Journal of Econometrics 22 (1983) 113–137. North-Holland Publishing Company

MODELS AND METHODS FOR THE ANALYSIS OF CORRELATION COEFFICIENTS

Jan DE LEEUW

Department of Data Theory, University of Leyden, 2311 CS Leyden, The Netherlands

The normal distribution plays a very prominant role in statistics, and in multivariate analysis the multinormal model is even more dominant than in the classical univariate or bivariate situations. In this paper we show that the assumption of continuous multinormal observations can be relaxed in various ways. For continuous multivariate analysis most of the existing large-sample theory can be applied without assuming multivariate normality. We also develop several discrete multinormal models. The first one is the classical Pearson model, which results from discreticizing multinormal variables. The second one is inspired by correspondence analysis, and the third one by loglinear analysis. The models are presented in terms of correlation coefficients, and applied to the Spearman model of factor analyses using an example from political science.

1. Introduction

In this paper we discuss a number of generalizations of classical multivariate analysis. These generalizations make it possible to analyze random samples from multivariate distributions which are not necessarily multivariate normal. In fact we concentrate on situations where the random variables are not even real valued, but assume their values in a partially ordered set. Of the three classical assumptions of multinormal, independent, and identically distributed observations we only retain the two last ones. And some of our results remain valid even if these last two assumptions are relaxed in appropriate ways.

The techniques we discuss try to remain as close as possible to the classical techniques of continuous multivariate analysis, discussed in a modernized form in this issue by Bentler (1983). We generalize the model as far as possible, with the restriction that the final parametrizations and estimators should remain very similar to those of continuous multivariate analysis. As a consequence our techniques are much more similar to the classical multinormal techniques than to the loglinear techniques discussed in this issue by Fienberg and Meyer (1983).

In order to stay close to multinormal analysis we shall concentrate on the analysis of correlation coefficients between pairs of variables. Our general approach can be used equally well, however, on means, covariances, and

0165–7410/83/$03.00 © Elsevier Science Publishers

even higher (product)-moments. The model we choose to illustrate our techniques is Spearman's (1904) two-factor model, which we consider to be by far the most important single model in the history of psychometrics. Of course our choice of the Spearman model as an illustration is not essential either, any other model framed in terms of product moments would have served equally well.

Because the Spearman model may not be familiar to all our readers we briefly review it here. In its simplest form the model is $\underline{x}_j = \theta_j \underline{z} + \delta_j \underline{u}_j$, with $\underline{x}_j$ the observed random variables, $\underline{z}$ the unobserved or latent common factor, and $\underline{u}_j$ the unobserved specific factors. We use the convention that random variables are underlined [Hemelrijk (1966)]. The θ_j and δ_j are constants, called, respectively, factor loadings and uniquenesses. All variables have zero expectation and unit variance, moreover the specific factors are supposed to be uncorrelated with the common factor. It follows that γ_{jl}, the correlation between $\underline{x}_j$ and $\underline{x}_l$, is given by $\gamma_{jl} = \theta_j \theta_l$. A correlation matrix satisfying these Spearman conditions is called a Spearman matrix. The model is called a two-factor model because each observed variable is composed of two unobserved ones: the common factor and a specific factor. More recently, however, the model is also called the single-factor model, because the variables have only a single common factor.

Our paper is related to the ones by Bentler and Muthén in this issue, but we have a somewhat different emphasis. We try to relax the assumptions of multivariate analysis and we try to simplify the computations, using results that have been familiar in the statistical literature for a very long time. Until recently, however, these results have been largely ignored in the applied psychometric and econometric literature. They make it possible to de-emphasize multivariate normality and maximum likelihood. We shall show that this can be done without any loss of either computational or statistical efficiency, in fact with a gain of efficiency in both respects. The paper starts with some general comments on the nature of multivariate data, and on linear multivariate analysis.

2. The nature of data

It is clear that in the strict operational sense all variables we ever observe in practice are categorical or discrete, i.e., assume only a finite number of values. This is due to the fact that our measurement procedures have a limited precision, which entails that even very precise measurement of physical variables does not really define continuous variation. Of course this point of view is especially relevant in the social and behavioural sciences, in which comparatively few variables can be realistically assumed to be 'approximately continuous'. This conclusion is independent of the metaphysical question if variation in nature is ever truly continuous, a

question which is at the root of the famous dispute between Pearson and Yule [cf. Norton (1978), MacKenzie (1978)]. The assumption that observations are independent realizations of multinormal variables is never true in a strict sense, and must thus always be rejected in very large samples. The distinction between discrete and continuous variables is consequently of limited value, although it may happen that a variable is 'sufficiently continuous' to make a multinormal model an interesting approximation. And this corresponds exactly with the historical origin of the normal distribution: multinormal models are numerical approximations to discrete probability models.

Of course it is generally true that models are approximations. The assumptions of independence is equally untrue in most practical situations, even if we carefully design our experiments. And even in discrete probability models we use various approximations, for example by assuming that there is an infinite number of observations. Models are never true, everything is significant if the sample is large enough. This is as it should be, and it should not prevent us from trying to make out models as realistic as possible.

In this paper we distinguish quantified variables, ordered variables, and unordered variables. This simply means variables which have values in a set of real numbers, in an ordered set, and in an arbitrary set. Of course a set of real numbers is a special ordered set, and an ordered set is a special arbitary set. Thus a quantified variable is a special ordered variable, and an ordered variable is a special unordered variable. If we have quantified variables we can ignore the numerical values and use only the order of these values in the model and/or the analysis. If we have an ordered variable we can replace the values by real numbers in some more or less rational way, and use these numbers in the model and/or analysis. Thus quantified, ordered, and unordered do not really refer to properties of the variables, but to aspects of the variables that we may or may not model and/or use in the analysis.

3. Linear analysis of quantified variables

In order to get a good perspective we shall start by discussing some of the usual linear multivariate analysis techniques for the analysis of correlation matrices. As is implied by our discussion in the previous section we can restrict ourselves (without any real loss of generality) to discrete variables which assume only a finite number of values. In this section these values are real numbers, i.e., the variables are quantified. Let us suppose that there are m variables and n observations, on the basis of these observations we can compute the correlation matrix $\underline{R}$. We want to fit a model of the form $\underline{R} \sim \Gamma(\theta)$, i.e., the population correlation coefficients are a specified function of the parameters θ. The observed correlation coefficients $\underline{R}$ must be approximately of this form.

What is the usual procedure in this case? In the first place it assumes a multinormal parent population. Fisher (1962) has derived the joint distribution of the correlation coefficients in this case. We can consequently use his results to compute maximum likelihood estimates of the parameters θ. Jöreskog (for example, 1978) proceeds differently. He computes the sample covariance matrix $\underline{C}$ and fits the model $\underline{C} \sim \Delta\Gamma(\theta)\Delta$, with Δ a diagonal matrix of standard deviations, also considered as free parameters which have to be estimated. The two approaches are essentially equivalent, but Fisher's method can be used in secondary analysis of correlations for which the corresponding standard deviations are unknown. Both methods use maximum likelihood estimation under normality assumptions, also known as full information maximum likelihood [cf. also McDonald (1975)].

A possible alternative, which sometimes leads to less computation, is to use generalized or weighted least squares. This method, which is also based on the assumption of multivariate normality, is applied to a class of models called 'patterned correlation matrices' in Jennrich (1970), Browne (1977) and Steiger (1980). If we say that the method is based on assuming multivariate normality, then we mean that the method gives estimates with the same first-order asymptotic properties as the maximum likelihood estimates if the data come from a multinormal distribution. In the case of multivariate normality both maximum likelihood and generalized least squares give estimates which are efficient in the usual sense, i.e., of all uniformly consistent asymptotically normal estimates they have the smallest possible dispersion matrix and consequently the maximum concentration around the true value. If we use the words good, better, best for estimates, we always mean them in this precise sense.

If the parent is non-normal both maximum likelihood and generalized least squares give consistent asymptotically normal estimates, but the asymptotic dispersions may differ, and either one of them could be better. Browne (1974) has remarked that generalized least squares is asymptotically first-order equivalent to multinormal maximum likelihood if the data are a random sample from a distribution with vanishing fourth order cumulants. Although this does not appear to be a major generalization, it clearly indicates in which direction we should look. In fact it is obvious from the Bentler paper in this issue that Browne (1982) goes beyond this result in his asymptotically distribution-free analysis of covariance matrices.

Both maximum likelihood and generalized least squares estimates are functions of the sample correlation coefficients, we can write $\underline{\theta} = \Phi(\underline{R})$. By an easy application of the maximum theorem [Berge (1963) or Dantzig et al. (1967)] the function Φ is continuous if the function Γ is continuous, in fact by the implicit function theorem the function Φ is continuously differentiable if Γ is. Both maximum likelihood and generalized least squares estimates are Fisher-consistent in the sense that $\Phi(\Gamma(\theta)) = \theta$ for all θ. 'A statistic is said to

be a consistent estimate of any parameter, if when calculated from an indefinitely large sample it tends to be accurately equal to that parameter' [Fisher (1925, p. 702)].

Now consider the class of all estimates of the form $\Phi(\underline{R})$, satisfying both continuous differentiability and Fisher-consistency. By the general theory of best asymptotically normal estimates [LeCam (1956), Wijsman (1959, 1960)] this class contains an estimate with the smallest possible dispersion, i.e., an estimate which is best in our sense. This estimate can be computed by minimizing the generalized least squares criterion

$$\delta(\theta)=\sum\sum\sum\sum v_{ijkl}(\underline{r}_{ij}-\gamma_{ij}(\theta))\underline{r}_{kl}-\tfrac{1}{7}(\theta)), \tag{1}$$

where $\underline{V}$ is the inverse of a consistent estimate of the dispersion matrix of the $\underline{r}_{ij}$ (thus $\underline{V}$ does not depend on θ, it is a function of the observations only). Efficiency in this case is based on the fact that the $\underline{r}_{ij}$ are differentiable functions of the sample product moments, they are consequently asymptotically normal if the sample moments are asymptotically normal. In this sense it is not necessary to assume that the observations are independent and identically distributed, the central limit theorem proves asymptotic normality of the sample moments under far more general conditions. Of course the form of the dependence between observations does have an influence, because it may imply that we need another estimate of the dispersion matrix of the $\underline{r}_{ij}$. This is what Van Praag (1982) calls population-sample decomposition: the characteristics of the sampling procedure are used in the consistent estimation of the dispersion, as soon as we have such an estimate we can estimate our population parameters by minimizing (1).

In the case in which we have independent and identically distributed replications we can estimate the covariance of the $\underline{r}_{ij}$ by using the formula [for example given by Hsu (1949)]

$$\begin{aligned}\underline{w}_{ijkl}=\underline{r}_{ijkl}&-\tfrac{1}{2}\underline{r}_{ij}(\underline{r}_{iikl}+\underline{r}_{jjkl})-\tfrac{1}{2}\underline{r}_{kl}(\underline{r}_{kkij}+\underline{r}_{llij})\\&+\tfrac{1}{4}\underline{r}_{ij}\underline{r}_{kl}(\underline{r}_{iikk}+\underline{r}_{iill}+\underline{r}_{jjkk}+\underline{r}_{jjll}),\end{aligned} \tag{2}$$

with

$$\underline{r}_{ijkl}=\underline{s}_{ijkl}\underline{s}_{ii}^{-\frac{1}{2}}\underline{s}_{jj}^{-\frac{1}{2}}\underline{s}_{kk}^{-\frac{1}{2}}\underline{s}_{ll}^{-\frac{1}{2}}, \tag{3}$$

and

$$\underline{s}_{ijkl}=n^{-1}\sum_{v=1}^{n}(\underline{x}_{vi}-\underline{m}_i)(\underline{x}_{vj}-\underline{m}_j)(\underline{x}_{vk}-\underline{m}_k)(\underline{x}_{vl}-\underline{m}_l), \tag{4}$$

with $\underline{m}_i$ the sample mean of variable i, and $\underline{s}_{ii}$ the sample variance. The data are collected in the $n\times m$ matrix $X=\{x_{vj}\}$. In the appendix we give a general

procedure which can be used to derive expressions like (2). Here we merely remark that (2) is asymptotically distribution free in the sense of Browne (1982) and Bentler in this issue, because if $n\rightarrow\infty$ the random variable $\underline{w}_{ijkl}$ converges in probability to the true covariance of $\underline{r}_{ij}$ and $\underline{r}_{kl}$, no matter what the population was we have been sampling from.

Estimates $\underline{\theta}$ minimizing (1) with $\underline{V}$ the inverse of $\underline{W}$ defined by (2) are best in the class $\Phi(\underline{R})$, at least if the observations are independent and identically distributed. In particular they are usually better that the maximum likelihood estimates and the generalized least squares estimates based on assuming multivariate normality. They are asymptotically equivalent to these multinormal estimates if the fourth-order population cumulants vanish. If this is the case we can use

$$\underline{r}_{ijkl}=\underline{r}_{ij}\underline{r}_{kl}+\underline{r}_{ik}\underline{r}_{jl}+\underline{r}_{jl}+\underline{r}_{il}\underline{r}_{jk} \tag{5}$$

in (2), and we get the more familiar formula for $\underline{w}_{ijkl}$ given by Pearson and Filon in 1898 [cf. Steiger (1980, eq. 3)]. The 'normal moment relations' (5) make it possible to compute $\underline{w}_{ijkl}$ from the observed correlations only. In this case Jennrich (1970) has given a computationally efficient method to compute the inverse $\underline{V}=\{\underline{v}_{ijkl}\}$. In the general case his formula does not apply. Thus if (5) is true the computational gain in using (5) can be considerable. The (first-order asymptotic) statistical gain is nil.

It is clear from the results so far that multinormal maximum likelihood and multinormal generalized least squares cannot be justified from first-order asymptotic considerations alone. We have outlined a method which is always at least as good, and usually better. For some strange reason psychometricians and econometricians largely have ignored this result. It has been suggested that there is a similar 'conspiracy of silence' in mathematical statistics too, led by people who are exceptionally and irrationally fond of maximum likelihood [Berkson (1980), LeCam (1980)]. Be this as it may, there have been a number of recent attempts to vindicate maximum likelihood by using the concept of second-order efficiency [again a good introduction is the discussion of Berkson (1980)]. This concept can be used, however, only if we first adjust efficient estimates to have the same first-order bias. It also assumes, of course, that the multinormal model is strictly true. We have already seen that any argument based on this last assumption has at most a limited practical value. But in the end, of course, large sample theory only suggests good estimates. In many cases the only practical way to compare them is by Monte Carlo. Steiger (1980) is a good example in this context.

Computationally the minimization of (1) can still be quite demanding. There are, however, alternative procedures which are simpler computationally and which give asymptotically equivalent estimates. We refer to the improvement methods for computing best asymptotically normal

estimates suggested by Neyman (1949), LeCam (1956) or Ferguson (1958). Again these methods have been largely ignored in the applied literature, and again for mysterious reasons. Mooijaart (1978) has used them for general latent structure analysis, Brown (1982) and Bentler in this issue are now applying them to covariance matrix analysis. Bentler calls them 2SADF estimates, which is short for two-stage asymptotically distribution-free. The general idea is easy to explain. Suppose $\underline{\theta}_0$ is any initial root-n-consistent estimate, i.e., $n^{\frac{1}{2}}(\underline{\theta}_0 - \theta)$ must be bounded in probability. In most cases such an initial estimate can easily be computed by solving some or all of the equations $\underline{R} = \Gamma(\theta)$, or by some other version of the classical method of moments. In the case of the two-factor theory we can use the methods developed by Spearman (1927, app.) himself, for example. Now define

$$\delta(\theta, \underline{\theta}_0) = \sum\sum\sum\sum \underline{v}_{ijkl}(\underline{r}_{ij} - \gamma_{ij}(\theta, \underline{\theta}_0))(\underline{r}_{kl} - \gamma_{kl}(\theta, \underline{\theta}_0)), \tag{6}$$

with

$$\gamma_{ij}(\theta, \underline{\theta}_0) = \gamma_{ij}(\underline{\theta}_0) + \sum \underline{g}_{ijs}(\theta_s - \underline{\theta}_{0s}), \tag{7}$$

and

$$\underline{g}_{ijs} = \partial\gamma_{ij}/\partial\theta_s \big|_{\theta=\theta_0}, \tag{8}$$

and improve the estimate $\underline{\theta}_0$ by minimizing (6) over θ. This can be interpreted as performing a single Gauss–Newton iteration starting from $\underline{\theta}_0$. Thus we have developed estimates which can be computed non-iteratively, by solving a single linear least squares problem, and which are superior from the first-order efficiency point of view to multinormal maximum likelihood or generalized least squares estimates. They are 'model-free' in the sense of Van Praag (1981), although the name 'asymptotically distribution-free' is more precise.

We apply this method to an example taken from the survey of parliament conducted by the Institute of Political Science of Leiden University in 1972. All members of the Second Chamber of the Dutch Parliament were asked to indicate his or her position with respect to six political issues. Responses were coded in three categories, indicating the amount by which they agreed with the statement made about the issue. The six statements were:

(a) The government should spend less money on aid to developing countries.
(b) The government should take stronger action against public disturbances.
(c) Income differences should remain as they are.
(d) Only management should decide important matters in industry.
(e) Taxes should be decreased so that people can decide for themselves how to spend their money.
(f) The government should insist on maintaining strong Western armies.

In fact the six statements were presented in a bipolar form together with their opposites. These opposites are more or less obvious: more money on aid, takes too strong action, should become much less, workers should participate, taxes increased for general welfare, shrinking Western armies. We only used the data from the 119 members of the Second Chamber who belonged to the seven largest political parties and who had no missing entries. Thus $m=6$, $n=119$, and each variable takes only three values.

We must emphasize that the example we use is not ideal in several respects. In the first place 119 observations is far too few for the asymptotic theory to become really relevant. But more importantly our 119 members of parliament cannot be considered to be a random sample from any well-defined population. They are certainly no random sample from the Dutch Second Chamber, which has a total of 150 members. No serious person will maintain that members of parliament are a random sample from the population of all voters. In fact we deal with a population here, the Dutch Second Chamber, with some members missing. And of course if we analyze populations inferential statistics becomes irrelevant: everything is significant, all estimates are exactly correct. We continue to use the example for pragmatic reasons. Our computer programs at the moment of writing this paper were in APL, and larger examples than this one could not be analyzed.

Our first analysis uses integer scoring of the categories. The correlations are given in table 1. Simple unweighted least squares estimates of the common factor loadings are in the first column of table 2 (for the interpretation we remember that loadings are correlations of the common factor with each of the observed variables). The loadings are differentiable functions of the correlations, thus they are asymptotically normal, and we can compute asymptotically distribution-free estimated standard deviations using (2). They are in the second column of table 2. If we assume that (5) is true, then we find other consistent estimates of the standard deviations, given in column 3 of table 2. We can compute asymptotically distribution-free efficient estimates by using formulas (2), (6), (7) and (8). They are in the first column of table 3, their estimated standard deviations are in the second column. Estimated loadings using in addition (5) are in the third column of table 3, and their estimated standard deviations are in the fourth column. The goodness-of-fit chi-square for the 2SADF estimates, which is simply the minimum value of (6) multiplied by the number of observations, is 25. If we use (5) it is 18, both chi-squares have $15-6=9$ degrees of freedom. Observe that we report chi-square values in integers, they must be taken with several grains of salt anyway, because of the peculiarities of the example.

If we compare tables 2 and 3 we see that the efficient estimates tend to be higher than the consistent initial estimates. One way to show this is by using the determinacy index of Guttman (1955), which is the multiple correlation

Table 1

Correlations using integer scores.

AID	1.000					
LAW	0.446	1.000				
INC	0.462	0.380	1.000			
PAR	0.398	0.241	0.589	1.000		
TAX	0.583	0.536	0.569	0.459	1.000	
ARM	0.516	0.483	0.417	0.403	0.514	1.000

Table 2

Least squares estimates of factor loadings.

		Standard deviations	
	Least squares	Non-parametric	Normal
AID	0.709	0.058	0.056
LAW	0.602	0.062	0.067
INC	0.707	0.071	0.057
PAR	0.601	0.069	0.068
TAX	0.807	0.060	0.045
ARM	0.677	0.065	0.060

Table 3

Improved least squares estimates of factor loadings.

	Non-parametric		Normal	
	WLS	SIG	WLS	SIG
AID	0.704	0.054	0.735	0.054
LAW	0.648	0.057	0.675	0.062
INC	0.763	0.064	0.786	0.051
PAR	0.683	0.058	0.708	0.060
TAX	0.839	0.054	0.823	0.044
ARM	0.718	0.057	0.711	0.058

between the variables and the common factor. For the least squares solution this is 0.854, for the non-parametric efficient solution it is 0.881, and for the multinormal efficient solution it is 0.886. On the average the variables have about 50% common and 50% unique variance. The results indicate that the two-factor model fits quite nicely, a fairly satisfactory 'left–right' scale can be constructed on the basis of these six questions. This is remarkable, since we

use the rather arbitrary 1/2/3-scoring system for the categories of our variables. Observe that the first-order asymptotic theory predicts that the standard deviations of the efficient estimates are lower than those of the consistent estimates. Comparing tables 2 and 3 shows that this is indeed true in our example. The theory also predicts that the standard deviations from the non-parametric theory should be lower than those from the multinormal theory. By comparing columns 2 and 4 of table 3 it is clear that this is not true in our example, or, more precisely, it is not reflected in the estimates of the standard deviations. This is unfortunate, but given the unrealistic example it need not worry us a great deal. It may indicate, however, a very real and important effect. The asymptotically non-parametric theory uses fourth-order moments, which have very large standard deviations. This could easily lead to larger small sample standard errors for the estimates, it could also lead to poor estimation of the standard errors of estimate in small samples. We need further study of more realistic examples, and possibly some Monte Carlo work, to see how serious this problem is.

4. Linear analysis of ordered variables

In the previous sections we have argued that the assumption of continuous multinormality is not very realistic in most situations in the social and behavioural sciences. Moreover it is not necessary to make this assumption in the case of quantified variables, because we can get results which are equally good or better from a statistical point of view by using a model-free or asymptotically distribution-free approach. On the other hand it is clear that in the previous section the multinormal distribution still looms large in the background, if only for the fact that we are interested in correlations. Now correlations are useful measures of association if both regressions are linear, i.e., for example in the case of multivariate normality. Moreover we ignore all moments and product moments of order higher than two in our modelling, which causes no loss of information in the case of multivariate normality. Thus in some respects we still act as if the data have many important properties in common with multinormal data.

Because of this it becomes interesting to investigate more realistic models, which are also based conceptually on the multinormal distribution. They take the essential discreteness of our measurements into account, however, and they can actually cope with situations in which some of the variables are measured in a very small number of categories, possibly even only two categories. For our first model we still assume (initially) that the observations are independent and identically distributed, the appropriate model for n observations which are categorical is thus multnomial. We assume that the multinomial parameters, the expected proportions in the cells, are of the form

$$\pi(i_1,\ldots,i_m)=\int\limits_{C(i_1,\ldots,i_m)} \mathrm{N}_{\Gamma(\theta)}(x)\,\mathrm{d}x, \tag{9}$$

where $C(i_1,\ldots,i_m)$ are m-dimensional rectangular parallelepipeds which partition m-space (some of them thus extend to plus or minus infinity). Moreover $\mathrm{N}_{\Gamma(\theta)}(x)$ is the density of a multivariate normal with means zero, variances equal to unity, and correlations $\Gamma(\theta)$. The observed proportions in the cells are $p(i_1,\ldots,i_m)$, they can be used in principle to estimate θ by maximum likelihood, minimum (modified) chi-square, and so on. This defines first-order efficient estimation for model (9): we want Fisher-consistent continuously differentiable functions of the observed proportions with minimal variance.

Observe that model (9) is distinctly Pearsonian in flavour. Variation in nature is 'truly' continuous and even 'truly' multinormal, unfortunately the imperfectness of our measuring instruments or our clumsy choice of scale prevents us from modelling this directly. Muthén (1983) models this by introducing a latent response variable which is categorized in order to produce the observed response variable, Muthén's contribution in this issue shows how the notion of a latent response variable can be integrated with existing analysis of covariance matrices literature to produce a general model. It is clear that this general model is also at the basis of the classical work on tetrachoric and polychoric correlations, started by Karl Pearson (1900), and followed up by many later psychometricians. We shall call model (9) the block-multinormal model, to distinguish it from other models for categorical variables that are derived from the multinormal distribution.

Relationships with other work reported in this issue becomes clear if we assume that in addition to (9) the two-factor model is true, i.e., $\gamma_{ij}(\theta)=\theta_i\theta_j$ for all $i\neq j$. We can now use a familiar reduction formula for the multivariate normal integral [Gupta (1963, pp. 799–801)] and reformulate (9) as

$$\pi(i_1,\ldots,i_m)=\int\limits_{-\infty}^{+\infty}\phi(z)\prod_{j=1}^{m}\{\Phi(\alpha_{i(j)}(z))-\Phi(\beta_{i(j)}(z))\}\,\mathrm{d}z. \tag{10}$$

In (10) we use ϕ and Φ for the univariate normal density and distribution. Moreover,

$$\alpha_{i(j)}(z)=(a_{i(j)}-\theta_j z)/(1-\theta_j^2)^{\frac{1}{2}},$$

and

$$\beta_{i(j)}(z)=(b_{i(j)}-\theta_j z)/(1-\theta_j^2)^{\frac{1}{2}},$$

with $b_{i(j)}$ the upper bound and $a_{i(j)}$ the lower bound of category i_j of variable j. This shows that (9), in the special case of a Spearman correlation matrix, is

equivalent to a multicategory probit model with a standard normal distribution on the latent trait. Several variations of this model have been discussed by Bartholomew in this issue [see also Bartholomew (1980)]. Compare also Samejima (1972), Muthén and Christofferson (1981), Bock and Aitkin (1981). If we replace Φ by Ψ, the cumulative logistic, and we replace ϕ by a general non-parametric density function, then (10) becomes the Rasch model for multicategory items, discussed by Andersen in this issue.

In the general model (9) we can distinguish two different cases. In the first one, the boundaries of the $C(i_1,\ldots,i_m)$ are known. This situation is closer to the analysis of quantified than to the analysis of ordered variables. The difference is that we take explicitly into account that the quantifications are not exact. In the social and behavioural sciences the model with known cell boundaries is of limited practical relevance, because it will often be too strong. It is a typical model for rounding errors. Nevertheless it is more realistic than the continuous multinormal model, which can be interpreted as a limiting case if we imbed the known-boundary model in a sequence of models with an increasingly fine partition of m-space. In the second special case, which is more interesting for our purposes, we do not suppose that the boundaries are known, only that their order is known. This makes the model considerably more flexible. It can deal not only with multinormal distributions, but also with distributions which can be made multinormal by separate monotone transformations on the marginals.

It is clear that first-order efficient estimation of the model (9) can be quite formidable from a computational point of view, and quite unwise from a statistical point of view. Computationally the use of (10) in the two-factor model is an enormous simplification, no multivariate integration is needed any more. Moreover, (10) can be combined with the EM-algorithm [Demptster et al. (1977)], as in Bock and Aitkin (1981), to compute full information maximum likelihood estimates. Eq. (10) can also be combined with the LeCam–Ferguson improvement method to compute modified minimum chi-square estimates by solving a single linear least squares problem. Again we need an initial consistent estimate which must be improved. Computing such an estimate is usually no big problem: we dichotomize the variables, compute tetrachoric correlations, and fit the Spearman model (or any other correlation model we are interested in) by familiar methods for the quantified case. The major problem for first-order efficient estimation, however, is the number of cells. In our small example there are already $3^6=729$ cells. Because there are only 119 observations, at least 610 cells will be empty. This makes it risky, in fact somewhat absurd, to rely on asymptotic theory which assumes that all cell entries are large. Fully efficient first-order information may not be practical or wise. We can, however, use a limited information method, which only uses the lower-order marginals. These lower-order marginal cells are obviously much better filled.

The argument for using limited information is also used by Muthén and Bartholomew in this issue.

We shall concentrate in this paper on procedures which use only the marginals up to the second order. Another clear influence of the multinormal distribution, of course. Variations of such procedures have been proposed by Muthén, Christofferson, Olsson, Bartholomew. References are in Muthén's paper in this issue or in Muthén and Christofferson (1981). Bartholomew (1980, and this issue) uses an approximate procedure, which was proposed earlier in the probit context by Lawley (1944). Muthén's methods are most easily explained if all variables are dichotomous. In that case he computes the tetrachoric correlations and their asymptotic dispersion matrix, this gives all the information that is needed for applying the general least squares criterion (1). For completeness we briefly repeat the definition of tetrachoric correlation: if both variables in a standardized bivariate normal distribution are dichotomized, then any estimate of the correlation parameter of the bivariate normal that is computed from the 2×2 table is called a tetrachoric correlation. As the transformation from all $\binom{m}{2}$ bivariate 2×2 tables to all m dichotomization points and all $\binom{m}{2}$ tetrachoric correlations is one-to-one, it follows that Muthén's method is efficient in the class of all Fisher-consistent continuously differentiable functions of the 2×2 tables. Muthén's method has similar desirable properties if some variables are binary and others are continuous. But as explained in our earlier sections we think that continuous variables are uninteresting, simply because they do not occur in practice. For polytomous variables (more than two categories) Muthén, Christofferson, and Olsson use a similar two-step approach: they first estimate polychoric correlations (similar definition as tetrachoric, but now the bivariate normal is polychotomized), and then use weighted least squares to estimate θ. In this issue Muthén seems to use an approximation to the weight matrix $\underline{v}_{ijkl}$ in (1) which involves (5). Moreover all polychoric correlations are estimated separately, one for each bivariate table, which causes loss of information because these tables are not independent.

We now outline a procedure which is efficient in the class of all Fisher-consistent estimates which are continuously differentiable functions of all bivariate marginals. In the case that all variables are dichotomous it is identical to Muthén's method, if the variables are polytomous it is better (in the first-order efficiency sense). More details are in De Leeuw (1983). The method can be thought of as an improved and efficient version of the classical procedure proposed, but not implemented, by Ritchie-Scott (1918). For notational convenience we suppose that all variables have the same number of categories, this number is k. In a bivariate table there are k^2-1 free parameters. We can parametrize by using the expected values of k^2-1 cells, but also by using the $2(k-1)$ cell boundaries and the $(k-1)^2$ tetrachoric correlation coefficients, one for each of the possible double dichotomies.

Estimated cell boundaries are computed by applying the inverse normal to the cumulative marginal proportions. Estimated tetrachoric correlations are computed for example as in Divgi (1979). Model (9) is equivalent to the hypothesis that for each bivariate distribution the $(k-1)^2$ tetrachoric correlations are equal, and that their common values have Spearman structure. The estimation has thus at least three steps. The first step computes $m(k-1)$ cell boundaries, and the $\binom{m}{2}(k-1)^2$ tetrachorics. For the tetrachorics we also compute their dispersion matrix, using the procedures in the appendix. In the second step we test the hypothesis that tetrachorics corresponding with the same bivariate table are equal by using weighted least squares. This also gives the polychoric correlation coefficients with their estimated dispersion. And in the third step we estimate the parameters θ by applying weighted least squares to the polychorics. If we use an improvement method in this last step, then of course the complete method consists of four steps. The chi-square is partitioned automatically: one part tests if the block-multinormal model is true, the second part tests if the parameteric specification of the correlations is true.

In our small and hopelessly inadequate example there are 4 tetrachoric correlations for each pair of variables, thus 60 tetrachorics in all. We illustrate two possible approaches. The first one estimates the 15 polychorics separately. They are given in table 4, the corresponding chi-squares, all of them with 3 degrees of freedom, are in table 5. The block-multinormal model fits rather poorly. The sum of the chi-squares is 244, but because the individual chi-squares are not independent this sum is not chi-square with 45 degrees of freedom. Because the block-multinormal model does not fit well, it does not make much sense to go on and estimate the factor loadings too. Nevertheless, purely for illustrative purposes, we have first estimated them by simple least squares in table 6. Estimates and standard deviations are in the first two columns of table 6. We have also improved the estimates to make them efficient in the class of all Fisher-consistent estimates which are differentiable functions of the polychorics in table 4. This gives the next two columns of table 6. The chi-square measuring the fit of the two-factor model, within the block-multinormal model (9), is 165, with 9 degrees of freedom. Again a poor fit. We compare this with a direct fit of the two-factor model on the 60 tetrachorics, using the transformation $\ln \gamma_{ij}(\theta) = \ln \theta_i + \ln \theta_j$. This skips a couple of steps compared with our three-step procedure, but it is essentially equivalent, the only difference being that we now have a model for the log-correlations which is linear. The loadings and estimated standard deviations are in the last two columns of table 6, the total chi-square is 427 with 54 degrees of freedom. All solutions in table 6 are Heywood cases, which means that they have loadings larger than one, which again illustrates the bad fit. This may be related to the fact that the correlation matrix in table 4 is not positive semidefinite, it has one negative eigenvalue of -0.06.

Table 4

Polychoric correlations, estimated separately.

AID	1.000					
LAW	0.477	1.000				
INC	0.644	0.516	1.000			
PAR	0.478	0.233	0.599	1.000		
TAX	0.826	0.750	0.742	0.800	1.000	
ARM	0.651	0.682	0.581	0.741	0.798	1.000

Table 5

Chi-squares for block-multinormal model.

AID	0.0					
LAW	8.2	0.0				
INC	4.6	9.7	0.0			
PAR	3.3	57.7	9.2	0.0		
TAX	39.0	10.9	3.7	14.2	0.0	
ARM	13.2	5.3	1.3	50.9	12.9	0.0

Table 6

Factor loadings for block-multinormal model.

	Separate polychorics				Direct	
	LS	SIG	WLS	SIG	WLS	SIG
AID	0.786	0.044	0.825	0.028	0.686	0.021
LAW	0.655	0.047	0.834	0.039	1.476	0.045
INC	0.750	0.043	1.014	0.026	0.778	0.030
PAR	0.716	0.032	0.770	0.029	0.766	0.026
TAX	1.035	0.034	0.907	0.028	0.910	0.021
ARM	0.846	0.044	0.870	0.041	1.022	0.039

The conclusion from the analysis of this example is, inevitably, that the block-multinormal model fits badly. The model may be of 'exceedingly doubtful validity' as Yule suggested in 1912 [compare Goodman (1981) or MacKenzie (1978) for a more extensive discussion]. On the other hand if the block-multinormal model is untrue, then it is very difficult to see how the continuous multinormal model could be true in the same situation. Of course the multinormal moment relations (5) can be true even if the multinormal model itself is not. We must also remember that the model has two

components, which both contribute to the fit. The first component says that all tetrachorics which can be computed for a pair of variables must be equal. The second component says that all tetrachorics must obey the Spearman conditions. The first component does not exist if all variables are binary, which is the case most commonly treated in the literature. For binary variables the fit is usually much better. But clearly for binary variables the assumption of block-normality becomes very arbitrary and contrived. This is what Yule had in mind. For binary variables the question is if the tetrachoric correlation is a valuable measure of association.

5. Linear analysis of unordered variables

The material in this section is closely related to recent papers of Goodman (1979, 1981a, b), of Fienberg (1982, and this issue), and to the work of Lancaster and his pupils and associates [summarized in Lancaster (1969)]. It is also similar in spirit to correspondence analysis (Deville and Saporta in this issue) and to the model implicit in fig. 1 of Keller and Wansbeek in this issue.

We have already seen in section 2 that if we say that a variable is unordered, then we merely mean that a possible order relation defined over the values it assumes is not used in the analysis or model. This is very useful in situations in which we are not sure about the order (think of occupational status and similar variables). The techniques we discuss in this section are still Pearsonian, in the sense that they compute a quantification of the variables, and thus assume implicitly that such a quantification makes sense. We can express this by saying that we assume the variables to be orderable or quantifiable. It is difficult to justify the techniques we discuss here for purely nominal variables (such as religion, state, nationality, sex) for which the idea of an order does not make much sense. For such variables the techniques of loglinear analysis seem more appropriate.

The model we discuss is related to the multiple correspondence analysis technique. There is an important difference between our approach and the one used by Heiser and Meulman or Deville and Saporta in this issue. They use correspondence analysis as a purely descriptive technique, without specification of a probabilistic or statistical model of any sort. Although we consider this to be perfectly legitimate, and in fact the only realistic way to proceed in most social and behavioural science situations, we also think that it is useful to discuss a statistical model in which correspondence analysis makes sense as a technique to compute parameter estimates. The particular model we use for this purpose has been discussed earlier by Gifi (1981) and De Leeuw (1982). It is important to remember, however, that our confirmatory approach to correspondence analysis is much more limited in scope and value than the technique itself. In the example used throughout

this paper, for instance, we 'need exploratory much more than confirmatory' [Tukey (1980)]. Or, to put it somewhat differently, only the exploratory approach is truly model free.

The model in this section will be explained for three discrete random variables with I, J and K categories. The expected values of the proportions are π_{ijk}, the marginals are π_{i**}, π_{*j*} and π_{**k}. Now suppose X is a square matrix of order I, with elements x_{is} $(i=1,\ldots,I;\ s=0,\ldots,I-1)$, which is orthonormal on the marginals of the first variable, i.e.,

$$\sum_{i=1}^{I} \pi_{i**} x_{is} x_{it} = \delta^{st}, \tag{11}$$

and moreover $x_{i0}=1$ for all i. Y is a similar matrix of order J, orthogonal on the marginals of the second variable, and Z is the same for the third variable. We can now use the elements of X, Y and Z to write π_{ijk} in the form

$$\pi_{ijk} = \pi_{i**}\pi_{*j*}\pi_{**k}\left(\sum_{s=0}^{I-1}\sum_{t=0}^{J-1}\sum_{u=0}^{K-1}\rho_{stu}x_{is}y_{jt}z_{ku}\right), \tag{12}$$

with

$$\rho_{stu} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\pi_{ijk}x_{is}y_{jt}z_{ku}. \tag{13}$$

This decomposition, which generalizes trivially to more than three variables, has been used a great deal by Lancaster (1969, ch. 12). It was also used, in the case of m binary variables, by Lazarsfeld (1961) and Bahadur (1961). It follows from our conventions that $\rho_{000}=1$, that $\rho_{00u}=0$ for $u>0$, $\rho_{0t0}=0$ for $t>0$, and $\rho_{s00}=0$ for $s>0$. Thus (12) can be written in more detail as

$$\begin{aligned}\pi_{ijk} = \pi_{i**}\pi_{*j*}\pi_{**k}\Bigg(1 &+ \sum_{s=1}^{I-1}\sum_{t=1}^{J-1}\rho_{st0}x_{is}y_{jt} + \sum_{s=1}^{I-1}\sum_{u=1}^{K-1}\rho_{s0u}x_{is}z_{ku} \\ &+ \sum_{t=1}^{J-1}\sum_{u=1}^{K-1}\rho_{0tu}y_{jt}z_{ku} + \sum_{s=1}^{I-1}\sum_{t=1}^{J-1}\sum_{u=1}^{K-1}\rho_{stu}x_{is}y_{jt}z_{ku}\Bigg).\end{aligned} \tag{14}$$

The model we propose assumes in the first place that $\rho_{stu}=0$ if $s>0$, $t>0$ and $u>0$. This is the additive definition of no second-order interaction, discussed by Darroch (1974), and Lancaster (1971, 1975a). In the second place it assumes that $\rho_{st0}=0$ if $s\neq t$, $\rho_{s0u}=0$ if $s\neq u$, and $\rho_{0tu}=0$ if $t\neq u$. This last assumption, which we call simultaneous diagonalisability, is again inspired by the multivariate normal distribution. If we decompose a continuous trivariate normal in the form (12)–(13), using the Hermite–

Chebyshev polynomials for X, Y and Z, then we have simultaneous diagonalisability. Lancaster's theory of Meixner classes (1975b) shows that it is true for many other multivariate distributions. It is of some interest to observe that the trivariate normal does not satisfy the assumption of vanishing second-order interaction.

With the two assumptions the decomposition (14) simplifies to

$$\dot{\pi}_{ijk} = \pi_{i**}\pi_{*j*}\pi_{**k}\left(1 + \sum_{s=1}^{I-1} \lambda_s x_{is} y_{js} + \sum_{s=1}^{I-1} \eta_s x_{is} z_{ks} + \sum_{t=1}^{J-1} \mu_t y_{jt} z_{kt}\right), \quad (15)$$

where we have assumed that $I \leqq J \leqq K$, and we have simplified the notation in an obvious way. There are two quite different interpretations of (15), which correspond with the boundaries known and boundaries unknown case in the previous section. The matrices X, Y and Z can be known, in which case the model says that they diagonalize the three bivariate distributions (and that there is no additive second-order interaction). They can also be unknown, in which case the model says that quantifications of the categories can be found which diagonalize the bivariate marginals (and that there is no additive second order interaction). The first model, with known quantifications, has been studied from a statistical point of view by O'Neill (1980). It does not seem to be very interesting for our purposes, for the same reason as the known boundary case was not very interesting in the previous section. Moreover, if we specify for example that the quantifications must be orthogonal polynomials on the marginals, which means essentially that they are known, it follows that we must use the order of the categories, i.e., the variables must be ordered. Thus we shall use the interpretation of (15) in which the quantifications are unknown. Observe that for the part of the model which specifies that the additive second-order interaction vanishes it is immaterial if the quantifications are known or unknown.

To make the connection with correspondence analysis more obvious we sum (15) over k and find

$$\pi_{ij*} = \pi_{i**}\pi_{*j*}\left(1 + \sum_{s=1}^{I-1} \lambda_2 x_{is} y_{js}\right). \quad (16a)$$

Eqs. (16b) and (16c) are not given, but they are of the same form, and are obtained by summing (15) over j and i. The model consisting of (16a)–(16b)–(16c) is weaker than (15), because it merely says that there are quantifications which diagonalize simultaneously, it does not say anything about the additive second order interaction. Actually (16a)–(16b)–(16c) may be more interesting than (15), for example because the trivariate normal satisfies (16a)–(16b)–(16c) but not (15). Taken separately (16a) is not restrictive: any bivariate table can be written in this form. This is the basic Fisher

decomposition of a contingency table [Maung (1942), Lancaster (1958)], on which correspondence analysis is based. It is closely related to the singular value decomposition, compare Heiser and Meulman and Deville and Saporta in this issue for details. In a similar way (16b) is not restrictive and (16c) is not, but taken together the three equations are restrictive because they specify that the three singular value decompositions are linked: instead of three matrices with left singular vectors and three matrices with right singular vectors we only have the three matrices X, Y and Z. That the bivariate distributions can be separately diagonalized is not restrictive, that they can be simultaneously diagonalized is.

Our proposal is to adopt (16) as a basic model for unordered categorical variables. In general this model consists of $\binom{m}{2}$ linked correspondence analyses, one for each pair of variables, each variable corresponds with a unique set of singular vectors. The algebra of the model is discussed in more detail by De Leeuw (1982), who also shows that techniques for multiple correspondence analysis and nonlinear principal component analysis produce consistent estimates of the parameters of model (16). Additional modelling can be done in terms of the singular values. To see this we rewrite (16) for m variables in matrix form

$$C_{jl} = D_j(uu' + \sum \lambda_{jls} x_{js} x_{ls}') D_l. \tag{17}$$

Here the C_{jl} are the bivariate marginals, the D_j are diagonal matrices with univariate marginals, u has all elements equal to one. The Spearman model can be formulated in this context as $\lambda_{jl1} = \theta_j \theta_l$. Or, in a stronger version, again inspired by the multinormal distribution, as $\lambda_{jls} = \theta_j^s \theta_l^s$, where superscript s stands for power s. Or $\lambda_{jls} = \theta_{js} \theta_{ls}$. Or $\lambda_{jl1} = \theta_j \theta_l$ and $\lambda_{jls} = 0$ for $s > 1$. There obviously is a great deal of useful flexibility here.

Actually fitting the model (17) is not difficult, and very similar to the procedure for fitting model (9) in the previous section. For the block-multinormal model (9) we used a one-to-one transformation of bivariate and univariate marginals to boundary points and tetrachorics. The model then transforms to equality of the tetrachorics corresponding with each pair, and Spearman structure of their common values. For model (17), which we could call the regression-multinormal model, we transform the marginals to singular vectors and singular values by performing all $\binom{m}{2}$ bivariate correspondence analyses. The model then says that the singular vectors should be equal if they correspond to the same variable, and that the singular values must have Spearman structure. Dispersion matrices for the various statistics are computed by using the general procedure in the appendix. Again the total chi-square can be partitioned in a component due to the inequality of the singular vectors and a component due to the lack of fit of the model for the correlations.

We have fitted the regression-multinormal model to our small example. The largest singular values (also called canonical correlations) are in table 7, the chi-square values indicating for each variable separately how well (17) fits are in table 8. The fit is quite satisfactory, certainly if we compare it with that of the block-multinormal model. We used table 7 again to estimate factor loadings by least squares. They are given, with their estimated standard deviations, in the first two colums of table 8. Efficient estimates of the factor loadings, using the log-transformation of the Spearman model as in the previous section, are given in the last two columns. The two-factor model now has chi-square 36, with 9 degrees of freedom. The Guttman determinacy of the least squares solution is 0.869, of the efficient solution it is 0.905. All in all the fit of the bilinear model (17) seems quite satisfactory, the fit of the two-factor model on the dominant singular values is more problematical. Of course analysis of a single example, certainly one as problematical as this, does not show that (17) is better than (9) in general. The results in this section, taken together with the algebraic and computational results of De Leeuw (1982), are encouraging however.

We have not indicated yet how (15) or (17) is related to the model discussed by Goodman, Fienberg, and Keller and Wansbeek, mentioned in the introduction to this section. In stead of (12) it uses the decomposition

$$\ln \pi_{ijk} = \sum_{s=0}^{I-1} \sum_{t=0}^{J-1} \sum_{u=0}^{K-1} \rho_{stu} x_{is} y_{jt} z_{ku}, \tag{18}$$

with

$$\rho_{stu} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{is} y_{jt} z_{ku} \pi_{i**} \pi_{*j*} \pi_{**k} \ln \pi_{ijk}. \tag{19}$$

This is the decomposition used in the ordinary loglinear model (Fienberg and Meyer in this issue). Now $\rho_{stu}=0$ for $s>0$ and $t>0$ and $u>0$ means that there is no multiplicative second-order interaction. Assuming that $\rho_{st0}=0$ if $s \neq t$, $\rho_{s0u}=0$ if $s \neq u$, and $\rho_{0tu}=0$ if $t \neq u$, gives the model

$$\pi_{ijk} = \sigma \alpha_i \beta_j \gamma_k \exp\left(\sum_{s=1}^{I-1} \lambda_s x_{is} y_{js} + \sum_{s=1}^{I-1} \eta_s x_{is} z_{ks} + \sum_{t=1}^{J-1} \mu_t y_{jt} z_{kt} \right). \tag{20}$$

If we assume in addition that $\lambda_s=0$ if $s>1$, $\eta_s=0$ if $s>1$, and $\mu_t=0$ if $t>1$, then (20) becomes

$$\pi_{ijk} = \sigma \alpha_i \beta_j \gamma_k \exp(\lambda x_i y_j + \eta x_i z_k + \mu y_j z_k). \tag{21}$$

This is the model suggested by Fienberg (1982) and by fig. 1 of Keller and Wansbeek in this issue. Using the analogy with their figure it could be called

Table 7

Maximal correlations.

AID	1.000					
LAW	0.448	1.000				
INC	0.503	0.440	1.000			
PAR	0.419	0.255	0.595	1.000		
TAX	0.593	0.618	0.570	0.479	1.000	
ARM	0.516	0.498	0.418	0.412	0.564	1.000

Table 8

Chi-squares for regression multinormal model.

AID	0.53
LAW	13.44
INC	2.23
PAR	0.09
TAX	2.01
ARM	0.63

Table 9

Factor loadings for regression multinormal model.

	LS	SIG	WLS	SIG
AID	0.718	0.061	0.738	0.059
LAW	0.643	0.073	0.634	0.055
INC	0.716	0.067	0.802	0.062
PAR	0.602	0.068	0.728	0.068
TAX	0.837	0.056	0.847	0.046
ARM	0.691	0.059	0.829	0.062

the point-multinormal model. It can be fitted in various ways, but we have not applied it to our example, and we have no definite algorithm yet. Two things are important. If (21) is true for the three-dimensional distribution, then it is not necessarily true that

$$\pi_{ij*} = \sigma\alpha_i\beta_j \exp(\lambda x_i y_j) \tag{22}$$

for the bivariate marginals. If we want to represent the continuous multinormal in the form (17) we need an infinite number of orthogonal

quantifications (the Hermite–Chebyshev polynomials) if we want to represent it in the form (21) we only need a single quantification. Other ways to compare the regression-multinormal and the point-multinormal model are suggested by the discussion in Darroch (1974) and Lancaster (1971, 1975). Of course the point-multinormal model can also be used with known or unknown quantifications, with known quantifications it is a technique for the analysis of quantified variables. A more detailed comparison of the three discrete multinormal models (point-multinormal, block-multinormal, and regression-multinormal) is called for. On the basis of the computational experience we have up till now, we think that the regression-multinormal model is more flexible and more likely to fit a fairly wide range of multivariate distributions.

Appendix

Suppose $\underline{x}$ is a discrete vector-valued random variable, assuming values x_v with probabilities π_v $(v=1,\ldots,N)$. The x_v are vectors with M elements. Moreover $\underline{x}_i$ are independent copies of $\underline{x}$ $(i=1,\ldots,n)$, $\underline{n}_v$ is the number of $\underline{x}_i$ equal to x_v, and $\underline{p}_v=\underline{n}_v/n$. Finally

$$\underline{C}=\sum_{v=1}^{N}\underline{p}_v x_v x_v', \tag{A.1}$$

and Γ is the expected value of $\underline{C}$. For the interpretation we must remember that in the case of quantified variables the x_v are simply the possible values of the multivariate variable. For discrete variables, ordered or unordered, the x_v are concatenated indicators or dummy variables. If there are m variables with k categories, then $M=mk$, each k-element subvector of x_v has one element equal to one and the others equal to zero. In this way $\underline{C}$ becomes the $mk\times mk$ table with all univariate and bivariate frequencies.

Now suppose ϕ_r are functions defined on the space of all positive semidefinite matrices of order M $(r=1,\ldots,t)$, and define the statistics $\underline{\alpha}_r=\phi_r(\underline{C})$. By a familiar result, given for example in Rao (1973, ch. 6),

$$n\operatorname{cov}(\underline{\alpha}_r,\underline{\alpha}_s)\rightarrow\sum_{v=1}^{N}\pi_v\frac{\partial\alpha_r}{\partial\pi_v}\frac{\partial\alpha_s}{\partial\pi_v}-\left(\sum_{v=1}^{N}\pi_v\frac{\partial\alpha_r}{\partial\pi_v}\right)\left(\sum_{v=1}^{N}\pi_v\frac{\partial\alpha_s}{\partial\pi_v}\right). \tag{A.2}$$

In (A.2) the partial derivatives of the α_r, considered as a function of the N-vector p, are evaluated at π. We assume that they exist and are continuous in a neighborhood of π. From (A.1) we find

$$\partial\alpha_r/\partial\pi_v=x_v' A_r x_v, \tag{A.3}$$

where $\Delta_r = \partial\phi_r/\partial C$, evaluated at $C = \Gamma$. Thus we can estimate the dispersion matrix of the α_r by first computing the $n \times t$ matrix $\underline{\Omega}$ with elements $\omega_{ir} = \underline{x}_i' \underline{\Delta}_r \underline{x}_i$, where $\underline{\Delta}_r$ is $\partial\phi_r/\partial C$ evaluated at $\underline{C}$, and by then computing the variance–covariance matrix of the ω_{ir}. In some case this can indeed be the most efficient or practical way to compute the required dispersion matrix, in some cases simplifications are possible, and in other cases computational tricks and short cuts are necessary. In any case the procedure can be used for all functions of second-order marginals with discrete multivariate data.

References

Andersen, E.B., 1983, Latent trait models, Journal of Econometrics, this issue.

Bahadur, R.R., 1961, A representation of the joint distribution of responses to n dichotomous items, in: H. Solomon, ed., Studies in item analysis and prediction (Stanford University Press, Stanford, CA).

Bartholomew, D.J., 1980, Factor analysis for categorical data, Journal of the Royal Statistical Society B 42, 293–321.

Bartholomew, D.J., 1983, Latent variable models for ordered categorical data, Journal of Econometrics, this issue.

Bentler, P.M., 1983, Simultaneous equation systems as moment structure models: With an introduction to latent variable models, Journal of Econometrics, this issue.

Berge, C., 1966, Espaces topologiques: Functions multivoques (Dunod, Paris).

Berkson, J., 1980, Minimum chi square, not maximum likelihood!, Annals of Statistics 8, 457–487.

Bock, R.D. and M. Aitkin, 1981, Marginal maximum likelihood estimation of items parameters: Application of an EM-algorithm, Psychometrika 46, 443–460.

Browne, M.W., 1974, Generalized least squares estimators in the analysis of covariance structures, South African Statistical Journal 8, 1–24.

Browne, M.W., 1977, The analysis of patterned correlation matrices by generalized least squares, British Journal of Mathematical and Statistical Psychology 30, 113–124.

Browne, M.W., 1982, Covariance structures, in: D.M. Hawkins, ed., Topics in applied multivariate analysis (Cambridge University Press, London).

Dantzig, G.B., J. Folkman and N. Shapiro, 1967, On the continuity of the minimum set of a continuous function, Journal of Mathematical Analysis and Applications 17, 519–549.

Darroch, J.N., 1974, Multiplicative and additive interaction in contingency tables, Biometrika 61, 207–214.

De Leeuw, J., 1982, Nonlinear principal component analysis, in: H. Caussinus, P. Ettinger and R. Tomassone, eds., COMPSTAT 1982 (Physica Verlag, Vienna).

De Leeuw, J., 1983, Discrete multinormal models (Department of Data Theory, University of Leyden, Leyden).

Dempster, A.P., N.M. Laird and D. B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society B 39, 1–38.

Deville, J.C. and G. Saporta, 1983, Correspondence analysis: Application to nominal time series, Journal of Econometrics, this issue.

Divgi, D.R., 1979, Calculation of tetrachoric correlation coefficient, Psychometrika 45, 169–172.

Ferguson, T.S., 1958, A method for generating best asymptotically normal estimates with application to the estimation of bacterial densities, Annals of Mathematical Statistics 29, 1046–1062.

Fienberg, S.E., 1982, Using information on ordering for loglinear model analysis of multidimensional contingency tables, Proceedings of the XIth International Biometrics Conference (Toulouse) 133–139.

Fienberg, S.E. and M.M. Meyer, 1983, Loglinear models and categorical data analysis with psychometric and econometric applications, Journal of Econometrics, this issue.

Fisher, R.A., 1925, Theory of statistical estimation, Proceedings of the Cambridge Philosophical Society 22, 700–725.
Fisher, R.A., 1962, The simultaneous distribution of correlation coefficients, Sankhya A 24, 1–8.
Gifi, A., 1981, Nonlinear multivariate analysis (Department of Data Theory, University of Leyden, Leyden).
Goodman, L.A., 1979, Simple models for the analysis of cross classifications having ordered categories, Journal of the American Statistical Association 74, 537–552.
Goodman, L.A., 1981a, Association models and canonical correlations in the analysis of cross-classifications having ordered categories, Journal of the American Statistical Association 76, 320–334.
Goodman, L.A., 1981b, Association models and the bivariate normal for contingency tables with ordered categories, Biometrika 68, 347–355.
Gupta, S.S., 1963, Probability integrals of multivariate normal and multivariate t, Annals of Mathematical Statistics 34, 792–838.
Guttman, L., 1955, The determinacy of factor score matrices with implications for five other basic problems of common-factor theory, British Journal of Statistical Psychology 8, 65–81.
Heiser, W.J. and J. Meulman, 1983, Analyzing rectangular tables by joint and constrained multidimensional scaling, Journal of Econometrics, this issue.
Hemelrijk, J., 1966, Underlining random variables, Statistica Neerlandica 20, 1–8.
Hsu, P.L., 1949, The limiting distribution of functions of sample means and application to testing hypotheses, Proceedings of the First Berkeley Symposium in Mathematical Statistics and Probability, 359–402.
Jennrich, R.I., 1970, An asymptotic χ^2 test for the equality of two correlation matrices, Journal of the American Statistical Association 65, 904–912.
Jöreskog, K.G., 1978, Structural analysis of covariance and correlation matrices, Psychometrika 43, 443–477.
Keller, W.J. and T. Wansbeek, 1983, Multivariate methods for quantitative and qualitative data, Journal of Econometrics, this issue.
Lancaster, H.O., 1958, The structure of bivariate distributions, Annals of Mathematical Statistics 29, 719–736.
Lancaster, H.O., 1969, The chi-squared distribution (Wiley, New York).
Lancaster, H.O., 1971, The multiplicative definition of interaction, Australian Journal of Statistics 13, 36–44.
Lancaster, H.O., 1975a, The multiplicative definition of interaction: An addendum, Australian Journal of Statistics 17, 34–35.
Lancaster, H.O., 1975b, Joint probability distributions in the Meixner classes, Journal of the Royal Statistical Society B 37, 434–443.
Lawley, D.N., 1944, The factorial analysis of multiple item tests, Proceedings of the Royal Society of Edinburgh 62, 74–82.
Lazarsfeld, P.F., 1961, The algebra of dichotomous systems, in: H. Solomon ed., Studies in item analysis and prediction (Stanford University Press, Stanford, CA).
LeCam, L., 1956, On the asymptotic theory of estimation and testing hypotheses, Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability 5.1, 129–156.
LeCam, L., 1980, Contribution to the discussion of Berkson (1980), Annals of Statistics 8, 473–478.
MacKenzie, D., 1978, Statistical theory and social interests: A case study, Social Studies of Science 8, 35–83.
Maung, K., 1942, Measurement of association in a contingency table with special reference to the pigmentation of hair and eye colours of Scottish school children, Annals of Eugenics 11, 189–223.
McDonald, R.P., 1975, Testing pattern hypotheses for correlation matrices, Psychometrika 40, 253–255.
Mooijaart, A., 1976, Latent structure models (Department of Psychology, University of Leyden, Leyden).
Muthén, B., 1983, Latent variable structural equation modeling with categorical data, Journal of Econometrics, this issue.

Muthén, B. and A. Christoffersson, 1981, Simultaneous factor analysis of dichotomous variables in several groups, Psychometrika 46, 407–419.
Neyman, J., 1949, Contribution to the theory of the χ^2 test, Proceedings of the First Berkeley Symposium in Mathematical Statistics and Probability, 239–273.
Norton, B.J., 1978, Karl Pearson and statistics: The social origin of scientific innovation, Social Studies of Science 8, 3–34.
O'Neill, M.E., 1980, The distribution of higher-order interactions in contingency tables, Journal of the Royal Statistical Society B 42, 357–365.
Pearson, K., 1900, Mathematical contributions to the theory of evolution: VII. On the correlation of characters not quantitatively measurable, Philosophical Transactions of the Royal Society A 195, 1–47.
Ritchie-Scott, A., 1918, The correlation coefficient of a polychoric table, Biometrika 12, 93–133.
Rubin, D.B. and D.T. Thayer, 1982, EM algorithms for ML factor analysis, Psychometrika 47, 69–76.
Samejima, F., 1972, A general model for free-response data, Psychometrika 37, monograph suppl.
Spearman, C., 1904, General intelligence, objectively determined and measured, American Journal of Psychology 15, 201–293.
Spearman, C., 1927, The abilities of man (MacMillan, New York).
Steiger, J.H., 1980, Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results, Multivariate Behavioural Research 15, 335–352.
Tukey, J.W., 1980, We need both exploratory and confirmatory, American Statistician 34, 185–191.
Van Praag, B.M.S., 1981, Model free regression, Economics Letters 7, 139–144.
Van Praag, B.M.S., 1982, The population-sample decomposition with an application to minimum distance estimators (Center for Research in Public Economics, University of Leyden, Leyden).
Wijsman, R.A., 1959, On the theory of BAN estimates, Annals of Mathematical Statistics 30, 185–191.
Wijsman, R.A., 1960, Correction to 'On the theory of BAN estimates', Annals of Mathematical Statistics 31, 1268–1270.