STATISTICAL PROPERTIES OF MULTIPLE CORRESPONDENCE ANALYSIS

Jan De Leeuw
Department of Data Theory FSW/RUL
Middelstegracht 4
2312 TW Leiden
The Netherlands

Paper presented at the conference on "New multivariate methods in statistics".
The 1984 Joint Summer Research Conference Series in the Mathematical Sciences.
June 10-16, Bowdoin College, Brunswick, Maine.

Abstract

Multiple correspondence analysis is a popular data analysis technique, but its statistical properties have not been studied so far. In this paper we linearize the statistics computed in multiple correspondence analysis, and we use these linearizations to compute delta method approximations to their standard errors. The results can be interpreted both in the standard framework of random sampling from a hypothetical infinite population, or in the 'non-stochastic' resampling framework based on postexperimental randomization.

Keywords

Correspondence analysis, delta method, randomization tests, distribution of eigenvalues.

INTRODUCTION

The technique studied in this paper is known under a large number of different names. We mention the most important ones. There is *principal components of scale analysis* (Guttman, 1941, 1950; Lord, 1958), *factorial analysis of qualitative data* (Burt, 1950), *second method of quantification* (Hayashi, 1956), *multiple correspondence analysis* (Benzécri, 1973, 1980; Cazes a.o., 1977; Lebart a.o., 1977), and *homogeneity analysis* (Gifi, 1981a). The technique has been derived from various data analytic points of view, starting either with ideas from multidimensional scaling, or from principal component analysis, or from scale analysis. We shall not discuss these various derivations here. We shall merely give a short outline of the one given by Gifi (1981a). Our paper starts with the equations defining the technique, and studies the statistical properties of the solutions to these equations. Various special cases have been studied before, but we shall attempt to derive them in a more uniform framework.

TECHNIQUE

The basic data we start with are n observations on m discrete variables. Suppose variable j can assume k_j different values. We can code our data using *indicator matrices* (also known as *dummies*). Indicator matrix G_j is n x k_j. It consists of zeroes and ones, and it has exactly one element equal to one in each row, indicating in which *category* of variable j *object* i belongs. By concatenating the G_j we obtain the n x K matrix G, with K the sum of the k_j. The row sums of G are all equal to m, the column sums are collected in the diagonal matrix D.

Multiple correspondence analysis can be defined as the application of *correspondence analysis* (Benzécri, 1973, 1980; Greenacre, 1984) or *dual scaling* (Nishisato, 1980), or *canonical analysis* (Kendall and Stuart, 1961, chapter 33) to the matrix G. This means that we want to find an n x p matrix X of *object scores* and a K x p matrix Y of *category quantifications* such that

$$GY = mX\Lambda, \tag{1a}$$

$$G'X = DY\Lambda, \tag{1b}$$

for some diagonal p x p matrix Λ. The number p is the *dimensionality* of the solution, it is chosen by the user.

Equations (1a) and (1b) can be interpreted in terms of *reciprocal averaging*. If we ignore the diagonal matrix Λ for the moment, then (1a) says that the score of an object is the average of the quantifications of the categories the object is in. And (1b) says that the quantifications of a category is the average of the scores of the objects in that category. The effect of Λ seems to be that these relations in terms of averages are both not precisely true. But we can rewrite (1a) and (1b) in such a way that at least one of these *centroid principles* is true. Define $Y = Y\Lambda$. Then

$$GY = mX\Lambda^2, \tag{2a}$$

$$G'X = DY. \tag{2b}$$

If we write (2b) as $Y = D^{-1}G'X$, we see that category quantifications are in the centroid of the scores of the objects in the categories. This is the scaling used in the multiple correspondence analysis program HOMALS (Gifi, 1981b), together with the normalization $X'X = nI$. Although scaling and normalization can be chosen in other ways too, for example in such a way that the dual centroid principle obtains, we shall use the Gifi-normalization in this paper too. Effectively (2a) and (2b) also define the *alternating least squares* algorithm used by Gifi to solve for object scores and category quantifications. An iteration starts with a current (X,Y). We then solve (2a) for a new X, with Y fixed at its current value. In this first step we also solve for a new Λ^2, but we do not require this matrix to be diagonal, only upper-triangular. It follows that the new X is the Gram-Schmidt orthogonalization of m^{-1} GY. This new X is then used in (2b) to solve for a new Y, which is not normalized in any way. It is easy to show that this procedure converges, and that Λ^2 in fact converges to a diagonal matrix. We shall briefly indicate a convergence proof below.

Equations (1a) and (1b) also show that multiple correspondence analysis is a (weighted) *singular value decomposition* of G. We know that with each singular value decomposition we can associate two dual eigenvalue problems in an natural way. If we eliminate Y from (1a) and (1b) we obtain

$$P_*X = X\Lambda^2, \tag{3}$$

with $P_* = m^{-1}GD^{-1}G'$. Matrix P_*, of order n x n, is the average of the m projectors $P_j = G_j (G'_jG_j)^{-1}G'_j$, which project orthogonally on the space spanned by the columns of G_j. Using

terminology and ideas from the analysis of variance or discriminant analysis field this could be called the *between-category space* of variable j. It follows from (3) that multiple correspondence analysis locates the n individuels in p-space in such a way that the within-category squared distances are small compared to the between-category squared distances. This relates the technique to multidimensional scaling, cluster analysis, and discriminant analysis. The Gifi-system of nonlinear multivariate analysis (Gifi, 1981a, De Leeuw, 1984b) is based on the basic geometric interpretation.

On the other hand we can also eliminate X from (1a) and (1b) and obtain the eigenvalue problem

$$CY = mDY\Lambda^2, \tag{4}$$

with C = G'G. The K x K matrix C is called the *Burt table* in the French correspondence analysis literature, after Burt (1950). It is really a supermatrix, with the k_j x k_l tables $C_{jl} = G'_jG_l$ as elements. The C_{jl} are, of course, the cross tables of variables j and l, or the *bivariate marginals*. We have D = diag(C), and the diagonal submatrices $D_j = G'_jG_j$, which are themselves diagonal matrices, contain the univariate marginals of the variables. Solving (4) means finding eigenvalues and eigenvectors of $m^{-1}D^{-1/2}CD^{-1/2}$. This relates multiple correspondence analysis to principal component analysis and other forms of canonical analysis. Compare De Leeuw (1982) for the details.

Before we proceed we must clear up one ambiguous point. The singular value problem (1a) and (1b) has, in general, min(n,K) one-dimensional solutions, from which we can pick p solutions in a great many ways. We are interested in the solution that corresponds with the p largest *nontrivial* singular values. The qualification 'nontrivial' is needed, because (1a) and (1b) have m trivial solutions. One of them, in which the corresponding column of both X and Y has all elements equal to one, has singular value equal to one. Inspection of (3) shows that this is actually the largest singular value. It is trivial, because it does not depend on the data in any way, only on our choice of coding the data. It is always there. The other m - 1 trivial solutions correspond with singular values equal to zero. They occur because the rank of G is always less than or equal to K - (m - 1). The m - 1 dimensions in the null space of G occur, because the vector with alle elements equal to one is in the column space of each of the individual G_j. This is, of course, all very well known in analysis of variance or dummy-variable contexts.

We now show that our alternating least squares algorithm, based on (2a) and (2b), actually converges to the solution we want, provided we introduce one small adjustment.

The proof is simple, because it merely consists of establishing the equivalence of our algorithm to the direct iteration method with orthogonalization for computing eigenvalues and eigenvectors of P_*. This method is very well known, and has been studied in great detail (Rutishauser, 1969, 1970; Stewart, 1969). In order to avoid convergence to the largest (trivial) singular value it suffices, theoretically, to start with an initial X orthogonal to the trivial solution, i.e. in deviations from the mean. Because of accumulation of round-off in actual computation we transform each new X to deviations from the mean during iteration. The known results, and this small adjustment, suffice to prove convergence to the solution we want. This is true even in the case in which there are multiple singular values.

In this paper we apply statistical large sample theory to the equations of multiple correspondence analysis, assuming that n is large. This makes it convenient to concentrate on (4), and to think of the elements of Y and Λ^2 as the basic 'parameters' that are estimated. There is a slightly different way to define C and D which is better for our purposes. The m variables define a single *multivariable* which can assume $R = k_1 \times \ldots \times k_m$ different values. The values are called *profiles*. Each profile corresponds with a binary vector of length K, indicating the categories of the m variables in the profile. The coding of profile r is a K-element vector g_r, which is the *direct sum* of the m subvectors of length k_j defined for each of the variables. We can compare this with the coding used in contingency table analysis and log-linear analysis, where profiles are represented as *direct products*. Now suppose profile r occurs n_r times in the data. Let $p_r = n_r/n$. Redefine C and D as

$$C = \sum_{r=1}^{R} p_r g_r g'_r, \tag{5a}$$

$$D = \text{diag}(C) = \sum_{r=1}^{R} p_r H_r, \tag{5b}$$

where $H_r = \text{diag}(g_r g'_r)$. Of course (5) redefines C and D merely by dividing by n. It follows from (1a) and (1b), by the way, that objects having the same profile get the same object score. Thus the singular value problem could be reformulated in terms of profile scores, using an R x K matrix G, with weights for the rows equal to the number of times the profile has occurred in the data. This formulation makes all the quantities computed by multiple correspondence analysis a function of the R-element vector of proportions p, which is very convenient for statistical purposes. For computational purposes we prefer the formulation in terms of an n x K matrix, because in most applications of multiple correspondence analysis (though certainly not in all) n is much smaller than R. In this paper

we start by ignoring this choice, and concentrate on Y and Λ^2. Statistical properties of the profile scores will be studied at a later stage.

LINEAR APPROXIMATION

For convenience we rewrite (4) as

$$CY = mDY\Omega, \tag{6}$$

with C and D now defined by (5a) and (5b), and with Ω short for Λ^2. We also fix the normalization of Y by requiring

$$Y'DY = m\Omega. \tag{7}$$

The matrices $\Delta_j = Y'_jD_jY_j$ are called *discrimination matrices* in De Leeuw (1984b). They measure the between-category variance for variable j. The average of the Δ_j is equal to Ω.

Standard large sample theory is based on linear approximation. We compute these linear approximations to Y and Ω, which are functions of p, in the neighborhood of a vector π. The formulas follow directly from perturbation theory of eigenvalue problems, which is reviewed, for example, by Wilkinson (1965, chapter 2), and which is discussed most fully by Kato (1970) or Baumgärtel (1972). In order to derive the formulas we define, for any nonzero isolated singular value λ_s, $x_{rs} = (g'_ry_s)/m\lambda_s$ and $z_{rs} = (y'_sH_ry_s)/m$. Of course all quantities we define are functions of p. In our expansions we shall indicate that they are evaluated at $p = \pi$ by using italic. For the linear approximation of an isolated nonzero eigenvalue we have the simple formula

$$\omega_s = \omega_s + \sum_{r=1}^{R} (x^2_{rs} - z_{rs})\,(p_r - \pi_r) + o(\|\,p - \pi\,\|). \tag{8}$$

Observe that, according to our normalizations,

$$\sum_{r=1}^{R} \pi_r x^2_{rs} = \sum_{r=1}^{R} \pi_r z_{rs} = \omega_s. \tag{9}$$

The formula for the approximation of the corresponding eigenvectors is somewhat more complicated. It is

$$y_s = y_s + \sum_{r=1}^{R} (p_r - \pi_r) \{\omega_s^{-1}(1/2x^2_{rs} - z_{rs})\, y_s -$$

$$\sum_{t \neq s} (\omega^2_t - \omega_s\omega_t)^{-1} (\lambda_s\lambda_t x_{rs}x_{rt} - \omega_s z_{rst})y_t\} + o(\| p - \pi \|). \qquad (10)$$

These linear approximations can be used for many purposes. Gifi (1981a) shows that we can use them to approximate the effect of leaving out objects, categories, or variables. This often gives more satisfactory results than the more exact, but also more conservative, approach based on inequalities also derived by Gifi. We shall use (8) and (10) for large sample purposes.

DELTA METHOD

If we want to compute asymptotic confidence regions for the eigenvalues we can do this directly from (8) by using the *delta method* (Rao, 1965, section 6a.2, or Tiago de Oliveira, 1982). If we assume simple multinomial sampling then the delta method, combined with the multivariate Moivre-Laplace central limit theorem, shows that $n^{1/2}$ ($\underline{\omega}_s$ - ω_s) is asymptotically multivariate normal, with means equal to zero, and disperion

$$\sum_{r=1}^{R} \pi_r (x^2_{rs} - z_{rs})(x^2_{rt} - z_{rt}). \qquad (11)$$

We use the convention of underlining random variables (Hemelrijk, 1966). Thus $\underline{\omega}_s$ is the observed eigenvalue under simple random sampling. Other assumptions which make $\underline{p}$ asymptotically normal could also be used, but generally they result in different expressions for the covariances (Van Praag, De Leeuw, and Kloek, 1986).

The covariance matrix in (11) can be estimated very easily. We can form the matrix with elements $\underline{x}^2_{is} - \underline{z}_{is}$, which has a row for each individuel, and compute its 'sample covariance matrix. Because of (9) the matrix is already in deviations from the column mean. We could also form $\underline{x}^2_{rs} - \underline{z}_{rs}$, a matrix with a row for each profile, and compute the dispersion matrix using weights $\underline{p}_r$. Wichever is more convenient. The dispersion matrix can be used to draw confidence ellipsoids or intervals, and to perform simple tests. If performing such tests we have to be careful, however. Our derivation of the linear

approximations assumed that the eigenvalues were isolated and nonzero. Thus hypotheses which assume that population eigenvalues are zero or are equal require additional analysis.

We now discuss the dispersion of the asymptotic normal distribution of $n^{1/2}$ ($\underline{y}_s - y_s$). Its form follows directly from (10), but it does not look very appetizing and we omit the precise formula. It is clear that we can again estimate it by computing the sample covariance matrix of a matrix, given essentially by (10), which has columns corresponding to the parameters and which has either n or R rows. Of course we may not need the compleet dispersion matrix. For p dimensions there are Kp parameters, and this defines a rather large matrix in most cases. It is suggested in Gifi (1981) that each of the K categories is used to compute its own ellipse, ignoring all between category covariances. This is very convenient for plotting, but in some cases it can be misleading. In other cases if may be preferable to compute the K x K dispersion matrices for each dimension separately, ignoring all between dimension covariance. We also have to remember, of course, that computing confidence ellipsoids for each point separately in general gives different information than computing a confidence ellipsoid for all parameters simultaneously and plotting its low-dimensional projections.

SPECIAL CASES

Some special cases of multiple correspondence analysis are mentioned separately in this section. This is either because statistical results have already been derived by others in these special cases, or because important simplifications are possible of our previous results.

First consider the case m = 2. Multiple correspondence analysis becomes equivalent to ordinary correspondence analysis or canonical analysis of a contingency table. Delta method results essentially equivalent to our results have been derived by O'Neill (1978a, 1978b, 1981) and Haberman (1981). In correspondence analysis we do not use the Burt table, but we compute singular values and singular vectors of $D_1^{-1/2}C_{12}D_2^{-1/2}$. This singular value problem is not exactly identical to our Burt table eigenproblem, but the relationship is very close indeed. Suppose $k_1 \geqq k_2$. Then there are k_2 singular values γs, all nonnegative, with the largest one equal to unity. The pairs of corresponding singular vectors are (y_{1s},y_{2s}). The eigenvalues of the Burt table are, in this case, the k_2 values $1/2\ (1 + \gamma_s)$ with eigenvector (y_{1s},y_{2s}), the k_2 values $1/2\ (1 - \gamma_s)$ with eigenvector $(y,0)$, where $y'D_1y_{1s} = 0$ for all s. Thus in a sense, there are actually only k_2 different solutions.

Another important special case has $k_j = 2$ for all j. In this case the Burt table has m + 1 nonzero eigenvalues, of which one is the trivial one equal to unity. The remaining m eigenvalues are the eigenvalues of the matrix of point correlations (phi coefficients) between the m binary variables. The eigenvectors y_s consist of m pieces of two elements each. The elements are determined, up to a scale factor, by the requirement that they must be orthogonal to the corresponding two-element vector of marginal frequencies. The scale factor turn out to be the elements of the eigenvector of the matrix of phi-coefficients. Thus in this case multiple correspondence analysis is principal component analysis of phi coefficients.

These two special cases have an important property in common. This is analyzed algebraically by De Leeuw (1982). Suppose that there exist k_j x k_j matrices Y_j such that $Y'_jD_jY_j = I$ and $C_{jl}Y_l = D_jY_j\Gamma_{jl}$ for all j and l, where Γ_{jl} is diagonal. If m = 2 or if $k_j = 2$ for all j, then this condition is trivially satisfied. De Leeuw shows that it is also true for a continuous multinormal distribution in a 'continuous Burt table', which means that it will by approximately true for a discreticized multinormal. If we take corresponding columns s from all Y_j, then the system of scores $(y_{1s},...,y_{ms})$ linearizes all bivariate regressions. It turns out that in many applications approximate linearizing systems exist. If they exist exactly, then multiple correspondence analysis will find them. The successive systems of scores with the regression linearizing properties are usually similar to orthogonal polynomials. If we order the scores in the appropriate way, we find that scores in y_{js} change sign exactly s times, and that zeroes of successive systems are interlaced. This is connected with the theory of *total positivity*, which has been applied to correspondence analysis by Schriever (1985).

There is another very important consequence of the existence of a system of scores that linearizes all bivariate regressions. Suppose $(y_1,...,y_m)$ is such a system, the dot indicates again that we are in the random sampling framework and have assumed the system to exist at $p = \pi$. The corresponding correlation coefficients are $\rho_{jl}=y'_jC_{jl}y_l$. We now linearize the correlation coefficient computed by using the multiple correspondence analysis scores at p, close to π. This gives

$$\rho_{jl} = \rho_{jl} + \sum_{r=1}^{R} (p_r - \pi_r) \{y'_jE^r_{jl}y_l - 1/2\rho_{jl} (y'_jE^r_{jj}y_j + y'_lE^r_{ll}y_l)\} + o (| | p - \pi | |). \qquad (12)$$

In formula (12) the E^r_{jl} are the (j,l) submatrices of $g_rg'_r$. The really interesting property of (12) is that we get the same linearization of the correlation coefficient if we consider the

scores as fixed in stead of as a function of the data. This is a consequence of the linearity, which was already discovered by Pearson in 1906 (cf De Leeuw, 1983a, for references and discussion).

Formula (12) has some important applications. Methods for the analysis of correlation coefficients usually assume multivariate normality, which means that they assume linear regressions. Pearson and his pupils already pointed out early in this century that the dispersion matrix of the correlation coefficients can be estmated quite easily without assuming normality by using fourth order product moments. The relevant formula have been given for the first time by Isserlis (1916), they are reviewed by De Leeuw (1983b). Of course they can also be used for categorical or ordinal variables, but they assume that the scores for the categories are known. Formula (12) implies that if we assume that the regressions can be linearized, then we can apply multiple correspondence analysis to compute the appropriate scores, and we can use Isserlis formula on the induced correlations. This gives consistent estimates of the parameters describing the correlation and regression coefficients in structural models, path models, and so on. Of course the assumption that the regressions can be linearized is much weaker than the assumption that the regressions are linear for given scores. It is very much weaker than the assumption of multivariate normality, which does not even make sense in situations with truly categorical data. The combination of multiple correspondence analysis with programs for fitting correlation models look promising.

TESTING PARTIAL INDEPENDENCE

Multiple correspondence analysis can also be used to test the hypothesis that one variable is pairwise independent of all others. The interpretation of multiple correspondence analysis in these terms may shed some light on the various statistics that are computed by the technique. The results in this section are a simple application of the general linearization formulas derived above.

If a variable, say variable j, is independent of all others, then multiple correspondence analysis will find $k_j - 1$ eigenvalues equal to 1/m. The corresponding eigenvectors y_s have zero elements for categories of all variables other than j. Of course we also have the 'trivial' eigenvector, with alle elements equal to one, and an additional m - 1 trivial eigenvalues equal to zero. There remain $(K - m) - (k_j - 1)$ nontrivial eigenvalues between zero and one, whose eigenvectors have zero elements for alle categories of variable j. We now apply (10)

to y_{js}, the segment of ys corresponding to the categories of variable j. We suppose that $\omega_s \neq 1/m$. Because $y_{js} = 0$ and, in fact, $y_{jt} \neq 0$ only if $\omega t = 1/m$, we find after some manipulation

$$y_{js} = -\omega_s (m^{-1} - \omega_s)^{-1} (D_j^{-1} - u_j u_j) \{ \sum_{r=1}^{R} (p_r - \pi_r) g_{jr} g'_r \} y_s + o(\| p - \pi \|). \qquad (13)$$

Using the delta method on (13) shows immediately that $n^{1/2} y_{js}$ is asymptotically normal with mean zero, and with asymptotic dispersion

$$V = \kappa^2_s (D_j^{-1} - u_j u'_j), \qquad (14)$$

where

$$\kappa_s = \omega_s/(m^{-1} - \omega_s). \qquad (15)$$

In both (13) and (14) u_j is a vector with all elements equal to one. From (14) we see that the dispersions of the category quantifications are proportional to the reciprocals of the marginal frequencies, with a proportionality factor that indicates haw far ω_s deviates from m^{-1}. An easy consequence of (14) is that

$$n y'_{js} D_j y_{js} \xrightarrow{L} \kappa^2_s \chi^2 (k_j - 1). \qquad (16)$$

Of course κ_s can be estimated consistently by inserting sample eigenvalues. Result (16), which holds in the case of independence of variable j from the rest (provided that $\lambda_s \neq m^{-1}$), is our final result in this section. It provides us with the asymptotic distribution of what Gifi (1981) calls the *discrimination measures*, i.e. the diagonal elements of the discrimination matrices. It suggests a simple test of the partial independence hypothesis, and a scale on which we can compare the category quantifications.

COMPLETE INDEPENDENCE

If all variables are independent of each other, then the results in the previous sections do not apply any more. All nontrivial population eigenvalues are equal to m^{-1}. A first result in this case was given by De Leeuw (1973). It is

$$n \sum_{s=1}^{K-m} (m\omega_s - 1)^2 = \sum_{j<l} \sum \chi^2_{jl}. \qquad (17)$$

This formula connects the sum of squares of the eigenvalues, or rather their variance, with the sum of all m over 2 off-diagonal bivariate chi-squares. In ordinary correspondence analysis the connection between the canonical correlations of a contingency table and the chi-square of that table is, of course, well known. Although (17) is interesting, its practical applicability seems limited.

If all variables are independent our asymptotics takes a different route. We study the matrix $D^{-1/2}CD^{-1/2}$. It is convenient to eliminate the trivial solutions first. We do this by using matrices W_j with $W'_jD_jW_j = I$. Moreover the first column of W_j consists of ones. Then form C, with submatrices $C_{jl} = W'_jC_{jl}W_l$. The eigenvalues of C/m are the same as our original ω_s, but now the diagonal submatrices of C are identities. Moreover the first row and column of each C_{jl} vanishes, except for the very first element, which is one. By leaving out the m first columns we get a matrix of order K - m, whose eigenvalues are all nontrivial ω_s. Our key result in this case, which is proved in the same way as the usual chi-squares partitioning results for ordinary two-way tables, is that the elements in the off-diagonal submatrices are all asymptotically independent standard normal. This result assumes complete independence of all variables, it is not enough to assume complete bivariate independence. This result easily implies (17), by the way.

Because eigenvalues are continuous functions of the matrix elements, it follows that the eigenvalues are distributed asymptotically as the eigenvalues of a matrix with independent standard normal variables in each of the off-diagonal blocks. This distribution obviously depends only on the vector $(k_1,...,k_m)$, but not on the marginals. This generalizes familiar results from the two-way case (given, for instance, in Lebart, 1976). Two-way results are simpler, because tabulated results for standard Wishart matrix eigenvalues can be used. Very few useful results are known about the asymptotic distribution of the eigenvectors in the case of complete independence.

VERY LARGE MATRICES

If K, the order of C and D, which is equal to the total number of categories over all variables, is very large, then it becomes interesting to study the limit distribution of the

eigenvalues. By this we mean the function F (x), which is equal to the number of nontrivial eigenvalues $\leqq$ x, divided by K - m. This type of limit distribution has been studied by many authors. We mention Wachter (1978) and Jonsson (1982), who also give many references to earlier work. We assume, in order to apply these results, that all variables are independent. Thus we remain in the 'null'-case. The basic result is that if (k - m)/n converges to some number τ if both K - m and n tend to infinity, then the distribution of the eigenvalues converges to a distribution with semi-circle density. The support and the scale of the density both depend on τ. Possible generalizations have been considered by Wachter (1976), who has also indicated the data analytic applications of his results (1975).

We do not have any experience yet with practical applications of eigenvalue limit distributions. This is also because our multiple correspondence analysis programs do not compute all eigenvalues, but only the first two or three. Moreover testing the null-hypothesis of complete independence is, in itself, rarely uninteresting, because it will usually be very far off the mark. It is too null.

NONSTOCHASTIC APPROACHES

Our results are based on linear approximation and the delta method. We assume that n is very large, and that the objects are a simple random sample from the population. In very many applications, certainly in the social and behavioural sciences, the model of a simple random sample (from an infinite hypothetical population) does not apply, or is very far-fetched indeed. Compare Freedman and Lane (1983), De Leeuw (1984a). For such situations some approaches have been developed which still assume that n is large, but which do not use the idea that the data 'are' in some sense a 'sample'. They are all based on some form of post-experimental randomization.

The delta method results on asymptotic normality, for instance, can be reformulated in the framework of the *bootstrap distribution*. The bootstrap distribution (Efron, 1979, 1982) is the distribution of the statistic under random sampling with replacement from the data. The bootstrap distribution for resamples of size n can be approximated by a multivariate normal, and this is what the delta method does. In more complicated cases it may be necessary to use the Monte Carlo version of the bootstrap, which actually draws a number of resamples from the data. The Monte Carlo version of the bootstrap has been used in (multiple) correspondence analysis by Gifi (1981) and Greenacre (1984). Gifi shows that there is in general a very good correspondence between Monte Carlo bootstrap

and delta method results, even if the number of resamples is as small as ten. In stead of the bootstrap distribution we can also use the *subsampling distribution* (Hartigan, 1971), which constructs its randomization by including each observation in the resample with probability 1/2.

We derived a test for partial independence from the delta method results. This means that we can imbed it in the same way in a resampling framework. But for significance testing ('null' situations) the alternative resampling framework that uses permutations is perhaps most natural. We study the distribution of our statistics under random permutations of the n observations on the variable whose independence from the other we want to test. The combinatorial or permutational central limit theorem gives us asymptotic normality of the 'off-diagonal' part of C (the cross table of variable j with all others). The diagonal parts are, of course, fixed. From this point the delta method takes over, and we find the same results as in the random sampling framework. The same thing is true for testing complete independence. We now use independent random permutitions of the observations for each of the m variables. This leaves the univariate marginals fixed, and gives a C with the same asymptotic distribution as under random sampling from an infinite population. Compare De Leeuw and Van der Burg (1986).

Freedman and Lane (1983) call this a *nonstochastic interpretation* of significance testing. It is very similar in spirit to the approach of randomization testing advocated, for instance, by Edgington (1980). It may not appeal to statisticians who emphasize primacy of, preferably parametric, models, but it seems a logically sound way to study a modes and well-defined question. In Gifi (1981) resampling and permutation methods are introduced as methods for *stability analysis*, i.e. as statistical or Monte Carlo methods to assess the size of derivatives and differentials. Basu (1980) criticizes the inferential properties of pest-experimental randomization procedures, but this criticism is hardly relevant in situations where no inference (from sample to population) is intended.

OBJECT SCORE ASYMPTOTICS

We have linearized category quantifications and eigenvalues, but not object scores. If we remember, however, the definition $x_{rs} = (g'_r y_s)/m\lambda_s$ of the profile scores, then it becomes immediately clear that these profile scores can be linearized too, and that we can use our previous formulas to obtain the required results. This does not add anything new, and we omit the details.

In some cases, however, another kind of asymptotics is called for. Suppose n is fixed, and m tends to infinity. We 'sample' from a universe of variables. It now becomes convenient to use the dual eigenproblem (3). Because of the discreteness of our variables projector P_j, of rank k_j, can assume only a finite number of values (the number of ways n balls can be placed in k_j cells). Suppose, for simplicity, that all variables have the same number of categories. Then $P_* = \sum q_a P_a$, where the P_a are the possible values and q_a are their relative frequencies. In this form the standard perturbation (or linearization) results for eigenanalysis apply again. No details are presented here, because by now the principles are clear. Nonstochastic interpretations can also be given.

The type of asymptotics we apply depends, obviously, on the relative size of n and m (or n and K). If both n and K are very large, which is a common situation in multiple correspondence analysis, the choice between the two becomes complicated. In fact here we only have results for the relatively unintersting null-case of complete independence, and these results are moreover limited to eigenvalues. It is clear that additional results are moreover limited to eigenvalues. It is clear that additional results in this direction are needed, but it is not yet obvious how one should proceed to derive useful approximation results in these double limit cases.

REFERENCES

Basu, D. (1980). Randomization analysis of experimental data: the Fisher randomization test. *Journal of the American Statistical Association*, 75, 575-595.

Baumgärtel, H. (1972). *Endlichdimensionale amalytische Störungstheorie*. Berlin, Akademie Verlag.

Benzécri, J.P. a.o. (1973). *L'Analyse des Données*. (2 vols). Paris, Dunod.

Benzécri, J.P. a.o. (1980). *Pratique de l'Analyse des Données*. (3 vols). Paris, Dunod.

Burt, C. (1950). The Factorial Analysis of Qualitative Data. *British Journal of Statistical Psychology*, 3, 166-185.

Cazes, P. Baumerder, A., Bonnefous, S., Pagès, J.P. (1977). Codage et Analyse des Tableaux Logiques. Introduction à la Pratique des Variables Qualitatives. *Cahiers du BURO*, n° 27, Paris, Université Pierre et Marie Curie.

De Leeuw, J. (1973). *Canonical Analysis of Categorical Data*. Unpublished doctoral dissertation, University of Leiden, Reissued by DSWO-Press, Leiden, 1984.

De Leeuw, J. (1982). Nonlinear Principal Component Analysis. In H. Caussinus, P. Ettinger, R. Tomassone (eds.) *COMPSTAT 1982*, Wien, Physika Verlag.

De Leeuw, J. (1983a). On the Prehistory of Correspondence Analysis. *Statistica Neerlandica*, 37, 161-164.

De Leeuw, J. (1983b). Models and Methods for the Analysis of Correlation Coefficients. *Journal of Econometrics*, 22, 113-138.

De Leeuw, J. (1984a). Models of Data. *Kwantitatieve Methoden*, 5, 17-30.

De Leeuw, J. (1984b). The Gifi System of Nonlinear Multivariate Analysis. In E. Diday a.o. (eds.) *Data Analysis and Informatics III*, Amsterdam, North Holland.

De Leeuw, J. and Van der Burg, E. (1986). The Permutation Distribution of Generalized Canonical Correlation Coefficients. In E. Diday a.o. (eds) *Data Analysis and Informatics IV*, Amsterdam, North Holland.

Edgington, E.D. (1980). *Randomization Tests*. New York, Marcel Dekker.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 7, 1-26.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*. Philadelphia, SIAM.

Freedman, D.A. and Lane, D. (1983). Significance Testing in a Nonstochastic Setting. In P.J. Bickel, K. Doksum, J.L. Hodges jr. (eds.) *A Festschrift for Erich L. Lehmann* . Belmont, Wadworth.

Gifi, A. (1981a). *Nonlinear Multivariate Analysis*. Leiden, Department of Data Theory FSW/RUL. New edition by DSWO-Press, Leiden, 1987.

Gifi, A. (1981b). *HOMALS Users Guide*. Leiden, Department of Data Theory, FSW/ RUL.

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

Guttman, L. (1941). The Quantification of a Class of Attributes: a Theory and Method of Scale Construction. In P. Horst (ed.) *The Prediction of Personal Adjustment*. New York, Social Science Research Council.

Guttman, L. (1950). The Principal Components of Scale Analysis. In S.A. Stouffer (ed.) *Measurement and Prediction*, Princeton, Princeton University Press.

Haberman, S.J. (1981). Tests for Independence in Two-way Contingency Tables based on Canonical Correlation and on Linear by Linear Interaction. *Annals of Statistics*, 9, 1178-1186.

Hartigan, J. (1971). Error Analysis by Replaced Samples. *Journal of the Royal Statistical Society*, B33, 98-110.

Hayashi, C. (1956). Theory and Examples of Quantification II (In Japanese). *Proceedings of the Institute of Statistical Mathematics*, 4, 19-30.

Hemelrijk, J. (1966). Underlining Random Variables. *Statistica Neerlandica*, 20, 1-8.

Isserlis, L. (1916). On Certain Probable Errors and Correlation Coefficients of Multiple Frequency Distributions with Skew Regression. *Biometrika*, 11, 185-190.

Jonsson, D. (1982). Some Limit Theorems for the Eigenvalues of a Sample Covariance Matrix. *Journal of Multivariate Analysis*, 12, 1-38.

Kato, T. (1970). *Perturbation Theory of Linear Operators*. Berlin, Springer.

Kendall, M.G. and Stuart, A. (1967). *The Advanced Theory of Statistics*. (3 vols.). London, Griffin.

Lebart, L. (1976). The Significance of Eigenvalues issued from Correspondence Analysis of Contingency Tables. In J. Gordesch and P. Naeve (eds.) *COMPSTAT 1976*. Wien, Physika Verlag.

Lebart, L., Morineau, A., Tabart, N. (1977). *Techniques de la Description Statistique: Méthodes et Logiciels pour l'Analyse des Grands Tableaux*. Paris, Dunod.

Lord, F.M. (1958). Some Relations between Guttman's Principal Components of Scale Analysis and other Psychometric Theory. *Psychometrika*, 23, 291-296.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto, University of Toronto Press.

O'Neill, M.E. (1978a). Asymptotic Distribution of the Canonical Correlations from Contingency Tables. *Australian Journal of Statistics*, 23, 58-66.

O'Neill, M.E. (1978b). Distributional Expansions for Canonical Correlations from Contingency Tables. *Journal of the Royal Statistical Society*, B40, 303-312.

O'Neill, M.E. (1981). A Note on the Canonical Correlations from Contingency Tables. *Australian Journal of Statistics*, 23, 58-66.

Rao, C.R. (1965). *Linear Statistical Inference and its Applications*. New York, Wiley.

Rutishauser, H. (1969). Computational Aspects of F.L. Bauers Simultaneous Iteration Method, *Numerische Mathematik*, 13, 4-13.

Rutishauser, H. (1970). Simultaneous Iteration for Symmetric Matrices. *Numerische Mathematik*, 16, 205-223.

Schriever, B.F. (1985). *Order Dependence*. Amsterdam, Center for Mathematics and Informatics.

Stewart, G.W. (1969). Accelerating the Orthogonal Iteration for the Eigenvectors of a Hermitian Matrix. *Numerische Mathematik*, 13, 362-376.

Tiago De Oliveira, J. (1982). The Delta-method for Obtention of Asymptotic Distribu tions; Applications. *Publications de l'Institut de Statistique de l'université de Paris* , 27, 49-70.

Van Praag, B.M.S., De Leeuw, J., & Kloek, T. (1986). The Population-Sample Decomposition Approach to multivariate estimation methods. *Applied Stochastic Models and Data Analysis*, 2, 99-119.

Wachter, K.W. (1975). *Users Guide and tables to Probability Plotting for Large-scale Principal Component Analysis*. Report W-75-1. Department of Statistics, Harvard University.

Wachter, K.W. (1976). Probability Plotting Points for Principal Components. In *Proceedings Ninth Interface Symposium Computer Science and Statistics*. Boston, Prindle, Weber, and Schmidt.

Wachter, K.W. (1978). The Strong Limits of Random Matrix Spectra for Sample Matrices of Independent Elements. *Annals of Probability*, 6, 1-18.

Wilkinson, J.H. (1965). *The Algebraic Eigenvalue Problem*. Oxford, Clarendon Press.