Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

CORRESPONDENCE ANALYSIS AND SERIATION

BY ROUH-JANE CHOU
National Tsing Hua University

The first nontrivial correspondence analysis (CA) solution of a two-way contingency table gives scores of the row and column categories so that the correlation between the row variable and the column variable is maximized. Hence it is natural to order the categories by the scores. In this paper, the appropriateness of this technique is investigated. Sufficient conditions are given. Sampling theories, when the data fail to satisfy the conditions because of the presence of random errors, are studied. As to whether one should use one or two CA solutions for the ordering, simulation study is used to show their difference.

1. Introduction. Ordering things in time has been the interest of archaeologists since a century ago when Petri tried to seriate chronologically some 900 graves by means of the numbers of various potteries in them. Mathematicians had developed methods of seriation since then. Among them, D.G. Kendall had made significant contributions in the sixties and seventies. Kendall (1963) proposed a model for the graves and made statistical inference about the model. In Kendall (1971a), a measure, called common content, of similarity between graves was suggested. An algorithm for achieving a right ordering of the graves satisfying the condition of being pre-Q was also given. Then in Kendall (1971b), the method of multidimensional scaling was introduced to seriate objects, given their similarity matrix.

Recently, the method of correspondence analysis (CA) for contingency tables has gained more and more attention, especially in its use of ordering objects (Greenacre (1984), Hill (1974) and Schriever (1986)). The idea is to order the rows by the elements of the first non-trivial eigenvector of a matrix obtained from the matrix of frequencies or proportions. Though this method looks suitable intuitively, a mathematical justification, however, is needed.

In this paper, we study the appropriateness of the first non-trivial correspondence analysis solution as a rule for seriating the rows of a Q-matrix. In Section 2, method of CA, ideas of Q-matrix and total positivity (TP) are

---

AMS 1991 Subject Classifications: Primary 62H20.

Key words and phrases: Contingency table, correspondence analysis, total positivity.

introduced. Sufficient conditions are obtained, in Section 3. The special case, when the solution is linear, is also investigated. In Section 4, we study the problem when the observed matrix is not a TP, but TP with some errors. Large sample results, such as consistency and asymptotic normality, are established. As to whether an ordering by the first two non-trivial solutions will do a better job when errors appear, simple simulation shows that the answer is negative. Section 5 contains some problems that need further investigation.

2. Correspondence Analysis and Total Positivity. Let $A = (a_{ij})$ be an $n \times p$ contingency table and $R = \text{diag}(a_{i+})$ and $C = \text{diag}(a_{+j})$ be diagonal matrices with elements the row and column totals, respectively. Correspondence analysis (CA), according to Hill (1974), is to assign scores to the row and column categories by reciprocal averaging process.

DEFINITION 2.1. *The triple (ρ, x, y) is a CA solution of A if*

$$\rho x = R^{-1}Ay \qquad \text{and} \qquad \rho y = C^{-1}A^t x. \tag{2.1}$$

The elements of the vector x are called "row scores" and the elements of y are called "column scores". The number ρ is the correlation of x and y with respect to the matrix A.

It is easy to see that x and y satisfy

$$\rho^2 x = R^{-1}AC^{-1}A^t x \qquad \text{and} \qquad \rho^2 y = C^{-1}A^t R^{-1}Ay. \tag{2.2}$$

Hence ρ^2 , x and y are the eigenvalue and eigenvectors of some matrix. Since the row sums of both $R^{-1}AC^{-1}A^t$ and $C^{-1}A^tR^{-1}A$ are all equal to unity, the maximal value of ρ^2 is 1 with corresponding eigenvectors $1 = (1, 1, \cdots, 1)$ and $1 = (1, 1, \cdots, 1)$ of dimensionality n and p, respectively.

The vector x corresponding to the next largest eigenvalue naturally offers an ordering of the row categories. As to whether this order can be interpreted as time, some conceptual background are needed.

The idea of Petri's "Concentration Principle", introduced by Kendall (1963), is that, over a certain period of time, new artifacts appeared, became popular, quickly or gradually, and then fell out of use.

DEFINITION 2.2. *The matrix $A = (a_{ij})$ is called a Q-matrix if for each $j = 1, \cdots, p$, the vector $(a_{ij}, i = 1, \cdots, n)$ is either increasing in i, decreasing in i or first increasing then decreasing in i.*

Any matrix that can be made a Q-matrix by permuting its rows is called a pre-Q matrix.

Consider a table whose each row consists of numbers of certain types of artifacts in a grave. According to the concentration principle, this table has the property of a Q-matrix provided that the graves are ordered in time.

Hence the rows of a Q-matrix are expected to be correctly ordered. We are to study whether the elements of the first non-trivial CA solution of a Q-matrix are in ascending or descending order.

However, we first note that it is not difficult to construct a 2-way table such that more than one ordering of the rows or columns result in Q-matrix. Consider, as an example, the 6×4 table

$$A = \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 1 & 7 & 5 \\ 0 & 3 & 43 & 22 \\ 2 & 9 & 64 & 32 \\ 2 & 12 & 94 & 35 \\ 1 & 6 & 62 & 18 \end{bmatrix}.$$

The row permutations (123456), (234561), (345621), (123546), (235461) and (354621) and column permutations (1234), (1243), (2341) all lead to Q matrices. Also note the famous example of Fisher's on the relationship between colors of eye and hair of Scottish school children, given in Greenacre (1984), shows that pattern-Q of a matrix alone is not enough to guarantee the monotonicity of its first non-trivial CA solution.

	Hair Color				
Eye Color	Fair	Red	Medium	Dark	Black
Blue	326	38	241	110	3
Light	688	116	584	188	4
Medium	343	84	909	412	26
Dark	98	48	403	681	85

This matrix satisfies the condition of being a Q, but the order given by CA solution is Light-Blue-Medium-Dark which leads to a non-Q matrix.

Other conditions, such as total positivity of the matrix, must be introduced. Let X and Y be subsets of R.

DEFINITION 2.3. *A real-valued function K defined on $X \times Y$ is said to be (strictly) totally positive of order s (TP_s or STP_s) if for every $t = 1, 2, \cdots, s$, all $x_1 < x_2 \cdots < x_t$ and $y_1 < y_2 < \cdots < y_t$ ($x_i \in X, y_i \in Y$, for $i = 1, 2, \cdots, t$), the determinant*

$$K\begin{bmatrix} x_1, x_2, \cdots, x_t \\ y_1, y_2, \cdots, y_t \end{bmatrix} = \begin{vmatrix} K(x_1, y_1) & K(x_1, y_2) & \cdots & K(x_1, y_t) \\ K(x_2, y_1) & K(x_2, y_2) & \cdots & K(x_2, y_t) \\ \vdots & \vdots & & \vdots \\ K(x_t, y_1) & K(x_t, y_2) & \cdots & K(x_t, y_t) \end{vmatrix}$$

is (positive) nonnegative.

Examples of TP functions include (See Marshall & Olkin (1979), for example)

(i) $K(x,y) = \exp(xy)$ is STP_∞ on R^2.

(ii) $K(x,y) = I(x \leq y)$ is TP_∞ on R^2.

(iii) The function

$$K(x,y) = \begin{cases} 1 & \text{if } x \leq y \leq x+1, \\ 0 & \text{otherwise.} \end{cases}$$

is TP_2 but not TP_3 on R^2.

(iv) $K(x,y) = [1 + (x-y)^2]^{-1}$ is not TP_2 on R^2.

In the case when $X = \{1,2,\cdots,n\}$ and $Y = \{1,2,\cdots,p\}$ K can be considered as an $n \times p$ matrix with elements $k_{ij} = K(i,j)$ for $i = 1,2,\cdots,n$ and $j = 1,2,\cdots,p$.

Note that all four examples share the property that, for any $x, K(x,y)$ first increases and then decreases in y. Hence, although the family of Q-matrices and the family of TP (or STP) matrices do not include each other, they overlap.

The following are some useful and important results taken from Marshall and Olkin (1979) and Gantmacher and Krein (1950), in the theory of total positivity.

LEMMA 2.1. *Let K be $(S)TP_s$ on $X \times Y$, then $f(x)g(y)K(x,y)$ is $(S)TP_s$ on $X \times Y$ whenever the functions f and g are (positive) nonnegative on X and Y, respectively.*

LEMMA 2.2. *Let K be TP_s on $X \times Y$ and let L be TP_r on $Y \times Z$, then the integral*

$$M(x,z) = \int K(x,y)L(y,z)d\sigma(y)$$

where σ is a sigma-finite measure, is $TP_{min(s,r)}$.

LEMMA 2.3. *If K is TP_∞ on R^2, then the integral equation*

$$\lambda\phi(x) = \int_{-\infty}^{\infty} K(x,y)\phi(y)\,dy \tag{2.3}$$

has the following properties:

(i) All the eigenvalues of (2.3) are positive and simple : $\lambda_0 > \lambda_1 > \lambda_2 > \cdots$.

(ii) The eigenfunction $\phi_0(x)$ corresponding to λ_0 does not have any zeroes.

(iii) The eigenfunction $\phi_j(x)$ corresponding to λ_j has exactly j nodes and no other zeroes, $j = 1,2,\cdots$

3. Sufficient Conditions. When $X = \{1, 2, \cdots, n\}$ and $Y = \{1, 2, \cdots, p\}$, Schriever (1986) gives sufficient conditions for the joint distribution of (X, Y) so that the first canonical functions are monotone. In this section, we study the continuous version of correspondence analysis.

Let $f_X(\cdot)$ and $f_{Y|X}(\cdot \mid x)$ denote the pdf of the random variable X and the conditional pdf of the random variable Y given $X = x$, respectively. The continuous analogue of the equations (2.1) and (2.2) are

$$\rho h(x) = E[g(Y) \mid X = x] \quad \text{and} \quad \rho g(y) = E[h(X) \mid Y = y], \tag{3.1}$$

and

$$\begin{aligned} \rho^2 h(x) &= E\{[h(X) \mid Y] \mid X = x\}, \\ \rho^2 g(y) &= E\{E[g(Y) \mid X] \mid Y = y\}, \end{aligned} \tag{3.2}$$

where $h(x)$ and $g(y)$ are the scoring functions of X and Y, respectively. The equations (3.2) say, assuming the validity of changing orders of integration,

$$\begin{aligned} \rho^2 g(y) &= \iint g(u) f_{Y|X}(u \mid x)\, du f_{X|Y}(x \mid y)\, dx \\ &= \int K(u, y) g(u)\, du \end{aligned} \tag{3.3}$$

where the kernel function $K(u, y)$ is defined by

$$K(u, y) = \int f_{Y|X}(u \mid x) f_{X|Y}(x \mid y)\, dx.$$

The scoring function $g(y)$ is thus the eigenfunction of $K(u, y)$.

3.1. An Example.

Let (X, Y) be jointly bivariate normal, $BN(0,\, 0;\, 1,\, 1,\, \rho)$, with density

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho x y + y^2)\right\}.$$

The conditional densities and hence the kernel function are

$$\begin{aligned} f_{Y|X}(y \mid x) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(y - \rho x)^2\right], \\ f_{Y|X}(x \mid y) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(y - \rho y)^2\right], \end{aligned}$$

and

$$\begin{aligned} K(u, y) &= \int f_{Y|X}(u \mid x) f_{X|Y}(x \mid y)\, dx \\ &= \frac{1}{\sqrt{2\pi(1-\rho^4)}} \exp\left[-\frac{1}{2(1-\rho^4)}(u - \rho^2 y)^2\right]. \end{aligned} \tag{3.4}$$

This $K(u,y)$ depends on u and y only through $u-\rho^2 y$. Denote $k^*(u-\rho^2 y) = K(u,y)$. Then we are solving the equation

$$\lambda^2 g(y) = \int_{-\infty}^{\infty} k^*(u-\rho^2 y) g(u)\, du. \tag{3.5}$$

Differentiating both sides of (3.5), we have

$$\begin{aligned}
\lambda^2 g'(y) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial y} k^*(u-\rho^2 y) g(u)\, du \\
&= g(u) \int_{-\infty}^{u} \frac{\partial k^*(s-\rho^2 y)}{\partial y}\, ds \bigg|_{u=-\infty} \\
&\quad - \int_{-\infty}^{\infty} \left[\int_{-\infty}^{u} \frac{\partial k^*(s-\rho^2 y)}{\partial y} ds \right] g'(u)\, du \\
&= -\rho^2 g(u) k^*(u-\rho^2 y)\big|_{u=-\infty} + \rho^2 \int_{-\infty}^{\infty} k^*(u-\rho^2 y) g'(u)\, du.
\end{aligned}$$

If

$$g(u) k^*(u-\rho^2 y)\big|_{u=-\infty}^{\infty} = 0 \qquad \text{for all} \quad y,$$

then

$$\lambda^2 g'(y) = \rho^2 \int_{-\infty}^{\infty} k^*(u-\rho^2 y) g'(u)\, du. \tag{3.6}$$

Hence $g'(y)$ is also an eigenfunction function of $k^*(u-\rho^2 y)$, its corresponding eigenvalue is λ^2/ρ^2.

Since $K(u,y)$ is TP_∞ and its trivial and largest eigenvalue is $\lambda_0 = 1$ with corresponding eigenfunction $g_0(y) \equiv 1$. By Lemma 2.3, the j th eigenfunction has j nodes. For $j = 1$, $g_1(y)$ has 1 node and no other zeros. According to (3.6), $g_1'(y)$ is also an eigenfunction, with eigenvalue λ_1^2/ρ^2 which is greater than λ_1^2. Hence

$$\lambda_1^2/\rho^2 = \lambda_0^2 = 1$$

and $\lambda_1^2 = \rho^2$. Furthermore, $g_1'(y)$ =constant $\times g_0(y)$=constant. That is, $g_1(y)$ is linear and hence monotone. In order to make g_1 orthogonal to g_0 and have unit variance, we take

$$g_1(y) = y.$$

A similar procedure gives us the following that

$$g_k(y) = H_k(y)/\sqrt{k!},$$

where $H_k(y)$ is the Hermite polynomial of degree k.

Since in the present case, X and Y are symmetric, we have also that

$$h_k(x) = H_k(x)/\sqrt{k!}.$$

Thus the reconstitution formula is

$$f(x,y) = \phi(x)\phi(y)\left\{1 + \sum_{k=1}^{\infty} \frac{\rho^k H_k(x)H_k(y)}{k!}\right\}, \tag{3.7}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right),$$

the marginal density function of X, is the standard normal density function.

The identity (3.7) is called Mehler identity in classical analysis (Lancaster (1958)).

3.2. General Results.

Kendall (1963) models a grave, in the spirit of concentration principle, by assuming that the numbers $n_x(y)$ of the xth artifact in the yth position are independent Poisson variables having expectation

$$En_x(y) = \mu_x \exp\left[-\psi(z_y)\right],$$

where

$$z_y = \begin{cases} \dfrac{y - \sigma_x}{\alpha_x} & \text{if } y \geq \sigma_x, \\ \dfrac{\sigma_x - y}{\beta_x} & \text{if } y < \sigma_x \end{cases}$$

and $0 \leq \psi(\cdot)$ is an increasing function, μ_x, α_x, β_x and σ_x are functions of x.

Several values of x and y then compose a random matrix whose expectation satisfies the condition of begin a Q matrix. As the elements of this Q matrix are all positive, we would like to view them as probabilities multiplied by a common constant. Thus each column may be viewed as the conditional probability of $Y = y$ given $X = x$ multiplied by the frequency of $X = x$.

It has been shown that the property of Q alone is not sufficient for the first non-trivial CA solution to give the correct ordering. We thus confine ourselves to totally positive conditional pdf's.

THEOREM 3.1. *Suppose that $f_{Y|X}(y \mid x)$ is TP_∞ on R^2, in solving* (3.3), *if*

(i) $L(u,y) = -\int_{-\infty}^{u} \frac{\partial}{\partial y} K(s,y)\, ds$ *is* TP_∞, *and*

(ii) $g(u)L(u,y)|_{u=-\infty}^{\infty} = 0$,

then the first non-trivial eigenfunction $g_1(y)$ is monotone.

PROOF. Since $f_{Y|X}(y \mid x)$ is TP_∞ and

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{f_Y(y)},$$

where

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x) dx$$

is the marginal pdf of Y, by Lemma 2.1, $f_{X|Y}(x \mid y)$ is TP_∞ . Hence, by Lemma 2.2,

$$K(u,y) = \int_{-\infty}^{\infty} f_{Y|X}(u \mid x) f_{X|Y}(x \mid y)\, dx$$

is TP_∞, and, as a result from Lemma 2.3, its first nontrivial eigenfunction $g_1(y)$ has one node and no other zeroes. Differentiating both sides of (3.3)

$$\begin{aligned}\lambda_1 g_1'(y) &= \int_{-\infty}^{\infty} \frac{\partial K(u,y)}{\partial y} g_1(u)\, du \\ &= -g_1(u) L(u,y)|_{u=-\infty}^{\infty} + \int_{-\infty}^{\infty} L(u,y) g_1'(u)\, du \\ &= \int_{-\infty}^{\infty} L(u,y) g_1'(u)\, du.\end{aligned}$$

by (ii). Thus $g_1'(y)$ is an eigenfunction of $L(u,y)$.

Since $L(u,y)$ is TP_∞ , its first eigenfunction has no sign-change. It must be $g_1'(y)$, because the derivatives of other eigenfunction of $K(u,y)$ have at least one node. Therefore, $g_1(y)$ is monotone. This completes the proof of the theorem.

It is easily checked that the kernel function (3.4) satisfies the conditions (i) and (ii) of the theorem.

3.3. Special Case when the First Non-trivial Eigenfunction Is Linear.

As the totally positive function that one would often think of is the exponential function, $\exp(xy)$, we now study the case when the conditional distribution of Y given $X = x$ belongs to the exponential family. If the marginal distribution of X is its conjugate prior distribution, then the first non-trivial CA solution is in fact linear. The interesting lemma by Diaconis and Ylvisaker (1979) is of key importance to this result.

LEMMA 3.1. *Given the exponential family of distribution with pdf*

$$dP_\theta(x) = \exp[x\theta - M(\theta)] d\mu(x).$$

Assuming the changeability of the order of differentiation and integration, then

(i) $E(X \mid \theta) = M'(\theta)$ and

(ii) $E(M'(\theta) \mid x) = ax + b (a, b$ are constants)

iff the conjugate prior density is

$$c(a,b) \exp\left[\frac{b}{a}\theta - \frac{1-a}{a} M(\theta)\right].$$

In order that $aX + b$ be a proper Bayes estimate of $M'(\theta)$ using square error loss, the value of a is restricted by $0 < a < 1$. For a proof of this lemma, see Diaconis and Ylvisaker (1979). Now comes the result:

THEOREM 3.2. *With conditional density of* Y *given* $X = x$*:*

$$f_{Y|X}(y \mid x) = e^{yx - M(x)}\mu(y)$$

and its conjugate prior density

$$f_X(x) = c(a,b)\exp\left[\frac{b}{a}x - \frac{1-a}{a}M(x)\right],$$

the function $g(y) = y - \frac{b}{1-a}$ *satisfies the equation*

$$\lambda g(y) = \int_{-\infty}^{\infty} K(u,y)g(u)du$$

with $\lambda = a$*.*

PROOF. Recall the equation (3.2)

$$\rho^2 g(y) = E\{E[g(Y) \mid X] \mid Y = y\}.$$

With $g(y) = y - \frac{b}{1-a}$, we have, by Lemma 3.1,

$$E(g(Y) \mid X) = M'(X) - \frac{b}{1-a}$$

and

$$E\left[M'(X) - \frac{b}{1-a}\middle| Y = y\right] = ay + b - \frac{b}{1-a} = a\left(y - \frac{b}{1-a}\right).$$

Note that $f_{Y|X}(y \mid x)$ is TP_∞ and hence so is $K(u,y)$, this linear function must be the first non-trivial eigenfunction of $K(u,y)$.

4. Asymptotic Theory and Simulation. The observed contingency tables may not satisfy the condition of being TP or Q matrices even when their matrices of expectations are, due to the presence of random errors. In this section, we assume that the elements in the contingency table are Poisson variables with positive means. The CA solutions of the observed table are then shown to be consistent estimators of their corresponding solutions of the mean matrix, when the total number of observations approaches infinity. The deviations between the observed and the estimated solutions are also shown to follow the multivariate normal distribution, asymptotically.

One may also suspect, when errors exist, whether using the first two nontrivial eigenfunctions will result in better ordering. When the underlying mean matrix is proportional to the standard bivariate normal density, simulation study shows that it is not the case.

4.1. Asymptotic Results.

Let $\Lambda = (\lambda_{ij})$ be a $p \times q (p \geq q)$ matrix with positive elements and $A = (a_{ij})$ be $p \times q$ where $a'_{ij}s$ are independent Poisson variables with means $n\lambda_{ij}$, respectively. Then, by the law of large numbers, we have

$$\frac{a_{ij}}{n} \xrightarrow{P} \lambda_{ij} \quad \text{as} \quad n \to \infty, \qquad i = 1, \cdots, p, \quad j = 1, \cdots, q,$$

where $\xrightarrow{P}$ means converging in probability. Hence

$$R^{1/2}AC^{-1/2} \xrightarrow{P} R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2} \quad \text{and}$$
$$(R^{-1/2}AC^{-1/2})^t(R^{-1/2}AC^{-1/2}) \xrightarrow{P} (R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2})$$

as $n \to \infty$, where $R(R_\Lambda)$ and $C(C_\Lambda)$ are diagonal matrix with elements row sums and column sums of $A(\Lambda)$, respectively.

Suppose that the nontrivial CA solutions for A and Λ are $(\ell_\alpha, x_\alpha, y_\alpha)$ and $(\lambda_\alpha, \xi_\alpha, \eta_\alpha)$, $\alpha = 1, \cdots, q-1$, $\ell_1 > \cdots > \ell_{q-1}$, $\lambda_1 > \cdots > \lambda_{q-1}$. By the following relationship, which are equivalent to the CA problem

$$\ell_\alpha^2(C^{*1/2}y_\alpha) = (R^{-1/2}AC^{-1/2})^t(R^{-1/2}AC^{-1/2})(C^{*1/2}y_\alpha),$$
$$\lambda_\alpha^2(C_\Lambda^{1/2}\eta_\alpha) = (R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2})^t(R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2})(C_\Lambda^{1/2}\eta_\alpha)$$

and the continuity of eigenvalues and eigenvectors as functions of the elements of the matrix, we have

$$\ell_\alpha^2 \xrightarrow{P} \lambda_\alpha^2 \quad \text{and} \quad C^{*1/2}y_\alpha \xrightarrow{P} C_\Lambda^{1/2}\eta_\alpha$$

as $n \to \infty$, where $C^* = n^{-1}C$. Also because $C^* \xrightarrow{P} C_\Lambda$, we have

THEOREM 4.1. *For* $\alpha = 1, 2, 3, \cdots, q-1$,

$$\ell_\alpha \xrightarrow{P} \lambda_\alpha \quad \textit{and} \quad y_\alpha \xrightarrow{P} \eta_\alpha \quad \text{as} \quad n \to \infty.$$

As to how big the deviations are, according to the central limit theorem, a_{ij} has the following Z-representation (Chou and Hu (1992)),

$$\frac{a_{ij}}{n} \stackrel{\mathrm{d}}{=} \lambda_{ij} + n^{-1/2}\sqrt{\lambda_{ij}}Z_{ij} + n^{-1}\frac{H_2(Z_{ij})}{6} + \cdots,$$

where $\stackrel{d}{=}$ means that both sides have the same distribution, and Z_{ij} denotes the standard normal variable. Hence

$$\frac{a_{ij}}{\sqrt{a_{i+}}\sqrt{a_{+j}}} \stackrel{d}{=} \frac{\lambda_{ij}}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}} + n^{-1/2}d_1(i,j) + n^{-1}d_2(i,j) + \cdots,$$

where

$$\begin{aligned} d_1(i,j) =& d_1(Z_{i1},\cdots,Z_{iq}; Z_{1j},\cdots,Z_{pj}) \\ =& \frac{a_1(Z_{ij})}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}} - \frac{\lambda_{ij}}{\lambda_i + \lambda_{+j}}\Big[\sqrt{\lambda_{i+}}c_1(Z_{1j},\cdots,Z_{pj}) \\ &+ \sqrt{\lambda_{+j}}b_1(Z_{i1},\cdots,Z_{iq})\Big] \\ d_2(i,j) =& d_2(Z_{i1},\cdots,Z_{iq}; Z_{1j},\cdots,Z_{pj}) \\ =& \frac{a_2(Z_{ij})}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}} - \frac{\lambda_{ij}}{\lambda_{i+}\lambda_{+j}}\Big[b_1(Z_{i1},\cdots,Z_{iq})c_1(Z_{1j},\cdots,Z_{pj}) \\ &+ \sqrt{\lambda_{i+}}c_2(Z_{1j},\cdots,Z_{pj}) + \sqrt{\lambda_{+j}}b_2(Z_{i1},\cdots,Z_{iq})\Big] \\ &- \frac{1}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}}d_1(i,j)\Big[\sqrt{\lambda_i}c_1(Z_{1j},\cdots,Z_{pj}) \\ &+ \sqrt{\lambda_{+j}}b_1(Z_{i1},\cdots,Z_{iq})\Big] \end{aligned}$$

with

$$a_1(Z_{ij}) = \sqrt{\lambda_{ij}}Z_{ij}, \qquad a_2(Z_{ij}) = \frac{1}{6}H_2(Z_{ij})$$

$$b_1(Z_{i1},\cdots,Z_{iq}) = \sum_{j=1}^{q} \sqrt{\lambda_{ij}}Z_{ij}/2\sqrt{\lambda_{i+}}$$

$$b_2(Z_{i1},\cdots,Z_{iq}) = \left[\sum_{j=1}^{q}\frac{1}{6}H_2(Z_{ij}) - \left(\sum_{j=1}^{q}\sqrt{\lambda_{ij}}Z_{ij}\right)^2/4\lambda_{i+}\right]/2\sqrt{\lambda_{i+}}$$

$$c_1(Z_{1j},\cdots,Z_{pj}) = \sum_{i=1}^{p} \sqrt{\lambda_{ij}}Z_{ij}/2\sqrt{\lambda_{+j}}$$

$$c_2(Z_{1j},\cdots,Z_{pj}) = \left[\sum_{i=1}^{p}\frac{1}{6}H_2(Z_{ij}) - \left(\sum_{i=1}^{p}\sqrt{\lambda_{ij}}Z_{ij}\right)^2/4\lambda_{+j}\right]/2\sqrt{\lambda_{+j}}$$

Therefore we have the following Z-representation:

LEMMA 4.1.

$$\begin{aligned} &(R^{-1/2}AC^{-1/2})^t(R^{-1/2}AC^{-1/2}) \\ \stackrel{d}{=}&(R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2})^t(R_\Lambda^{-1/2}\Lambda C_\Lambda^{-1/2}) + n^{-1/2}W + n^{-1}U\cdots, \end{aligned}$$

where

$$W = (W_{jk}) \quad \text{and} \quad U = (U_{jk}), \qquad \text{both are} \quad q \times q \quad \text{and}$$

$$W_{jk} = \sum_{i=1}^{p} \left[\frac{\lambda_{ij}}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}} d_1(i,k) + \frac{\lambda_{ik}}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+k}}} d_1(i,j) \right],$$

$$U_{jk} = \sum_{i=1}^{p} \left[d_1(i,j) d_1(i,k) + \frac{\lambda_{ij}}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+j}}} d_2(i,k) + \frac{\lambda_{ik}}{\sqrt{\lambda_{i+}}\sqrt{\lambda_{+k}}} d_2(i,j) \right].$$

Perturbation method (Courant and Hilbert (1953)) then gives the expansion of the eigenvalues and eigenvectors of $(R^{-1/2}AC^{-1/2})^t(R^{-1/2}AC^{-1/2})$. Denoting $y^*_\alpha = C^{*1/2} y_\alpha$ and $\eta^*_\alpha = C_\Lambda^{1/2} \eta_\alpha$, the results are

LEMMA 4.2.

$$\begin{aligned} \ell^2_\alpha &= \lambda^2_\alpha + n^{-1/2}\lambda^{(1)}_\alpha + n^{-1}\lambda^{(2)}_\alpha + \cdots , \\ y^*_\alpha &= \eta^*_\alpha + n^{-1/2}\eta^{*(1)}_\alpha + n^{-1}\eta^{*(2)}_\alpha + \cdots , \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} \lambda^{(1)}_\alpha &= \eta^{*t}_\alpha W \eta^*_\alpha, \\ \lambda^{(2)}_\alpha &= \eta^{*t}_\alpha U \eta^*_\alpha + \eta^{*t}_\alpha W \eta^{*(1)}_\alpha, \\ \eta^{*(1)}_\alpha &= H^{-1}_\alpha v_\alpha \quad \text{and} \quad \eta^{*(2)}_\alpha = H^{-1}_\alpha v^*_\alpha \end{aligned}$$

the rows of H_α *are*

$$\begin{aligned} h_{\alpha\beta} &= (\lambda^2_\alpha - \lambda^2_\beta)\eta^{*t}_\beta \quad \text{if} \quad \beta \neq \alpha \quad \text{and} \\ h_{\alpha\alpha} &= \eta^{*t}_\alpha \end{aligned}$$

and the elements of the vector v_α *and* v^*_α *are*

$$\begin{aligned} &v_{\alpha\beta} = \eta^{*t}_\beta W \, \eta^*_\alpha \quad \text{if} \quad \beta \neq \alpha, \quad \text{and} \quad v_{\alpha\alpha} = 0, \\ &v^*_{\alpha\beta} = \eta^{*t}_\beta (U\eta^*_\alpha + W\eta^{*(1)}_\alpha - \lambda^2_\alpha \eta^{*(1)}_\alpha) \quad \text{and} \quad v^*_{\alpha\alpha} = -\frac{1}{2}\eta^{*(1)t}_\alpha \eta^{*(1)}_\alpha. \end{aligned}$$

Since

$$\begin{aligned} y_\alpha =& C^{*-1/2} y^*_\alpha \\ =& (C_\Lambda^{-1/2} + n^{-1/2}C^{(1)}_\Lambda + n^{-1}C^{(2)}_\Lambda + \cdots)(\eta^*_\alpha + n^{-1/2}\eta^{*(1)}_\alpha + n^{-1}\eta^{*(2)}_\alpha \cdots) \\ =& \eta_\alpha + n^{-1/2}(C_\Lambda^{-1/2}\eta^{*(1)}_\alpha + C^{(1)}_\Lambda C_\Lambda^{1/2}\eta_\alpha) \\ &+ n^{-1}(C^{(2)}_\Lambda C_\Lambda^{1/2}\eta_\alpha + C_\Lambda^{-1/2}\eta^{*(2)}_\alpha + C^{(1)}_\Lambda \eta^{*(1)}_\alpha) + \cdots , \end{aligned} \tag{4.2}$$

where $C_\Lambda^{(1)} = diag(c_j^{(1)})$ *and* $C_\Lambda^{(2)} = diag(c_j^{(2)})$ *with*

$$c_j^{(1)} = \frac{\sum_{i=1}^{p} \sqrt{\lambda_{ij}} Z_{ij}}{2\lambda_{+j}^{3/2}},$$

$$c_j^{(2)} = \frac{3}{8} \frac{\left[\sum_{i=1}^{p} \sqrt{\lambda_{ij}} Z_{ij}\right]^2}{2\lambda_{+j}^{3/2}} - \frac{1}{2} \frac{\sum_{i=1}^{p} \frac{1}{6} H_2(Z_{ij})}{\lambda_{+j}^{3/2}}.$$

The $n^{-1/2}$ terms in (4.1) and (4.2) are both linear in $Z'_{ij}s$, hence we have

THEOREM 4.2. *Let the CA solutions for A and* Λ *are* $(\ell_\alpha, x_\alpha, y_\alpha)$ *and* $(\lambda_\alpha, \xi_\alpha, \eta_\alpha), \alpha = 1, 2, \cdots, q-1$, *with* $\ell_1 > \cdots > \ell_{q-1}$ *and* $\lambda_1 > \cdots > \lambda_{q-1}$. *Then, for* $\alpha = 1, 2, \cdots, q-1$, *as* $n \to \infty$

$$\sqrt{n}(\ell_\alpha - \lambda_\alpha) \xrightarrow{\text{d}} N(0, \sigma_\alpha^2),$$
$$\sqrt{n}(y_\alpha - \eta_\alpha) \xrightarrow{\text{d}} N(0, \Sigma_\alpha),$$

where

$$\Sigma_\alpha^2 = \frac{1}{4\lambda_\alpha^2} \text{var}\,(\lambda_\alpha^{(1)}) \qquad \text{and}$$
$$\Sigma_\alpha = \text{cov}\,(C_\Lambda^{-1/2} \eta_\alpha^{*(1)} + C_\Lambda^{(1)} C_\Lambda^{1/2} \eta_\alpha).$$

THEOREM 4.3. *Under the same conditions and notations as Theorem 4.2, we have for* $\alpha \neq \beta$, $\alpha, \beta = 1, 2, \cdots, q-1$,

$$\sqrt{n}(y_\alpha - n_\alpha, y_\beta - \eta_\beta) \xrightarrow{\text{d}} N(0, \Sigma),$$

where

$$\Sigma = \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\beta} \\ \Sigma_{\alpha\beta}^t & \Sigma_\beta \end{bmatrix}$$

and

$$\Sigma_{\alpha\beta} = E\left[C_\Lambda^{-1/2} \eta_\alpha^{*(1)} + C_\Lambda^{(1)} C_\Lambda^{1/2} \eta_\alpha\right]^t \left[C_\Lambda^{-1/2} \eta_\beta^{*(1)} + C_\Lambda^{(1)} C_\Lambda^{1/2} \eta_\beta\right].$$

4.2. Simulation Study.

Though the sample CA solutions are consistent for their corresponding population CA solutions, one is curious about whether the ordering of the rows obtained from first two CA solutions is better than that obtained from the first

CA solution only. A simulation is done when the cells of the population table are standard normal densities:

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right\}.$$

Remember that for this function, the first two non-trivial CA solutions are $H_1(x) = x$ and $H_2(x)/\sqrt{2} = (x^2-1)/\sqrt{2}$, respectively.

For each $\rho = 0.1(0.1)0.9$, an 11×11 table is generated where each cell is a Poisson observation with mean $nf(x_i, y_j)$, $i, j = 1, 2, \cdots, 11$. Sample first and second CA solutions, $(y_{1i}, y_{2i}), i = 1, 2, \cdots, 11$, are obtained. A first ordering is given by the orders of $y_{1i}, i = 1, 2, \cdots, 11$. A second one is obtained by projecting the points $(y_{1i}, y_{2i}), i = 1, 2, \cdots, 11$, to the parabola $y = (x^2-1)/\sqrt{2}$. Their respective rank correlations with the ordering of y_j, $j = 1, \cdots, 11$, are calculated. For each n=50, 100, 200 and 500, the experiments are repeated 500 times. The averages of the absolute values of the rank correlations, r_1 and r_2 , are given in Table 4.1.

From the table, it is seen that both r_1 and r_2 increase as n increases. The r_2 values are all smaller than their corresponding r_1 values. The interpretation is that, in the bivariate normal case, there is no need to consider the second eigenfunction. Using two eigenfunctions merely increases the sampling errors and hence decreases the precision.

5. Conclusion and Discussion. It is seen that in order to tackle the problem of ordering, total positivity is the right condition to consider. However, the sufficient condition given in Theorem 3.1 is not a pleasant one. It is difficult either to construct a bivariate distribution that satisfies this condition, to check whether commonly used distributions satisfy it or to prove whether bivariate normal is the only distribution that satisfies it. Also a question is about the order of TP. According to Fisher's example, given in Section 2, TP_2 is necessary. It is not difficult to check that the uniform density considered in Section 2 satisfies the condition of being TP_2 , but it is not TP_3. As CA leads to correct ordering, an interesting question to ask is whether TP_2 is also sufficient.

Another problem worthy of further study is that of convex curve fitting. It is related to the simulation of Section 4.2. According to Lemma 2.3, the second non-trivial eigenfunction of a TP kernel has two nodes. It is a quadratic function when the kernel is a bivariate normal density. For general TP kernels, it is natural to try to fit $\{(y_{1i}, y_{2i}), i = 1, 2, \cdots, p)$ with a convex curve to which the projections of the points offer an ordering.

TABLE 4.1. Comparison of orderings using first one or first two CA nontrivial solution

n =50 | n=100

rho	row r_1	col r_1	row r_2	col r_2	rho	row r_1	col r_1	row r_2	col r_2
.1	.3537	.3561	.3296	.3421	.1	.3829	.3607	.3608	.3409
.2	.4439	.4415	.4127	.4097	.2	.5523	.5633	.4974	.5892
.3	.5995	.5974	.5343	.5309	.3	.7975	.8101	.6921	.7088
.4	.7811	.7778	.6978	.6847	.4	.9087	.9047	.8094	.8103
.5	.8695	.8694	.7739	.7586	.5	.9539	.9606	.8470	.8338
.6	.9298	.9318	.8230	.8045	.6	.9763	.9767	.8392	.8293
.7	.9279	.9345	.8080	.8020	.7	.9870	.9860	.8101	.8224
.8	.9173	.9103	.8059	.7968	.8	.9840	.9888	.7779	.7991
.9	.8200	.8242	.7696	.7525	.9	.9584	.9571	.7953	.7796

n =200 | n =500

rho	row r_1	col r_1	row r_2	col r_2	rho	row r_1	col r_1	row r_2	col r_2
.1	.4242	.4522	.3886	.4119	.1	.6158	.6508	.5633	.5786
.2	.7625	.7585	.6809	.6707	.2	.9357	.9366	.8310	.8301
.3	.9220	.9179	.8127	.8169	.3	.9763	.9774	.8597	.8487
.4	.9675	.9586	.8439	.8368	.4	.9892	.9895	.8161	.8140
.5	.9823	.9815	.8354	.8377	.5	.9942	.9951	.7924	.7918
.6	.9892	.9888	.8161	.8144	.6	.9964	.9967	.8003	.7959
.7	.9944	.9942	.7747	.7764	.7	.9988	.9985	.8002	.7820
.8	.9959	.9968	.7835	.7616	.8	.9997	1.0000	.7811	.7792
.9	.9985	.9989	.7905	.7767	.9	.8200	.8242	.7696	.7525

r_1 = **correlation between the original order and the order determined by the first CA nontrival solution.**

r_2 = **correlation between the original order and the order determined by the first two CA nontrivial solutions.**

There is no doubt that the comparison between CA and MDS as seriation techniques remains to be of chief interest.

Acknowledgements. The author is indebted to Professor D.G. Kendall for presenting the problem of seriation during her visiting at the Statistical Laboratory, University of Cambridge in 1990. The referees' suggestions making this paper clearer are also appreciated.

REFERENCES

CHOU, R. and HU, F. J. (1992). On the Z-representation of a statistic. *Theory of Probability and Its Applications* **1**, 219–221.

COURANT, R. and HILBERT, D. (1953). *Methods of Mathematical Physics.* New York.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *The Ann. Statist.* **7**, 2, 269–281.

GANTMACHER, F. R. and KREIN, M. G. (1950). Oscillation matrices and kernels and small variations of mechanical systems, 2nd Russ. ed., Izdat. Tekh. -Teor. Lit., Moscow-Leningrad. (German transl. by A. Stohr. Akademie-Verlag, berlin, 1960.) [MR 14(1953) 178; MR 22 (1961) 5161; zbl. 41(a952)355].

GREENACRE, M. J. (1984). *Theory and Application of Correspondence Analysis.* Academic Press, New York.

HILL, M. O. (1974). Correspondence analysis: a neglected multi-variate method. *Appl. Statist.* **23**, 340–354.

KENDALL, D. G. (1963). A Statistical Approach to Finders Petrie's sequence dating. *Bull. Intern. Statist. Inst.* **40**, 657–680.

KENDALL, D. G. (1971a). Abundance matrices and seriation in archaeology. *Zeit Wahrscheinlick-keitsstheorie* **17**, 104–112.

KENDALL, D. G. (1971b). Seriation from abundance matrices. In Mathematics in the Archaeological and Historical Sciences. (Hodson, C.R., Kendall, D.G. and Tautu, P. Eds). Edinburgh, Edinburgh University Press. PP.215–252.

LANCASTER, H. O. (1958). The structure of bivariate distributions. *Ann Math. Statist.* **34**, 719–736.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Applications.* Academic Press, New York.

SCHRIEVER, B. F. (1986). *Order Dependence.* CWI Tract 20, The Netherlands.

INSITITUTE OF STATISTICS
NATIONAL TSING HUA UNIVERSITY
HSIN CHU, TAIWAN