Statistica & Applicazioni - Vol. VII, n. 1, 2009, pp. 25-44

A COMPARISON OF PRELIMINARY ESTIMATORS IN A CLASS OF ORDINAL DATA MODELS

Maria Iannario*

SUMMARY

In this paper, we propose several initial values for the EM algorithm of maximum likelihood estimates of the parameters in a class of models, called CUB, recently introduced for ordinal data. Specifically, we compare the algorithmic efficiency of each estimator with respect to a naive proposal through a vast simulation experiment. The results confirm a substantial gain in efficiency of the moments estimators over the whole parametric space. Then, some extensions are also discussed and several applications to real data sets are presented.

Keywords: *CUB models, Preliminary estimation, Algorithmic efficiency.*

1. INTRODUCTION

Statistical analysis of ordinal data is a frequent issue in different fields of research. Specifically, they arise when people are asked to select sequentially m ''objects'' (food, beverage, teams, songs, commuter transports, holidays, and so on) among a list of similar ones according to a criterion of preference: in this case, we are collecting *ranking data*. Instead, in different contexts, several items are submitted to a sample of subjects asking them to express a degree of liking/disliking on an ordinal scale (say, 1 to m) towards sentences, behaviours, problems, habits, and so on: in this case, we refer to evaluation or *rating data.*

Of course, the two situations are logically and statistically different. In *ranking analyses*, the experiment produces a discrete multivariate random variable whose components explain the stated preferences towards m given objects; instead, in *rating analyses* we are faced with a univariate random variable which expresses the level of consensus of several subjects towards a given item on the support $\{1, 2, \ldots, m\}$.

As a consequence, the observed components of a *ranking* study are not independent since any observable response vector is a permutation of the first integers; on the contrary, in a *rating* study any single answer expresses the subject's evaluation, and it can assume any value on the given support.

However, in both experiments the response is an ordinal expression of subject's perception towards objects/items/services; thus, the analysis of a single object or item may be expressed via the *marginal distribution* of ordinal random variables defined

* Dipartimento di Scienze Statistiche - Università degli Studi di Napoli Federico II - via L. Rodinò, 22 - 80138 NAPOLI (e-mail: maria.iannario@unina.it).

on the same support. In this regard, it is important to stress that our analysis is aimed at *univariate* distributions of data, also explained by means of possible subjects' covariates, and with this caveat we will present real data sets of both analyses.

In the statistical literature, the classical approach for modelling ordinal data is via the paradigm of Generalized Linear Models (GLM) introduced by McCullagh and Nelder (1989) and Nelder and Wedderburn (1972). Specifically, McCullagh (1980) proposed the cumulative probabilities of the ordered outcomes as a monotone transformation of a linear predictor onto the unit interval, assuming a logit or probit link function. Several variants of this proposal (adjacent categories, continuation-ratio logits, etc.) are discussed by Agresti (2002, pp. 274-293), Faraway (2006, pp. 106-112) and Dobson and Barnett (2008, pp. 157-162).

More specifically, rank modelling issues and related problems are raised by Fligner and Verducci (1993) and Marden (1995). With reference to rating data, a widespread approach is Item Response Theory (and its generalization) where simultaneous responses to several items are investigated by common latent traits and statistical models are able to account both for items' difficulty and subjects' ability (Bock and Moustaki, 2007). In this area, interesting procedures are discussed by Pagani and Zanarotti (2003), Mazzali (2004), Brentari *et al.* (2007), among others.

A different approach has been pursued by investigating the psychological components of subjects' decisions in order to assess explicitly the probability of a definite choice among a set of ordered alternatives. This class of model is called *CUB*[1]. In fact, this paradigm is more general than GLM one as we do not assume that random variables belong to the exponential class and the link function is not compelled to connect expectations to covariates. Since they are specified as a mixture of discrete distributions, effective tools for statistical inference have been derived by using Maximum Likelihood (ML) asymptotic theory; then, an EM algorithm for ML estimators has been implemented. Currently, a software in R and an updated list of references is available (Piccolo and Iannario, 2008).

However, the almost sure convergence of the EM algorithm is paid by a low convergence rate; thus, it is important to start the numerical procedure by accurate initial values[2]. Thus, we compare several estimators aimed at accelerating the convergence rate of EM algorithm. They belong to two groups: the first one (*grid, naive*, *fzero* estimators) is composed by estimators related to a simple searching of preliminary values; the second one (*robust* and *moments* estimators) consists of estimators derived

[1] The acronym stems from *C*ovariates in *U*niform and *B*inomial random variables. This class of models has been introduced by Piccolo (2003), D'Elia and Piccolo (2005) as discrete random variables (without covariates) and generalized by Piccolo (2006) and Piccolo and D'Elia (2008) in several contexts (with subjects' and objects' covariates). Further developments are in Iannario (2009b).

[2] This aspect is not really decisive for estimating few *CUB* models as the available software produces reasonable estimates in tenths of second; however, for extensive simulation experiments a sensible time reduction is often necessary. Further motivations are reported in the concluding remarks.

by statistical methods. Their performances are examined in terms of algorithmic efficiency and some extensions to *CUB* models with covariates are proposed.

The paper is organized as follows: in the next section, we briefly introduce *CUB* models and their main inferential issues. Then, in Sections 3-4, we list the rationale for introducing our estimators. In Section 5, we compare their relative efficiency with respect to the *naive* estimator through an extensive simulation experiment over the whole parametric space. Then, in Section 6, extensions to *CUB* models with a dummy covariate are discussed. In Section 7, applications of these findings to real data sets are performed. Finally, in Section 8 we consider problems generated by extreme distributions and some concluding remarks end the paper.

2. SPECIFICATION AND ESTIMATION OF *CUB* MODELS

The starting point of this approach is the evidence that human choices are the result of a very complex procedure that produces a selection from a discrete set of alternatives. If we limit ourselves to consider the marginal distribution of a single choice, the ultimate decision turns out to be an ordered value which is a synthesis of several latent variables that include *feeling* towards the objects/topics/items and an intrinsic *uncertainty/fuzziness* component. Thus, the act of choosing an object from a given list or the assignment of a score within a prefixed scale always produce an integer value that is a balanced mixture of a reflexive behaviour and a completely random one[3].

Since the support of respondents is generally referred to $\{1, 2, \ldots, m\}$, in order to formalize the previous considerations in a well defined probability model, the shifted Binomial and Uniform discrete random variables, respectively, are the components that should be conveniently weighted.

Thus, for a given integer[4] $m > 3$, we assume that the observed ordinal value r is the realization of a discrete random variable R. We will call R a *CUB* random variable if its probability distribution $P_r(R = r \mid \theta) = p_r$ is defined by:

$$p_r = \pi \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} + (1-\pi)\frac{1}{m}, \quad r = 1, 2, \ldots, m.$$

The parametric space $\Omega(\theta)$ for the vector $\theta = (\pi, \xi)'$ is the unit square:

$$\Omega(\theta) = \{(\pi, \xi) : \ 0 < \pi \leq 1; \ 0 \leq \xi \leq 1\}.$$

The identifiability of *CUB* models has been proven by Iannario (2009a) when $m > 3$ and further extensions for taking degenerate situations into account have been proposed (Iannario, 2009b).

---

[3] More extensive discussions of these and related aspects are in Piccolo (2006), Iannario and Piccolo (2009), also with reference to real data sets.

[4] The constraint $m > 3$ avoids to consider degenerate $(m = 1)$, indeterminate $(m = 2)$ or saturated $(m = 3)$ models.

As far as the interpretation of this model is concerned, the logic of mixture distributions assumes the presence of two clusters and people are selected from them with probability given by the weights of the combination. In fact, *we do not assume this as a real situation* but we only consider the result of a single respondent *as* if π and $(1-\pi)$ would be measures of its *propensity/inclination* to belong to first and second group, respectively.

In the previous model this quantity is constant but, by introducing subjects' characteristics, we may well consider π_i and $(1-\pi_i)$ as personal propensities characterizing the *i*-th subject. In fact, this class of models has been generalized by including covariates for explaining the effects of subjects' characteristics on the final choice.

Thus, we introduce $CUB(p,q)$ models when π and/or ξ parameters are related to p and/or q subjects' covariates, respectively, following both the logic of GLM approach and the simpler paradigm advocated by King *et al.* (2000). It requires a stochastic component (for a completely general random variable) and a deterministic link among the parameters (not necessarily the expectation) and the covariates.

Generally, we choose the link function as the logistic one; thus, the final specification for a general $CUB(p,q)$ model with covariates is, for any $i = 1, 2, \ldots, n$:

$$P_r(R = r \mid y_i, w_i) = \pi_i \binom{m-1}{r-1} (1-\xi_i)^{r-1} \xi_i^{m-r} + (1-\pi_i)\,\frac{1}{m}\,,$$

for $r = 1, 2, \ldots, m$ and

$$\pi_i = (\pi \mid y_i) = \frac{1}{1+\exp(-y_i\beta)}; \qquad \xi_i = (\xi \mid w_i) = \frac{1}{1+\exp(-w_i\gamma)}\,,$$

where $y_i = (1, y_{i1}, \ldots, y_{ip})$ is the *i*-th row of a design matrix[5] $\mathbf{Y}$ of $(p+1)$ explanatory variables related to the parameter π by means of the coefficient vector $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$, and $w_i = (1, w_{i1}, \ldots, w_{iq})$ is the *i*-th row of a design matrix $\mathbf{W}$ of $(q+1)$ explanatory variables related to ξ by means of the coefficient vector $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_q)'$. Of course, a $CUB(0,0)$ model is simply the mixture distribution as defined in first instance.

In the following sections, we will mainly consider models without covariates. Then, given a sample of ordinal values $r = (r_1, r_2, \ldots, r_n)'$, sufficient statistics are based on any $(m-1)$ subset of the absolute frequencies $(n_1, n_2, \ldots, n_m)'$. In fact, the log-likelihood function for a *CUB* model with parameter vector θ is:

$$\ell(\theta) = \sum_{r=1}^{m} n_r \log P_r(R = r \mid \theta).$$

and two $\ell(\theta)$ functions related to two samples are different if and only if at least one n_j is different in the two subsets.

[5] Notice that, in our setting, the *Y*'s variables (or a subset of them) may also coincide with the *W*'s variables (or a subset of them).

No closed solution is available for $\ell'(\theta) = 0$, and the most accredited numerical procedure[6] for achieving ML estimates for mixtures distributions is the EM algorithm (Dempster *et al.*, 1977). It is almost surely convergent in any regular problem but it is extremely slow (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000).

In this regard, an important feature for its effective implementation is to find convenient starting points based on preliminary estimators (Karlis and Xekalaki, 2003) as we will discuss in the next sections.

3. A FIRST GROUP OF PRELIMINARY ESTIMATORS

In this section, we present preliminary estimators of θ, that may be derived from the characteristics of the probability distribution. They will be called *grid, naive* and *fzero estimators,* respectively.

3.1 *The grid estimator*

Given the finiteness of the parametric space $\Omega(\theta)$, one of the simplest way to start iterations is to compute the log-likelihood function over a grid of few admissible values for the parameters, and then to choose as starting values the one who maximizes the log-likelihood. This approach is data independent and has been advocated by Laird (1978) for finite mixtures.

For $CUB(0,0)$ models, after some trial and error experiments, we limited the computation of the log-likelihood functions over the set:

$$\Omega_0(\theta) = \{\pi: \ \pi = 0.2,\, 0.4,\, 0.8\} \times \{\xi: \ \xi = 0.3,\, 0.7\}.$$

The rationale for this choice is related to the circumstance that π estimates cause more problems.

Thus, this estimator is defined as:

$$(\pi^*, \xi^*) = \Omega_0(\theta) \underset{\Omega_0(\theta)}{argmax}\ \ell(\theta)\,.$$

We call it the *grid estimator of* θ.

3.2 *The naive estimator*

For a sample of ordinal values $r = (r_1, r_2, \ldots, r_n)'$, let n_r and $f_r = n_r/n$, $r = 1, 2, \ldots, m$ be the observed absolute and relative frequencies of $(R = r)$, respec-

[6] A Newton-Raphson procedure might be pursued, and the circumstance that elements of information matrix may be recursively computed is worth of interest.

tively. We denote the sample mean by:

$$\overline{R}_n = \frac{1}{n}\sum_{i=1}^{n} r_i = \frac{1}{n}\sum_{r=1}^{m} r\, n_r = \sum_{r=1}^{m} r f_r\,.$$

A *naive estimator* of θ has been defined in D'Elia and Piccolo (2005) as:

$$\bar{\pi} = 0.5; \qquad \bar{\xi} = \frac{m - \bar{R}_n}{m-1}\,;$$

The estimator of π is the central value of its admissible range and the estimator of ξ is the ML estimator conditioned to $\pi = 1$ (that is, as if only a shifted Binomial distribution were considered).

Hitherto, the *naive* estimator has been implemented in the available software (up to version 1.1) and they have brought ML estimators to convergence in hundreds of real applications and thousands of simulation experiments. In this respect, we experienced that no modification has been required in order to reach convergence in acceptable times, although these estimates are not fully efficient with respect to the information contained in the sample data.

3.3 *The fzero estimator*

A further estimator may be introduced if we start with a preliminary guess of π, and produces an estimates of ξ based on the expectation.

Let us denote the shifted Binomial distribution as:

$$b_r(\xi) = \binom{m-1}{r-1}(1-\xi)^{r-1}\xi^{m-r}, \quad r = 1, 2, \ldots, m.$$

Then, we let $f_0 = min(f_1, f_2, \ldots, f_m)$, where $\{f_r,\, r = 1, 2, \ldots, m\}$ are the relative frequencies of the sample data. Since $b_r(\xi) > 0$, $\forall r = 1, 2, \ldots, m$, it turns out that: $(1-\pi)/m < f_0$ and this implies $\pi > (1 - m f_0)$. Notice that f_0 is a consistent estimator of $p_0 = min(p_1, p_2, \ldots, p_m)$ and this inequality is only true in probability[7].

Thus, we preliminarily assume as estimate: $\pi^o = 1 - m f_0$; then, remembering the relationship between expectation and parameters:

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

we may infer a value of the ξ parameter, that is:

$$\xi^o = \frac{1}{2} + \frac{1}{\pi_0(m-1)}\left(\frac{m+1}{2} - \overline{R}_n\right).$$

[7] If this assertion should not be verified in sample experiment, we adopt the solution $\pi^o = max(0, 1 - m f_0)$.

Notice that this is the ML estimate of ξ conditioned to $\pi = \pi^o$.

Since $\pi^o > 0$, the proposal is defined for $f_0 \in [0, 1/m)$. Thus, for avoiding singularities in ξ, we will insert a small quantity ($\delta = 0.01$) and use $\pi^o = 1 - m(f_0 - \delta)$.

Finally, the *fzero estimators* are defined by:

$$\pi^o = 1 - m\,(f_0 - \delta); \quad \xi^o = \frac{1}{2} + \frac{(m+1)/2 - \overline{R}_n}{(m-1)[1 - m\,(f_0 - \delta)]}\,.$$

4. A SECOND GROUP OF PRELIMINARY ESTIMATORS

The estimators of this section are derived by statistical approaches, based on a *robust* version of least squares and *moments* methods, respectively.

4.1 *A robust least squares estimator*

The probabilities p_r of the *CUB* random variable may be written as:

$$p_r - \frac{1}{m} = \pi \left[b_r(\xi) - \frac{1}{m} \right], \quad r = 1, 2, \ldots, m.$$

Since relative frequencies f_r are consistent estimators of the corresponding probabilities p_r, we may suppose: $f_r = p_r + \epsilon_r$ for some errors ϵ_r, $r = 1, 2, \ldots, m$.

For a given ξ, this assumption specifies the following regression model (without intercept):

$$f_r - \frac{1}{m} = \pi \left[b_r(\xi) - \frac{1}{m} \right] + \epsilon_r, \qquad r = 1, 2, \ldots, m\,.$$

Now, the errors ϵ_r are correlated since $\sum_r f_r = \sum_r b_r(\xi) = 1$ and this implies $\sum_r \epsilon_r = 0$.

Then, for a given ξ, the ordinary least squares (OLS) estimator of π:

$$\pi_{OLS} = \frac{\sum_{r=1}^{m} \left(f_r - \frac{1}{m} \right) \left(b_r(\xi) - \frac{1}{m} \right)}{\sum_{r=1}^{m} \left(b_r(\xi) - \frac{1}{m} \right)^2} = \frac{\sum_{r=1}^{m} \left(\frac{f_r - \frac{1}{m}}{b_r(\xi) - \frac{1}{m}} \right) \left(b_r(\xi) - \frac{1}{m} \right)^2}{\sum_{r=1}^{m} \left(b_r(\xi) - \frac{1}{m} \right)^2}$$

is not particularly efficient. Now, the second expression shows that π_{OLS} can be expressed as a weighted average of the ratios:

$$\frac{f_r - \frac{1}{m}}{b_r(\xi) - \frac{1}{m}}, \qquad r = 1, 2, \ldots, m.$$

Generally, m is quite small, and the OLS estimator is very sensible to local deviations between observed and expected frequencies. Moreover, in case $| b_r(\xi) - \frac{1}{m} | \simeq 0$, for some r, the π_{OLS} estimator becomes an average of ratios that are potentially infinite.

As a consequence, we consider as a *robust* estimator[8] of π, the quantity:

$$\underset{r=1,2,\ldots,m}{median} \left\{ \frac{f_r - \frac{1}{m}}{b_r(\xi) - \frac{1}{m}} \right\}.$$

Indeed, ξ is not known and it may be preliminarily estimated as for the *naive* estimators (subsection 3.2).

Finally, for the joint estimation of the parameters, we suggest the following 2-steps algorithm:

1. $\xi^0 = \frac{m - \bar{R}_n}{m-1}$; $\qquad \pi^0 = \underset{r=1,2,\ldots,m}{median} \left\{ \frac{f_r - 1/m}{b_r(\xi^0) - 1/m} \right\}$;

2. $\xi^1 = \frac{1}{2} - \frac{\bar{R}_n - (m+1)/2}{(m-1)\pi^0}$; $\qquad \pi^1 = \underset{r=1,2,\ldots,m}{median} \left\{ \frac{f_r - 1/m}{b_r(\xi^1) - 1/m} \right\}$;

and the proposed estimators[9] are (π^1, ξ^1).

We call it the *robust estimator* of θ.

4.2 *A moments estimator*

If we compute the sample moments m_1, m_2 of the observed ordinal data, then the solutions $(\tilde{\xi}, \tilde{\pi})$ of the two non-linear equations:

$$E(R) = \mu_1(\pi, \xi) = m_1; \qquad E\left(R^2\right) = \mu_2(\pi, \xi) = m_2;$$

are expressed by:

$$\begin{cases} \tilde{\xi} = \dfrac{2(1 + 3m_1 m - m^2) - 3(m_1 + m_2) \pm \sqrt{\Delta}}{3(m-2)(2m_1 - m - 1)}, & \text{if } m_1 \neq \dfrac{m+1}{2}; \\ \tilde{\pi} = \dfrac{2m_1 - (m+1)}{(m-1)(1 - 2\hat{\xi})}, & \text{if } \tilde{\xi} \neq \dfrac{1}{2}; \end{cases}$$

---

[8] Standard conventions for the specification of median for m even apply. Of course, other robust proposals may be explored; for instance, a trimmed estimator based on the ordered $(m-2)$ central ratios is an interesting alternative.

[9] Of course, it is necessary to consider that if $\pi^0 = 0$ we must slightly modify π^1, to avoid division by zero.

where

$$\Delta = d_0 + d_1 m_1 + d_2 m_1^2;$$
$$d_0 = m^4 + (3m_2 + 1)m^2 + 3(3m_2 + 2)m + (4 + 6m_2 + 9m_2^2);$$
$$d_1 = -3(m + 1)[m(2m - 1) + 6(m_2 + 1)];$$
$$d_2 = 3(4m^2 + 2m + 7).$$

The solutions are a couple of two quadratic expressions for each parameter, and only admissible solutions have to be considered as in Piccolo (2003, pp. 99-103).

Instead, if the empirical distribution is perfectly symmetric, the solutions are:

$$\begin{cases} \tilde{\xi} = \dfrac{1}{2}, & \text{if } m_1 = \dfrac{m+1}{2}; \\ \tilde{\pi} = \dfrac{2(2m^2 + 3m + 1 - 6m_2)}{(m-1)(m-2)}, & \text{if } \tilde{\xi} = \dfrac{1}{2}. \end{cases}$$

In both cases, we let $\tilde{\theta} = (\tilde{\pi}, \tilde{\xi})'$ and call it the *moments estimator* of θ. These solutions are consistent for π and ξ.

5. COMPARING THE EFFICIENCY OF ESTIMATORS

In order to check the efficiency gained in the estimation routine by starting from a given preliminary estimator, we will implement a simulation experiment by running the EM algorithm and in turn using each of the previous proposals as starting points for the numerical procedure. Our experiment will be performed throughout the whole parametric space and we measure the efficiency as the gain in the elapsed time for convergence: since this time is proportional to the steps required for convergence, we will introduce a measure of *algorithmic relative efficiency* based on the number of iterations. The plan of the experiment is the following:

1. We let $m = 9$. Then, we choose 500 models with parameters (π, ξ) specified by a bivariate Uniform random scatter on the parametric space $\Omega(\theta)$. This option has been suggested by the circumstance that a systematic scanning favours some estimators (as *grid* and *naive*, for instance).
2. For each *CUB* model characterized by (π, ξ), we generate a sample of $n = 300$ observations[10]. The parameters are estimated by EM algorithm using each of the preliminary estimators (*grid, naive, fzero, robust, moments*) defined in section 4 as starting values. In addition, we record the number of iterations required to converge to ML estimates with a parameter tolerance less than 10^{-6}.
3. Step 2 is replicated 1000 times so that, for each preliminary estimator, 1000 si-

[10] The experiment has been performed for several sample sizes and the results are substantially confirmed (except for larger variability when n is a limited size). Moreover, in choosing sample sizes, one should consider that, $n \geq 20m$ is a safe criterion, as confirmed by D'Elia (2004).

mulated values of iteration number are obtained and the related average is computed. These averages are then ordered and a vector of ranks is obtained.

4. The frequency distribution and some descriptive statistics of these ranks for the 500 models are finally evaluated (Table 1).

TABLE 1. - *Frequency distributions of ranks of the preliminary estimators*

Estimators	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	*Average*	*Stand. dev.*
Grid	17	48	123	92	220	3.9000	1.1699
Naive	6	17	39	274	164	4.1460	0.7937
Fzero	0	173	286	41	0	2.7360	0.5992
Robust	5	239	47	93	116	3.1520	1.2697
Moments	472	23	5	0	0	1.0660	0.2860

The simulation shows that the preferred initial estimator should be the *moments* estimator as it is ranked the best in 94% of the simulated models, and never fourth or fifth. A second choice (but never the best) is the *fzero* estimator, always ranked between the second and fourth position, followed by *robust* and *grid* estimators. As expected, the *naive* estimator is ranked last.

Results of Table 1 confirm that a convenient strategy for accelerating the convergence of the EM algorithm is to use information contained in the sample data. Thus, it is important to study how this convenience changes over the parametric space. In this respect, we introduce a synthetic measure of efficiency for each estimated model.

Given a *CUB* model, we compute the median of the number of replications required for the convergence of each proposed preliminary estimator. We notice that, although average numbers could be biased by few anomalous situations, in our experiment, they are completely consistent with conclusions obtained by using medians. Then, we define the *algorithmic relative efficiency* of the *j*-th proposal est_j in relation to the *naive* estimators est_{naive} (that is, the worst) by comparing their median numbers of iterations required for convergence:

$$algreff\,(est_j) = 1 - \frac{median\,\#(est_j)}{median\,\#(est_{naive})}\ .$$

Notice that this quantity may be locally negative if, for some model, the specific estimator is worst than *naive* estimator.

In Figures 1-2, we plot these quantities (for *grid, fzero, robust* and *moments* estimators) for varying π and ξ, respectively, together a *lowess* regression line (Cleveland, 1981). We prefer these marginal plots as they are more informative with respect to surfaces over the parametric space.

From Figure 1, we notice a symmetric behaviour of the efficiency around $\pi = 0.5$ except for *robust* estimator, whose efficiency is monotonically increasing with π, but is positive only for $\pi > 0.5$. Moreover, the *moments* estimator produces a stable gain (about 20%) if $\pi < 0.5$ and this gain regularly increases with π (up to 40%).

From Figure 2, we notice that the behaviour of the efficiency with respect to ξ is different, with an apparent symmetry around $\xi = 0.5$ and a reduced efficiency for extreme values of the parameter. The *grid* estimator is never convincing but for values around $\xi = 0.3,\ 0.7$, where log-likelihoods were indeed preliminarily computed. Then, the gain of *fzero* estimator seems regularly positioned around 0.12. As far as the *robust* estimator is concerned its efficiency is positive only for $\xi \in (0.3,\ 0.7)$ and it reaches a gain of 0.3 when $\xi = 0.5$. Instead, the *moments* estimator is regularly more efficient over the whole parametric space, exceeding a gain of 30% around $\xi = 0.5$.

For a comprehensive and synthetic visualization of the previous features, in Figures 3-4 we compare the smoothed efficiency of the estimators over the π and ξ ranges, respectively. It turns out that *moments* estimators are monotonically superior to any other proposal. The second best seems to be the *fzero* estimator but with an efficiency reduced by more than 50%.

The main conclusions have been confirmed when we replicated the simulation experiment for different values of m and n.

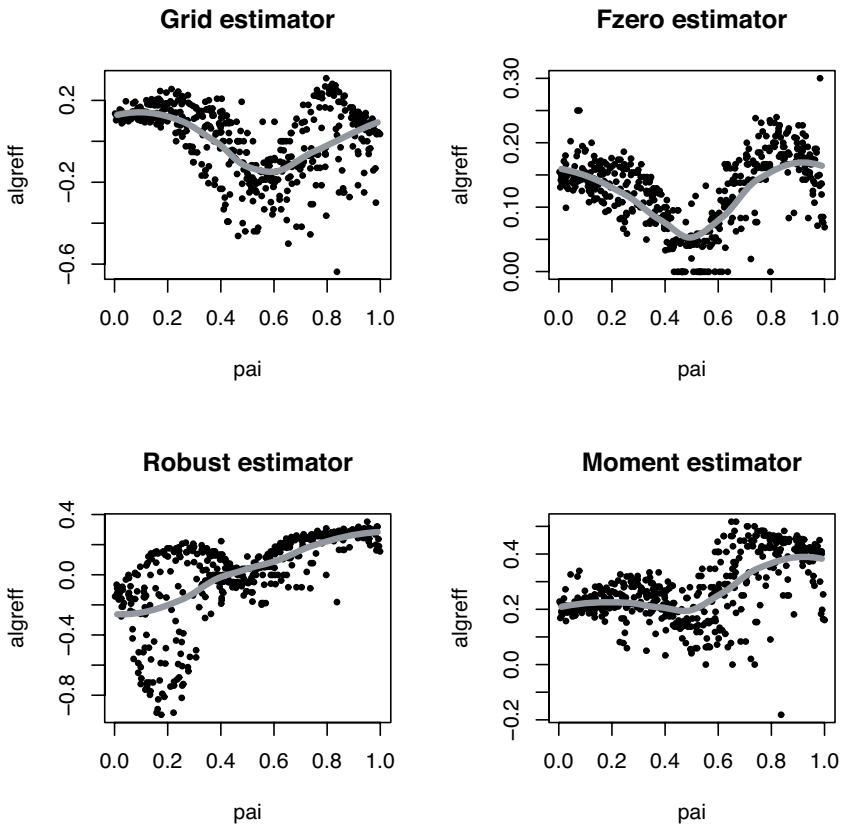

FIGURE 1. - *Algorithmic relative efficiency with respect to the parameter* π

6. SOME EXTENSIONS TO *CUB* MODELS WITH COVARIATES

It is not immediate to extend the previous results to general *CUB* models with covariates. Thus, we will discuss firstly the case where a dummy covariate is included for explaining π and/or ξ; then, we will add few considerations for a general approach (that deserves further studies).

A dummy covariate D_i is useful for explaining/testing different effects on the ordinal responses according to the membership of the i-th subject S_i to one of k groups (generally, $k = 2$) and also for discriminating purposes (Iannario, 2008). Specifically, in the case of two groups G_0 and G_1, we formally define:

$$D_i = \begin{cases} 0, & \text{if } S_i \in G_0; \\ 1, & \text{if } S_i \in G_1; \end{cases} \qquad i = 1, 2, \ldots, n\,.$$

Moreover, if this membership is relevant for explaining a different effect of the uncertainty and/or the feeling components, we specify a *CUB* model where the corre-

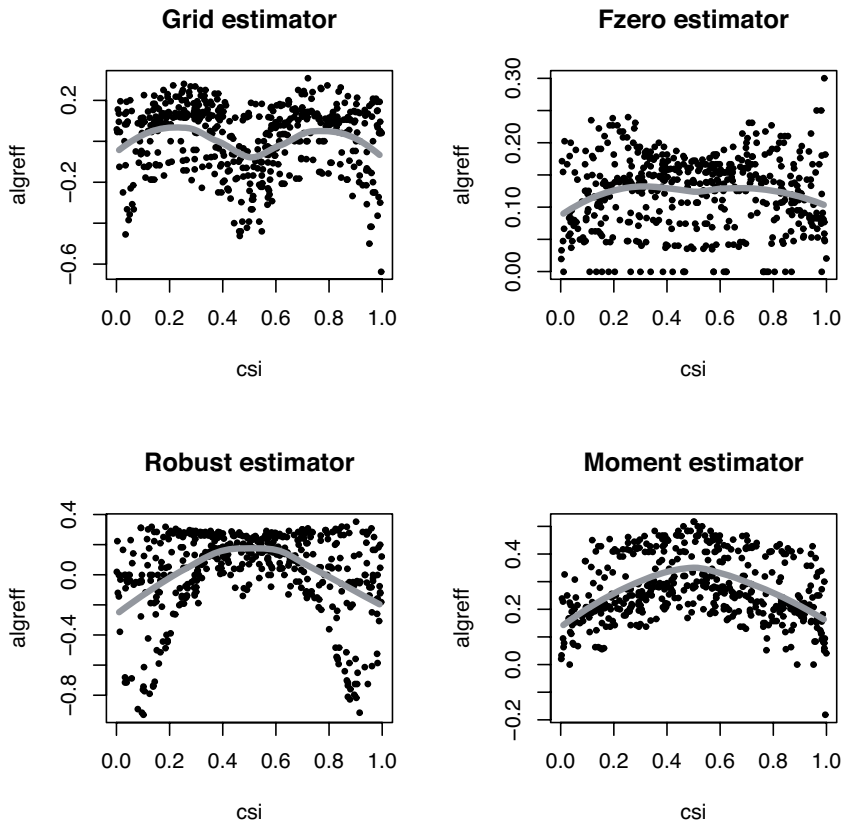

FIGURE 2. - *Algorithmic relative efficiency with respect to the parameter* ξ

sponding parameters are logistic functions of the dummy covariate[11], that is:

$$\pi_i = \frac{1}{1+e^{-(\beta_0+\beta_1 D_i)}}\,; \qquad \xi_i = \frac{1}{1+e^{-(\gamma_0+\gamma_1 D_i)}}\,, \qquad i = 1, 2, \ldots, n.$$

Previous definitions imply :

$$(\pi_i|S_i \in G_1) = \pi_1 = \frac{1}{1+e^{-(\beta_0+\beta_1)}}; \qquad (\pi_i|S_i \in G_0) = \pi_0 = \frac{1}{1+e^{-\beta_0}};$$

$$(\xi_i|S_i \in G_1) = \xi_1 = \frac{1}{1+e^{-(\gamma_0+\gamma_1)}}; \qquad (\xi_i|S_i \in G_0) = \xi_0 = \frac{1}{1+e^{-\gamma_0}};$$

After a simple algebra, we get:

$$\beta_0 = \log\left(\frac{\pi_0}{1-\pi_0}\right); \qquad \gamma_0 = \log\left(\frac{\xi_0}{1-\xi_0}\right).$$

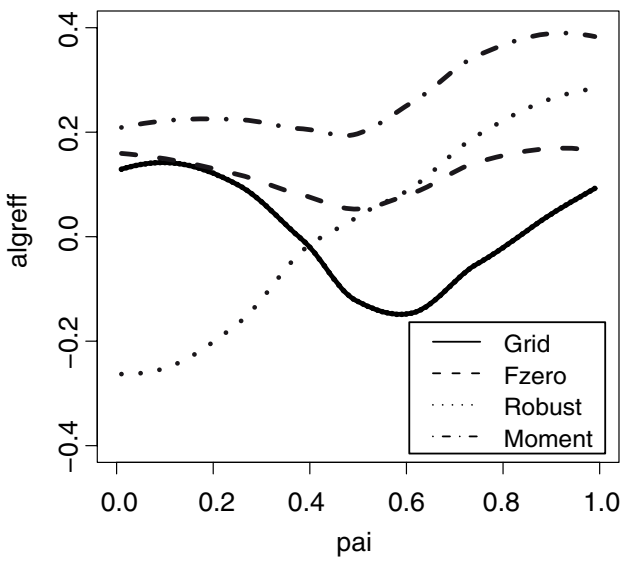

FIGURE 3. - *Smoothed algorithmic relative efficiency for varying* π

[11] In order to simplify the notation, we are supposing that the same covariate D_i is able to explain the behaviour of both parameters, if necessary. In fact, it will become evident that expressions for preliminary estimators are completely disjoint as far as different covariates of π and ξ parameters are concerned.

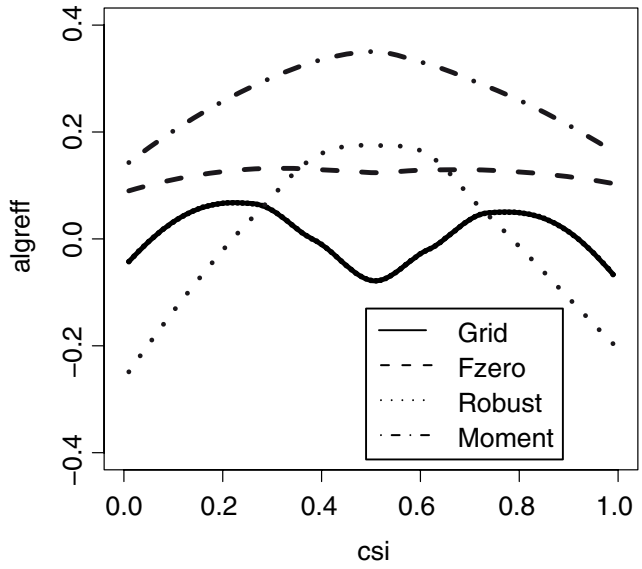

FIGURE 4. - *Smoothed algorithmic relative efficiency for varying* ξ

and, similarly:

$$\beta_1 = \log\left(\frac{\pi_1}{1-\pi_1}\right) - \log\left(\frac{\pi_0}{1-\pi_0}\right) = \log\left(\frac{\pi_1}{1-\pi_1}\right) - \beta_0\,;$$

$$\gamma_1 = \log\left(\frac{\xi_1}{1-\xi_1}\right) - \log\left(\frac{\xi_0}{1-\xi_0}\right) = \log\left(\frac{\xi_1}{1-\xi_1}\right) - \gamma_0\,.$$

Our proposal is to split the observed sample according to the groups and to get preliminary estimators of π_i, ξ_i, $i = 0, 1$, respectively; as a consequence of the findings of section 5, we will use the *moments* estimator. Then, from the previous equations, we obtain preliminary estimators for β_0, β_1 and γ_0, γ_1, respectively. This proposal may be easily implemented in current software.

It is not immediate to extend these results for getting preliminary estimators of β and γ vectors in a general *CUB* model. In this area, promising results have been obtained for γ by using a multiple regression device on observed r_i, $i = 1, 2, \ldots, n$ and further research is currently in progress. However, we think that cumbersome and *ad hoc* solutions are not always justified since the almost sure convergence of EM algorithm to the ML estimates is warranted even if we start with *naive* preliminary estimators.

7. EMPIRICAL EVIDENCES ON REAL DATA SETS

For checking the effectiveness of the proposed approach for dummy covariates, we report the results we have obtained on three different data sets. The first two cases

refer to ranking data whereas the last one is concerned with rating data; thus, interpretation of the estimated models should be cautiously considered since, in any case, we are considering the marginal distribution of interest. In any instance, we denote by n_0 and n_1 the sample size of the groups corresponding to value 0 and 1 of the dummy covariate, respectively.

7.1 *Preference towards Black color*

Data set *COLORS* has been collected during 1998 and it is composed by a sample of $n = 169$ people which ranked their preferences towards a list of $m = 12$ colors. In the experiment, $R = 1$ means the best and $R = m$ the worst.

In this case study, we present only the results obtained by the marginal distribution of the observed ranks of *Black* color. We found that a *CUB* model without covariates, with a log-likelihood of $\ell(\hat{\theta}) = -374.596$, gives significant estimates[12] of $\hat{\pi} = 0.421\,(0.056)$ and $\hat{\xi} = 0.949\,(0.014)$. This denotes a high preference towards *Black* color[13] with a limited uncertainty and a fairly good fitting[14] (measured by $Diss = 0.104$).

Among the relevant subjects' covariates we found that π parameter was related to Smoking habit ($= 0$, for no-smokers, $= 1$, for smokers). The resulting $CUB(1,0)$ model was significant[15] and the log-likelihood increased up to $\ell(\hat{\theta}) = -370.135$.

According to the findings of section 6, we report in Table 2 (*moments*) ($\tilde{\theta}$) and ML ($\hat{\theta}$) estimates. Although our sample size is not particularly large (with respect to m), the comparison seems particularly satisfactory and the usefulness of *moments* preliminary estimates is worth of consideration.

TABLE 2. - *Black color preferences: CUB(1,0) model*

Estimates	$(m = 12;\ n_0 = 111;\ n_1 = 58)$		
Preliminary (moments)	$\tilde{\beta}_0 = -0.642$	$\tilde{\beta}_1 = 1.667$	$\tilde{\xi} = 0.920$
Maximum Likelihood	$\hat{\beta}_0 = -0.791$	$\hat{\beta}_1 = 1.313$	$\hat{\xi} = 0.948$

[12] Values in parentheses denote standard errors of estimates.

[13] More than 45% of respondents ranked Black as the best or the second best color, with a sharp mode at $(R = 1)$. It is worthwhile to observe that the average rank of 4.248 would be unable to express such a high preference.

[14] For *CUB* models we prefer to assess the fitting by a normalized dissimilarity index $Diss \in [0,1]$ defined as one half the sum of absolute differences among estimated probabilities and observed relative frequencies.

[15] Hereafter, we are exploiting asymptotic results by comparing twice the difference in log-likelihoods functions with quantiles of a Chi-square random variable with degrees of freedom given by the difference of parameters of the models to be compared.

7.2 *Concern for urban environmental pollution*

Data set *EMERGENCIES* has been collected in the urban area of Naples during the years 2004-2006, and the whole sample size is $n = 773$ people. The respondents were asked to rank their concerns towards a list of $m = 9$ main problems (called ''emergencies'') they are usually faced to. In this case study, we limit ourselves to study the ordinal values which the sampled respondents expressed with regard to *Environmental pollution.* Then, observed data are the marginal distribution of this specific emergency.

The parameter estimates of the *CUB* model without covariates are $\hat{\pi} = 0.796\,(0.030)$ and $\hat{\xi} = 0.258\,(0.008)$, with a log-likelihood $\ell(\hat{\theta}) = -1483.315$ and a fairly good fitting (measured by $Diss = 0.102$). The model suggests low uncertainty in the responses and a moderate worry about *Environmental pollution.*

Specifically, it turned out that concern is different for men and women; thus, a dummy covariate Gender (=0, for men, =1, for women) has been introduced for explaining the behavior of the ξ parameter by means of a $CUB(0,1)$ model. Gender has been found significant yielding an increase in log-likelihood up to $\ell(\hat{\theta}) = -1465.270$.

In Table 3, we report both preliminary $(\tilde{\theta})$ and ML $(\hat{\theta})$ estimates of the parameters. Note that our sample is moderately large, thus, the proposed initial values are quite effective to improve the convergence towards ML estimates.

TABLE 3. - *Concern for urban Environmental pollution: CUB(1,0) model*

Estimates	$(m = 9;\ n_0 = 446;\ n_1 = 327)$		
Preliminary (moments)	$\tilde{\pi} = 0.826$	$\tilde{\gamma}_0 = -0.918$	$\tilde{\gamma}_1 = -0.417$
Maximum Likelihood	$\hat{\pi} = 0.804$	$\hat{\gamma}_0 = -0.857$	$\hat{\gamma}_1 = -0.487$

7.3 *Evaluation of University orientation services*

Data set *ORIENTATION* is the result of yearly surveys that during 2002-2004 were carried out in order to check the satisfaction of students towards Orientation services held at the University of Naples Federico II. The questionnaires refer to several aspects of the service (willingness, timetable, competence, and so on) and it was asked to rate the perceived satisfaction on a 7-point scale. Here, we will consider only the answers to the ''Global satisfaction'' item collected during 2003, and our sample consists of $n = 2457$ respondents.

The ordinal scores were fitted by a $CUB(0,0)$ model and we obtained the estimates $\hat{\pi} = 0.904\,(0.012)$ and $\hat{\xi} = 0.202\,(0.004)$. They denote a high preference towards the service and a very low uncertainty. The log-likelihood was $\ell(\hat{\theta}) = -3720.285$, with a sensible fitting of the proposed model to empirical data $(Diss = 0.050)$.

An important covariate for discriminating the answers was the attendance of the

service, as it turned out that a high frequency modified the expressed feeling. Thus, we introduce a dummy covariate for *Usage* (= 0, if rare and occasionally, = 1, if frequent and regularly) to explain the diversity of ξ_i parameters among respondents. The corresponding estimates were significant and the log-likelihood of the implied $CUB(0,1)$ model raised up to $\ell(\hat{\theta}) = -3612.769$, denoting a quite sensible improvement.

In Table 4 we report the main findings by comparing the preliminary and final (ML) estimates of the parameters of $CUB(0,1)$ model. In this case study, the sample size is very large and we find a high closeness between these estimates.

TABLE 4. - *Evaluation of University orientation services: CUB(0,1) model*

Estimates	$(m = 7;\ n_0 = 1657;\ n_1 = 800)$		
Preliminary (moments)	$\tilde{\pi} = 0.920$	$\tilde{\gamma}_0 = -1.141$	$\tilde{\gamma}_1 = -0.771$
Maximum Likelihood	$\hat{\pi} = 0.917$	$\hat{\gamma}_0 = -1.138$	$\hat{\gamma}_1 = -0.775$

7.4 *Closeness of preliminary estimates*

Results discussed in section 5 showed that the gain obtained by introducing the *moments* estimator is quite sensible as an accelerating device for the EM algorithm.

To quantify these aspects on the previous data sets, we compute the Euclidean distance $d(\theta_0, \hat{\theta})$ between the starting points:

$$\theta_0 = (0.1,\ 0.1,\ 0.1)'$$

and the final ML estimates $(\hat{\theta})$, and the Euclidean distance $d(\tilde{\theta}, \hat{\theta})$ between the proposed preliminary estimates $(\tilde{\theta})$ and the ML estimates. Indeed, the ratio:

$$\rho = 1 - \frac{d(\tilde{\theta}, \hat{\theta})}{d(\theta_0, \hat{\theta})}$$

is a relative measure of *ML-closeness* of our proposal. This quantity is shown in Table 5 for the data sets we have tested in previous subsections.

TABLE 5. - *Euclidean Distances and closeness among initial and ML estimates*

Data sets	$d(\theta_0, \hat{\theta})$	$d(\tilde{\theta}, \hat{\theta})$	ρ	n
COLORS $(m = 12)$	1.728	0.385	0.777	169
EMERGENCIES $(m = 9)$	1.325	0.095	0.928	773
ORIENTATION $(m = 7)$	1.722	0.006	0.997	2457

It is evident that consistent preliminary estimators improve convergence towards ML estimates by reducing the distance more than 70%, even for moderate sample

sizes. Although limited to few data sets, the results obtained when a single dummy covariate is present seem quite encouraging.

7.5 *Extreme data and preliminary estimates*

A different problem that motivates the use of more accurate preliminary estimates is generated by data characterized by behaviour of respondents not so sharp. In these cases, bad starting values may induce convergence to some parameter values that cause statistical problems in the inferential steps.

We experienced this problem with a real data set where $m = 7$ and frequency vector: $(69, 33, 63, 50, 40, 51, 44)'$. Then, *naive* estimates (given by $\bar{\pi} = 0.5$; $\bar{\xi} = 0.52952$) generate a sequence of iterations that converged to a solution with a non-definite positive information matrix.

Instead, by starting iterations at preliminary *moments* estimates (given by $\tilde{\pi} = 0.07562$; $\tilde{\xi} = 0.89280$), the EM algorithm converges in 46 iterations to the ML estimates of a very peculiar *CUB* model with $\hat{\pi} = 0.063$; $\hat{\xi} = 1.000$. In fact, the estimated model is a convex combination of a degenerate random variable at $(R = 1)$ and a discrete Uniform with $m = 7$. However, fitting is quite good $(Diss = 0.067)$ and the model correctly accounts for a diffuse uncertainty in the responses.

8. CONCLUSIONS

The gain in the number of iterations generally obtained by preliminarily using the *moments* estimator is a positive result of this work; moreover, the method may be easily implemented in current software when a dummy covariate is present.

Furthermore, good starting values are necessary when extreme data sets have to be analyzed by EM algorithm. This event is rare in real data sets but it should be well considered if one undertakes a massive simulation experiment.

A final aspect is related to the selection of covariates for explaining the parameters of models for ordinal variables. Although the main result of this paper may be directed to improve preliminary estimates *per se*, they might be also used for a quick inference on the importance of covariates for uncertainty and feeling components, *before* or even *without* considering final ML estimates.

In this perspective, the *moments* estimator should become a standard tool for preliminary specifications of *CUB* models and/or comparisons of parameters in different clusters.

ACKNOWLEDGEMENTS

The work has been partially supported by PRIN-2006 research project: "Stima e verifica di modelli statistici per l'analisi della soddisfazione degli studenti universitari" and benefited from research structures of CFEPSR, Portici. Critical suggestions by referees are gratefully acknowledged.

RIASSUNTO

Il contributo affronta il problema della scelta dei valori iniziali da utilizzare per incrementare l'efficienza della procedura EM per la stima dei parametri di modelli CUB per variabili ordinali. Nel lavoro sono presentati i risultati di un esteso esperimento predisposto per confrontare le performance di 5 stimatori preliminari -tre basati su una diretta interpretazione probabilistica relativa alla tipologia dei parametri, due su un approccio tipicamente inferenziale- in termini di numero di iterazioni necessarie per la convergenza dell'algoritmo EM. Dall'analisi emerge chiaramente che lo stimatore preliminare più efficiente è quello ottenuto con il metodo dei momenti. Il lavoro considera poi una prima estensione al caso dei modelli CUB con una variabile dummy e presenta alcuni risultati ottenuti con dataset reali.

REFERENCES

Agresti A. (2002). *Categorical Data Analysis*, 2nd edition, J. Wiley & Sons, New York.

Bock R.D., Moustaki I. (2007). Item response theory in a general framework. In C.R. Rao and S.Sinharay (Eds.), *Psychometrics*, Handbook of Statistics, **26**, 469-513.

Brentari E., Golia S., Manisera M. (2007). Models for categorical data: a comparison between the Rasch model and nonlinear principal component analysis. *Statistica & Applicazioni*, **V**, 53-77.

Cleveland W.S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, **35**, 54.

Dempster A.P., Laird N.M., Rubin D.B. (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society*, B, **39**, 1-38.

D'Elia A. (2004). Finite sample performance of the E-M algorithm for ranks data modelling. *Statistica*, **LXIII**, 41-51.

D'Elia A., Piccolo D. (2005). A mixture model for preference data analysis. *Computational Statistics & Data Analysis*, **49**, 917-934.

Dobson A.J., Barnett A.G. (2008). *An Introduction to Generalized Linear Models*, 3nd edition, Chapman & Hall/CRC, Boca Raton.

Faraway J.J. (2006). *Extending the Linear Model with R*, Chapman & Hall/ CRC, Boca Raton.

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

Iannario M. (2008). Dummy covariates in *CUB* models, *Statistica*, **LXVIII**, 2, forthcoming.

Iannario M. (2009a). On the identifiability of a mixture model for ordinal data, *submitted for publication.*

Iannario M. (2009b). Modelling shelter choices in ordinal data surveys, *submitted for publication.*

Iannario M., Piccolo D. (2009). A new statistical model for the analysis of Customer Satisfaction. *Quality Technology & Quantitative Management*, forthcoming.

Karlis D., Xekalaki E. (2003). Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, **41**, 577-590.

King G., Tomz M., Wittenberg J. (2000). Making the most of statistical analyses: improving interpretation and presentation. *American Journal of Political Science*, **44**, 341-355.

Laird N.M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805-811.

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

Mazzali A. (2004). Una generalizzazione della distanza di Cayley per l'analisi dei ranghi. *Statistica & Applicazioni*, **II**, 1, 73-88.

McCullagh P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society*, B, **42**, 109-142.

McCullagh P., Nelder J.A. (1998). *Generalized linear models*, 2^{nd} edition, Chapman & Hall, London.

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.

Nelder J.A., Wedderburn R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, A, **135**, 370-384.

Pagani L., Zanarotti M.C. (2003). Analisi della qualità di un servizio: un confronto tra scale mediante il modello di Rasch. *Statistica & Applicazioni*, **I**, 2, 20-39.

Piccolo D. (2003). On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, **5**, 85-104.

Piccolo D. (2006). Observed information matrix for *MUB* models. *Quaderni di Statistica*, **8**, 33-78.

Piccolo D., D'Elia A. (2008). A new approach for modelling consumers' preferences. *Food Quality and Preferences*, **19**, 247-259.

Piccolo D., Iannario M. (2008). A package in R for *CUB* models inference, Version 1.1, available at *www.dipstat.unina.it.*