Quaderni di Statistica
Vol. 5, 2003

Computational issues in the E-M algorithm for ranks model estimation with covariates

Domenico Piccolo
Dipartimento di Scienze Statistiche, Università di Napoli Federico II
E-mail: domenico.piccolo@unina.it

Summary: We discuss some computational issues arising in the maximum likelihood estimation of a statistical model for the ranks. The main result of the paper is a unified approach to the E-M algorithm for estimating both the parameters of the model and the coefficients of the raters' covariates. Emphasis is given to the implementation of the algorithms in a matrix-oriented language for an effective derivation of the estimates and of their asymptotic standard errors. In order to support the significance of the results, some experiences on a real data set are also reported.

Keywords: Rank modelling, E-M algorithm, Preference analysis.

1. Introduction

The usefulness of rank modelling for the analysis and/or the interpretation of statistical data arising from preferences or evaluation contexts has been established in many recent works (Marden, 1995; Fligner and Verducci, 1999). In this area, a more general approach has been derived by D'Elia and Piccolo (2003) that introduced a mixture modelling of both the uncertainty and liking/disliking feeling processes.

The random variable suggested for this process is a *M*ixture of a discrete *U*niform and a shifted *B*inomial random variables, and the model has been defined "*MUB model*". For a fixed number of objects, this model depends upon the values of two parameters. Moreover, D'Elia (2003b) derived the maximum likelihood (ML) estimators (and their asymptotic standard errors) for the parameters of the MUB model when one of the parameter is explicitly related to a set of rater's covariates.

In this paper, we derive the single steps which are necessary for an effective implementation of the estimation procedure in a completely general framework: that is, when none, one or both the model parameters are related to the raters' covariates.

The opportunity for assessing a unifying notation and algorithm could be considered as a useful byproduct of this work. Then, we limit ourselves to quote the structure of the MUB model and the E-M algorithm for an essential understanding of the single steps; more extensive discussions can be found in the references mentioned above.

The paper is organized as follows: in the next section we introduce notations for the MUB model. Then, in section 3 we reproduce the single steps for the parameters estimation and in section 4 we show the asymptotic expressions for the standard errors of the ML estimators. Some empirical evidences and final comments conclude the paper.

2. The MUB model: introduction and notation

Suppose that a set of m objects (or a set of m ordinal evaluation degrees) has been well defined and let r be the rank assigned by a single rater to a given item; we assume that $r = 1$ means "most preferred" while $r = m$ means "least preferred". We stress that m is a known and prefixed number, and the analysis concerns with the preference towards a single object or the evaluation of a single item.

Then, we interpret the value of r as the observed value of a discrete random variable R defined on the support $\{1, 2, ...m\}$. A probabilistic mixture model for R has been introduced by D'Elia and Piccolo (2003) that defined $R \sim MUB\,(m, \pi, \xi)$ if:

$$Pr\,(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1}\, \xi^{m-r} + (1-\pi)\,\frac{1}{m},\; r = 1, 2, ..., m.$$

Both the parameters π and ξ are defined on $[0, 1]$; however, π is a measure related to the uncertainty that it is generally associated with the elicitation mechanism, while ξ is positively related to the degree of liking expressed by the raters towards the prefixed object; further properties of this random variables, and the relations of these parameters with the first four moments are discussed by Piccolo (2003).

In the following, we let $\theta = (\pi, \xi)'$ for the set of parameters to be estimated by the ML method, and we define as $b\,(r;\xi)$ the probability distribution of a shifted Binomial random variable. Thus, we have:

$$b\,(r;\xi) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1}\, \xi^{m-r}, \;\; r = 1, 2, ..., m.$$

Then, the MUB random variable probability distribution becomes:

$$Pr\,(R = r|\theta) = \pi \left(b\,(r;\xi) - \frac{1}{m} \right) + \frac{1}{m}, \;\; r = 1, 2, ..., m.$$

After a simple algebra, the previous formula can be re-formulated in a different manner:

$$Pr\,(R = r) = (unc) + (imp) \binom{m-1}{r-1} (dis)^{r}\,, \, r = 1, \ldots, m;$$

where the quantities:

$$\textit{unc} = \frac{1-\pi}{m}, \quad \textit{imp} = \frac{\pi}{1-\xi}\, \xi^{m}, \quad \textit{dis} = \frac{1-\xi}{\xi}\,,$$

are defined for a more immediate meaning of the parameters.

In fact, we can introduce for these parameters the following interpretation:

unc is the *uncertainty share*, that is a constant baseline that is present in the probability of any rank, and thus in the elicitation process;

imp is a constant that raises the values of the probability distribution and thus it is an *impact coefficient* for a higher probability of a specific rank value;

dis is a sort of *disliking odd measure* since it is the ratio of a non-preference to a preference quantity.

As section 5 will confirm, the use of such parameters may help for appealing interpretations of the MUB models estimates, mainly in comparative studies.

In the following, we will distinguish the MUB model with no covariates (the distribution of ranks is accounted solely by the parameters in θ) from the MUB models with one or two sets of covariates (used for explaining the different preferences of subsets of raters).

In the first case, the information contained in the sample of the observed ranks for the n subjects: $\boldsymbol{r} = (r_1, r_2, ..., r_n)'$ is strictly equivalent to that contained in the vector of the observed frequencies of the ordered ranks: $\boldsymbol{n} = (n_1, n_2, ..., n_m)'$. This information is necessary and sufficient to make inference on the parameters in θ.

In the second case, the probabilistic models are able to explain the ranks distribution by means of π (which may be related to raters' covariates by a parameters vector $\boldsymbol{\beta}$) and/or by means of ξ (which may be related to raters' covariates by a parameters vector $\boldsymbol{\gamma}$). However, differently from the first case when the information consist of both the ranks and the raters' covariates, we cannot aggregate the subjects with the same rank since they generally have different values for the covariates. Thus, the sample data will be collected in the design matrix $\boldsymbol{D} = (\boldsymbol{r} \,|\, \mathbf{1}\,\boldsymbol{X})$ where $\mathbf{1}$ is an $n-$length vector of 1, and $\boldsymbol{X}$ is the covariates matrix.

Thus,

$$\boldsymbol{D} = \begin{pmatrix} r_1 & 1 & x_{11} & \dots & x_{1p} \\ r_2 & 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ r_n & 1 & x_{n1} & \dots & x_{np} \end{pmatrix} .$$

In the following, we denote the subset of variables related to π with $(\mathbf{1}\ \boldsymbol{Y})$ and those related to ξ with $(\mathbf{1}\,\boldsymbol{W})$. Now, the two subsets are not generally disjoint and we cannot assume that $\boldsymbol{X} = (\boldsymbol{Y}|\boldsymbol{W})$; moreover, we need to include the vector $\mathbf{1}$ in any set of variables.

Among the p covariates of $\boldsymbol{X}$, let us assume that one chooses[1] as variables for explaining the value of π only those whose indexes are in the

[1]Such selection can be based on *a priori* grounds or on some preliminary experimental evidences; for instance, on the comparison of the rank averages or patterns in different subsets of raters.

set: $I_\pi = \{s_1, s_2, \ldots, s_{p_\pi}\}$; then, we need $p_\pi + 1$ parameters to explain the π, and we assume -throughout the paper- that the length of the $\boldsymbol{\beta}$ vector is $p_\pi + 1$. Similarly, for explaining the value of ξ, let us assume that one chooses —among the p covariates of $\boldsymbol{X}$— only those whose indexes are in the set: $I_\xi = \left\{t_1, t_2, \ldots, t_{p_\xi}\right\}$; then, we need $p_\xi + 1$ parameters to explain the ξ, and we assume —throughout the paper— that the length of the $\boldsymbol{\gamma}$ vector is $p_\xi + 1$.

As a consequence, in order to simplify the notation we let ρ_i for the i-th row of the predictor in any case; specifically, $\rho_i = (1\,\boldsymbol{x}_i)'$ if the covariates are the $\boldsymbol{X}$ matrix, $\rho_i = (1\,\boldsymbol{y}_i)'$ for the $\boldsymbol{Y}$ matrix, and $\rho_i = (1\,\boldsymbol{w}_i)'$ for the $\boldsymbol{W}$ matrix. In this way, conditioning the parameter/parameters values to ρ_i may be expressed simply by $(\theta|\rho_i)$.

Then, as motivated by D'Elia (2003b), we let:

$$\pi|\rho_i = \frac{1}{1 + exp\,(-\rho_i\boldsymbol{\beta})}; \quad \xi|\rho_i = \frac{1}{1 + exp\,(-\rho_i\boldsymbol{\gamma})};$$

and $\boldsymbol{\beta} = (\beta_0,\, \beta_1,\, \ldots, \beta_p)'$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1,\, \ldots, \gamma_p)'$ are the parameter vectors to be estimated for the corresponding predictors. Of course, the role of β_0 and γ_0 consists of a baseline effect when the covariates are set to 0.

Finally, when both the parameters are related to the raters' covariates we apply the following notation:

$$\theta|\rho_i = \left(\frac{1}{1 + exp\,(-\boldsymbol{y}_i\beta)}\,, \;\; \frac{1}{1 + exp\,(-\boldsymbol{w}_i\gamma)}\right)'.$$

3. The maximum likelihood estimation via the E-M algorithm

In this section we discuss the computational steps involved in the E-M algorithm for the ML estimation of the parameters in the MUB model. First of all, we present the algorithm in absence of covariates (Table 1); then, the procedure is shown when only one set of covariates (Tables 2 and 3); finally, the general MUB model when both the parameters are functions of two (different or coincident) sets of covariates is presented (Table 4).

A fully discussion of the E-M algorithm is contained in McLachlan and Krishnan (1997) and specifically for a mixture model in McLachlan and Peel (2000). The formal derivation of the first model was obtained by D'Elia and Piccolo (2003), while for the MUB models with just one set of covariates the main results were obtained by D'Elia (2003b). Finally, computational issues related to the E-M algorithm (Table 4) and asymptotic derivations (section 4) for the general MUB model with covariates are derived in this paper.

The main advantage of the E-M algorithm derives from the splitting of the involved log-likelihood in two functions where the observed quantities and the missing ones (that is the probability π that the observation comes from one of the two sub-populations) are well defined and neatly separated. This situation implies that it is possible to activate two alternating but converging steps of Expectation and Maximization towards the ML estimates $\hat{\theta}$. Moreover, the derivation of the asymptotic standard errors of the ML estimators can be performed by straightforward steps.

For establishing a notation, we define the log-likelihood function of the MUB model without covariates, that is:

$$\begin{aligned} log\,L\,(\theta) &= \sum_{r=1}^{m} n_r\, log\,\{Pr\,(R=r|\theta)\} \\ &= \sum_{r=1}^{m} n_r\, log \left\{ \pi \left(b\,(r;\xi) - \frac{1}{m} \right) + \frac{1}{m} \right\}. \end{aligned}$$

When the covariates are present for explaining only the π parameter, that is $\pi = f\,(\boldsymbol{Y},\beta)$, the log-likelihood function of the MUB model with covariates for π is:

$$log\,L\,(\theta) = -\sum_{i=1}^{n} log\left(1+e^{(-\boldsymbol{y}_i\beta)}\right) - \sum_{i=1}^{n} log\left(b\,(r_i;\xi) + \frac{e^{(-\boldsymbol{y}_i\beta)}}{m} \right).$$

When the covariates are present for explaining only the ξ parameter, that is $\xi = g\,(\boldsymbol{W},\gamma)$, the log-likelihood function of the MUB model with covariates for ξ is:

$$log\,L\,(\theta) = \sum_{i=1}^{n} log \left\{ \pi \left(\binom{m-1}{r_i-1} \frac{e^{(-\boldsymbol{w}_i\gamma)(r_i-1)}}{\left(1+e^{(-\boldsymbol{w}_i\gamma)}\right)^{m-1}} \right) - \frac{1}{m} \right\}.$$

Finally, when the covariates are present for explaining both the π and the ξ parameters, that is $\pi = f(\boldsymbol{Y}, \beta)$ and $\xi = f(\boldsymbol{W}, \gamma)$, then the log-likelihood function of the general MUB model with covariates is:

$$log\, L(\theta) = \sum_{i=1}^{n} log \left\{ \frac{1}{1+e^{(-\boldsymbol{y}_i \beta)}} \left(\binom{m-1}{r_i - 1} \frac{e^{(-\boldsymbol{w}_i \gamma)(r_i-1)}}{(1+e^{(-\boldsymbol{w}_i \gamma)})^{m-1}} \right) - \frac{1}{m} \right\}.$$

All the steps required for the E-M procedures are explicitly reported in the Tables 1–4.

Table 1. E-M Algorithm for a MUB model without covariates.

$\boldsymbol{MUB\, Model\, (with\, \pi,\, \xi)}$	
$\theta = (\pi, \xi)'$; $\epsilon = 10^{-6}$; $dim\,(\theta) = 2.$	
$l\,(\theta) = log\, L\,(\theta) = \sum_{r=1}^{m} n_r\, log \left\{ \pi \left(b\,(r; \xi) - \frac{1}{m} \right) + \frac{1}{m} \right\}.$	
$Steps$	
0	$\theta^{(0)} = \left(\pi^{(0)}, \xi^{(0)}\right)' = \left(\frac{1}{2}, \frac{m-\overline{R}_n}{m-1} \right)'$; $l^{(0)} = log\, L\left(\theta^{(0)}\right).$
1	$b\left(r; \xi^{(k)}\right) = \binom{m-1}{r-1} \left(1-\xi^{(k)}\right)^{r-1} \left(\xi^{(k)}\right)^{m-r}$, $r = 1, 2, \ldots, m.$
2	$\tau\left(r; \theta^{(k)}\right) = \left[1 + \frac{1-\pi^{(k)}}{m\, \pi^{(k)}\, b\left(r; \xi^{(k)}\right)} \right]^{-1}$, $r = 1, 2, \ldots, m.$
3	$\overline{R}_n\left(\theta^{(k)}\right) = \frac{\sum_{r=1}^{m} r\, n_r\, \tau\left(r; \theta^{(k)}\right)}{\sum_{r=1}^{m} n_r\, \tau\left(r; \theta^{(k)}\right)}.$
4	$\pi^{(k+1)} = \frac{1}{n} \sum_{r=1}^{m} n_r\, \tau\left(r; \theta^{(k)}\right).$
5	$\xi^{(k+1)} = \frac{m - \overline{R}_n\left(\theta^{(k)}\right)}{m-1}.$
6	
7	$\theta^{(k+1)} = \left(\pi^{(k+1)},\ \xi^{(k+1)}\right)'.$
8	$l^{(k+1)} = log\, L\left(\theta^{(k+1)}\right).$
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \blacktriangleright k+1;\ go\,to\,1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\theta} = \theta^{(k+1)};\ stop. \end{cases}$

Table 2. E-M Algorithm of a MUB model with covariates for π.

$\boldsymbol{MUB\, Model\, with\, \pi = f\,(\beta; y)}$	
$\theta = (\beta\,', \xi)'$; $\epsilon = 10^{-6}$; $dim\,(\theta) = p_\pi + 2$.	
$l\,(\theta) = log\, L\,(\theta) = -\sum_{i=1}^{n} log\,\left(1 + e^{(-\boldsymbol{y}_i \beta)}\right) - \sum_{i=1}^{n} log\,\left(b\,(r_i; \xi) + \frac{e^{(-\boldsymbol{y}_i \beta)}}{m}\right)$.	
Steps	
0	$\theta^{(0)} = \left(\beta\,'^{(0)}, \xi^{(0)}\right)' = \left(0.1, \ldots, 0.1,\ \frac{m-\overline{R}_n}{m-1}\right)'$; $l^{(0)} = log\, L\,\left(\theta^{(0)}\right)$.
1	$b\,\left(r_i; \xi^{(k)}\right) = \binom{m-1}{r_i-1}\left(1-\xi^{(k)}\right)^{r_i-1}\left(\xi^{(k)}\right)^{m-r_i}$, $i = 1, 2, \ldots, n$.
2	$\pi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{y}_i{}'\beta^{(k)}}}$; $\tau\,\left(r_i; \theta^{(k)}\right) = \left[1 + \frac{e^{-\boldsymbol{y}_i{}'\beta^{(k)}}}{m\, b\left(r_i; \xi^{(k)}\right)}\right]^{-1}$, $i = 1, 2, \ldots, n$.
3	$\overline{R}_n\,\left(\theta^{(k)}\right) = \frac{\sum_{i=1}^{n} r_i\, \tau\left(r_i; \theta^{(k)}\right)}{\sum_{i=1}^{n} \tau\left(r_i; \theta^{(k)}\right)}$.
4	$Q_1\,\left(\beta^{(k)}\right) = -\sum_{i=1}^{n}\left\{log\left(1 + e^{-\boldsymbol{y}_i'\beta^{(k)}}\right) + \left(1 - \tau\,\left(r_i; \theta^{(k)}\right)\right) e^{-\boldsymbol{y}_i'\beta^{(k)}}\right\}$.
5	$\beta^{(k+1)} = \underset{\beta}{argmax}\, Q_1\,\left(\beta^{(k)}\right)$.
6	$\xi^{(k+1)} = \frac{m-\overline{R}_n\left(\theta^{(k)}\right)}{m-1}$.
7	$\theta^{(k+1)} = \left(\beta'^{(k+1)},\ \xi^{(k+1)}\right)'$.
8	$l^{(k+1)} = log\, L\,\left(\theta^{(k+1)}\right)$.
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \blacktriangleright k+1;\ go\, to\, 1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\theta} = \theta^{(k+1)};\ stop. \end{cases}$

Table 3. E-M Algorithm of a MUB model with covariates for ξ.

$\boldsymbol{MUB\ Model\ with\ \xi = g\left(\gamma; w\right)}$
$\theta = \left(\pi, \gamma\,'\right)'$; $\epsilon = 10^{-6}$; $dim\left(\theta\right) = p_\xi + 2.$
$l\left(\theta\right) = log\, L\left(\theta\right) = \sum_{i=1}^{n} log \left\{ \pi \left(\binom{m-1}{r_i-1} \frac{e^{\left(-\boldsymbol{w}_i\gamma\right)\left(r_i-1\right)}}{\left(1+e^{\left(-\boldsymbol{w}_i\gamma\right)}\right)^{m-1}} \right) - \frac{1}{m} \right\}$

Steps	
0	$\theta^{(0)} = \left(\pi^{(0)}, \gamma\,'^{(0)}\right)' = \left(\frac{1}{2}, 0.1, \ldots, 0.1\right)'$; $l^{(0)} = log\, L\left(\theta^{(0)}\right).$
1	$\xi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{w}_i'\gamma^{(k)}}}$; $b\left(r_i; \gamma^{(k)}\right) = \binom{m-1}{r_i-1} \frac{e^{-\left(r_i-1\right)\boldsymbol{w}_i\gamma^{(k)}}}{\left(1+e^{-\boldsymbol{w}_i\gamma^{(k)}}\right)^{m-1}}$, $i = 1, 2, \ldots, n.$
2	$\tau\left(r_i; \theta^{(k)}\right) = \left[1 + \frac{\pi^{(k)}}{m\left(1-\pi^{(k)}\right) b\left(r_i; \gamma^{(k)}\right)}\right]^{-1}$, $i = 1, 2, \ldots, n.$
3	
4	$\pi^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau\left(r_i; \theta^{(k)}\right).$
5	$Q_2\left(\gamma^{(k)}\right) = -\sum_{i=1}^{n} \tau\left(r_i; \theta^{(k)}\right) \left\{ \left(r_i - 1\right) \boldsymbol{w}_i \gamma^{(k)} + \left(m-1\right) log \left[1 + e^{-\boldsymbol{w}_i\gamma^{(k)}}\right] \right\}.$
6	$\gamma^{(k+1)} = \underset{\gamma}{argmax}\, Q_2\left(\gamma^{(k)}\right).$
7	$\theta^{(k+1)} = \left(\pi^{(k+1)},\ \gamma\,'^{(k+1)}\right)'.$
8	$l^{(k+1)} = log\, L\left(\theta^{(k+1)}\right).$
9	$\begin{cases} if\ l^{(k+1)} - l^{(k)} \geq \epsilon, & k \rightarrow k+1;\ go\, to\, 1; \\ if\ l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\theta} = \theta^{(k+1)};\ stop. \end{cases}$

Table 4. *E-M Algorithm for a general MUB model with covariates.*

$\boldsymbol{MUB\, Model\, with\, \pi = f\,(\beta; y)\,;\, \xi = g\,(\gamma; w)}$
$\theta = (\beta\,'; \gamma\,')'\,;\;\; \epsilon = 10^{-6};\; dim\,(\theta) = p_\pi + p_\xi + 2.$
$l\,(\theta) = log\, L\,(\theta) = \sum_{i=1}^{n} log \left\{ \frac{1}{1+e^{(-\boldsymbol{y}_i\beta)}} \left(\binom{m-1}{r_i-1} \frac{e^{(-\boldsymbol{w}_i\gamma)(r_i-1)}}{\left(1+e^{(-\boldsymbol{w}_i\gamma)}\right)^{m-1}} \right) - \frac{1}{m} \right\}$

Steps	
0	$\theta^{(0)} = \left(\beta\,'^{(0)}; \gamma\,'^{(0)}\right)' = (\,0.1, \ldots, 0.1;\; 0.1, \ldots, 0.1)'\,;\; l^{(0)} = log\, L\left(\theta^{(0)}\right).$
1	$\xi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{w}_i\gamma^{(k)}}};\;\; b\left(r_i; \gamma^{(k)}\right) = \binom{m-1}{r_i-1} \frac{e^{-\,(r_i-1)\,\boldsymbol{w}_i\gamma^{(k)}}}{\left(1+e^{-\boldsymbol{w}_i\gamma^{(k)}}\right)^{m-1}}\,,\; i = 1, 2, \ldots, n.$
2	$\pi_i^{(k)} = \frac{1}{1+e^{-\boldsymbol{y}_i\beta^{(k)}}};\;\; \tau\left(r_i; \theta^{(k)}\right) = \left[1 + \frac{e^{-\boldsymbol{y}_i\beta^{(k)}}}{m\, b\left(r_i; \xi^{(k)}\right)}\right]^{-1},\, i = 1, 2, \ldots, n.$
3	
4	$Q_1\left(\beta^{(k)}\right) = -\sum_{i=1}^{n} \left\{ log \left(1 + e^{-\boldsymbol{y}_i\beta^{(k)}}\right) + \left(1 - \tau\left(r_i; \theta^{(k)}\right)\right) e^{-\boldsymbol{y}_i\beta^{(k)}} \right\}.$
5	$Q_2\left(\gamma^{(k)}\right) = -\sum_{i=1}^{n} \tau\left(r_i; \theta^{(k)}\right) \left\{ (r_i - 1)\,\boldsymbol{w}_i\gamma^{(k)} + (m-1)\, log\, \left[1 + e^{-\boldsymbol{w}_i\gamma^{(k)}}\right] \right\}.$
6	$\beta^{(k+1)} = \underset{\beta}{argmax}\, Q_1\left(\beta^{(k)}\right);\;\; \gamma^{(k+1)} = \underset{\gamma}{argmax}\, Q_2\left(\gamma^{(k)}\right).$
7	$\theta^{(k+1)} = \left(\beta\,'^{(k+1)},\, \gamma\,'^{(k+1)}\right)'.$
8	$l^{(k+1)} = log\, L\left(\theta^{(k+1)}\right).$
9	$\begin{cases} if\; l^{(k+1)} - l^{(k)} \geq \epsilon, & k \blacktriangleright k+1;\; go\, to\, 1; \\ if\; l^{(k+1)} - l^{(k)} < \epsilon, & \hat{\theta} = \theta^{(k+1)};\; stop. \end{cases}$

Some words of comments may be useful for a correct interpretation of the previous steps:

i) in order to stress the correspondence among the four tables, we drew a blank line where the correspondent step does not apply;

ii) the initial values for the parameters in θ are derived by one of the following criteria: for π, we choose the midrange of the parameter space; for ξ, we choose the moment estimator; for the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ vectors we start from arbitrary[2] small values (e.g., 0.1);

iii) the conditioned average rank $\overline{R}_n\left(\theta^{(k)}\right)$ in Tables 1-2 is the average rank weighted with the *a posteriori* probability that each observed rank originates from the first component distribution $b\left(r;\xi\right)$;

iv) many proposals have been suggested for accelerating the convergence of the algorithm (McLachlan and Krishnan, 1997, pp. 70-73); in fact, the E-M procedure is generally slow, if compared with the second order convergence rates of the ML routines. However, in our experience (both for estimating models for real data sets and for running extensive simulations experiments), we never need such modifications[3].

4. Asymptotic standard errors of the estimators

In order to derive the asymptotic standard errors of the ML estimators, again we have to distinguish the MUB model without covariates from the models where one or both set of covariates are included.

In the first model (no covariates), the asymptotic standard errors can be obtained by exploiting a Rao (1973, pp. 367–368) proposal for grouped data. In fact, D'Elia and Piccolo (2003) proved that the asymptotic

[2]Of course, when some *a priori* information are available is better to start from more definite initial values.

[3]In fact, Piccolo (2003) showed that moment estimates of the two parameters are convenient starting values for accelerating the convergence of the E-M algorithm.

variance-covariance matrix of the ML estimators $\hat{\theta}$ is:

$$\boldsymbol{V} = \frac{1}{n} \begin{pmatrix} d_{\pi\pi} & d_{\pi\xi} \\ d_{\pi\xi} & d_{\xi\xi} \end{pmatrix}^{-1},$$

where:

$$\begin{aligned} d_{\pi\pi} &= \sum_{r=1}^{m} \frac{\left\{\frac{\partial\, p(r;\theta)}{\partial\, \pi}\right\}^2}{p\,(r;\theta)}; \\ d_{\xi\xi} &= \sum_{r=1}^{m} \frac{\left\{\frac{\partial\, p(r;\theta)}{\partial\, \xi}\right\}^2}{p\,(r;\theta)}; \\ d_{\pi\xi} &= \sum_{r=1}^{m} \frac{\left(\frac{\partial\, p(r;\theta)}{\partial\, \pi}\right)\left(\frac{\partial\, p(r;\theta)}{\partial\, \xi}\right)}{p\,(r;\theta)}. \end{aligned}$$

Exploiting the recursive nature of the Binomial distribution, these quantities can be easily obtained if one observes that:

$$\begin{aligned} \frac{\partial\, p\,(r;\theta)}{\partial\, \pi} &= b\,(r;\xi) - \frac{1}{m}; \\ \frac{\partial\, p\,(r;\theta)}{\partial\, \xi} &= \pi\, b\,(r;\xi)\, \frac{m - \xi\,(m-1) - r}{\xi\,(1-\xi)}. \end{aligned}$$

Further simplifications are possible for saving computing time.

When the MUB model includes one or two sets of covariates, then D'Elia (2003b) proved that the asymptotic variance-covariance of the ML estimators $\hat{\theta}$ may be conveniently estimated by the observed information matrix. In fact, thanks to the nature of the complete log-likelihood function, it is possible to separate the effect of the uncertainty parameter (π or the corresponding $\boldsymbol{\beta}$ parameters vector) from the preference parameter (ξ or the corresponding $\boldsymbol{\gamma}$ parameters vector). As a consequence, the asymptotic variance-covariance matrix of the ML estimators is block diagonal and thus the π estimator (or $\boldsymbol{\beta}$ estimators) and the ξ estimator (or $\boldsymbol{\gamma}$ estimators) are asymptotically jointly Gaussian and independent.

Specifically, relying on the quoted D'Elia (2003b) results, it is immediate to derive the corresponding asymptotic formulae:

- For the *MUB model with covariates for* π, the asymptotic variance-covariance matrix $\boldsymbol{V}$ of the ML estimators is:

$$\boldsymbol{V} = -\begin{pmatrix} \frac{1}{d_{\xi\xi}} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{B}^{-1} \end{pmatrix}$$

where:

$$d_{\xi\xi} = \left(\frac{1}{(1-\xi)^2} - \frac{m}{\xi^2}\right)\sum_{i=1}^{n} \tau_i + \frac{1-2\xi}{(1-\xi)^2\,\xi^2}\sum_{i=1}^{n} r_i\,\tau_i;$$

and the elements of the $\boldsymbol{B}$ matrix are defined by:

$$\{\boldsymbol{B}\}_{hj} = \frac{\partial^2}{\partial\beta_h\,\partial\beta_j} Q_1(\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n} y_{ih} y_{ij}\left[1+\cosh(\boldsymbol{y}_i\boldsymbol{\beta})\right]^{-1};$$

while the $Q_1(\boldsymbol{\beta})$ function is defined by the step 4 in Table 2.

- For the *MUB model with covariates for* ξ, the asymptotic variance-covariance matrix $\boldsymbol{V}$ of the ML estimators is:

$$\boldsymbol{V} = -\begin{pmatrix} \frac{1}{d_{\pi\pi}} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}^{-1} \end{pmatrix}$$

where:

$$d_{\pi\pi} = -\frac{n}{\pi(1-\pi)};$$

and the elements of the $\boldsymbol{\Gamma}$ matrix are defined by:

$$\{\boldsymbol{\Gamma}\}_{hj} = \frac{\partial^2}{\partial\gamma_h\,\partial\gamma_j} Q_2(\boldsymbol{\gamma}) = -\frac{1}{2}\sum_{i=1}^{n} \tau_i\, w_{ih} w_{ij}\left[1+\cosh(\boldsymbol{w}_i\boldsymbol{\gamma})\right]^{-1};$$

while the $Q_2(\boldsymbol{\gamma})$ function is defined by the step 5 in Table 3.

- For the *general MUB model with covariates both for* π *and* ξ, the asymptotic variance-covariance matrix $\boldsymbol{V}$ of the ML estimators is:

$$\boldsymbol{V} = -\begin{pmatrix} \boldsymbol{B}^{-1} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}^{-1} \end{pmatrix}$$

where the elements of the $\boldsymbol{B}$, $\boldsymbol{\Gamma}$ matrices are defined respectively by:

$$\{\boldsymbol{B}\}_{hj} = \frac{\partial^2}{\partial\beta_h\,\partial\beta_j}\,Q_1\,(\boldsymbol{\beta}) = -\frac{1}{2}\,\sum_{i=1}^{n} y_{ih}y_{ij}\,[1+\cosh\,(\boldsymbol{x}_i\boldsymbol{\beta})]^{-1}\,;$$

$$\{\boldsymbol{\Gamma}\}_{hj} = \frac{\partial^2}{\partial\gamma_h\,\partial\gamma_j}\,Q_2\,(\boldsymbol{\gamma}) = -\frac{1}{2}\,\sum_{i=1}^{n} \tau_i\,w_{ih}w_{ij}\,[1+\cosh\,(\boldsymbol{w}_i\boldsymbol{\gamma})]^{-1}\,;$$

and the $Q_1\,(\boldsymbol{\beta})$, $Q_2\,(\boldsymbol{\gamma})$ functions are defined by the steps 4–5 in Table 4.

5. Some empirical evidences

In this section we applied the previous algorithms to a real data set consisting of the preferences towards $m = 12$ colors expressed by a sample of $n = 169$ young people. No ties were allowed.

We choose to compare the ranks given for the *Pink color* as modelled by the proposed four MUB models, without and with the covariate Sex[4] ($= 0$, for Masculine; $= 1$, for Feminine).

These models will be denoted by MUB–00 (no covariates), MUB–10 (covariates only for π), MUB–01 (covariates only for ξ), MUB–11 (covariates for both π and ξ), respectively. Similarly, the values of their corresponding log-likelihood functions $l(\boldsymbol{\theta})$ computed at maximum will be denoted by $l_{00},\ l_{10},\ l_{01},\ l_{11}$, respectively.

The results –reported in Table 5– show that all the parameters are significant (standard errors are in parentheses). The values of the corresponding parameters conditioned to the covariate values ($= 0, 1,$ respectively) are also reported.

[4] In this case-study we found significant the effect of the same covariate on the raters' preferences. Of course, this situation is not strictly necessary, since the E-M algorithm presented in Table 4 allows for completely general set of covariates.

From the discussion of section 2, we deduce Table 6 where the implied distribution probabilities for the ranks expressed by the subjects towards the Pink color (for each of the four estimated models) are shown.

Table 5. Estimated MUB models for the Pink color.

$Models$	$Parameters\,estimates$		$l\,(\boldsymbol{\theta})$
$\mathbf{1.}\,\boldsymbol{MUB\,(\pi,\xi)}$	$\pi = 0.481\ (0.063)$	$\xi = 0.180\ (0.018)$	-398.995
$\mathbf{2.}\,\boldsymbol{MUB\,(\beta,\xi)}$	$\beta_0 = \quad 0.586\ (0.244)$	$\xi = 0.171\ (0.013)$	-397.082
	$\beta_1 = -1.171\ (0.324)$		
$Condit.\,estim.$	$(\pi \mid x_{i1} = 0) = 0.642$		
	$(\pi \mid x_{i1} = 1) = 0.358$		
$\mathbf{3.}\,\boldsymbol{MUB\,(\pi,\gamma)}$	$\pi = 0.498\ (0.038)$	$\gamma_0 = -1.831\ (0.141)$	-397.119
		$\gamma_1 = \quad 0.722\ (0.175)$	
$Condit.\,estim.$		$(\xi \mid x_{i1} = 0) = 0.138$	
		$(\xi \mid x_{i1} = 1) = 0.248$	
$\mathbf{4.}\,\boldsymbol{MUB\,(\beta,\gamma)}$	$\beta_0 = \quad 0.506\ (0.242)$	$\gamma_0 = -1.678\ (0.123)$	-396.246
	$\beta_1 = -0.853\ (0.318)$	$\gamma_1 = \quad 0.548\ (0.166)$	
$Condit.\,estim.$	$(\pi \mid x_{i1} = 0) = 0.624$	$(\xi \mid x_{i1} = 0) = 0.157$	
	$(\pi \mid x_{i1} = 1) = 0.414$	$(\xi \mid x_{i1} = 1) = 0.244$	

From Table 6, it is possible to assess that:

a) the *uncertainty share* (*unc*) is systematically higher for the women in all the models (except for the MUB–01 model where π is unchanged), thus explaining the larger variability of their choices;

b) the *impact coefficient* (*imp*) increases dramatically in moving from men to women (when we model preference by mean of the covariate Sex, as in the MUB–01 and MUB–11 models);

c) the '*disliking odds*' (*dis*) are in the ratio of about $2:1$ for the men with respect to the women[5].

Finally, we report in the last column of Table 6 the probability of a strong disliking towards the Pink color as measured by $Pr(R \geq 9)$; this quantity is the probability to put the Pink color in the worst quarter of the rank preferences distribution. Again, the results support the expected values. Indeed, in the data set, the relative frequencies of men and women that assigned a rank greater or equal to 9 are 0.671 and 0.542, respectively; the corresponding values predicted by the general MUB model are 0.699 and 0.497, that are almost coincident with the empirical ones, given the sampling variability.

Table 6. A comparison among the estimated MUB models.

MUB Models	*unc*	*imp*$\times 10^{12}$	*dis*	$Pr(R \geq 9)$
$[00]$ $no\,covariates$	0.04323	684	4.55185	0.596
$[10]$ $cov.\,for\,\pi\,(Sex=0)$	0.02980	486	4.84659	0.695
$[10]$ $cov.\,for\,\pi\,(Sex=1)$	0.05351	271	4.84659	0.535
$[01]$ $cov.\,for\,\xi\,(Sex=0)$	0.04181	28	6.24271	0.639
$[01]$ $cov.\,for\,\xi\,(Sex=1)$	0.04181	35861	3.03226	0.525
$[11]$ $cov.\,for\,\pi,\xi\,(Sex=0)$	0.03135	172	5.35042	0.699
$[11]$ $cov.\,for\,\pi,\xi\,(Sex=1)$	0.04882	24742	3.09366	0.497

Figure 1 shows the estimated MUB probability distribution as implied by the general MUB model where both the parameters are explained by the covariate Sex. The probability distributions of the four estimated MUB models are reported in the Figure 2. As a whole, they enhance both the $\pi-$role, in supporting the uncertainty of the elicitation process, and the $\xi-$role, in supporting the subjects' preferences.

From these evidences, some points deserve a discussion:

i) there is a substantial disliking towards the Pink color but it is sharper for the men (supported by a stronger asymmetry with a mode at

[5] Some caution should be used in order to avoid a strict interpretation of *dis*. Anyway, it is a useful quantity to compare models for sub-samples, as in this case.

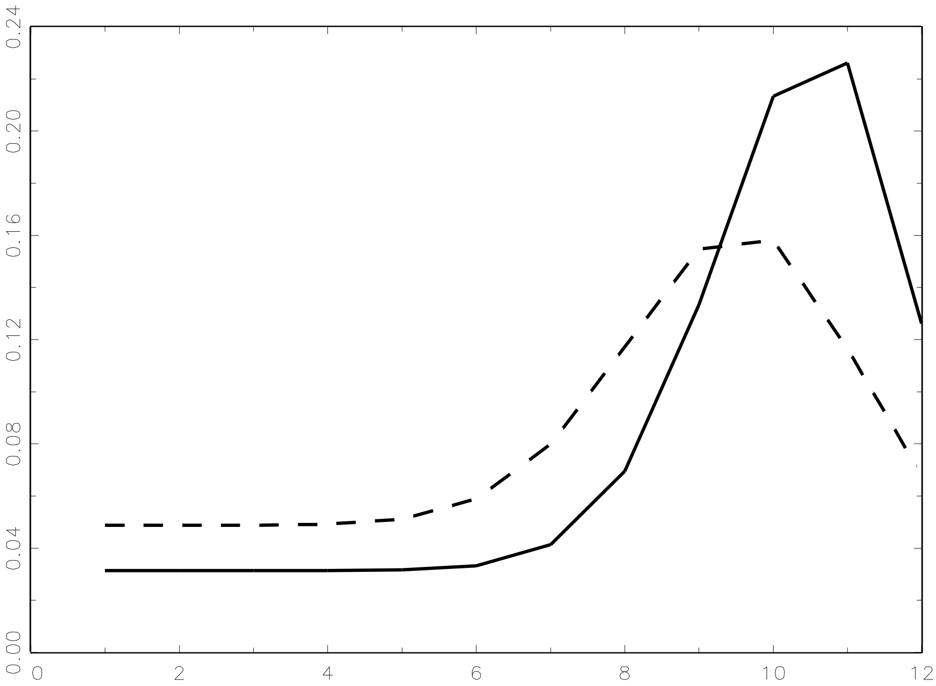

Figure 1. Rank preference distributions for Males (——) and Feminine (- - -) towards the Pink color.

$R = 11$) than for the women (moderate asymmetry with a mode at $R = 10$). These points are reflected in the estimated $\hat{\xi}$ parameters (Table 5), that are 0.157 and 0.244, respectively;

ii) it seems evident that the preferences are assessed in a more neat way for the men since $\hat{\pi}$ is significantly higher for the men than for the women (0.624 and 0.414, respectively). In fact, the *uncertainty shares*, as measured by $(1 - \hat{\pi})/m$, are estimated by 3.13% and 4.88%, respectively.

As a consequence, the women preferences distribution is less right-shifted, with a higher minimum level. On the contrary, the men distribution is clearly peaked upon high values of the ranks, denoting a strong disliking towards the Pink color.

From an inferential point of view, it is worth to note that the inclusion of covariates for both the parameters seems an important issue for improving the rank modelling of our data. In fact, as Table 7 confirms, the deviance is almost significant if we compare estimated MUB models with and without covariates[6].

Table 7. Deviances of the estimated MUB models for the rank preferences towards the Pink color.

Comparisons	*Deviance*	g	$\chi^2_{(g;0.05)}$
l_{10} vs. l_{00}	3.82490	3-2=1	3.841
l_{01} vs. l_{00}	3.75298	3-2=1	3.841
l_{11} vs. l_{00}	5.49882	4-2=2	6.635

As a further result, in Table 8 we present the (theoretical) expectations derived from the corresponding estimated MUB models. These values are to be compared with the averages of 8.775 (for the aggregate set), 9.233 (for the men) and 8.427 (for the women), respectively.

Table 8. Expectations for the estimated MUB models.

MUB models	*Expectations*	
MUB–00 (no-covariates)	8.193 (aggregate)	
	Masculine	*Feminine*
MUB–10 (covariates for π)	8.824	7.795
MUB–01(covariates for ξ)	8.484	7.881
MUB–11(covariates for π, ξ)	8.851	7.665

Some bias emerges from these results[7]; however, given the sample

[6]Of course, we are assuming that the asymptotic theory of likelihood ratio tests applies, that is: $-2\,(log\,(\boldsymbol{\theta}_0) - log\,(\boldsymbol{\theta}_1)) \stackrel{a}{\sim} \chi^2_{(g;0.05)}$, where g is the difference of dimensions between the vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$.

[7]Empirical averages are larger than the expectations: this kind of bias is mainly caused by a mode at $R = 12$ in the distribution of the observed ranks (more pronounced for the men than for the women) that increases the average ranks. In fact, the estimated MUB models imply a mode around $R = 10, 11$. In these situations it is worth to consider also the IHG models as discussed by D'Elia (1999, 2003a).

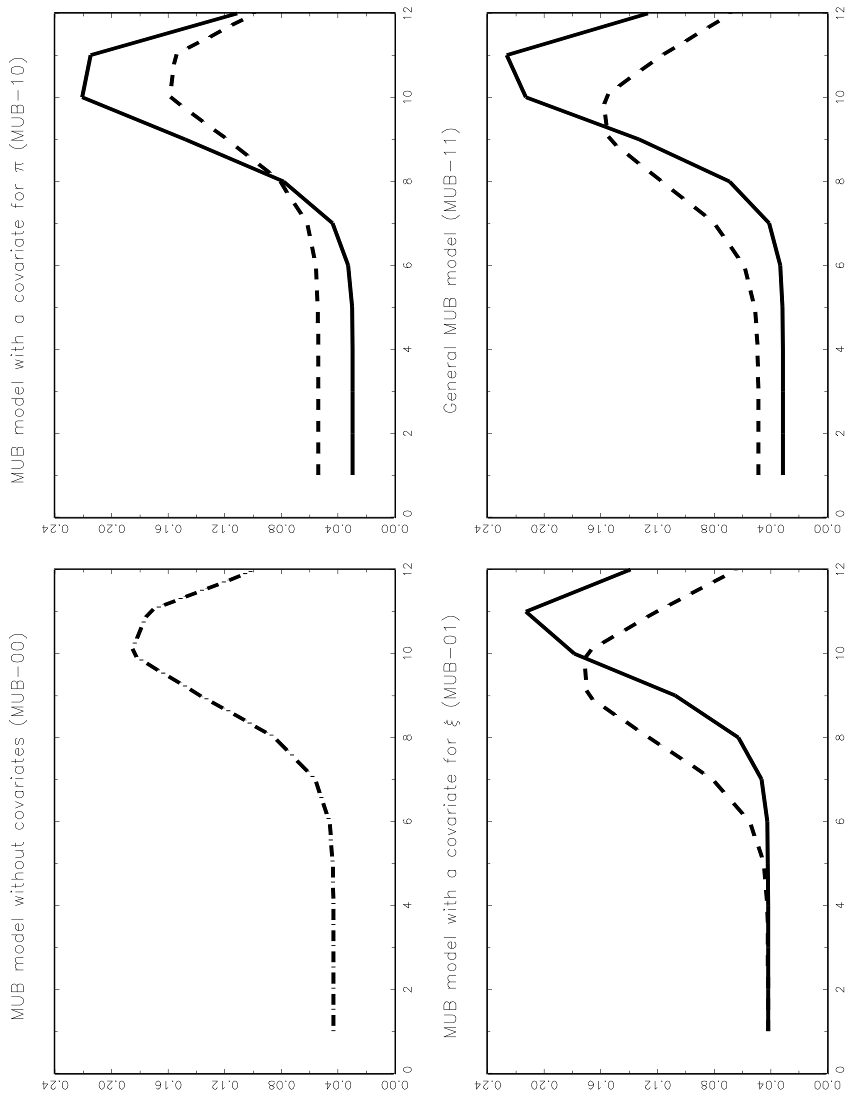

Figure 2. Probability distributions of the rank preferences towards the Pink color as implied by different estimated MUB models: Aggregate (.-.-.-.-.), Males (——) and Feminine (- - -) .

size, the differences of the sample averages and of the theoretical expectations are always in the same directions for both the genders .

Finally, the reported experience and more extensive experiments (here not discussed) support the idea that the MUB modelling with covariates is an approach valuable for catching the relevant preferences measures expressed by the subjects, also in the special cases of moderate sample sizes and dichotomous variables.

6. Concluding remarks

All the steps reported in the sections 3-4 have been translated into the Gauss© programming language and the speed performance has been checked on two PCs-Pentium.

The experiment[8] was performed by estimating a general MUB model with covariates (MUB-11) on the following data sets:

Data set 1: $n = 169$; Pink color preferences fitted to a MUB model where both parameters are expressed as logistic functions of the covariate Sex;

Data set 2: $n = 169$; Brown color preferences fitted to a MUB model where both parameters are expressed as logistic functions of the covariates Smoking and Sex, respectively;

Data set 3: $n = 1000$; Simulated rank data fitted to a MUB model where both parameters are expressed as logistic functions of two coincident and balanced dummy variables;

[8] The reported times include the E-M procedure until convergence and also the computations necessary for getting the asymptotic standard errors of the ML estimators. We used throughout the experiments a Gauss© DOS version 3.2.23. The new Gauss© 5.0, Windows version, reduced time significantly. For instance, on the first PC, the 16.7 seconds for Data set 3 become 12.9 seconds, with a reduction factor of 23%.

Although the elapsed time depends on the chosen data, Table 9 should give some idea of the efficiency of the implemented algorithm in different situations.

Table 9. Computing times (in seconds) for different data sets and PCs.

PCs (256 MB–RAM)	Data set 1	Data set 2	Data set 3
Pentium III (700Mhz)	2.47	3.52	16.70
Pentium IV (754Mhz)	1.56	2.65	12.79

As a concluding remark, we observe that the statistical models discussed in this paper seem consistent with the empirical evidences, although the use of rigorous asymptotic fitting measures is not adequate here.

In fact, notwithstanding the moderate sample size (we discussed of the preferences expressed by 96 women and 73 men), the results strongly support the usefulness of the general MUB model for enhancing the main features of one predictor variable (in our case, the gender of the subject) and for explaining substantial differences in the liking/disliking choices. Moreover, the applied algorithms are efficient procedures from a computational point of view.

Finally, the future steps in the MUB modelling approach with covariates should include some fitting measure of the adequacy of the model and the derivation of asymptotic tests for inferring about the parameters differences in selected sub-samples.

Acknowledgments: We thank the support received by the Dipartimento di Scienze Statistiche, University of Naples Federico II and the research structures offered by the Centro per la Formazione in Economia e Politica dello Sviluppo Rurale, Portici. We gratefully acknowledge some useful referees' suggestions that improved a preliminary version of the paper.

References

D'Elia A. (1999), A proposal for ranks statistical modelling, in: H. Friedl, A. Berghold, G. Kauermann (eds.), *Statistical Modelling*, Graz-Austria, 468-471.

D'Elia A. (2003a), Modelling ranks using the Inverse Hypergeometric distribution, *Statistical Modelling: an International Journal*, **3** (1), 65-78.

D'Elia A. (2003b), A mixture model with covariates for ranks data: some inferential developments, *Quaderni di Statistica*, **5**, *this issue*.

D'Elia A. and Piccolo D. (2003), A mixture model for preference data analysis, *submitted*.

Fligner M.A. and Verducci J. S. (1999), *Probability Models and Statistical Analyses of Ranking Data*, Springer-Verlag, New York.

McLachlan G., Krishnan G. J. (1997), *The EM Algorithm and Extensions*, J. Wiley & Sons, New York.

McLachlan G., Peel G. J. (2000), *Finite Mixture Models*, J. Wiley & Sons, New York.

Marden J. I. (1995), *Analyzing and Modelling Rank Data*, Chapman & Hall, London.

Piccolo D. (2003), A note on the Moments of a Mixture of Uniform and Shifted Binomial random variables, *Quaderni di Statistica*, **5**, *this issue*.

Rao C. R. (1973), *Linear Statistical Inference and its Applications*, 2^{nd} edition, J. Wiley & Sons, New York.