

# A comprehensive framework of regression models for ordinal data

Maria Iannario<sup>1</sup> · Domenico Piccolo<sup>1</sup>

Received: 10 December 2014 / Accepted: 15 June 2016 / Published online: 1 July 2016  
© Sapienza Università di Roma 2016

**Abstract** Literature on the models for ordinal variables grew very fast in the last decades and several proposals have been advanced when ordered data are expression of ratings, preferences, judgments, opinions, etc. A dichotomy has been emphasized between methods based on a latent variable which is behind the ordered selection and methods anchored to a probability distribution with a well defined pattern. In this paper, a comprehensive framework to regression models is proposed in case ordinal data come out from a discrete choice. The added value of this unifying perspective is the possibility to introduce further generalizations and also to deepen similarities and differences among the proposed models. A case study confirms the usefulness of this general framework. Some concluding remarks end the paper.

**Keywords** Ordinal data · Feeling · Uncertainty · Cumulative models · CUB models · Generalized mixture model with uncertainty

## 1 Introduction

Ordinal data are collected in many fields such as Sociology, Psychology, Medicine, Economics and Marketing [1]. In these contexts, quite often, they represent responses of interviewees as ratings, preferences, judgments, opinions, and so on. In this setting, the statistical analysis focusses on regression models able to interpret the subjects' selection process.

---

This work has been supported by FIRB2012 project at University of Perugia (code RBFR12SHVV) and the frame of Programme STAR (CUP E68C13000020003) at University of Naples Federico II, financially supported by UniNA and Compagnia di San Paolo.

---

✉ Domenico Piccolo  
domenico.piccolo@unina.it

Maria Iannario  
maria.iannario@unina.it

<sup>1</sup> Department of Political Sciences, University of Naples Federico II, Napoli, Italy

From a general point of view, ordinal data are usually analyzed by means of models based on the multinomial distribution, which do not imply that the probabilities of categories belong to a family of random variables. Different approaches rely on Item Response Theory [10, 63, 70], Underlying Response Variable models [6, 53–56] or, more recently, stochastic binary search algorithms [8]. These proposals have been introduced also to pursue clustering or classification objectives [71].

More specifically, the paper considers as paradigms for regression models of ordinal data those based on cumulative functions and discrete mixtures: proportional odds [49] and CUB models [58] are the simplest instances of the two approaches, respectively. Though derived from different principles, they are often compared in terms of fitting, parsimony, graphical tools and interpretation. In fact, it will be shown that such models are just two specifications of a more general paradigm for the analysis of ordinal data.

Then, a focal point of the discussion concerns the specification of “cutpoints” (or *thresholds*) which are real values able to transform a continuous phenomenon (whose existence is real or virtual) into a sequence of ordinal categories in a one-to-one correspondence with the first  $m$  integers, for some known  $m$ . In addition, we emphasize the need to parameterize the role of uncertainty in the responses as an inherent component of human decisions.

Several classifications are legitimate and we set apart models which are generated by a family of random variables characterized by few parameters from those which do not search explicitly for it. In the first instance a close relationship with the data generating process is required whereas in the second case more stringent fitting aptitudes are pursued.

The paper is organized as follows: in the next section, we examine the main components to be considered for the analysis of data generated by a selection of ordered categories. Then, a general framework and its main inferential issues are discussed in Sect. 3. Sections 4 and 5 consider different models according to discrete and latent variable paradigms, respectively. Section 6 investigates a real case study and the interpretative usefulness of the proposed approach. Some final remarks conclude the paper.

## 2 Data generating process for ordinal data

When data expresses opinions, evaluations, judgments, perceptions, etc. the selection of an ordinal category is the final outcome of a complex activity involving the human mind in a multifacet space filled by knowledge, memory, emotion, instinct, deduction, etc. where several factors related to subjects, items and circumstances affect the answer.

An important point of the debate is the distinction between people who *optimize* (those who do their best to give accurate responses) and people who *satisfice* (those who choose to do just enough to give a plausible answer that may not be the optimal one), as emphasized by [43, 45]. In any case, an authoritative model of question answering [66] distinguishes four steps in selecting a category for the response: (i) comprehension of the question; (ii) retrieval of information; (iii) derivation of a judgment; (iv) formulation of a response. This cognitive process includes both the *perceptual aspect* (the rater’s perception of the content of the item) and the *decisional aspect* (the rater’s use of the available ordinal scale). In all these steps, a subjective decision is mixed with a varying fuzziness.

To summarize, a simplified version of this mechanism assumes that at least two components drive the process leading to an ordinal choice: a substantive *personal attractiveness* towards the item and an *inherent indecision* to select a single category from a list of ordered options. Although these components should be better specified in each case study, hereafter, for simplicity, we qualify them as *feeling* and *uncertainty*.

## 2.1 Feeling

The first reaction of the mind towards an item/question may be positive, negative or indifferent: if a more accurate selection is requested then respondents should evaluate a finer subdivision. This human reaction will be called *feeling* but, according to different contexts, it may be better specified as attraction, preference, worry, concern, liking, agreement, level of perception, etc.

Feeling is the result of the personal history of the respondent with respect to the given item and, in this respect, genetic factors, emotional facts, family traditions, achieved education, environmental pressures, media influence, etc. are substantial and not independent facts. Thus, a probability distribution should be able to explain the continuous perception which derives from several agents (most of them are unconscious) and manifests itself as a discrete response.

A community of interviewees reacts to questions/items in different manner according to the factors which influence the final selection; however, people are positively, negatively or indifferently oriented in different proportions with some spread of preferences around the modal option. When positive, negative or indifferent proportions are prevalent, we should expect a left-skewed, right-skewed or almost symmetric distribution of the responses, respectively. That is, a homogeneous community should have a unique mode and the responses distribution may be captured by a well specified probability mass function. On the contrary, if subgroups are present (according to gender, nationality, education, marital status, working conditions, income distribution, etc.) different modal values may be expected.

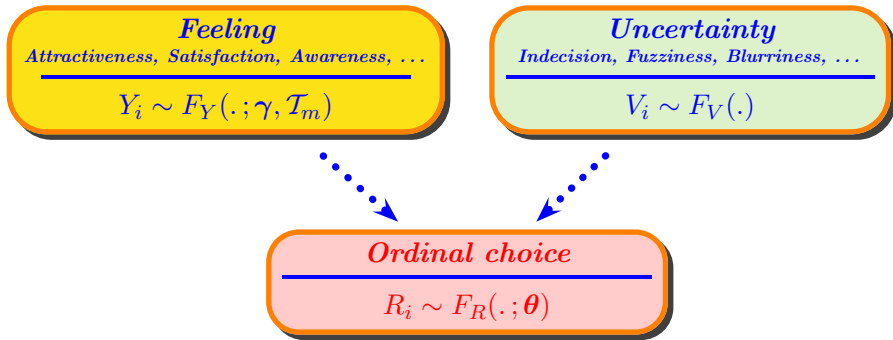
These arguments do not conflict with situations where some interviewees adhere to a specific category by laziness, ignorance, desire of privacy: this is a refuge option named *shelter effect* and represented as a degenerate distribution collapsed at the shelter category.

The psychological movement which generates an ordinal choice should be explained by a discrete random variable parsimoniously parameterized to include uni or multimodality, *shelter effect* and different skewness. If the output is conditional on manifest (concomitant) variables, the so-called subjects' covariates, one should insert these information in the modelling structure so as to improve interpretation and fitting of the observed data.

## 2.2 Uncertainty

Uncertainty is the result of convergent and related factors of different origins: limited set of information, knowledge/ignorance of properties and/or characteristics of the item to be judged, personal interest/engagement in activities related to the specific or similar items, amount of time devoted to the response, nature of the scale in terms of range and wording, tiredness or fatigue for a correct comprehension of the wording, willingness to joke and fake, lack of self-confidence of the respondent, apathy/boredom in the selection mechanism. Satisficing behaviour generates a varying degree of indecision [64] ranging from a complete lack of satisficing (totally accurate answer) to strong satisficing (totally random responses).

When a community of respondents manifests high uncertainty towards a given item a great heterogeneity in the distribution of responses is observed (the proportions of different categories tend to be similar), whereas if the uncertainty in the responses is extremely low most of the respondents select the same or surrounding categories (the distribution tends to be concentrated on a limited number of options). This component is not related to the variability of the responses (which is computed with the values  $r$ ,  $r = 1, \dots, m$  of the ordinal scores) but to the variability of the numeric values of the probabilities  $p_r$ ,  $r = 1, \dots, m$  (which



**Fig. 1** Stochastic mechanism of a discrete ordinal choice

involves the concept of *heterogeneity* of the distribution and derives from variability across individuals).

A further aspect is the *overdispersion*, that is the possibility that respondents manifest an inter-subject variability higher than those expected [36]. Overdispersion may be caused but not exhausted by uncertainty; thus, in some circumstances, a formal specification of the overdispersion may be effectively and parsimoniously achieved [37].

### 3 A unifying perspective

A framework which encompasses different regression models for ordinal data expresses the probability of a response as a finite mixture of (at least) two components which we have named as *feeling* and *uncertainty* (see Fig. 1).

Hereafter, with respect to a given item  $\mathcal{I}$  submitted to a sample of  $n$  respondents, we denote as  $R_i$  the final response expressed by the  $i$ -th subject and as  $Y_i$  and  $V_i$ ,  $i = 1, \dots, n$ , the corresponding random variables for feeling and uncertainty, respectively. All these random variables are defined over the discrete support  $\{1, \dots, m\}$ , for a known  $m$ . Let  $\theta = (\beta, \Psi)$  and  $\Psi = (\gamma, \mathcal{T}_m)$  the set of parameters characterizing the distribution of  $(R_1, \dots, R_n)$  and of the feeling component, respectively. Here,  $\mathcal{T}_m = (\tau_1, \dots, \tau_{m-1})$  denotes a vector of parameters, called thresholds and possibly involved in the distribution of  $Y_i$ , as specified in the following.

From a formal point of view, the feeling component  $Y_i$  assumes (categorical) values  $c_1, \dots, c_m$  and, for convenience, we score them by letting  $c_j = j$ , for  $j = 1, \dots, m$ . If the  $i$ -th subject is attracted by the  $j$ -th category as the preferred response to a given item we will denote this circumstance as:

$$Y_i = j, \quad \text{for } i = 1, \dots, n \text{ and } j = 1, \dots, m.$$

A *Generalized mixture model with uncertainty* (GEM) is defined as follows:

$$Pr(R_i = j | \theta) = \pi_i Pr(Y_i = j | t_i^{(\gamma)}, \Psi) + (1 - \pi_i) Pr(V_i = j), \quad (1)$$

for  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , where  $\pi_i = \pi(t_i^{(\beta)}, \beta) \in (0, 1]$  are introduced to weight the two components and  $t_i^{(\gamma)} \in T^{(\gamma)}$ ,  $t_i^{(\beta)} \in T^{(\beta)}$  include the selected covariates for the  $i$ -th subject. Available information on the subjects' characteristics (covariates) are collected in a

matrix  $T$ , and  $T^{(\beta)}$  and  $T^{(\gamma)}$  are submatrices of  $T$  containing the regressors for  $\pi_i$  and  $Y_i$ , respectively.

The probability distribution of the *uncertainty* component  $V_i$  is assumed known on the basis of *a priori* assumptions and it does not require parameters. Quite often,  $V_i$  is a discrete Uniform random variable over the support  $\{1, \dots, m\}$  although some different distributions without parameters may be usefully introduced to take account of different *response styles* (as discussed by [27], for instance). An increase of  $\pi_i$  implies a reduced impact of the uncertainty component; thus,  $(1 - \pi_i)$  is a (normalized) measure of the importance of the uncertainty in the model (1). The case  $\pi_i \equiv 0$  is logically possible if we conjecture that the selection process is equivalent to a totally random choice; however, this assumption rules out the identifiability of the parameters concerning the feeling. Thus, we prefer to set  $\pi_i > 0, \forall i$  and, to simplify notation, we drop in  $\pi_i$  the (possible) dependence on  $t_i^{(\beta)}$  and  $\beta$ .

Turning to the *feeling* component  $Y_i$ , ordinal phenomena may be interpreted as genuinely observed or derived by a continuous variable  $Y^*$  (a latent variable) which for convenience or necessity is examined in a discrete version by means of  $Y$ . This dichotomy, expressed by [4], and also discussed by [1] and [67], is a fundamental issue in ordinal data modelling since it may lead to different specifications. In the first case the correspondence with integers is immediate (although the scale is not metric) and the assumptions about a latent variable are not so important because the statistical procedures involve a direct consideration of  $Y$ . In the second case the discretization of  $Y$  is obtained by means of *cutpoints* (thresholds)  $\tau_j, j = 1, \dots, m - 1$  which transform the continuous support of  $Y^*$  into a sequence of ordered bins related to  $Y$ .

In general, the probability distribution of the *feeling component*  $Y_i$  is

$$\begin{cases} Pr(Y_i = j | \gamma, t_i^{(\gamma)}), & \text{if specified via a discrete distribution;} \\ F_{Y_i^*}(\tau_j; \gamma, t_i^{(\gamma)}) - F_{Y_i^*}(\tau_{j-1}; \gamma, t_i^{(\gamma)}), & \text{if specified via a latent variable distribution;} \end{cases}$$

where  $F_{Y_i^*}(\tau_j; \gamma, t_i^{(\gamma)}) = Pr(Y_i^* \leq \tau_j | \gamma, t_i^{(\gamma)})$  is the distribution function of the latent variable  $Y_i^*$ .

The interpretation of (1) is two-fold:

- according to the logic of latent class models [31,46], respondents split into two clusters in proportions  $(\pi_i)$  and  $(1 - \pi_i)$ , respectively, and one class consists of people whose selection is completely random;
- each respondent has *propensities*  $(\pi_i)$  and  $(1 - \pi_i)$  to select an ordinal category with a meditated and a random choice, respectively.

We strongly support the second interpretation. In fact, finite mixture models for ordinal data have been already discussed [12,28,29,72]; however, all these experiences concern combinations of probability distributions belonging to the same family and they implicitly assume that several subgroups justify such assumptions. As an instance, a proportion of Uniform distribution has been introduced to capture outliers (see: [5] for the continuous case).

A possible classification of regression models for ordinal data derives from the information available on the set of cutpoints:

$$T_m = \{-\infty = \tau_0 < \tau_1 < \dots < \tau_{m-1} < \tau_m = +\infty\}.$$

In this regard, we distinguish between *supervised* (classes I and II) and *unsupervised* discretization (class III) and, to simplify discussion and notation, we will assume fixed cutpoints.

- I There are no cutpoints to define the ordered sequence, thus  $\mathcal{T}_m \equiv \emptyset$ . In this case, the probability mass function of  $Y_i$  is directly specified.
- II Cutpoints are defined on *a priori* basis and are assumed to be known. A density function for  $Y_i^*$  is required and  $Y_i^*$  is transformed into a discrete version  $Y_i$  by means of some acceptable but arbitrary convention. A simple solution, valid for  $Y_i^*$  with a finite range  $(\tau_{min}, \tau_{max})$ , consists in selecting equi-distributed thresholds. It implies that, for each  $i = 1, \dots, n$ , the latent variable  $Y_i^*$  is discretized into  $Y_i$  by means of the transformation:

$$\tau_{min} + (j - 1) d < Y_i^* \leq \tau_{min} + j d \implies Y_i = j, \quad j = 1, \dots, m,$$

where  $d = (\tau_{max} - \tau_{min})/m$ . Alternative solutions are possible to take account of rounding effects, for instance.

- III Cutpoints are unknown and may be located everywhere on the real support of  $Y_i^*$ . The mapping from  $Y_i^*$  to  $Y_i$  is obtained by an *ad hoc* specification of cutpoints which are estimated by data. Notice that, since  $Y^*$  is a latent variable, both the scale and the origin cannot be identified ([2], pp.13–15). Then, for the continuous (latent) variable  $Y^* \in (\tau_{min}, \tau_{max})$ , where one or both  $\tau_{min}$  and  $\tau_{max}$  may be infinite, the correspondence between  $Y_i^*$  and  $Y_i$  becomes:

$$\tau_{j-1} < Y_i^* \leq \tau_j \implies Y_i = j, \quad \text{for } i = 1, \dots, n \text{ and } j = 1, \dots, m.$$

In all situations,  $Y_i$  is a discrete random variable characterized by a probability mass function  $Pr(Y_i = j \mid \boldsymbol{y}, \boldsymbol{t}_i^{(y)})$ . However, this discretization may be reached in a direct way (case I) or by means of the probability distribution function  $F_{Y_i^*}(\cdot)$  of a continuous latent variable  $Y_i^*$  and the knowledge (classes II) or the estimation (class III) of the cutpoints  $\tau_j \in \mathcal{T}_m$  which are necessary to transform  $Y_i^*$  into  $Y_i$ .

Discretization implies a different knowledge about the generating mechanism of the ordinal data. The uniform splitting (class II) reproduces the shape of the continuous variable  $Y^*$ ; as an instance, if a latent variable is strongly right-skewed we are assigning high probabilities to the first (low) categories. Instead, a general splitting (class III) may change the distribution of the latent variable  $Y^*$  by assigning the concentration of probabilities to selected categories thanks to a convenient subdivision of the original range. In a sense, the option to estimate cutpoints lets the observed data to become the criteria of subdivision of the range of  $Y^*$  (unsupervised choice). However, the main burden of estimable cutpoints is that the number of parameters increases by  $(m - 1)$ , and *ceteris paribus* this leads to a less parsimonious model.

The *deterministic* aspects of GEM model (1) concern the relationships of the weight of uncertainty (related to  $\pi_i$ ) and the intensity of feeling (denoted with  $\xi_i$ ) with respect to subjects' covariates. These links are expressed by:

$$\pi_i = g_\pi(\boldsymbol{t}_i^{(\beta)}, \boldsymbol{\beta}); \quad \xi_i = g_\xi(\boldsymbol{t}_i^{(y)}, \boldsymbol{\Psi}). \tag{2}$$

Then, according to the GLM jargon, the stochastic and systematic components of the response  $R_i$  will be specified by (1) and (2), respectively.

Here,  $\pi_i$  and  $\xi_i$  summarize the statistical content of the model (1) with respect to uncertainty and feeling, respectively. As it will be evident in Sects. 4, 5, the quantities  $\xi_i$  may be related to the expectation or distribution function of  $Y_i$  or directly to the parameters which characterize such a distribution.

Common specifications are the logistic, Gaussian or complementary log-log functions [1], for instance. Formally, we get:

$$\begin{cases} \text{logit}(\pi_i) = \mathbf{t}_i^{(\beta)} \boldsymbol{\beta}; \\ \text{logit}(\xi_i) = \mathbf{t}_i^{(\boldsymbol{\gamma})} \boldsymbol{\gamma}; \end{cases} \quad \text{or} \quad H^{-1}\left(\text{Pr}(Y_i \leq j \mid \mathcal{T}_m, \boldsymbol{\gamma})\right) = \tau_j - \mathbf{t}_i^{(\boldsymbol{\gamma})} \boldsymbol{\gamma}.$$

where  $\text{logit}(p) = \log(p/(1-p))$  and  $H(\cdot)$  is a monotone increasing function mapping  $\mathbb{R}$  into  $(0, 1)$ ; the  $\text{logit}(\cdot)$  function and the inverse of the Gaussian distribution function are just two instances of  $H^{-1}(\cdot)$ .

For inferential purposes, given a sample  $\mathbf{r} = (r_1, \dots, r_n)'$  of ratings and a matrix  $\mathbf{T}$  containing all subjects' covariates, the likelihood function is:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n \prod_{j=1}^m [\text{Pr}(R_i = j \mid \boldsymbol{\theta})]^{I(r_i=j)}$$

and the corresponding log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \log \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^m I(r_i = j) \log \left[ \pi_i \text{Pr}(Y_i = j \mid \mathbf{t}_i^{(\boldsymbol{\gamma})}, \boldsymbol{\Psi}) + (1 - \pi_i) p_j^V \right],$$

where  $I(r_i = j) = 1$  if the  $i$ -th respondent selects the  $j$  category and  $I(r_i = j) = 0$  otherwise. Hereafter, for simplicity, we let  $p_j^V = \text{Pr}(V_i = j)$ .

For this class of models, the EM procedure [25] is an effective algorithm to reach convergence almost everywhere for general mixture distributions [26,50]. Consistent starting values may be derived for most of the mixtures hereafter examined (as in [34], for instance) and convergence may be achieved when the increase in log-likelihood is less than a prefixed small value. Standard analyses based on deviance and likelihood ratio tests apply and *AIC*, *BIC* indexes are useful to compare non-nested models.

We mention the comparison of estimated probabilities  $\hat{p}_j$  with the corresponding relative frequencies  $f_j$  by means of a dissimilarity measure (*Diss*) or a fitting index ( $\mathcal{F}^2$ ) defined, respectively, by:

$$\text{Diss} = \frac{1}{2} \sum_{j=1}^m |f_j - \hat{p}_j|; \quad \mathcal{F}^2 = 1 - \text{Diss}.$$

Both of them are normalized in  $[0,1]$  and are especially useful in GEM models without covariates or when covariates have a discrete support.

In the next Sections, we will discuss some specifications of (1), (2) which lead to well known statistical models, so far presented as alternative approaches to the analysis of ordinal data.

### 4 Specification of models via discrete random variables

In case latent variables are not involved in the specification of the probability distribution, models for ordinal data are derived on the basis of an explicit data generating process without the knowledge of cutpoints (Class I). Considered as a whole, these probability distributions represent a sequential strategy aimed to fit the observed data with increasing complexity.

### 4.1 IHG models

The generating process of the marginal rank distribution of a given item (which can be considered as an ordinal evaluation of the item) has been studied by means of a discrete random variable [20,22,23]. More specifically, the inverse (or negative) hypergeometric distribution (IHG) has been introduced to mimic the waiting time of a white ball in the sequential drawing from an urn which contains black and white balls in some definite proportion [30,51,73].

No uncertainty is present (then,  $\pi_i \equiv 1$ ), and the *stochastic* and *systematic components* of the IHG model are:

$$\begin{cases} Pr(Y_i = j | \boldsymbol{\gamma}) = \xi_i(1 - \xi_i)^{j-1} \frac{m-1}{m} \prod_{s=1}^j \frac{m-s+1}{m-s+\xi_i(s-1)} ; \\ logit(\xi_i) = \boldsymbol{t}_i^{(\boldsymbol{\gamma})} \boldsymbol{\gamma} ; \end{cases}$$

for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ . The expectation of this random variable is  $\mathbb{E}(Y_i) = \frac{m - \xi_i}{1 + \xi_i(m-2)}$ . Then, a different parameterization which involves the mean value is possible.

This model is especially useful in case subjects judge the items as extreme (the worst or the best in a list) since the modal value of the IHG distribution may be 1 or  $m$ . The skewness is positive (negative) according to  $\xi_i < 1/m$  ( $\xi_i > 1/m$ ); if  $\xi_i = 1/m$  the IHG random variable becomes the discrete uniform distribution. A positive (negative) feeling towards the item is determined by  $\xi_i < 1/m$  ( $\xi_i > 1/m$ ); as a consequence, the interpretation of the parameter depends on the maximum of the support. This characteristic causes some difficulties in comparing IHG models when the number of categories is different.

### 4.2 Shifted binomial models

To take account of an intermediate modal value, a shifted binomial (SB) random variable has been introduced as a *proxy* of the generating process of paired comparison of categories [21]. If  $m = 2$ , the so-called Binomial regression is obtained.

Again, the proposal does not include uncertainty (then,  $\pi_i \equiv 1$ ) and the *stochastic* and *systematic components* of (1) are:

$$\begin{cases} Pr(Y_i = j | \boldsymbol{\gamma}) = \binom{m-1}{j-1} \xi_i^{m-j} (1 - \xi_i)^{j-1} ; & j = 1, \dots, m; \quad i = 1, \dots, n. \\ logit(\xi_i) = \boldsymbol{t}_i^{(\boldsymbol{\gamma})} \boldsymbol{\gamma} ; \end{cases}$$

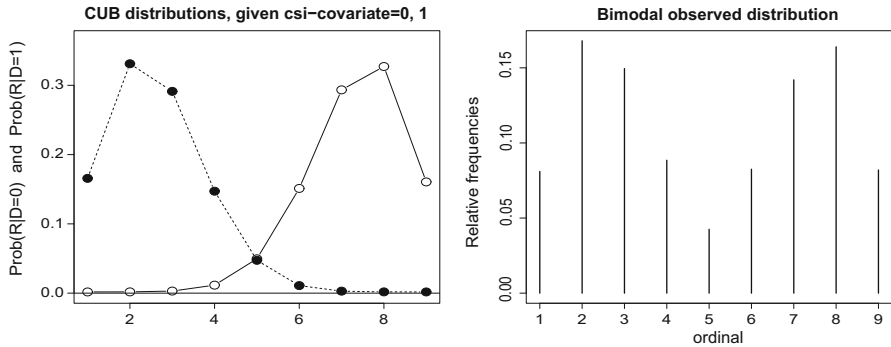
Since  $\mathbb{E}(Y_i) = 1 + (m-1)(1 - \xi_i) = m - (m-1)\xi_i$ , a direct parameterization by using the mean value is possible.

The SB random variable is a unimodal distribution but this model may well be exploited for fitting/interpreting multimodal observations if explanatory covariates (as a dichotomous one, for instance) are able to characterize the clusters. As an instance, for  $m = 9$ , Fig. 2 shows the implied distributions (conditional to  $D_i = 0, 1$ , respectively) of the following SB model:

$$\begin{cases} Pr(R_i = j) = \binom{8}{j-1} \xi_i^{8-j} (1 - \xi_i)^{j-1} ; & j = 1, 2, \dots, 9; \quad i = 1, 2, \dots, n. \\ logit(\xi_i) = -1.362 + 2.744 D_i ; \end{cases}$$

and a simulated frequency distribution generated by this model.





**Fig. 2** A bi-modal distribution well fitted by a shifted binomial model with a dichotomous covariate: conditional (left panel) and simulated (right panel) distributions

A relationship of the shifted binomial distribution with item response theory (IRT) has been advanced by [3] in the context of Likert-type personality measures to support the idea of a genuinely discrete generating process of the responses to a given item.

### 4.3 CUB models

The mixture defined CUB model is a (convex) Combination of a discrete Uniform and a (shifted) Binomial distribution. It has been introduced by [58] on the basis of psychological arguments and further discussed by [24,38,41,59], among others. A package in R [42] for building CUB models and several generalizations [40] is available.

This class of models is very flexible and parsimonious since observed data with different location, variability and skewness are adequately fitted with just two parameters. In addition, given the one-to-one correspondence between the parameters  $(\pi_i, \xi_i)$  and a probability distribution conditional on the observed covariates for the  $i$ -th subject, a CUB model allows effective interpretations when parameter estimates are examined as points in the parameter space (i.e., unit square). As a matter of fact, this visualization conveys immediate information in terms of uncertainty and feeling, respectively.

Formally, a CUB model is specified by the *stochastic* and *systematic components* given, respectively, by:

$$\begin{cases} Pr(R_i = j | \boldsymbol{\beta}, \boldsymbol{\gamma}) = \pi_i b_j(\xi_i) + (1 - \pi_i) p_j^V; \\ \text{logit}(\pi_i) = \boldsymbol{t}_i^{(\boldsymbol{\beta})} \boldsymbol{\beta}; & \text{logit}(\xi_i) = \boldsymbol{t}_i^{(\boldsymbol{\gamma})} \boldsymbol{\gamma}; \end{cases} \quad j = 1, \dots, m; \quad i = 1, \dots, n. \quad (3)$$

We have set  $b_j(\xi_i) = \binom{m-1}{j-1} \xi_i^{m-j} (1 - \xi_i)^{j-1}$  and  $p_j^V = 1/m, j = 1, \dots, m$ , for the probability mass functions of the shifted binomial and discrete Uniform random variable, respectively. Here,  $(\boldsymbol{t}_i^{(\boldsymbol{\beta})}, \boldsymbol{t}_i^{(\boldsymbol{\gamma})})$  are the information set for the  $i$ -th subject extracted from  $\boldsymbol{T}$  to specify the relationships of  $\pi_i$  and  $\xi_i$  with the corresponding subjects' covariates. Of course, the deterministic part of model (3) is not limited to the logistic link since any monotone mapping  $\mathbb{R}^p \leftrightarrow (0, 1)$  between linear combinations of covariates and parameters is legitimate.

CUB models have been also applied to a collection of items related to the same questionnaire to describe, classify or discriminate the behaviour of respondents with respect to the different items: these objectives may be pursued without any reference to subjects' covariates for clustering objectives [14,16] or for missing data imputation [19].

Furthermore, multilevel extensions [35] and multivariate generalizations [13, 15] have been advanced since the approach has been interpreted as a new paradigm for modelling ordinal data.

An important variant has been advanced by [48] who proposed non-linear CUB models specified by a non-constant transition probability from one category to the next. In this perspective, CUB models are the simplest solution for they assume a constant transition probability derived by some approximation.

### 4.4 CUBE models

To take account of a possible overdispersion in ordinal data, [36] proposed a mixture model, denoted as CUBE (Combination of Uniform and BEta-Binomial random variable), which has been further extended with the introduction of subjects' covariates [60]. The *stochastic* and the *systematic components* of this model are defined by:

$$\begin{cases} Pr(R_i = j | \theta) = \pi_i h_j(\xi_i, \phi_i) + (1 - \pi_i) p_j^V; \\ \text{logit}(\pi_i) = t_i^{(\beta)} \beta; \text{logit}(\xi_i) = t_i^{(\gamma)} \gamma; \log(\phi_i) = t_i^{(\alpha)} \alpha; \end{cases} \tag{4}$$

for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ . Generally,  $p_j^V = 1/m, j = 1, \dots, m$ . We let  $\theta = (\beta, \gamma, \alpha)$  whereas  $h_j(\xi_i, \phi_i)$  is the (shifted) Beta-Binomial distribution of the *feeling* of the  $i$ -th subject which has been parameterized as follows:

$$h_j(\xi_i, \phi_i) = \binom{m-1}{j-1} \frac{\prod_{k=1}^j [1 - \xi_i + \phi_i(k-1)] \prod_{k=1}^{m-j+1} [\xi_i + \phi_i(k-1)]}{[1 - \xi_i + \phi_i(j-1)] [\xi_i + \phi_i(m-j)] \prod_{k=1}^{m-1} [1 + \phi_i(k-1)]}, \tag{5}$$

for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ . If  $\phi_i \equiv 0, \forall i$  we get a CUB model; thus, CUB are nested into CUBE models and the selection between them is made effective by means of likelihood ratio tests.

Simulations and experimental evidence show that an explicit parameterization of a possible overdispersion allows a more correct analysis of the uncertainty in the estimated model [37].

### 4.5 Models with varying uncertainty

Hitherto, the discrete Uniform random variable has been considered as the building block to account for a subjective propensity to indecision: this solution is motivated by simplicity and parsimony.

In order to fit real situations generated by different *response styles* of interviewees, a greater flexibility has been suggested for  $p_j^V$  [27]; it can be achieved by selecting one of alternative discrete distributions over the support  $\{1, \dots, m\}$ , without the addition of further parameters. Formal tests have been introduced in this respect.

CUB models with a varying uncertainty (VCUB models) belong to the GEM models family (1) with a convenient specification of  $p_j^V$  and vice versa, all GEM models may be safely considered with a varying uncertainty in  $p_j^V$ .

### 4.6 The family of models with a *shelter* component

In psychological, sociological, political and marketing surveys an important proportion of respondents may choose a category which represents a sort of “refuge” to avoid a more demanding selection: this circumstance has been named a *shelter effect* and the corresponding option a *shelter category* [16]; then, a CUB model with a *shelter effect* as a sort of *c*-inflated distribution has been considered [32].

#### 4.6.1 GeCUB models

A CUB model with a *shelter effect* which includes subjects’ covariates has been referred to as a GeCUB model [39,41]. The *stochastic* and the *systematic components* are specified by:

$$\begin{cases} Pr(R_i = j | \theta) = \delta_i \left[ D_j^{(c)} \right] + (1 - \delta_i) \left[ \pi_i b_j(\xi_i) + (1 - \pi_i) p_j^V \right]; \\ \text{logit}(\pi_i) = \mathbf{t}_i^{(\beta)} \boldsymbol{\beta}; \quad \text{logit}(\xi_i) = \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}; \quad \text{logit}(\delta_i) = \mathbf{t}_i^{(\omega)} \boldsymbol{\omega}; \quad i = 1, \dots, n; \end{cases} \tag{6}$$

where  $\theta = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\omega})$  and the dummy (degenerate) random variable is defined by:  $D_j^{(c)} = I(j = c)$ , with  $I(E)$  the indicator function of  $E$ . The integer  $c \in [1, m]$  is assumed known on the basis of the nature of the selected option. Again, the logistic link may be substituted by a continuous monotone mapping from  $\mathbb{R}$  to  $(0, 1)$ .

The *stochastic component* of (6) may be written as

$$Pr(R_i = j | \theta) = \pi_i^* Pr(Y_i = j | \boldsymbol{\gamma}, \boldsymbol{\omega}) + (1 - \pi_i^*) p_j^V$$

where

$$\begin{cases} \pi_i^* & = \pi_i + \delta_i (1 - \pi_i); \\ Pr(Y_i = j | \boldsymbol{\gamma}, \boldsymbol{\omega}) & = \left( \frac{\delta_i}{\delta_i + \pi_i (1 - \delta_i)} \right) D_j^{(c)} + \left( \frac{\pi_i (1 - \delta_i)}{\delta_i + \pi_i (1 - \delta_i)} \right) b_j(\xi_i); \end{cases}$$

then, GeCUB models fit the GEM model family defined by (1), (2).

#### 4.6.2 “Don’t know” models

The family of CUB models has been exploited to cope with non-responses motivated by a “Don’t know” option [47]. The empirical results are really convincing and the new proposal captures respondents’ opinions with more fidelity than the simplistic solution to exclude “Don’t know” responses.

The specification of this model is:

$$Pr(R_i = j) = Pr(A = 0) Pr(R_i = j | A = 0) + Pr(A = 1) p_j^V,$$

where  $A$  is a dichotomous random variable indicating whether the respondent is able or not to formulate the requested rating  $R_i$ ; it assumes values 0, 1 with probabilities  $p_0$  and  $1 - p_0$ , respectively. Then, Authors argued that  $Pr(R_i = j | A = 0)$  may be the probability mass function of a CUB model.

These assumptions lead to a class of models included in (1):

$$Pr(R_i = j | \theta) = \pi_i^{**} Pr(Y_i = j | \boldsymbol{\gamma}, \boldsymbol{\omega}) + (1 - \pi_i^{**}) p_j^V, \quad i = 1, \dots, n; \quad j = 1, \dots, m,$$

where

$$Pr(Y_i = j | \boldsymbol{\gamma}, \boldsymbol{\omega}) = b_j(\xi_i); \quad p_j^V = 1/m, \forall j; \quad \pi_i^{**} = \pi_i p_0;$$

and  $\pi_i, \xi_i$ , are defined as in (3).

Although this model may be formally included in the family of CUB models with a *shelter effect* (“Don’t know” response is a sort of refuge caused by difficulties to express an opinion), a substantial difference should be noted: the *shelter category*  $c$  belongs to the support  $\{1, \dots, m\}$  whereas the “Don’t know” option is an added choice for the respondent. This characteristic justifies a two-step procedure for the estimation of  $p_0$  and the model parameters.

A possible extension includes subject’s covariates to characterize clusters of respondents who select the “Don’t know” option. In addition, to investigate the possible causes of the missing selection, we might consider objects’ covariates in a multi-item perspective, as pursued by [61], for instance. Thus, it would be possible to explain the reasons because some items have been considered by respondents more awkward than others.

### 5 Specification of models via latent variables

Some specifications of (1), which are based on the relationship with a latent variable and whose cutpoints are necessary to set the ordered categories, are now discussed. A difference between supervised (models of Class II: Sects. 5.1 and 5.2) and unsupervised discretization (models of Class III: Sects. 5.3 and 5.4) is considered. For these models,  $\xi_i = \text{logit}(\mu_i)$  and  $\xi_i = \text{logit}(F_{Y_i}(j))$ , respectively.

#### 5.1 CUN model

In a note relating CUB models and IRT, a direct discretization of the Normal distribution has been proposed [62] by specifying model (1) with

$$Pr(Y_i = j \mid \mu, \sigma) = \Phi\left(\frac{j + 1/2 - \mu}{\sigma}\right) - \Phi\left(\frac{j - 1/2 - \mu}{\sigma}\right),$$

where  $\Phi(\cdot)$  is the distribution function of a standard Gaussian random variable and  $(\mu, \sigma)$  are quantities estimable by data. Here  $\pi_i \equiv 1$  and  $\mathcal{T}_m = (\frac{1}{2}, 1 + \frac{1}{2}, \dots, m - \frac{1}{2})$ .

The introduction of the parameter  $\sigma$  for the variability improves the fitting of observed data; however, the uniform splitting and the symmetry of the Gaussian distribution around  $j = [\mu \pm \frac{1}{2}]$  are severe constraints to cope with observed data.

Simple generalizations include: (i) an *ad hoc* splitting of cutpoints by means of thresholds  $\mathcal{T}_m$  to be estimated by data as in probit/logit models; (ii) a link between parameters and subjects’ covariates by means of:

$$\text{logit}(\mu_i) = \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}; \quad \log(\sigma_i) = \mathbf{t}_i^{(\lambda)} \boldsymbol{\lambda};$$

where  $(\mathbf{t}_i^{(\gamma)}, \mathbf{t}_i^{(\lambda)})$  are the information set derived from  $\mathbf{T}^{(\gamma)}$  and  $\mathbf{T}^{(\lambda)}$ , the submatrices of  $\mathbf{T}$  containing regressors for the mean level and variability, respectively.

#### 5.2 Discretization of a Beta random variable

The noticeable flexibility of the shape of the Beta random variable over a finite continuous support has driven different Authors to address this distribution to get ordered values (updates references in [65]). Indeed, this random variable achieves “separation effects” in order to eliminate dependence of the variance from the expectation in robust experimental designs [11]. In a recent Ph.D. thesis [69], the split of the (0, 1) range of a Beta distribution into

$m$  equi-distributed cutpoints to obtain a probability mass function with a flexible shape has been proposed. The model does not consider uncertainty; thus,  $\pi_i \equiv 1, \forall i$  (see also [52]).

The parameterization for  $Y^*$  is:

$$h_{Y^*}(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\phi))} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad y \in (0, 1)$$

with  $\mu > 0$  and  $\phi > 0$ , as in Beta regression analysis [18]. As a consequence, the first two central moments are:

$$\mathbb{E}(Y^*) = \mu; \quad Var(Y^*) = \frac{\mu(1-\mu)}{1+\phi}.$$

Thus,  $\mu$  and  $\phi$  may be interpreted as *location* and *precision* parameter, respectively.

Similarly to [52],  $\mathcal{T}_m = \{0, 1/m, 2/m, \dots, 1\}$  has been selected as a supervised splitting. In presence of covariates, let us define:

$$\begin{aligned} Pr(\tau_{j-1} < Y_i^* < \tau_j) &= \int_{\tau_{j-1}}^{\tau_j} h_{Y^*}(y) dy \\ &= IB_{\mu_i\phi_i, (1-\mu_i)\phi_i}(\tau_j) - IB_{\mu_i\phi_i, (1-\mu_i)\phi_i}(\tau_{j-1}), \end{aligned}$$

for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ , where  $\tau_j = j/m$  for  $j = 0, 1, \dots, m$  and  $IB_{\mu\phi, (1-\mu)\phi}(\cdot)$  is the incomplete Beta function with parameters  $(\mu_i\phi_i, (1-\mu_i)\phi_i)$ .

Then, the *stochastic* and *systematic components* of this Discrete-Beta model become, for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ :

$$\begin{cases} Pr(Y_i = j | \boldsymbol{\gamma}, \boldsymbol{\lambda}) = IB_{\mu_i\phi_i, (1-\mu_i)\phi_i}(\tau_j) - IB_{\mu_i\phi_i, (1-\mu_i)\phi_i}(\tau_{j-1}); \\ \text{logit}(\mu_i) = \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}; \quad \log(\phi_i) = \mathbf{t}_i^{(\lambda)} \boldsymbol{\lambda}. \end{cases}$$

The proposed distribution may be underdispersed with respect to the (shifted) Binomial one whereas CUB models tend to have tails heavier than this discretized Beta random variable.

### 5.3 Cumulative models

In this class of models, the response  $R_i$  is interpreted as coincident with the expressed feeling  $Y_i$ . A distinctive feature of the approach is the presence of (unsupervised) cutpoints in  $\mathcal{T}_m$ . Since  $\pi_i \equiv 1$ , a possible uncertainty/heterogeneity of data is spread around all categories and captured by the estimated cutpoints.

Assume that  $p \geq 1$  covariates are relevant for explaining the latent regression model by means of

$$Y_i^* = \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma} + \epsilon_i, \quad i = 1, \dots, n,$$

where  $\epsilon_i \sim F_{Y_i^*}(\tau_j; \boldsymbol{\gamma})$ . Common choices for  $F_{Y^*}(\cdot)$  are Gaussian, logistic or Complementary log-log distributions: see [1,67], for details. Then,  $Pr(Y_i = j | \mathcal{T}_m, \boldsymbol{\gamma}) = Pr(\tau_{j-1} < Y_i^* \leq \tau_j)$ .

If a logistic random variable for  $\epsilon_i$  is considered, the *stochastic* and *systematic components* of the model are:

$$\begin{cases} Pr(Y_i = j | \mathcal{T}_m, \boldsymbol{\gamma}) = F_{Y_i^*}(\tau_j - \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}) - F_{Y_i^*}(\tau_{j-1} - \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}); \\ \text{logit}[Pr(Y_i \leq j | \mathcal{T}_m), \boldsymbol{\gamma}] = \tau_j - \mathbf{t}_i^{(\gamma)} \boldsymbol{\gamma}. \end{cases} \tag{7}$$

Of course,

$$Pr(Y_i \leq j \mid \mathcal{T}_m, \boldsymbol{\gamma}) = \left[ 1 + \exp(-[\tau_j - \mathbf{t}_i^{(y)} \boldsymbol{\gamma}]) \right]^{-1}, \quad i = 1, \dots, n; \quad j = 1, \dots, m.$$

The parameter vector  $(\mathcal{T}_m, \boldsymbol{\gamma})$  consists of intercepts  $\mathcal{T}_m = (\tau_1, \dots, \tau_{m-1})$  and covariate coefficients  $\boldsymbol{\gamma}$ . If there are no covariates the cumulative model is a saturated one and, given a sample of ordinal data, empirical and estimated distribution functions coincide.

Intercepts  $\tau_j$  characterize each category but model (7) assumes a constant effect  $\mathbf{t}_i^{(y)} \boldsymbol{\gamma}$  for all categories. This restraint has been removed by considering a varying effect  $\mathbf{t}_i^{(y)} \boldsymbol{\gamma}_j$  as in *stereotype models* [4], for instance. However, this extension increases the number of parameters to be estimated and requires *a priori* constraints to take the order of the categories into account: see ([1], pp 103–115) and ([67], p 260). Finally, a dispersion parameter may be added [49] and more general location-scale cumulative odds models may be considered [17]. Other generalizations include multilevel versions, varying choice of thresholds, multivariate setting, etc.

*Ceteris paribus*, model (7) presents constant log-odds for consecutive values of a single covariate and it has been denoted [49] as a *Proportional Odds Model* (POM) or, more correctly, a “proportional odds version of the cumulative logit model” ([1], p. 53). This assumption may be tested in practical applications by a formal score test [57] and/or a graphical device [44]. Other cumulative models (as *adjacent category* and *continuation ratio* models, for instance) have been introduced but their usage is more limited. Finally, *sequential models* support the selection of an ordinal score which implies a stepwise decision process ([67], pp 252–255). Notice that cumulative logit models can be also interpreted following both the IRT and URV approaches.

### 5.4 CUP models

Tutz *et al.* [68] have recently proposed to extend the class of cumulative models by introducing a component of uncertainty to improve the fitting. This mixture has been denoted as a CUP model (that is, a Combination of a discrete Uniform and a Preference random variable).

Formally, the *stochastic* and *systematic components* of a CUP model are:

$$\begin{cases} Pr(R_i = j \mid \boldsymbol{\theta}) = \pi_i \left[ F_{Y_i^*}(\tau_j - \mathbf{t}_i^{(y)} \boldsymbol{\gamma}) - F_{Y_i^*}(\tau_{j-1} - \mathbf{t}_i^{(y)} \boldsymbol{\gamma}) \right] + (1 - \pi_i) \frac{1}{m}; \\ \text{logit}(\pi_i) = \mathbf{t}_i^{(\beta)} \boldsymbol{\beta}; & \text{logit}(Pr(Y_i \leq j)) = \tau_j - \mathbf{t}_i^{(y)} \boldsymbol{\gamma}; \end{cases}$$

for  $j = 1, \dots, m$  and  $i = 1, \dots, n$ . Then, a CUP model may be considered a GEM model (1) where  $Pr(Y_i = j \mid \mathcal{T}_m, \boldsymbol{\gamma})$  coincides with (7) and  $p_j^V$  is the probability mass function of a discrete Uniform random variable.

The systematic part of the model saves the traditional definition of a predictor (as in cumulative models) but considers as well parameters  $\pi_i$  (as in the family of CUB models) to weight for the uncertainty component. Then, a plot of log-odds (which are independent of categories when the link is logistic) versus  $1 - \pi_i$  has been proposed as a graphical tool to interpret the effect of subjects’ covariates.

Authors emphasize the ability of CUP models to capture multimodality and empirical overdispersion with a better fitting. From an abstract point of view, the comparison with more parsimonious models (as those discussed in Sect. 4) should lead to prefer CUP models since they achieve more fidelity given a greater number of parameters. Empirical evidence shows that this is not a rule when one considers both fitting and parsimony criteria, as derived by *BIC* measures, for instance.

### 6 A real case study

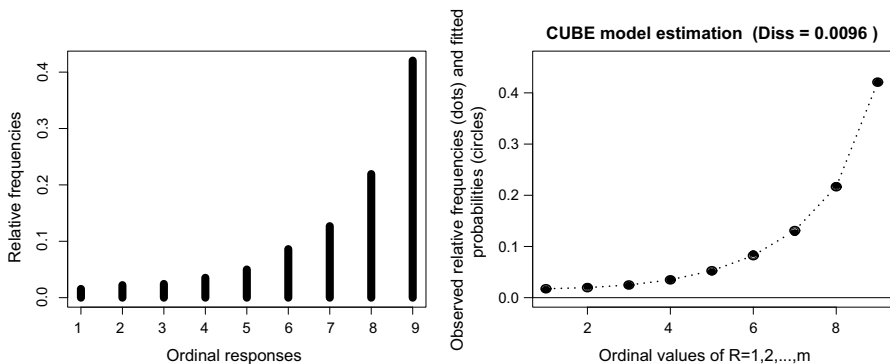
During 2007, a survey has been carried out in Naples, Italy, to score the most serious emergencies of the metropolitan area by using a Likert scale, with  $m = 9$  (where 1 = “completely unimportant” and 9 = “absolutely serious”). Hereafter, we will emphasize only the ordinal responses to the “worry for organized crime” item (*crime*) of  $n = 2381$  subjects. Sample data include the subjects’ covariates *gender* and *age*, among others.

Table 1 lists the main results based on log-likelihood and dissimilarities measures obtained by fitting parsimonious models based on discrete probability distributions (standard errors of estimated parameters are in parentheses). For the moment, we consider models without subjects’ covariates and for fitting purposes we compare the models which admits an explicit data generating process (see Sect. 4). More specifically, Fig. 3 shows the distribution of the observed responses (left panel) and compares them with the estimated probabilities obtained from a CUBE model (right panel).

We notice an almost perfect fit of CUBE model (dissimilarity index shows that less than 1% of responses should be changed to obtain a complete overlapping of the observed and estimated distributions) and this result is confirmed by a log-likelihood at the maximum which is very close to that of the saturated model (= -3994.56). Worry for this item is very strong ( $1 - \hat{\xi}$  is always greater than 0.83) and all estimated models imply an expected response in [7.1, 7.5]. In addition, uncertainty has been estimated around 0.30 by all models of CUB family. However, most of uncertainty may be attributable to overdispersion, as shown by the CUBE model where the inclusion of  $\phi$  significantly cuts uncertainty by half.

**Table 1** Models fitted to the *crime* response (without covariates)

<i>Models</i>	$1 - \hat{\pi}$	$1 - \hat{\xi}$	$\hat{\delta}$	$\hat{\phi}$	<i>Mean value</i>	<i>Log-lik</i>	<i>BIC</i>	<i>Diss</i>
IHG		0.963 (0.001)			7.108	-4111.0	8229.8	0.125
CUB	0.315 (0.015)	0.919 (0.004)			7.307	-4079.3	8174.1	0.105
CUB +sh(9)	0.319 (0.018)	0.828 (0.007)	0.289 (0.015)		7.425	-4014.4	8052.0	0.042
CUBE	0.151 (0.020)	0.863 (0.008)		0.193 (0.025)	7.465	-3995.8	8014.9	0.010



**Fig. 3** Distribution of responses to *crime* and probabilities of a fitted CUBE model

**Table 2** Models fitted to the crime response (with covariates)

<i>Models</i>	$1 - \hat{\pi}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\phi}$	<i>Log-lik</i>	<i>BIC</i>
IHG		-3.255 (0.004)	0.315 (0.004)	-0.255 (0.004)	0.097 (0.004)		-4098.7	8228.6
CUB	0.305 (0.004)	-2.362 (0.004)	0.602 (0.004)	-0.765 (0.004)	0.189 (0.004)		-4061.0	8160.9
D- BETA <sup>a</sup>		1.445 (0.036)	-0.213 (0.059)		-0.091 (0.049)		-4006.3	8056.1
POM			-0.411 (0.086)		-0.149 (0.075)		-3980.4	8046.3
CUP	0.119 (0.035)		-0.528 (0.105)		-0.192 (0.088)		-3978.8	8043.1
CUBE	0.052 (0.004)	-1.929 (0.004)	0.434 (0.004)		0.146 (0.004)	0.185 (0.023)	-3979.4	7997.6

<sup>a</sup> Dispersion link:  $\log(\hat{\sigma}_i) = 1.123 + 0.225 \text{ age}_i$   
(0.037) (0.085)

No significant covariates affect uncertainty whereas `gender` and the transformed covariate  $\text{age} = \log(\text{age}) - \overline{\log(\text{age})}$  are significant for the feeling component. Thus, we estimate different models with `gender`, `age` and `age`<sup>2</sup> and present results only for significant covariates. The (more complete) systematic component for the feeling turns out to be:

$$\text{logit}(\xi_i) = \gamma_0 + \gamma_1 \text{age}_i + \gamma_2 \text{age}_i^2 + \gamma_3 \text{gender}_i, \quad i = 1, \dots, n.$$

In addition, for the Discrete-Beta model we found a significant relationship of the dispersion parameter with `gender` and `age` expressed by:

$$\log(\sigma_i) = \lambda_0 + \lambda_1 \text{age}_i, \quad i = 1, \dots, n.$$

Table 2 lists the estimates obtained for the main models discussed so far. Opposite signs in D- BETA, POM and CUP models are expected given their different parameterizations.

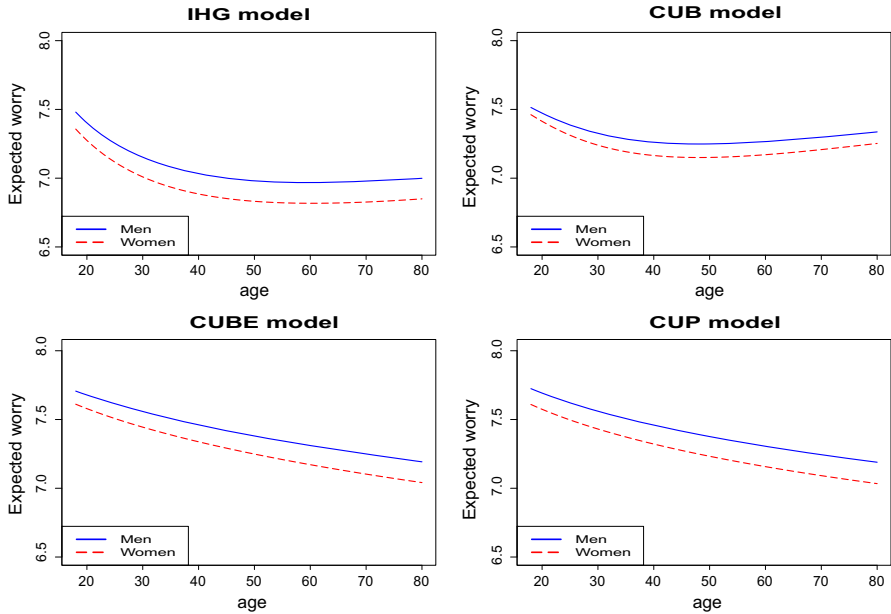
In our case study, the significance of `age` and `gender` is a common feature of all models with a similar impact whereas only IHG, CUB and partially CUBE (with a fairly significant estimate) are able to capture a parabolic effect of `age` on the responses. The unsupervised models (class III) improve the fitting and exclude `age`<sup>2</sup>; then, the parabolic effect is masked by (accounted for) the estimated cutpoints which adapt themselves to the specific shape of the categories. Observe that the maximum of log-likelihood is obtained for a CUP model (with 11 parameters) whereas a comparable fit is obtained by a CUBE model (with just 5 parameters) which turns out to achieve the minimum *BIC*.

A more effective picture displays the effect on the mean value  $\mathbb{E}(R_i)$  of a varying `age`, given the `gender`, for some of the previous models (Fig. 4). These plots show that the expected worry is regularly inferior for women; in addition, the effect of CUBE and CUP models are quite similar whereas some difference in level is implied by IHG and CUB models.

## 7 Conclusions

The specifications previously discussed may be further extended in several directions, by considering different distributions for the random variables  $Y_i$  and  $V_i$ . Given the many possibilities offered by the proposed mixture it seems difficult to be exhaustive in this regard.





**Fig. 4** Expected worry for crime as a function of gender and age according to IHG (top-left), CUB (top-right), CUBE (bottom-left) and CUP (bottom-right) models

Some of them are legitimate by a direct relationship with the data generating process and should be supported by empirical evidences.

According to the latent variable approach, possible candidates for the feeling are the negative exponential, Gamma and similar distributions: they are characterized by few parameters of location, variability and/or shape which can be related to subjects' covariates by adequate link functions. Similarly, if we jointly consider items and subjects, the approach of 3PL model in IRT [9], which implies a “guessing option”, may be explored within the class of GEM models.

Finally, the paradigm so far discussed emphasizes the role of a discrete support in modelling ordinal data and the importance to consider discrete states in the selection process, an approach fostered by [7] for longitudinal data and by [3] for personality measures.

A unifying perspective for modelling ordinal data is a useful framework to compare models, to discover unexpected similarities and to introduce new distributions. Most of all, the opportunity to see the same data from different points of view improves interpretation and prediction of ordinal responses by means of statistical models.

## References

1. Agresti, A.: Analysis of Ordinal Categorical Data, 2<sup>nd</sup> edition. Wiley, Hoboken (2010)
2. Agresti, A.: Foundations of Linear and Generalized Linear Models. Wiley, Hoboken (2015)
3. Allik, J.: A mixed-binomial model for Likert-type personality measure. *Front. Psychol.* **5**, 1–13 (2014)
4. Anderson, J.A.: Regression and ordered categorical variables. *J. R. Stat. Soc. Ser. B* **46**, 1–30 (1984)
5. Banfield, J.D., Raftery, A.E.: Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803–821 (1993)

6. Bartholomew, D.J., Knott, M.: *Latent Variable Models and Factor Analysis*. Hodder Arnold, London (1999)
7. Bartolucci, F., Farcomeni, A., Pennoni, F.: *Latent Markov Models for Longitudinal Data*. CRC Press, Boca Raton (2013)
8. Biernacki, C., Jacques, J.: Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm. Research Report, HAL-01052447 (2014)
9. Birnbaum, A.: Some latent trait models and their use in inferring on examinee's ability. In: Lord, F.M., Novick, M.R. (eds.) *Statistical Theories of Mental Test Scores*, pp. 397–472. Addison-Wesley, Reading (1968)
10. Bock, R.D., Moustaki, I.: Item response theory in a general framework. In: Rao, C.R., Sinharay, S. (eds.) *Psychometrics, Handbook of Statistics*, vol. 26, pp. 469–513. Elsevier, Amsterdam (2007)
11. Box, G.E.P.: Signal-to-noise ratios, performance criteria and transformation. *Technometrics* **30**, 1–31 (1988)
12. Breen, R., Luijkx, R.: Mixture models for ordinal data. *Sociological Methods and Research* **39**, 3–24 (2010)
13. Colombi, R., Giordano, S.: A mixture model for multidimensional ordinal data. In: Friedl, H., Wagner, H. (eds.) *Proceedings of 30th International Workshop on Statistical Modelling (IWSM)*, pp. 129–134. Johannes Kepler University Linz, Austria (2015)
14. Corduas, M.: Assessing similarity of rating distributions by Kullback-Leibler divergence. In: Fichet, B., et al. (eds.) *Classification and Multivariate Analysis for Complex Data Structures*, pp. 221–228. Springer, Berlin (2011)
15. Corduas, M.: Analyzing bivariate ordinal data with CUB margins. *Stat. Model. Int. J.* **15**, 411–432 (2015)
16. Corduas, M., Iannario, M., Piccolo, D.: A class of statistical models for evaluating services and performances. In: Bini, M., et al. (eds.) *Statistical methods for the evaluation of educational services and quality of products, Contribution to Statistics*, pp. 99–117. Springer, Berlin (2009)
17. Cox, C.: Location-scale cumulative odds models for ordinal data: a generalized non-linear model approach. *Stat. Med.* **14**, 1191–1203 (1995)
18. Cribari-Neto, F., Zeileis, A.: Beta regression in R. *J. Stat. Softw.* **34**, 1–24 (2010)
19. Cugnata, F., Salini, S.: Comparison of alternative imputation methods for ordinal data. *Commun. Stat. Simul. Comput.* (2014). doi:10.1080/03610918.2014.963611
20. D'Elia, A.: A proposal for ranks statistical modelling. In: Friedl, H. et al. (eds.) *Statistical Modelling—Proceedings of the 14th International Workshop on Statistical Modelling*, pp. 468–471. Graz, Austria (1999)
21. D'Elia, A.: Il meccanismo dei confronti appaiati nella modellistica per graduatorie: sviluppi statistici ed aspetti critici. *Quaderni di Statistica* **2**, 173–203 (2000)
22. D'Elia, A.: A comparison between two asymptotic tests for analysing preferences. *Quad. di Stat.* **3**, 127–143 (2001)
23. D'Elia, A.: Modelling ranks using the inverse hypergeometric distribution. *Stat. Model. Int. J.* **3**, 65–78 (2003)
24. D'Elia, A., Piccolo, D.: A mixture model for preference data analysis. *Comput. Stat. Data Anal.* **49**, 917–934 (2005)
25. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39**, 1–38 (1977)
26. Everitt, B.S., Hand, D.J.: *Finite Mixture Distributions*. Chapman & Hall, London (1981)
27. Gottard, A., Iannario, M., Piccolo, D.: Varying uncertainty in CUB models. *Adv. Data Anal. Classif.* **10**, 225–244 (2016)
28. Greene, W.H., Hensher, D.A.: A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp. Res. Part B* **37**, 681–689 (2003)
29. Grün, B., Leisch, F.: Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *J. Classif.* **25**, 225–247 (2008)
30. Guenther, W.C.: The inverse hypergeometric—a useful model. *Stat. Neerlandica* **29**, 129–144 (1975)
31. Heinen, T.: *Latent class and discrete latent traits models: similarities and differences*. Sage, Thousand Oaks (1996)
32. Iannario, M.: Modelling shelter choices in a class of mixture models for ordinal responses. *Stat. Methods Appl.* **21**, 1–22 (2012a)
33. Iannario, M.: CUBE models for interpreting ordered categorical data with overdispersion. *Quad. di Stat.* **14**, 137–140 (2012b)
34. Iannario, M.: Preliminary estimators for a mixture model of ordinal data. *Adv. Data Anal. Classif.* **6**, 163–184 (2012c)

35. Iannario, M.: Hierarchical CUB models for ordinal variables. *Commun. Stat. Theory Methods* **41**, 3110–3125 (2012d)
36. Iannario, M.: Modelling uncertainty and overdispersion in ordinal data. *Commun. Stat. Theory Methods* **43**, 771–786 (2014)
37. Iannario, M.: Detecting latent components in ordinal data with overdispersion by means of a mixture distribution. *Qual Quant.* **49**, 977–987 (2015)
38. Iannario, M., Piccolo, D.: CUB models: statistical methods and empirical evidence. In: Kenett, R.S., Salini, S. (eds.) *Modern Analysis of Customer Surveys: With Applications Using R*, pp. 231–258. Wiley, Chichester (2012a)
39. Iannario, M., Piccolo, D.: A framework for modelling ordinal data in rating surveys. *Proceedings of Joint Statistical Meetings, Section on statistics in marketing*, pp. 3308–3322. San Diego, California (2012b)
40. Iannario, M., Piccolo, D.: Inference for CUB models: a program in R. *Stat. Appl.* **XII**, 177–204 (2014)
41. Iannario, M., Piccolo, D.: A generalized framework for modelling ordinal data. *Stat. Methods Appl.* **25**, 163–189 (2016)
42. Iannario, M., Piccolo, D., Simone, R.: CUB : a class of mixture models for ordinal data. R package version 0.1. <http://CRAN.R-project.org/package=CUB> (2016)
43. Krosnick, J.A.: Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl. Cogn. Psychol.* **5**, 213–236 (1991)
44. Kim, J.-A.: Assessing practical significance of the proportional odds assumption. *Stat. Probab. Lett.* **65**, 233–239 (2003)
45. Krosnick, J.A.: Survey research. *Ann. Rev. Psychol.* **50**, 537–567 (1999)
46. Lazarsfeld, P.F., Henry, N.W.: *Latent Structure Analysis*. Houghton Mifflin, New York (1968)
47. Manisera, M., Zuccolotto, P.: Modeling “Don’t know” responses in rating scales. *Pattern Recognit. Lett.* **45**, 226–234 (2014a)
48. Manisera, M., Zuccolotto, P.: Modeling rating data with nonlinear CUB models. *Comput. Stat. Data Anal.* **78**, 100–118 (2014b)
49. McCullagh, P.: Regression models for ordinal data (with discussion). *J. R. Stat. Soc. Ser. B* **42**, 109–142 (1980)
50. McLachlan, G., Krishnan, T.: *The EM Algorithm and Extensions*, 2nd edn. Wiley, New York (2000)
51. Miller, G.K., Fridell, S.L.: A forgotten discrete distribution? Reviving the negative hypergeometric model. *Am. Stat.* **61**, 347–350 (2007)
52. Morrison, D.G.: Purchase intentions and purchase behavior. *J. Mark.* **43**, 65–74 (1979)
53. Moustaki, I.: A latent variable model for ordinal data. *Appl. Psychol. Meas.* **24**, 211–223 (2000)
54. Moustaki, I.: A generalized class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *Br. J. Math. Stat. Psychol.* **53**, 337–357 (2003)
55. Moustaki, I., Joreskog, K., Mavridis, D.: Factor models for ordinal variables with covariate effects on the manifest and latent variables: a comparison of LISREL and IRT approaches. *Struct. Equ. Model. J.* **11**, 487–513 (2004)
56. Moustaki, I., Knott, M.: Generalized latent trait models. *Psychometrika* **65**, 391–411 (2001)
57. Peterson, B., Harrell Jr., F.E.: Partial proportional odds models for ordinal response variables. *Appl. Stat.* **39**, 205–217 (1990)
58. Piccolo, D.: On the moments of a mixture of uniform and shifted binomial random variables. *Quad. di Stat.* **5**, 85–104 (2003)
59. Piccolo, D.: Observed information matrix for MUB models. *Quad. di Stat.* **8**, 33–78 (2006)
60. Piccolo, D.: Inferential issues in CUBE models with covariates. *Commun. Stat. Theory Methods* **44**, 5023–5036 (2015)
61. Piccolo, D., D’Elia, A.: A new approach for modelling consumers’ preferences. *Food Qual. Preference* **19**, 247–259 (2008)
62. Punzo, A.: CUN models. University of Catania, Personal communication, Preliminary note (2012)
63. Samejima, F.: Estimation of latent trait ability using a response pattern of graded scores. *Psychom. Monograph Suppl.* **17**, 1–139 (1969)
64. Simon, H.A.: *Models of Man*. Wiley, New York (1957)
65. Tamhane, A.C., Ankenman, B.E., Yang, Y.: The Beta distribution as a latent response model for ordinal data (I): estimation of location and dispersion parameters. *J. Stat. Simul. Comput.* **72**, 473–494 (2002)
66. Tourangeau, R., Rips, L.J., Rasinski, K.: *The Psychology of Survey Response*. Cambridge University Press, Cambridge (2000)
67. Tutz, G.: *Regression for Categorical Data*. Cambridge University Press, Cambridge (2012)
68. Tutz, G., Schneider, M., Iannario, M., Piccolo, D.: Mixture models for ordinal responses to account for uncertainty of choice. *Adv. Data Anal. Classif.* **10** (2016). doi:[10.1007/s11634-016-0247-9](https://doi.org/10.1007/s11634-016-0247-9)

69. Ursino, M.: Ordinal data: a new model with applications. Ph.D. Thesis, XXVI cycle, Polytechnic University of Turin (2014)
70. van der Linden, W.J., Hambleton, R.K. (eds.): Handbook of Modern Item Response Theory. Springer, New York (1997)
71. Vermunt, J.K., Magidson, J.: Technical guide for latent Gold 5.0: basic, advanced, and syntax. Statistical Innovations Inc., Belmont (2013)
72. Wedel, M., DeSarbo, W.S.: A mixture likelihood approach for generalized linear models. *J. Classif.* **12**, 21–55 (1995)
73. Wilks, S.S.: *Mathematical Statistics*. London, Wiley (1963)