MULTIVARIATE ANALYSIS OF ORDINAL DATA USING PROBITS

JAN DE LEEUW

ABSTRACT. We discuss algorithms to fit component and factor representations to multivariate multicategory ordinal data, using a probit threshold model. Fixed parameters to quantify objects or individuals are used to avoid high dimensional integration. Algorithms alternate Newton's method to fit thresholds and quadratic majorization to fit structural parameters. The basic

1. INTRODUCTION

Suppose we have measurements of n objects on each of $m \geq 1$ ordinal variables. Variable j has k_j ordered categories. Such data are very commonly collected in experiments involving rating scales, attitude scales, or Likert scales. In the special case in which $k_j = 2$ for all j we have binary variables, which are ubiquitous in test research, survey research, ecology, bioassay, archeology, political science and so on.

The classical multivariate analysis approach to data of this type was initiated by Pearson [1900]. He assumes that we were observing a rounded version of a multivariate normal random vector, and the probability of observing an observation in a particular cell is the integral of the multivariate normal over the cell. The theory was initially developed for two binary variables (tetrachoric correlation) and extended around 1920 to two general categorical variables (polychoric correlation). Multivariate data were

---

Date: April 1, 2006.

2000 *Mathematics Subject Classification.* 62-04,62P20,62P25.

Key words and phrases. Probit models, choice models, item reponse theory, majorization algorithms.

generally analyzed by estimating correlations separately from all bivariate tables, because a complete multivariate approach remains impractical.

The Pearsonian approach became and remains very popular in psychometrics and econometrics, because it allowed researchers to continue to work in the familiar multivariate normal framework and to freely formulate the structural models in terms of the unobserved normal variables. In econometrics [Maddala, 1983] and sociometrics [Long, 1987] choice-based probit regression modelling probits is much studied. In psychometrics multivariate probit models have been used in test theory ever since Lawley [1944]. Also in psychometrics the Fechner and Thurstone traditions in psychophysics largely relied on probit models, which were later systematized in bioassay by Finney [1947]. Probit regression was imbedded in the GLM framework by McCullagh and Nelder [1989, Chapter 4].

The major problem with a satisfactory treatment of the multivariate case is that we have to to evaluate high-dimensional integrals, and, except for some special cases, this leads to very demanding computational problems. Thus in the last 100 years a cottage industry has developed producing various shortcuts and approximations to circumvent the problems associated with high-dimensional integration.

In this paper we go in another direction, using incidental parameters for objects. Thus we change unobserved realizations of random variables to parameters that must be estimated. This avoids high-dimensional integration. From the statistical point of view estimating incidental parameters tends to introduce bias in the estimates of the structural parameters, but this argument only applies if we accept the usual repeated independent trials framework of multivariate analysis. Instead, we suggest that it is useful to think of multivariate analysis techniques as matrix approximation techniques, in which the question of bias does not arise because the probability model is just used as a heuristic way to define the distance between the observed and fitted matrix.

We use *majorization algorithms* [De Leeuw, 1994; Heiser, 1995; Lange et al., 2000] to minimize distances between observed and predicted. This leads to convenient computations and can handle high-dimensional problems. For binary data, in which $k_j = 2$ for all j, these techniques were discussed earlier by De Leeuw [2006].

2. Loss Function

The observations are collected in an $n \times m$ matrix of integers R, with $1 \leq r_{ij} \leq k_j$. Each row of R is an observed *profile*, the number of possible profiles is $k_1 \times \cdots \times k_m$. The $y_{ij\ell}$ are the observations, coded as *indicators*. Thus $y_{ij\ell} = 1$ if object i is in category ℓ of variable j, i.e. if $r_{ij} = \ell$, and $y_{ij\ell} = 0$ otherwise. In other words $y_{ij\ell}$ is either zero or one and satisfies

$$\sum_{\ell=1}^{k_j} y_{ij\ell} = 1$$

for all $i = 1, \cdots, n$ and $j = 1, \cdots, m$. We also suppose each object has a frequency f_i associated with it, so that objects with the same *profile* do not have to be presented as separate rows in the data matrix.

Because probit models are formulated in terms of normal distributions, it is convenient to introduce some shorthand notation. Let

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

and for $\phi_{0,1}(x)$ we simply write $\phi(x)$. In the same way

$$\Phi_{\mu\sigma}(z) = \int_{-\infty}^{z} \phi_{\mu\sigma}(x)dx$$

and $\Phi_{0,1}(z)$ is simply $\Phi(z)$. Thus

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}),$$
$$\Phi_{\mu,\sigma}(z) = \Phi(\sigma z + \mu).$$

The general problem we study in this paper is minimize the *deviance*, which is minus two times the logarithm of the probit likelihood. Thus we want to minimize a loss function of the form

(1)
$$\mathcal{D}(\tau,\theta) = -2\sum_{i=1}^{n} f_i \sum_{j=1}^{m}\sum_{\ell=1}^{k_j} y_{ij\ell}\log\left[\Phi(\tau_{j,\ell}+\eta_{ij}(\theta)) - \Phi(\tau_{j,\ell-1}+\eta_{ij}(\theta))\right],$$

where Φ is the cumulative standard normal and the τ_j are sequences of thresholds satisfying

$$-\infty = \tau_{j0} < \tau_{j1} < \cdots < \tau_{j,k_j-1} < \tau_{j,k_j} = +\infty,$$

The $\tau_{j\ell}$ can be either known, partially known, or completely unknown, depending on the application. In particular we distinguish the two cases $\tau_j \in \mathcal{K}_j$, the cone of strictly increasing vector, and $\tau_j \in \mathcal{L}_j$, a single ray starting at the origin (i.e. the set of all vectors which are a positive multiple of a given vector).

The η_{ij} are real valued functions, which we call *combination rules*. They are defined for each value of the parameter vector θ. They usually combine parameters quantifying the objects with parameters quantifying the variables.

3. Special Cases

We give some examples of different specifications of the combination rule η_{ij}.

3.1. Discrete Normal.

Suppose $x_1, \cdots, x_n$ are realizations of a one-dimensional random variable[1] $\underline{x}$ with a *discrete normal distribution* with parameters μ and σ. This means

$$\mathbf{prob}(\underline{x} = \ell) = \Phi_{\mu,\sigma}(\tau_\ell) - \Phi_{\mu,\sigma}(\tau_{\ell-1}).$$

In many cases, because of the inherent discreteness of actual observations, this is a more realistic representation than the usual continuous normal one.

[1] Observe that we follow the convention of underlining random varibles [Hemelrijk, 1966].

The deviance is

$$\mathcal{D}(\mu,\sigma,\tau) = -2\sum_{i=1}^{n} f_i \sum_{\ell=1}^{k} y_{i\ell}\log\left[\Phi(\sigma\tau_\ell+\mu)-\Phi(\sigma\tau_{\ell-1}+\mu)\right] =$$

$$= -2\sum_{\ell=1}^{k} n_\ell \log\left[\Phi(\sigma\tau_\ell+\mu)-\Phi(\sigma\tau_{\ell-1}+\mu)\right]$$

Here the n_ℓ are the frequencies in each of the intervals, i.e. $n_\ell = \sum_{=1}^{n} f_i y_{i\ell}$. In this discrete normal case η_{ij} does not depend on either i or j, in fact $m = 1$ and thus j does not vary at all.

In this example the thresholds are assumed known up tp a scale factor, i.e. restricted to be in $\mathcal{L}_j$. If they were unknown, i.e. only required to be in $\mathcal{K}_j$, then we could always attain perfect fit by setting $\mu = 0$, $\sigma = 1$ and

$$\tau_\ell = \Phi^{-1}(\frac{n_1+\cdots+n_\ell}{n}).$$

3.2. **Discrete Normal Regression.** Now suppose x_i is a realization of $\underline{x}_i$, which is discrete normal with mean $\eta_i(\theta)$ and variance σ^2. The deviance becomes

$$\mathcal{D}(\tau,\theta,\sigma) = -2\sum_{i=1}^{n} f_i \sum_{\ell=1}^{k} y_{i\ell}\log\left[\Phi(\sigma\tau_\ell+\eta_i(\theta))-\Phi(\sigma\tau_{\ell-1}+\eta_i(\theta))\right].$$

In the most common case, of course, $\eta_i(\theta) = x_i'\theta$, and we have the discrete linear regression model. Observe that in this case we can fit models in which either τ is known, which means we still to estimate σ, or τ is unknown, in which case we can absorb σ into τ.

In the binary case we have $\tau_0 = -\infty$, $\tau_2 = +\infty$, and only τ_1 must be estimated. We use τ for $\sigma\tau_1$ and z_i for $1 - 2y_{i2}$. The deviance is simply

$$\mathcal{D}(\tau,\theta) = -2\sum_{i=1}^{n} f_i \log\Phi(z_i(\tau+\eta_i(\theta))).$$

3.3. **Probit Item Analysis.** We can combine the probit model with thresholds with an additive combination rule, in which a parameter for the object is added to a parameter for the variable. This corresponds with the Rasch model for item analysis [Fischer and Molenaar, 1995].

$$\mathcal{D}(\tau, a) = -2\sum_{i=1}^{n} f_i \sum_{j=1}^{m}\sum_{\ell=1}^{k_j} y_{ij\ell}\log\left[\Phi(\tau_{j\ell}+a_i)-\Phi(\tau_{j,\ell-1}+a_i)\right].$$

In the binary case this simplifies as before to

$$\mathcal{D}(\tau, a) = -2\sum_{i=1}^{n} f_i \sum_{j=1}^{m}\log\Phi(z_{ij}(\tau_j+a_i)).$$

3.4. **Probit Principal Components.** In De Leeuw [2006] probit principal component analysis for binary data was discussed, and applied to roll-call data for the US House and Senate. It is straightforward to generalize the loss function to the case in which we have more than two response categories. It becomes

$$\mathcal{D}(\tau, A, B) = -2\sum_{i=1}^{n} f_i \sum_{j=1}^{m}\sum_{\ell=1}^{k_j} y_{ij\ell}\log\left[\Phi(\tau_{j\ell}+a_i'b_j)-\Phi(\tau_{j,\ell-1}+a_i'b_j)\right].$$

3.5. **Probit Factor Analysis.** The loss function is the same as for principal component analysis, but now

$$A = \begin{bmatrix} U & V \end{bmatrix},$$

where U is $n\times p$, V is $n\times m$, and $U'U = I$, $V'V = I$, and $U'V = 0$. Moreover

$$B' = \begin{bmatrix} C \\ D \end{bmatrix}$$

where C is $p\times m$ and D is $m\times m$ and diagonal. Thus $AB' = UC + VD$ and we fit the same bilinear approximation as in principal component analysis, except that we now have more components than variables and we have the diagonality restrictions on D. For details about fitting factor analysis models in this form to quantitative data, see De Leeuw [2004].

4. Geometry

The interpretation in terms of the probit likelihood tries to imbed the techniques of this paper in the framework of classical statistics. This is relatively straightforward in the case of a single variables, for instance for the discrete normal and for probit regression, in which $\eta_{ij}(\theta)$ is independent if i. it becomes less natural and convincing if $m > 1$ and there are additional parameters for each profile. In that case the likelihood is best interpreted as a convenient distance measure between the observed table $Y = \{y_{ij\ell}\}$ and the manifold of predicted tables $\Pi(\tau, \theta) = \{\pi_{ij\ell}(\tau, \theta)\}$. Our method then computes the metric projection, in this likelihood metric, of Y on the manifold.

In this context it is interesting to investigate the cases in which we have *perfect fit*.

5. Polychoric Factor Analysis

Acoording to the classic Pearsonian polychoric model the probability of a profile $(\ell_1, \cdots, \ell_m)$ is

$$\pi(\ell_1, \cdots, \ell_m) = \pi^{-\frac{1}{2}m}|\Sigma|^{-\frac{1}{2}} \int_{\tau_{1\ell_1}}^{\tau_{1,\ell_1-1}} \cdots \int_{\tau_{1\ell_m}}^{\tau_{1,\ell_m-1}} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}dx$$

6. Known Thresholds

In this section we develop the majorization algorithm for general combination rules in the case in which the thresholds are known. The majorization function is based on the following basic lemma.

Lemma 6.1. *Suppose* $-\infty \leq \alpha < \beta \leq +\infty$ *are fixed and define*

$$f(x) = -\log\left[\Phi(\beta + x) - \Phi(\alpha + x)\right].$$

Then $0 < f''(x) < 1$.

Proof. By direct calculation

$$f'(x) = -\frac{\phi(\beta + x) - \phi(\alpha + x)}{\Phi(\beta + x) - \Phi(\alpha + x)}$$

and

$$f''(x) = \frac{(\beta + x)\phi(\beta + x) - (\alpha + x)\phi(\alpha + x)}{\Phi(\beta + x) - \Phi(\alpha + x)} + \left[\frac{\phi(\beta + x) - \phi(\alpha + x)}{\Phi(\beta + x) - \Phi(\alpha + x)}\right]^2.$$

Now

$$\int_{\alpha+x}^{\beta+x} x\phi(x)dx = -(\phi(\beta+x)-\phi(\alpha+x)),$$

and

$$\int_{\alpha+x}^{\beta+x} x^2\phi(x)dx =$$
$$(\Phi(\beta+x)-\Phi(\alpha+x)) - [(\beta+x)\phi(\beta+x)-(\alpha+x)\phi(\alpha+x)]$$

Thus if $\underline{x}$ is a doubly truncated standard normal random variable $\underline{x}$, truncated on the right at $\beta+x$ and on the left at $\alpha+x$, we see that $\mathbf{E}(\underline{x}) = f'(x)$ and

$$\mathbf{E}(\underline{x}^2) = 1 - \frac{(\beta+x)\phi(\beta+x)-(\alpha+x)\phi(\alpha+x)}{\Phi(\beta+x)-\Phi(\alpha+x)}.$$

Thus $\mathbf{V}(\underline{x}) = 1 - f''(x)$. But this immediately implies the bounds given in the theorem. □

Lemma 6.1 generalizes the well-known result that f is strictly convex [Burridge, 1981; Pratt, 1981]. Strict convexity provides the lower bound, but not the upper one, which is what we need for majorization. By letting α or β tend to infinity, we obtain the previous bounding result [Böhning, 1999; De Leeuw, 2006] that $f(x) = -\log\Phi(x)$ satisfies $0 < f''(x) < 1$. This special case is basically all we need in the case of binary variables.

We now use Lemma 6.1 to construct a majorization function for the deviance. First define

$$g_{ij\ell}(\tau,\theta) = -\frac{\phi(\tau_{j\ell}+\eta_{ij}(\theta))-\phi(\tau_{j,\ell-1}+\eta_{ij}(\theta))}{\Phi(\tau_{j\ell}+\eta_{ij}(\theta))-\Phi(\tau_{j,\ell-1}+\eta_{ij}(\theta))},$$

and

$$h_{ij}(\tau,\theta) = \sum_{\ell=1}^{k_j} y_{ij\ell}\, g_{ij\ell}(\tau,\theta).$$

Theorem 6.2. *For all pairs θ and $\tilde\theta$*

$$\mathcal{D}(\tau,\theta) \leq \mathcal{D}(\tau,\tilde\theta)+\sum_{i=1}^n f_i \sum_{j=1}^m h_{ij}(\tau,\tilde\theta)(\eta_{ij}(\theta)-\eta_i(\tilde\theta))+\frac{1}{2}\sum_{i=1}^n f_i\sum_{j=1}^m(\eta_{ij}(\theta)-\eta_{ij}(\tilde\theta))^2$$

Proof. Apply Lemma 6.1 to the second-order Taylor expansion of the deviance. □

Now define the *target function* by

$$\gamma_{ij}(\tau,\theta) = \eta_{ij}(\theta) - h_{ij}(\tau,\theta).$$

Then the majorization algorithm corresponding with Theorem 6.2 is

$$\theta^{(k+1)} \in \mathop{\mathbf{argmin}}_{\theta} \sum_{i=1}^{n} f_i \sum_{j=1}^{m} (\eta_{ij}(\theta) - \gamma_{ij}(\tau,\theta^{(k)}))^2. \tag{2}$$

Of course the power of the majorization algorithm comes from the fact that we know how to solve the least squares problems (2) quickly and efficiently for the discrete normal, for probit regression, and for principal component and factor analysis.

7. Unknown Thresholds

If the thresholds are unknown, or partially known, they must be fitted as well. We use *block relaxation* [De Leeuw, 1994] in the obvious way. We start with initial estimates of the thresholds, then we find the optimal combination rules for these thresholds, then we find the optimal thresholds for these combination rules, and so on. Thus we alternate our majorization algorithm from the previous section with an algorithm that computes optimum thresholds. For this last algorithm we need another lemma.

Lemma 7.1. *Suppose x is fixed and define*

$$h(\alpha,\beta) = -\log\left[\Phi(\beta + x) - \Phi(\alpha + x)\right].$$

Then for all $\alpha < \beta$ we have $0 \prec \mathcal{D}^2 h(\alpha,\beta) \prec \infty$.

Proof. This is actually a consequence of the more general results of Pratt [1981] and Burridge [1981]. They prove strict convexity for all cumulative distribution functions with a concave log-density, and not just for the cumulative normal. □

Observe we cannot obtain an upper bound result in this case. In fact the second derivatives are unbounded if α and β tend to the same value. Thus we cannot use the same type of quadratic majorization we used previously for the structural parameters to find the thresholds. But the convexity of the loss as a function of the thresholds makes it natural and straightforward to apply Newton's method.

The part of $\mathcal{D}$ that depends on $\tau_{j\nu}$, with $1 \leq \nu \leq k_j - 1$, is

$$\sum_{i=1}^{n} f_i(y_{ij\nu} \log \pi_{ij\nu} + y_{ij,\nu+1} \log \pi_{ij,\nu+1}),$$

and thus

$$\frac{\partial \mathcal{D}}{\partial \tau_{j\nu}} = \sum_{i=1}^{n} f_i \phi(\eta_{ij} + \tau_{j\nu}) \left[\frac{y_{ij\nu}}{\pi_{ij\nu}} - \frac{y_{ij,\nu+1}}{\pi_{ij,\nu+1}} \right].$$

For the second derivatives we only have to consider two cases, because the Hessian is symmetric and tridiagonal. On the diagonal we have

$$\frac{\partial^2 f}{\partial \tau_\nu \partial \tau_\nu} = -\sum_{i=1}^{n} f_i(\eta_i + \tau_\nu)\phi(\eta_i + \tau_\nu) \left[\frac{y_{i\nu}}{\pi_{i\nu}} - \frac{y_{i,\nu+1}}{\pi_{i,\nu+1}} \right] + \\ - \sum_{i=1}^{n} f_i \phi^2(\eta_i + \tau_\nu) \left[\frac{y_{i\nu}}{\pi_{i\nu}^2} + \frac{y_{i,\nu+1}}{\pi_{i,\nu+1}^2} \right],$$

and on the subdiagonal and superdiagonal

$$\frac{\partial^2 f}{\partial \tau_\nu \partial \tau_{\nu-1}} = \sum_{i=1}^{n} f_i \phi(\eta_i + \tau_\nu)\phi(\eta_i + \tau_{\nu-1}) \frac{y_{i\nu}}{\pi_{i\nu}^2}.$$

These formulas can be easily adapted to the case in which we require the thresholds for all variables to be the same, or if we require the thresholds for a variable to be a linear functions of a given vector, or if we require the thresholds of all variables to be linear functions of a single unknown vector (the matrix of threshold values is of rank one). In the case in which we require that $\tau = \theta\overline{\tau}$, where $\overline{\tau}$ is known, we find

$$\frac{\partial f}{\partial \theta} = \overline{\tau}' \mathcal{D} f(\tau),$$
$$\frac{\partial^2 f}{\partial \theta \partial \theta} = \overline{\tau}' \mathcal{D}^2 f(\tau) \overline{\tau}.$$

8. Examples

8.1. **Discrete Normal.** We use data reported by Hargenvillers [Quetelet, 1842, page 59] on the height of 100,000 French conscripts.

under 1.570 meters	28,620
1.570 to 1.598 meters	11,580
1.598 to 1.624 meters	13,990
1.624 to 1.651 meters	14,410
1.651 to 1.678 meters	11,410
1.678 to 1.705 meters	8,780
1.705 to 1.732 meters	5,530
1.732 to 1.759 meters	3,190
above 1.759 meters	2,490
	100,000

TABLE 1. Quetelet/Hargenvliiers Height of French Conscripts

	Sera											
	w	w	w	(+)	w	(+)	?	w	w	(+)	w	w
	?	w	?	w	w	w	?	w	w	w	w	?
	w	w	w	w	w	w	w	w	w	w	w	w
	w	w	w	w	w	w	-	w	w	w	w	?
C	w	(+)	w	(+)	w	w	?	(+)	w	(+)	w	w
e	w	w	(+)	(+)	w	w	?	w	w	(+)	w	w
l	(+)	(+)	(+)	(+)	+	+	w	(+)	w	(+)	(+)	w
l	w	+	(+)	(+)	w	(+)	w	(+)	w	(+)	(+)	w
s	w	(+)	(+)	(+)	w	(+)	w	(+)	w	(+)	w	w
	?	?	w	w	w	w	?	w	w	w	w	?
	w	w	(+)	w	w	w	?	w	w	w	w	w
	w	(+)	+	(+)	(+)	(+)	w	(+)	w	(+)	+	w

TABLE 2. Fisher/Taylor Serological Readings

8.2. **Serological Readings.**

References

D. Böhning. The Lower Bound Method in Probit Regression. *Computational Statistics and Data Analysis*, 30:13–17, 1999.

J. Burridge. A Note on Maximum Likelihood Estimation for Regression Models Using Grouped Data. *Journal of the Royal Statistical Society B*, 43:41–45, 1981.

J. De Leeuw. Principal Component Analysis of Binary Data by Iterated Singular Value Decomposition. *Computational Statistics and Data Analysis*, 50(1):21–39, 2006.

J. De Leeuw. Block Relaxation Methods in Statistics. In H.H. Bock, W. Lenski, and M.M. Richter, editors, *Information Systems and Data Analysis*, Berlin, 1994. Springer Verlag.

J. De Leeuw. Least Squares Optimal Scaling of Partially Observed Linear Systems. In K. van Montfort, J. Oud, and A. Satorra, editors, *Recent Developments on Structural Equation Models*, chapter 7. Kluwer Academic Publishers, Dordrecht, Netherlands, 2004.

D. J. Finney. *Probit Analysis*. Cambridge University Press, Cambridge, U.K., 1947.

G.H. Fischer and I.W. Molenaar, editors. *Rasch Models. Foundations, Recent Developments, and Applications*. Springer, 1995.

W.J. Heiser. Convergent Computing by Iterative Majorization: Theory and Applications in Multidimensional Data Analysis. In W.J. Krzanowski, editor, *Recent Advanmtages in Descriptive Multivariate Analysis*, pages 157–189. Oxford: Clarendon Press, 1995.

J. Hemelrijk. Underlining Random Variables. *Statistica Neerlandica*, 20: 1–7, 1966.

K. Lange, D.R. Hunter, and I. Yang. Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9:1–20, 2000.

D.N. Lawley. The Factorial Analysis of Multiple Item tests. *Proceedings of the Royal Society of Edinburgh*, 62-A:74–82, 1944.

J. S. Long. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, 1987.

G.S. Maddala. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, 1983.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Number 37 in Monographs on Statistics and Applied Probability. Chapman and Hall, 2 edition, 1989.

K. Pearson. 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, 1900.

J.W. Pratt. Concavity of the Log Likelihood. *Journal of the American Statistical Association*, 76:103–106, 1981.

A. Quetelet. *A Treatise on Man and the Development of his Faculties*. William & Robert Chambers, Edinburgh, 1842.

Appendix A. Code

```
genProbit<-function(
        mat,
        eta,
        freq=rep(1,dim(mat)[1]),
        eps=1e-6,
        itmax=100,
        mod=1,
        verbose=FALSE,
        proj=pIdent,
        thres=repList("all",dim(mat)[2]),
        extra=NULL
        ) {
mat<-as.data.frame(mat); ncat<-as.vector(apply(mat,2,
   max)); tau<-thres
n<-dim(mat)[1]; m<-dim(mat)[2]; itel<-1; tht<-rep(1,m)
   ; fo<-Inf
for (j in 1:m) {
    cs<-cumsum(sapply(1:ncat[j],function(l) sum(freq[
       which(mat[,j]==l)])))
    tau[[j]]<-qnorm(cs[-length(cs)]/cs[length(cs)])
        if (is.double(thres[[j]])) {
                tht[j]<-sum(thres[[j]]*tau[[j]])/sum(
                   thres[[j]]*thres[[j]])
                tau[[j]]<-thres[[j]]*tht[j]
                }
        }
repeat {
        h<-matrix(0,n,m)
        for (j in 1:m) {
            op<-outer(eta[,j],tau[[j]],"+")
                dn<-dnorm(op); pn<-pnorm(op)
```

```
                    pp<-colDiff(cbind(0,pn,1)); pd
                      <-colDiff(cbind(0,dn,0))
                    gg<--(pd/pp)
                    h[,j]<-sapply(1:n,function(i) gg[i,
                      mat[[j]][i]])
                    }
               eta<-proj(eta-h,freq,extra); fs<-rep(0,n)
               for (j in 1:m) {
                  op<-outer(eta[,j],tau[[j]],"+"); pn
                    <-pnorm(op); pp<-colDiff(cbind(0,pn,1))
                    fs<-fs-2*log(sapply(1:n,function(i) pp
                      [i,mat[[j]][i]]))
                    }
          fe<-sum(freq*fs)
          if (itel%%mod == 0) {
               ft<-0
                    for (j in 1:m) {
                       if (is.double(thres[[j]])) {
                          th<-thresProp(mat[[j]],eta[,j
                            ],tht=tht[j],thres=thres[[j
                            ]],freq=freq,verbose=FALSE)
                          tht[j]<-th$tht; tau[[j]]<-tht[
                            j]*thres[[j]]; ft<-ft+th$f
                          }
                          else {
                                th<-thresFree(mat[[j
                                  ]],eta[,j],tau=tau
                                  [[j]],freq=freq,
                                  verbose=FALSE)
                          tau<-th$tau; ft<-ft+th$f
                          }
                          }
                    }
          else ft<-fe
```

```
          if (verbose) cat(
                  "Iteration: ",formatC(itel,width=3,
                     format="d"),
                  "O-Function: ",formatC(fo,digits=8,
                     width=12,format="f"),
                  "E-Function: ",formatC(fe,digits=8,
                     width=12,format="f"),
                  "T_Function: ",formatC(ft,digits=8,
                     width=12,format="f"),
                  "\n")
          if (((fo - ft) < eps) || (itel == itmax))
             break
          itel<-itel+1; fo<-ft
          }
return(list(tau=tau,tht=tht,eta=eta))
}

pIdent<-function(x,freq,extra) return(x)

pAve<-function(x,freq,extra) {
for (j in 1:dim(x)[2]) x[,j]<-weighted.mean(x[,j],freq
   )
return(x)
}

pRegres<-function(x,freq,extra) return(apply(x,2,
   function(y) y-(lsfit(extra,y,wt=freq)$residuals)))

pPCA<-function(x,freq,extra) return(rankApprox(x,freq,
   extra))

thresFree<-function(mat,eta,tau,freq,eps=1e-6,itmax
   =100,verbose=FALSE) {
k<-length(tau); n<-length(mat); itel<-1
```

```
repeat{
        ar<-outer(eta,tau,"+")
        pp<-colDiff(cbind(0,pnorm(ar),1))
        f<--2*sum(freq*log(sapply(1:n,function(i) pp[i,mat[i]])))
        dn<-dnorm(ar); rs<-matrix(0,n,k+1)
        for (i in 1:n) rs[i,mat[i]]<-1/pp[i,mat[i]]; ss<-rs/pp
        gg<--2*sapply(1:k,function(l) sum(freq*dn[,l]*(rs[,l]-rs[,l+1])))
        mg<-max(abs(gg))
        if (verbose) cat(
                "** Iteration: ",formatC(itel,width=3, format="d"),
                "** Function: ",formatC(f,digits=8,width=12,format="f"),
                "** MaxGrad: ",formatC(mg,digits=8,width=12,format="f"),
                "\n")
        if ((mg < eps) || (itel == itmax)) break
        dg<-2*sapply(1:k,function(l) sum(freq*ar[,l]*dn[,l]*(rs[,l]-rs[,l+1])))
        dg<-dg+2*sapply(1:k,function (l) sum(freq*(dn[,l]^2)*(ss[,l]+ss[,l+1])))
        dd<-diag(dg)
        for (i in 1:k-1) dd[i,i+1]<-dd[i+1,i]<--2*sum(freq*dn[,i]*dn[,i+1]*ss[,i+1])
        tau<-tau-solve(dd,gg)
        itel<-itel+1
        }
return(list(tau=tau,f=f))
}
```

```
thresProp<-function(mat,eta,tht,thres,freq,eps=1e-6,
    itmax=100,verbose=FALSE) {
k<-length(thres); n<-length(mat); tau<-tht*thres; itel
    <-1
repeat{
        ar<-tht*outer(eta,thres,"+")
        pp<-colDiff(cbind(0,pnorm(ar),1))
        f<--2*sum(freq*log(sapply(1:n,function(i) pp[i
            ,mat[i]])))
        dn<-dnorm(ar); rs<-matrix(0,n,k+1)
        for (i in 1:n) rs[i,mat[i]]<-1/pp[i,mat[i]];
            ss<-rs/pp
        gg<--sum(thres*sapply(1:k,function(l) sum(freq
            *dn[,l]*(rs[,l]-rs[,l+1]))))
        if (verbose) cat(
                "**␣Iteration:␣",formatC(itel,width
                    =3, format="d"),
                "**␣Function:␣",formatC(f,digits=8,
                    width=12,format="f"),
                "**␣AbsGrad:␣",formatC(abs(gg),digits
                    =8,width=12,format="f"),
                "\n")
        if ((abs(gg) < eps) || (itel == itmax)) break
        dd<-0;
        for (l in 1:k)
                dd<-dd+(thres[l]^2)*sum(freq*((ar[,l]
                    *dn[,l]*(rs[,l]-rs[,l+1]))+((dn[,l
                    ]^2)*(ss[,l]+ss[,l+1]))))
        for (l in 1:(k-1))
            dd<-dd-2*thres[l]*thres[l+1]*sum(freq*dn[,
                l]*dn[,l+1]*ss[,l+1])
    tht<-tht-(gg/dd)
        itel<-itel+1
        }
```

```
return(list(tht=tht,f=f))
}

rankApprox<-function(x,p) {
s<-svd(x)
return(s$u[,1:p]%*%(s$d[1:p]*t(s$v[,1:p])))
}

colDiff<-function(x) {
return(matrix(t(apply(x,1,diff)),dim(x)[1],dim(x)
    [2]-1))
}

repList<-function(x,n) {
z<-list()
for (i in 1:n)
    z<-c(z,list(x))
return(z)
}
```

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

URL, Jan de Leeuw: `http://gifi.stat.ucla.edu`