Adv Data Anal Classif (2012) 6:163–184
DOI 10.1007/s11634-012-0111-5

REGULAR ARTICLE

Preliminary estimators for a mixture model of ordinal data

Maria Iannario

Received: 14 November 2010 / Revised: 5 April 2012 / Accepted: 18 July 2012 /
Published online: 23 August 2012

Abstract In this paper, we propose preliminary estimators for the parameters of a mixture distribution introduced for the analysis of ordinal data where the mixture components are given by a Combination of a discrete Uniform and a shifted Binomial distribution (CUB model). After reviewing some preliminary concepts related to the meaning of parameters which characterize such models, we introduce estimators which are related to the location and heterogeneity of the observed distributions, respectively, in order to accelerate the EM procedure for the maximum likelihood estimation. A simulation experiment has been performed to investigate their main features and to confirm their usefulness. A check of the proposal on real case studies and some comments conclude the paper.

Keywords Ordinal data · CUB models · Preliminary estimators

Mathematics Subject Classification 6207 · 62E17 · 62F10

1 Introduction

Ordinal data are ubiquitous in several empirical studies concerning different disciplines: Economics and Politics, Medicine and Sociology, Marketing and Linguistics (Agresti 2010). Generally, such data are generated by respondents to a survey who select a graduated category in a definite Likert scale. Classical methodologies for fitting and interpreting such models are included in the framework of Generalized Linear Models (GLM). More specifically, cumulative probabilities are introduced for relating the responses to the characteristics (covariates) of subjects (McCullagh 1980; McCullagh and Nelder 1989).

M. Iannario (✉)
Department of Statistical Sciences, University of Naples Federico II,
Via Leopoldo Rodinò 22, 80138 Napoli, Italy
e-mail: maria.iannario@unina.it

A different approach for analysing ordinal data consists in finite mixtures, as in Wedel and DeSarbo (1995), Greene and Hensher (2003), Grün and Leisch (2008) and Breen and Luijkx (2010), among others. Generally, such proposals are introduced as generalizations of GLM and are obtained as convex combination of probability distributions belonging to the same class of models. These approaches generally assume the existence of some subgroups whose responses should be differently modelled.

Mainly motivated by psychological arguments, another mixture distribution has been proposed by Piccolo (2003). He interprets responses as the result of a complex mechanism whose main components are the *feeling* of the subject towards the item and an intrinsic *uncertainty* surrounding the modality of the choice. Thus, this class of models combines a Binomial distribution with a discrete Uniform distribution. These models are parsimonious with respect to parameters and flexible for fitting empirical data with different shapes. Inferential issues have been developed for their implementation and many applications support their usefulness in different fields (Iannario and Piccolo 2009).

In this paper we consider the EM algorithm for finding the maximum likelihood (ML) estimates of the parameters in this mixture model and propose using a special type of preliminary parameter estimates as initial values in the EM algorithm in order to increase the efficiency and speed of maximum likelihood (ML) estimation. To this end, we relate the two parameters of the mixture to a location and a heterogeneity measure, respectively. As a consequence, given the correspondence between the modal value of the mixture and the modal value of the first component we suggest the sample mode as the starting point of a possible estimate of the first parameter. Similarly, since we show that the weight of the mixture is dependent upon a heterogeneity measure (as Gini index) we derive a corresponding relationship for it, conditional on the value of the first parameter.

A peculiar characteristic of these proposals consists in the fact that both estimates are based on the frequency distribution of data. Then, these estimators are derived by taking the qualitative nature of ordinal data into account such that we do not require any arbitrary transformation of responses.

The paper is organized as follows: in the next section we set notations for the class of models we will discuss and briefly comment on their properties. Then, in Sect. 3 we derive a relationship between the mode and the feeling measure; similarly, in Sect. 4, we obtain a relationship between an index of heterogeneity and the uncertainty. In Sect. 5, we assess the joint preliminary estimation of model parameters and discuss their major distributional properties in terms of bias and variability. We also discuss several problems occurring in some extreme situations. In Sect. 6, the improvement gained by the new proposal with respect to some alternatives is compared in terms of computational efficiency. In Sect. 7, the performance of the proposed estimators has been checked on some real case studies. Section 8 provides some concluding remarks.

2 Specification and inference for CUB models

The stochastic structure we will deal with has been denoted CUB model (the acronym stems from the ***C***ombination of a discrete ***U***niform and a shifted ***B***inomial random variable). This class of models has been introduced by Piccolo (2003), further developed in D'Elia and Piccolo (2005) and Iannario (2012), and generalized by using a direct (logistic) link among parameters and subjects' covariates.

Although in this paper we limit ourselves to consider CUB models without covariates, we report that substantial improvements may be achieved both in interpretation and fitting of real data by using CUB models with covariates (Iannario and Piccolo 2012).

With respect to the classical GLM framework, CUB models are specified in terms of probabilities, require just two parameters to fit discrete distributions with different shapes and offer a simple visual interpretation of the results by means of a graphical display on the parametric space (Piccolo and D'Elia 2008).

Their main motivation derives from an interpretation of the behaviour of respondents when faced with multiple choices concerning evaluation and/or preferences. The selection of an ordered modality among several ones is a very complex mental process since it involves several factors influencing the final choice. Thus, a simplified version of such a psychological process should limit the analysis only to relevant components. In CUB models, the attractiveness (repulsion) towards the item and the indecision (fuzziness) in the response have been considered as the relevant ones. In fact, *feeling* is an internal/personal attitude concerning the opinion of the subject towards the object whereas *uncertainty* pertains to the operational modes of the final choice and sometimes to the external circumstances surrounding the final decision. In a sense, we may interpret feeling as a true latent variable while uncertainty is an inherent component of human decisions; thus, it deserves more consideration than just a residual analysis.

In the following, we refer to experimental surveys where people are asked to give ordered responses by means of a (Likert) m-point rating scale made up by numbers or ordinal attributes or categories. Then, for a given number of levels $m > 3$, we define a random variable R to be a CUB variable if and only if its probability mass function is defined by:

$$Pr\,(R = r \mid \boldsymbol{\theta}) = \pi\, b_r(\xi) \;+\; (1-\pi)\, U_r, \quad r = 1, 2, \ldots, m, \tag{1}$$

where we denoted the parameter vector by $\boldsymbol{\theta} = (\pi, \xi)'$ and

$$b_r(\xi) = \binom{m-1}{r-1}\, \xi^{m-r}\, (1-\xi)^{r-1}; \quad U_r = \frac{1}{m}, \quad r = 1, 2, \ldots, m, \tag{2}$$

are the shifted Binomial and discrete Uniform components of the mixture, respectively.

Since $Pr\,(R = r \mid \boldsymbol{\theta}) = p_r(\boldsymbol{\theta})$ is well defined when $\pi \in (0, 1]$ and $\xi \in [0, 1]$, the parametric space is the (left open) unit square:

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

It has been proved (Iannario 2010) that CUB models are identifiable for any $m > 3$.

The main interest in pursuing such an approach stems from the parsimony of parameters and their interpretation of the choice process. In fact, given the one-to-one correspondence among CUB models and points belonging to $(0, 1] \times [0, 1]$, such models may be graphically displayed in the unit square. This representation may help in several circumstances, and we list some of them:

- If CUB models are fitted for different subgroups of respondents and their parameter estimates are reported on $\Omega(\boldsymbol{\theta})$, the closeness or the distance among the points may help to detect similarity or dissimilarity in the behaviour of the subgroups.
- If CUB models are fitted before and after some interventions, the possible displacement of the points may be tested and assigned to a significant effect of the known actions.
- If CUB models are fitted for a given item in different times, the trajectory of the points during the time periods may easily show the dynamic direction of the feeling and/or uncertainty of respondents.

Thus, this representation may usefully synthesize hundreds or thousands of responses just by plotting their main indicators derived by modelling framework.

Although R is related to an ordinal variable (not necessarily a numeric one), we report that the expectation of R is given by:

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

The exact meaning of the parameter ξ changes with the setting of the analysis and, given that R is a reversible random variable, it mostly depends on how the ratings are codified (the larger the rating the larger the feeling, or vice versa). Thus, according to the context, the ξ parameter may be related to a *degree of perception*, a *measure of closeness*, a *level of satisfaction*, an *assessment of proficiency*, a *rating of concern*, an *index of selectiveness*, a *pain threshold*, a *subjective probability*, and so on.

Usually, high values of the responses imply high appreciation (rating) towards the object. Thus, the quantity $(1-\xi)$ increases with agreement towards the item whereas uncertainty of the choice increases with $(1-\pi)$. We may deduce that ξ is related to the predominance of unfavourable responses (that is, lower than the midrange) and this affects the skewness of the distribution (Lemma 1, Appendix).

The two parameters play a different role in determining the shape and the interpretation of the mixture. Briefly, the parameter ξ may be interpreted as mostly related to location measures and strongly determined by the skewness of responses: it increases when respondents prefer low ratings, and vice versa. On the contrary, the parameter π adds dispersion to the shifted Binomial distribution and it increases the amount of frequencies in each category; thus, it modifies the heterogeneity of the distribution.

Specification, estimation and testing issues for CUB models are discussed by Piccolo (2006). Moreover, asymptotically efficient estimators of the parameters are provided by maximum likelihood (ML) methods and, for discrete distributions, this is equivalent to minimize a discrepancy measure between the theoretical model and the observed data (Gourieroux and Monfort 1995, p.165). In this regard, the EM procedure may be considered a safe algorithm to reach convergence almost everywhere on $\Omega(\boldsymbol{\theta})$. Since this procedure is extremely slow, it is important to accelerate its convergence by giving adequate initial values; thus, the determination of starting values is a serious problem, as emphasized by McLachlan and Peel (2000, pp. 47–49) and Karlis and Xekalaki (2003) and Biernacki et al. (2003).

A focal point in this respect is the existence and uniqueness of ML estimators as we do, in fact, conjecture for CUB models. Although we do not know of a formal proof of this assertion, it may be confidently assumed as supported by a huge amount of real data set, several computations with extreme configuration of sample data and millions of simulations. Thus, the short run strategy advocated by Biernacki et al. (2003), and efficiently performed when multiple maxima are present, is not used in our context. In addition, given a set of sample data (generated by CUB models of different shapes), we obtain the same final estimates even with random starting values: as will be seen by our experiments in Sect. 6, random initial estimates slow down the iteration times but do not miss the convergence to ML estimates.

3 Location measures as estimators of parameter ξ

We denote by $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)'$ the collection of responses assigned by n raters to a given object/service/item. Then, $\boldsymbol{r}$ is the observed sample of size n generated by a *random sample* $(R_1, R_2, \ldots, R_n)$, where R_i, $i = 1, 2, \ldots, n$ are independent and identically distributed random variables as (1), each one characterized by the parameter vector $\boldsymbol{\theta}$.

As already discussed in Sect. 2, for a given $m > 3$, the parameter ξ completely characterizes the shifted Binomial component and it is inversely associated with feeling (positive evaluation increases with $1 - \xi$); thus, when $\xi \to 0$ ($\xi \to 1$) most of probability mass moves towards high (low) categories. As a consequence, preliminary estimators T_n of ξ may be obtained as linear functions of a location measure L_n, that is: $T_n = a + b\,L_n$, $b < 0$. If we specify L_n as the sample mean or mode (or even sample median or trimmed means), we get different estimators.

The simplest proposal is a (negative) linear function of the sample mean $\overline{R}_n$ of the observed ratings R_i, $i = 1, 2, \ldots, n$:

$$\overline{\xi} = \frac{m - \overline{R}_n}{m - 1}. \tag{3}$$

This statistic is both moments and maximum likelihood estimator of ξ in a CUB model *if and only if* $\pi = 1$ (that is, if the mixture degenerates into a shifted Binomial random variable).

The estimator $\overline{\xi} \in [0, 1]$ is always well defined; however, since

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

it is biased for ξ (even asymptotically), for any CUB model with $\pi < 1$. The bias has the sign of the skewness of distribution, since it depends on $(\frac{1}{2} - \xi)$. Although $Var(\overline{\xi}) = O(n^{-1})$, $\overline{\xi}$ is not a consistent estimator of ξ since the mean square error of $\overline{\xi}$ does not converge to 0.

To introduce an alternative proposal, we observe that the mode Mo of a CUB random variable coincides with the mode of the shifted Binomial component of the mixture (Lemma 2, Appendix):

$$Mo = 1 + [m(1-\xi)]\,;$$

here, $[x]$ denotes the maximum integer contained in x. Of course, Mo assumes only integer values on the support.

Since the mode of CUB models is invariant with respect to π, such location measure may be proposed for inferring on the parameter ξ. In fact, *if the mode is unique*, the correspondence among ξ and the population mode Mo is given by:

$$Mo = j \Longleftrightarrow \xi \in \left[\frac{m-j}{m}, \frac{m-j+1}{m}\right), \quad j = 1, 2, \ldots, m.$$

As a consequence, even if one knows exactly the mode Mo of this random variable, it is only possible to detect a range of admissible values of the true parameter ξ as determined by the real interval $\mathscr{I}(Mo)$:

$$\xi \in \mathscr{I}(Mo) = \left[\frac{m-Mo}{m}, \frac{m-Mo+1}{m}\right], \quad Mo = 1, 2, \ldots, m.$$

This approach based on the mode generates infinite solutions for ξ and we discuss only the simplest one as a convenient preliminary estimator of ξ.

Specifically, given a random sample $(R_1, \ldots, R_n)$ and the corresponding multinomial samples of absolute $(N_1, \ldots, N_m)$ and relative frequencies $(\mathscr{F}_1, \ldots, \mathscr{F}_n)$, respectively, we define the *sample mode* M_n as the unique integer value (if exists) of the support where the frequency is a maximum. Formally, we set:

$$M_n = j \quad \Longleftrightarrow \quad \mathscr{F}_j > \mathscr{F}_k, \quad \forall k \neq j = 1, 2, \ldots, m.$$

Given that the sample mode may be unique, coincident on adjacent values or coincident in different values, different strategies are necessary to cope with this problem:

- If the sample mode is *unique*, we propose as a point estimator the midpoint of the interval $\mathscr{I}(Mo)$, that is:

$$\tilde{\xi} = \frac{m - M_n + 0.5}{m} = 1 - \frac{M_n - 0.5}{m}. \tag{4}$$

- If $m(1-\xi)$ is an integer value, Binomial distribution supports two *adjacent modes* at $R = m(1-\xi)$ and $m(1-\xi)+1$ (as shown in Lemma 2, Appendix) and the parameter ξ is determined by $\xi = 1 - Mo/m$. Then, in the case of two adjacent sample modes at $R = j$ and $R = j+1$ we assume $\tilde{\xi} = 1 - \frac{M_n}{m}$.
- In the case of small sample size and/or high uncertainty in the responses, we may observe two *non-adjacent modes* at $R = j$ and $R = k$, respectively, for $j < k+1$. Then, as a practical suggestion, we choose estimator (4) where M_n is equal to j or k according to the category whose neighbours include most of the observed frequencies.

The last case may be considered uncommon in real surveys with ordinal data where the sample size ranges from some hundreds to several thousands. The previous solution may be motivated by the following consideration: since the parameter ξ is related to the predominance of unfavourable responses, and this modifies the skewness of the distribution (see Sect. 2), then the most reliable information about ξ should be derived from the category surrounded by most of the responses.

A different situation should be considered when two high frequencies are located at distant modalities. If this happens, there is room for seriously contending the validity of a standard CUB model, since the specification (1) implies a single or two adjacent modes. Indeed, high isolated frequencies should be interpreted as atypical responses and should be modelled by means of an extension of the class of CUB models, as obtained by inserting a *shelter* effect, for instance (Iannario 2012). In these cases, we should not automatically prefer the estimator (4) for a preliminary evaluation of the parameter ξ.

Now, if we concentrate on estimator (4), we notice that its admissible range is $[\frac{0.5}{m},\ 1-\frac{0.5}{m}]$; thus, the unit interval is now shortened by the amount $1/(2m)$ at both ends. The estimator $\tilde{\xi}$ can generate only m distinct values for the parameter ξ, which is instead intrinsically continuous: for moderate m, this circumstance produces some shortcomings. For instance, if $m=5$ the estimator (4) is constrained to $[0.1,\ 0.9]$ whereas the shape of the observed distribution might suggest more extreme values for ξ.

A solution to this problem might be obtained by defining: $\xi^* = \frac{m-M_n+U}{m} = 1+\frac{U-M_n}{m}$, where a previously generated value of a continuous Uniform random variable U (on the unit interval) should be determined. Although ξ^* is continuous and covers the whole unit interval for ξ, the maximum admissible error for any sub-interval is now doubled to $\frac{1}{m}$, and a preliminary random mechanism increases the variance of this alternative proposal.

4 Heterogeneity measures as estimators of parameter π

For given m categories, diversity (heterogeneity) among units increases if probabilities (or relative frequencies, for observed data) among categories become as equal as possible (that is, tends to the Uniform probability mass). Given a probability distribution, the addition of a constant value to each cell increases *per se* the diversity of the mixture distribution; since this constant quantity is strictly dependent on π, it seems relevant to connect diversity measures and preliminary π estimators. Specifically, we will use the Gini index whose computation and inference are immediate (Pardo 2006).

For a well defined probability distribution p_r, $r=1,2,\ldots,m$, the Gini index of heterogeneity G and its normalized version $\mathcal{G} \in [0,1]$ are defined, respectively, by:

$$G = 1-\sum_{r=1}^{m} p_r^2\,; \quad \mathcal{G} = \frac{m}{m-1}\left(1-\sum_{r=1}^{m} p_r^2\right).$$

If we denote by $G_{\text{CUB}}, G_{SB}, G_{UN}$ the (non-normalized) Gini indices of CUB, shifted Binomial and Uniform random variables, respectively, we get (Lemma 3, Appendix)

$$G_{\text{CUB}} = \pi^2\, G_{SB} \,+\, (1-\pi^2)\, G_{UN}. \tag{5}$$

This relationship holds true also for the normalized indices $\mathscr{G}_{\text{CUB}}, \mathscr{G}_{SB}, \mathscr{G}_{UN}$.

Since $G_{UN} = 1 - 1/m$ and $\mathscr{G}_{UN} = 1$, we obtain:

$$\mathscr{G}_{\text{CUB}} = 1 - \pi^2\, (1 - \mathscr{G}_{SB})\,. \tag{6}$$

Identity (6) shows that, for a given shifted Binomial component, heterogeneity (as measured by the Gini index) is inversely related to π and increases with uncertainty (related to $1-\pi$). Thus, it seems correct to introduce an estimator of π based on this heterogeneity measure.

More specifically, G_{SB} can be parametrically estimated by $\hat{G}_{SB}$, if one has a preliminary estimate of ξ (as given by $\tilde{\xi}$ in (4), for instance). In addition, G_{CUB} can be consistently and non-parametrically estimated by means of the relative frequencies $\mathscr{F}_r,\ r = 1, 2, \ldots, m$:

$$\hat{G}_{\text{CUB}} = 1 - \sum_{r=1}^{m} \mathscr{F}_r^2.$$

In this way, by solving for π in decompositions (5) and (6), we obtain a preliminary estimator of π, that is:

$$\tilde{\pi} = \sqrt{\frac{\hat{G}_{\text{CUB}} - G_{UN}}{\hat{G}_{SB} - G_{UN}}} = \sqrt{\frac{\hat{\mathscr{G}}_{\text{CUB}} - 1}{\hat{\mathscr{G}}_{SB} - 1}},$$

where the second expression considers normalized Gini indices.

This new estimator of π may be explicitly written as:

$$\tilde{\pi} = \sqrt{\frac{\sum_{r=1}^{m} \mathscr{F}_r^2 - \frac{1}{m}}{\sum_{r=1}^{m} [b_r(\tilde{\xi})]^2 - \frac{1}{m}}}, \tag{7}$$

and it is strictly dependent on a preliminary estimate of ξ.

In this regard, we remember that an unbiased estimator of Gini index is:

$$\hat{G} = \frac{n}{n-1}\left(1 - \sum_{r=1}^{m} \mathscr{F}_r^2\right)$$

and, for small sample sizes, it may be appropriate to modify (7) as:

$$\tilde{\pi}_u = \sqrt{\frac{\frac{n}{n-1}\, \sum_{r=1}^{m} \mathscr{F}_r^2 - \frac{1}{m} - \frac{1}{n-1}}{\frac{n}{n-1}\, \sum_{r=1}^{m} [b_r(\tilde{\xi})]^2 - \frac{1}{m} - \frac{1}{n-1}}}. \tag{8}$$

Notice that the radicand of the estimator (7) is non-negative since, for any well defined mass probability distribution $(q_1, q_2, \ldots, q_m)'$, we always get:

$$0 \leq \sum_{r=1}^{m} \left(q_r - \frac{1}{m}\right)^2 = \sum_{r=1}^{m} q_r^2 - \frac{1}{m},$$

and the equality sign holds if and only if q_r is the discrete Uniform distribution. However, given sampling variability, the radicand might be greater than 1; in this case, we must assume: $\tilde{\pi} = 1$.

5 Sampling distributions

In Sects. 3 and 4 we obtained preliminary estimators of parameters of a CUB model by using exploratory measures dependent on the frequency distribution. Operationally, given $m > 3$, and a relative frequencies distribution $\{f_1, f_2, \ldots, f_m\}$ derived from an observed sample $(r_1, r_2, \ldots, r_n)'$, we get the sample mode M_n and then compute sequentially the joint parameter estimators:

$$\tilde{\xi} = 1 - \frac{M_n - 0.5}{m}; \quad \tilde{\pi} = \min\left(\sqrt{\frac{\sum_{r=1}^{m} \mathscr{F}_r^2 - \frac{1}{m}}{\sum_{r=1}^{m} [b_r(\tilde{\xi})]^2 - \frac{1}{m}}},\ 1\right). \tag{9}$$

We observe that $\tilde{\xi}$ is computed without any reference to the parameter π (although its sampling distribution is determined by this quantity) whereas the $\tilde{\pi}$ estimator is strictly dependent on the observed $\tilde{\xi}$. This feature is not uncommon in statistical inference: for instance, for random samples generated by $X \sim N(\mu, \sigma^2)$, the estimator of μ is the sample mean (whose distribution is dependent on σ^2) whereas the estimator of σ^2 is the sample variance which is dependent on the sample mean.

Given the difficulty of assessing the bias, consistency and distributional shape of estimators (9) in a formal way, we will perform a Monte Carlo experiment to investigate their main properties. In all cases, 2000 replications are generated for each selected CUB model. A wide range of sample sizes and different m have been checked and in any case consistent results have been obtained. Thus, we present the main characteristics of estimators (9) when $m = 9$ and for random samples of size $n = 200$.

In planning the simulation experiment we select a wide range of CUB models in order to explore several features of this distribution in terms of skewness, location of mode, weight of tails, and so on. To save space, we limit ourselves showing and discussing only a subset of these results by choosing some typical shapes.

In the following pages, we consider results for both the bias and variance of estimators (9) since their behaviour is not always concordant; specifically, some emphasis is given to the bias since it mostly contributes to reduce the efficiency of estimators.

5.1 Sampling distributions of both estimators

If we examine the sampling distributions of $\tilde{\xi}$ by means of kernel histograms, for several selected CUB models (evenly scattered through the unit square), we observe that the discrete nature of estimator (4) induces multiple modal values, with the most prominent mode close to the true ξ.

Instead, the sampling distribution of the $\tilde{\pi}$ estimator is almost symmetric and unimodal for several CUB models, and this achievement is the joint result of the sums and of the square root which are present in definition (7); in fact, they enforce the application of Central Limit Theorem and induce more symmetric behaviour, respectively. Instead, a bump when $\pi \to 1$ and a secondary mode for some models with small ξ has been observed. Indeed, a modal value at $\pi = 1$ is induced by the assumption $\tilde{\pi} = 1$ when the radicand of (7) is greater than 1 (an event whose probability increases as $\pi \to 1$).

Finally, we observe that the correlation among the estimators $\tilde{\pi}$ and $\tilde{\xi}$ has been found proportional to $(\frac{1}{2} - \xi)$, and thus to the skewness of the CUB distribution. As a consequence, our preliminary estimators (9) are uncorrelated for a symmetric CUB model (characterized by $\xi = 0.5$), a feature they share with ML estimators (see Theorem 1, Appendix).

5.2 Sampling behaviour of the estimator of ξ

Firstly, we mention that the distribution of the estimator $\tilde{\xi}$ is strictly dependent on that of the sample mode M_n and we are not aware of general solutions in case of a discrete support. In this regard, we notice that several results are available in the statistical literature on the estimation of a mode of continuous random variables and most of them mimic the problem generated by stochastic processes whose periodogram is not a consistent estimator of the spectral density function; thus, smoothing the histogram by kernel based method is the basic tool for deriving consistent estimators of both density and mode (Parzen 1962; Silverman 1993). On the contrary, only few papers concern the sampling distribution of mode when the parent random variable is discrete, and also in this case some smoothing procedures are performed (Titterington 1980; Wang and van Ryzin 1980). A recent approach (Nettleton 2009) concerns tests for the dominance of a cell probability in multinomial samples and it is related to our issue but it does not solve the problem of the modal value distribution.

Since sample mode is a functional of relative frequencies (which are consistent for probabilities), we are in a position to expect similar properties by using an estimator derived from sample frequencies. However, the bias is a serious problem since it reduces the speed of the MSE convergence to 0. More specifically, the distributions of the sample mode collapse on some limited range given the discreteness of the support; thus, although more peaked distributions around parent mode arise with increasing n, it is not possible to rely on standard asymptotic results (like Normality, for instance). Moreover, as estimator (4) is restricted on the interval $[\frac{0.5}{m},\ 1 - \frac{0.5}{m}]$, we should expect some atypical behaviour in this case.

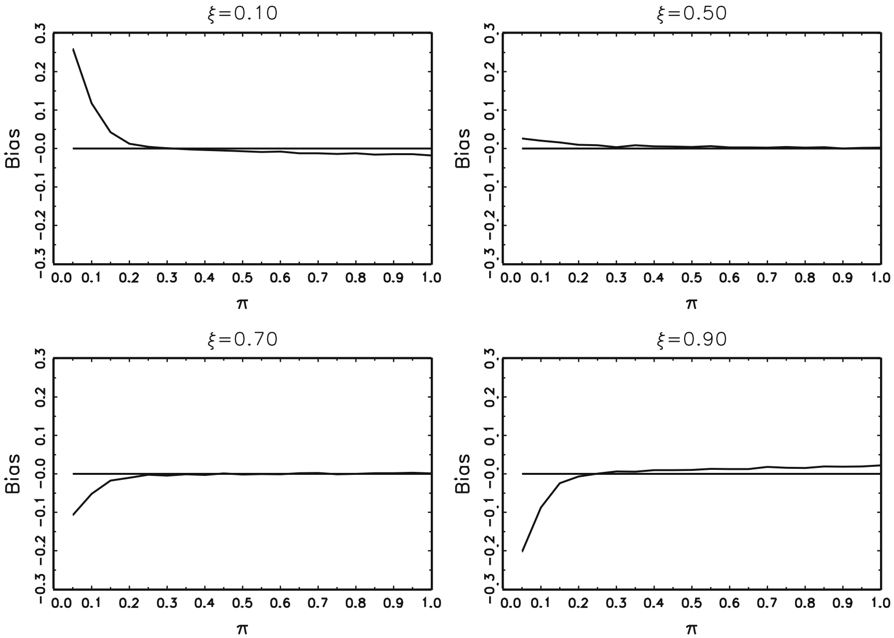

Fig. 1 Bias of $\tilde{\xi}$ estimator for given ξ and varying π $(m = 9)$

It turns out that the sampling distribution of the sample mode is determined by π although estimator (4) is invariant with respect to π. This is consistent with the following argument: if uncertainty is moderate $(\pi \to 1)$, the statistical information conveyed by relative frequencies $(\mathscr{F}_1, \mathscr{F}_2, \ldots, \mathscr{F}_n)$ is mainly related to the shifted Binomial distribution and thus it is characterized by the ξ parameter; moreover, the frequency heterogeneity is mostly due to the feeling component. On the contrary, when uncertainty is important $(\pi \to 0)$, the heterogeneity of the frequencies is mostly determined by the Uniform discrete distribution, and sample mode may change even for small variations in observed data; in these cases, convergence requires very large sample sizes.

As supported by Fig.1, the absolute bias of $\tilde{\xi}$ tends uniformly to 0 for $\pi \to 1$ (it is quite small for $\pi > 0.2$); instead, the bias is substantially large for $\pi \to 0$ and it is symmetric around $\xi = 1/2$. This feature is invariant with respect to the true ξ.

Figure 2 shows that the variance of $\tilde{\xi}$ drops to 0 for $\pi \to 1$ (mostly after $\pi > 0.2$) but it is extremely high for $\pi \to 0$: this is to be expected since, for small π, a CUB model approximates a Uniform discrete distribution which has no mode, and so sample mode becomes extremely erratic.

5.3 Sampling behaviour of the estimator of π

The sampling behaviour and distribution of $\tilde{\pi}$ are conditioned by $\tilde{\xi}$, since $\tilde{\pi} = \pi(\tilde{\xi})$. Fig. 3 shows that the bias of $\tilde{\pi}$ is moderate, symmetric with respect to ξ and extremely low when $\pi > 0.5$. Notice the presence of two hollows induced by the discrete nature of the conditioning $\tilde{\xi}$. In our experiment $m = 9$, and the hollows happen for

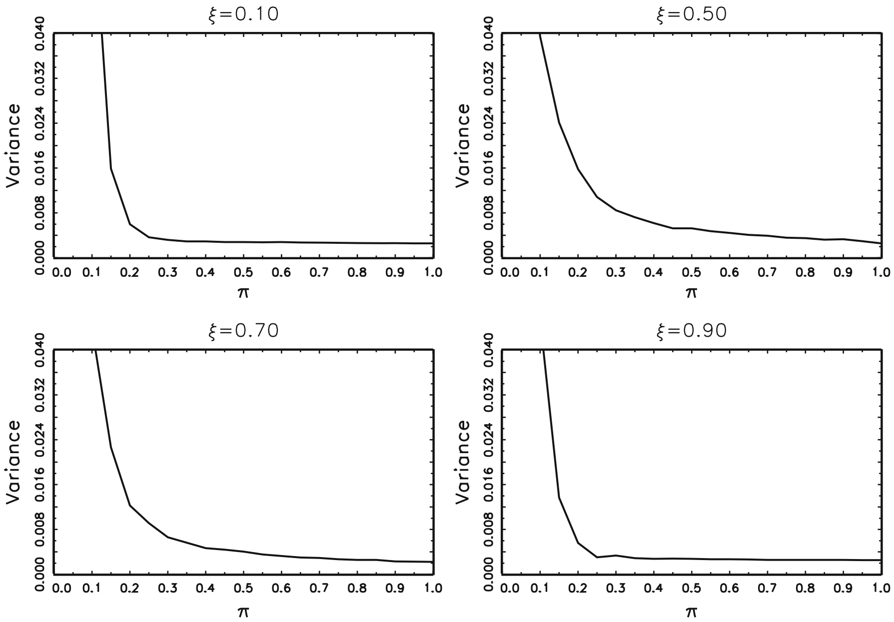

Fig. 2 Variance of $\tilde{\xi}$ estimator for given ξ and varying π ($m = 9$)

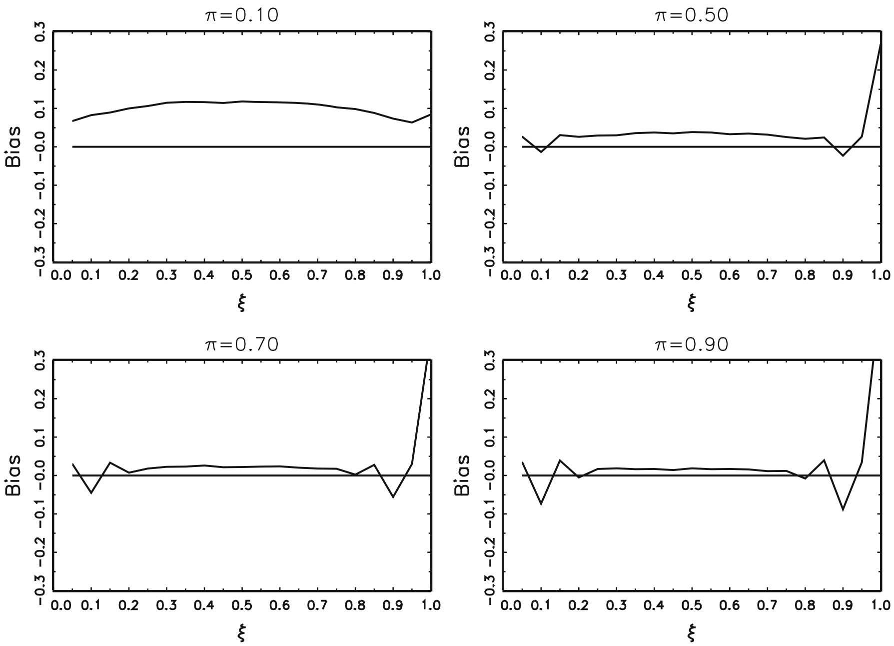

Fig. 3 Bias of $\tilde{\pi}$ estimator for given π and varying ξ ($m = 9$)

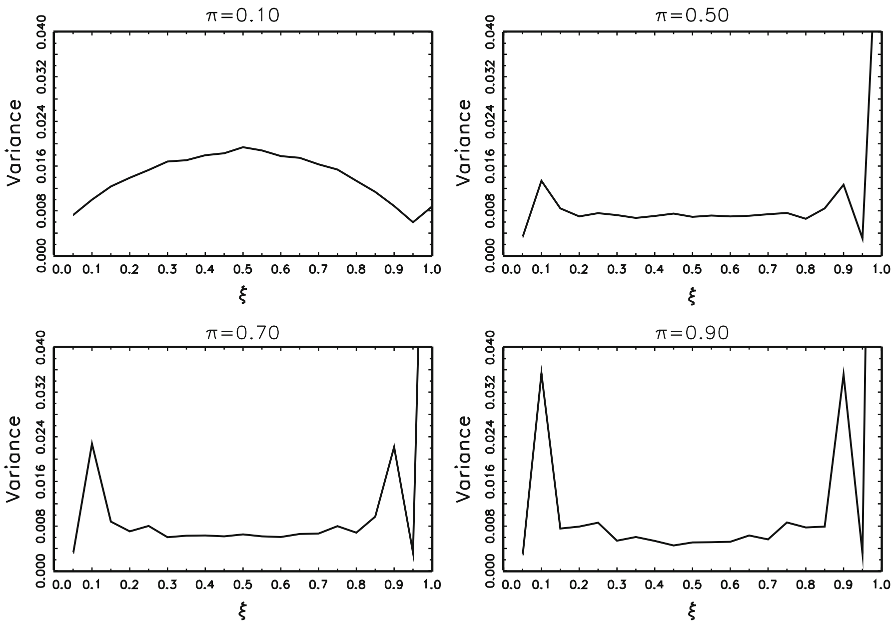

Fig. 4 Variance of $\tilde{\pi}$ estimator for given π and varying ξ $(m = 9)$

$\xi = 1/m = 0.11$ and $\xi = 1 - 1/m = 0.89$, respectively. Their size increases with π; in addition, the pile-up effect for $\xi = 1$ becomes more and more evident when $\pi > 0.5$.

The variance of the estimator $\tilde{\pi}$ is symmetric with respect to ξ and fairly constant for $\pi > 0.2$ (as confirmed by Fig. 4). Again, two hollows are present for $\xi = 1/m = 0.11$ and $\xi = 1-1/m = 0.89$, respectively, and their size as well a spike at $\xi = 1$ increase with π.

5.4 Discussion of an extreme case

In some CUB models where uncertainty is great, the main problem induced by estimators (9) is the bias caused by the sample mode since the variance of the estimators reduces with n but the bias lingers on a positive value.

As an example, the probability distribution of a CUB model with $\pi = 0.05$ and $\xi = 0.75$ is shown in Fig.5 (left panel). Then, even for sample sizes as large as $n = 1000,\ 2000, \ldots, 20000$ the bias of the mode estimator M_n is considerably high (right panel). Unfortunately, this difficulty is unavoidable since it derives from the very nature of the chosen estimators when faced with a distribution which is almost undistinguishable from a discrete Uniform random variable.

6 Computational efficiency

Estimators (9) may be useful to shorten the lengthy iterative EM procedure implemented for achieving the asymptotically efficient ML estimates. Thus, for finite sample

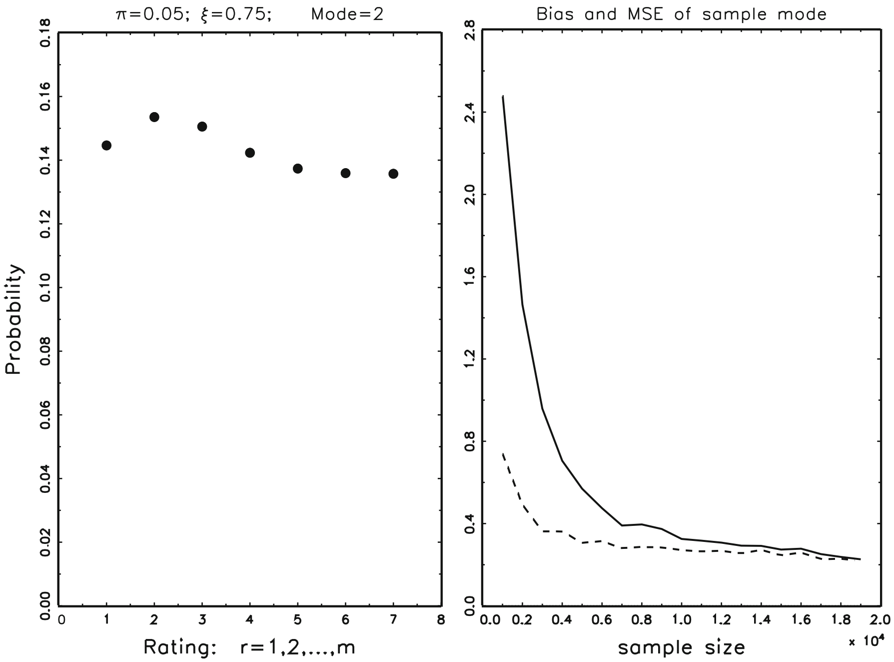

Fig. 5 An extreme CUB distribution ($m = 7$) and related bias (*dotted*) and MSE (*solid*) of the mode estimator M_n, obtained by simulation for varying sample sizes $n = 1000, 2000, \ldots, 20000$

Table 1 Alternative preliminary estimators for the simulation experiment

Estimators	Estimators of π	Estimators of ξ
Naive version	$\overline{\pi} = 0.5$	$\overline{\xi} = \frac{m - \overline{R}_n}{m-1}$
Random starting	$\pi^o = Unif(0, 1)$	$\xi^o = Unif(0, 1)$
New proposal	$\tilde{\pi} = \min\left(\sqrt{\frac{\sum_{r=1}^m \mathscr{F}_r^2 - \frac{1}{m}}{\sum_{r=1}^m [b_r(\tilde{\xi})]^2 - \frac{1}{m}}},\ 1\right)$	$\tilde{\xi} = 1 + \frac{0.5 - M_n}{m}$

sizes, we check if they really reduce time for convergence with respect to a naive proposal or some random starting.

Since time is proportional to the number of iterations necessary for convergence to ML estimates, we use these quantities for measuring the computational efficiency of different alternatives. In this way, results are independent from the numerical accuracy and speed of PC machines. Specifically, we will compare the preliminary estimators listed in Table 1 as starting values for the EM procedure.

Then, for a given $m > 3$, the simulation experiment has been conducted in the following way:

1. Specify a parameter vector $(\pi, \xi)'$ for a well defined CUB model.
2. For each given model, and for varying sample size $n = 100, \ldots, 1000$, generate 2000 random samples of ordinal values drawn by a CUB random variable (with parameters specified in step 1).

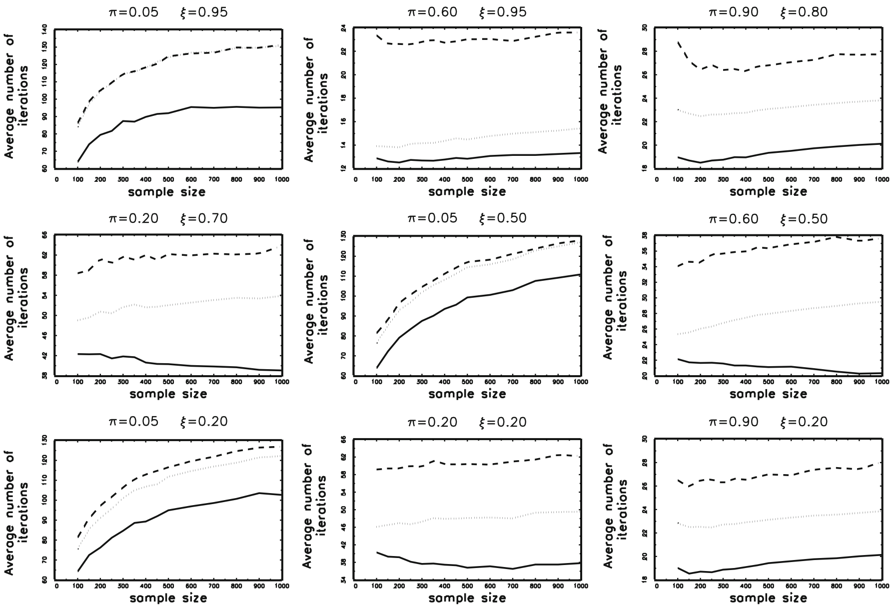

Fig. 6 Average number of iterations required for convergence: new (*solid*), random (*dashed*) and previous (*dotted*) preliminary estimators ($m = 9,\ \delta = 10^{-6}$)

3. For each random sample of step 2, set the three preliminary estimators of Table 1 as starting values of the EM procedure, and compute for each of them the average number of iterations required for the convergence to the ML solution.
4. Repeat steps 1–3 for different selected CUB models.
5. Plot summary results in a suitable format.

We checked the results by setting the tolerance required in the variation of log-likelihood function for reaching the convergence at $\delta = 10^{-3}$ and $\delta = 10^{-6}$, and obtained the same ranking of performance among the three estimators as shown in Fig. 6.

In the following we will comment on the most relevant results of an extensive experiment of simulation, based on the random selection of several CUB models evenly scattered throughout the parametric space. This choice is necessary since an *ad hoc* selection of models for the simulation experiment could favour the proposed estimators (given their discrete nature). We checked several values of $m = 7, 9, \ldots,$ but the general pattern of the results is substantially unmodified; thus, we will discuss the most significant findings when $m = 9$ (see Fig. 6).

It turns out that the proposed estimators (solid lines in the plots) outperform the others since they uniformly reduce the average number of required iterations for convergence. As expected, random starting is comparatively the worst method except in specific situations, thus it should not be chosen as a criterion for initial values. Finally, the naive version is generally intermediate but it generates a substantially longer time to achieve convergence with respect to the new proposal (9) for most of the parametric space.

Table 2 Ratings of a service

Items	Number of iterations		Relative gain	Euclidean distance	
	Naive	New	New vs. Naive	ϑ(Naive, ML)	ϑ(New, ML)
Information	37	33	0.108	0.302	0.083
Willingness	33	30	0.091	0.300	0.110
Opening Hours	47	39	0.170	0.296	0.042
Competence	37	33	0.108	0.302	0.099
Global	34	22	0.353	0.316	0.015

Comparisons of the number of iterations required for convergence

7 Some empirical evidence

As further confirmation of the results of the previous Section, we discuss some data set collected in real case studies.

Firstly, we present a case study where CUB models have been applied for interpreting the response behaviours of thousands of students attending the Orientation service of the University of Naples Federico II during the 2007 wave (Capecchi and Piccolo 2010; Iannario and Piccolo 2010). The purpose of the survey has been to measure the level of satisfaction towards several aspects of the service: *Information*, *Willingness*, *Opening Hours*, *Competence* and *Global* satisfaction. Students' opinions have been anonymously collected by means of a questionnaire with an ordinal (Likert) scale ranging from 1 (="completely unsatisfied") to 7 (="completely satisfied"). The results we will present refer to a sample of $n = 3,511$ validated responses.

We summarize the computational aspects of this research by reporting in Table 2 the number of iterations required to achieve convergence of the EM algorithm by using as different starting values the naive estimates and proposal (9). Then, we report the gain of the new estimates computed as the relative reduction of the number of iterations achieved by replacing naive with new estimators.

As already shown with the extensive simulations of Sect. 6, it is convenient to adopt (9) as starting points for the EM algorithm since they are closer to the final ML estimates. To measure this closeness, we have introduced the normalized Euclidean distance among the preliminary (naive or new) and the final (ML) estimates:

$$\vartheta(\hat{\boldsymbol{\theta}}_{pre},\ \hat{\boldsymbol{\theta}}_{ML}) = \sqrt{(\hat{\boldsymbol{\theta}}_{pre} - \hat{\boldsymbol{\theta}}_{ML})'(\hat{\boldsymbol{\theta}}_{pre} - \hat{\boldsymbol{\theta}}_{ML})/2}\,.$$

Table 2 confirms that such a measure is uniformly smaller for the proposal (9). Observe that for *Willingness* and *Global* items there is the maximum and minimum closeness among initial and final estimates, respectively, and we register the minimum and maximum relative gain in the number of iterations.

Unlike the first case study (characterized by a very low uncertainty among respondents), we analyze a further data set where more heterogeneity has been found in the sample. More precisely, we report a study on the preference ratings for living in two cities in Italy (Florence and Verona) expressed by means of an ordinal scale ($m = 12$)

Table 3 Preferences for living in a city

Items	Number of iterations		Relative gain	Euclidean distance	
	Naive	New	New vs. Naive	ϑ(Naive, ML)	ϑ(New, ML)
Living in Florence	20	13	0.350	0.238	0.016
Living in Verona	81	57	0.296	0.334	0.082

Comparisons of the number of iterations required for convergence

by a sample of $n = 214$ young people resident in Naples. Now, ratings 1 and m are the best and worst choices, respectively; thus, we should interpret ξ as a direct measure of preference towards the city. In this case, responses are characterized by an important uncertainty amount, estimated for Florence and Verona as $1 - \hat{\pi} = 0.166$ and $1 - \hat{\pi} = 0.834$, respectively.

Table 3 summarizes the main results and confirms a serious reduction in the number of iterations achieved by the new proposal. The reported distances show how estimates (9) accelerate the EM algorithm by starting the iterative procedure from a point far closer to the final ML estimates.

Both the case studies (and many others here unreported, for brevity) and the simulation experiments confirm the relevant role of uncertainty in determining the speed of convergence of the numerical routines for ML estimation. This consideration may be obvious from an interpretative point of view (since a tendency to maximum heterogeneity reduces information about the parameter ξ) but it strictly derives from the properties of the likelihood function of CUB models. In fact, the score function is proportional to $\left(1 - \frac{1}{m\, p_r(\boldsymbol{\theta})}\right)$ which is identically 0 if the probability distribution becomes a Uniform one, that is if $\pi = 0$. Thus, the algorithms become lengthy to converge when $\pi \to 0$ because of a borderline indeterminacy in the likelihood function.

8 Concluding remarks

Main usage of preliminary estimators in the framework of ordinal data analysis concerns both models specification and computational aspects.

First of all, the availability of quick and accurate estimators for subsets of sample characterized by relevant covariates (such as gender, job, education, socioeconomic status, and so on) allows inferring about their possible significance in a more elaborate CUB model: some experiments following these lines have been positively pursued by Iannario (2008), mainly with reference to feeling covariates. If this approach is repeatedly pursued for any available covariate, the method would lead to a selection criteria for choosing covariates (before submitting them to a fully specified model by ML estimate procedures).

Secondly, it is well known that the search for accurate preliminary estimators is a fundamental step in EM algorithm and speed and accuracy are relevant aspects if one needs to perform massive simulation experiments. In this direction, the results of this paper may orient towards the implementation of more efficient algorithms for ML estimation. Specifically, we would suggest a mixed strategy made up of the following steps:

- Start the search of the optimum by using the new estimators (9).
- Perform EM algorithm until the direction of the gradient does not change for some consecutive steps.
- Switch to scoring algorithm to accelerate the convergence towards ML estimates by a second-order rate.

This strategy might seem more expensive from a computational point of view, since it requires computations (and inversions) of the matrix of the second derivatives of the log-likelihood function in the final steps. However, the order of these matrices is generally low, the number of steps substantially decreases and the computation of the final variance-covariance matrix is necessary for assessing parameters significance and confidence regions.

Appendix

We will prove some formal results quoted in the paper.

Lemma 1 *A* CUB *random variable is symmetric if and only if* $\xi = \frac{1}{2}$.

The symmetry of a discrete distribution p_r defined over a support $\{r = 1, 2, \dots, m\}$ implies: $p_r = p_{m-r+1}$, $r = 1, 2, \dots, m$. These requirements, for a CUB distribution, become:

$$\pi\, b_r(\xi) + (1-\pi)\,\frac{1}{m} = \pi\, b_{m-r+1}(\xi) + (1-\pi)\,\frac{1}{m}\,, \quad r = 1, 2, \dots, m.$$

After a simple algebra, these equalities simplify to:

$$\xi^{m-r}\,(1-\xi)^{r-1} = \xi^{r-1}\,(1-\xi)^{m-r} \iff \left(\frac{\xi}{1-\xi}\right)^{m-1} = 1\,, \quad r = 1, 2, \dots, m.$$

whose unique solution is $\xi = \frac{1}{2}$.

To prove viceversa, we let $\xi = \frac{1}{2}$ in (1) and (2) and immediately obtain the symmetry of p_r, $r = 1, 2, \dots, m$.

This result may be also confirmed by a lengthy algebra, since the third central moment, computed by Piccolo (2003), is proportional to $(2\,\xi - 1)$. □

Lemma 2 *The modal value of a* CUB *model coincides with the modal value of its shifted Binomial component.*

If the modal value (mode) of a discrete distribution over the support $\{1, 2, \dots, m\}$ is unique, it is an integer $r \in \{1, 2, \dots, m\}$ such that:

$$(p_{r+1} - p_r)(p_r - p_{r-1}) < 0, \quad r = 2, 3, \dots, m-1.$$

For a CUB distribution, these inequalities solve into:

$$\left[b_{r+1}(\xi) - b_r(\xi)\right]\left[b_r(\xi) - b_{r-1}(\xi)\right] < 0, \quad r = 2, 3, \ldots, m-1,$$

which concern only the shifted Binomial distribution, and then lemma is proved.

In addition, when $\xi = 0$ ($\xi = 1$) the modal value of a CUB distribution is at $R = m$ ($R = 1$).

Finally, it is simply to show (by equating $p_r = p_{r+1}$, for instance) that in a CUB model two equal and adjacent modes at $R = r$ and $R = r + 1$ are possible if and only if $r = m(1 - \xi)$ is an integer. □

Lemma 3 *If we denote by $G_{\text{CUB}}, G_{SB}, G_{UN}$ the Gini and normalized Gini index of heterogeneity of* CUB*, shifted Binomial and discrete Uniform random variables, respectively, we obtain:*

$$G_{\text{CUB}} = \pi^2\, G_{SB} + (1 - \pi^2)\, G_{UN}$$

A dual relationship holds for the corresponding normalized Gini index of heterogeneity $\mathcal{G}_{\text{CUB}}, \mathcal{G}_{SB}, \mathcal{G}_{UN}$.

The first result is easily obtained since:

$$\begin{aligned}
\sum_{r=1}^{m} p_r^2 &= \sum_{r=1}^{m} \left[\pi\, b_r(\xi) + (1-\pi)\,\frac{1}{m}\right]^2 \\
&= \pi^2 \sum_{r=1}^{m} [b_r(\xi)]^2 + (1-\pi)^2 \sum_{r=1}^{m} \left(\frac{1}{m}\right)^2 + 2\,\pi\,(1-\pi)\,\frac{1}{m} \sum_{r=1}^{m} b_r(\xi) \\
&= \pi^2 \sum_{r=1}^{m} [b_r(\xi)]^2 + (1-\pi^2)\,\frac{1}{m}.
\end{aligned}$$

Then, from the definition of $G_{\text{CUB}},\ G_{SB},\ G_{UN}$, the final relationship holds. □

Lemma 4 *For any finite sequence $g_r,\ r = 1, 2, \ldots, m$, with $m \geq 1$, we have:*

$$g_r = g_{m-r+1},\ \ r = 1, 2, \ldots, m \Longrightarrow \sum_{r=1}^{m} \left(r - \frac{m+1}{2}\right) g_r = 0.$$

For m even, we get:

$$\begin{aligned}&\sum_{r=1}^{m/2}\left(r-\frac{m+1}{2}\right)g_r+\sum_{r=m/2+1}^{m}\left(r-\frac{m+1}{2}\right)g_r\\&=\sum_{r=1}^{m/2}\left(r-\frac{m+1}{2}\right)g_r+\sum_{r=1}^{m/2}\left(\frac{m+1}{2}-r\right)g_{m-r+1}\\&=\sum_{r=1}^{m/2}\left[\left(r-\frac{m+1}{2}\right)-\left(r-\frac{m+1}{2}\right)\right]g_r=0.\end{aligned}$$

For m odd, the central term of the sum is: $0\times g_{(m+1)/2}=0$, and previous result applies to the other terms.

This lemma is trivial if g_r is a mass probability distribution since the sum is the first central moment of a symmetric random variable. □

Theorem 1 *For any symmetric* CUB *model, ML estimators* $\hat{\boldsymbol{\theta}}$ *have asymptotically uncorrelated components.*

A classical result for grouped data, (Rao 1973, pp. 367–368) asserts that the variance and covariance matrix of ML estimators $\hat{\boldsymbol{\theta}}=(\hat{\pi},\hat{\xi})'$ is asymptotically given by:

$$\begin{aligned}\boldsymbol{V}(\boldsymbol{\theta})&=\frac{1}{n}\begin{bmatrix}\sum_{r=1}^{m}\frac{\left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial\pi}\right)^2}{p_r(\boldsymbol{\theta})} & \sum_{r=1}^{m}\frac{\left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial\pi}\right)\left(\frac{\partial p_r(\boldsymbol{\theta})}{d\xi}\right)}{p_r(\boldsymbol{\theta})}\\ \sum_{r=1}^{m}\frac{\left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial\pi}\right)\left(\frac{\partial p_r(\boldsymbol{\theta})}{\partial\xi}\right)}{p_r(\boldsymbol{\theta})} & \sum_{r=1}^{m}\frac{\left(\frac{d\,p_r(\boldsymbol{\theta})}{\partial\xi}\right)^2}{p_r(\boldsymbol{\theta})}\end{bmatrix}^{-1}\\&=\frac{1}{n}\begin{bmatrix}i_{11}&i_{12}\\i_{21}&i_{22}\end{bmatrix}^{-1}=\frac{1}{n}\begin{bmatrix}v_{11}&v_{12}\\v_{21}&v_{22}\end{bmatrix}.\end{aligned}$$

For CUB random variables, these quantities are:

$$\frac{\partial\,p_r(\boldsymbol{\theta})}{\partial\pi}=b_r(\xi)-\frac{1}{m};\quad\frac{\partial\,p_r(\boldsymbol{\theta})}{\partial\xi}=\pi\,b_r(\xi)\,\frac{A(\xi)-r}{\xi\,(1-\xi)};$$

where we put: $A(\xi)=m-\xi\,(m-1)$.

Then, ML estimators are uncorrelated if $v_{12}=v_{21}=0$ and this implies that $i_{12}=i_{21}=0$, that is they are *globally* (expected information) *orthogonal* (Lindsey 1996, pp. 236–238).

For a symmetric CUB model $\xi = 1/2$ (Lemma 1), and $p_r(\boldsymbol{\theta}) = \pi \left[\binom{m-1}{r-1} 2^{1-m} - \frac{1}{m}\right] + \frac{1}{m}$. Accordingly, previous derivatives simplifies to:

$$\frac{\partial\, p_r(\boldsymbol{\theta})}{\partial\, \pi}\Big|_{\xi=1/2} = \binom{m-1}{r-1} 2^{1-m} - \frac{1}{m};$$
$$\frac{\partial\, p_r(\boldsymbol{\theta})}{\partial\, \xi}\Big|_{\xi=1/2} = \pi \binom{m-1}{r-1} 2^{3-m} \left(\frac{m+1}{2} - r\right);$$

since: $A(1/2) = (m+1)/2$.

Then, for a symmetric CUB model, the quantities $i_{12} = i_{21}$ become:

$$\begin{aligned} i_{12} &= \sum_{r=1}^{m} \frac{\left(\frac{\partial\, p_r(\boldsymbol{\theta})}{\partial\, \pi}\right)\left(\frac{\partial\, p_r(\boldsymbol{\theta})}{d\, \xi}\right)}{p_r(\boldsymbol{\theta})} \\ &= -\pi\, 2^{3-m} \sum_{r=1}^{m} \left(r - \frac{m+1}{2}\right) \frac{\left[\binom{m-1}{r-1} 2^{1-m} - \frac{1}{m}\right]\binom{m-1}{r-1}}{\pi\left[\binom{m-1}{r-1} 2^{1-m} - \frac{1}{m}\right] + \frac{1}{m}} \\ &= \sum_{r=1}^{m} \left(r - \frac{m+1}{2}\right) g_r, \end{aligned}$$

where we let:

$$g_r = -\pi\, 2^{3-m} \frac{\left[\binom{m-1}{r-1} 2^{1-m} - \frac{1}{m}\right]\binom{m-1}{r-1}}{\pi\left[\binom{m-1}{r-1} 2^{1-m} - \frac{1}{m}\right] + \frac{1}{m}}, \quad r = 1, 2, \ldots, m.$$

Now, since: $\binom{m-1}{r-1} = \binom{m-1}{m-r}$, it is immediate to confirm that $g_r = g_{m-r+1}$, for $r = 1, 2, \ldots, m$.

Thus, according to Lemma 4, $i_{12} = 0$ and, as a consequence, the estimators are asymptotically uncorrelated. □

Acknowledgments This work has benefitted from a partial support of MIUR project PRIN2008: "Modelli per variabili latenti basati su dati ordinali" (CUP n.E61J10000020001), and of 2011 FARO project, University of Naples Federico II. The Author thanks the Associate Editor and referees for their constructive comments.

References

Agresti A (2010) Analysis of ordinal categorical data. Wiley, Hoboken

Biernacki C, Celeux G, Govaert G (2003) Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Comput Stat Data Anal 41:561–575

Breen R, Luijkx R (2010) Mixture models for ordinal data. Soc Methods Res 39:3–24

Capecchi S, Piccolo D (2010) Modelling approaches for ordinal data: the case of Orientation Service evaluation. Quaderni di Statistica 12:99–124

D'Elia A, Piccolo D (2005) A mixture model for preference data analysis. Comput Stat Data Anal 49: 917–934

Gourieroux C, Monfort A (1995) Statistics and econometrics models. Cambridge University Press, Cambridge
Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res B 37:681–689
Grün B, Leisch F (2008) Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. J Classif 25:225–247
Iannario M (2008) Selecting feeling covariates in rating surveys. Statistica Applicazioni VII:25–44
Iannario M (2010) On the identifiability of a mixture model for ordinal data. Metron LXVIII:87–94
Iannario M (2012) Modelling shelter choices in ordinal data surveys. Stat Methods Appl 21:1–22
Iannario M, Piccolo D (2009) A program in R for CUB models inference, version 2.0. http://www.dipstat.unina.it/CUBmodels1/
Iannario M, Piccolo D (2010) A new statistical model for the analysis of customer satisfaction. Quality Technol Quant Manag 7:149–168
Iannario M, Piccolo D (2012) CUB models: statistical methods and empirical evidence. In: Kenett RS, Salini S (eds) Modern analysis of customer surveys: with applications using R. Wiley, Chichester, pp 231–258
Karlis D, Xekalaki E (2003) Choosing initial values for the EM algorithm for finite mixtures. Comput Stat Data Anal 41:577–590
Lindsey JK (1996) Parametric statistical inference. Clarendon Press, Oxford
McCullagh P (1980) Regression models for ordinal data (with discussion). J R Stat Soc B 42:109–142
McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall, London
McLachlan G, Peel GJ (2000) Finite mixture models. Wiley, New York
Nettleton D (2009) Testing for the supremacy of a multinomial cell probability. J Am Stat Assoc 104:1052–1059
Pardo L (2006) Statistical inference based on divergence measures. Chapman & Hall/CRC, Boca Raton
Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065–1076
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 Preference 19:247–259
Rao CR (1973) Linear statistical inference and its applications, 2nd edn. Wiley, New York
Silverman BW (1993) Density estimation for statistics and data analysis. Chapman & Hall, London
Titterington DM (1980) A comparative study of kernel-based density estimates for categorical data. Technometrics 22:259–268
Wang M, Ryzin Jvan (1981) A class of smooth estimators for discrete distributions. Biometrika 68:301–309
Wedel M, DeSarbo WS (1995) A mixture likelihood approach for generalized linear models. J Classif 12:21–55