(x)$	<i>C</i> =0.2	-0.051(4)	0.163(1)	0.163(1)	0.143(3)	-0.418(5)	0.251
	<i>C</i> =0.4	-0.060(4)	0.160(1)	0.160(1)	0.120(3)	-0.381(5)	0.214

(): order

	Table 5: values of y in Formulation 2									
	Nishisato	C=0	<i>C</i> =0.2	<i>C</i> =0.4	Category sum					
1	-0.070(5)	0.783(1)	0.824(2)	0.861(2)	14(2)					
2	-0.413(10)	-0.453(8)	-0.458(8)	-0.461(8)	8(8)					
3	-0.329(8)	0.371(3)	0.317(3)	0.261(3)	10(4)					
4	-0.106(6)	-1.277(10)	-1.273(10)	-1.262(10)	5(10)					
5	0.362(3)	-0.041(5)	-0.031(6)	-0.020(6)	10(4)					
6	0.071(4)	0.783(1)	0.864(1)	0.941(1)	15(1)					
7	-0.255(7)	0.371(3)	0.277(4)	0.181(4)	9(7)					
8	0.433(1)	-0.041(5)	-0.031(6)	-0.020(6)	10(4)					
9	-0.387(9)	-0.041(5)	0.009(5)	0.060(5)	11(3)					
10	0.407(2)	-0.453(8)	-0.498(9)	-0.541(9)	7(9)					
V(y)		0.355	0.369	0.383						
V(x)		0.058	0.050	0.043						
V(y) + V(x)		0.413	0.419	0.426						
(): order										

Table 5: Values of *y* in Formulation 2

(): order

The *i*-th row shows comparison results of six pairs (A, B), (A, C), (A, D), (B, C), (B, D), (C, D). When F is given as follows, the first row means that

$$(A > B), (A < C), (A > D), (B < C), (B < D), (C > D)$$

where

The design matrix T is $M(M-1)/2 \times M$. At the above example

$$T = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{array}{c} AB \\ AC \\ AD \\ BC \\ BD \\ CD \end{array}$$
(27)

Let $x = (x_A, x_B, x_C, x_D)^t$. Then

$$Tx = (x_A - x_B, x_A - x_C, x_A - x_D, x_B - x_C, x_B - x_D, x_C - x_D)^t$$
(28)

$$FT = \begin{bmatrix} 1 & -3 & 3 & -1 \\ -2 & 0 & 3 & -1 \\ -3 & -1 & 3 & 1 \\ 0 & -1 & 1 & 0 \\ -1 & -2 & 3 & 0 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -3 \\ -1 & -3 & 1 & 3 \end{bmatrix}$$
(29)

The (i, j)-component of E(=FT) presents a judge of subject *i* for object *j*. For example the (1, 1)-component = 1 of *E* presents a judge of subject 1 for object 1: *A*. A comparison result of (A, B), (A, C), (A, D) is (1, -1, 1) that is, (A > B), (A < C), (A > D). This means that object *A* is superior to 2 objects and inferior to 1 object, that is, the balance is 1. The (1, 2)-component = -3 of *FT* presents a judge of subject 1 for object 2: *B*. A comparison result of (A, B), (B, C), (B, D) is (-1, -1, -1), that is, (A > B), (B < C), (B < D). This means that object *B* is inferior to 3 objects, that is, the balance is -3.

A solution x is obtained through maximization of variance among subjects, $x^t E^t E x$, under normalization of total variance, that is, x is obtained as an eigenvector corresponding to the maximum eigenvalue of $E^t E$.

6 Formulation different from Nishisato for paired comparison data

Let A_i be a comparison matrix of subject *i* in AHP (Analytic Hierarchy Process). If object h > object *j*, the (h, j)-th component of A_i is larger than 1 and the (h, j)-th component of D_i corresponding to A_i is equal to 1. For example, D_1 corresponding to A_1 is given by

$$D_{1} = \begin{bmatrix} 0 & 1 & -1 & 1 \\ -1 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 \\ -1 & 1 & -1 & 0 \end{bmatrix}$$
(30)

The following relation holds between D_i and FT.

$$E \equiv FT = \begin{bmatrix} D_1 \boldsymbol{e} & D_2 \boldsymbol{e} & D_3 \boldsymbol{e} & D_4 \boldsymbol{e} & D_5 \boldsymbol{e} & D_6 \boldsymbol{e} & D_7 \boldsymbol{e} & D_8 \boldsymbol{e} \end{bmatrix}^T$$
(31)

where $e = (1, 1, \dots, 1)^t$.

From these facts, the following treatments for A_i obtained in AHP can be considered.

(i) Obtain D_i from A_i . Calculate E by Eq.(31).

(ii) Let the (j, k)-th component of B_i be $\log\{\text{the } (j, k)\text{-th component of } A_i\}$. Use B_i instead of D_i . Calculate E by

$$E = \begin{bmatrix} B_1 \boldsymbol{e} & B_2 \boldsymbol{e} & B_3 \boldsymbol{e} & B_4 \boldsymbol{e} & B_5 \boldsymbol{e} & B_6 \boldsymbol{e} & B_7 \boldsymbol{e} & B_8 \boldsymbol{e} \end{bmatrix}^t$$
(32)

	object	1	2	3	4
j					
	AHP	0.311	0.389	0.141	0.159
1	(i)	0.356	0.389	0.232	0.159
	(ii)	0.366	0.389	0.238	0.159
	AHP	0.346	0.201	0.203	0.25
2	(i)	0.346	0.295	0.306	0.25
	(ii)	0.346	0.296	0.319	0.25
	AHP	0.188	0.143	0.149	0.52
3	(i)	0.249	0.167	0.149	0.52
	(ii)	0.368	0.143	0.145	0.52
	AHP	0.309	0.249	0.336	0.106
4	(i)	0.309	0.221	0.281	0.106
	(ii)	0.309	0.235	0.292	0.214
	AHP	0.286	0.225	0.387	0.102
5	(i)	0.333	0.25	0.387	0.102
	(ii)	0.361	0.236	0.387	0.102

Table 6: Comparison between AHP and Procedures (i) and (ii)

A solution x is obtained as an eigenvector corresponding to the maximum eigenvalue of $E^t E$ and a solution y is obtained as an eigenvector corresponding to the maximum eigenvalue of EE^t .

Example 2. Table 6 shows results of the case where N=19, J=5, M=4. For the purpose of comparison, AHP results are also shown in Table 6, where if the (j, k)-th component, $a_{i,jk}$, of A_i , the (j, k)-th component a paired comparison matrix is $(\prod_{i=1}^{19} a_{i,jk})^{1/19}$.

The same formulation 3 as formulation 2 can be considered.

(iii) Let a value of the (j, k)-th component of A_i be h.

If h > 1, let the (j, k)-th component, $b_{i,jk}$, of B_i be $\log t_h$.

If h = 0, let $b_{i,jk}$ be 0.

If h < 1, let $b_{i,jk}$ be $-\log t_h$.

Let f_{ij} be a mean of *j*-th row of B_i , that is,

$$f_{ij} = \sum_{k} b_{i,jk} / M. \tag{33}$$

Formulation 3

$$\max \sum_{j=1}^{M} (x_j - \mu)^2$$
 (34)

s.t.
$$\sum_{j=1}^{M} x_j / M = \mu = 0; \quad x_j = \sum_{i=1}^{N} f_{ij} / N$$
 (35)

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (f_{ij} - \mu)^2 / (NM) = 1$$
(36)

$$t_1 = 1 \le t_2 \le \dots \le t_K. \tag{37}$$

However, the same formulation as formulation 2 cannot be considered, because

$$y_i = \sum_{j=1}^M f_{ij}/M = 0.$$

Example 3. Suppose that

$$A_i = \begin{bmatrix} 1 & 5 & 3\\ 1/5 & 1 & 1/2\\ 1/3 & 2 & 1 \end{bmatrix}$$

Then,

$$B_{i} = \begin{bmatrix} 0 & \log t_{5} & \log t_{3} \\ -\log t_{5} & 0 & -\log t_{2} \\ -\log t_{3} & \log t_{2} & 0 \end{bmatrix}$$
$$f_{i1} = (\log t_{5} + \log t_{3})/3$$
$$f_{i2} = (-\log t_{5} + \log t_{2})/3$$
$$f_{i3} = (-\log t_{3} + \log t_{2})/3$$
$$y_{i} = \sum_{j} f_{ij}/3 = 0.$$

The last equation must be changed into

$$y_i = \sum_{j < k} b_{i,jk} \tag{38}$$

Formulation 4

$$\max \sum_{i=1}^{N} (y_i - \mu)^2$$
(39)

s.t.
$$\sum_{i=1}^{N} y_i / N = \mu = 0; \quad y_i = \sum_{j < k} b_{i,jk}$$
 (40)

$$\sum_{i=1}^{N} \sum_{j=1}^{M} (f_{ij} - \mu)^2 / (NM) = 1$$
(41)

$$t_1 = 1 \le t_2 \le \dots \le t_K \tag{42}$$

 μ : grand mean

Table 7: Values of t_h obtained by procedure (iv)									
h	1	2	3	4	5	6	7	8	9
t_h	1	2.113	6.014	8.185	8.185	8.185	8.185	8.185	16.000

Table 8: Aggregated evaluation by AHP and procedure (iv)AHP0.2700.1890.2300.311(iv)0.2880.1770.2110.324

The following procedure can be considered as a method of aggregating N subjects' judges A_1, \dots, A_N at AHP.

(iv) Let judge of objects k relating to criteria j by subject i be $f_{i,J+M(j-1)+k}$ $(i = 1, 2, \dots, N; j = 1, 2, \dots, J; k = 1, 2, \dots, M).$

Decide t_h which maximize the variance of $x_{J+M(j-1)+kj} = \sum_i f_{i,J+M(j-1)+k}/N$. Derive a value of criteria j as $x_j = \sum_{i=1}^N f_{ij}/N$ ($j = 1, 2, \dots, J$) and let

$$c_j = \exp(x_j) / \sum_k \exp(x_k)$$

Calculate $x_{J+M(j-1)+kj} = \sum_i f_{i,J+M(j-1)+k}/N$ $(k = 1, 2, \dots, M)$. Derive aggregated evaluation $\exp\{\sum_j c_j x_{J+M(j-1)+k}\}$. Calculate $\exp\{\sum_j c_j y_{ij}\}$ as evaluation of subject *i*.

Example 4. Data of Example 2 are analyzed, following to procedure (iv). Table 7 and Table 8 show a part of results. There are not large differences between them.

References

[1] S. Nishisato. *Analysis of categorical data: Dual scaling and its applications*, University of Toronto Press, 1980.