Quaderni di Statistica
Vol. 5, 2003

A mixture model with covariates for ranks data: some inferential developments

Angela D'Elia
Dipartimento di Scienze Statistiche
Università degli Studi di Napoli Federico II
E-mail: angela.delia@unina.it

Summary: The paper deals with a mixture model for ranks data with covariates. Special attention is devoted to the inferential and computational issues related with the main steps of an E-M algorithm for the estimation of the model's parameters. Effective expressions for computing asymptotic standard errors are also derived. Evidence from real data highlights the main results and supports the usefulness of the proposed approach.

Keywords: Covariates, E-M algorithm, Mixture model, Ranks data.

1. Introduction

The elicitation mechanism by means of which a rater expresses his/her preferences ranking towards *m* different items can be thought of as a composite procedure: indeed, the rank assigned to a given item results from the liking/disliking feeling for it, and depends also on the uncertainty of the ranking process itself. Of course, we expect that the uncertainty becomes greater for those items that do not excite strong liking or disliking feelings.

Thus, a convincing framework for modelling ranking procedures has to take into account both the selection mechanism and the uncertainty related to it. As a consequence, D'Elia and Piccolo (2003) proposed a Mixture model of a discrete Uniform and a shifted Binomial random variables (henceforth called MUB): here, the Uniform component describes the degree of uncertainty in the preferences, while the shifted Binomial random variable represents the behavior of the rater with respect to the liking/disliking feeling.

In this paper we extend the MUB model, allowing the inclusion of raters' covariates in the model's specification. In this way, (following the same strategy developed in other models for ranks data with covariates, as in D'Elia, 1999, 2000, 2003), we are able to link the main features of the raters (e.g. Sex, Age, Profession, etc.) to the rank they assigned to a given item; in particular, both the preference and the uncertainty feeling can be explained by means of subjects' specific covariates, yielding a deeper insight in the preferences data analysis.

The paper is organized as follows. In section 2, we recall the MUB framework, and we introduce some notation to be used throughout the paper. Then, in sections 3 and 4, we explicitly develop the MUB model with covariates, in order to explain how the degree of preference and the uncertainty feeling change with the covariates, respectively. Particular attention is paid to the computational aspects (e.g. the E-M algorithm for obtaining the maximum likelihood estimates of the parameters, in subsections 3.2 and 4.2); moreover, effective expressions for computing their asymptotic standard errors are also given (subsections 3.3 and 4.3). Then, several examples raised by real datasets are exploited (section 5) in order to highlight the potentialities of including the covariates in the MUB model. Final considerations about further developments on this issue end the paper.

2. The MUB model

Let *m* be the fixed and known number of items, and *r* be the rank assigned to a given item among *m*; in the following, we assume that $r = 1$ means "most preferred", while $r = m$ means "least preferred".

D'Elia and Piccolo (2003) consider *r* as an observed realization of the random variable $R \sim MUB(m, \pi, \xi)$ if:

$$Pr(R = r) = \pi P_B(r) + (1 - \pi) P_U(r), \qquad r = 1, 2, \ldots, m,$$

where:

$$P_B(r) = \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r}; \qquad P_U(r) = \frac{1}{m};$$

and $\pi \in [0,1]$; $\xi \in [0,1]$.

Then, we get:

$$Pr(R=1) = \pi \xi^{m-1} + \frac{1-\pi}{m};$$

and

$$E(R) = \pi(m-1)\left(\frac{1}{2} - \xi\right) + \frac{m+1}{2}. \tag{1}$$

For a fixed value of π, when the parameter ξ increases then $Pr(R=1)$ increases too, while $E(R)$ decreases: thus, ξ may be considered as a proxy of a *preference measure*. Indeed, $E(R) \geq (m+1)/2$ (that is, the midrange) if $\xi \leq 1/2$ (and, viceversa, $E(R) < (m+1)/2$ if $\xi > 1/2$): thus, a rater tends to a disliking or liking feeling - with respect to the central rank location $(m+1)/2$ - as long as ξ moves from 1/2 towards 0 or 1, respectively.

On the other hand, for a fixed value of ξ, the π parameter is inversely related to the uncertainty in the preferences, so that $(1-\pi)/m$ represents the *uncertainty share* spread out over the *m* items.

Even if we introduced the MUB model as a mixture for representing two aspects of the ranking procedures, of course it can be also seen as a mixture of two different populations of raters: the first (with probability π) is made of people that give ranks by means of a paired comparisons criterion (Bradley and Terry, 1952; and more recently: D'Elia, 2000); the second (with probability $1-\pi$) consists in people that give ranks with the maximum uncertainty attainable by a discrete random variable on the $\{1, 2, \ldots, m\}$ support. Anyway, also this interpretation confirms the role of ξ as a liking parameter and of π as a measure of uncertainty.

In order to include the *n* raters' covariates inside this framework, we need to link the parameters ξ and π to the $(p+1)$-length vector of a subject's specific covariates $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})'$, for $i = 1, 2, \ldots, n$; as it is standard in these contexts, the unit is included to represent a baseline effect.

Although we might need a model where both the parameters depend upon the covariates, in this paper we are going to develop a first step of a more complex framework, modelling a parameter at a time. In fact, we think that this preliminary step is meaningful in order to better understand the role of π and ξ.

To do that, we have specified two different structures:

- a model where π is kept fixed and only ξ depends upon the covariates (section 3);
- a model where ξ is fixed and only π is a function of the covariates (section 4).

It should be noticed that in the first case we are going to explain how the degree of preference changes with the covariates, while in the second situation we will be interested in studying their effect on the uncertainty feeling.

3. The MUB model with a subject specific ξ

In this section we develop an extension of the MUB model allowing the ξ parameter to depend upon the raters' covariates. Then, in subsection 3.2 we show the steps of the E-M algorithm for obtaining maximum likelihood estimates of the parameters (that is, π and the covariates coefficients), while in subsection 3.3 the expressions for computing their asymptotic standard errors are explicitly derived.

3.1 Model specification

Letting the ξ parameter to depend upon the raters' covariates means that we are considering a situation where the liking or disliking feeling towards a given item can be explained on the basis of some features of the raters; instead, throughout this section, the uncertainty (as measured by $1-\pi$) is not assumed to depend on the covariates.

In order to link the features of the *n* raters (as expressed in the $n \times p+1$ design matrix **X**) to the preference parameter ξ, we borrow the link function of a logistic regression model (McCullagh and Nelder, 1989), and then we let:

$$(\xi \mid \mathbf{X} = \mathbf{x}_i) = \frac{1}{1 + exp(-\mathbf{x}_i\gamma)}, \qquad i = 1, 2, ..., n,$$

where $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_p)'$ is the covariates coefficients vector. Of course, the rationale for this stems from the fact that the logistic function yields a smooth mapping from the entire real space to the interval [0, 1], as it is proper for ξ. Indeed, it can be easily seen that when the linear predictor $\mathbf{x}_i\gamma \to \infty$, then $(\xi \mid \mathbf{X} = \mathbf{x}_i) \to 1$; viceversa, when $\mathbf{x}_i\gamma \to -\infty$, then $(\xi \mid \mathbf{X} = \mathbf{x}_i) \to 0$.

Moreover, from (1) it results that:

$$E(R \mid \mathbf{X} = \mathbf{x}_i) = \frac{\pi(m-1)}{2} - \pi(m-1)\frac{1}{1 + exp(-\mathbf{x}_i\gamma)} + \left(\frac{m+1}{2}\right).$$

Without loosing in generality, if we let $\mathbf{x}_i = (1, x_{i1})'$, then $\mathbf{x}_i\gamma = \gamma_0 + \gamma_1 x_{i1}$, and we get:

$$E(R \mid \mathbf{X} = \mathbf{x}_i) = \frac{\pi(m-1)}{2} - \pi(m-1)\frac{1}{1 + exp(-\gamma_0 - \gamma_1 x_{i1})} + \left(\frac{m+1}{2}\right).$$

Thus, letting $\mathbf{u} = (0, 1)'$, it can be shown that:

$$E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) =$$

$$= \pi(m-1)\left[\frac{e^{-\gamma_0 - \gamma_1 x_{i1}}(e^{-\gamma_1} - 1)}{(1 + e^{-\gamma_0 - \gamma_1 x_{i1}})(1 + e^{-\gamma_0 - \gamma_1 x_{i1} - \gamma_1})}\right],$$

from which it results that $E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) > 0$ when $\gamma_1 < 0$.

This means that for a negative γ_1 estimate there is an increase in $E(R)$, which means a worsening in the preference feeling when the covariate x_{i1} augments; it happens viceversa for a positive γ_1 estimate.

The above results can be easily extended to any number $p > 1$ of covariates, leading to the following scheme, valid for a generic x_{ij} ($j = 1, 2, \ldots, p$), *ceteris paribus*:

- $\gamma_j < 0 \Rightarrow$ preference worsening, when x_{ij} increases;
- $\gamma_j = 0 \Rightarrow$ no effect on the preference;
- $\gamma_j > 0 \Rightarrow$ preference improvement, when x_{ij} increases.

3.2 The likelihood function and the E-M algorithm

Given a vector of observed independent ranks $\mathbf{r} = (r_1, r_2, \ldots, r_n)'$ for a pre-fixed item, and a $(n \times p + 1)$ design matrix $\mathbf{X}$, the log-likelihood function for the MUB model, using the above introduced specification for the parameter ξ, becomes:

$$log\, L(\gamma, \pi; \mathbf{r}, \mathbf{X}) =$$

$$= \sum_{i=1}^{n} log \left\{ \pi \left[\binom{m-1}{r_i-1} e^{-\mathbf{x}_i\gamma(r_i-1)} \frac{1}{(1+e^{-\mathbf{x}_i\gamma})^{m-1}} - \frac{1}{m} \right] + \frac{1}{m} \right\}.$$

In order to get the maximum likelihood estimates of both π and the coefficient vector γ, we rely on the E-M algorithm, whose effectiveness in estimating the parameters of mixture models has been widely proved in the literature (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000). Indeed, the rationale for the use of the E-M algorithm in fitting mixture models is that we can consider the observed data as incomplete, since the appropriate mixture component for each subject is unknown. As a result, the complete likelihood function factorizes in two separate components, allowing easy derivatives and computations, as it will be shown in the following.

The E-M algorithm involves the iteration until convergence of the following two steps (for any iteration $k = 0, 1, 2, \ldots$):

- *E-step*: compute the conditional expectation of the complete data log-likelihood, using the current parameters estimate; that is, in our case:

$$Q(\pi^{(k)}, \gamma^{(k)}) = \sum_{i=1}^{n} \left\{ \tau_i^{(k)} log(\pi^{(k)}) + (1 - \tau_i^{(k)}) log(1 - \pi^{(k)}) \right\} +$$

$$+ \sum_{i=1}^{n} \left\{ \tau_i^{(k)} log(P_B(r_i; \gamma^{(k)})) + (1 - \tau_i^{(k)}) log(1/m) \right\}.$$

In this expression $\pi^{(k)}$ and $\gamma^{(k)}$ denote the k-th iteration estimates of the parameters; then,

$$\tau_i^{(k)} = \left[1 + \frac{1-\pi^{(k)}}{m\pi^{(k)} P_B(r_i; \gamma^{(k)})}\right]^{-1}$$

is the estimated posterior probability that R_i belongs to the shifted Binomial component of the mixture model; and $P_B(r_i; \gamma^{(k)}) = Pr(R = r_i)$ is expressed as a function of the linear predictor $\mathbf{x}_i\gamma$.

- *M-step*: maximize the $Q(\pi^{(k)}, \gamma^{(k)})$ function in order to get updated estimates of the parameters: $\pi^{(k+1)}, \gamma^{(k+1)}$.

In fact, at a fixed k-th iteration, we get:

$$Q(\pi^{(k)}, \gamma^{(k)}) = Q_1(\pi^{(k)}) + Q_2(\gamma^{(k)}) + constant,$$

where we let:

$$Q_1(\pi^{(k)}) = \sum_{i=1}^{n} \left\{ \tau_i^{(k)} log(\pi) + (1-\tau_i^{(k)}) log(1-\pi) \right\};$$

$$Q_2(\gamma^{(k)}) = \sum_{i=1}^{n} \left\{ \tau_i^{(k)} log(P_B(r_i; \gamma^{(k)})) \right\} =$$

$$= -\sum_{i=1}^{n} \tau_i^{(k)} \{ (r_i - 1)\mathbf{x}_i\gamma^{(k)} + (m-1) log(1 + e^{-\mathbf{x}_i\gamma^{(k)}}) \}.$$

Thus, maximizing $Q_1(\pi^{(k)})$ with respect to π yields:

$$\pi^{(k+1)} = \frac{\sum_{i=1}^{n} \tau_i^{(k)}}{n},$$

while $\gamma^{(k+1)}$ is obtained by finding the maximum of $Q_2(\gamma^{(k)})$.

In detail, the implementation of the E-M algorithm for the MUB model with ξ depending on covariates is developed as follows:

1. Compute $logL(\pi^{(0)}, \gamma^{(0)}\mathbf{r}, \mathbf{X})$, using starting values $\pi^{(0)}$, $\gamma^{(0)}$ for initializing the algorithm;

2. $P_B(r_i;\gamma^{(k)}) = \begin{pmatrix} m-1 \\ r_i - 1 \end{pmatrix} \frac{\left[e^{-\mathbf{x}_i\gamma^{(k)}}\right]^{r_i-1}}{\left[1+e^{-\mathbf{x}_i\gamma^{(k)}}\right]^{m-1}};$ $\qquad i = 1, 2, \ldots, n;$

3. $\tau_i^{(k)} = \left[1 + \frac{1-\pi^{(k)}}{m\pi^{(k)} P_B(r_i;\gamma^{(k)})}\right]^{-1},$ $\qquad i = 1, 2, \ldots, n;$

4. $\pi^{(k+1)} = \frac{\sum_{i=1}^{n} \tau_i^{(k)}}{n};$ maximize $Q_2(\gamma^{(k)})$ for getting $\gamma^{(k+1)}$;

5. compute $logL(\pi^{(k+1)}, \gamma^{(k+1)}; \mathbf{r}, \mathbf{X})$;

6. if $L(\pi^{(k+1)}, \gamma^{(k+1)}) - L(\pi^{(k)}, \gamma^{(k)}) < \epsilon$ stop;
   else $k = k + 1$, and reiterate from 2.

3.3 Computing standard errors

In order to estimate the asymptotic covariance matrix of the maximum likelihood estimators we rely on the inverse of the observed information matrix: $I(\hat{\pi}, \hat{\gamma}; \mathbf{r})$, that is the negative of the Hessian of the log-likelihood function evaluated at $\pi = \hat{\pi}, \gamma = \hat{\gamma}$.

Now, the observed information matrix for the observed (incomplete) data can be computed in terms of the conditional moments of the gradient and curvature of the complete-data log-likelihood function obtained in the E-M algorithm (McLachlan and Peel, 2000, pp. 64-66; Pawitan, 341-362). In particular, for i.i.d. observations we can approximate $I(\hat{\pi}, \hat{\gamma}; \mathbf{r})$ by taking the second-order partial derivatives of the conditional expectation of the complete data log-likelihood:

$$Q(\pi, \gamma) = Q_1(\pi) + Q_2(\gamma) + constant$$

.

Then, it results that:

$$\mathbf{V} \simeq -\begin{bmatrix} d_{\pi\pi} & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}^{-1} = -\begin{bmatrix} d_{\pi\pi}^{-1} & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix},$$

where

$$d_{\pi\pi} = \frac{\partial^2}{\partial \pi^2} Q_1(\pi) = -\sum_{i=1}^{n} \frac{\tau_i}{\pi^2} - \sum_{i=1}^{n} \frac{1-\tau_i}{(1-\pi)^2} = \frac{-n}{\pi(1-\pi)},$$

and $\mathbf{\Gamma}$ is a $(p+1 \times p+1)$ matrix with generic element:

$$\{\mathbf{\Gamma}\}_{hj} = \frac{\partial^2}{\partial\gamma_h\partial\gamma_j} Q_2(\gamma) = -(m-1)\sum_{i=1}^{n} \tau_i x_{ih} x_{ij} \frac{e^{-\mathbf{x}_i\gamma}}{(1+e^{-\mathbf{x}_i\gamma})^2},$$

$(h = 0, 1, \ldots, p; j = 0, 1, \ldots, p)$, where we let $x_{i0} = 1$, for simplifying the notation.

For instance, when $\mathbf{x}_i\gamma = \gamma_0 + \gamma_1 x_{i1}$, the (3×3) asymptotic covariance matrix is:

$$\mathbf{V} \simeq -\begin{bmatrix} d_{\pi\pi} & d_{\pi\gamma_0} & d_{\pi\gamma_1} \\ d_{\gamma_0\pi} & d_{\gamma_0\gamma_0} & d_{\gamma_0\gamma_1} \\ d_{\gamma_1\pi} & d_{\gamma_1\gamma_0} & d_{\gamma_1\gamma_1} \end{bmatrix}^{-1},$$

where

$$d_{\gamma_0\gamma_0} = \frac{\partial^2}{\partial\gamma_0^2} Q_2(\gamma) = -(m-1)\sum_{i=1}^{n} \tau_i \left\{ \frac{e^{-\gamma_0-\gamma_1 x_{i1}}}{(1+e^{-\gamma_0-\gamma_1 x_{i1}})^2} \right\};$$

$$d_{\gamma_1\gamma_1} = \frac{\partial^2}{\partial\gamma_1^2} Q_2(\gamma) = -(m-1)\sum_{i=1}^{n} \tau_i \left\{ x_{i1}^2 \frac{e^{-\gamma_0-\gamma_1 x_{i1}}}{(1+e^{-\gamma_0-\gamma_1 x_{i1}})^2} \right\};$$

$$d_{\gamma_0\gamma_1} = d_{\gamma_1\gamma_0} = \frac{\partial^2}{\partial\gamma_0\gamma_1} Q_2(\gamma) = -(m-1)\sum_{i=1}^{n} \tau_i \left\{ x_{i1} \frac{e^{-\gamma_0-\gamma_1 x_{i1}}}{(1+e^{-\gamma_0-\gamma_1 x_{i1}})^2} \right\};$$

$$d_{\pi\gamma_0} = d_{\gamma_0\pi} = 0; d_{\pi\gamma_1} = d_{\gamma_1\pi} = 0.$$

Of course, in order to compute all the previous expressions, we use the ML estimates $\hat{\pi}, \hat{\gamma}$.

Several statistical and computing consequences stem from these results:

- the maximum likelihood estimators of π and of the coefficient vector γ are asymptotically uncorrelated, and then independent, as marginals of a Bivariate Normal distribution;
- it is possible to perform hypotheses tests on π and γ, separately;

- from a computing point of view, we get:

$$Var(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{n}.$$

4. The MUB model with a subject specific π

In this section we develop an extension of the MUB model allowing the π parameter to depend upon the raters'covariates. Then, following the same scheme of the previous section 3, we describe the related steps of the E-M algorithm, and the expressions for the asymptotic standard errors of the parameters estimates.

4.1 Model specification

Here we assume that the uncertainty in the ranking process depends upon a set of raters' covariates, and thus we link the π parameter to a linear predictor. On the other hand, ξ is kept fixed with respect to the raters' features. Following the same strategy of the previous section, we let:

$$(\pi \mid \mathbf{X} = \mathbf{x}_i) = \frac{1}{1 + exp(-\mathbf{x}_i \beta)}, \qquad i = 1, 2, ..., n,$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$ is the covariates coefficients vector. Again, we notice that when the linear predictor $\mathbf{x}_i\beta \to \infty$, then $(\pi \mid \mathbf{X}{=}\mathbf{x}_i) \to 1$; viceversa, when $\mathbf{x}_i\beta \to -\infty$, then $(\pi \mid \mathbf{X} = \mathbf{x}_i) \to 0$.

Moreover, using the same notation of section 3.1, if we let $\mathbf{x}_i\beta = \beta_0 + \beta_1 x_{i1}$, it results:

$$(\pi \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - (\pi \mid \mathbf{X} = \mathbf{x}_i) = \frac{e^{-\beta_0 - \beta_1 x_{i1}}(1 - e^{-\beta_1})}{(1 + e^{-\beta_0 - \beta_1 x_{i1}})(1 + e^{-\beta_0 - \beta_1 x_{i1} - \beta_1})},$$

showing that $(\pi \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - (\pi \mid \mathbf{X} = \mathbf{x}_i) > 0$ when $\beta_1 > 0$; that is, the uncertainty $(1 - \pi)$ decreases when x_{i1} augments.

Generalizing to $p > 1$ covariates, we have *ceteris paribus* (for $j = 1, 2, \ldots, p$):

- $\beta_j < 0 \Rightarrow$ uncertainty increases, when x_{ij} augments;
- $\beta_j = 0 \Rightarrow$ no effect on the uncertainty;
- $\beta_j > 0 \Rightarrow$ uncertainty decreases, when x_{ij} augments.

Instead, as far as it concerns the effect of the covariates on $E(R)$, it results:

$$E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) =$$
$$= \left(\frac{1}{2} - \xi\right)(m-1)\left[\frac{e^{-\beta_0-\beta_1 x_{i1}}(1-e^{-\beta_1})}{(1+e^{-\beta_0-\beta_1 x_{i1}})(1+e^{-\beta_0-\beta_1 x_{i1}-\beta_1})}\right].$$

Thus,

- $\xi < 1/2$ and $\beta_1 > 0 \Rightarrow E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) > 0$
- $\xi > 1/2$ and $\beta_1 < 0 \Rightarrow E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) > 0$;
- otherwise, $E(R \mid \mathbf{X} = \mathbf{x}_i + \mathbf{u}) - E(R \mid \mathbf{X} = \mathbf{x}_i) < 0$.

4.2 The likelihood function and the E-M algorithm

Given a vector of observed ranks $\mathbf{r} = (r_1, r_2, \ldots, r_n)'$, the log-likelihood function for the MUB model, using the above introduced specification for the parameter π, becomes:

$$log\, L(\beta, \xi; \mathbf{r}, \mathbf{X}) =$$
$$= \sum_{i=1}^{n} log \left\{ \frac{1}{1+e^{-\mathbf{x}_i \beta}} \left[P_B(r_i;\xi) - \frac{1}{m} \right] + \frac{1}{m} \right\}.$$

For obtaining the maximum likelihood estimates of both ξ and the coefficient vector β, we again rely on the E-M algorithm, by iterating the following two steps:

- *E-step*: compute the conditional expectation of the complete data log-likelihood, using the current parameters estimate; that is:

$$Q(\xi^{(k)},\beta^{(k)}) =$$

$$= \sum_{i=1}^{n}\left\{\tau_i^{(k)} log\left(\frac{1}{1+e^{-\mathbf{x}_i\beta^{(k)}}}\right) + (1-\tau_i^{(k)}) log\left(\frac{e^{-\mathbf{x}_i\beta^{(k)}}}{1+e^{-\mathbf{x}_i\beta^{(k)}}}\right)\right\} +$$

$$+\sum_{i=1}^{n}\left\{\tau_i^{(k)} log(P_B(r_i;\xi^{(k)})) + (1-\tau_i^{(k)}) log\left(\frac{1}{m}\right)\right\}.$$

- *M-step*: maximize the $Q(\xi^{(k)},\beta^{(k)})$ function in order to get updated estimates of the parameters: $\xi^{(k+1)},\beta^{(k+1)}$.

In fact, at a fixed k-th iteration, we get:

$$Q(\xi^{(k)},\beta^{(k)}) = Q_1(\xi^{(k)}) + Q_2(\beta^{(k)}) + constant,$$

where we let:

$$Q_1(\xi^{(k)}) = \sum_{i=1}^{n}\left\{\tau_i^{(k)} log(P_B(r_i;\xi^{(k)}))\right\};$$

$$Q_2(\beta^{(k)}) = \sum_{i=1}^{n}\left\{\tau_i^{(k)} log\left(\frac{1}{1+e^{-\mathbf{x}_i\beta^{(k)}}}\right) + (1-\tau_i^{(k)}) log\left(\frac{e^{-\mathbf{x}_i\beta^{(k)}}}{1+e^{-\mathbf{x}_i\beta^{(k)}}}\right)\right\} =$$

$$= -\sum_{i=1}^{n}\left\{log\left(1+e^{-\mathbf{x}_i\beta^{(k)}}\right) + (1-\tau_i^{(k)})\mathbf{x}_i\beta^{(k)}\right\}.$$

Maximizing $Q_1(\xi^{(k)})$ with respect to ξ, it yields:

$$\xi^{(k+1)} = \frac{m - \sum_{i=1}^{n} r_i\tau_i^{(k)} / \sum_{i=1}^{n}\tau_i^{(k)}}{m-1},$$

while $\beta^{(k+1)}$ is obtained by finding the maximum of $Q_2(\beta^{(k)})$.

In detail, the implementation of the E-M algorithm for the MUB model with a subject specific π is developed as follows:

1. Compute $logL(\xi^{(0)},\beta^{(0)};\mathbf{r},\mathbf{X})$, using starting values $\xi^{(0)}$, $\beta^{(0)}$ for initializing the algorithm;

2. $P_B(r_i;\xi^{(k)}) = \begin{pmatrix} m-1 \\ r_i-1 \end{pmatrix} (1-\xi^{(k)})^{r_i-1}(\xi^{(k)})^{m-r_i}$;
   $\pi_i^{(k)} = \frac{1}{1+e^{-\mathbf{x}_i\beta^{(k)}}}$, $\qquad i = 1, 2, \dots, n$;

3. $\tau_i^{(k)} = \left[1 + \frac{1-\pi_i^{(k)}}{m\pi_i^{(k)} P_B(r_i;\xi^{(k)})}\right]^{-1}$;
   $\bar{R}_n^{(k)} = \frac{\sum_{i=1}^n r_i \tau_i^{(k)}}{\sum_{i=1}^n \tau_i^{(k)}}$; $\qquad i = 1, 2, \dots, n$;

4. $\xi^{(k+1)} = \frac{m-\bar{R}_n^{(k)}}{m-1}$; $\qquad$ maximize $Q_2(\beta^{(k)})$ for getting $\beta^{(k+1)}$;

5. compute $logL(\xi^{(k+1)}, \beta^{(k+1)}; \mathbf{r}, \mathbf{X})$;

6. if $L(\xi^{(k+1)}, \beta^{(k+1)}) - L(\xi^{(k)}, \beta^{(k)}) < \epsilon$ stop;
   else $k = k + 1$, and reiterate from 2.

4.3 Computing standard errors

In this subsection we derive the inverse of the observed information matrix: $I(\hat{\xi}, \hat{\beta}; \mathbf{r})$, in order to estimate the asymptotic covariance matrix of the maximum likelihood estimators. As in the previous section, we approximate $I(\hat{\xi}, \hat{\beta}; \mathbf{r})$ by taking the second-order partial derivatives of the conditional expectation of the complete data log-likelihood:

$$Q(\xi, \beta) = Q_1(\xi) + Q_2(\beta) + constant.$$

Then, it results that:

$$\mathbf{V} \simeq - \begin{bmatrix} d_{\xi\xi} & \mathbf{0}' \\ \mathbf{0} & \mathbf{B} \end{bmatrix}^{-1} = - \begin{bmatrix} d_{\xi\xi}^{-1} & \mathbf{0}' \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix},$$

where

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

and $\mathbf{B}$ is a $(p+1 \times p+1)$ matrix with generic element:

$$\{\mathbf{B}\}_{hj} = \frac{\partial^2}{\partial\beta_h \partial\beta_j} Q_2(\beta) = -\sum_{i=1}^n x_{ih} x_{ij} \frac{e^{-\mathbf{x}_i\beta}}{(1+e^{-\mathbf{x}_i\beta})^2},$$

($h = 0, 1, \ldots, p; j = 0, 1, \ldots, p$), where again we let $x_{i0} = 1$.

For instance, when $\mathbf{x}_i\beta = \beta_0 + \beta_1 x_{i1}$, the ($3\times3$) asymptotic covariance matrix is:

$$\mathbf{V} \simeq - \begin{bmatrix} d_{\xi\xi} & d_{\xi\beta_0} & d_{\xi\beta_1} \\ d_{\beta_0\xi} & d_{\beta_0\beta_0} & d_{\beta_0\beta_1} \\ d_{\beta_1\xi} & d_{\beta_1\beta_0} & d_{\beta_1\beta_1} \end{bmatrix}^{-1},$$

where

$$d_{\beta_0\beta_0} = \frac{\partial^2}{\partial\beta_0^2} Q_2(\beta) = -\sum_{i=1}^{n} \left\{ \frac{e^{-\beta_0-\beta_1 x_{i1}}}{(1+e^{-\beta_0-\beta_1 x_{i1}})^2} \right\};$$

$$d_{\beta_1\beta_1} = \frac{\partial^2}{\partial\beta_1^2} Q_2(\beta) = -\sum_{i=1}^{n} \left\{ x_{i1}^2 \frac{e^{-\beta_0-\beta_1 x_{i1}}}{(1+e^{-\beta_0-\beta_1 x_{i1}})^2} \right\};$$

$$d_{\beta_0\beta_1} = d_{\beta_1\beta_0} = \frac{\partial^2}{\partial\beta_0\beta_1} Q_2(\beta) = -\sum_{i=1}^{n} \left\{ x_{i1} \frac{e^{-\beta_0-\beta_1 x_{i1}}}{(1+e^{-\beta_0-\beta_1 x_{i1}})^2} \right\};$$

$$d_{\xi\beta_0} = d_{\beta_0\xi} = 0; d_{\xi\beta_1} = d_{\beta_1\xi} = 0.$$

In fact, in order to compute all the previous expressions we use the ML estimates $\hat{\xi}$, $\hat{\beta}$.

Moreover, analogous considerations to those of subsection 3.3 hold.

5. Some evidence from real datasets

In this section we are going to show some evidence obtained by fitting the MUB model with covariates to real datasets. In particular, as illustrative examples, we choose only a sample of results from more wide case-studies. Besides, we limit ourselves to the statistical aspects of the estimation, avoiding any comments about the psychological issues that might be derived from our models, too.

The data refer to studies about the preferences of young people towards different colors (henceforth, **Colors** dataset, $m = 12$, $n = 169$), different professions (**Professions** dataset, $m = 14$, $n = 183$), and different places where to live (**Cities** dataset, $m = 12$, $n = 183$). In each case, the raters were asked to rank the items from the most preferred ($r = 1$) to the least loved ($r = m$), without ties.

In first instance, we illustrate the results obtained by fitting to the data the MUB model with a subject specific ξ (following the model proposed in section 3), that is the case where we assume that the degree of liking depends upon covariates, while the uncertainty feeling does not.

<u>Colors</u>

Tables 1, 2 and 3 show the ML estimates obtained via the method discussed in the previous sections, and concerning the colors Pink, Grey and Brown, respectively.

Table 1. Preferences towards Pink.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	-1.832	*0.141*
Sex (M=0, F=1)	$\hat{\gamma}_1$	0.722	*0.175*
	$\hat{\pi}$	0.498	*0.038*
$E(R \mid \text{Sex} = 0)$	8.483	$(\hat{\xi} \mid \text{Sex=0})$	0.138
$E(R \mid \text{Sex} = 1)$	7.881	$(\hat{\xi} \mid \text{Sex=1})$	0.248

Table 2. Preferences towards Grey.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	0.845	*0.116*
Left-handed (No=0, Yes=1)	$\hat{\gamma}_1$	1.017	*0.479*
	$\hat{\pi}$	0.211	*0.031*
$E(R \mid \text{Lh} = 0)$	6.036	$(\hat{\xi} \mid \text{Lh=0})$	0.699
$E(R \mid \text{Lh} = 1)$	5.650	$(\hat{\xi} \mid \text{Lh=1})$	0.866

Table 3. Preferences towards Brown.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	-3.465	*0.372*
Smoker (No=0, Yes=1)	$\hat{\gamma}_1$	4.391	*0.447*
	$\hat{\pi}$	0.175	*0.029*
$E(R \mid \text{Sm} = 0)$	7.403	$(\hat{\xi} \mid \text{Sm=0})$	0.030
$E(R \mid \text{Sm} = 1)$	6.084	$(\hat{\xi} \mid \text{Sm=1})$	0.716

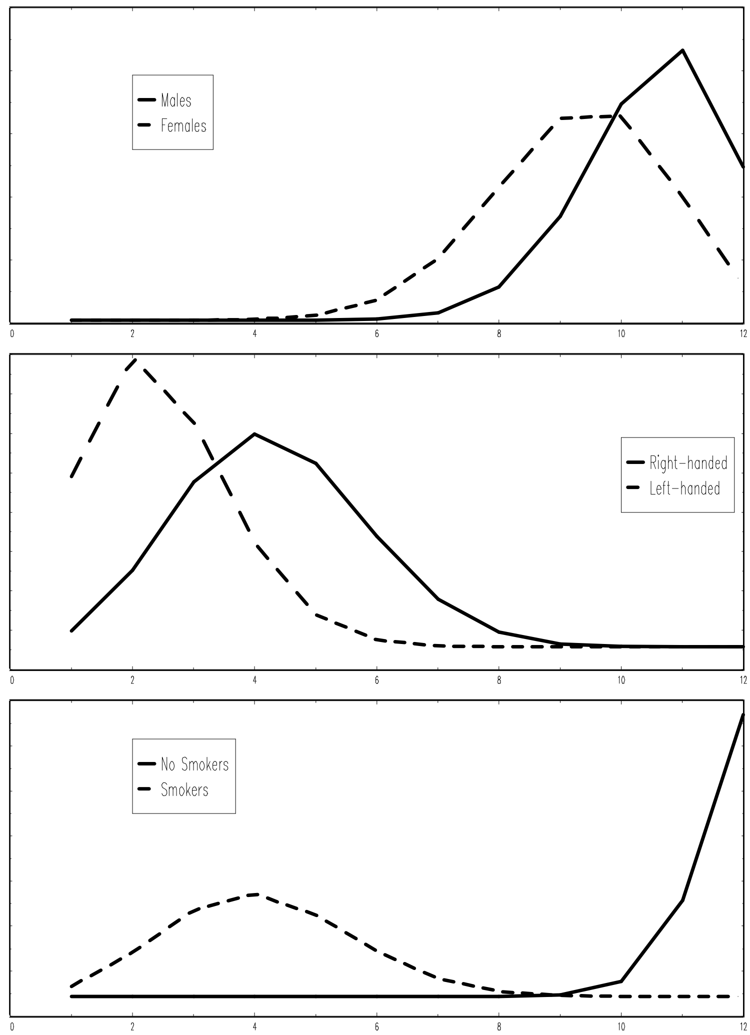

Figure 1. Estimated probability functions for preferences towards Pink, Grey and Brown,respectively.

In all these cases we obtained positive estimates of the coefficient γ_1, that implies (see also Figure 1):

- the preference towards Pink (generally low) is greater in the Females than in the Males;

- the preference towards Grey (generally quite high) is greater in Left-handed people than in Right-handed ones;

- the preference towards Brown is greater in the Smokers than in the no-Smokers.

Professions

Tables 4 and 5 show the results obtained for the preferences expressed towards Public Relations and Politics. Also in this case it emerges that the liking for this kind of jobs depends upon the Sex of the raters: indeed, the preference towards Public Relations is greater in the young women than in the young men, and it happens viceversa for the Politics.

Table 4. Preferences towards Public Relations.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	0.554	*0.085*
Sex (M=0, F=1)	$\hat{\gamma}_1$	1.305	*0.143*
	$\hat{\pi}$	0.525	*0.037*
$E(R \mid \text{Sex} = 0)$	6.579	$(\hat{\xi} \mid \text{Sex=0})$	0.635
$E(R \mid \text{Sex} = 1)$	5.009	$(\hat{\xi} \mid \text{Sex=1})$	0.865

Table 5. Preferences towards Politics.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	2.780	*0.384*
Sex (M=0, F=1)	$\hat{\gamma}_1$	-2.708	*0.442*
	$\hat{\pi}$	0.087	*0.020*
$E(R \mid \text{Sex} = 0)$	7.002	$(\hat{\xi} \mid \text{Sex=0})$	0.942
$E(R \mid \text{Sex} = 1)$	7.480	$(\hat{\xi} \mid \text{Sex=1})$	0.518

With respect to this example, we would like to stress that $(\hat{\xi} \mid \text{Sex=1}) = 0.518$: in fact, this happens since $\hat{\gamma}_0 \simeq -\hat{\gamma}_1$, and thus $(\hat{\xi} \mid \text{Sex=1}) = \frac{1}{1+exp(-2.78+2.71)} \simeq 0.5$. More in general, every time the estimated linear predictor $\mathbf{x}_i\hat{\gamma} \simeq 0$, we get $\hat{\xi} \simeq 1/2$, that is a symmetric distribution (see Figure 2). In particular, this happens when there a single dichotomic covariate and $\hat{\gamma}_0 \simeq -\hat{\gamma}_1$, as in the previous example.

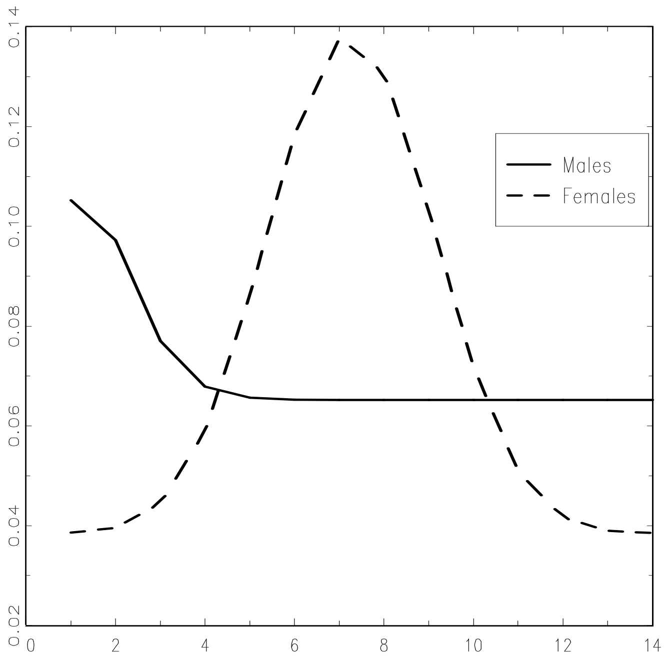

Figure 2. Estimated probability functions for preferences towards Politics.

Cities

Finally, from the Cities dataset, the greatest evidence of a significant difference in the preferences was obtained in the case of Verona (Table 6), for whom the disliking feeling of the men is much deeper than that of the women.

Table 6. Preferences towards Verona.

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	−3.637	*0.328*
Sex (M=0, F=1)	$\hat{\gamma}_1$	3.397	*0.343*
	$\hat{\pi}$	0.387	*0.036*
$E(R \mid \text{Sex} = 0)$	8.518	$(\hat{\xi} \mid \text{Sex=0})$	0.026
$E(R \mid \text{Sex} = 1)$	6.755	$(\hat{\xi} \mid \text{Sex=1})$	0.440

Now, we are going to illustrate the results obtained by fitting to the data the MUB model with a subject specific π (as proposed in section 4): this means that we assume that the uncertainty feeling in the ranking process may be explained by means of the raters' covariates; the ξ parameter is, instead, kept fixed.

Colors

Considering the same items as above, we notice that for Pink and Grey the same covariates found above are significant also for explaining the uncertainty (Tables 7 and 8). In particular, it emerges that:

- in the women there is more uncertainty as far as it concerns the ranking of Pink, with respect to the men;

- left-handed people exhibit less uncertainty in the ranking of Grey than right handed people.

Table 7. Preferences towards Pink.

Variable		estimate	standard error
Constant	$\hat{\beta}_0$	0.585	*0.244*
Sex (M=0, F=1)	$\hat{\beta}_1$	−1.171	*0.323*
	$\hat{\xi}$	0.171	*0.013*
$E(R \mid \text{Sex} = 0)$	8.825	$(\hat{\pi} \mid \text{Sex=0})$	0.642
$E(R \mid \text{Sex} = 1)$	7.795	$(\hat{\pi} \mid \text{Sex=1})$	0.357

Table 8. Preferences towards Grey.

Variable		estimate	standard error
Constant	$\hat{\beta}_0$	-2.165	*0.260*
Left-handed (No=0, Yes=1)	$\hat{\beta}_1$	3.879	*0.963*
	$\hat{\xi}$	0.787	*0.025*
$E(R \mid \text{Lh} = 0)$	6.175	$(\hat{\pi} \mid \text{Lh=0})$	0.103
$E(R \mid \text{Lh} = 1)$	3.333	$(\hat{\pi} \mid \text{Lh=1})$	0.847

Collecting this evidence with that of Tables 1 and 2, we can conclude:

- the covariate Sex has a significant effect on both the liking and the uncertainty in giving a rank to Pink: in particular, the females appear to be less disappointed with this color, but are also more doubtful in deciding a rank;
- the covariate Left-handed has a significant effect on both the liking and the uncertainty in giving a rank to Grey: in particular, left-handed people appear to prefer more this color, and are less doubtful in deciding a rank for it.

As far as it concerns the preferences towards Brown, the liking feeling resulted to depend upon the feature Smoker (Table 3), but this covariate was found not significant with respect to the uncertainty.

Professions

For this dataset, we get that the covariate Sex is significant also for explaining the uncertainty in the ranking of Public Relations (Table 9), where the women are less doubtful than the men.

Table 9. Preferences towards Public Relations.

Variable		estimate	standard error
Constant	$\hat{\beta}_0$	-1.815	*0.302*
Sex (M=0, F=1)	$\hat{\beta}_1$	2.208	*0.369*
	$\hat{\xi}$	0.848	*0.012*
$E(R \mid \text{Sex} = 0)$	6.866	$(\hat{\pi} \mid \text{Sex=0})$	0.140
$E(R \mid \text{Sex} = 1)$	4.799	$(\hat{\pi} \mid \text{Sex=1})$	0.597

On the other hand, there is no analogous evidence about the preferences toward Politics.

Cities

Here, in relation to Verona, the covariate Sex resulted significant again (Table 10), yielding a fewer uncertainty feeling in the ranking given by the women. In particular, the estimated value for the men $(\hat{\pi} \mid \text{Sex=0}) = 0.011$ highlights that the preference toward Verona in the males has an almost Uniform behavior (as it is depicted in Figure 3), with a constant probability for each rank: say, an equipreference feeling.

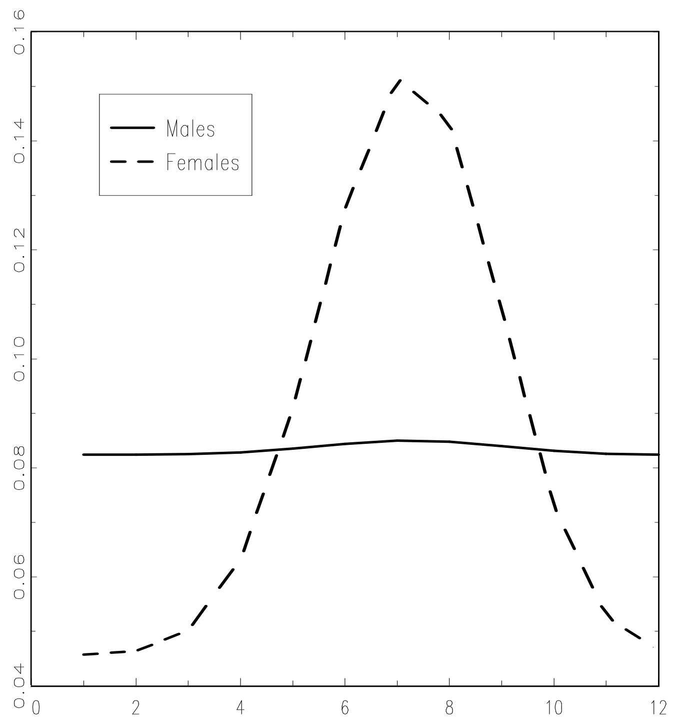

Figure 3. Estimated probability functions for preferences towards Verona.

Table 10. Preferences towards Verona.

Variable		estimate	standard error
Constant	$\hat{\beta}_0$	−4.483	*0.997*
Sex (M=0, F=1)	$\hat{\beta}_1$	4.290	*1.019*
	$\hat{\xi}$	0.436	*0.023*
$E(R \mid \text{Sex} = 0)$	6.508	$(\hat{\pi} \mid \text{Sex=0})$	0.011
$E(R \mid \text{Sex} = 1)$	6.816	$(\hat{\pi} \mid \text{Sex=1})$	0.452

Finally, we would like to show the case of Venice, for which the covariate Sex resulted significant in the MUB model with a subject specific π (Figure 4b), but not in that with ξ depending on the raters' covariates (Figure 4a, where the estimated probability functions for males and females can be hardly distinguished). In particular, it appears (Tables 11 and 11bis) that men and women have the same degree of liking towards Venice, but in the women there is less uncertainty in assigning a preference rank to this city.

Table 11. Preferences towards Venice (model for ξ).

Variable		estimate	standard error
Constant	$\hat{\gamma}_0$	0.268	*0.100*
Sex (M=0, F=1)	$\hat{\gamma}_1$	0.007	*0.050*
	$\hat{\pi}$	0.461	*0.037*
$E(R \mid \text{Sex} = 0)$	6.163	$(\hat{\xi} \mid \text{Sex=0})$	0.567
$E(R \mid \text{Sex} = 1)$	6.154	$(\hat{\xi} \mid \text{Sex=1})$	0.568

Table 11bis. Preferences towards Venice (model for π).

Variable		estimate	standard error
Constant	$\hat{\beta}_0$	−1.203	*0.249*
Sex (M=0, F=1)	$\hat{\beta}_1$	1.921	*0.333*
	$\hat{\xi}$	0.565	*0.016*
$E(R \mid \text{Sex} = 0)$	6.334	$(\hat{\pi} \mid \text{Sex=0})$	0.231
$E(R \mid \text{Sex} = 1)$	6.016	$(\hat{\pi} \mid \text{Sex=1})$	0.672

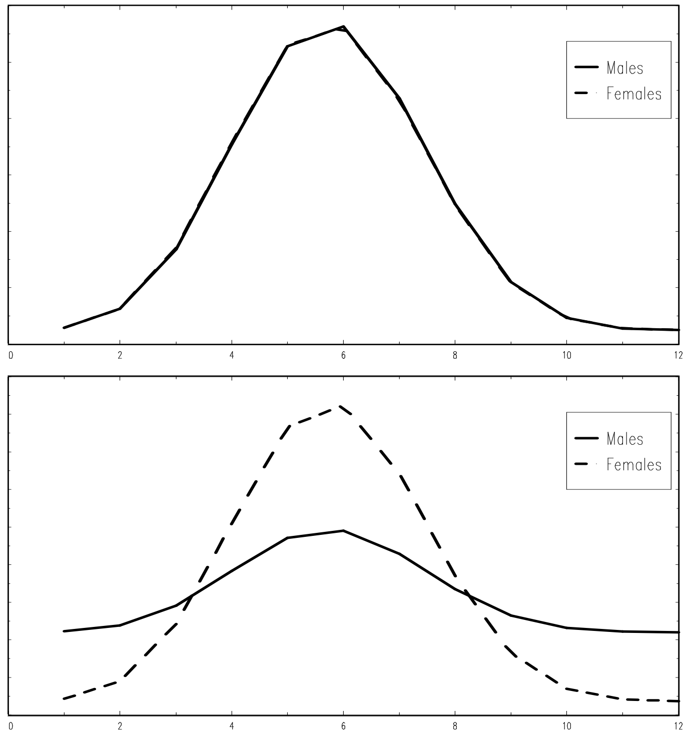

Figure 4 (a, b). Estimated probability functions for preferences towards Venice.

6. Concluding remarks and further developments

In this paper we developed an extension of a mixture model for ranks data (MUB), by introducing a link between the parameters and the raters' covariates. In this way, we are able to explain both the preference feeling and the uncertainty of the ranking process by means of some features of the raters. This allows a deeper interpretation of the liking/disliking behavior, to be used for inferential and predictive purposes.

In fact, this has been accomplished by modelling the two parameters of the MUB in function of a linear predictor, once at a time. Thus, a natural further step will be to develop a model where both the parameters depends upon a set (maybe disjoint) of covariates, and then to derive the corresponding E-M algorithm for obtaining the maximum likelihood estimates of the coefficients vectors β and γ together. In this vein, some preliminary results have been obtained by Piccolo (2003).

Indeed, the evidence obtained in this work is that, in many real applications, it is important to jointly consider the effect of the raters' covariates on both the liking and the uncertainty feeling, and that it may be the case where different covariates play significant roles with respect to them.

Moreover, it seems also important to propose some goodness of fit measures for this kind of models, by means, for instance, of a comparison between observed and expected quantities, and taking into account how the knowledge of a significant covariate may improve the estimation of a degree of preference towards a given item.

Acknowledgements. This research benefited from funds of Dipartimento di Scienze Statistiche, Università di Napoli Federico II. The Author has also exploited the facilities of CFEPSR, Portici. Useful comments by referees are gratefully acknowledged.

References

Bradley R. A., Terry M. A. (1952), Rank Analysis of Incomplete Block Designs I, *Biometrika*, **39**, 324-45.

D'Elia A. (1999), A Proposal for Ranks Statistical Modelling, in *Statistical Modelling* (Friedl H., Berghold A., Kauermann G. eds.), Graz, 468-471.

D'Elia A. (2000), Il Meccanismo dei Confronti Appaiati nella Modellistica per Graduatorie: Sviluppi Statistici ed Aspetti Critici, *Quaderni di Statistica*, **2**, 173-203.

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

D'Elia A., Piccolo D. (2003), A Mixture Model for Preferences Data Analysis, *submitted*.

McCullagh P., Nelder J. (1989), *Generalized Linear Models* (2^{nd} edition), Chapman & Hall. London.

McLachlan G., Krishnan T. (1997), *The E-M Algorithm and Extensions*, J. Wiley & Sons, New York.

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

Pawitan Y. (2001), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Clarendon Press, Oxford.

Piccolo D. (2003), Computational issues in the E-M algorithm for ranks model estimation with covariates, *Quaderni di Statistica*, **5**, this issue.