A class of models for ordinal data analysis: statistical issues and new developments*

Maria Iannario
Department of Political Sciences
University of Naples Federico II (IT)
e-mail: maria.iannario@unina.it

Abstract

`CUB` models are a general statistical framework for the analysis of ordinal data; for an up-to-date review see Iannario and Piccolo (2013). These models have found great interest in several contexts in which the rating has been considered in relation to the psychological mechanism surroundings the respondents' choice. More specifically, respondents' score is modelled as a weighted mixture of their propensity to adhere to a meditated choice (formally described by a shifted Binomial random variable) and/or a totally uninformative choice (described by a discrete Uniform distribution). The interest for this class of models increases with the introduction of subjects', objects' and contexts' covariates (it is pursued by means of link logit). This extension improves the fitting of models and allows to create the profiles of respondents. Specifically, the link is not related to expectation but to parameters characterizing uncertainty and feeling, respectively. This option has been motivated by the relationship between expectation and parameters defined for this random variable by:

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

thus, infinitely many pairs of parameters generate the same expectation. As a consequence, if ordinal data are generated by a `CUB` model mechanism, the mean value should not to be considered as a correct synthesis; indeed, it is immediate to show that completely different shapes give the same expected value. On the contrary, the interpretation and the comparison based on the points in the parameter space correctly cover the information contained in the `CUB` distribution.

Several other extensions and generalizations of basic `CUB` models have been proposed; some of them are currently in progress. Among these, we just performed the hierarchical structure with mixed effects for the *feeling* component (Iannario, 2012b), the generalization with *shelter* effect (Iannario, 2012c) and the analysis of complex sample design or clustering methods.

This mixture model has been also generalized by introducing a varying uncertainty component and a Beta-Binomial random variable by changing the *probability distributions* for *uncertainty* and *feeling* components, respectively. With respect to the standard model, the first proposal consists in the introduction of uncertainty distributions with different shapes (Trimmed

*Presented at the first meeting of the FIRB ("Futuro in ricerca" 2012) project "Mixture and latent variable models for causal-inference and analysis of socio-economic data", Perugia (IT), March 15-16, 2013

Uniform, Left/right Uniform bounded distribution, Symmetric discrete triangular distribution, among others) which could capture the real indecision of respondents with larger effectiveness. Since the alternative specifications are several and non nested, it is possible to implement a Vuong test for choosing among them. In this regard, some simulation experiment and real case studies seem to confirm the usefulness of the approach (Gottard et al., 2013). In addition, the presence of a possible overdispersion motivates the introduction of a Beta Binomial component instead of the shifted Binomial. In this case we introduced a more general class which encompasses the previous one. It is characterized by 3 parameters which offer a high degree of flexibility (Iannario, 2012a).

By considering a *multivariate approach* for the joint modelling of items, after the implementation of a multi-object (context), we will study a multivariate `CUB` models which may be tackled on the basis of latent variables approach. In this case the parameters may be consistently estimated and validated by maximum likelihood inference.

From a methodological point of view, such approaches require the study of numerical aspects of the algorithms for achieving the convergence in reasonable time, some extensive simulation to check the validity of assumed distributional results when working on limited data sets, and also a framework to compare adequately this different class of models with the classical ones.

From the empirical point of view, we are also checking if and when the family of models we are discussing about is strictly dependent on the choice of the number of categories (m) by analysing a possible m-equivalence. This topic is relevant when people plan questionnaires with too few or many ordered alternatives.

Finally, it is necessary to underline as the construction of `CUB` models -based on a discrete mixture- should be considered complementary to the other approaches. Indeed, one of the objectives of the project is to verify, on the same data set, analogies and differences among statistical models commonly performed for the analysis of ordinal data from both statistical and interpretative point of view.

References

Gottard, A., Iannario, M., and Piccolo, D. (2013). Varying uncertainty in CUB models. Submitted.

Iannario, M. (2012a). CUBE models for interpreting ordered categorical data with overdispersion. *Quaderni di Statistica*, 14:137–140.

Iannario, M. (2012b). Hierarchical `CUB` models for ordinal variables. *Communications in Statistics. Theory and Methods*, 41:3110–3125.

Iannario, M. (2012c). Modelling *shelter* choices in a class of mixture models for ordinal responses. *Statistical Methods and Applications*, 21:1–22.

Iannario, M. and Piccolo, D. (2013). A generalized framework for modelling ordinal data. Preliminary report.