

Principal Profiles

Trevor Hastie
AT&T Bell Laboratories
Murray Hill
New Jersey, 07974
and
Francesca Little
Department of Business Science
University of Cape Town
Rondebosch
7700 South Africa

Abstract

A variety of multivariate techniques have been modified to analyze categorical data represented as relative frequencies summing to one. Correspondence analysis is an appropriately weighted principal component analysis, and is useful for uncovering low dimensional structure in the compositional profiles. However, certain forms of "linear" structure are forced to appear non-linear by the constraints of the simplex, the so called "arch effect". This paper presents a new method for modelling compositional data. We finesse the arch effect by modelling the logit transform of the profiles, the natural parameter in the multinomial family. We discuss the formulation of the principal profiles model, its estimation by maximum likelihood, and its performance on a real data example.

Keywords: Compositional data, Correspondence Analysis, Scaling, Logistic regression, Arch effect.

AUTHORS' FOOTNOTE

Trevor Hastie is a Member of the Technical Staff at AT& T Bell Laboratories, Murray Hill, New Jersey 07974. Francesca Little is Lecturer, Department of Business Science, University of Cape Town. This work was partially completed while both authors were employed by the Institute for Biostatistics, South African Medical Research Council. The authors thank Michael Greenacre, Daryl Pregibon, Les Underhill and Walter Zuchini for valuable comments.

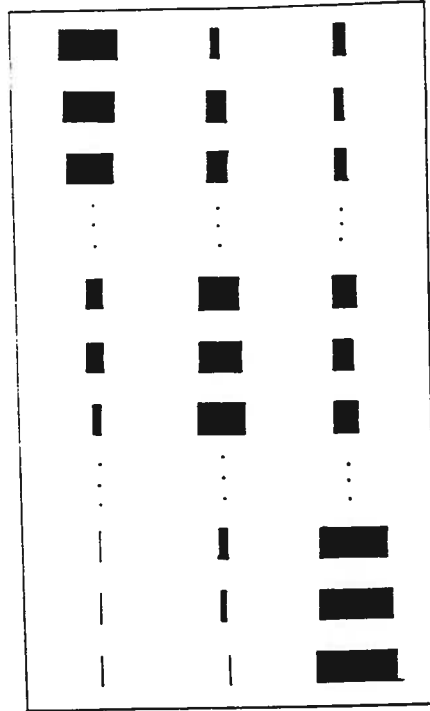


Figure 1 A shaded contingency table representation of the generated one dimensional data. The mass shifts “linearly” from the left corner to the right.

1. Introduction.

In this paper we present a new method for analyzing compositional data, where the observations are vectors which add to one. Data of this kind arise naturally if the response variable is categorical, or if there are several response variables and their relative values are of interest. Compositional data is often characterized by inherent curvature (Aitchison, 1983) as a result of the summation constraint. The following constructed example illustrates this curvature.

Consider the “linear” model,

$$\begin{aligned}
 \eta_1(t) &= -t/2 - 2 \\
 \eta_2(t) &= -t/2 \\
 p_1(t) &= \frac{\exp(\eta_1(t))}{1 + \exp(\eta_1(t))} \\
 p_1(t) + p_2(t) &= \frac{\exp(\eta_2(t))}{1 + \exp(\eta_2(t))} \\
 p_3(t) &= 1 - p_1(t) - p_2(t)
 \end{aligned} \tag{1}$$

for $t \in \mathbb{R}^1$.

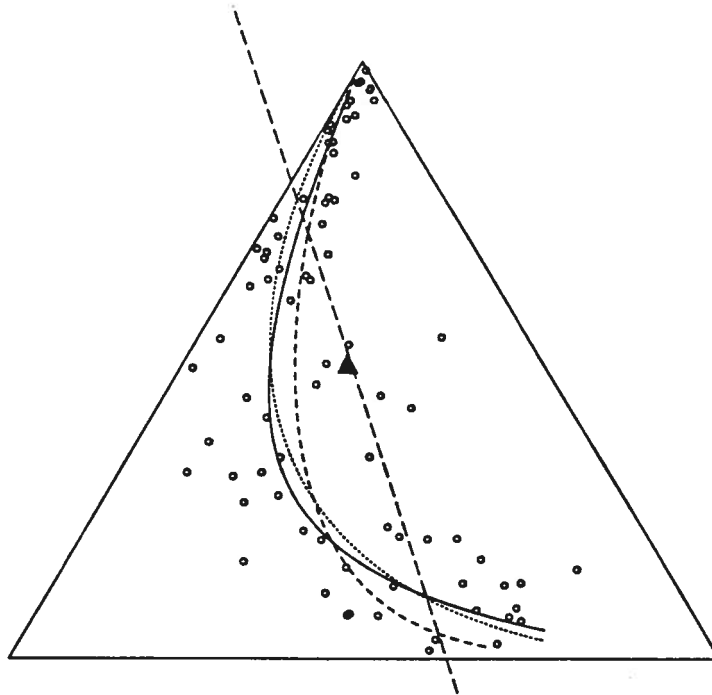


Figure 2 Artificially generated three dimensional profiles in the simplex. The solid curve is the one dimensional family of probability profiles (1). The points represent scaled multinomial realizations from profiles chosen at random from the curve. The straight line is the first correspondence analysis axis, passing through the mean profile (triangle). The dotted curve is the principal profile estimate, the dashed curve is the Aitchison estimate.

At $t = 0$, the profiles $(p_1(0), p_2(0), p_3(0))$ represent a categorization of the logistic distribution on the real line by cutting the line at -2 and 0 . Thus $p_1(0) = p(Z \leq -2)$ where Z has the standard logistic distribution. The distribution shifts linearly with t , but the cutpoints remain fixed. (This is the proportional odds model of McCullagh, 1980). Figure 1 demonstrates the diagonal nature of the model depicted as a contingency table for some given values of t . Figure 2 represents the one dimensional model $(p_1(t), p_2(t), p_3(t))$ (the solid line) and illustrates the curved nature of the data. For each of 100 values of t selected randomly from $N(0, 1)$, we randomly generated a multinomial *profile* (x_1, x_2, x_3) from the (scaled) multinomial distribution $M_3(20, p_1(t), p_2(t), p_3(t))/20$. The appropriate sample space for compositional data is the positive simplex; figure 2 represents our multinomial profile vectors in the unit simplex. Each corner of the simplex represents an extreme profile where all the mass is in one category. If the profiles can be ordered such that the mass shifts smoothly from one category to the next (as ours does), then its representation in the simplex will be curved.

Since compositional observations are at least two dimensional, it is often of interest to find lower dimensional summaries in the spirit of principal component or factor analysis. In real data examples the profiles might, for example, be percentages of a species of birds in a sample, artefact types in a ruin, or opinion categories in a poll. The purpose of the analysis is to find an ordination of the profiles over space, time or with respect to some factor to be determined from the analysis. Correspondence Analysis (C.A.) is a popular technique for doing just this (see e.g. Greenacre, 1984). C.A. does principal component analysis on the sample proportions, but uses a metric more appropriate than the standard Euclidean metric. The first axis determined by a C.A. of our generated data is included in Figure 2, (the straight broken line). Clearly, due to its linearity, it does not provide a good representation of the inherent one-dimensional nature of the data.

In the C.A. and multidimensional scaling literature this curvature is known as the “horse-shoe” or “arch” effect. Typically a two or higher dimensional C.A. model is required to capture this inherent one dimensional curvature in the data. Although a “horse-shoe” plot may give the correct one dimensional ordering of both observations and response categories, it presents a distortion (apparent shortening) of distances at the end points of the first axis. In addition, a horse-shoe plot that is almost a complete circle makes the derivation of a meaningful one-dimensional ordination very difficult. Hill and Gauch (1980) developed “Detrended Correspondence Analysis”, which is intended to eliminate the arch effect. Their method works better in situations where the curve is well defined by the data, but has problems with noisy data or outliers. In addition, Greenacre (1984) points out that in the detrending process, “control over the geometry is lost and it is possible that, just as the non-linearities can mask less important gradients, so the detrending might introduce further artifacts into the results.”

Aitchison (1983) approaches the curvature problem by transforming the compositional data. He uses the logistic transformation applied to the observed counts:

$$\begin{aligned} y_1 &= \log\left(\frac{x_1}{x_{d+1}}\right) \\ &\vdots \\ y_d &= \log\left(\frac{x_d}{x_{d+1}}\right) \end{aligned} \tag{2}$$

or in his notation

$$\mathbf{y}^{(d)} = \log(\mathbf{x}^{(d)}/x_{d+1}) \tag{3}$$

He then performs ordinary euclidean principal component analysis on the transformed data, and finds that this adequately accounts for the curvature. The Aitchison method, however, is unable to satisfactorily cope with zero proportions (see Aitchison 1982, and discussion). The somewhat ad hoc

solutions suggested by Aitchison are not very satisfactory since, as noted by Anderson (1982), the fit of the transformed normal model varies drastically with the treatment of the zeros. The analysis of data sets with a moderate number of zero components thus still remains a problem.

We develop a simple modification of the Aitchison method suggested by the logistic regression model. A simple version of the logistic regression model has a binomial response proportion R , and the true proportion P is modelled via the linear logistic regression model $\log(P/(1-P)) = a + bx$ where x is a covariate of interest. Initial attempts at estimation were centered on the (weighted) linear regression of $\log(Y/(1-Y))$ on x (see Cox 1970). The same problem of extreme proportions prompted the now accepted maximum binomial likelihood estimation for this model. In the same spirit, we propose the linear structural model:

$$\log(\mathbf{p}_i^{(d)}/p_{i,d+1}) = \alpha_0 + \alpha_1 t_i \quad (4)$$

where α_0 and α_1 are d -vectors, and $t_i, i = 1, \dots, n$, are incidental parameters, one for each profile.

Equation (4) has the factor analysis or structural model flavor, where the t_i are the instances or uniquenesses of the underlying one dimensional factor. We regard the data as multinomial realizations from (4). The left-hand side of (4) is the natural parameter vector for the multinomial family, and determines the structure of both the mean and the variance. Not all the parameters in (4) are identifiable, and just as in factor analysis, some standardization is required. We restrict α_1 to be of unit length, and define α_0 to be the point on the curve closest to the mean profile. Unlike factor analysis of unrestricted data, we do not wish to restrict the curve to contain the mean. Figure 2 clearly illustrates why not. We outline below a maximum likelihood scheme to estimate the parameters in (4). Our model thus sidesteps the extreme count problem just as the linear logistic model does for zero-one data. We will see, however, that for certain configurations of extreme counts, the maximum likelihood estimates of the incidental parameters t_i might not exist.

Figure 2 shows the results obtained by applying the maximum likelihood method to our generated data. The broken line in Figure 2 shows the curve fitted using Aitchison's approach; the curve estimated by maximum likelihood is drawn as a dotted line. The maximum likelihood curve certainly seems closer to the generating curve and data points than the Aitchison curve (note that the generating model for the fake data is not identical to our model). Its superior fit is confirmed by a comparison of deviances ($-2 \times \log$ -likelihood): 97,2 for the maximum likelihood curve, compared to 161,9 for Aitchison's curve. On the logistic scale it is apparent that the Aitchison curve is more influenced by the outlying "edge" points (Figure 3). These points cannot be represented exactly on the logit scale and so some truncation was needed to display them (discussed later).

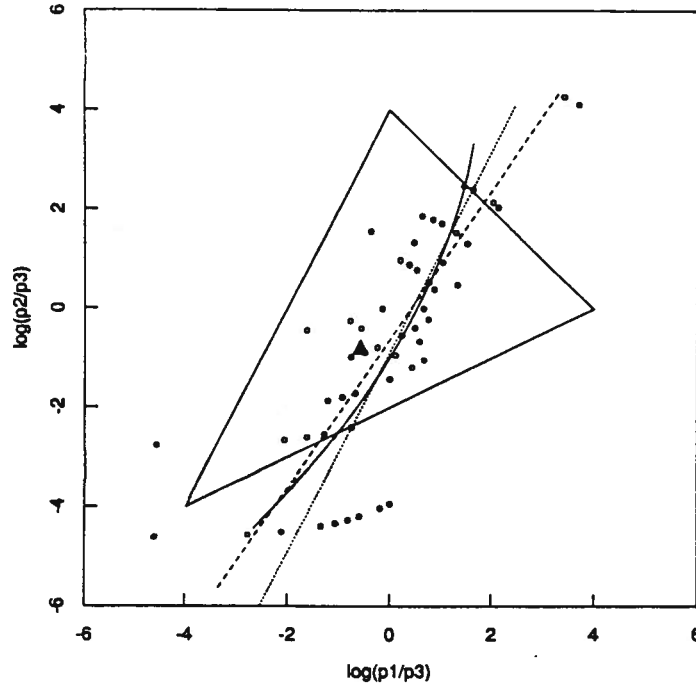


Figure 3 The various solutions in figure 2 displayed on the logit scale. The solid curve is the generating curve, the dotted curve is the principal profile estimate, and the dashed curve is the Aitchison estimate. The triangular outline represents a shrunken version of the simplex.

2. Estimation by maximum likelihood.

Let $X_{n \times (d+1)}$ be the matrix of observed counts for n observations, with $\mathbf{x}_i = (x_{i1}, \dots, x_{i,d+1})'$ the vector of counts for observation i . Let $c_i = \sum_{j=1}^{d+1} x_{ij}$. Denote the $n \times (d+1)$ matrix of observed profiles by R , where $\mathbf{r}_i = \mathbf{x}_i/c_i$ and let \mathbf{P} be the matrix of associated model probabilities, with i th row \mathbf{p}_i . The logit linear structural model (4) for the probabilities can be written:

$$\begin{aligned} \log\left(\frac{p_{i1}}{p_{i,d+1}}\right) &= \alpha_{01} + \alpha_{11}t_i \\ &\vdots \\ \log\left(\frac{p_{id}}{p_{i,d+1}}\right) &= \alpha_{0d} + \alpha_{1d}t_i \end{aligned}$$

For estimation purposes we assume \mathbf{x}_i has a multinomial distribution $M_{d+1}(c_i\mathbf{p}_i, c_i\boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \text{diag}(\mathbf{p}_i) - \mathbf{p}_i\mathbf{p}_i'$. We then propose to estimate $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ and t by maximizing the multinomial

likelihood

$$l(\alpha, t) = \sum_{i=1}^n \sum_{j=1}^{d+1} x_{ij} \log p_{ij} \quad (5)$$

where

$$p_{ij} = \frac{\exp(\eta_{ij})}{1 + \sum_{j=1}^d \exp(\eta_{ij})}$$

and

$$\eta_{ij} = \alpha_{0j} + \alpha_{1j} t_i.$$

Instead of attempting to maximize the likelihood over all the $n + 2d$ parameters, our algorithm alternates between estimating α and t .

Firstly, suppose that t is given and we wish to determine α_0 and α_1 . This is the usual multinomial regression model (see eg McCullagh and Nelder, 1983). It is not hard to show that the $2d$ score vector for α is

$$s_\alpha = \sum_{i=1}^n \begin{pmatrix} c_i(r_i^{(d)} - p_i^{(d)}) \\ c_i t_i (r_i^{(d)} - p_i^{(d)}) \end{pmatrix}. \quad (6)$$

Similarly, the Fisher information or Hessian is

$$\mathcal{I}_\alpha = - \sum_{i=1}^n \begin{pmatrix} c_i \Sigma_i^{(d)} & c_i t_i \Sigma_i^{(d)} \\ c_i t_i \Sigma_i^{(d)} & c_i t_i^2 \Sigma_i^{(d)} \end{pmatrix}. \quad (7)$$

where $\Sigma_i^{(d)}$ is $d \times d$ and represents the appropriate partition of Σ_i . The one-step Fisher scoring update, given an initial value α^0 , is then

$$\alpha^1 = \alpha^0 - \mathcal{I}_\alpha^{-1} s_\alpha. \quad (8)$$

Alternatively, suppose α is known. It is clear that the MLE of t_i involves only l_i . The one step Fisher scoring update for estimating t_i is given by

$$t_i^* = t_i - \frac{s_{t_i}}{i_{t_i}}, \quad (9)$$

where

$$s_{t_i} = c_i \alpha_1' (r_i^{(d)} - p_i^{(d)}), \quad (10)$$

$$\text{and } i_{t_i} = -c_i \alpha_1' \Sigma_i \alpha_1. \quad (11)$$

At this point several strategies are possible. One could iterate the α -step till convergence before proceeding to the t -step, and vice-versa. Empirically, we found it sufficient to alternate between one

step of each. Our algorithm for estimating α and t is thus:

Initialization: Use the Aitchison method to obtain initial parameter estimates. Hence, let

$$y_i = \log \left(\frac{x_i^{(d)} + c}{x_{i,d+1} + c} \right),$$

$z_i = y_i - \bar{y}$, and obtain the singular value decomposition of Z : $Z = UD_\lambda V^t$. Then $\alpha_0^0 = \bar{y}$, $\alpha_1^0 = v_1$, the first right-hand singular vector, and t^0 is the first column of UD_λ . The starting values thus define the largest principal component of Z .

Loop: over outer iteration counter m in

α - step: Keeping t fixed at t^{m-1} , let $\alpha^m = \alpha^{m-1} - \mathcal{I}_\alpha^{-1} s_\alpha$, where s_α and \mathcal{I}_α are given in (6) and (7) respectively. Normalize α_1 to have unit length.

t - step: Keeping α fixed at α^m , let $t_i^m = t_i^{m-1} - s_{t_i}/i_{t_i}$, where s_{t_i} and i_{t_i} are given by (10) and (11) respectively.

Let $\eta_i^m = \alpha_0^m + \alpha_1^m t_i^m$ and $p_{ij}^m = \exp(\eta_{ij}^m)/(1 + \sum_{k=1}^d \exp(\eta_{ik}^m))$. Compute the deviance

$$D^m = -2 \sum_{i=1}^n \sum_{j=1}^{d+1} x_{ij} \log(r_{ij}/p_{ij}^m),$$

Until: the deviance D^m fails to decrease.

At the converged α^m , use the t -step repeatedly to find the value $t_{\bar{r}}$ where the sample average profile \bar{r} projects. Finally let $t_i^* = t_i^m - t_{\bar{r}}$, $\alpha_1^* = \alpha_1^m$ and $\alpha_0^* = \alpha_0^m + \alpha_1^* t_{\bar{r}}$.

3. Discussion of the algorithm.

It is not necessary to assume the multinomial distribution for estimation. The algorithm could equally well have been derived on the basis of quasi-likelihood (McCullagh and Nelder 1983), which simply makes assumptions about the mean-covariance relationship. This relationship in turn reflects the summation constraint in the data. Likewise, the deviance is to first order the Mahalanobis distance from the points to the curve, measured in the metric of the covariance matrix.

The largest euclidean principal component of a data matrix Z is the solution of the minimization problem

$$\sum_{i=1}^n \|z_i - \alpha_0 - \alpha_1 t_i\|^2 = \min! \quad (12)$$

One can show that alternating between

- (1) for given t , performing d separate regressions of z_{ij} onto t_i , $i = 1, \dots, d$, to obtain α_{0j} and α_{1j} , and
- (2) for given α , projecting z_i onto the line $\alpha_0 + \alpha_1 t$ to obtain t_i ,

results in the usual eigenvector solution to (12). This is the Euclidean version of our alternating algorithm, and corresponds to the power method for finding the largest eigenvector of the covariance matrix derived from Z . Of course this procedure is not necessary, since $\alpha_0 = \bar{z}$ is optimal, and we can solve the projection step explicitly leaving

$$\sum_{i=1}^n \|z_i - \bar{z} - \alpha_1 \alpha_1' (z_i - \bar{z})\|^2 \quad (13)$$

to minimize over unit α_1 .

We now show that each pair of steps in the loop of our algorithm can be written in a form similar to (12). Let z_i denote the "adjusted dependent vector" (of dimension d),

$$z_i = \alpha_0^{m-1} + \alpha_1^{m-1} t_i^{m-1} + V_i^{-1} (r_i^{(d)} - p_i^{(d)}) \quad (14)$$

where $V_i = (\Sigma_i^m)^{(d)}$. Then it can be shown that the Fisher scoring update step for estimating α , given in (6) and (7), is equivalent to minimizing the weighted least squares regression problem

$$\begin{aligned} \chi_m^2 &= \sum_{i=1}^n (z_i - \alpha_0^m - \alpha_1^m t_i^m)' W_i (z_i - \alpha_0^m - \alpha_1^m t_i^m) \\ &= \sum_{i=1}^n \|z_i - \alpha_0^m - \alpha_1^m t_i^m\|_{W_i}^2 \end{aligned} \quad (15)$$

over α , where $W_i = c_i V_i$. As in logistic regression, z_i is the first order Taylor approximation of $\log(r_i^{(d)}/r_{i,d+1})$ about the current value for p_i .

The α -step of the power method applied to (12) required d separate regressions of z_{ij} onto t_i . For multinomial data however, the non-diagonal form of W_i requires the d regressions to be performed simultaneously.

Similarly, the t -step corresponds to projecting $z_i - \alpha_0^m$ onto α_1^m with metric W_i :

$$t_i^m = \frac{(z_i - \alpha_0^m)' W_i \alpha_1^m}{\alpha_1^m' W_i \alpha_1^m}. \quad (16)$$

It can then be shown that these form a pair of power method steps towards solving the eigenvalue

problem

$$\sum_{i=1}^n \left\| z_i - \alpha_0^m - \alpha_1^m \frac{\alpha_1^{m'} W_i (z_i - \alpha_0^m)}{\alpha_1^{m'} W_i \alpha_1^m} \right\|_{W_i}^2 = \min! \quad (17)$$

which is the weighted version of (13). Apart from drawing analogies to the euclidean principal components, this derivation does not help us much, since we know of no simpler way to solve (17). What does become apparant is that whereas the maximum-likelihood analysis performs a type of weighted principal component analysis on the logits, the Aitchison method performs an unweighted principal component analysis on the empirical logits. Since the weight matrices have $p_j(1-p_j)$ on their diagonals, extreme points (with large logits) will receive less weight in the former than in the latter. This phenomenon was evident in figure 2. Both Aitchison's method and detrended correspondence analysis suffer from outliers, whereas our method affords some protection through the weighting device. One could consider a more direct approach to outliers as in "m-estimation" in which large residuals are trimmed; we have not explored this avenue.

The Aitchison method is asymmetric, in that it depends on which element goes into the denominator (although in Aitchison (1983) an alternative formula is developed). Our method as stated is symmetric, since the estimation criterion is symmetric in the p 's, and not the transformation of them. Thus for any solution t , α_0 , α_1 , with say p_{d+1} in the denominator, we can transform to α_0^* and α_1^* with say p_1 in the denominator. The projections t^* will change appropriately, but the fitted p 's will be the same.

Before we can discuss convergence properties of the algorithm, we need to establish the existence (and uniqueness) of a minimum to (5). It is not hard to construct examples for which, given α , the optimal t for an extreme point (with at least one zero proportion) is infinite. This results in the point being fitted exactly, and the contribution to the deviance is zero. The constraints on α_1 , however, guarantee its existence, and if all the elements of \bar{r} are positive, we conjecture the same is true for α_0 . We have still to establish stronger existence properties for the model. In light of this it seems innappropriate to discuss convergence. In practise we set a threshold on $|t_i|$ (say 100), and in all the examples tried so far (over 10), the algorithm has converged in a few iterations. Points that would project with $|t_i|$ over 100 have a low weight and thus contribute a small amount to the information matrix.

Since each step of the alternating algorithm is a Newton-Raphson step, they are in the ascent direction. With step length control, the likelihood will increase after each iteration. The log-likelihood is bounded above by zero, and thus must converge. At this point let t^* and α^* denote the estimates. If $\partial l / \partial \alpha^* = 0$ and $\partial l / \partial t^* = 0$ (which is the case if $|t_i^*| < 100$) then the solution is a local maximum.

4. Higher dimensional models.

Extension to two or more dimensional models is straightforward, and we briefly sketch the idea here. The two-dimensional model has the form

$$\log p_i^{(d)} / p_{i,d+1} = \alpha_0 + \alpha_1 t_i + \alpha_2 s_i, \quad (18)$$

where $\{\alpha_1, \alpha_2\}$ describe a two dimensional factor plane. We have extended our alternating one dimensional algorithm in a natural way to estimate the additional parameters; the details are omitted. The score and information for α have dimension $3d$, and $\{t, s\}$ have $2n$ parameters. In addition α_1 and α_2 are standardized to be orthonormal.

5. Displaying the solutions.

The one dimensional solution consists of a set of scores for each profile, as well as the mean and factor vectors. The scores provide the ordering that is often the goal of such an analysis. Similarly the two dimensional model gives a pair of scores for each profile. A scatterplot of these pairs give a low dimensional representation of the data as in factor analysis.

In correspondence analysis it is customary to plot the projections of the extreme profiles (with all the mass in one category) together with the data. These extreme profiles can be considered as most representative of their respective columns, and allow for interpretation of the projected data. In factor analysis, this corresponds to examination of the factor loadings in order to give the factors names (Greenacre and Hastie 1987). For a given principal profile solution, the "t" step can be used repeatedly to project the extreme profiles (corners of the simplex) onto the solution axes.

In practice projecting extreme profiles can result in extreme factor scores, as discussed earlier. Our objective, however, in projecting them is to facilitate interpretation of the axes. The general direction in which these points lie, rather than their actual positions, is of importance. We thus suggest shrinking the corner points towards the mean profile and then finding projections for these "shrunk corner points". Typically the amount of shrinking is chosen so that they can be plotted jointly with the projections of the observed profiles.

6. Example: Questionnaire data.

Each row of table 1 gives the response profiles representing six possible answers to a question assessing the American public's attitude to recent and future business conditions, expressed as frequencies per thousand.

Table 1: Questionnaire Data

Question: Considering the country as a whole, do you think we will have good times or bad times or what during the next twelve months?

The table gives the results of a survey expressed as frequencies per thousand (Converse 1980)

Year	C1	C2	C3	C4	C5	C6
1961	121	71	87	205	341	175
1962	59	38	65	165	479	193
1963	53	36	57	144	517	193
1964	72	29	52	150	575	123
1965	49	24	33	18	639	137
1966	82	20	22	78	674	123
1967	116	48	49	114	519	154
1968	133	47	65	120	449	186
1969	111	23	55	109	528	174
1970	316	72	82	137	235	158
1971	248	53	80	145	255	219
1972	217	50	91	128	318	195
1973	298	69	91	81	280	181
1974	592	84	80	67	66	111
1975	324	88	26	169	184	109
1976	227	44	15	123	353	138

C1: Bad

C2: Bad (qualified)

C3: Pro-Con

C4: Good (qualified)

C5: Good

C6: Don't know/depends.

These are given for the sixteen - year period from 1961 to 1976 (Table 1; Converse 1980, table 7.1).

Figure 4 gives the two dimensional correspondence analysis solution for the rows of table 1. The one dimensional principal profile solution (circled years) is superimposed on the plot. It follows the arched representation of the projected years. In addition, the positions of the correspondence analysis corner points (column markers) indicate that the one dimensional solution provides an ordering from extreme pessimism on the left to optimism on the right. One can then simply read off the trends from this graph: an optimistic upswing from 61 to 66, followed by a a period of uncertainty, until eventually

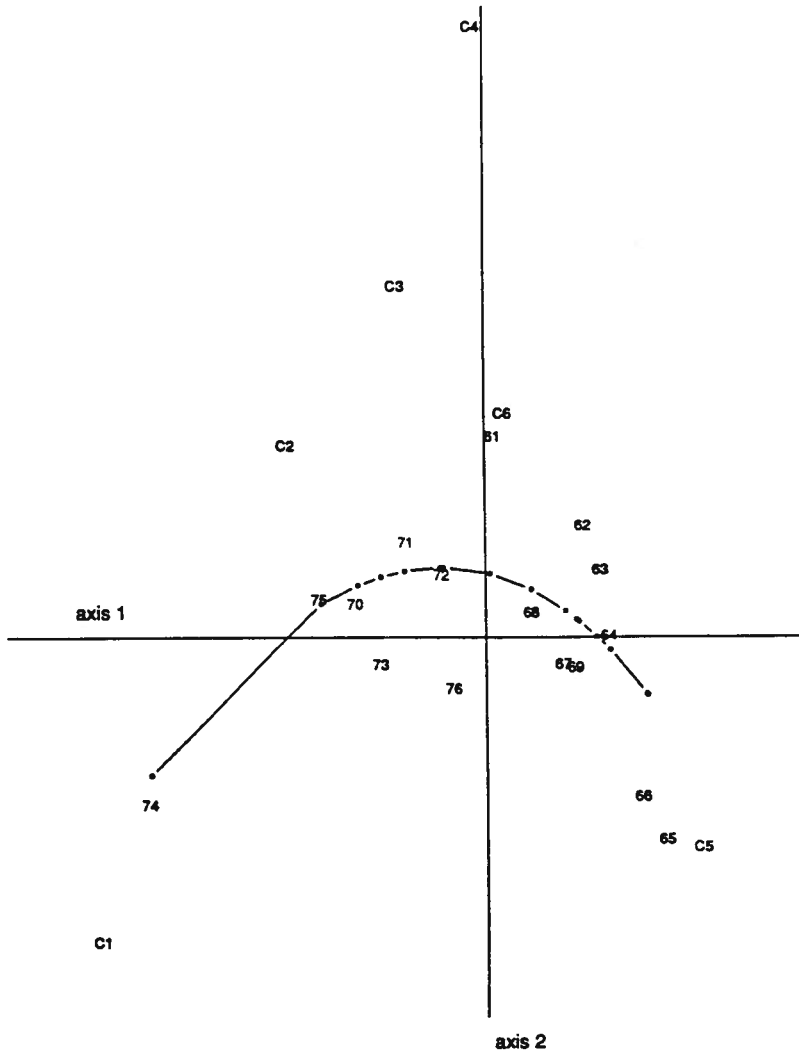


Figure 4a Correspondence analysis of the rows of the questionnaire data. The C_i labels refer to the projected extreme profiles or corners of the simplex. The curve represents the one dimensional principal profile solution. Figure 4b identifies the points on the curve

a period of pessimism in the mid 70's.

7. Discussion.

Principal profiles give a low dimensional approximation to high-dimensional compositional data, explicitly accounting for the inherent curvature in such data.

Our model is a simple modification of the Aitchison logit model, where we model the underlying "true" profiles instead of transforming the data. A priori our aim was to avoid zero count problems. The model is estimated by maximum likelihood, via an iterative algorithm similar to the "power"

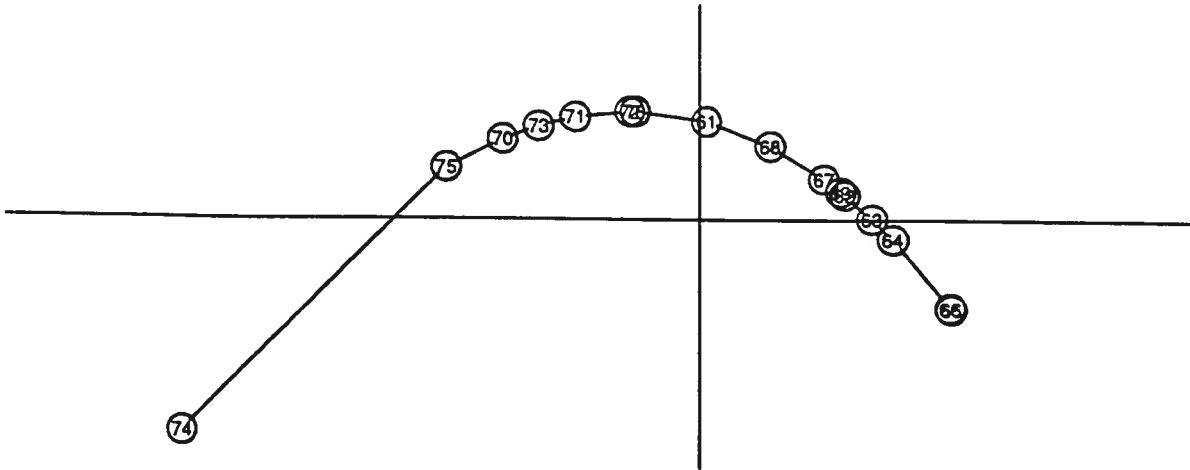


Figure 4b The curve in figure 4a with the projected years identified. The projections of 65 and 66 nearly coincide, as do 62 and 69 and 72 and 76.

method for finding eigenvectors of a symmetric matrix. An added bonus, and a natural consequence of this approach is that our algorithm uses weights in a sensible way; it penalizes extreme profiles which are high variance points on the logit scale.

The method has worked successfully on a number of data sets. We have implemented the algorithm in both one and two dimensions, and have had no convergence problems. In all cases, the model has been successful in accounting for the curvature present in correspondence analysis solutions.

Our formulation (4) of the principal profile model suggests an obvious generalization:

$$\log(p_i^{(d)}/p_{i,d+1}) = f(t_i) \quad (19)$$

where f is a vector of one dimensional functions. This model describes a general one dimensional curve in the simplex. In practise one could use polynomial functions or more adventurously estimate the $f_j(t_i)$ non-parametrically. The logit transformation guarantees that the estimates are probability profiles. Models of this type have been developed to generalize regular principal components(eg Carrol 1969 used polynomials, Hastie 1984 used non-parametric smooth functions). Our alternating algorithm is equally amenable to such generalizations: the multiple regression step (“ α step”) gets replaced by a polynomial or non-parametric regression, and the “ t step” gets replaced by a more general search routine (see Hastie 1984 for details).

This paper reports ongoing research. We have a method that works in practice; we need, however, to study its characteristics in more detail. These include conditions for existence and uniqueness of

solutions to the likelihood problem, convergence of the algorithm, and stability of the solutions.

Acknowledgements

This work was partially completed while both authors were employed by the Institute for Biostatistics, South African Medical Research Council. The authors thank Michael Greenacre, Daryl Pregibon, Les Underhill and Walter Zuchini for valuable comments.

References

- Aitchison, J. (1982), "The Statistical Analysis of Compositional Data (with discussion)." *Journal of the Royal Statistical Society, B* 44, No. 2, 139-177
- Aitchison, J. (1983), "Principle Component Analysis of Compositional Data." *Biometrika* 70, 1, 57-65
- Anderson, J. A. (1982), Discussion of "The Statistical Analysis of Compositional Data." *Journal of the Royal Statistical Society, B* 44, No. 2, 161-163.
- Carrol, D.J. (1969), "Polynomial Factor Analysis", *Proceedings of the 77th Annual Convention, APA*, 103-104.
- Converse, P.E. (1980), *American Social Attitudes Data Source Book, 1947-1978*, Harvard Academic Press.
- Cox, D. R. (1970) *The Analysis of Binary Data*, Methuen, London.
- Greenacre, M. J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press Inc. (London).
- Greenacre, M. J. and Hastie, T. J. (1987) "The Geometric Interpretation of Correspondence Analysis," *Journal of the American Statistical Association*, 82, in press.
- Hastie, T. J. (1984), *Principal Curves and Surfaces*, unpublished Ph.D thesis, Stanford University, California.
- Hill, M. O. and Gaugh, H. G., Jr. (1980), "An Improved Ordination Technique," *Vegetatio*, 42, 47-58
- Kendall, D. G. (1971), "Seriation". In *Mathematics in the Archaeological and Historical Sciences* (F. R. Hodson, D. G. Kendall and P. Tauto, eds.), University Press, Edinburgh.
- McCullagh, P. (1980), "Regression Models for Ordinal Data (with discussion)," *Journal of the Royal Statistical Society, B*, 42, 109-142
- McCullagh, P. and Nelder, J. (1983), *Generalized Linear Models*, Chapman Hall, London.