Supplementary material for the manuscript "Random effect models for multivariate mixed data: a Parafac-based finite mixture approach"

Marco Alfò [1] **and Paolo Giordani** [1]

[1] Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy

Address for correspondence: Marco Alfò, Dipartimento di Scienze Statistiche, Sapienza Università di Roma, P.le Aldo Moro, 5 I–00185 Rome, Italy.
E-mail: `marco.alfo@uniroma1.it`.
Phone: (+39) 06 4991 0763.
Fax: (+39) 06 4991 0072.

Abstract: Further details on the application and on the estimation algorithm

Key words: NMES data, EM algorithm, Parafac algorithm

1 The EM algorithm for ML parameter estimation

To discuss the EM algorithm for ML estimation of parameters in the proposed flexible approach, we start by recalling the likelihood function

$$\begin{aligned}
\ell(\cdot) &= \sum_{i=1}^{n} \log\left\{ \sum_{g_1,\ldots,g_p} \pi_{g_1,\ldots,g_p} \prod_j \left[f\left(y_{ij} \mid \mathbf{x}_{ij}, \zeta_{jg_j}\right)\right]\right\} = \qquad (1.1)\\
&= \sum_{i=1}^{n} \log\left\{ \sum_{g_1,\ldots,g_p} \pi_{g_1,\ldots,g_p} f\left(\mathbf{y}_i \mid \mathbf{x}_{i1},\ldots,\mathbf{x}_{ip}, \zeta_{1g_1},\ldots,\zeta_{pg_p}\right)\right\} = \\
&= \sum_{i=1}^{n} \log\left(\sum_{g_1,\ldots,g_p} \pi_{g_1,\ldots,g_p} f_{i,g_1,\ldots,g_p}\right).
\end{aligned}$$

Based on this form, we may define the elements of the score function as follows:

$$
\begin{aligned}
\frac{\partial \ell(\cdot)}{\partial \boldsymbol{\beta}_j} &= \sum_{i=1}^{n} \sum_{g_1,\dots,g_p} w_{ig_1,\dots,g_p} \frac{\partial}{\partial \boldsymbol{\beta}_j} \log\left[\pi_{g_1,\dots,g_p} f_{i,g_1,\dots,g_p}\right] \\
&= \sum_{i=1}^{n} \sum_{g_j=1}^{K_j} w_{ig_j} \frac{\partial}{\partial \boldsymbol{\beta}_j} \left[\log(f_{ijg_j})\right], \\
\frac{\partial \ell(\cdot)}{\partial \zeta_{jg_j}} &= \sum_{i=1}^{n} \sum_{g_1,\dots,g_{j-1},g_{j+1},\dots,g_p} w_{ig_1,\dots,g_p} \frac{\partial}{\partial \zeta_{jg_j}} \log\left[\pi_{g_1,\dots,g_p} f_{i,g_1,\dots,g_p}\right]. \\
&= \sum_{i=1}^{n} w_{ig_j} \frac{\partial}{\partial \zeta_{jg_j}} \left[\log(f_{ijg_j})\right], \\
\frac{\partial \ell(\cdot)}{\partial \pi_{g_1,\dots,g_p}} &= \sum_{i=1}^{n} w_{ig_1,\dots,g_p} \frac{\partial}{\partial \pi_{g_1,\dots,g_p}} \log(\pi_{g_1,\dots,g_p}),
\end{aligned}
$$

where

$$
\begin{aligned}
w_{ig_1,\dots,g_p} &= \frac{\pi_{g_1,\dots,g_p} f_{i,g_1,\dots,g_p}}{\sum_{h_1,\dots,h_p} \pi_{h_1,\dots,h_p} f_{i,h_1,\dots,h_p}}, \\
w_{ig_j} &= \sum_{g_1,\dots,g_{j-1},g_{j+1},\dots,g_p} w_{ig_1,\dots,g_p}
\end{aligned}
$$

represent joint and outcome-specific posterior probabilities of component membership, respectively. The weights for a specific outcome are simply defined by summing $w_{ig_1,\dots,g_p}$ over the remaining profiles; only at the E-step a $n \times \prod_j K_j$ calculation is needed, where the p densities and the priors are used to compute the posterior probabilities of component membership. Once the choice for K_j, $j = 1, \dots, p$, has been made, for example by fitting univariate regression models to each profile, the EM algorithm is quite simple to be implemented, by adding a factor variable with K_j levels to each outcome model. According to Karlis and Meligkotsidou (2007), we use AIC (Akaike, 1974) to choose the number of components in univariate models, and then retain such a number in the multivariate model. Different criteria can be used as well, with negligible impact on fixed parameter estimates.

2 Parafac approximation, additional steps for the EM algorithm

The idea is, here, to summarize the joint probabilities using a limited number of components, q $(< m)$, that approximate the original probability tensor at best:

$$
\pi_{g_1,\dots,g_p} \simeq \sum_{h=1}^{q} \tau_h \prod_{j=1}^{p} \pi_{jg_j|h}. \tag{2.1}
$$

The best Parafac approximation is achieved by least squares. It is convenient to reorder the elements of the tensor $\underline{\boldsymbol{\Pi}}$ in a matrix by using so-called matricization (or unfolding), see, e.g., Kiers (2000). There exist different kinds of matricization; the so-called 'Mode j' matricization of $\underline{\mathbf{\Pi}}$ is the matrix $\mathbf{\Pi}^{(j)}$ of order $(K_j \times K_1 \cdots K_{j-1}K_{j+1} \cdots K_p)$. Letting $\mathbf{\Pi}_j$, $j = 1, \ldots, p$, be the matrices of order $(k_j \times q)$ with generic column $\boldsymbol{\pi}_{jh} = [\pi_{j1|h} \cdots \pi_{jK_j|h}]'$, and $\mathbf{T}$ the diagonal matrix of order q with main diagonal equal to the vector $\boldsymbol{\tau} = [\tau_1 \cdots \tau_q]'$, the least squares estimates are found by solving the constrained problem

$$\begin{array}{rl} \underset{\mathbf{\Pi}_1,\ldots,\mathbf{\Pi}_p,\mathbf{T}}{\min} & = ||\mathbf{\Pi}^{(j)} - \mathbf{\Pi}_j\mathbf{T}(\mathbf{\Pi}_p \odot \cdots \odot \mathbf{\Pi}_{j+1} \odot \mathbf{\Pi}_{j-1} \odot \cdots \odot \mathbf{\Pi}_1)'||^2, \\ \text{s.t.} & \boldsymbol{\pi}'_{jh}\mathbf{1} = 1, \boldsymbol{\pi}_{jh} \geq \mathbf{0}, h = 1, \ldots, q, j = 1, \ldots, p, \\ & \mathbf{T} \text{ diagonal}, \boldsymbol{\tau}'\mathbf{1} = 1, \boldsymbol{\tau} \geq \mathbf{0}, \end{array} \tag{2.2}$$

where $\odot$ denotes the Khatri-Rao product of matrices (columnwise Kronecker product, i.e., given two matrices $\mathbf{V}$ and $\mathbf{W}$ with the same number q of columns $\mathbf{V} \odot \mathbf{W} = [\mathbf{v}_1 \otimes \mathbf{w}_1 \ldots \mathbf{v}_q \otimes \mathbf{w}_q]$ where $\otimes$ is the Kronecker product of matrices) and $\mathbf{1}$ and $\mathbf{0}$ are column vectors of 1's and 0's of appropriate order, respectively.

A column-wise Alternating Least Squares (ALS) algorithm can be used to find the optimal solution. It iteratively updates every column of the matrices $\mathbf{\Pi}_j$ and the diagonal of $\mathbf{T}$ keeping the remaining parameters fixed.

1. Update of $\boldsymbol{\pi}_{jh}$, $h = 1, \ldots, q$, $j = 1, \ldots, p$.

In order to update $\boldsymbol{\pi}_{jh}$ we first split the component matrix $\mathbf{\Pi}_j$ into two parts, $\boldsymbol{\pi}_{jh}$ and $\mathbf{\Pi}_{j-h}$, where $\mathbf{\Pi}_{j-h}$ is the matrix $\mathbf{\Pi}_j$ with the h-th column removed. Setting $\mathbf{Z}' = \mathbf{T}(\mathbf{\Pi}_p \odot \cdots \odot \mathbf{\Pi}_{j+1} \odot \mathbf{\Pi}_{j-1} \odot \cdots \odot \mathbf{\Pi}_1)'$, the loss function in equation (2.2) can then be written as

$$||\mathbf{\Pi}^{(j)} - \mathbf{\Pi}_{j-h}\mathbf{Z}'_{-h} - \boldsymbol{\pi}_{jh}\mathbf{z}'_h||^2, \tag{2.3}$$

where $\mathbf{z}_h$ denotes the h-th column of $\mathbf{Z}$ and $\mathbf{Z}_{-h}$ is equal to $\mathbf{Z}$ after removing the h-th column. Setting $\mathbf{\Pi}_h^{(j)} = \mathbf{\Pi}^{(j)} - \mathbf{\Pi}_{j-h}\mathbf{Z}'_{-h}$ and taking into account that, for any triple of matrices of appropriate order, $\text{vec}(\mathbf{UVW}') = (\mathbf{W} \otimes \mathbf{U})\text{vec}(\mathbf{V})$, where the vec operator converts a matrix into a column vector, we get the following minimization problem:

$$\begin{array}{rl} \underset{\boldsymbol{\pi}_{jh}}{\min} & ||\text{vec}(\mathbf{\Pi}_h^{(j)}) - (\mathbf{z}_s \otimes \mathbf{I})\boldsymbol{\pi}_{jh}||^2, \\ \text{s.t.} & \boldsymbol{\pi}'_{jh}\mathbf{1} = 1, \boldsymbol{\pi}_{jh} \geq \mathbf{0}, \end{array} \tag{2.4}$$

where $\mathbf{I}$ is the identity matrix. A solution to equation (2.4) is provided by ter Braak et al. (2009). If $\boldsymbol{\pi}_{jh}$ is reparametrized as $\mathbf{Yv} + K_j^{-1}\mathbf{1}$, being $\mathbf{Y}$ a columnwise orthonormal basis for the orthocomplement of $\mathbf{1}$ and $\mathbf{v}$ a vector, ter Braak et al. (2009) show that the above minimization problem boils down to a particular Least Squares with Inequality (LSI) problem. The LSI problem consists of minimizing

$$\begin{array}{rl} \underset{\mathbf{v}}{\min} & ||\mathbf{m} - \mathbf{Lv}||^2, \\ \text{s.t.} & \mathbf{Cv} \geq \mathbf{d}, \end{array} \tag{2.5}$$

where $\mathbf{L}$, $\mathbf{C}$, $\mathbf{m}$ and $\mathbf{d}$ are matrices and vectors of appropriate order. The solution to equation (2.5), satisfying the Kuhn-Tucker conditions, is found by Lawson and Hanson

(1995). In particular, it can be shown that the problem solved by ter Braak et al. (2009) and the LSI one coincide when $\mathbf{m} = \text{vec}(\mathbf{\Pi}_h^{(j)}) - K_j^{-1}\mathbf{1}$, $\mathbf{L} = (\mathbf{z}_h \otimes \mathbf{I})\mathbf{Y}$, $\mathbf{C} = \mathbf{Y}$ and $\mathbf{d} = -K_j^{-1}\mathbf{1}$. Given the optimal value of $\mathbf{v}$ minimizing equation (2.5), the optimal value of $\boldsymbol{\pi}_{jh}$ minimizing equation (2.4) is equal to $\mathbf{Y}\mathbf{v} + K_j^{-1}\mathbf{1}$.

2. Update of T.

The update of $\mathbf{T}$ is very similar to that of $\boldsymbol{\pi}_{jh}$. It can be expressed in terms of $\boldsymbol{\tau}$ bearing in mind that $\mathbf{T} = \text{diag}(\boldsymbol{\tau})$, where diag is the operator creating the matrix $\mathbf{T}$ with the elements of $\boldsymbol{\tau}$ along the diagonal. Starting from equation (2.2), we have

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & ||\text{vec}(\mathbf{\Pi}^{(j)}) - (\mathbf{Z} \otimes \mathbf{\Pi}_j)\text{vec}(\text{diag}(\boldsymbol{\tau}))||^2, \\ \text{s.t.} \quad & \boldsymbol{\tau}'\mathbf{1} = 1, \boldsymbol{\tau} \geq \mathbf{0}. \end{aligned} \tag{2.6}$$

getting a regression problem between the vector of the dependent variable $\text{vec}(\mathbf{\Pi}^{(j)})$ and the matrix of the explanatory variables $(\mathbf{Z} \otimes \mathbf{\Pi}_j)$ with regression coefficients $\text{vec}(\text{diag}(\boldsymbol{\tau}))$. Since $\text{diag}(\boldsymbol{\tau})$ is a diagonal matrix of order q, $\text{vec}(\text{diag}(\boldsymbol{\tau}))$ contains $q(q-1)$ elements equal to zero corresponding to the off-diagonal elements of $\text{diag}(\boldsymbol{\tau})$. It follows that $q(q-1)$ regression coefficients are equal to zero and, therefore, the corresponding explanatory variables do not play an active role and can be omitted. It is easy to see that, after a little algebra, equation (2.6) can equivalently be written as

$$\begin{aligned} \min_{\boldsymbol{\tau}} \quad & ||\text{vec}(\mathbf{\Pi}^{(j)}) - (\mathbf{Z} \odot \boldsymbol{\pi}_{jh})\boldsymbol{\tau}||^2, \\ \text{s.t.} \quad & \boldsymbol{\tau}'\mathbf{1} = 1, \boldsymbol{\tau} \geq \mathbf{0}, \end{aligned} \tag{2.7}$$

the solution of which is provided by ter Braak et al. (2009).

3 NMES data: univariate regression models

In this section, we consider two measures of utilization (*Emr* and *Hosp*) and define a regression model for these outcomes, as a function of socio-economic status, observed health status and private insurance coverage. We expect that, since the access to an emergency room is mainly driven by potentially unpredictable events, private insurance coverage may have a near null impact on this outcome. The opposite is true for *Hosp* as length of stay may well be a function of health care needs and of the individual capacity to cover hospitalization costs, see, e.g., Zweifel and Manning (2000). Both *Emr* and *Hosp* include a high proportion of zero counts, corresponding to zero recorded demand over the observed time interval, and a high degree of unconditional overdispersion. Therefore, we have defined two semi-parametric regression models, with (conditional) Poisson distributions for the responses, log link functions and discrete latent effects. The number of components in each profile have been chosen using AIC; it is worth noting that similar results can be observed by using BIC (Schwarz, 1978). Parameter estimates, derived using the `R` library `npmlreg` are reported in Table 1 below.

Table 1: Univariate regression models for *Emr* and *Hosp*. Parameter estimates and 90% bootstrap confidence intervals (CI's).

	Emr		*Hosp*	
Variable	Estimate	*CI*	Estimate	*CI*
Exchlth	-0.650	(-1.030-0,269)	-0.694	(-1.077,-0.310)
Poorhlth	0.500	(0.304,0,696)	0.485	(0.289,0.681)
Numchron	0.212	(0.162,0,262)	0.272	(0.222,0.323)
AdLim	0.403	(0.228,0,578)	0.338	(0.157,0.519)
West	0.201	(0.016,0,386)	0.096	(-0.086,0.279)
Age	0.101	(-0.017,0.219)	0.172	(0.058,0.287)
AfroAmer	0.144	(-0.072,0.360)	0.074	(-0.165,0.313)
Male	0.042	(-0.118,0.202)	0.175	(0.015, 0.334)
Married	-0.080	(-0.244,0,085)	-0.030	(-0.196,0.136)
EdYears	-0.017	(-0.038,0.004)	0.001	(-0.020,0.023)
Employed	0.161	(-0.092,0.414)	0.058	(-0.199,0.316)
Privins	0.078	(-0.125,0.281)	0.193	(-0.009,0.394)
Medicaid	0.197	(-0.060,0.454)	0.134	(-0.135,0.403)
ζ_1	-1.085	(-2.039,-0.131)	-5.321	(-8.179,-2.463)
ζ_2	-3.584	(-4.552,-2.616)	-1.524	(-2.492,-0.556)
ζ_3			-3.247	(-4.291,-2.202)
sd(RE)	1.142		1.356	

Looking the previous estimates and considering the context we are discussing, at least two comments arise. First, both utilization measures (*Emr* and *Hosp*) may depend on common, individual-specific, unobserved characteristics. This may be linked to health status and to the propensity to use a given service. Therefore, it could be interesting to define a bivariate model for the couple (*Emr*, *Hosp*) to have a measure of dependence. Second, as expected, the choice for a private insurance scheme is barely significant only in the equation for *Hosp*; however, since the choice for private insurance may depend on current individual health status, as well as represent an anticipation of future health care demand, its choice may be endogenous, as it may depend on unobservables that are among the drivers of utilization.

4 Parafac approximation of the joint probability tensor with $q = 1$

By fitting the Parafac using $q = 1$ to the joint probability tensor of Table 4 in the paper, we obtain the solution reported in Table 2.

The results in Table 2 lead to the approximated estimate of the prior probability tensor reported in Table 3. By comparing this table and Table 4 in the main text, we can conclude that the Parafac approximation we obtain for $q = 1$ is extremely poor.

Table 2: Parafac estimate (using $q = 1$) of $\underline{\mathbf{\Pi}}$.

$\hat{\mathbf{\Pi}}_1$ (*Emr*)			$\hat{\mathbf{\Pi}}_2$ (*Hosp*)			$\hat{\mathbf{\Pi}}_3$ (*Privins*)	
RE	*class* 1		RE	*class* 1		RE	*class* 1
comp. 1	0.61		*comp.* 1	0.49		*comp.* 1	0.58
comp. 2	0.39		*comp.* 2	0.30		*comp.* 2	0.42
			comp. 3	0.21			

Table 3: Parafac approximation (using $q = 1$) of the latent effects' joint probability distribution ($\underline{\mathbf{\Pi}}$).

		Privins					
		comp. 1			*comp.* 2		
		Emr			*Emr*		
		comp. 1	*comp.* 2	*comp.* 3	*comp.* 1	*comp.* 2	*comp.* 3
Hosp	*comp.* 1	0.175	0.107	0.076	0.125	0.076	0.054
	comp. 2	0.110	0.067	0.048	0.079	0.048	0.034

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Karlis, D., Meligkotsidou, L. (2007). Finite mixtures of multivariate Poisson distributions with application. *Journal of Statistical Planning and Inference*, **137**, 1942–1960.

Kiers, H.A.L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, **14**, 105–122.

Lawson, C.L, Hanson, R.J. (1995). *Solving Least Squares Problems, Classics in Applied Mathematics, Vol. 15.* SIAM, Philadelphia.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6** 461–464.

ter Braak, C.J.F., Kourmpetis, Y., Kiers, H.A.L., Bink, M.C.A.M. (2009). Approximating a similarity matrix by a latent class model: a reappraisal of additive fuzzy clustering. *Computational Statistics and Data Analysis*, **53**, 3183–3193.

Zweifel, P. and Manning, W.G. (2000) Moral Hazard and Consumer Incentives in Health Care, in Culyer, A.J. and Newhouse, J.P. (eds) *Handbook of Health Economics, part I A*, chapter 8, 409–459, Elsevier BV.