BEYOND CORRESPONDENCE ANALYSIS

Conference Presentation

Jan de Leeuw
Interdivisional Program in Statistics
UCLA

405 Hilgard Avenue
Los Angeles, CA 90024-1555
phone 310-206-8635 fax 310-206-5658
deleeuw@laplace.sscnet.ucla.edu

Version 1.0: September 1992

The basic motivation for this research that there is a wide gap in MVA between the multinomimial and the multinormal. There is discrete numerical variables, ordinal variables, non-normal numerical variables. in applied work, in the social, behavioural and life sciences, discrete numerical and ordinal variables seem to be the rule rather than the exception. Yet most MVA techniques are designed for either purely normal or purely nominal, which seems either too strong (bias) or too weak (precision).

This paper extends and summarizes:

Nonlinear Principal Component Analysis. In Caussinus et al. (eds.), COMPSTAT 1982, Wien, Physika Verlag.

Multivariate Analysis with Linearizable Regressions. Psychometrika, 53, 1988, 437-454.

Multivariate Analysis with Optimal Scaling. In Das Gupta and Sethuraman (eds), Progress in Multivariate Analysis, Calcutta, ISI, 1990.

In 1906 Pearson published a paper in *Biometrika* on the influence of scale order on correlation, in the case of *two characters which for some arrangement give a linear regression line.* What he basically proved is the following. Assign scores to the rows and columns of a cross table C, with marginals in the diagonal matrices D and E. Suppose the scores x and y are in deviations from the mean, with unit variance. Thus $r(x, y) = x'Cy$.

Now perturb the scores, again with vectors in deviations from the mean δ_x and δ_y. Then

$$\lim_{\epsilon \rightarrow 0} \frac{r(x + \epsilon\delta_x, y + \epsilon\delta_y) - r(x, y)}{\epsilon} =$$
$$= \delta_x'(Cy - r(x, y)Dx) + \delta_y'(C'x - r(x, y)Dy).$$

If both regressions are linear, then the rhs is zero, i.e. if both regressions are linear the correlation coefficient is relatively insensitive to details of scoring.

A version of the argument for more general rv's is in De Leeuw, On the prehistory of correspondence analysis, Statistica Neerlandica, 1984. We want to extend this to $m > 2$ variables as well.

In 1935 Hirschfeld (aka Hartley) published *A connection between correlation and contingency* in *PCPS* in which he proved (quite explicitly) the following.

Suppose we want to find scores that linearize the regressions in an $R \times C$ cross table. Thus we want

$$\begin{aligned} Cy &= \rho Dx, \\ C'x &= \rho Ey. \end{aligned}$$

This system always has $\min(R-1, C-1)$ non-trivial solutions, given by the generalized singular values and vectors of the triple (C, D, E). The vectors of scores are mutually orthogonal, etc. Moreover (generalizing Pearson) these scores give maxima, saddle points, minima of the correlation coefficient.

This was generalized to some extent by Fisher and Maung around 1940, and by Lancaster et al. since 1955, to general bivariate distributions. Also see Buja, Annals of Statistics, 1990. Again, we would like to look at $m > 2$.

It is clear that for $m > 2$ variables things are not so simple any more. In general not all bivariate regressions (let alone all multivariate regressions) can be linearized by scoring (or, if you prefer, transformation).

Let us call a multivariate distribution *bi-linearizable* if all bivariate regressions can be linearized by scoring. In obvious notation there exist m different vectors $\{y_1, \cdots, y_m\}$ such that

$$C_{j\ell}y_\ell = \lambda_{j\ell}D_j y_j.$$

Bi-linearizable distributions are

1) all variables are binary,
2) there are only two variables,
3) special cases, such as the multinormal (or elliptical),
4) the *strained multinormal* in the sense of Yule.

We see that assuming that the multivariate distribution is bi-linearizable is an important generalization from assuming it to be multinormal (or elliptical).

Actually, strained multinormals may not be very well known, so we describe them a bit more in detail.

Suppose $\underline{x} = (\underline{x}_1, \cdots, \underline{x}_m)$ is multivariate normal. Now suppose ϕ_j are strictly increasing, and define $\underline{y} = (\phi(\underline{x}_1), \cdots, \phi(\underline{x}_m))$. Then $\underline{y}$ is strained multinormal. Obviously we can unstrain $\underline{y}$ by applying the inverse transformations ϕ_j^{-1}.

This creates a fairly general family of multivariate distributions. In a strained multinormal we can linearize all regressions (not only the bivariate ones) by unstraining. Thus strained multinormal is stronger than bi-linearizable.

There is actually something in between strained normality and bi-linearizability: suppose orthonormal systems $\{y_{j1}, \cdots, y_{jp}\}$ exists such that

$$C_{j\ell}y_{\ell s} = \lambda_{j\ell s}D_j y_{js}.$$

This could be called bi-linearizable of order p. The multivariate normal is bi-linearizable of all orders.

It becomes interesting now to extend our results to a more general situation.

We give the formalism. Suppose we have m random variables $\underline{x}_j$ on the same probability space. We define $\mathcal{L}_j$ as the subspace of all measurable transformations of $\underline{x}_j$ with zero mean and finite variance. Suppose $\mathcal{K}_j$ is a p-dimensional subspace of $\mathcal{L}_j$, and suppose $\{\underline{y}_{j1}, \cdots, \underline{y}_{jp}\}$ is an orthonormal basis for $\mathcal{K}_j$. Just for ease of notation we suppose all subspaces have the same dimension p, but this is not essential.

We write $C_{j\ell}$ for the cross products of the bases. Thus $(C_{j\ell})_{st} = \mathbf{E}\,(\underline{y}_{js}\underline{y}_{\ell t})$. Also D_j is the diagonal matrix of variances, i.e. $D_j = C_{jj}$. The covariance between any two transformations of the form

$$\underline{z}_j = \sum_{s=1}^{p} \alpha_{js}\underline{y}_{js}$$

is simply

$$\mathbf{C}\,(\underline{z}_j, \underline{z}_\ell) = \alpha_j' C_{j\ell}\alpha_\ell.$$

Let us first study the bi-linearizable case. Immediately we are stuck with a number of questions about the linearizing transformations.

1) If they exist, how do we find them ? (estimation)
2) Do they exist ? (fit)
3) What do they do to the standard errors ? (precision)
4) How do they look ? Are they useful ? (data analysis)

We start with cross tables $C_{j\ell}$ and univariate marginals in diagonal matrices D_j. For standardized scores the correlations are

$$\rho_{j\ell} = \alpha_j C_{j\ell} \alpha_\ell,$$

and the correlation-ratios are

$$\eta^2_{j\ell} = \alpha_j C_{j\ell} D_\ell^{-1} C_{\ell j} \alpha_j.$$

Obviously

$$\rho^2_{j\ell} \leq \eta^2_{j\ell},$$

with equality if and only if the regression of variable ℓ on variable j is linear.

Now take any function Φ of the correlation coefficients and correlation ratios, and maximize it over the scores α_j. The stationary equations are

$$\sum_{\ell\neq j}^{m}\frac{\partial\Phi}{\partial\rho_{j\ell}}C_{j\ell}\alpha_\ell+\sum_{\ell\neq j}^{m}\frac{\partial\Phi}{\partial\eta_{j\ell}^2}C_{j\ell}D_\ell^{-1}C_{lj}\alpha_j=\lambda_j D_j\alpha_j.$$

The λ_j are Lagrange multipliers, taking care of the normalization of the scores. If the scores α_j linearize the bivariate regressions, then they solve the stationary equations with

$$\lambda_j=\sum_{\ell\neq j}^{m}\frac{\partial\Phi}{\partial\rho_{j\ell}}\rho_{j\ell}+\sum_{\ell\neq j}^{m}\frac{\partial\Phi}{\partial\eta_{j\ell}^2}\rho_{j\ell}^2.$$

But they are already many programs which maximize functions of the form Φ, such as multiple correspondence analysis, ACE, etc. We prefer to maximize

$$\Phi(x_1,\cdots,x_m)=\sum_{j=1}^{m}\sum_{\ell=1}^{m}(\eta_{jl}^2-\rho_{jl}^2),$$

because that seems most direct. From the point of view of consistency the choice does not matter: all give the same solution.

Suppose the α_j linearize the regressions. Complete the scores to matrices $A_j = (\,\alpha_j \quad | \quad \overline{A}_j\,)$, such that $A_j' D_j A_j = I$. Then

$$A_j' C_{j\ell} A_\ell = \begin{pmatrix} \rho_{jl} & 0 \\ 0 & \overline{A}_j' C_{j\ell} \overline{A}_\ell \end{pmatrix}.$$

Solving the equations gives us the parametric model (for the joint bivariate marginals)

$$D_j^{-1} C_{j\ell} D_\ell^{-1} = A_j \begin{pmatrix} \rho_{j\ell} & 0 \\ 0 & \Gamma_{jl} \end{pmatrix} A_\ell'.$$

This can be fitting by WLS directly to the bivariate marginals. For the bi-linearizability of order p we can strengthen the model to

$$D_j^{-1} C_{j\ell} D_\ell^{-1} = A_j \Delta_{j\ell} A_\ell',$$

with $\Delta_{j\ell}$ a diagonal matrix. For strained multinormality we combine this with no higher-order interactions, and we can even use likelihood methods.

In fact, a useful statistical procedure seems to be the two-step technique. First scale the variables by linearizing the regressions. Then apply standard techniques to the induced correlation coefficients. But what about the asymptotic normal distribution of these induced correlations ?

The nice result, again generalizing Pearson (1906), is that for linearizable distributions the asymptotic normal distribution is the same as the one we would derive if the scores had been known (fixed, not dependent on the data). Thus, if p is the distribution,

$$\alpha_j'(p)C_{j\ell}(p)\alpha_\ell(p) \overset{a.d.}{=} \alpha_j'C_{j\ell}(p)\alpha_\ell,$$

and we know since Isserlis (Biometrika, 1917) how to compute the asymptotic distribution of correlation coefficients.